REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000

# WAVELET ESTIMATION OF REGRESSION DERIVATIVES FOR BIASED AND NEGATIVELY ASSOCIATED DATA

Authors: Junke Kou
– School of Mathematics and Computational Science,
Guilin University of Electronic Technology,
Guilin, P.R. China (kjkou@guet.edu.cn)

Christophe Chesneau
– Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Normandie,
Caen, France (christophe.chesneau@gmail.com)

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- This paper considers the estimation of the derivatives of a regression function based on biased data. The main feature of the study is to explore the case where the data comes from a negatively associated process. In this context, two different wavelet estimators are introduced: a linear wavelet estimator and a nonlinear wavelet estimator using the hard thresholding rule. Their theoretical performance is evaluated by determining sharp rates of convergence under $L^p$ risk, assuming that the unknown function of interest belongs to a ball of Besov spaces $B^s_{p,q}(\mathbb{R})$. The obtained results extend some existing works on biased data in the independent case to the negatively associated case.

Key-Words:

- *Regression derivatives estimation; negatively associated; $L^p$ risk; wavelets.*

AMS Subject Classification:

- 62G07, 62G20, 42C40.

## 1. INTRODUCTION

In this paper, the biased nonparametric regression model is considered. It is formulated as follows. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the common density function

$$f(x, y) = \frac{\omega(x, y)g(x, y)}{\mu}, \ (x, y) \in [0, 1] \times \mathbb{R}, \tag{1.1}$$

where $\omega$ stands for a known positive function, $g$ denotes the density function of the unobserved random variables $(U, V)$ and $\mu := \mathbb{E}\omega(X, Y) < \infty$. In this setup $g$ and $f$ mean the target density and weighted density, respectively, and the resulting data are biased data. We want to estimate the $d$th derivative $r^{(d)}(x)$ of regression function

$$r(x) := \mathbb{E}(\rho(V)|U = x) = \int_{\mathbb{R}} \frac{\rho(y)g(x, y)}{h(x)} dy, \ x \in [0, 1]. \tag{1.2}$$

This above model arises in many applications. For example, in order to estimate the change rate of agricultural output $V$ when the input $U$ increase (decrease) in a country. We obtain data $(X_i, Y_i)(i = 1, 2, \ldots, n)$ from those regions where spend more in agriculture, then $X_i$ and $Y_i$ stands for the agricultural input and output. Because it is more likely to sample those special regions, the density $f$ of $(X_i, Y_i)$ satisfies $f(x, y) = \frac{\omega(x,y)g(x,y)}{\mu}$ with some weight function $\omega$ and the real density $g$ of $(U, V)$. Then we can estimate the change rate $r^{(d)}$ of the country by the given data $(X_i, Y_i)$. Hence, the work about this regression estimation model is very important.

The former works have developed kernel or modified local polynomials estimators for the problem of estimating $r(x)$, i.e., $r^{(d)}(x)$ with $d = 0$. See, for instance, [1], [20], [10], [21], [11], [12] and [5]. In order to obtain theoretical results, as optimal rates of convergence, in a general statistical setting or to reach the goal of adaptivity, wavelet methods have been developed by [9], [4] and [6]. Always focusing on wavelet methods, the estimation of $r(x)$ for (strongly mixing) dependent $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ has been explored by [7], [8] and [17]. Also, for the prime goal, the estimation of the derivative $r^{(d)}(x)$ has been considered by [3] and [14], but only for independent $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. More precisely, [3] provide an upper bound estimation over $L^p(\mathbb{R})(1 \leq p < \infty)$ risk for the derivative $r^{(d)}(x)$ of regression function with a linear wavelet estimator. Because this linear wavelet estimator is not adaptive, [14] construct a nonlinear wavelet estimator and study its convergence rate over $L^p(\mathbb{R})(1 \leq p < \infty)$ risk.

In this paper, we investigate a generalization of these works by considering the estimation of $r^{(d)}(x)$ from dependent $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$; the negatively associated case is considered. This kind of dependence naturally appear in many well-known multivariate distributions involved in a wide variety of applications. We refer to [2] and [16]. In this setting, a linear nonadaptive and nonlinear adaptive wavelet estimators are introduced. We determine their rates of convergence under the $L^p$ risk with $1 \leq p < \infty$, assuming that $r^{(d)}(x)$ belongs to Besov spaces $B^s_{\tilde{p},q}(\mathbb{R})$. We prove that, with mathematical efforts, the established results in the independent case can be transposed to the negatively associated case, showing the consistency of the wavelet methodology for this problem.

The rest of this paper is the following. The mathematical assumptions on the model are presented in Section 2. The necessary on the wavelets and Besov spaces are described in Section 3. The linear wavelet estimation is performed in Section 4. The nonlinear wavelet estimation is developed in Section 5. Some concluding remarks are postponed in Section 6.

## 2. ASUMPTIONS ON THE MODEL

In this section, we will introduce the definition and properties of negatively associated sample. In addition, some other assumptions for the model (1.1)-(1.2) are proposed.

**Definition 2.1.** [2] A sequence of random variable $X_1, X_2, \ldots,$ $X_n$ is said to be negatively associated, if for each pair of disjoint nonempty subsets $A$ and $B$ of $\{i = 1, 2, \ldots, n\}$,

$$\mathrm{Cov}\big(f(X_i, i \in A), g(X_j, j \in B)\big) \leq 0,$$

where $f$ and $g$ are real-valued coordinate-wise nondecreasing functions and the corresponding covariances exist.

This definition can be extended to random vectors (see [16]). It is well known that $\mathrm{Cov}\big(X_i, X_j\big) \equiv 0$ when the random variable $X_1, X_2, \ldots, X_n$ is independent. Hence, the independence case is a special case of negatively associated case. Also, let $X_1, X_2, \ldots, X_n$ be independent random variables with log concave densities. Then, if $\sum\limits_{i=1}^{n} X_i = c$ ($c$ is a constant), $X_1, X_2, \ldots, X_n$ are negatively associated.

For examples of negatively associated case, [16] showed that many well-known multivariate distributions process the negatively associated property. Some examples include: the multinomial distribution, the multivariate hypergeometric distribution, the Dirichlet compound multinomial distribution, the permutation distribution and so on. Because of its wide application in multivariate statistical analysis and system reliability, many research of negatively associated has already considered, see, e.g., [19], [24], [18], [23]. In addition, an important property of negative association is given in the following lemma. It will be at a center of one of our main result.

**Lemma 2.1.** *[16] Let $X_1, X_2, \ldots, X_n$ be a sequence of negatively associated random variables and $B_1, B_2, \ldots, B_m$ be some pairwise disjoint nonempty subsets of $\{i = 1, 2, \ldots, n\}$. If $f_i$ $(i = 1, 2, \ldots, m)$ are $m$ coordinate-wise nondecreasing (nonincreasing) functions, then $f_1\big(X_i,\ i \in B_1\big), f_2\big(X_i,\ i \in B_2\big), \ldots, f_m\big(X_i,\ i \in B_m\big)$ are also negatively associated.*

In this paper, $A \lesssim B$ denotes $A \leq cB$ with a positive constant $c$ which is independent of $A$ and $B$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $B \lesssim A$.

For the problem (1.1)-(1.2), in addition to assume that $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are negatively associated, we make the following other assumptions:

**A1**. The density function $h$ of the random variable $U$ is nonincreasing, and has a positive lower bound,

$$0 < c_1 \leq h(x), \quad x \in [0, 1].$$

**A2**. The weight function $\omega$ is coordinate-wise nonincreasing, and has both positive upper and lower bounds, i.e., for $(x, y) \in [0, 1] \times \mathbb{R}$,

$$\omega(x, y) \sim 1.$$

**A3**. The function $\rho$ is known, nondecreasing and $\rho \in L^\infty(\mathbb{R})$.

**A4**. We have $r^{(u)}(0) = r^{(u)}(1) = 0$ for any $u \in \{0, \ldots, d\}$.

**A5**. There exists a constant $c_2 > 0$ such that

$$\sup_{x \in [0,1]} |r^{(d)}(x)| \leq c_2.$$

These assumptions are quite standard for the considered problem (see [3] and [14]). Only those involving the non monotonicity of some functions are deeply link with the negatively associated dependence assumption. They will be used for technical purpose in the proofs.

## 3. WAVELETS AND BESOV SPACES

Throughout this paper, we work with the wavelet basis described below. A wavelet function $\psi$ can be constructed from the scaling function $\phi$ in a simple way such that $\{2^{j/2}\psi(2^j x - k), j \in \mathbb{Z}, k \in \mathbb{Z}\}$ constitutes an orthonormal basis (wavelet basis) of $L^2(\mathbb{R})$. Then, each $f \in L^2(\mathbb{R})$,

$$f = \sum_{k \in \mathbb{Z}} \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k}\psi_{j,k}$$

holds in $L^2(\mathbb{R})$ sense, where $\alpha_{j_0,k} = \langle f, \phi_{j_0,k}\rangle$, $\beta_{j,k} = \langle f, \psi_{j,k}\rangle$ and

$$\phi_{j_0,k}(x) = 2^{\frac{j_0}{2}}\phi(2^{j_0}x - k), \quad \psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k).$$

Let $P_j$ be the orthogonal projection operator from $L^2(\mathbb{R})$ onto the space $V_j$ with the orthonormal basis $\{\phi_{j,k}(\cdot) = 2^{j/2}\phi(2^j\cdot - k), k \in \mathbb{Z}\}$. Then, for $f \in L^2(\mathbb{R})$,

$$P_j f = \sum_{k\in\mathbb{Z}} \alpha_{j,k}\phi_{j,k}.$$

A scaling function $\phi$ is called $m$ regular, if $\phi \in C^m(\mathbb{R})$ and $|D^\alpha \phi(x)| \leq c(1+x^2)^{-l}$ for each $l \in \mathbb{Z}$ ($\alpha = 0, 1, \ldots, m$). In this paper, we choose Daubechies scaling function $D_{2N}$. Then, $\phi$ is $m$ regular when $N$ gets large enough. Furthermore, it can be shown that for $f \in L^p(\mathbb{R})$ $(1 \leq p < \infty)$,

$$P_j f(x) = \sum_{k\in\mathbb{Z}} \alpha_{j,k}\phi_{j,k}(x) \tag{3.1}$$

holds almost everywhere on $\mathbb{R}$ ([15]).

**Lemma 3.1.** *Let a scaling function $\phi \in L^2(\mathbb{R})$ satisfy $m$ regular and $\{\alpha_k\} \in l_p$ $(1 \leq p \leq \infty)$. Then*

$$\left\|\sum_{k\in\mathbb{Z}} \alpha_k 2^{\frac{j}{2}}\phi(2^j x - k)\right\|_p \sim 2^{j(\frac{1}{2}-\frac{1}{p})}\|(\alpha_k)\|_p.$$

The proof of lemma can be found in [15]. In addition, Lemma 3.1 holds if the scaling function $\phi$ is replaced by the corresponding wavelet $\psi$.

One advantage of wavelets is that it can characterize Besov spaces. Besov spaces are important in theory and applications, which contain Hölder and $L^2$ Sobolev spaces as special examples. The next lemma provides equivalent definition for Besov space.

**Lemma 3.2.** *Let $\phi$ be $m$ regular, $\psi$ be the corresponding wavelets and $f \in L^p(\mathbb{R})$. If $\alpha_{j,k} = \langle f, \phi_{j,k}\rangle$, $\beta_{j,k} = \langle f, \psi_{j,k}\rangle$ , $p, q \in [1, \infty]$ and $0 < s < m$, then the following assertions are equivalent:*

*(1)* $f \in B^s_{p,q}(\mathbb{R})$;

*(2)* $\{2^{js}\|P_j f - f\|_p\} \in l_q$;

*(3)* $\left\{2^{j(s-\frac{1}{p}+\frac{1}{2})}\|\beta_j\|_p\right\} \in l_q$.

*The Besov norm of $f$ can be defined by*

$$\|f\|_{B^s_{p,q}} := \|(\alpha_{j_0})\|_p + \|(2^{j(s-\frac{1}{p}+\frac{1}{2})}\|\beta_j\|_p)_{j\geq j_0}\|_q, \tag{3.2}$$

*where* $\|\beta_j\|_p^p = \sum\limits_{k\in\mathbb{Z}} |\beta_{j,k}|^p$.

In this paper, we will suppose the unknown function $r^{(d)}(x)$ belong to Besov balls $B^s_{p,q}(H)$ with $H > 0$, which means $f \in B^s_{p,q}(H) := \{f \in B^s_{p,q}(\mathbb{R}^d),\ \|f\|_{B^s_{p,q}} \leq H\}$.

## 4. LINEAR WAVELET ESTIMATION

This section will introduce a linear wavelet estimator and discuss its convergence rate over $L^p (1 \leq p < \infty)$ risk. Now our linear wavelet estimator is defined by

$$\widehat{r}^{(d)}_n(x) := \sum_{k\in\Omega} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(x). \tag{4.1}$$

In this definition, we have set

$$\widehat{\alpha}_{j_0,k} = (-1)^d \frac{\widehat{\mu}_n}{n} \sum_{i=1}^{n} \frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)} \phi^{(d)}_{j_0,k}(X_i), \tag{4.2}$$

$$\widehat{\mu}_n = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega(X_i,Y_i)}\right]^{-1} \tag{4.3}$$

and $\Omega = \{k \in \mathbb{Z}, \text{supp } r^{(d)} \cap \text{supp } \phi_{j_0,k} \neq \varnothing\}$. Then, it follows from the compactly supported properties of the function $r^{(d)}$ and $\phi_{j_0,k}$ that the cardinality of $\Omega$ satisfies $|\Omega| \sim 2^{j_0}$.

On the other hand, some existing results on these estimators in the independent case remain true. Indeed, according to the [14, Lemma 2.1], under Condition A4, we know that

$$\mathbb{E}\left(\frac{1}{\widehat{\mu}_n}\right) = \frac{1}{\mu} \tag{4.4}$$

and

$$\mathbb{E}\left[(-1)^d \frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)} \phi^{(d)}_{j_0,k}(X_i)\right] = \alpha_{j_0,k}. \tag{4.5}$$

These two equations mean that $\widehat{\mu}_n$ and $\widehat{\alpha}_{j_0,k}$ are unbiased estimators of $\mu$ and $\alpha_{j_0,k}$, respectively. Furthermore, the linear estimator $\widehat{r}^{(d)}_n(x)$ can also be as an unbiased estimator of $r^{(d)}(x)$. In the following, we present an important lemma, which will be used to prove our theorems.

**Lemma 4.1.** *For the problem (1.1)-(1.2) with Conditions A1—A5 hold. If $2^{j_0} \leq n$, then, for $1 \leq p < \infty$, we have*

$$\mathbb{E}\left|\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}\right|^p \lesssim 2^{j_0 dp} n^{-\frac{p}{2}}.$$

**Proof of Lemma 4.1:** According to the definition of $\widehat{\alpha}_{j_0,k}$, the following decomposition holds:

$$\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k}=\frac{\widehat{\mu}_n}{\mu}\left[(-1)^d\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\phi_{j_0,k}^{(d)}(X_i)-\alpha_{j_0,k}\right]+\alpha_{j_0,k}\cdot\widehat{\mu}_n\Big(\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\Big).$$

Furthermore, one has

$$\begin{aligned}\mathbb{E}\left|\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k}\right|^p\lesssim\ &\mathbb{E}\left|\frac{\widehat{\mu}_n}{\mu}\left[(-1)^d\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\phi_{j_0,k}^{(d)}(X_i)-\alpha_{j_0,k}\right]\right|^p\\ &+\mathbb{E}\left|\alpha_{j_0,k}\cdot\widehat{\mu}_n\left(\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right)\right|^p.\end{aligned}\tag{4.6}$$

Then, it follows from Condition A5, Hölder's inequality and the orthonormality of $\{\phi_{j_0,k}\}$ that $|\alpha_{j_0,k}|=|\int_{[0,1]}r^{(d)}(x)\phi_{j_0,k}(x)dx|\lesssim 1$. Moreover, Condition A2 and the definition of $\widehat{\mu}_n$ imply that $|\widehat{\mu}_n|\lesssim 1$. Hence, the inequality (4.6) reduces to

$$\begin{aligned}\mathbb{E}\left|\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k}\right|^p\lesssim\ &\mathbb{E}\left|\frac{\mu}{n}\sum_{i=1}^{n}(-1)^d\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\phi_{j_0,k}^{(d)}(X_i)-\alpha_{j_0,k}\right|^p+\mathbb{E}\left|\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right|^p\\ :=\ &Q_1+Q_2.\end{aligned}\tag{4.7}$$

Let us now bound $Q_1$ and $Q_2$ as sharp as possible.

- Upper bound of $Q_1$.

Define $\xi_i:=\frac{(-1)^d\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\phi_{j_0,k}^{(d)}(X_i)-\alpha_{j_0,k}$. Then, one gets

$$Q_1:=\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i\right|^p=\left(\frac{1}{n}\right)^p\mathbb{E}\left|\sum_{i=1}^{n}\xi_i\right|^p.$$

Because $\phi^{(d)}$ is a bounded variation function, one can assume

$$\phi^{(d)}:=\overline{\phi}-\widetilde{\phi},$$

where $\overline{\phi}$ and $\widetilde{\phi}$ are bounded, nonnegative and nondecreasing functions ([22]). Then, we can write

$$\phi_{j_0,k}^{(d)}:=2^{j_0d}(\overline{\phi}_{j_0,k}-\widetilde{\phi}_{j_0,k}).$$

Moreover, one defines

$$\overline{\alpha}_{j_0,k}:=\int(-1)^d2^{j_0d}\overline{\phi}_{j_0,k}(x)r(x)dx,\quad \widetilde{\alpha}_{j_0,k}:=\int(-1)^d2^{j_0d}\widetilde{\phi}_{j_0,k}(x)r(x)dx$$

and

$$\overline{\xi}_i:=\frac{(-1)^d2^{j_0d}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\phi}_{j_0,k}(X_i)-\overline{\alpha}_{j_0,k},\quad \widetilde{\xi}_i:=\frac{(-1)^d2^{j_0d}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\widetilde{\phi}_{j_0,k}(X_i)-\widetilde{\alpha}_{j_0,k}.$$

Then, we have $\alpha_{j_0,k} = \overline{\alpha}_{j_0,k} - \widetilde{\alpha}_{j_0,k},\ \xi_i = \overline{\xi}_i - \widetilde{\xi}_i$ and, by an elementary inequality of convexity, one gets

$$(4.8)\quad Q_1 = \left(\frac{1}{n}\right)^p \mathbb{E}\left|\sum_{i=1}^{n}\left(\overline{\xi}_i - \widetilde{\xi}_i\right)\right|^p \lesssim \left(\frac{1}{n}\right)^p \left[\mathbb{E}\left|\sum_{i=1}^{n}\overline{\xi}_i\right|^p + \mathbb{E}\left|\sum_{i=1}^{n}\widetilde{\xi}_i\right|^p\right].$$

Using (1.1), (1.2) and Condition A4, one knows that $\mathbb{E}\overline{\xi}_i = 0$. Note that $\frac{\rho(y)\overline{\phi}_{j_0,k}(x)}{\omega(x,y)h(x)}$ is a nondecreasing function by the monotonicity of $\overline{\phi}_{j_0,k}(x)$ and Conditions A1-A3. Furthermore, we get that $\{\overline{\xi}_i, i = 1, 2, \ldots, n\}$ is negatively associated by Lemma 2.1. On the other hand, $|\overline{\xi}_i|^p \lesssim \left|\frac{(-1)^d 2^{j_0 d}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\phi}_{j_0,k}(X_i)\right|^p + |\overline{\alpha}_{j_0,k}|^p$ and $|\overline{\alpha}_{j_0,k}|^p = \left|\mathbb{E}\left[\frac{(-1)^d 2^{j_0 d}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\phi}_{j_0,k}(X_i)\right]\right|^p \le \mathbb{E}\left|\frac{(-1)^d 2^{j_0 d}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\phi}_{j_0,k}(X_i)\right|^p$ thanks to Jensen's inequality. Then, one has

$$\begin{aligned}\mathbb{E}|\overline{\xi}_i|^p &\lesssim \mathbb{E}\left|\frac{(-1)^d 2^{j_0 d}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\phi}_{j_0,k}(X_i)\right|^p \\ &= \int_{\mathbb{R}}\int_{[0,1]}\left|\frac{(-1)^d 2^{j_0 d}\mu\rho(y)}{\omega(x,y)h(x)}\overline{\phi}_{j_0,k}(x)\right|^p f(x,y)dxdy.\end{aligned}$$

Using Conditions A1—A3 and (1.1), one finds that

$$(4.9)\qquad \mathbb{E}|\overline{\xi}_i|^p \lesssim 2^{j_0 dp}\int_{[0,1]}|\overline{\phi}_{j_0,k}(x)|^p dx \lesssim 2^{j_0[(d+\frac{1}{2})p-1]}.$$

In particular, $\mathbb{E}|\overline{\xi}_i|^2 \lesssim 2^{2j_0 d}$. Recall Rosenthal's inequality ([18]): If $X_1, X_2, \ldots, X_n$ are negatively associated random variables such that $\mathbb{E}X_i = 0$ and $\mathbb{E}|X_i|^p < \infty$, then

$$\mathbb{E}\left|\sum_{i=1}^{n}X_i\right|^p \lesssim \begin{cases}\sum\limits_{i=1}^{n}\mathbb{E}|X_i|^p + \left(\sum\limits_{i=1}^{n}\mathbb{E}X_i^2\right)^{\frac{p}{2}}, & p > 2;\\ \left(\sum\limits_{i=1}^{n}\mathbb{E}X_i^2\right)^{\frac{p}{2}}, & 1 \le p \le 2.\end{cases}$$

According to this inequality and (4.9), one gets

$$\mathbb{E}\left|\sum_{i=1}^{n}\overline{\xi}_i\right|^p \lesssim \begin{cases}[2^{j_0[(d+\frac{1}{2})p-1]}\cdot n + (n\cdot 2^{2j_0 d})^{\frac{p}{2}}], & p \ge 2;\\ 2^{j_0 dp}n^{p/2}, & 1 \le p < 2.\end{cases}$$

This with $2^{j_0} < n$ shows that $\mathbb{E}\left|\sum\limits_{i=1}^{n}\overline{\xi}_i\right|^p \lesssim 2^{j_0 dp}n^{p/2}$. Similarly, $\mathbb{E}\left|\sum\limits_{i=1}^{n}\widetilde{\xi}_i\right|^p \lesssim 2^{j_0 dp}n^{p/2}$. Combining those with (4.8), one knows that

$$(4.10)\qquad Q_1 \lesssim 2^{j_0 dp}n^{-p/2}.$$

- Upper bound of $Q_2$.

Using the definition of $\widehat{\mu}_n$, one has

$$\begin{aligned}\mathbb{E}\Big|\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\Big|^p &= \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega(X_i,Y_i)}-\frac{1}{\mu}\right|^p\\ &= \frac{1}{n^p}\mathbb{E}\left|\sum_{i=1}^{n}\left[\frac{1}{\omega(X_i,Y_i)}-\frac{1}{\mu}\right]\right|^p. \end{aligned}\tag{4.11}$$

Define $\eta_i := \frac{1}{\omega(X_i,Y_i)} - \frac{1}{\mu}$. Then, $\mathbb{E}(\eta_i) = 0$ by (4.4). The monotonicity of $\omega(x,y)$ in Condition A2 and Lemma 2.1 imply that $\eta_1,\ldots,\eta_n$ are negatively associated. In addition, $\mathbb{E}|\eta_i|^p \lesssim 1$ thanks to Condition A2. According to Rosenthal's inequality, one has

$$\mathbb{E}\left|\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right|^p \lesssim n^{-\frac{p}{2}}.\tag{4.12}$$

Now it is easy to see from (4.7), (4.10) and (4.12) that

$$\mathbb{E}\Big|\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k}\Big|^p \lesssim 2^{j_0 dp} n^{-\frac{p}{2}}.$$

This completes the proof of Lemma 4.1. □

In this position, we will state our first theorem.

**Theorem 4.1.** *For the problem (1.1)—(1.2) with Conditions A1—A5. Let $r^{(d)} \in B^s_{\widetilde{p},q}(H)(\widetilde{p},q \in [1,\infty),\ s>0)$, and $\widetilde{p} \geq p \geq 1$, or $\widetilde{p} \leq p < \infty$ and $s > \frac{1}{\widetilde{p}}$. The linear wavelet estimator $\widehat{r}_n^{(d)}$ be defined in (4.1) with $2^{j_0} \sim n^{\frac{1}{2s'+2d+1}}$ and $s' = s - \left(\frac{1}{\widetilde{p}}-\frac{1}{p}\right)_+$. Then, for $1 \leq p < \infty$, we have*

$$\mathbb{E}\int_{[0,1]}\Big|\widehat{r}_n^{(d)}(x)-r^{(d)}(x)\Big|^p dx \lesssim n^{-\frac{s'p}{2s'+2d+1}}.$$

**Proof of Theorem 4.1:** Note that

$$\mathbb{E}\int_{[0,1]}\Big|\widehat{r}_n^{(d)}(x)-r^{(d)}(x)\Big|^p dx \lesssim \mathbb{E}\left\|\sum_{k\in\Omega}(\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k})\phi_{j_0,k}\right\|_p^p + \left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_p^p.\tag{4.13}$$

It follows from Lemma 3.1 that

$$\mathbb{E}\left\|\sum_{k\in\Omega}(\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k})\phi_{j_0,k}\right\|_p^p \lesssim 2^{p(\frac{j_0}{2}-\frac{j_0}{p})}\sum_{k\in\Omega}\mathbb{E}\Big|\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k}\Big|^p.$$

Using Lemma 4.1, $|\Omega| \sim 2^{j_0}$ and $2^{j_0} \sim n^{\frac{1}{2s'+2d+1}}$, one knows

$$\mathbb{E}\left\|\sum_{k\in\Omega}(\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k})\phi_{j_0,k}\right\|_p^p \lesssim \left(\frac{2^{j_0(1+2d)}}{n}\right)^{\frac{p}{2}} \sim n^{-\frac{s'p}{2s'+2d+1}}. \tag{4.14}$$

Next, one estimates $\left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_p^p$. When $\widetilde{p}\le p$ and $s>\frac{1}{\widetilde{p}}$, $B^s_{\widetilde{p},q}(\mathbb{R})\subseteq B^{s'}_{p,q}(\mathbb{R})$. Then, $r^{(d)}\in B^{s'}_{p,q}(\mathbb{R})$ and

$$\left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_p^p \lesssim 2^{-j_0s'p} \tag{4.15}$$

thanks to Lemma 3.2. When $\widetilde{p}>p$, $s'=s$. Using Hölder's inequality and the compact support of $r^{(d)}$ and $\phi$, one gets

$$\left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_p^p \lesssim \left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_{\widetilde{p}}^p.$$

Then, it is easy to see from Lemma 3.2 and $r^{(d)}\in B^s_{\widetilde{p},q}(H)$ that $\left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_p^p \lesssim 2^{-j_0s'p}$. This result with (4.15) shows that, for $1\le p<\infty$,

$$\left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_p^p \lesssim 2^{-j_0s'p}. \tag{4.16}$$

Furthermore, by $2^{j_0}\sim n^{\frac{1}{2s'+2d+1}}$, one gets

$$\left\|P_{j_0}r^{(d)}-r^{(d)}\right\|_p^p \lesssim n^{-\frac{s'p}{2s'+2d+1}}. \tag{4.17}$$

Combining this with (4.13) and (4.14),

$$\mathbb{E}\int_{[0,1]}\left|\widehat{r}_n^{(d)}(x)-r^{(d)}(x)\right|^p dx \lesssim n^{-\frac{s'p}{2s'+2d+1}}.$$

This ends the proof of Theorem 4.1. □

Since $j_0$ depends on $s'$ which remains unknown, $\widehat{r}_n^{(d)}(x)$ is not adaptive. Theorem 4.1 is however of interest to determine in a simple manner sharp rates of convergence in our statistical setting. We do not however claim that they are optimal in the minimax sense; the lower bounds in this case are not proved in this study. Also, Theorem 4.1 can be viewed as generalization to the [3, Theorem 3.3] to the negatively associated case.

## 5. NONLINEAR WAVELET ESTIMATION

In this section, we will construct a adaptive nonlinear wavelet estimator and consider its upper bound over $L^p(1\le p<+\infty)$ risk. Now, we define our nonlinear wavelet estimator

$$\widetilde{r}_n^{(d)}(x) := \sum_{k\in\Omega} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{j_1}\sum_{k\in\Lambda_j} \widehat{\beta}_{j,k} I_{\{|\widehat{\beta}_{j,k}|\geq \kappa t_n\}}\psi_{j,k}(x), \tag{5.1}$$

where $t_n := 2^{jd}\sqrt{\frac{\ln n}{n}}$,

$$\widehat{\beta}_{j,k} = (-1)^d \frac{\widehat{\mu}_n}{n}\sum_{i=1}^{n} \frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi_{j,k}^{(d)}(X_i) \tag{5.2}$$

and $I_A$ denotes the indicator function over a set $A$, i.e., $I_A = 1$ if $A$ is satisfied and 0 otherwise. The positive integers $j_0, j_1$ (depend on $n$) and the positive number $\kappa$ will be given later on. The main difference between $\widetilde{r}^{(d)}$ and the linear wavelet estimator is the individual selection of the $\widehat{\beta}_{j,k}$'s done by the hard thresholding rule (formalized by the indicator function over $\{|\widehat{\beta}_{j,k}| \geq \kappa t_n\}$). We refer to [13] and [15] for the deep link between this selection technique and the intrinsic properties of the wavelets.

It should be pointed out that $\mathbb{E}\left[(-1)^d \frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi_{j,k}^{(d)}(X_i)\right] = \beta_{j,k}$ thanks to [14, Lemma 2.1] (which uses Condition A4).

Note that Lemma 4.1 is still true if $\widehat{\alpha}_{j_0,k}$ is replaced by $\widehat{\beta}_{j,k}$, which leads to the following lemma.

**Lemma 5.1.** *For the problem (1.1)-(1.2) with Conditions A1—A5 hold. If $2^j \leq n$, then for $1 \leq p < \infty$, we have*

$$\mathbb{E}\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right|^p \lesssim 2^{jdp} n^{-\frac{p}{2}}.$$

**Lemma 5.2.** *For the problem (1.1)-(1.2) with Conditions A1-A5. Then, for $j2^j \leq n$ and each $w > 0$, there exists a constant $\kappa > 1$ such that*

$$\mathbb{P}\left(\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right| \geq \kappa t_n\right) \lesssim 2^{-wj}.$$

**Proof of Lemma 5.2:** Via similar arguments to those used in (4.7), we obtain

$$\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right| \lesssim \left|\frac{\mu}{n}\sum_{i=1}^{n}(-1)^d \frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi_{j,k}^{(d)}(X_i) - \beta_{j,k}\right| + \left|\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n}\right|.$$

Hence, it suffices to prove

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(-1)^d\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi_{j,k}^{(d)}(X_i) - \beta_{j,k}\right]\right| \geq \frac{\kappa}{2}t_n\right) \lesssim 2^{-wj} \tag{5.3}$$

and

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\left[\frac{1}{\omega(X_i,Y_i)}-\frac{1}{\mu}\right]\right|\geq\frac{\kappa}{2}t_n\right)\lesssim 2^{-wj}.$$

One shows the first inequality (5.3) only, the second one is similar and even simpler.

Define $\gamma_i := \frac{(-1)^d\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi_{j,k}^{(d)}(X_i)-\beta_{j,k}$. Then, one has

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(-1)^d\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi_{j,k}^{(d)}(X_i)-\beta_{j,k}\right]\right|\geq\frac{\kappa}{2}t_n\right)=\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\right|\geq\frac{\kappa}{2}t_n\right).$$

Because $\psi^{(d)}$ is a bounded variation function, one can assume

$$\psi^{(d)} := \overline{\psi}-\widetilde{\psi},$$

where $\overline{\psi}$ and $\widetilde{\psi}$ are bounded, nonnegative and nondecreasing functions ([22]). Then,

$$\psi_{j,k}^{(d)} := 2^{jd}(\overline{\psi}_{j,k}-\widetilde{\psi}_{j,k}).$$

Moreover, one defines

$$\overline{\beta}_{j,k} := \int(-1)^d2^{jd}\overline{\psi}_{j,k}(x)r(x)dx,\ \widetilde{\beta}_{j,k} := \int(-1)^d2^{jd}\widetilde{\psi}_{j,k}(x)r(x)dx,$$

and

$$\overline{\gamma}_i := \frac{(-1)^d2^{jd}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\psi}_{j,k}(X_i)-\overline{\beta}_{j,k},\ \widetilde{\gamma}_i := \frac{(-1)^d2^{jd}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\widetilde{\psi}_{j,k}(X_i)-\widetilde{\beta}_{j,k}.$$

Then, $\beta_{j,k}=\overline{\beta}_{j,k}-\widetilde{\beta}_{j,k},\ \gamma_i=\overline{\gamma}_i-\widetilde{\gamma}_i$ and

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\right|\geq\frac{\kappa}{2}t_n\right)\\ &\lesssim\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\overline{\gamma}_i\right|\geq\frac{\kappa}{4}t_n\right)+\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\gamma}_i\right|\geq\frac{\kappa}{4}t_n\right).\end{aligned}\tag{5.4}$$

According to (1.1), (1.2) and Condition A4, one gets $\mathbb{E}\overline{\gamma}_i=\overline{\beta}_{j,k}$. Moreover, $\overline{\gamma}_1,\overline{\gamma}_2,\ldots,\overline{\gamma}_n$ are negatively associated by Conditions A1-A3, Lemma 2.1 and the nondecreasing property of $\overline{\psi}_{j,k}$. On the other hand, by the bounded properties of functions in Conditions A1-A3, $\left|\frac{(-1)^d2^{jd}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\psi}_{j,k}(X_i)\right|\lesssim 2^{j(d+\frac{1}{2})}$ and

$$|\overline{\gamma}_i|\lesssim\left|\frac{(-1)^d2^{jd}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\psi}_{j,k}(X_i)\right|+\mathbb{E}\left|\frac{(-1)^d2^{jd}\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\overline{\psi}_{j,k}(X_i)\right|\lesssim 2^{j(d+\frac{1}{2})}.$$

Similar to the arguments of (4.9) with $p=2$, $\mathbb{E}(\overline{\gamma}_i)^2 \lesssim 2^{2jd}$. Recall Bernstein's inequality: Let $X_1,\dots,X_n$ be negatively associated random variables such that $\mathbb{E}X_i=0$, $|X_i|\le M$ and $\mathbb{E}X_i^2=\sigma^2$. Then, for each $v\ge 0$,

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n} X_i\right| \ge v\right) \le 2\cdot \exp\left\{-\frac{nv^2}{2(\sigma^2+\frac{vM}{3})}\right\}.$$

It follows from Bernstein's inequality, $t_n = 2^{jd}\sqrt{\frac{\ln n}{n}}$ and $j2^j \le n$ that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\overline{\gamma}_i\right| \ge \frac{\kappa}{4}t_n\right) \lesssim \exp\left\{-\frac{n(\frac{\kappa t_n}{4})^2}{2(2^{2jd}+\frac{\kappa t_n}{12}\ 2^{j(d+\frac{1}{2})})}\right\} \lesssim \exp\left\{-\frac{\ln n\ \kappa^2}{32(1+\frac{\kappa}{12})}\right\}.$$

Obviously, there exists sufficiently large $\kappa>1$ such that $\exp\{-\frac{\ln n\ \kappa^2}{32(1+\frac{\kappa}{12})}\} \lesssim 2^{-wj}$. Hence,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\overline{\gamma}_i\right| \ge \frac{\kappa}{4}t_n\right) \lesssim 2^{-wj}.$$

Similarly, $\mathbb{P}\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n}\widetilde{\gamma}_i\right| \ge \frac{\kappa}{4}t_n\right) \lesssim 2^{-wj}$. Those results with (5.4) show that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(-1)^d\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi_{j,k}^{(d)}(X_i)-\beta_{j,k}\right]\right| \ge \frac{\kappa}{2}t_n\right) \lesssim 2^{-wj}.$$

This ends the proof of Lemma 5.2. □

Now we will give our last theorem in this position.

**Theorem 5.1.** *For the problem (1.1)—(1.2) with Conditions A1—A5. Let* $r^{(d)} \in B^s_{\widetilde{p},q}(H)(\widetilde{p},q\in[1,\infty),\ s>0)$, *and* $\widetilde{p}\ge p\ge 1$, *or* $\widetilde{p}\le p<\infty$ *and* $s>\frac{1}{\widetilde{p}}$. *Then, the nonlinear wavelet estimator* $\widetilde{r}_n^{(d)}$ *defined in (5.1) with* $2^{j_0}\sim n^{\frac{1}{2m+2d+1}}$ $(m>s)$ *and* $2^{j_1}\sim(\frac{n}{\ln n})^{\frac{1}{2d+1}}$ *satisfies*

$$\mathbb{E}\int_{[0,1]}|\widetilde{r}_n^{(d)}(x)-r^{(d)}(x)|^p dx \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p}, \tag{5.5}$$

*where*

$$\alpha = \begin{cases} \frac{s}{2s+2d+1}, & \widetilde{p}\ge\frac{p(2d+1)}{2s+2d+1}, \\ \frac{s-1/\widetilde{p}+1/p}{2(s-1/\widetilde{p})+2d+1}, & \widetilde{p}<\frac{p(2d+1)}{2s+2d+1}. \end{cases} \tag{5.6}$$

**Proof of Theorem 5.1:** For the proof of Theorem 5.1, we will prove it under two case respectively.

**(i)** Upper bound estimation under $\widetilde{p}\le p<\infty$ and $s>\frac{1}{\widetilde{p}}$.

In this case, (5.6) can be rewritten as

$$\alpha = \min\left\{\frac{s}{2s+2d+1},\ \frac{s-1/\widetilde{p}+1/p}{2(s-1/\widetilde{p})+2d+1}\right\}.$$

By the definition of $\widetilde{r}_n^{(d)}(x)$,

$$\mathbb{E}\int_{[0,1]}|\widetilde{r}_n^{(d)}(x)-r^{(d)}(x)|^p dx \lesssim \mathbb{E}\left\|\sum_{k\in\Omega}(\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k})\phi_{j_0,k}\right\|_p^p + \left\|r^{(d)}-P_{j_1+1}r^{(d)}\right\|_p^p$$
$$+\,\mathbb{E}\left\|\sum_{j=j_0}^{j_1}\sum_{k\in\Lambda j}\left[\widehat{\beta}_{j,k}I_{\{|\widehat{\beta}_{j,k}|\geq\kappa t_n\}}-\beta_{j,k}\right]\psi_{j,k}\right\|_p^p. \tag{5.7}$$

It follows from Lemma 3.1 that

$$\mathbb{E}\left\|\sum_{k\in\Omega}(\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k})\phi_{j_0,k}\right\|_p^p \lesssim 2^{p(\frac{j_0}{2}-\frac{j_0}{p})}\sum_{k\in\Omega}\mathbb{E}\left|\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k}\right|^p.$$

1 Using Lemma 4.1, $|\Omega|\sim 2^{j_0}$ and $2^{j_0}\sim n^{\frac{1}{2m+2d+1}}$ $(m>s)$, one knows

$$\mathbb{E}\left\|\sum_{k\in\Omega}(\widehat{\alpha}_{j_0,k}-\alpha_{j_0,k})\phi_{j_0,k}\right\|_p^p \lesssim n^{-\frac{mp}{2m+2d+1}} < n^{-\frac{sp}{2s+2d+1}} \leq n^{-\alpha p}. \tag{5.8}$$

2 Similar to the arguments of (4.15), when $\widetilde{p}\leq p$ and $s>\frac{1}{\widetilde{p}}$, one gets that

$$\left\|P_{j_1+1}r^{(d)}-r^{(d)}\right\|_p \lesssim 2^{-j_1(s-1/\widetilde{p}+1/p)}. \tag{5.9}$$

On the other hand, $s-\frac{1}{\widetilde{p}}+\frac{1}{p}\geq\alpha$ thanks to $\widetilde{p}\leq p$ and $s>\frac{1}{\widetilde{p}}$. Then, it follows from $2^{j_1}\sim(\frac{n}{\ln n})^{\frac{1}{2d+1}}$ that

$$\left\|P_{j_1+1}r^{(d)}-r^{(d)}\right\|_p^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{(s-1/\widetilde{p}+1/p)p}{2d+1}} \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p}.$$

3 The main work for the proof of Theorem 5.1 is to show

$$Z := \mathbb{E}\left\|\sum_{j=j_0}^{j_1}\sum_{k\in\Lambda j}\left[\widehat{\beta}_{j,k}I_{\{|\widehat{\beta}_{j,k}|\geq\kappa t_n\}}-\beta_{j,k}\right]\psi_{j,k}\right\|_p^p \lesssim (\ln n)^{\frac{3p}{2}}n^{-\alpha p}. \tag{5.10}$$

It is easy to see from Lemma 3.1 that

$$Z \lesssim (j_1-j_0+1)^{p-1}\sum_{j=j_0}^{j_1}2^{p(\frac{j}{2}-\frac{j}{p})}\sum_{k\in\Lambda_j}\mathbb{E}\left|\widehat{\beta}_{j,k}I_{\{|\widehat{\beta}_{j,k}|\geq\kappa t_n\}}-\beta_{j,k}\right|^p.$$

1 Then, the classical technique ([13]) gives

$$Z \lesssim (j_1 - j_0 + 1)^{p-1}(Z_1 + Z_2 + Z_3), \tag{5.11}$$

where

$$Z_1 = \sum_{j=j_0}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} \mathbb{E}\left[|\widehat{\beta}_{j,k} - \beta_{j,k}|^p I_{\{|\widehat{\beta}_{j,k}-\beta_{j,k}|>\frac{\kappa t_n}{2}\}}\right],$$

$$Z_2 = \sum_{j=j_0}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} \mathbb{E}\left[|\widehat{\beta}_{j,k} - \beta_{j,k}|^p I_{\{|\widehat{\beta}_{j,k}|\geq\kappa t_n,\ |\beta_{j,k}|\geq\frac{\kappa t_n}{2}\}}\right],$$

$$Z_3 = \sum_{j=j_0}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |\beta_{j,k}|^p I_{\{|\widehat{\beta}_{j,k}|<\kappa t_n,\ |\beta_{j,k}|\leq 2\kappa t_n\}}.$$

2 • Upper bound of $Z_1$.

It follows from Hölder's inequality that

$$\mathbb{E}\left[\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right|^p I_{\{|\widehat{\beta}_{j,k}-\beta_{j,k}|>\frac{\kappa t_n}{2}\}}\right] \leq \left[\mathbb{E}\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right|^{2p}\right]^{\frac{1}{2}} \left[\mathbb{P}\left(\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right| > \frac{\kappa t_n}{2}\right)\right]^{\frac{1}{2}}.$$

3 Furthermore, Lemmas 5.1 and 5.2 imply that

$$\mathbb{E}\left[\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right|^p I_{\{|\widehat{\beta}_{j,k}-\beta_{j,k}|>\frac{\kappa t_n}{2}\}}\right] \lesssim 2^{jdp} n^{-\frac{p}{2}}\ 2^{-\frac{wj}{2}},$$

where $\kappa > 1$ is chosen for $w > p + 2dp$ in Lemma 5.2. This with the choice $2^{j_0} \sim n^{\frac{1}{2m+2d+1}}(m > s)$ shows that

$$\begin{aligned} Z_1 &\lesssim n^{-\frac{p}{2}} \sum_{j=j_0}^{j_1} 2^{j(\frac{p}{2}+dp-\frac{w}{2})} \lesssim n^{-\frac{p}{2}} 2^{j_0(\frac{p}{2}+dp)} \lesssim n^{-\frac{mp}{2m+2d+1}} \\ &\leq n^{-\frac{sp}{2s+2d+1}} \leq n^{-\alpha p}. \end{aligned} \tag{5.12}$$

4 • Upper bound of $Z_2$.

Taking

$$2^{j_0^*} \sim \left(\frac{n}{\ln n}\right)^{\frac{1-2\alpha}{2d+1}}.$$

Because $0 < \alpha \leq \frac{s}{2s+2d+1}$ and $2^{j_0} \sim n^{\frac{1}{2m+2d+1}}(m > s)$, $2^{j_0^*} \leq 2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{2d+1}}$ and $2^{j_0^*} \geq (\frac{n}{\ln n})^{\frac{1-\frac{2s}{2s+2d+1}}{2d+1}} = (\frac{n}{\ln n})^{\frac{1}{2s+2d+1}} \gtrsim n^{\frac{1}{2m+2d+1}} \sim 2^{j_0}$. Furthermore, it follows from Lemma 5.1 that

$$\begin{aligned} Z_{21} &:= \sum_{j=j_0}^{j_0^*} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} \mathbb{E}\left[\left|\widehat{\beta}_{j,k} - \beta_{j,k}\right|^p I_{\{|\widehat{\beta}_{j,k}|\geq\kappa t_n,\ |\beta_{j,k}|\geq\frac{\kappa t_n}{2}\}}\right] \lesssim \\ &\lesssim \sum_{j=j_0}^{j_0^*} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} 2^{jdp} n^{-\frac{p}{2}} \lesssim 2^{j_0^*(\frac{p}{2}+dp)} n^{-\frac{p}{2}} \lesssim n^{-\alpha p}. \end{aligned} \tag{5.13}$$

On the other hand, by Lemmas 5.1 and 3.2, and $t_n = 2^{jd}\sqrt{\frac{\ln n}{n}}$, one has

$$
\begin{aligned}
Z_{22} &:= \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} \mathbb{E}\left[\left|\widehat{\beta}_{j,k}-\beta_{j,k}\right|^p I_{\{|\widehat{\beta}_{j,k}|\geq \kappa t_n,\ |\beta_{j,k}|\geq \frac{\kappa t_n}{2}\}}\right] \\
&\lesssim \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} \mathbb{E}\left|\widehat{\beta}_{j,k}-\beta_{j,k}\right|^p \left(\frac{|\beta_{j,k}|}{\kappa t_n/2}\right)^{\widetilde{p}} \\
&\lesssim \sum_{j=j_0^*+1}^{j_1} (\ln n)^{-\widetilde{p}/2} n^{-\frac{p-\widetilde{p}}{2}} 2^{-j[s\widetilde{p}-\frac{(p-\widetilde{p})(2d+1)}{2}]}. \qquad (5.14)
\end{aligned}
$$

Define

$$\varepsilon := s\widetilde{p} - \frac{(p-\widetilde{p})(2d+1)}{2}.$$

Then, $\varepsilon > 0$ holds if and only if $\widetilde{p} > \frac{p(2d+1)}{2s+2d+1}$, and (5.14) can be rewritten as

$$Z_{22} \lesssim (\ln n)^{-\widetilde{p}/2} n^{-\frac{p-\widetilde{p}}{2}} \sum_{j=j_0^*+1}^{j_1} 2^{-j\varepsilon}. \qquad (5.15)$$

When $\varepsilon > 0$, $\widetilde{p} > \frac{p(2d+1)}{2s+2d+1}$ and $\alpha = \frac{s}{2s+2d+1}$ thanks to (5.6). Moreover, it can be easily checked that $\frac{p-\widetilde{p}}{2} + \frac{1-2\alpha}{2d+1}[s\widetilde{p} - \frac{(p-\widetilde{p})(2d+1)}{2}] = \alpha p$. This with the choice of $2^{j_0^*}$ leads to

$$
\begin{aligned}
Z_{22} &\lesssim (\ln n)^{-\widetilde{p}/2} n^{-\frac{p-\widetilde{p}}{2}} 2^{-j_0^*\varepsilon} \leq (\ln n)\left(\frac{1}{n}\right)^{\frac{p-\widetilde{p}}{2}+\frac{1-2\alpha}{2d+1}[s\widetilde{p}-\frac{(p-\widetilde{p})(2d+1)}{2}]} \\
&= (\ln n) n^{-\alpha p}. \qquad (5.16)
\end{aligned}
$$

For the case $\varepsilon \leq 0$, $\widetilde{p} \leq \frac{p(2d+1)}{2s+2d+1}$ and $\alpha = \frac{s-\frac{1}{\widetilde{p}}+\frac{1}{p}}{2(s-\frac{d}{\widetilde{p}})+2d+1}$. Define $p_1 := (1-2\alpha)p$. Then, $\alpha \leq \frac{s}{2s+2d+1}$ and $\widetilde{p} \leq \frac{p(2d+1)}{2s+2d+1} < (1-2\alpha)p = p_1$. Similarly to (5.14), one has

$$
\begin{aligned}
Z_{22} &\lesssim \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} \mathbb{E}\left|\widehat{\beta}_{j,k}-\beta_{j,k}\right|^p \left(\frac{|\beta_{j,k}|}{\kappa t_n/2}\right)^{p_1} \\
&\lesssim \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} 2^{jdp} n^{-\frac{p}{2}} t_n^{-p_1} \|\beta_j\|_{p_1}^{p_1}.
\end{aligned}
$$

Because $\widetilde{p} \leq p_1$ and $r^{(d)} \in B^s_{\widetilde{p},q}(H)$, we get $\|\beta_j\|_{p_1}^{p_1} \leq \|\beta_j\|_{\widetilde{p}}^{p_1} \lesssim 2^{-j(s-\frac{1}{\widetilde{p}}+\frac{1}{2})p_1}$ and

$$
\begin{aligned}
Z_{22} &\lesssim \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} 2^{jdp} n^{-\frac{p}{2}} t_n^{-p_1} 2^{-j(s-\frac{1}{\widetilde{p}}+\frac{1}{2})p_1} \\
&\leq \left(\frac{1}{n}\right)^{\frac{p-p_1}{2}} \sum_{j=j_0^*+1}^{j_1} 2^{-j(sp_1-\frac{p_1}{\widetilde{p}}+\frac{p_1}{2}+dp_1-dp-\frac{p}{2}+1)}.
\end{aligned}
$$

By the definitions of $p_1$ and $\alpha$, $sp_1 - \frac{p_1}{\widetilde{p}} + \frac{p_1}{2} + dp_1 - dp - \frac{p}{2} + 1 = 0$ and $Z_{22} \lesssim (\frac{1}{n})^{\frac{p-p_1}{2}}(\ln n) = (\ln n)(\frac{1}{n})^{\alpha p}$. This with (5.13) and (5.16) shows in both cases,

$$Z_2 = Z_{21} + Z_{22} \lesssim (\ln n)\, n^{-\alpha p}. \tag{5.17}$$

• Upper bound of $Z_3$.

It is easy to see that

$$\begin{aligned} Z_{31} &:= \sum_{j=j_0}^{j_0^*} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |\beta_{j,k}|^p I_{\{|\widehat{\beta}_{j,k}|<\kappa t_n,\ |\beta_{j,k}|\le 2\kappa t_n\}} \\ &\le \sum_{j=j_0}^{j_0^*} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |2\kappa t_n|^p \lesssim \sum_{j=j_0}^{j_0^*} 2^{j(\frac{p}{2}+dp)} \left(\frac{\ln n}{n}\right)^{\frac{p}{2}} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}} 2^{j_0^*(\frac{p}{2}+dp)} \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p}. \end{aligned} \tag{5.18}$$

On the other hand, one has

$$\begin{aligned} Z_{32} &:= \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |\beta_{j,k}|^p I_{\{|\widehat{\beta}_{j,k}|<\kappa t_n,\ |\beta_{j,k}|\le 2\kappa t_n\}} \\ &\le \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |\beta_{j,k}|^p \left|\frac{2\kappa t_n}{\beta_{j,k}}\right|^{p-\widetilde{p}} \\ &\lesssim \sum_{j=j_0^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} t_n^{p-\widetilde{p}} \|\beta_j\|_{\widetilde{p}}^{\widetilde{p}} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\widetilde{p}}{2}} \sum_{j=j_0^*+1}^{j_1} 2^{-j\varepsilon}. \end{aligned} \tag{5.19}$$

The same arguments as (5.15) shows that, for $\varepsilon > 0$,

$$Z_{32} \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p}. \tag{5.20}$$

For the case of $\varepsilon \le 0$, one defines

$$2^{j_1^*} \sim \left(\frac{n}{\ln n}\right)^{\frac{\alpha}{s-1/\widetilde{p}+1/p}}.$$

Note that $\varepsilon \le 0$ and $s > \frac{1}{\widetilde{p}}$. Then, $\widetilde{p} \le \frac{p(2d+1)}{2s+2d+1}$, $\alpha = \frac{s-\frac{1}{\widetilde{p}}+\frac{1}{p}}{2(s-\frac{1}{\widetilde{p}})+2d+1}$ and $\alpha \le s-\frac{1}{\widetilde{p}}+\frac{1}{p}$. Hence, $n^{\frac{1-2\alpha}{2d+1}} \lesssim 2^{j_0^*} \le 2^{j_1^*} \le 2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{2d+1}}$ and $Z_{32} = Z_{321} + Z_{322}$, where

$$Z_{321} := \sum_{j=j_0^*+1}^{j_1^*} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |\beta_{j,k}|^p I_{\{|\widehat{\beta}_{j,k}|<\kappa t_n,\ |\beta_{j,k}|\le 2\kappa t_n\}},$$

$$Z_{322} := \sum_{j=j_1^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |\beta_{j,k}|^p I_{\{|\widehat{\beta}_{j,k}|<\kappa t_n,\ |\beta_{j,k}|\le 2\kappa t_n\}}.$$

By the arguments of (5.15) and the choice of $2^{j_1^*}$, one has

$$Z_{321} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\widetilde{p}}{2}} 2^{-j_1^*\varepsilon} = \left(\frac{\ln n}{n}\right)^{\frac{p-\widetilde{p}}{2}+\frac{\alpha\varepsilon}{s-1/\widetilde{p}+1/p}}.$$

It is easy to check that $\frac{p-\widetilde{p}}{2}+\frac{\alpha\varepsilon}{s-1/\widetilde{p}+1/p}=\alpha p$. Then,

$$Z_{321} \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p}.$$

On the other hand, using $\|\beta_j\|_{\widetilde{p}} \lesssim 2^{-j(s-\frac{1}{\widetilde{p}}+\frac{1}{2})}$, $s>\frac{1}{\widetilde{p}}$ and $2^{j_1^*}\sim(\frac{n}{\ln n})^{\frac{\alpha}{s-1/\widetilde{p}+1/p}}$.

$$\begin{aligned} Z_{322} &\le \sum_{j=j_1^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \sum_{k\in\Lambda_j} |\beta_{j,k}|^p \le \sum_{j=j_1^*+1}^{j_1} 2^{p(\frac{j}{2}-\frac{j}{p})} \|\beta_j\|_{\widetilde{p}}^p \\ &\lesssim \sum_{j=j_1^*+1}^{j_1} 2^{-j(1+sp-p/\widetilde{p})} \lesssim 2^{-j_1^*(1+sp-p/\widetilde{p})} \sim \left(\frac{\ln n}{n}\right)^{\alpha p}. \end{aligned}$$

Now, it follows that for $\varepsilon\le 0$,

$$Z_{32} = Z_{321}+Z_{322} \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p}.$$

Combining this with (5.18) and (5.20), one knows

$$Z_3 \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p}. \tag{5.21}$$

Then, it follows from (5.11), (5.12), (5.17) and (5.21) that

$$Z \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p}.$$

Hence,

$$E\int_{[0,1]} |\widetilde{r}_n^{(d)}(x) - r^{(d)}(x)|^p dx \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p} \tag{5.22}$$

in the case of $\widetilde{p}\le p<\infty$ and $s>\frac{1}{\widetilde{p}}$.

**(ii)** Upper bound estimation under $\widetilde{p}>p$.

From the above arguments, one finds that when $\widetilde{p}=p$, the inequality (5.22) still holds without the assumption $s>\frac{1}{\widetilde{p}}$. It remains to conclude (5.22) for $\widetilde{p}>p\ge 1$. By Hölder's inequality,

$$\int_{[0,1]} |\widetilde{r}_n^{(d)}(x) - r^{(d)}(x)|^p dx \lesssim \left[\int_{[0,1]} |\widetilde{r}_n^{(d)}(x) - r^{(d)}(x)|^{\widetilde{p}} dx\right]^{\frac{p}{\widetilde{p}}}.$$

Using Jensen's inequality and (5.22) with $\widetilde{p} = p$, one gets

$$\mathbb{E}\int_{[0,1]} |\widetilde{r}_n^{(d)}(x) - r^{(d)}(x)|^p dx \lesssim \left[\mathbb{E}\int_{[0,1]} |\widetilde{r}_n^{(d)}(x) - r^{(d)}(x)|^{\widetilde{p}} dx\right]^{\frac{p}{\widetilde{p}}} \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p}.$$

This completes the proof of Theorem 5.1. □

Contrary to the linear wavelet estimator given by (4.1), $\widetilde{r}_n^{(d)}(x)$ is fully adaptive; its construction does not depend on $s$. The convergence rate of the nonlinear estimator keeps the same as that of the linear one up to a logarithmic factor when $\widetilde{p} > p$. However, it gets better in the case of $\widetilde{p} \leq p$. This aspect remains standard in nonlinear wavelet estimation in the standard regression (or density) estimation framework (see [15]). Also, Theorem 5.1 can be viewed as generalization to the [14, Theorem 1] to the negatively associated case.

## 6. Concluding remarks

In this paper, the estimation of the derivatives of a regression function for biased data is considered. The feature of the study is to investigate the negatively dependent assumption on the data, beyond the independent assumption, opening new perspective of applications. Two wavelet estimators are introduced. The first estimator is based on wavelet projection of wavelet coefficient estimators only, the second estimator is nonlinear; a selection of the wavelet coefficient estimators are applied according to their magnitude via a hard thresholding rule. Sharp rates of convergence are obtained under the $L^p$ risk with $1 \leq p < \infty$, assuming that the function of interest belongs to a ball of Besov spaces $B^s_{\widetilde{p},q}(\mathbb{R})$. These rates correspond to those obtained in the independent setting, showing that the wavelet methodology is consistent for this problem. Perspectives of this work are to prove the optimal lower bounds in the minimax sense, to relax some assumptions on the model, mainly the compact support of $r^{(d)}$ and explore the practical aspects of the proposed estimators. These points needs further investigations that we leave for a future work.

## ACKNOWLEDGMENTS

The authors would like to thank the referees for their important comments and suggestions.

This paper is supported by the National Natural Science Foundation of China (No. 71961004), Guangxi Natural Science Foundation (Nos. 2018GXNSFBA281076, 2019GXNSFFA245012).

## REFERENCES


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

[2] Alam, K. and Saxena, K.M.L. (1981). Positive dependence in multivariate distributions, *Communications in Statistics—Theory and Methods*, **10**, 1183-1196.

[3] Chaubey, Y.P., Chesneau, C. and Navarro, F. (2017). Linear wavelet estimation of the derivatives of a regression function based on biased data, *Communications in Statistics—Theory and Methods*, **46**, 19, 9541-9556.

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

[5] Chaubey, Y.P., Laïb, N. and Li, J. (2012). Generalized kernel regression estimator for dependent size-biased data, *Journal of Statistical and Planning Inference*, **142**, 708-727.

[6] Chaubey, Y.P. and Shirazi, E. (2015). On MISE of a non linear wavelet estimator of the regression function based on biased data under strong mixing, *Communications in Statistics—Theory and Methods*, **44**, 5, 885-899.

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

[8] Chaubey, Y.P. and Shirazi, E. (2015). On MISE of a non linear wavelet estimator of the regression function based on biased data under strong mixing, *Communications in Statistics-Theory and Methods*, **44**, 5, 885-899.

[9] Chesneau, C. and Shirazi, E. (2014). Nonparametric wavelet regression based on biased data, *Communications in Statistics—Theory and Methods*, **43**, 13, 2642-2658.

[10] Cristóbal, J.A. and Alcalá, J.T. (2000). Nonparametric regression estimators for length biased data, *Journal of Statistical and Planning Inference*, **89**, 145-168.

[11] Cristóbal, J.A. and Alcalá, J.T. (2001). An overview of nonparametric contributions to the problem of functional estimation from biased data, *Test*, **10**, 2, 309-332.

[12] Cristòbal, J.A., Ojeda, J.L. and Alcalà, J.T. (2004). Confidence bands in nonparametric regression with length biased data, *Annals of the Institute of Statistical Mathematics*, **56**, 3, 475-496.

[13] Donoho, D.L., Johnstone, M. I., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *The Annals of Statistics*, **24**, 2, 508-539.

[14] Guo, H.J. and Kou, J.K. (2019). Non linear wavelet estimation of regression derivatives based on biased data, *Communications in Statistics—Theory and Methods*, **48**, 13, 3219-3235.

[15] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1997). *Wavelets, Approximation and Statistical Application*, New York: Springer-Verlag.

[16] Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications, *Annals of Statistics*, **11**, 286-295.

[17] Kou, J. and Liu, Y. (2018). Wavelet regression estimations with strong mixing data, *Statistical Methods and Applications*, **27**, 4, 667-688.

[18] Liu, Y. M. and Xu, J.L. (2014). Wavelet density estimation for negatively associated stratified size-biased sample, *Journal of Nonparametric Statistics*, **26**, 3, 537-554.

[19] Roussas, G.G. (1999) *Positive and negative dependence with some statistical application*, Ghosh. ed., Asymptotics, Nonparametrics and Time Series. New York: Marcel Dekker.

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

[21] Wu, C.O. (2000). Local polynomial regression with selection biased data, *Statistica Sinica*, **10**, 3, 789-817.

[22] Xu, J.L. (2014). Wavelet linear estimations of density derivatives from a negatively associated stratified size-biased sample, *Frontiers of Mathematics in China*, **9**, 3, 623-640.

[23] Xu, F., Wang, B.H. and Hou, Y.W. (2019). Almost sure local central limit theorem for the product of some partial sums of negatively associated sequences, *Journal of Inequalities and Applications*, **1**, 309.

[24] Zhang, Y., Yang X.Y., Dong Z.S. and Wang, D.H. (2011). The limit theorem for dependent random variables with applications to autoregression models, *Journal of Systems Science and Complexity*, **24**, 565-579.