




Wavelet regression estimations with strong mixing data

Junke Kou¹ · Youming Liu¹ 

Accepted: 24 June 2018 / Published online: 29 June 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Using a wavelet basis, we establish in this paper upper bounds of wavelet estimation on $L^p(\mathbb{R}^d)$ risk of regression functions with strong mixing data for $1 \leq p < \infty$. In contrast to the independent case, these upper bounds have different analytic formulae for $p \in [1, 2]$ and $p \in (2, +\infty)$. For $p = 2$, it turns out that our result reduces to a theorem of Chaubey et al. (J Nonparametr Stat 25:53–71, 2013); and for $d = 1$ and $p = 2$, it becomes the corresponding theorem of Chaubey and Shirazi (Commun Stat Theory Methods 44:885–899, 2015).

Keywords Regression estimation · L^p risk · Convergence rate · Strong mixing · Wavelet

Mathematics Subject Classification 62G07 · 42C40 · 62G20

1 Introduction and preliminary

Nonparametric regression estimation plays important roles in practical applications. The classical approach uses the Nadaraya–Watson type kernel estimators (Ahmad 1995). Because wavelet bases have the local property in both time and frequency domains, a wavelet provides a new method for analyzing functions (signals) with discontinuities or sharp spikes. Therefore it is natural to get better estimations than the kernel method for some cases. Great achievements have been made in this area, see Delyon and Judisky (1996), Hall and Patil (1996), Masry (2000), Chaubey et al. (2013), Chaubey and Shirazi (2015), Chesneau and Shirazi (2014) and Chesneau et al. (2015).

In this paper we consider the following model: Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a strictly stationary random process defined on a probability space (Ω, \mathcal{F}, P) with the common density function

✉ Youming Liu
liuym@bjut.edu.cn

¹ Department of Applied Mathematics, Beijing University of Technology, Beijing 100124, People's Republic of China

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

where ω stands for a known positive function, g denotes the density function of the unobserved random variables (U, V) and $\mu = E\omega(U, V) < \infty$. Then we want to estimate the unknown regression function

$$r(x) = E(\rho(V)|U = x), \quad x \in \mathbb{R}^d$$

from a sequence of strong mixing data $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$.

Chaubey et al. (2013) provide an upper bound of the mean integrated squared error for a linear wavelet estimator when $\rho(V) = V$ and $V \in [a, b]$; Chaubey and Shirazi (2015) consider the case of a nonlinear wavelet estimator with $d = 1$.

In this paper, we further extend these work to the d -dimensional setting over L^p risk for $1 \leq p < \infty$. When $p = 2$, our result reduces to Theorem 4.1 of Chaubey et al. (2013); in the case of $d = 1$ and $p = 2$, it becomes Theorem 5.1 of Chaubey and Shirazi (2015).

1.1 Wavelets and Besov spaces

As a central notion in wavelet analysis, Multiresolution Analysis plays an important role for constructing a wavelet basis, which means a sequence of closed subspaces $\{V_j\}_{j \in \mathbb{Z}}$ 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) $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}^d)$. 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 $\varphi(x) \in L^2(\mathbb{R}^d)$ 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 φ of tensor products, we find $M = 2^d - 1$ wavelet functions ψ^ℓ ($\ell = 1, 2, \dots, M$) 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

$$\varphi_{j_0, k}(x) = 2^{\frac{dj_0}{2}} \varphi(2^{j_0}x - k), \quad \psi_{j, k}^\ell(x) = 2^{\frac{dj}{2}} \psi^\ell(2^jx - k).$$

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

$$P_j f = \sum_{k \in \mathbb{Z}^d} \alpha_{j,k} \varphi_{j,k}. \quad (1)$$

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 φ is called m -regular, if $\varphi \in C^m(\mathbb{R}^d)$ and $|D^\alpha \varphi(x)| \leq c(1 + |x|^2)^{-\ell}$ for each $\ell \in \mathbb{Z}$ and each multi-index $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq m$.

Lemma 1.1 (Meyer 1990) *Let φ be m -regular, ψ^ℓ ($\ell = 1, 2, \dots, M, M = 2^d - 1$) 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) $\{2^{js} \|P_{j+1}f - P_j f\|_p\} \in l_q$;
- (3) $\{2^{j(s - \frac{d}{p} + \frac{d}{2})} \|\beta_j\|_p\} \in l_q$.

The Besov norm of f can be defined by

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

We also need a lemma (Härdle et al. 1998) in our later discussions.

Lemma 1.2 *Let a scaling function $\varphi \in L^2(\mathbb{R}^d)$ be m -regular and $\{\alpha_k\} \in l_p, 1 \leq p \leq \infty$. Then there exist $c_2 \geq c_1 > 0$ such that*

$$c_1 2^{j(\frac{d}{2} - \frac{d}{p})} \|(\alpha_k)\|_p \leq \left\| \sum_{k \in \mathbb{Z}^d} \alpha_k 2^{\frac{jd}{2}} \varphi(2^j x - k) \right\|_p \leq c_2 2^{j(\frac{d}{2} - \frac{d}{p})} \|(\alpha_k)\|_p.$$

In Härdle et al. (1998), the authors assume a weaker condition than m -regularity. For $d = 1$, the proof of the lemma can be found in Härdle et al. (1998). Similar arguments work as well for $d \geq 2$. In addition, Lemma 1.2 holds if the scaling function φ is replaced by the corresponding wavelet ψ .

1.2 Problem and main theorem

In this paper we aim to estimate the unknown regression function

$$r(x) = E(\rho(V)|U = x), \quad x \in \mathbb{R}^d \quad (2)$$

from a strong mixing sequence (see Definition 1.1) of bivariate random variables $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ with the common density function

$$f(x, y) = \frac{\omega(x, y)g(x, y)}{\mu}, \quad (x, y) \in \mathbb{R}^d \times \mathbb{R}, \tag{3}$$

where ω denotes a weight function, g stands for the density function of the unobserved bivariate random variable (U, V) and $\mu = E\omega(U, V) < \infty$. In addition, $h(x)$ is assumed to be the known density of U with compact support on $[0, 1]^d$, as in Chaubey et al. (2013) and Chaubey and Shirazi (2015). Throughout the paper, we always require $\text{supp } X_i \subseteq [0, 1]^d$.

Definition 1.1 A strictly stationary sequence of random vectors $\{X_i\}_{i \in \mathbb{Z}}$ 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 σ field generated by $\{X_i\}_{i \leq 0}$ and F_k^∞ does by $\{X_i\}_{i \geq k}$.

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. In addition, $\{X_i\}$ is said to be geometrically strong mixing, when $\alpha(k) \leq \gamma\delta^k$ for some $\gamma > 0$ and $0 < \delta < 1$. Now, we provide two examples for geometrically strong mixing data.

Example 1 (Kulik 2008) Let $X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$ with

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

Then it can be proved by Theorem 2 and Corollary 1 of Doukhan (1994) on Page 58 that $\{X_t, t \in \mathbb{Z}\}$ is a geometrically strong mixing sequence.

Example 2 (Mokkadem 1988) Let $\{\varepsilon(t), t \in \mathbb{Z}\} \stackrel{i.i.d.}{\sim} N_r(\mathbf{0}, \Sigma)$ (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 geometrically strong mixing.

Those two important examples tell us that the strong mixing data does not reduce to the classical *i.i.d*, although $\alpha(k)$ goes to zero in the so fast way, $\alpha(k) = O(\delta^k)$.

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 ω, ρ, h in (2) and (3) to be Borel measurable. A Borel measurable function f on \mathbb{R}^d means $\{x \in \mathbb{R}^d, f(x) > c\}$ being a Borel set for each $c \in \mathbb{R}$. In that case, we can prove easily that $\{f(X_i)\}$ remains strong mixing and $\alpha_{f(X)}(k) \leq \alpha_X(k)$ ($k = 1, 2, \dots$), if $\{X_i\}$ has the same property, see Guo (2016). This note is important for the proofs of Propositions 2.2 and 2.3.

Before introducing our estimators, we formulate the following assumptions:

H1. The density function h of the random variable U has a positive lower bound,

$$\inf_{x \in [0, 1]^d} h(x) \geq c_1 > 0.$$

H2. The weight function ω has both positive upper and lower bounds, i.e., for $(x, y) \in [0, 1]^d \times \mathbb{R}$,

$$0 < c_2 \leq \omega(x, y) \leq c_3 < +\infty.$$

H3. There exists a constant $c_4 > 0$ such that,

$$\sup_{y \in \mathbb{R}} |\rho(y)| \leq c_4, \quad \int_{\mathbb{R}} |\rho(y)| dy \leq c_4.$$

H4. The strong mixing coefficient of $\{(X_i, Y_i), i = 1, 2, \dots, n\}$ satisfies $\alpha(k) \leq \gamma e^{-c_5 k}$ with $\gamma > 0, c_5 > 0$.

H5. The density $f_{(X_1, Y_1, X_{k+1}, Y_{k+1})}$ of $(X_1, Y_1, X_{k+1}, Y_{k+1})$ ($k \geq 1$) and the density $f_{(X_1, Y_1)}$ of (X_1, Y_1) satisfy that for $(x, y, x^*, y^*) \in [0, 1]^d \times \mathbb{R} \times [0, 1]^d \times \mathbb{R}$,

$$\sup_{k \geq 1} \sup_{(x, y, x^*, y^*) \in [0, 1]^d \times \mathbb{R} \times [0, 1]^d \times \mathbb{R}} |h_k(x, y, x^*, y^*)| \leq c_6,$$

where $h_k(x, y, x^*, y^*) = f_{(X_1, Y_1, X_{k+1}, Y_{k+1})}(x, y, x^*, y^*) - f_{(X_1, Y_1)}(x, y) f_{(X_{k+1}, Y_{k+1})}(x^*, y^*)$ and $c_6 > 0$.

The assumptions H1 and H2 are standard for the nonparametric regression model with biased data (Chaubey et al. 2013; Chesneau and Shirazi 2014). In Chaubey and Shirazi (2015), the authors assume $h \equiv 1$. While $Y \in [a, b]$ is required by Chaubey et al. (2013). Condition H5 can be viewed as a ‘Castellana-Leadbetter’ type condition in Masry (2000).

We choose d -dimensional scaling function

$$\varphi(x) = \varphi(x_1, \dots, x_d) := D_{2N}(x_1) \cdots D_{2N}(x_d)$$

with $D_{2N}(\cdot)$ being the one-dimensional Daubechies scaling function. Then φ 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 $r \in L^2(\mathbb{R}^d)$ with $\text{supp } r \subseteq [0, 1]^d$ and $M = 2^d - 1$,

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

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

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

We introduce

$$\widehat{\mu}_n = \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{\omega(X_i, Y_i)} \right]^{-1}, \tag{4}$$

$$\widehat{\alpha}_{j_0,k} = \frac{\widehat{\mu}_n}{n} \sum_{i=1}^n \frac{\rho(Y_i)}{\omega(X_i, Y_i) h(X_i)} \varphi_{j_0,k}(X_i) \tag{5}$$

and

$$\widehat{\beta}_{j,k}^{\ell} = \frac{\widehat{\mu}_n}{n} \sum_{i=1}^n \frac{\rho(Y_i)}{\omega(X_i, Y_i) h(X_i)} \psi_{j,k}^{\ell}(X_i). \tag{6}$$

By H1–H3, the definitions in (4)–(6) are all well-defined. When $\rho(Y) = Y$, $\widehat{\mu}_n$ and $\widehat{\alpha}_{j_0,k}$ in (4) and (5) are the same as those of Chaubey et al. (2013). If $d = 1$ and $h(x) = I_{[0,1]}(x)$, then $\widehat{\mu}_n$, $\widehat{\alpha}_{j_0,k}$ and $\widehat{\beta}_{j,k}^{\ell}$ in (4)–(6) reduce completely to those in Chaubey and Shirazi (2015).

We define our linear wavelet estimator

$$\widehat{r}_n^{lin}(x) = \sum_{k \in \Lambda_{j_0}} \widehat{\alpha}_{j_0,k} \varphi_{j_0,k}(x) \tag{7}$$

and the nonlinear wavelet estimator

$$\widehat{r}_n^{non}(x) = \widehat{r}_n^{lin}(x) + \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \widehat{\beta}_{j,k}^{\ell} I_{\{|\widehat{\beta}_{j,k}^{\ell}| \geq \kappa t_n\}} \psi_{j,k}^{\ell}(x), \tag{8}$$

where t_n , j_0 and j_1 are specified in the Main Theorem, while the constant κ will be chosen in the proof of the theorem.

Comparing with the wavelet estimators in Chaubey et al. (2013) and Chaubey and Shirazi (2015), we use a wavelet basis in the whole space instead of wavelets on an interval. In the later case, boundary elements need be treated appropriately.

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

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

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 $B \lesssim A$. The indicator function on a set G is denoted by I_G as usual.

Main Theorem Consider the problem defined by (2) and (3) under the assumptions H1–H5. Let $r \in B_{\tilde{p},q}^s(H)$, $\tilde{p}, q \in [1, \infty)$, $s > 0$, $\text{supp } r \subseteq [0, 1]^d$ and either $\tilde{p} \geq p$ or $\tilde{p} \leq p < \infty$ and $s > \frac{d}{\tilde{p}}$. Then for $1 \leq p < +\infty$, the linear wavelet estimator \hat{r}_n^{lin} defined in (7) with $2^{j_0} \sim n^{\frac{1}{2s'+d+dI_{\{p>2\}}}}$ and $s' = s - d(\frac{1}{\tilde{p}} - \frac{1}{p})_+$ satisfies

$$E \int_{[0,1]^d} \left| \hat{r}_n^{lin}(x) - r(x) \right|^p dx \lesssim n^{-\frac{s'p}{2s'+d+dI_{\{p>2\}}}}; \tag{9a}$$

The nonlinear estimator in (8) with $2^{j_0} \sim n^{\frac{1}{2m+d+dI_{\{p>2\}}}}$ ($m > s$), $2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{d}}$ and $t_n = \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{jd}{2}} I_{\{p > 2\}} \right] \sqrt{\frac{\ln n}{n}}$ satisfies

$$E \int_{[0,1]^d} \left| \hat{r}_n^{non}(x) - r(x) \right|^p dx \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p}, \tag{9b}$$

where

$$\alpha = \begin{cases} \frac{s}{2s+d+dI_{\{p>2\}}}, & \tilde{p} > \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}}, \\ \frac{s-d/\tilde{p}+d/p}{2(s-d/\tilde{p})+d+dI_{\{p>2\}}}, & \tilde{p} \leq \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}}. \end{cases} \tag{9c}$$

Remark 1 When $p = 2$, (9a) reduces to Theorem 4.1 of Chaubey et al. (2013); If $p = 2$ and $d = 1$, (9b) becomes Theorem 5.1 in Chaubey and Shirazi (2015) up to a $\ln n$ factor.

In contrast to the linear wavelet estimator \hat{r}_n^{lin} , the nonlinear estimator \hat{r}_n^{non} is adaptive, which means both j_0 and j_1 do not depend on s, \tilde{p} and q . On the other hand, the convergence rate of the nonlinear estimator remains the same as that of the linear estimator up to $(\ln n)^{\frac{3p}{2}}$, when $\tilde{p} \geq p$. However, it gets better for $\tilde{p} < p$. The same situation happens for *i.i.d.* case.

Remark 2 Compared with the estimation for *i.i.d.* data in Kou and Liu (2017), the convergence rate of Main Theorem keeps the same (up to a $\ln n$ factor), when $p \in [1, 2]$. However, it becomes worse for $p > 2$. This exhibits a major difference between those two types of data.

For *i.i.d.* case, a lower bound estimation under L^p risk is provided by Kou and Liu (2016). It is a challenging problem for strong mixing data.

Remark 3 From (9a)–(9c) in our Main Theorem, we find that the convergence rates close to zero, when the dimension d gets very large (curse of dimensionality). In fact, the same situation happens for the classical *i.i.d.* case (Delyon and Judisky 1996; Kou and Liu 2017). To reduce the influence of the dimension d on the accuracy of estimation, a known method is to assume some independent structure of the samples

(Rebelles 2015a, b; Lepski 2013). Since the strong mixing data is much more complicated than the *i.i.d* one, it would be a challenging problem to do the same in our setting. We shall study it in future.

2 Three propositions

In this section, we provide three propositions for the proof of the Main Theorem which is given in Sect. 3. Clearly, $\mu := E\omega(U, V) > 0$ under the condition H2. Moreover, the following simple (but important) lemma holds.

Lemma 2.1 *For the problem defined in (2) and (3) and $\widehat{\mu}_n$ given by (4),*

$$E(\widehat{\mu}_n^{-1}) = \mu^{-1}, \tag{10a}$$

$$E \left[\frac{\mu\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \varphi_{j_0,k}(X_i) \right] = \alpha_{j_0,k}, \tag{10b}$$

$$E \left[\frac{\mu\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \psi_{j,k}^\ell(X_i) \right] = \beta_{j,k}^\ell, \tag{10c}$$

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

Proof One includes a simple proof for completeness, although it is essentially the same as that of Chaubey et al. (2013). By (4),

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

This with (3) leads to

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

which concludes (10a). Using (3) and (2), one knows that

$$\begin{aligned} E \left[\frac{\mu\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \varphi_{j_0,k}(X_i) \right] &= \int_{[0,1]^d \times \mathbb{R}} \frac{\mu\rho(y)}{\omega(x, y)h(x)} \varphi_{j_0,k}(x) f(x, y) dx dy \\ &= \int_{[0,1]^d} \varphi_{j_0,k}(x) \int_{\mathbb{R}} \frac{\rho(y)}{h(x)} g(x, y) dy dx \\ &= \int_{[0,1]^d} r(x)\varphi_{j_0,k}(x) dx = \alpha_{j_0,k}. \end{aligned}$$

This completes the proof of (10b). Similar arguments show (10c). □

To establish the next two propositions, we need an important Lemma.

Lemma 2.2 (Davydov 1970) *Let $\{X_i\}_{i \in \mathbb{Z}}$ be strong mixing with the mixing coefficient $\alpha(k)$, f and g be two measurable functions. If $E|f(X_1)|^p$ and $E|g(X_1)|^q$ exist for $p, q > 0$ and $\frac{1}{p} + \frac{1}{q} < 1$, then there exists a constant $c > 0$ such that*

$$\left| \text{cov}\left(f(X_1), g(X_{k+1})\right) \right| \leq c \left[\alpha(k) \right]^{1 - \frac{1}{p} - \frac{1}{q}} \left[E|f(X_1)|^p \right]^{\frac{1}{p}} \left[E|g(X_1)|^q \right]^{\frac{1}{q}}.$$

Proposition 2.1 *Let (X_i, Y_i) ($i = 1, 2, \dots, n$) be strong mixing, H1–H5 hold and $2^{jd} \leq n$. Then*

$$\text{var} \left[\sum_{i=1}^n \frac{1}{\omega(X_i, Y_i)} \right] \lesssim n \quad \text{and} \quad \text{var} \left[\sum_{i=1}^n \frac{\rho(Y_i) \psi_{j,k}^\ell(X_i)}{\omega(X_i, Y_i) h(X_i)} \right] \lesssim n. \tag{11}$$

Proof Note that Condition H2 implies $\text{var} \left(\frac{1}{\omega(X_i, Y_i)} \right) \leq E \left(\frac{1}{\omega(X_i, Y_i)} \right)^2 \lesssim 1$ and

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

Then it suffices to show

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

for the first inequality of (11). By the strict stationarity of (X_i, Y_i) ,

$$\begin{aligned} & \left| \sum_{v=2}^n \sum_{i=1}^{v-1} \text{cov} \left(\frac{1}{\omega(X_v, Y_v)}, \frac{1}{\omega(X_i, Y_i)} \right) \right| \\ &= \left| \sum_{m=1}^n (n - m) \text{cov} \left(\frac{1}{\omega(X_1, Y_1)}, \frac{1}{\omega(X_{m+1}, Y_{m+1})} \right) \right| \\ &\leq n \sum_{m=1}^n \left| \text{cov} \left(\frac{1}{\omega(X_1, Y_1)}, \frac{1}{\omega(X_{m+1}, Y_{m+1})} \right) \right|. \end{aligned}$$

On the other hand, Lemma 2.2 and H2 show that

$$\left| \text{cov} \left(\frac{1}{\omega(X_1, Y_1)}, \frac{1}{\omega(X_{m+1}, Y_{m+1})} \right) \right| \lesssim \sqrt{\alpha(m)} \sqrt{E \left| \frac{1}{\omega(X_1, Y_1)} \right|^4} \lesssim \sqrt{\alpha(m)}.$$

These with H4 give the desired conclusion (12),

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

To prove the second inequality of (11), one observes

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

By (3) and H1–H3, the first term of the above inequality is bounded by

$$nE \left(\frac{\rho(Y_i)\psi_{j,k}^\ell(X_i)}{\omega(X_i, Y_i)h(X_i)} \right)^2 \lesssim n \int_{[0,1]^d \times \mathbb{R}} \left[\psi_{j,k}^\ell(x) \right]^2 \frac{g(x, y)}{h(x)} dy dx = n.$$

It remains to show

$$\begin{aligned} &\left| \sum_{v=2}^n \sum_{i=1}^{v-1} \text{cov} \left(\frac{\rho(Y_v)\psi_{j,k}^\ell(X_v)}{\omega(X_v, Y_v)h(X_v)}, \frac{\rho(Y_i)\psi_{j,k}^\ell(X_i)}{\omega(X_i, Y_i)h(X_i)} \right) \right| \\ &\leq n \left(\sum_{m=1}^{2^{jd}-1} + \sum_{m=2^{jd}}^n \right) \left| \text{cov} \left[\frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)}, \frac{\rho(Y_{m+1})\psi_{j,k}^\ell(X_{m+1})}{\omega(X_{m+1}, Y_{m+1})h(X_{m+1})} \right] \right| \lesssim n \end{aligned} \tag{13}$$

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

According to H5 and H1–H3,

$$\begin{aligned} &\left| \text{cov} \left(\frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)}, \frac{\rho(Y_{m+1})\psi_{j,k}^\ell(X_{m+1})}{\omega(X_{m+1}, Y_{m+1})h(X_{m+1})} \right) \right| \\ &\leq \int_{[0,1]^d \times \mathbb{R} \times [0,1]^d \times \mathbb{R}} \left| \frac{\rho(y)\psi_{j,k}^\ell(x)}{\omega(x, y)h(x)} \cdot \frac{\rho(y^*)\psi_{j,k}^\ell(x^*)}{\omega(x^*, y^*)h(x^*)} \right| |h_m(x, y, x^*, y^*)| dx dy dx^* dy^* \\ &\lesssim \left(\int_{\mathbb{R}} |\rho(y)| dy \right)^2 \left(\int_{[0,1]^d} |\psi_{j,k}^\ell(x)| dx \right)^2 \lesssim 2^{-jd}. \end{aligned}$$

Hence,

$$\sum_{m=1}^{2^{jd}-1} \left| \text{cov} \left(\frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)}, \frac{\rho(Y_{m+1})\psi_{j,k}^\ell(X_{m+1})}{\omega(X_{m+1}, Y_{m+1})h(X_{m+1})} \right) \right| \lesssim \sum_{m=1}^{2^{jd}-1} 2^{-jd} \lesssim 1. \tag{14}$$

On the other hand, Lemma 2.2, H1–H3 and the arguments before (13) tell that

$$\begin{aligned} & \left| \text{cov} \left(\frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)}, \frac{\rho(Y_{m+1})\psi_{j,k}^\ell(X_{m+1})}{\omega(X_{m+1}, Y_{m+1})h(X_{m+1})} \right) \right| \lesssim \sqrt{\alpha(m)} \sqrt{E \left| \frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)} \right|^4} \\ & \lesssim \sqrt{\alpha(m)} \sup \left| \frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)} \right| \sqrt{E \left| \frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)} \right|^2} \lesssim \sqrt{\alpha(m)} 2^{\frac{jd}{2}}. \end{aligned}$$

Moreover, $\sum_{m=2^{jd}}^n \left| \text{cov} \left(\frac{\rho(Y_1)\psi_{j,k}^\ell(X_1)}{\omega(X_1, Y_1)h(X_1)}, \frac{\rho(Y_{m+1})\psi_{j,k}^\ell(X_{m+1})}{\omega(X_{m+1}, Y_{m+1})h(X_{m+1})} \right) \right| \lesssim \sum_{m=2^{jd}}^n \sqrt{\alpha(m)} 2^{\frac{jd}{2}} \lesssim \sum_{m=1}^n \sqrt{m\alpha(m)} \leq \sum_{m=1}^{+\infty} m^{\frac{1}{2}} \gamma e^{-\frac{cm}{2}} < +\infty$. This with (14) shows (13). \square

To estimate $E \left| \widehat{\alpha}_{j_0,k} - \alpha_{j_0,k} \right|^p$ and $E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p$, we introduce a moment bound, which can be found in Yokoyama (1980), Kim (1993) and Shao and Yu (1996).

Lemma 2.3 *Let $\{X_i, i = 1, 2, \dots, n\}$ be a strong mixing sequence of random variables with the mixing coefficients $\alpha(n) \leq cn^{-\theta}$ ($c > 0, \theta > 0$). If $EX_i = 0, \|X_i\|_\eta := (E|X_i|^\eta)^{\frac{1}{\eta}} < \infty, 2 < p < \eta < +\infty$ and $\theta > \frac{p}{2(\eta-p)}$, then there exists $K = K(p, \eta, \theta) < \infty$ such that*

$$E \left| \sum_{i=1}^n X_i \right|^p \leq K \|X_i\|_\eta^p n^{\frac{p}{2}}.$$

Proposition 2.2 *Let $r \in B_{\tilde{p},q}^s(H)$ ($\tilde{p}, q \in [1, \infty), s > 0$) and $\widehat{\alpha}_{j_0,k}, \widehat{\beta}_{j,k}^\ell$ be defined by (5) and (6). If H1–H5 hold, then*

$$E \left| \widehat{\alpha}_{j_0,k} - \alpha_{j_0,k} \right|^p \lesssim \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{j_0 dp}{2}} I_{\{p > 2\}} \right] n^{-\frac{p}{2}}, \tag{15a}$$

$$E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p \lesssim \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{j dp}{2}} I_{\{p > 2\}} \right] n^{-\frac{p}{2}}. \tag{15b}$$

Proof One proves (15b) only, (15a) is similar. By the definition of $\widehat{\beta}_{j,k}^\ell$,

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

and $E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p \lesssim E \left| \frac{\widehat{\mu}_n}{\mu} \left[\frac{\mu}{n} \sum_{i=1}^n \frac{\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \psi_{j,k}^\ell(X_i) - \beta_{j,k}^\ell \right] \right|^p + E \left| \beta_{j,k}^\ell \widehat{\mu}_n \left(\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n} \right) \right|^p$. Since Condition H3 implies the boundedness of $r, \left| \beta_{j,k}^\ell \right| := \left| \int_{[0,1]^d} r(x) \psi_{j,k}^\ell(x) dx \right| \lesssim 1$ thanks to Hölder inequality and orthonormality of $\{\psi_{j,k}^\ell\}$.

On the other hand, $\left| \frac{\widehat{\mu}_n}{\mu} \right| \lesssim 1$ and $|\widehat{\mu}_n| \lesssim 1$ because of H2. Hence,

$$E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p \lesssim E \left| \frac{\mu}{n} \sum_{i=1}^n \frac{\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \psi_{j,k}^\ell(X_i) - \beta_{j,k}^\ell \right|^p + E \left| \frac{1}{\mu} - \frac{1}{\widehat{\mu}_n} \right|^p. \tag{16}$$

When $p = 2$, it is easy to see from Lemma 2.1 and Proposition 2.1 that $E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^2 \lesssim \text{var} \left[\frac{1}{n} \sum_{i=1}^n \frac{\rho(Y_i)\psi_{j,k}^\ell(X_i)}{\omega(X_i, Y_i)h(X_i)} \right] + \text{var} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{\omega(X_i, Y_i)} \right] \lesssim \frac{1}{n}$. This with Jensen’s inequality shows $E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p \lesssim n^{-\frac{p}{2}}$ for $1 \leq p \leq 2$.

It remains to show (15b) true for $p > 2$. By the definition of $\widehat{\mu}_n$,

$$E \left| \frac{1}{\mu} - \frac{1}{\widehat{\mu}_n} \right|^p = E \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{\omega(X_i, Y_i)} - \frac{1}{\mu} \right|^p = E \left| \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\omega(X_i, Y_i)} - \frac{1}{\mu} \right] \right|^p. \tag{17}$$

Let $\eta_i := \frac{1}{\omega(X_i, Y_i)} - \frac{1}{\mu}$. Then $E(\eta_i) = 0$ thanks to (10a). Furthermore, η_1, \dots, η_n are strong mixing by the same property of $\{(X_i, Y_i)\}$ and the Borel measurability of the function $\frac{1}{\omega(x, y)} - \frac{1}{\mu}$ (Guo 2016). On the other hand, Condition H2 implies $|\eta_i| \lesssim 1$ and $\|\eta_i\|_\eta^p \lesssim 1$. By Condition H4, θ in Lemma 2.3 can be taken large enough so that $\theta > \frac{pn}{2(\eta-p)}$ for fixed p and η with $2 < p < \eta < +\infty$. Then it follows from Lemma 2.3 and (17) that

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

Finally, one needs only to show

$$Q_n := E \left| \frac{\mu}{n} \sum_{i=1}^n \frac{\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \psi_{j,k}^\ell(X_i) - \beta_{j,k}^\ell \right|^p \lesssim 2^{\frac{jd p}{2}} n^{-\frac{p}{2}}. \tag{18}$$

Define $\xi_i := \frac{\mu\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \psi_{j,k}^\ell(X_i) - \beta_{j,k}^\ell$. Then similar arguments to η_i conclude that $E(\xi_i) = 0$, $Q_n = \frac{1}{n^p} E \left| \sum_{i=1}^n \xi_i \right|^p$ and ξ_1, \dots, ξ_n are strong mixing with the mixing coefficients $\alpha(k) \leq \gamma e^{-ck}$. According to H1–H3 and $|\psi^\ell(x)| \lesssim 1$,

$$\left| \frac{\mu\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \psi_{j,k}^\ell(X_i) \right| \lesssim 2^{\frac{jd}{2}}.$$

This with $\beta_{j,k}^\ell = E \left[\frac{\mu\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \psi_{j,k}^\ell(X_i) \right]$ leads to $E|\xi_i|^\eta \lesssim 2^{\frac{jd\eta}{2}}$ and $\|\xi_i\|_\eta^p \lesssim 2^{\frac{jd p}{2}}$. Using Lemma 2.3 again, one obtains the desired conclusion (18). \square

To prove the last proposition in this section, we need the following Bernstein-type inequality (Liebscher 1996, 2001; Rio 1995).

Lemma 2.4 *Let $(X_i)_{i \in \mathbb{Z}}$ be a strong mixing process with mixing coefficient $\alpha(k)$, $EX_i = 0$, $|X_i| \leq M < \infty$ and $D_m = \max_{1 \leq j \leq 2m} \text{var} \left(\sum_{i=1}^j X_i \right)$. Then for $\varepsilon > 0$ and $n, m \in \mathbb{N}$ with $0 < m \leq \frac{n}{2}$,*

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

From the next proposition, we realize the reason for choosing $2^{j_1} \sim \left(\frac{n}{(\ln n)^3} \right)^{\frac{1}{d}}$ in our Main Theorem. The classical choice is $2^{j_1} \sim \left(\frac{n}{\ln n} \right)^{\frac{1}{d}}$, see Chesneau and Shirazi (2014).

Proposition 2.3 *Let $r \in B_{\tilde{p},q}^s(H)$ ($\tilde{p}, q \in [1, \infty)$, $s > 0$), $\widehat{\beta}_{j,k}^\ell$ be defined in (6) and $t_n = \sqrt{\frac{\ln n}{n}}$. If H1–H5 hold and $2^{jd} \leq \frac{n}{(\ln n)^3}$, then for $w > 0$, there exists a constant $\kappa > 1$ such that*

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

Proof According to the arguments of (16), $\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right| \lesssim \frac{1}{n} \left| \sum_{i=1}^n \left[\frac{1}{\omega(X_i, Y_i)} - \frac{1}{\mu} \right] \right| + \left| \frac{1}{n} \sum_{i=1}^n \frac{\mu \rho(Y_i)}{\omega(X_i, Y_i) h(X_i)} \psi_{j,k}^\ell(X_i) - \beta_{j,k}^\ell \right|$. Hence, it suffices to prove

$$\begin{aligned} \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \left[\frac{1}{\omega(X_i, Y_i)} - \frac{1}{\mu} \right] \right| \geq \frac{\kappa}{2} t_n \right) &\lesssim 2^{-wj} \quad \text{and} \\ \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \left[\frac{\mu \rho(Y_i)}{\omega(X_i, Y_i) h(X_i)} \psi_{j,k}^\ell(X_i) - \beta_{j,k}^\ell \right] \right| \geq \frac{\kappa}{2} t_n \right) &\lesssim 2^{-wj}. \end{aligned} \tag{19}$$

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

Define $\xi_i := \frac{\mu \rho(Y_i)}{\omega(X_i, Y_i) h(X_i)} \psi_{j,k}^\ell(X_i) - \beta_{j,k}^\ell$. Then $E(\xi_i) = 0$ thanks to (10c), and ξ_1, \dots, ξ_n are strong mixing with the mixing coefficients $\alpha(k) \leq \gamma e^{-ck}$ because of Condition H4. By H1–H3, $\left| \frac{\mu \rho(Y_i)}{\omega(X_i, Y_i) h(X_i)} \psi_{j,k}^\ell(X_i) \right| \lesssim 2^{\frac{jd}{2}}$ and

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

According to Proposition 2.1, $D_m = \max_{1 \leq j \leq 2m} \text{var} \left(\sum_{i=1}^j \xi_i \right) \lesssim m$. Then it follows from Lemma 2.4 with $m = u \ln n$ (the constant u will be chosen later on) that

$$\begin{aligned} \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \xi_i \right| \geq \frac{\kappa}{2} t_n \right) &= \mathbb{P} \left(\left| \sum_{i=1}^n \xi_i \right| \geq \frac{\kappa}{2} n t_n \right) \\ &\lesssim \exp \left\{ -\frac{(\kappa n t_n)^2}{64} \left(n m^{-1} D_m + \frac{1}{6} \kappa n t_n 2^{\frac{jd}{2}} m \right)^{-1} \right\} \\ &\quad + 64 \frac{2^{\frac{jd}{2}}}{\kappa n t_n} n \gamma e^{-cm}. \end{aligned} \tag{20}$$

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 $1 - cu < -\frac{w}{d}$, then the second term of (20) is bounded by 2^{-wj} . On the other hand, the first one of (20) 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 $m = u \ln n$. For this estimation, $2^{jd} \leq \frac{n}{(\ln n)^3}$ is essential, one can not replace that condition by $2^{jd} \leq \frac{n}{\ln n}$. Obviously, there exists sufficiently large $\kappa > 1$ such that $\exp \left\{ -\frac{\kappa^2 \ln n}{64} \left(1 + \frac{1}{6} \kappa u \right)^{-1} \right\} \lesssim 2^{-wj}$. Finally, the desired conclusion (19) follows. \square

3 Proof of main theorem

This section proves the Main Theorem. We rewrite it as Theorem 3.1. The main idea of the proof comes from Donoho et al. (1996). When $p = 2$, the corresponding estimates seems easier (see Chaubey et al. 2013; Chaubey and Shirazi 2015), because $L^2(\mathbb{R}^d)$ is a Hilbert space.

Theorem 3.1 *Consider the problem defined by (2) and (3) with the assumptions H1–H5. Let $r \in B_{\tilde{p},q}^s(H)(\tilde{p}, q \in [1, \infty), s > 0)$, $\text{supp } r \subseteq [0, 1]^d$ and either $\tilde{p} \geq p$ or $\tilde{p} \leq p < \infty$ and $s > \frac{d}{\tilde{p}}$. Then for $1 \leq p < +\infty$, the linear wavelet estimator \hat{r}_n^{lin} defined in (7) with $2^{j_0} \sim n^{\frac{1}{2s'+d+dI_{(p>2)}}$ and $s' = s - d \left(\frac{1}{\tilde{p}} - \frac{1}{p} \right)_+$ satisfies*

$$E \int_{[0,1]^d} \left| \hat{r}_n^{\text{lin}}(x) - r(x) \right|^p dx \lesssim n^{-\frac{s'p}{2s'+d+dI_{(p>2)}}}; \tag{21a}$$

The nonlinear one in (8) with $2^{j_0} \sim n^{\frac{1}{2m+d+dI_{\{p>2\}}}}$ ($m > s$), $2^{j_1} \sim \left(\frac{n}{(\ln n)^3}\right)^{\frac{1}{d}}$ and $t_n = \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{jd}{2}} I_{\{p>2\}} \right] \sqrt{\frac{\ln n}{n}}$ satisfies

$$E \int_{[0,1]^d} \left| \widehat{r}_n^{non}(x) - r(x) \right|^p dx \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p}, \tag{21b}$$

where

$$\alpha = \begin{cases} \frac{s}{2s+d+dI_{\{p>2\}}}, & \widetilde{p} > \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}}, \\ \frac{s-d/\widetilde{p}+d/p}{2(s-d/\widetilde{p})+d+dI_{\{p>2\}}}, & \widetilde{p} \leq \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}}. \end{cases} \tag{21c}$$

Proof When $\widetilde{p} \leq p$ and $s > \frac{d}{\widetilde{p}}$, $s' - \frac{d}{\widetilde{p}} = s - \frac{d}{\widetilde{p}} > 0$. By the equivalence of (1) and (3) in Lemma 1.1, $B_{\widetilde{p},q}^s(H) \subseteq B_{p,q}^{s'}(H')$ for some $H' > 0$. Then $r \in B_{p,q}^{s'}(H')$ and $\|P_{j_0}r - r\|_p = \|\sum_{j=j_0}^\infty (P_{j+1}r - P_jr)\|_p \lesssim \sum_{j=j_0}^\infty \|P_{j+1}r - P_jr\|_p \lesssim \sum_{j=j_0}^\infty 2^{-js'} \lesssim 2^{-j_0s'}$ thanks to (2) of Lemma 1.1. Moreover,

$$\|P_{j_0}r - r\|_p^p \lesssim 2^{-j_0s'p}. \tag{22}$$

It is easy to see from Lemma 1.2 that

$$E \left\| \widehat{r}_n^{lin} - P_{j_0}r \right\|_p^p = E \left\| \sum_{k \in \Lambda_{j_0}} (\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}) \varphi_{j_0,k} \right\|_p^p \lesssim 2^{pd\left(\frac{j_0}{2} - \frac{j_0}{\widetilde{p}}\right)} \sum_{k \in \Lambda_{j_0}} E \left| \widehat{\alpha}_{j_0,k} - \alpha_{j_0,k} \right|^p.$$

According to Proposition 2.2 and $|\Lambda_{j_0}| \sim 2^{j_0d}$,

$$E \left\| \widehat{r}_n^{lin} - P_{j_0}r \right\|_p^p \lesssim 2^{\frac{j_0dp}{2}} \left(I_{\{1 \leq p \leq 2\}} + 2^{\frac{j_0dp}{2}} I_{\{p>2\}} \right) n^{-\frac{p}{2}}. \tag{23}$$

This with (22) shows that $E \int_{[0,1]^d} \left| \widehat{r}_n^{lin}(x) - r(x) \right|^p dx \leq E \int_{\mathbb{R}^d} \left| \widehat{r}_n^{lin}(x) - r(x) \right|^p dx \lesssim E \left\| \widehat{r}_n^{lin} - P_{j_0}r \right\|_p^p + \|P_{j_0}r - r\|_p^p \lesssim 2^{-j_0s'p} + 2^{\frac{j_0dp}{2}} \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{j_0dp}{2}} I_{\{p>2\}} \right] n^{-\frac{p}{2}}$. To get a balance, one chooses $2^{j_0} \sim n^{\frac{1}{2s'+d+dI_{\{p>2\}}}}$. Then

$$E \int_{[0,1]^d} \left| \widehat{r}_n^{lin}(x) - r(x) \right|^p dx \lesssim n^{-\frac{s'p}{2s'+d+dI_{\{p>2\}}}}, \tag{24}$$

which is the desired conclusion (21a) for $\widetilde{p} \leq p$ and $s > \frac{d}{\widetilde{p}}$.

From the above arguments, one finds that when $p = \widetilde{p}$, $s' = s > 0$ and the inequality (24) still holds without the assumption $s > \frac{d}{\widetilde{p}}$. It remains to conclude (21a) for $\widetilde{p} > p \geq 1$. By Hölder inequality,

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

Using Jensen inequality and (24) with $p = \tilde{p}$, one gets

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

This completes the proof of (21a).

Similar to the arguments of (21a), it suffices to prove (21b) for $\tilde{p} \leq p$ and $s > \frac{d}{\tilde{p}}$. In this case, (21c) can be rewritten as

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

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

$$E \int_{[0,1]^d} \left| \widehat{r}_n^{non}(x) - r(x) \right|^p dx \lesssim I_1 + I_2 + Z, \tag{25}$$

where $I_1 := E \left\| \widehat{r}_n^{lin} - P_{j_0}r \right\|_p^p$, $I_2 := \left\| r - P_{j_1+1}r \right\|_p^p$ and

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

According to (23), $2^{j_0} \sim n^{\frac{1}{2m+d+dI_{\{p>2\}}}}$ ($m > s$) and the definition of α in (21c),

$$I_1 = E \left\| \widehat{r}_n^{lin} - P_{j_0}r \right\|_p^p \lesssim 2^{\frac{j_0 d p}{2}} \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{j_0 d p}{2}} I_{\{p > 2\}} \right] n^{-\frac{p}{2}} \leq n^{-\alpha p}.$$

The same arguments as (22) shows $\left\| P_{j_1+1}r - r \right\|_p^p \lesssim 2^{-j_1 s' p}$. On the other hand, $\frac{s'}{d} = \frac{s}{d} - \frac{1}{\tilde{p}} + \frac{1}{p} \geq \alpha$ thanks to $\tilde{p} \leq p$ and $s > \frac{d}{\tilde{p}}$. Then it follows from $2^{j_1} \sim \left(\frac{n}{(\ln n)^3} \right)^{\frac{1}{d}}$ and $0 < \alpha < \frac{1}{2}$ that

$$I_2 = \left\| P_{j_1+1}r - r \right\|_p^p \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p}.$$

The main work for the proof of (21b) is to show

$$Z = E \left\| \sum_{j=j_0}^{j_1} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \left[\widehat{\beta}_{j,k}^\ell I_{\{|\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n\}} - \beta_{j,k}^\ell \right] \psi_{j,k}^\ell \right\|_p^p \lesssim (\ln n)^{\frac{3p}{2}} n^{-\alpha p}. \tag{26}$$

Note that Lemma 1.2 tells

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

Then the classical technique (see Donoho et al. 1996) gives

$$Z \lesssim (j_1 - j_0 + 1)^{p-1} (Z_1 + Z_2 + Z_3), \tag{27}$$

where

$$\begin{aligned} Z_1 &= \sum_{j=j_0}^{j_1} 2^{pd(\frac{j}{2} - \frac{j}{p})} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E \left[\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p I_{\{|\widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell| > \frac{\kappa t_n}{2}\}} \right], \\ Z_2 &= \sum_{j=j_0}^{j_1} 2^{pd(\frac{j}{2} - \frac{j}{p})} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E \left[\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p I_{\{|\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n, |\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2}\}} \right], \\ Z_3 &= \sum_{j=j_0}^{j_1} 2^{pd(\frac{j}{2} - \frac{j}{p})} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \left| \beta_{j,k}^\ell \right|^p I_{\{|\widehat{\beta}_{j,k}^\ell| < \kappa t_n, |\beta_{j,k}^\ell| \leq 2\kappa t_n\}}. \end{aligned}$$

For Z_1 , one observes that

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

thanks to Hölder inequality. By Proposition 2.3,

$$\mathbb{P} \left(\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right| > \kappa 2^{\frac{jd}{2}} \sqrt{\frac{\ln n}{n}} \right) \leq \mathbb{P} \left(\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right| > \kappa \sqrt{\frac{\ln n}{n}} \right) \lesssim 2^{-wj}.$$

On the other hand, Proposition 2.2 implies $E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p \lesssim 2^{\frac{jd p}{2}} n^{-\frac{p}{2}}$ for $1 \leq p < +\infty$. Therefore,

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

Then for $w > 2pd$ in Proposition 2.3, $Z_1 \lesssim \sum_{j=j_0}^{j_1} 2^{pd(\frac{j}{2} - \frac{j}{p})} 2^{jd} 2^{\frac{jd p}{2}} n^{-\frac{p}{2}} 2^{-\frac{wj}{2}} \lesssim \left(\frac{1}{n}\right)^{\frac{p}{2}} 2^{-j_0(\frac{w}{2} - pd)} \lesssim \left(\frac{1}{n}\right)^{\frac{p}{2}} \left(\frac{1}{n}\right)^{\frac{\frac{w}{2} - pd}{2m+d+1(p>2)}} \leq \left(\frac{1}{n}\right)^{\alpha p}$, where one uses $\alpha < \frac{1}{2}$ and the choice $2^{j_0} \sim n^{\frac{1}{2m+d+1(p>2)}}$ ($m > s$). Hence,

$$Z_1 \leq n^{-\alpha p}. \tag{28}$$

To estimate Z_2 , one rewrites

$$Z_2 = \left(\sum_{j=j_0}^{j_0^*} + \sum_{j=j_0^*+1}^{j_1} \right) \left\{ 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E \left[\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p I_{\left\{ |\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n, |\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2} \right\}} \right] \right\}$$

$$:= Z_{21} + Z_{22}$$

with the integer $j_0^* \in [j_0, j_1]$ being specified later on. By Proposition 2.2,

$$Z_{21} := \sum_{j=j_0}^{j_0^*} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E \left[\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p I_{\left\{ |\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n, |\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2} \right\}} \right]$$

$$\lesssim \sum_{j=j_0}^{j_0^*} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{jdp}{2}} I_{\{p > 2\}} \right] n^{-\frac{p}{2}}$$

$$\lesssim \left[2^{\frac{j_0^*pd}{2}} I_{\{1 \leq p \leq 2\}} + 2^{j_0^*pd} I_{\{p > 2\}} \right] \left(\frac{1}{n} \right)^{\frac{p}{2}}. \tag{29}$$

On the other hand, it follows from Proposition 2.2, Lemma 1.1(3) and $t_n = \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{jd}{2}} I_{\{p > 2\}} \right] \sqrt{\frac{\ln n}{n}}$ that

$$Z_{22} := \sum_{j=j_0^*+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E \left[\left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p I_{\left\{ |\widehat{\beta}_{j,k}^\ell| \geq \kappa t_n, |\beta_{j,k}^\ell| \geq \frac{\kappa t_n}{2} \right\}} \right]$$

$$\lesssim \sum_{j=j_0^*+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p \left(\frac{|\beta_{j,k}^\ell|}{\kappa t_n/2} \right)^{\widetilde{p}}$$

$$\lesssim \sum_{j=j_0^*+1}^{j_1} (\ln n)^{-\frac{\widetilde{p}}{2}} \left(\frac{1}{n} \right)^{\frac{p-\widetilde{p}}{2}} 2^{-j\left(s\widetilde{p} + \frac{\widetilde{p}d}{2} + \frac{\widetilde{p}d}{2} I_{\{p > 2\}} - \frac{pd}{2} - \frac{pd}{2} I_{\{p > 2\}}\right)}.$$

Define

$$\varepsilon := s\widetilde{p} + \frac{\widetilde{p}d}{2} + \frac{\widetilde{p}d}{2} I_{\{p > 2\}} - \frac{pd}{2} - \frac{pd}{2} I_{\{p > 2\}}.$$

Then for $\varepsilon > 0$, $Z_{22} \lesssim (\ln n)^{-\frac{\widetilde{p}}{2}} \left(\frac{1}{n} \right)^{\frac{p-\widetilde{p}}{2}} \sum_{j=j_0^*+1}^{j_1} 2^{-j\varepsilon} \lesssim (\ln n)^{-\frac{\widetilde{p}}{2}} \left(\frac{1}{n} \right)^{\frac{p-\widetilde{p}}{2}} 2^{-j_0^*\varepsilon}$. To balance this with (29), one takes

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

Note that $0 < \alpha \leq \frac{s}{2s+d+dI_{\{p>2\}}} < \frac{1}{2}$ and $2^{j_0} \sim n^{\frac{1}{2m+d+dI_{\{p>2\}}}}$ ($m > s$). Then $2^{j_0^*} \leq 2^{j_1} \sim (\frac{n}{(\ln n)^3})^{\frac{1}{d}}$ and $2^{j_0^*} \gtrsim (\frac{n}{\ln n})^{\frac{1}{2s+d+dI_{\{p>2\}}}} \gtrsim n^{\frac{1}{2m+d+dI_{\{p>2\}}}} \sim 2^{j_0}$ ($m > s$). Since $\varepsilon > 0$, $\tilde{p} > \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}}$ and $\alpha = \frac{s}{2s+d+dI_{\{p>2\}}}$ thanks to (21c). Moreover, it can be checked that

$$\frac{p - \tilde{p}}{2} + \frac{1 - 2\alpha}{d + dI_{\{p>2\}}} \left(s\tilde{p} + \frac{\tilde{p}d}{2} + \frac{\tilde{p}d}{2} I_{\{p>2\}} - \frac{pd}{2} - \frac{pd}{2} I_{\{p>2\}} \right) = \alpha p$$

by considering $p > 2$ and $p \in [1, 2]$ respectively. This with the choice of $2^{j_0^*}$ leads to

$$Z_{22} \lesssim (\ln n)^{-\frac{\tilde{p}}{2}} \left(\frac{1}{n}\right)^{\frac{p-\tilde{p}}{2}} \left(\frac{\ln n}{n}\right)^{\frac{1-2\alpha}{d+dI_{\{p>2\}}}} \left(s\tilde{p} + \frac{\tilde{p}d}{2} + \frac{\tilde{p}d}{2} I_{\{p>2\}} - \frac{pd}{2} - \frac{pd}{2} I_{\{p>2\}} \right) \lesssim \left(\frac{1}{n}\right)^{\alpha p} \tag{30}$$

On the other hand, (29) with the choice of $2^{j_0^*}$ implies

$$Z_{21} \lesssim \left(\frac{1}{n}\right)^{\alpha p} \tag{31}$$

for each $\varepsilon \in \mathbb{R}$.

For the case $\varepsilon \leq 0$, $\tilde{p} \leq \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}}$ and $\alpha = \frac{s - \frac{d}{\tilde{p}} + \frac{d}{p}}{2(s - \frac{d}{\tilde{p}}) + d + dI_{\{p>2\}}}$ (see (21c)). Define $p_1 := (1 - 2\alpha)p$. Then $\alpha \leq \frac{s}{2s+d+dI_{\{p>2\}}}$ and $\tilde{p} \leq \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}} \leq (1 - 2\alpha)p = p_1$. Similar to the case $\varepsilon > 0$, $Z_{22} \lesssim \sum_{j=j_0^*+1}^{j_1} 2^{pd(\frac{j}{2} - \frac{j}{p})} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} E \left| \widehat{\beta}_{j,k}^\ell - \beta_{j,k}^\ell \right|^p \left(\frac{|\beta_{j,k}^\ell|}{\kappa t_n / 2} \right)^{p_1}$. Because $\tilde{p} \leq p_1$ and $r \in B_{\tilde{p},q}^s(H)$, $\|\beta_j\|_{p_1}^{p_1} \leq \|\beta_j\|_{\tilde{p}}^{p_1} \lesssim 2^{-j(s - \frac{d}{\tilde{p}} + \frac{d}{2})p_1}$ and

$$Z_{22} \lesssim (\ln n)^{-\frac{p_1}{2}} \left(\frac{1}{n}\right)^{\frac{p-p_1}{2}} \sum_{j=j_0^*+1}^{j_1} 2^{-j(s p_1 - \frac{d p_1}{p} + \frac{d p_1}{2} + \frac{d p_1}{2} I_{\{p>2\}} - \frac{d p}{2} - \frac{d p}{2} I_{\{p>2\}} + d)}$$

By the definitions of p_1 and α , $s p_1 - \frac{d p_1}{p} + \frac{d p_1}{2} + \frac{d p_1}{2} I_{\{p>2\}} - \frac{d p}{2} - \frac{d p}{2} I_{\{p>2\}} + d = 0$ and $Z_{22} \lesssim (\ln n)^{-\frac{p_1}{2}} \left(\frac{1}{n}\right)^{\frac{p-p_1}{2}} (\ln n) \lesssim (\ln n) \left(\frac{1}{n}\right)^{\alpha p}$. This with (30) and (31) shows in both cases,

$$Z_2 = Z_{21} + Z_{22} \lesssim \ln n \left(\frac{1}{n}\right)^{\alpha p} \tag{32}$$

Finally, one estimates Z_3 . Clearly,

$$\begin{aligned} Z_{31} &:= \sum_{j=j_0}^{j_0^*} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \left| \beta_{j,k}^\ell \right|^p I_{\left\{ |\widehat{\beta}_{j,k}^\ell| < \kappa t_n, \quad |\beta_{j,k}^\ell| \leq 2\kappa t_n \right\}} \\ &\leq \sum_{j=j_0}^{j_0^*} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \left| 2\kappa t_n \right|^p \lesssim \sum_{j=j_0}^{j_0^*} 2^{\frac{jdp}{2}} \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{jdp}{2}} I_{\{p > 2\}} \right] \left(\frac{\ln n}{n} \right)^{\frac{p}{2}} \\ &\lesssim 2^{\frac{j_0^* dp}{2}} \left[I_{\{1 \leq p \leq 2\}} + 2^{\frac{j_0^* dp}{2}} I_{\{p > 2\}} \right] \left(\frac{\ln n}{n} \right)^{\frac{p}{2}}. \end{aligned}$$

This with the choice of $2^{j_0^*}$ shows

$$Z_{31} \lesssim (\ln n)^{\frac{p}{2}} \left(\frac{1}{n} \right)^{\alpha p}. \tag{33}$$

On the other hand,

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

The same arguments as (30) shows that for $\varepsilon > 0$,

$$Z_{32} \lesssim (\ln n)^{\frac{p}{2}} \left(\frac{1}{n} \right)^{\alpha p}. \tag{34}$$

To prove (34) true when $\varepsilon \leq 0$, one writes

$$\begin{aligned} Z_{32} &= \left(\sum_{j=j_0^*+1}^{j_1^*} + \sum_{j=j_1^*+1}^{j_1} \right) \left\{ 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} \left| \beta_{j,k}^\ell \right|^p I_{\left\{ |\widehat{\beta}_{j,k}^\ell| < \kappa t_n, \quad |\beta_{j,k}^\ell| \leq 2\kappa t_n \right\}} \right\} \\ &:= Z_{321} + Z_{322}, \end{aligned}$$

where the integer $j_1^* \in [j_0^* + 1, j_1]$ will be determined in the following. Similar to the case $\varepsilon > 0$,

$$Z_{321} \lesssim \left(\frac{\ln n}{n} \right)^{\frac{p-\tilde{p}}{2}} 2^{-j_1^* \varepsilon} \tag{35}$$

holds for $\varepsilon \leq 0$. By $s > \frac{d}{\tilde{p}}$ and $\|\beta_j\|_{\tilde{p}} \lesssim 2^{-j(s-\frac{d}{\tilde{p}}+\frac{d}{2})}$,

$$\begin{aligned} Z_{322} &\leq \sum_{j=J_1^*+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{\tilde{p}}\right)} \sum_{\ell=1}^M \sum_{k \in \Lambda_j} |\beta_{j,k}^\ell|^p \leq \sum_{j=J_1^*+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{\tilde{p}}\right)} \|\beta_j\|_{\tilde{p}}^p \\ &\lesssim \sum_{j=J_1^*+1}^{j_1} 2^{-j(d+sp-pd/\tilde{p})} \lesssim 2^{-J_1^*(d+sp-pd/\tilde{p})}. \end{aligned}$$

To make a balance with (35), one takes $\left(\frac{\ln n}{n}\right)^{\frac{p-\tilde{p}}{2}} 2^{-J_1^*\varepsilon} \sim 2^{-J_1^*(d+sp-pd/\tilde{p})}$, which means

$$2^{J_1^*} \sim \left(\frac{n}{\ln n}\right)^{\frac{\alpha}{s-d/\tilde{p}+d/p}},$$

where $\varepsilon \leq 0$ is equivalent to $\tilde{p} \leq \frac{p(d+dI_{\{p>2\}})}{2s+d+dI_{\{p>2\}}}$ and $\alpha = \frac{s-\frac{d}{\tilde{p}}+\frac{d}{p}}{2(s-\frac{d}{\tilde{p}})+d+dI_{\{p>2\}}}$. In that case, $\left(\frac{n}{\ln n}\right)^{\frac{1-2\alpha}{d+dI_{\{p>2\}}}} \sim 2^{J_0^*} \leq 2^{J_1^*} \leq 2^{j_1} \sim \left(\frac{n}{(\ln n)^3}\right)^{\frac{1}{d}}$. Note that

$$\frac{p-\tilde{p}}{2} + \frac{\alpha\varepsilon}{s-d/\tilde{p}+d/p} = \frac{p-\tilde{p}}{2} + \frac{s\tilde{p} + \frac{\tilde{p}d}{2} + \frac{\tilde{p}d}{2}I_{\{p>2\}} - \frac{pd}{2} - \frac{pd}{2}I_{\{p>2\}}}{2\left(s-\frac{d}{\tilde{p}}\right) + d + dI_{\{p>2\}}} = \alpha p.$$

Then $Z_{321} \lesssim (\ln n)^{\frac{p}{2}} \left(\frac{1}{n}\right)^{\alpha p}$ and $Z_{322} \lesssim (\ln n)^{\frac{p}{2}} \left(\frac{1}{n}\right)^{\alpha p}$. Therefore, $Z_{32} = Z_{321} + Z_{322} \lesssim (\ln n)^{\frac{p}{2}} \left(\frac{1}{n}\right)^{\alpha p}$ for $\varepsilon \leq 0$. Combining this with (33) and (34), one knows $Z_3 \lesssim (\ln n)^{\frac{p}{2}} \left(\frac{1}{n}\right)^{\alpha p}$ in both case. This with (27), (28) and (32) shows

$$Z \lesssim (\ln n)^{p-1} \left[n^{-\alpha p} + (\ln n) \left(\frac{1}{n}\right)^{\alpha p} + (\ln n)^{\frac{p}{2}} \left(\frac{1}{n}\right)^{\alpha p} \right] \lesssim (\ln n)^{\frac{3p}{2}} \left(\frac{1}{n}\right)^{\alpha p},$$

which is the desired conclusion. Although $(\ln n)^{\frac{3p}{2}}$ can be replaced by $(\ln n)^{\max\{p, \frac{3p}{2}-1\}}$ in the last inequality, it can not be in (21b) because one uses Jensen inequality for $\tilde{p} \geq p$. \square

Acknowledgements This work is supported by the National Natural Science Foundation of China [11771030] and the Beijing Natural Science Foundation [1172001]. The authors would like to thank professor Wenqing Xu for his suggestions in writing this manuscript, and the referees for their comments which improve the manuscript.

References

Ahmad IA (1995) On multivariate kernel estimation for samples from weighted distributions. *Stat Probab Lett* 22:121–129

- Chaubey YP, Chesneau C, Shirazi E (2013) Wavelet-based estimation of regression function for dependent biased data under a given random design. *J Nonparametr Stat* 25:53–71
- Chaubey YP, Shirazi E (2015) On MISE of a nonlinear wavelet estimator of the regression function based on biased data under strong mixing. *Commun Stat Theory Methods* 44:885–899
- Chesneau C, Shirazi E (2014) Nonparametric wavelet regression based on biased data. *Commun Stat Theory Methods* 43:2642–2658
- Chesneau C, Fadili J, Maillot B (2015) Adaptive estimation of an additive regression function from weakly dependent data. *J Multivar Anal* 133:77–94
- Davydov YA (1970) The invariance principle for stationary processes. *Theory Probab Appl* 3:487–498
- Delyon B, Judisky A (1996) On minimax wavelet estimators. *Appl Comput Harmonic Anal* 3:215–228
- Donoho DL, Johnstone MI, Kerkyacharian G et al (1996) Density estimation by wavelet thresholding. *Ann Stat* 24:508–539
- Doukhan P (1994) *Mixing: properties and examples*. Springer, New York
- Guo HJ (2016) *Wavelet estimations for a class of regression functions with errors-in-variables*. Dissertation, Beijing University of Technology
- Hall P, Patil P (1996) On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods. *J Roy Stat Soc B* 58:361–377
- Härdle W, Kerkyacharian G, Picard D et al (1998) *Wavelets, approximation and statistical application*. Springer, New York
- Kim TY (1993) A note on moment bounds for strong mixing sequences. *Stat Probab Lett* 16:163–168
- Kou JK, Liu YM (2017) Nonparametric regression estimations over L^p risk based on biased data. *Commun Stat Theory Methods* 46(5):2375–2395
- Kou JK, Liu YM (2016) A extension of Chesneau's theorem. *Stat Probab Lett* 108:23–32
- Kulik R (2008) Nonparametric deconvolution problem for dependent sequences. *Electron J Stat* 2:722–740
- Lepski O (2013) Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *Ann Stat* 41(2):1005–1034
- Liebscher E (1996) Strong convergence of sums of α mixing random variables with applications to density estimation. *Stoch Process Appl* 65:69–80
- Liebscher E (2001) Estimation of the density and regression function under mixing conditions. *Stat Decis* 19:9–26
- Masry E (2000) Wavelet-based estimation of multivariate regression function in Besov spaces. *J Nonparametr Stat* 12:283–308
- Meyer Y (1990) *Wavelets and operators*. Hermann, Paris
- Mokkadem A (1988) Mixing properties of ARMA processes. *Stoch Process Appl* 29:309–315
- Rebelles G (2015a) L^p adaptive estimation of an anisotropic density under independence hypothesis. *Electron J Stat* 9:106–134
- Rebelles G (2015b) Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli* 21(4):1984–2023
- Rio E (1995) The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann Probab* 23:1188–1203
- Shao QM, Yu H (1996) Weak convergence for weighted empirical processes of dependent sequences. *Ann Probab* 24:2098–2127
- Yokoyama R (1980) Moment bounds for stationary mixing sequences. *Probab Theory Relat Fields* 52:45–57