



HAL
open science

Adaptive estimation of an additive regression function from weakly dependent data

Christophe Chesneau, Jalal M. Fadili, Bertrand Maillot

► **To cite this version:**

Christophe Chesneau, Jalal M. Fadili, Bertrand Maillot. Adaptive estimation of an additive regression function from weakly dependent data. 2011. hal-00641912v1

HAL Id: hal-00641912

<https://hal.science/hal-00641912v1>

Preprint submitted on 17 Nov 2011 (v1), last revised 6 Aug 2012 (v2)

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

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

Adaptive estimation of an additive regression function from weakly dependent data

Christophe Chesneau*, Jalal Fadili†, Bertrand Maillot*

Abstract

A d -dimensional nonparametric additive regression model with dependent observations is considered. Using the marginal integration and the methods of wavelets, we develop a new adaptive estimator for a component of the additive regression function. Its asymptotic properties are investigated via the minimax approach under the \mathbb{L}_2 risk over Besov balls. We prove that it attains a sharp rate of convergence, close to the one obtained in the one-dimensional case. In particular, it is both independent of d and slightly deteriorated by the dependence of the observations.

Key words and phrases: Additive regression, Adaptivity, Wavelets, Hard thresholding.

AMS 2000 Subject Classifications: 62G07, 62G20.

*LMNO, CNRS-Université de Caen, Campus II, Science 3, 14032, Caen, France

†GREYC, CNRS-ENSICAEN-Université de Caen, 6, Bd du Maréchal Juin, 14050 Caen Cedex, France

1 Introduction

Let d be a positive integer, $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$ be a $\mathbb{R} \times [0, 1]^d$ -valued stationary processes on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and ρ be a given real measurable function. The unknown regression function associated to $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$ and ρ is defined by

$$g(\mathbf{x}) = \mathbb{E}(\rho(Y)|\mathbf{X} = \mathbf{x}), \quad \mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d.$$

The additive case is considered: there exist d unknown real measurable functions g_1, \dots, g_d and an unknown real number μ such that

$$g(\mathbf{x}) = \mu + \sum_{\ell=1}^d g_\ell(x_\ell). \quad (1.1)$$

For any $\ell \in \{1, \dots, d\}$, our goal is to estimate g_ℓ from n observations $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n)$ of $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$.

In the case where $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$ is a *i.i.d.* process, this additive regression model becomes the standard one. In this case, Stone [28, 29, 30] proved that g can be estimated with the same rate of estimation error as in the one-dimensional case. The estimation of the component g_ℓ has been investigated in several papers via various methods (kernel, splines, wavelets, ...). See e.g. [4], [16], [18], [23, 24], [1], [2] [27], [33], [26] and [13].

In some applications, the *i.i.d.* assumptions on the observations is often too stringent. For this reason, some authors have explored the estimation of g_ℓ in the dependent case. When $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$ is a strongly mixing process, this problem has been addressed by [5], [9] and results for continuous time processes under strong mixing condition have been obtained by [10, 11]. In particular, they have developed non-adaptive kernel estimators for g_ℓ and studied its asymptotic properties. However, to the best of our knowledge, the adaptive estimation of g_ℓ for dependant processes has been addressed only by [14]. The lack of results for adaptive estimation motivates this study.

To reach our goal, as in [33], we combine the marginal integration technique introduced by [22] with wavelet methods. Wavelets enable us to construct adaptive estimator which attains sharp rates of convergence under

mild assumption on the smoothness of the unknown function. For a survey on wavelets in nonparametric estimation, we refer to [15]. In this study, we develop a wavelet estimator based on an individual selection named “hard thresholding”. The idea of this estimator is simple: we estimate unknown wavelet coefficients of g_ℓ based on the observations, we select the greatest and ignore the others, then we reconstruct the chosen wavelet coefficients estimators on the considered wavelet basis. Adopting the minimax point of view under the \mathbb{L}_2 risk over Besov balls, we prove that our adaptive estimator attains a sharp rate of convergence; it corresponds to the optimal one in the *i.i.d.* case for the univariate regression estimation problem up to an extra logarithmic term. And only this logarithmic term is slightly deteriorated by the dependence assumption.

The rest of the paper is organized as follows. Section 2 presents additional assumptions on the model. In Section 3, we describe wavelet basis on $[0, 1]$, Besov balls and wavelets tensor-products on $[0, 1]^d$. Our wavelet hard thresholding estimator is defined in Section 4. Its rate of convergence under the \mathbb{L}_2 risk over Besov balls are determined Section 5. The proofs are postponed in Section 6.

2 Notations and assumptions

We formulate the following assumptions:

On the variables.

- For any $i \in \{1, \dots, n\}$, we set $\mathbf{X}_i = (X_{1,i}, \dots, X_{d,i})$. We suppose that
 - for any $i \in \{1, \dots, n\}$, $X_{1,i}, \dots, X_{d,i}$ are identically distributed with the common distribution $\mathcal{U}([0, 1])$,
 - $\mathbf{X}_1, \dots, \mathbf{X}_n$ are identically distributed with the common known density f .
- We suppose that the following identifiability condition is satisfied: for any $\ell \in \{1, \dots, d\}$ and $i \in \{1, \dots, n\}$, we have

$$\mathbb{E}(g_\ell(X_{\ell,i})) = 0. \tag{2.1}$$

Strongly mixing. Throughout this work, we use the strong mixing dependence structure on $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$. For any $m \in \mathbb{Z}$, we define the m -th strongly mixing coefficient of $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$ by

$$\alpha_m = \sup_{(A,B) \in \mathcal{F}_{-\infty,0}^{(Y,\mathbf{X})} \times \mathcal{F}_{m,\infty}^{(Y,\mathbf{X})}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where $\mathcal{F}_{-\infty,0}^{(Y,\mathbf{X})}$ is the σ -algebra generated by $\dots, (Y_{-1}, \mathbf{X}_{-1}), (Y_0, \mathbf{X}_0)$ and $\mathcal{F}_{m,\infty}^{(Y,\mathbf{X})}$ is the σ -algebra generated by $(Y_m, \mathbf{X}_m), (Y_{m+1}, \mathbf{X}_{m+1}), \dots$

We suppose that there exist three constants $\gamma > 0$, $c > 0$ and $\theta > 0$ such that, for any integer $m \geq 1$,

$$\alpha_m \leq \gamma \exp(-cm^\theta). \quad (2.2)$$

Notice that, for (2.2), the standard *i.i.d.* case corresponds to $\theta \rightarrow \infty$. Further details on strongly mixing dependence can be found in [3], [32], [12], [21] and [6].

Boundedness assumptions.

- We suppose that there exists a known constant $C > 0$ such that

$$\sup_{y \in \mathbb{R}} |\rho(y)| \leq C. \quad (2.3)$$

- We suppose that there exists a known constant $C > 0$ such that

$$\int_{-\infty}^{\infty} |\rho(y)| dy \leq C. \quad (2.4)$$

- We suppose that there exists a known constant $c > 0$ such that

$$\inf_{\mathbf{x} \in [0,1]^d} f(\mathbf{x}) \geq c. \quad (2.5)$$

- For any $m \in \{1, \dots, n\}$, let $f_{(Y_0, \mathbf{X}_0, Y_m, \mathbf{X}_m)}$ be the density of $(Y_0, \mathbf{X}_0, Y_m, \mathbf{X}_m)$, $f_{(Y_0, \mathbf{X}_0)}$ the density of (Y_0, \mathbf{X}_0) and, for any $(y, \mathbf{x}, y_*, \mathbf{x}_*) \in \mathbb{R} \times [0, 1]^d \times \mathbb{R} \times [0, 1]^d$,

$$\begin{aligned} h_m(y, \mathbf{x}, y_*, \mathbf{x}_*) = \\ f_{(Y_0, \mathbf{X}_0, Y_m, \mathbf{X}_m)}(y, \mathbf{x}, y_*, \mathbf{x}_*) - f_{(Y_0, \mathbf{X}_0)}(y, \mathbf{x})f_{(Y_0, \mathbf{X}_0)}(y_*, \mathbf{x}_*). \end{aligned} \quad (2.6)$$

We suppose that there exists a known constant $C > 0$ such that

$$\sup_{m \in \{1, \dots, n\}} \sup_{(y, \mathbf{x}, y_*, \mathbf{x}_*) \in \mathbb{R} \times [0, 1]^d \times \mathbb{R} \times [0, 1]^d} |h_m(y, \mathbf{x}, y_*, \mathbf{x}_*)| \leq C. \quad (2.7)$$

Such boundedness assumptions are standard for the estimation of g_ℓ from a strongly mixing process. See e.g. [10, 11].

3 Wavelets tensor-products and Besov balls

3.1 Wavelet bases on $[0, 1]$

Let N be a positive integer. We consider an orthonormal wavelet basis generated by dilations and translations of a "father" Daubechies-type wavelet ϕ and a "mother" Daubechies-type wavelet ψ of the family $db2N$. In particular, ϕ and ψ have compact supports, $\int \phi(x) dx = 1$ and, for any $r \in \{0, \dots, N-1\}$, $\int x^r \psi(x) dx = 0$.

Set

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

Then, with an appropriate treatment at the boundaries, there exists an integer τ satisfying $2^\tau \geq 2N$ such that, for any integer $j_* \geq \tau$, the system

$$\{\phi_{j_*,k}(\cdot), k \in \{0, \dots, 2^{j_*}-1\}; \psi_{j,k}(\cdot); j \in \mathbb{N} - \{0, \dots, j_*-1\}, k \in \{0, \dots, 2^j-1\}\},$$

is an orthonormal basis of $\mathbb{L}_2([0, 1]) = \{h : [0, 1] \rightarrow \mathbb{R}; \int_0^1 h^2(x) dx < \infty\}$. See [7] and [19].

For any integer $j_* \geq \tau$, any $h \in \mathbb{L}_2([0, 1])$ can be expanded into a wavelet series as

$$h(x) = \sum_{k=0}^{2^{j_*}-1} \alpha_{j_*,k} \phi_{j_*,k}(x) + \sum_{j=j_*}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x), \quad x \in [0, 1],$$

where

$$\alpha_{j,k} = \int_0^1 h(x) \phi_{j,k}(x) dx, \quad \beta_{j,k} = \int_0^1 h(x) \psi_{j,k}(x) dx. \quad (3.1)$$

3.2 Besov balls

Let $M > 0$, $s > 0$, $p \geq 1$ and $q \geq 1$. A function h in $\mathbb{L}_2([0, 1])$ belongs to $B_{p,q}^s(M)$ if, and only if, there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (3.1) satisfy

$$\left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^q \right)^{1/q} \leq M^*.$$

In this expression, s is a smoothness parameter and p and q are norm parameters. For a particular choice of s , p and q , the Besov balls contain the standard Hölder and Sobolev balls. See [20].

3.3 Wavelet tensor-product bases on $[0, 1]^d$

For the purpose of this paper, we use a compactly supported wavelet-tensor product basis on $[0, 1]^d$ based on the Daubechies wavelets. For any $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$, we construct 2^d functions as follows: a scale function

$$\Phi(\mathbf{x}) = \prod_{v=1}^d \phi(x_v)$$

and $2^d - 1$ wavelet functions:

$$\Psi_u(\mathbf{x}) = \begin{cases} \psi(x_u) \prod_{\substack{v=1 \\ v \neq u}}^d \phi(x_v) & \text{when } u \in \{1, \dots, d\}, \\ \prod_{v \in A_u} \psi(x_v) \prod_{v \notin A_u} \phi(x_v) & \text{when } u \in \{d+1, \dots, 2^d - 1\}, \end{cases}$$

where $(A_u)_{u \in \{d+1, \dots, 2^d - 1\}}$ forms the set of all the non void subsets of $\{1, \dots, d\}$ of cardinal superior or equal to 2.

For any integer j and any $\mathbf{k} = (k_1, \dots, k_d)$, we set

$$\Phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2} \Phi(2^j x_1 - k_1, \dots, 2^j x_d - k_d),$$

for any $u \in \{1, \dots, 2^d - 1\}$,

$$\Psi_{j,\mathbf{k},u}(\mathbf{x}) = 2^{jd/2} \Psi_u(2^j x_1 - k_1, \dots, 2^j x_d - k_d).$$

We set $D_j = \{0, \dots, 2^j - 1\}^d$. Then, with an appropriate treatment at the boundaries, there exists an integer τ such that the system

$$\{\Phi_{\tau, \mathbf{k}}, \mathbf{k} \in D_\tau; (\Psi_{j, \mathbf{k}, u})_{u \in \{1, \dots, 2^{d-1}\}}, \quad j \in \mathbb{N} - \{0, \dots, \tau - 1\}, \mathbf{k} \in D_j\}$$

forms an orthonormal basis of $\mathbb{L}_2([0, 1]^d) = \{h : [0, 1]^d \rightarrow \mathbb{R}; \int_{[0, 1]^d} h^2(x) dx < \infty\}$.

For any integer j_* such that $j_* \geq \tau$, a function $h \in \mathbb{L}_2([0, 1]^d)$ can be expanded into a wavelet series as

$$h(\mathbf{x}) = \sum_{\mathbf{k} \in D_{j_*}} \alpha_{j_*, \mathbf{k}} \Phi_{j_*, \mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^{d-1}} \sum_{j=j_*}^{\infty} \sum_{\mathbf{k} \in D_j} \beta_{j, \mathbf{k}, u} \Psi_{j, \mathbf{k}, u}(\mathbf{x}), \quad \mathbf{x} \in [0, 1]^d,$$

where

$$\alpha_{j, \mathbf{k}} = \int_{[0, 1]^d} h(x) \Phi_{j, \mathbf{k}}(\mathbf{x}) d\mathbf{x}, \quad \beta_{j, \mathbf{k}, u} = \int_{[0, 1]^d} h(\mathbf{x}) \Psi_{j, \mathbf{k}, u}(\mathbf{x}) d\mathbf{x}. \quad (3.2)$$

For further details about wavelet bases on $[0, 1]^d$, we refer to [7] and [19].

4 Estimators

4.1 Wavelet coefficients estimators

The following proposition present a wavelet decomposition of g_ℓ based on the ‘‘marginal integration’’ method (introduced by [22]) and the wavelet-tensor product basis on $[0, 1]^d$.

Proposition 4.1 *Suppose that (2.1) holds. Then, for any $j_* \geq \tau$ and $\ell \in \{1, \dots, d\}$, we can write*

$$g_\ell(x) = \sum_{k=1}^{2^{j_*}-1} a_{j_*, k, \ell} \phi_{j_*, k}(x) + \sum_{j=j_*}^{\infty} \sum_{k=1}^{2^j-1} b_{j, k, \ell} \psi_{j, k}(x) - \mu, \quad x \in [0, 1],$$

where

$$a_{j, k, \ell} = a_{j, k_\ell, \ell} = 2^{-j(d-1)/2} \int_{[0, 1]^d} g(\mathbf{x}) \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) d\mathbf{x}, \quad (4.1)$$

$$b_{j,k,\ell} = \hat{b}_{j,k,\ell} = 2^{-j(d-1)/2} \int_{[0,1]^d} g(\mathbf{x}) \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) d\mathbf{x} \quad (4.2)$$

$\mathbf{k}_{-\ell} = (k_1, \dots, k_{\ell-1}, k_{\ell+1}, \dots, k_d)$ and $D_j^* = \{0, \dots, 2^j - 1\}^{d-1}$.

Remark that, due to the definitions of g and $\Psi_{j,\mathbf{k},\ell}$, $b_{j,k,\ell}$ is exactly the “mother-wavelet coefficient” of g_ℓ i.e.

$$b_{j,k,\ell} = \int_0^1 g_\ell(x) \psi_{j,k}(x) dx = \beta_{j,k}. \quad (4.3)$$

Thanks to Proposition 4.1, the first step to estimate g_ℓ consists in estimating the unknown coefficients $a_{j,k,\ell}$ (4.1) and $b_{j,k,\ell}$ (4.2). In this study, we consider respectively

$$\hat{a}_{j,k,\ell} = \hat{a}_{j,k_\ell,\ell} = 2^{-j(d-1)/2} \frac{1}{n} \sum_{i=1}^n \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_i) \quad (4.4)$$

and

$$\hat{b}_{j,k,\ell} = \hat{b}_{j,k_\ell,\ell} = 2^{-j(d-1)/2} \frac{1}{n} \sum_{i=1}^n \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{X}_i). \quad (4.5)$$

These estimators enjoy powerful statistical properties. Some of them are presented in Proposition 4.2, 4.3, 4.4 and 4.5 below.

Proposition 4.2 *Suppose that (2.1) holds. For any $j \geq \tau$, $\ell \in \{1, \dots, d\}$ and $k \in \{0, \dots, 2^j - 1\}$, $\hat{a}_{j,k,\ell}$ (4.4) and $\hat{b}_{j,k,\ell}$ (4.5) are unbiased estimators of $a_{j,k,\ell}$ and $b_{j,k,\ell}$ respectively.*

Proposition 4.3 (Moment inequality I) *Suppose that the assumptions of Section 2 hold. Let $j \geq \tau$ such that $2^j \leq n$, $k \in \{0, \dots, 2^j - 1\}$, $\ell \in \{1, \dots, d\}$, $\hat{a}_{j,k,\ell}$ be (4.4) and $\hat{b}_{j,k,\ell}$ be (4.5). Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left((\hat{a}_{j,k,\ell} - a_{j,k,\ell})^2 \right) \leq C \frac{1}{n}, \quad \mathbb{E} \left((\hat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \right) \leq C \frac{1}{n}.$$

Mention that, in the proof of Proposition 4.3, we only need to have the existence of two constants $C > 0$ and $q \in (0, 1)$ such that $\sum_{m=1}^n m^q \alpha_m^q \leq C < \infty$. And this inequality is obviously satisfied by (2.2).

Proposition 4.4 (Moment inequality II) *Suppose that the assumptions of Section 2 hold. Let $j \geq \tau$ such that $2^j \leq n$, $k \in \{0, \dots, 2^j - 1\}$, $\ell \in \{1, \dots, d\}$ and $\widehat{b}_{j,k,\ell}$ be (4.5). Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^4 \right) \leq C \frac{2^j}{n}.$$

Proposition 4.5 (Concentration inequality) *Suppose that the assumptions of Section 2 hold. Let $j \geq \tau$ such that $2^j \leq n/(\ln n)^{1+1/\theta}$, $k \in \{0, \dots, 2^j - 1\}$, $\ell \in \{1, \dots, d\}$, $\widehat{b}_{j,k,\ell}$ be (4.5) and $\lambda_n = ((\ln n)^{1+1/\theta}/n)^{1/2}$. Then there exist two constants $C > 0$ and $\kappa > 0$ such that*

$$\mathbb{P} \left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \geq \kappa \lambda_n / 2 \right) \leq C \frac{1}{n^4}.$$

4.2 Hard thresholding estimator

Let $\ell \in \{1, \dots, d\}$. We investigate the estimation of g_ℓ by the hard thresholding estimator:

$$\widehat{g}_\ell(x) = \sum_{k=0}^{2^\tau-1} \widehat{a}_{\tau,k,\ell} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \widehat{b}_{j,k,\ell} \mathbf{1}_{\{|\widehat{b}_{j,k,\ell}| \geq \kappa \lambda_n\}} \psi_{j,k}(x) - \widehat{\mu}, \quad (4.6)$$

$x \in [0, 1]$ where $\widehat{a}_{j,k,\ell}$ is (4.4), $\widehat{b}_{j,k,\ell}$ is (4.5),

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^n \rho(Y_i), \quad (4.7)$$

j_1 is the integer satisfying $2^{j_1} = \lceil n/(\ln n)^{1+1/\theta} \rceil$, κ is a large enough constant (the one in Proposition 4.5) and

$$\lambda_n = \sqrt{\frac{(\ln n)^{1+1/\theta}}{n}}.$$

Note that, due to the assumptions on the model, our wavelet hard thresholding estimator (4.6) is simpler than the one of [33]. However, contrary to the one of [33], it takes into account the dependence of the observations. In its construction, this is characterized by the presence of the factor θ of (2.2) in j_1 and the threshold λ_n .

For further details on wavelet hard thresholding estimators for standard nonparametric estimation, we refer to [15].

5 Result

Theorem 5.1 *Let $\ell \in \{1, \dots, d\}$. Suppose that the assumptions of Section 2 hold. Let \widehat{g}_ℓ be (4.6). Suppose that $g_\ell \in B_{p,q}^s(M)$ with $q \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 1/p\}$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left(\int_0^1 (\widehat{g}_\ell(x) - g_\ell(x))^2 dx \right) \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}.$$

The proof of Theorem 5.1 is based on a suitable decomposition of the \mathbb{L}_2 risk and the statistical properties of (4.4) and (4.5) presented in Propositions 4.2, 4.3, 4.4 and 4.5 below.

6 Relation to prior work

The rate $((\ln n)^{1+1/\theta}/n)^{2s/(2s+1)}$ is the optimal one for the standard one-dimensional regression model with uniform random design up to an extra logarithmic term. See e.g. [15] and [31].

Theorem 5.1 provides an “adaptive contribution” to the results of [5], [9] and [10, 11]. Furthermore, if we confine our study to the *i.i.d.*, we have a similar result to [33, Theorem 3] but without the condition $s > \max(d/2, d/p)$. However, we have more restrictive assumptions on the model (ρ is bounded from above, the density of \mathbf{X} is known, etc.).

7 Proofs

In this section, the quantity 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 .

7.1 Technical results on wavelets

Proof of Proposition 4.1. Because of (2.3), we have $g \in \mathbb{L}_2([0, 1]^d)$. For any $j_* \geq \tau$, we can expand g on our wavelet-tensor product basis as

$$g(\mathbf{x}) = \sum_{\mathbf{k} \in D_{j^*}} \alpha_{j^*, \mathbf{k}} \Phi_{j^*, \mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^{d-1}} \sum_{j=j^*}^{\infty} \sum_{\mathbf{k} \in D_j} \beta_{j, \mathbf{k}, u} \Psi_{j, \mathbf{k}, u}(\mathbf{x}), \quad \mathbf{x} \in [0, 1]^d \quad (7.1)$$

where

$$\alpha_{j, \mathbf{k}} = \int_{[0, 1]^d} g(\mathbf{x}) \Phi_{j, \mathbf{k}}(\mathbf{x}) d\mathbf{x}, \quad \beta_{j, \mathbf{k}, u} = \int_{[0, 1]^d} g(\mathbf{x}) \Psi_{j, \mathbf{k}, u}(\mathbf{x}) d\mathbf{x}.$$

Moreover, using the ‘‘marginal integration’’ method based on (2.1), we can write

$$g_\ell(x_\ell) = \int_{[0, 1]^{d-1}} g(\mathbf{x}) \prod_{\substack{v=1 \\ v \neq \ell}}^d dx_v - \mu, \quad x_\ell \in [0, 1]. \quad (7.2)$$

Since $\int_0^1 \phi_{j, k}(x) dx = 2^{-j/2}$ and $\int_0^1 \psi_{j, k}(x) dx = 0$, observe that

$$\int_{[0, 1]^{d-1}} \Phi_{j^*, \mathbf{k}}(\mathbf{x}) \prod_{\substack{v=1 \\ v \neq \ell}}^d dx_v = 2^{-j^*(d-1)/2} \phi_{j^*, k_\ell}(x_\ell)$$

and

$$\int_{[0, 1]^{d-1}} \Psi_{j, \mathbf{k}, u}(\mathbf{x}) \prod_{\substack{v=1 \\ v \neq \ell}}^d dx_v = \begin{cases} 2^{-j(d-1)/2} \psi_{j, k_\ell}(x_\ell) & \text{if } u = \ell, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, putting (7.1) in (7.2) and writing $x = x_\ell$, we obtain

$$g_\ell(x) = \sum_{\mathbf{k} \in D_{j^*}} 2^{-j^*(d-1)/2} \alpha_{j^*, \mathbf{k}} \phi_{j^*, k_\ell}(x) + \sum_{j=j^*}^{\infty} \sum_{\mathbf{k} \in D_j} 2^{-j(d-1)/2} \beta_{j, \mathbf{k}, \ell} \psi_{j, k_\ell}(x) - \mu.$$

Or, equivalently,

$$g_\ell(x) = \sum_{k=1}^{2^{j^*-1}} a_{j^*, k, \ell} \phi_{j^*, k}(x) + \sum_{j=j^*}^{\infty} \sum_{k=1}^{2^{j-1}} b_{j, k, \ell} \psi_{j, k}(x) - \mu,$$

where

$$a_{j, k, \ell} = a_{j, k_\ell, \ell} = 2^{-j(d-1)/2} \int_{[0, 1]^d} g(\mathbf{x}) \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) d\mathbf{x}$$

and

$$b_{j,k,\ell} = b_{j,k_\ell,\ell} = 2^{-j(d-1)/2} \int_{[0,1]^d} g(\mathbf{x}) \sum_{\mathbf{k}-\ell \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) d\mathbf{x}.$$

Proposition 4.1 is proved. ■

Proposition 7.1 *For any $\ell \in \{1, \dots, d\}$, $j \geq \tau$ and $k = k_\ell \in \{0, \dots, 2^j - 1\}$, set*

$$h_{j,k}^{(1)}(\mathbf{x}) = \sum_{\mathbf{k}-\ell \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}), \quad h_{j,k}^{(2)}(\mathbf{x}) = \sum_{\mathbf{k}-\ell \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}), \quad \mathbf{x} \in [0, 1]^d.$$

Then there exists a constant $C > 0$ such that, for any $a \in \{1, 2\}$,

$$\sup_{\mathbf{x} \in [0,1]^d} |h_{j,k}^{(a)}(\mathbf{x})| \leq C2^{jd/2}, \quad \int_{[0,1]^d} |h_{j,k}^{(a)}(\mathbf{x})| d\mathbf{x} \leq C2^{-j/2}2^{j(d-1)/2}$$

and

$$\int_{[0,1]^d} (h_{j,k}^{(a)}(\mathbf{x}))^2 d\mathbf{x} = 2^{j(d-1)}.$$

Proof of Proposition 7.1.

- Since $\sup_{x \in [0,1]} |\phi_{j,k}(x)| \leq C2^{j/2}$ and $\sup_{x \in [0,1]} \sum_{k=0}^{2^j-1} |\phi_{j,k}(x)| \leq C2^{j/2}$, we obtain

$$\sup_{\mathbf{x} \in [0,1]^d} |h_{j,k}^{(1)}(\mathbf{x})| = \left(\sup_{x \in [0,1]} |\phi_{j,k}(x)| \right) \left(\sup_{x \in [0,1]} \sum_{k=0}^{2^j-1} |\phi_{j,k}(x)| \right)^{d-1} \leq C2^{jd/2}.$$

- Using $\int_0^1 |\phi_{j,k}(x)| dx = C2^{-j/2}$, we obtain

$$\begin{aligned} \int_{[0,1]^d} |h_{j,k}^{(1)}(\mathbf{x})| d\mathbf{x} &\leq \left(\int_0^1 |\phi_{j,k}(x)| dx \right) \left(\sum_{k=0}^{2^j-1} \int_0^1 |\phi_{j,k}(x)| dx \right)^{d-1} \\ &= C2^{-j/2}2^{j(d-1)/2}. \end{aligned}$$

- Since, for any $(u_{\mathbf{k}})_{\mathbf{k} \in D_j}$, $\int_{[0,1]^d} \left(\sum_{\mathbf{k} \in D_j} u_{\mathbf{k}} \Phi_{j,\mathbf{k}}(\mathbf{x}) \right)^2 d\mathbf{x} = \sum_{\mathbf{k} \in D_j} u_{\mathbf{k}}^2$, we obtain

$$\int_{[0,1]^d} (h_{j,k}^{(1)}(\mathbf{x}))^2 d\mathbf{x} = \int_{[0,1]^d} \left(\sum_{\mathbf{k}-\ell \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) \right)^2 d\mathbf{x} = 2^{j(d-1)}.$$

Proceeding in a similar fashion, using $\sup_{x \in [0,1]} |\psi_{j,k}(x)| \leq C2^{j/2}$, $\int_0^1 |\psi_{j,k}(x)| dx = C2^{-j/2}$ and, for any $(u_{\mathbf{k}})_{\mathbf{k} \in D_j}$, $\int_{[0,1]^d} \left(\sum_{\mathbf{k} \in D_j} u_{\mathbf{k}} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) \right)^2 d\mathbf{x} = \sum_{\mathbf{k} \in D_j} u_{\mathbf{k}}^2$, we obtain the same results for $h_{j,k}^{(2)}$.

This ends the proof of Proposition 7.1. ■

7.2 Statistical properties of the coefficients estimators

Proof of Proposition 4.2. We have

$$\begin{aligned} \mathbb{E}(\widehat{a}_{j,k,\ell}) &= 2^{-j(d-1)/2} \mathbb{E} \left(\frac{\rho(Y_1)}{f(\mathbf{X}_1)} \sum_{\mathbf{k}-\ell \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_1) \right) \\ &= 2^{-j(d-1)/2} \mathbb{E} \left(\mathbb{E}(\rho(Y_1) | \mathbf{X}_1) \frac{1}{f(\mathbf{X}_1)} \sum_{\mathbf{k}-\ell \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_1) \right) \\ &= 2^{-j(d-1)/2} \mathbb{E} \left(\frac{g(\mathbf{X}_1)}{f(\mathbf{X}_1)} \sum_{\mathbf{k}-\ell \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_1) \right) \\ &= 2^{-j(d-1)/2} \int_{[0,1]^d} \frac{g(\mathbf{x})}{f(\mathbf{x})} \sum_{\mathbf{k}-\ell \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= 2^{-j(d-1)/2} \int_{[0,1]^d} g(\mathbf{x}) \sum_{\mathbf{k}-\ell \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) d\mathbf{x} = a_{j,k,\ell}. \end{aligned}$$

Proceeding in a similar fashion, we prove that $\mathbb{E}(\widehat{b}_{j,k,\ell}) = b_{j,k,\ell}$.

Proposition 4.2 is proved. ■

Proof of Proposition 4.3. For the sake of simplicity, for any $i \in \{1, \dots, n\}$, set

$$Z_i = \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_i).$$

Thanks to Proposition 4.2, we have

$$\mathbb{E}((\widehat{a}_{j,k,\ell} - a_{j,k,\ell})^2) = \mathbb{V}(\widehat{a}_{j,k,\ell}) = 2^{-j(d-1)} \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^n Z_i\right). \quad (7.3)$$

An elementary covariance decomposition gives

$$\begin{aligned} \mathbb{V}\left(\sum_{i=1}^n Z_i\right) &= n\mathbb{V}(Z_1) + 2 \sum_{v=2}^n \sum_{u=1}^{v-1} \text{Cov}(Z_v, Z_u) \\ &\leq n\mathbb{V}(Z_1) + 2 \left| \sum_{v=2}^n \sum_{u=1}^{v-1} \text{Cov}(Z_v, Z_u) \right|. \end{aligned} \quad (7.4)$$

Using (2.3), (2.5) and Proposition 7.1, we have

$$\begin{aligned} \mathbb{V}(Z_1) &\leq \mathbb{E}(Z_1^2) \leq \frac{\sup_{y \in \mathbb{R}} \rho^2(y)}{\inf_{\mathbf{x} \in [0,1]^d} f(\mathbf{x})} \mathbb{E}\left(\frac{1}{f(\mathbf{X}_1)} \left(\sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_1)\right)^2\right) \\ &\leq C \int_{[0,1]^d} \frac{1}{f(\mathbf{x})} \left(\sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x})\right)^2 f(\mathbf{x}) d\mathbf{x} \\ &= C \int_{[0,1]^d} \left(\sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x})\right)^2 d\mathbf{x} = C2^{j(d-1)}. \end{aligned} \quad (7.5)$$

It follows from the stationarity of $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$ and $2^j \leq n$ that

$$\left| \sum_{v=2}^n \sum_{u=1}^{v-1} \text{Cov}(Z_v, Z_u) \right| = \left| \sum_{m=1}^n (n-m) \text{Cov}(Z_0, Z_m) \right| \leq R_1 + R_2, \quad (7.6)$$

where

$$R_1 = n \sum_{m=1}^{2^j-1} |\text{Cov}(Z_0, Z_m)|, \quad R_2 = n \sum_{m=2^j}^n |\text{Cov}(Z_0, Z_m)|.$$

Bound for R_1 . Let, for any $(y, \mathbf{x}, y_*, \mathbf{x}_*) \in \mathbb{R} \times [0, 1]^d \times \mathbb{R} \times [0, 1]^d$, $h_m(y, \mathbf{x}, y_*, \mathbf{x}_*)$ be (2.6). Using (2.7), (2.4) and Proposition 7.1, we obtain

$$\begin{aligned}
& |\text{Cov}(Z_0, Z_m)| \\
&= \left| \int_{-\infty}^{\infty} \int_{[0,1]^d} \int_{-\infty}^{\infty} \int_{[0,1]^d} h_m(y, \mathbf{x}, y_*, \mathbf{x}_*) \times \right. \\
&\quad \left. \left(\frac{\rho(y)}{f(\mathbf{x})} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) \frac{\rho(y_*)}{f(\mathbf{x}_*)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}_*) \right) dy d\mathbf{x} dy_* d\mathbf{x}_* \right| \\
&\leq \int_{-\infty}^{\infty} \int_{[0,1]^d} \int_{-\infty}^{\infty} \int_{[0,1]^d} |h_m(y, \mathbf{x}, y_*, \mathbf{x}_*)| \times \\
&\quad \left| \frac{\rho(y)}{f(\mathbf{x})} \right| \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) \right| \left| \frac{\rho(y_*)}{f(\mathbf{x}_*)} \right| \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}_*) \right| dy d\mathbf{x} dy_* d\mathbf{x}_* \\
&\leq C \left(\int_{-\infty}^{\infty} |\rho(y)| dy \right)^2 \left(\int_{[0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) \right| d\mathbf{x} \right)^2 \leq C 2^{-j} 2^{j(d-1)}.
\end{aligned}$$

Therefore

$$R_1 \leq C n 2^{-j} 2^{j(d-1)} 2^j = C n 2^{j(d-1)}. \quad (7.7)$$

Bound for R_2 . By the Davydov inequality for strongly mixing processes (see [8]), for any $q \in (0, 1)$, we have

$$\begin{aligned}
|\text{Cov}(Z_0, Z_m)| &\leq 10 \alpha_m^q \left(\mathbb{E} \left(|Z_0|^{2/(1-q)} \right) \right)^{1-q} \\
&\leq 10 \alpha_m^q \left(\frac{\sup_{y \in \mathbb{R}} |\rho(y)|}{\inf_{\mathbf{x} \in [0,1]^d} f(\mathbf{x})} \sup_{\mathbf{x} \in [0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) \right| \right)^{2q} (\mathbb{E}(Z_0^2))^{1-q}.
\end{aligned}$$

By (2.3), (2.5) and Proposition 7.1, we have

$$\begin{aligned}
\frac{\sup_{y \in \mathbb{R}} |\rho(y)|}{\inf_{\mathbf{x} \in [0,1]^d} f(\mathbf{x})} \sup_{\mathbf{x} \in [0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) \right| &\leq C \sup_{\mathbf{x} \in [0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j, \mathbf{k}}(\mathbf{x}) \right| \\
&\leq C 2^{jd/2}.
\end{aligned}$$

By (7.5), we have

$$\mathbb{E}(Z_0^2) \leq C 2^{j(d-1)}.$$

Therefore

$$|\mathbb{Cov}(Z_0, Z_m)| \leq C2^{qj}2^{j(d-1)}\alpha_m^q.$$

Observe that $\sum_{m=1}^{\infty} m^q \alpha_m^q = \gamma^q \sum_{m=1}^{\infty} m^q \exp(-cqm^\theta) < \infty$. Hence

$$R_2 \leq Cn2^{qj}2^{j(d-1)} \sum_{m=2^j}^n \alpha_m^q \leq Cn2^{j(d-1)} \sum_{m=2^j}^n m^q \alpha_m^q \leq Cn2^{j(d-1)}. \quad (7.8)$$

Putting (7.6), (7.7) and (7.8) together, we have

$$\left| \sum_{v=2}^n \sum_{u=1}^{v-1} \mathbb{Cov}(Z_v, Z_u) \right| \leq Cn2^{j(d-1)}. \quad (7.9)$$

Combining (7.3), (7.4), (7.5) and (7.9), we obtain

$$\mathbb{E} \left((\widehat{a}_{j,k,\ell} - a_{j,k,\ell})^2 \right) \leq C2^{-j(d-1)} \frac{1}{n^2} n 2^{j(d-1)} = C \frac{1}{n}.$$

Proceeding in a similar fashion, we prove that

$$\mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \right) \leq C \frac{1}{n}.$$

This ends the proof of Proposition 4.3. ■

Proof of Proposition 4.4. It follows from (2.3), (2.5) and Proposition 7.1 that

$$\begin{aligned} |\widehat{b}_{j,k,\ell}| &\leq 2^{-j(d-1)/2} \frac{1}{n} \sum_{i=1}^n \frac{|\rho(Y_i)|}{|f(\mathbf{X}_i)|} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{X}_i) \right| \\ &\leq 2^{-j(d-1)/2} \frac{\sup_{y \in \mathbb{R}} |\rho(y)|}{\inf_{\mathbf{x} \in [0,1]^d} f(\mathbf{x})} \sup_{\mathbf{x} \in [0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) \right| \\ &\leq C2^{-j(d-1)/2} 2^{jd/2} = C2^{j/2}. \end{aligned}$$

Because of (2.3), we have $\sup_{\mathbf{x} \in [0,1]^d} |g(\mathbf{x})| \leq C$. It follows from Proposition 7.1 that

$$\begin{aligned} |b_{j,k,\ell}| &\leq 2^{-j(d-1)/2} \int_{[0,1]^d} |g(\mathbf{x})| \left| \sum_{\mathbf{k}-\ell \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) \right| d\mathbf{x} \\ &\leq C 2^{-j(d-1)/2} \int_{[0,1]^d} \left| \sum_{\mathbf{k}-\ell \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) \right| d\mathbf{x} \\ &\leq C 2^{-j(d-1)/2} 2^{-j} 2^{jd/2} = C 2^{-j/2}. \end{aligned} \quad (7.10)$$

Hence

$$|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \leq |\widehat{b}_{j,k,\ell}| + |b_{j,k,\ell}| \leq C 2^{j/2}. \quad (7.11)$$

It follows from (7.11) and Proposition 4.3 that

$$\mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^4 \right) \leq C 2^j \mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \right) \leq C \frac{2^j}{n}.$$

The proof of Proposition 4.4 is complete. ■

Proof of Proposition 4.5. Let us present a Bernstein inequality for exponentially strongly mixing process. See [25] and [17, Theorem 2.1 with (2.1)].

Lemma 7.1 ([25] and [17]) *Let $\gamma > 0$, $c > 0$, $\theta > 0$ and $(Y_i)_{i \in \mathbb{Z}}$ be a stationary process such that, for any integer $m \geq 1$, the associated m -th strongly mixing coefficient satisfies*

$$\alpha_m \leq \gamma \exp(-cm^\theta).$$

Let n be a positive integer, $h : \mathbb{R} \rightarrow \mathbb{C}$ be a measurable function and, for any $i \in \mathbb{Z}$, $U_i = h(Y_i)$. We assume that $\mathbb{E}(U_1) = 0$ and there exists a constant $M > 0$ satisfying $|U_1| \leq M$. Then, for any $m \in \{1, \dots, n\}$ and $\lambda > 4mM/n$, we have

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n U_i \right| \geq \lambda \right) \leq 4 \exp \left(- \frac{\lambda^2 n}{m(64\mathbb{E}(U_1^2) + 8\lambda M/3)} \right) + 4\gamma \frac{n}{m} \exp(-cm^\theta).$$

For any $i \in \{1, \dots, n\}$, set

$$U_i = 2^{-j(d-1)/2} \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{X}_i) - b_{j,k,\ell}.$$

Then we can write

$$\widehat{b}_{j,k,\ell} - b_{j,k,\ell} = \frac{1}{n} \sum_{i=1}^n U_i.$$

So

$$\mathbb{P} \left(\left| \widehat{b}_{j,k,\ell} - b_{j,k,\ell} \right| \geq \kappa \lambda_n / 2 \right) = \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n U_i \right| \geq \kappa \lambda_n / 2 \right),$$

where U_1, \dots, U_n are identically distributed, depend on $(Y_i, \mathbf{X}_i)_{i \in \mathbb{Z}}$ satisfying (2.2),

- by Proposition 4.2, we have $\mathbb{E}(U_1) = 0$,
- by the elementary inequality: $(x + y)^2 \leq 2(x^2 + y^2)$, $(x, y) \in \mathbb{R}^2$, (7.5) and (7.10), we have

$$\begin{aligned} \mathbb{E}(U_1^2) &\leq 2 \cdot 2^{-j(d-1)} \mathbb{E} \left(\left(\frac{\rho(Y_1)}{f(\mathbf{X}_1)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{X}_1) \right)^2 \right) + 2b_{j,k,\ell}^2 \\ &\leq C 2^{j(d-1)} 2^{-j(d-1)} + C 2^{-j} \leq C, \end{aligned}$$

- proceeding in a similar fashion to (7.11), we obtain $|U_1| \leq C 2^{j/2}$.

Lemma 7.1 applied with the random variables U_1, \dots, U_n , $\lambda = \kappa \lambda_n / 2$, $\lambda_n = ((\ln n)^{1+1/\theta} / n)^{1/2}$, $m = (u \ln n)^{1/\theta}$ with $u > 0$ (chosen later), $M = C 2^{j/2}$

and $2^j \leq n/(\ln n)^{1+1/\theta}$ gives

$$\begin{aligned}
& \mathbb{P} \left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \geq \kappa \lambda_n / 2 \right) \\
& \leq 4 \exp \left(-C \frac{\kappa^2 \lambda_n^2 n}{m(1 + \kappa \lambda_n M)} \right) + 4\gamma \frac{n}{m} \exp(-cm^\theta) \\
& \leq 4 \exp \left(-C \frac{\kappa^2 (\ln n)^{1+1/\theta}}{(u \ln n)^{1/\theta} (1 + \kappa 2^{j/2} ((\ln n)^{1+1/\theta}/n)^{1/2})} \right) \\
& + 4\gamma \frac{n}{(u \ln n)^{1/\theta}} \exp(-cu \ln n) \\
& = 4 \exp \left(-C \frac{\kappa^2 \ln n}{u^{1/\theta} (1 + \kappa 2^{j/2} ((\ln n)^{1+1/\theta}/n)^{1/2})} \right) + 4\gamma \frac{1}{(u \ln n)^{1/\theta}} n^{1-cu} \\
& \leq C \left(n^{-C\kappa^2/(u^{1/\theta}(1+\kappa))} + n^{1-cu} \right).
\end{aligned}$$

Therefore, for large enough κ and u , we have

$$\mathbb{P} \left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \geq \kappa \lambda_n / 2 \right) \leq C \frac{1}{n^4}.$$

This ends the proof of Proposition 4.5. ■

7.3 Proof of the main result

Proof of Theorem 5.1. Using Proposition 4.1, we have

$$\begin{aligned}
& \widehat{g}_\ell(x) - g_\ell(x) \\
& = \sum_{k=0}^{2^\tau-1} (\widehat{\alpha}_{\tau,k,\ell} - \alpha_{\tau,k,\ell}) \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} (\widehat{b}_{j,k,\ell} \mathbf{1}_{\{|\widehat{b}_{j,k,\ell}| \geq \kappa \lambda_n\}} - b_{j,k,\ell}) \psi_{j,k}(x) \\
& - \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} b_{j,k,\ell} \psi_{j,k}(x) - (\widehat{\mu} - \mu).
\end{aligned}$$

Using the elementary inequality: $(x+y)^2 \leq 2(x^2+y^2)$, $(x,y) \in \mathbb{R}^2$, and the orthonormality of the wavelet basis, we have

$$\mathbb{E} \left(\int_0^1 (\widehat{g}_\ell(x) - g_\ell(x))^2 dx \right) \leq 2(T + U + V + W), \quad (7.12)$$

where

$$T = \mathbb{E}((\widehat{\mu} - \mu)^2), \quad U = \sum_{k=0}^{2^\tau-1} \mathbb{E}((\widehat{\alpha}_{\tau,k,\ell} - \alpha_{\tau,k,\ell})^2),$$

$$V = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left((\widehat{b}_{j,k,\ell} \mathbf{1}_{\{|\widehat{b}_{j,k,\ell}| \geq \kappa \lambda_n\}} - b_{j,k,\ell})^2 \right), \quad W = \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2.$$

Bound for T . We proceed as in the proof of Proposition 4.3. By (2.1), we have $\mathbb{E}(\rho(Y_1)) = \mu$. Thanks to the stationarity of $(Y_i)_{i \in \mathbb{Z}}$, we have

$$T = \mathbb{V}(\widehat{\mu}) \leq \frac{1}{n} \mathbb{V}(\rho(Y_1)) + 2 \frac{1}{n} \sum_{m=1}^n |\text{Cov}(\rho(Y_0), \rho(Y_m))|.$$

Using (2.3), the Davydov inequality (see [8]) and (2.2), we obtain

$$T \leq C \frac{1}{n} \left(1 + \sum_{m=1}^n \alpha_m^q \right) \leq C \frac{1}{n} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \quad (7.13)$$

Bound for U . Using Proposition 4.3, we obtain

$$U \leq C 2^\tau \frac{1}{n} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \quad (7.14)$$

Bound for W . For $q \geq 1$ and $p \geq 2$, we have $B_{p,q}^s(M) \subseteq B_{2,\infty}^s(M)$. Hence, by (4.3),

$$\begin{aligned} W &\leq C \sum_{j=j_1+1}^{\infty} 2^{-2js} \leq C 2^{-2j_1 s} \leq C n^{-2s} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s} \\ &\leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

For $q \geq 1$ and $p \in [1, 2)$, we have $B_{p,q}^s(M) \subseteq B_{2,\infty}^{s+1/2-1/p}(M)$. Since $s > 1/p$, we have $s + 1/2 - 1/p > s/(2s + 1)$. So, by (4.3),

$$\begin{aligned} W &\leq C \sum_{j=j_1+1}^{\infty} 2^{-2j(s+1/2-1/p)} \leq C 2^{-2j_1(s+1/2-1/p)} \\ &\leq C n^{-2(s+1/2-1/p)} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2(s+1/2-1/p)} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

Hence, for $q \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 1/p\}$, we have

$$W \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \quad (7.15)$$

Bound for V . We have

$$V = V_1 + V_2 + V_3 + V_4, \quad (7.16)$$

where

$$V_1 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\{|\widehat{b}_{j,k,\ell}| \geq \kappa \lambda_n\}} \mathbf{1}_{\{|b_{j,k,\ell}| < \kappa \lambda_n/2\}} \right),$$

$$V_2 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\{|\widehat{b}_{j,k,\ell}| \geq \kappa \lambda_n\}} \mathbf{1}_{\{|b_{j,k,\ell}| \geq \kappa \lambda_n/2\}} \right),$$

$$V_3 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(b_{j,k,\ell}^2 \mathbf{1}_{\{|\widehat{b}_{j,k,\ell}| < \kappa \lambda_n\}} \mathbf{1}_{\{|b_{j,k,\ell}| \geq 2\kappa \lambda_n\}} \right)$$

and

$$V_4 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(b_{j,k,\ell}^2 \mathbf{1}_{\{|\widehat{b}_{j,k,\ell}| < \kappa \lambda_n\}} \mathbf{1}_{\{|b_{j,k,\ell}| < 2\kappa \lambda_n\}} \right).$$

Bounds for V_1 and V_3 . The following inclusions hold:

$$\left\{ |\widehat{b}_{j,k,\ell}| < \kappa \lambda_n, |b_{j,k,\ell}| \geq 2\kappa \lambda_n \right\} \subseteq \left\{ |\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa \lambda_n/2 \right\},$$

$$\left\{ |\widehat{b}_{j,k,\ell}| \geq \kappa \lambda_n, |b_{j,k,\ell}| < \kappa \lambda_n/2 \right\} \subseteq \left\{ |\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa \lambda_n/2 \right\}$$

$$\text{and } \left\{ |\widehat{b}_{j,k,\ell}| < \kappa \lambda_n, |b_{j,k,\ell}| \geq 2\kappa \lambda_n \right\} \subseteq \left\{ |b_{j,k,\ell}| \leq 2|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \right\}.$$

So

$$\max(V_1, V_3) \leq C \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\{|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa \lambda_n/2\}} \right).$$

Applying the Cauchy-Schwarz inequality and using Propositions 4.4, 4.5 and $2^j \leq n$, we have

$$\begin{aligned} & \mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\{|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa \lambda_n/2\}} \right) \\ & \leq \left(\mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^4 \right) \right)^{1/2} \left(\mathbb{P} \left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa \lambda_n/2 \right) \right)^{1/2} \\ & \leq C \left(\frac{2^j}{n} \right)^{1/2} \left(\frac{1}{n^4} \right)^{1/2} \leq C \frac{1}{n^2}. \end{aligned}$$

Therefore

$$\begin{aligned} \max(V_1, V_3) &\leq C \frac{1}{n^2} \sum_{j=\tau}^{j_1} 2^j \leq C \frac{1}{n^2} 2^{j_1} \leq C \frac{1}{n} \\ &\leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \end{aligned} \quad (7.17)$$

Bound for V_2 . Using Proposition 4.3, we obtain

$$\mathbb{E} \left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \right) \leq C \frac{1}{n} \leq C \frac{(\ln n)^{1+1/\theta}}{n}.$$

Hence

$$V_2 \leq C \frac{(\ln n)^{1+1/\theta}}{n} \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|b_{j,k,\ell}| > \kappa \lambda_n / 2\}}.$$

Let j_2 be the integer defined by

$$2^{j_2} = \left\lceil \left(\frac{n}{(\ln n)^{1+1/\theta}} \right)^{1/(2s+1)} \right\rceil. \quad (7.18)$$

We have

$$V_2 \leq V_{2,1} + V_{2,2},$$

where

$$V_{2,1} = C \frac{(\ln n)^{1+1/\theta}}{n} \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|b_{j,k,\ell}| > \kappa \lambda_n / 2\}}$$

and

$$V_{2,2} = C \frac{(\ln n)^{1+1/\theta}}{n} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|b_{j,k,\ell}| > \kappa \lambda_n / 2\}}.$$

We have

$$V_{2,1} \leq C \frac{(\ln n)^{1+1/\theta}}{n} \sum_{j=\tau}^{j_2} 2^j \leq C \frac{(\ln n)^{1+1/\theta}}{n} 2^{j_2} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}.$$

For $q \geq 1$ and $p \geq 2$, we have $B_{p,q}^s(M) \subseteq B_{2,\infty}^s(M)$. So, by (4.3),

$$\begin{aligned} V_{2,2} &\leq C \frac{(\ln n)^{1+1/\theta}}{n\lambda_n^2} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2 \leq C \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \leq C 2^{-2j_2s} \\ &\leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

For $q \geq 1$, $p \in [1, 2)$ and $s > 1/p$, using (4.3), $\mathbf{1}_{\{|b_{j,k,\ell}| > \kappa\lambda_n/2\}} \leq C|b_{j,k,\ell}|^p/\lambda_n^p = C|\beta_{j,k}|^p/\lambda_n^p$, $B_{p,q}^s(M) \subseteq B_{2,\infty}^{s+1/2-1/p}(M)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} V_{2,2} &\leq C \frac{\ln n}{n\lambda_n^p} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+1/2-1/p)p} \\ &\leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{(2-p)/2} 2^{-j_2(s+1/2-1/p)p} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

So, for $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 1/p\}$, we have

$$V_2 \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \quad (7.19)$$

Bound for V_4 . We have

$$V_4 \leq \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2 \mathbf{1}_{\{|b_{j,k,\ell}| < 2\kappa\lambda_n\}}.$$

Let j_2 be the integer (7.18). Then

$$V_4 \leq V_{4,1} + V_{4,2},$$

where

$$V_{4,1} = \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2 \mathbf{1}_{\{|b_{j,k,\ell}| < 2\kappa\lambda_n\}}, \quad V_{4,2} = \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2 \mathbf{1}_{\{|b_{j,k,\ell}| < 2\kappa\lambda_n\}}.$$

We have

$$\begin{aligned} V_{4,1} &\leq C \sum_{j=\tau}^{j_2} 2^j \lambda_n^2 = C \frac{(\ln n)^{1+1/\theta}}{n} \sum_{j=\tau}^{j_2} 2^j \leq C \frac{(\ln n)^{1+1/\theta}}{n} 2^{j_2} \\ &\leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

For $q \geq 1$ and $p \geq 2$, we have $B_{p,q}^s(M) \subseteq B_{2,\infty}^s(M)$. Hence, by (4.3),

$$V_{4,2} \leq \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \leq C 2^{-2j_2s} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}.$$

For $q \geq 1$, $p \in [1, 2)$ and $s > 1/p$, using (4.3), $b_{j,k,\ell}^2 \mathbf{1}_{\{|b_{j,k,\ell}| < 2\kappa\lambda_n\}} \leq C \lambda_n^{2-p} |b_{j,k,\ell}|^p = C \lambda_n^{2-p} |\beta_{j,k}|^p$, $B_{p,q}^s(M) \subseteq B_{2,\infty}^{s+1/2-1/p}(M)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} V_{4,2} &\leq C \lambda_n^{2-p} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p = C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{(2-p)/2} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \\ &\leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+1/2-1/p)p} \\ &\leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{(2-p)/2} 2^{-j_2(s+1/2-1/p)p} \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

Thus, for $q \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 1/p\}$, we have

$$V_4 \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \quad (7.20)$$

It follows from (7.16), (7.17), (7.19) and (7.20) that

$$V \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}. \quad (7.21)$$

Combining (7.12), (7.13), (7.14), (7.15) and (7.21), we have, for $q \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 1/p\}$,

$$\mathbb{E} \left(\int_0^1 (\widehat{g}_\ell(x) - g_\ell(x))^2 dx \right) \leq C \left(\frac{(\ln n)^{1+1/\theta}}{n} \right)^{2s/(2s+1)}.$$

The proof of Theorem 5.1 is complete. ■

Acknowledgement. This work is supported by ANR grant NatImages, ANR-08-EMER-009.

References

- [1] Amato, U. and Antoniadis, A. (2001). Adaptive wavelet series estimation in separable nonparametric regression models. *Statistics and Computing*, 11, 373-394.
- [2] Amato, U., Antoniadis, A. and De Feis, I. (2002). Fourier series approximation of separable models. *Journal of Computational and Applied Mathematics*, 146, 459-479.
- [3] Bradley, R.C. (2007). *Introduction to strong mixing conditions*. Vol. 1,2,3. Kendrick Press.
- [4] Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics*, 17, 453-555.
- [5] Camlong-Viot, C., Rodriguez-Po, J. M. and Vieu, P. (2006). Nonparametric and semiparametric estimation of additive models with both discrete and continuous variables under dependence. In *The art of semi-parametrics*, Contrib. Statist., pages 155-178. Physica-Verlag/Springer, Heidelberg.
- [6] Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.
- [7] 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.

- [8] Davydov, Y. (1970). The invariance principle for stationary processes. *Theor. Probab. Appl.*, 15, 3, 498-509.
- [9] Debbbarh, M. (2006). Asymptotic normality for the wavelets estimator of the additive regression components. *Comptes Rendus Mathematique*, 343, 9, 1, 601-606.
- [10] Debbbarh, M. and Maillot, B. (2008a). Additive regression model for continuous time processes, *Comm. Statist. Theory Methods*, 37,13-15, 2416-2432.
- [11] Debbbarh, M. and Maillot, B. (2008b). Asymptotic normality of the additive regression components for continuous time processes, *C. R. Math. Acad. Sci. Paris* 346 , 15-16, 901-906.
- [12] Doukhan, P. (1994). Mixing. Properties and Examples. Lecture Notes in Statistics 85. *Springer Verlag*, New York.
- [13] Fan, J. and Jiang, J. (2005). Nonparametric inferences for additive models. *Journal of the Americal Statistical Association*, 100, 890-907.
- [14] Gao, J., Tong, H. and Wolff, R. (2002). Adaptive orthogonal series estimation in additive stochastic regression models. *Statist. Sinica*, 12 (2), 409-428.
- [15] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). Wavelet, Approximation and Statistical Applications. Lectures Notes in Statistics. 129, *Springer Verlag*, New York.
- [16] Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized additive models*. London: Chapman and Hall.
- [17] Liebscher, E. (1996). Strong convergence of sums of α -mixing random variables with applications to density estimation, *Stochastic Processes and their Applications*, 65, 69-80.
- [18] Linton, O.B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, 84, 469-473.

- [19] Mallat, S. (2009). A wavelet tour of signal processing. Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.
- [20] Meyer, Y., *Wavelets and Operators*. Cambridge University Press, Cambridge, 1992.
- [21] Modha, D. and Masry, E. (1996). Minimum complexity regression estimation with weakly dependent observations. *IEEE Trans. Inform. Theory*, 42, 2133-2145.
- [22] Newey, W.K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10 (2), 233-253.
- [23] Opsomer, J.D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, 25, 186-211.
- [24] Opsomer, J.D. and Ruppert, D. (1998). A fully automated bandwidth selection method for fitting additive models. *Journal of the American Statistical Association*, 93, 605-619.
- [25] Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences, *Annals Prob.*, 23, 1188-1203.
- [26] Sardy, S. and Tseng, P. (2004). AMlet, RAMlet, and GAMlet: automatic nonlinear fitting of additive models, robust and generalized, with wavelets. *Journal of Computational and Graphical Statistics*, 13, 283-309.
- [27] Sperlich, S., Tjostheim, D. and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory*, 18, 197-251.
- [28] Stone, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, 13, 689-705.
- [29] Stone, C.J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.*, 14, 590-606.

- [30] Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.*, 22, 118-184.
- [31] Tsybakov, A.B., *Introduction l'estimation non paramtrique*, Springer, 2004.
- [32] Withers, C.S. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57, 477-480.
- [33] Zhang, S. and Wong, M.-Y. (2003). Wavelet threshold estimation for additive regression models. *Annals of Statistics*, 31, 152-173.