On the estimation of density-weighted average derivative by wavelet methods under various dependence structures

Christophe Chesneau, Maher Kachour

On the estimation of density-weighted average derivative by wavelet methods under various dependence structures

Christophe Chesneau[1] and Maher Kachour[2]
[1]*LMNO, Université de Caen Basse-Normandie, 14032, France*
[2]*LMJL, Université de Nantes, 44322, France.*

February 23, 2012

Abstract

The problem of estimating the density-weighted average derivative of a regression function is considered. We present a new consistent estimator based on a plug-in approach and wavelet projections. Its performances are explored under various dependence structures on the observations: the independent case, the ρ-mixing case and the α-mixing case. More precisely, denoting n the number of observations, in the independent case, we prove that it attains $1/n$ under the mean squared error, in the ρ-mixing case, $1/\sqrt{n}$ under the mean absolute error, and in the α-mixing case, and $\sqrt{\ln n/n}$ under the mean absolute error.

Key words and phrases: Nonparametric estimation of density-weighted average derivative, 'Plug-in' approach, Wavelets, Consistency, ρ-mixing, α-mixing.

AMS 2000 Subject Classifications: 62G07, 62G20.

1 Introduction

We consider the following nonparametric regression model.

$$Y_i = f(X_i) + \xi_i, \qquad i \in \{1, \ldots, n\}, \tag{1.1}$$

where the design variables (or input variables) $X_1, \ldots, X_n$ are n identically distributed random variables with common unknown density function g, the noise $\xi_1, \ldots, \xi_n$ are n identically distributed random variables with $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}(\xi_1^4) < \infty$, and f is an unknown regression function. Moreover, it is understood that ξ_i is independent of X_i, for

any $i \in \{1, \ldots, n\}$. In this paper, we are interested in the pointwise estimation of the density-weighted average derivative, which is defined as follows

$$\delta = \mathbb{E}\left(g(X_1)f'(X_1)\right) = \int g^2(x)f'(x)dx. \tag{1.2}$$

It is known that the estimation of δ is of interest in some econometric problems, especially in the context of estimation of coefficients in index models (see e.g. Stoker (1986, 1989), Powell *et al.* (1989) and Härdle and Stoker (1989)).

When $(X_1, Y_1), \ldots, (X_n, Y_n)$ are *i.i.d.*, the most frequently used nonparametric techniques are based on kernel estimators. Three different approaches can be found in Härdle and Stoker (1989), Powell *et al.* (1989) and Stoker (1991). Their consistency are established. Recent theoretical and practical developments related to these estimators can be found in e.g. Härdle *et al.* (1992), Türlach (1994), Powell and Stoker (1996), Banerjee (2007), Schafgans and Zinde-Walsh (2010) and Cattaneo *et al.* (2010, 2011). A new estimator based on orthogonal series methods has been introduced in Prakasa Rao (1995). More precisely, using the same plug-in approach of Powell *et al.* (1989), $\hat{\delta}$ the estimator of the density-weighted average derivative has the following form

$$\hat{\delta} = -\frac{2}{n}\sum_{i=1}^{n} Y_i \hat{g}_i'(X_i), \tag{1.3}$$

where $\hat{g}_i'$ denotes an orthogonal series estimator of g' constructed from $(X_1, Y_1), \ldots, (X_n, Y_n)$. Moreover, the consistency of this estimator is proved.

In this study, we develop a new estimator based on a different plug-in approach to the one in Powell *et al.* (1989) and a particular orthogonal series method: the wavelet series method. The main advantage of this method is its adaptability to the varying degrees of smoothness of the underlying unknown curves. For a complete discussion of wavelets and their applications in statistics, we refer to Antoniadis (1997), Härdle *et al.* (1998) and Vidakovic (1999).

When $(X_1, Y_1), \ldots, (X_n, Y_n)$ are *i.i.d.*, we prove that our estimator attains the parametric rate of convergence $1/n$ under the Mean Square Error (MSE). This rate is a bit better to the one attains by the estimator in Prakasa Rao (1995). Moreover, the flexibility of our approach enables us to consider possible dependent observations, thus opening new perspectives of applications. This is illustrated by the considerations of the ρ-mixing dependence introduced by Kolmogorov and Rozanov (1960) and the α-mixing dependence introduced by Rosenblatt (1956). Adopting the Mean Absolute Error (MAE), we prove that our estimator attains the rate of convergence $1/\sqrt{n}$ in the ρ-mixing case, and $\sqrt{\ln n/n}$ in the α-mixing case. All these results prove the consistency of our estimator and its robustness in term of dependence on the observations. Mention that, to the best of our knowledge, the estimation of δ in such a dependent setting has never been explored earlier.

The remainder of the paper is set out as follows. Next, in Section 2, we discuss the preliminaries of the wavelet orthogonal bases and we recall the definition of some mixing conditions. Our wavelet estimator is described in Section 3. Assumptions on (1.1) are described in Section 4. Section 5 presents our main theoretical results. Finally, the proofs are postponed to Section 6.

2 Preliminaries and Definitions

2.1 Orthonormal bases of compactly supported wavelets

Fix a $b > 0$. Let the following set of functions

$$\mathbb{L}^2([-b,b]) = \left\{ h : [-b,b] \to \mathbb{R};\ ||h||_2^2 = \int_{-b}^{b} (h(x))^2 dx \right\}.$$

For the purposes of this paper, we use the compactly supported wavelet bases on $[-b,b]$ briefly described below.

Let $N \geq 2$ be a fixed integer, and ϕ and ψ be the initial wavelet functions of the Daubechies wavelets $db2N$ (see e.g. Mallat (2009)). These functions have the features to be compactly supported and $\mathcal{C}^{N-1}$.

Set

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

Then there exists an integer τ and a set of consecutive integers Λ_j with a length proportional to 2^j such that the collection

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

is an orthonormal basis of $\mathbb{L}^2([-b,b])$.

Hence, for any $h \in \mathbb{L}^2([-b,b])$ can be expanded on $\mathcal{B}$ as

$$h(x) = \sum_{k\in\Lambda_\tau} \alpha_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty}\sum_{k\in\Lambda_j} \beta_{j,k}\psi_{j,k}(x), \tag{2.1}$$

where

$$\alpha_{\tau,k} = \int_{-b}^{b} h(x)\phi_{\tau,k}(x)dx, \qquad \beta_{j,k} = \int_{-b}^{b} h(x)\psi_{j,k}(x)dx.$$

For more details, we refer to Meyer (1992), Cohen *et al.* (1993) and Mallat (2009).

2.2 Mixing conditions

In this subsection, we will recall the definition of two standard kinds of dependence for random sequences: the ρ-mixing dependence and the α-mixing dependence.
Suppose $Z = (Z_t)_{t\in\mathbb{Z}}$ is a strictly stationary random sequence defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For $j \in \mathbb{Z}$, define the σ-fields

$$\mathcal{F}^Z_{-\infty,j} = \sigma(Z_k, k \leq j), \qquad \mathcal{F}^Z_{j,\infty} = \sigma(Z_k, k \geq j).$$

Definition 2.1 (ρ-mixing dependence) *For any $m \in \mathbb{Z}$, we define the m-th maximal correlation coefficient of $(Z_t)_{t\in\mathbb{Z}}$ by*

$$\rho_m = \sup_{(U,V)\in\mathbb{L}^2(\mathcal{F}^Z_{-\infty,0})\times\mathbb{L}^2(\mathcal{F}^Z_{m,\infty})} \frac{|Cov(U,V)|}{\sqrt{\mathbb{V}(U)\mathbb{V}(V)}},$$

where $Cov(.,.)$ denotes the covariance function and $\mathbb{L}^2(\mathcal{A}) = \left\{U \in \mathcal{A};\ \ \mathbb{E}(U^2) < \infty\right\}$ for any $\mathcal{A} \in \{\mathcal{F}^Z_{-\infty,0}, \mathcal{F}^Z_{m,\infty}\}$.

We say that $(Z_t)_{t\in\mathbb{Z}}$ is ρ-mixing if and only if $\lim_{m\to\infty} \rho_m = 0$.

Full details on ρ-mixing can be found in e.g. Kolmogorov and Rozanov (1960), Doukhan (1994), Shao (1995) and Zhengyan and Lu (1996).

Definition 2.2 (α-mixing dependence) *For any $m \in \mathbb{Z}$, we define the m-th strong mixing coefficient of $(Z_t)_{t\in\mathbb{Z}}$ by*

$$\alpha_m = \sup_{(A,B)\in\mathcal{F}^Z_{-\infty,0}\times\mathcal{F}^Z_{m,\infty}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|.$$

We say that $(Z_t)_{t\in\mathbb{Z}}$ is α-mixing if and only if $\lim_{m\to\infty} \alpha_m = 0$.

Full details on α-mixing can be found in e.g. Rosenblatt (1956), Doukhan (1994), Carrasco and Chen (2002) and Fryzlewicz and Subba Rao (2011).

3 A new wavelet-based estimator for δ

Let τ be the integer mentioned in Section 2. Recall that, for any integer j, Λ_j denotes a set of consecutive integers, with a length proportional to 2^j.
Proposition 3.1 below provides another expression of the density-weighted average derivative (1.2) in terms of wavelet coefficients.

Proposition 3.1 *Consider the regression model with random design* (1.1). *Suppose that* $\mathrm{supp}(X_1) = [-b, b]$ *for a known* $b > 0$, $fg \in \mathbb{L}^2([-b, b])$, $g' \in \mathbb{L}^2([-b, b])$ *and* $g(-b) = g(b) = 0$. *Then the density-weighted average derivative* (1.2) *can be expressed as*

$$\delta = -2 \left(\sum_{k \in \Lambda_\tau} \alpha_{\tau,k} c_{\tau,k} + \sum_{j=\tau}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k} d_{j,k} \right),$$

where

$$\alpha_{\tau,k} = \int_{-b}^{b} f(x) g(x) \phi_{\tau,k}(x) dx, \qquad c_{\tau,k} = \int_{-b}^{b} g'(x) \phi_{\tau,k}(x) dx, \tag{3.1}$$

$$\beta_{j,k} = \int_{-b}^{b} f(x) g(x) \psi_{j,k}(x) dx, \qquad d_{j,k} = \int_{-b}^{b} g'(x) \psi_{j,k}(x) dx. \tag{3.2}$$

We consider the following plug-in estimator for δ:

$$\hat{\delta} = -2 \left(\sum_{k \in \Lambda_\tau} \hat{\alpha}_{\tau,k} \hat{c}_{\tau,k} + \sum_{j=\tau}^{j_0} \sum_{k \in \Lambda_j} \hat{\beta}_{j,k} \hat{d}_{j,k} \right), \tag{3.3}$$

where

$$\hat{\alpha}_{\tau,k} = \frac{1}{n} \sum_{i=1}^{n} Y_i \phi_{\tau,k}(X_i), \qquad \hat{c}_{\tau,k} = -\frac{1}{n} \sum_{i=1}^{n} (\phi_{\tau,k})'(X_i), \tag{3.4}$$

$$\hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} Y_i \psi_{j,k}(X_i), \qquad \hat{d}_{j,k} = -\frac{1}{n} \sum_{i=1}^{n} (\psi_{j,k})'(X_i) \tag{3.5}$$

and j_0 is an integer which will be chosen a posteriori.

Remark 3.1 *The construction of our estimator* (3.3) *uses a plug-in approach derived to Proposition 3.1. Note that it completely differs to the estimator* (1.3) *of Prakasa Rao (1995).*

Remark 3.2 *Mention that* $\hat{c}_{\tau,k}$ (3.4) *and* $\hat{d}_{j,k}$ (3.5) *have been introduced by Prakasa Rao (1996) in the derivative density estimation problem via wavelets. In the context of dependent observations, see Chaubey et al. (2005) and Chaubey et al. (2006).*

Proposition 3.2 *Suppose that* $\mathrm{supp}(X_1) = [-b, b]$. *Then*

- $\hat{\alpha}_{\tau,k}$ (3.4) *and* $\hat{\beta}_{j,k}$ (3.5) *are unbiased estimators for* $\alpha_{\tau,k}$ (3.1) *and* $\beta_{j,k}$ (3.2) *respectively.*
- *under* $g(-b) = g(b) = 0$, $\hat{c}_{\tau,k}$ (3.4) *and* $\hat{d}_{j,k}$ (3.5) *are unbiased estimators for* $c_{\tau,k}$ (3.1) *and* $d_{j,k}$ (3.2) *respectively.*

4 Model assumptions

4.1 Assumptions on f and g

We formulate the following assumptions on f and g:

H1. The support of X_1, denoted by $\text{supp}(X_1)$, is compact. In order to fix the notations, we suppose that $\text{supp}(X_1) = [-b, b]$ for a known $b > 0$.

H2. There exists a known constant $C_1 > 0$ such that

$$\sup_{x\in[-b,b]} f(x) \leq C_1.$$

H3. The function g satisfies $g(-b) = g(b) = 0$ and there exist two known constants $C_2 > 0$ and $C_3 > 0$ such that

$$\sup_{x\in[-b,b]} g(x) \leq C_2, \qquad \sup_{x\in[-b,b]} |g'(x)| \leq C_3.$$

The assumption **H1** is realistic; as noted in Banerjee (2007), in many applications $\text{supp}(X_1)$ is often bounded. The others assumptions are satisfied for a wide variety of functions.

4.2 Assumptions on the wavelet coefficients of fg and g'

Let $s_1 > 0$, $s_2 > 0$ and $\beta_{j,k}$ and $d_{j,k}$ be given by (3.2). We formulate the following assumptions on $\beta_{j,k}$ and $d_{j,k}$:

H4(s_1). There exists a constant $C_4 > 0$ such that

$$|\beta_{j,k}| \leq C_4 2^{-j(s_1+1/2)}.$$

H5(s_2). There exists a constant $C_5 > 0$ such that

$$|d_{j,k}| \leq C_5 2^{-j(s_2+1/2)}.$$

Remark 4.1 *In terms of function sets,* **H4(s_1)** *and* **H5(s_2)** *are equivalent to $fg \in \mathcal{L}_{s_1}(M_1)$ and $g' \in \mathcal{L}_{s_2}(M_2)$ with $M_1 > 0$ and $M_2 > 0$ respectively, where*

$$\mathcal{L}_s(M) = \left\{h; [0,1] \to \mathbb{R};\ |h^{(\lfloor s\rfloor)}(x) - h^{(\lfloor s\rfloor)}(y)| \leq M|x-y|^\alpha,\ s = \lfloor s\rfloor + \alpha,\ \alpha \in (0,1]\right\},$$

$M > 0$, $\lfloor s\rfloor$ is the integer part of s and $h^{(\lfloor s\rfloor)}$ the $\lfloor s\rfloor$-th derivatives of h. We refer to (Härdle et al., 1998, Chapter 8).

5 Main results

5.1 The independent case

In this subsection, we suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent.

Proposition 5.1 *Consider the nonparametric regression model, defined by* (1.1)*. Assume that* **H1**, **H2** *and* **H3** *hold. Let* $\beta_{j,k}$ *and* $d_{j,k}$ *be given by* (3.2)*, and* $\hat{\beta}_{j,k}$ *and* $\hat{d}_{j,k}$ *be given by* (3.5) *with* j *such that* $2^j \leq n$*. Then*

- *there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^4\right) \leq C\frac{1}{n^2}, \tag{5.1}$$

- *there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^4\right) \leq C\frac{2^{4j}}{n^2}. \tag{5.2}$$

These inequalities hold with $(\hat{\alpha}_{\tau,k}, \hat{c}_{\tau,k})$ *in* (3.4) *instead of* $(\hat{\beta}_{j,k}, \hat{d}_{j,k})$*, and* $(\alpha_{\tau,k}, c_{\tau,k})$ *in* (3.1) *instead of* $(\beta_{j,k}, d_{j,k})$ *for* $j = \tau$*.*

Proposition 5.2 *Consider the nonparametric regression model, defined by* (1.1)*.*

- *Suppose that* **H1**, **H2**, **H3**, **H4**(s_1) *and* **H5**(s_2) *hold. Let* $\beta_{j,k}$ *and* $d_{j,k}$ *be given by* (3.2)*, and* $\hat{\beta}_{j,k}$ *and* $\hat{d}_{j,k}$ *be given by* (3.5) *with* j *such that* $2^j \leq n$*. Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k})^2\right) \leq C\left(\frac{2^{-j(2s_1-1)}}{n} + \frac{2^{-j(2s_2+1)}}{n} + \frac{2^{2j}}{n^2}\right).$$

- *Suppose that* **H1**, **H2** *and* **H3** *hold. Let* $\alpha_{\tau,k}$ *and* $c_{\tau,k}$ *be given by* (3.1)*, and* $\hat{\alpha}_{\tau,k}$ *and* $\hat{c}_{\tau,k}$ *be given by* (3.4)*. Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k} - \alpha_{\tau,k}c_{\tau,k})^2\right) \leq C\frac{1}{n}.$$

The following theorem establishes the upper bound of the MSE of our estimator.

Theorem 5.1 *Assume that* **H1**, **H2**, **H3**, **H4**(s_1) *with* $s_1 > 3/2$ *and* **H5**(s_2) *with* $s_2 > 1/2$ *hold. Let* δ *be given by* (1.2) *and* $\hat{\delta}$ *be given by* (3.3) *with* j_0 *such that* $n^{1/4} < 2^{j_0+1} \leq 2n^{1/4}$*. Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{\delta} - \delta)^2\right) \leq C\frac{1}{n}.$$

Remark 5.1 *Theorem 5.1 shows that, under some assumptions, our estimator* (3.3) *has a better MSE than the one in Prakasa Rao (1995), i.e.* $q^2(n)/n$, *where* $q(n)$ *satifies* $\lim_{n\to\infty} q(n) = \infty$ *(*$q(n)$ *is the tuning parameter of the considered projection orthogonal series estimator).*

The rest of the study is devoted to the estimation of δ in the ρ-mixing case and the α-mixing case. For details and applications of dependent nonparametric regression model (1.1), see White and Domowitz (1984) and the references there in. For technical convenience, the performance of (3.3) is explored via the MAE (not the MSE).

5.2 The ρ-mixing case

Now, we assume that $(X_1, Y_1), \ldots, (X_n, Y_n)$ coming from a ρ-mixing strictly stationary process $(X_t, Y_t)_{t\in\mathbb{Z}}$ process (for details see Definition 2.1).

Proposition 5.3 *Consider the nonparametric regression model, defined by* (1.1). *Suppose that* **H1**, **H2**, **H3** *and* (5.5) *hold. Let* $\beta_{j,k}$ *and* $d_{j,k}$ *be given by* (3.2), *and* $\hat{\beta}_{j,k}$ *and* $\hat{d}_{j,k}$ *be given by* (3.5). *Then*

- *there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2\right) \leq C\frac{1}{n}, \tag{5.3}$$

- *there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^2\right) \leq C\frac{2^{2j}}{n}. \tag{5.4}$$

These inequalities hold with $(\hat{\alpha}_{\tau,k}, \hat{c}_{\tau,k})$ *in* (3.4) *instead of* $(\hat{\beta}_{j,k}, \hat{d}_{j,k})$, *and* $(\alpha_{\tau,k}, c_{\tau,k})$ *in* (3.1) *instead of* $(\beta_{j,k}, d_{j,k})$ *for* $j = \tau$.

Proposition 5.4 *Consider the nonparametric regression model, defined by* (1.1).

- *Suppose that* **H1**, **H2**, **H3**, **H4**(s_1), **H5**(s_2) *and* (5.5) *hold, Let* $\beta_{j,k}$ *and* $d_{j,k}$ *be given by* (3.2), *and* $\hat{\beta}_{j,k}$ *and* $\hat{d}_{j,k}$ *be given by* (3.5). *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(|\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k}|\right) \leq C\left(\frac{2^{-j(s_1-1/2)}}{\sqrt{n}} + \frac{2^{-j(s_2+1/2)}}{\sqrt{n}} + \frac{2^j}{n}\right).$$

- *Suppose that* **H1**, **H2**, **H3** *and* (5.5) *hold. Let* $\alpha_{\tau,k}$ *and* $c_{\tau,k}$ *be given by* (3.1), *and* $\hat{\alpha}_{\tau,k}$ *and* $\hat{c}_{\tau,k}$ *be given by* (3.4). *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(|\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k} - \alpha_{\tau,k}c_{\tau,k}|\right) \leq C\frac{1}{\sqrt{n}}.$$

Theorem 5.2 determines the upper bound of the MAE of our estimator in the ρ-mixing case.

Theorem 5.2 (The ρ-mixing case) *Consider the nonparametric regression model, defined by* (1.1). *Suppose that*

- *there exists a constant* $C_* > 0$ *such that*

$$\sum_{m=1}^{\infty} \rho_m \leq C_*, \tag{5.5}$$

- **H1**, **H2**, **H3**, **H4**(s_1) *with* $s_1 > 3/2$ *and* **H5**(s_2) *with* $s_2 > 1/2$ *hold.*

Let δ *be given by* (1.2) *and* $\hat{\delta}$ *be given by* (3.3) *with* j_0 *such that* $n^{1/4} < 2^{j_0+1} \leq 2n^{1/4}$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(|\hat{\delta} - \delta|\right) \leq C\frac{1}{\sqrt{n}}.$$

5.3 The α-mixing case

Here, we assume that $(X_1, Y_1), \ldots, (X_n, Y_n)$ coming from a α-mixing strictly stationary process $(X_t, Y_t)_{t\in\mathbb{Z}}$ process (for details see Definition 2.2).

Proposition 5.5 *Consider the nonparametric regression model, defined by* (1.1). *Suppose that* **H1**, **H2**, **H3**, (5.8) *and* (5.9) *hold. Let* $\beta_{j,k}$ *and* $d_{j,k}$ *be given by* (3.2), *and* $\hat{\beta}_{j,k}$ *and* $\hat{d}_{j,k}$ *be given by* (3.5) *with* j *such that* $2^j \leq n$. *Then*

- *there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2\right) \leq C\frac{\ln n}{n}, \tag{5.6}$$

- *there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^2\right) \leq C\frac{2^{2j}\ln n}{n}. \tag{5.7}$$

These inequalities hold with $(\hat{\alpha}_{\tau,k}, \hat{c}_{\tau,k})$ *in* (3.4) *instead of* $(\hat{\beta}_{j,k}, \hat{d}_{j,k})$, *and* $(\alpha_{\tau,k}, c_{\tau,k})$ *in* (3.1) *instead of* $(\beta_{j,k}, d_{j,k})$ *for* $j = \tau$.

Proposition 5.6 *Consider the nonparametric regression model, defined by* (1.1).

- *Suppose that* **H1**, **H2**, **H3**, **H4**(s_1), **H5**(s_2), (5.8) *and* (5.9) *hold. Let* $\beta_{j,k}$ *and* $d_{j,k}$ *be given by* (3.2), *and* $\hat{\beta}_{j,k}$ *and* $\hat{d}_{j,k}$ *be given by* (3.5) *with* j *satisfying* $2^j \leq n$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(|\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k}|\right) \leq C\left(2^{-j(s_1-1/2)}\sqrt{\frac{\ln n}{n}} + 2^{-j(s_2+1/2)}\sqrt{\frac{\ln n}{n}} + 2^j\frac{\ln n}{n}\right).$$

- *Suppose that* **H1**, **H2**, **H3**, (5.8) *and* (5.9) *hold. Let* $\alpha_{\tau,k}$ *and* $c_{\tau,k}$ *be given by* (3.1), *and* $\hat{\alpha}_{\tau,k}$ *and* $\hat{c}_{\tau,k}$ *be given by* (3.4). *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(|\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k} - \alpha_{\tau,k}c_{\tau,k}|\right) \leq C\sqrt{\frac{\ln n}{n}}.$$

Theorem 5.3 investigates the upper bound of the MAE of our estimator in the α-mixing case.

Theorem 5.3 (The α-mixing case) *Consider the nonparametric regression model, defined by* (1.1). *Suppose that*

- *there exists a constant* C_* *such that*

$$|Y_1| \leq C_*, \tag{5.8}$$

- *there exist two constant* $a > 0$ *and* $b > 0$ *such that the strong mixing coefficient satisfies*

$$\alpha_m \leq ab^{-m}, \tag{5.9}$$

- **H1**, **H2**, **H3**, **H4**(s_1) *with* $s_1 > 3/2$ *and* **H5**(s_2) *with* $s_2 > 1/2$ *hold.*

Let δ *be given by* (1.2) *and* $\hat{\delta}$ *be given by* (3.3) *with* j_0 *such that* $(n/\ln n)^{1/4} < 2^{j_0+1} \leq 2(n/\ln n)^{1/4}$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(|\hat{\delta} - \delta|\right) \leq C\sqrt{\frac{\ln n}{n}}.$$

Conclusion

In this paper we introduce a new density-weighted average derivative estimator using wavelet methods. We evaluate its theoretical performances under various dependence assumptions on the observations. In particular, Theorems 5.1, 5.2 and 5.3 imply the consistency of our estimator (3.3), i.e. $\lim_{n\to\infty}\hat{\delta} \stackrel{p}{=} \delta$, for the considered dependence structures. This illustrates the flexibility of our approach. Our results could be useful to econometricians and statisticians working with density-weighted average derivative estimation, as a simple theory using dependent observations has been absent in this literature until now.

6 Proofs

6.1 On the construction of $\hat{\delta}$

Proof of Proposition 3.1

Using $\text{supp}(X_1) = [-b, b]$, $g(-b) = g(b) = 0$ and an integration by part, we obtain

$$\delta = \left[g^2(x)f(x)\right]_{-b}^{b} - 2\int_{-b}^{b} f(x)g(x)g'(x)dx = -2\int_{-b}^{b} f(x)g(x)g'(x)dx. \tag{6.1}$$

Since $fg \in \mathbb{L}^2([-b, b])$ and $g' \in \mathbb{L}^2([-b, b])$, we can expand these two functions on $\mathcal{B}$ as (2.1):

$$f(x)g(x) = \sum_{k\in\Lambda_\tau} \alpha_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty}\sum_{k\in\Lambda_j} \beta_{j,k}\psi_{j,k}(x),$$

where $\alpha_{\tau,k}$ and $\beta_{j,k}$ are (3.1), and

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

where $c_{\tau,k}$ and $d_{j,k}$ are (3.2). Observing that the integral term in (6.1) is the scalar product of fg and g', the orthonormality of $\mathcal{B}$ on $\mathbb{L}^2([-b, b])$ yields

$$\delta = -2\int_{-b}^{b} f(x)g(x)g'(x)dx = -2\left(\sum_{k\in\Lambda_\tau} \alpha_{\tau,k}c_{\tau,k} + \sum_{j=\tau}^{\infty}\sum_{k\in\Lambda_j} \beta_{j,k}d_{j,k}\right).$$

Proposition 3.1 is proved. □

Proof of Proposition 3.2

- Since $(Y_1, X_1), \ldots, (Y_n, X_n)$ are identically distributed, ξ_i and X_i are independent for any $i \in \{1, \ldots, n\}$, and $\mathbb{E}(\xi_1) = 0$, we have

  $$\mathbb{E}(\hat{\beta}_{j,k}) = \mathbb{E}(Y_1\psi_{j,k}(X_1)) = \mathbb{E}(f(X_1)\psi_{j,k}(X_1)) = \int_{-b}^{b} f(x)g(x)\psi_{j,k}(x)dx = \beta_{j,k}.$$

  Similarly, we prove that $\mathbb{E}(\hat{\alpha}_{\tau,k}) = \alpha_{\tau,k}$.

- Using the identical distribution of $X_1, \ldots, X_n$, $\mathbb{E}(\xi_1) = 0$, an integration by parts and $g(b) = g(-b) = 0$, we obtain

$$\begin{aligned}\mathbb{E}(\hat{d}_{j,k}) &= -\mathbb{E}((\psi_{j,k})'(X_1)) = -\int_{-b}^{b} g(x)(\psi_{j,k})'(x)dx \\ &= -\left([g(x)\psi_{j,k}(x)]_{-b}^{b} - \int_{-b}^{b} g'(x)\psi_{j,k}(x)dx\right) = \int_{-b}^{b} g'(x)\psi_{j,k}(x)dx = d_{j,k}.\end{aligned}$$

  Similarly, we prove that $\mathbb{E}(\hat{c}_{\tau,k}) = c_{\tau,k}$.

This ends the proof of Proposition 3.2. □

6.2 Proof of the main results

6.2.1 The independent case

In the sequel, we assume that $(X_1, Y_1), \cdots, (X_n, Y_n)$ are independent. To bound the fouth cental moment of the estimators, defined by (3.4) and (3.5), we use the following version of the Rosenthal inequality. The proof can be found in Rosenthal (1970).

Lemma 6.1 *Let n be a positive integer, $p \geq 2$ and $U_1, \ldots, U_n$ be n zero mean independent random variables such that $\sup_{i\in\{1,\ldots,n\}} \mathbb{E}(|U_i|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} U_i\right|^p\right) \leq C\left(\sum_{i=1}^{n} \mathbb{E}\left(|U_i|^p\right) + \left(\sum_{i=1}^{n} \mathbb{E}\left(U_i^2\right)\right)^{p/2}\right).$$

Proof of Proposition 5.1

- Observe that

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^4\right) = \frac{1}{n^4}\mathbb{E}\left(\left(\sum_{i=1}^{n}(Y_i\psi_{j,k}(X_i) - \beta_{j,k})\right)^4\right).$$

  Set

$$U_i = Y_i\psi_{j,k}(X_i) - \beta_{j,k}, \qquad i \in \{1, \ldots, n\}.$$

  Since $(X_1, Y_1), \ldots, (X_n, Y_n)$ are *i.i.d.*, we get that $U_1, \ldots, U_n$ are also *i.i.d.*. Moreover, from Proposition 3.2, we have $\mathbb{E}(U_1) = 0$. Thus, Lemma (with $p = 4$) 6.1 yields

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^4\right) \leq C\frac{1}{n^4}\left(n\mathbb{E}(U_1^4) + n^2(\mathbb{E}(U_1^2))^2\right).$$

Using **H1**, **H2**, the Hölder inequality, **H3**, the independence between ξ_1 and X_1, $\mathbb{E}(\xi_1^4) < \infty$, applying the change of variables $y = 2^j x - k$, and using the fact that ψ is compactly supported, we have for any $u \in \{2, 4\}$,

$$\begin{aligned}\mathbb{E}(U_1^u) &\leq C\mathbb{E}((Y_1\psi_{j,k}(X_1))^u) \leq C(C_1^u + \mathbb{E}(\xi_1^u))\mathbb{E}((\psi_{j,k}(X_1))^u) \\ &= C\int_{-b}^{b}(\psi_{j,k}(x))^u g(x)dx \leq C\int_{-b}^{b}(\psi_{j,k}(x))^u dx \\ &= C2^{j(u-2)/2}\int_{-b}^{b}(\psi(x))^u dx \leq C2^{j(u-2)/2}. \end{aligned} \tag{6.2}$$

Therefore, since $2^j \leq n$, we obtain

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^4\right) \leq C\left(\frac{1}{n^3}2^j + \frac{1}{n^2}\right) \leq C\frac{1}{n^2}.$$

- We have that

$$\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^4\right) = \frac{1}{n^4}\mathbb{E}\left(\left(\sum_{i=1}^{n}((\psi_{j,k})'(X_i) - d_{j,k})\right)^4\right).$$

Now, set

$$U_i = (\psi_{j,k})'(X_i) - d_{j,k}, \qquad i \in \{1, \ldots, n\}.$$

Since $X_1, \ldots, X_n$ are *i.i.d.*, it is clear that $U_1, \ldots, U_n$ are also *i.i.d.*. Moreover, by Propostion 3.2, we have $\mathbb{E}(U_1) = 0$. Hence, Lemma 6.1 (with $p = 4$) yields

$$\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^4\right) \leq C\frac{1}{n^4}\left(n\mathbb{E}(U_1^4) + n^2(\mathbb{E}(U_1^2))^2\right).$$

Using **H2**, the Hölder inequality, **H3**, $(\psi_{j,k})'(x) = 2^{3j/2}\psi'(2^j x - k)$, applying the change of variables $y = 2^j x - k$, and using the fact that ψ is compactly supported and $\mathcal{C}^{N-1}$, where $N \geq 2$, we have for any $u \in \{2, 4\}$,

$$\begin{aligned}\mathbb{E}(U_1^u) &\leq C\mathbb{E}(((\psi_{j,k})'(X_1))^u) = C\int_{-b}^{b}((\psi_{j,k})'(x))^u g(x)dx \leq C\int_{-b}^{b}((\psi_{j,k})'(x))^u dx \\ &= C2^{j(3u-2)/2}\int_{-b}^{b}(\psi'(x))^u dx \leq C2^{j(3u-2)/2}. \end{aligned} \tag{6.3}$$

Putting these inequalities together and using $2^j \leq n$, we obtain

$$\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^4\right) \leq C\left(\frac{2^{5j}}{n^3} + \frac{2^{4j}}{n^2}\right) \leq C\frac{2^{4j}}{n^2}.$$

Proposition 5.1 is proved. □

Proof of Proposition 5.2

- We have the following decomposition

$$\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k} = \beta_{j,k}(\hat{d}_{j,k} - d_{j,k}) + d_{j,k}(\hat{\beta}_{j,k} - \beta_{j,k}) + (\hat{\beta}_{j,k} - \beta_{j,k})(\hat{d}_{j,k} - d_{j,k}).$$

Therefore

$$\mathbb{E}\left((\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k})^2\right) \leq 3(T_1 + T_2 + T_3)$$

where

$$T_1 = \beta_{j,k}^2\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^2\right), \qquad T_2 = d_{j,k}^2\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2\right)$$

and

$$T_3 = \mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2(\hat{d}_{j,k} - d_{j,k})^2\right).$$

Upper bound for T_1. It follows from the Cauchy-Schwarz inequality, the second point in Proposition 5.1 and $\mathbf{H4}(s_1)$ that

$$T_1 \leq C_4^2 2^{-2j(s_1+1/2)}\sqrt{\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^4\right)} \leq C2^{-2j(s_1+1/2)}\frac{2^{2j}}{n} = C\frac{2^{-j(2s_1-1)}}{n}.$$

Upper bound for T_2. By the Cauchy-Schwarz inequality, the first point in Proposition 5.1 and $\mathbf{H5}(s_2)$, we obtain

$$T_2 \leq C_5^2 2^{-2j(s_2+1/2)}\sqrt{\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^4\right)} \leq C\frac{2^{-j(2s_2+1)}}{n}.$$

Upper bound for T_3. The Cauchy-Schwarz inequality and Proposition 5.1 yield

$$T_3 \leq \sqrt{\left(\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^4\right)\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^4\right)\right)} \leq C\sqrt{\frac{1}{n^2}\frac{2^{4j}}{n^2}} = C\frac{2^{2j}}{n^2}.$$

Combining the inequalities above, we obtain

$$\mathbb{E}\left((\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k})^2\right) \leq C\left(\frac{2^{-j(2s_1-1)}}{n} + \frac{2^{-j(2s_2+1)}}{n} + \frac{2^{2j}}{n^2}\right).$$

- The proof of the second point is identical to the first one but with the bounds $|\alpha_{\tau,k}| \leq C$ and $|c_{\tau,k}| \leq C$ thanks to **H2** and **H3**.

This ends the proof of Proposition 5.2. □

The following Lemma will be very usefull for the proof of Theorem 5.1.

Lemma 6.2 *Let n be a positive integer and $U_1, \ldots, U_n$ be n zero mean independent random variables such that $\sup_{i\in\{1,\ldots,n\}} \mathbb{E}(U_i^2) < \infty$. Then*

$$\mathbb{E}\left(\left(\sum_{i=1}^{n} U_i\right)^2\right) \leq \left(\sum_{i=1}^{n} \sqrt{\mathbb{E}(U_i^2)}\right)^2.$$

Proof of Theorem 5.1

It follows from Proposition 3.1 that

$$\begin{aligned}\hat{\delta} - \delta &= -2\sum_{k\in\Lambda_\tau} (\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k} - \alpha_{\tau,k}c_{\tau,k}) - 2\sum_{j=\tau}^{j_0}\sum_{k\in\Lambda_j} (\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k}) \\ &+ 2\sum_{j=j_0+1}^{\infty}\sum_{k\in\Lambda_j} \beta_{j,k}d_{j,k}.\end{aligned}$$

Therefore

$$\mathbb{E}\left((\hat{\delta} - \delta)^2\right) \leq 12(W_1 + W_2 + W_3), \tag{6.4}$$

where

$$W_1 = \mathbb{E}\left(\left(\sum_{k\in\Lambda_\tau} (\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k} - \alpha_{\tau,k}c_{\tau,k})\right)^2\right), \qquad W_2 = \mathbb{E}\left(\left(\sum_{j=\tau}^{j_0}\sum_{k\in\Lambda_j} (\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k})\right)^2\right),$$

and

$$W_3 = \left(\sum_{j=j_0+1}^{\infty}\sum_{k\in\Lambda_j} \beta_{j,k}d_{j,k}\right)^2.$$

Now, to bound W_1 and W_2, we use Lemma 6.2, see above.
Upper bound for W_1. Using Lemma 6.2, the second point of Proposition 5.2 and $\mathrm{Card}(\Lambda_\tau) \leq C$, we obtain

$$W_1 \leq \left(\sum_{k\in\Lambda_\tau} \sqrt{\mathbb{E}\left((\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k} - \alpha_{\tau,k}c_{\tau,k})^2\right)}\right)^2 \leq C\frac{1}{n}. \tag{6.5}$$

Upper bound for W_2. It follows from Lemma 6.2, the first point of Proposition 5.2, $\mathrm{Card}(\Lambda_j) \leq C2^j$, the elementary inequality: $\sqrt{a+b+c} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$, $s_1 > 3/2$,

$s_2 > 1/2$ and $2^{j_0} \leq n^{1/4}$ that

$$\begin{aligned}
W_2 &\leq \left(\sum_{j=\tau}^{j_0} \sum_{k \in \Lambda_j} \sqrt{\mathbb{E}\left((\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k} d_{j,k})^2 \right)} \right)^2 \\
&\leq C \left(\sum_{j=\tau}^{j_0} 2^j \sqrt{\frac{2^{-j(2s_1-1)}}{n} + \frac{2^{-j(2s_2+1)}}{n} + \frac{2^{2j}}{n^2}} \right)^2 \\
&\leq C \left(\sum_{j=\tau}^{j_0} \left(\frac{2^{-j(s_1-3/2)}}{\sqrt{n}} + \frac{2^{-j(s_2-1/2)}}{\sqrt{n}} + \frac{2^{2j}}{n} \right) \right)^2 \\
&\leq C \left(\frac{1}{\sqrt{n}} \sum_{j=\tau}^{j_0} 2^{-j(s_1-3/2)} + \frac{1}{\sqrt{n}} \sum_{j=\tau}^{j_0} 2^{-j(s_2-1/2)} + \frac{1}{n} \sum_{j=\tau}^{j_0} 2^{2j} \right)^2 \\
&\leq C \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} + \frac{2^{2j_0}}{n} \right)^2 \leq C \frac{1}{n}. \qquad (6.6)
\end{aligned}$$

Upper bound for W_3. By $\mathbf{H4}(s_1)$ with $s_1 > 3/2$, $\mathbf{H5}(s_2)$ with $s_2 > 1/2$ and $2^{j_0+1} > n^{1/4}$, we have

$$\begin{aligned}
W_3 &\leq \left(\sum_{j=j_0+1}^{\infty} \sum_{k \in \Lambda_j} |\beta_{j,k}||d_{j,k}| \right)^2 \leq C \left(\sum_{j=j_0+1}^{\infty} 2^j 2^{-j(s_1+1/2)} 2^{-j(s_2+1/2)} \right)^2 \leq C 2^{-2j_0(s_1+s_2)} \\
&\leq C 2^{-4j_0} \leq C \frac{1}{n}. \qquad (6.7)
\end{aligned}$$

Putting (6.4), (6.5), (6.6) and (6.7) together, we obtain

$$\mathbb{E}\left((\hat{\delta} - \delta)^2 \right) \leq C \frac{1}{n}.$$

This ends the proof of Theorem 5.1. □

6.2.2 The ρ-mixing case

In the sequel, we assume that $(X_1, Y_1), \ldots, (X_n, Y_n)$ coming from a ρ-mixing strictly stationary process $(X_t, Y_t)_{t \in \mathbb{Z}}$ process (see Definition 2.1).

Proof of Proposition 5.3

- From Proposition 3.2, we have $\mathbb{E}(\hat{\beta}_{j,k}) = \beta_{j,k}$. It follows that

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2 \right) = \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^{n} Y_i \psi_{j,k}(X_i) \right) = S_1 + S_2,$$

where

$$S_1 = \frac{1}{n}\mathbb{V}(Y_1\psi_{j,k}(X_1)), \qquad S_2 = \frac{2}{n^2}\sum_{v=2}^{n}\sum_{\ell=1}^{v-1}\mathrm{Cov}\left(Y_v\psi_{j,k}(X_v), Y_\ell\psi_{j,k}(X_\ell)\right).$$

Upper bound for S_1. It follows from (6.2) with $u = 2$ that

$$S_1 \leq \frac{1}{n}\mathbb{E}\left((Y_1\psi_{j,k}(X_1))^2\right) \leq C\frac{1}{n}.$$

Upper bound for S_2. The stationarity of $(X_t, Y_t)_{t\in\mathbb{Z}}$ implies that

$$\begin{aligned} S_2 &= \frac{1}{n^2}\sum_{m=1}^{n-1}(n-m)\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1)\right) \\ &\leq \frac{1}{n}\sum_{m=1}^{n-1}\left|\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1)\right)\right|. \end{aligned}$$

A standard covariance inequality for ρ-mixing gives

$$\left|\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1)\right)\right| \leq \mathbb{E}((Y_1\psi_{j,k}(X_1))^2)\rho_m$$

(see, for instance, (Zhengyan and Lu, 1996, Lemma 1.2.7.)).

Equation (6.2) with $u = 2$ yields

$$\mathbb{E}((Y_1\psi_{j,k}(X_1))^2) \leq C.$$

Therefore, using (5.5),

$$S_2 \leq C\frac{1}{n}\sum_{m=1}^{n-1}\rho_m \leq C\frac{1}{n}\sum_{m=1}^{\infty}\rho_m \leq C\frac{1}{n}.$$

Combining the inequalities above, we obtain

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2\right) \leq C\frac{1}{n}.$$

- Proceeding as for the first point but with $(\psi_{j,k})'(X_i)$ instead of $Y_i\psi_{j,k}(X_i)$ and (6.3) instead of (6.2).

Proposition 5.3 is proved. □

Proof of Proposition 5.4

- We have the following decomposition

  $$\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k} = \beta_{j,k}(\hat{d}_{j,k} - d_{j,k}) + d_{j,k}(\hat{\beta}_{j,k} - \beta_{j,k}) + (\hat{\beta}_{j,k} - \beta_{j,k})(\hat{d}_{j,k} - d_{j,k}).$$

  Therefore

  $$\mathbb{E}\left(|\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k}|\right) \leq T_1 + T_2 + T_3$$

  where

  $$T_1 = |\beta_{j,k}|\mathbb{E}\left(|\hat{d}_{j,k} - d_{j,k}|\right), \qquad T_2 = |d_{j,k}|\mathbb{E}\left(|\hat{\beta}_{j,k} - \beta_{j,k}|\right)$$

  and

  $$T_3 = \mathbb{E}\left(|(\hat{\beta}_{j,k} - \beta_{j,k})(\hat{d}_{j,k} - d_{j,k})|\right).$$

  Upper bound for T_1. Using the Cauchy-Schwarz inequality, the second point in Proposition 5.3 and **H4**(s_1), we obtain

  $$T_1 \leq C_4 2^{-j(s_1+1/2)}\sqrt{\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^2\right)} \leq C2^{-j(s_1+1/2)}\frac{2^j}{\sqrt{n}} = C\frac{2^{-j(s_1-1/2)}}{\sqrt{n}}.$$

  Upper bound for T_2. By the Cauchy-Schwarz inequality, the first point in Proposition 5.3 and **H5**(s_2), we obtain

  $$T_2 \leq C_5 2^{-j(s_2+1/2)}\sqrt{\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2\right)} \leq C\frac{2^{-j(s_2+1/2)}}{\sqrt{n}}.$$

  Upper bound for T_3. The Cauchy-Schwarz inequality and Proposition 5.3 yield

  $$T_3 \leq \sqrt{\left(\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2\right)\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^2\right)} \leq C\sqrt{\frac{1}{n}\frac{2^{2j}}{n}} = C\frac{2^j}{n}.$$

  The above inequalities imply that

  $$\mathbb{E}\left(|\hat{\beta}_{j,k}\hat{d}_{j,k} - \beta_{j,k}d_{j,k}|\right) \leq C\left(\frac{2^{-j(s_1-1/2)}}{\sqrt{n}} + \frac{2^{-j(s_2+1/2)}}{\sqrt{n}} + \frac{2^j}{n}\right).$$

- The proof of the second point is identical to the first one but with the bounds $|\alpha_{\tau,k}| \leq C$ and $|c_{\tau,k}| \leq C$ thanks to **H2** and **H3**.

This ends the proof of Proposition 5.4. □

Proof of Theorem 5.2

Using Proposition 3.1, we have

$$\begin{aligned}\hat{\delta}-\delta &= -2\sum_{k\in\Lambda_\tau}(\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k}-\alpha_{\tau,k}c_{\tau,k})-2\sum_{j=\tau}^{j_0}\sum_{k\in\Lambda_j}(\hat{\beta}_{j,k}\hat{d}_{j,k}-\beta_{j,k}d_{j,k})\\ &+ 2\sum_{j=j_0+1}^{\infty}\sum_{k\in\Lambda_j}\beta_{j,k}d_{j,k}.\end{aligned}$$

Therefore

$$\mathbb{E}\left(|\hat{\delta}-\delta|\right)\leq W_1+W_2+W_3, \tag{6.8}$$

where

$$W_1=\sum_{k\in\Lambda_\tau}\mathbb{E}\left(|\hat{\alpha}_{\tau,k}\hat{c}_{\tau,k}-\alpha_{\tau,k}c_{\tau,k}|\right),\qquad W_2=\sum_{j=\tau}^{j_0}\sum_{k\in\Lambda_j}\mathbb{E}\left(|\hat{\beta}_{j,k}\hat{d}_{j,k}-\beta_{j,k}d_{j,k}|\right),$$

and

$$W_3=\sum_{j=j_0+1}^{\infty}\sum_{k\in\Lambda_j}|\beta_{j,k}||d_{j,k}|.$$

Upper bound for W_1. The second point of Proposition 5.4 and $\mathrm{Card}(\Lambda_\tau)\leq C$ give

$$W_1\leq C\frac{1}{\sqrt{n}}. \tag{6.9}$$

Upper bound for W_2. It follows from the first point of Proposition 5.4, $\mathrm{Card}(\Lambda_j)\leq C2^j$, $s_1>3/2$, $s_2>1/2$ and $2^{j_0}\leq n^{1/4}$ that

$$\begin{aligned}W_2 &\leq C\sum_{j=\tau}^{j_0}2^j\left(\frac{2^{-j(s_1-1/2)}}{\sqrt{n}}+\frac{2^{-j(s_2+1/2)}}{\sqrt{n}}+\frac{2^j}{n}\right)\\ &\leq C\left(\frac{1}{\sqrt{n}}\sum_{j=\tau}^{j_0}2^{-j(s_1-3/2)}+\frac{1}{\sqrt{n}}\sum_{j=\tau}^{j_0}2^{-j(s_2-1/2)}+\frac{1}{n}\sum_{j=\tau}^{j_0}2^{2j}\right)\\ &\leq C\left(\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{n}}+\frac{2^{2j_0}}{n}\right)\leq C\frac{1}{\sqrt{n}}.\end{aligned} \tag{6.10}$$

Upper bound for W_3. By $\mathbf{H4}(s_1)$ with $s_1>3/2$, $\mathbf{H5}(s_2)$ with $s_2>1/2$ and $2^{j_0+1}>n^{1/4}$, we have

$$W_3\leq C\sum_{j=j_0+1}^{\infty}2^j2^{-j(s_1+1/2)}2^{-j(s_2+1/2)}\leq C2^{-j_0(s_1+s_2)}\leq C2^{-2j_0}\leq C\frac{1}{\sqrt{n}}. \tag{6.11}$$

Putting (6.8), (6.9), (6.10) and (6.11) together, we obtain

$$\mathbb{E}\left(|\hat{\delta}-\delta|\right) \leq C\frac{1}{\sqrt{n}}.$$

This ends the proof of Theorem 5.2. □

6.2.3 The α-mixing case

Recall that, here, we assume that $(X_1,Y_1),\ldots,(X_n,Y_n)$ coming from a α-mixing strictly stationary process $(X_t,Y_t)_{t\in\mathbb{Z}}$ process (see Definition 2.2).

Proof of Proposition 5.5

- Proposition 3.2 yields $\mathbb{E}(\hat{\beta}_{j,k})=\beta_{j,k}$. Therefore,

$$\mathbb{E}\left((\hat{\beta}_{j,k}-\beta_{j,k})^2\right)=\frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n}Y_i\psi_{j,k}(X_i)\right)=S_1+S_2,$$

where

$$S_1=\frac{1}{n}\mathbb{V}(Y_1\psi_{j,k}(X_1)),\qquad S_2=\frac{2}{n^2}\sum_{v=2}^{n}\sum_{\ell=1}^{v-1}\mathrm{Cov}\left(Y_v\psi_{j,k}(X_v),Y_\ell\psi_{j,k}(X_\ell)\right).$$

Upper bound for S_1. It follows from (6.2) with $u=2$ that

$$S_1\leq\frac{1}{n}\mathbb{E}\left((Y_1\psi_{j,k}(X_1))^2\right)\leq C\frac{1}{n}.$$

Upper bound for S_2. The stationarity of $(X_t,Y_t)_{t\in\mathbb{Z}}$ implies that

$$\begin{aligned}S_2 &= \frac{1}{n^2}\sum_{m=1}^{n-1}(n-m)\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}),Y_1\psi_{j,k}(X_1)\right)\\ &\leq \frac{1}{n}\sum_{m=1}^{n-1}|\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}),Y_1\psi_{j,k}(X_1)\right)|.\end{aligned}$$

Let $[c\ln n]$ be the integer part of $c\ln n$ where $c=1/\ln b$. We have

$$\sum_{m=1}^{n-1}|\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}),Y_1\psi_{j,k}(X_1)\right)|=$$
$$\sum_{m=1}^{[c\ln n]}|\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}),Y_1\psi_{j,k}(X_1)\right)|+\sum_{m=[c\ln n]+1}^{n-1}|\mathrm{Cov}\left(Y_{m+1}\psi_{j,k}(X_{m+1}),Y_1\psi_{j,k}(X_1)\right)|.$$

On the one hand, the Cauchy-Schwarz inequality and (6.2) with $u = 2$ yield

$$|\mathrm{Cov}\,(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1))\,| \le \mathbb{E}((Y_1\psi_{j,k}(X_1))^2) \le C.$$

Hence

$$\sum_{m=1}^{[c\ln n]} |\mathrm{Cov}\,(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1)) \le C\ln n.$$

On the other hand, a standard covariance inequality for α-mixing gives, for any $\gamma \in (0,1)$,

$$|\mathrm{Cov}\,(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1))\,| \le 10\alpha_m^{\gamma}\left(\mathbb{E}\left(|Y_1\psi_{j,k}(X_1)|^{2/(1-\gamma)}\right)\right)^{1-\gamma}.$$

(See, for instance, Davydov (1970)).

Taking $\gamma = 1/2$ and using (5.8), (5.9), again (6.2) with $u = 2$ and $2^j \le n$, we obtain

$$\begin{aligned}
&\sum_{m=[c\ln n]+1}^{n-1} |\mathrm{Cov}\,(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1))\,| \\
&\quad\le C(\sup_{x\in[-b,b]}|\psi_{j,k}(x)|)\sqrt{\mathbb{E}\left((Y_1\psi_{j,k}(X_1))^2\right)}\sum_{m=[c\ln n]+1}^{n-1}\sqrt{\alpha_m} \\
&\quad\le C2^{j/2}\sum_{m=[c\ln n]+1}^{\infty} b^{-m/2} \le C\sqrt{n}b^{-c\ln n/2} \le C.
\end{aligned}$$

Hence

$$\sum_{m=[c\ln n]+1}^{n-1} |\mathrm{Cov}\,(Y_{m+1}\psi_{j,k}(X_{m+1}), Y_1\psi_{j,k}(X_1))\,| \le C.$$

Then

$$S_2 \le C\frac{\ln n}{n}.$$

Combining the inequalities above, we obtain

$$\mathbb{E}\left((\hat{\beta}_{j,k} - \beta_{j,k})^2\right) \le C\frac{\ln n}{n}.$$

- The proof is similar to the first point. It is enough to replace $Y_i\psi_{j,k}(X_i)$ by $(\psi_{j,k})'(X_i)$,

apply (6.3) instead of (6.2) and observe that

$$\begin{aligned}
&\sum_{m=[c\ln n]+1}^{n-1} \left|\mathrm{Cov}\left((\psi_{j,k})'(X_{m+1}),(\psi_{j,k})'(X_1)\right)\right| \\
&\leq C(\sup_{x\in\mathbb{R}}\left|(\psi_{j,k})'(x)\right|)\sqrt{\mathbb{E}\left(((\psi_{j,k})'(X_1))^2\right)}\sum_{m=[c\ln n]+1}^{n-1}\sqrt{\alpha_m} \\
&\leq C2^{3j/2}2^j\sum_{m=[c\ln n]+1}^{\infty} b^{-m/2} \leq C2^{2j}\sqrt{n}b^{-c\ln n/2} \leq C2^{2j}.
\end{aligned}$$

Proposition 5.5 is proved. □

Proof of Proposition 5.6

The proof of Proposition 5.6 is identical to the one of Proposition 5.4. It is enough to use Proposition 5.5 instead of Proposition 5.3 and to replace $1/n$ by $\ln n/n$. □

Proof of Theorem 5.3

The proof of Theorem 5.3 is identical to the one of Theorem 5.2. It suffices to use Proposition 5.6 instead of Proposition 5.4 and to replace $1/n$ by $\ln n/n$. □

References

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *Journal of the Italian Statistical Society*, Series B, 6, 97-144.

Banerjee, A.N. (2007). A method of estimating the Average Derivative. *Journal of Econometrics*, 136,(1), 65-88.

Cattaneo, M.D., Crump, R.K. and Jansson, M. (2010). Robust Data-Driven Inference for Density-Weighted Average Derivatives. *Journal of the American Statistical Association*, 105, (491), 1070-1083.

Cattaneo, M.D., Crump, R.K. and Jansson, M. (2011). Small Bandwidth Asymptotics for Density-Weighted Average Derivatives. *Econometric Theory*, (to appear).

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.

Chaubey, Y.P. and Doosti, H. (2005). Wavelet based estimation of the derivatives of a density for m-dependent random variables. *Journal of the Iranian Statistical Society*, 4, (2), 97-105.

Chaubey, Y.P., Doosti, H. and 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.

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.

Davydov, Y. (1970). The invariance principle for stationary processes. *Theor. Probab. Appl.*, 15, 3, 498-509.

Doukhan, P. (1994). Mixing. Properties and Examples. Lecture Notes in Statistics 85. Springer Verlag, New York.

Fryzlewicz, P. and Subba Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes, *Bernoulli*, Volume 17, Number 1 (2011), 320-346.

Hardle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.*, 84, 986-995.

Härdle, W., Hart, J., Marron, J.S. and Tsybakov, A.B. (1992), Bandwidth Choice for Average Derivative Estimation. *Journal of the American Statistical Association*, 87, 417. 218-226.

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

Kolmogorov, A.N. and Rozanov, Yu.A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theor. Probab. Appl.*, 5, 204-208.

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.

Powell, U.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57, 1403-1430.

Powell, U.L. and Stoker, T.M. (1996). Optimal Bandwidth Choice For Density-Weighted Averages. *Journal of Econometrics*, 75, 291-316.

Prakasa Rao, B.L.S. (1995). Consistent estimation of density-weighted average derivative by orthogonal series method. *Statistics and Probability Letters*, 22, 205-212.

Prakasa Rao, B.L.S. (1996). Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bull. Inform. Cyb*, 28, 91-10.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Ac. Sc. USA*, 42, 43-47.

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.

Schafgans, M. and Zinde-Walsh, V. (2010). Smoothness adaptive average derivative estimation. *Econometrics journal*, 13, (1), 40-62.

Shao, Q.-M. (1995). Maximal inequality for partial sums of ρ-mixing sequences. *Ann. Probab.*, 23, 948-965.

Stoker, T.M. (1986). Consistent estimation of scaled coefficients. *Econometrica*, 54, 1461-1481.

Stoker, T.M. (1989). Tests of additive derivative constraints. *Rev. Econom. Stud.*, 56, 535-552.

Stoker, T.M. (1991), *Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives, Nonparametric and Semiparametric Methods in Econometrics and Statistics*. W.A. Barnett, J. Powell and G. Tauchen, eds., Cambridge University Press

Türlach, B.A. (1994). Fast implementation of density-weighted average derivative estimation. *Computationally Intensive Statistical Methods*, 26, 28-33.

White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica*, 52, 143-162.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons, Inc., New York, 384 pp.

Zhengyan, L. and Lu, C. (1996). Limit Theory for Mixing Dependent Random Variables, Kluwer, Dordrecht.