

RESEARCH

Open Access



# Wavelet density estimation for mixing and size-biased data

Junke Kou<sup>1</sup> and Huijun Guo<sup>1\*</sup> 

\*Correspondence:

gkmath17@163.com

<sup>1</sup>School of Mathematics and Computational Science, Guilin University of Electronic Technology, Guilin, P.R. China

## Abstract

This paper considers wavelet estimation for a multivariate density function based on mixing and size-biased data. We provide upper bounds for the mean integrated squared error (MISE) of wavelet estimators. It turns out that our results reduce to the corresponding theorem of Shirazi and Doosti (*Stat. Methodol.* 27:12–19, 2015), when the random sample is independent.

**Keywords:** Density estimation; Strong mixing; Size-biased; Wavelets

## 1 Introduction

Let  $\{Y_i, i \in \mathbb{Z}\}$  be a strictly stationary random process defined on a probability space  $(\Omega, \mathcal{F}, P)$  with the common density function

$$g(y) = \frac{\omega(y)f(y)}{\mu}, \quad y \in \mathbb{R}^d, \quad (1)$$

where  $\omega$  denotes a known positive function,  $f$  stands for an unknown density function of the unobserved random variable  $X$  and  $\mu = E\omega(X) = \int_{\mathbb{R}^d} \omega(y)f(y) dy < +\infty$ . We want to estimate the unknown density function  $f$  from a sequence of strong mixing data  $Y_1, Y_2, \dots, Y_n$ .

When  $Y_1, Y_2, \dots, Y_n$  are independent and  $d = 1$ , Ramírez and Vidakovic [13] propose a linear wavelet estimator and show it to be  $L^2$  consistent; Chesneau [1] considers the optimal convergence rates of wavelet block thresholding estimator; Shirazi and Doosti [16] expand Ramírez and Vidakovic's [13] work to  $d \geq 1$ . Chesneau et al. [2] extend the independence to both positively and negatively associated cases. They show a convergence rate for mean integrated squared error (MISE). An upper bound of wavelet estimation on  $L^p$  ( $1 \leq p < +\infty$ ) risk in negatively associated case is given by Liu and Xu [9].

This paper deals with the  $d$ -dimensional density estimate problem (1), when  $Y_1, Y_2, \dots, Y_n$  are strong mixing. We give upper bounds for the mean integrated squared error (MISE) of wavelet estimators. It turns out that our linear result reduces to Shirazi and Doosti's [16] theorem, when the random sample is independent.

### 1.1 Wavelets and Besov spaces

As a central notion in wavelet analysis, Multiresolution Analysis (MRA, Meyer [11]) plays an important role in constructing a wavelet basis, which means a sequence of closed sub-

spaces  $\{V_j\}_{j \in \mathbb{Z}}$  of the square integrable function space  $L^2(\mathbb{R}^d)$  satisfying the following properties:

- (i)  $V_j \subseteq V_{j+1}, j \in \mathbb{Z}$ . Here and after,  $\mathbb{Z}$  denotes the integer set and  $\mathbb{N} := \{n \in \mathbb{Z}, n \geq 0\}$ ;
- (ii)  $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}^d)$ . This means the space  $\bigcup_{j \in \mathbb{Z}} V_j$  being dense in  $L^2(\mathbb{R}^d)$ ;
- (iii)  $f(2 \cdot) \in V_{j+1}$  if and only if  $f(\cdot) \in V_j$  for each  $j \in \mathbb{Z}$ ;
- (iv) There exists a scaling function  $\varphi \in L^2(\mathbb{R}^d)$  such that  $\{\varphi(\cdot - k), k \in \mathbb{Z}^d\}$  forms an orthonormal basis of  $V_0 = \overline{\text{span}}\{\varphi(\cdot - k)\}$ .

When  $d = 1$ , there is a simple way to define an orthonormal wavelet basis. Examples include the Daubechies wavelets with compact supports. For  $d \geq 2$ , the tensor product method gives an MRA  $\{V_j\}$  of  $L^2(\mathbb{R}^d)$  from one-dimensional MRA. In fact, with a scaling function  $\varphi$  of tensor products, we find  $M = 2^d - 1$  wavelet functions  $\psi^\ell$  ( $\ell = 1, 2, \dots, M$ ) such that, for each  $f \in L^2(\mathbb{R}^d)$ , the following decomposition

$$f = \sum_{k \in \mathbb{Z}^d} \alpha_{j_0, k} \varphi_{j_0, k} + \sum_{j=j_0}^{\infty} \sum_{\ell=1}^M \sum_{k \in \mathbb{Z}^d} \beta_{j, k}^\ell \psi_{j, k}^\ell$$

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

$$\varphi_{j_0, k}(y) = 2^{\frac{j_0 d}{2}} \varphi(2^{j_0} y - k), \quad \psi_{j, k}^\ell(y) = 2^{\frac{j d}{2}} \psi^\ell(2^j y - k).$$

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

$$P_j f = \sum_{k \in \mathbb{Z}^d} \alpha_{j, k} \varphi_{j, k}. \tag{2}$$

A wavelet basis can be used to characterize Besov spaces. The next lemma provides equivalent definitions for those spaces, for which we need one more notation: a scaling function  $\varphi$  is called  $m$ -regular if  $\varphi \in C^m(\mathbb{R}^d)$  and  $|D^\alpha \varphi(y)| \leq c(1 + |y|^2)^{-\ell}$  for each  $\ell \in \mathbb{Z}$  and each multi-index  $\alpha \in \mathbb{N}^d$  with  $|\alpha| \leq m$ .

**Lemma 1.1** (Meyer [11]) *Let  $\varphi$  be  $m$ -regular,  $\psi^\ell$  ( $\ell = 1, 2, \dots, M, M = 2^d - 1$ ) be the corresponding wavelets and  $f \in L^p(\mathbb{R}^d)$ . If  $\alpha_{j, k} = \langle f, \varphi_{j, k} \rangle, \beta_{j, k}^\ell = \langle f, \psi_{j, k}^\ell \rangle, p, q \in [1, \infty]$ , and  $0 < s < m$ , then the following assertions are equivalent:*

- (1)  $f \in B_{p, q}^s(\mathbb{R}^d)$ ;
- (2)  $\{2^{j s} \|P_{j+1} f - P_j f\|_p\} \in l_q$ ;
- (3)  $\{2^{j(s - \frac{d}{p} + \frac{d}{q})} \|\beta_j\|_p\} \in l_q$ .

The Besov norm of  $f$  can be defined by

$$\|f\|_{B_{p, q}^s} := \left\| (\alpha_{j_0}) \right\|_p + \left\| (2^{j(s - \frac{d}{p} + \frac{d}{q})} \|\beta_j\|_p)_{j \geq j_0} \right\|_q \quad \text{with } \|\beta_j\|_p^p = \sum_{\ell=1}^M \sum_{k \in \mathbb{Z}^d} |\beta_{j, k}^\ell|^p.$$

### 1.2 Estimators and result

In this paper, we require  $\text{supp } Y_i \subseteq [0, 1]^d$  in model (1). This is similar to Chesneau [1], Chesneau et al. [2], Liu and Xu [9]. Now we give the definition of strong mixing.

**Definition 1.1** (Rosenblatt [15]) A strictly stationary sequence of random vectors  $\{Y_i\}_{i \in \mathbb{Z}}$  is said to be strong mixing if

$$\lim_{k \rightarrow \infty} \alpha(k) = \lim_{k \rightarrow \infty} \sup \{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in F_{-\infty}^0, B \in F_k^\infty \} = 0,$$

where  $F_{-\infty}^0$  denotes the  $\sigma$  field generated by  $\{Y_i\}_{i \leq 0}$  and  $F_k^\infty$  does by  $\{Y_i\}_{i \geq k}$ .

Obviously, the independent and identically distributed (*i.i.d.*) data are strong mixing since  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$  and  $\alpha(k) \equiv 0$  in that case. Now, we provide two examples for strong mixing data.

*Example 1* Let  $X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$  with

$$\{\varepsilon_t, t \in \mathbb{Z}\} \stackrel{i.i.d.}{\sim} N(0, \sigma^2) \quad \text{and} \quad a_k = \begin{cases} 2^{-k}, & k \geq 0, \\ 0, & k < 0. \end{cases}$$

Then it can be proved by Theorem 2 and Corollary 1 of Doukhan [5] on p. 58 that  $\{X_t, t \in \mathbb{Z}\}$  is a strong mixing sequence.

*Example 2* Let  $\{\varepsilon(t), t \in \mathbb{Z}\} \stackrel{i.i.d.}{\sim} N_r(\vec{0}, \Sigma)$  ( $r$ -dimensional normal distribution) and  $\{Y(t), t \in \mathbb{Z}\}$  satisfy the auto-regression moving average equation

$$\sum_{i=0}^p B(i)Y(t-i) = \sum_{k=0}^q A(k)\varepsilon(t-k)$$

with  $l \times r$  and  $l \times l$  matrices  $A(k), B(i)$  respectively, as well as  $B(0)$  being the identity matrix. If the absolute values of the zeros of the determinant  $\det P(z) := \det \sum_{i=0}^p B(i)z^i$  ( $z \in \mathbb{C}$ ) are strictly greater than 1, then  $\{Y(t), t \in \mathbb{Z}\}$  is strong mixing (Mokkadem [12]).

It is well known that a Lebesgue measurable function maps *i.i.d.* data to *i.i.d.* data. When dealing with strong mixing data, it seems necessary to require the functions  $\omega$  in (1) to be Borel measurable. A Borel measurable function  $f$  on  $\mathbb{R}^d$  means  $\{y \in \mathbb{R}^d, f(y) > c\}$  being a Borel set for each  $c \in \mathbb{R}$ . In that case, we can prove easily that  $\{f(Y_i)\}$  remains strong mixing and  $\alpha_{f(Y)}(k) \leq \alpha_Y(k)$  ( $k = 1, 2, \dots$ ) if  $\{Y_i\}$  has the same property, see Guo [6]. This note is important for the proofs of the lemmas in the next section.

Before introducing our estimators, we formulate the following assumptions:

A1. The weight function  $\omega$  has both positive upper and lower bounds, i.e., for  $y \in [0, 1]^d$ ,

$$0 < c_1 \leq \omega(y) \leq c_2 < +\infty.$$

A2. The strong mixing coefficient of  $\{Y_i, i = 1, 2, \dots, n\}$  satisfies  $\alpha(k) = O(\gamma e^{-c_3 k})$  with  $\gamma > 0, c_3 > 0$ .

A3. The density  $f_{(Y_1, Y_{k+1})}$  of  $(Y_1, Y_{k+1})$  ( $k \geq 1$ ) and the density  $f_{Y_1}$  of  $Y_1$  satisfy that for  $(y, y^*) \in [0, 1]^d \times [0, 1]^d$ ,

$$\sup_{k \geq 1} \sup_{(y, y^*) \in [0, 1]^d \times [0, 1]^d} |h_k(y, y^*)| \leq c_4,$$

where  $h_k(y, y^*) = f_{(Y_1, Y_{k+1})}(y, y^*) - f_{Y_1}(y)f_{Y_{k+1}}(y^*)$  and  $c_4 > 0$ .

Assumption A1 is standard for the nonparametric density model with size-biased data, see Ramírez and Vidakovic [13], Chesneau [1], Liu and Xu [9]. Condition A3 can be viewed as a ‘Castellana–Leadbetter’ type condition in Masry [10].

We choose a  $d$ -dimensional scaling function

$$\varphi(y) = \varphi(y_1, \dots, y_d) := D_{2N}(y_1) \cdots D_{2N}(y_d)$$

with  $D_{2N}(\cdot)$  being the one-dimensional Daubechies scaling function. Then  $\varphi$  is  $m$ -regular ( $m > 0$ ) when  $N$  gets large enough. Note that  $D_{2N}$  has compact support  $[0, 2N - 1]$  and the corresponding wavelet has compact support  $[-N + 1, N]$ . Then, for  $f \in L^2(\mathbb{R}^d)$  with  $\text{supp} f \subseteq [0, 1]^d$  and  $M = 2^d - 1$ ,

$$f(y) = \sum_{k \in \Lambda_{j_0}} \alpha_{j_0,k} \varphi_{j_0,k}(y) + \sum_{j=j_0}^{\infty} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \beta_{j,k}^{\ell} \psi_{j,k}^{\ell}(y),$$

where  $\Lambda_{j_0} = \{1 - 2N, 2 - 2N, \dots, 2^{j_0}\}^d$ ,  $\Lambda_j = \{-N, -N + 1, \dots, 2^j + N - 1\}^d$  and

$$\alpha_{j_0,k} = \int_{[0,1]^d} f(y) \varphi_{j_0,k}(y) dy, \quad \beta_{j,k}^{\ell} = \int_{[0,1]^d} f(y) \psi_{j,k}^{\ell}(y) dy.$$

We introduce

$$\widehat{\mu}_n = \left[ \frac{1}{n} \sum_{i=1}^n \frac{1}{\omega(Y_i)} \right]^{-1}, \tag{3}$$

$$\widehat{\alpha}_{j_0,k} = \frac{\widehat{\mu}_n}{n} \sum_{i=1}^n \frac{\varphi_{j_0,k}(Y_i)}{\omega(Y_i)}, \tag{4}$$

and

$$\widehat{\beta}_{j,k}^{\ell} = \frac{\widehat{\mu}_n}{n} \sum_{i=1}^n \frac{\psi_{j,k}^{\ell}(Y_i)}{\omega(Y_i)}. \tag{5}$$

Now, we define our linear wavelet estimator

$$\widehat{f}_n^{\text{lin}}(y) = \sum_{k \in \Lambda_{j_0}} \widehat{\alpha}_{j_0,k} \varphi_{j_0,k}(y) \tag{6}$$

and the nonlinear wavelet estimator

$$\widehat{f}_n^{\text{non}}(y) = \widehat{f}_n^{\text{lin}}(y) + \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \widehat{\beta}_{j,k}^{\ell} I_{\{|\widehat{\beta}_{j,k}^{\ell}| \geq \kappa t_n\}} \psi_{j,k}^{\ell}(y) \tag{7}$$

with  $t_n := \sqrt{\frac{\ln n}{n}}$ . The positive integers  $j_0$  and  $j_1$  are specified in the theorem, while the constant  $\kappa$  will be chosen in the proof of the theorem.

The following notations are needed to state our theorem: For  $H > 0$ ,

$$B_{p,q}^s(H) := \{f \in B_{p,q}^s(\mathbb{R}^d), \|f\|_{B_{p,q}^s} \leq H\}$$

and  $x_+ := \max\{x, 0\}$ . In addition,  $A \lesssim B$  denotes  $A \leq cB$  for some constant  $c > 0$ ;  $A \gtrsim B$  means  $B \lesssim A$ ;  $A \sim B$  stands for both  $A \lesssim B$  and  $B \lesssim A$ .

**Main theorem** Consider the problem defined by (1) under assumptions A1–A3. Let  $f \in B_{p,q}^s(H)$  ( $p, q \in [1, \infty)$ ,  $s > \frac{d}{p}$ ) and  $\text{supp} f \subseteq [0, 1]^d$ . Then the linear wavelet estimator  $\widehat{f}_n^{\text{lin}}$  defined in (6) with  $2^{j_0} \sim n^{\frac{1}{2s'+d}}$  and  $s' = s - d(\frac{1}{p} - \frac{1}{2})_+$  satisfies

$$E \int_{[0,1]^d} |\widehat{f}_n^{\text{lin}}(y) - f(y)|^2 dy \lesssim n^{-\frac{2s'}{2s'+d}}; \tag{8a}$$

the nonlinear estimator in (7) with  $2^{j_0} \sim n^{\frac{1}{2m+d}}$  ( $m > s$ ),  $2^{j_1} \sim (\frac{n}{(\ln n)^3})^{\frac{1}{d}}$  satisfies

$$E \int_{[0,1]^d} |\widehat{f}_n^{\text{non}}(y) - f(y)|^2 dy \lesssim (\ln n)^3 n^{-\frac{2s}{2s+d}}. \tag{8b}$$

*Remark 1* When  $d = 1$ ,  $n^{-\frac{2s}{2s+1}}$  is the optimal convergence rate in the minimax sense for the standard nonparametric density model, see Donoho et al. [4].

*Remark 2* When the strong mixing data  $Y_1, Y_2, \dots, Y_n$  reduce to independent and identically distributed (*i.i.d.*) data, the convergence rate of our linear estimator is the same as that of Theorem 3.1 in Shirazi and Doosti [16].

*Remark 3* Compared with the linear wavelet estimator  $\widehat{f}_n^{\text{lin}}$ , the nonlinear estimator  $\widehat{f}_n^{\text{non}}$  is adaptive, which means both  $j_0$  and  $j_1$  do not depend on  $s, p$ , and  $q$ . On the other hand, the convergence rate of the nonlinear estimator remains the same as that of the linear one up to  $(\ln n)^3$ , when  $p \geq 2$ . However, it gets better for  $1 \leq p < 2$ .

### 2 Some lemmas

In this section, we provide some lemmas for the proof of the theorem. The following simple (but important) lemma holds.

**Lemma 2.1** For the model defined in (1),

$$E(\widehat{\mu}_n^{-1}) = \mu^{-1}, \tag{9a}$$

$$E \left[ \frac{\mu \varphi_{j_0,k}(Y_i)}{\omega(Y_i)} \right] = \alpha_{j_0,k}, \tag{9b}$$

$$E \left[ \frac{\mu \psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right] = \beta_{j,k}^\ell, \tag{9c}$$

where  $\alpha_{j_0,k} = \int_{[0,1]^d} f(y) \varphi_{j_0,k}(y) dy$  and  $\beta_{j,k}^\ell = \int_{[0,1]^d} f(y) \psi_{j,k}^\ell(y) dy$  ( $\ell = 1, 2, \dots, M$ ).

*Proof* One includes a simple proof for completeness. By (3),

$$E(\widehat{\mu}_n^{-1}) = E \left[ \frac{1}{n} \sum_{i=1}^n \frac{1}{\omega(Y_i)} \right] = E \left[ \frac{1}{\omega(Y_i)} \right].$$

This with (1) leads to

$$E(\widehat{\mu}_n^{-1}) = \int_{[0,1]^d} \frac{g(y)}{\omega(y)} dy = \frac{1}{\mu} \int_{[0,1]^d} f(y) dy = \frac{1}{\mu},$$

which concludes (9a). Using (1), one knows that

$$E\left[\frac{\mu\varphi_{j_0,k}(Y_i)}{\omega(Y_i)}\right] = \int_{[0,1]^d} \frac{\mu\varphi_{j_0,k}(y)}{\omega(y)} g(y) dy = \int_{[0,1]^d} f(y)\varphi_{j_0,k}(y) dy = \alpha_{j_0,k}.$$

This completes the proof of (9b). Similar arguments show (9c). □

To estimate  $E|\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2$  and  $E|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2$ , we introduce an important inequality, which can be found in Davydov [3].

**Davydov’s inequality** Let  $\{Y_i\}_{i \in \mathbb{Z}}$  be strong mixing with mixing coefficient  $\alpha(k)$ ,  $f$  and  $g$  be two measurable functions. If  $E|f(Y_1)|^p$  and  $E|g(Y_1)|^q$  exist for  $p, q > 0$  and  $\frac{1}{p} + \frac{1}{q} < 1$ , then there exists a constant  $c > 0$  such that

$$|\text{cov}(f(Y_1), g(Y_{k+1}))| \leq c[\alpha(k)]^{1-\frac{1}{p}-\frac{1}{q}} [E|f(Y_1)|^p]^{\frac{1}{p}} [E|g(Y_1)|^q]^{\frac{1}{q}}.$$

**Lemma 2.2** Let  $f \in B_{p,q}^s(H)$  ( $p, q \in [1, \infty)$ ,  $s > \frac{d}{p}$ ) and  $\widehat{\alpha}_{j_0,k}$ ,  $\widehat{\beta}_{j,k}^\ell$  be defined by (4) and (5). If A1–A3 hold, then

$$E|\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2 \lesssim n^{-1}, \quad E|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 \lesssim n^{-1}.$$

*Proof* One proves the second inequality only, the first one is similar. By the definition of  $\widehat{\beta}_{j,k}^\ell$ ,

$$\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell = \frac{\widehat{\mu}_n}{\mu} \left[ \frac{\mu}{n} \sum_{i=1}^n \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} - \beta_{j,k}^\ell \right] + \beta_{j,k}^\ell \cdot \widehat{\mu}_n \left( \frac{1}{\mu} - \frac{1}{\widehat{\mu}_n} \right)$$

and  $E|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 \lesssim E|\frac{\widehat{\mu}_n}{\mu} [\frac{\mu}{n} \sum_{i=1}^n \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} - \beta_{j,k}^\ell]|^2 + E|\beta_{j,k}^\ell \widehat{\mu}_n (\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n})|^2$ . Note that  $B_{p,q}^s(\mathbb{R}^d) \subseteq B_{\infty,\infty}^{s-\frac{d}{p}}(\mathbb{R}^d)$  with  $s > \frac{d}{p}$ . Then  $f \in B_{\infty,\infty}^{s-\frac{d}{p}}(\mathbb{R}^d)$  and  $\|f\|_\infty \lesssim 1$ . Moreover,  $|\beta_{j,k}^\ell| := |\int_{[0,1]^d} f(y) \times \psi_{j,k}^\ell(y) dy| \lesssim 1$  thanks to Hölder’s inequality and orthonormality of  $\{\psi_{j,k}^\ell\}$ . On the other hand,  $|\frac{\widehat{\mu}_n}{\mu}| \lesssim 1$  and  $|\widehat{\mu}_n| \lesssim 1$  because of A1. Hence,

$$E|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 \lesssim E\left| \frac{\mu}{n} \sum_{i=1}^n \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} - \beta_{j,k}^\ell \right|^2 + E\left| \frac{1}{\mu} - \frac{1}{\widehat{\mu}_n} \right|^2. \tag{10}$$

It follows from Lemma 2.1 and the definition of variance that

$$\begin{aligned} E|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 &\lesssim \text{var} \left[ \frac{1}{n} \sum_{i=1}^n \frac{\psi_{j,k}^\ell(X_i)}{\omega(Y_i)} \right] + \text{var} \left[ \frac{1}{n} \sum_{i=1}^n \frac{1}{\omega(X_i, Y_i)} \right] \\ &= \frac{1}{n^2} \text{var} \left[ \sum_{i=1}^n \frac{\psi_{j,k}^\ell(X_i)}{\omega(Y_i)} \right] + \frac{1}{n^2} \text{var} \left[ \sum_{i=1}^n \frac{1}{\omega(X_i, Y_i)} \right]. \end{aligned} \tag{11}$$

Note that Condition A1 implies  $\text{var}(\frac{1}{\omega(Y_i)}) \leq E(\frac{1}{\omega(Y_i)})^2 \lesssim 1$  and

$$\text{var} \left[ \sum_{i=1}^n \frac{1}{\omega(Y_i)} \right] \lesssim n \text{var} \left( \frac{1}{\omega(Y_i)} \right) + \left| \sum_{v=2}^n \sum_{i=1}^{v-1} \text{cov} \left( \frac{1}{\omega(Y_v)}, \frac{1}{\omega(Y_i)} \right) \right|.$$

Then it suffices to show

$$\left| \sum_{v=2}^n \sum_{i=1}^{v-1} \text{cov} \left( \frac{1}{\omega(Y_v)}, \frac{1}{\omega(Y_i)} \right) \right| \lesssim n. \tag{12}$$

By the strict stationarity of  $Y_i$ ,

$$\begin{aligned} \left| \sum_{v=2}^n \sum_{i=1}^{v-1} \text{cov} \left( \frac{1}{\omega(Y_v)}, \frac{1}{\omega(Y_i)} \right) \right| &= \left| \sum_{m=1}^n (n-m) \text{cov} \left( \frac{1}{\omega(Y_1)}, \frac{1}{\omega(Y_{m+1})} \right) \right| \\ &\leq n \sum_{m=1}^n \left| \text{cov} \left( \frac{1}{\omega(Y_1)}, \frac{1}{\omega(Y_{m+1})} \right) \right|. \end{aligned}$$

On the other hand, Davydov’s inequality and A1 show that

$$\left| \text{cov} \left( \frac{1}{\omega(Y_1)}, \frac{1}{\omega(Y_{m+1})} \right) \right| \lesssim \sqrt{\alpha(m)} \sqrt{E \left| \frac{1}{\omega(Y_1)} \right|^4} \lesssim \sqrt{\alpha(m)}.$$

These with A2 give the desired conclusion (12),

$$\left| \sum_{v=2}^n \sum_{i=1}^{v-1} \text{cov} \left( \frac{1}{\omega(Y_v)}, \frac{1}{\omega(Y_i)} \right) \right| \lesssim n \sum_{m=1}^n \sqrt{\alpha(m)} \lesssim n.$$

Now, the main work is to show

$$\text{var} \left[ \sum_{i=1}^n \frac{\psi_{j,k}^\ell(X_i)}{\omega(Y_i)} \right] \lesssim n. \tag{13}$$

Clearly,

$$\text{var} \left[ \sum_{i=1}^n \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right] \lesssim n \text{var} \left( \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right) + \left| \sum_{v=2}^n \sum_{i=1}^{v-1} \text{cov} \left( \frac{\psi_{j,k}^\ell(Y_v)}{\omega(Y_v)}, \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right) \right|.$$

By A1–A3 and (1), the first term of the above inequality is bounded by

$$nE \left( \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right)^2 \lesssim n \int_{[0,1]^d} [\psi_{j,k}^\ell(y)]^2 f(y) dy \lesssim n.$$

It remains to show

$$\begin{aligned} & \left| \sum_{v=2}^n \sum_{i=1}^{v-1} \operatorname{cov} \left( \frac{\psi_{j,k}^\ell(Y_v)}{\omega(Y_v)}, \frac{\psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right) \right| \\ & \leq n \left( \sum_{m=1}^{2^{jd}-1} + \sum_{m=2^{jd}}^n \right) \left| \operatorname{cov} \left[ \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)}, \frac{\psi_{j,k}^\ell(Y_{m+1})}{\omega(Y_{m+1})} \right] \right| \lesssim n, \end{aligned} \tag{14}$$

where the assumption  $2^{jd} \leq n$  is needed.

According to A1 and A3,

$$\begin{aligned} \left| \operatorname{cov} \left( \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)}, \frac{\psi_{j,k}^\ell(Y_{m+1})}{\omega(Y_{m+1})} \right) \right| & \leq \int_{[0,1]^d \times [0,1]^d} \left| \frac{\psi_{j,k}^\ell(y)}{\omega(y)} \cdot \frac{\psi_{j,k}^\ell(y^*)}{\omega(y^*)} \right| |h_m(y, y^*)| dy dy^* \\ & \lesssim \left( \int_{[0,1]^d} |\psi_{j,k}^\ell(y)| dy \right)^2 \lesssim 2^{-jd}. \end{aligned}$$

Hence,

$$\sum_{m=1}^{2^{jd}-1} \left| \operatorname{cov} \left( \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)}, \frac{\psi_{j,k}^\ell(Y_{m+1})}{\omega(Y_{m+1})} \right) \right| \lesssim \sum_{m=1}^{2^{jd}-1} 2^{-jd} \lesssim 1. \tag{15}$$

On the other hand, Davydov’s inequality and A1–A3 tell that

$$\begin{aligned} \left| \operatorname{cov} \left( \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)}, \frac{\psi_{j,k}^\ell(Y_{m+1})}{\omega(Y_{m+1})} \right) \right| & \lesssim \sqrt{\alpha(m)} \sqrt{E \left| \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)} \right|^4} \\ & \lesssim \sqrt{\alpha(m)} \sup \left| \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)} \right| \sqrt{E \left| \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)} \right|^2} \lesssim \sqrt{\alpha(m)} 2^{\frac{jd}{2}}. \end{aligned}$$

Moreover,  $\sum_{m=2^{jd}}^n \left| \operatorname{cov} \left( \frac{\psi_{j,k}^\ell(Y_1)}{\omega(Y_1)}, \frac{\psi_{j,k}^\ell(Y_{m+1})}{\omega(Y_{m+1})} \right) \right| \lesssim \sum_{m=2^{jd}}^n \sqrt{\alpha(m)} 2^{\frac{jd}{2}} \lesssim \sum_{m=1}^n \sqrt{m\alpha(m)} \leq \sum_{m=1}^{+\infty} m^{\frac{1}{2}} \gamma e^{-\frac{cm}{2}} < +\infty$ . This with (15) shows (14).  $\square$

To prove the last lemma in this section, we need the following Bernstein-type inequality (Liebscher [7, 8], Rio [14]).

**Bernstein-type inequality** Let  $(Y_i)_{i \in \mathbb{Z}}$  be a strong mixing process with mixing coefficient  $\alpha(k)$ ,  $EY_i = 0$ ,  $|Y_i| \leq M < \infty$ , and  $D_m = \max_{1 \leq j \leq 2m} \operatorname{var}(\sum_{i=1}^j Y_i)$ . Then, for  $\varepsilon > 0$  and  $n, m \in \mathbb{N}$  with  $0 < m \leq \frac{n}{2}$ ,

$$\mathbb{P} \left( \left| \sum_{i=1}^n Y_i \right| \geq \varepsilon \right) \leq 4 \cdot \exp \left\{ -\frac{\varepsilon^2}{16} \left( nm^{-1} D_m + \frac{1}{3} \varepsilon M m \right)^{-1} \right\} + 32 \frac{M}{\varepsilon} n \alpha(m).$$

**Lemma 2.3** Let  $f \in B_{p,q}^s(H)$  ( $p, q \in [1, \infty)$ ,  $s > \frac{d}{p}$ ),  $\widehat{\beta}_{j,k}^\ell$  be defined in (5) and  $t_n = \sqrt{\frac{\ln n}{n}}$ . If A1–A3 hold and  $2^{jd} \leq \frac{n}{(\ln n)^3}$ , then there exists a constant  $\kappa > 1$  such that

$$\mathbb{P} \left( \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right| \geq \kappa t_n \right) \lesssim n^{-4}.$$

*Proof* According to the arguments of (10),  $|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell| \lesssim \frac{1}{n} |\sum_{i=1}^n [\frac{1}{\omega(Y_i)} - \frac{1}{\mu}]| + \frac{1}{n} \times \sum_{i=1}^n \frac{\mu \psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} - \beta_{j,k}^\ell$ . Hence, it suffices to prove

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n \left[ \frac{1}{\omega(Y_i)} - \frac{1}{\mu} \right] \right| \geq \frac{\kappa}{2} t_n\right) \lesssim n^{-4} \quad \text{and} \\ & \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \left[ \frac{\mu \psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} - \beta_{j,k}^\ell \right] \geq \frac{\kappa}{2} t_n\right) \lesssim n^{-4}. \end{aligned} \tag{16}$$

One shows the second inequality only, because the first one is similar and even simpler.

Define  $\eta_i := \frac{\mu \psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} - \beta_{j,k}^\ell$ . Then  $E(\eta_i) = 0$  thanks to (9c), and  $\eta_1, \dots, \eta_n$  are strong mixing with the mixing coefficients  $\alpha(k) \leq \gamma e^{-ck}$  because of Condition A2. By A1–A3,  $|\frac{\mu \psi_{j,k}^\ell(Y_i)}{\omega(Y_i)}| \lesssim 2^{\frac{jd}{2}}$  and

$$|\eta_i| \leq \left| \frac{\mu \psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right| + E \left| \frac{\mu \psi_{j,k}^\ell(Y_i)}{\omega(Y_i)} \right| \lesssim 2^{\frac{jd}{2}}.$$

According to the arguments of (13),  $D_m = \max_{1 \leq j \leq 2m} \text{var}(\sum_{i=1}^j \eta_i) \lesssim m$ . Then it follows from Bernstein-type inequality with  $m = u \ln n$  (the constant  $u$  will be chosen later on) that

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n \eta_i \right| \geq \frac{\kappa}{2} t_n\right) &= \mathbb{P}\left(\left| \sum_{i=1}^n \eta_i \right| \geq \frac{\kappa}{2} n t_n\right) \\ &\lesssim \exp\left\{-\frac{(\kappa n t_n)^2}{64} \left( n m^{-1} D_m + \frac{1}{6} \kappa n t_n 2^{\frac{jd}{2}} m \right)^{-1}\right\} \\ &\quad + 64 \frac{2^{\frac{jd}{2}}}{\kappa n t_n} n \gamma e^{-cm}. \end{aligned} \tag{17}$$

Clearly,  $64 \frac{2^{\frac{jd}{2}}}{\kappa n t_n} n \gamma e^{-cm} \lesssim n e^{-cu \ln n}$  holds due to  $t_n = \sqrt{\frac{\ln n}{n}}$ ,  $2^{jd} \leq \frac{n}{(\ln n)^3}$  and  $m = u \ln n$ . Choose  $u$  such that  $1 - cu < -4$ , then the second term of (17) is bounded by  $n^{-4}$ . On the other hand, the first one of (17) has the following upper bound:

$$\exp\left\{-\frac{\kappa^2 \ln n}{64} \left(1 + \frac{1}{6} \kappa \sqrt{\frac{\ln n}{n}} \left(\frac{n}{(\ln n)^3}\right)^{\frac{1}{2}} m\right)^{-1}\right\} \lesssim \exp\left\{-\frac{\kappa^2 \ln n}{64} \left(1 + \frac{1}{6} \kappa u\right)^{-1}\right\}$$

thanks to  $D_m \lesssim m$ ,  $2^{jd} \leq \frac{n}{(\ln n)^3}$  and  $m = u \ln n$ . Obviously, there exists sufficiently large  $\kappa > 1$  such that  $\exp\{-\frac{\kappa^2 \ln n}{64} (1 + \frac{1}{6} \kappa u)^{-1}\} \lesssim n^{-4}$ . Finally, the desired conclusion (16) follows.  $\square$

### 3 Proof of the theorem

This section proves the theorem. The main idea of the proof comes from Donoho et al. [4].

*Proof of (8a)* Note that

$$\begin{aligned} E \int_{[0,1]^d} |\widehat{f}_n^{\text{lin}}(y) - f(y)|^2 dy &\leq E \int_{\mathbb{R}^d} |\widehat{f}_n^{\text{lin}}(y) - f(y)|^2 dy \\ &= E \|\widehat{f}_n^{\text{lin}} - P_{j_0} f\|_2^2 + \|P_{j_0} f - f\|_2^2. \end{aligned} \tag{18}$$

It is easy to see that

$$E \|\widehat{f}_n^{\text{lin}} - P_{j_0} f\|_2^2 = E \left\| \sum_{k \in \Lambda_{j_0}} (\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}) \varphi_{j_0,k} \right\|_2^2 = \sum_{k \in \Lambda_{j_0}} E |\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2.$$

According to Lemma 2.2,  $|\Lambda_{j_0}| \sim 2^{j_0 d}$  and  $2^{j_0} \sim n^{\frac{1}{2s'+d}}$ ,

$$E \|\widehat{f}_n^{\text{lin}} - P_{j_0} f\|_2^2 \lesssim \frac{2^{j_0 d}}{n} \sim n^{-\frac{2s'}{2s'+d}}. \tag{19}$$

When  $p \geq 2, s' = s$ . By Hölder’s inequality,  $f \in B_{p,q}^s(H)$ , and Lemma 1.1,

$$\|P_{j_0} f - f\|_2^2 \lesssim \|P_{j_0} f - f\|_p^2 \lesssim 2^{-2j_0 s} \sim n^{-\frac{2s}{2s+d}}. \tag{20}$$

When  $1 \leq p < 2$  and  $s > \frac{d}{p}, B_{p,q}^s(\mathbb{R}^d) \subseteq B_{2,\infty}^s(\mathbb{R}^d)$ . Then it follows from Lemma 1.1 and  $2^{j_0} \sim n^{\frac{1}{2s'+d}}$  that

$$\|P_{j_0} f - f\|_2^2 \lesssim \sum_{j=j_0}^{\infty} 2^{-2js'} \lesssim 2^{-2j_0 s'} \sim n^{-\frac{2s'}{2s'+d}}. \tag{21}$$

This with (20) shows in both cases

$$\|P_{j_0} f - f\|_2^2 \lesssim n^{-\frac{2s'}{2s'+d}}. \tag{22}$$

By (18), (19), and (22),

$$E \int_{[0,1]^d} |\widehat{f}_n^{\text{lin}}(y) - f(y)|^2 dy \lesssim n^{-\frac{2s'}{2s'+d}}. \quad \square$$

*Proof of (8b)* By the definitions of  $\widehat{f}_n^{\text{lin}}$  and  $\widehat{f}_n^{\text{non}}, \widehat{f}_n^{\text{non}}(y) - f(y) = [\widehat{f}_n^{\text{lin}}(y) - P_{j_0} f(y)] - [f(y) - P_{j_1+1} f(y)] + \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} [\widehat{\beta}_{j,k}^\ell I_{\{|\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n\}} - \beta_{j,k}^\ell] \psi_{j,k}^\ell(y)$ . Hence,

$$E \int_{[0,1]^d} |\widehat{f}_n^{\text{non}}(y) - f(y)|^2 dy \lesssim T_1 + T_2 + Q, \tag{23}$$

where  $T_1 := E \|\widehat{f}_n^{\text{lin}} - P_{j_0} f\|_2^2, T_2 := \|f - P_{j_1+1} f\|_2^2$  and

$$Q := E \left\| \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} [\widehat{\beta}_{j,k}^\ell I_{\{|\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n\}} - \beta_{j,k}^\ell] \psi_{j,k}^\ell \right\|_2^2.$$

According to (19) and  $2^{j_0} \sim n^{\frac{1}{2m+d}}$  ( $m > s$ ),

$$T_1 = E \|\widehat{f}_n^{\text{lin}} - P_{j_0} f\|_2^2 \lesssim \frac{2^{j_0 d}}{n} \sim n^{-\frac{2m}{2m+d}} < n^{-\frac{2s}{2s+d}}.$$

When  $p \geq 2$ , the same arguments as (20) shows  $T_2 = \|f - P_{j_1+1}f\|_2^2 \lesssim 2^{-2j_1 s}$ . This with  $2^{j_1} \sim (\frac{n}{(\ln n)^3})^{\frac{1}{d}}$  leads to

$$T_2 \lesssim 2^{-2j_1 s} \sim \left(\frac{(\ln n)^3}{n}\right)^{\frac{2s}{d}} \leq (\ln n)^3 n^{-\frac{2s}{2s+d}}. \tag{24}$$

On the other hand,  $B_{p,q}^s(\mathbb{R}^d) \subseteq B_{2,\infty}^{s+d/2-d/p}(\mathbb{R}^d)$  when  $1 \leq p < 2$  and  $s > \frac{d}{p}$ . Then

$$T_2 \lesssim 2^{-2j_1(s+\frac{d}{2}-\frac{d}{p})} \sim \left(\frac{(\ln n)^3}{n}\right)^{\frac{2(s+\frac{d}{2}-\frac{d}{p})}{d}} \leq (\ln n)^3 n^{-\frac{2s}{2s+d}}.$$

Hence,

$$T_2 \lesssim (\ln n)^3 n^{-\frac{2s}{2s+d}}$$

for each  $1 \leq p < +\infty$ .

The main work for the proof of (8b) is to show

$$Q = E \left\| \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} [\widehat{\beta}_{j,k}^\ell I_{\{|\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n\}} - \beta_{j,k}^\ell] \psi_{j,k}^\ell \right\|_2^2 \lesssim (\ln n)^3 n^{-\frac{2s}{2s+d}}. \tag{25}$$

Note that

$$Q = \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E |\widehat{\beta}_{j,k}^\ell I_{\{|\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n\}} - \beta_{j,k}^\ell|^2 \lesssim Q_1 + Q_2 + Q_3, \tag{26}$$

where

$$\begin{aligned} Q_1 &= \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E [|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 I_{\{|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell| > \frac{\kappa t_n}{2}\}}], \\ Q_2 &= \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E [|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}}], \\ Q_3 &= \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} |\beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \leq 2\kappa t_n\}}. \end{aligned}$$

For  $Q_1$ , one observes that

$$E [|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 I_{\{|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell| > \frac{\kappa t_n}{2}\}}] \leq [E |\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^4]^{\frac{1}{2}} \left[ \mathbb{P} \left( |\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell| > \frac{\kappa t_n}{2} \right) \right]^{\frac{1}{2}}$$

thanks to Hölder’s inequality. By Lemmas 2.1–2.3 and  $2^{jd} \leq n$ ,

$$E [|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 I_{\{|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell| > \frac{\kappa t_n}{2}\}}] \lesssim \left[ \frac{2^{jd}}{n} \right]^{\frac{1}{2}} \left[ \frac{1}{n^4} \right]^{\frac{1}{2}} \lesssim \frac{1}{n^2}.$$

Then  $Q_1 \lesssim \sum_{j=j_0}^{j_1} \frac{2^{jd}}{n^2} \lesssim \frac{2^{j_1 d}}{n^2} \lesssim \frac{1}{n} \leq n^{-\frac{2s}{2s+d}}$ , where one uses the choice  $2^{j_1} \sim (\frac{n}{(\ln n)^3})^{\frac{1}{d}}$ . Hence,

$$Q_1 \leq n^{-\frac{2s}{2s+d}}. \tag{27}$$

To estimate  $Q_2$ , one defines

$$2^{j'} \sim n^{\frac{1}{2s+d}}.$$

It is easy to see that  $2^{j_0} \sim n^{\frac{1}{2m+d}} \leq 2^{j'} \sim n^{\frac{1}{2s+d}} \leq 2^{j_1} \sim (\frac{n}{(\ln n)^3})^{\frac{1}{d}}$ . Furthermore, one rewrites

$$\begin{aligned} Q_2 &= \left( \sum_{j=j_0}^{j'} + \sum_{j=j'+1}^{j_1} \right) \left\{ \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E[|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}}] \right\} \\ &:= Q_{21} + Q_{22}. \end{aligned}$$

By Lemma 2.2 and  $2^{j'} \sim n^{\frac{1}{2s+d}}$ ,

$$\begin{aligned} Q_{21} &:= \sum_{j=j_0}^{j'} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E[|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}}] \\ &\lesssim \sum_{j=j_0}^{j'} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \frac{1}{n} \lesssim \sum_{j=j_0}^{j'} \frac{2^{jd}}{n} \lesssim \frac{2^{j'd}}{n} \sim n^{-\frac{2s}{2s+d}}. \end{aligned} \tag{28}$$

On the other hand, it follows from Lemma 2.2 that

$$\begin{aligned} Q_{22} &:= \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E[|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}}] \\ &\lesssim \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \frac{1}{n} I_{\{|\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}}. \end{aligned}$$

When  $p \geq 2$ ,

$$\begin{aligned} Q_{22} &\lesssim \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \frac{1}{n} I_{\{|\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}} \lesssim \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \frac{1}{n} \left( \frac{\beta_{j,k}^\ell}{\kappa t_n/2} \right)^2 \\ &\lesssim \sum_{j=j'+1}^{j_1} 2^{-2js} \lesssim 2^{-2j's} \sim n^{-\frac{2s}{2s+d}} \end{aligned} \tag{29}$$

with  $f \in B_{p,q}^s(H)$ , Lemma 1.1, Lemma 2.2, and  $t_n = \sqrt{\frac{\ln n}{n}}$ . When  $1 \leq p < 2$  and  $s > \frac{d}{p}$ ,  $B_{p,q}^s(\mathbb{R}^d) \subseteq B_{2,\infty}^{s+d/2-d/p}(\mathbb{R}^d)$ . Then

$$\begin{aligned} Q_{22} &\lesssim \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \frac{1}{n} I_{\{|\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}} \lesssim \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \frac{1}{n} \left( \frac{\beta_{j,k}^\ell}{\kappa t_n/2} \right)^p \\ &\lesssim \sum_{j=j'+1}^{j_1} n^{\frac{p}{2}-1} 2^{-j(s+d/2-d/p)p} \lesssim n^{\frac{p}{2}-1} 2^{-j'(s+d/2-d/p)p} \sim n^{-\frac{2s}{2s+d}}. \end{aligned} \tag{30}$$

Hence, this with (28) and (29) shows

$$Q_2 \lesssim n^{-\frac{2s}{2s+d}}. \tag{31}$$

Finally, one estimates  $Q_3$ . Clearly,

$$\begin{aligned} Q_{31} &:= \sum_{j=j_0}^{j'} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} |\beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \leq 2\kappa t_n\}} \\ &\leq \sum_{j=j_0}^{j'} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} |2\kappa t_n|^2 \lesssim \sum_{j=j_0}^{j'} \frac{\ln n}{n} 2^{jd} \lesssim \frac{\ln n}{n} 2^{j'd}. \end{aligned}$$

This with the choice of  $2^{j'}$  shows

$$Q_{31} \lesssim (\ln n) n^{-\frac{2s}{2s+d}}. \tag{32}$$

On the other hand,  $Q_{32} := \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} |\beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \leq 2\kappa t_n\}}$ . According to the arguments of (29),

$$Q_{32} \lesssim \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} |\beta_{j,k}^\ell|^2 \lesssim n^{-\frac{2s}{2s+d}} \tag{33}$$

for  $p \geq 2$ . When  $1 \leq p < 2$ ,  $|\beta_{j,k}^\ell|^2 I_{\{|\beta_{j,k}^\ell| \leq 2\kappa t_n\}} \leq |\beta_{j,k}^\ell|^p |2\kappa t_n|^{2-p}$ . Then similar to the arguments of (30),

$$\begin{aligned} Q_{32} &\lesssim \sum_{j=j'+1}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} |\beta_{j,k}^\ell|^p |2\kappa t_n|^{2-p} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{2-p}{2}} \sum_{j=j'+1}^{j_1} 2^{-j(s+d/2-d/p)p} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{2-p}{2}} 2^{-j'(s+d/2-d/p)p} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{2-p}{2}} \left(\frac{1}{n}\right)^{\frac{(s+d/2-d/p)p}{2s+d}} \leq (\ln n) n^{-\frac{2s}{2s+d}}. \end{aligned} \tag{34}$$

Combining this with (33) and (32), one knows  $Q_3 \lesssim (\ln n) n^{-\frac{2s}{2s+d}}$  in both cases. This with (26), (27), and (31) shows

$$Q \lesssim (\ln n)^3 n^{-\frac{2s}{2s+d}},$$

which is the desired conclusion. □

**Acknowledgements**

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

**Funding**

This paper is supported by the National Natural Science Foundation of China (No. 11771030), Guangxi Natural Science Foundation (No. 2017GXNSFAA198194), and Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation.

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 February 2018 Accepted: 18 July 2018 Published online: 25 July 2018

**References**

1. Chesneau, C.: Wavelet block thresholding for density estimation in the presence of bias. *J. Korean Stat. Soc.* **39**, 43–53 (2010)
2. Chesneau, C., Dewan, I., Doosti, H.: Wavelet linear density estimation for associated stratified size-biased sample. *J. Nonparametr. Stat.* **2**, 429–445 (2012)
3. Davydov, Y.A.: The invariance principle for stationary processes. *Theory Probab. Appl.* **3**, 487–498 (1970)
4. Donoho, D.L., Johnstone, M.I., Kerkycharian, G., Picard, D.: Density estimation by wavelet thresholding. *Ann. Stat.* **24**, 508–539 (1996)
5. Doukhan, P.: *Mixing: Properties and Examples*. Springer, New York (1994)
6. Guo, H.J.: *Wavelet estimations for a class of regression functions with errors-in-variables*. Dissertation, Beijing University of Technology (2016)
7. Liebscher, E.: Strong convergence of sums of  $\alpha$  mixing random variables with applications to density estimation. *Stoch. Process. Appl.* **65**, 69–80 (1996)
8. Liebscher, E.: Estimation of the density and regression function under mixing conditions. *Stat. Decis.* **19**, 9–26 (2001)
9. Liu, Y.M., Xu, J.L.: Wavelet density estimation for negatively associated stratified size-biased sample. *J. Nonparametr. Stat.* **26**, 537–554 (2014)
10. Masry, E.: Wavelet-based estimation of multivariate regression function in Besov spaces. *J. Nonparametr. Stat.* **12**, 283–308 (2000)
11. Meyer, Y.: *Wavelets and Operators*. Hermann, Paris (1990)
12. Mokkadem, A.: Mixing properties of ARMA processes. *Stoch. Process. Appl.* **29**, 309–315 (1988)
13. Ramírez, P., Vidakovic, B.: Wavelet density estimation for stratified size-biased sample. *J. Stat. Plan. Inference* **140**, 419–432 (2010)
14. Rio, E.: The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann. Probab.* **23**, 1188–1203 (1995)
15. Rosenblatt, M.: A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42**, 43–47 (1970)
16. Shirazi, E., Doosti, H.: Multivariate wavelet-based density estimation with size-biased data. *Stat. Methodol.* **27**, 12–19 (2015)

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---