

RESEARCH

Open Access

Nonlinear wavelet density estimation for biased data in Sobolev spaces

Jinru Wang*, Meng Wang and Yuan Zhou

*Correspondence:
wangjinru@bjut.edu.cn
Department of Applied
Mathematics, Beijing University of
Technology, Beijing, 100124, China

Abstract

In this paper, we consider the density estimation problem from independent and identically distributed (i.i.d.) biased observations. We develop an adaptive wavelet hard thresholding rule and evaluate its performance by considering L_p risk over Sobolev balls. We prove that our estimation attains a sharp rate of convergence and show the optimality.

MSC: 49K40; 90C29; 90C31

Keywords: Sobolev spaces; wavelet density estimation; hard thresholding; convergence rate

1 Introduction

In practice, it usually happens that drawing a direct sample from a random variable X is impossible. In this paper, we consider the problem of estimating the density functions $f^X(x)$ without observing directly the i.i.d. sample X_1, X_2, \dots, X_n . We observe the samples Y_1, Y_2, \dots, Y_n from biased data with the following density function:

$$f^Y(x) = \frac{g(x)f^X(x)}{\mu}, \quad (1.1)$$

where $g(x)$ is the so-called weight or bias function, $\mu = E(g(X))$. The purpose of this paper is to estimate the density function $f^X(x)$ from the samples Y_1, Y_2, \dots, Y_n .

Several examples of this biased data can be found in the literature. For instance, in paper [1], it is shown that the distribution of the concentration of alcohol in the blood of intoxicated drivers is of interest, since the drunken driver has a larger chance of being arrested, the collected data are size-biased.

The density estimation problem for biased data (1.1) has been discussed in several papers. In 1982, Vardi [2] considered the nonparametric maximum likelihood estimation for $f^X(x)$. In 1991, Jones [3] discussed the mean squared error properties of the kernel density estimation. In 2004, Efromovich [4] developed the Efromovich-Pinsker adaptive Fourier estimator. It was based on a blockwise shrinkage algorithm and achieved the minimax rate of convergence under the L_2 risk over a Besov class $B_{2,2}^s$.

In 2010, Ramírez and Vidakovic [5] proposed a linear wavelet estimator and discussed the consistency of a function in $L_2[0,1]$ under the mean integrated squared error (MISE) sense. But the wavelet estimator in paper [5] contained the unknown parameter μ . In the same year, Christophe [6] constructed a nonlinear wavelet estimator and evaluated the L_p

risk in the Besov space $B_{r,q}^s$. However, Sobolev spaces W_r^N ($N \in \mathbb{N}^+$) except $r = 2$ is not a special case in the Besov space $B_{r,q}^s$.

In this paper, we consider the nonlinear hard thresholding wavelet density estimation for biased data in Sobolev spaces W_r^N ($N \in \mathbb{N}^+$). We mainly give the upper bound of minimax rate of convergence under the L_p risk without particular restriction on the parameters r and p , and the convergence rate is optimal.

2 Preliminaries

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

2.1 Wavelets

In this paper, we always assume that the scaling wavelet φ is orthonormal, compactly supported and $N + 1$ regular.

Definition 2.1 The scaling function $\varphi(x)$ is called m regular if $\varphi(x)$ has continuous derivatives of order m and its corresponding wavelet $\psi(x)$ has vanishing moments of order m , i.e., $\int x^k \psi(x) dx = 0, k = 0, 1, \dots, m - 1$.

The following conditions about the scaling function φ and the kernel function $K(x, y)$ will be very useful in the third section.

Condition (θ) The function $\theta_\varphi(x) = \sum_{k \in \mathbb{Z}} |\varphi(x - k)|$ is such that $\text{ess sup}_{x \in \mathbb{R}} \theta_\varphi(x) < \infty$.

Condition $H(N)$ There exists an integrable function $F(x)$ such that for any $x, y \in \mathbb{R}$, $|K(x, y)| \leq F(x - y)$, where $\int |x|^N F(x) dx < \infty$.

Condition $M(N)$ Condition $H(N)$ is satisfied and $\int K(x, y)(y - x)^k dy = \delta_{0k}, k = 0, \dots, N, x \in \mathbb{R}$.

For any $x \in \mathbb{R}, j, k \in \mathbb{Z}$, denoted by $\varphi_{jk}(x) := 2^{\frac{j}{2}} \varphi(2^j x - k), \psi_{jk}(x) := 2^{\frac{j}{2}} \psi(2^j x - k)$, then for any $f(x) \in L_r(\mathbb{R}) := \{f(x) | \int_{\mathbb{R}} |f(x)|^r dx < \infty\}$, where $1 \leq r < \infty$, we have the following equation [7]:

$$f(x) = \sum_{k \in \mathbb{Z}} \alpha_{J,k} \varphi_{J,k}(x) + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} \beta_{j,k} \psi_{j,k}(x), \quad \text{a.e.,} \tag{2.1}$$

where

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

2.2 Sobolev space

The Sobolev space $W_r^N(\mathbb{R})$ ($N \in \mathbb{N}^+$) is defined by $W_r^N(\mathbb{R}) := \{f : f \in L_r(\mathbb{R}), f^{(N)} \in L_r(\mathbb{R})\}$, which is equipped with the norm $\|f\|_{W_r^N} := \|f\|_r + \|f^{(N)}\|_r$. The Sobolev balls $\tilde{W}_r^N(A, L)$ are defined as follows:

$$\tilde{W}_r^N(A, L) := \{f \in W_r^N(\mathbb{R}) : f \text{ is a probability density function, } \text{supp} f \leq A, \|f^{(N)}\|_r \leq L\}.$$

Between a Sobolev space and a Besov space, the following embedding conclusions are established.

Lemma 2.1 [8] *Let $s > 0, 1 \leq p, q, r \leq \infty$, then*

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

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

2.3 Auxiliary lemmas

The following lemmas given by [9] will be used in the next section.

Lemma 2.2 *If the scaling function φ satisfies Condition (θ) , then for any sequence $\{\lambda_k\}_{k \in \mathbb{Z}}$ satisfying $\|\lambda\|_{l_p} := (\sum_k |\lambda_k|^p)^{1/p} < \infty$, we have $C_1 \|\lambda\|_{l_p} 2^{(\frac{j}{2}-\frac{j}{p})} \leq \|\sum_k \lambda_k \varphi_{j,k}\|_p \leq C_2 \|\lambda\|_{l_p} 2^{(\frac{j}{2}-\frac{j}{p})}$, where $C_1 = (\|\theta_\varphi\|_\infty^{\frac{1}{p}} \|\varphi\|_1^{\frac{1}{q}})^{-1}$, $C_2 = (\|\theta_\varphi\|_\infty^{\frac{1}{q}} \|\varphi\|_1^{\frac{1}{p}})^{-1}$, $1 \leq p \leq \infty, \frac{1}{p} + \frac{1}{q} = 1$.*

Lemma 2.3 *For some integer $N \geq 0$, if the kernel function $K(x, y)$ satisfies Conditions $M(N)$ and $H(N+1), f \in B_{pq}^s(\mathbb{R})$, where $1 \leq p, q \leq \infty, 0 < s < N+1$, then we have $\|K_j f - f\|_p = 2^{-js} \varepsilon_j$, where $\varepsilon_j \in l_q$.*

Lemma 2.4 (Rosenthal inequality) *Let X_1, \dots, X_n be independent random variables such that $E(X_i) = 0$ and $E(|X_i|^p) < \infty$, then there exists a constant $C(p) > 0$ such that*

$$E\left(\left|\sum_{i=1}^n X_i\right|^p\right) \leq C(p) \left(\sum_{i=1}^n E(|X_i|^p) + \left(\sum_{i=1}^n E(X_i^2)\right)^{p/2}\right), \quad p > 2,$$

$$E\left(\left|\sum_{i=1}^n X_i\right|^p\right) \leq \left(\sum_{i=1}^n E(X_i^2)\right)^{p/2}, \quad 0 < p \leq 2.$$

Lemma 2.5 (Bernstein inequality) *Let X_1, X_2, \dots, X_n be independent random variables such that $E(X_i) = 0, E(X_i^2) \leq \sigma^2, |X_i| \leq M < \infty$. Then*

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

Remark In this paper, we often use the notation $A \lesssim B$ to indicate 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 $A \sim B$.

3 Main results

In this paper, our hard thresholding wavelet density estimator is defined as follows:

$$\hat{f}_n^{X\text{non}}(x) = \sum_k \hat{\alpha}_{j_0 k} \varphi_{j_0 k}(x) + \sum_{j=j_0}^{\hat{j}} \sum_k \hat{\beta}_{jk}^* \psi_{jk}(x), \tag{3.1}$$

where

$$\hat{\alpha}_{j_0 k} := \frac{\hat{\mu}}{n} \sum_{i=1}^n \frac{\varphi_{j_0 k}(Y_i)}{g(Y_i)}, \quad \hat{\beta}_{jk} := \frac{\hat{\mu}}{n} \sum_{i=1}^n \frac{\psi_{jk}(Y_i)}{g(Y_i)}, \quad \hat{\mu} := \frac{n}{\sum_{i=1}^n \frac{1}{g(Y_i)}}.$$

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

$$I\{|\hat{\beta}_{jk}| \geq \lambda\} := \begin{cases} 1, & |\hat{\beta}_{jk}| \geq \lambda, \\ 0, & |\hat{\beta}_{jk}| < \lambda. \end{cases}$$

Suppose that the parameters j_0, j_1, λ of the wavelet thresholding estimator (3.1) satisfy the assumptions:

$$2^{j_0} \sim \begin{cases} ((\ln n)^{\frac{p-r}{r}} n)^{\frac{1}{2N+1}}, & r > \frac{p}{2N+1}, \\ n^{\frac{1-2/p}{2(N-1/r)+1}}, & r \leq \frac{p}{2N+1}, \end{cases} \quad (3.2)$$

$$2^{j_1} \sim \begin{cases} n^{\frac{N}{N'(2N+1)}}, & r > \frac{p}{2N+1}, \\ (\frac{n}{\ln n})^{\frac{1}{2(N-1/r)+1}}, & r \leq \frac{p}{2N+1}, \end{cases} \quad (3.3)$$

$$\lambda = c \sqrt{\frac{j}{n}}, \quad (3.4)$$

where c is a suitably chosen positive constant.

Lemma 3.1 *Suppose that there exist two constants g_1 and g_2 such that $0 < g_1 \leq g(x) \leq g_2 < \infty$ for $x \in \mathbb{R}$. Let α_{jk}, β_{jk} be the coefficients in the expansion (2.1) and let $\hat{\alpha}_{jk}, \hat{\beta}_{jk}$ be defined by estimator in (3.1). If $2^j \leq n$, then for any $1 \leq p < \infty$, we have*

- (i) $E|\alpha_{jk} - \hat{\alpha}_{jk}|^p \lesssim n^{-\frac{p}{2}}$;
- (ii) $E|\beta_{jk} - \hat{\beta}_{jk}|^p \lesssim n^{-\frac{p}{2}}$.

Proof (i) From the definition of $\hat{\alpha}_{jk}$ and the triangular inequality, we have

$$\begin{aligned} |\hat{\alpha}_{j,k} - \alpha_{j,k}| &= \left| \frac{\hat{\mu}}{n} \sum_{i=1}^n \frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k} \right| \\ &= \left| \frac{\hat{\mu}}{\mu} \left(\frac{\mu}{n} \sum_{i=1}^n \frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k} \right) + \hat{\mu} \alpha_{j,k} \left(\frac{1}{\mu} - \frac{1}{\hat{\mu}} \right) \right| \\ &\leq \left| \frac{\hat{\mu}}{\mu} \right| \left| \frac{\mu}{n} \sum_{i=1}^n \frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k} \right| + |\hat{\mu} \alpha_{j,k}| \left| \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right|. \end{aligned}$$

Since $g_1 \leq g(y) \leq g_2$, we have

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

and

$$|\alpha_{j,k}| \leq \int |f^X(y)| |\varphi_{j,k}(y)| dy \leq \left(\int |f^X(y)|^2 dy \right)^{\frac{1}{2}} \left(\int |\varphi_{j,k}(y)|^2 dy \right)^{\frac{1}{2}} \leq A^{1/2} \|f\|_{\infty}.$$

Furthermore, a Sobolev space and a Besov space have the following embedding theorem, $W_r^N \hookrightarrow B_{r\infty}^N \hookrightarrow B_{\infty\infty}^{N-1/r}$, for any integer $N > 1/r$, then we have $\|f\|_{\infty} \leq \|f\|_{B_{\infty\infty}^{N-1/r}} \leq$

$\|f\|_{W_r^N} = c$. Therefore, by the convexity inequality, we get

$$\begin{aligned} & E|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p \\ & \leq E\left(\frac{g_2}{g_1}\left|\frac{\mu}{n}\sum_{i=1}^n\frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k}\right| + g_2A^{1/2}c\left|\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right|\right)^p \\ & \leq 2^{p-1}\max\left\{\frac{g_2}{g_1}, g_2A^{1/2}c\right\}^p E\left(\left|\frac{\mu}{n}\sum_{i=1}^n\frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k}\right|^p + \left|\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right|^p\right) \\ & \lesssim E\left|\frac{\mu}{n}\sum_{i=1}^n\frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k}\right|^p + E\left|\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right|^p \\ & =: T_1 + T_2, \end{aligned}$$

where $T_1 := E\left|\frac{\mu}{n}\sum_{i=1}^n\frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k}\right|^p$, $T_2 := E\left|\frac{1}{\hat{\mu}} - \frac{1}{\mu}\right|^p$.

The term T_i is estimated as follows. Firstly, let $\xi_i := \mu\frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k}$, we can see that they are i.i.d., and $E(\xi_i) = 0$. Moreover, for any $m \geq 2$,

$$\begin{aligned} E|\xi_i|^m & = E\left|\mu\frac{\varphi_{j,k}(Y_i)}{g(Y_i)} - \alpha_{j,k}\right|^m \\ & \leq 2^{m-1}\left(E\left|\mu\frac{\varphi_{j,k}(Y_i)}{g(Y_i)}\right|^m + |\alpha_{j,k}|^m\right), \end{aligned}$$

where

$$\begin{aligned} E\left|\mu\frac{\varphi_{j,k}(Y_i)}{g(Y_i)}\right|^m & = \mu^{m-1}\frac{\varphi_{j,k}^{m-2}(Y_i)}{g(Y_i)^{m-1}}E\left|\mu\frac{\varphi_{j,k}^2(Y_i)}{g(Y_i)}\right| \\ & \leq g_2^{m-1}g_1^{-m+1}2^{\frac{j}{2}(m-2)}\|\varphi\|_\infty^{m-2}E\left|\mu\frac{\varphi_{j,k}^2(Y_i)}{g(Y_i)}\right|, \end{aligned}$$

and

$$\begin{aligned} E\left|\mu\frac{\varphi_{j,k}^2(Y_i)}{g(Y_i)}\right| & = \int \mu\frac{\varphi_{j,k}^2(y)}{g(y)}f^Y(y)dy = \int \mu\frac{\varphi_{j,k}^2(y)}{g(y)}\frac{g(y)f^X(y)}{\mu}dy \\ & \leq \|f\|_\infty \leq \|f\|_{W_r^N} = c. \end{aligned}$$

So, we have

$$E\left|\mu\frac{\varphi_{j,k}(Y_i)}{g(Y_i)}\right|^m \leq g_2^{m-1}g_1^{-m+1}2^{\frac{j}{2}(m-2)}\|\varphi\|_\infty^{m-2}c.$$

Since $2^j \leq n$, we obtain

$$E|\xi_i|^m \leq C2^{j(m-2)/2} \lesssim n^{(m-2)/2}.$$

By Rosenthal's inequality, we have

$$T_1 = E\left|\frac{1}{n}\sum_{i=1}^n\xi_i\right|^p \lesssim n^{-p}(nE|\xi_i|^p + n^{p/2}(E|\xi_i|^2)^{p/2}) \lesssim n^{-p/2}. \tag{3.5}$$

To estimate the term T_2 , let $\eta_i = \frac{1}{g(Y_i)} - \frac{1}{\mu}$. We can compute $E(\eta_i) = 0$ easily, and for any $m \geq 2$, $E|\eta_i|^m \leq C$.

If $p \geq 2$, i.e., $1 - p < -p/2$, using Rosenthal's inequality, we have

$$T_2 = E \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right|^p \lesssim n^{-p} (nE|\eta_i|^p + n^{p/2} (E|\eta_i|^2)^{p/2}) \lesssim n^{-p+1} + n^{-p/2} \lesssim n^{-p/2}. \tag{3.6}$$

If $1 \leq p < 2$, we get

$$T_2 = E \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right|^p \leq n^{-p} (n^{p/2} (E|\eta_i|^2)^{p/2}) \leq n^{-p/2}. \tag{3.7}$$

By (3.5), (3.6) and (3.7), we obtain

$$E|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p \lesssim T_1 + T_2 \lesssim n^{-p/2}.$$

(ii) It is similar to (i), we omit it. □

Lemma 3.2 *If $j^2 \leq n$, then for any $\omega > 0$, there exists a constant $c > 0$ such that*

$$P\left(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda = c\sqrt{\frac{j}{n}}\right) \lesssim 2^{-\omega j}. \tag{3.8}$$

Proof We can easily get

$$\begin{aligned} \hat{\mu} &\leq g_2, & \mu &\geq g_1, & \frac{1}{\mu} &\leq g_1^{-1}, \\ |\beta_{j,k}| &\leq A^{1/2} \|f\|_{\infty} \leq A^{1/2} \|f\|_{W_r^N}. \end{aligned}$$

Therefore,

$$\begin{aligned} |\hat{\beta}_{j,k} - \beta_{j,k}| &= \left| \frac{\hat{\mu}}{\mu} \left(\frac{\mu}{n} \sum_{i=1}^n \frac{\psi_{j,k}(Y_i)}{g(Y_i)} - \beta_{j,k} \right) + \hat{\mu} \beta_{j,k} \left(\frac{1}{\mu} - \frac{1}{\hat{\mu}} \right) \right| \\ &\leq \frac{g_2}{g_1} \left| \frac{1}{n} \sum_{i=1}^n \left(\mu \frac{\psi_{j,k}(Y_i)}{g(Y_i)} - \beta_{j,k} \right) \right| + g_2 A^{1/2} \|f^X\|_{W_r^N} \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{g(Y_i)} - \frac{1}{\mu} \right) \right| \\ &=: \frac{g_2}{g_1} \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| + g_2 A^{1/2} \|f^X\|_{W_r^N} \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right|, \end{aligned}$$

where $\xi_i = \mu \frac{\psi_{j,k}(Y_i)}{g(Y_i)} - \beta_{j,k}$, $\eta_i = \frac{1}{g(Y_i)} - \frac{1}{\mu}$. So, we get

$$\begin{aligned} P(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda) &\leq P\left(\frac{g_2}{g_1} \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| + g_2 A^{1/2} \|f^X\|_{W_r^N} \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right| > \lambda\right) \\ &\leq P\left(\frac{g_2}{g_1} \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > \lambda/2\right) + P\left(g_2 A^{1/2} \|f^X\|_{W_r^N} \left| \frac{1}{n} \sum_{i=1}^n \eta_i \right| > \lambda/2\right) \end{aligned}$$

$$\begin{aligned}
 &= P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| > \frac{\lambda g_1}{2g_2}\right) + P\left(\left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \frac{\lambda}{2g_2 A^{1/2} \|f^X\|_{W_r^N}}\right) \\
 &=: P_1 + P_2,
 \end{aligned}$$

where $P_1 := P(|\frac{1}{n} \sum_{i=1}^n \xi_i| > \frac{\lambda g_1}{2g_2})$, $P_2 := P(|\frac{1}{n} \sum_{i=1}^n \eta_i| > \frac{\lambda}{2g_2 A^{1/2} \|f^X\|_{W_r^N}})$.

Now, we estimate P_1 . Clearly, $E\xi_i = 0$, and

$$\begin{aligned}
 E\xi_i^2 &= E\left(\mu \frac{\psi_{j,k}(Y_i)}{g(Y_i)} - \beta_{j,k}\right)^2 \\
 &\leq 2\left(E\left|\mu \frac{\psi_{j,k}(Y_i)}{g(Y_i)}\right|^2 + \beta_{j,k}^2\right) \\
 &= 2\left(E\left|\mu \frac{\psi_{j,k}^2(Y_i)}{g(Y_i)}\right| \frac{\mu}{g(Y_i)} + \beta_{j,k}^2\right) \\
 &\leq 2\left(\frac{g_2}{g_1} \|f^X\|_{W_r^N} + A \|f^X\|_{W_r^N}^2\right) \\
 &:= \sigma^2.
 \end{aligned}$$

Furthermore, we have

$$|\xi_i| = \left| \mu \frac{\psi_{j,k}(Y_i)}{g(Y_i)} - E\left(\mu \frac{\psi_{j,k}(Y_i)}{g(Y_i)}\right) \right| \leq 2 \cdot 2^{j/2} g_2 g_1^{-1} \|\psi\|_{\infty}.$$

By Bernstein's inequality, we obtain

$$\begin{aligned}
 P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| > \frac{\lambda g_1}{2g_2}\right) &\leq 2 \exp\left(-\frac{n \lambda^2 g_1^2 / 4g_2^2}{2(\sigma^2 + \frac{\lambda g_1}{2g_2} \cdot 2^{j/2} g_2 g_1^{-1} \|\psi\|_{\infty} / 3)}\right) \\
 &= 2 \exp\left(-\frac{n \cdot c^2 \frac{j}{n} \cdot g_1^2 / 4g_2^2}{2(\sigma^2 + g_2 c \sqrt{j/n} 2^{j/2} g_2 g_1^{-1} \|\psi\|_{\infty} g_1 / 3)}\right) \\
 &= 2 \exp\left(-\frac{c^2 j g_1^2 / 4g_2^2}{2(\sigma^2 + \sqrt{j 2^j / n} \|\psi\|_{\infty} c / 3)}\right).
 \end{aligned}$$

Since $j 2^j \leq n$, then

$$\begin{aligned}
 P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| > \frac{\lambda g_1}{2g_2}\right) &\leq 2 \exp\left(-\frac{c^2 j g_1^2 / 4g_2^2}{2(\sigma^2 + \|\psi\|_{\infty} c / 3)}\right) \\
 &= 2 \exp\left(-\frac{c^2 g_1^2 / 4g_2^2}{2(\sigma^2 + \|\psi\|_{\infty} c / 3)} j\right).
 \end{aligned}$$

Taking $c_1 > 0$ such that $\frac{c^2 g_1^2 / 4g_2^2}{2(\sigma^2 + \|\psi\|_{\infty} c_1 / 3)} \geq \omega$, then

$$P_1 = P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| > \frac{\lambda g_1}{2g_2}\right) \leq 2e^{-\omega j} \lesssim 2^{-\omega j}. \tag{3.9}$$

Next, we estimate P_2 . We compute that $E\eta_i = 0$, i.e.,

$$\begin{aligned} E\eta_i &= E\left(\frac{1}{g(Y_i)}\right) - E\left(\frac{1}{\mu}\right) = \int \frac{1}{g(y)} f^Y(y) dy - \frac{1}{\mu} \\ &= \int \frac{1}{g(y)} \frac{g(y) f^X(y)}{\mu} dy - \frac{1}{\mu} \\ &= \frac{1}{\mu} \int f^X(y) dy - \frac{1}{\mu} = 0, \end{aligned}$$

and

$$\begin{aligned} E\eta_i^2 &= E\left(\frac{1}{g(Y_i)} - \frac{1}{\mu}\right)^2 \leq 2\left(E\left|\frac{1}{g(Y_i)}\right|^2 + \frac{1}{\mu^2}\right) \leq \frac{4}{g_1^2}, \\ |\eta_i| &= \left|\frac{1}{g(Y_i)} - \frac{1}{\mu}\right| = \left|\frac{1}{g(Y_i)} - E\frac{1}{g(Y_i)}\right| \leq 2g_1^{-1}. \end{aligned}$$

By Bernstein's inequality, we obtain

$$\begin{aligned} &P\left(\left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \frac{\lambda}{2g_2 A^{1/2} \|f^X\|_{W_r^N}}\right) \\ &\leq 2 \exp\left(-\frac{n(\lambda/(2g_2 A^{1/2} \|f^X\|_{W_r^N}))^2}{2(\frac{4}{g_1^2} + \lambda g_1^{-1}/(3g_2 A^{1/2} \|f^X\|_{W_r^N}))}\right) \\ &= 2 \exp\left(-\frac{nc^2 j/(4ng_2^2 A \|f^X\|_{W_r^N}^2)}{2(\frac{4}{g_1^2} + c\sqrt{j/n}/(3g_1 g_2 A^{1/2} \|f^X\|_{W_r^N}))}\right). \end{aligned}$$

Since $j \leq n$, then

$$\begin{aligned} &P\left(\left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \frac{\lambda}{2g_2 A^{1/2} \|f^X\|_{W_r^N}}\right) \\ &\leq 2 \exp\left(-\frac{c^2/(4g_2^2 A \|f^X\|_{W_r^N}^2)}{2(\frac{4}{g_1^2} + c/(3g_1 g_2 A^{1/2} \|f^X\|_{W_r^N}))} j\right). \end{aligned}$$

Taking $c_2 > 0$ such that $\frac{c_2^2/(4g_2^2 A \|f^X\|_{W_r^N}^2)}{2(\frac{4}{g_1^2} + c_2/(3g_1 g_2 A^{1/2} \|f^X\|_{W_r^N}))} \geq \omega$, we have

$$P_2 = P\left(\left|\frac{1}{n} \sum_{i=1}^n \eta_i\right| > \frac{\lambda}{2g_2 A^{1/2} \|f^X\|_{W_r^N}}\right) \leq 2e^{-\omega j} \lesssim 2^{-\omega j}. \tag{3.10}$$

Taking $c = \max\{c_1, c_2\}$, by (3.9) and (3.10), we have

$$P(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda) \leq P_1 + P_2 \lesssim 2^{-\omega j}.$$

□

Lemma 3.3 *Suppose that there exist two constants g_1 and g_2 such that $0 < g_1 \leq g(x) \leq g_2 < \infty$, for $x \in \mathbb{R}$, and $\hat{\beta}_{jk}, \hat{\beta}_{jk}^*$ are given by (3.1). Then*

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

where c_3, c_4 are constants.

Proof By Lemma 2.2, we obtain

$$\begin{aligned} E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p &\leq E \sum_{j=j_0}^{j_1} \left\| \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p \\ &\lesssim E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk}^* - \beta_{jk}|^p \right)^{\frac{1}{p}}. \end{aligned}$$

Furthermore, since $\hat{\beta}_{jk}^* = \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| > \lambda\}$, we have

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

Note that

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

and if $|\hat{\beta}_{jk}| \leq \lambda, |\beta_{jk}| > 2\lambda$, we get $|\hat{\beta}_{jk} - \beta_{jk}| \geq |\beta_{jk}| - |\hat{\beta}_{jk}| > \frac{|\beta_{jk}|}{2}$, i.e., $|\beta_{jk}| < 2|\hat{\beta}_{jk} - \beta_{jk}|$; therefore, we have

$$\begin{aligned} E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p &\lesssim E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I\left\{|\beta_{jk}| \geq \frac{\lambda}{2}\right\} \right)^{\frac{1}{p}} \\ &\quad + E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I\left\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2}\right\} \right)^{\frac{1}{p}} \end{aligned}$$

$$\begin{aligned}
 & + \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\beta_{jk}|^p I\{|\beta_{jk}| \leq 2\lambda\} \right)^{\frac{1}{p}} \\
 & =: W_1 + W_2 + W_3,
 \end{aligned}$$

where

$$\begin{aligned}
 W_1 & := E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I\left\{|\beta_{jk}| \geq \frac{\lambda}{2}\right\} \right)^{\frac{1}{p}}, \\
 W_2 & := E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I\left\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2}\right\} \right)^{\frac{1}{p}}, \\
 W_3 & := \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\beta_{jk}|^p I\{|\beta_{jk}| \leq 2\lambda\} \right)^{\frac{1}{p}}.
 \end{aligned}$$

(i) Firstly, we estimate

$$W_1 := E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I\left\{|\beta_{jk}| \geq \frac{\lambda}{2}\right\} \right)^{\frac{1}{p}}.$$

By Lemma 3.1, we have

$$E|\hat{\beta}_{jk} - \beta_{jk}|^p \lesssim n^{-\frac{p}{2}}.$$

Using $I\{|\beta_{jk}| \geq \frac{\lambda}{2}\} \leq \left(\frac{|\beta_{jk}|}{\frac{\lambda}{2}}\right)^r$ and Jensen's inequality, we obtain

$$\begin{aligned}
 W_1 & = E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I\left\{|\beta_{jk}| \geq \frac{\lambda}{2}\right\} \right)^{\frac{1}{p}} \\
 & \leq \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k E|\hat{\beta}_{jk} - \beta_{jk}|^p I\left\{|\beta_{jk}| > \frac{\lambda}{2}\right\} \right)^{\frac{1}{p}} \\
 & \lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k n^{-\frac{p}{2}} \left(\frac{|\beta_{jk}|}{\lambda/2}\right)^r \right)^{\frac{1}{p}} \\
 & = \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} n^{-\frac{1}{2}} \lambda^{-\frac{r}{p}} \|\beta_j\|_r^{\frac{r}{p}}.
 \end{aligned}$$

By $\|\beta_j\|_r \lesssim 2^{-j(N+\frac{1}{2}-\frac{1}{p})}$ and $\lambda \sim c\sqrt{\frac{\ln n}{n}}$, we have

$$\begin{aligned}
 W_1 & \lesssim \sum_{j=j_0}^{j_1} n^{-\frac{1}{2}} 2^{j(\frac{1}{2}-\frac{1}{p})} 2^{-j(N+\frac{1}{2}-\frac{1}{p})\frac{r}{p}} \left(\frac{n}{\ln n}\right)^{\frac{r}{2p}} \\
 & = n^{-\frac{1}{2}} \left(\frac{n}{\ln n}\right)^{\frac{r}{2p}} \sum_{j=j_0}^{j_1} 2^{-j\xi} \\
 & \leq n^{\frac{r-p}{2p}} \ln n^{-\frac{r}{2p}} (2^{-j_0\xi} I\{\xi > 0\} + 2^{-j_1\xi} I\{\xi < 0\} + (j_1 - j_0 + 1)I\{\xi = 0\}).
 \end{aligned}$$

Using Lemma 3.1 and (3.4), we obtain

$$W_1 \lesssim \begin{cases} (\ln n)^{c_3} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ (\ln n)^{c_4} \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r = \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r < \frac{p}{2N+1}, \end{cases} \quad (3.11)$$

where c_3, c_4 are constants.

(ii) For

$$W_3 := \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\beta_{jk}|^p I\{|\beta_{jk}| \leq 2\lambda\} \right)^{\frac{1}{p}},$$

let $\xi := \frac{1}{2}(\frac{r}{p}(2N+1) - 1)$. By $I\{|\beta_{jk}| \leq 2\lambda\} \leq (\frac{2\lambda}{|\beta_{jk}|})^{p-r}$ ($r < p$), we have

$$\begin{aligned} W_3 &\lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\beta_{jk}|^p \left(\frac{2\lambda}{|\beta_{jk}|}\right)^{p-r} \right)^{\frac{1}{p}} \\ &= \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} (2\lambda)^{\frac{p-r}{p}} \left(\sum_k |\beta_{jk}|^r \right)^{\frac{1}{p}} \\ &= \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} (2\lambda)^{\frac{p-r}{p}} \|\beta_j\|_r^{\frac{r}{p}}. \end{aligned}$$

Since $f^X \in W_r^N(\mathbb{R})$, then $\|\beta_j\|_r \lesssim 2^{-j(N+\frac{1}{2}-\frac{1}{r})}$. Taking $\lambda = c\sqrt{\frac{j}{n}} \sim c\sqrt{\frac{\ln n}{n}}$, $j_1 - j_0 \sim C(\ln n)$, we have

$$\begin{aligned} W_3 &\lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \lambda^{\frac{p-r}{p}} 2^{-j(N+\frac{1}{2}-\frac{1}{r})\frac{r}{p}} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2p}} \sum_{j=j_0}^{j_1} 2^{-j\frac{1}{2}[\frac{r}{p}(2N+1)-1]} \\ &= \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2p}} \sum_{j=j_0}^{j_1} 2^{-j\xi} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2p}} (2^{-j_0\xi} I\{\xi > 0\} + 2^{-j_1\xi} I\{\xi < 0\} + (j_1 - j_0 + 1) I\{\xi = 0\}). \end{aligned}$$

Note that $\xi > 0$ if and only if $r > \frac{p}{2N+1}$. When $\xi = 0$, i.e., $p = r(2N+1)$, we can compute $\frac{N'}{2(N-\frac{1}{r})+1} = \frac{p-r}{2p}$. Using (3.2), (3.3), we obtain

$$W_3 \lesssim \begin{cases} \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2p}} 2^{-j_0\xi} = (\ln n)^{\frac{p-r}{2r(2N+1)}} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2p}} (j_1 - j_0 + 1) \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2p}}, & r = \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2p}} 2^{-j_1\xi} = \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r < \frac{p}{2N+1}. \end{cases} \quad (3.12)$$

(iii) Finally, we estimate

$$W_2 := E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\hat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{\frac{1}{p}}.$$

Let $1 < q', q < \infty$, and $\frac{1}{q} + \frac{1}{q'} = 1$. Using Jensen's inequality and Hölder's inequality, we have

$$\begin{aligned} W_2 &= E \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\hat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{\frac{1}{p}} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k E \left(|\hat{\beta}_{jk} - \beta_{jk}|^p I \left\{ |\hat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right) \right)^{\frac{1}{p}} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k (E |\hat{\beta}_{jk} - \beta_{jk}|^{qp})^{\frac{1}{q}} \left(E I^q \left\{ |\hat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right\} \right)^{\frac{1}{q'}} \right)^{\frac{1}{p}} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k (E |\hat{\beta}_{jk} - \beta_{jk}|^{qp})^{\frac{1}{q}} \left(P \left(|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\lambda}{2} \right) \right)^{\frac{1}{q'}} \right)^{\frac{1}{p}}. \end{aligned}$$

By Lemma 3.1 and Lemma 3.2, we obtain

$$W_2 \leq \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} (2^j n^{-\frac{p}{2}} 2^{-\frac{\omega j}{q}})^{\frac{1}{p}} = n^{-\frac{1}{2}} \sum_{j=j_0}^{j_1} 2^{j(\frac{1}{2}-\frac{\omega}{pq})}.$$

Taking large enough ω such that $\frac{1}{2} < \frac{\omega}{pq}$, we get

$$W_2 \lesssim n^{-\frac{1}{2}} 2^{j_0(\frac{1}{2}-\frac{\omega}{pq})} \leq \sqrt{n^{-1} 2^{j_0}}.$$

Taking 2^{j_0} as in (3.2), we have

$$W_2 \lesssim \begin{cases} (\ln n)^{\frac{p-r}{2r(2N+1)}} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ n^{-\frac{N'}{2(N-1/r)+1}}, & r \leq \frac{p}{2N+1}. \end{cases} \quad (3.13)$$

Putting (3.11), (3.12) and (3.13) together, we can obtain

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

where c_3, c_4 are constants. □

Theorem 3.4 *Let the scaling function $\varphi(x)$ be orthonormal, compactly supported and $N + 1$ regular. There exist two positive constants g_1 and g_2 such that $g_1 \leq g(x) \leq g_2, x \in \mathbb{R}$. If*

$\hat{f}_n^{X_{\text{non}}}$ is the nonlinear wavelet estimator in (3.1), and assumptions (3.2), (3.3) and (3.4) are satisfied, then for any $f^X \in \tilde{W}_r^N(A, L)$, where $1 \leq r < p < \infty$, $N > \frac{1}{r}$, we have

$$\sup_{f^X \in \tilde{W}_r^N(A, L)} E \|\hat{f}_n^{X_{\text{non}}} - f^X\|_p \lesssim \begin{cases} (\ln n)^{c_3} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ (\ln n)^{c_4} \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r = \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r < \frac{p}{2N+1}, \end{cases}$$

where $N' = N - 1/r + 1/p$, c_3, c_4 are constants.

Proof By the definition of $\hat{f}_n^{X_{\text{non}}}$ in (3.1) and the expansion of f^X in (2.1), one has

$$\hat{f}_n^{X_{\text{non}}} - f^X = \sum_k (\hat{\alpha}_{j_0k} - \alpha_{j_0k}) \varphi_{j_0k} + \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} + P_{j_1+1} f^X - f^X.$$

Then

$$\begin{aligned} E \|\hat{f}_n^{X_{\text{non}}} - f^X\|_p &\leq E \left\| \sum_k (\hat{\alpha}_{j_0k} - \alpha_{j_0k}) \varphi_{j_0k} \right\|_p + E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p + \|P_{j_1+1} f^X - f^X\|_p \\ &=: I_1 + I_2 + I_3, \end{aligned}$$

where

$$\begin{aligned} I_1 &:= E \left\| \sum_k (\hat{\alpha}_{j_0k} - \alpha_{j_0k}) \varphi_{j_0k} \right\|_p, \\ I_2 &:= E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p, \\ I_3 &:= \|P_{j_1+1} f^X - f^X\|_p. \end{aligned}$$

Firstly, we estimate

$$I_1 := E \left\| \sum_k (\hat{\alpha}_{j_0k} - \alpha_{j_0k}) \varphi_{j_0k} \right\|_p.$$

By Lemma 2.2 and Jensen's inequality,

$$I_1 \lesssim 2^{j_0(\frac{1}{2}-\frac{1}{p})} E \left(\sum_k |\hat{\alpha}_{j_0k} - \alpha_{j_0k}|^p \right)^{\frac{1}{p}} \leq 2^{j_0(\frac{1}{2}-\frac{1}{p})} \left(\sum_k E |\hat{\alpha}_{j_0k} - \alpha_{j_0k}|^p \right)^{\frac{1}{p}}.$$

Since $f^X(x)$ and $\varphi(x)$ are compactly supported, then the number of elements in $\{k : \alpha_{j_0k} \neq 0\}$ is $O(2^{j_0})$. By Lemma 3.1, we have $E |\hat{\alpha}_{j_0k} - \alpha_{j_0k}|^p \lesssim n^{-\frac{p}{2}}$.

Therefore

$$I_1 \lesssim 2^{j_0(\frac{1}{2}-\frac{1}{p})} (2^{j_0} n^{-\frac{p}{2}})^{\frac{1}{p}} = \sqrt{n^{-1} 2^{j_0}}.$$

Using (3.2), we have

$$I_1 \lesssim \begin{cases} (\ln n)^{\frac{p-r}{2r(2N+1)}} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ n^{-\frac{N'}{2(N-1/r)+1}}, & r \leq \frac{p}{2N+1}, \end{cases} \quad (3.14)$$

where $N' = N - \frac{1}{r} + \frac{1}{p}$.

Next, we estimate

$$I_3 := \|P_{j_1+1}f^X - f^X\|_p.$$

In reference [9], it turns out that if the scaling function $\varphi(x)$ is orthonormal, compactly supported and $N + 1$ regular, then the associated kernel function $K(x, y) := \sum_k \varphi(x-k)\varphi(y-k)$ satisfies Conditions $H(N + 1)$ and $M(N)$, and $K_j f(x) = P_j f(x)$.

Since a Sobolev space and a Besov space have the following embedding theorem: $\tilde{W}_r^N \hookrightarrow \tilde{B}_{p\infty}^N \hookrightarrow \tilde{B}_{p\infty}^{N'}$, where $N' = N - \frac{1}{r} + \frac{1}{p}$, then $f^X \in \tilde{B}_{p\infty}^{N'}$. By Lemma 2.3, we have

$$\|P_{j_1+1}f^X - f^X\|_p \lesssim 2^{-j_1 N'}.$$

Taking 2^{j_1} as in (3.3), we have

$$I_3 \lesssim \begin{cases} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r \leq \frac{p}{2N+1}. \end{cases} \quad (3.15)$$

Finally, we estimate

$$I_2 := E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p.$$

Using Lemma 3.3, we obtain

$$E \left\| \sum_{j=j_0}^{j_1} \sum_k (\hat{\beta}_{jk}^* - \beta_{jk}) \psi_{jk} \right\|_p \lesssim \begin{cases} (\ln n)^{c_3} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ (\ln n)^{c_4} \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r = \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r < \frac{p}{2N+1}. \end{cases} \quad (3.16)$$

By (3.14), (3.15) and (3.16), we obtain

$$\sup_{f^X \in \tilde{W}_r^N(\Delta, L)} E \|\hat{f}_n^{X, \text{non}} - f^X\|_p \lesssim \begin{cases} (\ln n)^{c_3} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ (\ln n)^{c_4} \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r = \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r < \frac{p}{2N+1}. \end{cases} \quad \square$$

4 Optimality

Now, we discuss the optimality of the rates of convergence. Using similar techniques as those in reference [10], we can obtain the following lower bound theorem.

Theorem 4.1 *Let the scaling function $\varphi(x)$ be orthonormal, compactly supported and $N + 1$ regular, $f^X \in \tilde{W}_r^N(\mathbb{A}, L)$. If there exist two positive constants g_1 and g_2 such that $g_1 \leq g(x) \leq g_2$, $x \in \mathbb{R}$, then for any estimator \hat{f}_n^X , we have*

$$\inf_{\hat{f}_n^X} \sup_{f^X \in \tilde{W}_r^N(\mathbb{A}, L)} E \|\hat{f}_n^X - f^X\|_p \gtrsim \begin{cases} n^{-\frac{N}{2N+1}}, & r > \frac{p}{2N+1}, \\ \left(\frac{\ln n}{n}\right)^{\frac{N'}{2(N-1/r)+1}}, & r \leq \frac{p}{2N+1}, \end{cases}$$

where $1 \leq r, p < \infty$, $N > \frac{1}{r}$.

Remark The proof is very similar to that in reference [10], in which the author studied the lower bound of the convergence rates in Besov spaces for the samples without bias data.

According to Theorem 4.1, we can see that:

- (i) When $r < \frac{p}{2N+1}$, our nonlinear estimator can attain the optimal rate.
- (ii) When $r = \frac{p}{2N+1}$, our convergence rate and the optimal rate of convergence differ in a logarithmic. So, it is sub-optimal.
- (iii) When $r > \frac{p}{2N+1}$, the logarithmic factor is an extra penalty for the chosen wavelet thresholding, our convergence rate is sub-optimal.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

WJR participated in the sequence alignment and drafted the manuscript. WM participated in the design of the study and performed the statistical analysis. ZY conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China (No. 11271038) and Foundation of BJUT (No. 006000542213501).

Received: 12 January 2013 Accepted: 14 June 2013 Published: 3 July 2013

References

1. Efromovich, S: Nonparametric Curve Estimation. Methods, Theory, and Applications. Springer Series in Statistics. Springer, New York (1999)
2. Vardi, Y: Nonparametric estimation in the presence of length bias. *Ann. Stat.* **10**(2), 616-620 (1982)
3. Jones, MC: Kernel density estimation for length-biased data. *Biometrika* **78**(3), 511-519 (1991)
4. Efromovich, S: Density estimation for biased data. *Ann. Stat.* **32**, 1137-1161 (2004)
5. Ramirez, P, Vidakovic, B: Wavelet density estimation for stratified size-biased sample. *J. Stat. Plan. Inference* **140**(2), 419-432 (2010)
6. Christophe, C: Wavelet block thresholding for density estimation in the presence of bias. *J. Korean Stat. Soc.* **39**, 43-53 (2010)
7. Kelly, C, Kon, MA, Rapheal, LA: Local convergence for wavelet expansion. *J. Funct. Anal.* **126**, 102-138 (1994)
8. Triebel, H: Theory of Function Spaces. Birkhäuser, Basel (1983)
9. Härdle, W, Kerkycharian, G, Picard, D, Tsybakov, A: Wavelets, Approximation and Statistical Applications. Springer, Berlin (1997)
10. Wang, HY: Convergence rates of density estimation in Besov spaces. *Appl. Math.* **2**(10), 1258-1262 (2011)

doi:10.1186/1029-242X-2013-308

Cite this article as: Wang et al.: Nonlinear wavelet density estimation for biased data in Sobolev spaces. *Journal of Inequalities and Applications* 2013 **2013**:308.