



**HAL**  
open science

# Adaptive wavelet multivariate regression with errors in variables

Michaël Chichignoud, van Ha Hoang, Thanh Mai Pham Ngoc, Vincent Rivoirard

► **To cite this version:**

Michaël Chichignoud, van Ha Hoang, Thanh Mai Pham Ngoc, Vincent Rivoirard. Adaptive wavelet multivariate regression with errors in variables. 2016. hal-01253508v1

**HAL Id: hal-01253508**

**<https://hal.science/hal-01253508v1>**

Preprint submitted on 11 Jan 2016 (v1), last revised 17 Mar 2017 (v2)

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

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

# Adaptive wavelet multivariate regression with errors in variables

Michaël Chichignoud <sup>\*</sup>, Van Ha Hoang <sup>†</sup>, Thanh Mai Pham Ngoc <sup>‡</sup> and Vincent Rivoirard <sup>§</sup>

January 11, 2016

## 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 estimator 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.

**Keywords :** Adaptive wavelet estimator. Anisotropic regression. Deconvolution. Measurement errors.

**Primary subjects.** 62G08

## 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  $\delta_l$  are i.i.d unobservable random variables having error density  $g$ . We assume that  $g$  is known. The  $\delta_l$ 's are independent of the  $X_l$ 's and  $Y_l$ 's. The  $\varepsilon_l$ 's are i.i.d standard normal random variables with variance  $s^2$ . 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 [Whittemore and Keller \(1988\)](#), [Fan and Masry \(1992\)](#)) or in environmental science where air quality is measured with errors ([Delaigle et al. \(2015\)](#)).

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

---

<sup>\*</sup>ETH Zürich

<sup>†</sup>Laboratoire Paul Painlevé UMR CNRS 8524, Université Lille 1- Sciences et Technologies.

<sup>‡</sup>Laboratoire de Mathématiques, UMR 8628, Université Paris Sud.

<sup>§</sup>CEREMADE, UMR CNRS 7534, Université Paris Dauphine.

Meister (2009) 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 Fan and Masry (1992), Fan and Truong (1993), Ioannides and Alevizos (1997), Koo and Lee (1998), Meister (2009), Comte and Taupin (2007), Chesneau (2010), Du et al. (2011), Carroll et al. (2009), Delaigle et al. (2015). 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 Fan and Masry (1992) where the authors explored the asymptotic normality for mixing processes. In the one dimensional setting, Chesneau (2010) 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  $[0, 1]$ . Comte and Taupin (2007) 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 taking into account the noise in the covariates. When using wavelets, a crucial point lies in the choice of the resolution level. But 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 Gach et al. (2013) 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 and fully data-driven. It is automatically selected by a method inspired from Goldenshluger and Lepski (2011) to tackle anisotropy problems. This method has been used recently in various contexts (see Doumic et al. (2012), Comte and Lacour (2013) and Bertin et al. (2013)). 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  $k = (k_1, \dots, k_d) \in \mathbb{Z}^d$ ,

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

for any given function  $h$ . We denote by  $\mathcal{F}$  the Fourier transform of any function  $f$  defined on  $\mathbb{R}^d$  by

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

where  $\langle \cdot, \cdot \rangle$  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  $y$ :  $\lfloor y \rfloor \leq y < \lfloor y \rfloor + 1$ .

## 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

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

So, 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 2.1 describes the construction of the projection kernel on wavelet bases depending on a maximal resolution level. Section 2.2 describes the Goldenshluger-Lepski procedure to select the resolution level adaptively.

### 2.1 Approximation kernels and family of estimators for $p$

We consider noise densities  $g = (g_1, \dots, g_d)$  which satisfy the following relationship (see [Fan and Koo \(2002\)](#)) :

$$\mathcal{F}(g)(t) = \prod_{l=1}^d \mathcal{F}(g_l)(t_l), \quad t_l \in \mathbb{R}. \quad (1)$$

In the sequel, we consider a father wavelet  $\varphi$  on the real line satisfying the following conditions:

- (A1) The father wavelet  $\varphi$  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)  $\varphi$  is of class  $\mathcal{C}^r$ , where  $r \geq 2$ .

These properties are satisfied for instance by Daubechies and Coiflets wavelets (see [Härdle et al. \(1998\)](#), chapter 8). The associated projection kernel on the space

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

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 any 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 2.2. The natural approach is based on unbiased estimation of the projection coefficients  $p_{jk}$ . To do so, we adapt the kernel approach proposed by [Fan and Truong \(1993\)](#) in our wavelets context. To this purpose, we set

$$\hat{p}_{jk} := \frac{1}{n} \sum_{u=1}^n Y_u \times (\mathcal{D}_j \varphi)_{j,k}(W_u) = 2^{\frac{s_j}{2}} \frac{1}{n} \sum_{u=1}^n Y_u \int e^{-i \langle t, 2^j W_u - k \rangle} \prod_{l=1}^d \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l} t_l)} dt,$$

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

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

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

Lemma 3, proved in section 5.2.1 states that  $\mathbb{E}[\hat{p}_j(x)] = p_j(x)$  which justifies our approach. Furthermore, the deconvolution operator  $(\mathcal{D}_j f)(w)$  in (2) is the multidimensional wavelet analogous of the operator  $K_n(x)$  defined in (2.4) in [Fan and Truong \(1993\)](#): the Fourier transform of their kernel  $K$  has been replaced in our procedure by the Fourier transform of the wavelet  $\varphi_{jk}$  and their bandwidth  $h$  by  $2^{-j}$ .

Note that 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  $\mathcal{F}(g)(t) = 1$ .

## 2.2 Selection rule by using the Goldenshluger-Lepski methodology

The second and final step consists in selecting the multidimensional resolution level  $j$  depending on  $x$  and based on a data-driven selection rule inspired from a method exposed in [Goldenshluger and Lepski \(2011\)](#). To define this latter we have to introduce some quantities. 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  $\sigma_j^2 := \text{Var}(U_j(Y_1, W_1))$  and the variance of  $\hat{p}_j$  is thus equal to  $\frac{\sigma_j^2}{n}$ . We set :

$$\hat{\sigma}_j^2 := \frac{1}{n(n-1)} \sum_{l=2}^n \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - U_j(Y_v, W_v))^2, \quad (3)$$

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

$$\tilde{\sigma}_{j,\tilde{\gamma}}^2 := \hat{\sigma}_j^2 + 2C_j \sqrt{2\tilde{\gamma} \hat{\sigma}_j^2 \frac{\log n}{n}} + 8\tilde{\gamma} C_j^2 \frac{\log n}{n}, \quad (4)$$

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

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

For any  $\varepsilon > 0$ , let  $\gamma > 0$  and

$$\Gamma_\gamma(j) := \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}_{j,\tilde{\gamma}}^2 \log n}{n}} + \frac{c_j \gamma \log n}{n},$$

where

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

Let

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

and

$$\Gamma_\gamma^*(j) := \sup_{j'} \Gamma_\gamma(j, j'). \quad (5)$$

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

$$\hat{R}_j := \sup_{j'} \left\{ |\hat{p}_{j \wedge j'}(x) - \hat{p}_{j'}(x)| - \Gamma_\gamma(j', j) \right\}_+ + \Gamma_\gamma^*(j). \quad (6)$$

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

$$\hat{j} := \arg \min_{j \in J} \hat{R}_j, \quad (7)$$

where the set  $J$  is defined as

$$J := \left\{ j \in \mathbb{N}^d : 2^{S_j} \leq \left\lfloor \frac{n}{\log^2 n} \right\rfloor \right\}. \quad (8)$$

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  $|\hat{p}_{j \wedge j'}(x) - \hat{p}_{j'}(x)|$  is classical when using the Goldenshluger Lepski method (see sections 2.1 and 5.2 in Bertin et al. (2013)). 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  $\varepsilon$ ,  $\tilde{\gamma}$  and  $\gamma$ . Actually we prove that  $\Gamma_\gamma^2(j)$  is of order  $\frac{\log n}{n} \sigma_j^2$  (see Lemma 6 and 10). The dependance of  $\tilde{\sigma}_{j,\tilde{\gamma}}^2$  (4) in  $\|m\|_\infty$  appears only in smaller order terms. In conclusion, up to the knowledge of  $\|m\|_\infty$  the procedure is completely data-driven. Next section explains how to choose the constants  $\gamma$  and  $\tilde{\gamma}$ . Our approach is non asymptotic and based on sharp concentration inequalities.

### 3 Rates of convergence

There exists  $C_1 > 0$  such that for any  $x \in [0, 1]^d$ ,  $f_X(x) \geq C_1$ .

As we face a deconvolution problem, we need to define the assumptions made on the smoothness of the density of the errors covariates  $g$ . There exist positive constants  $c_g$  and  $C_g$  such that

$$c_g(1 + |t_l|)^{-\nu} \leq |\mathcal{F}(g_l)(t_l)| \leq C_g(1 + |t_l|)^{-\nu}, \quad 0 \leq \nu \leq r - 2, \quad t_l \in \mathbb{R}. \quad (9)$$

We also require a condition for the derivative of the Fourier transform of  $g$ . There exists a positive constant  $\mathcal{C}_g$  such that

$$|\mathcal{F}'(g_l)(t_l)| \leq \mathcal{C}_g(1 + |t_l|)^{-\nu-1}, \quad t_l \in \mathbb{R}. \quad (10)$$

Laplace and Gamma distributions satisfy the above assumptions (9) and (10). Assumptions (9) and (10) control the decay of the Fourier transform of  $g$  at a polynomial rate. Hence we deal with a mildly ill-posed inverse problem. The index  $\nu$  is usually known as the degree of ill-posedness of the deconvolution problem at hand.

#### 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  $\tilde{\gamma} > 2q(\nu + 2)$ ,*

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

where

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

and  $R_1$  a constant depending only on  $q$ .

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  $|\mathbb{E}[\hat{p}_{\eta \wedge j'}(x)] - \mathbb{E}[\hat{p}_{j'}(x)]|$ , 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

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

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

We study the rates of convergence of the estimators over anisotropic Hölder Classes. 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  $f : [0, 1]^d \rightarrow \mathbb{R}$  belongs to the anisotropic Hölder class  $\mathbb{H}_d(\vec{\beta}, L)$  of functions if  $f$  is bounded and for any  $l = 1, \dots, d$  and for all  $z \in \mathbb{R}$*

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

The following theorem gives the rate of convergence of the estimator  $\hat{p}_{\hat{j}}(x)$  and justifies the optimality of our oracle inequality.

**Theorem 2.** *Let  $q \geq 1$  be fixed and let  $\hat{j}$  be the adaptive index defined in (7). Then, for any  $\vec{\beta} \in (0, 1]^d$  and  $L > 0$ , it holds*

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

with  $\vec{\beta} = \frac{1}{\frac{1}{\beta_1} + \dots + \frac{1}{\beta_d}}$  and  $R_2$  a constant depending on  $\gamma, q, \varepsilon, \tilde{\gamma}, \|m\|_\infty, s, \|f_X\|_\infty, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$ .

**Remark 1.** *The estimate  $\hat{p}$  achieves the optimal rate of convergence up to a logarithmic term (see section 3.3 in Comte and Lacour (2013)). This logarithmic loss, due to adaptation, is known to be nevertheless unavoidable for  $d = 1$  and one can conjecture that it is also the case for higher dimension (see Remark 1 in Comte and Lacour (2013)).*

### 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 Bertin et al. (2013)). For this purpose, we use an estimate introduced by Comte and Lacour (2013) (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 Goldenshluger and Lepski (2011). 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  $[0, 1]^d$ :

$$\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}}. \quad (11)$$

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, for any  $\vec{\beta} \in (0, 1]^d$  and  $L > 0$ , it holds*

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

with  $R_3$  a constant depending on  $\gamma, q, \varepsilon, \tilde{\gamma}, \|m\|_\infty, s, \|f_X\|_\infty, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$ .

The estimate  $\hat{m}$  is again optimal up to a logarithmic term (see Remark 1).

## 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  $\varepsilon_l$ '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  $x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$ . In Figure 2, we plot the noisy regression Doppler function according to the three design scenario. For



the covariate errors  $\delta_i$ '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 (9), 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 [Fan and Truong \(1993\)](#)) which is given by

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

$R_r$  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, [Fan and Truong \(1993\)](#) chose  $R_r = 0.70$ . We computed the reliability ratio in [Table 1](#) for the considered simulations.

$\sigma_{g_L}$	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

Table 1: Reliability ratio.

We recall that our estimator of  $m(x)$  is given by the ratio of two estimators (see (11)) :

$$\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}}. \quad (12)$$

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 [Comte and Lacour \(2013\)](#). 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  $\hat{\sigma}_{\hat{j}, \hat{\gamma}}^2$  instead of  $\hat{\sigma}_{\hat{j}, \hat{\gamma}}^2$  and  $\frac{2 \max_i \|Y_i\| \|T_j\|_\infty}{3}$  instead of  $c_j$ . It remains to settle the value of the constant  $\gamma$ . To do so, we compute the pointwise risk of  $\hat{p}_{\hat{j}}(x)$  in function of  $\gamma$ : [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.

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

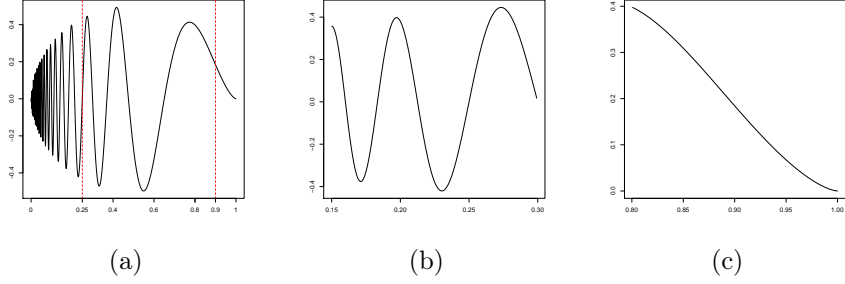


Figure 1: a/ Representation of Doppler function. b/ A zoom of Doppler function on  $[0.15, 0.30]$ . c/ A zoom of Doppler function on  $[0.80, 1]$ .

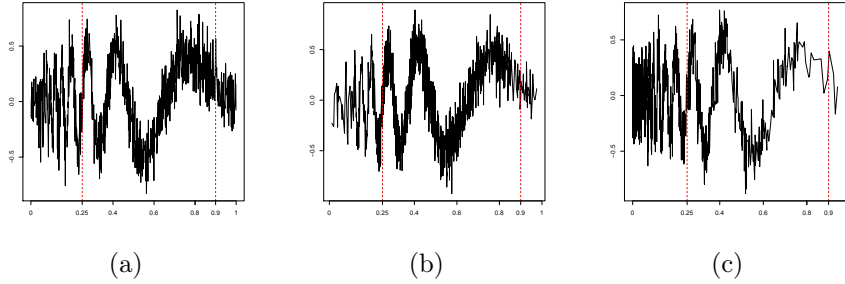


Figure 2: a/ Noisy Doppler with  $X_i \sim \mathcal{U}[0, 1]$ . b/ Noisy Doppler with  $X_i \sim \text{Beta}(2, 2)$ . c/ Noisy Doppler function with  $X_i \sim \text{Beta}(0.5, 2)$ .

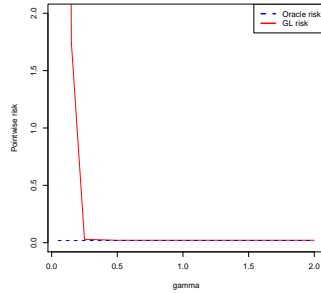


Figure 3: Pointwise risk of  $\hat{p}_j$  at  $x_0 = 0.25$  in function of parameter  $\gamma$  for the  $\text{Beta}(2, 2)$  design and  $\sigma_{g_L} = 0.075$ .

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  $\hat{p}_j(x)$  (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.

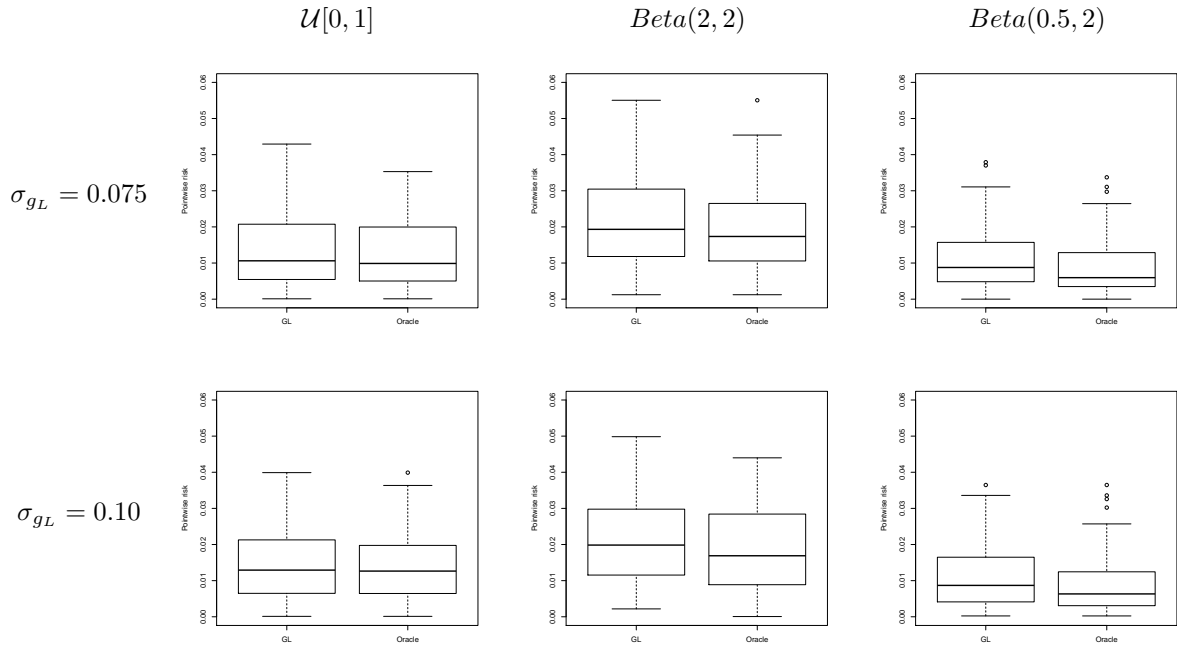


Figure 4: Estimation of  $p(x)$  at  $x_0 = 0.25$

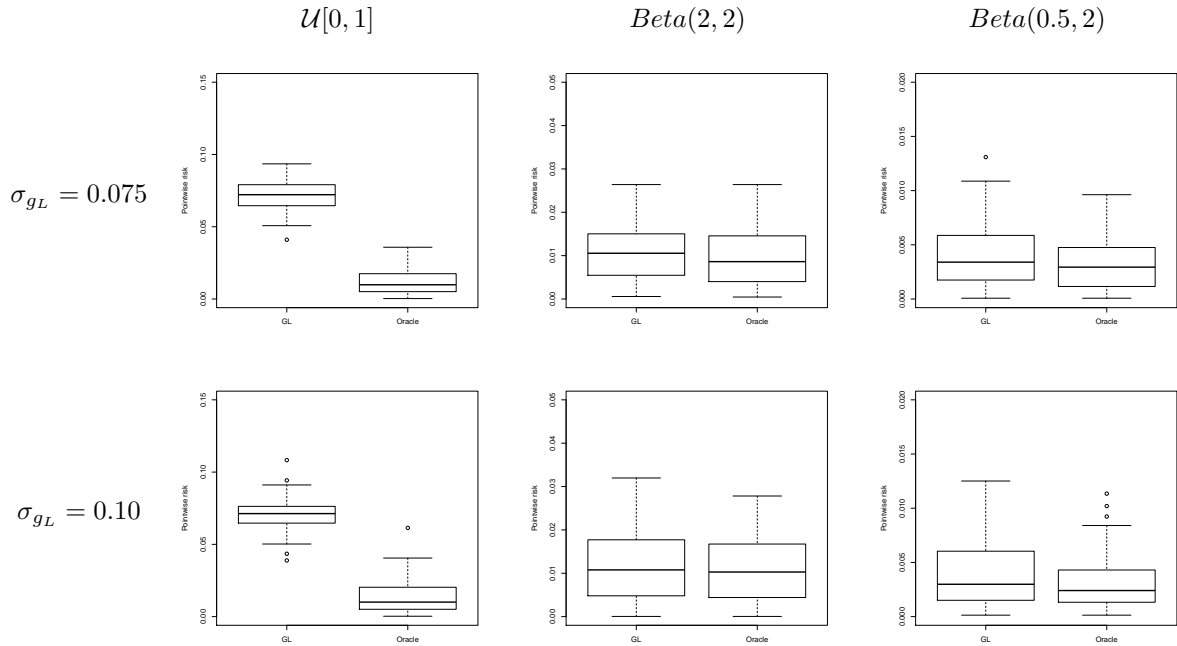


Figure 5: Estimation of  $p(x)$  at  $x_0 = 0.90$

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  $\sigma_{g_L}$  deteriorates slightly 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 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 2. 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 expected as the two densities charge points near 0 and 1 differently.

## 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  $(a_1 + \dots + a_p)^q \leq p^{q-1}(a_1^q + \dots + a_p^q)$  for all  $a_1, \dots, a_p \in \mathbb{R}_+^p$ ,  $p \in \mathbb{N}$  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  $\eta \in \mathbb{N}^d$ :

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

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

$$\hat{R}_{\eta} \leq \sup_{j, j'} \left\{ |\hat{p}_{j \wedge j'} - \mathbb{E}[\hat{p}_{j \wedge j'}]| - \Gamma_{\gamma}(j \wedge j') \right\}_+ + \sup_{j'} \left\{ |\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]| - \Gamma_{\gamma}(j') \right\}_+ + \sup_{j'} |\mathbb{E}[\hat{p}_{\eta \wedge j'}] - \mathbb{E}[\hat{p}_{j'}]| + \Gamma_{\gamma}^*(\eta).$$

Hence

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

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

$$|\hat{p}_{\hat{j}} - p(x)| \leq 2 \sup_{j, j'} \left\{ |\hat{p}_{j \wedge j'} - \mathbb{E}[\hat{p}_{j \wedge j'}]| - \Gamma_{\gamma}(j \wedge j') \right\}_+ + 3 \sup_{j'} \left\{ |\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]| - \Gamma_{\gamma}(j') \right\}_+ + 3B(\eta) + 3\Gamma_{\gamma}^*(\eta)$$

Consequently

$$|\hat{p}_{\hat{j}} - p(x)|^q \leq 3^{2q-1} \left( [B(\eta) + \Gamma_{\gamma}^*(\eta)]^q + \sup_{j'} \left\{ |\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]| - \Gamma_{\gamma}(j') \right\}_+^q + \sup_{j, j'} \left\{ |\hat{p}_{j \wedge j'} - \mathbb{E}[\hat{p}_{j \wedge j'}]| - \Gamma_{\gamma}(j \wedge j') \right\}_+^q \right).$$

Using Proposition 2, we have

$$\mathbb{E} \left| \hat{p}_j - p(x) \right|^q \leq C \left( \mathbb{E} \left[ (B(\eta) + \Gamma_\gamma^*(\eta))^q \right] \right) + o(n^{-q}).$$

Then, we get

$$\mathbb{E} \left| \hat{p}_j - p(x) \right|^q \leq R_1 \left( \inf_\eta \mathbb{E} \left[ (B(\eta) + \Gamma_\gamma^*(\eta))^q \right] \right) + o(n^{-q}),$$

where  $R_1$  is a constant only depending on  $q$ .

□

### 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:

$$\mathbb{E} \left[ \left| \hat{p}_j(x) - p(x) \right|^q \right] \leq C \left( \inf_\eta \mathbb{E} \left[ (B(\eta) + \Gamma_\gamma^*(\eta))^q \right] \right) + o(n^{-q}).$$

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

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

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

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

using Lemma 6. But

$$c_\eta = 16(2\|m\|_\infty + s) \|T_\eta\|_\infty \leq C 2^{S_\eta(\nu+1)},$$

using Lemma 10. Hence

$$\mathbb{E} \left[ \Gamma_\gamma(\eta)^q \right] \leq C \left( \left( \frac{\log n}{n} \right)^{\frac{q}{2}} 2^{S_\eta(2\nu+1)\frac{q}{2}} + \left( \frac{\log n}{n} \right)^q 2^{S_\eta(\nu+1)q} \right).$$

We have

$$\left( \frac{\log n}{n} \right)^{\frac{q}{2}} 2^{S_\eta(2\nu+1)\frac{q}{2}} \geq \left( \frac{\log n}{n} \right)^q 2^{S_\eta(\nu+1)q} \iff 2^{S_\eta} \leq \frac{n}{\log n},$$

which is true since by (8),  $2^{S_\eta} \leq \frac{n}{\log^2 n}$ .

This yields

$$\mathbb{E} \left[ \Gamma_\gamma^*(\eta)^q \right] \leq C \left( \frac{2^{S_\eta(2\nu+1)} \log n}{n} \right)^{\frac{q}{2}}.$$

Eventually, we obtain the bound for the pointwise risk:

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

Setting the gradient of the right hand side of the inequality above with respect to  $\eta$  it turns out that the optimal  $\eta_l$  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

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

with  $R_2$  a constant depending on  $\gamma, q, \varepsilon, \tilde{\gamma}, \|m\|_\infty, s, \|f_X\|_\infty, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$ . The proof of Theorem 2 is completed.  $\square$

### 5.1.3 Proof of Theorem 3

**Proof.** We recall that  $m(x) = \frac{p(x)}{f_X(x)}$  and  $\hat{m}(x) = \frac{\hat{p}_j(x)}{\hat{f}_X(x) \vee n^{-1/2}}$ . We now state the main properties of the adaptive estimate  $\hat{f}_X$  showed by Comte and Lacour (2013) (Theorem 2): for all  $q \geq 1$ , all  $\vec{\beta} \in (0, 1]^d$ , all  $L > 0$  and  $n$  large enough, it holds

$$\mathbb{P}(E_1) := \mathbb{P} \left( |\hat{f}_X(x) - f_X(x)| \geq C \phi_n(\vec{\beta}) \right) \leq n^{-2q}, \quad (13)$$

and

$$\mathbb{P} \left( |\hat{f}_X(x) - f_X(x)| \leq Cn \right) = 1, \quad (14)$$

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 Comte and Lacour (2013) Section 3.4 for further details). From (13), we easily deduce, since  $f_X(x) \geq C_1 > 0$ , for  $n$  large enough that

$$\mathbb{P}(E_2) := \mathbb{P} \left( \hat{f}_X(x) < \frac{C_1}{2} \right) \leq n^{-2q}. \quad (15)$$

We now start the proof of the theorem. We have together with (14)

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

**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

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

Then, using Theorem 2 and (15), 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 (14) and the inequality  $\hat{f}_X(x) \vee n^{-1/2} \geq n^{-1/2}$ , it holds for  $n$  large enough

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

Then, using the definition of  $\mathcal{A}_2$ , (13) and (15), 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  $\gamma, q, \varepsilon, \tilde{\gamma}, \|m\|_\infty, s, \|f_X\|_\infty, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$ . This completes the proof of Theorem 3.  $\square$

## 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  $u > 0$ ,*

$$\mathbb{P} \left( \left| \hat{p}_j(x) - p_j(x) \right| \geq \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n} \right) \leq 2e^{-u}, \quad (16)$$

where

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

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 (4).

**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

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

we have for any  $q \geq 2$ ,

$$A_q := \sum_{l=1}^n \mathbb{E}[|U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]|^q] = \sum_{l=1}^n \mathbb{E}[|m(X_l)T_j(W_l) + \varepsilon_l T_j(W_l) - \mathbb{E}[m(X_l)T_j(W_l)]|^q]. \quad (17)$$

With  $q = 2$ ,

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

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

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

Furthermore,

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

Finally,

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

Besides we have (see page 23 in [Patel and Read \(1982\)](#)) denoting  $\Gamma$  the Gamma function

$$\mathbb{E}[|Z|^q] = \frac{2^{q/2}}{\sqrt{\pi}} \Gamma\left(\frac{q+1}{2}\right) \leq 2^{q/2} 2^{-1/2} q! \leq 2^{(q-1)/2} q!, \quad (18)$$

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

$$A_q \leq \frac{q!}{2} \times A_2 \times c_j^{q-2}, \quad (19)$$

with

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

We can now apply Proposition 2.9 of Massart (2007). 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

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

Therefore, with

$$\sigma_j^2 = \frac{A_2}{n} = \text{Var}(Y_1 T_j(W_1)), \quad (20)$$

$$\sigma_j^2 = \sigma_\varepsilon^2 \mathbb{E}[T_j^2(W_1)] + \text{Var}(m(X_1)T_j(W_1)) \quad (21)$$

$$\begin{aligned}&\leq \sigma_\varepsilon^2 \|f_X\|_\infty \|T_j\|_2^2 + \|m\|_\infty^2 \|f_X\|_\infty \|T_j\|_2^2 \\ &\leq \|f_X\|_\infty \|T_j\|_2^2 (\sigma_\varepsilon^2 + \|m\|_\infty^2).\end{aligned} \quad (22)$$

We conclude that for any  $u > 0$ ,

$$\mathbb{P}\left(|\hat{p}_j(x) - p_j(x)| \geq \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}\right) \leq 2e^{-u}. \quad (23)$$

Now, we can write

$$\begin{aligned}\hat{\sigma}_j^2 &= \frac{1}{n(n-1)} \sum_{l=2}^n \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - U_j(Y_v, W_v))^2 \\ &= \frac{1}{n(n-1)} \sum_{l=2}^n \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)] - U_j(Y_v, W_v) + \mathbb{E}[U_j(Y_v, W_v)])^2 \\ &= s_j^2 - \frac{2}{n(n-1)} \xi_j,\end{aligned}$$

with

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

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  $\tilde{\gamma} > 0$ ,

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

We have that

$$\mathbb{P}(\Omega_n(\tilde{\gamma})^c) \leq n^{1-\tilde{\gamma}}. \quad (24)$$

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

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

we recall that

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

**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

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

**Proof.**

We denote

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

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

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

Furthermore, as

$$\begin{aligned} \mathbb{E}_{\Omega_n(\tilde{\gamma})}[U_j(Y_1, W_1)] &= \mathbb{E}[m(X_1)T_j(W_1) | \Omega_n(\tilde{\gamma})] + \mathbb{E}[\varepsilon_1 T_j(W_1) | \Omega_n(\tilde{\gamma})] \\ &= \mathbb{E}[m(X_1)T_j(W_1)] \\ &= \mathbb{E}[U_j(Y_1, W_1)] \end{aligned} \tag{25}$$

we get

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

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

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

where  $\tilde{\epsilon}_n = n^{1-\tilde{\gamma}} + o(n^{1-\tilde{\gamma}})$ . Using (21) we have

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})} \left[ (U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2 \right] \leq (1 + e_{n,j}) \sigma_j^2, \tag{26}$$

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

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

$$\begin{aligned} \mathbb{E}_{\Omega_n(\tilde{\gamma})} [(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2] &= \text{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)] \frac{\mathbb{E}[\varepsilon_1^2 \mathbf{1}_{\Omega_n(\tilde{\gamma})}]}{\mathbb{P}(\Omega_n(\tilde{\gamma}))} \\ &\geq \text{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)] \mathbb{E}[\varepsilon_1^2 \mathbf{1}_{\Omega_n(\tilde{\gamma})}] \\ &= \text{Var}(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)] \mathbb{E}[\varepsilon_1^2 (1 - \mathbf{1}_{\Omega_n^c(\tilde{\gamma})})] \\ &= \sigma_j^2 - \mathbb{E}[T_j^2(W_1)] \mathbb{E}[\varepsilon_1^2 \mathbf{1}_{\Omega_n^c(\tilde{\gamma})}]. \end{aligned}$$

Now using Cauchy Scharwz, (18) and (24) we have

$$\begin{aligned} \mathbb{E}_{\Omega_n(\tilde{\gamma})} [(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2] &\geq \sigma_j^2 - \mathbb{E}[T_j^2(W_1)] (\mathbb{E}[\varepsilon_1^4])^{\frac{1}{2}} (\mathbb{P}(\Omega_n^c(\tilde{\gamma})))^{\frac{1}{2}} \\ &\geq \sigma_j^2 - Cs^2 \mathbb{E}[T_j^2(W_1)] n^{\frac{1-\tilde{\gamma}}{2}} \\ &= \sigma_j^2 (1 + \tilde{e}_{n,j}), \end{aligned} \tag{27}$$

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  $\mathbb{E}_{\Omega_n(\tilde{\gamma})} [(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2]$  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  $\xi_j$ .

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

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

**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 (25), (26) and (27) we have

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})} [U_j(Y_1, W_1)] = \mathbb{E}[U_j(Y_1, W_1)] \tag{28}$$

and

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

The  $\xi_j$  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 (2005) are satisfied. So that we are able to apply Theorem 3.1 of Houdré and Reynaud-Bouret (2005): there exist absolute constants  $c_1, c_2, c_3$  and  $c_4$  such that for any  $u > 0$ ,

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

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:

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

Finally,

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

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

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

□

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

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

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

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

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

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

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

so

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

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  $\delta$ ,  $\eta$ , and  $\tau'$  depending only on  $\tilde{\gamma}$  so that for  $n$  large enough,

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

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

$$\mathbb{P}(\sigma_j^2 \geq (1 + \tilde{\varepsilon}')\tilde{\sigma}_{j,\tilde{\gamma}}^2) \leq R_4n^{-\tilde{\gamma}}.$$

Combining this inequality with (23), we obtain the desired result of Proposition 1.  $\square$

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

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

$$\mathbb{E}\left[\sup_{j \in J} (|\hat{p}_j(x) - p_j(x)| - \Gamma_\gamma(j))_+^q\right] = o(n^{-q}). \quad (29)$$

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

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

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

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

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

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

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

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 (16), we get

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

Now using Lemma 10, we have that  $\sigma_j^2 \leq R_{10}2^{S_j(2\nu+1)}$  and  $c_j \leq C2^{S_j(\nu+1)}$ . Hence,

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

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

It remains to find an upperbound for the following quantity:

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

We have

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

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

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

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

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

$$\mathbb{E} \left[ \sup_{j \in J} (|p_j(x)|^q \mathbf{1}_{\tilde{\Omega}_{\tilde{\gamma}}^c}) \right] \leq \sup_{j \in J} (\|p\|_2 2^{\frac{S_j}{2}})^q \mathbb{P}(\tilde{\Omega}_{\tilde{\gamma}}^c) \tag{30}$$

$$\leq C \left( \frac{n}{\log^2 n} \right)^{\frac{q}{2}} n^{-\tilde{\gamma}}. \tag{31}$$

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

$$\begin{aligned}
\mathbb{E} \left[ \sup_{j \in J} (|\hat{p}_j(x)|)^q \mathbf{1}_{\tilde{\Omega}_{\tilde{\gamma}}^c} \right] &= \mathbb{E} \left[ \sup_{j \in J} \left| \frac{1}{n} \sum_{l=1}^n Y_l T_j(W_l) \right|^q \mathbf{1}_{\tilde{\Omega}_{\tilde{\gamma}}^c} \right] \\
&\leq \frac{1}{n^q} \mathbb{E} \left[ \sup_{j \in J} \left( \sum_{l=1}^n |m(X_l) + \varepsilon_l| |T_j(W_l)| \right)^q \mathbf{1}_{\tilde{\Omega}_{\tilde{\gamma}}^c} \right] \\
&\leq \frac{n^{q-1}}{n^q} \mathbb{E} \left[ \sup_{j \in J} \sum_{l=1}^n |m(X_l) + \varepsilon_l|^q |T_j(W_l)|^q \mathbf{1}_{\tilde{\Omega}_{\tilde{\gamma}}^c} \right] \\
&\leq \frac{C}{n} \mathbb{E} \left[ \sup_{j \in J} \sum_{l=1}^n (\|m\|_\infty^q + |\varepsilon_l|^q) |T_j(W_l)|^q \mathbf{1}_{\tilde{\Omega}_{\tilde{\gamma}}^c} \right] \\
&\leq C \left( \sup_{j \in J} (\|T_j\|_\infty^q) \mathbb{P}(\tilde{\Omega}_{\tilde{\gamma}}^c) + \sup_{j \in J} (\|T_j\|_\infty^q) \mathbb{E} \left[ |\varepsilon_1|^q \mathbf{1}_{\tilde{\Omega}_{\tilde{\gamma}}^c} \right] \right) \\
&\leq C \left( \sup_{j \in J} (\|T_j\|_\infty^q) \mathbb{P}(\tilde{\Omega}_{\tilde{\gamma}}^c) + \sigma_\varepsilon^q \sup_{j \in J} (\|T_j\|_\infty^q) (\mathbb{E} [|Z|^{2q}])^{\frac{1}{2}} (\mathbb{P}(\tilde{\Omega}_{\tilde{\gamma}}^c))^{\frac{1}{2}} \right),
\end{aligned}$$

where  $Z \sim \mathcal{N}(0, 1)$ . Using (18) and  $\|T_j\|_\infty \leq T_4 2^{S_j(\nu+1)}$ , we get

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

We have

$$\begin{aligned}
E' &\leq C n^{-\frac{\tilde{\gamma}}{2}} \left( \left( \frac{n}{\log^2 n} \right)^{\frac{q}{2}} + \left( \frac{n}{\log^2 n} \right)^{(\nu+1)q} \right) \\
&= o(n^{-q}),
\end{aligned}$$

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

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  $\varphi$  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

$$\hat{p}_{jk} := \frac{1}{n} \sum_{u=1}^n Y_u \times (\mathcal{D}_j \varphi)_{j,k}(W_u) = 2^{\frac{S_j}{2}} \frac{1}{n} \sum_{u=1}^n Y_u \int e^{-i \langle t, 2^j W_u - k \rangle} \prod_{l=1}^d \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l} t_l)} dt.$$



Let us prove now that  $\mathbb{E}[\hat{p}_{jk}] = p_{jk}$ .

We have

$$\mathbb{E}[\hat{p}_{jk}] = 2^{\frac{s_j}{2}} \left( \int \mathbb{E}(Y_1 e^{-i\langle t, 2^j W_1 - k \rangle}) \prod_{l=1}^d \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l} t_l)} dt \right).$$

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

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

Consequently

$$\begin{aligned} \mathbb{E}[\hat{p}_{jk}] &= 2^{\frac{s_j}{2}} \int e^{i\langle t, k \rangle} \mathcal{F}(p)(2^j t) \mathcal{F}(g)(2^j t) \prod_{l=1}^d \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l} t_l)} dt \\ &= 2^{\frac{s_j}{2}} \int e^{i\langle t, k \rangle} \mathcal{F}(p)(2^j t) \prod_{l=1}^d \overline{\mathcal{F}(\varphi)(t_l)} dt \\ &= \int \mathcal{F}(p)(t) \overline{\mathcal{F}(\varphi_{jk})(t)} dt. \end{aligned}$$

Since by Parseval equality, we have

$$p_{jk} = \int p(t) \varphi_{jk}(t) dt = \int \mathcal{F}(p)(t) \overline{\mathcal{F}(\varphi_{jk})(t)} dt,$$

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

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

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

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

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

Therefore, using lemma 3

$$\begin{aligned} \mathbb{E}[\hat{p}_j(x)] - p(x) &= p_j(x) - p(x) = \int K_j(x, y) (p(y) - p(x)) dy \\ &= \sum_k \varphi_{jk}(x) \int \varphi_{jk}(y) (p(y) - p(x)) dy \\ &= \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \dots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \varphi_{jk}(x) \int \prod_{l=1}^d 2^{\frac{j_l}{2}} \varphi(2^{j_l} y_l - k_l) (p(y) - p(x)) dy. \end{aligned}$$

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  $u_l \in [0; 1]$ ,

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

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

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

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

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

□

**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  $i = 1, 2$ ,  $j_i \leq j'_i$  the result is obvious. It is also the case if for  $l = 1, 2$ ,  $j'_l \leq j_l$ . So, without loss of generality, we assume that  $j_1 \leq j'_1$  and  $j'_2 \leq j_2$ . We have:

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

Since  $j_1 \leq j'_1$ , 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  $j'_2 \leq j_2$ , 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 :

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

Therefore,

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

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

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

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  $q, \tilde{\gamma}, \|m\|_\infty, s, \|f_X\|_\infty, \varphi, c_g, \mathcal{C}_g$ .

**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 [Härdle et al. \(1998\)](#)) 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 (17). Hence

$$\mathbb{E}[|B_l|^{\frac{q}{2}}] \leq C \left( \frac{A_q}{n} + \sigma_j^q \right). \quad (32)$$

Using the control of  $A_q$  in (19), equation (20) and Lemma 10 we have

$$\begin{aligned} A_q &\leq Cn\sigma_j^2 \|T_j\|_\infty^{q-2} \\ &\leq Cn2^{S_j(q\nu+q-1)}. \end{aligned} \quad (33)$$

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 (32) and (33) we get

$$\begin{aligned}\mathbb{E}\left[\left(\sum_l B_l\right)^{\frac{q}{2}}\right] &\leq C\left(\sum_l\left(\frac{A_q}{n}+\sigma_j^q\right)+\left(\sum_l\left(\frac{A_4}{n}+\sigma_j^4\right)\right)^{\frac{q}{4}}\right) \\ &\leq C\left(A_q+n\sigma_j^q+(A_4)^{\frac{q}{4}}+n^{\frac{q}{4}}\sigma_j^q\right) \\ &\leq C\left(n2^{S_j(q\nu+q-1)}+n2^{S_j(2\nu+1)\frac{q}{2}}+(n2^{S_j(4\nu+3)})^{\frac{q}{4}}\right).\end{aligned}$$

Consequently

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

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 (8). 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

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

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 (8). Furthermore we have

$$2^{S_j\frac{q(2\nu+1)}{2}}\geq\left(\frac{\log^2 n}{n}2^{2S_j(\nu+1)}\right)^{\frac{q}{2}}\iff 2^{S_j(q\nu+\frac{q}{2})}\geq\left(\frac{\log^2 n}{n}\right)^{\frac{q}{2}}2^{S_j(q\nu+q)}\iff 2^{S_j}\leq\frac{n}{\log^2 n}, \quad (34)$$

which is true again by (8). Consequently

$$\mathbb{E}[\tilde{\sigma}_{j,\tilde{\gamma}}^q]\leq R_52^{S_j(2\nu+1)\frac{q}{2}},$$

with  $R_5$  a constant depending on  $q, \tilde{\gamma}, \|m\|_\infty, s, \|f_X\|_\infty, \varphi, c_g, C_g$  and the lemma is proved for  $q \geq 2$ . For the case  $q \leq 2$  the result follows from Jensen inequality.  $\square$

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

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

**Proof.** Let  $x \in \mathbb{R}^d$  be fixed. We set for any  $j$  and any  $l \in \{1, \dots, d\}$ ,

$$\mathcal{Z}_{j,l}(x) = \{k_l : |2^{j_l} x_l - k_l| \leq A\},$$

whose cardinal is smaller or equal to  $(2A+1)$ . Since

$$\varphi_{jk}(x) = \prod_{l=1}^d 2^{\frac{j_l}{2}} \varphi(2^{j_l} x_l - k_l),$$

then

$$\varphi_{jk}(x) \neq 0 \Rightarrow \forall l \in \{1, \dots, d\}, k_l \in \mathcal{Z}_{j,l}(x).$$

Now,

$$\begin{aligned} \sum_k |\varphi_{jk}(x)| &= \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \prod_{l=1}^d 2^{\frac{j_l}{2}} |\varphi(2^{j_l} x_l - k_l)| \\ &\leq \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \|\varphi\|_\infty^d 2^{\frac{S_j}{2}} \\ &\leq (2A+1)^d \|\varphi\|_\infty^d 2^{\frac{S_j}{2}}. \end{aligned}$$

□

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

$$|\mathcal{F}(\varphi)(t)| \leq R_6(1+|t|)^{-r}, \quad \text{for any } t. \quad (35)$$

and

$$\left| \overline{\mathcal{F}(\varphi)(t)}' \right| \leq R_7(1+|t|)^{-r}, \quad \text{for any } t. \quad (36)$$

**Proof.** First, let us focus on the case  $|t| \geq 1$ .

We have by integration by parts that

$$\mathcal{F}(\varphi)(t) = \int e^{-itx} \varphi(x) dx = \left[ -\frac{1}{it} e^{-itx} \varphi(x) \right]_{-\infty}^{\infty} + \frac{1}{it} \int e^{-itx} \varphi'(x) dx.$$

Using that the father wavelet  $\varphi$  is compactly supported on  $[-A, A]$ , we get

$$\mathcal{F}(\varphi)(t) = \frac{1}{it} \int e^{-itx} \varphi'(x) dx.$$

By successive integration by parts and using that  $|t| \geq 1$  one gets

$$|\mathcal{F}(\varphi)(t)| = \left| \frac{1}{(it)^r} \int e^{-itx} \varphi^{(r)}(x) dx \right| \leq \frac{2^r}{(1+|t|)^r} \int |\varphi^{(r)}(x)| dx,$$

the integral  $\int_{-A}^A |\varphi^{(r)}(x)| dx$  being finite.

For the derivative we have

$$\overline{\mathcal{F}(\varphi)(t)}' = i \int e^{itx} x \varphi(x) dx.$$

Following the same scheme as for  $\mathcal{F}(\varphi)(t)$ , one gets by integration by parts and using the Leibniz formula that

$$\begin{aligned} \left| \overline{\mathcal{F}(\varphi)(t)}' \right| &= \left| \frac{1}{(it)^r} \int e^{itx} \frac{d^r}{dx^r} (x\varphi(x)) dx \right| = \left| \frac{1}{(it)^r} \int e^{itx} \sum_{k=0}^r \binom{r}{k} x^{(k)} \varphi(x)^{(r-k)} dx \right| \\ &\leq \frac{2^r}{(1+|t|)^r} \sum_{k=0}^r \binom{r}{k} \int |x^{(k)} \varphi(x)^{(r-k)}| dx, \end{aligned}$$

the quantity  $\sum_{k=0}^r \binom{r}{k} \int_{-A}^A |x^{(k)} \varphi(x)^{(r-k)}| dx$  being finite.

Hence the lemma is proved for  $|t| \geq 1$ .

The result for  $|t| \leq 1$  is obvious since

$$|\mathcal{F}(\varphi)(t)| = \left| \int e^{-itx} \varphi(x) dx \right| \leq \int |\varphi(x)| dx < \infty,$$

and

$$\left| \overline{\mathcal{F}(\varphi)(t)}' \right| = \left| i \int e^{itx} x \varphi(x) dx \right| \leq \int |x \varphi(x)| dx < \infty.$$

Then the lemma is proved for any  $t$ . □

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

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

where  $R_8$  is a constant depending on  $\varphi$ ,  $C_g$  and  $c_g$ .

**Proof.** If all the  $|w_l| < 1$  then using (9), Lemma 8 and  $r \geq \nu + 2$  with  $\nu \geq 0$  we have

$$|(\mathcal{D}_j \varphi)(w)| \leq \prod_{l=1}^d \int \frac{|\mathcal{F}(\varphi)(t_l)|}{|\mathcal{F}(g_l)(2^{j_l} t_l)|} dt_l \tag{37}$$

$$\leq C \prod_{l=1}^d \int |\mathcal{F}(\varphi)(t_l) (1 + 2^{j_l} |t_l|)^\nu| dt_l \tag{38}$$

$$\leq C 2^{S_j \nu} \prod_{l=1}^d \int (1 + |t_l|)^{\nu-r} dt_l \tag{39}$$

$$\leq C 2^{S_j \nu} \leq C 2^{S_j \nu} \prod_{l=1}^d (1 + |w_l|)^{-1}. \tag{40}$$

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 (40). 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  $|\eta_l(t_l)| \rightarrow 0$  when  $t_l \rightarrow \pm\infty$ , 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, (9) and (10)

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

Therefore,

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

with  $D_1$ ,  $D_2$  and  $D_3$  defined below.

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

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



as soon as  $\nu > 0$ .

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

since  $\nu - r \leq -2$ .

When  $\nu = 0$  we still have

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

Indeed when  $\nu = 0$

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

and

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

using Lemma 8 and  $r \geq 2$ . □

**Lemma 10.** *There exist constants  $T_3$  depending on  $\|m\|_\infty$ ,  $\sigma_\varepsilon$ ,  $\|f_X\|_\infty$ ,  $\varphi$ ,  $c_g$ ,  $\mathcal{C}_g$  and  $T_4$  depending on  $\varphi$ ,  $c_g$ ,  $\mathcal{C}_g$  such that*

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

**Proof.** We have

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

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

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

where  $R_{10}$  is a constant depending on  $\|m\|_\infty, s, \|f_X\|_\infty, \varphi, c_g, \mathcal{C}_g$ . This gives the bound for  $\sigma_j^2$ .

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

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

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

**Acknowledgements:** 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

- Bertin, K., Lacour, C., and Rivoirard, V. (2013). Adaptive pointwise estimation of conditional density function. *Submitted*.
- Carroll, R. J., Delaigle, A., and Hall, P. (2009). Nonparametric prediction in measurement error models. *J. Amer. Statist. Assoc.*, 104(487):993–1003.
- 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(2):185–208.
- Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2):569–609.
- Comte, F. and Taupin, M.-L. (2007). Adaptive estimation in a nonparametric regression model with errors-in-variables. *Statist. Sinica*, 17(3):1065–1090.
- 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(1):149–169.
- 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(2):925–950.
- Du, L., Zou, C., and Wang, Z. (2011). Nonparametric regression function estimation for errors-in-variables models with validation data. *Statist. Sinica*, 21(3):1093–1113.
- Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory*, 48(3):734–747.

- Fan, J. and Masry, E. (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.*, 43(2):237–271.
- Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.*, 21(4):1900–1925.
- Gach, F., Nickl, R., and Spokoiny, V. (2013). Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(3):900–914.
- Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632.
- Härdle, W., Kerkycharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- Ioannides, D. A. and Alevizos, P. D. (1997). Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.*, 32(1):35–43.
- Koo, J.-Y. and Lee, K.-W. (1998). *B*-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.*, 40(1):57–66.
- Meister, A. (2009). *Deconvolution problems in nonparametric statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.
- Patel, J. K. and Read, C. B. (1982). *Handbook of the normal distribution*, volume 40 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York.
- Whittemore, A. S. and Keller, J. B. (1988). Approximations for regression with covariate measurement error. *J. Amer. Statist. Assoc.*, 83(404):1057–1066.