Série des Documents de Travail

n° 2017-70

Linear wavelet estimation of the derivatives of a regression function based on biased data

Y-P. CHAUBEY[1]
C. CHESNEAU[2]
F. NAVARRO[3]

[1] Department of Mathematics and Statistics, Concordia University, E-mail: yogen.chaubey@concordia.ca

[2] Université de Caen - LMNO, E-mail: christophe.chesneau@unicaen.fr

[3] CREST-ENSAI, E-mail : fabien.navarro@ensai.fr

Linear wavelet estimation of the derivatives of a regression function based on biased data

Yogendra P. Chaubey

Department of Mathematics and Statistics,
Concordia University, 1455 de Maisonneuve Blvd. West, Montréal (Québec), H3G 1M8, Canada e-mail: `yogen.chaubey@concordia.ca`

Christophe Chesneau

Laboratoire de Mathématiques Nicolas Oresme,
Université de Caen BP 5186, F 14032 Caen Cedex, FRANCE. e-mail: `christophe.chesneau@unicaen.fr`

Fabien Navarro

Department of Mathematics and Statistics,
Concordia University, 1455 de Maisonneuve Blvd. West, Montréal (Québec), H3G 1M8, Canada e-mail: `fabien.navarro@concordia.ca`

Abstract: This paper deals with the problem of estimating the derivatives of a regression function based on biased data. We develop two different linear wavelet estimators according to the knowledge of the "biased density" of the design. The new estimators are analyzed with respect to their $\mathbb{L}^p$ risk with $p \geq 1$ over Besov balls. Fast polynomial rates of convergence are obtained.

AMS 2000 subject classifications: 62G07, 62G20.
Keywords and phrases: Nonparametric regression, Biased data, Derivatives function estimation, Wavelets, Besov balls.

1. Problem statement

Here we consider the standard biased nonparametric regression model in which case we observe n *i.i.d.* bivariate random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ with the common density

$$f(x, y) = \frac{w(x, y)g(x, y)}{\mu}, \qquad (x, y) \in \mathbb{R}^2,$$

where w is a known positive function, g is the density of an unobserved bivariate random variable (U, V) and $\mu = \mathbb{E}(w(U, V))$ is an unknown real number. We suppose that the support of U is a finite interval, say $[0, 1]$ for the sake of simplicity. In this setting, the unknown regression function of interest is defined by

$$\varphi(x) = \mathbb{E}(V|U = x), \qquad x \in [0, 1].$$

The general aim is to estimate some functionals of φ from $(X_1, Y_1), \ldots, (X_n, Y_n)$.

The direct estimation of φ is a well known problem. It has been considered via different kinds of estimation methods. The most popular of them are the kernel methods. Important results on their performances can be found in, e.g., Ahmad (1995), Sköld (1999), Cristóbal and Alcalá (2000), Wu (2000), Cristóbal and Alcalá (2001), Cristóbal *et al.* (2004), Ojeda *et al.* (2007), Ojeda-Cabrera and Van Keilegom (2009) and Chaubey *et al.* (2012). Recently, wavelet methods based on a multiresolution analysis has been developed for the estimation of φ. Thanks to its powerful local adaptivity against discontinuities, they enjoy nice asymptotic properties for a wide class of unknown regression functions φ. See, e.g., Chesneau and Shirazi (2014), Chaubey *et al.* (2013) and Chaubey and Shirazi (2014). Another recent estimation study via wavelet methods related to the estimation of φ can be found in Chesneau *et al.* (2014).

This study offers three new theoretical contributions. The first one is the estimation of the m^{th} derivative $\varphi^{(m)}$ (assuming that it exists), not just $\varphi = \varphi^{(0)}$. This is of interest in the detection of structures in φ as jump detection and discontinuities, constructions of confidence intervals, and many other statistical aspects. See, for instance, Hall (2010) and the references therein. The second contribution is the construction of an efficient linear wavelet estimator in the case when the density of U is unknown. The consideration of this case is new in the context of wavelet estimation. The third contribution concerns the evaluation of the performances of our estimators: we adopt the $\mathbb{L}^p$-risk with $p \geq 1$, more general to the $\mathbb{L}^2$-risk (or Mean Integrated Squared Error). To the best of our knowledge, it has never been investigated in this setting, despite its potential interest to exhibit new phenomena in terms of rates of convergence. In this study, they are determined assuming that $\varphi^{(m)}$ belongs to the Besov balls; a wide class of homogeneous and inhomogeneous functions.

The organization of this paper is as follows. The next section describes the considered wavelet basis, Besov balls and basics on linear wavelet estimation. The problem of estimating the derivatives of a regression function from biased data is considered in Section 3, distinguishing the estimation of φ when the density of U is known or not. Here we have constructed efficient linear wavelet estimators and their performances are demonstrated in terms of rates of convergence under the $\mathbb{L}^p$ risk over Besov balls, with $p \geq 1$. The proofs are carried out in Section 4.

2. Preliminaries

This section is devoted to the presentation of the main notions of the study, i.e., the wavelet basis, the Besov balls and the linear wavelet estimation in general.

2.1. Wavelet basis

For any $p \geq 1$, we defined the set $\mathbb{L}^p([0,1])$ by

$$\mathbb{L}^p([0,1]) = \left\{ t : [0,1] \to \mathbb{R};\ ||t||_p = \left(\int_{[0,1]} |t(x)|^p dx \right)^{1/p} < \infty \right\}.$$

Among the existing constructions of wavelet basis on the unit interval, we consider the one introduced by Cohen *et al.* (1993). It is briefly described below.

Let ϕ and ψ be the initial wavelet functions of the Daubechies wavelets family *db2N* with $N \geq 5m$. These functions are interesting as they are compactly supported and belong to the class $\mathcal{C}^m$. For any $j \geq 0$, we set $\Lambda_j = \{0, \ldots, 2^j - 1\}$ and, for $k \in \Lambda_j$,

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

With appropriate treatment at the boundaries, there exists an integer τ such that, for any integer $\ell \geq \tau$, the family

$$\mathcal{S} = \{\phi_{\ell,k},\ k \in \Lambda_\ell;\ \psi_{j,k};\ j \in \mathbb{N} - \{0, \ldots, \ell - 1\},\ k \in \Lambda_j\}$$

forms an orthonormal basis of $\mathbb{L}^2([0,1])$.

Therefore, for any integer $\ell \geq \tau$ and $t \in \mathbb{L}^2([0,1])$, we have the following wavelet expansion:

$$t(x) = \sum_{k \in \Lambda_\ell} c_{\ell,k}\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k \in \Lambda_j} d_{j,k}\psi_{j,k}(x), \qquad x \in [0,1], \tag{2.1}$$

where $c_{j,k}$ and $d_{j,k}$ are defined by

$$c_{j,k} = \int_{[0,1]} t(x)\phi_{j,k}(x)dx, \qquad d_{j,k} = \int_{[0,1]} t(x)\psi_{j,k}(x)dx. \tag{2.2}$$

These are approximation and detail wavelet coefficients of t respectively; see, e.g., Cohen *et al.* (1993) and Mallat (2009).

Let us now introduce a $\mathbb{L}^p$-norm result related to the approximation term.

Lemma 2.1. *Let $p \geq 1$. For any sequence of real numbers $(\theta_{j,k})_{j,k}$, there exists a constant $C > 0$ such that, for any $j \geq \tau$,*

$$\int_{[0,1]} \left(\sum_{k \in \Lambda_j} \theta_{j,k}\phi_{j,k}(x) \right)^p dx \leq C 2^{j(p/2-1)} \sum_{k \in \Lambda_j} |\theta_{j,k}|^p.$$

The proof can be found in, e.g., (Härdle *et al.*, 1998, Proposition 8.3).

2.2. Besov balls

For the sake of simplicity, we consider the following wavelet sequential definition of the Besov balls. We say that $t \in B^s_{q,r}(M)$ with $s \in (0, N)$, $q \geq 1$, $r \geq 1$ and $M > 0$ if there exists a constant $C > 0$ (depending on M) such that $c_{j,k}$ and $d_{j,k}$ (2.2) satisfy

$$2^{\tau(1/2-1/q)} \left(\sum_{k \in \Lambda_\tau} |c_{\tau,k}|^q \right)^{1/q} + \left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/q)} \left(\sum_{k \in \Lambda_j} |d_{j,k}|^q \right)^{1/q} \right)^r \right)^{1/r} \leq C,$$

with the usual modifications if $q = \infty$ or $r = \infty$.

In wavelet estimation, the Besov balls are particularly interesting because they contain a wide variety of homogeneous and inhomogeneous functions. For particular choices of s, p and r, $B^s_{q,r}(M)$ correspond to standard balls of function spaces, as the Hölder and Sobolev balls (see, e.g., Meyer (1992) and Härdle *et al.* (1998)).

The following lemma presents a standard inclusion for Besov balls which will be useful in the proofs of our main results.

Lemma 2.2. *For any* $p \geq 1$, $q \geq 1$, $M > 0$ *and* $s \in (\max(1/q - 1/p, 0), N)$, *we have*

$$B^s_{q,r}(M) \subseteq B^{s_*}_{p,r}(M),$$

with $s_* = s + \min(1/p - 1/q, 0)$.

See (Härdle *et al.*, 1998, Corollary 9.2).

2.3. Linear wavelet estimation

The idea of the linear wavelet estimation is to estimate the approximation wavelet coefficients $c_{j,k}$ of an unknown function t and project these estimators on $\mathcal{S}$ at a suitable level j_0. They are of the form:

$$\hat{t}(x) = \sum_{k \in \Lambda_{j_0}} \hat{c}_{j_0,k} \phi_{j_0,k}(x), \tag{2.3}$$

where $\hat{c}_{j,k}$ denotes an estimator for $c_{j,k}$ constructed from n observations.

Such estimators generally enjoy good theoretical properties under the $\mathbb{L}^p$-risk; see, for instance, Härdle *et al.* (1998), Chapter 10 and Chaubey *et al.* (2011).

In this study, this $\mathbb{L}^p$-risk is considered: we aim to construct linear wavelet estimators $\hat{\varphi}^{(m)}$ of the form (2.3) such that, for any $\varphi^{(m)} \in B^s_{q,r}(M)$,

$$\lim_{n \to \infty} \mathbb{E}\left(\|\hat{\varphi}^{(m)} - \varphi^{(m)}\|_p^p \right) = 0,$$

as fast as possible.

3. Results

This section is devoted to the linear wavelet estimation of the three following related problems:

1. the estimation of $\varphi^{(m)}$ when h is known,
2. the estimation of h and
3. the estimation of $\varphi^{(m)}$ when h is unknown,

where h denotes the marginal probability density function of the random variable U.

3.1. Assumptions

The following assumptions will be used in our main results:

- We have

$$\varphi^{(u)}(0) = \varphi^{(u)}(1) = 0, \qquad u \in \{0, \ldots, m\}. \tag{3.1}$$

- There exists a constant $C_1 > 0$ such that

$$\sup_{x \in [0,1]} |\varphi^{(m)}(x)| \leq C_1. \tag{3.2}$$

- There exist two constants $C_2 > 0$ and $c_2 > 0$ such that

$$\inf_{(x,y) \in [0,1] \times \mathbb{R}} w(x,y) \geq c_2, \qquad \sup_{(x,y) \in [0,1] \times \mathbb{R}} w(x,y) \leq C_2. \tag{3.3}$$

- There exist two constants $c_3 > 0$ and $C_3 > 0$ such that

$$c_3 \leq \inf_{x \in [0,1]} h(x), \qquad \sup_{x \in [0,1]} h(x) \leq C_3. \tag{3.4}$$

- There exists a constant $C_4 > 0$ such that

$$\sup_{x \in [0,1]} \int_{\mathbb{R}} y^{2p} g(x,y) dy \leq C_4. \tag{3.5}$$

Despite their restrictive natures, these assumptions are satisfied by wide class of functions $\varphi^{(m)}$, $h(x)$, $w(x,y)$ and $g(x,y)$.

3.2. Estimation of $\varphi^{(m)}$ when h is known

When h is known, we consider the linear wavelet estimator $\hat{\varphi}_1^{(m)}$ of $\varphi^{(m)}$ defined by

$$\hat{\varphi}_1^{(m)}(x) = \sum_{k \in \Lambda_{j_0}} \hat{c}_{j_0,k}^{(m)} \phi_{j_0,k}(x), \qquad x \in [0,1], \tag{3.6}$$

where

$$\hat{c}_{j,k}^{(m)} = (-1)^m \frac{\hat{\mu}}{n} \sum_{i=1}^{n} \frac{Y_i}{w(X_i,Y_i)h(X_i)} (\phi_{j,k})^{(m)}(X_i), \tag{3.7}$$

$(\phi_{j,k})^{(m)}(x) = 2^{j/2}2^{mj}\phi^{(m)}(2^j x - k)$,

$$\hat{\mu} = \left(\frac{1}{n} \sum_{i=1} \frac{1}{w(X_i,Y_i)} \right)^{-1} \tag{3.8}$$

and j_0 is an integer chosen a posteriori. The form of the estimator $\hat{c}_{j,k}^{(m)}$ is motivated by writing $c_{j,k}^{(m)} = \int_{[0,1]} \varphi^{(m)}(x)\phi_{j,k}(x)dx$ in the present context as an appropriate expectation with respect to density f.

The estimator $\hat{c}_{j,k}^{(m)}$ satisfies the moment inequality described below.

Proposition 3.1. *Let $p \geq 1$. Suppose that the assumptions in Subsection 3.1 hold. Let $\hat{c}_{j,k}^{(m)}$ be given by (3.7) with j such that $2^j \leq n$ and $c_{j,k}^{(m)} = \int_{[0,1]} \varphi^{(m)}(x)\phi_{j,k}(x)dx$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left((\hat{c}_{j,k}^{(m)} - c_{j,k}^{(m)})^{2p} \right) \leq C \left(\frac{2^{2jm}}{n} \right)^p .$$

Theorem 3.1 below investigates the rate of convergence attained by $\hat{\varphi}_1^{(m)}$ under the $\mathbb{L}^p$-risk assuming that $\varphi^{(m)} \in B_{q,r}^s(M)$.

Theorem 3.1. *Let $p \geq 1$. Suppose that the assumptions in Subsection 3.1 hold and that $\varphi^{(m)} \in B_{q,r}^s(M)$ with $M > 0$, $q \geq 1$, $r \geq 1$ and $s \in (\max(1/q - 1/p, 0), N)$. Let $\hat{\varphi}_1^{(m)}$ be defined by (3.6) with j_0 such that*

$$2^{j_0} = [n^{1/(2s_*+2m+1)}], \tag{3.9}$$

$s_* = s + \min(1/p - 1/q, 0)$ *(where $[a]$ denotes the integer part of a).*

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

$$\mathbb{E}\left(\|\hat{\varphi}_1^{(m)} - \varphi^{(m)}\|_p^p \right) \leq Cn^{-s_*p/(2s_*+2m+1)}.$$

The integer j_0 is chosen to minimize the $\mathbb{L}^p$-risk of $\hat{\varphi}_1^{(m)}$. Note that, for $m = 0$ and $p = 2$, Theorem 3.1 becomes (Chesneau and Shirazi, 2014, Theorem 4.1, $p = 2$).

Remark 3.1. *It follows from Theorem 3.1, the Markov inequality and the Borel-Cantelli lemma that, for $p > 2 + (2m + 1)/s_*$, we have*

$$\lim_{n\to\infty} \|\hat{\varphi}_1^{(m)} - \varphi^{(m)}\|_p^p = 0 \text{ almost surely.}$$

When h is unknown, the estimator $\hat{\varphi}_1^{(m)}$ (3.6) is not appropriate since it depends on h in its construction. To solve this problem, a first step is to investigate the estimation of h from $(X_1, Y_1), \ldots (X_n, Y_n)$. This is done in the next section.

3.3. Estimation of h

This problem of estimating h from $(X_1, Y_1), \dots (X_n, Y_n)$ is close to the standard weighted density estimation problem. See, e.g., Ahmad (1995) for kernel methods and Ramirez and Vidakovic (2010) for wavelet methods. However, to the best of our knowledge, it has never been considered in our bivariate context.

We define the linear wavelet estimator $\hat{h}$ of h by

$$\hat{h}(x) = \sum_{k \in \Lambda_{j_1}} \hat{c}_{j_1,k} \phi_{j_1,k}(x), \qquad x \in [0,1], \tag{3.10}$$

where

$$\hat{c}_{j,k} = \frac{\hat{\mu}}{n} \sum_{i=1}^{n} \frac{1}{w(X_i, Y_i)} \phi_{j,k}(X_i), \tag{3.11}$$

$\hat{\mu}$ is given by (3.8) and j_1 is an integer chosen a posteriori.

Theorem 3.2 below investigates the rate of convergence attained by $\hat{h}$ under the $\mathbb{L}^p$ risk assuming that $h \in B^s_{q,r}(M)$.

Theorem 3.2. *Let $p \geq 1$. Suppose that the assumptions* (3.3) *and* (3.4) *hold and that $h \in B^s_{q,r}(M)$ with $M > 0$, $q \geq 1$, $r \geq 1$ and $s \in (\max(1/q - 1/p, 0), N)$. Let $\hat{h}$ be defined by* (3.10) *with j_0 such that*

$$2^{j_1} = [n^{1/(2s_*+1)}], \tag{3.12}$$

$s_* = s + \min(1/p - 1/q, 0)$.

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

$$\mathbb{E}\left(\|\hat{h} - h\|_p^p\right) \leq C n^{-s_* p/(2s_*+1)}.$$

The rate of convergence $n^{-s_* p/(2s_*+1)}$ corresponds to the one obtained for standard density estimation under the $\mathbb{L}^p$-risk. See, for instance, Donoho *et al.* (1996) and (Härdle *et al.*, 1998, Chapter 10).

We are now able to investigate the estimation of $\varphi^{(m)}$ when h is unknown via a plug-in approach using $\hat{\varphi}_1^{(m)}$ (3.6) and $\hat{h}$ (3.10).

3.4. Estimation of $\varphi^{(m)}$ when h is unknown

In the case where h is unknown, we propose the linear wavelet estimator $\hat{\varphi}_2^{(m)}$ of $\varphi^{(m)}$ defined by

$$\hat{\varphi}_2^{(m)}(x) = \sum_{k \in \Lambda_{j_2}} \tilde{c}^{(m)}_{j_2,k} \phi_{j_2,k}(x), \qquad x \in [0,1], \tag{3.13}$$

where

$$\tilde{c}^{(m)}_{j,k} = (-1)^m \frac{\hat{\mu}}{[n/2]} \sum_{i=1}^{[n/2]} \frac{Y_i}{w(X_i, Y_i)\hat{h}(X_i)} \mathbf{1}_{\left\{|\hat{h}(X_i)| \geq c_3/2\right\}} (\phi_{j,k})^{(m)}(X_i), \tag{3.14}$$

j_2 is an integer chosen a posteriori, $\hat{\mu}$ is given by (3.8), $\mathbf{1}$ denotes the indicator function, c_3 refers to (3.4), $\hat{h}$ is given by (3.10) but defined with the random variables $((X_{[n/2]+1}, Y_{[n/2]+1}), \ldots, (X_n, Y_n))$ and an integer j_2 chosen a posteriori.

The construction of $\tilde{c}_{j,k}^{(m)}$ follows the "plug-in spirit" of the NES estimator introduced by Pensky and Vidakovic (2001). It is an adaptation of the version developed in (Chesneau, 2014, Subsection 3.3) in the present context.

Theorem 3.3 below investigates the rate of convergence attained by $\hat{\varphi}_2^{(m)}$ under the $\mathbb{L}^p$ risk assuming that $\varphi^{(m)} \in B^s_{q,r}(M)$.

Theorem 3.3. *Let $p \geq 1$ and $p_* = \max(p, 2)$. Suppose that the assumptions in Subsection 3.1 hold, $\varphi^{(m)} \in B^{s_1}_{q_1,r_1}(M_1)$ with $M_1 > 0$, $q_1 \geq 1$, $r_1 \geq 1$, $s \in (\max(1/q_1 - 1/p_*, 0), N)$, and $h \in B^{s_2}_{q_2,r_2}(M_2)$ with $M_2 > 0$, $q_2 \geq 1$, $r_2 \geq 1$ and $s_2 \in (\max(1/q_2 - 1/p_*, 0), N)$. Let $\hat{\varphi}_2^{(m)}$ be defined by (3.13) and (3.14) with j_1, j_2 such that*

$$2^{j_1} = [n^{1/(2s_o+1)}], \qquad s_o = s_2 + \min(1/p_* - 1/q_2, 0), \tag{3.15}$$

and

$$2^{j_2} = [n^{2s_o/((2s_o+1)(2s_*+2m+1))}], \qquad s_* = s_1 + \min(1/p_* - 1/q_1, 0) \tag{3.16}$$

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

$$\mathbb{E}\left(\|\hat{\varphi}_2^{(m)} - \varphi^{(m)}\|_p^p\right) \leq C n^{-2s_* s_o p/((2s_o+1)(2s_*+2m+1))}.$$

Again, the definitions of the integers j_1 and j_2 are based on theoretical consideration; they are chosen to minimize the $\mathbb{L}^p$-risk of $\hat{\varphi}_2^{(m)}$. An interest of Theorem 3.3 is to measure the influences of the smoothness of h in the linear wavelet estimation of $\varphi^{(m)}$. For $p = 2$, note that the obtained rate of convergence corresponds to the one obtained in the unbiased case (Chesneau, 2014, Theorem 3).

Remark 3.2. *Similar arguments to Remark 3.1 give, for p such that $2s_* s_o p/((2s_o + 1)(2s_* + 2m + 1)) > 1$,*

$$\lim_{n\to\infty} \|\hat{\varphi}_2^{(m)} - \varphi^{(m)}\|_p^p = 0 \text{ almost surely.}$$

4. Proofs

In this section, C denotes any constant that does not depend on j, k and n. Its value may change from one term to another and may depend on ϕ or ψ.

Proof of Proposition 3.1 The proof is a generalization of (Chesneau *et al.*, 2014, Proposition 4 (ii)) to the m^{th} derivatives and the $\mathbb{L}^p$-norm. We obtain the desired result via the Rosenthal inequality presented below (see Rosenthal (1970)).

Lemma 4.1 (Rosenthal's inequality)**.** *Let n be a positive integer, $\gamma \geq 2$ and $U_1, \ldots, U_n$ be n i.i.d. random variables such that $\mathbb{E}(U_1) = 0$ and $\mathbb{E}(|U_1|^\gamma) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} U_i\right|^\gamma\right) \leq C \max\left(n\mathbb{E}(|U_1|^\gamma), n^{\gamma/2}\left(\mathbb{E}(U_1^2)\right)^{\gamma/2}\right).$$

Observe that

$$\begin{aligned} &\hat{c}_{j,k}^{(m)} - c_{j,k}^{(m)} \\ &= \frac{\hat{\mu}}{\mu}\left((-1)^n \frac{\mu}{n}\sum_{i=1}^{n} \frac{Y_i}{w(X_i, Y_i)h(X_i)}(\phi_{j,k})^{(m)}(X_i) - c_{j,k}^{(m)}\right) + c_{j,k}^{(m)}\hat{\mu}\left(\frac{1}{\mu} - \frac{1}{\hat{\mu}}\right). \end{aligned}$$

Using the triangular inequality, by (3.2) and (3.3): $|\hat{\mu}/\mu| \leq C_2/c_2$, $|\hat{\mu}| \leq C_2$, and $|c_{j,k}^{(m)}| \leq C_1$, we obtain

$$|\hat{c}_{j,k}^{(m)} - c_{j,k}^{(m)}| \leq C\left(\left|(-1)^n \frac{\mu}{n}\sum_{i=1}^{n} \frac{Y_i}{w(X_i, Y_i)h(X_i)}(\phi_{j,k})^{(m)}(X_i) - c_{j,k}^{(m)}\right| + \left|\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right|\right).$$

The inequality: $(x + y)^{2p} \leq 2^{2p-1}(x^{2p} + y^{2p})$, $(x, y) \in \mathbb{R}^2$, gives

$$\mathbb{E}\left((\hat{c}_{j,k}^{(m)} - c_{j,k}^{(m)})^{2p}\right) \leq C(Q_1 + Q_2), \tag{4.1}$$

where

$$Q_1 = \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left((-1)^m \mu \frac{Y_i}{w(X_i, Y_i)h(X_i)}(\phi_{j,k})^{(m)}(X_i) - c_{j,k}^{(m)}\right)\right)^{2p}\right)$$

and

$$Q_2 = \mathbb{E}\left(\left(\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right)^{2p}\right).$$

Now we investigate upper bounds for Q_1 and Q_2.

Upper bound for Q_1. Note that

$$Q_1 = \frac{1}{n^{2p}}\mathbb{E}\left(\left(\sum_{i=1}^{n} U_i\right)^{2p}\right),$$

with

$$U_i = (-1)^m \mu \frac{Y_i}{w(X_i, Y_i)h(X_i)}(\phi_{j,k})^{(m)}(X_i) - c_{j,k}^{(m)}, \qquad i \in \{1, \ldots, n\}.$$

Since $(X_1, Y_1), \ldots, (X_n, Y_n)$ are *i.i.d.*, $U_1, \ldots, U_n$ are also *i.i.d.*. Let us now show that $\mathbb{E}(U_1) = 0$. Using the definition of $f(x, y)$, the equality $\int_{\mathbb{R}} yg(x, y)dy =$

$\varphi(x)h(x)$ and m integrations by parts with (3.1), we obtain

$$\begin{aligned}
&\mathbb{E}\left((-1)^m\mu\frac{Y_1}{w(X_1,Y_1)h(X_1)}(\phi_{j,k})^{(m)}(X_1)\right)\\
&= \int_{\mathbb{R}}\int_{[0,1]}(-1)^m\mu\frac{y}{w(x,y)h(x)}(\phi_{j,k})^{(m)}(x)f(x,y)dxdy\\
&= (-1)^m\int_{\mathbb{R}}\int_{[0,1]}\mu\frac{y}{w(x,y)h(x)}(\phi_{j,k})^{(m)}(x)\frac{w(x,y)g(x,y)}{\mu}dxdy\\
&= (-1)^m\int_{[0,1]}\frac{1}{h(x)}(\phi_{j,k})^{(m)}(x)\left(\int_{\mathbb{R}}yg(x,y)dy\right)dx\\
&= (-1)^m\int_{[0,1]}\frac{1}{h(x)}(\phi_{j,k})^{(m)}(x)\varphi(x)h(x)dx\\
&= (-1)^m\int_{[0,1]}\varphi(x)(\phi_{j,k})^{(m)}(x)dx = \int_{[0,1]}\varphi^{(m)}(x)\phi_{j,k}(x)dx = c_{j,k}^{(m)}.
\end{aligned}$$

Therefore $\mathbb{E}(U_1) = 0$.

Let $u \in \{2, 2p\}$. Using the inequality: $(x+y)^u \le 2^{u-1}(x^u + y^u)$, $(x,y) \in \mathbb{R}^2$, the Hölder inequality, (3.3), (3.4), (3.5), the definition of $f(x,y)$, $(\phi_{j,k})^{(m)}(x) = 2^{j/2}2^{mj}\phi^{(m)}(2^j - k)$, a change of variables and $2^j \le n$, we have

$$\begin{aligned}
\mathbb{E}(U_1^u) &\le 2^{u-1}\mathbb{E}\left(\left((-1)^m\mu\frac{Y_1}{w(X_1,Y_1)h(X_1)}(\phi_{j,k})^{(m)}(X_1)\right)^u + (c_{j,k}^{(m)})^u\right)\\
&\le 2^u\mathbb{E}\left(\left((-1)^m\mu\frac{Y_1}{w(X_1,Y_1)h(X_1)}(\phi_{j,k})^{(m)}(X_1)\right)^u\right)\\
&\le C\mathbb{E}\left(\left(Y_1(\phi_{j,k})^{(m)}(X_1)\right)^u\frac{1}{w(X_1,Y_1)}\right)\\
&= C\int_{\mathbb{R}}\int_{[0,1]}\left(y(\phi_{j,k})^{(m)}(x)\right)^u\frac{1}{w(x,y)}f(x,y)dxdy\\
&= C\int_{\mathbb{R}}\int_{[0,1]}\left(y(\phi_{j,k})^{(m)}(x)\right)^u\frac{1}{w(x,y)}\frac{w(x,y)g(x,y)}{\mu}dxdy\\
&\le C\int_{[0,1]}\left(\int_{\mathbb{R}}y^u g(x,y)dy\right)\left((\phi_{j,k})^{(m)}(x)\right)^u dx\\
&\le C\int_{[0,1]}\left((\phi_{j,k})^{(m)}(x)\right)^u dx = C2^{jmu}2^{j(u-2)/2}\int_{[0,1]}(\phi^{(m)}(x))^u dx\\
&\le C2^{jmu}n^{(u-2)/2}. \qquad (4.2)
\end{aligned}$$

It follows from Lemma 4.1 with $U_1,\dots,U_n$ and $\gamma = 2p$, and (4.2) that

$$\begin{aligned}
Q_1 &\le C\frac{1}{n^{2p}}\max\left(n\mathbb{E}(U_1^{2p}), n^p(\mathbb{E}(U_1^2))^p\right)\\
&\le C\frac{1}{n^{2p}}\max\left(n2^{2jmp}n^{p-1}, n^p(2^{2jm})^p\right) \le C\frac{2^{2jmp}}{n^p}. \qquad (4.3)
\end{aligned}$$

Upper bound for Q_2. We can write

$$Q_2 = \frac{1}{n^{2p}}\mathbb{E}\left(\left(\sum_{i=1}^{n} U_i\right)^{2p}\right),$$

with

$$U_i = \frac{1}{w(X_i, Y_i)} - \frac{1}{\mu}, \qquad i \in \{1, \ldots, n\}.$$

Since $(X_1, Y_1), \ldots, (X_n, Y_n)$ are *i.i.d.*, $U_1, \ldots, U_n$ are also *i.i.d.*. Moreover, from (Chesneau *et al.*, 2014, Proposition 2 (i)), we have $\mathbb{E}(U_1) = 0$. Using (3.3), for any $u \in \{2, 2p\}$, we arrive at $\mathbb{E}(U_1^u) \leq C$. Thus, Lemma 4.1 with $\gamma = 2p$ yields

$$Q_2 \leq C\frac{1}{n^{2p}} \max\left(n\mathbb{E}(U_1^{2p}), n^p(\mathbb{E}(U_1^2))^p\right) \leq C\frac{1}{n^p}. \tag{4.4}$$

It follows from (4.1), (4.3) and (4.4) that

$$\mathbb{E}\left((\hat{c}_{j,k}^{(m)} - c_{j,k}^{(m)})^{2p}\right) \leq C\left(\frac{2^{2jmp}}{n^p} + \frac{1}{n^p}\right) \leq C\left(\frac{2^{2jm}}{n}\right)^p.$$

Thus Proposition 3.1 is proved.

□

Proof of Theorem 3.1. We expand $\varphi^{(m)}$ on $\mathcal{S}$ as in (2.1) at the level $\ell = j_0$ given by (3.9):

$$\varphi^{(m)}(x) = \sum_{k\in\Lambda_{j_0}} c_{j_0,k}^{(m)}\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\Lambda_j} d_{j,k}^{(m)}\psi_{j,k}(x),$$

where $c_{j_0,k}^{(m)} = \int_{[0,1]} \varphi^{(m)}(x)\phi_{j_0,k}(x)dx$ and $d_{j,k}^{(m)} = \int_{[0,1]} \varphi^{(m)}(x)\psi_{j,k}(x)dx$.

Using the inequality: $||f + g||_p^p \leq 2^{p-1}(||f||_p^p + ||g||_p^p)$, $f, g \in \mathbb{L}^p([0,1])$, we have

$$\mathbb{E}\left(\|\hat{\varphi}_1^{(m)} - \varphi^{(m)}\|_p^p\right) \leq C(A_1 + A_2), \tag{4.5}$$

where

$$A_1 = \mathbb{E}\left(\|\sum_{k\in\Lambda_{j_0}}\left(\hat{c}_{j_0,k}^{(m)} - c_{j_0,k}^{(m)}\right)\phi_{j_0,k}\|_p^p\right), \qquad A_2 = \|\sum_{j=j_0}^{\infty}\sum_{k\in\Lambda_j} d_{j,k}^{(m)}\psi_{j,k}\|_p^p.$$

Let us now investigate upper bounds for A_1 and A_2.

Upper bound for A_1. It follows from Lemma 2.1, the Hölder inequality, Proposition 3.1 and $\mathrm{Card}(\Lambda_j) = 2^j$ that

$$\begin{aligned} A_1 &\le C 2^{j_0(p/2-1)} \sum_{k\in\Lambda_{j_0}} \mathbb{E}\left(|\hat{c}^{(m)}_{j_0,k} - c^{(m)}_{j_0,k}|^p\right) \\ &\le C 2^{j_0(p/2-1)} \sum_{k\in\Lambda_{j_0}} \left(\mathbb{E}\left((\hat{c}^{(m)}_{j_0,k} - c^{(m)}_{j_0,k})^{2p}\right)\right)^{1/2} \\ &\le C 2^{j_0(p/2-1)} 2^{j_0} \left(\frac{2^{2j_0 m}}{n}\right)^{p/2} \le C \left(\frac{2^{j_0(1+2m)}}{n}\right)^{p/2}. \end{aligned} \tag{4.6}$$

Upper bound for A_2. Using Lemma 2.2 and proceeding as in (Donoho *et al.*, 1996, eq (24)), we have

$$A_2 \le C 2^{-j_0 s_* p}. \tag{4.7}$$

It follows from (4.5), (4.6) and (4.7) that

$$\mathbb{E}\left(\|\hat{\varphi}^{(m)}_1 - \varphi^{(m)}\|^p_p\right) \le C\left(\left(\frac{2^{j_0(1+2m)}}{n}\right)^{p/2} + 2^{-j_0 s_* p}\right) \le C n^{-s_* p/(2s_*+2m+1)}.$$

Hence, Theorem 3.1 is proved.

□

Proof of Theorem 3.2. We use a similar approach here as in the proof of Theorem 3.1. We expand h on $\mathcal{S}$ as (2.1) at the level $\ell = j_1$ given by (3.12):

$$h(x) = \sum_{k\in\Lambda_{j_1}} c_{j_1,k}\phi_{j_1,k}(x) + \sum_{j=j_1}^{\infty}\sum_{k\in\Lambda_j} d_{j,k}\psi_{j,k}(x),$$

where $c_{j_1,k} = \int_{[0,1]} h(x)\phi_{j_1,k}(x)dx$ and $d_{j,k} = \int_{[0,1]} h(x)\psi_{j,k}(x)dx$.

The inequality: $||f+g||^p_p \le 2^{p-1}(||f||^p_p + ||g||^p_p)$, $f, g \in \mathbb{L}^p([0,1])$, yields

$$\mathbb{E}\left(\|\hat{h} - h\|^p_p\right) \le C(B_1 + B_2), \tag{4.8}$$

where

$$B_1 = \mathbb{E}\left(\|\sum_{k\in\Lambda_{j_1}} (\hat{c}_{j_1,k} - c_{j_1,k})\,\phi_{j_1,k}\|^p_p\right), \qquad B_2 = \|\sum_{j=j_1}^{\infty}\sum_{k\in\Lambda_j} d_{j,k}\psi_{j,k}\|^p_p.$$

Let us now investigate upper bounds for B_1 and B_2.

Upper bound for B_1. First of all, by the definition of $f(x,y)$ and (3.11), observe that

$$\begin{aligned}\mathbb{E}\left(\frac{\mu}{\hat{\mu}}\hat{c}_{j,k}\right) &= \mathbb{E}\left(\frac{\mu}{w(X_1,Y_1)}\phi_{j,k}(X_1)\right)\\ &= \int_{\mathbb{R}}\int_{[0,1]}\frac{\mu}{w(x,y)}\phi_{j,k}(x)f(x,y)dxdy\\ &= \int_{\mathbb{R}}\int_{[0,1]}\frac{\mu}{w(x,y)}\phi_{j,k}(x)\frac{w(x,y)g(x,y)}{\mu}dxdy\\ &= \int_{[0,1]}\phi_{j,k}(x)\left(\int_{\mathbb{R}}g(x,y)dy\right)dx = \int_{[0,1]}\phi_{j,k}(x)h(x)dx = c_{j,k}.\end{aligned}$$

Proceeding as in the proof of Proposition 3.1 but with "1" instead of "Y_i" and $m=0$, under (3.3) and (3.4) only, we prove the existence of a constant $C>0$ such that

$$\mathbb{E}\left((\hat{c}_{j_1,k}-c_{j_1,k})^{2p}\right) \le C\frac{1}{n^p}, \tag{4.9}$$

It follows from Lemma 2.1, the Hölder inequality, (4.9) and $\mathrm{Card}(\Lambda_j)=2^j$ that

$$\begin{aligned}B_1 &\le C2^{j_1(p/2-1)}\sum_{k\in\Lambda_{j_1}}\mathbb{E}\left(|\hat{c}_{j_1,k}-c_{j_1,k}|^p\right)\\ &\le C2^{j_1(p/2-1)}\sum_{k\in\Lambda_{j_1}}\left(\mathbb{E}\left((\hat{c}_{j_1,k}-c_{j_1,k})^{2p}\right)\right)^{1/2}\\ &\le C2^{j_1(p/2-1)}2^{j_1}\frac{1}{n^{p/2}} \le C\left(\frac{2^{j_1}}{n}\right)^{p/2}.\end{aligned} \tag{4.10}$$

Upper bound for B_2. Proceeding as in (4.7), we obtain

$$B_2 \le C2^{-j_1 s_* p}. \tag{4.11}$$

It follows from (4.8), (4.10) and (4.11) that

$$\mathbb{E}\left(\|\hat{h}-h\|_p^p\right) \le C\left(\left(\frac{2^{j_1}}{n}\right)^{p/2}+2^{-j_1 s_* p}\right) \le Cn^{-s_* p/(2s_*+1)}.$$

Thus Theorem 3.2 is proved.

□

Proof of Theorem 3.3. Firstly, let us consider the case $p\ge 2$. We expand $\varphi^{(m)}$ on $\mathcal{S}$ as (2.1) at the level $\ell=j_2$ given by (3.16):

$$\varphi^{(m)}(x) = \sum_{k\in\Lambda_{j_2}}c^{(m)}_{j_2,k}\phi_{j_2,k}(x)+\sum_{j=j_2}^{\infty}\sum_{k\in\Lambda_j}d^{(m)}_{j,k}\psi_{j,k}(x).$$

Using the inequality: $||f+g||_p^p \leq 2^{p-1}(||f||_p^p + ||g||_p^p)$, $f, g \in \mathbb{L}^p([0,1])$, we get

$$\mathbb{E}\left(\|\hat{\varphi}_2^{(m)} - \varphi^{(m)}\|_p^p\right) \leq C(D+E), \tag{4.12}$$

where

$$D = \mathbb{E}\left(\|\sum_{k\in\Lambda_{j_2}}\left(\tilde{c}_{j_2,k}^{(m)} - c_{j_2,k}^{(m)}\right)\phi_{j_0,k}\|_p^p\right), \qquad E = \|\sum_{j=j_2}^{\infty}\sum_{k\in\Lambda_j} d_{j,k}^{(m)}\psi_{j,k}\|_p^p.$$

Upper bound for E. Proceeding as in (4.7), we obtain

$$E \leq C2^{-j_2 s_* p}. \tag{4.13}$$

Upper bound for D. Let $\hat{c}_{j_2,k}^{(m)}$ be (3.7) with $n=[n/2]$ and $j=j_2$ (3.16). The inequality $|x+y|^p \leq 2^{p-1}(|x|^p+|y|^p)$, $(x,y)\in\mathbb{R}^2$, and Lemma 2.1 yield

$$D \leq C(D_1+D_2), \tag{4.14}$$

where

$$D_1 = 2^{j_2(p/2-1)}\sum_{k\in\Lambda_{j_2}}\mathbb{E}\left(|\tilde{c}_{j_2,k}^{(m)} - \hat{c}_{j_2,k}^{(m)}|^p\right)$$

and

$$D_2 = 2^{j_2(p/2-1)}\sum_{k\in\Lambda_{j_2}}\mathbb{E}\left(|\hat{c}_{j_2,k}^{(m)} - c_{j_2,k}^{(m)}|^p\right).$$

Upper bound for D_2. Proceeding as in (4.6), we obtain

$$D_2 \leq C2^{j_2(p/2-1)}\operatorname{Card}(\Lambda_{j_2})\frac{2^{j_2 mp}}{[n/2]^{p/2}} \leq C2^{j_2 p/2}2^{j_2 mp}\frac{1}{n^{p/2}}. \tag{4.15}$$

Upper bound for D_1. Using the triangular inequality, the definition of $\tilde{c}_{j_2,k}^{(m)}$ (3.14) and (3.3), we arrive at

$$\begin{aligned}
&|\tilde{c}_{j_2,k}^{(m)} - \hat{c}_{j_2,k}^{(m)}| \\
&= \left|(-1)^m\frac{\hat{\mu}}{[n/2]}\sum_{i=1}^{[n/2]}\frac{Y_i}{w(X_i,Y_i)}(\phi_{j,k})^{(m)}(X_i)\left(\frac{1}{\hat{h}(X_i)}\mathbf{1}_{\{|\hat{h}(X_i)|\geq c_3/2\}} - \frac{1}{h(X_i)}\right)\right| \\
&\leq C\frac{1}{[n/2]}\sum_{i=1}^{[n/2]}\frac{|Y_i|}{w(X_i,Y_i)}|(\phi_{j,k})^{(m)}(X_i)|\left|\frac{1}{\hat{h}(X_i)}\mathbf{1}_{\{|\hat{h}(X_i)|\geq c_3/2\}} - \frac{1}{h(X_i)}\right|.
\end{aligned}$$

Owing to the triangular inequality, $\left\{|\hat{h}(X_i)| < c_3/2\right\} \subseteq \left\{|\hat{h}(X_i) - h(X_i)| > c_3/2\right\}$, (3.4) and the Markov inequality, we have

$$\begin{aligned}
&\left|\frac{1}{\hat{h}(X_i)}\mathbf{1}_{\left\{|\hat{h}(X_i)|\geq c_3/2\right\}} - \frac{1}{h(X_i)}\right| \\
&\leq \frac{1}{h(X_i)}\left(\left|\frac{\hat{h}(X_i)-h(X_i)}{\hat{h}(X_i)}\right|\mathbf{1}_{\left\{|\hat{h}(X_i)|\geq c_3/2\right\}} + \mathbf{1}_{\left\{|\hat{h}(X_i)|< c_3/2\right\}}\right) \\
&\leq \frac{1}{c_3}\left(\frac{2}{c_3}\left|\hat{h}(X_i)-h(X_i)\right| + \mathbf{1}_{\left\{|\hat{h}(X_i)-h(X_i)|> c_3/2\right\}}\right) \\
&\leq C|\hat{h}(X_i)-h(X_i)|.
\end{aligned}$$

Hence

$$|\tilde{c}^{(m)}_{j_2,k} - \hat{c}^{(m)}_{j_2,k}| \leq CF_{j_2,k,n},$$

where

$$F_{j,k,n} = \frac{1}{[n/2]}\sum_{i=1}^{[n/2]}\frac{|Y_i|}{w(X_i,Y_i)}|(\phi_{j,k})^{(m)}(X_i)||\hat{h}(X_i)-h(X_i)|.$$

Let us now introduce $W_n = ((X_{[n/2]+1}, Y_{[n/2]+1})\ldots,(X_n,Y_n))$. Using the inequality: $|x+y|^p \leq 2^{p-1}(|x|^p+|y|^p)$, $(x,y)\in\mathbb{R}^2$, we arrive at

$$D_1 \leq C2^{j_2(p/2-1)}\sum_{k\in\Lambda_{j_2}}\mathbb{E}(|F_{j_2,k,n}|^p) \leq C(D_{1,1}+D_{1,2}), \tag{4.16}$$

where

$$D_{1,1} = 2^{j_2(p/2-1)}\sum_{k\in\Lambda_{j_2}}\mathbb{E}\left(\mathbb{E}\left(|F_{j_2,k,n} - \mathbb{E}\left(F_{j_2,k,n}|W_n\right)|^p|W_n\right)\right)$$

and

$$D_{1,2} = 2^{j_2(p/2-1)}\sum_{k\in\Lambda_{j_2}}\mathbb{E}\left(|\mathbb{E}\left(F_{j_2,k,n}|W_n\right)|^p\right).$$

Before bounding $D_{1,1}$ and $D_{1,2}$, let us prove a general moment inequality.

General moment inequality. Let $u\in[1,p]$. Using (3.3), the definition of $f(x,y)$

and (3.5), we arrive at

$$\begin{aligned}
&\mathbb{E}\left(\frac{|Y_1|^u}{w(X_1,Y_1)^u}|(\phi_{j_2,k})^{(m)}(X_1)|^u|\hat{h}(X_1)-h(X_1)|^u\middle|W_n\right)\\
&\leq C\mathbb{E}\left(\frac{|Y_1|^u}{w(X_1,Y_1)}|(\phi_{j_2,k})^{(m)}(X_1)|^u|\hat{h}(X_1)-h(X_1)|^u\middle|W_n\right)\\
&= C\int_{\mathbb{R}}\int_{[0,1]}\frac{|y|^u}{w(x,y)}|(\phi_{j,k})^{(m)}(x)|^u|\hat{h}(x)-h(x)|^u f(x,y)dxdy\\
&= C\int_{\mathbb{R}}\int_{[0,1]}\frac{|y|^u}{w(x,y)}|(\phi_{j,k})^{(m)}(x)|^u|\hat{h}(x)-h(x)|^u \frac{w(x,y)g(x,y)}{\mu}dxdy\\
&\leq C\int_{[0,1]}|(\phi_{j,k})^{(m)}(x)|^u|\hat{h}(x)-h(x)|^u\left(\int_{\mathbb{R}}|y|^u g(x,y)dy\right)dx\\
&\leq C\int_{[0,1]}|(\phi_{j,k})^{(m)}(x)|^u|\hat{h}(x)-h(x)|^u dx.
\end{aligned}$$

The Hölder inequality with the exponents $(p/u, p/(p-u))$ (and the usual modification if $u=p$), $(\phi_{j,k})^{(m)}(x) = 2^{j/2}2^{mj}\phi^{(m)}(2^j-k)$ and a change of variables imply that

$$\begin{aligned}
&\int_{[0,1]}|(\phi_{j,k})^{(m)}(x)|^u|\hat{h}(x)-h(x)|^u dx\\
&\leq \left(\int_{[0,1]}|(\phi_{j,k})^{(m)}(x)|^{pu/(p-u)}dx\right)^{(p-u)/p}||\hat{h}-h||_p^u\\
&= 2^{ju/2}2^{jmu}\left(\int_{[0,1]}|\phi^{(m)}(2^jx-k)|^{pu/(p-u)}dx\right)^{(p-u)/p}||\hat{h}-h||_p^u\\
&\leq C2^{ju/2}2^{jmu}2^{-j(p-u)/p}||\hat{h}-h||_p^u.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{E}\left(\frac{|Y_1|^u}{w(X_1,Y_1)^u}|(\phi_{j_2,k})^{(m)}(X_1)|^u|\hat{h}(X_1)-h(X_1)|^u\middle|W_n\right)\\
&\leq C2^{ju/2}2^{jmu}2^{-j(p-u)/p}||\hat{h}-h||_p^u. \qquad (4.17)
\end{aligned}$$

Let us now bound $D_{1,2}$.

Upper bound for $D_{1,2}$. By (4.17) with $u=1$, we have

$$\begin{aligned}
\mathbb{E}\left(F_{j_2,k,n}|W_n\right) &= \mathbb{E}\left(\frac{|Y_1|}{w(X_1,Y_1)}|(\phi_{j_2,k})^{(m)}(X_1)||\hat{h}(X_1)-h(X_1)|\middle|W_n\right)\\
&\leq C2^{j_2/2}2^{j_2m}2^{-j_2(p-1)/p}||\hat{h}-h||_p.
\end{aligned}$$

Hence

$$\begin{aligned}
D_{1,2} &\leq C2^{j_2(p/2-1)}\operatorname{Card}(\Lambda_{j_2})2^{j_2p/2}2^{j_2mp}2^{-j_2(p-1)}\mathbb{E}\left(\|\hat{h}-h\|_p^p\right)\\
&\leq C2^{(mp+1)j_2}\mathbb{E}\left(\|\hat{h}-h\|_p^p\right). \qquad (4.18)
\end{aligned}$$

Upper bound for $D_{1,1}$. Note that

$$\mathbb{E}\left(|F_{j_2,k,n} - \mathbb{E}\left(F_{j_2,k,n}|W_n\right)|^p|W_n\right) = \frac{1}{[n/2]^p}\mathbb{E}\left(\left|\sum_{i=1}^{[n/2]} U_i\right|^p \middle| W_n\right),$$

with

$$U_i = \frac{|Y_i|}{w(X_i,Y_i)}|(\phi_{j_2,k})^{(m)}(X_i)||\hat{h}(X_i)-h(X_i)|-\mathbb{E}\left(F_{j_2,k,n}|W_n\right), \qquad i \in \{1,\dots,n\}.$$

We aim to apply Lemma 4.1 to $U_1,\dots,U_{[n/2]}$ with the expectation conditionally to W_n.

First of all, note that, conditionally to W_n, $U_1,\dots,U_{[n/2]}$ are *i.i.d.* with $\mathbb{E}(U_1|W_n) = 0$.

Let $u \in \{2,p\}$. The inequality: $(x+y)^u \leq 2^{u-1}(x^u+y^u)$, $(x,y) \in \mathbb{R}^2$, the Hölder inequality and (4.17) imply that

$$\begin{aligned}\mathbb{E}\left(U_1^u|W_n\right) &\leq 2^u\mathbb{E}\left(\frac{|Y_1|^u}{w(X_1,Y_1)^u}|(\phi_{j_2,k})^{(m)}(X_1)|^u|\hat{h}(X_1)-h(X_1)|^u\middle|W_n\right)\\ &\leq C2^{j_2u/2}2^{j_2mu}2^{-j_2(p-u)/p}||\hat{h}-h||_p^u.\end{aligned}$$

Thus, thanks to Lemma 4.1 with $\gamma = p$, we have

$$\begin{aligned}&\mathbb{E}\left(|F_{j_2,k,n} - \mathbb{E}\left(F_{j_2,k,n}|W_n\right)|^p|W_n\right)\\ &\leq C\frac{1}{n^p}\max\left(n\mathbb{E}(U_1^p|W_n), n^{p/2}(\mathbb{E}(U_1^2|W_n))^{p/2}\right)\\ &\leq C\frac{1}{n^p}\max\left(n2^{j_2p/2}2^{jmp}||\hat{h}-h||_p^p, n^{p/2}(2^{j_2}2^{2mj_2}2^{-j_2(p-2)/p}||\hat{h}-h||_p^2)^{p/2}\right)\\ &\leq C\frac{1}{n^p}2^{j_2mp}\max\left(n2^{j_2p/2}, n^{p/2}2^{j_2}\right)||\hat{h}-h||_p^p.\end{aligned}$$

Hence, by $2^{j_2} \leq n$,

$$\begin{aligned}D_{1,1} &\leq C\frac{1}{n^p}2^{j_2(p/2-1)}\,\mathrm{Card}(\Lambda_{j_2})2^{j_2mp}\max\left(n2^{j_2p/2}, n^{p/2}2^{j_2}\right)\mathbb{E}\left(||\hat{h}-h||_p^p\right)\\ &\leq C2^{(mp+1)j_2}\left(\frac{1}{n^p}\max\left(n2^{j_2(p-1)}, n^{p/2}2^{j_2p/2}\right)\right)\mathbb{E}\left(||\hat{h}-h||_p^p\right)\\ &\leq C2^{(mp+1)j_2}\mathbb{E}\left(||\hat{h}-h||_p^p\right).\end{aligned} \tag{4.19}$$

Putting (4.16), (4.18) and (4.19) together and using $p \geq 2$, we get

$$D_1 \leq C2^{(mp+1)j_2}\mathbb{E}\left(||\hat{h}-h||_p^p\right) \leq C2^{j_2p/2}2^{j_2mp}\mathbb{E}\left(||\hat{h}-h||_p^p\right). \tag{4.20}$$

By (4.14), (4.15) and (4.20) together, we arrive at

$$D \leq C2^{j_2p/2}2^{j_2mp}\max\left(\mathbb{E}\left(\|\hat{h}-h\|_p^p\right), \frac{1}{n^{p/2}}\right). \tag{4.21}$$

Combining (4.12), (4.13) and (4.21), we obtain

$$\mathbb{E}\left(\|\hat{\varphi}_2^{(m)} - \varphi^{(m)}\|_p^p\right) \leq C\left(2^{j_2 p/2} 2^{j_2 mp} \max\left(\mathbb{E}\left(\|\hat{h} - h\|_p^p\right), \frac{1}{n^{p/2}}\right) + 2^{-j_2 s_* p}\right) \quad (4.22)$$

Since $h \in B^{s_2}_{q_2,r_2}(M_2)$ with $M_2 > 0$, $q_2 \geq 1$, $r_2 \geq 1$ and $s_2 \in (\max(1/q_2 - 1/p, 0), N)$, with j_1 as (3.15), Theorem 3.2 ensures the existance of a constant $C > 0$ such that

$$\mathbb{E}\left(\|\hat{h} - h\|_p^p\right) \leq C(n - [n/2])^{-s_o p/(2s_o+1)} \leq Cn^{-s_o p/(2s_o+1)}.$$

Therefore, chosing j_2 as (3.16), it follows from (4.22) that

$$\begin{aligned}\mathbb{E}\left(\|\hat{\varphi}_2^{(m)} - \varphi^{(m)}\|_p^p\right) &\leq C\left(2^{j_2 p/2} 2^{j_2 mp} n^{-s_o p/(2s_o+1)} + 2^{-j_2 s_* p}\right) \\ &\leq Cn^{-2s_* s_o p/((2s_o+1)(2s_*+2m+1))}. \quad (4.23)\end{aligned}$$

The case $p \in [1, 2)$ is an immediate consequence: using the Hölder inequality with the exponent $2/p \geq 1$ and (4.23) with $p = 2$, we obtain

$$\begin{aligned}\mathbb{E}\left(\|\hat{\varphi}_2^{(m)} - \varphi^{(m)}\|_p^p\right) &\leq \left(\mathbb{E}\left(\|\hat{\varphi}_2^{(m)} - \varphi^{(m)}\|_2^2\right)\right)^{p/2} \\ &\leq C\left(n^{-4s_* s_o/((2s_o+1)(2s_*+2m+1))}\right)^{p/2} \\ &= Cn^{-2s_* s_o p/((2s_o+1)(2s_*+2m+1))}.\end{aligned}$$

This completes the proof of Theorem 3.3.

□

References

Ahmad, I.A. (1995). On multivariate kernel estimation for samples from weighted distributions. *Statistics and Probability Letters*, **22**, 121-129.

Chaubey, Y.P., Chesneau, C. and Doosti (2011). On linear wavelet density estimation: Some recent developments. *J. Ind. Soc. Agr. Stat.*, **65**, 169-179.

Chaubey, Y.P., Laïb, N. and Li, J. (2012). Generalized kernel regression estimator for dependent size-biased data. *J. Statist. Plann. Inference*, **142**, 708-727.

Chaubey, Y., Chesneau, C. and Shirazi, E. (2013). Wavelet-based estimation of regression function for dependent biased data under a given random design. *Journal of Nonparametric Statistics*, **25**, 1, 53-71.

Chaubey, Y.P. and Shirazi, E. (2014). On MISE of a nonlinear wavelet estimator of the regression function based on biased data under strong mixing. *Communications in Statistics: Theory and Methods*, (to appear).

Chesneau, C. (2014). A note on wavelet estimation of the derivatives of a regression function in a random design setting. *International Journal of Mathematics and Mathematical Sciences*, Volume 2014, Article ID 195765, 8 pages.

Chesneau, C., Kachour, M. and Navarro, F. (2014). Average derivative estimation from biased data. *ISRN Probability and Statistics*, Volume 2014, Article ID 864530, 7 pages.

Chesneau, C. and Shirazi, E. (2014). Nonparametric wavelet regression based on biased data. *Communications in Statistics - Theory and Methods*, **43**, 2642-2658.

Cohen, A., Daubechies, I., Jawerth, B. and Vial , P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, **24**, 54-81.

Cristóbal, J.A. and Alcalá, J.T. (2000). Nonparametric regression estimators for length biased data. *J. Statist. Plann. Inference*, **89**, 145-168.

Cristóbal, J.A. and Alcalá, J.T. (2001). An overview of nonparametric contributions to the problem of functional estimation from biased data. *Test*, **10**, 09-332.

Cristòbal, J.A., Ojeda, J.L. and Alcalà, J.T. (2004). Confidence bands in nonparametric regression with length biased data. *Ann. Inst. Statist. Math.*, **56,** 475-496.

Donoho, D., Johnstone, I., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, **24**, 508-539.

Hall, B. (2010). *Nonparametric estimation of derivatives with applications*, University of Kentucky Doctoral Dissertations. Paper 114.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics, New York 129, Springer Verlag.

Mallat, S. (2009). *A wavelet tour of signal processing (Third Edition): The Sparse Way.* Elsevier/ Academic Press, Amsterdam.

Meyer, Y. (1992). *Wavelets and Operators.* Cambridge University Press, Cambridge.

Ojeda, J.L., Gonzáles-Manteiga, W. and Cristóbal, J.A. (2007). it A bootstrap based model checking for selection-biased data, Universidade de Santiago de Compostela, Report 07-05.

Ojeda-Cabrera, J. and Van Keilegom, I. (2009). Goodness-of-fit tests for parametric regression with selection biased data. *J. Statist. Planning Inf.*, **139**, 2836-2850.

Pensky, M. and Vidakovic, B. (2001). On non-equally spaced wavelet regression. *Annals of the Institute of Statistical Mathematics*, **53**, 681-690.

Ramirez, P. and Vidakovic, B. (2010). Wavelet density estimation for stratified size-biased sample. *Journal of Statistical Planning and Inference*, **140**, 419-432.

Rosenthal, H.P. (1970). On the subspaces of $\mathbb{L}^p$ $(p \geq 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics*, **8**, 273-303.

Sköld, M. (1999). Kernel regression in the presence of size-bias. *Journal of Nonparametric Statistics*, **12**, 41-51.

Wu, C. O. (2000). Local polynomial regression with selection biased data. *Statist. Sinica*, **10**, 789-817.