CONSISTENCY OF WAVELET REGRESSION ESTIMATOR BASED ON BIASED DATA

HUIJUN GUO and JUNKE KOU*

Communicated by Ionel Popescu

When regression function belongs to Besov space, wavelet estimations are investigated by Chesneau and Shirazi (2014), Kou and Liu (2016, 2017). However, in many practical applications, one does not know whether the regression function is smooth or not. It makes sense to discuss consistency of wavelet estimator. This paper considers the mean $L^p (1 \le p \le \infty)$ consistency of wavelet estimator for multivariate regression functions based on biased data.

AMS 2020 Subject Classification: 62G07, 62G20, 42C40.

Key words: consistency, regression estimation, biased data, wavelets.

1. INTRODUCTION

Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be independent and identically distributed ($i.i.d.$) random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with common density function

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

where ω stands for a known positive function, g denotes the density function of the unobserved random variable (U, V) and

$$\mu = \mathbb{E}[\omega(U, V)] = \int_{\mathbb{R}} \int_{\mathbb{R}^d} \omega(x, y) g(x, y) \mathrm{d}x \mathrm{d}y < \infty.$$

In this setup g and f mean the target density and weighted density, respectively, and the resulting data are biased data. Then the problem is to estimate the unknown regression function

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

from the given biased data $(X_i, Y_i) (i = 1, 2, \ldots, n)$ in some sense.

The above model arises in many applications ([10]). An example is the estimation of the correlation $r(x)$ of agricultural output V and input U of a

*Corresponding author

country. We can obtain (X_i, Y_i) data from different agricultural regions of a country with X_i the input and Y_i the output. Since it is more likely to sample data from special regions, it means that the data we obtain is biased. Therefore the model we propose for the estimation of the true regression function of the output versus the input is important and valuable.

The aim of this paper is to develop wavelet estimations for the regression model with weaker conditions. Wavelet methods are of interest in nonparametric statistics thanks to their ability to estimate efficiently a wide variety of unknown functions, especially for those with discontinuties or sharp spikes. For this regression model, when the observed data (X_i, Y_i) $(i = 1, 2, \ldots, n)$ are independent, Chesneau and Shirazi [3] construct wavelet estimators of the regression function and discuss its mean integrated squared error (L^2-risk) over Besov space. An upper bound over L^p $(1 \leq p < \infty)$ risk of wavelet estimators are established by Kou and Liu [6]. Moreover, a lower bound estimations over L^p $(1 \leq p < \infty)$ risk are proved by Kou and Liu [7]. When the (X_i, Y_i) $(i = 1, 2, \ldots, n)$ is extended to strong mixing case, Chaubey, Chesneau and Shirazi [1, 2] study the L^2-risk of linear and nonlinear wavelet estimator respectively. It should be pointed out that those above results all require that the regression function belongs to Besov space, which means the regression function is smooth. However, in many practical cases we do not have smoothness of the regression function. As in the Example 2.3 and 2.4 of [6], the regression functions $r(x)$ have sharp spikes. Therefore, it is natural to consider the mean consistency of wavelet estimator $\hat{r}_n$, which means that $\mathbb{E}\|\hat{r}_n - r\|_p^p$ converges to zero as the sample size n tends to infinity.

In this article, we consider the mean L^p $(1 \leq p \leq \infty)$ consistency of linear wavelet estimator for the regression model. This work is an important supplement to the regression problems.

2. WAVELETS AND THREE LEMMAS

This section provides some concepts and important lemmas, which are needed for proving our main results in the next section. A multiresolution analysis (MRA) ([9]) is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of the square integrable function space $L^2(\mathbb{R}^d)$ satisfying:

(i) $V_j \subseteq V_{j+1}$, $j \in \mathbb{Z}$.

(ii) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}^d)$ (the space $\bigcup_{j\in\mathbb{Z}} V_j$ is 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 $\varphi(x) \in L^2(\mathbb{R}^d)$ (scaling function) such that

$$\{\varphi(\cdot - k), k \in \mathbb{Z}^d\}$$

forms an orthonormal basis of $V_0 = \overline{span}\{\varphi(\cdot - k)\}$.

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 $f \in L^2(\mathbb{R}^d)$,

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

If a scaling function φ satisfies Condition (θ), i.e.

$$\sum_{k \in \mathbb{Z}^d} |\varphi(x - k)| \in L^\infty(\mathbb{R}^d),$$

then the function $\varphi \in L(\mathbb{R}^d) \bigcap L^\infty(\mathbb{R}^d)$ (so that $\varphi \in L^p$ for $1 \le p \le \infty$), and $\sum_{k \in \mathbb{Z}^d} \varphi(x-k)\overline{\varphi(y-k)}$ converges absolutely almost everywhere. It can be shown that for $f \in L^p(\mathbb{R}^d)$ $(1 \le p \le \infty)$,

$$P_j f(x) = \sum_{k \in \mathbb{Z}^d} \alpha_{j,k}\varphi_{j,k}(x) = \int_{\mathbb{R}^d} K_j(x, y) f(y)\mathrm{d}y, \tag{3}$$

where $K(x, y) = \sum_{k \in \mathbb{Z}^d} \varphi(x - k)\overline{\varphi(y - k)}$ and $K_j(x, y) = 2^{jd}K(2^j x, 2^j y)$. Furthermore, the following lemma is true.

LEMMA 2.1. *Suppose that a scaling function $\varphi(x)$ satisfies Condition (θ) and $\alpha_k \in l_p$ $(1 \le p \le +\infty)$. Then there exist $0 < c_1 < c_2$ such that*

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

When $d = 1$, the proof of this lemma can be found in Härdle, *et al.* [5]. Similar arguments work as well for $d \ge 2$.

Now we introduce another concept, which is a little stronger than Condition (θ).

Condition S. There exists a bounded nonincreasing function $\Phi(x)$ such that $|\varphi(x)| \le \Phi(|x|)$ *(a.e.)* and $\int_{\mathbb{R}^d} \Phi(|x|)\mathrm{d}x < \infty$.

In this paper, we choose a compactly supported and bounded scaling function φ (*supp* $\varphi \subseteq \{x \in \mathbb{R}^d,\ |x| \le T\}$), such as the tensor product of Daubechies scaling function ([5,9]), so that φ satisfies Condition S. The following lemmas are proved by Liu and Xu [8] and will be used later on.

Lemma 2.2. *If a scaling function $\varphi(x)$ satisfies Condition S, then*

(1) $\int_{\mathbb{R}^d} K(x,y)\mathrm{d}y = 1$ *(a.e.);*

(2) *There exists $F(x) \in L^\infty(\mathbb{R}^d) \cap L(\mathbb{R}^d)$ such that*

$$|K(x,y)| \le F(x-y) \ (a.e.).$$

Lemma 2.3. *If a scaling function $\varphi(x)$ satisfies Condition S, then for $f(x) \in L^p(\mathbb{R}^d)$ $(1 \le p < \infty)$,*

$$\lim_{j\to\infty} \|P_j f(x) - f(x)\|_p = 0.$$

When $f(x)$ is uniformly continuous, this result is also true for $p = \infty$.

We also need the following inequality, which can be found in Härdle *et al.* [5].

Rosenthal's inequality. *Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbb{E}X_i = 0$ and $\mathbb{E}|X_i|^p < \infty$. Then*

$$\mathbb{E}\left|\sum_{i=1}^n X_i\right|^p \lesssim \begin{cases} \sum\limits_{i=1}^n \mathbb{E}|X_i|^p + \left(\sum\limits_{i=1}^n \mathbb{E}X_i^2\right)^{\frac{p}{2}}, & p \ge 2; \\ \left(\sum\limits_{i=1}^n \mathbb{E}X_i^2\right)^{\frac{p}{2}}, & 1 \le p < 2. \end{cases}$$

Here and after, we use $A \lesssim B$ to denote $A \le cB$ for some constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $B \lesssim A$.

3. WAVELET ESTIMATOR AND THEOREMS

In this section, we will construct a wavelet estimator and discuss its mean L^p $(1 \le p \le \infty)$ consistency.

For the regression model, a straightforward wavelet estimator of r can be constructed by estimating the projection $P_j r$ of r on V_j. Then a wavelet estimator is defined by

$$\widehat{r}_n(t) := \sum_k \widehat{\alpha}_{j,k}\varphi_{j,k}(t), \tag{4}$$

where

$$\widehat{\alpha}_{j,k} = \frac{\widehat{\mu}_n}{n}\sum_{i=1}^n \frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j,k}(X_i), \quad \widehat{\mu}_n = \left[\frac{1}{n}\sum_{i=1}^n \frac{1}{\omega(X_i,Y_i)}\right]^{-1} \tag{5}$$

and $h(x)$ denotes the density function of the random variable U.

It is easy to see from (1) and (2) that

$$\mathbb{E}\left(\frac{1}{\widehat{\mu}_n}\right)=\mathbb{E}\left(\frac{1}{\omega(X_i,Y_i)}\right)=\int_{\mathbb{R}^d}\int_{\mathbb{R}}\frac{f(x,y)}{\omega(x,y)}\mathrm{d}y\mathrm{d}x=\int_{\mathbb{R}^d}\int_{\mathbb{R}}\frac{g(x,y)}{\mu}\mathrm{d}y\mathrm{d}x=\frac{1}{\mu}$$

and

$$\begin{aligned}\mathbb{E}\left[\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j,k}(X_i)\right] &= \mathbb{E}\left[\frac{\mu\ \rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j,k}(X_i)\right]\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}}\frac{\mu\ \rho(y)}{\omega(x,y)h(x)}f(x,y)\varphi_{j,k}(x)\mathrm{d}y\mathrm{d}x\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}}\frac{\rho(y)g(x,y)}{h(x)}\mathrm{d}y\varphi_{j,k}(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^d}r(x)\varphi_{j,k}(x)\mathrm{d}x=\alpha_{j,k}.\end{aligned}$$

On the other hand, by the definitions of $\widehat{\alpha}_{j,k}$ and $K_j(x,y)$, one gets

$$\widehat{r}_n(t)=\frac{\widehat{\mu}_n}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i). \tag{6}$$

THEOREM 1. *Consider the regression model with* $\omega(x,y)\sim 1$, $h(x)\gtrsim 1$ *and* $\rho(y)\in L^\infty(\mathbb{R})$. *Then for* $r\in L^p(\mathbb{R}^d)$ $(1\le p<\infty)$ *and the estimator* $\widehat{r}_n$ *defined in (4) with* $2^{jd}\sim n^{\frac{1}{3}}$,

$$\lim_{n\to\infty}\mathbb{E}\|\widehat{r}_n-r\|_p^p=0.$$

Remark 1. In this theorem, we assume $1\le p<\infty$. Similar result is obtained in Theorem 2 for $p=\infty$. On the other hand, Chesneau and Shirazi [3], Kou and Liu [6] all require that the regression function r has compact support. However, we do not need this condition in Theorem 1 and Theorem 2.

Remark 2. The assumptions $\omega(x,y)\sim 1$ and $h(x)\gtrsim 1$ are standard for the nonparametric regression model with biased data, see Chaubey, Chesneau and Shirazi [1], Chesneau and Shirazi [3], Kou and Liu [6, 7].

Proof. It is easy to see that

$$\mathbb{E}\|\widehat{r}_n-r\|_p^p=\mathbb{E}\|\widehat{r}_n-P_jr+P_jr-r\|_p^p\lesssim\mathbb{E}\|\widehat{r}_n-P_jr\|_p^p+\|P_jr-r\|_p^p. \tag{7}$$

Using Lemma 2.3, $\lim_{j\to\infty}\|P_jr-r\|_p^p=0$. When $n\to\infty$, $j\to\infty$ because of $2^{jd}\sim n^{\frac{1}{3}}$. Hence,

$$\lim_{n\to\infty}\|P_jr-r\|_p^p=0.$$

The main work for the proof of Theorem 1 is to show

$$\lim_{n\to\infty}\mathbb{E}\|\widehat{r}_n-P_jr\|_p^p=0. \tag{8}$$

Note that

$$
\begin{aligned}
|\widehat{r}_n - P_j r| &= \left|\frac{\widehat{\mu}_n}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i) - P_j r\right| \\
&\leq \left|\frac{\widehat{\mu}_n}{\mu}\left[\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i) - P_j r\right]\right| \\
&\quad + \left|\widehat{\mu}_n\cdot P_j r\cdot\left(\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right)\right|.
\end{aligned}
$$

This, together with $\omega(x,y)\sim 1$ and the definition of $\widehat{\mu}_n$, shows

$$
\begin{aligned}
Q: &= \mathbb{E}\|\widehat{r}_n - P_j r\|_p^p = \mathbb{E}\int_{\mathbb{R}^d}|\widehat{r}_n - P_j r|^p \mathrm{d}t \\
&\lesssim \mathbb{E}\int_{\mathbb{R}^d}\left|\left[\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i) - P_j r\right]\right|^p \mathrm{d}t \\
&\quad + \|P_j r\|_p^p\ \mathbb{E}\left|\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right|^p.
\end{aligned}
$$

According to (3) and Lemma 2.2,

$$
\begin{aligned}
\|P_j r\|_p^p &\leq \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}|K_j(t,y)r(y)|\,\mathrm{d}y\right]^p \mathrm{d}t \\
&\lesssim \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}2^{jd}\left|F(2^j(t-y))r(y)\right|\mathrm{d}y\right]^p \mathrm{d}t \\
&\lesssim \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}\left|F(y)r(t-2^{-j}y)\right|\mathrm{d}y\right]^p \mathrm{d}t. \qquad (9)
\end{aligned}
$$

Take $\widetilde{F}(y) := \frac{F(y)}{\|F\|_1}$. Then $\widetilde{F}(y)\in L(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d}\widetilde{F}(y)\mathrm{d}y = 1$. By Jensen inequality and Fubini theorem, (9) reduces to

$$
\begin{aligned}
\|P_j r\|_p^p &\lesssim \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\widetilde{F}(y)|r(t-2^{-j}y)|^p \mathrm{d}y\mathrm{d}t \\
&\lesssim \int_{\mathbb{R}^d}\widetilde{F}(y)\int_{\mathbb{R}^d}|r(t-2^{-j}y)|^p \mathrm{d}t\mathrm{d}y \lesssim 1.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
Q &\lesssim \mathbb{E}\int_{\mathbb{R}^d}\left|\left[\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i) - P_j r\right]\right|^p \mathrm{d}t + \mathbb{E}\left|\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right|^p \\
&:= Q_1 + Q_2. \qquad (10)
\end{aligned}
$$

First, one estimates Q_2. By the definition of $\widehat{\mu}_n$,

$$
Q_2 = \mathbb{E}\left|\frac{1}{\mu}-\frac{1}{\widehat{\mu}_n}\right|^p = \frac{1}{n^p}\mathbb{E}\left|\sum_{i=1}^{n}\eta_i\right|^p \qquad (11)
$$

with $\eta_i = \frac{1}{\omega(X_i,Y_i)} - \frac{1}{\mu}$. Note that $\{\eta_i\}$ are $i.i.d$ and $\mathbb{E}\eta_i = 0$. Since $\omega(x,y) \sim 1$, $\mathbb{E}|\eta_i|^p \lesssim 1$. By Rosenthal's inequality,

$$Q_2 = \frac{1}{n^p}\mathbb{E}\left|\sum_{i=1}^{n} \eta_i\right|^p \lesssim \begin{cases} \frac{1}{n^p}[n + n^{\frac{p}{2}}], & p \geq 2; \\ n^{-\frac{p}{2}}, & 1 \leq p < 2. \end{cases}$$

Hence, for $1 \leq p < \infty$,

$$\lim_{n\to\infty} Q_2 = 0. \tag{12}$$

To estimate Q_1, one defines $\xi_i = \frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t, X_i) - P_j r(t)$. Note that

$$\begin{aligned} \mathbb{E}\left[\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i)\right] &= \int_{\mathbb{R}^d}\int_{\mathbb{R}} \frac{\mu\rho(y)}{\omega(x,y)h(x)}K_j(t,x)f(x,y)\mathrm{d}y\mathrm{d}x \\ &= \int_{\mathbb{R}^d} K_j(t,x)\int_{\mathbb{R}} \frac{\rho(y)g(x,y)}{h(x)}\mathrm{d}y\mathrm{d}x \\ &= \int_{\mathbb{R}^d} K_j(t,x)r(x)\mathrm{d}x = P_j r(t). \end{aligned}$$

Then $\{\xi_i\}$ are $i.i.d.$ and $\mathbb{E}\xi_i = 0$. It is easy to see that

$$\begin{aligned} Q_1 &= \mathbb{E}\int_{\mathbb{R}^d}\left|\left[\frac{\mu}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i) - P_j r\right]\right|^p \mathrm{d}t \\ &= \frac{1}{n^p}\int_{\mathbb{R}^d}\mathbb{E}\left|\sum_{i=1}^{n}\xi_i\right|^p \mathrm{d}t. \end{aligned} \tag{13}$$

Using Jensen inequality, $|P_j r(t)|^p \leq \mathbb{E}\left|\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i)\right|^p$. Hence,

$$\mathbb{E}|\xi_i|^p \lesssim \mathbb{E}\left|\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}K_j(t,X_i)\right|^p.$$

It follows from $\omega(x,y) \sim 1$, $h(x) \gtrsim 1$, $\rho(y) \in L^\infty(\mathbb{R})$ and Lemma 2.2 that

$$\begin{aligned} \mathbb{E}|\xi_i|^p &\lesssim \int_{\mathbb{R}}\int_{\mathbb{R}^d}\left|\frac{\mu\rho(y)}{\omega(x,y)h(x)}K_j(t,x)\right|^p f(x,y)\mathrm{d}x\mathrm{d}y \\ &\lesssim \int_{\mathbb{R}}\int_{\mathbb{R}^d} 2^{jdp}F^p(2^j t - 2^j x)f(x,y)\mathrm{d}x\mathrm{d}y \lesssim 2^{jdp}. \end{aligned} \tag{14}$$

When $p \geq 2$, the Rosenthal's inequality and Jensen inequality shows that

$$\mathbb{E}\left|\sum_{i=1}^{n}\xi_i\right|^p \lesssim \sum_{i=1}^{n}\mathbb{E}|\xi_i|^p + \left(\sum_{i=1}^{n}\mathbb{E}\xi_i^2\right)^{\frac{p}{2}} \lesssim n^{\frac{p}{2}}\mathbb{E}|\xi_i|^p. \tag{15}$$

Combining this with (13), one obtains

$$Q_1 \lesssim \frac{1}{n^{\frac{p}{2}}} \int_{\mathbb{R}^d} \mathbb{E}|\xi_i|^p \mathrm{d}t. \tag{16}$$

By (14), Lemma 2.2 and Fubini theorem,

$$\int_{\mathbb{R}^d} \mathbb{E}|\xi_i|^p \mathrm{d}t \lesssim \int_{\mathbb{R}} \int_{\mathbb{R}^d} f(x,y) \int_{\mathbb{R}^d} 2^{jdp} F^p(2^j t - 2^j x) \mathrm{d}t\mathrm{d}x\mathrm{d}y \lesssim 2^{jd(p-1)}. \tag{17}$$

This with (16) shows $Q_1 \lesssim \frac{2^{jd(p-1)}}{n^{\frac{p}{2}}}$. Hence,

$$\lim_{n\to\infty} Q_1 = \lim_{n\to\infty} \frac{1}{n^p} \int_{\mathbb{R}^d} \mathbb{E}\left|\sum_{i=1}^n \xi_i\right|^p \mathrm{d}t = 0 \tag{18}$$

with $2^{jd} \sim n^{\frac{1}{3}}$ and $2 \le p < \infty$.

For $1 \le p < 2$, by (13),

$$Q_1 \le \frac{1}{n} \int_{\mathbb{R}^d} \mathbb{E}\left|\sum_{i=1}^n \xi_i\right| \mathrm{d}t + \frac{1}{n^2} \int_{\mathbb{R}^d} \mathbb{E}\left|\sum_{i=1}^n \xi_i\right|^2 \mathrm{d}t. \tag{19}$$

Note that $\lim\limits_{n\to\infty} \frac{1}{n^2} \int_{\mathbb{R}^d} \mathbb{E}|\sum\limits_{i=1}^n \xi_i|^2 \mathrm{d}t = 0$ thanks to (18). It remains to prove

$$\lim_{n\to\infty} \frac{1}{n} \int_{\mathbb{R}^d} \mathbb{E}\left|\sum_{i=1}^n \xi_i\right| \mathrm{d}t = 0. \tag{20}$$

According to Rosenthal's inequality and (14),

$$\begin{aligned} \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^n \xi_i\right| &\lesssim \frac{1}{n}\left[\sum_{i=1}^n \mathbb{E}\xi_i^2\right]^{\frac{1}{2}} \lesssim n^{-\frac{1}{2}} \left(\mathbb{E}\xi_i^2\right)^{\frac{1}{2}} \\ &\lesssim n^{-\frac{1}{2}} \left[\int_{\mathbb{R}} \int_{\mathbb{R}^d} 2^{2jd} F^2(2^j t - 2^j x) f(x,y) \mathrm{d}x\mathrm{d}y\right]^{\frac{1}{2}} := A(t). \end{aligned} \tag{21}$$

On the other hand,

$$\begin{aligned} \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^n \xi_i\right| &\le \mathbb{E}|\xi_i| \lesssim \int_{\mathbb{R}} \int_{\mathbb{R}^d} 2^{jd} F(2^j t - 2^j x) f(x,y) \mathrm{d}x\mathrm{d}y \\ &\lesssim \int_{\mathbb{R}} \int_{\mathbb{R}^d} F(x) f(t - \frac{x}{2^j}, y) \mathrm{d}x\mathrm{d}y \\ &\lesssim \int_{\mathbb{R}} \int_{\mathbb{R}^d} F(x) f(t,y) \mathrm{d}x\mathrm{d}y \\ &\quad + \int_{\mathbb{R}} \int_{\mathbb{R}^d} F(x) \left| f(t - \frac{x}{2^j}, y) - f(t,y)\right| \mathrm{d}x\mathrm{d}y. \end{aligned}$$

$$:= B(t) + C(t). \tag{22}$$

Then we get that

$$\frac{1}{n}\int_{\mathbb{R}^d} \mathbb{E}\left|\sum_{i=1}^{n} \xi_i\right| \mathrm{d}t \lesssim \int_{\mathbb{R}^d} \min\{A(t), B(t)+C(t)\}\mathrm{d}t$$
$$\lesssim \int_{\mathbb{R}^d} \min\{A(t), B(t)\}\mathrm{d}t + \int_{\mathbb{R}^d} C(t)\mathrm{d}t. \tag{23}$$

One knows that

$$\int_{\mathbb{R}^d} C(t)\mathrm{d}t = \int_{\mathbb{R}^d}\int_{\mathbb{R}}\int_{\mathbb{R}^d} F(x)\left|f(t-\frac{x}{2^j}, y) - f(t,y)\right| \mathrm{d}x\mathrm{d}y\mathrm{d}t$$
$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}}\int_{\mathbb{R}^d} F(x)\left|f(t-\frac{x}{2^j}, y) - f(t,y)\right| \mathrm{d}t\mathrm{d}y\mathrm{d}x.$$

Since $\int_{\mathbb{R}}\int_{\mathbb{R}^d} F(x)|f(t-\frac{x}{2^j}, y) - f(t,y)|\mathrm{d}t\mathrm{d}y \le 2F(x) \in L(\mathbb{R}^d)$ and

$$\lim_{j\to\infty}\int_{\mathbb{R}}\int_{\mathbb{R}^d} |f(t-\frac{x}{2^j}, y) - f(t,y)|\mathrm{d}t\mathrm{d}y = 0,$$

then by the Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} C(t)\mathrm{d}t = 0. \tag{24}$$

It is easy to see that $\int_{\mathbb{R}^d} B(t)\mathrm{d}t = \int_{\mathbb{R}^d}\int_{\mathbb{R}}\int_{\mathbb{R}^d} F(x)f(t,y)\mathrm{d}x\mathrm{d}y\mathrm{d}t = \|F\|_1$. So, $B(t) \in L(\mathbb{R}^d)$. Moreover, $\lim_{n\to\infty} A(t) = 0$ thanks to Lemma 2.2 and $2^{jd} \sim n^{\frac{1}{3}}$. Hence,

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} \min\{A(t), B(t)\}\mathrm{d}t = \int_{\mathbb{R}^d} \lim_{n\to\infty} \min\{A(t), B(t)\}\mathrm{d}t = 0 \tag{25}$$

with the Lebesgue dominated convergence theorem. Combining this with (23) and (24),

$$\lim_{n\to\infty}\frac{1}{n}\int_{\mathbb{R}^d} \mathbb{E}|\sum_{i=1}^{n} \xi_i|\mathrm{d}t = 0.$$

□

Next we give the mean L^∞ consistency.

THEOREM 2. *Consider the regression model with $\omega(x,y) \sim 1$, $h(x) \gtrsim 1$ and $\rho(y) \in L^\infty(\mathbb{R})$. Let $r(x)$ is uniformly continuous and $2^{jd} \sim n^{\frac{1}{6}}$. If there exists a bounded non-increasing function γ such that $|r(x)| \le \gamma(|x|)$ and $\int_{\mathbb{R}^d}[\gamma(|x|)]^{\frac{1}{2}}\mathrm{d}x < +\infty$, then*

$$\lim_{n\to\infty}\mathbb{E}\|\widehat{r}_n - r\|_\infty = 0.$$

Remark 3. In Theorem 2, condition $\rho \in L^\infty(\mathbb{R})$ implies the boundedness of regression function r, this means $r \in L^\infty(\mathbb{R}^d)$. On the other hand, to estimate $\mathbb{E}\|\widehat{r}_n - r\|_\infty$, the continuity of r is essential, which can be found in Geng and Wang [4], Zeng and Wang [11].

Remark 4. The function ρ is assumed to be an identity function ($\rho(y) = y$) and $y \in [a, b]$ in Chaubey, *et al.* [1]. However, we only require $\rho(y) \in L^\infty(\mathbb{R})$. In addition, the estimation model is considered in a compact support ($(x, y) \in [0, 1]^d \times [a, b]$) by Chaubey, *et al.* [1]. But in this paper, we do not have this restrictive condition.

Proof. Similar to the arguments of Theorem 1,

$$\mathbb{E}\|\widehat{r}_n - r\|_\infty \lesssim \mathbb{E}\|\widehat{r}_n - P_j r\|_\infty + \|P_j r - r\|_\infty.$$

Then $\lim\limits_{n\to\infty} \|P_j r - r\|_\infty = 0$ by Lemma 2.3. We only need to prove

$$\lim_{n\to\infty} \mathbb{E}\|\widehat{r}_n - P_j r\|_\infty = 0. \tag{26}$$

To estimate $\mathbb{E}\|\widehat{r}_n - P_j r\|_\infty$, one defines

$$\widetilde{\alpha}_{j,k} := \frac{\mu}{n} \sum_{i=1}^{n} \frac{\rho(Y_i)}{\omega(X_i, Y_i)h(X_i)} \varphi_{j,k}(X_i), \qquad \widetilde{r}_n(x) := \sum_k \widetilde{\alpha}_{j,k} \varphi_{j,k}(x).$$

Then $\widehat{\alpha}_{j,k} = \frac{\widehat{\mu_n}}{\mu}\ \widetilde{\alpha}_{j,k}$ and $\widehat{r}_n = \frac{\widehat{\mu_n}}{\mu}\ \widetilde{r}_n$ thanks to (4) and (5). Furthermore, one knows that

$$\mathbb{E}\|\widehat{r}_n - P_j r\|_\infty = \mathbb{E}\left\| \frac{\widehat{\mu_n}}{\mu} (\widetilde{r}_n - P_j r) + \widehat{\mu_n}\ P_j r \left(\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n}\right)\right\|_\infty.$$

Obviously,

$$\|P_j r\|_\infty \le \int_{\mathbb{R}^d} |K_j(x, y) r(y)| \mathrm{d}y \le \int_{\mathbb{R}^d} F(y) |r(x - 2^j y)| \mathrm{d}y \lesssim 1$$

thanks to $r \in L^\infty(\mathbb{R}^d)$ and $F \in L(\mathbb{R}^d)$. On the other hand, $\omega(x, y) \sim 1$ implies $|\widehat{\mu_n}| \lesssim 1$. Hence

$$\mathbb{E}\|\widehat{r}_n - P_j r\|_\infty \lesssim \mathbb{E}\|\widetilde{r}_n - P_j r\|_\infty + \mathbb{E}\left|\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n}\right|.$$

This with the definition of $\widetilde{r}_n$ and Lemma 2.1 ($p = \infty$) shows

$$\mathbb{E}\|\widehat{r}_n - P_j r\|_\infty \lesssim 2^{\frac{jd}{2}} \sum_k \mathbb{E}|\widetilde{\alpha}_{j,k} - \alpha_{j,k}| + \mathbb{E}\left|\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n}\right|. \tag{27}$$

It is known that $\lim\limits_{n\to\infty} \mathbb{E}|\frac{1}{\mu} - \frac{1}{\widehat{\mu}_n}| = 0$ by the proof of (12).

Finally, one estimates $2^{\frac{jd}{2}}\sum\limits_k \mathbb{E}|\widetilde{\alpha}_{j,k}-\alpha_{j,k}|$. According to Rosenthal's inequality and $\mathbb{E}\widetilde{\alpha}_{j,k}=\alpha_{j,k}$,

$$\begin{aligned}\mathbb{E}|\widetilde{\alpha}_{j,k}-\alpha_{j,k}| &= \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^{n}\left(\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j,k}(X_i)-\alpha_{j,k}\right)\right| \\ &\lesssim \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}\left(\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j,k}(X_i)-\alpha_{j,k}\right)^2\right]^{\frac{1}{2}} \\ &\lesssim n^{-\frac{1}{2}}\left[\mathbb{E}\left(\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j,k}(X_i)\right)^2\right]^{\frac{1}{2}}. \qquad (28)\end{aligned}$$

It follows from (1), (2), $\omega(x,y)\sim 1$ and $h(x)\gtrsim 1$ that

$$\mathbb{E}\left(\frac{\mu\rho(Y_i)}{\omega(X_i,Y_i)h(X_i)}\varphi_{j,k}(X_i)\right)^2\lesssim\int_{\mathbb{R}^d}|r(x)||\varphi_{j,k}(x)|^2\mathrm{d}x.$$

Moreover,

$$\int_{\mathbb{R}^d}|r(x)||\varphi_{j,k}(x)|^2\mathrm{d}x\lesssim\int_{|x-k|\le T}|r(\frac{x}{2^j})|\mathrm{d}x$$

because φ is compactly supported and bounded. Furthermore, one has

$$2^{\frac{jd}{2}}\sum_k \mathbb{E}|\widetilde{\alpha}_{j,k}-\alpha_{j,k}|\lesssim\left(\frac{2^{jd}}{n}\right)^{\frac{1}{2}}\sum_k\left[\int_{|x-k|\le T}\left|r\left(\frac{x}{2^j}\right)\right|\mathrm{d}x\right]^{\frac{1}{2}}. \qquad (29)$$

Note that

$$\sum_k\left[\int_{|x-k|\le T}\left|r\left(\frac{x}{2^j}\right)\right|\mathrm{d}x\right]^{\frac{1}{2}}=\left(\sum_{|k|\le T+1}+\sum_{|k|\ge T+1}\right)\left[\int_{|x-k|\le T}\left|r\left(\frac{x}{2^j}\right)\right|\mathrm{d}x\right]^{\frac{1}{2}}.$$

Since $\rho(y)\in L^\infty(\mathbb{R})$ implies the boundedness of r,

$$\sum_{|k|\le T+1}\left[\int_{|x-k|\le T}|r(\frac{x}{2^j})|\mathrm{d}x\right]^{\frac{1}{2}}\lesssim 1.$$

On the other hand, $\sum\limits_{|k|\ge1}\gamma^{\frac{1}{2}}(\frac{|k|}{2^j})\lesssim 2^{jd}$ by the property of γ. This with $|r(x)|\le\gamma(|x|)$ leads to

$$\sum_k\left[\int_{|x-k|\le T}\left|r\left(\frac{x}{2^j}\right)\right|\mathrm{d}x\right]^{\frac{1}{2}}\lesssim 1+\sum_{|k|\ge T+1}\left[\int_{|x-k|\le T}\left|r\left(\frac{x}{2^j}\right)\right|\mathrm{d}x\right]^{\frac{1}{2}}$$

$$
\begin{aligned}
&\lesssim 1+\sum_{|k|\geq T+1}\left[\int_{|x-k|\leq T}\gamma\left(\frac{|x|}{2^j}\right)\mathrm{d}x\right]^{\frac{1}{2}}\\
&\lesssim 1+\sum_{|k|\geq T+1}\left[\int_{|x-k|\leq T}\gamma\left(\frac{|k|-T}{2^j}\right)\mathrm{d}x\right]^{\frac{1}{2}}\\
&\lesssim 1+\sum_{|k|\geq 1}\left[\gamma\left(\frac{|k|}{2^j}\right)\right]^{\frac{1}{2}}\lesssim 2^{jd}.
\end{aligned}
$$

This with (29) and $2^{jd}\sim n^{\frac{1}{6}}$ shows that

$$
2^{\frac{jd}{2}}\sum_k \mathbb{E}|\widetilde{\alpha}_{j,k}-\alpha_{j,k}|\lesssim n^{-\frac{1}{4}} \tag{30}
$$

and

$$
\lim_{n\to\infty} 2^{\frac{jd}{2}}\sum_k \mathbb{E}|\widetilde{\alpha}_{j,k}-\alpha_{j,k}|=0.
$$

The desired conclusion (26) is proved. □

Acknowledgments. The authors would like to thank the referee for insightful comments that helped us improve the paper significantly. This paper is supported by the National Natural Science Foundation of China (Nos. 12001132, 12001133), Guangxi Natural Science Foundation (No. 2019GXNSFFA245012), and Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation.

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.* J. Nonparametr. Stat. **25** (2013), 53-71.

[2] Y.P. Chaubey and E. Shirazi, *On MISE of a nonlinear wavelet estimation of regression function based on biased data under strong mixing.* Comm. Statist. Theory Methods **44** (2015), 885-899.

[3] C. Chesneau and E. Shirazi, *Nonparametric wavelet regression based on biased data.* Comm. Statist. Theory Methods **43** (2014), 2642-2658.

[4] Z.J. Geng and J.R. Wang, *The mean consistency of wavelet density estimators.* J. Inequal. Appl. 2015 (2015), paper no. 111, 14 p.

[5] W. Härdle, G. Kerkyacharian, D. Picard and A. Tsybakov, *Wavelets, Approximation and Statistical Applications.* Lect. Notes Stat. **129**, New York, Springer-Verlag, 1998.

[6] J.K. Kou and Y.M. Liu, *Nonparametric regression estimations over L^p risk based on biased data.* Comm. Statist. Theory Methods **46** (2017), 2375-2395.

[7] J.K. Kou and Y.M. Liu, *An extension of Chesneau's theorem.* Statist. Probab. Lett. **108** (2016), 23-32.

[8] Y.M. Liu and J.L. Xu, *On the L^p-consistency of wavelet estimators.* Acta Math. Sin. (Engl. Ser.) **32** (2016), 765-782.

[9] Y. Meyer, *Wavelets and Operators.* Actualités Mathématiques, Paris, Hermann, 1990.

[10] E. Sköld, *Kernel regression in the presence of size-bias.* J. Nonparametr. Stat. **12** (1999), *1*, 41-51.

[11] X.C. Zeng and J.R. Wang, *Wavelet density deconvolution estimations with heteroscedastic measurement errors.* Statist. Probab. Lett. **134** (2018), 79-85.

Received September 16, 2018

Guilin University of Electronic Technology
School of Mathematics and Computational Science
Guilin, Guangxi, P. R. China
guohuijun@guet.edu.cn

Guilin University of Electronic Technology
School of Mathematics and Computational Science
Guilin, Guangxi, P. R. China
kjkou@guet.edu.cn