Electronic Journal of Statistics
Vol. 11 (2017) 682–724
ISSN: 1935-7524
DOI: 10.1214/17-EJS1238

Adaptive wavelet multivariate regression with errors in variables

Michaël Chichignoud*

ETH Zürich
e-mail: chichignoud@stat.math.ethz.ch

Van Ha Hoang

Laboratoire Paul Painlevé UMR CNRS 8524
Université des Sciences et Technologies de Lille 1
59655 Villeneuve d'Ascq Cedex, France
e-mail: van-ha.hoang@math.univ-lille1.fr

Thanh Mai Pham Ngoc†

Laboratoire de Mathématiques, UMR 8628, Université Paris Sud
91405 Orsay Cedex France
e-mail: thanh.pham_ngoc@math.u-psud.fr

and

Vincent Rivoirard†

CEREMADE, CNRS, UMR 7534, Université Paris Dauphine, PSL Research University
75016 Paris, France
e-mail: rivoirard@ceremade.dauphine.fr

Abstract: In the multidimensional setting, we consider the errors-in-variables model. We aim at estimating the unknown nonparametric multivariate regression function with errors in the covariates. We devise an adaptive estimators based on projection kernels on wavelets and a deconvolution operator. We propose an automatic and fully data driven procedure to select the wavelet level resolution. We obtain an oracle inequality and optimal rates of convergence over anisotropic Hölder classes. Our theoretical results are illustrated by some simulations.

MSC 2010 subject classifications: Primary 62G08.
Keywords and phrases: Adaptive wavelet estimator, anisotropic regression, deconvolution, measurement errors.

Received January 2016.

*Supported in part as member of the German-Swiss Research Group FOR916 (Statistical Regularization and Qualitative Constraints) with grant number 20PA20E-134495/1.
†Supported in part by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

Contents

1. Introduction

We consider the problem of multivariate nonparametric regression with errors in variables. We observe the i.i.d dataset

$$(W_1, Y_1), \dots, (W_n, Y_n)$$

where

$$Y_l = m(X_l) + \varepsilon_l$$

and

$$W_l = X_l + \delta_l,$$

with $Y_l \in \mathbb{R}$. The covariates errors δ_l are i.i.d unobservable random variables having error density g. We assume that g is known. The δ_l's are independent of the X_l's and Y_l's. The ε_l's are i.i.d standard normal random variables, independent of the X_l's with variance s^2 which is assumed to be known. We wish to estimate the regression function $m(x), x \in [0,1]^d$, but direct observations of the covariates X_l are not available. Instead due to the measuring mechanism or the nature of the environment, the covariates X_l are measured with errors. Let us denote f_X the density of the X_l's assumed to be positive and f_W the density of the W_l's.

Use of errors-in-variables models appears in many areas of science such as medicine, econometry or astrostatistics and is appropriate in a lot of practical experimental problems. For instance, in epidemiologic studies where risk factors are partially observed (see [23], [10]) or in environmental science where air quality is measured with errors ([6]).

In the error-free case, that is $\delta_l = 0$, one retrieves the classical multivariate nonparametric regression problem. Estimating a function in a nonparametric way from data measured with error is not an easy problem. Indeed, constructing a consistent estimator in this context is challenging as we have to face to a deconvolution step in the estimation procedure. Deconvolution problems arise in many fields where data are obtained with measurement errors and has attracted a lot of attention in the statistical literature, see [21] for an excellent source of references. The nonparametric regression with errors-in-variables model has been the object of a lot of attention as well, we may cite the works of [10], [11], [17], [19], [21], [5], [3], [8], [2], [6]. The literature has mainly to do with kernel-based approaches, based on the Fourier transform. All the works cited have tackled the univariate case except for [10] where the authors explored the asymptotic normality for mixing processes. In the one dimensional setting, [3] used Meyer wavelets in order to devise his statistical procedure but his assumptions on the model are strong since the corrupted observations W_l follow a uniform density on $[0, 1]$. [5] investigated the mean integrated squared error with a penalized estimator based on projection methods upon Shannon basis. But the authors do not give any clue about how to choose the resolution level of the Shannon basis. Furthermore, the constants in the penalized term are calibrated via intense simulations.

In the present article, our aim is to study the multidimensional setting and the pointwise risk. We would like to take into account the anisotropy for the function to estimate. Our approach relies on the use of projection kernels on wavelets bases combined with a deconvolution operator involving the noise in the covariates. When using wavelets, a crucial point lies in the choice of the resolution level. Actually, the main goal of the paper focuses on how to choose in a calibrated way the multiresolution analysis. It is well-known that theoretical results in adaptive estimation do not provide the way to choose the numerical constants in the resolution level and very often lead to conservative choices. We may cite the work of [12] which attempts to tackle this problem. For the density estimation problem and the sup-norm loss, the authors based their statistical procedure on Haar projection kernels and provide a way to choose locally the resolution level. Nonetheless, in practice, their procedure relies on heavy Monte Carlo simulations to calibrate the constants. In our paper the resolution level of our estimator is optimal, partially data-driven and varies x by x. It is automatically selected by a method inspired from [14] to tackle anisotropy problems. This method has been used recently in various contexts (see [7], [4] and [1]). Furthermore, we do not resort to thresholding which is very popular when using wavelets and our selection rule is adaptive to the unknown regularity of the regression function. We obtain oracle inequalities and provide optimal rates of convergence for anisotropic Hölder classes. The performances of our adaptive estimator, the negative impact of the errors in the covariates, the effects of the design density are assessed by examples based on simulations.

The paper is organized as follows. In Section 2, we describe our estimation procedure. In Section 3, we provide an oracle inequality and rates of convergences of our estimator for the pointwise risk. Section 4 gives some numerical illustrations. Proofs of theorems, propositions and technical lemmas are to be found in Section 5.

Notation Let $\mathbb{N} = \{0, 1, 2, \dots\}$ and $j = (j_1, \dots, j_d) \in \mathbb{N}^d$, we set $S_j = \sum_{i=1}^d j_i$ and for any $y \in \mathbb{R}^d$, we set, with a slight abuse of notation,

$$2^j y := (2^{j_1} y_1, \dots, 2^{j_d} y_d)$$

and for any $k = (k_1, \cdots, k_d) \in \mathbb{Z}^d$,

$$h_{j,k}(y) := 2^{\frac{S_j}{2}} h(2^j y - k) = 2^{\frac{S_j}{2}} h(2^{j_1} y_1 - k_1, \dots, 2^{j_d} y_d - k_d),$$

for any function $h : \mathbb{R}^d \to \mathbb{R}$. We denote by $\mathcal{F}$ the Fourier transform of any Lebesgue integrable function $f \in \mathbb{L}_1(\mathbb{R}^d)$ by

$$\mathcal{F}(f)(t) = \int_{\mathbb{R}^d} e^{-i<t,y>} f(y) dy, \quad t \in \mathbb{R}^d,$$

where $< ., . >$ denotes the usual scalar product.

For two integers a, b, we denote $a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$. And $\lfloor y \rfloor$ denotes the largest integer smaller than y: $\lfloor y \rfloor \le y < \lfloor y \rfloor + 1$.

2. The estimation procedure

For estimating the regression function m, the idea consists in writing m as the ratio

$$m(x) = \frac{m(x) f_X(x)}{f_X(x)}, \quad x \in [0, 1]^d.$$

In the sequel, we denote

$$p(x) := m(x) \times f_X(x).$$

First, we estimate p, then f_X. Since estimating f_X is a classical deconvolution problem, the main task consists in estimating p. We propose a wavelet-based procedure with an automatic choice of the maximal resolution level. Section 2.2 describes the construction of the projection kernel on wavelet bases depending on a maximal resolution level. Section 2.3 describes the Goldenshluger-Lepski procedure to select the resolution level adaptively.

2.1. Technical conditions

To facilitate the presentation, we collect in this subsection all the conditions that we need throughout the paper.

First, some conditions are imposed on the regression function m and the design density f_X. We suppose that

$$m \in \mathcal{M}(\mathfrak{m}) = \{S : [0,1]^d \to \mathbb{R} : \|S\|_\infty \le \mathfrak{m}\},\ \mathfrak{m} > 0, \tag{1}$$

and

$$f_X \in \mathcal{M}(\mathfrak{d}) = \{f \text{ density on } [0,1]^d \text{ and } \|f\|_\infty \le \mathfrak{d}\},\ \mathfrak{d} > 0. \tag{2}$$

Futhermore, there exists $C_1 > 0$ such that for any $x \in [0,1]^d$, $f_X(x) \ge C_1$. To ensure the existence of all Fourier transforms, we also suppose that $m \cdot f_X$ and $\mathcal{F}(m \cdot f_X) \in \mathbb{L}_1(\mathbb{R}^d)$.

To derive rates of convergence and lower bounds as we have to face a deconvolution step, we need some assumptions on the smoothness of the density of the errors covariates g. We suppose that

$$\mathcal{F}(g)(t) = \prod_{l=1}^{d} \mathcal{F}(g_l)(t_l),$$

and there exist positive constants c_g and C_g such that

$$c_g(1+|t_l|)^{-\nu} \le |\mathcal{F}(g_l)(t_l)| \le C_g(1+|t_l|)^{-\nu}, \quad \nu \ge 0, \quad t_l \in \mathbb{R}. \tag{3}$$

The left hand side of the above inequality is usual when proving upper bounds. But here as we use compactly supported wavelets, we also need the right hand side to prove upper bounds. This supplementary assumption has been already used in deconvolution density estimation problem (see [9]). The right hand side of inequality (3) also appears in the proofs of lower bounds.

We require another condition on the derivative of the Fourier transform of g to prove lower bounds. There exists a positive constant $\mathcal{C}_g$ such that

$$|\mathcal{F}'(g_l)(t_l)| \le \mathcal{C}_g(1+|t_l|)^{-\nu-1}, \quad t_l \in \mathbb{R}. \tag{4}$$

Laplace and Gamma distributions satisfy the above Assumptions (3) and (4). Assumptions (3) and (4) control the decay of the Fourier transform of each components of g at a polynomial rate controlled by the degree of ill-posedness ν. Hence we deal with a midly ill-posed inverse problem.

We consider a father wavelet φ on the real line satisfying the following conditions:

- (A1) The father wavelet φ is compactly supported on $[-A, A]$, where A is a positive integer.
- (A2) There exists a positive integer N, such that for any x

$$\int \sum_{k\in\mathbb{Z}} \varphi(x-k)\varphi(y-k)(y-x)^\ell dy = \delta_{0\ell}, \quad \ell = 0, \dots, N.$$

- (A3) φ is of class $\mathcal{C}^r$, where $r > \nu + 1$.

Conditions (A1), (A2) and (A3) are satisfied for instance by Coiflets wavelets (see [15], Chapter 8). Condition (A3) has already been encountered in the literature (see condition (A2) in [9]). It ensures that our estimator is well-defined (more explanations about this are given in Section 2.2). Condition (A3) is also useful to prove Lemma 9.

Remark 1. *Note that most of our results remain valid by using wavelets with compactly supported Fourier transform such as Meyer wavelets. However, in this case, the summation in* (5) *is not finite, which leads to some difficulties in practice. [9] also used compactly supported wavelets such as Daubechies ones when dealing with deconvolution density problem.*

2.2. Approximation kernels and family of estimators for p

The associated projection kernel on the space

$$V_j := \operatorname{span}\{\varphi_{jk}, k \in \mathbb{Z}^d\}, \quad j \in \mathbb{N}^d,$$

is given for any x and y by

$$K_j(x,y) = \sum_k \varphi_{jk}(x)\varphi_{jk}(y),$$

where for any x,

$$\varphi_{jk}(x) = \prod_{l=1}^{d} 2^{\frac{j_l}{2}} \varphi(2^{j_l}x_l - k_l), \quad j \in \mathbb{N}^d,\ k \in \mathbb{Z}^d.$$

Therefore, the projection of p on V_j can be written for any z,

$$p_j(z) := K_j(p)(z) := \int K_j(z,y)p(y)dy = \sum_k p_{jk}\varphi_{jk}(z)$$

with

$$p_{jk} = \int p(y)\varphi_{jk}(y)dy.$$

First we estimate unbiasedly projection p_j. Secondly to obtain the final estimate of p, it will remain to select a convenient value of j which will be done in Section 2.3. The natural approach is based on unbiased estimation of the projection coefficients p_{jk}. To do so, we adapt the kernel approach proposed by [11] in our wavelets context. To this purpose, we set

$$\begin{aligned}\hat{p}_{jk} :=& \frac{1}{n}\sum_{u=1}^{n} Y_u \times (\mathcal{D}_j\varphi)_{j,k}(W_u)\\ =& \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\frac{1}{n}\sum_{u=1}^{n} Y_u \int e^{-i<t,2^jW_u-k>} \prod_{l=1}^{d} \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)} dt_l,\end{aligned}$$

then

$$\hat{p}_j(x) = \sum_k \hat{p}_{jk}\varphi_{jk}(x) = \frac{1}{n}\sum_k \sum_{u=1}^{n} Y_u \times (\mathcal{D}_j\varphi)_{j,k}(W_u)\varphi_{jk}(x), \tag{5}$$

where the deconvolution operator $\mathcal{D}_j$ is defined as follows for a function f defined on $\mathbb{R}$

$$(\mathcal{D}_j f)(w) = \frac{1}{(2\pi)^d}\int e^{-i<t,w>}\prod_{l=1}^{d}\frac{\overline{\mathcal{F}(f)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}dt, w \in \mathbb{R}^d. \tag{6}$$

Lemma 3, proved in Section 5.2.1 states that $\mathbb{E}[\hat{p}_j(x)] = p_j(x)$ which justifies our approach. Note that as φ has compact support, the summation in k is finite for all x (see the expression of estimator $\hat{p}_j(x)$ in (5)).

The deconvolution operator $(\mathcal{D}_j f)(w)$ in (6) is the multidimensional wavelet analogous of the operator $K_n(x)$ defined in (2.4) in [11]: the Fourier transform of their kernel K has been replaced in our procedure by the Fourier transform of the wavelet φ_{jk} and their bandwith h by 2^{-j}. Eventually, our estimator is well-defined: using Lemma 8 and Assumption (3) we have that, for C a constant,

$$\prod_{l=1}^{d}\left|\frac{\mathcal{F}(\varphi)(t_l)}{\mathcal{F}(g_l)(2^{j_l}t_l)}\right| \le C\prod_{l=1}^{d}(1+|t_l|)^{-r}(1+|2^{j_l}t_l|)^{\nu} \le C2^{S_j\nu}\prod_{l=1}^{d}(1+|t_l|)^{\nu-r},$$

which is integrable using condition (A3).

The definition of the estimator $\hat{p}_j(x)$ still makes sense when we do not have any noise on the variables X_l i.e $g(x) = \delta_0(x)$ because in this case $\mathcal{F}(g)(t) = 1$.

2.3. Selection rule by using the Goldenshluger-Lepski methodology

The second and final step consists in selecting the multidimensional resolution level j depending on x thanks to a data-driven selection rule. This selection rule is a modification in the light of [14] of a method exposed in [18]. First we have to introduce some quantities which will intervene in the rule. In the sequel we denote for any $w \in \mathbb{R}^d$,

$$T_j(w) := \sum_k (\mathcal{D}_j\varphi)_{j,k}(w)\varphi_{jk}(x)$$

and

$$U_j(y,w) := y\sum_k (\mathcal{D}_j\varphi)_{j,k}(w)\varphi_{jk}(x) = y \times T_j(w),$$

so we have

$$\hat{p}_j(x) = \frac{1}{n}\sum_{u=1}^{n} U_j(Y_u, W_u).$$

Proposition 1 in Section 5.2.1 shows that $\hat{p}_j(x)$ concentrates around $p_j(x)$. So, the idea is to find a maximal resolution $\hat{j}$ that mimics the oracle index. The oracle index minimizes a bias variance trade-off. So we have to find an estimation for the bias-variance decomposition of $\hat{p}_j(x)$. We denote $\sigma_j^2 := \mathrm{Var}(U_j(Y_1, W_1))$ and the variance of $\hat{p}_j$ is thus equal to $\frac{\sigma_j^2}{n}$. We set:

$$\hat{\sigma}_j^2 := \frac{1}{n(n-1)} \sum_{l=2}^{n} \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - U_j(Y_v, W_v))^2, \tag{7}$$

and since $\mathbb{E}(\hat{\sigma}_j^2) = \sigma_j^2$, $\hat{\sigma}_j^2$ is a natural estimator of σ_j^2. To devise our procedure, we introduce a slightly overestimate of σ_j^2 given by:

$$\tilde{\sigma}_{j,\tilde{\gamma}}^2 := \hat{\sigma}_j^2 + 2C_j \sqrt{2\tilde{\gamma}\hat{\sigma}_j^2 \frac{\log n}{n}} + 8\tilde{\gamma} C_j^2 \frac{\log n}{n}, \tag{8}$$

where $\tilde{\gamma}$ is a positive constant and

$$C_j := \left(\mathfrak{m} + s\sqrt{2\tilde{\gamma}\log n}\right) \|T_j\|_\infty.$$

Let $\gamma > 0$ and

$$\Gamma_\gamma(j) := \sqrt{\frac{2\gamma \tilde{\sigma}_{j,\tilde{\gamma}}^2 \log n}{n}} + \frac{c_j \gamma \log n}{n},$$

where

$$c_j := 16\,(2\mathfrak{m} + s)\,\|T_j\|_\infty.$$

Let

$$\Gamma_\gamma(j, j') := \Gamma_\gamma(j) + \Gamma_\gamma(j \wedge j'),$$

and

$$\Gamma_\gamma^*(j) := \sup_{j' \in J} \Gamma_\gamma(j, j'). \tag{9}$$

We now define the selection rule for the resolution index. Let

$$\hat{R}_j := \sup_{j' \in J} \left\{ |\hat{p}_{j \wedge j'}(x) - \hat{p}_{j'}(x)| - \Gamma_\gamma(j', j) \right\}_+ + \Gamma_\gamma^*(j). \tag{10}$$

Then $\hat{p}_{\hat{j}}(x)$ is the final estimator of $p(x)$ with $\hat{j}$ such that

$$\hat{j} := \arg\min_{j \in J} \hat{R}_j, \tag{11}$$

where the set J is defined as

$$J := \left\{ j \in \mathbb{N}^d : \quad 2^{S_j} \leq \left\lfloor \frac{n}{\log^2 n} \right\rfloor \right\}. \tag{12}$$

Now, we shall highlight how the above quantities interplay in the estimation of the risk decomposition of $\hat{p}_j$. An inspection of the proof of Theorem 1 shows that a control of the bias of $\hat{p}_j$ is provided by:

$$\sup_{j'} \left\{ |\hat{p}_{j \wedge j'}(x) - \hat{p}_{j'}(x)| - \Gamma_\gamma(j', j) \right\}_+.$$

The term $|\hat{p}_{j\wedge j'}(x)-\hat{p}_{j'}|$ is classical when using the Goldenshluger Lepski method (see Sections 2.1 and 5.2 in [1]). Furthermore for technical reasons (see proof of Theorem 1), we do not estimate the variance of $\hat{p}_j(x)$ by $\frac{\hat{\sigma}_j^2}{n}$ but we replace it by $\Gamma_\gamma^2(j)$. Note that we have the straightforward control

$$\Gamma_\gamma(j) \leq C\left(\hat{\sigma}_j\sqrt{\frac{\log n}{n}} + (C_j + c_j)\frac{\log n}{n}\right),$$

where C is a constant depending on ε, $\tilde{\gamma}$ and γ. Actually we prove that $\Gamma_\gamma^2(j)$ is of order $\frac{\log n}{n}\sigma_j^2$ (see Lemma 6 and 10). The dependence of $\tilde{\sigma}^2_{j,\tilde{\gamma}}$ (8) in $\mathfrak{m}$ appears only in smaller order terms. In conclusion, up to the knowledge of $\mathfrak{m}$ and s^2 the procedure is completely data-driven. Next section explains how to choose the constants γ and $\tilde{\gamma}$. Our approach is non asymptotic and based on sharp concentration inequalities.

3. Rates of convergence

3.1. Oracle inequality and rates of convergence for $p(\cdot)$

First, we state an oracle inequality which highlights the bias-variance decomposition of the risk.

Theorem 1. *Let $q \geq 1$ be fixed and let $\hat{j}$ be the adaptive index defined as above. Then, it holds for any $\gamma > q(\nu+1)$ and $\tilde{\gamma} > 2q(\nu+2)$,*

$$\mathbb{E}\left[\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^q\right] \leq R_1\left(\inf_\eta \mathbb{E}\left[\left\{B(\eta) + \Gamma^*_\gamma(\eta)\right\}^q\right]\right) + R'_1 n^{-q},$$

where

$$B(\eta) := \max\left(\sup_{j'}\left|\mathbb{E}\left[\hat{p}_{\eta\wedge j'}(x)\right] - \mathbb{E}\left[\hat{p}_{j'}(x)\right]\right|, \left|\mathbb{E}[\hat{p}_\eta(x)] - p(x)\right|\right)$$

R_1 a constant depending only on q and R'_1 a constant depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ, c_g, $\mathcal{C}_g$.

The oracle inequality in Theorem 1 illustrates a bias-variance decomposition of the risk. The term $B(\eta)$ is a bias term. Indeed, one recognizes on the right side the classical bias term

$$|\mathbb{E}[\hat{p}_\eta(x)] - p(x)| = |p_\eta(x) - p(x)|.$$

Concerning $|\mathbb{E}\left[\hat{p}_{\eta\wedge j'}(x)\right] - \mathbb{E}\left[\hat{p}_{j'}(x)\right]|$, for sake of clarity let us consider for instance the univariate case: if $j' \leq \eta$ this term is equal to zero. If $j' \geq \eta$, it turns to be

$$\left|\mathbb{E}\left[\hat{p}_\eta(x)\right] - \mathbb{E}\left[\hat{p}_{j'}(x)\right]\right| = |p_\eta(x) - p_{j'}(x)| \leq |p_\eta(x) - p(x)| + |p_{j'}(x) - p(x)|.$$

As we have the following inclusion for the projection spaces $V_\eta \subset V_{j'}$, the term $p_{j'}$ is closer to p than p_η for the L_2-distance. Hence we expect a good control of $|p_{j'}(x) - p(x)|$ with respect to $|p_\eta(x) - p(x)|$. Finally, the third term is a remain term and is negligible.

We study the rates of convergence of the estimators over anisotropic Hölder classes which are adapted to local estimation. Let us define them.

Definition 1 (Anisotropic Hölder Space)**.** *Let $\vec{\beta} = (\beta_1, \beta_2, \dots, \beta_d) \in (\mathbb{R}_+^*)^d$ and $L > 0$. We say that $f : [0,1]^d \to \mathbb{R}$ belongs to the anisotropic Hölder class $\mathbb{H}_d(\vec{\beta}, L)$ of functions if f is bounded and for any $l = 1, ..., d$ and for all $z \in \mathbb{R}$*

$$\sup_{x\in[0,1]^d} \left| \frac{\partial^{\lfloor \beta_l \rfloor} f}{\partial x_l^{\lfloor \beta_l \rfloor}}(x_1, \dots, x_l + z, \dots, x_d) - \frac{\partial^{\lfloor \beta_l \rfloor} f}{\partial x_l^{\lfloor \beta_l \rfloor}}(x_1, \dots, x_l, \dots, x_d) \right| \le L|z|^{\beta_l - \lfloor \beta_l \rfloor}.$$

The following theorem gives the rate of convergence of the estimator $\hat{p}_{\hat{j}}(x)$ for the pointwise $\mathbb{L}_q$ risk with $q \ge 1$. Of course, one gets the usual pointwise $\mathbb{L}_2$ risk for $q = 2$.

Theorem 2. *Let $q \ge 1$ be fixed and let $\hat{j}$ be the adaptive index defined in (11). Then, if for any l, $\lfloor \beta_l \rfloor \le N$ and $L > 0$, it holds*

$$\sup_{p\in\mathbb{H}_d(\vec{\beta},L)} \mathbb{E} \left| \hat{p}_{\hat{j}}(x) - p(x) \right|^q \le L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_2 \left(\frac{\log n}{n} \right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with $\bar{\beta} = \frac{1}{\frac{1}{\beta_1} + \cdots + \frac{1}{\beta_d}}$ and R_2 a constant depending on $\gamma, q, \tilde{\gamma}, \mathfrak{m}, \mathfrak{d}, s, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$.

Remark 2. *The estimate $\hat{p}_{\hat{j}}(x)$ achieves the optimal rate of convergence up to a logarithmic term (see Section 3.3 in [4]). This logarithmic loss is due to adaptation.*

The next section presents convergence rates for the estimator $\hat{m}(x)$ of the regression function m.

3.2. Rates of convergence for $m(\cdot)$

As mentioned above, the estimation of m requires an adaptive estimate of f_X. This is due to kernel estimators, e.g. projection estimators do not need the additional estimate (see [1]). For this purpose, we use an estimate introduced by [4] (Section 3.4) denoted by $\hat{f}_X$. This estimate is constructed from a deconvolution kernel and the bandwidth is selected via a method described in [14]. We will not give the explicit expression of $\hat{f}_X$ for ease of exposition. Then, we define the estimate of m for all x in $[0,1]^d$:

$$\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}}. \tag{13}$$

The term $n^{-1/2}$ is added to avoid the drawback when $\hat{f}_X$ is closed to 0.

Theorem 3. *Let $q \geq 1$ be fixed and let $\hat{m}$ defined as above. Then, if for any l, $\lfloor \beta_l \rfloor \leq N$ and $L > 0$, it holds*

$$\sup_{(m,f_X)\in\mathbb{H}_d(\vec{\beta},L)\times\mathbb{H}_d(\vec{\beta},L)} \mathbb{E}\,|\hat{m}(x) - m(x)|^q \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_3 \left(\frac{\log n}{n}\right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with R_3 a constant depending on $\gamma, q, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$.

The following theorem gives a lower bound for the pointwise risk:

Theorem 4. *Let $q \geq 1$, $L > 0$ and for any l, $\lfloor \beta_l \rfloor \leq N$. Then for any estimator $\tilde{m}$ of m and for n large enough we have*

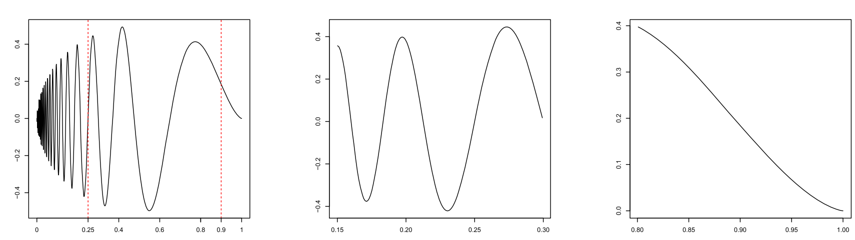
$$\sup_{(m,f_X)\in\mathbb{H}_d(\vec{\beta},L)\times\mathbb{H}_d(\vec{\beta},L)} \mathbb{E}\,|\tilde{m}(x) - m(x)|^q \geq R_4 n^{-q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with R_4 a positive constant depending on $\vec{\beta}$, L, s, C_g and $\mathcal{C}_g$.

Consequently, the estimate $\hat{m}$ achieves the optimal rate of convergence up to a logarithmic term and oracle inequality derived in Theorem 1 is then optimal.

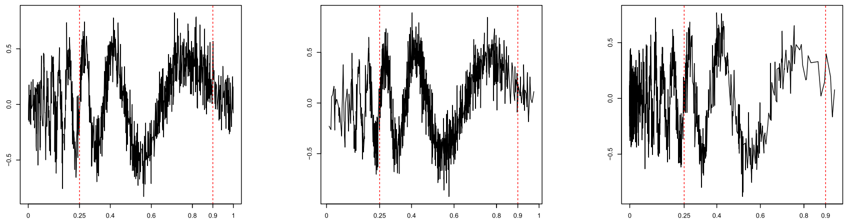
4. Numerical results

In this section, we implement some simulations to illustrate the theoretical results. We aim at estimating the Doppler regression function m at two points $x_0 = 0.25$ and $x_0 = 0.90$ (see Figure 1). We have $n = 1024$ observations and the regression errors ε_l's follow a standard normal density with variance $s^2 = 0.15^2$. As for the design density of the X_l's, we consider the Beta density and the uniform density on $[0,1]$. The uniform distribution is quite classical in regression with random design. The $Beta(2,2)$ and $Beta(0.5,2)$ distributions reflect two very different behaviors on $[0,1]$. Indeed, we recall that the Beta density with parameters (a,b) (denoted here by $Beta(a,b)$) is proportional to $x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$. Moreover, despite the fact that Beta densities vanish in



(a) (b) (c)

FIG 1. *a) Representation of Doppler function. b) A zoom of Doppler function on* $[0.15, 0.30]$. *c) A zoom of Doppler function on* $[0.80, 1]$.



(a) (b) (c)

Fig 2. *a/ Noisy Doppler with* $X_i \sim \mathcal{U}[0,1]$*. b/ Noisy Doppler with* $X_i \sim Beta(2,2)$*. c/ Noisy Doppler function with* $X_i \sim Beta(0.5,2)$*.*

0 and 1 and the design density f_X is assumed to be bounded from below, the choice of Beta distributions is still reasonable for simulations on any compact strictly included into $[0,1]$. Our numerical study illustrates the deteriorated performances of the estimator at points very closed to 0 and 1. This is justified in Table 3.

In Figure 2, we plot the noisy regression Doppler function according to the three design scenario. For the covariate errors δ_i's, we focus on the centered Laplace density with scale parameter $\sigma_{g_L} > 0$ that we denote g_L. This latter has the following expression:

$$g_L(x) = \frac{1}{2\sigma_{g_L}} e^{-\frac{|x|}{\sigma_{g_L}}}.$$

The choice of the centered Laplace noise is motivated by the fact that the Fourier transform of g_L is given by

$$\mathcal{F}(g_L)(t) = \frac{1}{1+\sigma_{g_L}^2 t^2},$$

and according to Assumption (3), it gives an example of an ordinary smooth noise with degree of ill-posedness $\nu = 2$. Furthermore, when facing regression problems with errors in the design, it is common to compute the so-called reliability ratio (see [11]) which is given by

$$R_r := \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + 2\sigma_{g_L}^2}.$$

R_r permits to assess the amount of noise in the covariates. The closer to 0 R_r is, the bigger the amount of noise in the covariates is and the more difficult the deconvolution step will be. For instance, [11] chose $R_r = 0.70$. We computed the reliability ratio in Table 1 for the considered simulations.

We recall that our estimator of $m(x)$ is given by the ratio of two estimators (see (13)):

$$\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}}. \tag{14}$$

TABLE 1
Reliability ratio.

σ_{g_L}	Design of the X_i		
	$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$
0.075	0.88	0.81	0.80
0.10	0.80	0.71	0.69

First, we compute $\hat{p}_{\hat{j}}(x)$ an estimator of $p(x) = m(x) \times f_X(x)$ which is denoted "GL" in the graphics below. We use coiflet wavelets of order 5. Then we divide $\hat{p}_{\hat{j}}(x)$ by the adaptive deconvolution density estimator $\hat{f}_X(x)$ of [4]. This latter is constructed with a deconvolution kernel and an adaptive bandwidth. For the selection of the coiflet level $\hat{j}$ in $\hat{p}_{\hat{j}}(x)$, we advise to use $\hat{\sigma}_j^2$ instead of $\tilde{\sigma}_{j,\tilde{\gamma}}^2$ and $\frac{2 \max_i |Y_i| \|T_j\|_\infty}{3}$ instead of c_j. It remains to settle the value of the constant γ. To do so, we compute the pointwise risk of $\hat{p}_{\hat{j}}(x)$ in function of γ: Figure 3 shows a clear "dimension jump" and accordingly the value $\gamma = 0.5$ turns to be reasonable. Hence we fix $\gamma = 0.5$ for all simulations and our selection rule is completely data-driven.

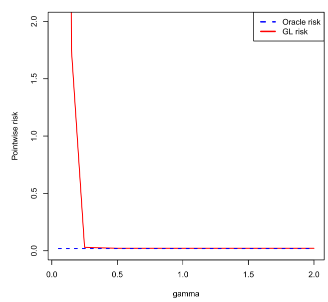


FIG 3. *Pointwise risk of $\hat{p}_{\hat{j}}$ at $x_0 = 0.25$ in function of parameter γ for the $Beta(2,2)$ design and $\sigma_{g_L} = 0.075$.*

Boxplots in Figure 4 and 5 summarize our numerical experiments. Theorem 1 gives an oracle inequality for the estimation of $p(x)$. We compare the pointwise risk error of $\hat{p}_{\hat{j}}(x)$ (computed with 100 Monte Carlo repetitions) with the oracle risk one. The oracle is $\hat{p}_{j_{oracle}}$ with the index j_{oracle} defined as follows:

$$j_{oracle} := \arg\min_{j \in J} |\hat{p}_j(x) - p(x)|.$$

In Table 2, we have computed the MAE (Mean Absolute Error) of $\hat{m}(x)$ over 100 Monte Carlo runs.

Our performances are close to those of the oracle (see Figure 4 and 5) and are quite satisfying both at $x_0 = 0.25$ and $x_0 = 0.90$. When going deeper into details, increasing the Laplace noise parameter σ_{g_L} deteriorates sligthly the performances. Hence it seems that our procedure is robust to the noise in the covariates and accordingly to the deconvolution step. Concerning the role of the design density, when considering the $Beta(0.5,2)$ distribution, we expect

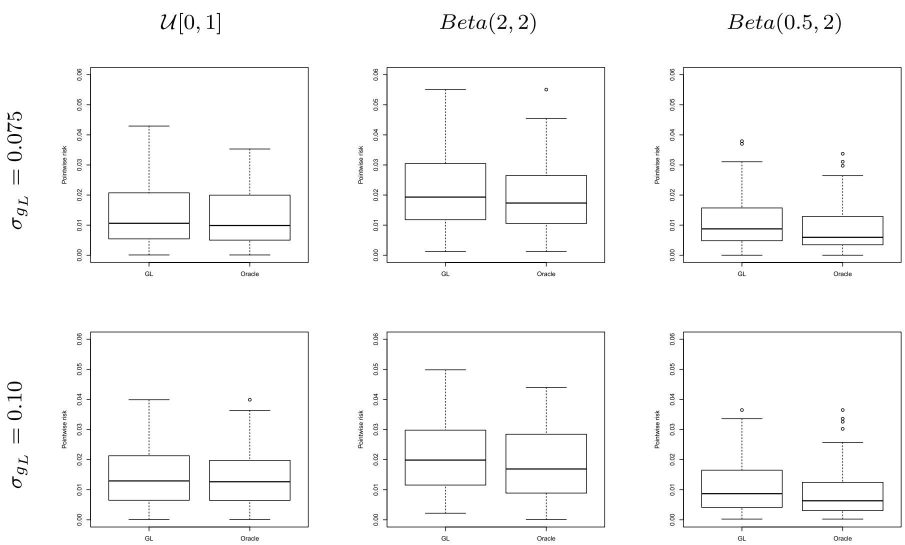


FIG 4. *Estimation of $p(x)$ at $x_0 = 0.25$*

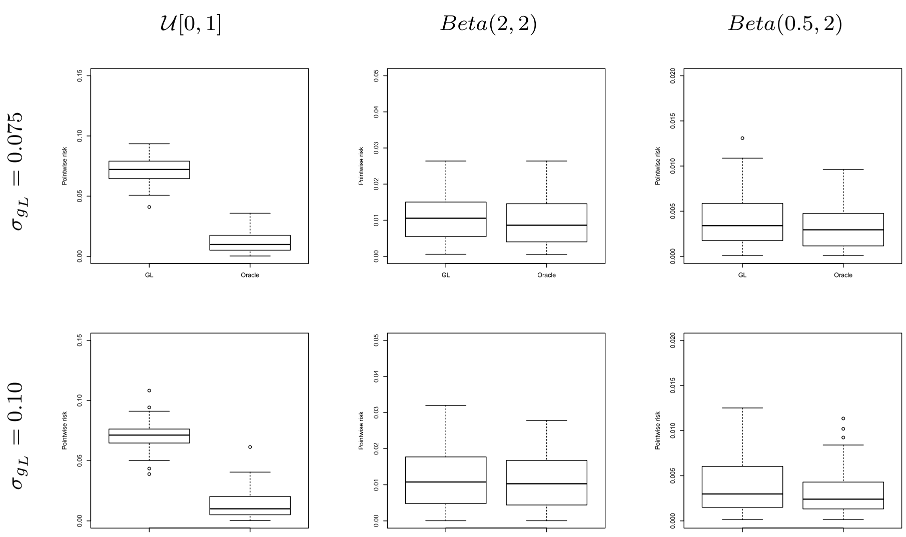


FIG 5. *Estimation of $p(x)$ at $x_0 = 0.90$*

the performances to be better near 0 as the observations tend to concentrate near 0 and to be bad close to 1. Indeed, this phenomenon is confirmed by Table 3. And when comparing the $Beta(2,2)$ and $Beta(0.5,2)$ distributions, the performances are much better for the $Beta(0.5,2)$ at $x_0 = 0.25$ whereas the $Beta(2,2)$ distribution yields better results at $x_0 = 0.90$. This is what is

Table 2
MAE of $\hat{m}(x)$: on the left at $x_0 = 0.25$ and on the right $x_0 = 0.90$.

σ_{g_L}	design of the X_i			σ_{g_L}	design of the X_i		
	$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$		$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$
0.075	0.0144	0.0204	0.0071	0.075	0.0212	0.0177	0.1012
0.10	0.0156	0.0206	0.0072	0.10	0.0192	0.0195	0.104

Table 3
MAE of $\hat{m}(x)$ at the points very closed to 0 and 1: on the left: $x_0 = 0.01$ and on the right: $x_0 = 0.98$.

σ_{g_L}	design of the X_i			σ_{g_L}	design of the X_i		
	$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$		$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$
0.075	0.3461	0.5312	0.3445	0.075	0.2153	0.3429	0.5130
0.10	0.3668	0.5493	0.3589	0.10	0.2191	0.3453	0.5293

expected as the two densities charge points near 0 and 1 differently.

For our simulations, we have chosen coiflets of order $K = 5$. The Fourier transform for the coiflet φ is given by:

$$\mathcal{F}(\varphi)(t) = \int_{\mathbb{R}} \exp(-itx)\varphi(x)dx, \quad t \in \mathbb{R}.$$

In theory, the Fourier transform is defined for all $t \in \mathbb{R}$. But in practice, it is sufficient to select $t \in [-L, L]$ since $\mathcal{F}(\varphi)(t)$ almost vanishes for all t outside the interval $[-L, L]$ where L is chosen to be large enough. Preliminary simulations allowed us to select $L = 50$ and we partitioned the interval $[-L, L]$ into $M = 4096$ points $t_k = -L + k\triangle t$, $k = 0, \dots, M-1$ with $\triangle t = 2L/(M-1)$. Then we approximated $\mathcal{F}(\varphi)(t_k)$ by a Riemann sum:

$$\mathcal{F}(\varphi)(t_k) \approx \sum_{j=0}^{N-1} \exp(-it_k x_j)\varphi(x_j)\triangle x.$$

Since the support of the coiflet φ is $[-2K, 4K-1]$, we approximated $\mathcal{F}(\varphi)(t_k)$ on the grid $[x_0, x_1, \dots, x_{N-1}]$ where $x_j = -2K + j\triangle x$, $j = 0, \dots, N-1$ and $\triangle x = (6K-1)/(N-1)$, $N = 2048$.

In a similar way, we approximated the integral

$$\big(\mathcal{D}_j\varphi\big)(w) = \int_{\mathbb{R}} \exp(-iwt)\frac{\overline{\mathcal{F}(\varphi)(t)}}{\mathcal{F}(g_L)(2^j t)}dt, \quad w \in \mathbb{R},$$

by

$$\sum_{k=0}^{M-1} \exp(-it_k w)\big(1 + \sigma_{g_L}^2 (2^j t_k)^2\big)\overline{\mathcal{F}(\varphi)(t_k)}\triangle t,$$

since $\mathcal{F}(g_L)(t) = 1/(1 + \sigma_{g_L}^2 t^2)$ by the choice of the centered Laplace noise.

5. Proofs

5.1. Proofs of theorems

This section is devoted to the proofs of theorems. These proofs use some propositions and technical lemmas which are respectively in Section 5.2.1 and 5.2.2. In the sequel, C is a constant which may vary from one line to another one.

5.1.1. Proof of Theorem 1

Proof. We firstly recall the basic inequality $(a_1+\cdots+a_p)^q \leq p^{q-1}(a_1^q+\cdots+a_p^q)$ for all $a_1,\ldots,a_p \in \mathbb{R}_+^p$, $p \in \mathbb{N}$ and $q \geq 1$. For ease of exposition, we denote $\hat{p}_{\hat{j}}(x) = \hat{p}_{\hat{j}}$. So, we can show for any $\eta \in \mathbb{N}^d$:

$$\begin{aligned}
\left|\hat{p}_{\hat{j}} - p(x)\right| &\leq \left|\hat{p}_{\hat{j}} - \hat{p}_{\hat{j}\wedge\eta}\right| + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| + |\hat{p}_\eta - p(x)| \\
&\leq \left|\hat{p}_{\eta\wedge\hat{j}} - \hat{p}_{\hat{j}}\right| - \Gamma_\gamma(\hat{j},\eta) + \Gamma_\gamma(\hat{j},\eta) + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| \\
&\qquad - \Gamma_\gamma(\eta,\hat{j}) + \Gamma_\gamma(\eta,\hat{j}) + |\hat{p}_\eta - p(x)| \\
&\leq \left|\hat{p}_{\eta\wedge\hat{j}} - \hat{p}_{\hat{j}}\right| - \Gamma_\gamma(\hat{j},\eta) + \Gamma_\gamma(\eta,\hat{j}) + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| \\
&\qquad - \Gamma_\gamma(\eta,\hat{j}) + \Gamma_\gamma(\hat{j},\eta) + |\hat{p}_\eta - p(x)| \\
&\leq \left|\hat{p}_{\eta\wedge\hat{j}} - \hat{p}_{\hat{j}}\right| - \Gamma_\gamma(\hat{j},\eta) + \Gamma^*_\gamma(\eta) + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| \\
&\qquad - \Gamma_\gamma(\eta,\hat{j}) + \Gamma^*_\gamma(\hat{j}) + |\hat{p}_\eta - p(x)| \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\hat{p}_\eta - p(x)| \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\mathbb{E}[\hat{p}_\eta] - p(x)| + |\hat{p}_\eta - \mathbb{E}[\hat{p}_\eta]| \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\mathbb{E}[\hat{p}_\eta] - p(x)| + |\hat{p}_\eta - \mathbb{E}[\hat{p}_\eta]| - \Gamma_\gamma(\eta) + \Gamma_\gamma(\eta) \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\mathbb{E}[\hat{p}_\eta] - p(x)| + \sup_{j'}\Big\{ |\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]| - \Gamma_\gamma(j')\Big\}_+ + \Gamma^*_\gamma(\eta).
\end{aligned}$$

By definition of $\hat{j}$, we recall that $\hat{R}_{\hat{j}} \leq \inf_\eta \hat{R}_\eta$ and

$$\begin{aligned}
\hat{R}_\eta \leq& \sup_{j,j'}\Big\{ |\hat{p}_{j\wedge j'} - \mathbb{E}[\hat{p}_{j\wedge j'}]| - \Gamma_\gamma(j\wedge j')\Big\}_+ \\
&+ \sup_{j'}\Big\{ |\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]| - \Gamma_\gamma(j')\Big\}_+ + \sup_{j'} |\mathbb{E}[\hat{p}_{\eta\wedge j'}] - \mathbb{E}[\hat{p}_{j'}]| + \Gamma^*_\gamma(\eta).
\end{aligned}$$

Hence

$$\begin{aligned}
\left|\hat{p}_{\hat{j}} - p(x)\right| \leq 2\Bigg[&\sup_{j,j'}\Big\{ |\hat{p}_{j\wedge j'} - \mathbb{E}[\hat{p}_{j\wedge j'}]| - \Gamma_\gamma(j\wedge j')\Big\}_+ \\
&+ \sup_{j'}\Big\{ |\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]| - \Gamma_\gamma(j')\Big\}_+ + \sup_{j'} |\mathbb{E}[\hat{p}_{\eta\wedge j'}] - \mathbb{E}[\hat{p}_{j'}]|\Bigg]
\end{aligned}$$

$$+ 2\,\Gamma^*_\gamma(\eta) + |\mathbb{E}[\hat p_\eta] - p(x)| + \sup_{j'}\Big\{ |\hat p_{j'} - \mathbb{E}[\hat p_{j'}]| - \Gamma_\gamma(j')\Big\}_+ + \Gamma^*_\gamma(\eta).$$

By definition of $B(\eta) = \max\left(\sup_{j'} |\mathbb{E}\hat p_{\eta\wedge j'} - \mathbb{E}\hat p_{j'}|, |\mathbb{E}\hat p_\eta - p(x)|\right)$, we get

$$\begin{aligned}\Big|\hat p_{\hat j} - p(x)\Big| \leq & 2\sup_{j,j'}\Big\{ |\hat p_{j\wedge j'} - \mathbb{E}[\hat p_{j\wedge j'}]| - \Gamma_\gamma(j\wedge j')\Big\}_+ \\ & + 3\sup_{j'}\Big\{ |\hat p_{j'} - \mathbb{E}[\hat p_{j'}]| - \Gamma_\gamma(j')\Big\}_+ + 3B(\eta) + 3\Gamma^*_\gamma(\eta).\end{aligned}$$

Consequently

$$\begin{aligned}\Big|\hat p_{\hat j} - p(x)\Big|^q \leq & 3^{2q-1}\Bigg(\big[B(\eta) + \Gamma^*_\gamma(\eta)\big]^q + \sup_{j'}\Big\{ |\hat p_{j'} - \mathbb{E}\hat p_{j'}| - \Gamma_\gamma(j')\Big\}_+^q \\ & + \sup_{j,j'}\Big\{ |\hat p_{j\wedge j'} - \mathbb{E}\hat p_{j\wedge j'}| - \Gamma_\gamma(j\wedge j')\Big\}_+^q\Bigg).\end{aligned}$$

Using Proposition 2, we have

$$\mathbb{E}\Big|\hat p_{\hat j} - p(x)\Big|^q \leq C\left(\mathbb{E}\left[\left(B(\eta) + \Gamma^*_\gamma(\eta)\right)^q\right]\right) + R_1' n^{-q}.$$

Then, we get

$$\mathbb{E}\Big|\hat p_{\hat j} - p(x)\Big|^q \leq R_1\left(\inf_\eta \mathbb{E}\left[\left(B(\eta) + \Gamma^*_\gamma(\eta)\right)^q\right]\right) + R_1' n^{-q},$$

where R_1 is a constant only depending on q and R_1' a constant depending on $\mathfrak{m}$, $\mathfrak{d}$, s, φ, c_g, $\mathcal{C}_g$. □

5.1.2. Proof of Theorem 2

Proof. The proof is a direct application of Theorem 1 together with a standard bias-variance trade-off. We first recall the assertion of this theorem:

$$\mathbb{E}\left[\Big|\hat p_{\hat j}(x) - p(x)\Big|^q\right] \leq C\left(\inf_\eta \mathbb{E}\left[\left(B(\eta) + \Gamma^*_\gamma(\eta)\right)^q\right]\right) + R_1' n^{-q}.$$

For the bias term, we use Proposition 3 to get:

$$B(\eta) \leq CL\sum_{l=1}^d 2^{-\eta_l\beta_l}, \text{for all } \eta\in J.$$

Now let us focus on $\mathbb{E}\left[\Gamma^*_\gamma(\eta)^q\right]$. We have

$$\mathbb{E}\left[\Gamma_\gamma(\eta)^q\right] = \mathbb{E}\left[\left(\sqrt{\frac{2\gamma(1+\varepsilon)\tilde\sigma^2_{\eta,\tilde\gamma}\log n}{n}} + \frac{c_\eta\gamma\log n}{n}\right)^q\right]$$

$$\leq 2^{q-1}\left(\left(\frac{2\gamma(1+\varepsilon)\log n}{n}\right)^{\frac{q}{2}}\mathbb{E}[\tilde{\sigma}^{q}_{\eta,\tilde{\gamma}}]+\left(\frac{c_\eta\gamma\log n}{n}\right)^{q}\right)$$
$$\leq C\left(\left(\frac{\log n}{n}\right)^{\frac{q}{2}}2^{(2S_\eta\nu+S_\eta)\frac{q}{2}}+\left(\frac{c_\eta\log n}{n}\right)^{q}\right),$$

using Lemma 6. But

$$c_\eta = 16\,(2\mathfrak{m}+s)\,\|T_\eta\|_\infty \leq C2^{S_\eta\nu+S_\eta},$$

using Lemma 10. Hence

$$\mathbb{E}\left[\Gamma_\gamma(\eta)^q\right]\leq C\left(\left(\frac{\log n}{n}\right)^{\frac{q}{2}}2^{(2S_\eta\nu+S_\eta)\frac{q}{2}}+\left(\frac{\log n}{n}\right)^{q}2^{(S_\eta\nu+S_\eta)q}\right).$$

We have

$$\left(\frac{\log n}{n}\right)^{\frac{q}{2}}2^{(2S_\eta\nu+S_\eta)\frac{q}{2}} \quad\geq\quad \left(\frac{\log n}{n}\right)^{q}2^{(S_\eta\nu+S_\eta)q}\Longleftrightarrow 2^{S_\eta}\leq\frac{n}{\log n},$$

which is true since by (12), $2^{S_\eta}\leq\frac{n}{\log^2 n}$.

This yields

$$\mathbb{E}[\Gamma^*_\gamma(\eta)^q]\leq C\left(\frac{2^{(2S_\eta\nu+S_\eta)}\log n}{n}\right)^{\frac{q}{2}}.$$

Eventually, we obtain the bound for the pointwise risk:

$$\mathbb{E}\left|\hat{p}_{\hat{j}}(x)-p(x)\right|^q\leq C\left(\inf_\eta\left\{L\sum_{l=1}^{d}2^{-\eta_l\beta_l}+\sqrt{\frac{2^{(2S_\eta\nu+S_\eta)}\log(n)}{n}}\right\}^q\right)+R_1'n^{-q}.$$

Setting the gradient of the right hand side of the inequality above with respect to η it turns out that the optimal η_l is proportional to $\frac{2}{\log 2}\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+2\nu+1)}(\log L+\frac{1}{2}\log(\frac{n}{\log(n)}))$, which leads for n large enough to

$$\mathbb{E}\left|\hat{p}_{\hat{j}}(x)-p(x)\right|^q\leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}}R_2\left(\frac{\log(n)}{n}\right)^{\frac{\bar{\beta}q}{2\bar{\beta}+2\nu+1}},$$

with R_2 a constant depending on $\gamma,q,\tilde{\gamma},\mathfrak{m},s,\mathfrak{d},\varphi,c_g,\mathcal{C}_g,\vec{\beta}$. The proof of Theorem 2 is completed. □

5.1.3. *Proof of Theorem 3*

Proof. We recall that $m(x)=\frac{p(x)}{f_X(x)}$ and $\hat{m}(x)=\frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x)\vee n^{-1/2}}$. We now state the main properties of the adaptive estimate $\hat{f}_X$ showed by [4] (Theorem 2): for all $q\geq 1$, all $\vec{\beta}\in(0,1]^d$, all $L>0$ and n large enough, it holds

$$\mathbb{P}\left(E_1\right):=\mathbb{P}\left(|\hat{f}_X(x)-f_X(x)|\geq C\phi_n(\vec{\beta})\right)\leq n^{-2q}, \tag{15}$$

and

$$\mathbb{P}\left(|\hat{f}_X(x) - f_X(x)| \leq Cn\right) = 1, \tag{16}$$

where $\phi_n(\vec{\beta}) := (\log(n)/n)^{\bar{\beta}/(2\bar{\beta}+2\nu+1)}$. Although the construction of the estimate $\hat{f}_X(x)$ depends on q, we remove the dependency for ease of exposition (see [4] Section 3.4 for further details). From (15), we easily deduce, since $f_X(x) \geq C_1 > 0$, for n large enough that

$$\mathbb{P}(E_2) := \mathbb{P}\left(\hat{f}_X(x) < \frac{C_1}{2}\right) \leq n^{-2q}. \tag{17}$$

We now start the proof of the theorem. We have together with (16)

$$\begin{aligned}
|\hat{m}(x) - m(x)| &= \left| \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}} - \frac{p(x)}{f_X(x)} \right| \\
&\leq \left| \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}} - \frac{p(x)}{\hat{f}_X(x) \vee n^{-1/2}} \right| + \left| \frac{p(x)}{\hat{f}_X(x) \vee n^{-1/2}} - \frac{p(x)}{f_X(x)} \right| \\
&\leq \left| \frac{\hat{p}_{\hat{j}}(x) - p(x)}{\hat{f}_X(x) \vee n^{-1/2}} \right| + \|m\|_\infty \|f_X\|_\infty \left| \frac{(\hat{f}_X(x) \vee n^{-1/2}) - f_X(x)}{f_X(x)(\hat{f}_X(x) \vee n^{-1/2})} \right| \\
&:= \mathcal{A}_1 + \|m\|_\infty \|f_X\|_\infty \mathcal{A}_2.
\end{aligned}$$

Control of $\mathbb{E}[\mathcal{A}_1^q]$. Using Cauchy-Schwarz inequality and the inequality $\hat{f}_X(x) \vee n^{-1/2} \geq n^{-1/2}$, we obtain for n large enough

$$\begin{aligned}
\mathbb{E}[\mathcal{A}_1^q] &= \mathbb{E}[\mathcal{A}_1^q \mathbb{1}_{E_2^c}] + \mathbb{E}[\mathcal{A}_1^q \mathbb{1}_{E_2}] \\
&\leq \mathbb{E}[\mathcal{A}_1^q \mathbb{1}_{E_2^c}] + \sqrt{\mathbb{E}[\mathcal{A}_1^{2q}]} \sqrt{\mathbb{P}(E_2)} \\
&\leq C\mathbb{E}\left[\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^q\right] + n^{q/2} \sqrt{\mathbb{E}\left[\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^{2q}\right]} \sqrt{\mathbb{P}(E_2)}.
\end{aligned}$$

Then, using Theorem 2 and (17), we finally have $\mathbb{E}[\mathcal{A}_1^q] \leq C\phi_n^q(\vec{\beta})$.

Control of $\mathbb{E}[\mathcal{A}_2^q]$. Using (16) and the inequality $\hat{f}_X(x) \vee n^{-1/2} \geq n^{-1/2}$, it holds for n large enough

$$\begin{aligned}
\mathbb{E}[\mathcal{A}_2^q] &\leq \mathbb{E}[\mathcal{A}_2^q \mathbb{1}_{E_1^c \cap E_2^c}] + \mathbb{E}[\mathcal{A}_2^q (\mathbb{1}_{E_1} + \mathbb{1}_{E_2})] \\
&\leq \mathbb{E}[\mathcal{A}_2^q \mathbb{1}_{E_1^c \cap E_2^c}] + Cn^{3q/2}(\mathbb{P}(E_1) + \mathbb{P}(E_2)).
\end{aligned}$$

Then, using the definition of $\mathcal{A}_2$, (15) and (17), we obtain $\mathbb{E}[\mathcal{A}_2^q] \leq C\phi_n^q(\vec{\beta})$.

Eventually, by definitions of $\mathcal{A}_1$ and $\mathcal{A}_2$, the proof is completed and

$$\mathbb{E}[|\hat{m}(x) - m(x)|^q] \leq C(\mathbb{E}[\mathcal{A}_1^q] + \mathbb{E}[\mathcal{A}_2^q]) \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_3 \left(\frac{\log(n)}{n}\right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)}$$

where R_3 is a constant depending on $\gamma, q, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$. This completes the proof of Theorem 3. □

5.1.4. Proof of Theorem 4

Following Meister [21], the proof is straightforward. Indeed, for the regression problem with errors in variables, Theorem 3.5 in [21] proves a lower bound in probability for the pointwise risk which claims that the minimax rate in dimension 1 is $n^{-\frac{2\beta}{2\beta+2\nu+1}}$ for Hölder class of index β and noise degree-of-ill-posedness parameter ν. Following step by step the proof of Theorem 3.5 in [21] in dimension 2 (the extension to general case can be easily deduced), one obtains the lower bound of Theorem 4. In fact, Meister uses densities such as Cauchy distributions which admit multivariate counterparts.

5.2. Statements and proofs of auxiliary results

This section is devoted to statements and proofs of auxiliary results used in Section 5.1

5.2.1. Statements and proofs of propositions

Let us start with Proposition 1 which states a concentration inequality of $\hat{p}_j$ around p_j.

Proposition 1. *Let j be fixed. For any $u > 0$,*

$$\mathbb{P}\left(|\hat{p}_j(x) - p_j(x)| \geq \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}\right) \leq 2e^{-u}, \tag{18}$$

where

$$\sigma_j^2 = \mathrm{Var}(Y_1 T_j(W_1)).$$

For any $\tilde{\gamma} > 1$ we have for any $\tilde{\varepsilon} > 0$ that there exists R_4 only depending on $\tilde{\gamma}$ and $\tilde{\varepsilon}$ such that

$$\mathbb{P}(\sigma_j^2 \geq (1+\tilde{\varepsilon})\tilde{\sigma}_{j,\tilde{\gamma}}^2) \leq R_4 n^{-\tilde{\gamma}},$$

$\tilde{\sigma}_{j,\tilde{\gamma}}^2$ being defined in (8).

Proof. First, note that

$$\hat{p}_j(x) = \sum_k \hat{p}_{jk}\varphi_{jk}(x) = \frac{1}{n}\sum_{l=1}^n Y_l \sum_k (\mathcal{D}_j\varphi)_{j,k}(W_l)\varphi_{jk}(x) = \frac{1}{n}\sum_{l=1}^n U_j(Y_l, W_l).$$

To prove Proposition 1, we apply the Bernstein inequality to the variables $U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]$ that are independent. Since,

$$U_j(Y_l, W_l) = Y_l T_j(W_l),$$

and

$$\mathbb{E}\left[\varepsilon_l T_j(W_l)\right] = 0,$$

we have for any $q \geq 2$,

$$A_q := \sum_{l=1}^{n} \mathbb{E}[|U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]|^q]$$
$$= \sum_{l=1}^{n} \mathbb{E}\left[|m(X_l)T_j(W_l) + \varepsilon_l T_j(W_l) - \mathbb{E}[m(X_l)T_j(W_l)]|^q\right]. \quad (19)$$

With $q = 2$,

$$\begin{aligned} A_2 &= \sum_{l=1}^{n} \mathbb{E}[|U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]|^2] \\ &= n\mathrm{Var}(Y_1 T_j(W_1)) \\ &= n\mathbb{E}[(m(X_1)T_j(W_1) + \varepsilon_1 T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)])^2] \\ &= n\mathbb{E}[\varepsilon_1^2 T_j^2(W_1)] + n\mathrm{Var}(m(X_1)T_j(W_1)) \\ &= n\left(s^2\mathbb{E}[T_j^2(W_1)] + \mathrm{Var}(m(X_1)T_j(W_1))\right). \end{aligned}$$

Now, for any $q \geq 3$, with $Z \sim \mathcal{N}(0,1)$,

$$\begin{aligned} A_q &\leq n2^{q-1}\left(\mathbb{E}[|m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q] + \mathbb{E}[|\varepsilon_1 T_j(W_1)|^q]\right) \\ &\leq n2^{q-1}\left(\mathbb{E}[|m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q] + s^q\mathbb{E}[|Z|^q]\mathbb{E}[|T_j(W_1)|^q]\right) \\ &\leq n2^{q-1}\left(\mathbb{E}[|m(X_1)T_j W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q]\right. \\ &\qquad \left. + s^q\mathbb{E}[|Z|^q]\mathbb{E}[T_j^2(W_1)]\|T_j\|_\infty^{q-2}\right). \end{aligned}$$

Furthermore,

$$\begin{aligned} &\mathbb{E}[|m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q] \\ &\qquad \leq \mathbb{E}[(m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)])^2] \times (2\|m\|_\infty \|T_j\|_\infty)^{q-2} \\ &\qquad = \mathrm{Var}(m(X_1)T_j(W_1)) \times (2\|m\|_\infty \|T_j\|_\infty)^{q-2}. \end{aligned}$$

Finally,

$$\begin{aligned} A_q &\leq n2^{q-1}\|T_j\|_\infty^{q-2}\left(\mathrm{Var}(m(X_1)T_j(W_1)) \times (2\|m\|_\infty)^{q-2} + s^q\mathbb{E}[|Z|^q]\mathbb{E}[T_j^2(W_1)]\right) \\ &\leq n2^{q-1}\|T_j\|_\infty^{q-2}\mathbb{E}[|Z|^q]\left(\mathrm{Var}(m(X_1)T_j(W_1)) \times (2\|m\|_\infty)^{q-2} + s^q\mathbb{E}[T_j^2(W_1)]\right) \\ &\leq n2^{q-1}\|T_j\|_\infty^{q-2}\mathbb{E}[|Z|^q]\left(\mathrm{Var}(m(X_1)T_j(W_1))\right. \\ &\qquad \left. + s^2\mathbb{E}[T_j^2(W_1)]\right) \times \left((2\|m\|_\infty)^{q-2} + s^{q-2}\right) \\ &\leq 2^{q-1}\|T_j\|_\infty^{q-2}\mathbb{E}[|Z|^q] \times A_2 \times (2\|m\|_\infty + s)^{q-2}. \end{aligned}$$

Besides we have (see page 23 in [22]) denoting Γ the Gamma function

$$\mathbb{E}[|Z|^q] = \frac{2^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right) \leq 2^{q/2}2^{-1/2}q! \leq 2^{(q-1)/2}q!, \quad (20)$$

as $\frac{1}{\sqrt{\pi}} \leq \frac{1}{\sqrt{2}}$ and $\Gamma(\frac{q+1}{2}) \leq \Gamma(q+1) = q!$. So, for $q \geq 3$,

$$\begin{aligned} A_q &\leq 2^{q-1}\|T_j\|_\infty^{q-2} 2^{(q-1)/2} q! \times A_2 \times (2\|m\|_\infty + s)^{q-2} \\ &\leq \frac{q!}{2} \times A_2 \times \left(2^{\frac{3q-1}{2(q-2)}}\|T_j\|_\infty (2\|m\|_\infty + s)\right)^{q-2}, \end{aligned}$$

The function $\frac{3q-1}{2(q-2)}$ is decreasing in q. Hence for any $q \geq 3$, $2^{\frac{3q-1}{2(q-2)}} \leq 16$. Thus

$$A_q \leq \frac{q!}{2} \times A_2 \times {c_j}^{q-2}, \tag{21}$$

with

$$c_j := 16\|T_j\|_\infty (2\mathfrak{m} + s).$$

We can now apply Proposition 2.9 of Massart [20]. We denote f_W the density of the W_l's. We have

$$\begin{aligned} \mathbb{E}[T_j^2(W_1)] &= \int T_j^2(w) f_W(w) dw \\ &\leq \|f_X\|_\infty \|T_j\|_2^2, \end{aligned}$$

since the density f_W is the convolution of f_X and g, $\|f_W\|_\infty = \|f_X \star g\|_\infty \leq \|f_X\|_\infty$. We have

$$\begin{aligned} \mathrm{Var}(m(X_1)T_j(W_1)) &\leq \mathbb{E}[m^2(X_1)T_j^2(W_1)] \\ &\leq \|m\|_\infty^2 \int T_j^2(w) f_W(w) dw \\ &\leq \|m\|_\infty^2 \|f_X\|_\infty \|T_j\|_2^2. \end{aligned}$$

Therefore, with

$$\sigma_j^2 = \frac{A_2}{n} = \mathrm{Var}(Y_1 T_j(W_1)), \tag{22}$$

$$\begin{aligned} \sigma_j^2 &= \sigma_\varepsilon^2 \mathbb{E}[T_j^2(W_1)] + \mathrm{Var}(m(X_1)T_j(W_1)) \qquad (23) \\ &\leq \sigma_\varepsilon^2 \|f_X\|_\infty \|T_j\|_2^2 + \|m\|_\infty^2 \|f_X\|_\infty \|T_j\|_2^2 \\ &\leq \|f_X\|_\infty \|T_j\|_2^2 (s^2 + \|m\|_\infty^2). \end{aligned}$$

We conclude that for any $u > 0$,

$$\mathbb{P}\left(|\hat{p}_j(x) - p_j(x)| \geq \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}\right) \leq 2e^{-u}. \tag{24}$$

Now, we can write

$$\hat{\sigma}_j^2 = \frac{1}{n(n-1)} \sum_{l=2}^{n} \sum_{v=1}^{l-1} \Big(U_j(Y_l, W_l) - U_j(Y_v, W_v)\Big)^2$$

$$= \frac{1}{n(n-1)} \sum_{l=2}^{n} \sum_{v=1}^{l-1} \Big(U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)] - U_j(Y_v, W_v) + \mathbb{E}[U_j(Y_v, W_v)] \Big)^2$$

$$= s_j^2 - \frac{2}{n(n-1)} \xi_j,$$

with

$$s_j^2 := \frac{1}{n(n-1)} \sum_{l=2}^{n} \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)])^2 + (U_j(Y_v, W_v) - \mathbb{E}[U_j(Y_v, W_v)])^2$$

$$= \frac{1}{n} \sum_{l=1}^{n} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)])^2$$

and

$$\xi_j := \sum_{l=2}^{n} \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]) \times (U_j(Y_v, W_v) - \mathbb{E}[U_j(Y_v, W_v)]).$$

In the sequel, we denote for any $\tilde{\gamma} > 0$,

$$\Omega_n(\tilde{\gamma}) = \left\{ \max_{1 \le l \le n} |\varepsilon_l| \le s\sqrt{2\tilde{\gamma} \log n} \right\}.$$

We have that

$$\mathbb{P}(\Omega_n(\tilde{\gamma})^c) \le n^{1-\tilde{\gamma}}. \tag{25}$$

Note that on $\Omega_n(\tilde{\gamma})$,

$$\|U_j(\cdot, \cdot)\|_\infty \le C_j,$$

we recall that

$$C_j = (\mathfrak{m} + s\sqrt{2\tilde{\gamma} \log n}) \|T_j\|_\infty.$$

Lemma 1. *For any $\tilde{\gamma} > 1$ and any $u > 0$, there exists a sequence $e_{n,j} > 0$ such that* $\limsup_j e_{n,j} = 0$ *and*

$$\mathbb{P}\left(\sigma_j^2 \ge s_j^2 + 2C_j\sigma_j \sqrt{\frac{2u(1+e_{n,j})}{n}} + \frac{\sigma_j^2 u}{3n} \middle| \Omega_n(\tilde{\gamma}) \right) \le e^{-u}.$$

Proof. We denote

$$\mathbb{P}_{\Omega_n(\tilde{\gamma})}(\cdot) = \mathbb{P}\left(\cdot | \Omega_n(\tilde{\gamma})\right), \quad \mathbb{E}_{\Omega_n(\tilde{\gamma})}(\cdot) = \mathbb{E}\left(\cdot | \Omega_n(\tilde{\gamma})\right).$$

Note that conditionally to $\Omega_n(\tilde{\gamma})$ the variables $U_j(Y_1, W_1), \dots, U_j(Y_n, W_n)$ are independent. So, we can apply the classical Bernstein inequality to the variables

$$V_l := \frac{\sigma_j^2 - (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)])^2}{n} \le \frac{\sigma_j^2}{n}.$$

Furthermore, as

$$\begin{aligned}\mathbb{E}_{\Omega_n(\tilde{\gamma})}[U_j(Y_1,W_1)] &= \mathbb{E}[m(X_1)T_j(W_1)|\Omega_n(\tilde{\gamma})] + \mathbb{E}[\varepsilon_1 T_j(W_1)|\Omega_n(\tilde{\gamma})] \\ &= \mathbb{E}[m(X_1)T_j(W_1)] \\ &= \mathbb{E}[U_j(Y_1,W_1)] \end{aligned} \tag{26}$$

we get

$$\begin{aligned}\sum_{l=1}^{n}\mathbb{E}_{\Omega_n(\tilde{\gamma})}[V_l^2] &= \frac{\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[\left(\sigma_j^2 - (U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right)^2\right]}{n} \\ &= \frac{\sigma_j^4 + \mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^4\right]}{n} \\ &\qquad - \frac{2\sigma_j^2\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]}{n} \\ &\le \frac{\sigma_j^4 + (4C_j^2 - 2\sigma_j^2)\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]}{n}.\end{aligned}$$

We shall find an upperbound for $\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]$:

$$\begin{aligned}&\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1) - \mathbb{E}[U_j(Y_1,W_1)])^2\right] \\ &\quad= \mathrm{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[\varepsilon_1^2 T_j^2(W_1)|\Omega_n(\tilde{\gamma})] \\ &\quad= \mathrm{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\frac{\mathbb{E}[\varepsilon_1^2 \mathbb{1}_{\Omega_n(\tilde{\gamma})}]}{\mathbb{P}(\Omega_n(\tilde{\gamma}))} \\ &\quad\le \mathrm{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\frac{s^2}{\mathbb{P}(\Omega_n(\tilde{\gamma}))} \\ &\quad\le \mathrm{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\frac{s^2}{1-n^{1-\tilde{\gamma}}} \\ &\quad= \mathrm{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]s^2(1+\tilde{e}_n),\end{aligned}$$

where $\tilde{e}_n = n^{1-\tilde{\gamma}} + o(n^{1-\tilde{\gamma}})$. Using (23) we have

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right] \le (1+e_{n,j})\sigma_j^2, \tag{27}$$

where $(e_{n,j})$ is a sequence such that $\limsup_j e_{n,j} = 0$.

Now let us find a lower bound for $\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]$:

$$\begin{aligned}&\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1) - \mathbb{E}[U_j(Y_1,W_1)])^2\right] \\ &\quad= \mathrm{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\frac{\mathbb{E}[\varepsilon_1^2 \mathbb{1}_{\Omega_n(\tilde{\gamma})}]}{\mathbb{P}(\Omega_n(\tilde{\gamma}))} \\ &\quad\ge \mathrm{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\mathbb{E}[\varepsilon_1^2 \mathbb{1}_{\Omega_n(\tilde{\gamma})}]\end{aligned}$$

$$= \operatorname{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\mathbb{E}[\varepsilon_1^2(1 - \mathbb{1}_{\Omega_n^c(\tilde{\gamma})})]$$
$$= \sigma_j^2 - \mathbb{E}[T_j^2(W_1)]\mathbb{E}[\varepsilon_1^2 \mathbb{1}_{\Omega_n^c(\tilde{\gamma})}].$$

Now using Cauchy Scharwz, (20) and (25) we have

$$\begin{aligned}\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1) - \mathbb{E}[U_j(Y_1,W_1)])^2\right] &\geq \sigma_j^2 - \mathbb{E}[T_j^2(W_1)](\mathbb{E}[\varepsilon_1^4])^{\frac{1}{2}}(\mathbb{P}(\Omega_n^c(\tilde{\gamma})))^{\frac{1}{2}} \\ &\geq \sigma_j^2 - Cs^2\mathbb{E}[T_j^2(W_1)]n^{\frac{1-\tilde{\gamma}}{2}} \\ &= \sigma_j^2(1+\tilde{e}_{n,j}), \qquad (28)\end{aligned}$$

where $(\tilde{e}_{n,j})$ is a sequence such that $\limsup_j \tilde{e}_{n,j} = 0$.

Finally, using the bounds we just got for $\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1) - \mathbb{E}[U_j(Y_1,W_1)])^2\right]$ yields

$$\begin{aligned}\sum_{l=1}^{n} \mathbb{E}_{\Omega_n(\tilde{\gamma})}[V_l^2] &\leq \frac{\sigma_j^4 + 4C_j^2\sigma_j^2(1+e_{n,j}) - 2\sigma_j^4(1+\tilde{e}_{n,j})}{n} \\ &\leq \frac{4C_j^2\sigma_j^2(1+e_{n,j}) - \sigma_j^4(1+2\tilde{e}_{n,j})}{n} \\ &\leq \frac{4C_j^2\sigma_j^2(1+e_{n,j})}{n}.\end{aligned}$$

We obtain the claimed result. □

Now, we deal with ξ_j.

Lemma 2. *There exists an absolute constant $c > 0$ such that for any $u > 1$,*

$$\mathbb{P}\left(\xi_j \geq c(n\sigma_j^2 u + C_j^2 u^2) \,\middle|\, \Omega_n(\tilde{\gamma})\right) \leq 3e^{-u}.$$

Proof. Note that conditionally to $\Omega_n(\tilde{\gamma})$, the vectors $(Y_l, W_l)_{1 \leq l \leq n}$ are independent. We remind that by (26), (27) and (28) we have

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})}[U_j(Y_1,W_1)] = \mathbb{E}[U_j(Y_1,W_1)] \qquad (29)$$

and

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1) - \mathbb{E}[U_j(Y_1,W_1)])^2\right] = (1+e_{n,j})\sigma_j^2.$$

The ξ_j can be written as

$$\xi_j = \sum_{l=2}^{n}\sum_{v=1}^{l-1} g_j(Y_l, W_l, Y_v, W_v),$$

with

$$g_j(y,w,y',w') = (U_j(y,w) - \mathbb{E}[U_j(Y_1,W_1)])) \times (U_j(y',w') - \mathbb{E}[U_j(Y_1,W_1)]).$$

Previous computations show that conditions (2.3) and (2.4) of Houdré and Reynaud-Bouret [16] are satisfied. So that we are able to apply Theorem 3.1 of [16]: there exist absolute constants c_1, c_2, c_3 and c_4 such that for any $u > 0$,

$$\mathbb{P}_{\Omega_n(\tilde{\gamma})}\left(\xi_j \geq c_1 C\sqrt{u} + c_2 Du + c_3 Bu^{3/2} + c_4 Au^2\right) \leq 3e^{-u},$$

where A, B, C, and D are defined and controlled as follows. We have:

$$A = \|g_j\|_\infty \leq 4C_j^2.$$

$$C^2 = \sum_{l=2}^{n} \sum_{v=1}^{l-1} \mathbb{E}_{\Omega_n(\tilde{\gamma})}[g_j^2(Y_l, W_l, Y_v, W_v)] = \frac{n(n-1)}{2} \sigma_j^4 (1 + e_{n,j})^2.$$

Let

$$\mathcal{A} = \left\{ (a_l)_l, (b_v)_v : \ \mathbb{E}_{\Omega_n(\tilde{\gamma})} \left[\sum_{l=2}^{n} a_l^2(Y_l, W_l) \right] \leq 1, \ \mathbb{E}_{\Omega_n(\tilde{\gamma})} \left[\sum_{l=1}^{n-1} b_l^2(Y_l, W_l) \right] \leq 1 \right\}.$$

We have:

$$\begin{aligned}
D &= \sup_{(a_l)_l, (b_v)_v \in \mathcal{A}} \mathbb{E}_{\Omega_n(\tilde{\gamma})} \left[\sum_{l=2}^{n} \sum_{v=1}^{l-1} g_j(Y_l, W_l, Y_v, W_v) a_l(Y_l, W_l) b_v(Y_v, W_v) \right] \\
&= \sup_{(a_l)_l, (b_v)_v \in \mathcal{A}} \Bigg[\sum_{l=2}^{n} \sum_{v=1}^{l-1} \mathbb{E}_{\Omega_n(\tilde{\gamma})} \left[(U_j(Y_l, W_l) - [U_j(Y_l, W_l)])) a_l(Y_l, W_l) \right] \\
&\qquad \times \mathbb{E}_{\Omega_n(\tilde{\gamma})} \left[(U_j(Y_v, W_v) - \mathbb{E}[U_j(Y_v, W_v)])) b_v(Y_v, W_v) \right] \Bigg] \\
&\leq \sup_{(a_l)_l, (b_v)_v \in \mathcal{A}} \sum_{l=2}^{n} \sum_{v=1}^{l-1} \sigma_j^2 (1 + e_{n,j}) \sqrt{\mathbb{E}_{\Omega_n(\tilde{\gamma})}[a_l^2(Y_l, W_l)] \mathbb{E}_{\Omega_n(\tilde{\gamma})}[b_v^2(Y_v, W_v)]} \\
&\leq \sigma_j^2 (1 + e_{n,j}) \sup_{(a_l)_l, (b_v)_v \in \mathcal{A}} \sum_{l=2}^{n} \sqrt{l-1} \\
&\qquad \times \sqrt{\mathbb{E}_{\Omega_n(\tilde{\gamma})}[a_l^2(Y_l, W_l)] \sum_{v=1}^{l-1} \mathbb{E}_{\Omega_n(\tilde{\gamma})}[b_v^2(Y_v, W_v)]} \\
&\leq \sigma_j^2 (1 + e_{n,j}) \sqrt{\frac{n(n-1)}{2}}.
\end{aligned}$$

Finally,

$$\begin{aligned}
B^2 &= \sup_{y,w} \sum_{v=1}^{n-1} \mathbb{E}_{\Omega_n(\tilde{\gamma})} \Big[(U_j(y, w) - \mathbb{E}[U_j(Y_1, W_1)]))^2 \\
&\qquad \times (U_j(Y_v, W_v) - \mathbb{E}[U_j(Y_1, W_1)])^2 \Big] \\
&\leq 4(n-1) C_j^2 \sigma_j^2 (1 + e_{n,j}).
\end{aligned}$$

Therefore, there exists an absolute constant $c > 0$ such that for any $u > 1$,

$$c_1 C \sqrt{u} + c_2 D u + c_3 B u^{3/2} + c_4 A u^2 \leq c(n\sigma_j^2 u + C_j^2 u^2). \qquad \square$$

Let us go back to the proof of Proposition 1. We apply Lemmas 1 and 2 with $u > 1$ and we obtain, by setting

$$M_j(u) = \hat{\sigma}_j^2 + 2C_j\sigma_j\sqrt{\frac{2u(1+e_{n,j})}{n}} + \frac{\sigma_j^2 u}{3n} + \frac{2c(n\sigma_j^2 u + C_j^2 u^2)}{n(n-1)},$$

$$\begin{aligned}\mathbb{P}\left(\sigma_j^2 \geq M_j(u)\right) &\leq \mathbb{P}\Bigg(\sigma_j^2 \geq s_j^2 - \frac{2}{n(n-1)}\xi_j + 2C_j\sigma_j\sqrt{\frac{2u(1+e_{n,j})}{n}} \\ &\qquad + \frac{\sigma_j^2 u}{3n} + \frac{2c(n\sigma_j^2 u + C_j^2 u^2)}{n(n-1)}\Bigg) \\ &\leq \mathbb{P}\left(\sigma_j^2 \geq s_j^2 + 2C_j\sigma_j\sqrt{\frac{2u(1+e_{n,j})}{n}} + \frac{\sigma_j^2 u}{3n}\,\middle|\, \Omega_n(\tilde{\gamma})\right) \\ &\qquad + \mathbb{P}\left(\xi_j \geq c(n\sigma_j^2 u + C_j^2 u^2) \,\middle|\, \Omega_n(\tilde{\gamma})\right) + 1 - \mathbb{P}(\Omega_n(\tilde{\gamma})).\end{aligned}$$

Therefore, with $u = \tilde{\gamma}\log n$ and $\tilde{\gamma} > 1$, we obtain for n large enough:

$$\mathbb{P}\left(\sigma_j^2 \geq M_j(\tilde{\gamma}\log n)\right) \leq 5n^{-\tilde{\gamma}}.$$

And there exist a and b two absolute constants such that

$$\mathbb{P}\left(\sigma_j^2 \geq \hat{\sigma}_j^2 + 2C_j\sigma_j\sqrt{\frac{2\tilde{\gamma}\log n(1+e_{n,j})}{n}} + \frac{\sigma_j^2 a\tilde{\gamma}\log n}{n} + \frac{C_j^2 b^2\tilde{\gamma}^2\log^2 n}{n^2}\right) \leq 5n^{-\tilde{\gamma}}.$$

Now, we set

$$\theta_1 = \left(1 - \frac{a\tilde{\gamma}\log n}{n}\right),\ \theta_2 = C_j\sqrt{\frac{2\tilde{\gamma}\log n(1+e_{n,j})}{n}},\ \theta_3 = \hat{\sigma}_j^2 + \frac{C_j^2 b^2\tilde{\gamma}^2\log^2 n}{n^2}$$

so

$$\mathbb{P}\left(\theta_1\sigma_j^2 - 2\theta_2\sigma_j - \theta_3 \geq 0\right) \leq 5n^{-\tilde{\gamma}}.$$

We study the polynomial

$$p(\sigma) = \theta_1\sigma^2 - 2\theta_2\sigma - \theta_3.$$

Since $\sigma \geq 0$, $p(\sigma) \geq 0$ means that

$$\sigma \geq \frac{1}{\theta_1}\left(\theta_2 + \sqrt{\theta_2^2 + \theta_1\theta_3}\right),$$

which is equivalent to

$$\sigma^2 \geq \frac{1}{\theta_1^2}\left(2\theta_2^2 + \theta_1\theta_3 + 2\theta_2\sqrt{\theta_2^2 + \theta_1\theta_3}\right).$$

Hence

$$\mathbb{P}\left(\sigma_j^2 \geq \frac{1}{\theta_1^2}\left(2\theta_2^2 + \theta_1\theta_3 + 2\theta_2\sqrt{\theta_2^2 + \theta_1\theta_3}\right)\right) \leq 5n^{-\tilde{\gamma}}.$$

So,

$$\mathbb{P}\left(\sigma_j^2 \geq \frac{\theta_3}{\theta_1} + \frac{2\theta_2\sqrt{\theta_3}}{\theta_1\sqrt{\theta_1}} + \frac{4\theta_2^2}{\theta_1^2}\right) \leq 5n^{-\tilde{\gamma}}.$$

So, there exist absolute constants δ, η, and τ' depending only on $\tilde{\gamma}$ so that for n large enough,

$$\mathbb{P}\Bigg(\sigma_j^2 \geq \hat{\sigma}_j^2\left(1+\delta\frac{\log n}{n}\right) + \left(1+\eta\frac{\log n}{n}\right)2C_j\sqrt{2\tilde{\gamma}\hat{\sigma}_j^2(1+e_{n,j})\frac{\log n}{n}} + 8\tilde{\gamma}C_j^2\frac{\log n}{n}\left(1+\tau'\left(\frac{\log n}{n}\right)^{1/2}\right)\Bigg) \leq 5n^{-\tilde{\gamma}}.$$

Finally, for all $\tilde{\varepsilon} > 0$ there exists R_4 depending on ε' and $\tilde{\gamma}$ such that for n large enough

$$\mathbb{P}(\sigma_j^2 \geq (1+\varepsilon')\tilde{\sigma}_{j,\tilde{\gamma}}^2) \leq R_4 n^{-\tilde{\gamma}}.$$

Combining this inequality with (24), we obtain the desired result of Proposition 1. □

Proposition 2 shows that the residual term in the oracle inequality is negligible.

Proposition 2. *We have for any $q \geq 1$,*

$$\mathbb{E}\left[\sup_{j\in J}\left(|\hat{p}_j(x) - p_j(x)| - \Gamma_\gamma(j)\right)_+^q\right] \leq R_1' n^{-q}, \tag{30}$$

with R_1' a constant depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ, c_g, $\mathcal{C}_g$ and φ.

Proof. We recall that $J = \left\{j \in \mathbb{N}^d : \quad 2^{S_j} \leq \lfloor \frac{n}{\log^2 n}\rfloor\right\}$.

Let $\tilde{\gamma} > 0$ and let us consider the event

$$\tilde{\Omega}_{\tilde{\gamma}} = \left\{\sigma_j^2 \leq (1+\varepsilon)\tilde{\sigma}_{j,\tilde{\gamma}}^2,\ \forall j \in J\right\}.$$

Let $\gamma > 0$. We set in the sequel

$$E := \mathbb{E}\left[\sup_{j\in J}\left(|\hat{p}_j(x) - p_j(x)| - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}_{j,\tilde{\gamma}}^2\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)_+^q \mathbb{1}_{\tilde{\Omega}_{\tilde{\gamma}}}\right],$$

and $R_j := |\hat{p}_j(x) - p_j(x)|$. We have:

$$\begin{aligned} E &= \int_0^\infty \mathbb{P}\left[\sup_{j\in J}\left(R_j - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}_{j,\tilde{\gamma}}^2\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)_+^q \mathbb{1}_{\tilde{\Omega}_{\tilde{\gamma}}} > y\right] dy \\ &\leq \sum_{j\in J}\int_0^\infty \mathbb{P}\left[\left(R_j - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}_{j,\tilde{\gamma}}^2\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)_+^q \mathbb{1}_{\tilde{\Omega}_{\tilde{\gamma}}} > y\right] dy \end{aligned}$$

$$\leq \sum_{j\in J}\int_0^\infty \mathbb{P}\left[\left(R_j - \sqrt{\frac{2\gamma\sigma_j^2\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)^q > y\right]dy.$$

Let us take u such that

$$y = h(u)^q,$$

where

$$h(u) = \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}.$$

Note that for any $u > 0$,

$$h'(u) \leq \frac{h(u)}{u}.$$

Hence

$$\begin{aligned} E &\leq C\sum_{j\in J}\int_0^\infty \mathbb{P}\left[R_j > \sqrt{\frac{2\gamma\sigma_j^2\log n}{n}} + \frac{c_j\gamma\log n}{n} + \sqrt{\frac{2u\sigma_j^2}{n}} + \frac{uc_j}{n}\right] h(u)^{q-1}h'(u)du \\ &\leq C\sum_{j\in J}\int_0^\infty \mathbb{P}\left[R_j > \sqrt{\frac{2\sigma_j^2(\gamma\log n + u)}{n}} + \frac{c_j(\gamma\log n + u)}{n}\right] h(u)^{q-1}h'(u)du. \end{aligned}$$

Now using concentration inequality (18), we get

$$\begin{aligned} E &\leq C\sum_{j\in J}\int_0^\infty e^{-(\gamma\log n + u)}h(u)^{q-1}h'(u)du \\ &\leq C\sum_{j\in J}\int_0^\infty e^{-(\gamma\log n + u)}h(u)^q\frac{1}{u}du \\ &\leq Ce^{-\gamma\log n}\sum_{j\in J}\int_0^\infty e^{-u}\left(\sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}\right)^q\frac{1}{u}du \\ &\leq C\left(e^{-\gamma\log n}\sum_{j\in J}\left(\frac{\sigma_j^2}{n}\right)^{q/2}\int_0^\infty e^{-u}u^{\frac{q}{2}-1}du + \left(\frac{c_j}{n}\right)^q\int_0^\infty e^{-u}u^{q-1}du\right). \end{aligned}$$

Now using Lemma 10, we have that $\sigma_j^2 \leq R_{10}2^{(2S_j\nu+S_j)}$ and $c_j \leq C2^{S_j\nu+S_j}$. Hence,

$$\begin{aligned} E &\leq C\left(e^{-\gamma\log n}\sum_{j\in J}\left(\frac{2^{(2S_j\nu+S_j)}}{n}\right)^{q/2} + \left(\frac{2^{(S_j\nu+S_j)}}{n}\right)^q\right) \\ &\leq Cn^{-\gamma+q\nu}(\log n)^{-(2\nu+1)q} \leq Cn^{-q}, \end{aligned}$$

as soon as $\gamma > q(\nu+1)$.

It remains to find an upperbound for the following quantity:

$$E' := \mathbb{E}\left[\sup_{j\in J}\left(|\hat{p}_j(x)-p_j(x)| - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}^2_{j,\tilde{\gamma}}\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)_+^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right].$$

We have

$$\begin{aligned}E' &\leq \mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)-p_j(x)|^q\,\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\\ &\leq 2^{q-1}\left(\mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]+\mathbb{E}\left[\sup_{j\in J}(|p_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\right).\end{aligned}$$

First, let us deal with the term $\mathbb{E}\left[\sup_{j\in J}(|p_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]$.

Following the lines of the proof of Lemma 7 we easily get that $\sum_k \varphi^2_{jk}(x) \leq C2^{S_j}$, hence

$$\begin{aligned}|p_j(x)| = \left|\sum_k p_{jk}\varphi_{jk}(x)\right| &\leq \left(\sum_k p^2_{jk}\right)^{\frac{1}{2}}\left(\sum_k \varphi^2_{jk}(x)\right)^{\frac{1}{2}}\\ &\leq C\|p\|_2 2^{\frac{S_j}{2}}.\end{aligned}$$

Now using Proposition 1 which states that $\mathbb{P}(\tilde{\Omega}^c_{\tilde{\gamma}}) \leq Cn^{-\tilde{\gamma}}$

$$\begin{aligned}\mathbb{E}\left[\sup_{j\in J}(|p_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] &\leq \sup_{j\in J}(\|p\|_2 2^{\frac{S_j}{2}})^q\mathbb{P}(\tilde{\Omega}^c_{\tilde{\gamma}}) &(31)\\ &\leq C\left(\frac{n}{\log^2 n}\right)^{\frac{q}{2}} n^{-\tilde{\gamma}}. &(32)\end{aligned}$$

It remains to find an upperbound for $\mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]$. We have

$$\begin{aligned}&\mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\\ &\quad= \mathbb{E}\left[\sup_{j\in J}\left|\frac{1}{n}\sum_{l=1}^n Y_l T_j(W_l)\right|^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\\ &\quad\leq \frac{1}{n^q}\mathbb{E}\left[\sup_{j\in J}\left(\sum_{l=1}^n |m(X_l)+\varepsilon_l|\,|T_j(W_l)|\right)^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\\ &\quad\leq \frac{n^{q-1}}{n^q}\mathbb{E}\left[\sup_{j\in J}\sum_{l=1}^n |m(X_l)+\varepsilon_l|^q\,|T_j(W_l)|^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\\ &\quad\leq \frac{C}{n}\mathbb{E}\left[\sup_{j\in J}\sum_{l=1}^n (\|m\|^q_\infty+|\varepsilon_l|^q)|T_j(W_l)|^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\end{aligned}$$

$$\leq C\left(\sup_{j\in J}(\|T_j\|_\infty^q)\mathbb{P}(\tilde\Omega^c_{\tilde\gamma}) + \sup_{j\in J}(\|T_j\|_\infty^q)\mathbb{E}\left[|\varepsilon_1|^q \mathbb{1}_{\tilde\Omega^c_{\tilde\gamma}}\right]\right)$$
$$\leq C\left(\sup_{j\in J}(\|T_j\|_\infty^q)\mathbb{P}(\tilde\Omega^c_{\tilde\gamma}) + s^q \sup_{j\in J}(\|T_j\|_\infty^q)\left(\mathbb{E}\left[|Z|^{2q}\right]\right)^{\frac12}\left(\mathbb{P}(\tilde\Omega^c_{\tilde\gamma})\right)^{\frac12}\right),$$

where $Z \sim \mathcal{N}(0,1)$. Using (20) and $\|T_j\|_\infty \leq T_4 2^{S_j(\nu+1)}$, we get

$$\mathbb{E}\left[\sup_{j\in J}(|\hat p_j(x)|)^q \mathbb{1}_{\tilde\Omega^c_{\tilde\gamma}}\right] \leq C\left(\frac{n}{\log^2 n}\right)^{(\nu+1)q} n^{-\frac{\tilde\gamma}{2}},$$

We have

$$E' \leq Cn^{-\frac{\tilde\gamma}{2}}\left(\left(\frac{n}{\log^2 n}\right)^{\frac q2} + \left(\frac{n}{\log^2 n}\right)^{(\nu+1)q}\right)$$
$$\leq Cn^{-q},$$

as soon as $\tilde\gamma > 2q(\nu+2)$. This ends the proof of Proposition 2. □

Proposition 3 controls the bias term in the oracle inequality.

Proposition 3. *For any $j = (j_1,\dots,j_d) \in \mathbb{Z}^d$ and $j' = (j'_1,\dots,j'_d) \in \mathbb{Z}^d$ and any x, if $p \in \mathbb{H}_d(\vec\beta, L)$*

$$|p_{j\wedge j'}(x) - p_{j'}(x)| \leq R_{12} L \sum_{l=1}^d 2^{-j_l\beta_l},$$

where R_{12} is a constant only depending on φ and $\vec\beta$. We have denoted

$$j\wedge j' = (j_1\wedge j'_1,\dots,j_d\wedge j'_d).$$

Proof. We first state three lemmas.

Lemma 3. *For any j and any k, we have:*

$$\mathbb{E}[\hat p_{jk}] = p_{jk}.$$

Proof. Recall that

$$\hat p_{jk} := \frac1n\sum_{u=1}^n Y_u \times (\mathcal{D}_j\varphi)_{j,k}(W_u) = \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\frac1n\sum_{u=1}^n Y_u \int e^{-i<t,2^jW_u-k>}\frac{\overline{\mathcal{F}(\varphi)(t)}}{\mathcal{F}(g)(2^jt)}dt.$$

Let us prove now that $\mathbb{E}(\hat p_{jk}) = p_{jk}$.

We have

$$\mathbb{E}(\hat p_{jk}) = \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\left(\int \mathbb{E}(Y_1 e^{-i<t,2^jW_1-k>})\frac{\overline{\mathcal{F}(\varphi)(t)}}{\mathcal{F}(g)(2^jt)}dt\right).$$

We shall develop the right member of the last equality. We have:

$$\begin{aligned}
\mathbb{E}\left[Y_1 e^{-i<t,2^j W_1-k>}\right] &= \mathbb{E}\left[(m(X_1)+\varepsilon_1)e^{-i<t,2^j W_1-k>}\right]\\
&= \mathbb{E}\left[m(X_1)e^{-i<t,2^j W_1-k>}\right]\\
&= \mathbb{E}\left[m(X_1)e^{-i<t,2^j X_1-k>}\right]\mathbb{E}\left[e^{-i<t,2^j\delta_1>}\right]\\
&= \int m(x)e^{-i<t,2^j x-k>} f_X(x)dx \times \mathcal{F}(g)(2^j t)\\
&= e^{i<t,k>}\mathcal{F}(p)(2^j t)\mathcal{F}(g)(2^j t).
\end{aligned}$$

Consequently

$$\begin{aligned}
\mathbb{E}\left[\hat{p}_{jk}\right] &= \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\int e^{i<t,k>}\mathcal{F}(p)(2^j t)\mathcal{F}(g)(2^j t)\frac{\overline{\mathcal{F}(\varphi)(t)}}{\mathcal{F}(g)(2^j t)}dt\\
&= \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\int e^{i<t,k>}\mathcal{F}(p)(2^j t)\overline{\mathcal{F}(\varphi)(t)}dt\\
&= \frac{1}{(2\pi)^d}\int \mathcal{F}(p)(t)\overline{\mathcal{F}(\varphi_{jk})(t)}dt.
\end{aligned}$$

Since by Parseval equality, we have

$$p_{jk} = \int p(t)\varphi_{jk}(t)dt = \frac{1}{(2\pi)^d}\int \mathcal{F}(p)(t)\overline{\mathcal{F}(\varphi_{jk})(t)}dt,$$

the result follows.

Note that in the case where we don't have any noise on the variable i.e $g(x) = \delta_0(x)$, since $\mathcal{F}(g)(t) = 1$, the proof above remains valid and we get $\mathbb{E}[\hat{p}_{jk}] = p_{jk}$. □

Lemma 4. *If for any l, $\lfloor\beta_l\rfloor \le N$, the following holds: for any $j \in \mathbb{Z}^d$ and any $p \in \mathbb{H}_d(\vec{\beta}, L)$,*

$$|\mathbb{E}[\hat{p}_j(x)] - p(x)| \le L(\|\varphi\|_\infty\|\varphi\|_1)^d(2A+1)^d\sum_{l=1}^d \frac{(2A\times 2^{-j_l})^{\beta_l}}{\lfloor\beta_l\rfloor!}.$$

Proof. Let x be fixed and $j = (j_1, \dots, j_d) \in \mathbb{Z}^d$. We have:

$$\int K_j(x,y)dy = \int \sum_{k_1}\cdots\sum_{k_d}\prod_{l=1}^d [2^{j_l}\varphi(2^{j_l}x_l - k_l)\varphi(2^{j_l}y_l - k_l)dy_l] = 1.$$

Therefore, using lemma 3

$$\begin{aligned}
&\mathbb{E}[\hat{p}_j(x)] - p(x)\\
&\qquad = p_j(x) - p(x)
\end{aligned}$$

$$= \int K_j(x,y)(p(y)-p(x))dy$$
$$= \sum_k \varphi_{jk}(x) \int \varphi_{jk}(y)(p(y)-p(x))dy$$
$$= \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \varphi_{jk}(x) \int \prod_{l=1}^{d} 2^{\frac{j_l}{2}} \varphi(2^{j_l} y_l - k_l)(p(y)-p(x))dy.$$

Now, we use that

$$p(y)-p(x) = \sum_{l=1}^{d} p(x_1,\dots,x_{l-1},y_l,y_{l+1},\dots,y_d) - p(x_1,\dots x_{l-1},x_l,y_{l+1},\dots,y_d),$$

with $p(x_1,\dots,x_l,y_{l+1},\dots,y_d) = p(x_1,\dots,x_d)$ if $l = d$ and $p(x_1,\dots,x_{l-1}, y_l,\dots,y_d) = p(y_1,\dots,y_d)$ if $l = 1$. Furthermore, the Taylor expansion gives: for any $l \in \{1,\dots,d\}$, for some $u_l \in [0;1]$,

$$p(x_1,\dots,x_{l-1},y_l,y_{l+1},\dots,y_d) - p(x_1,\dots x_{l-1},x_l,y_{l+1},\dots,y_d) =$$
$$\sum_{k=1}^{\lfloor \beta_l \rfloor} \frac{\partial^k p}{\partial x_l^k}(x_1,\dots x_{l-1},x_l,y_{l+1},\dots,y_d) \times \frac{(y_l-x_l)^k}{k!} +$$
$$\frac{\partial^{\lfloor \beta_l \rfloor} p}{\partial x_l^{\lfloor \beta_l \rfloor}}(x_1,\dots x_{l-1},x_l+(y_l-x_l)u_l,y_{l+1},\dots,y_d) \times \frac{(y_l-x_l)^{\lfloor \beta_l \rfloor}}{\lfloor \beta_l \rfloor !}$$
$$-\frac{\partial^{\lfloor \beta_l \rfloor} p}{\partial x_l^{\lfloor \beta_l \rfloor}}(x_1,\dots x_{l-1},x_l,y_{l+1},\dots,y_d) \times \frac{(y_l-x_l)^{\lfloor \beta_l \rfloor}}{\lfloor \beta_l \rfloor !}.$$

Using vanishing moments of K_j and $p \in \mathbb{H}_d(\vec{\beta}, L)$, we obtain:

$$|p_j(x)-p(x)|$$
$$\leq \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} |\varphi_{jk}(x)| \int \prod_{l=1}^{d} 2^{\frac{j_l}{2}} |\varphi(2^{j_l} y_l - k_l)| \sum_{l=1}^{d} L \frac{|y_l - x_\ell|^{\beta_l}}{\lfloor \beta_l \rfloor !} dy$$
$$\leq \|\varphi\|_\infty^d \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \int_{[-A;A]^d} \prod_{l=1}^{d} |\varphi(u_l)| \sum_{l=1}^{d} L \frac{|2^{-j_l}(u_l+k_l) - x_l|^{\beta_l}}{\lfloor \beta_l \rfloor !} du.$$

Since for any l, $k_l \in \mathcal{Z}_{j,l}(x)$, we finally obtain

$$|p_j(x)-p(x)|$$
$$\leq \|\varphi\|_\infty^d \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \int_{[-A;A]^d} \prod_{l=1}^{d} |\varphi(u_l)| \sum_{l=1}^{d} L \frac{(2A \times 2^{-j_l})^{\beta_l}}{\lfloor \beta_l \rfloor !} du$$
$$\leq L(\|\varphi\|_\infty \|\varphi\|_1)^d (2A+1)^d \sum_{l=1}^{d} \frac{(2A \times 2^{-j_l})^{\beta_l}}{\lfloor \beta_l \rfloor !}. \qquad \square$$

Lemma 5. *We have for any $j = (j_1, \dots, j_d) \in \mathbb{Z}^d$ and $j' = (j'_1, \dots, j'_d) \in \mathbb{Z}^d$ and any x,*

$$K_{j'}(p_j)(x) = p_{j \wedge j'}(x).$$

Proof. We only deal with the case $d = 2$. The extension to the general case can be easily deduced. If for $i = 1, 2$, $j_i \le j'_i$ the result is obvious. It is also the case if for $l = 1, 2$, $j'_l \le j_l$. So, without loss of generality, we assume that $j_1 \le j'_1$ and $j'_2 \le j_2$. We have:

$$\begin{aligned}
K_{j'}(p_j)(x) &= \int K_{j'}(x,y) p_j(y) dy \\
&= \int \sum_k \varphi_{j'k}(x)\varphi_{j'k}(y) p_j(y) dy \\
&= \int \sum_{k_1} \sum_{k_2} \varphi_{j'_1 k_1}(x_1)\varphi_{j'_2 k_2}(x_2)\varphi_{j'_1 k_1}(y_1)\varphi_{j'_2 k_2}(y_2) p_j(y) dy_1 dy_2 \\
&= \int \sum_{k_1} \sum_{k_2} \varphi_{j'_1 k_1}(x_1)\varphi_{j'_2 k_2}(x_2)\varphi_{j'_1 k_1}(y_1)\varphi_{j'_2 k_2}(y_2) \\
&\times \sum_{\ell_1} \sum_{\ell_2} \varphi_{j_1 \ell_1}(y_1)\varphi_{j_2 \ell_2}(y_2)\varphi_{j_1 \ell_1}(u_1)\varphi_{j_2 \ell_2}(u_2) p(u_1,u_2) du_1 du_2 dy_1 dy_2.
\end{aligned}$$

Since $j_1 \le j'_1$, we have in the one-dimensional case, by a slight abuse of notation, $V_{j_1} \subset V_{j'_1}$ and

$$\int \sum_{k_1} \varphi_{j'_1 k_1}(x_1)\varphi_{j'_1 k_1}(y_1)\varphi_{j_1 \ell_1}(y_1) dy_1 = \int K_{j'_1}(x_1, y_1)\varphi_{j_1 \ell_1}(y_1) dy_1 = \varphi_{j_1 \ell_1}(x_1).$$

Similarly, since $j'_2 \le j_2$, we have $V_{j'_2} \subset V_{j_2}$ and

$$\int \sum_{\ell_2} \varphi_{j_2 \ell_2}(y_2)\varphi_{j_2 \ell_2}(u_2)\varphi_{j'_2 k_2}(y_2) dy_2 = \int K_{j_2}(u_2, y_2)\varphi_{j'_2 k_2}(y_2) dy_2 = \varphi_{j'_2 k_2}(u_2).$$

Therefore, with $\tilde{j} = j \wedge j'$,

$$\begin{aligned}
K_{j'}(p_j)(x) &= \int \sum_{k_2} \sum_{\ell_1} \varphi_{j'_2 k_2}(x_2)\varphi_{j_1 \ell_1}(u_1)\varphi_{j_1 \ell_1}(x_1)\varphi_{j'_2 k_2}(u_2) p(u_1,u_2) du_1 du_2 \\
&= \int \sum_{\ell_1} \sum_{\ell_2} \varphi_{\tilde{j}_2 \ell_2}(x_2)\varphi_{\tilde{j}_1 \ell_1}(u_1)\varphi_{\tilde{j}_1 \ell_1}(x_1)\varphi_{\tilde{j}_2 \ell_2}(u_2) p(u_1,u_2) du_1 du_2 \\
&= \int \sum_{\ell} \varphi_{\tilde{j}\ell}(x)\varphi_{\tilde{j}\ell}(u) p(u) du \\
&= p_{\tilde{j}}(x),
\end{aligned}$$

which ends the proof of the lemma. □

Now, we shall go back to the proof of Proposition 3. We easily deduce the result:

$$p_{j\wedge j'}(x) - p_{j'}(x) = K_{j'}(p_j)(x) - K_{j'}(p)(x) \\ = \int K_{j'}(x,y)(p_j(y) - p(y))dy.$$

Therefore,

$$|p_{j\wedge j'}(x) - p_{j'}(x)| \leq \int |K_{j'}(x,y)||p_j(y) - p(y)|dy \\ \leq R_{12} L \sum_{l=1}^{d} 2^{-j_l \beta_l} \times \int |K_{j'}(x,y)|dy,$$

where R_{12} is a constant only depending on φ and $\vec{\beta}$. We conclude by observing that

$$\int |K_{j'}(x,y)|dy = \int \sum_{k_1} \cdots \sum_{k_d} \prod_{l=1}^{d} [2^{j_l'} |\varphi(2^{j_l'} x_l - k_l)||\varphi(2^{j_l'} y_l - k_l)|dy_i] \\ \leq \|\varphi\|_\infty^d \sum_{k_1 \in \mathcal{Z}_{j',1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j',d}(x)} \left(\int |\varphi(v)|dv \right)^d \\ \leq \left(\|\varphi\|_\infty \|\varphi\|_1 (2A+1) \right)^d.$$

We thus obtain the claimed result of Proposition 3. □

5.2.2. Appendix

Technical lemmas are stated and proved below.

Lemma 6. *We have*

$$\mathbb{E}[(\tilde{\sigma}_{j,\tilde{\gamma}})^q] \leq R_5 2^{S_j(2\nu+1)\frac{q}{2}},$$

with R_5 a constant depending on $q, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g$.

Proof. First, let us focus on the case $q \geq 2$. We recall the expression of $\tilde{\sigma}_{j,\tilde{\gamma}}^2$

$$\tilde{\sigma}_{j,\tilde{\gamma}}^2 = \hat{\sigma}_j^2 + 2C_j \sqrt{2\tilde{\gamma}\hat{\sigma}_j^2 \frac{\log n}{n}} + 8\tilde{\gamma} C_j^2 \frac{\log n}{n}.$$

We shall first prove that

$$\mathbb{E}[(\hat{\sigma}_j)^q] \leq C 2^{S_j(2\nu+1)\frac{q}{2}}.$$

Let us remind that

$$\hat{\sigma}_j^2 = \frac{1}{2n(n-1)} \sum_{l \neq v} (U_j(Y_l, W_l) - U_j(Y_v, W_v))^2.$$

We easily get

$$\hat{\sigma}_j^2 \leq \frac{C}{n} \sum_l (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_1, W_1)])^2.$$

First let us remark that

$$\left(\sum_l (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_1, W_1)])^2 \right)^{\frac{q}{2}} \\ \leq C \left(\left(\sum_l ((U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_1, W_1)])^2 - \sigma_j^2) \right)^{\frac{q}{2}} + n^{\frac{q}{2}} \sigma_j^q \right)$$

We will use Rosenthal inequality (see [15]) to find an upper bound for

$$\mathbb{E}\left[\left(\sum_l ((U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_1, W_1)])^2 - \sigma_j^2) \right)^{\frac{q}{2}} \right].$$

We set

$$B_l := (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_1, W_1)])^2 - \sigma_j^2.$$

The variables B_l are i.i.d and centered. We have to check that $\mathbb{E}[|B_l|^{\frac{q}{2}}] < \infty$. We have

$$\mathbb{E}[|B_l|^{\frac{q}{2}}] \leq C(\mathbb{E}[|(U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_1, W_1)]|^q] + \sigma_j^q),$$

but

$$\mathbb{E}[|(U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_1, W_1)]|^q] = \frac{A_q}{n},$$

with A_q defined in (19). Hence

$$\mathbb{E}[|B_l|^{\frac{q}{2}}] \leq C\left(\frac{A_q}{n} + \sigma_j^q \right). \tag{33}$$

Using the control of A_q in (21), equation (22) and Lemma 10 we have

$$\begin{aligned} A_q &\leq Cn\sigma_j^2 \|T_j\|_\infty^{q-2} \\ &\leq Cn 2^{S_j(q\nu+q-1)}. \end{aligned} \tag{34}$$

Now, we are able to apply the Rosenthal inequality to the variables B_l which yields

$$\mathbb{E}\left[\left(\sum_l B_l \right)^{\frac{q}{2}} \right] \leq C\left(\sum_l \mathbb{E}[|B_l|^{\frac{q}{2}}] + \left(\sum_l \mathbb{E}[B_l^2] \right)^{\frac{q}{4}} \right),$$

and using (33) and (34) we get

$$\mathbb{E}\left[\left(\sum_l B_l \right)^{\frac{q}{2}} \right] \leq C\left(\sum_l \left(\frac{A_q}{n} + \sigma_j^q \right) + \left(\sum_l \left(\frac{A_4}{n} + \sigma_j^4 \right) \right)^{\frac{q}{4}} \right)$$

$$\leq C\left(A_q + n\sigma_j^q + (A_4)^{\frac{q}{4}} + n^{\frac{q}{4}}\sigma_j^q\right)$$
$$\leq C\left(n2^{S_j(q\nu+q-1)} + n2^{S_j(2\nu+1)\frac{q}{2}} + (n2^{S_j(4\nu+3)})^{\frac{q}{4}}\right).$$

Consequently

$$\mathbb{E}[\hat{\sigma}_j^q] \leq Cn^{-\frac{q}{2}}\left(n2^{S_j(q\nu+q-1)} + n2^{S_j(2\nu+1)\frac{q}{2}} + (n2^{S_j(4\nu+3)})^{\frac{q}{4}} + n^{\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{2}}\right)$$
$$\leq C(n^{1-\frac{q}{2}}2^{S_j(q\nu+q-1)} + n^{1-\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{2}} + n^{-\frac{q}{4}}2^{S_j(4\nu+3)\frac{q}{4}} + 2^{S_j(2\nu+1)\frac{q}{2}}).$$

Let us compare each term of the r.h.s of the last inequality. We have

$$n^{1-\frac{q}{2}}2^{S_j(q\nu+q-1)} \leq 2^{S_j(2\nu+1)\frac{q}{2}} \iff 2^{S_j} \leq n,$$

which is true by (12). Similarly we have

$$n^{-\frac{q}{4}}2^{S_j(4\nu+3)\frac{q}{4}} \leq 2^{S_j(2\nu+1)\frac{q}{2}} \iff 2^{S_j} \leq n,$$

and obviously

$$n^{1-\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{2}} \leq 2^{S_j(2\nu+1)\frac{q}{2}}.$$

Thus we get that the dominant term in r.h.s is $2^{S_j(2\nu+1)\frac{q}{2}}$. Hence

$$\mathbb{E}[\hat{\sigma}_j^q] \leq C2^{S_j(2\nu+1)\frac{q}{2}}.$$

Now using that

$$\mathbb{E}[\tilde{\sigma}_{j,\tilde{\gamma}}^q] \leq C\left(\mathbb{E}[\hat{\sigma}_j^q] + \left(2C_j\sqrt{2\tilde{\gamma}\frac{\log n}{n}}\right)^{\frac{q}{2}}\mathbb{E}[\hat{\sigma}_j^{\frac{q}{2}}] + \left(8\tilde{\gamma}C_j^2\frac{\log n}{n}\right)^{\frac{q}{2}}\right),$$

and since $C_j \leq C\sqrt{\log n}2^{S_j(\nu+1)}$, we have

$$\mathbb{E}[\tilde{\sigma}_{j,\tilde{\gamma}}^q] \leq C\Bigg(2^{S_j(2\nu+1)\frac{q}{2}} + ((\log n)n^{-\frac{1}{2}}2^{S_j(\nu+1)})^{\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{4}}$$
$$+ \left(\frac{\log^2 n}{n}2^{2S_j(\nu+1)}\right)^{\frac{q}{2}}\Bigg).$$

Let us compare the three terms of the right hand side. We have

$$2^{S_j\frac{q(2\nu+1)}{2}} \geq ((\log n)n^{-\frac{1}{2}}2^{S_j(\nu+1)})^{\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{4}}$$
$$\iff 2^{S_j(q\nu+\frac{q}{2})} \geq (\log n)^{\frac{q}{2}}n^{-\frac{q}{4}}2^{S_j(q\nu+\frac{3q}{4})} \iff 2^{S_j} \leq \frac{n}{\log^2 n},$$

which is true by (12). Furthermore we have

$$2^{S_j\frac{q(2\nu+1)}{2}} \geq \left(\frac{\log^2 n}{n}2^{2S_j(\nu+1)}\right)^{\frac{q}{2}}$$

$$\iff 2^{S_j(q\nu+\frac{q}{2})} \geq \left(\frac{\log^2 n}{n}\right)^{\frac{q}{2}} 2^{S_j(q\nu+q)} \iff 2^{S_j} \leq \frac{n}{\log^2 n}, \tag{35}$$

which is true again by (12). Consequently

$$\mathbb{E}[\tilde{\sigma}^q_{j,\tilde{\gamma}}] \leq R_5 2^{S_j(2\nu+1)\frac{q}{2}},$$

with R_5 a constant depending on $q, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g$ and the lemma is proved for $q \geq 2$.

For the case $q \leq 2$ the result follows from Jensen inequality. □

Lemma 7. *Under Assumption (A1) on the father wavelet φ, we have for any $j = (j_1, \ldots, j_d)$ and any $x \in \mathbb{R}^d$,*

$$\sum_k |\varphi_{jk}(x)| \leq (2A+1)^d \|\varphi\|^d_\infty 2^{\frac{S_j}{2}}.$$

Proof. A proof of this standard result can be found in Section 4.2 in [13]. □

Lemma 8. *Under condition (A1) and φ is $\mathcal{C}^r$, there exist constants R_6 and R_7 depending on φ such that*

$$|\mathcal{F}(\varphi)(t)| \leq R_6(1+|t|)^{-r}, \quad \textit{for any } \ t \in \mathbb{R}. \tag{36}$$

and

$$\left|\overline{\mathcal{F}(\varphi)(t)}'\right| \leq R_7(1+|t|)^{-r}, \quad \textit{for any } \ t \in \mathbb{R}. \tag{37}$$

Proof. A proof of this result can be found in Section 4.2 of [13]. □

Lemma 9. *Under conditions (A1) and (A3), for $\nu \geq 0$, we have*

$$|(\mathcal{D}_j\varphi)(w)| \leq R_8 2^{S_j\nu} \prod_{l=1}^d (1+|w_l|)^{-1}, \ w \in \mathbb{R}^d$$

where R_8 is a constant depending on φ, $\mathcal{C}_g$ and c_g.

Proof. If all the $|w_l| < 1$ then using (3), Lemma 8 and $r > \nu + 1$ with $\nu \geq 0$ we have

$$|(\mathcal{D}_j\varphi)(w)| \leq \int \frac{\prod_{l=1}^d |\mathcal{F}(\varphi)(t_l)|}{|\mathcal{F}(g)(2^j t)|} dt \tag{38}$$

$$\leq C \prod_{l=1}^d \int \left|\mathcal{F}(\varphi)(t_l)(1+2^{j_l}|t_l|)^{\nu_l}\right| dt_l \tag{39}$$

$$\leq C 2^{S_j\nu} \prod_{l=1}^d \int (1+|t_l|)^{\nu_l - r} dt_l \tag{40}$$

$$\leq C 2^{S_j\nu} \leq C 2^{S_j\nu} \prod_{l=1}^d (1+|w_l|)^{-1}. \tag{41}$$

Now we consider the case where there exists at least one w_l such that $|w_l| \geq 1$. We have

$$(\mathcal{D}_j\varphi)(w) = \prod_{l=1,|w_l|\leq 1}^{d} \int e^{-it_l w_l} \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)} dt_l \times \prod_{l=1,|w_l|\geq 1}^{d} \int e^{-it_l w_l} \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)} dt_l.$$

For the left-hand product on $|w_l| \leq 1$ we use the result (41). Now let us consider the right-hand product with $|w_l| \geq 1$. We set in the sequel

$$\eta_l(t_l) := \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}.$$

We have

$$\prod_{l=1,|w_l|\geq 1}^{d} \int e^{-it_l w_l} \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)} dt_l = \prod_{l=1,|w_l|\geq 1}^{d} \int e^{-it_l w_l} \eta_l(t_l) dt_l.$$

Since $|\eta_l(t_l)| \to 0$ when $t_l \to \pm\infty$, an integration by part yields

$$\int e^{-it_l w_l} \eta_l(t_l) dt_l = i w_l^{-1} \int e^{-it_l w_l} \eta_l'(t_l) dt_l.$$

Let us compute the derivative of $\eta_l(t_l)$

$$\eta_l'(t_l) = \frac{\overline{\mathcal{F}(\varphi)(t_l)}' \mathcal{F}(g)(2^{j_l}t_l) - 2^{j_l} \mathcal{F}'(g)(2^{j_l}t_l) \overline{\mathcal{F}(\varphi)(t_l)}}{(\mathcal{F}(g)(2^{j_l}t_l))^2}.$$

Using Lemma 8, (3) and (4)

$$\begin{aligned}
|\eta_l'(t_l)| &\leq \left| \frac{\overline{\mathcal{F}(\varphi)(t_l)}'}{\mathcal{F}(g)(2^{j_l}t_l)} \right| + 2^{j_l} \left| \frac{\mathcal{F}'(g)(2^{j_l}t_l)\mathcal{F}(\varphi)(t_l)}{(\mathcal{F}(g)(2^{j_l}t_l))^2} \right| \\
&\leq C \left((1+|t_l|)^{-r}(1+2^{j_l}|t_l|)^{\nu} + 2^{j_l}(1+2^{j_l}|t_l|)^{-\nu-1}(1+|t_l|)^{-r}(1+2^{j_l}|t_l|)^{2\nu} \right) \\
&\leq C \left(2^{j_l\nu}(1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + 2^{j_l}(1+2^{j_l}|t_l|)^{\nu-1}(1+|t_l|)^{-r} \right) \\
&\leq C \left(2^{j_l\nu}(1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + 2^{j_l\nu}(2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r} \right) \\
&\leq C 2^{j_l\nu} \left((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r} \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left| \int e^{-it_l w_l} \eta_l(t_l) dt_l \right| \\
&\quad \leq |w_l|^{-1} \int |\eta_l'(t_l)| dt_l \\
&\quad \leq C|w_l|^{-1} 2^{j_l\nu} \int \left((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r} \right) dt_l \\
&\quad \leq C|w_l|^{-1} 2^{j_l\nu}(D_1+D_2+D_3),
\end{aligned}$$

with D_1, D_2 and D_3 defined below.

$$\begin{aligned} D_1 &:= \int_{|t_l|\le 2^{-j_l}} \big((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\big)\,dt_l \\ &\le C\int_{|t_l|\le 2^{-j_l}} \big((2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}\big)\,dt_l \\ &\le C2^{-j_l}(2^{-j_l\nu}+2^{-j_l(\nu-1)}) \\ &\le C. \end{aligned}$$

$$\begin{aligned} D_2 &:= \int_{2^{-j_l}\le|t_l|\le 1} \big((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\big)\,dt_l \\ &\le C\int_{2^{-j_l}\le|t_l|\le 1} \big((2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}\big)\,dt_l \\ &\le C\int_1^{2^{j_l}} ((2^{-j_l}+2^{-j_l}s)^{\nu} + (2^{-j_l}+2^{-j_l}s)^{\nu-1})2^{-j_l}ds \\ &\le C2^{-j_l(\nu+1)}\int_1^{2^{j_l}} s^{\nu}ds + C2^{-j_l\nu}\int_1^{2^{j_l}} s^{\nu-1}ds \\ &\le C, \end{aligned}$$

as soon as $\nu > 0$.

$$\begin{aligned} D_3 &:= \int_{|t_l|\ge 1} \big((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\big)\,dt_i \\ &\le C\int_{|t_l|\ge 1} \big(|t_l|^{\nu-r} + |t_l|^{\nu-1-r}\big)\,dt_l \\ &\le C, \end{aligned}$$

since $\nu - r < -1$.

When $\nu = 0$ we still have

$$\left|\int e^{-it_lw_l}\eta_l(t_l)dt_l\right| \le C|w_l|^{-1}2^{j_l\nu} = C|w_l|^{-1}.$$

Indeed when $\nu = 0$

$$\eta_l(t_l) = \overline{\mathcal{F}(\varphi)(t_l)},$$

and

$$\begin{aligned} \left|iw_l^{-1}\int e^{-it_lw_l}\eta_l'(t_l)dt_l\right| &= \left|iw_l^{-1}\int e^{-it_lw_l}\overline{\mathcal{F}(\varphi)(t_l)}'dt_l\right| \\ &\le |w_l|^{-1}\int\left|\overline{\mathcal{F}(\varphi)(t_l)}'\right|dt_l \\ &\le C|w_l|^{-1}\int(1+|t|)^{-r}dt < C|w_l|^{-1}, \end{aligned}$$

using Lemma 8 and $r \ge 2$. □

Lemma 10. *There exist constants R_{10} depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ, c_g, $\mathcal{C}_g$ and R_{11} depending on φ, c_g, $\mathcal{C}_g$ such that*

$$\sigma_j^2 \le R_{10} 2^{S_j(2\nu+1)}, \quad \|T_j\|_\infty \le R_{11} 2^{S_j(\nu+1)}.$$

Proof. We have

$$\begin{aligned}
\sigma_j^2 &= \operatorname{Var}(U_j(Y_1, W_1)) \\
&\le \mathbb{E}\left[|U_j(Y_1,W_1)|^2\right] \\
&= \mathbb{E}\left[\left|Y_1 \sum_k (\mathcal{D}_j\varphi)_{j,k}(W_1)\varphi_{jk}(x)\right|^2\right] \\
&= \mathbb{E}\left[\left|(m(X_1)+\varepsilon_1) \sum_k (\mathcal{D}_j\varphi)_{j,k}(W_1)\varphi_{jk}(x)\right|^2\right] \\
&\le 2(\|m\|_\infty^2 + \sigma_\varepsilon^2)\mathbb{E}\left[\left|\sum_k (\mathcal{D}_j\varphi)_{j,k}(W_1)\varphi_{jk}(x)\right|^2\right] \\
&\le 2(\|m\|_\infty^2 + \sigma_\varepsilon^2)\int \left|\sum_k (\mathcal{D}_j\varphi)_{j,k}(w)\varphi_{jk}(x)\right|^2 f_W(w)dw \\
&\le 2(\|m\|_\infty^2 + \sigma_\varepsilon^2)\|f_X\|_\infty \int 2^{S_j}\left|\sum_k (\mathcal{D}_j\varphi)(2^j w - k)\varphi_{jk}(x)\right|^2 dw.
\end{aligned}$$

Now making the change of variable $z = 2^j w - k$, we get using Lemma 7 and Lemma 9 to bound $(\mathcal{D}_j\varphi)(z)$

$$\begin{aligned}
\sigma_j^2 &\le 2(\|m\|_\infty^2 + \sigma_\varepsilon^2)\|f_X\|_\infty \int \left|\sum_k (\mathcal{D}_j\varphi)(z)\varphi_{jk}(x)\right|^2 dz \\
&\le C\int 2^{2S_j\nu}\prod_{i=l}^{d}\frac{1}{(1+|z_l|)^2}\left(\sum_k|\varphi_{jk}(x)|\right)^2 dz \\
&\le R_{10}2^{S_j(2\nu+1)},
\end{aligned}$$

where R_{10} is a constant depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ, c_g, $\mathcal{C}_g$. This gives the bound for σ_j^2.

For $\|T_j\|_\infty$, using again Lemma 7 and Lemma 9, we have

$$\begin{aligned}
\|T_j\|_\infty &\le \max_k \|(\mathcal{D}_j\varphi)_{j,k}\|_\infty \sum_k |\varphi_{jk}(x)| \le 2^{\frac{S_j}{2}}\|(\mathcal{D}_j\varphi)\|_\infty \sum_k |\varphi_{jk}(x)| \\
&\le R_{11}2^{S_j(\nu+1)},
\end{aligned}$$

where R_{11} is a constant depending on φ, c_g, $\mathcal{C}_g$. □

Acknowledgements

We would like to thank three anonymous referees for their helpful suggestions which helped us to improve the presentation of the paper. The research of Thanh Mai Pham Ngoc and Vincent Rivoirard is partly supported by the french Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration). Michaël Chichignoud now works at Winton Capital Management, supported in part as member of the German-Swiss Research Group FOR916 (Statistical Regularization and Qualitative Constraints) with grant number 20PA20E-134495/1.

References

[1] Bertin, K., Lacour, C. and Rivoirard, V. (2016). Adaptive pointwise estimation of conditional density function. *Les Annales de l'IHP* **2** 939-980. MR3498017

[2] Carroll, R. J., Delaigle, A. and Hall, P. (2009). Nonparametric prediction in measurement error models. *J. Amer. Statist. Assoc.* **104** 993–1003. MR2562002 (2011b:62073)

[3] Chesneau, C. (2010). On adaptive wavelet estimation of the regression function and its derivatives in an errors-in-variables model. *Curr. Dev. Theory Appl. Wavelets* **4** 185–208. MR2788969 (2011m:62112)

[4] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 569–609. MR3088382

[5] Comte, F. and Taupin, M. L. (2007). Adaptive estimation in a nonparametric regression model with errors-in-variables. *Statist. Sinica* **17** 1065–1090. MR2397388 (2009e:62167)

[6] Delaigle, A., Hall, P. and Jamshidi, F. (2015). Confidence bands in non-parametric errors-in-variables regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 149–169. MR3299403

[7] Doumic, M., Hoffmann, M., Reynaud-Bouret, P. and Rivoirard, V. (2012). Nonparametric estimation of the division rate of a size-structured population. *SIAM J. Numer. Anal.* **50** 925–950. MR2914292

[8] Du, L., Zou, C. and Wang, Z. (2011). Nonparametric regression function estimation for errors-in-variables models with validation data. *Statist. Sinica* **21** 1093–1113. MR2827516 (2012h:62153)

[9] Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory* **48** 734–747. MR1889978 (2003e:94017)

[10] Fan, J. and Masry, E. (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.* **43** 237–271. MR1193614 (94d:62121)

[11] Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925. MR1245773 (94i:62081)

[12] Gach, F., Nickl, R. and Spokoiny, V. (2013). Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 900–914. MR3112439

[13] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-dimensional Statistical Models*. Cambridge University Press. MR1618204 (99f:42065)

[14] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214 (2012k:62091)

[15] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*. *Lecture Notes in Statistics* **129**. Springer-Verlag, New York. MR1618204 (99f:42065)

[16] Houdré, C. and Reynaud-Bouret, P. (2003). Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic inequalities and applications*. *Progr. Probab.* **56** 55–69. Birkhäuser, Basel. MR2073426 (2005d:60031)

[17] Ioannides, D. A. and Alevizos, P. D. (1997). Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.* **32** 35–43. MR1439495 (98f:62097)

[18] Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. MR1863916

[19] Koo, J.-Y. and Lee, K.-W. (1998). *B*-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.* **40** 57–66. MR1650520 (99j:62047)

[20] Massart, P. (2007). *Concentration Inequalities and Model Selection. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, 6-23, 2003*. Springer.

[21] Meister, A. (2009). *Deconvolution problems in nonparametric statistics*. *Lecture Notes in Statistics* **193**. Springer-Verlag, Berlin. MR2768576 (2012h:62001)

[22] Patel, J. K. and Read, C. B. (1982). *Handbook of the normal distribution*. *Statistics: Textbooks and Monographs* **40**. Marcel Dekker, Inc., New York. MR664762 (83j:62002)

[23] Whittemore, A. S. and Keller, J. B. (1988). Approximations for regression with covariate measurement error. *J. Amer. Statist. Assoc.* **83** 1057–1066. MR997582 (90f:62225)