Nonparametric estimation in a regression model with additive and multiplicative noise

Christophe Chesneau, Salima El Kolei, Junke Kou, Fabien Navarro, Salima Kolei

HAL Id: hal-02159579
https://hal.archives-ouvertes.fr/hal-02159579v2

Submitted on 20 Jun 2020

Nonparametric estimation in a regression model with additive and multiplicative noise

Christophe Chesneau[*], Salima El Kolei[†], Junke Kou[‡], Fabien Navarro[§]

June 20, 2020

Abstract

In this paper, we consider an unknown functional estimation problem in a general nonparametric regression model with the feature of having both multiplicative and additive noise.We propose two new wavelet estimators in this general context. We prove that they achieve fast convergence rates under the mean integrated square error over Besov spaces. The obtained rates have the particularity of being established under weak conditions on the model. A numerical study in a context comparable to stochastic frontier estimation (with the difference that the boundary is not necessarily a production function) supports the theory.

Keywords: Nonparametric regression, multiplicative regression models, nonparametric frontier, rates of convergence, wavelets.

[*]Christophe Chesneau

Université de Caen - LMNO, France, E-mail: christophe.chesneau@unicaen.fr

[†]Salima El Kolei

CREST, ENSAI, France, E-mail: salima.el-kolei@ensai.fr

[‡]Junke Kou

Guilin University of Electronic Technology, China, E-mail: kjkou@guet.edu.cn

[§]Fabien Navarro

CREST, ENSAI, France, E-mail: fabien.navarro@ensai.fr

1   Introduction

We consider a nonparametric regression model with both multiplicative and additive noise. It is defined by n random variables $Y_1, \ldots, Y_n$, where

$$Y_i = f(\boldsymbol{X}_i)U_i + V_i, \quad i \in \{1, \ldots, n\}, \tag{1}$$

f is an unknown regression function defined on a subset Δ of $\mathbb{R}^d$, with $d \geq 1$, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are n identically distributed random vectors with support on Δ, $U_1, \ldots, U_n$ are n identically distributed random variables and $V_1, \ldots, V_n$ are n identically distributed random variables. Moreover, it is supposed that $\boldsymbol{X}_i$ and U_i are independent, and U_i and V_i are independent for any $i \in \{1, \ldots, n\}$. We are interested in the estimation of the unknown function $r := f^2$ from $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$; the random vectors $(U_1, V_1), \ldots, (U_n, V_n)$ form the multiplicative-additive noise. We consider the general formulation of model given by (1) since besides the theoretical interest it embodies several potential applications. For example, for $U_i = 1$, (1) becomes the standard nonparametric regression model with additive noise. It has been studied in many papers via various nonparametric methods, including kernel, splines, projection and wavelets methods. See, for instance, the books of Härdle et al. (2012), Tsybakov (2009) and Comte (2015), and the references therein. For $V_i = 0$, (1) becomes the standard nonparametric regression model with multiplicative noise. Recent studies can be found in Chichignoud (2012); Comte (2015) and the references therein. For $V_i \neq 0$ with the same variance across i a first study, based on a linear wavelet estimator, was proposed by Chesneau et al. (2019). In the case where V_i is a function of $\boldsymbol{X}_i$, (1) becomes the popular nonparametric regression model with multiplicative heteroscedastic noises:

$$Y_i = g(X_i) + f(X_i)U_i$$

In particular, this model is widely used in financial applications, where the aim is to estimate the variance function $r := f^2$ from the returns of an asset, for instance, to establish confidence intervals/bands for the mean function g. Variance estimation is a fundamental statistical problem with wide applications (see Muller et al. (1987); Hall and Marron (1990); Härdle and Tsybakov (1997); Wang et al. (2008); Brown et al. (2007); Cai et al. (2008) for fixed design and recently Kulik et al. (2011); Verzelen et al. (2018); Shen et al. (2019) for random design).

This multiplicative regression model is also popular in various application areas. For example, in econometrics, within deterministic ($V_i = 0$) and stochastic ($V_i \neq 0$) non-parametric frontier models. These models can be interpreted as a special case of the model (1), where the random variable U_i represents the technical inefficiency of the company and V_i represents noise that disrupts its performance, the nature of which comes from unanticipated events such that machine failure, strikes, staff strikes, etc. Under monotonicity and concavity assumptions, the regression function r can be viewed in this case as a function of the production set of a firm and its estimation is therefore of paramount importance in production econometrics. Specific estimation methods have been developed, see for instance Farrell (1957); De Prins et al. (1984); Gijbels et al. (1999); Daouia and Simar (2005) for deterministic frontiers models and Fan et al. (1996); Kumbhakar et al. (2007); Simar and Zelenyuk (2011) for stochastic frontier models. For general regression function and general nonparametric setting, we refer to Girard and Jacob (2008); Girard et al. (2013) and Jirak et al. (2014) for the definitions and properties of robust estimators.

Applications also exist in signal and image processing (*e.g.*, for Global Positioning System signal detection Huang et al. (2013) as well as in speckle noise reduction encounter in particular in synthetic-aperture radar images Kuan et al. (1985) or in medical ultrasound images Rabbani et al. (2008); Mateo and Fernández-Caballero (2009)), where noise sources can be both additive and multiplicative. In this context, one can also cite Korostelev and Tsybakov (2012) where the author deals with the estimation of the function's support.

The aim of this paper is to develop wavelet methods for the general model (1), with a special focus on mild assumptions on the distributions of U_1 and V_1 (moments of order 4 will be required, including Gaussian noise). Wavelet methods are of interest in nonparametric statistics thanks to their ability to estimate efficiently a wide variety of unknown functions, including those with spikes and bumps. We refer to Abramovich et al. (2000) and Härdle et al. (2012), and the references therein. To the best of our knowledge, their development for (1) taking in full generality is new in the literature. First of all, we construct a linear wavelet estimator using projections of wavelet coefficients estimators. We evaluate its rate of convergence under the mean integrated square error (MISE) under mild assumptions on the smoothness of r; it is assumed that r be-

longs to Besov spaces. The linear wavelet estimator has the advantage to be simple, but the knowledge of the smoothness of r is necessary to calibrate a tuning parameter which plays a crucial role in the determination of fast rates of convergence. For this reason, an alternative is given by a nonlinear wavelet estimator. Using a thresholding rule of wavelet coefficients estimators, we develop a nonlinear wavelet estimator. To reach the goal of mildness assumptions on the model, we use a truncation rule in the definition of the wavelet coefficients estimators. This technique was introduced by Delyon and Juditsky (1996) in the nonparametric regression estimation setting, and recently improved in Chesneau (2013) (in a multidimensional regression function under mixing dependence framework) and Chaubey et al. (2015) (for a density estimation under a multiplicative censoring problem). The construction of the hard wavelet estimator does not depend on the smoothness of r and we prove that, from a global point of view, it achieves a better rate of convergence under the MISE. In practice, the empirical performance of the estimators developed in this paper depends on the choice of several parameters, the truncation level of the linear estimator as well as the threshold parameter of the non-linear estimator. We propose here a method of automatic selection of these two parameters based on the 2-fold cross-validation method (2FCV) introduced by Nason (1996). A numerical study, in a context similar to stochastic frontier estimation, is being carried out to demonstrate the applicability of this approach.

The rest of the paper is organized as follows. Preliminaries on wavelets are described in Section 2. Section 3 specifies some assumptions on the model, presents our wavelet estimators and the main results on their performances. Numerical experiments are presented in Section 4. Section 5 is devoted to the proofs of the main result.

2   Preliminaries on wavelets

2.1   Wavelet bases on $[0, 1]^d$

We begin with a classical notation in wavelet analysis. A multiresolution analysis (MRA) is a sequence of closed subspaces $\{V_j\}_{j \in \mathbb{Z}}$ of the square integrable function space $L^2(\mathbb{R})$ satisfying the following properties:

   (i) $V_j \subseteq V_{j+1}$, $j \in \mathbb{Z}$. $\mathbb{Z}$ denotes the integer set and $\mathbb{N} := \{n \in \mathbb{Z}, n \geq 0\}$;

(ii) $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$ (the space $\bigcup_{j \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$);

(iii) $f(2\cdot) \in V_{j+1}$ if and only if $f(\cdot) \in V_j$ for each $j \in \mathbb{Z}$;

(iv) There exists $\phi \in L^2(\mathbb{R})$ (scaling function) such that $\{\phi(\cdot - k), \ k \in \mathbb{Z}\}$ forms an orthonormal basis of $V_0 = \overline{\mathrm{span}}\{\phi(\cdot - k)\}$.

See Meyer (1992) for further details. For the purpose of this paper, we use the compactly supported scaling function ϕ of the Daubechies family, and the associated compactly supported wavelet function ψ (see Daubechies (1992)). Then we consider the wavelet tensor product bases on $[0,1]^d$ as described in Cohen et al. (1993). The main lines and notations are described below. We set $\Phi(\boldsymbol{x}) = \prod_{v=1}^d \phi(x_v)$ and wavelet functions: $\Psi_u(\boldsymbol{x}) = \psi(x_u) \prod_{\substack{v=1 \\ v \neq u}}^d \phi(x_v)$ when $u \in \{1, \ldots, d\}$, and $\Psi_u(\boldsymbol{x}) = \prod_{v \in A_u} \psi(x_v) \prod_{v \notin A_u} \phi(x_v)$ when $u \in \{d+1, \ldots, 2^d - 1\}$, where $(A_u)_{u \in \{d+1, \ldots, 2^d - 1\}}$ forms the set of all the non-void subsets of $\{1, \ldots, d\}$ of cardinal superior or equal to 2. For any integer j and any $\boldsymbol{k} = (k_1, \ldots, k_d)$, we set $\Phi_{j,\boldsymbol{k}}(\boldsymbol{x}) = 2^{jd/2} \Phi(2^j x_1 - k_1, \ldots, 2^j x_d - k_d)$, for any $u \in \{1, \ldots, 2^d - 1\}$, $\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}) = 2^{jd/2} \Psi_u(2^j x_1 - k_1, \ldots, 2^j x_d - k_d)$. Now, let us set $\Lambda_j = \{0, \ldots, 2^j - 1\}^d$. Then, with an appropriate treatment on the elements which step on the boundaries 0 and 1, there exists an integer τ such that the system $\mathcal{S} = \{\Phi_{\tau,\boldsymbol{k}}, \boldsymbol{k} \in \Lambda_\tau; \ (\Psi_{j,\boldsymbol{k},u})_{u \in \{1, \ldots, 2^d - 1\}}, \ j \geq \tau, \ \boldsymbol{k} \in \Lambda_j\}$ forms an orthonormal basis of $L^2([0,1]^d)$. For any integer $j_* \geq \tau$, a function $h \in L^2([0,1]^d)$ can be expressed via $\mathcal{S}$ by the following wavelet series:

$$h(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \Lambda_{j_*}} \alpha_{j_*,\boldsymbol{k}} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) + \sum_{j=j_*}^{\infty} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \beta_{j,\boldsymbol{k},u} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}), \quad \boldsymbol{x} \in [0,1]^d, \qquad (2)$$

where $\alpha_{j,\boldsymbol{k}} = \langle h, \Phi_{j,\boldsymbol{k}} \rangle_{[0,1]^d}$ and $\beta_{j,\boldsymbol{k},u} = \langle h, \Psi_{j,\boldsymbol{k},u} \rangle_{[0,1]^d}$.

Also, let us mention that, by construction, $\int_{[0,1]^d} \Phi_{j,\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} = 2^{-jd/2}$ and $\int_{[0,1]^d} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}) d\boldsymbol{x} = 0$.

Let P_j be the orthogonal projection operator from $L^2([0,1]^d)$ onto the space V_j with the orthonormal basis $\{\Phi_{j,\boldsymbol{k}}(\cdot) = 2^{jd/2} \Phi(2^j \cdot - \boldsymbol{k}), \boldsymbol{k} \in \Lambda_j\}$. Then, for any $h \in L^2([0,1]^d)$,

$$P_j h(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \Lambda_j} \alpha_{j,\boldsymbol{k}} \Phi_{j,\boldsymbol{k}}(\boldsymbol{x}), \quad \boldsymbol{x} \in [0,1]^d.$$

2.2 Besov spaces

Besov spaces are important in theory and applications. They have the features to a wide variety of function spaces as Hölder and L^2 Sobolev spaces. Definitions of those spaces are given below. Suppose that ϕ is m regular (*i.e.*, $\phi \in C^m$ and $|D^\alpha \phi(x)| \le c(1+|x|^2)^{-l}$ for each $l \in \mathbb{Z}$, with $\alpha = 0, 1, \dots, m$) and consider the wavelet framework defined in Subsection 2.1. Let $h \in L^p([0,1]^d)$, $p, q \in [1, \infty]$ and $0 < s < m$. Then the following assertions are equivalent:

(1) $h \in B^s_{p,q}([0,1]^d)$;   (2) $\{2^{js}\|P_{j+1}h - P_j h\|_p\} \in l_q$;   (3) $\{2^{j(s-\frac{d}{p}+\frac{d}{2})}\|\beta_{j,\cdot,\cdot}\|_p\} \in l_q$.

The Besov norm of h can be defined by

$$\|f\|_{B^s_{p,q}} := \|(\alpha_{\tau,\cdot})\|_p + \|(2^{j(s-\frac{d}{p}+\frac{d}{2})}\|\beta_{j,\cdot,\cdot}\|_p)_{j \ge \tau}\|_q, \text{ where } \|\beta_{j,\cdot,\cdot}\|_p^p = \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} |\beta_{j,\boldsymbol{k},u}|^p.$$

Further details on Besov spaces are given in Meyer (1992), Triebel (1994) and Härdle et al. (2012).

3 Assumptions, estimators and main result

We consider the model (1) with $\Delta = [0,1]^d$ for the sake of simplicity. Additional technical assumptions are formulated below.

A.1 We suppose that $f : [0,1]^d \to \mathbb{R}$ is bounded from above.

A.2 We suppose that $\boldsymbol{X}_1 \sim \mathcal{U}([0,1]^d)$.

A.3 We suppose that U_1 is reduced (mainly for the sake of simplicity in exposition) and has a moment of order 4.

A.4 We suppose that V_1 has a moment of order 4.

The two following assumptions involving V_i and $\boldsymbol{X}_i$ are complementary and will be considered separately in the study:

A.5 We suppose that $\boldsymbol{X}_i$ and V_i are independent for any $i \in \{1, \dots, n\}$, and U_1 is centered or V_1 is centered.

A.6 We suppose that $V_i = g(\boldsymbol{X}_i)$ where $g : [0,1]^d \to \mathbb{R}$ is known and bounded from above, and U_1 is centered.

These assumptions will be discussed later; some of them can be relaxed. In our main results, we will consider the two following sets of assumptions:

$$\mathbf{H.1} = \{\mathbf{A.1}, \mathbf{A.2}, \mathbf{A.3}, \mathbf{A.4}, \mathbf{A.5}\}, \quad \mathbf{H.2} = \{\mathbf{A.1}, \mathbf{A.2}, \mathbf{A.3}, \mathbf{A.4}, \mathbf{A.6}\}.$$

As usual in wavelet methods, the first step towards the estimation of r is to consider its wavelet series given by (2). Then we aim to estimate the unknown wavelet coefficients $\alpha_{j,\mathbf{k}} = \langle r, \Phi_{j,\mathbf{k}} \rangle_{[0,1]^d}$ and $\beta_{j,\mathbf{k},u} = \langle r, \Psi_{j,\mathbf{k},u} \rangle_{[0,1]^d}$ by efficient estimators. In this study, we propose to estimate $\alpha_{j,\mathbf{k}}$ by

$$\hat{\alpha}_{j,\mathbf{k}} := \frac{1}{n} \sum_{i=1}^{n} Y_i^2 \Phi_{j,\mathbf{k}}(\mathbf{X}_i) - v_{j,\mathbf{k}}, \tag{3}$$

where

$$v_{j,\mathbf{k}} := \begin{cases} \mathbf{E}\left[V_1^2\right] 2^{-jd/2} & \text{under } \mathbf{A.5}, \\ \displaystyle\int_{[0,1]^d} g^2(\mathbf{x}) \Phi_{j,\mathbf{k}}(\mathbf{x}) d\mathbf{x} & \text{under } \mathbf{A.6}. \end{cases}$$

This is an unbiased estimator of $\alpha_{j,\mathbf{k}}$ and it converges to $\alpha_{j,\mathbf{k}}$ in L^2 (see Lemmas 5.1 and 5.2 in Section 5). On the other side, we propose to estimate $\beta_{j,\mathbf{k},u}$ by

$$\hat{\beta}_{j,\mathbf{k},u} := \frac{1}{n} \sum_{i=1}^{n} \left(Y_i^2 \Psi_{j,\mathbf{k},u}(\mathbf{X}_i) - w_{j,\mathbf{k},u}\right) \mathbb{1}_{\left\{\left|Y_i^2 \Psi_{j,\mathbf{k},u}(\mathbf{X}_i) - w_{j,\mathbf{k},u}\right| \leq \rho_n\right\}}, \tag{4}$$

where $\mathbb{1}_A$ denotes the indicator function over an event A, $\rho_n := \sqrt{n/\ln n}$ and

$$w_{j,\mathbf{k},u} := \begin{cases} 0 & \text{under } \mathbf{A.5}, \\ \displaystyle\int_{[0,1]^d} g^2(\mathbf{x}) \Psi_{j,\mathbf{k},u}(\mathbf{x}) d\mathbf{x} & \text{under } \mathbf{A.6}. \end{cases}$$

Due to the thresholding in its definition, this estimator is not unbiased of $\beta_{j,\mathbf{k},u}$ but it converges to $\beta_{j,\mathbf{k},u}$ in L^2 (see Lemmas 5.1 and 5.2 in Section 5). The role of the thresholding is to relax assumptions on U_1 and V_1; note that only moments of order 4 is required in $\mathbf{A.3}$ and $\mathbf{A.4}$ including uniform or Gaussian distribution. This selection rule has been introduced in a wavelet setting in Delyon and Juditsky (1996). It has been recently improved in Chesneau (2013) (in a multidimensional regression function under mixing dependence framework) and Chaubey et al. (2015) (in a density estimation under multiplicative censoring setting). In this study, we adapt it to the general nonparametric regression model (1).

The next step in the construction of our wavelet estimators for r is to expand the most informative of the wavelet coefficients estimators using the initial wavelet basis. We then define the linear wavelet estimator by

$$\hat{r}_n^{\text{lin}}(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \Lambda_{j_*}} \hat{\alpha}_{j_*,\boldsymbol{k}} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}), \quad \boldsymbol{x} \in [0,1]^d.$$

We thus have projected the $\hat{\alpha}_{j,\boldsymbol{k}}$'s on the father wavelet basis at a certain level j_*. Despite the simplicity of its construction, this estimator has a serious drawback: its performance highly depends on the choice of the level j_*. A suitable choice of j_*, but depending on the smoothness of r, will be specified in our main result. To address this problem, an alternative is proposed by using a hard-thresholding rule that performs a term-by-term selection of the wavelet coefficient estimators $\hat{\beta}_{j,\boldsymbol{k},u}$ and to project them on the original wavelet basis. We define the nonlinear wavelet estimator by

$$\hat{r}_n^{\text{non}}(\boldsymbol{x}) := \sum_{\boldsymbol{k} \in \Lambda_{j_*}} \hat{\alpha}_{j_*,\boldsymbol{k}} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) + \sum_{j=j_*}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \hat{\beta}_{j,\boldsymbol{k},u} \mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}| \geq \kappa t_n\}} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}), \ \boldsymbol{x} \in [0,1]^d,$$

where $t_n := \sqrt{\ln n / n} = \rho_n^{-1}$. The positive integer j_1 is specified in our main result, while the constant κ will be chosen in its proof (see the proof of Lemma 5.3). The idea of keeping the estimators $\hat{\beta}_{j,\boldsymbol{k},u}$ with magnitude greater to t_n is not new; it is a well-known wavelet techniques with strong mathematical and practical results for numerous nonparametric problems; t_n is so-called "universal threshold". We refer to Donoho et al. (1995), Delyon and Juditsky (1996) and Härdle et al. (2012). In this study, we describe how to calibrate such estimator when we deal with the general model (1).

In the sequel, we adopt the following notations: $x_+ := \max\{x, 0\}$. $A \lesssim B$ denotes $A \leq cB$ for some constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $B \lesssim A$.

Theorem 3.1 below determines the rates of convergence attained by $\hat{r}_n^{\text{lin}}$ and $\hat{r}_n^{\text{non}}$ over the MISE.

Theorem 3.1. *Consider the problem defined by* (1) *under the assumptions* **H.1** *or* **H.2**, *let* $r \in B_{p,q}^s([0,1]^d)$ *with* $p, q \in [1, \infty)$, $s > d/p$. *Then*

- *the linear wavelet estimator* $\hat{r}_n^{\text{lin}}$ *with* $2^{j_*} \sim n^{\frac{1}{2s'+d}}$ *and* $s' = s - d(1/p - 1/2)_+$ *satisfies*

$$\mathbf{E}\left[\int_{[0,1]^d} \left(\hat{r}_n^{\text{lin}}(\boldsymbol{x}) - r(\boldsymbol{x})\right)^2 d\boldsymbol{x}\right] \lesssim n^{-\frac{2s'}{2s'+d}}; \tag{5a}$$

- *the nonlinear estimator with $2^{j*} \sim n^{\frac{1}{2m+d}}$ $(m > s)$, $2^{j_1} \sim (n/\ln n)^{\frac{1}{d}}$ satisfies*

$$\mathbf{E}\left[\int_{[0,1]^d} (\hat{r}_n^{\mathrm{non}}(\boldsymbol{x}) - r(\boldsymbol{x}))^2 \, d\boldsymbol{x}\right] \lesssim (\ln n) n^{-\frac{2s}{2s+d}}. \tag{5b}$$

The obtained rates of convergence are those obtained in the standard density estimation problem or the regression function estimation problem under the MISE over Besov spaces (see Härdle et al. (2012)). Under some strong conditions on the model as $U_i := 1$, the rate of convergence $n^{-\frac{2s}{2s+d}}$ is proved to be optimal in the minimax sense (see Härdle et al. (2012) and Tsybakov (2009)). So our nonlinear wavelet estimator can be optimal in the minimax sense up to a $\ln n$. However, in full generality, without specifying the distributions of U_i and V_i, the optimal lower bounds for the MISE are difficult to determine via standard techniques (Fano's lemma, . . .) and the optimality of our estimators remains an open question.

Remark 3.1. *Some assumptions used in Theorem 3.1 can be relaxed without changing the result. In particular, one can consider the domain $\Delta = [a,b]^d$ with $(a,b) \in \mathbb{R}^2$ and $a < b$ with an adaptation of the wavelet basis. In this case, we can also replace **A.2** by $\boldsymbol{X}_1$ with density function $h : [a,b]^d \to \mathbb{R}$ bounded from below, with the following wavelet coefficient estimators:*

$$\hat{\alpha}_{j,\boldsymbol{k}} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i^2}{h(\boldsymbol{X}_i)} \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_i) - v_{j,\boldsymbol{k}},$$

and

$$\hat{\beta}_{j,\boldsymbol{k},u} = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i^2}{h(\boldsymbol{X}_i)} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{X}_i) - w_{j,\boldsymbol{k},u}\right) \mathbb{1}_{\left\{\left|\frac{Y_i^2}{h(\boldsymbol{X}_i)} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{X}_i) - w_{j,\boldsymbol{k},u}\right| \leq \rho_n\right\}}.$$

*Finally, note that, in **A.6** can be improved by assuming g unknown. To the best of our knowledge, only Cai et al. (2008) have developed wavelet methods in this case for $d = 1$, deterministic design $(\boldsymbol{X}_i := i/n)$ and infinite moments for U_i. Extension of these methods in the general setting of (1) needs further developments that we leave for a future work.*

4   Numerical Experiments

To illustrate the empirical performance of the estimators proposed in this work, we carried out a simulation study. The objective is to highlight some of the theoretical
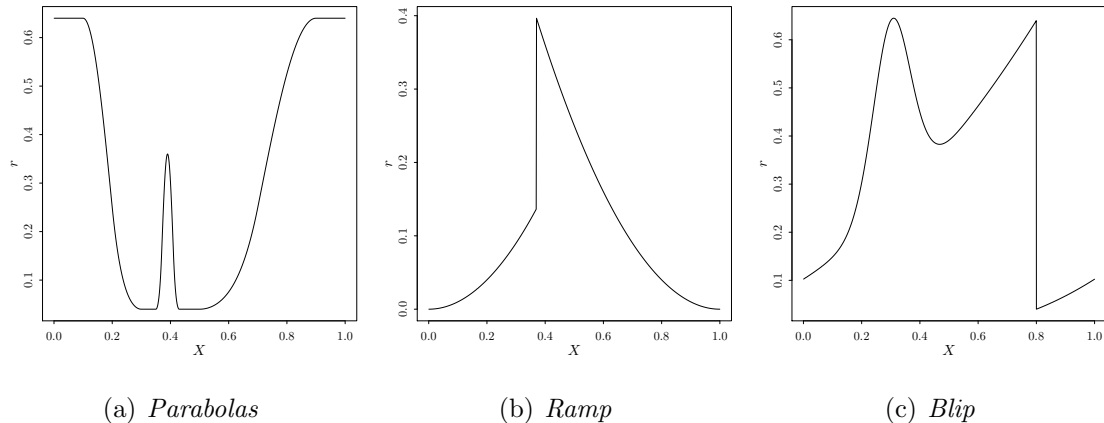
(a) *Parabolas*   (b) *Ramp*   (c) *Blip*

Figure 1: The three test functions considered.

findings using numerical examples. We begin by giving some details about the specificities inherent in wavelet estimators in a non-deterministic design framework. We also try to propose a realistic simulation setting using an adaptive selection method to select both the truncation parameter of the linear estimator and the threshold parameter of the non-linear estimator. In this context, we compare their empirical performances in the model with both multiplicative and additive noise. Simulations were performed using R and in particular the **rwavelet** package Navarro and Chesneau (2019) (available from https://github.com/fabnavarro/rwavelet).

4.1 Computational aspect of wavelets and parameters selection

For fixed design, thanks to Mallat's pyramidal algorithm (Mallat, 2008), the computation of wavelet-based estimators is simple and fast. When considering uniform random design, the implementation requires some changes and several strategies have been developed in the literature (see, *e.g.*, Cai et al. (1998); Hall et al. (1997)). For uniform design regression, Cai and Brown (1999) has proposed to use an approach in which the wavelet coefficients are computed by a simple application of Mallat's algorithm using the ordered Y_i's as input variables. We have followed this approach because it preserves the simplicity of calculation and the efficiency of the equispaced algorithm. In the context of wavelet regression in random design with heteroscedastic noise, Kulik et al. (2009) and Navarro and Saumard (2017) also adopted this approach.

Nason successfully adjusted the standard two-Fold Cross Validation (2FCV) method

to select the threshold parameter in wavelet shrinkage (see, Nason (1996)). For the calibration of linear wavelet estimators, his strategy was used by Navarro and Saumard (2017). We have chosen to apply this approach to select both the threshold and truncation parameter of linear and non-linear estimators. More precisely, in the linear case, we built a collection of linear estimators $\hat{r}^{\mathrm{lin}}_{j_*,n}, j_* = 0, 1, \ldots, \log 2(n) - 1$ (by successively adding whole resolution levels of wavelet coefficients), and select the best among this collection by minimizing a 2FCV criterion denoted by $2\mathrm{FCV}(j_*)$. The resulting estimator of the truncation level is denoted by $\hat{\hat{j}}_*$ and the corresponding estimator of r by $\hat{r}^{\mathrm{lin}}_{\hat{\hat{j}}_*,n}$ (see, Navarro and Saumard (2017, 2018) for more details). For the nonlinear estimator, the same estimator $\hat{\hat{j}}_*$ of the truncation parameter j_* obtained for the linear is used. The estimator of the thresholding parameter is obtained using the 2FCV method developed in Nason (1996). The parameter j_1 is fixed *a priori* as the maximum level allowed by the wavelet decomposition (*i.e.*, $j_1 = \log 2(n) - 1$). It is a classic choice that allows the coefficients to be selected down to the smallest scale. In addition, in order to facilitate and not to overburden the implementation of the nonlinear estimator, we perform a standard hard thresholding of the wavelet coefficient estimators (rather than the double threshold used in its definition). In order to be able to evaluate the performance of these two criteria, the mean square error (MSE) is used (*i.e.*, $\mathrm{MSE}(\hat{r}_{j_*,n}, r) = \frac{1}{n} \sum_{i=1}^{n} (r(X_i) - \hat{r}_{j_*,n}(X_i))^2)$). We consider three test functions for r (see Figure 1), commonly used in the wavelet literature, *Parabolas, Ramp* and *Blip* (see, *e.g.*, Donoho et al. (1995)). In all simulations, we examine the case $d = 1$, the design is chosen to satisfy **A.2** (*i.e.*, $\mathcal{U}([0,1])$) and the choice of the wavelet family used is also fixed (*i.e.*, Daubechies compactly supported wavelet with 8 vanishing moments).

4.2 Additive-multiplicative regression

This subsection examines the behaviour and performance of linear and non-linear estimators in the context of additive and multiplicative regression by considering $V_1 \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2 = 0.01$ and $U_1 \sim U([-1, 1])$. Thus, the goal is to estimate the frontier r from (X_i, Y_i) sample simulated from one of the test functions. By applying one of the linear or nonlinear methods developed above to the estimation of r, one can construct an estimator whose rate of convergence is given by (5a) and (5b) respectively.

Note that here, the nature of the frontier function is not necessarily the same as that commonly found in the literature on stochastic boundary estimation. Indeed, here r is not necessarily a production function (*e.g.*, r is concave), the only assumption we make is given by **A.1**. Thus, the application here can be seen as the estimation of the boundary or frontier of a sample affected by some additive (positive) noise (see Jirak et al. (2014)).
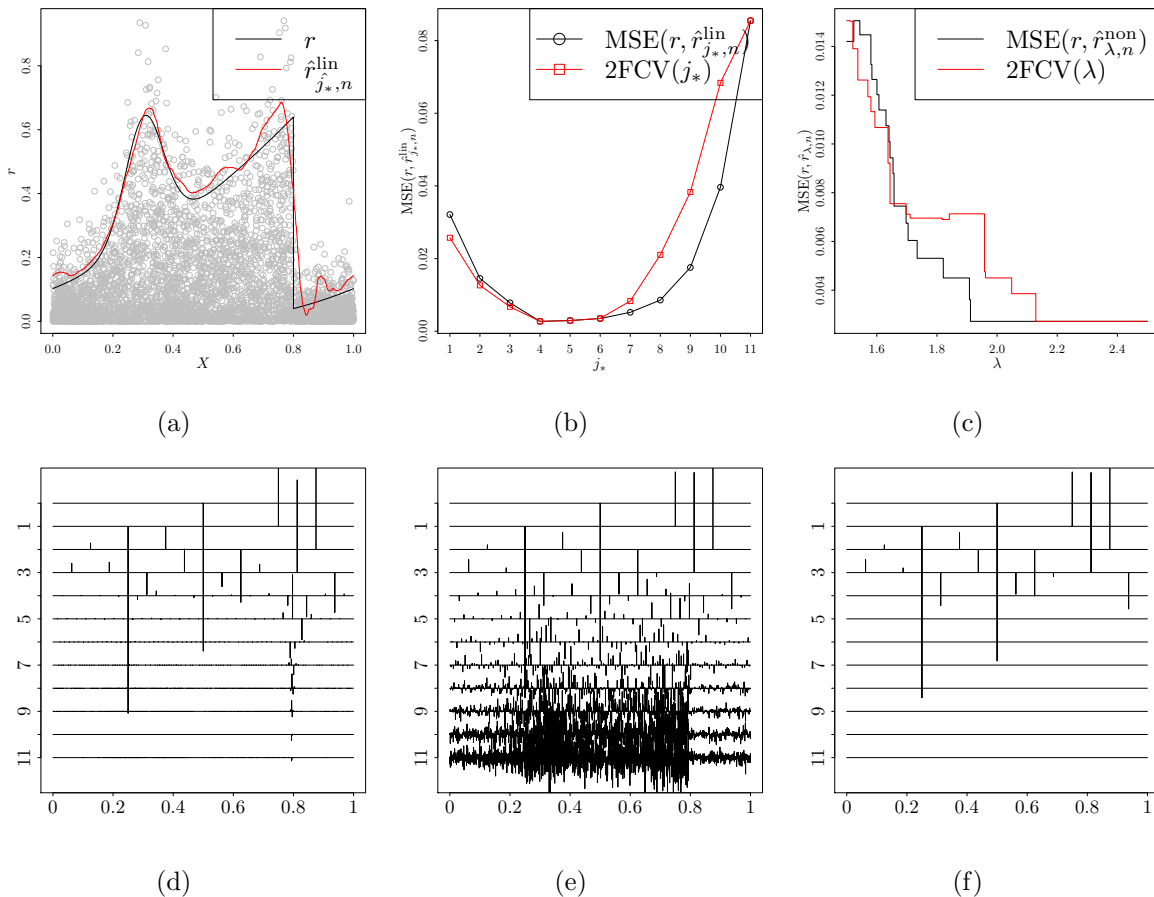

Figure 2: Typical estimation from a single simulation for $n = 4096$ and $\sigma^2 = 0.01$. Noisy observations (X, Y^2) (grey circles), true function (black line) and $\hat{r}^{\mathrm{lin}}_{\hat{j}_*,n}$ (a). Graphs of the MSE (black line) against j_* and (rescaled) 2FCV($\cdot$) criterion (red line) for the linear (b) and nonlinear (c) cases respectively. Original wavelet coefficients (d). Noisy wavelet coefficients (e). Estimated wavelet coefficients (f).

A typical example of estimation for the Blip function, with $n = 4096$ is given in Figure 2. It can be seen that the minimum of 2FCV(j_*) criteria coincides with that of unknown risk (*i.e.*, $\hat{j}_* = 4$) and therefore provides the best possible linear estimator for the collection under consideration (*i.e.*, MSE$(r, \hat{r}^{\mathrm{lin}}_{\hat{j}_*,n}) = 0.0027$). We have not included
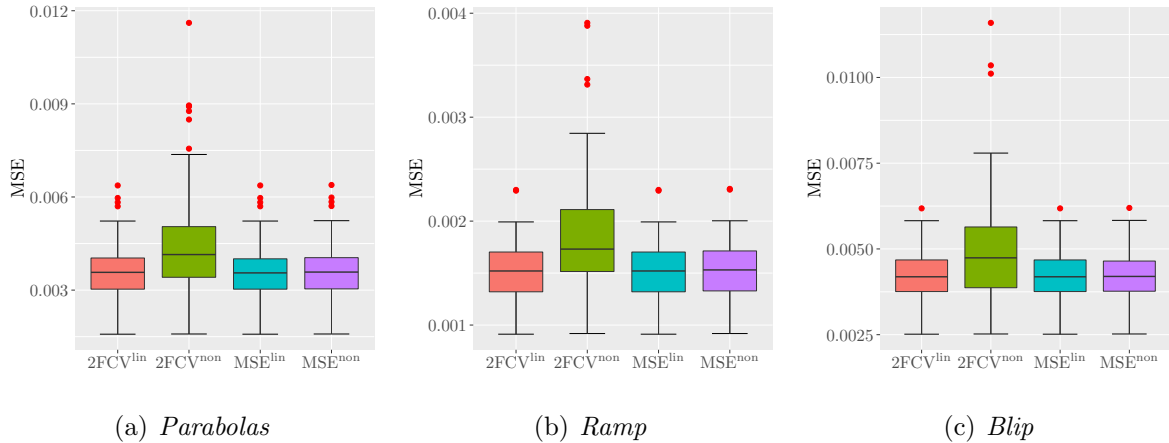
(a) *Parabolas* (b) *Ramp* (c) *Blip*

Figure 3: Box plots of the MSE for the three-test functions with $n = 4096$ and $\sigma^2 = 0.01$.

the results of the non-linear estimator in Figure 2. Indeed, in this case, the value of the threshold obtained by minimizing the cross validated criterion leads to the elimination of all the thresholded coefficients (*i.e.*, going from $\hat{j}_*$ to j_1) and therefore leads to the same estimate and the same risk as the linear estimator. We can see (Figure 2(c)) that the unknown risk behaves in the same way here. This is partly because the amplitude of the coefficients at the fine scales is so large and variable from one scale to another that it is not possible to obtain an overall optimal threshold value that makes it possible to maintain certain important coefficients and that, on the contrary, keeps coefficients associated with noise, with this specific thresholding policy (*i.e.*, a 'keep' or 'kill' rule). In particular the important coefficients located on scales larger than $\hat{j}_*$ (especially those encoding the discontinuity of r) are too small in amplitude to be maintained by a global threshold.

In order to determine whether this phenomenon observed for a single function and a single realization is confirmed in a more general context, we compare the performance in terms of MSE (computed on the functions after reconstruction) for both estimators and for the three-test functions. For each function, a sample of $N = 100$ is generated and we compare the average behavior of the MSE for parameters selected with the oracle obtained by minimizing the MSE using the original signal r (denoted by MSE$^{\text{lin}}$ and MSE$^{\text{non}}$ respectively), the linear 2FCV$^{\text{lin}}$ strategy and the non-linear 2FCV$^{\text{non}}$ (*i.e.*, calculated from $\hat{j}_*$ and the threshold that minimizes the 2FCV(λ)) strategy. Figure 3 presents the results in the form of boxplots, one for each function. On the one hand, for all three functions, we can see that the performance of 2FCV$^{\text{lin}}$ is at the MSE$^{\text{lin}}$

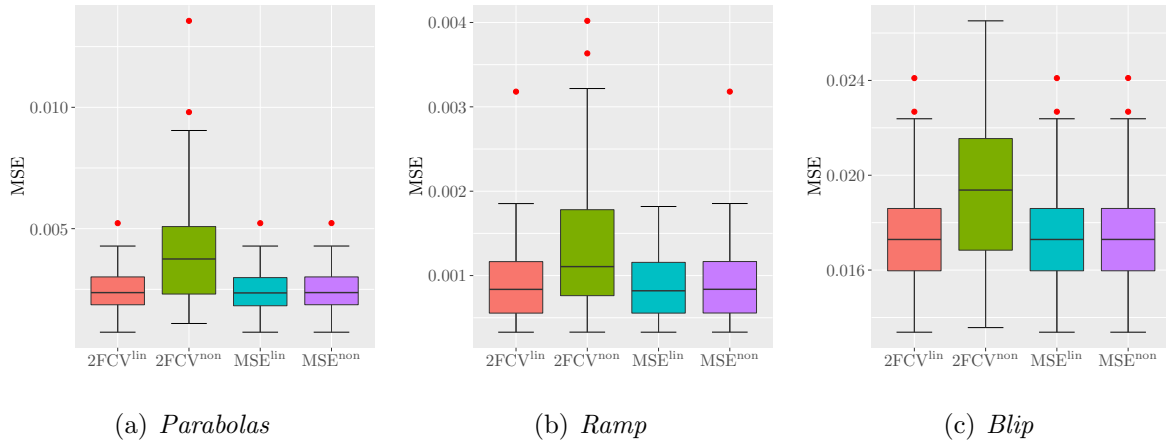(a) *Parabolas*   (b) *Ramp*   (c) *Blip*

Figure 4: Box plots of the MSE for the three-test functions with $n = 2048$ and $\sigma^2 = 0.025$.

level. This procedure therefore provides a remarkable surrogate of the unknown risk. On the other hand, the non-linear $2FCV^{non}$ oracle is similar to MSE^{lin}, which means that the optimal threshold here leads systematically to the suppression of all threshold coefficients — which corresponds to the selection of values of the threshold parameter which is greater than the largest noisy wavelet coefficient in absolute value. Finally, the variability of the $2FCV^{non}$ is high as a result of selected threshold values that are sometimes too small, resulting in the conservation of unnecessary coefficients in the reconstruction. This is because the curves associated with non-linear criteria do not generally allow a single global minimum, but the minimum is reached in the form of a plateau (see Figure 2(c)). In practice, when the minimum is reached on such a plateau, the first element that constitutes it is selected first. This has no influence on MSE^{non} but generates this variability of $2FCV^{non}$, *i.e.*, when the abscissa of the first point constituting the plateau associated with $2FCV^{non}$ is lower than that of MSE^{non}) Note that to overcome this problem, in the presence of a plateau, we could for example select a threshold value in the middle of it. We have not done so here to emphasize the fact that a cross validation strategy of the global threshold seems ineffective in this setting. It should also be noted that in our simulations, this finding is also verified for other noise levels or sample sizes (the results are generally very similar, so we give only another example by considering a lower number of samples, $n = 2048$ and a lower additive noise level $\sigma^2 = 0.025$).

In conclusion, the linear approach seems more appropriate than the non-linear ap-

proach in the context of the simulations considered in this study. One way to fully benefit from the non-linear approach would be to consider an optimal threshold selection strategy on a scale by scale basis. The selection procedure used here, based on an interpolation performed in the original domain, does not facilitate this extension. For this purpose it would be necessary, for example, to define an interpolated version of the cross validation method in the wavelet coefficients domain.

5 Auxiliary results and proof of the main result

5.1 Auxiliary results

In this section, we provide some lemmas for the proof of the main Theorem.

Lemma 5.1. *Let* $j \geq \tau$, $\boldsymbol{k} \in \Lambda_j$, $\hat{\alpha}_{j,\boldsymbol{k}}$ *be* (3). *Then, under* ***H.1*** *or* ***H.2***, *we have*

$$\mathbf{E}[\hat{\alpha}_{j,\boldsymbol{k}}] = \alpha_{j,\boldsymbol{k}}, \quad \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \Psi_{j,\boldsymbol{k},u}(\boldsymbol{X}_i) - w_{j,\boldsymbol{k},u}\right] = \beta_{j,\boldsymbol{k},u}.$$

Proof of Lemma 5.1. Using the independence assumptions on the random variables, **H.1** or **H.2**, observe that

$$\mathbf{E}\left[U_1 V_1 f(\boldsymbol{X}_1) \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] = \begin{cases} \mathbf{E}[U_1]\mathbf{E}[V_1]\mathbf{E}\left[f(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] & \text{under } \mathbf{A.5}, \\ \mathbf{E}[U_1]\mathbf{E}\left[V_1 f(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] & \text{under } \mathbf{A.6}, \end{cases}$$

$$= 0,$$

and

$$v_{j,\boldsymbol{k}} = \begin{cases} \mathbf{E}\left[V_1^2\right] 2^{-jd/2} = \mathbf{E}\left[V_1^2\right] \int_{[0,1]^d} \Phi_{j,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x} = \mathbf{E}\left[V_1^2\right]\mathbf{E}\left[\Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] & \text{under } \mathbf{A.5}, \\ \int_{[0,1]^d} g^2(\boldsymbol{x})\Phi_{j,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x} & \text{under } \mathbf{A.6}, \end{cases}$$

$$= \mathbf{E}\left[V_1^2 \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right].$$

Therefore

$$\mathbf{E}[\hat{\alpha}_{j,\boldsymbol{k}}] = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_i) - v_{j,\boldsymbol{k}}\right] = \mathbf{E}\left[Y_1^2 \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] - v_{j,\boldsymbol{k}}$$

$$= \mathbf{E}\left[U_1^2 r(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] + 2\mathbf{E}\left[U_1 V_1 f(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] + \mathbf{E}\left[V_1^2 \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] - v_{j,\boldsymbol{k}}$$

$$= \mathbf{E}\left[U_1^2\right]\mathbf{E}\left[r(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] = \int_{[0,1]^d} r(\boldsymbol{x})\Phi_{j,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x} = \alpha_{j,\boldsymbol{k}}.$$

Using similar mathematical arguments, since $\int_{[0,1]^d} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})d\boldsymbol{x} = 0$, we have

$$
w_{j,\boldsymbol{k},u} = \begin{cases} 0 = \mathbf{E}\left[V_1^2\right] \displaystyle\int_{[0,1]^d} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})d\boldsymbol{x} = \mathbf{E}\left[V_1^2\right] \mathbf{E}\left[\Psi_{j,\boldsymbol{k},u}(\boldsymbol{X}_1)\right] & \text{under } \mathbf{A.5}, \\[4mm] \displaystyle\int_{[0,1]^d} g^2(\boldsymbol{x})\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})d\boldsymbol{x} & \text{under } \mathbf{A.6}, \end{cases}
$$
$$
= \mathbf{E}\left[V_1^2 \Psi_{j,\boldsymbol{k},u}(\boldsymbol{X}_1)\right].
$$

We prove the second equality. The proof of Lemma 5.1 is complete. □

Lemma 5.2. *Let* $j \geq \tau$ *such that* $2^j \leq n$, $\boldsymbol{k} \in \Lambda_j$, $\hat{\alpha}_{j,\boldsymbol{k}}$ *and* $\hat{\beta}_{j,\boldsymbol{k},u}$ *be* (3) *and* (4) *respectively. Then, under* **H.1** *or* **H.2***,*

$$
\mathbf{E}\left[(\hat{\alpha}_{j,\boldsymbol{k}} - \alpha_{j,\boldsymbol{k}})^2\right] \lesssim \frac{1}{n}, \quad \mathbf{E}\left[(\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u})^2\right] \lesssim \frac{\ln n}{n}.
$$

Proof of Lemma 5.2. Owing to Lemma 5.1 we have $\mathbf{E}[\hat{\alpha}_{j,\boldsymbol{k}}] = \alpha_{j,\boldsymbol{k}}$. Therefore

$$
\begin{aligned}
\mathbf{E}\left[(\hat{\alpha}_{j,\boldsymbol{k}} - \alpha_{j,\boldsymbol{k}})^2\right] = \mathbf{V}\left[\hat{\alpha}_{j,\boldsymbol{k}}\right] &= \mathbf{V}\left[\frac{1}{n}\sum_{i=1}^n Y_i^2 \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_i) - v_{j,\boldsymbol{k}}\right] \\
&= \mathbf{V}\left[\frac{1}{n}\sum_{i=1}^n Y_i^2 \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_i)\right] = \frac{1}{n}\mathbf{V}\left[Y_1^2 \Phi_{j,\boldsymbol{k}}(\boldsymbol{X}_1)\right] \leq \frac{1}{n}\mathbf{E}\left[Y_1^4 \Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right] \\
&\lesssim \frac{1}{n}\left(\mathbf{E}\left[U_1^4 f^4(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right] + \mathbf{E}\left[V_1^4 \Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right]\right) \\
&= \frac{1}{n}\left(\mathbf{E}\left[U_1^4\right] \mathbf{E}\left[f^4(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right] + \mathbf{E}\left[V_1^4 \Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right]\right). \quad (6)
\end{aligned}
$$

By **A.1** and $\mathbf{E}\left[\Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right] = \int_{[0,1]^d} \left(\Phi_{j,\boldsymbol{k}}(\boldsymbol{x})\right)^2 d\boldsymbol{x} = 1$, we have $\mathbf{E}\left[f^4(\boldsymbol{X}_1)\Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right] \lesssim 1$. On the other hand, we have

$$
\mathbf{E}\left[V_1^4 \Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right] = \begin{cases} \mathbf{E}\left[V_1^4\right] \mathbf{E}\left[\Phi_{j,\boldsymbol{k}}^2(\boldsymbol{X}_1)\right] \lesssim 1 & \text{under } \mathbf{A.5}, \\[4mm] \displaystyle\int_{[0,1]^d} g^4(\boldsymbol{x})\Phi_{j,\boldsymbol{k}}^2(\boldsymbol{x})d\boldsymbol{x} \lesssim \int_{[0,1]^d} \Phi_{j,\boldsymbol{k}}^2(\boldsymbol{x})d\boldsymbol{x} = 1 & \text{under } \mathbf{A.6}. \end{cases}
$$

Thus all the terms in the brackets of (6) are bounded from above. The first inequality in Lemma 5.2 is proved.

Now, by the definition of $\hat{\beta}_{j,\boldsymbol{k},u}$, taking $K_i := Y_i^2 \Psi_{j,\boldsymbol{k},u}(\boldsymbol{X}_i) - w_{j,\boldsymbol{k},u}$ and $D_i := K_i \mathbb{1}_{\{|K_i| \leq \rho_n\}} - \mathbf{E}\left[K_i \mathbb{1}_{\{|K_i| \leq \rho_n\}}\right]$ for the sake of simplicity, the second equation in Lemma 5.1 yields

$$
\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u} = \frac{1}{n}\sum_{i=1}^n D_i - \mathbf{E}\left[K_1 \mathbb{1}_{\{|K_1| > \rho_n\}}\right].
$$

Hence, using $\mathbf{E}\left[\left((1/n)\sum_{i=1}^{n}D_i\right)^2\right] = (1/n^2)\mathbf{V}\left[\sum_{i=1}^{n}D_i\right] = \mathbf{V}[D_1]/n \leq \mathbf{E}\left[D_1^2\right]/n$, we have

$$\mathbf{E}\left[(\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u})^2\right] \lesssim \mathbf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}D_i\right)^2\right] + \left(\mathbf{E}\left[|K_1|\mathbb{1}_{\{|K_1|>\rho_n\}}\right]\right)^2.$$

$$\lesssim \frac{1}{n}\mathbf{E}\left[D_1^2\right] + \left(\mathbf{E}\left[|K_1|\mathbb{1}_{\{|K_1|>\rho_n\}}\right]\right)^2.$$

Proceeding as for the proof of the first inequality, using the assumptions **H.1** or **H.2**, note that $\mathbf{E}\left[K_1^2\right] \lesssim \mathbf{E}\left[Y_1^4\Psi_{j,\boldsymbol{k},u}^2(\boldsymbol{X}_1)\right] + w_{j,\boldsymbol{k},u}^2 \lesssim 1$ and

$$\left(\mathbf{E}\left[K_1\mathbb{1}_{\{|K_1|>\rho_n\}}\right]\right)^2 \lesssim \left(\mathbf{E}\left[K_1^2\right]/\rho_n\right)^2 \lesssim \frac{\ln n}{n}. \tag{7}$$

Therefore,

$$\mathbf{E}\left[(\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u})^2\right] \lesssim \frac{1}{n} + \frac{\ln n}{n} \lesssim \frac{\ln n}{n}.$$

The second inequality in Lemma 5.2 is proved. This ends the proof of Lemma 5.2. $\square$

Lemma 5.3. *Let $j \geq \tau$ such that $2^{jd} \lesssim n/\ln n$, $\boldsymbol{k} \in \Lambda_j$, $\hat{\beta}_{j,\boldsymbol{k},u}$ be (4). Then there exists a constant $\kappa > 1$ such that*

$$\mathbb{P}(|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| \geq \kappa t_n) \lesssim n^{-4}.$$

Proof of Lemma 5.3. By the definition of $\hat{\beta}_{j,\boldsymbol{k},u}$, taking $K_i := Y_i^2\Psi_{j,\boldsymbol{k},u}(\boldsymbol{X}_i) - w_{j,\boldsymbol{k},u}$ and $D_i := K_i\mathbb{1}_{\{|K_i|\leq\rho_n\}} - E\left[K_i\mathbb{1}_{\{|K_i|\leq\rho_n\}}\right]$ for the sake of simplicity, the second equation in Lemma 5.1 yields

$$|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| \lesssim \frac{1}{n}\left|\sum_{i=1}^{n}D_i\right| + E\left[|K_1|\mathbb{1}_{\{|K_1|>\rho_n\}}\right].$$

Using (7), there exists $c > 0$ such that $\mathbf{E}\left[|K_1|\mathbb{1}_{\{|K_1|>\rho_n\}}\right] \leq c\sqrt{\ln n/n}$. Then

$$\{|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| \geq \kappa t_n\} \subseteq \left\{\frac{1}{n}\left|\sum_{i=1}^{n}D_i\right| \geq (\kappa - c)t_n\right\}.$$

Note that $\mathbf{E}[D_i] = 0$ thanks to Lemma 5.1. According to the proof of Lemma 5.2, $\mathbf{E}[D_i^2] := \delta^2 \lesssim 1$. This with $|D_i| \lesssim \sqrt{n/\ln n}$ and Bernstein inequality shows

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}D_i\right| \geq (\kappa - c)t_n\right) \lesssim \exp\left\{-\frac{n(\kappa - c)^2t_n^2}{2(\delta^2 + (\kappa - c)t_n\rho_n/3)}\right\}$$

$$\lesssim \exp\left\{-\ln n\frac{(\kappa - c)^2}{2(\delta^2 + (\kappa - c)/3)}\right\}$$

$$\lesssim n^{-\frac{(\kappa-c)^2}{2(\delta^2+(\kappa-c)/3)}}.$$

Then one choose large enough κ such that

$$\mathbb{P}(|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| \geq \kappa t_n) \lesssim n^{-\frac{(\kappa-1)^2}{2(\delta^2+(\kappa-1)/3)}} \lesssim n^{-4}.$$

This is the desired conclusion. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

5.2   Proof of the main result

This section is devoted to the proof of Theorem 3.1. We prove (5a) and (5b) in turn.

Proof of (5a)   Note that

$$\mathbf{E}\left[\int_{[0,1]^d}\left|\hat{r}_n^{\mathrm{lin}}(\boldsymbol{x}) - r(\boldsymbol{x})\right|^2 d\boldsymbol{x}\right] = \mathbf{E}\left[\left\|\hat{r}_n^{\mathrm{lin}} - P_{j_*}r\right\|_2^2\right] + \left\|P_{j_*}r - r\right\|_2^2. \qquad (8)$$

It is easy to see that

$$\mathbf{E}\left[\left\|\hat{r}_n^{\mathrm{lin}} - P_{j_*}r\right\|_2^2\right] = \mathbf{E}\left[\left\|\sum_{\boldsymbol{k}\in\Lambda_{j_*}}(\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}})\Phi_{j_*,\boldsymbol{k}}\right\|_2^2\right] = \sum_{\boldsymbol{k}\in\Lambda_{j_*}}\mathbf{E}\left[\left|\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}}\right|^2\right].$$

According to Lemma 5.2, $|\Lambda_{j_*}| \sim 2^{j_*d}$ and $2^{j_*} \sim n^{\frac{1}{2s'+d}}$,

$$\mathbf{E}\left[\left\|\hat{r}_n^{\mathrm{lin}} - P_{j_*}r\right\|_2^2\right] \lesssim \frac{2^{j_*d}}{n} \sim n^{-\frac{2s'}{2s'+d}}. \qquad (9)$$

When $p \geq 2$, $s' = s$. By Hölder inequality and $r \in B_{p,q}^s([0,1]^d)$,

$$\|P_{j_*}r - r\|_2^2 \lesssim \|P_{j_*}r - r\|_p^2 \lesssim 2^{-2j_*s} \sim n^{-\frac{2s}{2s+d}}.$$

When $1 \leq p < 2$ and $s > d/p$, $B_{p,q}^s([0,1]^d) \subseteq B_{2,\infty}^{s'}([0,1]^d)$

$$\|P_{j_*}r - r\|_2^2 \lesssim \sum_{j=j_*}^{\infty} 2^{-2js'} \lesssim 2^{-2j_*s'} \sim n^{-\frac{2s'}{2s'+d}}.$$

Therefore, in both cases,

$$\|P_{j_*}r - r\|_2^2 \lesssim n^{-\frac{2s'}{2s'+d}}. \qquad (10)$$

By (8), (9) and (10),

$$\mathbf{E}\left[\int_{[0,1]^d}\left|\hat{r}_n^{\mathrm{lin}}(\boldsymbol{x}) - r(\boldsymbol{x})\right|^2 d\boldsymbol{x}\right] \lesssim n^{-\frac{2s'}{2s'+d}}.$$

Proof of (5b) We now follow the lines of (Delyon and Juditsky, 1996, Theorem 2) with adaptation to our statistical setting, by using the definitions of our estimators and the

auxiliary results of Section 5.1. By the definitions of $\hat{r}_n^{\mathrm{lin}}$ and $\hat{r}_n^{\mathrm{non}}$, we have

$$\hat{r}_n^{\mathrm{non}}(\boldsymbol{x}) - r(\boldsymbol{x}) = \left(\hat{r}_n^{\mathrm{lin}}(\boldsymbol{x}) - P_{j_*}r(\boldsymbol{x})\right) - \left(r(\boldsymbol{x}) - P_{j_1+1}r(\boldsymbol{x})\right)$$

$$+ \sum_{j=j_*}^{j_1}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}\left(\hat{\beta}_{j,\boldsymbol{k},u}\mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\kappa t_n\}} - \beta_{j,\boldsymbol{k},u}\right)\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}).$$

Hence,

$$\mathbf{E}\left[\int_{[0,1]^d}\left|\hat{r}_n^{\mathrm{non}}(\boldsymbol{x}) - r(\boldsymbol{x})\right|^2 d\boldsymbol{x}\right] \lesssim T_1 + T_2 + Q,$$

where $T_1 := \mathbf{E}\left[\left\|\hat{r}_n^{\mathrm{lin}} - P_{j_*}r\right\|_2^2\right], \quad T_2 := \left\|r - P_{j_1+1}r\right\|_2^2$ and

$$Q := \mathbf{E}\left[\left\|\sum_{j=j_*}^{j_1}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}\left(\hat{\beta}_{j,\boldsymbol{k},u}\mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\kappa t_n\}} - \beta_{j,\boldsymbol{k},u}\right)\Psi_{j,\boldsymbol{k},u}\right\|_2^2\right].$$

According to (9) and $2^{j_*} \sim n^{\frac{1}{2m+d}}$ $(m > s)$,

$$T_1 = \mathbf{E}\left[\left\|\hat{r}_n^{\mathrm{lin}} - P_{j_*}r\right\|_2^2\right] \lesssim \frac{2^{j_*d}}{n} \sim n^{-\frac{2m}{2m+d}} < n^{-\frac{2s}{2s+d}}.$$

When $p \geq 2$, by the same arguments as (10) shows $T_2 = \left\|r - P_{j_1+1}r\right\|_2^2 \lesssim 2^{-2j_1s}$. This with $2^{j_1} \sim (n/\ln n)^{\frac{1}{d}}$ leads to

$$T_2 \lesssim 2^{-2j_1s} \sim \left(\frac{\ln n}{n}\right)^{\frac{2s}{d}} \leq (\ln n)n^{-\frac{2s}{2s+d}}.$$

On the other hand, $B_{p,q}^s([0,1]^d) \subseteq B_{2,\infty}^{s+d/2-d/p}([0,1]^d)$ when $1 \leq p < 2$ and $s > d/p$. Then

$$T_2 \lesssim 2^{-2j_1(s+\frac{d}{2}-\frac{d}{p})} \sim \left(\frac{\ln n}{n}\right)^{\frac{2(s+\frac{d}{2}-\frac{d}{p})}{d}} \leq (\ln n)n^{-\frac{2s}{2s+d}}.$$

Hence,

$$T_2 \lesssim (\ln n)n^{-\frac{2s}{2s+d}},$$

for each $1 \leq p < +\infty$.

The main work for the proof of (5b) is to show

$$Q = \mathbf{E}\left[\left\|\sum_{j=j_*}^{j_1}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}\left(\hat{\beta}_{j,\boldsymbol{k},u}\mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\kappa t_n\}} - \beta_{j,\boldsymbol{k},u}\right)\Psi_{j,\boldsymbol{k},u}\right\|_2^2\right] \lesssim (\ln n)n^{-\frac{2s}{2s+d}}.$$

Note that

$$Q = \sum_{j=j_*}^{j_1}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}\mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}\mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\kappa t_n\}} - \beta_{j,\boldsymbol{k},u}\right|^2\right] \lesssim Q_1 + Q_2 + Q_3, \qquad (11)$$

where

$$Q_1 = \sum_{j=j_*}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^2 \mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\frac{\kappa t_n}{2}\}}\right],$$

$$Q_2 = \sum_{j=j_*}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^2 \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}|\geq\frac{\kappa t_n}{2}\}}\right],$$

$$Q_3 = \sum_{j=j_*}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^2 \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}|\leq 2\kappa t_n\}}.$$

For Q_1, one observes that

$$\mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^2 \mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\frac{\kappa t_n}{2}\}}\right] \leq \left(\mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^4\right]\right)^{\frac{1}{2}} \left(\mathbb{P}\left(|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| > \frac{\kappa t_n}{2}\right)\right)^{\frac{1}{2}}$$

thanks to Hölder inequality. By Lemma 5.2, Lemma 5.3 and $|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}|^2 \lesssim n/\ln n$,

$$\mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^2 \mathbb{1}_{\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\frac{\kappa t_n}{2}\}}\right] \lesssim \frac{1}{n^2}.$$

Then $Q_1 \lesssim \sum_{j=j_*}^{j_1} 2^{jd}/n^2 \lesssim 2^{j_1 d}/n^2 \lesssim 1/n \leq n^{-\frac{2s}{2s+d}}$, where one uses the choice $2^{j_1} \sim (n/\ln n)^{\frac{1}{d}}$. Hence,

$$Q_1 \leq n^{-\frac{2s}{2s+d}}. \tag{12}$$

To estimate Q_2, one defines

$$2^{j'} \sim n^{\frac{1}{2s+d}}.$$

It is easy to see that $2^{j_*} \sim n^{\frac{1}{2m+d}} \leq 2^{j'} \sim n^{\frac{1}{2s+d}} \leq 2^{j_1} \sim (n/\ln n)^{\frac{1}{d}}$. Furthermore, one rewrites

$$Q_2 = \left(\sum_{j=j_*}^{j'} + \sum_{j=j'+1}^{j_1}\right) \left\{\sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^2 \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}|\geq\frac{\kappa t_n}{2}\}}\right]\right\}$$

$$:= Q_{21} + Q_{22}.$$

By Lemma 5.2 and $2^{j'} \sim n^{\frac{1}{2s+d}}$,

$$Q_{21} := \sum_{j=j_*}^{j'} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^2 \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}|\geq\frac{\kappa t_n}{2}\}}\right]$$

$$\lesssim \sum_{j=j_*}^{j'} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \frac{\ln n}{n} \lesssim \sum_{j=j_*}^{j'} (\ln n)\frac{2^{jd}}{n} \lesssim (\ln n)\frac{2^{j'd}}{n} \sim (\ln n)n^{-\frac{2s}{2s+d}}.$$

On the other hand, it follows from Lemma 5.2 that

$$Q_{22} := \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \mathbf{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^2 \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{\kappa t_n}{2}\}}\right]$$

$$\lesssim \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \frac{\ln n}{n} \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{\kappa t_n}{2}\}}.$$

When $p \geq 2$, since $r \in B_{p,q}^s([0,1]^d)$, Lemma 5.2 and $t_n = \sqrt{\ln n/n}$,

$$Q_{22} \lesssim \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \frac{\ln n}{n} \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{\kappa t_n}{2}\}} \lesssim \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \frac{\ln n}{n} \left(\frac{\beta_{j,\boldsymbol{k},u}}{\kappa t_n/2}\right)^2$$

$$\lesssim \sum_{j=j'+1}^{j_1} 2^{-2js} \lesssim 2^{-2j's} \sim n^{-\frac{2s}{2s+d}}. \tag{13}$$

When $1 \leq p < 2$ and $s > d/p$, $B_{p,q}^s([0,1]^d) \subseteq B_{2,\infty}^{s+d/2-d/p}([0,1]^d)$. Then

$$Q_{22} \lesssim \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \frac{\ln n}{n} \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{\kappa t_n}{2}\}} \lesssim \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \frac{\ln n}{n} \left(\frac{\beta_{j,\boldsymbol{k},u}}{\kappa t_n/2}\right)^p$$

$$\lesssim \sum_{j=j'+1}^{j_1} (\ln n) n^{\frac{p}{2}-1} 2^{-j(s+d/2-d/p)p}$$

$$\lesssim (\ln n) n^{\frac{p}{2}-1} 2^{-j'(s+d/2-d/p)p} \sim (\ln n) n^{-\frac{2s}{2s+d}}. \tag{14}$$

It follows from the upper bounds above that

$$Q_2 \lesssim (\ln n) n^{-\frac{2s}{2s+d}}. \tag{15}$$

Finally, one evaluates Q_3. Clearly,

$$Q_{31} := \sum_{j=j_*}^{j'} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} |\beta_{j,\boldsymbol{k},u}|^2 \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}| \leq 2\kappa t_n\}}$$

$$\leq \sum_{j=j_*}^{j'} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} \left|2\kappa t_n\right|^2 \lesssim \sum_{j=j_*}^{j'} \frac{\ln n}{n} 2^{jd} \lesssim \frac{\ln n}{n} 2^{j'd}.$$

This with the choice of $2^{j'}$ shows

$$Q_{31} \lesssim (\ln n) n^{-\frac{2s}{2s+d}}.$$

On the other hand, $Q_{32} := \sum\limits_{j=j'+1}^{j_1} \sum\limits_{u=1}^{2^d-1} \sum\limits_{\boldsymbol{k} \in \Lambda_j} |\beta_{j,\boldsymbol{k},u}|^2 \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}| \leq 2\kappa t_n\}}$. According to the arguments of (13), for $p \geq 2$,

$$Q_{32} \lesssim \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} |\beta_{j,\boldsymbol{k},u}|^2 \lesssim n^{-\frac{2s}{2s+d}}.$$

When $1 \le p < 2$, $|\beta_{j,\boldsymbol{k},u}|^2 \, \mathbb{1}_{\{|\beta_{j,\boldsymbol{k},u}| \le 2\kappa t_n\}} \le |\beta_{j,\boldsymbol{k},u}|^p \, |2\kappa t_n|^{2-p}$. Then similar to the arguments of (14),

$$
\begin{aligned}
Q_{32} &\lesssim \sum_{j=j'+1}^{j_1} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k} \in \Lambda_j} |\beta_{j,\boldsymbol{k},u}|^p \, |2\kappa t_n|^{2-p} \\
&\lesssim \left(\frac{\ln n}{n}\right)^{\frac{2-p}{2}} \sum_{j=j'+1}^{j_1} 2^{-j(s+d/2-d/p)p} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{2-p}{2}} 2^{-j'(s+d/2-d/p)p} \\
&\lesssim \left(\frac{\ln n}{n}\right)^{\frac{2-p}{2}} \left(\frac{1}{n}\right)^{\frac{(s+d/2-d/p)p}{2s+d}} \le (\ln n) n^{-\frac{2s}{2s+d}}.
\end{aligned}
$$

It follows from the inequalities above that

$$
Q_3 \lesssim (\ln n) n^{-\frac{2s}{2s+d}}, \tag{16}
$$

in both cases. Owing to (11), (12), (15), and (16), we prove that

$$
Q \lesssim (\ln n) n^{-\frac{2s}{2s+d}},
$$

which is the desired conclusion.   □

Acknowledgements

References

Abramovich, F., T. C. Bailey, and T. Sapatinas (2000). Wavelet analysis and its statistical applications. *Journal of the Royal Statistical Society: Series D (The Statistician) 49*(1), 1–29.

Brown, L. D., M. Levine, et al. (2007). Variance estimation in nonparametric regression via the difference sequence method. *The Annals of Statistics 35*(5), 2219–2232.

Cai, T. T. and L. D. Brown (1999). Wavelet estimation for samples with random uniform design. *Statistics & Probability Letters 42*(3), 313–321.

Cai, T. T., L. D. Brown, et al. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics 26*(5), 1783–1799.

Cai, T. T., L. Wang, et al. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics 36*(5), 2025–2054.

Chaubey, Y. P., C. Chesneau, and H. Doosti (2015). Adaptive wavelet estimation of a density from mixtures under multiplicative censoring. *Statistics 49*(3), 638–659.

Chesneau, C. (2013). On the adaptive wavelet estimation of a multidimensional regression function under α-mixing dependence: Beyond the standard assumptions on the noise. *Commentationes Mathematicae Universitatis Carolinae 4*, 527–556.

Chesneau, C., J. Kou, and F. Navarro (2019). Linear wavelet estimation in regression with additive and multiplicative noise. working paper or preprint.

Chichignoud, M. (2012). Minimax and minimax adaptive estimation in multiplicative regression: locally bayesian approach. *Probability Theory and Related Fields 153*(3-4), 543–586.

Cohen, A., I. Daubechies, and P. Vial (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis 1*(1), 54–81.

Comte, F. (2015). *Estimation non-paramétrique*. Spartacus-IDH.

Daouia, A. and L. Simar (2005). Robust nonparametric estimators of monotone boundaries. *Journal of Multivariate Analysis 96*(2), 311–331.

Daubechies, I. (1992). *Ten lectures on wavelets*, Volume 61. Siam.

De Prins, D., L. Simar, and H. Tulkens (1984). Measuring labour efficiency. *Post Offices in The Performance of Public Enterprises: Concepts and Measurement (P. Pestieau ve H. Tulkens M. Marchand)*, 243–267.

Delyon, B. and A. Juditsky (1996). On minimax wavelet estimators. *Applied and Computational Harmonic Analysis 3*(3), 215–228.

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, 301–369.

Fan, Y., Q. Li, and A. Weersink (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business & Economic Statistics 14*(4), 460–468.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General) 120*(3), 253–281.

Gijbels, I., E. Mammen, B. U. Park, and L. Simar (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association 94*(445), 220–228.

Girard, S., A. Guillou, and G. Stupfler (2013). Frontier estimation with kernel regression on high order moments. *Journal of Multivariate Analysis 116*, 172–189.

Girard, S. and P. Jacob (2008). Frontier estimation via kernel regression on high power-transformed data. *Journal of Multivariate Analysis 99*(3), 403–420.

Hall, P. and J. Marron (1990). On variance estimation in nonparametric regression. *Biometrika 77*(2), 415–419.

Hall, P., B. A. Turlach, et al. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *The Annals of Statistics 25*(5), 1912–1925.

Härdle, W., G. Kerkyacharian, D. Picard, and A. Tsybakov (2012). *Wavelets, approximation, and statistical applications*, Volume 129. Springer Science & Business Media.

Härdle, W. and A. Tsybakov (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics 81*(1), 223–242.

Huang, P., Y. Pi, and I. Progri (2013). Gps signal detection under multiplicative and additive noise. *The Journal of Navigation 66*(4), 479–500.

Jirak, M., A. Meister, M. Reiß, et al. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *The Annals of Statistics 42*(5), 1970–2002.

Korostelev, A. P. and A. B. Tsybakov (2012). *Minimax theory of image reconstruction*, Volume 82. Springer Science & Business Media.

Kuan, D. T., A. A. Sawchuk, T. C. Strand, and P. Chavel (1985). Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2), 165–177.

Kulik, R., M. Raimondo, et al. (2009). Wavelet regression in random design with heteroscedastic dependent errors. *The Annals of Statistics 37*(6A), 3396–3430.

Kulik, R., C. Wichelhaus, et al. (2011). Nonparametric conditional variance and error density estimation in regression models with dependent errors and predictors. *Electronic Journal of Statistics 5*, 856–898.

Kumbhakar, S. C., B. U. Park, L. Simar, and E. G. Tsionas (2007). Nonparametric stochastic frontiers: a local maximum likelihood approach. *Journal of Econometrics 137*(1), 1–27.

Mallat, S. (2008). *A wavelet tour of signal processing: the sparse way.* Academic press.

Mateo, J. L. and A. Fernández-Caballero (2009). Finding out general tendencies in speckle noise reduction in ultrasound images. *Expert Systems with Applications 36*(4), 7786–7797.

Meyer, Y. (1992). *Wavelets and operators*, Volume 1. Cambridge university press.

Muller, H.-G., U. Stadtmuller, et al. (1987). Estimation of heteroscedasticity in regression analysis. *The Annals of Statistics 15*(2), 610–625.

Nason, G. P. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 463–479.

Navarro, F. and C. Chesneau (2019). *R package rwavelet: Wavelet Analysis.* (Version 0.4.0).

Navarro, F. and A. Saumard (2017). Slope heuristics and v-fold model selection in heteroscedastic regression using strongly localized bases. *ESAIM: Probability and Statistics 21*, 412–451.

Navarro, F. and A. Saumard (2018). Efficiency of the v -fold model selection for local-ized bases. In P. Bertail, D. Blanke, P.-A. Cornillon, and E. Matzner-Løber (Eds.), *Nonparametric Statistics*, Cham, pp. 53–68. Springer International Publishing.

Rabbani, H., M. Vafadust, P. Abolmaesumi, and S. Gazor (2008). Speckle noise reduc-tion of medical ultrasound images in complex wavelet domain using mixture priors. *IEEE Transactions on Biomedical Engineering 55*(9), 2152–2160.

Shen, Y., C. Gao, D. Witten, and F. Han (2019). Optimal estimation of variance in nonparametric regression with random design. *arXiv preprint arXiv:1902.10822*.

Simar, L. and V. Zelenyuk (2011). Stochastic fdh/dea estimators for frontier analysis. *Journal of Productivity Analysis 36*(1), 1–20.

Triebel, H. (1994). Theory of function spaces ii. *Bull. Amer. Math. Soc 31*, 119–125.

Tsybakov, A. B. (2009). Introduction to nonparametric estimation. revised and ex-tended from the 2004 french original. translated by vladimir zaiats.

Verzelen, N., E. Gassiat, et al. (2018). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli 24*(4B), 3683–3710.

Wang, L., L. D. Brown, T. T. Cai, M. Levine, et al. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics 36*(2), 646–664.