

Research Article

Optimal Wavelet Estimation of Density Derivatives for Size-Biased Data

Jinru Wang, Zijuan Geng, and Fengfeng Jin

Department of Applied Mathematics, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Fengfeng Jin; jinfengfeng@emails.bjut.edu.cn

Received 10 November 2013; Accepted 17 December 2013; Published 30 January 2014

Academic Editor: D. Baleanu

Copyright © 2014 Jinru Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A perfect achievement has been made for wavelet density estimation by Dohono et al. in 1996, when the samples without any noise are independent and identically distributed (i.i.d.). But in many practical applications, the random samples always have noises, and estimation of the density derivatives is very important for detecting possible bumps in the associated density. Motivated by Dohono's work, we propose new linear and nonlinear wavelet estimators $\hat{f}_{\text{lin}}^{(m)}$, $\hat{f}_{\text{non}}^{(m)}$ for density derivatives $f^{(m)}$ when the random samples have size-bias. It turns out that the linear estimation $E(\|\hat{f}_{\text{lin}}^{(m)} - f^{(m)}\|_p)$ for $f^{(m)} \in B_{r,q}^s(A, L)$ attains the optimal convergence rate when $r \geq p$, and the nonlinear one $E(\|\hat{f}_{\text{lin}}^{(m)} - f^{(m)}\|_p)$ does the same if $r < p$.

1. Introduction

Wavelet analysis plays important roles in both pure and applied mathematics such as signal processing, image compress, numerical solution, and local fractional calculus [1, 2]. One of which is to estimate an unknown density function based on random samples [3–8]. The perfect achievement was made by Dohono et al. [9], when the i.i.d. samples X_1, X_2, \dots, X_n have not any noise. On the other hand, Besov spaces contain many functional spaces (e.g., Hölder spaces and Sobolev spaces with noninteger exponents) as their special examples. In some statistical models, the error is measured in L^p norm [9–13].

In practice, it usually happens that getting the direct sample from a random variable is impossible. In this paper, we want to consider the true density function $f_X(x)$. But we can only observe the samples Y_i , $i = 1, 2, \dots, n$, for the size-biased data; that is,

$$f_Y(y) = \frac{g(y)f_X(y)}{\mu}, \quad (1)$$

where $g(y)$ is the so-called bias function, $\mu = E(g(X)) < \infty$.

In many cases, a linear g is recommended, but, in general, the form of g should be studied via additional experiments. The purpose of this paper is to estimate the derivatives of

the true density functions $f_X^{(m)}$, $m = 0, 1, 2, \dots$; we study the optimal convergence rate of wavelet estimators in L^p norm over Besov spaces.

Size-biased data arise when an observation depends on samples magnitude. Several examples of model (1) can be found in the literature [14]. For instance, in [15], it is shown that the distribution of the concentration of alcohol in the blood of intoxicated drivers is of interest; since the drunken drivers have a larger chance of being arrested, the collected data are size-biased.

The estimation problem for biased data (1) has been discussed in some papers. In 1982, Vardi [16] considered the nonparametric maximum likelihood estimation for $f^X(x)$. In 1991, Jones [17] discussed the mean squared error properties of the kernel density estimation. In 2004, Efromovich [18] developed the Efromovich-Pinsker adaptive Fourier estimator. It was based on a block shrinkage algorithm and achieved the minimax rate of convergence under the L_2 risk over the Besov class $B_{2,2}^s$.

In 2010, Ramírez and Vidakovic [14] proposed a linear wavelet estimator and discussed the consistency of function in $L_2[0, 1]$ under the mean integrated squared error (MISE) sense. But the wavelet estimator in paper [14] contained the unknown parameter μ . In the same year, Chesneau [10] constructed a nonlinear wavelet estimator and evaluated the

L_p risk in the Besov space $B_{r,q}^s$. But about the estimation of the density derivatives about model (1), to our knowledge, we have not seen any result. Estimation of the derivatives of a density is very important in detecting possible bumps.

The current paper is organized as follows. In Section 2, we briefly describe the preliminaries on wavelets and Besov space. The linear estimator and its convergence rate are presented in Section 3. In order to discuss optimality, Section 4 is devoted to give the lower bound for an arbitrary estimator. In Section 5, we consider nonlinear wavelet estimator and its optimal convergence rate. Our estimations improve the theorems in [10, 13, 14, 18].

2. Wavelets and Besov Spaces

In this section, we will recall some useful and well-known concepts and lemmas.

In order to construct a wavelet basis, we need a structure in $L^2(\mathbb{R})$ which can decompose $L^2(\mathbb{R})$ into a direct sum of mutually orthogonal spaces.

Definition 1 (see [19]). A multiresolution analysis (MRA) of $L^2(\mathbb{R})$ is a set of increasing, closed linear subspaces $V_j \subset V_{j+1}$, for all $j \in \mathbb{Z}$, called scaling spaces, satisfying

- (a) $\bigcap_{-\infty}^{\infty} V_j = \{0\}$, $\overline{\bigcup_{-\infty}^{\infty} V_j} = L^2(\mathbb{R})$;
- (b) $f(\cdot) \in V_0$ if and only if $f(2^j \cdot) \in V_j$ all $j \in \mathbb{Z}$;
- (c) $f(\cdot) \in V_0$ if and only if $f(\cdot - k) \in V_0$ for all $k \in \mathbb{Z}$;
- (d) there exists a function $\phi(\cdot) \in V_0$ such that $\{\phi(\cdot - k)\}$ is an orthogonal basis in V_0 . The function $\phi(\cdot)$ is called the scaling function of the multiresolution analysis.

With the standard notation f in wavelet analysis, there exists a corresponding wavelet function $\psi(x) = \sum_k (-1)^k \overline{h_{1-k}} \varphi_{1k}(x)$, $h_k = \langle \varphi, \varphi_{1k} \rangle$, such that for fixed j , $\{\psi_{j,k}(t)\}_{k \in \mathbb{Z}}$ is an orthonormal basis of W_j which is the orthogonal complement of the space V_j in V_{j+1} ($V_{j+1} = V_j \oplus W_j$). For fixed $J \in \mathbb{N}$, both $\{\varphi_{Jk}(x), \psi_{j,k}(x)\}_{j \geq J, k \in \mathbb{Z}}$ and $\{\psi_{j,k}(x)\}_{j,k \in \mathbb{Z}}$ are orthonormal bases of $L^2(\mathbb{R})$.

As usual, $L_p(\mathbb{R})$ ($p \geq 1$) denotes the classical Lebesgue space on the real line \mathbb{R} . Although wavelet bases are constructed for $L_2(\mathbb{R})$, most of them constitute unconditional bases for $L_p(\mathbb{R})$.

Lemma 2 (see [20]). *Let φ be a compactly supported, orthonormal scaling function and ψ the corresponding wavelet. Then for any $f(x) \in L_p(\mathbb{R})$ with $1 \leq p \leq \infty$, the following expansion:*

$$\sum_{k \in \mathbb{Z}} \alpha_{J,k} \varphi_{J,k}(x) + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} \beta_{j,k} \psi_{j,k}(x) \quad (2)$$

converges to $f(x)$ for almost everywhere $x \in \mathbb{R}$, where

$$\alpha_{j,k} = \int_{\mathbb{R}} f(x) \varphi_{j,k}(x) dx, \quad (3)$$

$$\beta_{j,k} = \int_{\mathbb{R}} f(x) \psi_{j,k}(x) dx.$$

Lemma 3 (see [3]). *If the scaling function φ satisfies $\text{ess sup} \sum_{k \in \mathbb{Z}} |\varphi(x-k)| < \infty$, then for any sequence $\{\lambda_k\}_{k \in \mathbb{Z}} \in l_p$, one has*

$$C_1 \|\lambda\|_{l_p} 2^{(j/2-j/p)} \leq \left\| \sum_k \lambda_k \varphi_{j,k} \right\|_p \leq C_2 \|\lambda\|_{l_p} 2^{(j/2-j/p)}, \quad (4)$$

where $C_1 = (\|\theta_\varphi\|_\infty^{1/p} \|\varphi\|_1^{1/q})^{-1}$, $C_2 = (\|\theta_\varphi\|_\infty^{1/q} \|\varphi\|_1^{1/p})^{-1}$, $1 \leq p \leq \infty$, $1/p + 1/q = 1$.

Letting $1 \leq p, q \leq \infty$, $s := n + \alpha$, $\alpha \in (0, 1]$, the Besov spaces $B_{pq}^s(\mathbb{R})$ are defined by

$$B_{pq}^s(\mathbb{R}) := \left\{ f \mid f \in L_p(\mathbb{R}), f^{(n)} \in L_p(\mathbb{R}), \omega_p^2(f^n, t) = \varepsilon(t) t^\alpha, \|\varepsilon\|_q^* < \infty \right\} \quad (5)$$

with the associated norm $\|f\|_{B_{pq}^s} := \|f\|_p + \|f^{(n)}\|_p + \|\omega_p^2(f^n, t)/t^\alpha\|_q^*$, where $\omega_p^2(f, t) := \sup_{|h| \leq t} \|f(x+2h) - 2f(x+h) + f(x)\|_p$ denotes the smoothness modulus of f , and $\|\varepsilon\|_q^* := \begin{cases} (\int_0^\infty |\varepsilon(t)|^q (dt/t))^{1/q}, & 1 \leq q < \infty, \\ \text{ess sup}_t |\varepsilon(t)|, & q = \infty. \end{cases}$

Between the different Besov spaces, the following embedding conclusions are established [3]. Let $s > 0$, $1 \leq p, q, r \leq \infty$; then

- (i) $B_{pq}^s \hookrightarrow B_{p\infty}^s \hookrightarrow B_{\infty\infty}^{s-1/p}$, $s > 1/p$;
- (ii) $B_{r'q}^s \hookrightarrow B_{pq}^{s'}$, $r < p$, $s' = s - 1/r + 1/p$,

where $A \hookrightarrow B$ denotes that the Banach space A is continuously embedding in the Banach space B ; that is, there exists a constant $c > 0$ such that, for any $u \in A$, we have $\|u\|_B \leq c\|u\|_A$.

A scaling function $\varphi(x)$ is called t -regular, if $\varphi(x)$ has continuous derivatives of order t , and its corresponding wavelet $\psi(x)$ has vanishing moments of order t ; that is,

$$\int x^k \psi(x) dx = 0, \quad k = 0, 1, \dots, t-1. \quad (6)$$

One of advantages of wavelets is that they can characterize Besov spaces.

Lemma 4 (see [3]). *Let $\varphi(x)$ be a compact supported, t -regular orthonormal scaling function with the corresponding wavelet $\psi(x)$ and $0 < s < t$. If $f \in L_p(\mathbb{R})$, $\alpha_{0k} := \langle f, \varphi_{0k} \rangle$, $\beta_{jk} := \langle f, \psi_{jk} \rangle$, and $1 \leq p, q \leq \infty$, then the following are equivalent:*

- (i) $f \in B_{pq}^s(\mathbb{R})$;
- (ii) $(2^{js} \|P_j f - f\|_p)_{j \geq 0} \in l^q$, where P_j is the projection operator to V_j ; that is, $P_j f := \sum_k \langle f, \varphi_{j,k} \rangle \varphi_{j,k}$;
- (iii) $\|\alpha_0\|_p + \|\{2^{j(s+1/2-1/p)} \|\beta_{j,k}\|_p\}_{j \geq 0}\|_q < +\infty$. In this case,

$$\|f\|_{B_{pq}^s} \sim \|\alpha_0\|_p + \left\| \{2^{j(s+1/2-1/p)} \|\beta_{j,k}\|_p\}_{j \geq 0} \right\|_q. \quad (7)$$

Note 1. The notation $A \lesssim B$ indicates that $A \leq cB$ with a positive constant c , which is independent of A and B . If $A \lesssim B$ and $B \lesssim A$, we write $A \sim B$.

In this paper, the Besov balls $\tilde{B}_{r,q}^s(A, L)$ are defined by

$$\tilde{B}_{r,q}^s(A, L) := \left\{ f \mid f \in B_{r,q}^s(\mathbb{R}), \|f\|_{B_{r,q}^s} \leq L, |\text{supp } f| \leq A \right\}. \tag{8}$$

3. Linear Estimator

In this section, we will give a linear estimator for density derivatives $f_X^{(d)}$ in Besov spaces $\tilde{B}_{r,q}^s(A, L)$.

The linear wavelet estimator of the derivative of a density $f_X(y)$ is defined as follows:

$$\hat{f}_{X,\text{lin}}^{(d)}(y) = \sum_k \hat{\alpha}_{J,k} \phi_{J,k}(y), \tag{9}$$

where

$$\hat{\alpha}_{J,k} := (-1)^d \frac{\hat{\mu}}{n} \sum_{i=1}^n \frac{\phi_{J,k}^{(d)}(Y_i)}{g(Y_i)}, \quad \hat{\mu} := \frac{n}{\sum_{i=1}^n (1/g(Y_i))}. \tag{10}$$

The following inequalities play important roles in this paper.

Lemma 5 (see [3] (Rosenthal inequality)). *Let X_1, \dots, X_n be independent random variables such that $E(X_i) = 0$ and $|X_i| < M$; then there exists a constant $C(p) > 0$ such that*

- (i) $E(|\sum_{i=1}^n X_i|^p) \leq C(p)(M^{p-2} \sum_{i=1}^n E(|X_i|^2) + (\sum_{i=1}^n E(X_i^2))^{p/2}), p > 2,$
- (ii) $E(|\sum_{i=1}^n X_i|^p) \leq C(p)(\sum_{i=1}^n E(X_i^2))^{p/2}, 0 < p \leq 2.$

About the defined coefficients in (9), although $E(\hat{\alpha}_{J,k}) \neq \alpha_{J,k}$, we have the following estimation.

Lemma 6. *If $2^j \leq n$, then, for any $1 \leq p < \infty$, one has $E|\hat{\alpha}_{Jk} - \alpha_{Jk}|^p \leq n^{-p/2} 2^{jd^p}$.*

Proof. By the definitions of $\hat{\alpha}_{Jk}$, α_{Jk} and triangular inequality, one observes that

$$\begin{aligned} & |\hat{\alpha}_{Jk} - \alpha_{Jk}| \\ &= \left| (-1)^d \frac{\hat{\mu}}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right| \\ &= \left| \frac{\hat{\mu}}{\mu} \left((-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right) + \hat{\mu} \alpha_{Jk} \left(\frac{1}{\mu} - \frac{1}{\hat{\mu}} \right) \right| \\ &\leq \left| \frac{\hat{\mu}}{\mu} \right| \cdot \left| (-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right| + |\hat{\mu} \alpha_{Jk}| \cdot \left| \frac{1}{\mu} - \frac{1}{\hat{\mu}} \right|. \end{aligned} \tag{11}$$

Since $g_1 \leq g(y) \leq g_2$ for any $y \in \mathbb{R}$, one has that

$$\hat{\mu} = \frac{n}{\sum_{i=1}^n (1/g(Y_i))} \leq g_2, \quad \mu = Eg(X) \geq g_1, \tag{12}$$

and $\alpha_{Jk} = \int_A f_X^{(d)}(y) \phi_{Jk}(y) dy$. Thanks to embedding theorem $B_{r,\infty}^s \hookrightarrow 1B_{\infty,\infty}^{s-1/r}$, for any $s > 1/r$, one gets $\|f\|_{\infty} \leq \|f\|_{B_{\infty,\infty}^{s-1/r}} \leq \|f\|_{B_{r,\infty}^s}$. It is easy to see $|\hat{\mu}/\mu|, |\hat{\mu}|, |\alpha_{Jk}|$ are bounded. Using the convexity inequality, one obtains

$$\begin{aligned} & E|\hat{\alpha}_{Jk} - \alpha_{Jk}|^p \\ &\leq E \left(\left| (-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right| + \left| \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right| \right)^p \\ &\leq 2^{p-1} E \left(\left| (-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right|^p + \left| \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right|^p \right) \\ &\leq E \left| (-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right|^p + E \left| \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right|^p \\ &=: T_1 + T_2, \end{aligned} \tag{13}$$

where

$$T_1 := E \left| (-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right|^p, \quad T_2 := E \left| \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right|^p. \tag{14}$$

(i) To estimate T_1 :

$$\begin{aligned} T_1 &=: E \left| (-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right|^p \\ &=: E \left| \frac{1}{n} \sum_{i=1}^n \left((-1)^d \mu \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{Jk} \right) \right|^p, \end{aligned} \tag{15}$$

denote $\xi_i := (-1)^d \mu (\phi_{Jk}^{(d)}(Y_i)/g(Y_i)) - \alpha_{Jk}$. Note that $\{\xi_i\}$ are i.i.d. samples, and $E\xi_i = 0$. Moreover, for any given integer $m \geq 2$, one has

$$\begin{aligned} & E \left| (-1)^d \mu \frac{\phi_{Jk}^{(d)}(Y_i)}{g(Y_i)} \right|^m \\ &= E \left| \frac{\mu}{g(Y_i)} (\phi_{Jk}^{(d)}(Y_i))^2 \right| \cdot \left| \left(\frac{\mu}{g(Y_i)} \right)^{m-1} (\phi_{Jk}^{(d)}(Y_i))^{m-2} \right| \\ &\leq \left(\frac{g_2}{g_1} \right)^{m-1} 2^{(J/2+dJ)(m-2)} \|\phi^{(d)}\|_{\infty}^{m-2} E \left| \frac{\mu}{g(Y_i)} (\phi_{Jk}^{(d)}(Y_i))^2 \right| \\ &\leq 2^{(J/2+dJ)(m-2)} E \left| \frac{\mu}{g(Y_i)} (\phi_{Jk}^{(d)}(Y_i))^2 \right|, \\ & E \left| \frac{\mu}{g(Y_i)} (\phi_{Jk}^{(d)}(Y_i))^2 \right| \\ &= \int_A \frac{\mu}{g(y)} (\phi_{Jk}^{(d)}(y))^2 f_Y(y) dy \\ &= \int_A \frac{\mu}{g(y)} (\phi_{Jk}^{(d)}(y))^2 \frac{g(y) f_X(y)}{\mu} dy \\ &= \int_A f_X(y) (\phi_{Jk}^{(d)}(y))^2 dy \end{aligned}$$

$$\begin{aligned}
&= \int_A f_X(y) 2^{J(1/2+d)2} (\phi_{jk}^{(d)}(2^J y - k))^2 dy \\
&\leq 2^{2dJ}.
\end{aligned} \tag{16}$$

Therefore,

$$\begin{aligned}
E|\xi_i|^m &= E \left| (-1)^d \mu \frac{\phi_{jk}^{(d)}(Y_i)}{g(Y_i)} - \alpha_{jk} \right|^m \\
&\leq 2^{m-1} \left(E \left| (-1)^d \mu \frac{\phi_{jk}^{(d)}(Y_i)}{g(Y_i)} \right|^m + |\alpha_{jk}|^m \right) \\
&\leq 2^{J(1/2+d)m-J}.
\end{aligned} \tag{17}$$

By Rosenthal's inequality Lemma 5,

(a) if $1 \leq p \leq 2$, one has

$$\begin{aligned}
T_1 &= E \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right|^p \leq n^{-p} \left(\sum_{i=1}^n E|\xi_i|^2 \right)^{p/2} \\
&= n^{-p} (n E|\xi_i|^2)^{p/2} \leq n^{-p/2} 2^{Jdp};
\end{aligned} \tag{18}$$

(b) if $p \geq 2$, one gets

$$\begin{aligned}
T_1 &= E \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right|^p \leq n^{-p} C \left[\sum_{i=1}^n E|\xi_i|^p + \left(\sum_{i=1}^n E|\xi_i|^2 \right)^{p/2} \right] \\
&\leq n^{-p} (n 2^{J(1/2+d)p-J}) + n^{-p} \left(\sum_{i=1}^n E|\xi_i|^2 \right)^{p/2} \leq n^{-p/2} 2^{Jdp}.
\end{aligned} \tag{19}$$

(ii) To estimate the term T_2 , since

$$\begin{aligned}
T_2 &=: E \left| \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right|^p = E \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{g(Y_i)} - \frac{1}{\mu} \right|^p \\
&= E \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{g(Y_i)} - \frac{1}{\mu} \right) \right|^p,
\end{aligned} \tag{20}$$

let $\eta_i := 1/g(Y_i) - 1/\mu$. It is easy to see $E\eta_i = 0$, and, for any integer $m \geq 2$, one obtains

$$\begin{aligned}
E|\eta_i|^m &= E \left| \frac{1}{g(Y_i)} - \frac{1}{\mu} \right|^m \leq 2^{m-1} \left(E \left| \frac{1}{g(Y_i)} \right|^m + \mu^{-m} \right) \\
&\leq 2^{m-1} (g_1^{-m} + g_1^m) := C(m).
\end{aligned} \tag{21}$$

Similarly, by Rosenthal's inequality Lemma 5,

(a) for $1 \leq p < 2$, $m = 2$, one gets

$$T_2 = E \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right|^p \leq n^{-p} \left(\sum_{i=1}^n E|\eta_i|^2 \right)^{p/2} \leq n^{-p} \cdot n^{p/2} = n^{-p/2}; \tag{22}$$

(b) for $p \geq 2$, that is, $1 - p \leq -p/2$, and $m = p$, one has

$$\begin{aligned}
T_2 &= E \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right|^p \leq n^{-p} C(p) \left(\sum_{i=1}^n E|\eta_i|^p + \left(\sum_{i=1}^n E|\eta_i|^2 \right)^{p/2} \right) \\
&\leq n^{-p+1} + n^{-p/2} \leq n^{-p/2}.
\end{aligned} \tag{23}$$

Summarizing the above estimation about T_1, T_2 , one obtains that $E|\hat{\alpha}_{jk} - \alpha_{jk}|^p \leq n^{-p/2} 2^{Jdp}$. \square

Theorem 7. Let scaling function $\phi(x)$ be compactly supported and t -regular. If $\hat{f}_{X,\text{lin}}^{(d)}$ be the estimator defined in (9), then for $1 \leq r, p, q < \infty$, $1/r < s < t - d$, one has

$$\sup_{f_X^{(d)} \in \tilde{B}_{r,q}^s(A,L)} E \left\| \hat{f}_{X,\text{lin}}^{(d)}(y) - f_X^{(d)}(y) \right\|_p^p \leq n^{-s'p/(1+2(s'+d))}, \tag{24}$$

where $s' = s - (1/r - 1/p)_+$.

Proof. Firstly, using triangular inequality and convexity inequality, we decompose $E \left\| \hat{f}_{X,\text{lin}}^{(d)}(y) - f_X^{(d)}(y) \right\|_p^p$ into the bias term and the stochastic term; that is,

$$\begin{aligned}
&E \left\| \hat{f}_{X,\text{lin}}^{(d)}(y) - f_X^{(d)}(y) \right\|_p^p \\
&\leq E \left(\left\| \hat{f}_{X,\text{lin}}^{(d)}(y) - P_J f_X^{(d)}(y) \right\|_p \right. \\
&\quad \left. + \left\| P_J f_X^{(d)}(y) - f_X^{(d)}(y) \right\|_p \right)^p \\
&\leq 2^{p-1} \left(E \left\| \hat{f}_{X,\text{lin}}^{(d)}(y) - P_J f_X^{(d)}(y) \right\|_p \right. \\
&\quad \left. + \left\| P_J f_X^{(d)}(y) - f_X^{(d)}(y) \right\|_p \right).
\end{aligned} \tag{25}$$

For the bias term $\left\| P_J f_X^{(d)}(y) - f_X^{(d)}(y) \right\|_p$, one can estimate it as follows.

(i) When $r = p$, Lemma 4 reduces to

$$\left\| P_J f_X^{(d)}(y) - f_X^{(d)}(y) \right\|_p^p \leq 2^{-Jsp}. \tag{26}$$

(ii) When $r < p$, using Besov space embedding theorems

$$B_{r,q}^s \hookrightarrow B_{p,q}^{s'}, \quad s' = s - 1/r + 1/p, \quad \text{one has}$$

$$\begin{aligned}
&\sup_{f_X^{(d)} \in \tilde{B}_{r,q}^s(A,L)} \left\| P_J f_X^{(d)}(y) - f_X^{(d)}(y) \right\|_p^p \\
&\leq \sup_{f_X^{(d)} \in \tilde{B}_{p,q}^{s'}(A,L)} \left\| P_J f_X^{(d)}(y) - f_X^{(d)}(y) \right\|_p^p \leq 2^{-J s' p}.
\end{aligned} \tag{27}$$

(iii) When $r > p$, Hölder's inequality and Lemma 4 tell us that

$$\begin{aligned} & \|P_J f_X^{(d)}(y) - f_X^{(d)}(y)\|_p^p \\ &= \int_A |P_J f_X^{(d)}(y) - f_X^{(d)}(y)|^p dy \\ &\leq \left(\int_A |P_J f_X^{(d)}(y) - f_X^{(d)}(y)|^{p \cdot (r/p)} dy \right)^{p/r} \\ &\quad \times \left(\int_A 1 dy \right)^{1-p/r} \\ &\leq \|P_J f_X^{(d)}(y) - f_X^{(d)}(y)\|_r^p \leq 2^{-J^s p}. \end{aligned} \tag{28}$$

Hence, for $1 \leq r < \infty$, one obtains that

$$\sup_{f_X^{(d)} \in \tilde{B}_{r,q}^s(A,L)} \|P_J f_X^{(d)}(y) - f_X^{(d)}(y)\|_p^p \leq 2^{-J^s p}. \tag{29}$$

Next, we estimate the stochastic term $E\|\tilde{f}_X^{(d)}(y) - P_J f_X^{(d)}(y)\|_p^p$. Clearly, due to Lemmas 3 and 6, one gets

$$\begin{aligned} E\|\tilde{f}_{X,\text{lin}}^{(d)}(y) - P_J f_X^{(d)}(y)\|_p^p &= E\left\| \sum_k (\tilde{\alpha}_{jk} - \alpha_{jk}) \phi_{j,k}(y) \right\|_p^p \\ &\leq 2^{(J/2-J/p)p} \sum_k E|(\tilde{\alpha}_{jk} - \alpha_{jk})|^p \\ &\leq n^{-p/2} 2^{J(1/2+d)p}. \end{aligned} \tag{30}$$

By choosing J such that $2^{J(n)} \sim n^{1/(1+2(s'+d))}$, one obtains that

$$\begin{aligned} & \sup_{f_X^{(d)} \in \tilde{B}_{r,q}^s(A,L)} E\|\tilde{f}_{X,\text{lin}}^{(d)}(y) - f_X^{(d)}(y)\|_p^p \\ & \leq 2^{-J^s p} + n^{-p/2} 2^{J(1/2+d)p} \\ & \leq n^{-s' p / (1+2(s'+d))}. \end{aligned} \tag{31}$$

□

Remark 8. Theorem 7 can be considered as natural extension of [14] if $d = 0, p = r = 2$. Moreover, the next part shows the optimality of our linear estimation for $r \geq p$.

4. Lower Bound and Optimality

This section is devoted to showing that the linear estimator defined in (9) attains the optimal convergence rate for $r \geq p$. The idea of proof is motivated by [21].

Lemma 9 (Varshamov-Gilbert Lemma [5]). *Let $\Theta := \{\varepsilon = (\varepsilon_1, \dots, \varepsilon_m)\}$, $\varepsilon_i \in \{0, 1\}$; then there exists a subset $\{\varepsilon^0, \dots, \varepsilon^M\}$ of Θ with $\varepsilon^0 = (0, \dots, 0)$ such that $M \geq 2^{m-3}$ and $\sum_{k=1}^M |\varepsilon_k^i - \varepsilon_k^j| \geq m/8, 0 \leq i \neq j \leq M$.*

Lemma 10 (Fano's lemma [22]). *Let $(\Omega, \mathfrak{N}, P_k)$ be probability measurable spaces and $A_k \in \mathfrak{N}, k = 0, 1, \dots, m$. If $A_k \cap A_v = \emptyset$ for $\forall k \neq v$, one has*

$$\sup_{0 \leq k \leq m} P_k(A_k^c) \geq \min \left\{ \frac{1}{2}, \sqrt{m} \exp(-3e^{-1} - \kappa_m) \right\}, \tag{32}$$

where A_k^c stands for the complement of $A_k, \kappa_m := \inf_{0 \leq v \leq m} (1/m) \sum_{k \neq v} K(P_k, P_v)$ and $K(P_k, P_v)$ stands for Kullback distance in [5].

Based on the above lemmas, we have the following lower bound estimation.

Theorem 11. *Let $f_X^{(d)} \in \tilde{B}_{r,q}^s(A, L)$ with $1 \leq r, q \leq \infty, 1 \leq p < \infty$ and $s > 1/r$; there exist two constants g_1 and g_2 such that $0 < g_1 \leq g(x) \leq g_2 < \infty$. If $\tilde{f}_X^{(d)}$ is any estimator of $f_X^{(d)}$ with i.i.d. random samples, then*

$$\begin{aligned} & \sup_{f_X^{(d)} \in \tilde{B}_{r,q}^s(A,L)} E\|\tilde{f}_n^{(d)} - f_X^{(d)}\|_p \\ & \geq \max \left\{ \left(\frac{\ln n}{n} \right)^{(s-1/r+1/p)/(2(s-1/r+d)+1)}, n^{-s/(2(s+d)+1)} \right\} \\ & = \begin{cases} \left(\frac{\ln n}{n} \right)^{(s-1/r+1/p)/(2(s-1/r+d)+1)}, & r \leq \frac{(1+2d)p}{2(s+d)+1}, \\ n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \end{aligned} \tag{33}$$

Proof. (i) Firstly, we prove

$$\sup_{f_X^{(d)} \in \tilde{B}_{r,q}^s(A,L)} E\|\tilde{f}_n^{(d)} - f_X^{(d)}\|_p \geq \left(\frac{\ln n}{n} \right)^{(s-1/r+1/p)/(2(s-1/r+d)+1)}. \tag{34}$$

It is sufficient to construct $h_k(x)$ such that $h_k^{(d)} \in \tilde{B}_{r,q}^s(A, L)$ and

$$\sup_k E\|\tilde{f}_n^{(d)} - h_k^{(d)}\|_p \geq \left(\frac{\ln n}{n} \right)^{(s-1/r+1/p)/(2(s-1/r+d)+1)}. \tag{35}$$

Suppose that φ is a compactly supported, $t(t > s + d)$ regular and orthonormal scaling function and ψ is the corresponding wavelet with $\text{supp } \psi \subseteq [0, l], l \in \mathbb{N}^+$. Assume $h(x) \in \tilde{B}_{r,q}^{s+d}(A, L/2), h(x)|_{[0,l]} = C_0 > 0$. Define $\Delta_j := \{0, l, 2l, \dots, (2^j - 1)l, 2^j l\}, a_j := 2^{-j(s+d+1/2-1/r)}$ and

$$h_k(x) := h(x) + a_j \psi_{jk}(x) I\{k \neq 2^j l\}, \quad k \in \Delta_j. \tag{36}$$

Obviously, $h_{2^j l} = h(x)$ and $h_k(x) \in \tilde{B}_{r,q}^{s+d}(A, L)$; that is, $h_k^{(d)}(x) \in \tilde{B}_{r,q}^s(A, L)$. Moreover, $\text{supp } \psi_{jk} \cap \text{supp } \psi_{j'k'} = \emptyset$ for $k, k' \in \Delta_j$ and $k \neq k'$. So, one gets

$$\begin{aligned} \|h_k^{(d)} - h_{k'}^{(d)}\|_p &= \|a_j (\psi_{jk}^{(d)} - \psi_{j'k'}^{(d)})\|_p \\ &\geq a_j \|\psi_{jk}^{(d)}\|_p = 2^{-j(s+1/p-1/r)} \|\psi^{(d)}\|_p := \eta_j. \end{aligned} \tag{37}$$

Clearly, $A_k := \{\|\widehat{f}_n^{(d)} - h_k^{(d)}\|_p < \eta_j/2\}$ satisfies $A_k \cap A_{k'} = \emptyset$ for $k \neq k'$. Then, Fano's Lemma 10 tells us that

$$\sup_{k \in \Delta_j} P_{h_k}^n(A_k^c) \geq \min \left\{ \frac{1}{2}, \sqrt{2^j} \exp(-3e^{-1} - \kappa_{2j}) \right\}. \quad (38)$$

On the other hand, one has

$$\begin{aligned} E \|\widehat{f}_n^{(d)}(x) - h_k^{(d)}(x)\|_p &\geq \frac{\eta_j}{2} P_{h_{y_k}}^n \left(\|\widehat{f}_n^{(d)} - h_k^{(d)}\|_p \geq \frac{\eta_j}{2} \right) \\ &= \frac{\eta_j}{2} P_{h_{y_k}}^n(A_k^c). \end{aligned} \quad (39)$$

Then

$$\begin{aligned} \sup_{k \in \Delta_j} E \left(\|\widehat{f}_n^{(d)}(x) - h_k^{(d)}(x)\|_p \right) &\geq \frac{\eta_j}{2} \sup_{k \in \Delta_j} P_{h_{y_k}}^n(A_k^c) \\ &\geq \frac{\eta_j}{2} \min \left\{ \frac{1}{2}, \sqrt{2^j} \exp(-3e^{-1} - \kappa_{2j}) \right\}, \end{aligned} \quad (40)$$

where $\kappa_{2j} := \inf_{k \in \Delta_j} \sum_{k \neq v} K(P_{h_{y_k}}^n, P_{h_{y_v}}^n)$. Next, one shows that $\kappa_{2j} \leq g_2 C_0^{-1} n a_j^2 / g_1$.

Recall that

$$\begin{aligned} K(P_1^n, P_2^n) &:= \int_{f_1^n, f_2^n > 0} f_1^n(x) \ln \frac{f_1^n(x)}{f_2^n(x)} dx \\ &= \sum_{i=1}^n \int f_1(x_i) \ln \frac{f_1(x_i)}{f_2(x_i)} dx_i = nK(P_1^1, P_2^1), \end{aligned} \quad (41)$$

where $f_i^n(x) := f_i(x_1) f_i(x_2) \cdots f_i(x_n)$, $i = 1, 2$. Note that, $\ln u \leq u - 1$, if $u > 0$, one has

$$\begin{aligned} K(P_1^n, P_2^n) &= n \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx \\ &\leq n \int \frac{1}{f_2(x)} [f_1(x) - f_2(x)]^2 dx. \end{aligned} \quad (42)$$

Since $h_k^{(d)}(x) = g(x)h_k(x)/\mu$ and $\text{supp } |h_k(x) - h(x)| \subset [0, l]$,

$$\begin{aligned} \kappa_{2j} &:= \inf_{k \in \Delta_j} 2^{-j} \sum_{k \neq v} K(P_{h_{y_k}}^n, P_{h_{y_v}}^n) \\ &\leq 2^{-j} \sum_{k \in \Delta_j, k \neq 2^j l} K(P_{h_{y_k}}^n, P_{h_{y_v}}^n) \\ &\leq 2^{-j} n \sum_{k \in \Delta_j, k \neq 2^j l} \int \frac{\mu}{g(x)h(x)} \\ &\quad \times \left[\frac{g(x)h_k(x)}{\mu} - \frac{g(x)h(x)}{\mu} \right]^2 dx \\ &\leq \frac{g_2}{g_1} 2^{-j} n \sum_{k \in \Delta_j, k \neq 2^j l} C_0^{-1} \int a_j^2 |\psi_{jk}|^2 dx \\ &= \frac{g_2}{g_1} C_0^{-1} n a_j^2. \end{aligned} \quad (43)$$

Taking $2^j \sim (n/\ln n)^{1/(2(s+d-1/r)+1)}$, then $n a_j^2 = n 2^{-2j(s+d+1/2-1/r)} \sim \ln n$. One can choose C such that $n a_j^2 \leq C \ln n$ and $C g_2 [4(s+d-1/r)+2] < C_0 g_1$. Therefore

$$\begin{aligned} \sqrt{2^j} e^{-\kappa_{2j}} &\geq \sqrt{2^j} e^{-g_2 C_0^{-1} n a_j^2 / g_1} \\ &\geq \left(\frac{n}{\ln n} \right)^{[4(s+d-1/r)+2]^{-1}} n^{-g_2 C_0^{-1} C / g_1} \geq 1. \end{aligned} \quad (44)$$

One has

$$\sup_{k \in \Delta_j} E \left(\|\widehat{f}_n^{(d)}(x) - h_k^{(d)}(x)\|_p \right) \geq \frac{\eta_j}{2} \sup_{k \in \Delta_j} P_{h_{y_k}}^n(A_k^c) \geq C \eta_j. \quad (45)$$

Noting that $\eta_j = 2^{-j(s+1/p-1/r)} \|\psi^{(d)}\|_p$, one gets

$$\begin{aligned} \sup_{k \in \Delta_j} E \left(\|\widehat{f}_n^{(d)}(x) - h_k^{(d)}(x)\|_p \right) &\geq \left(\frac{\ln n}{n} \right)^{(s+1/p-1/r)/(2(s+d-1/r)+1)} \\ &\geq \left(\frac{\ln n}{n} \right)^{(s+1/p-1/r)/(2(s+d-1/r)+1)}. \end{aligned} \quad (46)$$

(ii) Next, we prove

$$\sup_{f_X^{(d)} \in \widetilde{B}_{r,q}^s(A,L)} E \|\widehat{f}_n^{(d)} - f_X^{(d)}\|_p \geq n^{-s/(2(s+d)+1)}. \quad (47)$$

Similarly, it is sufficient to construct $h_{\varepsilon^i}(x)$, $i = 0, 1, \dots, M$ such that $h_{\varepsilon^i}^{(d)}(x) \in \widetilde{B}_{r,q}^s(A, L)$ and

$$\sup_i E \|\widehat{f}_n^{(d)}(x) - h_{\varepsilon^i}^{(d)}(x)\|_p \geq n^{-s/(2(s+d)+1)}. \quad (48)$$

Similarly to prove (i), suppose that $h(x) \in \widetilde{B}_{r,q}^{s+d}(A, L/2)$, $h(x)|_{[0,l]} = C_0 > 0$. Defining $\widetilde{a}_j := 2^{-j(s+d+1/2)}$, $\widetilde{\Delta}_j := \{0, l, 2l, \dots, (2^j - 1)l\}$ and

$$h_{\varepsilon^i}(x) := h(x) + \widetilde{a}_j \sum_{k \in \widetilde{\Delta}_j} \varepsilon_k^i \psi_{jk}(x), \quad i = 0, 1, \dots, M, \quad (49)$$

with $\varepsilon^i = (\varepsilon_k^i)_{k \in \widetilde{\Delta}_j} \in \{0, 1\}^{2^j}$. Moreover, since $\varepsilon_k \in \{0, 1\}$, one knows that $\sum_{k \in \widetilde{\Delta}_j} |\varepsilon_k^i|^r \leq 2^j$ and

$$2^{j(s+d+1/2-1/r)} \widetilde{a}_j \left(\sum_{k \in \widetilde{\Delta}_j} |\varepsilon_k^i|^r \right)^{1/r} \leq 1. \quad (50)$$

By Lemma 10, $\|\widetilde{a}_j \sum_{k \in \widetilde{\Delta}_j} \varepsilon_k^i \psi_{jk}(x)\|_{\widetilde{B}_{r,q}^{s+d}} \leq C$. Hence $h_{\varepsilon^i}^{(d)}(x) \in \widetilde{B}_{r,q}^s(A, L)$. According to Lemma 9, there exist $\{\varepsilon^0, \dots, \varepsilon^M\}$ such that $M \geq 2^{2^j-3}$ and

$$\sum_{k \in \widetilde{\Delta}_j} |\varepsilon_k^l - \varepsilon_k^i| \geq \frac{2^j}{8} = 2^{j-3}, \quad 0 \neq l \neq i \leq M. \quad (51)$$

Since $\text{supp } \psi_{jk} \cap \text{supp } \psi_{j'k'} = \emptyset$ for $k \neq k' \in \tilde{\Delta}_j$, this leads to $\|h_{\varepsilon^i}^{(d)} - h_{\varepsilon^i}^{(d)}\|_p^p \geq 2^{-js} \cdot 2^{-3} \|\psi\|_p^p$ and

$$\|h_{\varepsilon^i}^{(d)} - h_{\varepsilon^i}^{(d)}\|_p \geq 2^{-js} \cdot 8^{-1/p} \|\psi\|_p := \eta_j. \quad (52)$$

Clearly, the sets $A_{\varepsilon^i} := \{\|\tilde{f}_n^{(d)} - h_{\varepsilon^i}^{(d)}\|_p < \eta_j/2\}$, $i = 0, 1, \dots, M$, satisfy $A_{\varepsilon^i} \cap A_{\varepsilon^i} = \emptyset$ for $\forall i \neq i$. Using Fano's Lemma 10, one has

$$\sup_{0 \leq i \leq M} P_{h_{\varepsilon^i}}^n(A_{\varepsilon^i}^c) \geq \min \left\{ \frac{1}{2}, \sqrt{M} \exp(-3e^{-1} - \kappa_M) \right\}, \quad (53)$$

where $\kappa_M := \inf_{0 \leq v \leq M} (1/M) \sum_{i \neq v} K(P_{h_{\varepsilon^i}}^n, P_{h_{\varepsilon^v}}^n)$, and one can get $\kappa_M \leq (g_2/g_1)C_0^{-1}n\tilde{a}_j^2 2^j$ due to the similar arguments as (i).

Taking $2^j \sim n^{1/(2(s+d)+1)}$, then $n\tilde{a}_j^2 = n2^{-2j(s+d+1/2)} \sim 1$. One can choose a constant $C > 0$ such that $n\tilde{a}_j^2 \leq C < (g_1/32g_2)C_0$. By $M \geq 2^{2j-3}$, then one obtains

$$\sqrt{M}e^{-\kappa_M} \geq 2^{(1/2)(2^j-3)} e^{-g_2C_0^{-1}n\tilde{a}_j^2 2^j g_1^{-1}} \geq 1. \quad (54)$$

On the other hand, $\eta_j := 8^{-1/p} \|\psi\|_p 2^{-js} \sim n^{-s/(2(s+d)+1)}$ reduces to

$$\sup_{0 \leq i \leq M} E \left(\|\tilde{f}_n^{(d)} - h_{\varepsilon^i}^{(d)}\|_p \right) \geq \sup_{0 \leq i \leq M} \frac{\eta_j}{2} P_{h_{\varepsilon^i}}^n(A_{\varepsilon^i}^c) \geq C\eta_j. \quad (55)$$

Therefore, one gets the following desired result:

$$\sup_{f_X^{(d)} \in \tilde{B}_{\text{eq}}^s(A, L)} E \|\tilde{f}_n^{(d)} - f_X^{(d)}\|_p \geq n^{-s/(2(s+d)+1)}. \quad (56)$$

□

Note that $s' = s$, if $r \geq p$. Then we have the following corollary.

Corollary 12. *If $r \geq p$, the linear estimator (9) attains the optimal convergence rate.*

5. Nonlinear Estimator

In this paper, the nonlinear wavelet estimator is defined as follows:

$$\tilde{f}_{X, \text{non}}^{(d)}(y) = \sum_k \tilde{\alpha}_{j_0 k} \phi_{j_0 k}(y) + \sum_{j=j_0}^{j_1} \sum_k \tilde{\beta}_{jk}^* \psi_{jk}(y), \quad (57)$$

where

$$\begin{aligned} \tilde{\alpha}_{j_0 k} &:= (-1)^d \frac{\hat{\mu}}{n} \sum_{i=1}^n \frac{\phi_{j_0 k}^{(d)}(Y_i)}{g(Y_i)}, & \tilde{\beta}_{jk} &:= (-1)^d \frac{\hat{\mu}}{n} \sum_{i=1}^n \frac{\psi_{jk}^{(d)}(Y_i)}{g(Y_i)}, \\ \hat{\mu} &:= \frac{n}{\sum_{i=1}^n (1/g(Y_i))}. \end{aligned} \quad (58)$$

The hard thresholding wavelet coefficients are $\hat{\beta}_{jk}^* := \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq \lambda\}$, where

$$I\{|\hat{\beta}_{jk}| \geq \lambda\} := \begin{cases} 1, & |\hat{\beta}_{jk}| \geq \lambda, \\ 0, & |\hat{\beta}_{jk}| < \lambda. \end{cases} \quad \lambda = c2^{jd} \sqrt{\frac{j}{n}}. \quad (59)$$

About the wavelet coefficients, we can get the following lemmas whose proof is very similar to Lemma 6 and we omit it.

Lemma 13. *If $2^j \leq n$, then, for any $1 \leq p < \infty$, one has $E|\beta_{jk} - \hat{\beta}_{jk}|^p \leq n^{-p/2} \cdot 2^{jd p}$.*

Lemma 14 (see [3] (Bernstein inequality)). *Let X_1, X_2, \dots, X_n be independent random variables such that $|X_i| \leq M < \infty$, $E(X_i) = 0$, $b_n^2 := \sum_{i=1}^n E(X_i^2)$; then*

$$P\left(\left|\sum_{i=1}^n X_i\right| > \lambda\right) \leq 2 \exp\left(-\frac{\lambda^2}{2(b_n^2 + M\lambda/3)}\right), \quad \forall \lambda > 0. \quad (60)$$

Lemma 15. *If $j2^j \leq n$, then, for any $\omega > 0$, there exists a constant $c > 0$ such that*

$$P\left(\left|\hat{\beta}_{jk} - \beta_{jk}\right| > \lambda = c2^{jd} \sqrt{\frac{j}{n}}\right) \leq 2^{-\omega j}. \quad (61)$$

Proof. One can easily get

$$\hat{\mu} \leq g_2, \quad \mu \geq g_1, \quad \frac{1}{\mu} \leq g_1^{-1},$$

$$|\beta_{j,k}| \leq A^{1/2} \|f_X^{(d)}\|_{\infty} \leq A^{1/2} \|f_X^{(d)}\|_{B_{\text{eq}}^s},$$

$$\begin{aligned} & \left| \hat{\beta}_{j,k} - \beta_{j,k} \right| \\ &= \left| \frac{\hat{\mu}}{\mu} \left((-1)^d \frac{\mu}{n} \sum_{i=1}^n \frac{\psi_{j,k}^{(d)}(Y_i)}{g(Y_i)} - \beta_{j,k} \right) + \hat{\mu} \beta_{j,k} \left(\frac{1}{\mu} - \frac{1}{\hat{\mu}} \right) \right| \\ &\leq \left| \frac{1}{n} \sum_{i=1}^n \left((-1)^d \frac{\mu \psi_{j,k}^{(d)}(Y_i)}{g(Y_i)} - \beta_{j,k} \right) \right| \\ &\quad + \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{g(Y_i)} - \frac{1}{\mu} \right) \right| \\ &=: \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| + \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right|, \end{aligned} \quad (62)$$

where $\xi_i = (-1)^d (\mu \psi_{j,k}^{(d)}(Y_i)/g(Y_i)) - \beta_{j,k}$, $\eta_i = 1/g(Y_i) - 1/\mu$. So, one obtains that

$$\begin{aligned} & P\left(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| > \lambda\right) \\ &\leq P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| + \left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \lambda\right) \\ &\leq P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| > \frac{\lambda}{2}\right) + P\left(\left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \frac{\lambda}{2}\right). \end{aligned} \quad (63)$$

Now, we estimate $P(|(1/n) \sum_{i=1}^n \xi_i| > \lambda/2)$. Clearly, $E\xi_i = 0$ and $E\xi_i^2 \leq 2^{2jd} =: \sigma_1^2$. Moreover, one has

$$\begin{aligned} |\xi_i| &= \left| (-1)^d \frac{\mu \psi_{j,k}^{(d)}(Y_i)}{g(Y_i)} - (-1)^d E \left(\frac{\mu \psi_{j,k}^{(d)}(Y_i)}{g(Y_i)} \right) \right| \\ &\leq 2g_2 g_1^{-1} \cdot 2^{j(1/2+d)} \|\psi^{(d)}\|_\infty \\ &=: M_1 2^{j(1/2+d)}, \end{aligned} \tag{64}$$

where $M_1 := 2g_2 g_1^{-1} \|\psi^{(d)}\|_\infty$. By Bernstein's inequality and $\lambda = c2^{jd} \sqrt{j/n}$, $j2^j \leq n$, one knows that

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| > \frac{\lambda}{2}\right) &\leq 2 \exp\left(-\frac{n(\lambda/2)^2}{2(\sigma_1^2 + M_1 \cdot 2^{j(1/2+d)} \lambda/6)}\right) \\ &= 2 \exp\left(-\frac{nc^2 2^{2jd} (j/n) / 4}{2(2^{2jd} + M_1 \cdot 2^{j(1/2+d)} c 2^{jd} \sqrt{j/n/6})}\right) \\ &= 2 \exp\left(-\frac{c^2/4}{2(1 + M_1 c/6)} j\right). \end{aligned} \tag{65}$$

Taking $c_1 > 0$ such that $(c^2/4)/2(1 + M_1 c/6) \geq \omega$, then

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| > \frac{\lambda}{2}\right) \leq 2e^{-\omega j} \leq 2^{-\omega j}. \tag{66}$$

Next, we estimate $P(|(1/n) \sum_{i=1}^n \eta_i| > \lambda/2)$. It is easy to know that

$$\begin{aligned} E\eta_i &= E\left(\frac{1}{g(Y_i)}\right) - E\left(\frac{1}{\mu}\right) = \int \frac{1}{g(y)} f_Y(y) dy - \frac{1}{\mu} \\ &= \int \frac{1}{g(y)} \frac{g(y) f_X(y)}{\mu} dy - \frac{1}{\mu} \\ &= \frac{1}{\mu} \int f_X(y) dy - \frac{1}{\mu} = 0, \\ E\eta_i^2 &= E\left(\frac{1}{g(Y_i)} - \frac{1}{\mu}\right)^2 \\ &\leq 2\left(E\left|\frac{1}{g(Y_i)}\right|^2 + \frac{1}{\mu^2}\right) \leq \frac{4}{g_1^2} =: \sigma_2^2, \\ |\eta_i| &= \left|\frac{1}{g(Y_i)} - \frac{1}{\mu}\right| = \left|\frac{1}{g(Y_i)} - E\frac{1}{g(Y_i)}\right| \leq 2g_1^{-1} =: M_2. \end{aligned} \tag{67}$$

Since $j2^j \leq n$, Bernstein's inequality tells us that

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \frac{\lambda}{2}\right) &\leq 2 \exp\left(-\frac{n(\lambda/2)^2}{2(\sigma_2^2 + M_2 \lambda/3)}\right) \\ &= 2 \exp\left(-\frac{nc^2 2^{2jd} (j/n) / 4}{2(\sigma_2^2 + M_2 c 2^{jd} \sqrt{j/n/3})}\right) \\ &\leq 2 \exp\left(-\frac{c^2/4}{2(\sigma_2^2 + M_2 c/3)} j\right). \end{aligned} \tag{68}$$

Taking $c_2 > 0$ such that $(c^2/4)/2(\sigma_2^2 + M_2 c/3) \geq \omega$, one gets

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \frac{\lambda}{2}\right) \leq 2e^{-\omega j} \leq 2^{-\omega j}. \tag{69}$$

Letting $c = \max\{c_1, c_2\}$, by (66) and (69), one obtains that $P(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda) \leq 2^{-\omega j}$. \square

Lemma 16. Suppose that the parameters j_0, j_1, λ of the wavelet estimator defined in (57) satisfy the assumptions:

$$2^{j_0} \sim \begin{cases} ((\ln n)^{(p-r)/r} n)^{1/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ n^{(1-2/p)/(2(s-1/r+d)+1)}, & r \leq \frac{(1+2d)p}{2(s+d)+1}, \end{cases} \tag{70}$$

$$2^{j_1} \sim \begin{cases} n^{s/s'(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{n}{\ln n}\right)^{1/(2(s-1/r+d)+1)}, & r \leq \frac{(1+2d)p}{2(s+d)+1}; \end{cases} \tag{71}$$

then one has

$$\begin{aligned} E\left\|\sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk}\right\|_p &\leq \begin{cases} (\ln n)^{c_3} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ (\ln n)^{c_4} n^{-s'/(2(s-1/r+d)+1)}, & r = \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r < \frac{(1+2d)p}{2(s+d)+1}, \end{cases} \end{aligned} \tag{72}$$

where c_3, c_4 are positive constants.

Proof. According to Lemma 3, one obtains that

$$\begin{aligned}
 & E \left\| \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \Psi_{jk} \right\|_p \\
 & \leq E \sum_{j=j_0}^{j_1} \left\| \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \Psi_{jk} \right\|_p \\
 & \leq E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk}^* - \beta_{jk}|^p \right)^{1/p}.
 \end{aligned} \tag{73}$$

Recalling that $\widehat{\beta}_{jk}^* = \widehat{\beta}_{jk} I\{|\widehat{\beta}_{jk}| > \lambda\}$, one has

$$\begin{aligned}
 & E \left\| \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \Psi_{jk} \right\|_p \\
 & \leq E \left(\sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p \right)^{1/p} \right. \\
 & \quad \times \left(I \left\{ |\widehat{\beta}_{jk}| > \lambda, |\beta_{jk}| \geq \frac{\lambda}{2} \right\} \right. \\
 & \quad \left. \left. + I \left\{ |\widehat{\beta}_{jk}| > \lambda, |\beta_{jk}| < \frac{\lambda}{2} \right\} \right) \right. \\
 & \quad \left. + \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p \right)^{1/p} \right. \\
 & \quad \times \left(I \left\{ |\widehat{\beta}_{jk}| \leq \lambda, |\beta_{jk}| \leq 2\lambda \right\} \right. \\
 & \quad \left. \left. + I \left\{ |\widehat{\beta}_{jk}| \leq \lambda, |\beta_{jk}| > 2\lambda \right\} \right) \right).
 \end{aligned} \tag{74}$$

Note that

$$\begin{aligned}
 & I \left\{ |\widehat{\beta}_{jk}| > \lambda, |\beta_{jk}| < \frac{\lambda}{2} \right\} \leq I \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\}, \\
 & I \left\{ |\widehat{\beta}_{jk}| \leq \lambda, |\beta_{jk}| > 2\lambda \right\} \leq I \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\},
 \end{aligned} \tag{75}$$

and one gets $|\widehat{\beta}_{jk} - \beta_{jk}| \geq |\beta_{jk}| - |\widehat{\beta}_{jk}| > |\beta_{jk}|/2$; that is, $|\beta_{jk}| < 2|\widehat{\beta}_{jk} - \beta_{jk}|$, if $|\widehat{\beta}_{jk}| \leq \lambda, |\beta_{jk}| > 2\lambda$. Therefore,

$$\begin{aligned}
 & E \left\| \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \Psi_{jk} \right\|_p \\
 & \leq E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\beta_{jk}| \geq \frac{\lambda}{2} \right\} \right)^{1/p} \\
 & \quad + E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{1/p} \\
 & \quad + \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p I \left\{ |\beta_{jk}| \leq 2\lambda \right\} \right)^{1/p} \\
 & =: e_{21} + e_{22} + e_{23},
 \end{aligned} \tag{76}$$

where

$$\begin{aligned}
 e_{21} & := E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\beta_{jk}| \geq \frac{\lambda}{2} \right\} \right)^{1/p}, \\
 e_{22} & := E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \\
 & \quad \times \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{1/p}, \\
 e_{23} & := \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p I \left\{ |\beta_{jk}| \leq 2\lambda \right\} \right)^{1/p}.
 \end{aligned} \tag{77}$$

(i) Firstly, we estimate

$$e_{23} := \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p I \left\{ |\beta_{jk}| \leq 2\lambda \right\} \right)^{1/p}. \tag{78}$$

Let $\xi := (1/2)((r/p)(2s+1+2d) - (1+2d))$. By $I\{|\beta_{jk}| \leq 2\lambda\} \leq (2\lambda/|\beta_{jk}|)^{p-r}$ ($r < p$), one obtains that

$$\begin{aligned}
 e_{23} & \leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p \left(\frac{2\lambda}{|\beta_{jk}|} \right)^{p-r} \right)^{1/p} \\
 & = \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} (2\lambda)^{(p-r)/p} \left(\sum_k |\beta_{jk}|^r \right)^{1/p} \\
 & = \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} (2\lambda)^{(p-r)/p} \|\beta_j\|_r^{r/p}.
 \end{aligned} \tag{79}$$

On the other hand, $f_X^{(d)} \in \widetilde{B}_{r,q}^s(\mathbb{R})$ implies $\|\beta_j\|_r \leq 2^{-j(s+1/2-1/r)}$. Taking $\lambda = c2^{jd} \sqrt{j/n} \sim c2^{jd} \sqrt{(\ln n)/n}$, $j_1 - j_0 \sim \ln n$, one has

$$\begin{aligned}
 e_{23} & \leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \lambda^{(p-r)/p} 2^{-j(s+1/2-1/r)(r/p)} \\
 & \leq \left(\frac{\ln n}{n} \right)^{(p-r)/2p} \sum_{j=j_0}^{j_1} 2^{-j(1/2)[(r/p)(2s+1+2d) - (1+2d)]} \\
 & = \left(\frac{\ln n}{n} \right)^{(p-r)/2p} \sum_{j=j_0}^{j_1} 2^{-j\xi} \\
 & \leq \left(\frac{\ln n}{n} \right)^{(p-r)/2p} \left(2^{-j_0\xi} I\{\xi > 0\} + 2^{-j_1\xi} I\{\xi < 0\} \right. \\
 & \quad \left. + (j_1 - j_0 + 1) I\{\xi = 0\} \right).
 \end{aligned} \tag{80}$$

Note that $\xi > 0$ if and only if $r > (1+2d)p/(2(s+d)+1)$. When $\xi = 0$, that is, $p = r(2s+1+2d)/(1+2d)$, one can

compute $s'/(2(s-1/r+d)+1) = (p-r)/2p$. From (70), (71), one obtains that

$$e_{23} \leq \begin{cases} (\ln n)^{c_3} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ (\ln n)^{c_4} n^{-s'/(2(s-1/r+d)+1)}, & r = \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r < \frac{(1+2d)p}{2(s+d)+1}, \end{cases} \quad (81)$$

where c_3, c_4 are positive constants.

(ii) We estimate

$$e_{21} := E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\beta_{jk}| \geq \frac{\lambda}{2} \right\} \right)^{1/p}. \quad (82)$$

Due to Lemma 13, one has

$$E|\widehat{\beta}_{jk} - \beta_{jk}|^p \leq n^{-p/2} 2^{jd p}. \quad (83)$$

Using $I\{|\beta_{jk}| \geq \lambda/2\} \leq (|\beta_{jk}|/(\lambda/2))^r$ and Jensen's inequality, it can be proved that

$$\begin{aligned} e_{21} &= E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\beta_{jk}| \geq \frac{\lambda}{2} \right\} \right)^{1/p} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k E|\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{1/p} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k n^{-p/2} \left(\frac{|\beta_{jk}|}{\lambda/2} \right)^r \right)^{1/p} \\ &= \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} n^{-1/2} 2^{jd} \lambda^{-r/p} \|\beta_j\|_r^{r/p}. \end{aligned} \quad (84)$$

Similar to the proof of e_{23} , by $\|\beta_j\|_r \leq 2^{-j(s+1/2-1/r)}$ and $\lambda \sim c2^{jd} \sqrt{(\ln n)/n}$, $j_1 - j_0 \sim \ln n$, one gets

$$\begin{aligned} e_{21} &\leq \sum_{j=j_0}^{j_1} n^{-1/2} 2^{j(1/2-1/p+d)} 2^{-jd(r/p)} 2^{-j(s+1/2-1/r)(r/p)} \\ &\quad \times \left(\frac{n}{\ln n} \right)^{r/2p} \\ &= n^{-(p-r)/2p} (\ln n)^{-r/2p} \sum_{j=j_0}^{j_1} 2^{-j\xi}. \end{aligned} \quad (85)$$

Observing that $e_{21} \leq e_{23}$, one obtains that

$$e_{21} \leq \begin{cases} (\ln n)^{c_3} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ (\ln n)^{c_4} n^{-s'/(2(s-1/r+d)+1)}, & r = \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r < \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \quad (86)$$

(iii) Finally, we estimate

$$\begin{aligned} e_{22} &:= E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \\ &\quad \times \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{1/p}. \end{aligned} \quad (87)$$

Let $1 < q', q < \infty$ and $1/q + 1/q' = 1$. Thanks to Jensen's inequality and Höler's inequality, one has

$$\begin{aligned} e_{22} &= E \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{1/p} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k E \left(|\widehat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right) \right)^{1/p} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k (E|\widehat{\beta}_{jk} - \beta_{jk}|^{qp})^{1/q} \right. \\ &\quad \left. \times \left(E I^{q'} \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{1/q'} \right)^{1/p} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k (E|\widehat{\beta}_{jk} - \beta_{jk}|^{qp})^{1/q} \right. \\ &\quad \left. \times \left(P \left(|\widehat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right) \right)^{1/q'} \right)^{1/p}. \end{aligned} \quad (88)$$

According to Lemmas 13 and 15, one knows that

$$\begin{aligned} e_{22} &\leq \sum_{j=j_0}^{j_1} 2^{j(1/2-1/p)} \left(2^j n^{-p/2} 2^{jd p} 2^{-\omega j/q'} \right)^{1/p} \\ &= n^{-1/2} \sum_{j=j_0}^{j_1} 2^{j(1/2+d-\omega/pq')}. \end{aligned} \quad (89)$$

Choosing large enough ω such that $1/2 + d < \omega/pq'$, one gets

$$e_{22} \leq 2n^{-1/2} 2^{j_0(1/2+d-\omega/pq')} \leq n^{-1/2} 2^{j_0(1/2+d)}. \quad (90)$$

Taking 2^{j_0} as in (70), one has

$$e_{22} \leq \begin{cases} (\ln n)^{c_3} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ n^{-s'/(2(s-1/r+d)+1)}, & r \leq \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \quad (91)$$

Putting (81), (86), and (91) together, one obtains that

$$E \left\| \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p \leq \begin{cases} (\ln n)^{c_3} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ (\ln n)^{c_4} n^{-s'/(2(s-1/r+d)+1)}, & r = \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r < \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \quad (92)$$

Theorem 17. Let the scaling function $\phi(x)$ be orthonormal, compactly supported, and t -regular. If $\widehat{f}_{X,non}^{(d)}(x)$ is the nonlinear wavelet estimator defined in (57), and the assumptions (70), (71), and (81) are satisfied. Then for any $1 \leq r < p < \infty$, $1/r < s < t - d$, one has

$$\sup_{f_X^{(d)} \in \widetilde{B}_{r,q}^s(A,L)} E \left\| \widehat{f}_{X,non}^{(d)} - f_X^{(d)} \right\|_p \leq \begin{cases} (\ln n)^{\theta_1} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ (\ln n)^{\theta_2} n^{s'/(2(s-1/r+d)+1)}, & r = \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r < \frac{(1+2d)p}{2(s+d)+1}, \end{cases} \quad (93)$$

where $s' = s - 1/r + 1/p$, θ_1, θ_2 are positive constants.

Proof. By the definition of $\widehat{f}_{X,non}^{(d)}$ in (57) and the expansion of $f_X^{(d)}$ in (2), one has

$$\begin{aligned} \widehat{f}_{X,non}^{(d)} - f_X^{(d)} &= \sum_k (\widehat{\alpha}_{j_0k} - \alpha_{j_0k}) \phi_{j_0k} \\ &+ \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} + P_{j_1+1} f_X^{(d)} - f_X^{(d)}. \end{aligned} \quad (94)$$

Hence,

$$\begin{aligned} E \left\| \widehat{f}_{X,non}^{(d)} - f_X^{(d)} \right\|_p &\leq E \left\| \sum_k (\widehat{\alpha}_{j_0k} - \alpha_{j_0k}) \phi_{j_0k} \right\|_p \\ &+ E \left\| \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p + \left\| P_{j_1+1} f_X^{(d)} - f_X^{(d)} \right\|_p \\ &=: e_1 + e_2 + e_3, \end{aligned} \quad (95)$$

where

$$\begin{aligned} e_1 &:= E \left\| \sum_k (\widehat{\alpha}_{j_0k} - \alpha_{j_0k}) \phi_{j_0k} \right\|_p, \\ e_2 &:= E \left\| \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p, \\ e_3 &:= \left\| P_{j_1+1} f_X^{(d)} - f_X^{(d)} \right\|_p. \end{aligned} \quad (96)$$

Firstly, we estimate

$$e_1 := E \left\| \sum_k (\widehat{\alpha}_{j_0k} - \alpha_{j_0k}) \phi_{j_0k} \right\|_p. \quad (97)$$

Thanks to Lemma 3 and Jensen's inequality, one obtains that

$$\begin{aligned} e_1 &\leq 2^{j_0(1/2-1/p)} E \left(\sum_k |\widehat{\alpha}_{j_0k} - \alpha_{j_0k}|^p \right)^{1/p} \\ &\leq 2^{j_0(1/2-1/p)} \left(\sum_k E |\widehat{\alpha}_{j_0k} - \alpha_{j_0k}|^p \right)^{1/p}. \end{aligned} \quad (98)$$

Due to Lemma 6, one has $E |\widehat{\alpha}_{j_0k} - \alpha_{j_0k}|^p \leq n^{-p/2} 2^{j_0 d p}$. Since $f_X^{(d)}(x)$ and $\phi(x)$ are compactly supported, then the number of elements in $\{k : \alpha_{j_0k} \neq 0\}$ is $O(2^{j_0})$. Therefore,

$$e_1 \leq 2^{j_0(1/2-1/p)} (2^{j_0} n^{-p/2} 2^{j_0 d p})^{1/p} = n^{-1/2} 2^{j_0(1/2+d)}. \quad (99)$$

From (70), one has

$$e_1 \leq \begin{cases} (\ln n)^{c_3} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ n^{-(s'+2d/p)/(2(s-1/r+d)+1)}, & r \leq \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \quad (100)$$

Next, we estimate term

$$e_3 := \left\| P_{j_1+1} f_X^{(d)} - f_X^{(d)} \right\|_p. \quad (101)$$

When $r < p$, $s' = s - 1/r + 1/p$, one has $\widetilde{B}_{r,q}^s \hookrightarrow \widetilde{B}_{p,q}^{s'}$, thanks to the Besov spaces embedding theorem. By Lemma 4, one has

$$\left\| P_{j_1+1} f_X^{(d)} - f_X^{(d)} \right\|_p \leq 2^{-j_1 s'}. \quad (102)$$

Taking 2^{j_1} as in (71), it can be proved that

$$e_3 \leq \begin{cases} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r \leq \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \quad (103)$$

Finally, we estimate

$$e_2 := E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p. \tag{104}$$

Using Lemma 16, one has

$$E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p \leq \begin{cases} (\ln n)^{c_3} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ (\ln n)^{c_4} n^{-s'/(2(s-1/r+d)+1)}, & r = \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r < \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \tag{105}$$

By (100), (103), and (105), one obtains that

$$\sup_{f_X^{(d)} \in \tilde{B}_{p,q}^s(A,L)} E \left\| \hat{f}_{X,\text{non}}^{(d)} - f_X^{(d)} \right\|_p \leq \begin{cases} (\ln n)^{\theta_1} n^{-s/(2(s+d)+1)}, & r > \frac{(1+2d)p}{2(s+d)+1}, \\ (\ln n)^{\theta_2} n^{s'/(2(s-1/r+d)+1)}, & r = \frac{(1+2d)p}{2(s+d)+1}, \\ \left(\frac{\ln n}{n}\right)^{s'/(2(s-1/r+d)+1)}, & r < \frac{(1+2d)p}{2(s+d)+1}. \end{cases} \tag{106}$$

□

Remark 18. Since $r < p$ implies that $s' = s - 1/r + 1/p$, one can easily find $s'/(2(s - 1/r + d) + 1) > s'/(2(s' + d) + 1)$. Then, Theorems 17 and 7 tell us that the nonlinear estimator does better than the linear one. Moreover, Theorems 17 and 11 show that our nonlinear one attains the optimal. coverage rate

Corollary 19. *When $r < p$, the nonlinear estimator (57) attains the optimal coverage rate up to a $\ln n$ factor.*

Remark 20. If $d = 0, r = p = 2$, our estimation reduces to Ramírez and Vidaković's result [14].

Remark 21. If $d = 0$, our results can be considered as the natural extension of paper [10].

Remark 22. If $g(x) \equiv 1$, model (1) reduces to the samples without any noise; our results reduce to the estimations in [13]. Even for $p/(2(s+d)+1) < r < p$, our result is also optimal coverage rate.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper is supported by CSC Foundation and Fundamental Research Fund of BJUT.

References

- [1] X. J. Yang, D. Baleanu, H. M. Srivastava, and M. J. A. Tenreiro, "On local fractional continuous wavelet transform," *Abstract and Applied Analysis*, vol. 2013, Article ID 725416, 5 pages, 2013.
- [2] Y. Zhao, D. Baleanu, C. Cattani, D. F. Cheng, and X. J. Yang, "Local fractional discrete wavelet transform for solving signal on cantor sets," *Mathematical Problems in Engineering*, vol. 2013, Article ID 560932, 6 pages, 2013.
- [3] W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov, *Wavelets, Approximation and Statistical Applications*, Springer, New York, NY, USA, 1997.
- [4] B. L. S. P. Rao, "Nonparametric estimation of the derivatives of a density by the method of wavelets," *Bulletin of Informatics and Cybernetics*, vol. 28, no. 1, pp. 91-100, 1996.
- [5] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer, Berlin, Germany, 2009.
- [6] K. Tribouley, "Practical estimation of multivariate densities using wavelet methods," *Statistica Neerlandica*, vol. 49, no. 1, pp. 41-62, 1995.
- [7] B. Vidakovic, *Statistical Modeling by Wavelets*, John Wiley & Sons, New York, NY, USA, 1999.
- [8] G. Walter and J. Ghorai, "Advantages and disadvantages of density estimation with wavelets," in *Proceedings of the 24th Symposium on the Interface*, vol. 24, pp. 234-243, 1992.
- [9] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, "Density estimation by wavelet thresholding," *The Annals of Statistics*, vol. 24, no. 2, pp. 508-539, 1996.
- [10] C. Chesneau, "Wavelet block thresholding for density estimation in the presence of bias," *Journal of the Korean Statistical Society*, vol. 39, no. 1, pp. 43-53, 2010.
- [11] Y. P. Chaubey, H. Doosti, and B. L. S. Prakasa Rao, "Wavelet based estimation of the derivatives of a density for a negatively associated process," *Journal of Statistical Theory and Practice*, vol. 2, no. 3, pp. 453-463, 2008.
- [12] G. Kerkycharian and D. Picard, "Density estimation in Besov spaces," *Statistics & Probability Letters*, vol. 13, no. 1, pp. 15-24, 1992.
- [13] Y. Liu and H. Wang, "Wavelet estimations for density derivatives," *Science China Mathematics*, vol. 56, no. 3, pp. 483-495, 2013.
- [14] P. Ramírez and B. Vidakovic, "Wavelet density estimation for stratified size-biased sample," *Journal of Statistical Planning and Inference*, vol. 140, no. 2, pp. 419-432, 2010.
- [15] S. Efromovich, *Nonparametric Curve Estimation. Methods, Theory, and Applications*, Springer Series in Statistics, Springer, New York, NY, USA, 1999.
- [16] Y. Vardi, "Nonparametric estimation in the presence of length bias," *The Annals of Statistics*, vol. 10, no. 2, pp. 616-620, 1982.
- [17] M. C. Jones, "Kernel density estimation for length biased data," *Biometrika*, vol. 78, no. 3, pp. 511-519, 1991.
- [18] S. Efromovich, "Density estimation for biased data," *The Annals of Statistics*, vol. 32, no. 3, pp. 1137-1161, 2004.
- [19] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 1992.

- [20] S. E. Kelly, M. A. Kon, and L. A. Raphael, "Local convergence for wavelet expansions," *Journal of Functional Analysis*, vol. 126, no. 1, pp. 102–138, 1994.
- [21] G. Beylkin, "Wavelets and fast numerical algorithms," *Proceedings of Symposia in Applied Mathematics*, vol. 47, pp. 89–117, 1993.
- [22] G. Beylkin, "On the representation of operators in bases of compactly supported wavelets," *SIAM Journal on Numerical Analysis*, vol. 29, no. 6, pp. 1716–1740, 1992.