Hindawi
Journal of Function Spaces
Volume 2021, Article ID 1599286, 10 pages
https://doi.org/10.1155/2021/1599286

*Research Article*

# Nonparametric Pointwise Estimation for a Regression Model with Multiplicative Noise

**Jia Chen and Junke Kou**

*School of Mathematics and Computational Science, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China*

Correspondence should be addressed to Junke Kou; kjkou@guet.edu.cn

Received 5 August 2021; Accepted 23 September 2021; Published 11 October 2021

Academic Editor: Andrea Scapellato



In this paper, we consider a general nonparametric regression estimation model with the feature of having multiplicative noise. We propose a linear estimator and nonlinear estimator by wavelet method. The convergence rates of those regression estimators under pointwise error over Besov spaces are proved. It turns out that the obtained convergence rates are consistent with the optimal convergence rate of pointwise nonparametric functional estimation.

## 1. Introduction

The regression estimation plays important roles in practical applications. The classical regression estimation model considers a strictly stationary random process $\{(U_i, V_i), i \in \mathbb{Z}\}$, which defined on $[0, 1]^d \times \mathbb{R}$ and has a common density function $g$. Then, a regression function is defined by

$$r(x) := \mathbb{E}[\rho(V) \mid U = x] = \frac{\int_{\mathbb{R}} \rho(y) g(x, y) dy}{h(x)}, x \in [0, 1]^d, \quad (1)$$

where $\rho(y)$ be a known function and $h(x)$ stands for the density function of random variable $U$. This classical model is aimed at estimating the unknown regression function $r(x)$ by the data $(U_1, V_1), (U_2, V_2), \cdots, (U_n, V_n)$.

However, the above data $(U_1, V_1), (U_2, V_2), \cdots, (U_n, V_n)$ is not easily observed in practical applications. In other words, the observed data contains some noise. In view of this case, this paper considers a regression estimation model with multiplicative noise. Let $\{(X_i, Y_i), i \in \mathbb{Z}\}$ be a strictly stationary random process with a common density function

$$f(x, y) = \frac{\omega(x, y) g(x, y)}{\mu}, (x, y) \in [0, 1]^d \times \mathbb{R}, \quad (2)$$

where $\omega(x, y)$ be a multiplicative noise function, $g(x, y)$ stands for the density function of the unobserved random variable $(U, V)$, and $\mu := \mathbb{E}[\omega(U, V)] < \infty$. Then, we want to estimate the unknown regression function $r(x)$ based on the observed noisy data $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$.

This paper is aimed at considering the pointwise error of wavelet regression estimators under some mild assumptions. Chaubey et al. [1, 2] construct linear and nonlinear wavelet estimators based on dependent data and evaluate its rate of convergence under the mean integrated square error, respectively. The convergence rates of wavelet estimators in independent case are discussed by Chesneau and Shirazi [3]. An optimal convergence rate of regression estimator in independent case is proved by Kou and Liu [4]. Guo and Kou [5] consider wavelet estimations for negatively associated case. The mean integrated square error of regression derivative estimators based on strong mixing data are discussed by Kou et al. [6]. But those above results all only focus on global error. There is a lack of result on pointwise error for the regression model with multiplicative noise ((1) and (2)). Hence, this paper will study the convergence rates of wavelet estimators under pointwise risk.

*1.1. Strong Mixing and Assumptions.* In many practical applications, the observed data is usual not independent.

This paper will consider strong mixing case, which is an important dependent relationship in time series. Now we will introduce the definition of strong mixing in the following.

*Definition 1.* Let $\{X_i, i \in \mathbb{Z}\}$ be a strict stationary sequence of random vectors and

$$\lim_{k\longrightarrow\infty} \alpha(k) = \lim_{k\longrightarrow\infty} \sup \left\{ |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A\in F_{-\infty}^{0}, B \in F_k^{\infty} \right\} = 0, \tag{3}$$

where $F_{-\infty}^{0}$ denotes the $\sigma$ field generated by $\{X_i, i<0\}$ and $F_k^{\infty}$ does by $\{X_i, i \geq k\}$. Then, the sequence $\{X_i, i \in \mathbb{Z}\}$ is strong mixing.

Note that when the random sample is independent, $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$ and $\alpha(k) \equiv 0$. Hence, the independent and identically distributed (*i.i.d.*) data must be strong mixing. Gorodetskii [7] shows that linear processes are strong mixing under certain conditions. Genon-Catahot et al. [8] prove that some continuous time diffusion models and stochastic volatility models are strong mixing, which include some most popular models in the pricing theories of financial assets, such as the Black Scholes pricing theory of options. Moreover, strong mixing also naturally appears in a large class of stationary time series such as ARMA and GARCH models.

Next, we formulate some assumptions for the above regression model. Note that the assumptions C1 and C2 are the standard conditions for nonparametric regression estimation model with multiplicative noise [1, 3]. C3 is a technical condition, which plays an important role in the proof of some lemmas. For the assumptions C4 and C5, the functions $\alpha(k) \equiv 0$ and $h_k(x, y, x', y') \equiv 0$ when the random sample is independent. Hence, C4 and C5 may be considered as other explanation of strong mixing.

*C1.* The density function $h(x)$ satisfies $\inf_{x\in[0,1]^d} h(x) \geq c_1$ with a positive constant $c_1$.

*C2.* For any $(x, y) \in [0,1]^d \times \mathbb{R}$, there exist two positive constant $c_2 < c_3 < \infty$ such that

$$c_2 \leq \omega(x, y) \leq c_3. \tag{4}$$

*C3.* The function $\rho(y)$ satisfies $\rho(y) \in L(\mathbb{R}) \cap L^{\infty}(\mathbb{R})$.

*C4.* The strong mixing coefficient of $\{(X_i, Y_i), i=1,2,\cdots,n\}$ satisfies $\alpha(k) \leq \gamma e^{-c_4 k}$ with $\gamma > 0$ and $c_4 > 0$.

*C5.* Let $f_{(X_1,Y_1,X_{k+1},Y_{k+1})}$ be the density function of $(X_1, Y_1, X_{k+1}, Y_{k+1})(k \geq 1)$ and $f_{(X_1,Y_1)}$ be the density function of $(X_1, Y_1)$. Then, for $(x, y, x', y') \in [0,1]^d \times \mathbb{R} \times [0,1]^d \times \mathbb{R}$,

$$\sup_{k\geq 1} \sup_{(x,y,x',y')\in[0,1]^d\times\mathbb{R}\times[0,1]^d\times\mathbb{R}} \left| h_k\left(x, y, x', y'\right) \right| \leq c_4, \tag{5}$$

with $c_4 > 0$ and $h_k := f_{(X_1,Y_1,X_{k+1},Y_{k+1})} - f_{(X_1,Y_1)} f_{(X_{k+1},Y_{k+1})}$.

*1.2. Wavelets and Besov Space.* In this paper, we construct regression estimators by wavelet method. Now let us briefly present the wavelet basis of $L^2(\mathbb{R}^d)$ by tensor product method. For a scaling function $\varphi(x)$, there is $M = 2^d - 1$ wavelet function $\psi^l (l = 1, 2, \cdots, M)$ such that for each $f(x) \in L^2(\mathbb{R}^d)$,

$$f(x) = \sum_{k\in\mathbb{Z}^d} \alpha_{j_0,k}\varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{l=1}^{M} \sum_{k\in\mathbb{Z}^d} \beta_{j,k}^l \psi_{j,k}^l(x), \tag{6}$$

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

$$\varphi_{j_0,k}(x) = 2^{j_0 d/2}\varphi\left(2^{j_0 d}x - k\right), \psi_{j,k}^l(x) = 2^{jd/2}\psi^l\left(2^{jd}x - k\right). \tag{7}$$

If a scaling function $\varphi(x)$ satisfies Condition $\theta$, i.e., $\sum_{k\in\mathbb{Z}^d} |\varphi(x-k)| \in L^{\infty}(\mathbb{R}^d)$, then the functions $\varphi(x) \in L(\mathbb{R}^d) \cap L^{\infty}(\mathbb{R}^d)$ and $\sum_{k\in\mathbb{Z}^d} \varphi(x-k)\varphi(\bar{y}k)$ converge absolutely almost everywhere. In this paper, we use the Daubechies wavelet, which has compact supports and satisfies Condition $\theta$. Then, it can be shown that for $f(x) \in L^p(\mathbb{R}^d)(1 \leq p<\infty)$,

$$\mathbf{P}_j f(x) = \sum_{k\in\mathbb{Z}^d} \alpha_{j,k}\varphi_{j,k}(x), \tag{8}$$

holds almost everywhere on $\mathbb{R}^d$.

Besov spaces are important in theory and applications, which contain Hölder and $L^2$ lebesgue spaces as special examples [9–11]. The next lemma provides equivalent definitions of Besov space [12].

**Lemma 2.** *Let $\varphi(x)$ be a scaling function, $\psi^l(x)(l = 1, 2, \cdots, M)$ be the corresponding wavelet function, and $f(x) \in L^p(\mathbb{R}^d)(1 \leq p<\infty)$. For $1 \leq p, q \leq +\infty$ and $s > 0$, then the following assertions are equivalent:*

$$\begin{gathered} f \in B_{p,q}^s\left(\mathbb{R}^d\right), \\ \left\{2^{js}\left\|\mathbf{P}_j f - f\right\|_p\right\} \in l_q, \\ \left\{2^{j(s-d/p+d/2)}\left\|\beta_{j\cdot}\right\|_p\right\} \in l_q. \end{gathered} \tag{9}$$

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

$$\|f\|_{B_{p,q}^s} = \left\|\alpha_{j_0}\right\|_p + \left\|\left(2^{j(s-d/p+d/2)}\left\|\beta_{j\cdot}\right\|_p\right)_{j\geq j_0}\right\|_q, \tag{10}$$

*with $\|\alpha_{j_0}\|_p^p = \sum_{k\in\mathbb{Z}^d} |\alpha_{j_0,k}|^p$ and $\|\beta_{j\cdot}\|_p^p = \sum_{l=1}^{M}\sum_{k\in\mathbb{Z}^d} |\beta_{j,k}^l|^p$.*

*In this paper, we assume that the regression function $r(x)$ belongs to Besov ball with $H > 0$, i.e.,*

$$r(x) \in B^s_{p,q}(H) := \left\{ f \in B^s_{p,q}\left(\mathbb{R}^d\right), \|f\|_{B^s_{p,q}} \le H \right\}. \quad (11)$$

## 2. Wavelet Estimators and Theorem

Let us define linear wavelet estimator as follows:

$$\widehat{r}_n^{\text{lin}}(x) := \sum_{k\in\Omega} \widehat{\alpha}_{j_0,k}\varphi_{j_0,k}(x), \quad (12)$$

where

$$\widehat{\mu}_n = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega(X_i, Y_i)}\right]^{-1},$$
$$\widehat{\alpha}_{j_0,k} = \frac{\widehat{\mu}_n}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i, Y_i)h(X_i)}. \quad (13)$$

Using hard thresholding method, our nonlinear wavelet estimator is defined by

$$\widehat{r}_n^{\text{non}}(x) := \sum_{k\in\Omega} \widehat{\alpha}_{j_0,k}\varphi_{j_0,k}(x) + \sum_{j=j_0}^{j_1}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\widehat{\beta}^l_{j,k}\mathbf{I}_{\left\{\left|\widehat{\beta}^l_{j,k}\right|\ge\kappa t_n\right\}}\psi^l_{j,k}(x), \quad (14)$$

with $t_n = \sqrt{\ln n/n}$, the indicator function $\mathbf{I}(x)$ and

$$\widehat{\beta}^l_{j,k} = \frac{\widehat{\mu}_n}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)\psi^l_{j,k}(X_i)}{\omega(X_i, Y_i)h(X_i)}. \quad (15)$$

The positive parametrics $j_0$, $j_1$, and $\kappa$ will be given in later discussion. In addition, $\Omega := \{k \in \mathbb{Z}, \text{supp } r \cap \text{supp } \varphi_{j_0,k} \ne \varnothing\}$ and $\Omega_j := \{k \in \mathbb{Z}, \text{supp } r \cap \text{supp } \psi^l_{j,k} \ne \varnothing\}$. Because of the compactly supported of regression functions $r(x)$, $\varphi_{j_0,k}(x)$ and $\psi^l_{j,k}(x)$, the cardinality of $\Omega$ and $\Omega_j$ satisfies that $\Omega \sim 2^{j_0 d}$ and $\Omega_j \sim 2^{jd}$, respectively. Here and after, $AB$ denotes $A \le cB$ for some positive constant $c$, and $A \sim B$ stands for both $AB$ and $BA$.

Now we state our main theorem in the following, which investigates the performance of linear and nonlinear wavelet estimators in terms of rates of convergence under pointwise risk over Besov space.

**Theorem 3.** *Consider the regression problems (1) and (2) under the assumptions C1-C5 and $r(x) \in B^s_{p,q}(H)(p, q \in [1,+\infty), s > d/p)$. Then, the linear wavelet estimator $\widehat{r}_n^{lin}(x)$ with $2^{j_0} \sim n^{1/2(s-d/p)+d}$ satisfies*

$$\mathbb{E}\left[\left|r\wedge_n^{lin}(x) - r(x)\right|^2\right] n^{-2(s-d/p)/2(s-d/p)+d}. \quad (16)$$

*The nonlinear one with $2^{j_0} \sim n^{1/2m+d}(m > s)$ and $2^{j_1} \sim \left(n/(\ln n)^3\right)^{1/d}$ does*

$$\mathbb{E}\left[|r\wedge_n^{non}(x) - r(x)|^2\right](\ln n)^3 n^{-2(s-d/p)/2(s-d/p)+d}. \quad (17)$$

*Remark 4.* Note that the linear and nonlinear wavelet estimators all can attain the optimal convergence rate for pointwise risk [13].

*Remark 5.* Compared with the linear wavelet estimator $\widehat{r}_n^{\text{lin}}(x)$, the convergence rate of the nonlinear estimator $\widehat{r}_n^{\text{non}}(x)$ keeps the same as the linear one up to a $\ln n$ factor. Moreover, the nonlinear wavelet estimator is adaptive, which means that it only depends on the observed data.

## 3. Auxiliary Results

This section will provide some results for the proof of Theorem 3.

**Lemma 6.** *For the problem defined in (1) and (2), we have*

$$\mathbb{E}\left[\frac{1}{\widehat{\mu}_n}\right] = \frac{1}{\mu},$$
$$\mathbb{E}\left[\frac{\mu\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i, Y_i)h(X_i)}\right] = \alpha_{j_0,k}, \quad (18)$$
$$\mathbb{E}\left[\frac{\mu\rho(Y_i)\psi^l_{j,k}(X_i)}{\omega(X_i, Y_i)h(X_i)}\right] = \beta^l_{j,k}.$$

**Lemma 7.** *Let $\{(X_i, Y_i), i = 1, 2, \cdots, n\}$ be a stationary strong mixing process. If C1-C5 hold and $2^{jd} \le n$,*

$$Var\left[\sum_{i=1}^{n}\frac{1}{\omega(X_i, Y_i)}\right] \lesssim n, \quad (19)$$

$$Var\left[\sum_{i=1}^{n}\frac{\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i, Y_i)h(X_i)}\right] \lesssim n, \quad (20)$$

$$Var\left[\sum_{i=1}^{n}\frac{\rho(Y_i)\psi^l_{j,k}(X_i)}{\omega(X_i, Y_i)h(X_i)}\right] \lesssim n. \quad (21)$$

*Proof.* According to the properties of variance, we know

$$\begin{aligned}\operatorname{Var}\left[\sum_{i=1}^{n}\frac{1}{\omega(X_i,Y_i)}\right] &= \sum_{i=1}^{n}\operatorname{Var}\left[\frac{1}{\omega(X_i,Y_i)}\right] + 2\sum_{1\le i<v\le n}\operatorname{Cov}\left(\frac{1}{\omega(X_i,Y_i)},\frac{1}{\omega(X_v,Y_v)}\right)\\ &\lesssim n\operatorname{Var}\left[\frac{1}{\omega(X_i,Y_i)}\right] + \left|\sum_{v=2}^{n}\sum_{i=1}^{v-1}\operatorname{Cov}\left(\frac{1}{\omega(X_i,Y_i)},\frac{1}{\omega(X_v,Y_v)}\right)\right|.\end{aligned} \tag{22}$$

Now we estimate the first term. It follows from the property of variance and C2 that

$$\operatorname{Var}\left(\frac{1}{\omega(X_i,Y_i)}\right) \le \mathbb{E}\left(\frac{1}{\omega(X_i,Y_i)}\right)^2 = \int_{[0,1]^d\times\mathbb{R}}\left(\frac{1}{\omega(x,y)}\right)^2 f(x,y)dxdy \lesssim 1. \tag{23}$$

Using the property of strictly stationary,

$$\begin{aligned}&\left|\sum_{v=2}^{n}\sum_{i=1}^{v-1}\operatorname{Cov}\left(\frac{1}{\omega(X_i,Y_i)},\frac{1}{\omega(X_v,Y_v)}\right)\right|\\ &\quad\lesssim \sum_{i=1}^{n-1}(n-i)\left|\operatorname{Cov}\left(\frac{1}{\omega(X_1,Y_1)},\frac{1}{\omega(X_{i+1},Y_{i+1})}\right)\right|\\ &\quad\lesssim n\sum_{i=1}^{n-1}\left|\operatorname{Cov}\left(\frac{1}{\omega(X_1,Y_1)},\frac{1}{\omega(X_{i+1},Y_{i+1})}\right)\right|.\end{aligned} \tag{24}$$

Then, we need the following Davydov inequality [14]. Let $\{X_i\}_{i\in Z}$ be strong mixing with mixing coefficient $\alpha(k)$ and $f$ and $g$ be two measurable functions. If $\mathbb{E}|f(X_1)|^p$ and $\mathbb{E}|g(X_1)|^q$ exist for $p, q > 0$ and $1/p + 1/q < 1$,

$$|\operatorname{Cov}(f(X_1), g(X_{k+1}))| \lesssim \alpha(k)^{1-1/p-1/q}\left[\mathbb{E}\,|\,f(X_1)|^p\right]^{1/p}\left[\mathbb{E}\,|\,g(X_1)|^q\right]^{1/q}. \tag{25}$$

Using the Davydov inequality with $p = q = 4$,

$$\left|\operatorname{Cov}\left(\frac{1}{\omega(X_1,Y_1)},\frac{1}{\omega(X_{i+1},Y_{i+1})}\right)\right| \lesssim \alpha(i)^{1/2}\left(\mathbb{E}\left|\frac{1}{\omega(X_1,Y_1)}\right|^4\right)^{1/2}. \tag{26}$$

This with C2 shows $|\operatorname{Cov}(1/\omega(X_1,Y_1), 1/\omega(X_{i+1},Y_{i+1}))| \lesssim \alpha(i)^{1/2}$. Furthermore, it follows from C4 that

$$\begin{aligned}\sum_{i=1}^{n-1}\left|\operatorname{Cov}\left(\frac{1}{\omega(X_1,Y_1)},\frac{1}{\omega(X_{i+1},Y_{i+1})}\right)\right| &\lesssim \sum_{i=1}^{n-1}\alpha(i)^{1/2}\\ &\le \sum_{i=1}^{n-1}\gamma^{1/2}e^{-(c_4/2)i} \lesssim \sum_{i=1}^{\infty}e^{-(c_4/2)i} \lesssim 1.\end{aligned} \tag{27}$$

Combining this result with and (22)–(24) and ((27)), one gets that

$$\operatorname{Var}\left[\sum_{i=1}^{n}\frac{1}{\omega(X_i,Y_i)}\right] \lesssim n. \tag{28}$$

For the estimation of (20), we know that $\operatorname{Var}[\sum_{i=1}^{n}\rho(Y_i)\varphi_{j_0,k}(X_i)/\omega(X_i,Y_i)h(X_i)] \lesssim n\operatorname{Var}[\rho(Y_i)\varphi_{j_0,k}(X_i)/\omega(X_i,Y_i)h(X_i)] + |\sum_{v=2}^{n}\sum_{i=1}^{v-1}\operatorname{Cov}(\rho(Y_i)\varphi_{j_0,k}(X_i)/\omega(X_i,Y_i)h(X_i), \rho(Y_v)\varphi_{j_0,k}(X_v)/\omega(X_v,Y_v)h(X_v))|$.Then, it is easy to see from the property of variance and the assumptions C1-C3 that

$$\begin{aligned}n\operatorname{Var}\left[\frac{\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i,Y_i)h(X_i)}\right] &\le n\mathbb{E}\left(\frac{\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i,Y_i)h(X_i)}\right)^2\\ &= n\int_{[0,1]^d\times\mathbb{R}}\left(\frac{\rho(y)\varphi_{j_0,k}(x)}{\omega(x,y)h(x)}\right)^2 f(x,y)dxdy\\ &= n\int_{[0,1]^d\times\mathbb{R}}\frac{(\rho(y))^2\left(\varphi_{j_0,k}(x)\right)^2 g(x,y)}{\mu\omega(x,y)(h(x))^2}dxdy\\ &\lesssim n\int_{[0,1]^d\times\mathbb{R}}\frac{\left(\varphi_{j_0,k}(x)\right)^2 g(x,y)}{h(x)}dxdy\\ &= n\int_{[0,1]^d}\left(\varphi_{j_0,k}(x)\right)^2 dx \lesssim n.\end{aligned} \tag{29}$$

Now it remains to prove

$$\left|\sum_{v=2}^{n}\sum_{i=1}^{v-1}\operatorname{Cov}\left(\frac{\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i,Y_i)h(X_i)},\frac{\rho(Y_v)\varphi_{j_0,k}(X_v)}{\omega(X_v,Y_v)h(X_v)}\right)\right| \lesssim n. \tag{30}$$

Similar to the arguments of (24), one gets that $|\sum_{v=2}^{n}\sum_{i=1}^{v-1}\operatorname{Cov}(\rho(Y_i)\varphi_{j_0,k}(X_i)/\omega(X_i,Y_i)h(X_i), \rho(Y_v)\varphi_{j_0,k}(X_v)/\omega(X_v,Y_v)h(X_v))| n\sum_{i=1}^{n-1}|\operatorname{Cov}(\rho(Y_1)\varphi_{j_0,k}(X_1)/\omega(X_1,Y_1)h(X_1), \rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})/\omega(X_{i+1},Y_{i+1})h(X_{i+1}))|$. Because of $2^{jd} \le n$, we can obtain that

$$\left|\sum_{v=2}^{n}\sum_{i=1}^{v-1}\operatorname{Cov}\left(\frac{\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i,Y_i)h(X_i)},\frac{\rho(Y_v)\varphi_{j_0,k}(X_v)}{\omega(X_v,Y_v)h(X_v)}\right)\right| \lesssim n(T_1+T_2), \tag{31}$$

$$T_1 := \sum_{i=1}^{2^{j_0d}-1}\left|\operatorname{Cov}\left(\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)},\frac{\rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})}{\omega(X_{i+1},Y_{i+1})h(X_{i+1})}\right)\right|, \tag{32}$$

$$T_2 := \sum_{i=2^{j_0d}}^{n-1}\left|\operatorname{Cov}\left(\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)},\frac{\rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})}{\omega(X_{i+1},Y_{i+1})h(X_{i+1})}\right)\right|. \tag{33}$$

First, estimate $T_1$. By the property of covariance, $|\operatorname{Cov}(\rho(Y_1)\varphi_{j_0,k}(X_1)/\omega(X_1,Y_1)h(X_1), \rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})/\omega(X_{i+1},$

$Y_{i+1})h(X_{i+1}))| = |\mathbb{E}(\rho(Y_1)\varphi_{j_0,k}(X_1)/\omega(X_1, Y_1)h(X_1)\rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})/\omega(X_{i+1}, Y_{i+1})h(X_{i+1})) - \mathbb{E}(\rho(Y_1)\varphi_{j_0,k}(X_1)/\omega(X_1, Y_1)h(X_1))\mathbb{E}(\rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})/\omega(X_{i+1}, Y_{i+1})h(X_{i+1}))|$. Furthermore, using the assumptions C1-C3 and C5, one can get that

$$\begin{aligned}
&\left|\mathrm{Cov}\left(\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)}, \frac{\rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})}{\omega(X_{i+1},Y_{i+1})h(X_{i+1})}\right)\right| \\
&\quad \le \le \int_{[0,1]^d\times\mathbb{R}\times[0,1]^d\times\mathbb{R}} \left|\frac{\rho(y)\varphi_{j_0,k}(x)}{\omega(x,y)h(x)}\right| \left|\frac{\rho(y')\varphi_{j_0,k}(x')}{\omega(x',y')h(x')}\right| \left|h_i(x,y,x',y')\right| dxdydx'dy' \\
&\quad \lesssim \left(\int_{\mathbb{R}} |\rho(y)|\,dy\right)^2 \left(\int_{[0,1]^d} |\varphi_{j_0,k}(x)|\,dx\right)^2 \\
&\quad \lesssim \left(\int_{[0,1]^d} |\varphi_{j_0,k}(x)|\,dx\right)^2 \lesssim 2^{-j_0 d}.
\end{aligned} \tag{34}$$

Then, one have

$$\begin{aligned}
T_1 &= \sum_{i=1}^{2^{j_0 d}-1} \left|\mathrm{Cov}\left(\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)}, \frac{\rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})}{\omega(X_{i+1},Y_{i+1})h(X_{i+1})}\right)\right| \\
&\lesssim \sum_{i=1}^{2^{j_0 d}-1} 2^{-j_0 d} \lesssim 1.
\end{aligned} \tag{35}$$

Now we estimate $T_2$. Using the above Davydov inequality with $p = q = 4$,

$$\begin{aligned}
&\left|\mathrm{Cov}\left(\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)}, \frac{\rho(Y_{i+1})\varphi_{j_0,k}(X_{i+1})}{\omega(X_{i+1},Y_{i+1})h(X_{i+1})}\right)\right| \\
&\quad \lesssim \alpha(i)^{1/2}\left(\mathbb{E}\left|\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)}\right|^4\right)^{1/2} \\
&\quad \lesssim \le \alpha(i)^{1/2} \sup\left|\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)}\right| \left(\mathbb{E}\left|\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)}\right|^2\right)^{1/2}.
\end{aligned} \tag{36}$$

It follows from C1-C3 that

$$\sup\left|\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_1,Y_1)h(X_1)}\right| \lesssim 2^{j_0 d}. \tag{37}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}\left|\frac{\rho(Y_1)\varphi_{j_0,k}(X_1)}{\omega(X_i,Y_i)h(X_1)}\right|^2 &= \int_{[0,1]^d\times\mathbb{R}} \left(\frac{\rho(y)\varphi_{j_0,k}(x)}{\omega(x,y)h(x)}\right)^2 \frac{\omega(x,y)g(x,y)}{\mu}\,dxdy \\
&\lesssim \int_{[0,1]^d} \left(\varphi_{j_0,k}(x)\right)^2 dx \lesssim 1.
\end{aligned} \tag{38}$$

Hence,

$$T_2 \lesssim \sum_{i=2^{j_0 d}}^{n-1} \alpha(i)^{\frac{1}{2}} 2^{j_0 d} \le \sum_{i=2^{j_0 d}}^{n-1} \gamma\frac{1}{2} e^{-(c_4/2)i} 2^{j_0 d} \lesssim \sum_{i=1}^{\infty} e^{-(c_4/2)i} i^{\frac{1}{2}} \lesssim 1. \tag{39}$$

Now the above results (31), (35), and (39) imply the desired inequality

$$\left|\sum_{v=2}^{n}\sum_{i=1}^{v-1} \mathrm{Cov}\left(\frac{\rho(Y_i)\varphi_{j_0,k}(X_i)}{\omega(X_i,Y_i)h(X_i)}, \frac{\rho(Y_v)\varphi_{j_0,k}(X_v)}{\omega(X_v,Y_v)h(X_v)}\right)\right| \lesssim n. \tag{40}$$

The prove of the last inequality (21) is similar to (20). □

**Lemma 8.** *Let assumptions C1-C5 hold and $2^{jd} \le n$. Then,*

$$\mathbb{E}\left|\alpha\wedge_{j_0,k} - \alpha_{j_0,k}\right|^2 \lesssim n^{-1}, \mathbb{E}\left|\beta\wedge_{j,k}^{l} - \beta_{j,k}^{l}\right|^2 \lesssim n^{-1}. \tag{41}$$

*Proof.* By the definition of $\widehat{\alpha}_{j_0,k}$,

$$\begin{aligned}
\left|\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}\right| &= \left|\frac{\widehat{\mu}_n}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j_0,k}(X_i) - \alpha_{j_0,k}\right| \\
&\le \left|\frac{\widehat{\mu}_n}{\mu}\left(\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j_0,k}(X_i) - \alpha_{j_0,k}\right)\right| \\
&\quad + \left|\alpha_{j_0,k}\widehat{\mu}_n\left(\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n}\right)\right|.
\end{aligned} \tag{42}$$

Note that C3 and the definition of regression function $r(x)$ imply the boundedness of $r(x)$. Furthermore, $|\alpha_{j_0,k}| = |\int_{[0,1]^d} r(x)\varphi_{j_0,k}(x)dx| \lesssim 1$ by Hölder inequality and the orthonormality of $\varphi_{j_0,k}(x)$. On the other hand, $|\widehat{\mu}_n/\mu| \lesssim 1$ and $|\mu| \lesssim 1$ because of C2. Then, one can obtain that

$$\begin{aligned}
\mathbb{E}\left|\alpha\wedge_{j_0,k} - \alpha_{j_0,k}\right|^2 &\lesssim \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j_0,k}(X_i) - \alpha_{j_0,k}\right|^2 \\
&\quad + \mathbb{E}\left|\frac{1}{\mu} - \frac{1}{\mu\wedge_n}\right|^2.
\end{aligned} \tag{43}$$

It is easy to see from Lemmas 6 and 7 that

$$\mathbb{E}\left|\alpha\wedge_{j_0,k}-\alpha_{j_0,k}\right|^2 \lesssim \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j_0,k}(X_i)\right)+\operatorname{Var}\left(\frac{1}{\widehat{\mu}_n}\right) \lesssim \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j_0,k}(X_i)\right) + \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n}\frac{1}{\omega(X_i,Y_i)}\right)\lesssim n^{-1}. \quad (44)$$

For the second inequality, according to the definition of $\widehat{\beta}^l_{j,k}$,

$$\left|\widehat{\beta}^l_{j,k}-\beta^l_{j,k}\right| \le \left|\frac{\widehat{\mu}_n}{\mu}\left(\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi^l_{j,k}(X_i)-\beta^l_{j,k}\right)\right| + \left|\beta^l_{j,k}\widehat{\mu}_n\left(\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right)\right|. \quad (45)$$

Then, it follows from the similar arguments of (43) that

$$\mathbb{E}\left|\beta\wedge^l_{j,k}-\beta^l_{j,k}\right|^2 \lesssim \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi^l_{j,k}(X_i)-\beta^l_{j,k}\right|^2+\mathbb{E}\left|\frac{1}{\mu}-\frac{1}{\mu\wedge_n}\right|^2. \quad (46)$$

Using Lemmas 6 and 7, one can get

$$\mathbb{E}\left|\beta\wedge^l_{j,k}-\beta^l_{j,k}\right|^2 \lesssim \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi^l_{j,k}(X_i)\right)+\operatorname{Var}\left(\frac{1}{\widehat{\mu}_n}\right)\lesssim n^{-1}. \quad (47)$$

□

**Lemma 9.** *Let $\widehat{\beta}^l_{j,k}$ be defined by (15). If C1-C5 hold, $t_n=\sqrt{\ln n/n}$ and $2^{jd}\le n/(\ln n)^3$; then for each $w>0$, there exists a constant $\kappa>1$ such that*

$$\mathbb{P}\left(\left|\widehat{\beta}^l_{j,k}-\beta^l_{j,k}\right|\ge\kappa t_n\right)\lesssim 2^{-wj}. \quad (48)$$

*Proof.* According to the arguments of (43) and the definition of $\widehat{\beta}^l_{j,k}$,

$$\left|\widehat{\beta}^l_{j,k}-\beta^l_{j,k}\right| \le \left|\frac{\widehat{\mu}_n}{\mu}\left(\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi^l_{j,k}(X_i)-\beta^l_{j,k}\right)\right| + \left|\beta^l_{j,k}\widehat{\mu}_n\left(\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right)\right| \lesssim \left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi^l_{j,k}(X_i)-\beta^l_{j,k}\right)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\omega(X_i,Y_i)}-\frac{1}{\mu}\right)\right|. \quad (49)$$

Then, we only need to prove the following two inequalities, respectively.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi^l_{j,k}(X_i)-\beta^l_{j,k}\right)\right|\ge\frac{\kappa}{2}t_n\right)\lesssim 2^{-wj}, \quad (50)$$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\omega(X_i,Y_i)}-\frac{1}{\mu}\right)\right|\ge\frac{\kappa}{2}t_n\right)\lesssim 2^{-wj}. \quad (51)$$

(i) *Proof of* (50). Define $\eta_i := (\mu\rho(Y_i)/\omega(X_i,Y_i)h(X_i))\psi^l_{j,k}(X_i)-\beta^l_{j,k}$. Then, $\mathbb{E}(\eta_i)=0$ thanks to Lemma 6. Moreover, $\eta_1,\eta_2,\cdots,\eta_n$ are strong mixing with the mixing coefficients $\alpha(k)\le\gamma e^{-c_4k}$. Using the assumptions C1-C3, one gets that $|\eta_i|\le|(\mu\rho(Y_i)/\omega(X_i,Y_i)h(X_i))\psi^l_{j,k}(X_i)|+|\beta^l_{j,k}|\lesssim 2^{jd/2}$. On the other hand, it follows from Lemma 7 that

$$\max_{1\le j\le 2m}\operatorname{Var}\left(\sum_{i=1}^{j}\eta_i\right)=\max_{1\le j\le 2m}\operatorname{Var}\left(\sum_{i=1}^{j}\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\psi^l_{j,k}(X_i)\right)\lesssim m. \quad (52)$$

Because strong mixing is much more complicated than independent, we need to use a new technique. In this paper, we use the following Bernstein type inequality [15]. Let $\{X_i\}_{i\in Z}$ be a strong mixing process with mixing coefficient $\alpha(k)$, $\mathbb{E}(X_i)=0$, $|X_i|\le M<\infty$, and $D_m:=\max_{1\le j\le 2m}\operatorname{Var}(\sum_{i=1}^{j}X_i)$. Then, for $\varepsilon>0$ and $n,m\in N$ with $0<m\le n/2$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}X_i\right|\ge\varepsilon\right)\exp\left(-\frac{\varepsilon^2}{16}\left(nm^{-1}D_m+\frac{1}{3}\varepsilon Mm\right)^{-1}\right)+\frac{M}{\varepsilon}n\alpha(m). \quad (53)$$

According to the above Bernstein type inequality with $m = u \ln n$ and $M = 2^{jd/2}$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i\right| \geq \frac{\kappa}{2}t_n\right) &= \mathbb{P}\left(\left|\sum_{i=1}^{n}\eta_i\right| \geq \frac{\kappa}{2}nt_n\right) \\ &\lesssim \exp\left(-\frac{(\kappa/2nt_n)^2}{16}\left(nm^{-1}m + \frac{1}{3}\frac{\kappa}{2}nt_n 2^{jd/2}m\right)^{-1}\right) \\ &\quad + \frac{2^{jd/2}}{(\kappa/2)nt_n}n\alpha(m) \\ &\lesssim \exp\left(-\frac{(\kappa nt_n)^2}{64}\left(n + \frac{\kappa u}{6}nt_n 2^{jd/2}\ln n\right)^{-1}\right) \\ &\quad + \frac{2^{jd/2}}{t_n}e^{-cu\ln n}.\end{aligned} \tag{54}$$

Then, it follows from $t_n = \sqrt{\ln n/n}$ and $2^{jd} \leq n/(\ln n)^3 \leq n$ that

$$\exp\left(-\frac{(\kappa nt_n)^2}{64}\left(n + \frac{\kappa u}{6}nt_n 2^{jd/2}\ln n\right)^{-1}\right) \lesssim \exp\left(-\frac{\kappa^2\ln n}{64}\left(1 + \frac{\kappa u}{6}\right)^{-1}\right). \tag{55}$$

Obviously, there exists sufficiently large $\kappa > 1$ such that

$$\exp\left(-\frac{\kappa^2\ln n}{64}\left(1 + \frac{\kappa u}{6}\right)^{-1}\right) \lesssim 2^{-wj}. \tag{56}$$

On the other hand, note that $(2^{jd/2}/t_n)e^{-cu\ln n} \lesssim n^{1-cu} \lesssim 2^{jd(1-cu)}$. Choosing $u$ such that $d(1-cu) < -w$, then

$$\frac{2^{jd/2}}{t_n}e^{-cu\ln n} \lesssim 2^{-wj}. \tag{57}$$

Together with the above results, one can obtain the desired result (50)

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i\right| \geq \frac{\kappa}{2}t_n\right) \lesssim 2^{-wj}. \tag{58}$$

(ii) *Proof of* (51). Define $\xi_i := 1/\omega(X_i, Y_i) - 1/\mu$. Then, $\mathbb{E}(\xi_i) = 0$ thanks to Lemma 6. Moreover, it follows from the assumption C4 that $\xi_1, \xi_2, \cdots, \xi_n$ are strong mixing with the mixing coefficients $\alpha(k) \leq \gamma e^{-c_4 k}$. Note that $|\xi_i| \leq |1/\omega(X_i, Y_i)| + |1/\mu| \lesssim 1$ with C2. According to Lemma 7, $\max_{1\leq j\leq 2m}\mathrm{Var}(\sum_{i=1}^{j}\xi_i) = \max_{1\leq j\leq 2m}\mathrm{Var}(\sum_{i=1}^{j}1/\omega(X_i, Y_i)) \lesssim m$. Then, similar to the proof of (50), one can get the desired result (51).

□

## 4. Proof of Theorem

This last section is devoted to proving Theorem 3.

*Proof of* (16). It is easy to see that

$$\begin{aligned}\mathbb{E}\left|r\wedge_n^{\text{lin}}(x) - r(x)\right|^2 &\leq \mathbb{E}\left(\left|r\wedge_n^{\text{lin}}(x) - \mathbf{P}_{j_0}r(x)\right| + \left|\mathbf{P}_{j_0}r(x) - r(x)\right|\right)^2 \\ &\lesssim \mathbb{E}\left|r\wedge_n^{\text{lin}}(x) - \mathbf{P}_{j_0}r(x)\right|^2 + \mathbb{E}\left|\mathbf{P}_{j_0}r(x) - r(x)\right|^2.\end{aligned} \tag{59}$$

(i) Upper bound of $\mathbb{E}|r\wedge_n^{\text{lin}}(x) - \mathbf{P}_{j_0}r(x)|^2$. Note that

$$\mathbb{E}\left|r\wedge_n^{\text{lin}}(x) - \mathbf{P}_{j_0}r(x)\right|^2 \leq \mathbb{E}\left(\sum_{k\in\Omega}\left|\alpha\wedge_{j_0,k} - \alpha_{j_0,k}\right|\left|\varphi_{j_0,k}(x)\right|\right)^2. \tag{60}$$

Then, it follows from Cauchy-Schwarz inequality and Condition $\theta$ that

$$\begin{aligned}\mathbb{E}\left|r\wedge_n^{\text{lin}}(x) - \mathbf{P}_{j_0}r(x)\right|^2 &\leq \mathbb{E}\left(\sum_{k\in\Omega}\left|\alpha\wedge_{j_0,k} - \alpha_{j_0,k}\right|^2\left|\varphi_{j_0,k}(x)\right|\right) \\ &\quad\cdot\left(\sum_{k\in\Omega}\left|\varphi_{j_0,k}(x)\right|\right) \\ &\lesssim 2^{j_0 d/2}\sum_{k\in\Omega}\left(\left|\varphi_{j_0,k}(x)\right|\mathbb{E}\left|\alpha\wedge_{j_0,k} - \alpha_{j_0,k}\right|^2\right).\end{aligned} \tag{61}$$

Using Lemma 8 and $2^{j_0} \sim n^{1/2(s-d/p)+d}$,

$$\begin{aligned}\mathbb{E}\left|r\wedge_n^{\text{lin}}(x) - \mathbf{P}_{j_0}r(x)\right|^2 &\lesssim 2^{j_0 d/2}n^{-1}\sum_{k\in\Omega}\left|\varphi_{j_0,k}(x)\right| \\ &\lesssim 2^{j_0 d}n^{-1} \lesssim n^{-2(s-d/p)/2(s-d/p)+d}.\end{aligned} \tag{62}$$

(ii) Upper bound of $\mathbb{E}|\mathbf{P}_{j_0}r(x) - r(x)|^2$. It is easy to see from Hölder inequality that

$$\begin{aligned}\mathbb{E}\left|\mathbf{P}_{j_0}r(x) - r(x)\right|^2 &= \left(\sum_{j=j_0}^{\infty}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta_{j,k}^l\right|\left|\psi_{j,k}^l(x)\right|\right)^2 \\ &\leq \left\{\sum_{j=j_0}^{\infty}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta_{j,k}^l\right|^p\left|\psi_{j,k}^l(x)\right|\right)^{1/p} \right. \\ &\quad\left.\cdot\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^l(x)\right|\right)^{1/p'}\right\}^2.\end{aligned} \tag{63}$$

Furthermore, Condition $\theta$ and the boundedness of $\psi^l_{j,k}(x)$ imply that

$$\mathbb{E}\left|\mathbf{P}_{j_0}r(x)-r(x)\right|^2\lesssim\left\{\sum_{j=j_0}^{\infty}2^{jd/2}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta^l_{j,k}\right|^p\right)^{1/p}\right\}^2=\left\{\sum_{j=j_0}^{\infty}2^{jd/2}\left\|\beta_{j\cdot}\right\|_p\right\}^2. \quad (64)$$

Now using $r(x)\in B^s_{p,q}(H)$, Lemma 2, and $2^{j_0}\sim n^{1/2(s-d/p)+d}$, one can get that

$$\mathbb{E}\left|\mathbf{P}_{j_0}r(x)-r(x)\right|^2\lesssim\left(\sum_{j=j_0}^{\infty}2^{-j(s-d/p)}\right)^2\lesssim 2^{-2j_0(s-d/p)}\sim n^{-2(s-d/p)/2(s-d/p)+d}. \quad (65)$$

Combining with (59), (62), and (65), the desired result is obtained by

$$\mathbb{E}\left|r\wedge^{\text{lin}}_n(x)-r(x)\right|^2\lesssim n^{-2(s-d/p)/2(s-d/p)+d}. \quad (66)$$

□

*Proof of* (17). By the definition of $\widehat{r}^{\text{non}}_n(x)$,

$$|\widehat{r}^{\text{non}}_n(x)-r(x)|\left|\widehat{r}^{\text{lin}}_n(x)-\mathbf{P}_{j_0}r(x)\right|+\left|\mathbf{P}_{j_1}r(x)-r(x)\right| ++\left|\sum_{j=j_0}^{j_1}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left(\widehat{\beta}^l_{j,k}\mathbf{I}_{\left\{\left|\widehat{\beta}^l_{j,k}\right|\geq\kappa t_n\right\}}-\beta^l_{j,k}\right)\psi^l_{j,k}(x)\right|. \quad (67)$$

Hence,

$$\mathbb{E}|r\wedge^{\text{non}}_n(x)-r(x)|^2\lesssim\mathbb{E}\left|r\wedge^{\text{lin}}_n(x)-\mathbf{P}_{j_0}r(x)\right|^2+\mathbb{E}\left|\mathbf{P}_{j_1}r(x)-r(x)\right|^2++\mathbb{E}\left|\sum_{j=j_0}^{j_1}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left(\beta\wedge^l_{j,k}\mathbf{I}_{\left\{|\beta\wedge^l_{j,k}|\geq\kappa t_n\right\}}-\beta^l_{j,k}\right)\psi^l_{j,k}(x)\right|^2. \quad (68)$$

According to the arguments of (62), $2^{j_0}\sim n^{1/2m+d}$ and $s>d/p$,

$$\mathbb{E}\left|r\wedge^{\text{lin}}_n(x)-\mathbf{P}_{j_0}r(x)\right|^2 2^{j_0d}n^{-1}\lesssim n^{-2m/2m+d}\lesssim n^{-2(s-d/p)/2(s-d/p)+d}. \quad (69)$$

Similar to the arguments of (65) and $2^{j_1}\sim\left(n/(\ln n)^3\right)^{1/d}$,

$$\mathbb{E}\left|\mathbf{P}_{j_1}r(x)-r(x)\right|^2\lesssim\left\{\sum_{j=j_1}^{\infty}2^{jd/2}\left\|\beta_{j\cdot}\right\|_p\right\}^2\lesssim 2^{-2j_1(s-d/p)}\lesssim\left(\frac{n}{(\ln n)^3}\right)^{-2(s-d/p)/d}(\ln n)^3\lesssim n^{-2(s-d/p)/2(s-d/p)+d}. \quad (70)$$

Then, we only need to estimate the upper bound of

$$Q:=\mathbb{E}\left|\sum_{j=j_0}^{j_1}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left(\beta\wedge^l_{j,k}\mathbf{I}_{\left\{|\beta\wedge^l_{j,k}|\geq\kappa t_n\right\}}-\beta^l_{j,k}\right)\psi^l_{j,k}(x)\right|^2. \quad (71)$$

Note that

$$Q\leq(j_1-j_0+1)\mathbb{E}\sum_{j=j_0}^{j_1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge^l_{j,k}\mathbf{I}_{\left\{|\beta\wedge^l_{j,k}|\geq\kappa t_n\right\}}-\beta^l_{j,k}\right|\left|\psi^l_{j,k}(x)\right|\right\}^2\lesssim(j_1-j_0+1)(Q_1+Q_2+Q_3+Q_4), \quad (72)$$

where

$$Q_1:=\mathbb{E}\sum_{j=j_0}^{j_1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge^l_{j,k}-\beta^l_{j,k}\right|\mathbf{I}_{\left\{|\beta\wedge^l_{j,k}|\geq\kappa t_n\right\}}\mathbf{I}_{\left\{|\beta^l_{j,k}|\geq\kappa/2t_n\right\}}\left|\psi^l_{j,k}(x)\right|\right\}^2,$$

$$Q_2:=\mathbb{E}\sum_{j=j_0}^{j_1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge^l_{j,k}-\beta^l_{j,k}\right|\mathbf{I}_{\left\{|\beta\wedge^l_{j,k}|\geq\kappa t_n\right\}}\mathbf{I}_{\left\{|\beta^l_{j,k}|<\kappa/2t_n\right\}}\left|\psi^l_{j,k}(x)\right|\right\}^2,$$

$$Q_3:=\mathbb{E}\sum_{j=j_0}^{j_1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta^l_{j,k}\right|\mathbf{I}_{\left\{|\beta\wedge^l_{j,k}|\leq\kappa t_n\right\}}\mathbf{I}_{\left\{|\beta^l_{j,k}|\geq 2\kappa t_n\right\}}\left|\psi^l_{j,k}(x)\right|\right\}^2,$$

$$Q_4:=\mathbb{E}\sum_{j=j_0}^{j_1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta^l_{j,k}\right|\mathbf{I}_{\left\{|\beta\wedge^l_{j,k}|\leq\kappa t_n\right\}}\mathbf{I}_{\left\{|\beta^l_{j,k}|\leq 2\kappa t_n\right\}}\left|\psi^l_{j,k}(x)\right|\right\}^2. \quad (73)$$

(i) *Upper Bound of $Q_2$ and $Q_3$*. It follows from Cauchy-Schwarz inequality that

$$\max\{Q_2,Q_3\} \lesssim \mathbb{E}\sum_{j=j_0}^{j_1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|\mathbf{I}_{\left\{\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|\geq\kappa/2t_n\right\}}\left|\psi_{j,k}^{l}(x)\right|\right\}^2$$
$$\lesssim \mathbb{E}\sum_{j=j_0}^{j_1}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\mathbf{I}_{\left\{\left|\widehat{\beta}_{j,k}^{l}-\beta_{j,k}^{l}\right|\geq\kappa/2t_n\right\}}\left|\psi_{j,k}^{l}(x)\right|\right)\cdot\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^{l}(x)\right|\right)$$
$$\lesssim \sum_{j=j_0}^{j_1}2^{jd/2}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left\{\left|\psi_{j,k}^{l}(x)\right|\mathbb{E}\left(\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\mathbf{I}_{\left\{\left|\widehat{\beta}_{j,k}^{l}-\beta_{j,k}^{l}\right|\geq\kappa/2t_n\right\}}\right)\right\}. \quad (74)$$

Using Lemmas 8 and 9, one gets

$$\mathbb{E}\left(\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\mathbf{I}_{\left\{\left|\widehat{\beta}_{j,k}^{l}-\beta_{j,k}^{l}\right|\geq\kappa/2t_n\right\}}\right)$$
$$\leq\left[\mathbb{E}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^4\right]^{1/2}\left[\mathbb{P}\left(\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|\geq\frac{\kappa}{2}t_n\right)\right]^{1/2}$$
$$\leq 2^{jd/2}\left[\mathbb{E}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\right]^{1/2}\left[\mathbb{P}\left(\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|\geq\frac{\kappa}{2}t_n\right)\right]^{1/2}$$
$$\lesssim 2^{j(d/2-w/2)}n^{-1/2}. \quad (75)$$

Then, choosing $w$ such that $w>2d+2m$, one has that

$$\max\{Q_2,Q_3\}\lesssim\sum_{j=j_0}^{j_1}2^{j(d/2-w/2)}n^{-1/2}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^{l}(x)\right|\sum_{j=j_0}^{j_1}2^{j(3d/2-w/2)}n^{-1/2}$$
$$\lesssim 2^{j_0(3d/2-w/2)}n^{-1/2}n^{-w/2+m-d/2m+d}n^{-2m/2m+d}n^{-2(s-d/p)/2(s-d/p)+d}. \quad (76)$$

(ii) *Upper Bound of* $Q_1$. To estimate $Q_1$, one defines

$$2^{j^*}\sim n^{1/2(s-d/p)+d}. \quad (77)$$

Note that $Q_1$ can be decomposed into the following two terms:

$$Q_1=\mathbb{E}\left(\sum_{j=j_0}^{j^*-1}+\sum_{j=j^*}^{j_1}\right)\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|\mathbf{I}_{\left\{\left|\beta\wedge_{j,k}^{l}\right|\geq\kappa t_n\right\}}\mathbf{I}_{\left\{\left|\beta_{j,k}^{l}\right|\geq k/2t_n\right\}}\left|\psi_{j,k}^{l}(x)\right|\right\}^2$$
$$:=Q_{11}+Q_{12}. \quad (78)$$

Using Cauchy-Schwarz inequality and Lemma 8,

$$Q_{11}\lesssim\mathbb{E}\sum_{j=j_0}^{j^*-1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|\left|\psi_{j,k}^{l}(x)\right|\right\}^2$$
$$\lesssim\mathbb{E}\sum_{j=j_0}^{j^*-1}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\left|\psi_{j,k}^{l}(x)\right|\right)\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^{l}(x)\right|\right)$$
$$\lesssim\sum_{j=j_0}^{j^*-1}2^{jd/2}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^{l}(x)\right|\mathbb{E}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\right)$$
$$\lesssim\sum_{j=j_0}^{j^*-1}2^{jd/2}n^{-1}\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^{l}(x)\right|$$
$$\lesssim\sum_{j=j_0}^{j^*-1}2^{jd}n^{-1}n^{-2(s-d/p)/2(s-d/p)+d}. \quad (79)$$

On the other hand, it is easy to see that

$$Q_{12}\lesssim\mathbb{E}\sum_{j=j^*}^{j_1}\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|\frac{\left|\beta_{j,k}^{l}\right|}{t_n}\left|\psi_{j,k}^{l}(x)\right|\right\}^2$$
$$\lesssim\frac{n}{\ln n}\mathbb{E}\sum_{j=j^*}^{j_1}\left\|\beta_{j.}\right\|_\infty^2\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\left|\psi_{j,k}^{l}(x)\right|\right)\cdot\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^{l}(x)\right|\right)$$
$$\lesssim n\sum_{j=j^*}^{j_1}2^{\frac{jd}{2}}\left\|\beta_{j.}\right\|_\infty^2\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^{l}(x)\right|\mathbb{E}\left|\beta\wedge_{j,k}^{l}-\beta_{j,k}^{l}\right|^2\right)$$
$$\lesssim\sum_{j=j^*}^{j_1}2^{jd}\left\|\beta_{j.}\right\|_p^2. \quad (80)$$

Then, one knows

$$Q_{12}\lesssim\sum_{j=j^*}^{j_1}\lesssim 2^{-2j(s-d/p)}2^{-2j^*(s-d/p)}\lesssim n^{-2(s-d/p)/2(s-d/p)+d}, \quad (81)$$

with Lemma 2. Now together with (78), (79), and (81),

$$Q_1\lesssim n^{-2(s-d/p)/2(s-d/p)+d}. \quad (82)$$

(iii) *Upper Bound of* $Q_4$. Note that

$$Q_4 = \mathbb{E}\left(\sum_{j=j_0}^{j^*-1} + \sum_{j=j^*}^{j_1}\right)\left\{\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta_{j,k}^l\right|\mathbf{I}_{\left\{\left|\hat{\beta}_{j,k}^l\right|\le\kappa t_n\right\}}\mathbf{I}_{\left\{\left|\beta_{j,k}^l\right|\le 2\kappa t_n\right\}}\left|\psi_{j,k}^l(x)\right|\right\}^2 \\ \coloneqq Q_{41} + Q_{42}. \quad (83)$$

By Condition $\theta$ and $t_n = \sqrt{\ln n/n}$,

$$Q_{41} \lesssim \sum_{j=j_0}^{j^*-1}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j} t_n\left|\psi_{j,k}^l(x)\right|\right)^2 \\ \lesssim \frac{\ln n}{n}\sum_{j=j_0}^{j^*-1} 2^{jd} \frac{\ln n}{n} 2^{j^* d}(\ln n) n^{-2(s-d/p)/2(s-d/p)+d}. \quad (84)$$

Moreover, according to Hölder inequality and Lemma 2, we can get

$$Q_{42} \lesssim \sum_{j=j^*}^{j_1}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta_{j,k}^l\right|\left|\psi_{j,k}^l(x)\right|\right)^2 \\ \lesssim \sum_{j=j^*}^{j_1}\left\{\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\beta_{j,k}^l\right|^p\left|\psi_{j,k}^l(x)\right|\right)^{1/p}\left(\sum_{l=1}^{M}\sum_{k\in\Omega_j}\left|\psi_{j,k}^l(x)\right|\right)^{1/p'}\right\}^2 \\ \lesssim \sum_{j=j^*}^{j_1} 2^{jd}\left\|\beta_{j.}\right\|_p^2 \sum_{j=j^*}^{j_1} 2^{-2j(s-d/p)} n^{-2(s-d/p)/2(s-d/p)+d}. \quad (85)$$

Combining this with (83) and (84),

$$Q_4 \lesssim (\ln n) n^{-2(s-d/p)/2(s-d/p)+d}. \quad (86)$$

Now the results (72), (76), (82), and (86) imply that

$$Q \lesssim (\ln n)^3 n^{-2(s-d/p)/2(s-d/p)+d}. \quad (87)$$

This with (68)–(70) shows

$$\mathbb{E}\left|\hat{r}_n^{\text{non}}(x) - r(x)\right|^2 \lesssim (\ln n)^3 n^{-2(s-d/p)/2(s-d/p)+d}. \quad (88)$$

□

## Data Availability

No data were used in this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This paper is supported by the National Natural Science Foundation of China (Nos. 12001133, 12001132, and 71961004), Guangxi Natural Science Foundation (Nos. 2018GXNSFBA281076 and 2019GXNSFFA245012), Guangxi Science and Technology Project (Nos. Guike AD18281058 and Guike AD18281019), and Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation. The authors would like to thank the two reviewers for their thoughtful comments.

## References

[1] Y. P. Chaubey, C. Chesneau, and E. Shirazi, "Wavelet-based estimation of regression function for dependent biased data under a given random design," *Journal of Nonparametric Statistics*, vol. 25, no. 1, pp. 53–71, 2013.

[2] Y. P. Chaubey and E. Shirazi, "On MISE of a non linear wavelet estimator of the regression function based on biased data under strong mixing," *Communications in Statistics-Theory and Methods*, vol. 44, no. 5, pp. 885–899, 2015.

[3] C. Chesneau and E. Shirazi, "Nonparametric wavelet regression based on biased data," *Communications in Statistics-Theory and Methods*, vol. 43, no. 13, pp. 2642–2658, 2014.

[4] J. K. Kou and Y. M. Liu, "An extension of Chesneau's theorem," *Statistics and Probability Letters*, vol. 108, pp. 23–32, 2016.

[5] H. J. Guo and J. K. Kou, "Wavelet regression estimation over $L^p$ risk based on negatively associated sample," *Applicable Analysis*, vol. 18, no. 4, pp. 2050024–2052667, 2020.

[6] J. K. Kou, J. Chen, and H. J. Guo, "MISE of wavelet estimators for regression derivatives with biased strong mixing data," *Communications in Statistics-Theory and Methods*, vol. 50, no. 14, pp. 3436–3452, 2021.

[7] V. V. Gorodetskii, "On the strong mixing property for linear sequences," *Theory of Probability and Its Applications*, vol. 22, no. 2, pp. 411–413, 1978.

[8] V. Genon-Catahot, T. Jeantheau, and C. Laredo, "Stochastic volatility models as hidden Markov models and statistical applications," *Bernoulli*, vol. 6, no. 6, pp. 1051–1079, 2000.

[9] S. Gala, E. Galakhov, M. A. Ragusa, and O. Salieva, "Beale–Kato–Majda regularity criterion of smooth solutions for the Hall-MHD equations with zero viscosity," *Bulletin of the Brazilian Mathematical Society, New Series*, 2021.

[10] M. Giacofci, S. Lambert-Lacroix, and F. Picard, "Minimax wavelet estimation for multisample heteroscedastic nonparametric regression," *Journal of Nonparametric Statistics*, vol. 30, no. 1, pp. 238–261, 2018.

[11] W. Jiang, *Topics in High-Dimensional Inference*, Rutgers University Community Repository, 2009.

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

[13] T. T. Cai, "Rates of convergence and adaptation over Besov space under pointwise risk," *Statistica Sinica*, vol. 13, no. 3, pp. 881–902, 2003.

[14] Y. A. Davydov, "The invariance principle for stationary processes," *Theory of Probability and Its Applications*, vol. 15, no. 3, pp. 487–498, 1970.

[15] E. Rio, "The functional law of the iterated logarithm for stationary strongly mixing sequences," *Annals of Probability*, vol. 23, no. 3, pp. 1188–1203, 1995.