AIMS Mathematics
http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(6): 14340–14361.
DOI: 10.3934/math.2023734
Received: 22 February 2023
Revised: 01 April 2023
Accepted: 06 April 2023
Published: 18 April 2023

Research article

Wavelet estimations of the derivatives of variance function in heteroscedastic model

Junke Kou*and Hao Zhang

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

* **Correspondence:** Email: kjkou@guet.edu.cn.

Abstract: This paper studies nonparametric estimations of the derivatives $r^{(m)}(x)$ of the variance function in a heteroscedastic model. Using a wavelet method, a linear estimator and an adaptive nonlinear estimator are constructed. The convergence rates under $L^{\tilde{p}}(1 \leq \tilde{p} < \infty)$ risk of those two wavelet estimators are considered with some mild assumptions. A simulation study is presented to validate the performances of the wavelet estimators.

Keywords: nonparametric wavelet estimation; derivative function; heteroscedastic model; $L^{\tilde{p}}$ risk
Mathematics Subject Classification: 62G07, 62G20, 42C40

1. Introduction

This paper considers the following heteroscedastic model:

$$Y_i = f(X_i)U_i + g(X_i), i \in \{1, \cdots, n\}. \tag{1.1}$$

In this equation, $g(x)$ is a known mean function, and the variance function $r(x)(r(x) := f^2(x))$ is unknown. Both the mean function $g(x)$ and variance function $r(x)$ are defined on $[0, 1]$. The random variables $U_1, \ldots, U_n$ are independent and identically distributed (*i.i.d.*) with $\mathrm{E}[U_i] = 0$ and $V[U_i] = 1$. Furthermore, the random variable X_i is independent of U_i for any $i \in \{1, \cdots, n\}$. The purpose of this paper is to estimate the mth derivative functions $r^{(m)}(x)(m \in N)$ from the observed data $(X_1, Y_1), \cdots, (X_n, Y_n)$ by a wavelet method.

Heteroscedastic models are widely used in economics, engineering, biology, physical sciences and so on; see Box [1], Carroll and Ruppert [2], Härdle and Tsybakov [3], Fan and Yao [4], Quevedo and Vining [5] and Amerise [6]. For the above estimation model (1.1), the most popular method is the kernel method. Many important and interesting results of kernel estimators have been obtained by

Wang et al. [7], Kulik and Wichelhaus [8] and Shen et al. [9]. However, the optimal bandwidth parameter of the kernel estimator is not easily obtained in some cases, especially when the function has some sharp spikes. Because of the good local properties in both time and frequency domains, the wavelet method has been widely used in nonparametric estimation problems; see Donoho and Johnstone [10], Cai [11], Nason et al. [12], Cai and Zhou [13], Abry and Didier [14] and Li and Zhang [15]. For the estimation problem (1.1), Kulik and Raimondo [16] studied the adaptive properties of warped wavelet nonlinear approximations over a wide range of Besov scales. Zhou et al. [17] developed wavelet estimators for detecting and estimating jumps and cusps in the mean function. Palanisamy and Ravichandran [18] proposed a data-driven estimator by applying wavelet thresholding along with the technique of sparse representation. The asymptotic normality for wavelet estimators of variance function under $\alpha-$mixing condition was obtained by Ding and Chen [19].

In this paper, we focus on nonparametric estimation of the derivative function $r^{(m)}(x)$ of the variance function $r(x)$. It is well known that derivative estimation plays an important and useful role in many practical applications (Woltring [20], Zhou and Wolfe, [21], Chacón and Duong [22], Wei et al. [23]). For the estimation model (1.1), a linear wavelet estimator and an adaptive nonlinear wavelet estimator for the derivative function $r^{(m)}(x)$ are constructed. Moreover, the convergence rates over $L^{\tilde{p}}(1 \leq \tilde{p} < \infty)$ risk of two wavelet estimators are proved in Besov space $B^s_{p,q}(\mathbb{R})$ with some mild conditions. Finally, numerical experiments are carried out, where an automatic selection method is used to obtain the best parameters of two wavelet estimators. According to the simulation study, both wavelet estimators can efficiently estimate the derivative function. Furthermore, the nonlinear wavelet estimator shows better performance than the linear estimator.

This paper considers wavelet estimations of a derivative function in Besov space. Now, we first introduce some basic concepts of wavelets. Let ϕ be an orthonormal scaling function, and the corresponding wavelet function is denoted by ψ. It is well known that $\{\phi_{\tau,k} := 2^{\tau/2}\phi(2^\tau x - k), \psi_{j,k} := 2^{j/2}\psi(2^j x - k), j \geq \tau, k \in \mathbb{Z}\}$ forms an orthonormal basis of $L^2(\mathbb{R})$. This paper uses the Daubechies wavelet, which has a compactly support. Then, for any integer j_*, a function $h(x) \in L^2([0,1])$ can be expanded into a wavelet series as

$$h(x) = \sum_{k\in\Lambda_{j_*}} \alpha_{j_*,k}\phi_{j_*,k}(x) + \sum_{j=j_*}^{\infty}\sum_{k\in\Lambda_j}\beta_{j,k}\psi_{j,k}(x), x \in [0,1]. \tag{1.2}$$

In this equation, $\Lambda_j = \{0, 1, \ldots, 2^j - 1\}$, $\alpha_{j_*,k} = \langle h, \phi_{j_*,k}\rangle_{[0,1]}$ and $\beta_{j,k} = \langle h, \psi_{j,k}\rangle_{[0,1]}$.

Lemma 1.1. *Let a scaling function ϕ be t-regular (i.e., $\phi \in \mathscr{C}^t$ and $|D^\alpha\phi(x)| \leq c(1 + |x|^2)^{-l}$ for each $l \in \mathbb{Z}$ and $\alpha = 0, 1, \ldots, t$). If $\{\alpha_k\} \in l_p$ and $1 \leq p \leq \infty$, there exist $c_2 \geq c_1 > 0$ such that*

$$c_1 2^{j(\frac{1}{2}-\frac{1}{p})}\left\|(\alpha_k)\right\|_p \leq \left\|\sum_{k\in\Lambda_j}\alpha_k 2^{\frac{j}{2}}\phi(2^j x - k)\right\|_p \leq c_2 2^{j(\frac{1}{2}-\frac{1}{p})}\left\|(\alpha_k)\right\|_p.$$

Besov spaces contain many classical function spaces, such as the well known Sobolev and Hölder spaces. The following lemma gives an important equivalent definition of a Besov space. More details about wavelets and Besov spaces can be found in Meyer [24] and Härdle et al. [25].

Lemma 1.2. *Let ϕ be t-regular and $h \in L^p([0,1])$. Then, for $p, q \in [1,\infty)$ and $0 < s < t$, the following assertions are equivalent:*

(i) $h \in B^s_{p,q}([0,1])$*;*
(ii) $\{2^{js}\|h - P_j h\|_p\} \in l_q$*;*
(iii) $\{2^{j(s-\frac{1}{p}+\frac{1}{2})}\|\beta_{j,k}\|_p\} \in l_q$.
The Besov norm of h can be defined by

$$\|h\|_{B^s_{p,q}} = \left\|(\alpha_{\tau,k})\right\|_p + \left\|(2^{j(s-\frac{1}{p}+\frac{1}{2})}\|\beta_{j,k}\|_p)_{j\geq\tau}\right\|_q,$$

where $\left\|\beta_{j,k}\right\|_p^p = \sum_{k\in\Lambda_j} \left|\beta_{j,k}\right|^p$.

2. Wavelet estimators and main theorem

In this section, we will construct our wavelet estimators, and give the main theorem of this paper. The main theorem shows the convergence rates of wavelet estimators under some mild assumptions. Now, we first give the technical assumptions of the estimation model (1.1) in the following.

A1: The variance function $r : [0,1] \to \mathbb{R}$ is bounded.

A2: For any $i \in \{0, \ldots, m-1\}$, variance function r satisfies $r^{(i)}(0) = r^{(i)}(1) = 0$.

A3: The mean function $g : [0,1] \to \mathbb{R}$ is bounded and known.

A4: The random variable X satisfies $X \sim U([0,1])$.

A5: The random variable U has a moment of order $2\tilde{p}$ $(\tilde{p} \geq 1)$.

In the above assumptions, A1 and A3 are conventional conditions for nonparametric estimations. The condition A2 is used to prove the unbiasedness of the following wavelet estimators. In addition, A4 and A5 are technique assumptions, which will be used in Lemmas 4.3 and 4.5.

According to the model (1.1), our linear wavelet estimator is constructed by

$$\hat{r}_n^{lin}(x) := \sum_{k\in\Lambda_{j_*}} \hat{\alpha}_{j_*,k}\phi_{j_*,k}(x). \tag{2.1}$$

In this definition, the scale parameter j_* will be given in the following main theorem, and

$$\hat{\alpha}_{j,k} := \frac{1}{n}\sum_{i=1}^{n} Y_i^2(-1)^m\phi_{j,k}^{(m)}(X_i) - \int_0^1 g^2(x)(-1)^m\phi_{j,k}^{(m)}(x)dx. \tag{2.2}$$

More importantly, it should be pointed out that this linear wavelet estimator is an unbiased estimator of the derivative function $r^{(m)}(x)$ by Lemma 4.1 and the properties of wavelets.

On the other hand, a nonlinear wavelet estimator is defined by

$$\hat{r}_n^{non}(x) := \sum_{k\in\Lambda_{j_*}} \hat{\alpha}_{j_*,k}\phi_{j_*,k}(x) + \sum_{j=j_*}^{j_1}\hat{\beta}_{j,k}\mathbb{I}_{\{|\hat{\beta}_{j,k}|\geq\kappa t_n\}}\psi_{j,k}(x). \tag{2.3}$$

In this equation, $\mathbb{I}_A$ denotes the indicator function over an event A, $t_n = 2^{mj}\sqrt{\ln n/n}$,

$$\hat{\beta}_{j,k} := \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^2(-1)^m\psi_{j,k}^{(m)}(X_i) - w_{j,k}\right)\mathbb{I}_{\left\{\left|Y_i^2(-1)^m\psi_{j,k}^{(m)}(X_i) - w_{j,k}\right|\leq\rho_n\right\}}, \tag{2.4}$$

$\rho_n = 2^{mj}\sqrt{n/\ln n}$, and $w_{j,k} = \int_0^1 g^2(x)(-1)^m \psi_{j,k}^{(m)}(x)dx$. The positive integer j_* and j_1 will also be given in our main theorem, and the constant κ will be chosen in Lemma 4.5. In addition, we adopt the following symbol: $x_+ := \max\{x, 0\}$. $A \lesssim B$ denotes $A \le cB$ for some constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $B \lesssim A$.

In this position, the convergence rates of two wavelet estimators are given in the following main theorem.

Main theorem For the estimation model (1.1) with the assumptions **A1**-**A5**, $r^{(m)}(x) \in B^s_{p,q}([0,1])(p, q \in [1,\infty), s > 0)$ and $1 \le \tilde{p} < \infty$, if $\{p > \tilde{p} \ge 1, s > 0\}$ or $\{1 \le p \le \tilde{p}, s > 1/p\}$.

(a) the linear wavelet estimator $\hat{r}_n^{lin}(x)$ with $s' = s - (\frac{1}{p} - \frac{1}{\tilde{p}})_+$ and $2^{j_*} \sim n^{\frac{1}{2s'+2m+1}}$ satisfies

$$E\left[\left\|\hat{r}_n^{lin}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \lesssim n^{-\frac{\tilde{p}s'}{2s'+2m+1}}. \tag{2.5}$$

(b) the nonlinear wavelet estimator $\hat{r}_n^{non}(x)$ with $2^{j_*} \sim n^{\frac{1}{2t+2m+1}}$ $(t > s)$ and $2^{j_1} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2m+1}}$ satisfies

$$E\left[\left\|\hat{r}_n^{non}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \lesssim (\ln n)^{\tilde{p}-1}\left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}, \tag{2.6}$$

where

$$\delta = min\left\{\frac{s}{2s+2m+1}, \frac{s-1/p+1/\tilde{p}}{2(s-1/p)+2m+1}\right\} = \begin{cases} \frac{s}{2s+2m+1} & p > \frac{\tilde{p}(2m+1)}{2s+2m+1} \\ \frac{s-1/p+1/\tilde{p}}{2(s-1/p)+2m+1} & p \le \frac{\tilde{p}(2m+1)}{2s+2m+1}. \end{cases}$$

Remark 1. Note that $n^{-\frac{s\tilde{p}}{2s+1}}$ $(n^{-\frac{(s-1/p+1/\tilde{p})\tilde{p}}{2(s-1/p)+1}})$ is the optimal convergence rate over $L^{\tilde{p}}(1 \le \tilde{p} < +\infty)$ risk for nonparametric wavelet estimations (Donoho et al. [26]). The linear wavelet estimator can obtain the optimal convergence rate when $p > \tilde{p} \ge 1$ and $m = 0$.

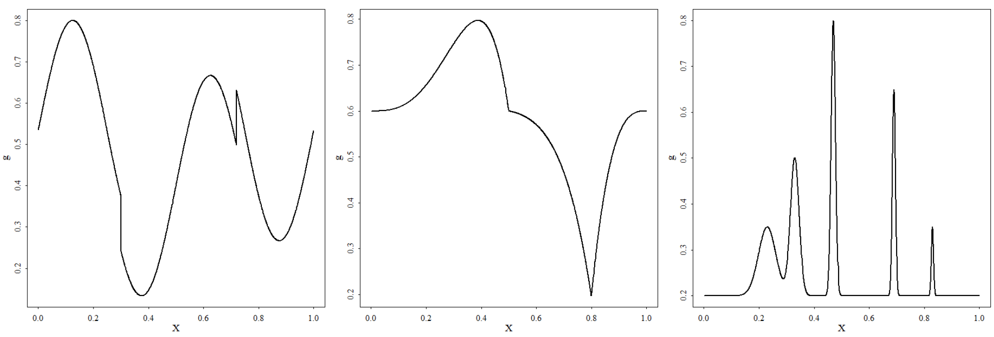
Remark 2. When $m = 0$, this derivative estimation problem reduces to the classical variance function estimation. Then, the convergence rates of the nonlinear wavelet estimator are same as the optimal convergence rates of nonparametric wavelet estimation up to a $\ln n$ factor in all cases.

Remark 3. According to main theorem (a) and the definition of the linear wavelet estimator, it is easy to see that the construction of the linear wavelet estimator depends on the smooth parameter s of the unknown derivative function $r^{(m)}(x)$, which means that the linear estimator is not adaptive. Compared with the linear estimator, the nonlinear wavelet estimator only depends on the observed data and the sample size. Hence, the nonlinear estimator is adaptive. More importantly, the nonlinear wavelet estimator has a better convergence rate than the linear estimator in the case of $p \le \tilde{p}$.

3. Simulation study

In order to illustrate the empirical performance of the proposed estimators, we produce a numerical illustration using an adaptive selection method, which is used to obtain the best parameters of the wavelet estimators. For the problem (1.1), we choose three common functions, *HeaviSine*, *Corner* and *Spikes*, as the mean function $g(x)$; see Figure 1. Those functions are usually used in wavelet literature.

On the other hand, we choose the function $f(x)$ by $f_1(x) = 3(4x-2)^2e^{-(4x-2)^2}$, $f_2(x) = sin(2\pi sin\pi x)$ and $f_3(x) = -(2x-1)^2 + 1$, respectively. In addition, we assume that the random variable U satisfies $U \sim N[0, 1]$. The aim of this paper is to estimate the derivative function $r^{(m)}(x)$ of the variance function $r(x)(r = f^2)$ by the observed data $(X_1, Y_1), \ldots, (X_n, Y_n)$. In this section, we adopt $r_1(x) = [f_1(x)]^2$, $r_2(x) = [f_2(x)]^2$ and $r_3(x) = [f_3(x)]^2$. For the sake of simplicity, our simulation study focuses on the derivative function $r'(x)(m = 1)$ and $r(x)(m = 0)$ by the observed data $(X_1, Y_1), \ldots, (X_n, Y_n)$ $(n = 4096)$. Furthermore, we use the mean square error ($MSE\ (\hat{r}(x), r(x)) = \frac{1}{n}\sum_{i=1}^{n}(\hat{r}(X_i) - r(X_i))^2$) and the average magnitude of error ($AME\ (\hat{r}(x), r(x)) = \frac{1}{n}\sum_{i=1}^{n}|\hat{r}(X_i) - r(X_i)|$) to evaluate the performances of the wavelet estimators separately.

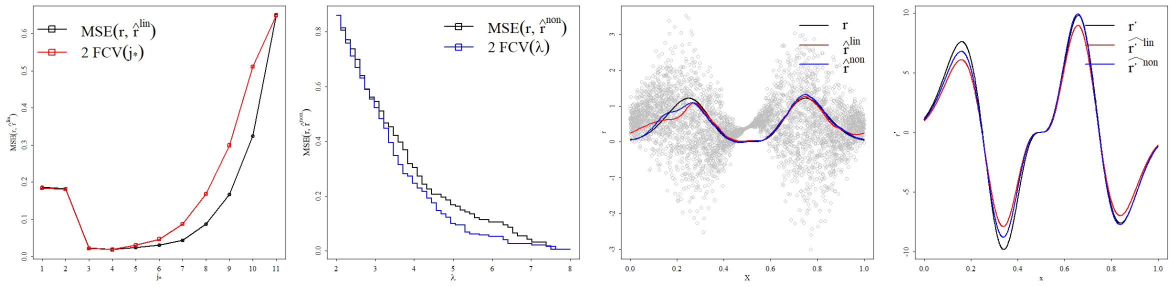


(a) (b) (c)

Figure 1. Three mean functions. (a) $HeaviSine$, (b) $Corner$, (c) $Spikes$.

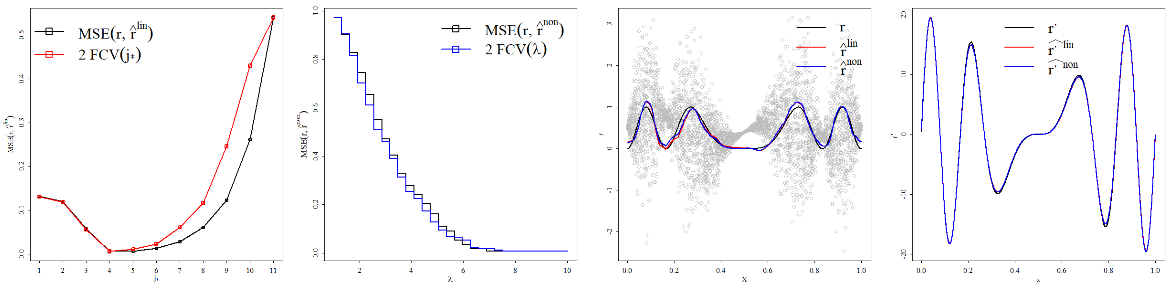
For the linear and nonlinear wavelet estimators, the scale parameter j_* and threshold value λ ($\lambda = \kappa t_n$) play important roles in the function estimation problem. In order to obtain the optimal scale parameter and threshold value of wavelet estimators, this section uses the two-fold cross validation (2FCV) approach (Nason [27], Navarro and Saumard [28]). During the first example of simulation study, we choose $HeaviSine$ as the mean function $g(x)$, and $f_1(x) = 3(4x-2)^2e^{-(4x-2)^2}$. The estimation results of two wavelet estimators are presented by Figure 2. For the optimal scale parameter j_* of the linear wavelet estimator, we built a collection of j_* and $j_* = 1, \ldots, log2(n) - 1$. The best parameter j_* is selected by minimizing a 2FCV criterion denoted by 2FCV(j_*); see Figure 2(a). According to Figure 2(a), it is easy to see that the 2FCV(j_*) and MSE both can get the minimum value when $j_* = 4$. For the nonlinear wavelet estimator, the best threshold value λ is also obtained by the 2FCV(λ) criterion in Figure 2(b). Meanwhile, the parameter j_* is same as the linear estimator, and the parameter j_1 is chosen as the maximum scale parameter $log2(n)-1$. From Figure 2(c) and 2(d), the linear and nonlinear wavelet estimators both can get a good performance with the best scale parameter and threshold value. More importantly, the nonlinear wavelet estimator shows better performance than the linear estimator.

In the following simulation study, more numerical experiments are presented to sufficiently verify the performance of the wavelet method. According to Figures 3–10, the wavelet estimators both can obtain good performances in different cases. Especially, the nonlinear wavelet estimator gets better estimation results than the linear estimator. Also, the MSE and AME of the wavelet estimators in all examples are provided by Table 1. Meanwhile, it is easy to see from Table 1 that the nonlinear wavelet estimators can have better performance than the linear estimators.



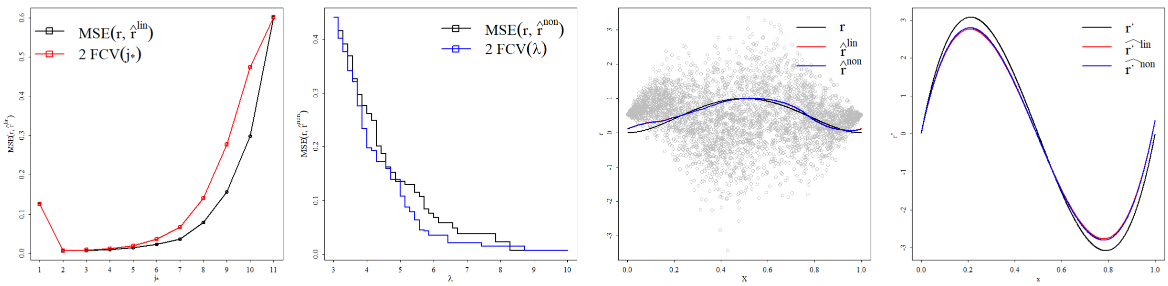
(a) (b) (c) (d)

Figure 2. The estimation results of wavelet estimators when $g(x)$ is $HeaviSine$ and $r(x) = r_1(x)$. (a) Graphs of the MSE (black line) and 2FCV criterion (red line) of the linear estimator. (b) Graphs of the MSE (black line) and 2FCV criterion (blue line) of the nonlinear estimator. (c) Fluctuating data (X, Y) (gray circles), the true variance $r(x)$ (black line), the linear estimator $\hat{r}^{lin}$ (red line) and the nonlinear estimator $\hat{r}^{non}$ (blue line). (d) The estimation results of the linear (red line) and nonlinear (blue line) for derivative function $r'(x)$.



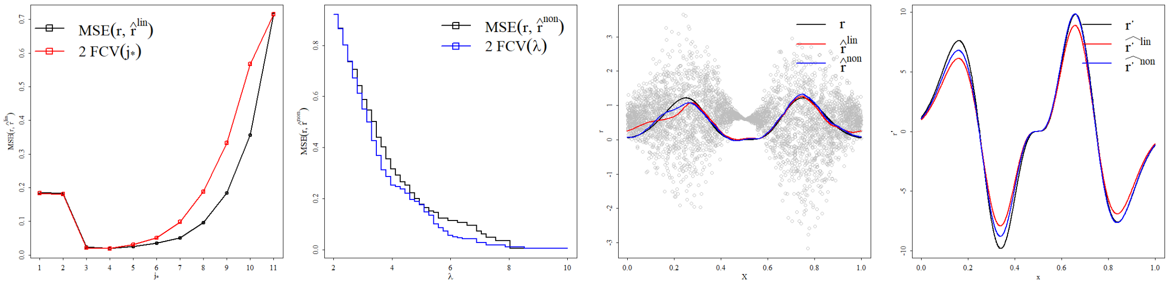
(a) (b) (c) (d)

Figure 3. The estimation results of wavelet estimators when $g(x)$ is $HeaviSine$ and $r(x) = r_2(x)$.



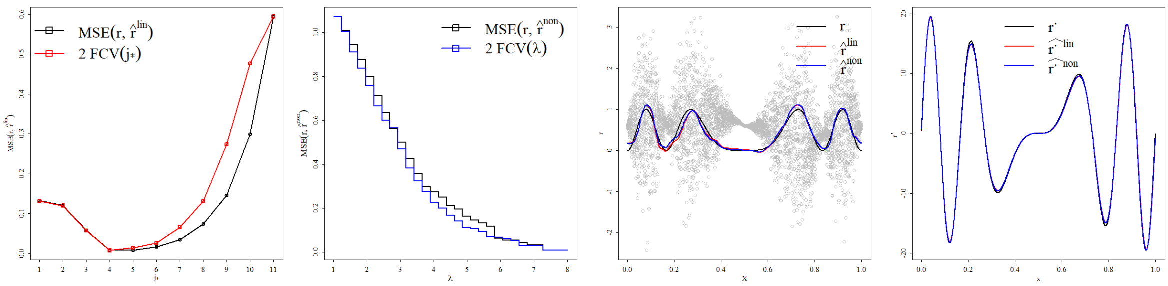
(a) (b) (c) (d)

Figure 4. The estimation results of wavelet estimators when $g(x)$ is $HeaviSine$ and $r(x) = r_3(x)$.



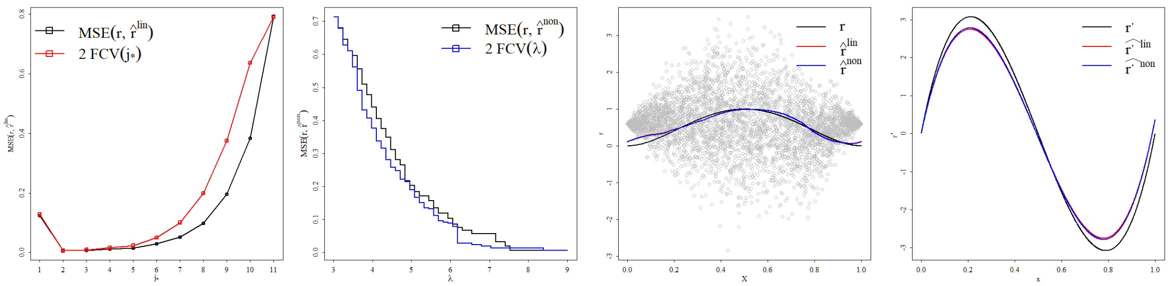
(a) (b) (c) (d)

Figure 5. The estimation results of wavelet estimators when $g(x)$ is *Corner* and $r(x) = r_1(x)$.



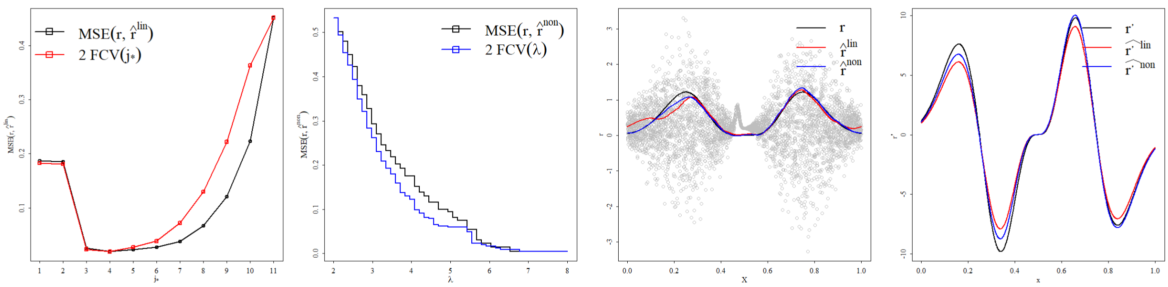
(a) (b) (c) (d)

Figure 6. The estimation results of wavelet estimators when $g(x)$ is *Corner* and $r(x) = r_2(x)$.



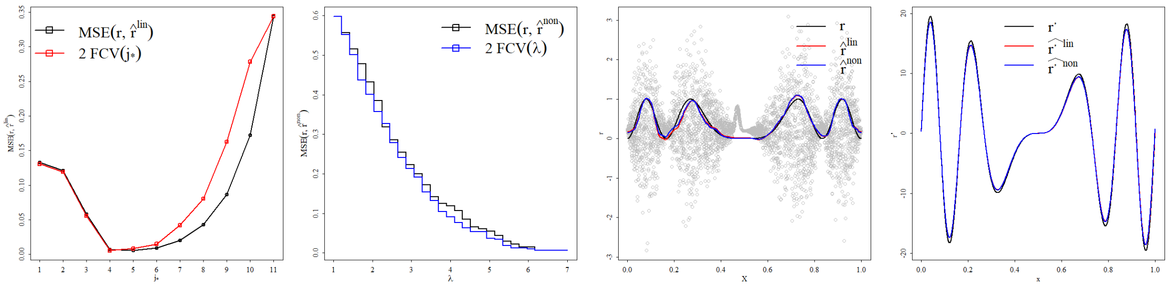
(a) (b) (c) (d)

Figure 7. The estimation results of wavelet estimators when $g(x)$ is *Corner* and $r(x) = r_3(x)$.



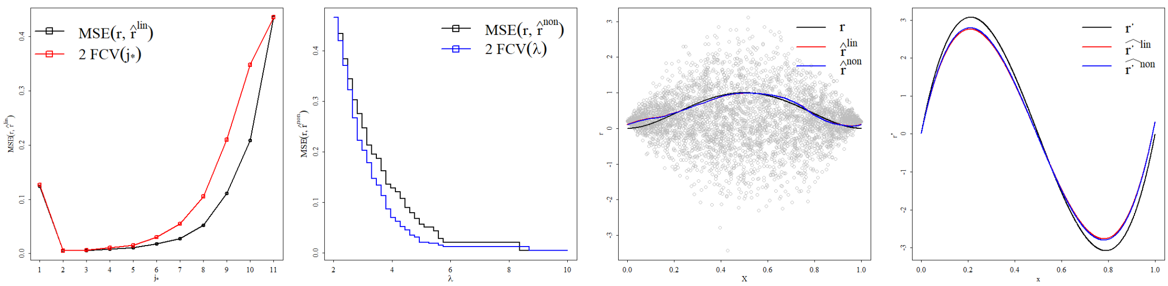
(a) (b) (c) (d)

Figure 8. The estimation results of wavelet estimators when $g(x)$ is *Spikes* and $r(x) = r_1(x)$.



(a) (b) (c) (d)

Figure 9. The estimation results of wavelet estimators when $g(x)$ is $Spikes$ and $r(x) = r_2(x)$.



(a) (b) (c) (d)

Figure 10. The estimation results of wavelet estimators when $g(x)$ is $Spikes$ and $r(x) = r_3(x)$.

Table 1. The MSE and AME of the wavelet estimators.

	$HeaviSine$			$Corner$			$Spikes$		
	r_1	r_2	r_3	r_1	r_2	r_3	r_1	r_2	r_3
$MSE(\hat{r}^{lin}, r)$	0.0184	0.0073	0.0071	0.0189	0.0075	0.0064	0.0189	0.0069	0.0052
$MSE(\hat{r}^{non}, r)$	0.0048	0.0068	0.0064	0.0044	0.0070	0.0057	0.0042	0.0061	0.0046
$MSE(\hat{r}'^{lin}, r')$	0.7755	0.0547	0.0676	0.7767	0.1155	0.0737	0.7360	0.2566	0.0655
$MSE(\hat{r}'^{non}, r')$	0.2319	0.0573	0.0560	0.2204	0.0644	0.0616	0.2406	0.2868	0.0539
$AME(\hat{r}^{lin}, r)$	0.0935	0.0653	0.0652	0.0973	0.0667	0.0615	0.0964	0.0621	0.0550
$AME(\hat{r}^{non}, r)$	0.0506	0.0641	0.0619	0.0486	0.0649	0.0583	0.0430	0.0595	0.0518
$AME(\hat{r}'^{lin}, r')$	0.6911	0.1876	0.2348	0.7021	0.2686	0.2451	0.6605	0.4102	0.2320
$AME(\hat{r}'^{non}, r')$	0.3595	0.1862	0.2125	0.3450	0.2020	0.2229	0.3696	0.4198	0.2095

4. Proof of main theorem

4.1. Auxiliary results

Now, we provide some lemmas for the proof of the main Theorem.

Lemma 4.1. *For the model (1.1) with A2 and A4,*

$$E[\hat{\alpha}_{j,k}] = \alpha_{j,k}, \tag{4.1}$$

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^2(-1)^m\psi_{j,k}^{(m)}(X_i)-w_{j,k}\right)\right]=\beta_{j,k}. \tag{4.2}$$

Proof. According to the definition of $\hat{\alpha}_{j,k}$,

$$\begin{aligned}
\mathrm{E}[\hat{\alpha}_{j,k}] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}Y_i^2(-1)^m\phi_{j,k}^{(m)}(X_i)-\int_0^1 g^2(x)(-1)^m\phi_{j,k}^{(m)}(x)dx\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[Y_i^2(-1)^m\phi_{j,k}^{(m)}(X_i)\right]-\int_0^1 g^2(x)(-1)^m\phi_{j,k}^{(m)}(x)dx\\
&= \mathrm{E}\left[Y_1^2(-1)^m\phi_{j,k}^{(m)}(X_1)\right]-\int_0^1 g^2(x)(-1)^m\phi_{j,k}^{(m)}(x)dx\\
&= \mathrm{E}\left[r(X_1)U_1^2(-1)^m\phi_{j,k}^{(m)}(X_1)\right]+2\mathrm{E}[f(X_1)U_1g(X_1)(-1)^m\phi_{j,k}^{(m)}(X_1)]\\
&+ \mathrm{E}\left[g^2(X_1)(-1)^m\phi_{j,k}^{(m)}(X_1)\right]-\int_0^1 g^2(x)(-1)^m\phi_{j,k}^{(m)}(x)dx.
\end{aligned}$$

Then, it follows from A4 that

$$\mathrm{E}\left[g^2(X_1)(-1)^m\phi_{j,k}^{(m)}(X_1)\right]=\int_0^1 g^2(x)(-1)^m\phi_{j,k}^{(m)}(x)dx.$$

Using the assumption of independence between U_i and X_i,

$$\mathrm{E}\left[r(X_1)U_1^2(-1)^m\phi_{j,k}^{(m)}(X_1)\right]=\mathrm{E}[U_1^2]\mathrm{E}\left[r(X_1)(-1)^m\phi_{j,k}^{(m)}(X_1)\right],$$

$$\mathrm{E}[f(X_1)U_1g(X_1)(-1)^m\phi_{j,k}^{(m)}(X_1)]=\mathrm{E}[U_1]\mathrm{E}[f(X_1)g(X_1)(-1)^m\phi_{j,k}^{(m)}(X_1)].$$

Meanwhile, the conditions $\mathrm{V}[U_1]=1$ and $\mathrm{E}[U_1]=0$ imply $\mathrm{E}[U_1^2]=1$. Hence, one gets

$$\begin{aligned}
\mathrm{E}[\hat{\alpha}_{j,k}] &= \mathrm{E}\left[r(X_1)(-1)^m\phi_{j,k}^{(m)}(X_1)\right]\\
&= \int_0^1 r(x)(-1)^m\phi_{j,k}^{(m)}(x)dx=(-1)^m\int_0^1 r(x)\phi_{j,k}^{(m)}(x)dx\\
&= \int_0^1 r^{(m)}(x)\phi_{j,k}(x)dx=\alpha_{j,k}
\end{aligned}$$

by the assumption A2.

On the other hand, one takes ψ instead of ϕ, and $w_{j,k}$ instead of $\int_0^1 g^2(x)(-1)^m\phi_{j,k}^{(m)}(x)dx$. The second equation will be proved by the similar mathematical arguments. □

Lemma 4.2. *(Rosenthal's inequality) Let* $X_1,\ldots,X_n$ *be independent random variables such that* $\mathrm{E}[X_i]=0$ *and* $\mathrm{E}[|X_i|^p]<\infty$. *Then,*

$$\mathrm{E}\left[\left|\sum_{i=1}^{n}X_i\right|^p\right]\lesssim\begin{cases}\sum_{i=1}^{n}\mathrm{E}\left[|X_i|^p\right]+\left(\sum_{i=1}^{n}\mathrm{E}\left[|X_i|^2\right]\right)^{\frac{p}{2}}, & p>2,\\ \left(\sum_{i=1}^{n}\mathrm{E}\left[|X_i|^2\right]\right)^{\frac{p}{2}}, & 1\le p\le 2.\end{cases}$$

Lemma 4.3. *For the model (1.1) with A1–A5, $2^j \le n$ and $1 \le \tilde{p} < \infty$,*

$$E\left[\left|\hat{\alpha}_{j,k} - \alpha_{j,k}\right|^{\tilde{p}}\right] \lesssim n^{-\frac{\tilde{p}}{2}} 2^{\tilde{p}mj}, \tag{4.3}$$

$$E\left[\left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^{\tilde{p}}\right] \lesssim \left(\frac{\ln n}{n}\right)^{-\frac{\tilde{p}}{2}} 2^{\tilde{p}mj}. \tag{4.4}$$

Proof. By (4.1) and the independence of random variables X_i and U_i, one has

$$\begin{aligned}
\left|\hat{\alpha}_{j,k} - \alpha_{j,k}\right| &= \left|\frac{1}{n}\sum_{i=1}^{n} Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i) - \int_0^1 g^2(x)(-1)^m \phi_{j,k}^{(m)}(x)dx - E\left[\hat{\alpha}_{j,k}\right]\right| \\
&= \frac{1}{n}\left|\sum_{i=1}^{n}\left(Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i) - E\left[Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i)\right]\right)\right| \\
&= \frac{1}{n}\left|\sum_{i=1}^{n} A_i\right|.
\end{aligned}$$

In this above equation, $A_i := Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i) - E\left[Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i)\right]$.

According to the definition of A_i, one knows that $E[A_i] = 0$ and

$$\begin{aligned}
E\left[|A_i|^{\tilde{p}}\right] &= E\left[\left|Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i) - E\left[Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i)\right]\right|^{\tilde{p}}\right] \\
&\lesssim E\left[\left|Y_i^2(-1)^m \phi_{j,k}^{(m)}(X_i)\right|^{\tilde{p}}\right] \\
&\lesssim E\left[\left|(r(X_1)U_1^2 + g^2(X_1))(-1)^m \phi_{j,k}^{(m)}(X_i)\right|^{\tilde{p}}\right] \\
&\lesssim E\left[U_1^{2\tilde{p}}\right] E\left[\left|r(X_1)\phi_{j,k}^{(m)}(X_i)\right|^{\tilde{p}}\right] + E\left[\left|g^2(X_1)\phi_{j,k}^{(m)}(X_i)\right|^{\tilde{p}}\right].
\end{aligned}$$

The assumption A5 shows $E[U_1^{2\tilde{p}}] \lesssim 1$. Furthermore, it follows from A1 and A3 that

$$\begin{aligned}
E[U_1^{2\tilde{p}}]E\left[|r(X_1)\phi_{j,k}^{(m)}(X_1)|^{\tilde{p}}\right] &\lesssim E\left[|\phi_{j,k}^{(m)}(X_1)|^{\tilde{p}}\right], \\
E\left[g^{2\tilde{p}}(X_1)|\phi_{j,k}^{(m)}(X_1)|^{\tilde{p}}\right] &\lesssim E\left[|\phi_{j,k}^{(m)}(X_1)|^{\tilde{p}}\right].
\end{aligned}$$

In addition, and the properties of wavelet functions imply that

$$\begin{aligned}
E\left[\left|\phi_{j,k}^{(m)}(X_i)\right|^{\tilde{p}}\right] = \int_0^1 |\phi_{j,k}^{(m)}(x)|^{\tilde{p}}dx &= 2^{j(\tilde{p}/2+m\tilde{p}-1)}\int_0^1 |\phi^{(m)}(2^jx-k)|^{\tilde{p}}d(2^jx-k) \\
&= 2^{j(\tilde{p}/2+m\tilde{p}-1)}\|\phi^{(m)}\|_{\tilde{p}}^{\tilde{p}} \lesssim 2^{j(\tilde{p}/2+m\tilde{p}-1)}.
\end{aligned}$$

Hence,

$$E\left[|A_i|^{\tilde{p}}\right] \lesssim 2^{j(\tilde{p}/2+m\tilde{p}-1)}.$$

Especially in $\tilde{p} = 2$, $E\left[|A_i|^2\right] \lesssim 2^{2mj}$.

Using Rosenthal's inequality and $2^j \le n$,

$$\begin{aligned}
\mathrm{E}\left[\left|\hat{\alpha}_{j,k}-\alpha_{j,k}\right|^{\tilde{p}}\right] &= \frac{1}{n^{\tilde{p}}}\mathrm{E}\left[\left|\sum_{i=1}^{n} A_i\right|^{\tilde{p}}\right] \\
&\lesssim \begin{cases} \frac{1}{n^{\tilde{p}}}\left(\sum_{i=1}^{n}\mathrm{E}\left[|A_i|^{\tilde{p}}\right]+(\sum_{i=1}^{n}\mathrm{E}\left[|A_i|^2\right])^{\frac{\tilde{p}}{2}}\right), & \tilde{p}>2, \\ \frac{1}{n^{\tilde{p}}}\left(\sum_{i=1}^{n}\mathrm{E}\left[|A_i|^2\right]\right)^{\frac{\tilde{p}}{2}}, & 1\le\tilde{p}\le 2, \end{cases} \\
&\lesssim \begin{cases} \frac{1}{n^{\tilde{p}}}\left(n\cdot 2^{j(\frac{\tilde{p}}{2}+m\tilde{p}-1)}+(n\cdot 2^{2mj})^{\frac{\tilde{p}}{2}}\right), & \tilde{p}>2, \\ \frac{1}{n^{\tilde{p}}}\left(n\cdot 2^{2mj}\right)^{\frac{\tilde{p}}{2}}, & 1\le\tilde{p}\le 2, \end{cases} \\
&\lesssim n^{-\frac{\tilde{p}}{2}}2^{\tilde{p}mj}.
\end{aligned}$$

Then, the first inequality is proved.

For the second inequality, note that

$$\begin{aligned}
\beta_{j,k} &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^2(-1)^m\psi_{j,k}^{(m)}(X_i)-w_{j,k}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\left(Y_i^2(-1)^m\psi_{j,k}^{(m)}(X_i)-\int_0^1 g^2(x)(-1)^m\psi_{j,k}^{(m)}(x)dx\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[K_i\right]
\end{aligned}$$

with (4.2) and $K_i := Y_i^2(-1)^m\psi_{j,k}^{(m)}(X_i)-\int_0^1 g^2(x)(-1)^m\psi_{j,k}^{(m)}(x)dx$.

Let $B_i := K_i\mathbb{I}_{\{|K_i|\le\rho_n\}}-\mathrm{E}\left[K_i\mathbb{I}_{\{|K_i|\le\rho_n\}}\right]$. Then, by the definition of $\hat{\beta}_{j,k}$ in (2.4),

$$|\hat{\beta}_{j,k}-\beta_{j,k}| = |\frac{1}{n}\sum_{i=1}^{n}K_i\mathbb{I}_{\{|K_i|\le\rho_n\}}-\beta_{j,k}| \le \frac{1}{n}\left|\sum_{i=1}^{n}B_i\right|+\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[|K_i|\mathbb{I}_{\{|K_i|>\rho_n\}}\right]. \tag{4.5}$$

Similar to the arguments of A_i, it is easy to see that $\mathrm{E}\left[B_i\right]=0$ and

$$\mathrm{E}\left[|B_i|^{\tilde{p}}\right] \lesssim \mathrm{E}\left[\left|K_i\mathbb{I}_{\{|K_i|\le\rho_n\}}\right|^{\tilde{p}}\right] \lesssim \mathrm{E}\left[|K_i|^{\tilde{p}}\right] \lesssim 2^{j(\frac{\tilde{p}}{2}+m\tilde{p}-1)}.$$

Especially in the case of $\tilde{p}=2$, one can obtain $\mathrm{E}\left[|B_i|^2\right] \lesssim 2^{2mj}$. On the other hand,

$$\mathrm{E}\left[|K_i|\mathbb{I}_{\{|K_i|>\rho_n\}}\right] \lesssim \mathrm{E}\left[|K_i|\cdot\frac{|K_i|}{\rho_n}\right] = \frac{\mathrm{E}\left[K_1^2\right]}{\rho_n} \lesssim \frac{2^{2mj}}{\rho_n} = t_n = 2^{mj}\sqrt{\frac{\ln n}{n}}. \tag{4.6}$$

According to Rosenthal's inequality and $2^j \le n$,

$$\begin{aligned}
\mathrm{E}\left[|\hat{\beta}_{j,k}-\beta_{j,k}|^{\tilde{p}}\right] &\lesssim \frac{1}{n^{\tilde{p}}}\mathrm{E}\left[\left|\sum_{i=1}^{n} B_i\right|^{\tilde{p}}\right]+(t_n)^{\tilde{p}} \\
&\lesssim \begin{cases} \frac{1}{n^{\tilde{p}}}\left(\sum_{i=1}^{n}\mathrm{E}\left[|B_i|^{\tilde{p}}\right]+(\sum_{i=1}^{n}\mathrm{E}\left[|B_i|^2\right])^{\frac{\tilde{p}}{2}}\right)+(t_n)^{\tilde{p}}, & \tilde{p}>2, \\ \frac{1}{n^{\tilde{p}}}\left(\sum_{i=1}^{n}\mathrm{E}\left[|B_i|^2\right]\right)^{\frac{\tilde{p}}{2}}+(t_n)^{\tilde{p}}, & 1\le \tilde{p}\le 2, \end{cases} \\
&\lesssim \begin{cases} \frac{1}{n^{\tilde{p}}}\left(n\cdot 2^{j(\frac{\tilde{p}}{2}+m\tilde{p}-1)}+(n\cdot 2^{2mj})^{\frac{\tilde{p}}{2}}\right)+\left(\frac{\ln n}{n}\right)^{-\frac{\tilde{p}}{2}}\cdot 2^{\tilde{p}mj}, & \tilde{p}>2, \\ \frac{1}{n^{\tilde{p}}}\left(n\cdot 2^{2mj}\right)^{\frac{\tilde{p}}{2}}+\left(\frac{\ln n}{n}\right)^{-\frac{\tilde{p}}{2}}\cdot 2^{\tilde{p}mj}, & 1\le \tilde{p}\le 2, \end{cases} \\
&\lesssim \left(\frac{\ln n}{n}\right)^{-\frac{\tilde{p}}{2}} 2^{\tilde{p}mj}.
\end{aligned}$$

Then, the second inequality is proved. □

Lemma 4.4. *(Bernstein's inequality) Let* $X_1,\ldots,X_n$ *be independent random variables such that* $\mathrm{E}[X_i]=0$*,* $|X_i|<M$ *and* $\mathrm{E}[|X_i|^2]:=\sigma^2$*. Then, for each* $v>0$

$$\mathrm{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}X_i\right|\ge v\right)\le 2\exp\left\{-\frac{nv^2}{2(\sigma^2+vM/3)}\right\}.$$

Lemma 4.5. *For the model (1.1) with A1–A5 and* $1\le\tilde{p}<+\infty$*, there exists a constant* $\kappa>1$ *such that*

$$\mathrm{P}\left(\left|\hat{\beta}_{j,k}-\beta_{j,k}\right|\ge \kappa t_n\right)\lesssim n^{-\tilde{p}}. \tag{4.7}$$

Proof. According to (4.5), one gets $K_i=Y_i^2(-1)^m\psi_{j,k}^{(m)}(X_i)-\int_0^1 g^2(x)(-1)^m\psi_{j,k}^{(m)}(x)dx$, $B_i=K_i\mathbb{I}_{\{|K_i|\le\rho_n\}}-\mathrm{E}\left[K_i\mathbb{I}_{\{|K_i|\le\rho_n\}}\right]$ and

$$|\hat{\beta}_{j,k}-\beta_{j,k}|\le\frac{1}{n}\left|\sum_{i=1}^{n}B_i\right|+\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[|K_i|\mathbb{I}_{\{|K_i|>\rho_n\}}\right].$$

Meanwhile, (4.6) shows that there exists $c>0$ such that $\mathrm{E}\left[|K_i|\mathbb{I}_{\{|K_i|>\rho_n\}}\right]\le ct_n$. Furthermore, the following conclusion is true.

$$\begin{aligned}
\left\{|\hat{\beta}_{j,k}-\beta_{j,k,u}|\ge\kappa t_n\right\} &\subseteq \left\{\left[\frac{1}{n}\left|\sum_{i=1}^{n}B_i\right|+\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(|K_i|\mathbb{I}_{\{|K_i|>\rho_n\}}\right)\right]\ge\kappa t_n\right\} \\
&\subseteq \left\{\frac{1}{n}\left|\sum_{i=1}^{n}B_i\right|\ge(\kappa-c)t_n\right\}.
\end{aligned}$$

Note that the definition of B_i implies that $|B_i|\lesssim\rho_n$ and $\mathrm{E}[B_i]=0$. Using the arguments of Lemma 4.3, $\mathrm{E}[B_i^2]:=\sigma^2\lesssim 2^{2mj}$. Furthermore, by Bernstein's inequality,

$$\begin{aligned}
\mathrm{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n} B_i\right| \geq (\kappa-c)t_n\right) &\lesssim \exp\left\{-\frac{n(\kappa-c)^2 {t_n}^2}{2(\sigma^2+(\kappa-c)t_n\rho_n/3)}\right\} \\
&\lesssim \exp\left\{-\frac{n(\kappa-c)^2 2^{2mj}\cdot\frac{\ln n}{n}}{2(2^{2mj}+(\kappa-c)\cdot 2^{2mj}/3)}\right\} \\
&= \exp\left\{-(\ln n)\frac{(\kappa-c)^2}{2(1+(\kappa-c)/3)}\right\} \\
&= n^{-\frac{(\kappa-c)^2}{2(1+(\kappa-c)/3)}}.
\end{aligned}$$

Then, one can choose large enough κ such that

$$\mathrm{P}\left(\left|\hat{\beta}_{j,k}-\beta_{j,k}\right| \geq \kappa t_n\right) \lesssim n^{-\frac{(\kappa-c)^2}{2(1+(\kappa-c)/3)}} \lesssim n^{-\tilde{p}}.$$

□

4.2. *Proof of main theorem*

Proof of (a): Note that

$$\left\|\hat{r}_n^{lin}(x)-r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} \lesssim \left\|\hat{r}_n^{lin}(x)-P_{j_*}r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} + \left\|P_{j_*}r^{(m)}(x)-r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}$$

Hence,

$$\mathrm{E}\left[\left\|\hat{r}_n^{lin}(x)-r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \lesssim \mathrm{E}\left[\left\|\hat{r}_n^{lin}(x)-P_{j_*}r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] + \left\|P_{j_*}r^{(m)}(x)-r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}. \tag{4.8}$$

■ The stochastic term $\mathrm{E}\left[\left\|\hat{r}_n^{lin}(x)-P_{j_*}r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right]$.
It follows from Lemma 1.1 that

$$\begin{aligned}
\mathrm{E}\left[\left\|\hat{r}_n^{lin}(x)-P_{j_*}r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] &= \mathrm{E}\left[\left\|\sum_{k\in\Lambda_{j_*}}\left(\hat{\alpha}_{j_*,k}-\alpha_{j_*,k}\right)\phi_{j_*,k}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \\
&\sim 2^{j_*(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_{j_*}} \mathrm{E}\left[\left|\hat{\alpha}_{j_*,k}-\alpha_{j_*,k}\right|^{\tilde{p}}\right].
\end{aligned}$$

Then, according to (4.3), $|\Lambda_{j_*}| \sim 2^{j_*}$ and $2^{j_*} \sim n^{\frac{1}{2s'+2m+1}}$, one gets

$$\mathrm{E}\left[\left\|\hat{r}_n^{lin}(x)-P_{j_*}r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \sim 2^{j_*\frac{\tilde{p}}{2}(2m+1)}\cdot n^{-\frac{\tilde{p}}{2}} \sim n^{-\frac{\tilde{p}s'}{2s'+2m+1}}. \tag{4.9}$$

■ The bias term $\left\|P_{j_*}r^{(m)}(x)-r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}$.
When $p > \tilde{p} \geq 1$, $s' = s-(\frac{1}{p}-\frac{1}{\tilde{p}})_+ = s$. Using Hölder inequality, Lemma 1.2 and $r^{(m)} \in B^s_{p,q}([0,1])$,

$$\left\|P_{j_*}r^{(m)}(x)-r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} \lesssim \left\|P_{j_*}r^{(m)}(x)-r^{(m)}(x)\right\|_{p}^{\tilde{p}} \lesssim 2^{-j_*\tilde{p}s} = 2^{-j_*\tilde{p}s'} \sim n^{-\frac{\tilde{p}s'}{2s'+2m+1}}.$$

When $1 \le p \le \tilde{p}$ and $s > \dfrac{1}{p}$, one knows that $B^s_{p,q}([0,1]) \subseteq B^{s'}_{\tilde{p},\infty}([0,1])$ and

$$\left\|P_{j_*} r^{(m)}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} \lesssim 2^{-j_* \tilde{p} s'} \sim n^{-\frac{\tilde{p}s'}{2s'+2m+1}}.$$

Hence, the following inequality holds in both cases.

$$\left\|P_{j_*} r^{(m)}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} \lesssim n^{-\frac{\tilde{p}s'}{2s'+2m+1}}. \tag{4.10}$$

Finally, the results (4.8)–(4.10) show

$$\mathrm{E}\left[\left\|\hat{r}_n^{lin}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \lesssim n^{-\frac{\tilde{p}s'}{2s'+2m+1}}.$$

Proof of (b): By the definitions of $\hat{r}_n^{lin}(x)$ and $\hat{r}_n^{non}(x)$, one has

$$\begin{aligned}\left\|\hat{r}_n^{non}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} \lesssim{}& \left\|\hat{r}_n^{lin}(x) - P_{j_*} r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} + \left\|r^{(m)}(x) - P_{j_1+1} r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} \\ &+ \left\|\sum_{j=j_*}^{j_1} \sum_{k\in\Lambda_j} \left(\hat{\beta}_{j,k} \mathbb{I}_{\{|\hat{\beta}_{j,k}| \ge \kappa t_n\}} - \beta_{j,k}\right)\psi_{j,k}(x)\right\|_{\tilde{p}}^{\tilde{p}}.\end{aligned}$$

Furthermore,

$$\mathrm{E}\left[\left\|\hat{r}_n^{non}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \lesssim T_1 + T_2 + Q. \tag{4.11}$$

In this above inequality,

$$\begin{aligned}T_1 &:= \mathrm{E}\left[\left\|\hat{r}_n^{lin}(x) - P_{j_*} r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right], \\ T_2 &:= \left\|r^{(m)}(x) - P_{j_1+1} r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}, \\ Q &:= \mathrm{E}\left[\left\|\sum_{j=j_*}^{j_1} \sum_{k\in\Lambda_j} \left(\hat{\beta}_{j,k} \mathbb{I}_{\{|\hat{\beta}_{j,k}| \ge \kappa t_n\}} - \beta_{j,k}\right)\psi_{j,k}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right].\end{aligned}$$

■ For T_1. According to (4.9) and $2^{j_*} \sim n^{\frac{1}{2t+2m+1}}$ $(t > s)$,

$$T_1 \sim 2^{j_* \frac{\tilde{p}}{2}(2m+1)} \cdot n^{-\frac{\tilde{p}}{2}} \sim n^{-\frac{\tilde{p}t}{2t+2m+1}} < n^{-\frac{\tilde{p}s}{2s+2m+1}} \le n^{-\tilde{p}\delta}. \tag{4.12}$$

■ For T_2. Using similar mathematical arguments as (4.10), when $p > \tilde{p} \ge 1$, one can obtain $T_2 := \left\|r^{(m)}(x) - P_{j_1+1} r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}} \lesssim 2^{-j_1 \tilde{p} s}$. This with $2^{j_1} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2m+1}}$ leads to

$$T_2 \lesssim 2^{-j_1 \tilde{p} s} < \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2m+1}} \le \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2s+2m+1}} \le \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

On the other hand, when $1 \le p \le \tilde{p}$ and $s > \dfrac{1}{p}$, one has $B^s_{p,q}([0,1]) \subseteq B^{s-\frac{1}{p}+\frac{1}{\tilde{p}}}_{\tilde{p},\infty}([0,1])$ and

$$T_2 \lesssim 2^{-j_1 \tilde{p}(s-1/p+1/\tilde{p})} \sim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}(s-1/p+1/\tilde{p})}{2m+1}} < \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}(s-1/p+1/\tilde{p})}{2(s-1/p)+2m+1}} \le \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

Therefore, for each $1 \le \tilde{p} < \infty$,

$$T_2 \lesssim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \tag{4.13}$$

■ For Q. According to Hölder inequality and Lemma 1.1,

$$\begin{aligned} Q &\lesssim (j_1 - j_* + 1)^{\tilde{p}-1} \sum_{j=j_*}^{j_1} \mathrm{E}\left[\left\|\sum_{k\in\Lambda_j}\left(\hat{\beta}_{j,k}\mathbb{I}_{\{|\hat{\beta}_{j,k}|\ge\kappa t_n\}} - \beta_{j,k}\right)\psi_{j,k}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \\ &\lesssim (j_1 - j_* + 1)^{\tilde{p}-1} \sum_{j=j_*}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k}\mathbb{I}_{\{|\hat{\beta}_{j,k}|\ge\kappa t_n\}} - \beta_{j,k}|^{\tilde{p}}\right]. \end{aligned}$$

Note that

$$\begin{aligned} |\hat{\beta}_{j,k}\mathbb{I}_{\{|\hat{\beta}_{j,k}|\ge\kappa t_n\}} - \beta_{j,k}|^{\tilde{p}} &= |\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\hat{\beta}_{j,k}|\ge\kappa t_n, |\beta_{j,k}|<\frac{\kappa t_n}{2}\}} + |\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\hat{\beta}_{j,k}|\ge\kappa t_n, |\beta_{j,k}|\ge\frac{\kappa t_n}{2}\}} \\ &\quad + |\beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\hat{\beta}_{j,k}|<\kappa t_n, |\beta_{j,k}|>2\kappa t_n\}} + |\beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\hat{\beta}_{j,k}|<\kappa t_n, |\beta_{j,k}|\le 2\kappa t_n\}}. \end{aligned}$$

Meanwhile,

$$\{|\hat{\beta}_{j,k}| \ge \kappa t_n, |\beta_{j,k}| < \frac{\kappa t_n}{2}\} \subseteq \{|\hat{\beta}_{j,k} - \beta_{j,k}| > \frac{\kappa t_n}{2}\},$$
$$\{|\hat{\beta}_{j,k}| < \kappa t_n, |\beta_{j,k}| > 2\kappa t_n\} \subseteq \{|\hat{\beta}_{j,k} - \beta_{j,k}| > \kappa t_n\} \subseteq \{|\hat{\beta}_{j,k} - \beta_{j,k}| > \frac{\kappa t_n}{2}\}.$$

Then, Q can be decomposed as

$$Q \lesssim (j_1 - j_* + 1)^{\tilde{p}-1}\left(Q_1 + Q_2 + Q_3\right), \tag{4.14}$$

where

$$Q_1 := \sum_{j=j_*}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\hat{\beta}_{j,k}-\beta_{j,k}|>\frac{\kappa t_n}{2}\}}\right],$$
$$Q_2 := \sum_{j=j_*}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\beta_{j,k}|\ge\frac{\kappa t_n}{2}\}}\right],$$
$$Q_3 := \sum_{j=j_*}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\beta_{j,k}|\le 2\kappa t_n\}}.$$

■ For Q_1. It follows from the Hölder inequality that

$$\mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}}\mathbb{I}_{\{|\hat{\beta}_{j,k}-\beta_{j,k}|>\frac{\kappa t_n}{2}\}}\right] \le \left(\mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{2\tilde{p}}\right]\right)^{\frac{1}{2}}\left[\mathrm{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| > \frac{\kappa t_n}{2}\right)\right]^{\frac{1}{2}}.$$

By Lemma 4.3, one gets

$$\mathrm{E}\left[\left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^{2\tilde{p}}\right] \lesssim \left(\frac{\ln n}{n}\right)^{-\tilde{p}} \cdot 2^{2\tilde{p}mj}.$$

This with Lemma 4.5, $|\Lambda_j| \sim 2^j$ and $2^{j_1} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2m+1}}$ shows that

$$Q_1 \lesssim \sum_{j=j_*}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} 2^j \cdot \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} 2^{\tilde{p}mj} \cdot n^{-\frac{\tilde{p}}{2}} \lesssim n^{-\frac{\tilde{p}}{2}} < n^{-\tilde{p}\delta}. \tag{4.15}$$

■ For Q_2. One defines

$$2^{j'} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2s+2m+1}}.$$

Clearly, $2^{j_*} \sim n^{\frac{1}{2t+2m+1}}$ $(t > s) \le 2^{j'} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2s+2m+1}} < 2^{j_1} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2m+1}}$. Furthermore, one rewrites

$$Q_2 = \left(\sum_{j=j_*}^{j'} + \sum_{j=j'+1}^{j_1}\right) 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}| \ge \frac{\kappa t_n}{2}\}}\right] := Q_{21} + Q_{22}. \tag{4.16}$$

■ For Q_{21}. By Lemma 4.3 and $2^{j'} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2s+2m+1}}$,

$$\begin{aligned} Q_{21} &:= \sum_{j=j_*}^{j'} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}| \ge \frac{\kappa t_n}{2}\}}\right] \\ &\le \sum_{j=j_*}^{j'} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}}\right] \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} \sum_{j=j_*}^{j'} 2^{j(2m+1)\frac{\tilde{p}}{2}} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} 2^{j'(2m+1)\frac{\tilde{p}}{2}} \sim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2s+2m+1}} \le \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \end{aligned} \tag{4.17}$$

■ For Q_{22}. Using Lemma 4.3, one has

$$\begin{aligned} Q_{22} &:= \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}| \ge \frac{\kappa t_n}{2}\}}\right] \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}+\tilde{p}mj} \sum_{k\in\Lambda_j} \mathbb{I}_{\{|\beta_{j,k}| \ge \frac{\kappa t_n}{2}\}}. \end{aligned}$$

When $p > \tilde{p} \ge 1$, by the Hölder inequality, $t_n = 2^{mj}\sqrt{\ln n/n}$, $2^{j'} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2s+2m+1}}$ and Lemma 1.2, one can obtain that

$$\begin{aligned} Q_{22} &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}+\tilde{p}mj} \sum_{k\in\Lambda_j} \left(\frac{|\beta_{j,k}|}{\frac{\kappa t_n}{2}}\right)^{\tilde{p}} \\ &\lesssim \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} = \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \left\|\beta_{j,k}\right\|_{\tilde{p}}^{\tilde{p}} \\ &\le \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \cdot 2^{j(1-\frac{\tilde{p}}{p})} \left\|\beta_{j,k}\right\|_{p}^{\tilde{p}} \end{aligned}$$

$$\lesssim \sum_{j=j'+1}^{j_1} 2^{-j\tilde{p}s} \lesssim 2^{-j'\tilde{p}s} \sim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2s+2m+1}} \leq \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \tag{4.18}$$

When $1 \leq p \leq \tilde{p}$, it follows from Lemma 1.2 that

$$\begin{aligned} Q_{22} &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}+\tilde{p}mj} \sum_{k\in\Lambda_j} \left(\frac{|\beta_{j,k}|}{\frac{\kappa t_n}{2}}\right)^p \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}+j(\tilde{p}-p)m} \left\|\beta_{j,k}\right\|_p^p \\ &\leq \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j_1} 2^{-j(sp+\frac{p}{2}-\frac{\tilde{p}}{2}-(\tilde{p}-p)m)}. \end{aligned} \tag{4.19}$$

Take

$$\epsilon := sp - \frac{\tilde{p}-p}{2}(2m+1).$$

Then, (4.19) can be rewritten as

$$Q_{22} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j_1} 2^{-j\epsilon}. \tag{4.20}$$

When $\epsilon > 0$ holds if and only if $p > \frac{\tilde{p}(2m+1)}{2s+2m+1}$, $\delta = \frac{s}{2s+2m+1}$ and

$$Q_{22} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} 2^{-j'\epsilon} \sim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2s+2m+1}} = \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \tag{4.21}$$

When $\epsilon \leq 0$ holds if and only if $p \leq \frac{\tilde{p}(2m+1)}{2s+2m+1}$, $\delta = \frac{s-1/p+1/\tilde{p}}{2(s-1/p)+2m+1}$. Define

$$2^{j''} \sim \left(\frac{n}{\ln n}\right)^{\frac{\delta}{s-1/p+1/\tilde{p}}} = \left(\frac{n}{\ln n}\right)^{\frac{1}{2(s-1/p)+2m+1}},$$

and obviously, $2^{j'} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2s+2m+1}} < 2^{j''} \sim \left(\frac{n}{\ln n}\right)^{\frac{\delta}{s-1/p+1/\tilde{p}}} < 2^{j_1} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2m+1}}$. Furthermore, one rewrites

$$\begin{aligned} Q_{22} &= \left(\sum_{j=j'+1}^{j''} + \sum_{j=j''+1}^{j_1}\right) 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} \mathrm{E}\left[|\hat{\beta}_{j,k}-\beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}|\geq\frac{\kappa t_n}{2}\}}\right] \\ &:= Q_{221} + Q_{222}. \end{aligned} \tag{4.22}$$

For Q_{221}. Note that $\frac{\tilde{p}-p}{2} + \frac{\delta\epsilon}{s-1/p+1/\tilde{p}} = \tilde{p}\delta$ in the case of $\epsilon \leq 0$. Then, by the same arguments of (4.20), one gets

$$Q_{221} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j''} 2^{-j\epsilon} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} 2^{-j''\epsilon} \sim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \tag{4.23}$$

For Q_{222}. The conditions $1 \le p \le \tilde{p}$ and $s > 1/p$ imply $B^s_{p,q}([0,1]) \subset B^{s-\frac{1}{p}+\frac{1}{\tilde{p}}}_{\tilde{p},q}([0,1])$. Similar to (4.18), one obtains

$$\begin{aligned} Q_{222} &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} \sum_{j=j''+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}+\tilde{p}mj} \sum_{k\in\Lambda_j} \left(\frac{|\beta_{j,k}|}{\frac{\kappa t_n}{2}}\right)^{\tilde{p}} \\ &\lesssim \sum_{j=j''+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \left\|\beta_{j,k}\right\|^{\tilde{p}}_{\tilde{p}} \lesssim \sum_{j=j''+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \cdot 2^{-j(s-\frac{1}{p}+\frac{1}{2})\tilde{p}} \\ &\lesssim 2^{-j''(s-\frac{1}{p}+\frac{1}{\tilde{p}})\tilde{p}} \sim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \end{aligned} \tag{4.24}$$

Combining (4.18), (4.21), (4.23) and (4.24),

$$Q_{22} \lesssim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

This with (4.16) and (4.17) shows that

$$Q_2 \lesssim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \tag{4.25}$$

■ For Q_3. According to the definition of $2^{j'}$, one can write

$$Q_3 = \left(\sum_{j=j_*}^{j'} + \sum_{j=j'+1}^{j_1}\right) 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}|\le 2\kappa t_n\}} := Q_{31} + Q_{32}.$$

■ For Q_{31}. It is easy to see that

$$\begin{aligned} Q_{31} &:= \sum_{j=j_*}^{j'} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}|\le 2\kappa t_n\}} \le \sum_{j=j_*}^{j'} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} (2\kappa t_n)^{\tilde{p}} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}}{2}} \cdot 2^{(2m+1)j'\frac{\tilde{p}}{2}} \sim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2s+2m+1}} \le \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \end{aligned}$$

■ For Q_{32}. One rewrites $Q_{32} = \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}|\le 2\kappa t_n\}}$. When $p > \tilde{p} \ge 1$, using the Hölder inequality and Lemma 1.2,

$$Q_{32} \le \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} \lesssim 2^{-j'\tilde{p}s} \sim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2s+2m+1}} \le \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

When $1 \le p \le \tilde{p}$, one has

$$\begin{aligned}
Q_{32} &\le \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} \left(\frac{2\kappa t_n}{|\beta_{j,k}|}\right)^{\tilde{p}-p} \\
&\lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}+j(\tilde{p}-p)m} \left\|\beta_{j,k}\right\|_p^p \\
&\le \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j_1} 2^{-j(sp+\frac{p}{2}-\frac{\tilde{p}}{2}-(\tilde{p}-p)m)} \\
&= \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j_1} 2^{-j\epsilon}.
\end{aligned}$$

For the case of $\epsilon > 0$, one can easily obtain that $\delta = \frac{s}{2s+2m+1}$ and

$$Q_{32} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} 2^{-j'\epsilon} \sim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}s}{2s+2m+1}} = \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

When $\epsilon \le 0$, $\delta = \frac{s-1/p+1/\tilde{p}}{2(s-1/p)+2m+1}$. Moreover, by the definition of $2^{j''}$, one rewrites

$$Q_{32} = \left(\sum_{j=j'+1}^{j''} + \sum_{j=j''+1}^{j_1}\right) 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} \mathbb{I}_{\{|\beta_{j,k}|\le 2\kappa t_n\}} := Q_{321} + Q_{322}.$$

Note that

$$Q_{321} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} \sum_{j=j'+1}^{j''} 2^{-j\epsilon} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{\tilde{p}-p}{2}} 2^{-j''\epsilon} \sim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

On the other hand, similar to the arguments of (4.24), one has

$$Q_{322} \le \sum_{j=j''+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \sum_{k\in\Lambda_j} |\beta_{j,k}|^{\tilde{p}} = \sum_{j=j''+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{\tilde{p}})\tilde{p}} \left\|\beta_{j,k}\right\|_{\tilde{p}}^{\tilde{p}} \lesssim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

Therefore, in all of the above cases,

$$Q_3 \lesssim \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}. \tag{4.26}$$

Finally, combining the above results (4.14), (4.15), (4.25) and (4.26), one gets

$$Q \lesssim (j_1 - j_* + 1)^{\tilde{p}-1} \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta} \lesssim (\ln n)^{\tilde{p}-1} \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

This with (4.11)–(4.13) shows

$$\mathrm{E}\left[\left\|\hat{r}_n^{non}(x) - r^{(m)}(x)\right\|_{\tilde{p}}^{\tilde{p}}\right] \lesssim (\ln n)^{\tilde{p}-1} \left(\frac{\ln n}{n}\right)^{\tilde{p}\delta}.$$

5. Conclusions

This paper considers wavelet estimations of the derivatives $r^{(m)}(x)$ of the variance function $r(x)$ in a heteroscedastic model. The upper bounds over $L^{\tilde{p}}(1 \leq \tilde{p} < \infty)$ risk of the wavelet estimators are discussed under some mild assumptions. The results show that the linear wavelet estimator can obtain the optimal convergence rate in the case of $p > \tilde{p} \geq 1$. When $p \leq \tilde{p}$, the nonlinear wavelet estimator has a better convergence rate than the linear estimator. Moreover, the nonlinear wavelet estimator is adaptive. Finally, some numerical experiments are presented to verify the good performances of the wavelet estimators.

Acknowledgments

We would like to thank the reviewers for their valuable comments and suggestions, which helped us to improve the quality of the manuscript. This paper is supported by the Guangxi Natural Science Foundation (No. 2022JJA110008), National Natural Science Foundation of China (No. 12001133), Center for Applied Mathematics of Guangxi (GUET), and Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation.

Conflict of interest

All authors declare that they have no conflicts of interest.

References

1. G. Box, Signal-to-noise ratios, performance criteria, and transformations, *Technometrics*, **30** (1988), 1–17.
2. R. J. Carroll, D. Ruppert, *Transformation and wighting in regression*, Boca Raton: CRC Press, 1988
3. W. Härdle, A. Tsybakov, Local polynomial estimators of the volatility function in nonparametric autoregression, *J. Econometrics*, **81** (1997), 223–242. https://doi.org/10.1016/S0304-4076(97)00044-4
4. J. Q. Fan, Q. W. Yao, Efficient estimation of conditional variance functions in stochastic regression, *Biometrika*, **85** (1998), 645–660. https://doi.org/10.1093/biomet/85.3.645
5. A. V. Quevedo, G. G. Vining, Online monitoring of nonlinear profiles using a Gaussian process model with heteroscedasticity, *Qual. Eng.*, **34** (2022), 58–74. https://doi.org/10.1080/08982112.2021.1998530
6. I. L. Amerise, Constrained quantile regression and heteroskedasticity, *J. Nonparamet. Stat.*, **34** (2022), 344–356. https://doi.org/10.1080/10485252.2022.2053536
7. L. Wang, L. D. Brown, T. T. Cai, M. Levine, Effect of mean on variance function estimation in nonparametric regression, *Ann. Statist.*, **36** (2008), 646–664. https://doi.org/10.1214/009053607000000901

8. R. Kulik, C. Wichelhaus, Nonparametric conditional variance and error density estimation in regression models with dependent errors and predictors, *Electron. J. Statist.*, **5** (2011), 856–898. https://doi.org/10.1214/11-EJS629
9. Y. D. Shen, C. Gao, D. Witten, F. Han, Optimal estimation of variance in nonparametric regression with random design, *Ann. Statist.*, **48** (2020), 3589–3618. https://doi.org/10.1214/20-AOS1944
10. D. L. Donoho, I. M. Johnstone, Minimax estimation via wavelet shrinkage, *Ann Statist.*, **26** (1998), 879–921. https://doi.org/10.1214/aos/1024691081
11. T. T. Cai, Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *Ann. Statist.*, **27** (1999), 898–924. https://doi.org/10.1214/aos/1018031262
12. G. P. Nason, R. Von Sachs, G. Kroisandt, Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, *J. Roy. Statist. Soc. B*, **62** (2000), 271–292. https://doi.org/10.1111/1467-9868.00231
13. T. T. Cai, H. H. Zhou, A data-driven block thresholding approach to wavelet estimation, *Ann. Statist.*, **37** (2009), 569–595. https://doi.org/10.1214/07-AOS538
14. P. Abry, G. Didier, Wavelet estimation for operator fractional Brownian motion, *Bernoulli*, **24** (2018), 895–928. https://doi.org/10.3150/15-BEJ790
15. L. Y. Li, B. Zhang, Nonlinear wavelet-based estimation to spectral density for stationary non-Gaussian linear processes, *Appl. Comput. Harmon. Anal.*, **60** (2022), 176–204. https://doi.org/10.1016/j.acha.2022.03.001
16. R. Kulik, M. Raimondo, Wavelet regression in random design with heteroscedastic dependent errors, *Ann. Statist.*, **37** (2009), 3396–3430. https://doi.org/10.1214/09-AOS684
17. Y. Zhou, A. T. K. Wan, S. Y. Xie, X. J. Wang, Wavelet analysis of change-points in a non-parametric regression with heteroscedastic variance, *J. Econometrics*, **159** (2010), 183–201. https://doi.org/10.1016/j.jeconom.2010.06.001
18. T. Palanisamy, J. Ravichandran, A wavelet-based hybrid approach to estimate variance function in heteroscedastic regression models, *Stat. Paper*, **56** (2015), 911—932. https://doi.org/10.1007/s00362-014-0614-6
19. L. Ding, P. Chen, Wavelet estimation in heteroscedastic regression models with $\alpha-$mixing random errorson, *Lith. Math. J.*, **61** (2021), 13–36. https://doi.org/10.1007/s10986-021-09508-x
20. H. J. Woltring, On optimal smoothing and derivative estimation from noisy displacement data in biomechanics, *Hum. Movement Sci.*, **4** (1985), 229–245. https://doi.org/10.1016/0167-9457(85)90004-1
21. S. G. Zhou, D. A. Wolfe, On derivative estimation in spline regression, *Stat. Sinica,* **10** (2000), 93–108. http://www.jstor.org/stable/24306706
22. J. E. Chacón, T. Duong, Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting, *Electron. J. Stat.*, **7** (2013), 499–532. https://doi.org/10.1214/13-EJS781
23. Y. Q. Wei, D. Y. Liu, D. Boutat, Innovative fractional derivative estimation of the pseudo-state for a class of fractional order linear systems, *Automatica*, **99** (2019), 157–166. https://doi.org/10.1016/j.automatica.2018.10.028
24. Y. Meyer, *Wavelets and operators*, London: Cambridge university press, 1992.

25. W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, *Wavelets, approximation, and statistical applications*, New York: Springer, 1998.
26. D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, *Ann. Statist.*, **24** (1996), 508–539. http://www.jstor.org/stable/2242660
27. G. P. Nason, Wavelet shrinkage using cross-validation, *J. Roy. Statist. Soc. B*, **58** (1996), 463–479. https://doi.org/10.1111/j.2517-6161.1996.tb02094.x
28. F. Navarro, A. Saumard, Slope heuristics and V-Fold model selection in heteroscedastic regression using strongly localized bases, *ESAIM Probab. Stat.*, **21** (2017), 412–451. https://doi.org/10.1051/ps/2017005

AIMS Press