



HAL
open science

A note on wavelet estimation of the derivatives of a regression function in a random design setting

Christophe Chesneau

► **To cite this version:**

Christophe Chesneau. A note on wavelet estimation of the derivatives of a regression function in a random design setting. 2014. hal-00925546

HAL Id: hal-00925546

<https://hal.science/hal-00925546>

Preprint submitted on 8 Jan 2014

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

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

A note on wavelet estimation of the derivatives of a regression function in a random design setting

Christophe Chesneau

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

Abstract: We investigate the estimation of the derivatives of a regression function in the nonparametric regression model with random design. New wavelet estimators are developed. Their performances are evaluated via the mean integrated squared error. Fast rates of convergence are obtained for a wide class of unknown functions.

AMS 2000 subject classifications: 62G07, 62G20.

Keywords and phrases: Nonparametric regression, Derivatives function estimation, Wavelets, Besov balls, Hard thresholding.

1. Introduction

We consider the nonparametric regression model with random design described as follows. Let $(Y_1, X_1), \dots, (Y_n, X_n)$ be n random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \dots, n, \quad (1.1)$$

ξ_1, \dots, ξ_n are n i.i.d. random variables such that $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}(\xi_1^2) < \infty$, X_1, \dots, X_n are n i.i.d. random variables with common density $g : [0, 1] \rightarrow [0, \infty)$ and $f : [0, 1] \rightarrow \mathbb{R}$ is an unknown regression function. It is assumed that X_i and ξ_i are independent for any $i = 1, \dots, n$. We aim to estimate $f^{(m)}$, i.e. the m -th derivative of f , for any integer m , from $(Y_1, X_1), \dots, (Y_n, X_n)$.

In the literature, various estimation methods have been proposed and studied. The main ones are the kernel methods (see, e.g., Gasser and Müller (1984), Härdle and Gasser (1985), Mack and Müller (1989), Ruppert and Wand (1994) and Wand and Jones (1995)), the smoothing splines and local polynomial methods (see, e.g., Stone (1985), Wahba and Wang (1990), Zhou and Wolfe (2000) and Jarrow *et al.* (2004)). The object of this note is to introduce new efficient estimators based on wavelet methods. Contrary to the others, they have the benefit of enjoying local adaptivity against discontinuities thanks to the use of a multiresolution analysis. Reviews on wavelet methods can be found in, e.g., Antoniadis (1997), Härdle *et al.* (1998) and Vidakovic (1999). To the best of our knowledge, only Cai (2002) and Petsa and Sapatinas (2011) have proposed

wavelet estimators for $f^{(m)}$ from (1.1) but defined with a deterministic equidistant design, i.e., $X_i = i/n$. The consideration of a random design complicates significantly the problem and no wavelet estimators exist in this case. This motivates our study.

In a first part, assuming that g is known, we propose two wavelet estimators: the first one is linear nonadaptive and the second one, nonlinear adaptive. Both use the approach of Prakasa Rao (1996) initially developed in the context of the density estimation problem. Then we determine their rates of convergence by considering the mean integrated squared error (MISE) and assuming that $f^{(m)}$ belongs to Besov balls. In a second part, we develop a linear wavelet estimator in the case where g is unknown. It is derived from the one introduced by Pensky and Vidakovic (2001) considering the estimation of $f^{(0)} = f$ from (1.1). We evaluate its rate of convergence again under the MISE over Besov balls. The obtained rates of convergence are similar those attained by wavelet estimators for the derivatives of a density (see, e.g., Prakasa Rao (1996), Chaubey *et al.* (2006, 2008)).

The organization of this note is as follows. The next section describes some basics on wavelets and Besov balls. Our estimators and their rates of convergence are presented in Section 3. The proofs are carried out in Section 4.

2. Preliminaries

This section is devoted to the presentation of the considered wavelet basis and the Besov balls.

2.1. Wavelet basis

We set

$$\mathbb{L}^2([0, 1]) = \left\{ h : [0, 1] \rightarrow \mathbb{R}; \|h\|_2 = \left(\int_0^1 (h(x))^2 dx \right)^{1/2} < \infty \right\}.$$

We consider the wavelet basis on $[0, 1]$ introduced by Cohen *et al.* (1993). Let ϕ and ψ be the initial wavelet functions of the Daubechies wavelets family $db2N$ with $N \geq 1$ (see, e.g., Daubechies (1992)). These functions have the distinction of being compactly supported and belong to the class C^a for $N > 5a$. For any $j \geq 0$, we set $\Lambda_j = \{0, \dots, 2^j - 1\}$ and, for $k \in \Lambda_j$,

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

With appropriated treatments at the boundaries, there exists an integer τ such that, for any integer $\ell \geq \tau$,

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

forms an orthonormal basis of $\mathbb{L}^2([0, 1])$. For any integer $\ell \geq \tau$ and $h \in \mathbb{L}^2([0, 1])$, we have the following wavelet expansion:

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

where

$$c_{j,k} = \int_0^1 h(x) \phi_{j,k}(x) dx, \quad d_{j,k} = \int_0^1 h(x) \psi_{j,k}(x) dx. \quad (2.1)$$

These quantities are called the wavelet coefficients of h . See, e.g., Cohen *et al.* (1993) and Mallat (2009).

2.2. Besov balls

We consider the following wavelet sequential definition of the Besov balls. We say that $h \in B_{p,r}^s(M)$ with $s > 0$, $p \geq 1$, $r \geq 1$ and $M > 0$ if there exists a constant $C > 0$ such that $c_{j,k}$ and $d_{j,k}$ (2.1) satisfy

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

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

The interest of Besov balls is to contain various kinds of homogeneous and inhomogeneous functions h . For particular choices of s , p and r , $B_{p,r}^s(M)$ correspond to standard balls of function spaces, as the Hölder and Sobolev balls (see, e.g., Meyer (1992) and Härdle *et al.* (1998)).

3. Results

In this section, we set the assumptions on the model, present our wavelet estimators and determine their rates of convergence under the MISE over Besov balls.

3.1. Assumptions

We formulate the following assumptions:

- (K1) We have $f^{(q)}(0) = f^{(q)}(1) = 0$ for any $q \in \{0, \dots, m\}$.
- (K2) There exists a constant $C_1 > 0$ such that

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

(K3) There exists a constant $c_2 > 0$ such that

$$c_2 \leq \inf_{x \in [0,1]} g(x).$$

(K4) There exists a constant $C_3 > 0$ such that

$$\sup_{x \in [0,1]} g(x) \leq C_3.$$

3.2. Wavelet estimators: when g is known

We consider the wavelet basis \mathcal{B} with $N > 5m$ to ensure that ϕ and ψ belongs to \mathcal{C}^m .

Linear wavelet estimator. We define the linear wavelet estimator $\hat{f}_1^{(m)}$ by

$$\hat{f}_1^{(m)}(x) = \sum_{k \in \Lambda_{j_0}} \hat{c}_{j_0,k}^{(m)} \phi_{j_0,k}(x), \quad x \in [0, 1], \quad (3.1)$$

where

$$\hat{c}_{j,k}^{(m)} = \frac{(-1)^m}{n} \sum_{i=1}^n \frac{Y_i}{g(X_i)} (\phi_{j,k})^{(m)}(X_i) \quad (3.2)$$

and j_0 is an integer chosen a posteriori.

The definition of $\hat{c}_{j,k}^{(m)}$ is motivated by the following unbiased property: using the independence between X_1 and ξ_1 , $\mathbb{E}(\xi_1) = 0$, and m integrations by parts with **(K1)**, we obtain

$$\begin{aligned} \mathbb{E}(\hat{c}_{j,k}^{(m)}) &= \mathbb{E} \left((-1)^m \frac{Y_1}{g(X_1)} (\phi_{j,k})^{(m)}(X_1) \right) \\ &= \mathbb{E} \left((-1)^m \frac{f(X_1)}{g(X_1)} (\phi_{j,k})^{(m)}(X_1) \right) + \mathbb{E}(\xi_1) \mathbb{E} \left((-1)^m \frac{1}{g(X_1)} (\phi_{j,k})^{(m)}(X_1) \right) \\ &= (-1)^m \mathbb{E} \left(\frac{f(X_1)}{g(X_1)} (\phi_{j,k})^{(m)}(X_1) \right) = (-1)^m \int_0^1 \frac{f(x)}{g(x)} (\phi_{j,k})^{(m)}(x) g(x) dx \\ &= (-1)^m \int_0^1 f(x) (\phi_{j,k})^{(m)}(x) dx = \int_0^1 f^{(m)}(x) \phi_{j,k}(x) dx = c_{j,k}^{(m)}, \end{aligned} \quad (3.3)$$

which is the wavelet coefficient of $f^{(m)}$ associated to $\phi_{j,k}$.

This approach was initially introduced by Prakasa Rao (1996) for the estimation of the derivatives of a density. Its adaptation to (1.1) gives a suitable alternative to the wavelet methods developed by Cai (2002) and Petsa and Sapatinas (2011) in the case $X_i = i/n$, specially in the treatment of the random design.

Note that, for the standard case $m = 0$, this estimator has been considered and studied in Chesneau (2007).

Theorem 3.1 below investigates the rate of convergence attained by $\hat{f}_1^{(m)}$ under the MISE assuming that $f^{(m)}$ belongs to Besov balls.

Theorem 3.1. *Suppose that (K1), (K2) and (K3) are satisfied and that $f^{(m)} \in B_{p,r}^s(M)$ with $M > 0$, $p \geq 1$, $r \geq 1$ and $s \in (\max(1/p - 1/2, 0), N)$. Let $\hat{f}_1^{(m)}$ be defined by (3.1) with j_0 such that*

$$2^{j_0} = \lceil n^{1/(2s_*+2m+1)} \rceil, \quad (3.4)$$

$s_* = s + \min(1/2 - 1/p, 0)$ and $[a]$ denotes the integer part of a .

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

$$\mathbb{E} \left(\|\hat{f}_1^{(m)} - f^{(m)}\|_2^2 \right) \leq C n^{-2s_*/(2s_*+2m+1)}.$$

The rate of convergence $n^{-2s_*/(2s_*+2m+1)}$ corresponds to the one obtained in the derivatives density estimation framework. See, e.g., Prakasa Rao (1996), Chaubey *et al.* (2006, 2008). For $m = 0$, Theorem 3.1 becomes (Chesneau, 2007, Theorem 3.1, $p = 2$).

Hard thresholding wavelet estimator. We define the hard thresholding wavelet estimator $\hat{f}_2^{(m)}$ by

$$\hat{f}_2^{(m)}(x) = \sum_{k \in \Lambda_\tau} \hat{c}_{\tau,k}^{(m)} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} \hat{d}_{j,k}^{(m)} \mathbf{1}_{\{|\hat{d}_{j,k}^{(m)}| \geq \kappa \lambda_j\}} \psi_{j,k}(x), \quad (3.5)$$

$x \in [0, 1]$, where $\hat{c}_{j,k}^{(m)}$ is defined by (3.2),

$$\hat{d}_{j,k}^{(m)} = \frac{(-1)^m}{n} \sum_{i=1}^n \frac{Y_i}{g(X_i)} (\psi_{j,k})^{(m)}(X_i) \mathbf{1}_{\left\{ \left| \frac{Y_i}{g(X_i)} (\psi_{j,k})^{(m)}(X_i) \right| \leq \varsigma_j \right\}}, \quad (3.6)$$

$\mathbf{1}$ is the indicator function, $\kappa > 0$ is a large enough constant, j_1 is the integer satisfying

$$2^{j_1} = \lceil n^{1/(2m+1)} \rceil,$$

$$\varsigma_j = \theta_\psi 2^{mj} \sqrt{\frac{n}{\ln n}}, \quad \lambda_j = \theta_\psi 2^{mj} \sqrt{\frac{\ln n}{n}}$$

and $\theta_\psi = \sqrt{(2/c_2)(C_1^2 + \mathbb{E}(\xi_1^2)) \|\psi^{(m)}\|_2^2}$.

The construction of $\hat{f}_2^{(m)}$ is an adaptation of the hard thresholding wavelet estimator introduced by Delyon and Juditsky (1996) to the estimation of $f^{(m)}$ from (1.1). It used the modern version developed by Chaubey *et al.* (2013). The advantage of $\hat{f}_2^{(m)}$ over $\hat{f}_1^{(m)}$ (3.1) is that $\hat{f}_2^{(m)}$ is adaptive ; thanks to the thresholding in (3.6), its performance does not depend on the knowledge of the smoothness of $f^{(m)}$. The second thresholding in (3.6) enables us to relax some assumptions on the model, and, in particular, to only suppose $\mathbb{E}(\xi_1^2) < \infty$ on ξ_1 (its density can be unknown). Basics and important results on hard thresholding wavelet estimators can be found in, e.g., Donoho and Johnstone (1994, 1995), Donoho *et al.* (1995, 1996) and Delyon and Juditsky (1996).

Theorem 3.2 below determines the rate of convergence attained by $\hat{f}_2^{(m)}$ under the MISE assuming that $f^{(m)}$ belongs to Besov balls.

Theorem 3.2. *Suppose that **(K1)**, **(K2)** and **(K3)** are satisfied and that $f^{(m)} \in B_{p,r}^s(M)$ with $M > 0$, $r \geq 1$, $\{p \geq 2 \text{ and } s \in (0, N)\}$ or $\{p \in [1, 2] \text{ and } s \in ((2m + 1)/p, N)\}$. Let $\hat{f}_2^{(m)}$ be defined by (3.5). Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left(\|\hat{f}_2^{(m)} - f^{(m)}\|_2^2 \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+2m+1)}.$$

The proof is based on a general result proved by (Chaubey *et al.*, 2013, Theorem 6.1). Let us observe that, for the case $p \geq 2$, $(\ln n/n)^{2s/(2s+2m+1)}$ is equal to the rate of convergence attained by $\hat{f}_1^{(m)}$ up to a logarithmic factor (see Theorem 3.1). However, for the case $p \in [1, 2)$, it is significantly better in terms of power.

3.3. Wavelet estimators: when g is unknown

In the case where g is unknown, we propose the linear wavelet estimator $\hat{f}_3^{(m)}$ defined by

$$\hat{f}_3^{(m)}(x) = \sum_{k \in \Lambda_{j_2}} \tilde{c}_{j_2,k}^{(m)} \phi_{j_2,k}(x), \quad x \in [0, 1], \quad (3.7)$$

where

$$\tilde{c}_{j,k}^{(m)} = \frac{(-1)^m}{a_n} \sum_{i=1}^{a_n} \frac{Y_i}{\hat{g}(X_i)} \mathbf{1}_{\{|\hat{g}(X_i)| \geq c_2/2\}} (\phi_{j,k})^{(m)}(X_i), \quad (3.8)$$

$a_n = \lfloor n/2 \rfloor$, j_2 is an integer chosen a posteriori, c_2 refers to **(K3)** and \hat{g} is an estimator of g constructed from the random variables $U_n = (X_{a_n+1}, \dots, X_n)$. For instance, we can consider the linear wavelet estimator \hat{g} by

$$\hat{g}(x) = \sum_{k \in \Lambda_{j_3}} \bar{c}_{j_3,k} \phi_{j_3,k}(x), \quad x \in [0, 1], \quad (3.9)$$

where

$$\bar{c}_{j,k} = \frac{1}{n - a_n} \sum_{i=1}^{n-a_n} \phi_{j,k}(X_{a_n+i}),$$

and j_3 is an integer chosen a posteriori.

The estimator $\hat{f}_3^{(m)}$ is close to the ‘‘NES linear wavelet estimator’’ proposed by Pensky and Vidakovic (2001) for $m = 0$. However, there are notable differences in the thresholding in (3.8), the partitioning of the variables and the definition of \hat{g} , making the study of its performance under the MISE more simpler (see the proofs of Theorem 3.3 below).

Theorem 3.3 below determines an upper bound of the MISE of $\hat{f}_3^{(m)}$, then exhibits its rate of convergence when $f^{(m)}$ belongs to Besov balls.

Theorem 3.3. *Suppose that **(K1)**, **(K2)** and **(K3)** are satisfied and that $f^{(m)} \in B_{p_1, r_1}^{s_1}(M_1)$ with $M_1 > 0$, $p_1 \geq 1$, $r_1 \geq 1$ and $s_1 \in (\max(1/p_1 - 1/2, 0), N)$. Let $\hat{f}_3^{(m)}$ be defined by (3.7). Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left(\|\hat{f}_3^{(m)} - f^{(m)}\|_2^2 \right) \leq C \left(2^{(2m+1)j_2} \max \left(\mathbb{E} (\|\hat{g} - g\|_2^2), \frac{1}{n} \right) + 2^{-2j_2 s_*} \right),$$

with $s_* = s_1 + \min(1/2 - 1/p_1, 0)$.

In addition, suppose that **(K4)** is satisfied, $g \in B_{p_2, r_2}^{s_2}(M_2)$ with $M_2 > 0$, $p_2 \geq 1$, $r_2 \geq 1$ and $s_2 \in (\max(1/p_2 - 1/2, 0), N)$, consider $\hat{f}_3^{(m)}$ with the estimator \hat{g} defined by (3.9) with j_3 such that

$$2^{j_3} = \lfloor (n - a_n)^{1/(2s_o+1)} \rfloor, \quad (3.10)$$

$s_o = s_2 + \min(1/2 - 1/p_2, 0)$ and j_2 such that

$$2^{j_2} = \lfloor n^{2s_o/((2s_o+1)(2s_*+2m+1))} \rfloor. \quad (3.11)$$

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

$$\mathbb{E} \left(\|\hat{f}_3^{(m)} - f^{(m)}\|_2^2 \right) \leq C n^{-4s_* s_o/((2s_o+1)(2s_*+2m+1))}.$$

The first point of Theorem 3.3 is proved for any estimator \hat{g} of g depending on U_n . Taking $\hat{g} = g$, it corresponds to the upper bound of the MISE for $\hat{f}_1^{(m)}$ established in the proof of Theorem 3.1. Note that the rate of convergence described in the second point is slower to the one attained by $\hat{f}_1^{(m)}$ (see Theorem 3.1). The fact that the smoothness of g influences the performance of \hat{g} and, a fortiori, $\hat{f}_3^{(m)}$, seems natural. This phenomenon also appears in (Pensky and Vidakovic, 2001, Theorem 2.1) for $m = 0$.

Remark 3.1. *If c_2 exists but is unknown, we can defined $\hat{f}_3^{(m)}$ as (3.7) with $1/\ln n$ instead of c_2 in the threshold of (3.8). The impact of this modification is a logarithmic term in Theorem 3.3, i.e.,*

$$\mathbb{E} \left(\|\hat{f}_3^{(m)} - f^{(m)}\|_2^2 \right) \leq C \left(2^{(2m+1)j_2} \max \left((\ln n) \mathbb{E} (\|\hat{g} - g\|_2^2), \frac{1}{n} \right) + 2^{-2j_2 s_*} \right).$$

Moreover, choosing j_2 such that

$$2^{j_2} = \lfloor n^{2s_o/((2s_o+1)(2s_*+2m+1))} (\ln n)^{-1/(2s_*+2m+1)} \rfloor,$$

there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\|\hat{f}_3^{(m)} - f^{(m)}\|_2^2 \right) \leq C n^{-4s_* s_o/((2s_o+1)(2s_*+2m+1))} (\ln n)^{2s_*/(2s_*+2m+1)}.$$

Remark 3.2. *Note that the assumption **(K4)** has been only used in the second point of Theorem 3.3.*

Conclusion and perspective. We explore the estimation of $f^{(m)}$ from (1.1). Distinguishing the cases where g is known or not, we propose wavelet methods and prove that they attain fast rates of convergence under the MISE assuming that $f^{(m)} \in B_{p,r}^s(M)$.

A perspective of this work will be to develop an adaptive wavelet estimator, as the hard thresholding one, for the estimation of $f^{(m)}$ in the case where g is unknown. The extension of $\hat{f}_3^{(m)}$ in this sense is not trivial and new theoretical problems raised. Another perspective is the consideration of dependent $(Y_1, X_1), \dots, (Y_n, X_n)$. These two aspects need further investigations that we leave for a future work.

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 Theorem 3.1. First of all, we expand the function $f^{(m)}$ on \mathcal{B} at the level j_0 given by (3.4) :

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

where $c_{j_0,k}^{(m)} = \int_0^1 f^{(m)}(x) \phi_{j_0,k}(x) dx$ and $d_{j,k}^{(m)} = \int_0^1 f^{(m)}(x) \psi_{j,k}(x) dx$.

Since \mathcal{B} forms an orthonormal basis of $\mathbb{L}^2([0, 1])$, we get

$$\mathbb{E} \left(\|\hat{f}_1^{(m)} - f^{(m)}\|_2^2 \right) = \sum_{k \in \Lambda_{j_0}} \mathbb{E} \left(\left(\hat{c}_{j_0,k}^{(m)} - c_{j_0,k}^{(m)} \right)^2 \right) + \sum_{j=j_0}^{\infty} \sum_{k \in \Lambda_j} (d_{j,k}^{(m)})^2. \quad (4.1)$$

Using the fact that $\hat{c}_{j_0,k}^{(m)}$ is an unbiased estimator of $c_{j_0,k}^{(m)}$ (see (3.3)), $(Y_1, X_1), \dots, (Y_n, X_n)$ are i.i.d., the inequalities : $\mathbb{V}(D) \leq \mathbb{E}(D^2)$ for any random variable D and $(a+b)^2 \leq 2(a^2 + b^2)$, $(a, b) \in \mathbb{R}^2$, **(K2)** and **(K3)**, we have

$$\begin{aligned} & \mathbb{E} \left(\left(\hat{c}_{j_0,k}^{(m)} - c_{j_0,k}^{(m)} \right)^2 \right) = \mathbb{V} \left(\hat{c}_{j_0,k}^{(m)} \right) \\ & = \frac{1}{n} \mathbb{V} \left(\frac{Y_1}{g(X_1)} (\phi_{j_0,k})^{(m)}(X_1) \right) \\ & \leq \frac{1}{n} \mathbb{E} \left(\left(\frac{Y_1}{g(X_1)} (\phi_{j_0,k})^{(m)}(X_1) \right)^2 \right) \\ & \leq \frac{1}{n} 2 \mathbb{E} \left(\frac{(f(X_1))^2 + \xi_1^2}{(g(X_1))^2} ((\phi_{j_0,k})^{(m)}(X_1))^2 \right) \\ & \leq \frac{1}{n} \frac{2}{c_2} (C_1^2 + \mathbb{E}(\xi_1^2)) \mathbb{E} \left(\frac{1}{g(X_1)} ((\phi_{j_0,k})^{(m)}(X_1))^2 \right). \end{aligned} \quad (4.2)$$

Using $(\phi_{j_0,k})^{(m)}(x) = 2^{(j_0/2)(2m+1)}\phi^{(m)}(2^{j_0}x - k)$, the change of variables $y = 2^{j_0}x - k$ and the fact that ϕ is compactly supported, we obtain

$$\begin{aligned} \mathbb{E}\left(\frac{1}{g(X_1)}((\phi_{j_0,k})^{(m)}(X_1))^2\right) &= \int_0^1 \frac{1}{g(x)}((\phi_{j_0,k})^{(m)}(x))^2 g(x) dx \\ &= 2^{2mj_0} \int_0^1 2^{j_0} (\phi^{(m)}(2^{j_0}x - k))^2 dx \\ &\leq 2^{2mj_0} \|\phi^{(m)}\|_2^2. \end{aligned} \quad (4.3)$$

Therefore

$$\mathbb{E}\left(\left(\hat{c}_{j_0,k}^{(m)} - c_{j_0,k}^{(m)}\right)^2\right) \leq C 2^{2mj_0} \frac{1}{n}$$

and, for j_0 satisfying (3.4), it holds

$$\sum_{k \in \Lambda_{j_0}} \mathbb{E}\left(\left(\hat{c}_{j_0,k}^{(m)} - c_{j_0,k}^{(m)}\right)^2\right) \leq C 2^{(2m+1)j_0} \frac{1}{n} \leq C n^{-2s_*/(2s_*+2m+1)}. \quad (4.4)$$

On the other hand, we have $f^{(m)} \in B_{p,r}^s(M) \subseteq B_{2,\infty}^{s_*}(M)$ [see Härdle *et al.* (1998), Corollary 9.2], which implies

$$\sum_{j=j_0}^{\infty} \sum_{k \in \Lambda_j} (d_{j,k}^{(m)})^2 \leq C 2^{-2j_0 s_*} \leq C n^{-2s_*/(2s_*+2m+1)}. \quad (4.5)$$

It follows from (4.1), (4.4) and (4.5) that

$$\mathbb{E}\left(\|\hat{f}_1^{(m)} - f^{(m)}\|_2^2\right) \leq C n^{-2s_*/(2s_*+2m+1)}.$$

Theorem 3.1 is proved.

Proof of Theorem 3.2. Observe that, for $\gamma \in \{\phi, \psi\}$, any integer $j \geq \tau$ and any $k \in \Lambda_j$,

- using arguments similar to (3.3), we obtain

$$\mathbb{E}\left(\frac{(-1)^m}{n} \sum_{i=1}^n \frac{Y_i}{g(X_i)} (\gamma_{j,k})^{(m)}(X_i)\right) = \int_0^1 f^{(m)}(x) \gamma_{j,k}(x) dx.$$

- using arguments similar to (4.2) and (4.3), we have

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}\left(\left((-1)^m \frac{Y_i}{g(X_i)} (\gamma_{j,k})^{(m)}(X_i)\right)^2\right) &= n \mathbb{E}\left(\left(\frac{Y_1}{g(X_1)} (\gamma_{j,k})^{(m)}(X_1)\right)^2\right) \\ &\leq C_*^2 n 2^{2mj}, \end{aligned}$$

with $C_*^2 = (2/c_2)(C_1^2 + \mathbb{E}(\xi_1^2)) \|\gamma^{(m)}\|_2^2$.

Applying (Chaubey *et al.*, 2013, Theorem 6.1) (presented in Appendix) with $\mu_n = \nu_n = n$, $\delta = m$, $\theta_\gamma = C_*$, $W_i = (Y_i, X_i)$,

$$q_i(\gamma, (y, x)) = (-1)^m \frac{y}{g(x)} \gamma^{(m)}(x)$$

and $f^{(m)} \in B_{p,r}^s(M)$ with $M > 0$, $r \geq 1$, either $\{p \geq 2$ and $s \in (0, N)\}$ or $\{p \in [1, 2)$ and $s \in (1/p, N)\}$, we prove the existence of a constant $C > 0$ such that

$$\mathbb{E} \left(\|\hat{f}_2^{(m)} - f^{(m)}\|_2^2 \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+2m+1)}.$$

Theorem 3.2 is proved. \square

Proof of Theorem 3.3. As in the proof of Theorem 3.1, we first expand the function $f^{(m)}$ on \mathcal{B} at the level j_2 given by (3.11) :

$$f^{(m)}(x) = \sum_{k \in \Lambda_{j_2}} c_{j_2,k}^{(m)} \phi_{j_2,k}(x) + \sum_{j=j_2}^{\infty} \sum_{k \in \Lambda_j} d_{j,k}^{(m)} \psi_{j,k}(x).$$

Since \mathcal{B} forms an orthonormal basis of $\mathbb{L}^2([0, 1])$, we get

$$\mathbb{E} \left(\|\hat{f}_3^{(m)} - f^{(m)}\|_2^2 \right) = \sum_{k \in \Lambda_{j_2}} \mathbb{E} \left(\left(\tilde{c}_{j_2,k}^{(m)} - c_{j_2,k}^{(m)} \right)^2 \right) + \sum_{j=j_2}^{\infty} \sum_{k \in \Lambda_j} (d_{j,k}^{(m)})^2. \quad (4.6)$$

Using $f^{(m)} \in B_{p,r}^s(M) \subseteq B_{2,\infty}^{s^*}(M)$ [see Härdle *et al.* (1998), Corollary 9.2], we have

$$\sum_{j=j_2}^{\infty} \sum_{k \in \Lambda_j} (d_{j,k}^{(m)})^2 \leq C 2^{-2j_2 s^*}. \quad (4.7)$$

Let $\hat{c}_{j_2,k}^{(m)}$ be (3.2) with $n = a_n$ and $j = j_2$ (3.11). The elementary inequality: $(a + b)^2 \leq 2(a^2 + b^2)$, $(a, b) \in \mathbb{R}^2$, yields

$$\sum_{k \in \Lambda_{j_2}} \mathbb{E} \left(\left(\tilde{c}_{j_2,k}^{(m)} - \hat{c}_{j_2,k}^{(m)} \right)^2 \right) \leq 2(T_1 + T_2), \quad (4.8)$$

where

$$T_1 = \sum_{k \in \Lambda_{j_2}} \mathbb{E} \left(\left(\tilde{c}_{j_2,k}^{(m)} - \hat{c}_{j_2,k}^{(m)} \right)^2 \right), \quad T_2 = \sum_{k \in \Lambda_{j_2}} \mathbb{E} \left(\left(\hat{c}_{j_2,k}^{(m)} - c_{j_2,k}^{(m)} \right)^2 \right).$$

Proceeding as in (4.4), we get

$$T_2 \leq C 2^{(2m+1)j_2} \frac{1}{a_n} \leq C 2^{(2m+1)j_2} \frac{1}{n}. \quad (4.9)$$

Let us now investigate the upper bound for T_1 .

The triangular inequality gives

$$\begin{aligned} \left| \tilde{c}_{j_2,k}^{(m)} - \hat{c}_{j_2,k}^{(m)} \right| &= \left| \frac{(-1)^m}{a_n} \sum_{i=1}^{a_n} Y_i (\phi_{j,k})^{(m)}(X_i) \left(\frac{1}{\hat{g}(X_i)} \mathbf{1}_{\{|\hat{g}(X_i)| \geq c_2/2\}} - \frac{1}{g(X_i)} \right) \right| \\ &\leq \frac{1}{a_n} \sum_{i=1}^{a_n} |Y_i| |(\phi_{j,k})^{(m)}(X_i)| \left| \frac{1}{\hat{g}(X_i)} \mathbf{1}_{\{|\hat{g}(X_i)| \geq c_2/2\}} - \frac{1}{g(X_i)} \right|. \end{aligned}$$

Moreover, we have

$$\begin{aligned} &\frac{1}{\hat{g}(X_i)} \mathbf{1}_{\{|\hat{g}(X_i)| \geq c_2/2\}} - \frac{1}{g(X_i)} \\ &= \frac{1}{g(X_i)} \left(\left(\frac{g(X_i) - \hat{g}(X_i)}{\hat{g}(X_i)} \right) \mathbf{1}_{\{|\hat{g}(X_i)| \geq c_2/2\}} - \mathbf{1}_{\{|\hat{g}(X_i)| < c_2/2\}} \right). \end{aligned}$$

It follows from the triangular inequality, the indicator function, **(K3)**, $\{|\hat{g}(X_i)| < c_2/2\} \subseteq \{|\hat{g}(X_i) - g(X_i)| > c_2/2\}$ and the Markov inequality that

$$\begin{aligned} &\left| \frac{1}{\hat{g}(X_i)} \mathbf{1}_{\{|\hat{g}(X_i)| \geq c_2/2\}} - \frac{1}{g(X_i)} \right| \\ &\leq \frac{1}{g(X_i)} \left(\frac{2}{c_2} |\hat{g}(X_i) - g(X_i)| + \mathbf{1}_{\{|\hat{g}(X_i) - g(X_i)| > c_2/2\}} \right) \\ &\leq \frac{4}{c_2} \frac{|\hat{g}(X_i) - g(X_i)|}{g(X_i)}. \end{aligned}$$

Hence

$$\left| \tilde{c}_{j_2,k}^{(m)} - \hat{c}_{j_2,k}^{(m)} \right| \leq C A_{j_2,k,n},$$

where

$$A_{j,k,n} = \frac{1}{a_n} \sum_{i=1}^{a_n} |Y_i| |(\phi_{j,k})^{(m)}(X_i)| \frac{|\hat{g}(X_i) - g(X_i)|}{g(X_i)}.$$

Let us now consider $U_n = (X_{a_n+1}, \dots, X_n)$. For any random variable D , we have the equality:

$$\mathbb{E}(D^2) = \mathbb{E}(\mathbb{E}(D^2|U_n)) = \mathbb{E}(\mathbb{V}(D|U_n)) + \mathbb{E}((\mathbb{E}(D|U_n))^2),$$

where $\mathbb{E}(D|U_n)$ denotes the expectation of D conditionally to U_n and $\mathbb{V}(D|U_n)$, the variance of D conditionally to U_n . Therefore

$$T_1 \leq C \sum_{k \in \Lambda_{j_2}} \mathbb{E}(A_{j_2,k,n}^2) = C(W_{j_2,n} + Z_{j_2,n}), \quad (4.10)$$

where

$$W_{j_2,n} = \sum_{k \in \Lambda_{j_2}} \mathbb{E}(\mathbb{V}(A_{j_2,k,n}|U_n)), \quad Z_{j_2,n} = \sum_{k \in \Lambda_{j_2}} \mathbb{E}((\mathbb{E}(A_{j_2,k,n}|U_n))^2).$$

Let us now observe that, owing to the independence of $(Y_1, X_1), \dots, (Y_n, X_n)$, the random variables $|Y_1| |(\phi_{j_2, k})^{(m)}(X_1)| |\hat{g}(X_1) - g(X_1)| / g(X_1), \dots,$

$|Y_{a_n}| |(\phi_{j_2, k})^{(m)}(X_{a_n})| |\hat{g}(X_{a_n}) - g(X_{a_n})| / g(X_{a_n})$ conditionally to U_n are independent. This remark combines with the inequalities: $\mathbb{V}(D|U_n) \leq \mathbb{E}(D^2|U_n)$ for any random variable D and $(a+b)^2 \leq 2(a^2 + b^2)$, $(a, b) \in \mathbb{R}^2$, the independence between X_1 and ξ_1 , **(K2)** and **(K3)**, yields

$$\begin{aligned} \mathbb{V}(A_{j_2, k, n} | U_n) &= \frac{1}{a_n} \mathbb{V} \left(|Y_1| |(\phi_{j_2, k})^{(m)}(X_1)| \frac{|\hat{g}(X_1) - g(X_1)|}{g(X_1)} \Big| U_n \right) \\ &\leq \frac{1}{a_n} \mathbb{E} \left(Y_1^2 \left((\phi_{j_2, k})^{(m)}(X_1) \right)^2 \left(\frac{\hat{g}(X_1) - g(X_1)}{g(X_1)} \right)^2 \Big| U_n \right) \\ &\leq \frac{1}{a_n} \frac{2}{c_2} (C_1^2 + \mathbb{E}(\xi_1^2)) \mathbb{E} \left(\left((\phi_{j_2, k})^{(m)}(X_1) \right)^2 \frac{(\hat{g}(X_1) - g(X_1))^2}{g(X_1)} \Big| U_n \right) \\ &= \frac{2}{c_2} (C_1^2 + \mathbb{E}(\xi_1^2)) \frac{1}{a_n} \int_0^1 \left((\phi_{j_2, k})^{(m)}(x) \right)^2 \frac{(\hat{g}(x) - g(x))^2}{g(x)} g(x) dx \\ &\leq C \frac{1}{n} \int_0^1 \left((\phi_{j_2, k})^{(m)}(x) \right)^2 (\hat{g}(x) - g(x))^2 dx. \end{aligned}$$

Thanks to the support compact of $\phi^{(m)}$, we have $\sum_{k \in \Lambda_{j_2}} (\phi^{(m)}(2^{j_2} x - k))^2 \leq C$. Therefore, using $(\phi_{j_2, k})^{(m)}(x) = 2^{(j_2/2)(2m+1)} \phi^{(m)}(2^{j_2} x - k)$,

$$\begin{aligned} W_{j_2, n} &\leq C \frac{1}{n} \mathbb{E} \left(\int_0^1 (\hat{g}(x) - g(x))^2 \sum_{k \in \Lambda_{j_2}} \left((\phi_{j_2, k})^{(m)}(x) \right)^2 dx \right) \\ &\leq C 2^{(2m+1)j_2} \frac{1}{n} \mathbb{E} (\|\hat{g} - g\|_2^2). \end{aligned} \quad (4.11)$$

On the other hand, by the Hölder inequality for conditional expectations, arguments similar to (4.2) and (4.3), we get

$$\begin{aligned} \mathbb{E}(A_{j_2, k, n} | U_n) &= \mathbb{E} \left(|Y_1| |(\phi_{j_2, k})^{(m)}(X_1)| \frac{|\hat{g}(X_1) - g(X_1)|}{g(X_1)} \Big| U_n \right) \\ &\leq \left(\mathbb{E} \left(\frac{Y_1^2}{g(X_1)} \left((\phi_{j_2, k})^{(m)}(X_1) \right)^2 \Big| U_n \right) \right)^{1/2} \left(\mathbb{E} \left(\frac{(\hat{g}(X_1) - g(X_1))^2}{g(X_1)} \Big| U_n \right) \right)^{1/2} \\ &= \left(\mathbb{E} \left(\frac{Y_1^2}{g(X_1)} \left((\phi_{j_2, k})^{(m)}(X_1) \right)^2 \right) \right)^{1/2} \left(\int_0^1 \frac{(\hat{g}(x) - g(x))^2}{g(x)} g(x) dx \right)^{1/2} \\ &\leq C 2^{m j_2} \|\hat{g} - g\|_2. \end{aligned}$$

Hence

$$Z_{j_2, n} \leq C 2^{(2m+1)j_2} \mathbb{E} (\|\hat{g} - g\|_2^2). \quad (4.12)$$

It follows from (4.10), (4.11) and (4.12) that

$$T_1 \leq C2^{(2m+1)j_2} \mathbb{E} (\|\hat{g} - g\|_2^2). \quad (4.13)$$

Putting (4.8), (4.9) and (4.13) together, we get

$$\sum_{k \in \Lambda_{j_2}} \mathbb{E} \left(\left(\tilde{c}_{j_2, k}^{(m)} - c_{j_2, k}^{(m)} \right)^2 \right) \leq C2^{(2m+1)j_2} \max \left(\mathbb{E} (\|\hat{g} - g\|_2^2), \frac{1}{n} \right). \quad (4.14)$$

Combining (4.6), (4.7) and (4.14), we obtain

$$\mathbb{E} \left(\|\hat{f}_3^{(m)} - f^{(m)}\|_2^2 \right) \leq C \left(2^{(2m+1)j_2} \max \left(\mathbb{E} (\|\hat{g} - g\|_2^2), \frac{1}{n} \right) + 2^{-2j_2 s_*} \right). \quad (4.15)$$

A slight adaptation of (Donoho *et al.*, 1996, Proposition 1) gives the following result. Suppose that **(K4)** is satisfied and $g \in B_{p_2, r_2}^{s_2}(M_2)$ with $M_2 > 0$, $p_2 \geq 1$, $r_2 \geq 1$ and $s_2 \in (\max(1/p_2 - 1/2, 0), N)$. Let \hat{g} be defined by (3.9) with j_3 as (3.10). Then there exists a constant $C > 0$ such that

$$\mathbb{E} (\|\hat{g} - g\|_2^2) \leq C(n - a_n)^{-2s_o/(2s_o+1)} \leq Cn^{-2s_o/(2s_o+1)}.$$

Therefore, choosing j_2 as (3.11) and using (4.15), we have

$$\begin{aligned} \mathbb{E} \left(\|\hat{f}_3^{(m)} - f^{(m)}\|_2^2 \right) &\leq C \left(2^{(2m+1)j_2} n^{-2s_o/(2s_o+1)} + 2^{-2j_2 s_*} \right) \\ &\leq Cn^{-4s_* s_o / ((2s_o+1)(2s_*+2m+1))}. \end{aligned}$$

Theorem 3.3 is proved. □

Appendix

Let us now present in details (Chaubey *et al.*, 2013, Theorem 6.1) used in the proof of Theorem 3.2.

We consider a general form of the hard thresholding wavelet estimator denoted by \hat{f}_H for estimating an unknown function $f \in \mathbb{L}^2([0, 1])$ from n independent random variables W_1, \dots, W_n :

$$\hat{f}_H(x) = \sum_{k \in \Lambda_\tau} \hat{\alpha}_{\tau, k} \phi_{\tau, k}(x) + \sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} \hat{\beta}_{j, k} \mathbf{1}_{\{|\hat{\beta}_{j, k}| \geq \kappa \vartheta_j\}} \psi_{j, k}(x), \quad (4.16)$$

where

$$\hat{\alpha}_{j, k} = \frac{1}{v_n} \sum_{i=1}^n q_i(\phi_{j, k}, W_i),$$

$$\hat{\beta}_{j, k} = \frac{1}{v_n} \sum_{i=1}^n q_i(\psi_{j, k}, W_i) \mathbf{1}_{\{|q_i(\psi_{j, k}, W_i)| \leq \varsigma_j\}},$$

$$\varsigma_j = \theta_\psi 2^{\delta j} \frac{v_n}{\sqrt{\mu_n \ln \mu_n}}, \quad \vartheta_j = \theta_\psi 2^{\delta j} \sqrt{\frac{\ln \mu_n}{\mu_n}},$$

$\kappa \geq 2 + 8/3 + 2\sqrt{4 + 16/9}$ and j_1 is the integer satisfying

$$2^{j_1} = \lceil \mu_n^{1/(2\delta+1)} \rceil.$$

Here, we suppose that there exist

- n functions q_1, \dots, q_n with $q_i : \mathbb{L}^2([0, 1]) \times W_i(\Omega) \rightarrow \mathbb{C}$ for any $i \in \{1, \dots, n\}$,
- two sequences of real numbers $(v_n)_{n \in \mathbb{N}}$ and $(\mu_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \rightarrow \infty} v_n = \infty$ and $\lim_{n \rightarrow \infty} \mu_n = \infty$

such that, for $\gamma \in \{\phi, \psi\}$,

(A1) any integer $j \geq \tau$ and any $k \in \Lambda_j$,

$$\mathbb{E} \left(\frac{1}{v_n} \sum_{i=1}^n q_i(\gamma_{j,k}, W_i) \right) = \int_0^1 f(x) \gamma_{j,k}(x) dx.$$

(A2) there exist two constants, $\theta_\gamma > 0$ and $\delta \geq 0$, such that, for any integer $j \geq \tau$ and any $k \in \Lambda_j$,

$$\sum_{i=1}^n \mathbb{E} \left(|q_i(\gamma_{j,k}, W_i)|^2 \right) \leq \theta_\gamma^2 2^{2\delta j} \frac{v_n^2}{\mu_n}.$$

Let \hat{f}_H be (4.16) under (A1) and (A2). Suppose that $f \in B_{p,r}^s(M)$ with $r \geq 1$, $\{p \geq 2 \text{ and } s \in (0, N)\}$ or $\{p \in [1, 2) \text{ and } s \in ((2\delta+1)/p, N)\}$. Then there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\|\hat{f}_H - f\|_2^2 \right) \leq C \left(\frac{\ln \mu_n}{\mu_n} \right)^{2s/(2s+2\delta+1)}.$$

□

References

- Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society Series B*, 6, 97-144.
- Cai, T. (2002). On adaptive wavelet estimation of a derivative and other related linear inverse problems, *J. Statistical Planning and Inference*, 108, 329-349.
- Chaubey, Y.P., Doosti, H., Prakasa Rao, B.L.S. (2006). Wavelet based estimation of the derivatives of a density with associated variables, *International Journal of Pure and Applied Mathematics*, 27, 1, 97-106.
- Chaubey, Y.P., Doosti, H., Prakasa Rao, B.L.S. (2008). Wavelet based estimation of the derivatives of a density for a negatively associated process, *Journal of Statistical Theory and Practice*, 2, 3, 453-463.

- Chaubey, Y.P., Chesneau, C. and Doosti, H. (2013). Adaptive wavelet estimation of a density from mixtures under multiplicative censoring, *preprint*.
- Chesneau, C. (2007). Regression with random design: a minimax study, *Statistics and Probability Letters*, 77, 1, 40-53.
- Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms, *Applied and Computational Harmonic Analysis*, 24, 1, 54-81.
- Daubechies, I. (1992). *Ten lectures on wavelets*, SIAM.
- Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*, 3, 215-228.
- Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 81, 425-455.
- Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, 90, 1200-1224.
- Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia?, *Journal of the Royal Statistical Society, Ser. B*, 57, 301-369.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Annals of Statistics*, 24, 508-539.
- Gasser, T. and Müller, H.G. (1984). Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics*, 11, 3, 171-185.
- Härdle, W. and Gasser, T. (1985). On robust kernel estimation of derivatives of regression functions, *Scandinavian Journal of Statistics*, 12, 3, 233-240.
- Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics New York, 129, Springer Verlag.
- Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the interest rate term structure of corporate debt with a semiparametric penalized spline model, *Journal of the American Statistical Association*, 99, 465, 57-66.
- Mack, Y. P. and Müller, H.G. (1989). Derivative estimation in nonparametric regression with random predictor variable, *Sankhya: The Indian Journal of Statistics*, 51, Series A, 59-72.
- Mallat, S. (2009). *A wavelet tour of signal processing*, Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.
- Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.
- Pensky, M. and Vidakovic, B. (2001). On non-equally spaced wavelet regression, *Annals of the Institute of Statistical Mathematics*, 53, 681-690.
- Petsa, A. and Sapatinas, T. (2011). On the estimation of the function and its derivatives in nonparametric regression: a Bayesian estimation approach, *Sankhya, Series A*, 73, 231-244.
- Prakasa Rao, B.L.S. (1996). Nonparametric estimation of the derivatives of a density by the method of wavelets, *Bull. Inform. Cyb.*, 28, 91-100.

- Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression, *Annals of Statistics*, 22, 1346-1370.
- Stone, C. (1985). Additive regression and other nonparametric models, *Annals of Statistics*, 13, 2689-705.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.
- Wahba, G. and Wang, Y. (1990). When is the optimal regularization parameter insensitive to the choice of loss function? *Communications in Statistics: Theory Methods*, 19, 5, 1685-1700.
- Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, London: Chapman and Hall.
- Zhou, S. and Wolfe, D.A. (2000). On derivative estimation in spline regression, *Statistica Sinica*, 10, 1, 93-108.