Acta Universitatis Apulensis No. 29/2012
ISSN: 1582-5329 pp. 65-84

ADAPTIVE WAVELET REGRESSION IN RANDOM DESIGN AND GENERAL ERRORS WITH WEAKLY DEPENDENT DATA

Christophe Chesneau

Abstract. We investigate the function estimation in a nonparametric regression model having the following particularities: the design is random, the errors admit finite moments of order two and the data are weakly dependent. In this general framework, we construct a new adaptive estimator. It is based on wavelets and the combination of two hard thresholding rules. We determine an upper bound of the associated mean integrated squared error and prove that it is sharp for a wide class of regression functions.

2000 *Mathematics Subject Classification*: 62G07, 62G20.

1. Introduction

Let $(X_i, Y_i)_{i\in\mathbb{Z}}$ be a bivariate stationary random process where, for any $i \in \mathbb{Z}$,

$$Y_i = f(X_i) + \xi_i, \tag{1}$$

$f : [0,1] \to \mathbb{R}$ is the unknown regression function, $(X_i)_{i\in\mathbb{Z}}$ is a sequence of identically distributed random variables having the common known density $g : [0,1] \to [0,\infty)$ and $(\xi_i)_{i\in\mathbb{Z}}$ is a sequence of identically distributed variables independent of $(X_i)_{i\in\mathbb{Z}}$ satisfying $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}(\xi_1^2) < \infty$. Furthermore, we suppose that $(X_i, Y_i)_{i\in\mathbb{Z}}$ is a strongly mixing process (to be defined in Section 2). This dependence structure is reasonably weak and has many practical applications (see e.g. [33], [20] and [28]). Given n observations $(X_1, Y_1), \dots, (X_n, Y_n)$ drawn from $(X_i, Y_i)_{i\in\mathbb{Z}}$, we aim to estimate f globally on $[0,1]$.

To measure the performance of an estimator $\widehat{f}$ of f, we use the Mean Integrated Squared Error (MISE) defined by:

$$\mathbf{R}(\widehat{f}, f) = \mathbb{E}\left(\int_0^1 (\widehat{f}(x) - f(x))^2 dx\right).$$

Our goal is to construct $\widehat{f}$ such that the associated MISE is as small as possible. Many methods can be considered (kernel, spline, wavelets,...) (see e.g. [32]). In this study, we focus our attention on the wavelet methods. They are attractive for nonparametric function estimation because of their virtues from the viewpoints of spatial adaptivity, computational efficiency and asymptotic optimality properties. Further details can be found in [1] and [22].

In the literature, when $(X_1, Y_1), \ldots, (X_n, Y_n)$ are *i.i.d.*, various wavelet methods have been developed. See e.g. [14-16], [17], [21], [2, 3], [4], [31], [5, 6], [7, 8], [10], [29], [13], [34], [23] and [9]. When $\xi_1, \ldots, \xi_n$ have some kind of dependence (long memory, ρ-mixing,...), see e.g. [18], [19], [25], [26] and [24]. To the best of our knowledge, the wavelet estimation of f when $(X_1, Y_1), \ldots, (X_n, Y_n)$ are weakly dependent has only been investigated by [26] and [30]. More precisely, with a non-necessarily bounded Y_1, [26] has constructed a linear non-adaptive wavelet estimator of f which attains a sharp rate of convergence under the uniform risk over Besov balls. Considering a bounded Y_1, [30] have developed a non-linear non-adaptive wavelet estimator of f and studied its asymptotic MISE properties. However, the adaptive wavelet estimation of f, more realistic, has never been addressed earlier and motivates this study. In addition to this new challenge, we relax some classical assumptions on the errors: Y_1 can be non-bounded and the common distribution of $\xi_1, \ldots, \xi_n$ can be unknown; only $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}(\xi_1^2) < \infty$ are required. Furthermore, let us mention that no "Castellana-Leadbetter"-type condition on the density of $((X_0, Y_0), (X_m, Y_m))$, $m \in \mathbb{N}$, is supposed.

We construct a new adaptive wavelet estimator based on the following steps: we estimate the unknown wavelet coefficients of f by a new thresholded versions of the empirical ones, we operate a term-by-term selection of these estimators via a hard thresholding rule, then we reconstruct the selected estimators by taking the initial wavelet basis and choosing appropriate levels. Naturally, the definitions of both thresholds take into account the dependence of the data and are chosen to minimize the associated MISE. Assuming that f belongs to a Besov balls $B^s_{p,q}(H)$ (to be defined in Section 3), we prove that our estimator $\widehat{f}$ satisfies

$$\mathbf{R}(\widehat{f}, f) \leq C \left(\frac{\ln n_\theta}{n_\theta} \right)^{2s/(2s+1)},$$

where $C > 0$ is a constant (independent of n), $n_\theta = n^{\theta/(\theta+1)}$ and θ refers to the exponentially strong mixing case. The obtained rate of convergence is sharp.

The paper is organized as follows. Section 2 clarifies the assumptions on the model and introduces some notations. Section 3 describes the considered wavelet basis and the Besov balls $B^s_{p,q}(H)$. Our wavelet hard thresholding estimator is presented in Section 4. Its asymptotic performances are explored in Section 5.

Section 6 is devoted to the proofs.

2. Assumptions on the model

- Set, for any $i \in \mathbb{Z}$, $Z_i = (X_i, Y_i)$. We suppose that $Z = (Z_i)_{i\in\mathbb{Z}}$ is strictly stationary and exponentially strongly mixing. Let us now clarify this kind of dependence.

  For any $m \in \mathbb{Z}$, we define the m-th strongly mixing coefficient of Z by

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

  where $\mathcal{F}^Z_{-\infty,0}$ is the σ-algebra generated by $\ldots, Z_{-1}, Z_0$ and $\mathcal{F}^Z_{m,\infty}$ is the σ-algebra generated by $Z_m, Z_{m+1}, \ldots$. The bivariate random process Z is said to be strongly mixing if $\lim_{m\to\infty}\alpha_m = 0$.

  The exponentially strongly mixing condition is characterized by the following inequality: there exists three known constants, $\gamma > 0$, $c > 0$ and $\theta > 0$, such that, for any integer $m \geq 1$,

  $$\alpha_m \leq \gamma \exp(-cm^\theta). \tag{2}$$

  This assumption is satisfied by a large class of processes (GARCH, ARMA, ARMA-GARCH, ...). See e.g. [33], [20] and [28].

  Remark that, if $Z = (Z_i)_{i\in\mathbb{Z}}$ is a bivariate sequence of *i.i.d.* random variables, we can take $\theta \to \infty$.

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

  $$\sup_{x\in[0,1]} |f(x)| \leq C_*. \tag{3}$$

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

  $$\inf_{x\in[0,1]} g(x) \geq c_*. \tag{4}$$

3. Wavelet bases and Besov balls

Throughout the paper, we work with a compactly supported wavelet basis on $[0,1]$. A concise mathematical description of this basis is given below. Let N be a positive integer, and ϕ and ψ be the initial wavelet functions of the Daubechies wavelets $db2N$. 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, with an appropriate treatment at the boundaries, there exists a positive integer τ such that the system

$$\mathcal{W} = \{\phi_{\tau,k},\ k \in \{0,\dots,2^\tau - 1\};\ \psi_{j,k};\ j \in \mathbb{N} - \{0,\dots,\tau-1\},\ k \in \{0,\dots,2^j-1\}\},$$

forms an orthonormal basis of $\mathbb{L}^2([0,1]) = \{h : [0,1] \to \mathbb{R};\ \int_0^1 h^2(x)dx < \infty\}$. See [11].

Any $h \in \mathbb{L}^2([0,1])$ can be expanded on $\mathcal{W}$ as

$$h(x) = \sum_{k=0}^{2^\tau-1} c_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x),$$

where $c_{j,k}$ and $d_{j,k}$ are the wavelet coefficients of h defined by

$$c_{j,k} = \int_0^1 h(x)\phi_{j,k}(x)dx, \qquad d_{j,k} = \int_0^1 h(x)\psi_{j,k}(x)dx. \tag{5}$$

As is traditional in the wavelet estimation literature, we shall investigate the performances of our estimator $\widehat{f}$ by assuming that the unknown regression function f belongs to Besov balls. Their definitions are given below.

Let $H > 0$, $s > 0$, $p \geq 1$, $r \geq 1$ and $\mathbb{L}^p([0,1]) = \{h : [0,1] \to \mathbb{R};\ \int_0^1 |h(x)|^p dx < \infty\}$. Set, for every measurable function h on $[0,1]$ and $\epsilon \geq 0$, $\Delta_\epsilon(h)(x) = h(x+\epsilon) - h(x)$, $\Delta_\epsilon^2(h)(x) = \Delta_\epsilon(\Delta_\epsilon h)(x)$ and identically, for $N \in \mathbb{N}^*$, $\Delta_\epsilon^N(h)(x) = \Delta_\epsilon^{N-1}(\Delta_\epsilon h)(x) = \Delta_\epsilon(\Delta_\epsilon^{N-1}h)(x)$. Let

$$\rho^N(t,h,p) = \sup_{\epsilon\in[-t,t]} \left(\int_0^1 |\Delta_\epsilon^N(h)(u)|^p du\right)^{1/p}.$$

Then, for $s > 0$, we define the Besov ball $B^s_{p,r}(H)$ by

$$B^s_{p,r}(H) = \left\{ h \in \mathbb{L}^p([0,1]);\ \left(\int_0^1 \left(\frac{\rho^N(t,h,p)}{t^s}\right)^r \frac{dt}{t}\right)^{1/r} \leq H \right\}.$$

Besov balls can be expressed in terms of wavelet coefficients. We have the following equivalence: $h \in B^s_{p,r}(H)$ with $s \in (0,N)$ if and only if there exists a constant $H^* > 0$ (depending on H) such that the associated wavelet coefficients (5) satisfy

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

In this expression, s is a smoothness parameter and p and r are norm parameters. For a particular choice of s, p and r, the Besov balls contain the standard Hölder and Sobolev balls. See [22, Chapter 9] and [27].

4. Estimators

The first step to estimate f in (1) from $(X_1, Y_1), \ldots, (X_n, Y_n)$ consists in expanding f on the wavelet basis $\mathcal{W}$. Then we aim to estimate the unknown wavelet coefficients: $c_{j,k} = \int_0^1 f(x)\phi_{j,k}(x)dx$ and $d_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$. The considered estimators are described below.

Set

$$\gamma_n = \mu\sqrt{\frac{n_\theta}{\ln n_\theta}},$$

where $n_\theta = n^{\theta/(\theta+1)}$, θ is the one in (2) and $\mu = \sqrt{(C_*^2 + \mathbb{E}(\xi_1^2))/c_*}$.

For any integer $j \geq \tau$ and any $k \in \{0, \ldots, 2^j - 1\}$,

- we estimate $c_{j,k}$ by
$$\widehat{c}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{g(X_i)}\phi_{j,k}(X_i)\mathbb{I}_{\left\{\left|\frac{Y_i}{g(X_i)}\phi_{j,k}(X_i)\right| \leq \gamma_n\right\}}, \tag{6}$$
where, for any random event $\mathcal{A}$, $\mathbb{I}_{\mathcal{A}}$ is the indicator function on $\mathcal{A}$.
- we estimate $d_{j,k}$ by
$$\widehat{d}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{g(X_i)}\psi_{j,k}(X_i)\mathbb{I}_{\left\{\left|\frac{Y_i}{g(X_i)}\psi_{j,k}(X_i)\right| \leq \gamma_n\right\}}. \tag{7}$$

Remark that $\widehat{c}_{j,k}$ and $\widehat{d}_{j,k}$ are thresholded versions of the standard empirical wavelet estimators for (1) (see e.g. [14-16]). Such a thresholding has been introduced by [13] for (1) when $(X_1, Y_1), \ldots, (X_n, Y_n)$ are *i.i.d.*. In our study, it allows us to have non restrictive assumptions on $\xi_1, \ldots, \xi_n$ and treat the weak dependence of $(X_1, Y_1), \ldots, (X_n, Y_n)$.

We define the hard thresholding estimator $\widehat{f}$ by

$$\widehat{f}(x) = \sum_{k=0}^{2^\tau - 1} \widehat{c}_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \widehat{d}_{j,k}\mathbb{I}_{\left\{|\widehat{d}_{j,k}| \geq \kappa\lambda_n\right\}}\psi_{j,k}(x), \tag{8}$$

where $\widehat{c}_{\tau,k}$ is defined by (6) with $j = \tau$, $\widehat{d}_{j,k}$ by (7), j_1 is the integer satisfying

$$\frac{1}{2}n_\theta < 2^{j_1} \leq n_\theta,$$

$\kappa \geq 2 + 16/(3u) + 4\sqrt{(1/u)(16/9u^2 + 2)}$ with $u = (1/2)(c/8)^{1/(\theta+1)}$ and

$$\lambda_n = \mu\sqrt{\frac{\ln n_\theta}{n_\theta}}.$$

The main idea of the hard thresholding rule used in (8) is to estimate only the "large" unknown wavelet coefficients of f (and remove the other). Indeed, they are those which contain its main characteristics. The definition of the threshold λ_n is based on theoretical consideration (see Proposition 3 below).

Let us mention that $\widehat{f}$ is adaptive i.e. its construction does not depend on the smoothness parameter of f. However, it depends on the factor θ related to (2).

More details on hard thresholding estimators in wavelet estimation are given in [22, Chapter 11].

5. Results

5.1. Auxiliary results

Propositions 1 and 2 below show moments properties for (6) and (7).

Proposition 1. *Consider* (1) *under the assumptions of Section* 2. *For any integer* $j \geq \tau$ *and any* $k \in \{0, \ldots, 2^j - 1\}$, *let* $d_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$ *and* $\widehat{d}_{j,k}$ *be* (7). *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^2\right) \leq C\frac{\ln n_\theta}{n_\theta}.$$

This inequality holds with $c_{j,k} = \int_0^1 f(x)\phi_{j,k}(x)dx$ *instead of* $d_{j,k}$ *and* $\widehat{c}_{j,k}$ *defined by* (6) *instead of* $\widehat{d}_{j,k}$.

Proposition 2. *Consider* (1) *under the assumptions of Section* 2. *For any integer* $j \geq \tau$ *and any* $k \in \{0, \ldots, 2^j - 1\}$, *let* $d_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$ *and* $\widehat{d}_{j,k}$ *be* (7). *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^4\right) \leq C.$$

Proposition 3 below determines a sharp concentration inequality for (7).

Proposition 3. *Consider* (1) *under the assumptions of Section* 2. *For any integer* $j \geq \tau$ *and any* $k \in \{0, \ldots, 2^j - 1\}$, *let* $d_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$, $\widehat{d}_{j,k}$ *be* (7) *and*

$$\lambda_n = \mu\sqrt{\frac{\ln n_\theta}{n_\theta}}.$$

Then, for any $\kappa \geq 2 + 16/(3u) + 4\sqrt{(1/u)(16/9u^2 + 2)}$ *with* $u = (1/2)(c/8)^{1/(\theta+1)}$, *we have*

$$\mathbb{P}\left(|\widehat{d}_{j,k} - d_{j,k}| \geq \kappa\lambda_n/2\right) \leq 2(1 + 4e^{-2}\gamma)\frac{1}{n_\theta^4}.$$

5.2. Main result

Theorem 1 below investigates the performance of $\widehat{f}$ under the MISE over Besov balls.

Theorem 1. *Consider* (1) *under the assumptions of Section* 2. *Let* $\widehat{f}$ *be* (8). *Suppose that* $f \in B^s_{p,r}(H)$ *with* $r \geq 1$, *either* $p \geq 2$ *and* $s \in (0, N)$ *or* $p \in [1, 2)$ *and* $s \in (1/p, N)$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbf{R}(\widehat{f}, f) \leq C\left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)},$$

where $n_\theta = n^{\theta/(\theta+1)}$.

The proof of Theorem 1 uses a suitable decomposition of the MISE with the results in Propositions 1, 2 and 3.

If we restrict our study to independent $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.e. $\theta \to \infty$, our rate of convergence becomes $(\ln n/n)^{2s/(2s+1)}$ which is the standard "near optimal" one in the minimax sense. See e.g. [22, Chapter 11] and [13].

Note that the rate of convergence $(\ln n_\theta/n_\theta)^{2s/(2s+1)}$ is also attained by the abstract minimum complexity regression estimator in [28, Theorem 2.1] but for a slightly different regression problem with more restrictions on $(X_1, Y_1), \ldots, (X_n, Y_n)$, $\xi_1, \ldots, \xi_n$ and f (see [28, Section II.B]).

Mention that the obtained rate of convergence can perhaps be improved by considering another thresholding rule as the block thresholding ones (see e.g. [5,6] and [9]) or with more restrictive assumption on the model as "Castellana-Leadbetter"-type condition (see e.g. [26]). Another possible perspective of this work is to consider the case where g is unknown. A pertinent approach could be to use warped wavelets in the construction of our hard thresholding estimator as it is developed in [23] for the *i.i.d.* case. All these aspects need further investigations that we leave for a future work.

6. Proofs

In this section, C represents a positive constant which may differ from one term to another.

6.1. Proof of Proposition 1

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

$$W_{i,j,k} = \frac{Y_i}{g(X_i)} \psi_{j,k}(X_i).$$

Since X_1 and ξ_1 are independent and $\mathbb{E}(\xi_1) = 0$, we have

$$\begin{aligned}
\mathbb{E}(W_{1,j,k}) &= \mathbb{E}\left(\frac{Y_1}{g(X_1)}\psi_{j,k}(X_1)\right) \\
&= \mathbb{E}\left(\frac{f(X_1)}{g(X_1)}\psi_{j,k}(X_1)\right) + \mathbb{E}(\xi_1)\mathbb{E}(\frac{1}{g(X_1)}\psi_{j,k}(X_1)) \\
&= \mathbb{E}\left(\frac{f(X_1)}{g(X_1)}\psi_{j,k}(X_1)\right) = \int_0^1 \frac{f(x)}{g(x)}\psi_{j,k}(x)g(x)dx \\
&= \int_0^1 f(x)\psi_{j,k}(x)dx = d_{j,k}.
\end{aligned}$$

Hence, since $W_{1,j,k}, \ldots, W_{n,j,k}$ are identically distributed,

$$\begin{aligned}
d_{j,k} &= \mathbb{E}\left(W_{1,j,k}\right) = \mathbb{E}\left(W_{1,j,k}\mathbb{1}_{\{|W_{1,j,k}| \le \gamma_n\}}\right) + \mathbb{E}\left(W_{1,j,k}\mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}}\right) \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} W_{i,j,k}\mathbb{1}_{\{|W_{i,j,k}| \le \gamma_n\}}\right) + \mathbb{E}\left(W_{1,j,k}\mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}}\right). \qquad (9)
\end{aligned}$$

This with the elementary inequality $(x+y)^2 \le 2(x^2+y^2)$, $(x,y) \in \mathbb{R}^2$, imply that

$$\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^2\right) \le 2(A+B), \qquad (10)$$

where

$$A = \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left(W_{i,j,k}\mathbb{1}_{\{|W_{i,j,k}| \le \gamma_n\}} - \mathbb{E}(W_{i,j,k}\mathbb{1}_{\{|W_{i,j,k}| \le \gamma_n\}})\right)\right)^2\right)$$

and

$$B = \left(\mathbb{E}\left(|W_{1,j,k}|\mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}}\right)\right)^2.$$

Let us bound B and A (by order of difficulty).

Upper bound for B. Since $\mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}} \le |W_{1,j,k}|/\gamma_n$, we have

$$\mathbb{E}\left(|W_{1,j,k}|\mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}}\right) \le \frac{\mathbb{E}(W_{1,j,k}^2)}{\gamma_n}.$$

Let us now bound $\mathbb{E}(W_{1,j,k}^2)$. Since X_1 and ξ_1 are independent and $\mathbb{E}(\xi_1) = 0$, we have

$$\begin{aligned}
\mathbb{E}(W_{1,j,k}^2) &= \mathbb{E}\left(\frac{Y_1^2}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) = \mathbb{E}\left(\frac{(f(X_1)+\xi_1)^2}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) \\
&= \mathbb{E}\left(\frac{f^2(X_1)}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) + 2\mathbb{E}(\xi_1)\mathbb{E}\left(\frac{f(X_1)}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) \\
&+ \mathbb{E}\left(\xi_1^2\right)\mathbb{E}\left(\frac{1}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) \\
&= \mathbb{E}\left(\frac{f^2(X_1)}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) + \mathbb{E}(\xi_1^2)\mathbb{E}\left(\frac{1}{g^2(X_1)}\psi_{j,k}^2(X_1)\right).
\end{aligned}$$

Using (3), (4) and $\int_0^1 \psi_{j,k}^2(x)dx = 1$, we obtain

$$\begin{aligned}
\mathbb{E}\left(\frac{f^2(X_1)}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) &\leq C_*^2\mathbb{E}\left(\frac{1}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) \\
&= C_*^2\int_0^1 \frac{1}{g^2(x)}\psi_{j,k}^2(x)g(x)dx = C_*^2\int_0^1 \frac{1}{g(x)}\psi_{j,k}^2(x)dx \\
&\leq \frac{C_*^2}{c_*}\int_0^1 \psi_{j,k}^2(x)dx = \frac{C_*^2}{c_*}.
\end{aligned}$$

And, in a similar way,

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{g^2(X_1)}\psi_{j,k}^2(X_1)\right) &= \int_0^1 \frac{1}{g^2(x)}\psi_{j,k}^2(x)g(x)dx = \int_0^1 \frac{1}{g(x)}\psi_{j,k}^2(x)dx \\
&\leq \frac{1}{c_*}\int_0^1 \psi_{j,k}^2(x)dx = \frac{1}{c_*}.
\end{aligned}$$

So

$$\mathbb{E}(W_{1,j,k}^2) \leq \frac{1}{c_*}(C_*^2 + \mathbb{E}(\xi_1^2)) = \mu^2 = C. \tag{11}$$

Therefore

$$B = \left(\mathbb{E}\left(|W_{1,j,k}|\mathbb{1}_{\{|W_{1,j,k}|>\gamma_n\}}\right)\right)^2 \leq \left(\frac{\mathbb{E}(W_{1,j,k}^2)}{\gamma_n}\right)^2 \leq C\frac{1}{\gamma_n^2} \leq C\frac{\ln n_\theta}{n_\theta}. \tag{12}$$

Upper bound for A. We have

$$\begin{aligned} A &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} W_{i,j,k}\mathbb{I}_{\{|W_{i,j,k}|\leq\gamma_n\}}\right) \\ &= \frac{1}{n^2}\sum_{v=1}^{n}\sum_{\ell=1}^{n}\mathbb{C}ov\left(W_{v,j,k}\mathbb{I}_{\{|W_{v,j,k}|\leq\gamma_n\}}, W_{\ell,j,k}\mathbb{I}_{\{|W_{\ell,j,k}|\leq\gamma_n\}}\right) \\ &\leq \frac{1}{n}\mathbb{V}\left(W_{1,j,k}\mathbb{I}_{\{|W_{1,j,k}|\leq\gamma_n\}}\right) \\ &+ \frac{2}{n^2}\left|\sum_{v=2}^{n}\sum_{\ell=1}^{v-1}\mathbb{C}ov\left(W_{v,j,k}\mathbb{I}_{\{|W_{v,j,k}|\leq\gamma_n\}}, W_{\ell,j,k}\mathbb{I}_{\{|W_{\ell,j,k}|\leq\gamma_n\}}\right)\right|. \qquad (13) \end{aligned}$$

Using (11), we obtain

$$\mathbb{V}\left(W_{1,j,k}\mathbb{I}_{\{|W_{1,j,k}|\leq\gamma_n\}}\right) \leq \mathbb{E}\left(W_{1,j,k}^2\mathbb{I}_{\{|W_{1,j,k}|\leq\gamma_n\}}\right) \leq \mathbb{E}(W_{1,j,k}^2) \leq \mu^2 = C. \qquad (14)$$

Let us now bound the covariance term. It follows from the strict stationarity of $(X_i, Y_i)_{i\in\mathbb{Z}}$ that

$$\begin{aligned} &\left|\sum_{v=2}^{n}\sum_{\ell=1}^{v-1}\mathbb{C}ov\left(W_{v,j,k}\mathbb{I}_{\{|W_{v,j,k}|\leq\gamma_n\}}, W_{\ell,j,k}\mathbb{I}_{\{|W_{\ell,j,k}|\leq\gamma_n\}}\right)\right| \\ &= \left|\sum_{m=1}^{n}(n-m)\mathbb{C}ov\left(W_{0,j,k}\mathbb{I}_{\{|W_{0,j,k}|\leq\gamma_n\}}, W_{m,j,k}\mathbb{I}_{\{|W_{m,j,k}|\leq\gamma_n\}}\right)\right| \\ &\leq n\sum_{m=1}^{n}\left|\mathbb{C}ov\left(W_{0,j,k}\mathbb{I}_{\{|W_{0,j,k}|\leq\gamma_n\}}, W_{m,j,k}\mathbb{I}_{\{|W_{m,j,k}|\leq\gamma_n\}}\right)\right|. \end{aligned}$$

By the Davydov inequality for strongly mixing processes (see [12]), the inequality $|W_{0,j,k}|\mathbb{I}_{\{|W_{0,j,k}|\leq\gamma_n\}} \leq \max(\gamma_n, |W_{0,j,k}|)$ and (11), for any $q \in (0,1)$, we have

$$\begin{aligned} &\left|\mathbb{C}ov\left(W_{0,j,k}\mathbb{I}_{\{|W_{0,j,k}|\leq\gamma_n\}}, W_{m,j,k}\mathbb{I}_{\{|W_{m,j,k}|\leq\gamma_n\}}\right)\right| \\ &\leq 10\alpha_m^q\left(\mathbb{E}\left(|W_{0,j,k}|^{2/(1-q)}\mathbb{I}_{\{|W_{0,j,k}|\leq\gamma_n\}}\right)\right)^{1-q} \\ &= 10\alpha_m^q\left(\mathbb{E}\left(|W_{0,j,k}|^{2q/(1-q)}\mathbb{I}_{\{|W_{0,j,k}|\leq\gamma_n\}}W_{0,j,k}^2\right)\right)^{1-q} \\ &\leq C\alpha_m^q\left(\gamma_n^{2q/(1-q)}\right)^{1-q}\left(\mathbb{E}\left(W_{0,j,k}^2\right)\right)^{1-q} \leq C\alpha_m^q\left(\frac{n_\theta}{\ln n_\theta}\right)^q. \end{aligned}$$

Thanks to (2), we have $\sum_{m=1}^{n}\alpha_m^q \leq \gamma^q\sum_{m=1}^{\infty}\exp\left(-cqm^\theta\right) = C$. So

$$\left|\sum_{v=2}^{n}\sum_{\ell=1}^{v-1}\mathbb{C}ov\left(W_{v,j,k}\mathbb{I}_{\{|W_{v,j,k}|\leq\gamma_n\}}, W_{\ell,j,k}\mathbb{I}_{\{|W_{\ell,j,k}|\leq\gamma_n\}}\right)\right| \leq Cn\left(\frac{n_\theta}{\ln n_\theta}\right)^q. \qquad (15)$$

Taking $q = 1/\theta$, we have $n_\theta^q/n = 1/n_\theta$. The inequalities (13), (14) and (15) imply

$$A \leq C\left(\frac{1}{n} + \frac{1}{n^2}n\left(\frac{n_\theta}{\ln n_\theta}\right)^q\right) \leq C\frac{n_\theta^q}{n} = C\frac{1}{n_\theta} \leq C\frac{\ln n_\theta}{n_\theta}. \quad (16)$$

It follows from (10), (12) and (16) that

$$\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^2\right) \leq C\frac{\ln n_\theta}{n_\theta}. \quad (17)$$

Replacing ϕ instead of ψ in the previous proof, one can show that (17) holds with $c_{j,k} = \int_0^1 f(x)\phi_{j,k}(x)dx$ instead of $d_{j,k}$ and $\widehat{c}_{j,k}$ defined by (6) instead of $\widehat{d}_{j,k}$.

The proof of Proposition 1 is complete.

6.2. Proof of Proposition 2

We have

$$|\widehat{d}_{j,k} - d_{j,k}| \leq |\widehat{d}_{j,k}| + |d_{j,k}|.$$

We have

$$|\widehat{d}_{j,k}| \leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i}{g(X_i)}\psi_{j,k}(X_i)\right| \mathbb{1}_{\left\{\left|\frac{Y_i}{g(X_i)}\psi_{j,k}(X_i)\right| \leq \gamma_n\right\}} \leq \gamma_n = \mu\sqrt{\frac{n_\theta}{\ln n_\theta}}.$$

It follows from (3), the Cauchy-Schwarz inequality and $\int_0^1 \psi_{j,k}^2(x)dx = 1$ that

$$\begin{aligned} |d_{j,k}| &\leq \int_0^1 |f(x)||\psi_{j,k}(x)|dx \leq C_* \int_0^1 |\psi_{j,k}(x)|dx \\ &\leq C_*\left(\int_0^1 \psi_{j,k}^2(x)dx\right)^{1/2} = C_* \leq C\sqrt{\frac{n_\theta}{\ln n_\theta}}. \end{aligned}$$

Therefore

$$|\widehat{d}_{j,k} - d_{j,k}| \leq C\sqrt{\frac{n_\theta}{\ln n_\theta}}. \quad (18)$$

By (18) and Proposition 1, we have

$$\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^4\right) \leq C\frac{n_\theta}{\ln n_\theta}\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^2\right) \leq C\frac{n_\theta}{\ln n_\theta}\frac{\ln n_\theta}{n_\theta} = C.$$

Proposition 2 is proved

6.3. Proof of Proposition 3

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

$$W_{i,j,k} = \frac{Y_i}{g(X_i)} \psi_{j,k}(X_i).$$

By (9), we have

$$\begin{aligned} |\widehat{d}_{j,k} - d_{j,k}| \\ &= \left| \frac{1}{n} \sum_{i=1}^{n} \left(W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \leq \gamma_n\}} - \mathbb{E}\left(W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \leq \gamma_n\}} \right) \right) - \mathbb{E}\left(W_{1,j,k} \mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}} \right) \right| \\ &\leq \left| \frac{1}{n} \sum_{i=1}^{n} \left(W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \leq \gamma_n\}} - \mathbb{E}\left(W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \leq \gamma_n\}} \right) \right) \right| + \mathbb{E}\left(|W_{1,j,k}| \mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}} \right). \end{aligned}$$

Using (11), we obtain

$$\mathbb{E}\left(|W_{1,j,k}| \mathbb{1}_{\{|W_{1,j,k}| > \gamma_n\}} \right) \leq \frac{\mathbb{E}(W_{1,j,k}^2)}{\gamma_n} \leq \mu^2 \frac{1}{\mu} \sqrt{\frac{\ln n_\theta}{n_\theta}} = \lambda_n.$$

Hence

$$\begin{aligned} \mathbb{P}\left(|\widehat{d}_{j,k} - d_{j,k}| \geq \kappa \lambda_n / 2 \right) \\ \leq \quad \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \left(W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \leq \gamma_n\}} - \mathbb{E}\left(W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \leq \gamma_n\}} \right) \right) \right| \geq (\kappa/2 - 1)\lambda_n \right). \end{aligned} \tag{19}$$

Let us now present a Bernstein inequality for exponentially strongly mixing process. This is a slightly modified version of [28, Theorem 4.2].

Lemma 1. **[28]** *Let $\gamma > 0$, $c > 0$, $\theta > 1$ and $(S_i)_{i \in \mathbb{Z}}$ be a stationary process such that, for any integer $m \geq 1$, the associated m-th strongly mixing coefficient satisfies*

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

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

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

where $u = (1/2)(c/8)^{1/(\theta+1)}$.

Set, for any $i \in \{1, \dots, n\}$,

$$U_{i,j,k} = W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \le \gamma_n\}} - \mathbb{E}\left(W_{i,j,k} \mathbb{1}_{\{|W_{i,j,k}| \le \gamma_n\}}\right).$$

Then $U_{1,j,k}, \dots, U_{n,j,k}$ are identically distributed, depend on the stationary strongly mixing process $(X_i, Y_i)_{i \in \mathbb{Z}}$ which satisfies (2), $\mathbb{E}(U_{1,j,k}) = 0$,

$$|U_{1,j,k}| \le |W_{1,j,k}| \mathbb{1}_{\{|W_{1,j,k}| \le \gamma_n\}} + \mathbb{E}\left(|W_{1,j,k}| \mathbb{1}_{\{|W_{1,j,k}| \le \gamma_n\}}\right) \le 2\gamma_n$$

and, by (11),

$$\mathbb{E}(U_{1,j,k}^2) = \mathbb{V}\left(W_{1,j,k} \mathbb{1}_{\{|W_{1,j,k}| \le \gamma_n\}}\right) \le \mathbb{E}(W_{1,j,k}^2) \le \mu^2.$$

It follows from Lemma 1 that

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_{i,j,k}\right| \ge (\kappa/2 - 1)\lambda_n\right) \\ &\le \quad 2(1 + 4e^{-2}\gamma) \exp\left(-\frac{u(\kappa/2-1)^2 \lambda_n^2 n_\theta}{2\left(\mu^2 + 2(\kappa/2-1)\lambda_n \gamma_n/3\right)}\right). \end{aligned} \tag{20}$$

We have

$$\lambda_n \gamma_n = \mu \sqrt{\frac{\ln n_\theta}{n_\theta}} \mu \sqrt{\frac{n_\theta}{\ln n_\theta}} = \mu^2, \qquad \lambda_n^2 = \mu^2 \frac{\ln n_\theta}{n_\theta}.$$

Combining (19) and (20), for any $\kappa \ge 2 + 16/(3u) + 4\sqrt{(1/u)(16/9u^2 + 2)}$, we have

$$\begin{aligned} &\mathbb{P}\left(|\widehat{d}_{j,k} - d_{j,k}| \ge \kappa \lambda_n / 2\right) \\ &\le \quad 2(1 + 4e^{-2}\gamma) \exp\left(-\frac{u(\kappa/2-1)^2 \ln n_\theta}{2\left(1 + 2(\kappa/2-1)/3\right)}\right) \\ &= \quad 2(1 + 4e^{-2}\gamma) n_\theta^{-\frac{u(\kappa/2-1)^2}{2(1+2(\kappa/2-1)/3)}} \le 2(1 + 4e^{-2}\gamma) \frac{1}{n_\theta^4}. \end{aligned}$$

This ends the proof of Proposition 3.

6.4. Proof of Theorem 1

We expand the function f on $\mathcal{W}$ as

$$f(x) = \sum_{k=0}^{2^\tau - 1} c_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j - 1} d_{j,k} \psi_{j,k}(x),$$

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

Since $\mathcal{W}$ is an orthonormal basis of $\mathbb{L}^2([0,1])$, we can write

$$\mathbf{R}(\widehat{f}, f) = \mathbb{E}\left(\int_0^1 (\widehat{f}(x) - f(x))^2 dx\right) = U + V + W, \tag{21}$$

where

$$U = \sum_{k=0}^{2^\tau - 1} \mathbb{E}\left((\widehat{c}_{\tau,k} - c_{\tau,k})^2\right), \qquad V = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j - 1} \mathbb{E}\left((\widehat{d}_{j,k} \mathbb{I}_{\left\{|\widehat{d}_{j,k}| \geq \kappa\lambda_n\right\}} - d_{j,k})^2\right)$$

and

$$W = \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j - 1} d_{j,k}^2.$$

Let us now bound U, W and V.

Upper bound for U. Using Proposition 1 and $2s/(2s+1) < 1$, we obtain

$$U \leq C 2^\tau \frac{\ln n_\theta}{n_\theta} \leq C \left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)}. \tag{22}$$

Upper bound for W. For $r \geq 1$ and $p \geq 2$, we have $B^s_{p,r}(H) \subseteq B^s_{2,\infty}(H)$. Since $2s/(2s+1) < 2s$, we have

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

For $r \geq 1$ and $p \in [1,2)$, we have $B^s_{p,r}(H) \subseteq B^{s+1/2-1/p}_{2,\infty}(H)$. Since $s \in (1/p, N)$, we have $s + 1/2 - 1/p > s/(2s+1)$. So

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

Hence, for $r \geq 1$, either $p \geq 2$ and $s \in (0, N)$ or $p \in [1,2)$ and $s \in (1/p, N)$, we have

$$W \leq C \left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)}. \tag{23}$$

Upper bound for V. We have

$$V = V_1 + V_2 + V_3 + V_4, \qquad (24)$$

where

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

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

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

and

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

Let us now investigate the bounds of V_1, V_2, V_3 and V_4.
Upper bounds for $V_1 + V_3$. Remark that:
$\left\{|\widehat{d}_{j,k}| < \kappa\lambda_n,\ |d_{j,k}| \geq 2\kappa\lambda_n\right\} \subseteq \left\{|\widehat{d}_{j,k} - d_{j,k}| > \kappa\lambda_n/2\right\}$, $\left\{|\widehat{d}_{j,k}| \geq \kappa\lambda_n,\ |d_{j,k}| < \kappa\lambda_n/2\right\} \subseteq \left\{|\widehat{d}_{j,k} - d_{j,k}| > \kappa\lambda_n/2\right\}$ and $\left\{|\widehat{d}_{j,k}| < \kappa\lambda_n,\ |d_{j,k}| \geq 2\kappa\lambda_n\right\} \subseteq \left\{|d_{j,k}| \leq 2|\widehat{d}_{j,k} - d_{j,k}|\right\}$.

So

$$V_1 + V_3 \leq C \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^2 \mathbb{I}_{\left\{|\widehat{d}_{j,k} - d_{j,k}| > \kappa\lambda_n/2\right\}} \right).$$

It follows from the Cauchy-Schwarz inequality, Propositions 2 and 3 that

$$\begin{aligned} &\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^2 \mathbb{I}_{\left\{|\widehat{d}_{j,k} - d_{j,k}| > \kappa\lambda_n/2\right\}} \right) \\ \leq\ & \left(\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^4 \right) \right)^{1/2} \left(\mathbb{P}\left(|\widehat{d}_{j,k} - d_{j,k}| > \kappa\lambda_n/2 \right) \right)^{1/2} \\ \leq\ & C \left(\frac{1}{n_\theta^4} \right)^{1/2} = C \frac{1}{n_\theta^2}. \end{aligned}$$

Since $2s/(2s+1) < 1$, we have

$$V_1 + V_3 \leq C \frac{1}{n_\theta^2} \sum_{j=\tau}^{j_1} 2^j \leq C \frac{1}{n_\theta^2} 2^{j_1} \leq C \frac{1}{n_\theta} \leq C \left(\frac{\ln n_\theta}{n_\theta} \right)^{2s/(2s+1)}. \qquad (25)$$

Upper bound for V_2. Using again Proposition 1, we obtain

$$\mathbb{E}\left((\widehat{d}_{j,k} - d_{j,k})^2\right) \leq C\frac{\ln n_\theta}{n_\theta}.$$

Hence

$$V_2 \leq C\frac{\ln n_\theta}{n_\theta}\sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1}\mathbb{1}_{\left\{|d_{j,k}|>\kappa\lambda_n/2\right\}}.$$

Let j_2 be the integer defined by

$$\frac{1}{2}\left(\frac{n_\theta}{\ln n_\theta}\right)^{1/(2s+1)} < 2^{j_2} \leq \left(\frac{n_\theta}{\ln n_\theta}\right)^{1/(2s+1)}. \tag{26}$$

We have

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

where

$$V_{2,1} = C\frac{\ln n_\theta}{n_\theta}\sum_{j=\tau}^{j_2}\sum_{k=0}^{2^j-1}\mathbb{1}_{\left\{|d_{j,k}|>\kappa\lambda_n/2\right\}}$$

and

$$V_{2,2} = C\frac{\ln n_\theta}{n_\theta}\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}\mathbb{1}_{\left\{|d_{j,k}|>\kappa\lambda_n/2\right\}}.$$

We have

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

For $r \geq 1$ and $p \geq 2$, since $B^s_{p,r}(H) \subseteq B^s_{2,\infty}(H)$, we have

$$\begin{aligned} V_{2,2} &\leq C\frac{\ln n_\theta}{n_\theta\lambda_n^2}\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}d^2_{j,k} \leq C\sum_{j=j_2+1}^{\infty}\sum_{k=0}^{2^j-1}d^2_{j,k} \leq C2^{-2j_2s} \\ &\leq C\left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)}. \end{aligned}$$

For $r \geq 1$, $p \in [1,2)$ and $s \in (1/p, N)$, using $\mathbb{I}_{\left\{|d_{j,k}| > \kappa\lambda_n/2\right\}} \leq C|d_{j,k}|^p/\lambda_n^p$, $B^s_{p,r}(H) \subseteq B^{s+1/2-1/p}_{2,\infty}(H)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

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

So, for $r \geq 1$, either $p \geq 2$ and $s \in (0, N)$ or $p \in [1,2)$ and $s \in (1/p, N)$, we have

$$V_2 \leq C\left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)}. \tag{27}$$

Upper bound for V_4. We have

$$V_4 \leq \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} d_{j,k}^2 \mathbb{I}_{\left\{|d_{j,k}| < 2\kappa\lambda_n\right\}}.$$

Let j_2 be the integer (26). Then

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

where

$$V_{4,1} = \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} d_{j,k}^2 \mathbb{I}_{\left\{|d_{j,k}| < 2\kappa\lambda_n\right\}}, \qquad V_{4,2} = \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} d_{j,k}^2 \mathbb{I}_{\left\{|d_{j,k}| < 2\kappa\lambda_n\right\}}.$$

We have

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

For $r \geq 1$ and $p \geq 2$, since $B^s_{p,r}(H) \subseteq B^s_{2,\infty}(H)$, we have

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

For $r \geq 1$, $p \in [1,2)$ and $s \in (1/p, N)$, using $d_{j,k}^2 \mathbb{1}_{\{|d_{j,k}|<2\kappa\lambda_n\}} \leq C\lambda_n^{2-p}|d_{j,k}|^p$, $B_{p,r}^s(H) \subseteq B_{2,\infty}^{s+1/2-1/p}(H)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

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

So, for $r \geq 1$, either $p \geq 2$ and $s \in (0, N)$ or $p \in [1,2)$ and $s \in (1/p, N)$, we have

$$V_4 \leq C\left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)}. \tag{28}$$

It follows from (24), (25), (27) and (28) that

$$V \leq C\left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)}. \tag{29}$$

Combining (21), (22), (23) and (29), we have, for $r \geq 1$, either $p \geq 2$ and $s \in (0, N)$ or $p \in [1,2)$ and $s \in (1/p, N)$,

$$\mathbf{R}(\widehat{f}, f) \leq C\left(\frac{\ln n_\theta}{n_\theta}\right)^{2s/(2s+1)}.$$

The proof of Theorem 1 is complete.

References

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

[2] A. Antoniadis, J. Bigot and T. Sapatinas, *Wavelet estimators in nonparametric regression: a comparative simulation study*, Journal of Statistical Software, 6 (2001), i06.

[3] A. Antoniadis, D. Leporini and J.-C. Pequet, *Wavelet thresholding for some classes of non-Gaussian noise*, Statistica Neerlandica, 56 **4** (2002), 434-453.

[4] R. Averkamp and C. Houdré, *Wavelet thresholding for non-necessarily Gaussian noise: functionality*, The Annals of Statistics, 33 (2005), 2164-2193.

[5] T. Cai, *Adaptive wavelet estimation: a block thresholding and oracle inequality approach*, The Annals of Statistics, 27 (1999), 898-924.

[6] T. Cai, *On block thresholding in wavelet regression: adaptivity, block size and threshold level*, Statistica Sinica, 12 (2002), 1241-1273.

[7] T. Cai and L. D. Brown, *Wavelet shrinkage for nonequispaced samples*, The Annals of Statistics, 26 (1998), 1783-1799.

[8] T. Cai and L. D. Brown, *Wavelet estimation for samples with random uniform design*, Statistics and Probability Letters, 42 (1999), 313-321.

[9] C. Chesneau, *Wavelet block thresholding for samples with random design: a minimax approach under the Lp risk*, Electronic Journal of Statistics, 1 (2007), 331-346.

[10] E. Chicken, *Block thresholding and wavelet estimation for nonequispaced samples*, Journal of Statistical Planning and Inference, 116 (2003), 113-129.

[11] A. Cohen, I. Daubechies, B. Jawerth and P. Vial, *Wavelets on the interval and fast wavelet transforms*, Applied and Computational Harmonic Analysis, 24 **1** (1993), 54-81.

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

[13] B. Delyon and A. Juditsky, *On minimax wavelet estimators*, Applied Computational Harmonic Analysis, 3 (1996), 215-228.

[14] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), 425-455.

[15] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, Journal of the American Statistical Association, 90 432 (1995), 1200- 1224.

[16] D. L. Donoho and I. M. Johnstone, *Minimax estimation via wavelet shrinkage*, The Annals of Statistics, 26 (1998), 879-921.

[17] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard, *Wavelet shrinkage: asymptopia ? (with discussion)*, Journal of the Royal Statistical Society Serie B, 57 (1995), 301-369.

[18] H. Doosti, M. Afshari and H.A. Niroumand, *Wavelets for nonparametric stochastic regression with mixing stochastic process*, Communication in Statistics: Theory and Methods, 37 **3** (2008), 373-385.

[19] H. Doosti and H.A. Niroumand, *Multivariate Stochastic Regression Estimation by Wavelets for Stationary Time Series*, Pakistan journal of Statistics, 27 **1** (2009), 37-46.

[20] P. Doukhan, *Mixing. Properties and Examples*, Lecture Notes in Statistics 85. Springer Verlag, New York, 1994.

[21] P. Hall and B. A. Turlach, *Interpolation methods for nonlinear wavelet regres-*

sion with irregularly spaced design, The Annals of Statistics, 25 (1997), 1912-1925.

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

[23] G. Kerkyacharian and D. Picard, *Regression in random design and warped wavelets*, Bernoulli, 10 **6** (2004), 1053-1105.

[24] R. Kulik and M. Raimondo, *Lp wavelet regression with correlated errors and inverse problems*, Statistica Sinica, 19 (2009), 1479-1489.

[25] L. Li, J. Liu and Y. Xiao, *On wavelet regression with long memory infinite moving average errors*, Journal of Applied Probability and Statistics, 4 (2009), 183-211.

[26] E. Masry, *Wavelet-Based estimation of multivariate regression functions in besov spaces*, Journal of Nonparametric Statistics, 12 **2** (2000), 283-308.

[27] Y. Meyer, *Ondelettes et Opérateurs*, Hermann, Paris, 1990.

[28] D. Modha and E. Masry, *Minimum complexity regression estimation with weakly dependent observations*, IEEE Trans. Inform. Theory, 42 (1996), 2133-2145.

[29] M. Neumann and V. Spokoiny, *On the efficiency of wavelet estimators under arbitrary error distributions*, Mathematical Methods of Statistics, 4 (1995), 137-166.

[30] P.N. Patil and Y.K. Truong, *Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series*, Annals of the Institute of Statistical Mathematics, 53 **1** (2001), 159-178.

[31] M. Pensky and T. Sapatinas, *Frequentist optimality of Bayes factor estimators in wavelet regression models*, Statistica Sinica, 17 (2007), 599-633.

[32] A. Tsybakov, *Introduction à l'estimation nonparamétrique*, Springer Verlag, Berlin, 2004.

[33] C.S. Withers, *Conditions for linear processes to be strong-mixing*, Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57 (1981), 477-480.

[34] S. Zhang and Z. Zheng, *Nonlinear wavelet estimation of regression function with random design*, Science in China Series A: Mathematics, 42 **8** (1999), 825-833.

Christophe Chesneau
Department of Mathematics, LMNO
University of Caen
UFR de Sciences, F-14032, Caen
email: *chesneau@math.unicaen.fr*