

Optimal convergence rates of wavelet estimators for a hidden density in a mixture model

JUNKE KOU*  AND DAN LIANG 

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

Abstract. This paper investigates nonparametric estimations of a density function in a mixture model. Firstly, a lower bound estimation under L^p ($1 \leq p < +\infty$) error of an arbitrary density estimator is discussed. Secondly, a linear estimator and an adaptive nonlinear estimator of the unknown density function are constructed by the wavelet method. The rates of convergence of those two wavelet estimators are discussed with some mild conditions. Combining with the lower bound estimations, two wavelet estimators can attain the optimal convergence rate. Finally, numerical examples are given to verify the performance of the two wavelet estimators.

AMS subject classifications: 62G07, 62G20, 42C40

Keywords: nonparametric estimation; density function; mixture model; L^p risk

Received October 14, 2025; accepted March 5, 2026

1. Introduction

This paper considers the following mixture density model:

$$g(\mathbf{x}) = \theta h(\mathbf{x}) + (1 - \theta)f(\mathbf{x}), \mathbf{x} \in \Omega. \quad (1)$$

In the above equation, $g(\mathbf{x})$ is the corresponding density function of independent and identically distributed (i.i.d) random vectors X_1, X_2, \dots, X_n . Other two functions $h(\mathbf{x})$ and $f(\mathbf{x})$ all are bounded density functions, and the function $h(\mathbf{x})$ is known. Ω denotes a compact support subset of \mathbb{R}^d . The parameter θ is a known mixing ratio and $\theta \in (0, 1)$. The aim of this model is to estimate the unknown density $f(\mathbf{x})$ by the observed data X_1, X_2, \dots, X_n .

As a powerful statistical tool, the above mixture model plays a significant role in statistics, economics, big data processing, and other fields, see [27, 34, 20, 4, 25]. In the context of dealing with contamination problems [24], model (1) describes the case where the density function $f(\mathbf{x})$ of an unknown distribution is contaminated by an arbitrary distribution $h(\mathbf{x})$ with a proportion θ , which constitutes a particular instance of the pollution model [13]. In addition, Maiboroda and Sugakova [21], Chen et al. [7], and Liu and Gao [18] made different assumptions about the contamination model to study the problem of nonparametric density estimation.

The mixture model is also widely used in multiple testing problems, such as microarray analysis [3, 22], neuroimaging [30, 33], and other related fields. In the model, $h(\mathbf{x})$ and $f(\mathbf{x})$ represent the densities of the observed values under the null hypothesis and the alternative hypothesis, respectively, and θ is the asymptotic proportion of the true null hypotheses. Efron et al. [11] employed model (1) to estimate the local false discovery rate (FDR), which is defined as the posterior probability derived from the mixture model. Robin et al. [28] and Nguyen and Matias [26] further proposed stochastic weighted kernel estimators for $f(\mathbf{x})$, with weights based on the posterior probabilities.

For a nonparametric mixture density model, the kernel function method is commonly used by many researchers, such as Marzio and Taylo [23], Liu and Yu [19], Zhang et al [35], Chagny et al [6]. Nevertheless, when the function has sharp peaks and prominent parts, the corresponding kernel estimator is not particularly effective. Wavelet estimation is characterized by its multiresolution analysis and adaptivity. The wavelet transform refines various unknown functions and signals progressively across multiple scales through

Email addresses: kjkou@guet.edu.cn. (J. Kou), liangdan2012@163.com (D. Liang)

*Corresponding author.

dilation and translation. For more details, refer to [14, 1]. Therefore, it has achieved relatively significant accomplishments in the field of nonparametric estimation. Baldi et al. [2] proposed an adaptive density estimation method based on spherical wavelets (needlets), and proved that it has the theoretically optimal rate of convergence. Wang [32] provided the optimal rate of convergence for density estimation in Besov space with Fano's lemma. Chesneau et al. [8] proposed linear and nonlinear wavelet estimators for 2D continuous-discrete density functions, and studied integrated mean squared error of wavelet estimators in Besov balls. Chesneau et al. [9] investigated the nonparametric estimation of regression models with additive and multiplicative noise. Juditsky and Lambert-Lacroix [15] studied the optimal rate of convergence of the L^p ($1 \leq p < +\infty$) risk using biquadrature wavelet estimation of the density function in one-dimensional Hölder space. Shen et al. [29] studied the consistency and asymptotic normality of wavelet estimators of the regression function under the φ -mixed case. Cao and Zeng [5] proposed a data-driven wavelet estimation method for estimating the derivatives of the density function.

The main contributions of this paper are as follows. We consider nonparametric optimal estimations of a density function in a mixture model. Firstly, a lower bound estimation over L^p ($1 \leq p < +\infty$) risk of an arbitrary possible density estimator of the unknown density function $f(\mathbf{x})$ is proved. Secondly, a linear density estimator is constructed by using the wavelet method. A convergence rate of this linear estimator is discussed with $f(\mathbf{x}) \in B_{\eta,q}^s(\Omega)$ and some other mild conditions. Then, we can easily see that this linear wavelet can attain the optimal convergence rate when $p < \eta$. Furthermore, in order to overcome the shortage of this nonadaptive linear estimator, a nonlinear wavelet-based estimator is proposed by using the hard thresholding algorithm. According to the lower bound estimations and the corresponding theory results of this nonlinear wavelet-based estimator, this nonlinear estimator can get the optimal convergence rates up to an $\ln n$ factor in the cases of $p < \eta$ or $p \geq \eta$. Finally, numerical experiments indicate that both wavelet-based estimators can effectively estimate the unknown density function $f(\mathbf{x})$.

This paper is structured as follows. Wavelet theory and Besov space are introduced in Section 2. A lower bound estimation of any possible density estimator of the unknown density function in Besov spaces $B_{\eta,q}^s(\Omega)$ is proved in Section 3. The upper bound estimations over L^p ($1 \leq p < +\infty$) risk of two wavelet-based density estimators are discussed in Section 4. The numerical simulation studies are shown in Section 5.

2. Wavelet and Besov space

The aim of this paper is to study wavelet estimation of the Besov spaces density function based on a mixture model. We begin with the notion of multiresolution analysis (MRA) by assuming that $\{V_j\}_{j \in \mathbb{Z}}$ is a sequence of closed subspaces of the space $L^2(\mathbb{R}^d)$, and that for any $j \in \mathbb{Z}$, it holds that $V_j \subset V_{j+1}$ if the following conditions hold:

- (1) $\bigcap V_j = \{0\}$ and $\overline{\bigcup V_j} = L^2(\mathbb{R}^d)$;
- (2) $f(\mathbf{x}) \in V_0$ if and only if $f(2^j \mathbf{x}) \in V_j$;
- (3) There is a function $\Phi(\mathbf{x}) \in V_0$ such that $\{\Phi(\mathbf{x} - \mathbf{k}), \mathbf{k} \in \mathbb{Z}^d\}$ is an orthonormal basis in V_0 .

In the present study, we adopt the compactly supported scalar function $\Phi(\mathbf{x})$ from the family of Daubechies. For any $u \in \{0, 1, \dots, 2^d - 1\}$, we introduce the associated compactly supported wavelet function $\Psi_u(\mathbf{x})$. Furthermore, for a positive integer τ and $\Lambda_j = \{0, 1, \dots, 2^j - 1\}^d$, it is obvious that $S = \{\Phi_{\tau,k} := 2^{\tau d/2} \Phi(2^\tau \mathbf{x} - \mathbf{k}), \Psi_{j,k,u} := 2^{j d/2} \Psi_u(2^j \mathbf{x} - \mathbf{k}), j \geq \tau, \mathbf{k} \in \Lambda_j\}$ constitutes a standard orthogonal basis for $L^2(\Omega)$. For any positive integer $j_* \geq \tau$, the function $F(\mathbf{x}) \in L^2(\Omega)$ and can be expanded into the following wavelet level using wavelet basis S expanded into the following wavelet series:

$$F(\mathbf{x}) = \sum_{\mathbf{k} \in \Lambda_{j_*}} \alpha_{j_*,\mathbf{k}} \Phi_{j_*,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_*}^{\infty} \sum_{u=1}^{2^d-1} \sum_{\mathbf{k} \in \Lambda_j} \beta_{j,\mathbf{k},u} \Psi_{j,\mathbf{k},u}(\mathbf{x}), \mathbf{x} \in \Omega. \quad (2)$$

In this equation, $\alpha_{j_*,\mathbf{k}} = \int_{\Omega} F(\mathbf{x}) \Phi_{j_*,\mathbf{k}}(\mathbf{x}) d\mathbf{x}$ and $\beta_{j,\mathbf{k},u} = \int_{\Omega} F(\mathbf{x}) \Psi_{j,\mathbf{k},u}(\mathbf{x}) d\mathbf{x}$.

Let P_{V_j} denote the orthogonal projection operator. This operator maps functions from the space $L^2(\Omega)$ onto the subspace V_j . Then, for any $F(\mathbf{x}) \in L^2(\Omega)$,

$$P_{V_j} F(\mathbf{x}) = \sum_{\mathbf{k} \in \Lambda_j} \alpha_{j,\mathbf{k}} \Phi_{j,\mathbf{k}}(\mathbf{x}), \mathbf{x} \in \Omega.$$

Lemma 1. Assume that a scaling function $\Phi(\mathbf{x})$ is m -regular, meaning that $\Phi(\mathbf{x}) \in C^m$ and $|D^a \Phi(\mathbf{x})| \leq c(1 + |\mathbf{x}|^2)^{-l}$ for each $l \in \mathbb{N}$ and $a = 0, 1, \dots, m$. For any sequence $\{a_k\} \in l_p$ (i.e., $\|(a_k)\|_p := (\sum_k |a_k|^p)^{\frac{1}{p}} < \infty$), when $1 \leq p \leq \infty$, there exist constants $0 < c_1 \leq c_2$ satisfying:

$$c_1 2^{j(\frac{d}{2} - \frac{d}{p})} \|(a_k)\|_p \leq \left\| \sum_{\mathbf{k} \in \Lambda_j} a_k 2^{\frac{j d}{2}} \Phi(2^j \mathbf{x} - \mathbf{k}) \right\|_p \leq c_2 2^{j(\frac{d}{2} - \frac{d}{p})} \|(a_k)\|_p.$$

It is well known that Besov spaces are very important function spaces that can be characterized by wavelet bases. The equivalent definition of Besov spaces given in terms of wavelet coefficients is given below [14].

Lemma 2. Suppose the scale function $\Phi(\mathbf{x})$ is regular of order m , $0 < s < m$; let $F(\mathbf{x}) \in L^p(\Omega)$, $1 \leq p, q \leq \infty$. Then the following inferences are equivalent:

- (1) $F(\mathbf{x}) \in B_{p,q}^s(\Omega)$;
- (2) $\{2^{js} \|\mathbb{P}_{V_{j+1}} F(\mathbf{x}) - \mathbb{P}_{V_j} F(\mathbf{x})\|_p\} \in l_q$;
- (3) $\{2^{j(s - \frac{d}{p} + \frac{d}{2})} \|\beta_{j,k,u}\|_p\} \in l_q$.

The Besov norm of $F(\mathbf{x})$ is defined as

$$\|F(\mathbf{x})\|_{B_{p,q}^s} := \left\| (\alpha_{\tau,k}) \right\|_p + \left\| (2^{j(s - \frac{d}{p} + \frac{d}{2})} \|\beta_{j,k,u}\|_p)_{j \geq \tau} \right\|_q,$$

where $\|(a_{\tau,k})\|_p := (\sum_{\mathbf{k} \in \Lambda_j} |a_{\tau,k}|^p)^{\frac{1}{p}}$ and $\|\beta_{j,k,u}\|_p^p = \sum_{u=1}^{2^d-1} \sum_{\mathbf{k} \in \Lambda_j} |\beta_{j,k,u}|^p$.

3. Lower bound estimation

In this section, we will discuss the lower bound estimations of the density function in the mixture model. We firstly introduce the Kullback-Leibler distance [31], which plays a key role in the following discussions. Let $(\Omega, \mathcal{F}, P_k)$ be a measurable space, and U and W are two probability measures defined on it. If U is absolutely continuous with respect to W (denoted as $U \ll W$), the corresponding Kullback-Leibler distance is defined as

$$\mathbb{K}(U, W) := \int u(x) \ln \frac{u(x)}{w(x)} dx.$$

Here, u and w represent the probability density functions of U and W , respectively.

Based on the Kullback-Leibler distance, we can derive the following lemma. In the studies described below, $a \lesssim b$ means that $a \leq cb$ with a positive constant c , which is independent of a and b . $a \gtrsim b$ means that $b \lesssim a$. In addition, $a \succ b$ means $a \geq cb$, $a \sim b$ means that $a \lesssim b$ and $b \lesssim a$ hold.

Lemma 3. For the estimation model (1), g_1 and g_2 are two measurable functions defined on Ω . X_1, X_2, \dots, X_n is a set of i.i.d. random vectors. Then, we have

$$\mathbb{K}(P_{g_1}^n, P_{g_2}^n) \lesssim n \|f_1 - f_2\|_2.$$

Proof. According to model (1), $g_1(\mathbf{x}) = \theta h(\mathbf{x}) + (1 - \theta) f_1(\mathbf{x})$ and $g_2(\mathbf{x}) = \theta h(\mathbf{x}) + (1 - \theta) f_2(\mathbf{x})$. Due to the fact that X_1, X_2, \dots, X_n are independent and identically distributed random vectors, we can easily get the corresponding joint density function:

$$g_1^n(\mathbf{x}) = \prod_{i=1}^n g_1(\mathbf{x}_i), g_2^n(\mathbf{x}) = \prod_{i=1}^n g_2(\mathbf{x}_i).$$

Due to the definition of the Kullback-Leibler distance,

$$\begin{aligned}
\mathbb{K}(P_{g_1}^n, P_{g_2}^n) &= \int g_1^n(\mathbf{x}) \ln \frac{g_1^n(\mathbf{x})}{g_2^n(\mathbf{x})} d\mathbf{x} \\
&= \int_{\Omega} \prod_{j=1}^n g_1(\mathbf{x}_j) \ln \frac{\prod_{i=1}^n g_1(\mathbf{x}_i)}{\prod_{i=1}^n g_2(\mathbf{x}_i)} d\mathbf{x} \\
&= \sum_{i=1}^n \int_{\Omega} g_1(\mathbf{x}_i) \ln \frac{g_1(\mathbf{x}_i)}{g_2(\mathbf{x}_i)} d\mathbf{x}_i \prod_{j \neq i} \int_{\Omega} g_1(\mathbf{x}_j) d\mathbf{x}_j \\
&= \sum_{i=1}^n \int_{\Omega} g_1(\mathbf{x}_i) \ln \frac{g_1(\mathbf{x}_i)}{g_2(\mathbf{x}_i)} d\mathbf{x}_i \\
&= n \int_{\Omega} g_1(\mathbf{x}_1) \ln \frac{g_1(\mathbf{x}_1)}{g_2(\mathbf{x}_1)} d\mathbf{x}_1 \\
&= n \mathbb{K}(P_{g_1}, P_{g_2}).
\end{aligned} \tag{3}$$

Next, we derive the upper bound of $\mathbb{K}(P_{g_1}, P_{g_2})$. Note that $\ln(Q+1) \leq |Q|$ holds for all $Q > -1$. In addition,

$$\frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} = \frac{\theta h(\mathbf{x}) + (1-\theta)f_1(\mathbf{x})}{\theta h(\mathbf{x}) + (1-\theta)f_2(\mathbf{x})} = 1 + \frac{(1-\theta)[f_1(\mathbf{x}) - f_2(\mathbf{x})]}{\theta h(\mathbf{x}) + (1-\theta)f_2(\mathbf{x})} := 1 + Q(\mathbf{x}).$$

Hence,

$$\begin{aligned}
\mathbb{K}(P_{g_1}, P_{g_2}) &= \int_{\Omega} g_1(\mathbf{x}) \ln \frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} d\mathbf{x} \\
&\leq \int_{\Omega} g_1(\mathbf{x}) |Q(\mathbf{x})| d\mathbf{x} \\
&= (1-\theta) \int_{\Omega} g_1(\mathbf{x}) \frac{|f_1(\mathbf{x}) - f_2(\mathbf{x})|}{g_2(\mathbf{x})} d\mathbf{x}.
\end{aligned}$$

Due to $g_1(\mathbf{x}) = \theta h(\mathbf{x}) + (1-\theta)f_1(\mathbf{x})$, we can get

$$\begin{aligned}
\mathbb{K}(P_{g_1}, P_{g_2}) &\leq \theta(1-\theta) \int_{\Omega} \frac{h(\mathbf{x})|f_1(\mathbf{x}) - f_2(\mathbf{x})|}{g_2(\mathbf{x})} d\mathbf{x} + (1-\theta)^2 \int_{\Omega} \frac{f_1(\mathbf{x})|f_1(\mathbf{x}) - f_2(\mathbf{x})|}{g_2(\mathbf{x})} d\mathbf{x} \\
&:= A_1 + A_2.
\end{aligned} \tag{4}$$

For A_1 , the regularity of the density function shows that $g_2(\mathbf{x}) \geq \theta h(\mathbf{x})$, and $\frac{h(\mathbf{x})}{g_2(\mathbf{x})} \leq \frac{1}{\theta}$. Therefore,

$$A_1 \leq (1-\theta) \int_{\Omega} |f_1(\mathbf{x}) - f_2(\mathbf{x})| d\mathbf{x}. \tag{5}$$

For A_2 , it follows from the properties of the density function that $g_2(\mathbf{x}) > (1-\theta)f_2(\mathbf{x})$. The boundness property of $f(\mathbf{x})$ implies that $\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \leq C_1$. Furthermore, $\frac{f_1(\mathbf{x})}{g_2(\mathbf{x})} \leq \frac{f_1(\mathbf{x})}{(1-\theta)f_2(\mathbf{x})} \leq \frac{C_1}{1-\theta}$ and

$$A_2 \leq C_1(1-\theta) \int_{\Omega} |f_1(\mathbf{x}) - f_2(\mathbf{x})| d\mathbf{x}. \tag{6}$$

Combining (4), (5), and (6), we know that

$$\mathbb{K}(P_{g_1}, P_{g_2}) \leq (1+C_1)(1-\theta) \int_{\Omega} |f_1(\mathbf{x}) - f_2(\mathbf{x})| d\mathbf{x}. \tag{7}$$

By using the Cauchy-Schwarz inequality, we can get

$$\int_{\Omega} |f_1(\mathbf{x}) - f_2(\mathbf{x})| d\mathbf{x} \leq \left(\int_{\Omega} |f_1(\mathbf{x}) - f_2(\mathbf{x})|^2 d\mathbf{x} \right)^{\frac{1}{2}} \left(\int_{\Omega} 1 d\mathbf{x} \right)^{\frac{1}{2}} \lesssim \|f_1 - f_2\|_2.$$

This, together with (7) and (3), shows that

$$\mathbb{K} \left(P_{g_1}^n, P_{g_2}^n \right) \lesssim n \|f_1 - f_2\|_2.$$

□

Fano's lemma [31] *Let $(\Omega, \mathcal{F}, P_k)$ be a probability measure space and $G_k \in \mathcal{F} (k = 0, 1, \dots, D)$ a sequence of measurable sets on it. If $G_k \cap G_l = \emptyset$ holds with $k \neq l$, then*

$$\sup_{0 \leq k \leq D} P_k^n(G_k^c) \geq \min \left\{ \sqrt{D} \exp(-K_D), \frac{1}{2} \right\},$$

where G_k^c denotes the complement set of G_k and $K_D := \inf_{0 \leq l \leq D} \frac{1}{D} \sum_{k \neq l} \mathbb{K}(P_{g_k}, P_{g_l})$.

In this position, the lower bound estimation of any possible density estimator will be stated. Throughout this paper, we assume the density function $f(\mathbf{x}) \in B_{\eta, q}^s(\Omega, T)$ with $B_{\eta, q}^s(\Omega, T) := \{f \in B_{\eta, q}^s(\Omega), \|f\|_{B_{\eta, q}^s} \leq T\}$.

Theorem 1. *For the estimation model (1), $f(\mathbf{x}) \in B_{\eta, q}^s(\Omega, T)$ with $0 \leq \eta, q < +\infty$, $p \in [1, +\infty)$ and $s > \frac{d}{\eta}$. Suppose the region satisfies $\Omega = [0, b]^d$. If $\hat{f}_n(\mathbf{x})$ is any possible density estimator of the density function $f(\mathbf{x})$, we have*

$$\sup_{f(\mathbf{x}) \in B_{\eta, q}^s(\Omega, T)} \mathbb{E} \left[\|\hat{f}_n(\mathbf{x}) - f(\mathbf{x})\|_p \right] \gtrsim \max \left\{ n^{-\frac{s}{2s+d}}, \left(\frac{\ln n}{n} \right)^{\frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d}} \right\}.$$

Remark 1. According this theorem, note that

$$\max \left\{ n^{-\frac{s}{2s+d}}, \left(\frac{\ln n}{n} \right)^{\frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d}} \right\} = \begin{cases} n^{-\frac{s}{2s+d}}, & \eta > \frac{pd}{2s+d}, \\ \left(\frac{\ln n}{n} \right)^{\frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d}}, & \eta \leq \frac{pd}{2s+d}. \end{cases}$$

This conclusion means that the lower bound estimations have different results under distinct cases.

Proof. In order to prove the above theorem, we divide a theoretical result into two parts as follows:

$$\sup_{f(\mathbf{x}) \in B_{\eta, q}^s(\Omega, T)} \mathbb{E} \left[\|\hat{f}_n(\mathbf{x}) - f(\mathbf{x})\|_p \right] \gtrsim n^{-\frac{s}{2s+d}}$$

and

$$\sup_{f(\mathbf{x}) \in B_{\eta, q}^s(\Omega, T)} \mathbb{E} \left[\|\hat{f}_n(\mathbf{x}) - f(\mathbf{x})\|_p \right] \gtrsim \left(\frac{\ln n}{n} \right)^{\frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d}}.$$

The proof of the first part. In order to prove the first part, we need to construct $f_{\lambda^i}(\mathbf{x}) (i = 0, 1, \dots, D)$ such that $f_{\lambda^i}(\mathbf{x}) \in B_{\eta, q}^s(\Omega, T)$ and

$$\sup_i \mathbb{E} \left[\|\hat{f}_n(\mathbf{x}) - f_{\lambda^i}(\mathbf{x})\|_p \right] \gtrsim n^{-\frac{s}{2s+d}}, \quad (8)$$

with $\hat{f}_n(\mathbf{x})$ being any possible estimator of the density function $f(\mathbf{x})$.

Let the orthogonal scale function Φ have compact support and satisfy the m -order ($m > s$) regularity condition; the corresponding wavelet function is denoted by Ψ_u and $\text{supp} \Psi_u \subseteq [0, b]^d = \Omega$. Assuming that there exists a compactly supported density function $y(\mathbf{x}) \in B_{\eta, q}^s(\Omega, T)$, $\min_{\mathbf{x} \in \Omega} y(\mathbf{x}) = c_0 > 0$ and $\int_{\Omega} y(\mathbf{x}) d\mathbf{x} = 1$.

Now, we define $\delta_j := 2^{-j(s+\frac{d}{2})}$ and $f_\lambda(\mathbf{x}) := y(\mathbf{x}) + \delta_j \sum_{k \in \Lambda_j} \lambda_k \Psi_{j,k,u}(\mathbf{x})$. In those definitions, $\lambda = (\lambda_k)_{k \in \Lambda_j} \in \{0, 1\}^{2^{jd}}$ and $\Lambda_j := \{0, b, 2b, \dots, (2^j - 1)b\}^d$. Then we can easily get

$$\begin{aligned} \int_{\Omega} f_\lambda(\mathbf{x}) d\mathbf{x} &= \int_{\Omega} \left[y(\mathbf{x}) + \delta_j \sum_{k \in \Lambda_j} \lambda_k \Psi_{j,k,u}(\mathbf{x}) \right] d\mathbf{x} \\ &= \int_{\Omega} y(\mathbf{x}) d\mathbf{x} + \delta_j \sum_{k \in \Lambda_j} \lambda_k \int_{\Omega} \Psi_{j,k,u}(\mathbf{x}) d\mathbf{x} \\ &= 1 + \delta_j \sum_{k \in \Lambda_j} \lambda_k \cdot 0 = 1. \end{aligned}$$

Moreover, we can choose sufficiently large j such that

$$f_\lambda(\mathbf{x}) \geq c_0 - \delta_j 2^{-js} \|\Psi_u\|_{\infty} \geq 0.$$

Hence, by the above discussions, the functions $f_\lambda(\mathbf{x})$ are density functions.

On the other hand, due to $\lambda_k \in \{0, 1\}$ and Lemma 2, we can obtain

$$\sum_{k \in \Lambda_j} |\lambda_k|^\eta \leq 2^{jd}, 2^{j(s+\frac{d}{2}-\frac{d}{\eta})} \delta_j \left(\sum_{k \in \Lambda_j} |\lambda_k|^\eta \right)^{\frac{1}{\eta}} \leq 1.$$

Hence, $\left\| \delta_j \sum_{k \in \Lambda_j} \lambda_k \Psi_{j,k,u} \right\|_{B_{\eta,q}^s} \lesssim 1$ and $f_\lambda(\mathbf{x}) \in B_{\eta,q}^s(\Omega, T)$. Then, a series of density functions $f_\lambda(\mathbf{x})$ has been constructed, which belong to Besov spaces $B_{\eta,q}^s(\Omega, T)$. Next, we prove that those density functions $f_\lambda(\mathbf{x})$ satisfy the result (8) with any possible density estimators. The Varshamov-Gilbert Lemma will be used in later discussions. Now, we introduce it as follows.

Varshamov-Gilbert Lemma [10, 31] For the set $(\lambda_k)_{k \in \Lambda_j} \in \{0, 1\}^{2^{jd}}$, there exist a subset $\{\lambda^0, \lambda^1, \dots, \lambda^D\}$ and an element $\lambda^0 = (0, 0, \dots, 0)$ such that $D \geq e^{2^{jd-3}}$ and

$$\sum_{k \in \Lambda_j} |\lambda_k^\omega - \lambda_k^\mu| \geq 2^{jd-3} (0 \leq \omega \neq \mu \leq D).$$

According to the Varshamov-Gilbert Lemma, when $0 \leq \omega \neq \mu \leq D$, we have

$$\begin{aligned} \|f_{\lambda^\omega}(\mathbf{x}) - f_{\lambda^\mu}(\mathbf{x})\|_p^p &= \left\| \delta_j \sum_{k \in \Lambda_j} (\lambda_k^\omega - \lambda_k^\mu) \Psi_{j,k,u}(\mathbf{x}) \right\|_p^p \\ &= \int_{\Omega} \left| \delta_j \sum_{k \in \Lambda_j} (\lambda_k^\omega - \lambda_k^\mu) \Psi_{j,k,u}(\mathbf{x}) \right|^p d\mathbf{x} \\ &= \delta_j^p \sum_{k \in \Lambda_j} |\lambda_k^\omega - \lambda_k^\mu|^p \int_{\Omega} |\Psi_{j,k,u}(\mathbf{x})|^p d\mathbf{x} \\ &= \delta_j^p \sum_{k \in \Lambda_j} |\lambda_k^\omega - \lambda_k^\mu|^p \|\Psi_{j,k,u}\|_p^p \\ &= 2^{-j(sp+d)} \|\Psi_u\|_p^p \sum_{k \in \Lambda_j} |\lambda_k^\omega - \lambda_k^\mu|^p \geq \frac{1}{8} 2^{-jsp} \|\Psi_u\|_p^p. \end{aligned} \tag{9}$$

We rewrite the above result as follows:

$$\|f_{\lambda^\omega}(\mathbf{x}) - f_{\lambda^\mu}(\mathbf{x})\|_p \geq 8^{-\frac{1}{p}} 2^{-js} \|\Psi_u\|_p := \sigma_j.$$

For an arbitrary density estimator $\hat{f}_n(\mathbf{x})$, we define $G_{\lambda^\mu} := \{\|\hat{f}_n(\mathbf{x}) - f_{\lambda^\mu}(\mathbf{x})\|_p < \frac{\sigma_j}{2}\} (\mu = 1, 2, \dots, D)$. Then, $G_{\lambda^\omega} \cap G_{\lambda^\mu} = \emptyset$ for any $\omega \neq \mu$. Furthermore, according to Fano's lemma, we can obtain

$$\sup_{0 \leq \mu \leq D} P_{g,\lambda^\mu}^n(G_{\lambda^\mu}^c) \geq \min \left\{ \sqrt{D} \exp(-K_D), \frac{1}{2} \right\}, \tag{10}$$

where their common distribution is characterized by probability measure P_g^n with corresponding joint density function $g^n(\mathbf{x}) := \prod_{i=1}^n g(\mathbf{x}_i)$. Using the arguments of (9),

$$\|f_{\lambda^\omega}(\mathbf{x}) - f_{\lambda^\mu}(\mathbf{x})\|_2^2 = \delta_j^2 \sum_{k \in \Lambda_j} |\lambda_k^\omega - \lambda_k^\mu|^2 \|\Psi_{j,k,u}\|_2^2 \leq \delta_j^2 2^{jd}.$$

Furthermore, by the definition of K_D and Lemma 3,

$$K_D = \inf_{0 \leq \mu \leq D} \frac{1}{D} \sum_{\omega \neq \mu} \mathbb{K} \left(P_{g_{\lambda^\omega}}^n, P_{g_{\lambda^\mu}}^n \right) \leq n \delta_j^2 2^{jd}.$$

Taking $2^j \sim n^{\frac{1}{2s+d}}$, we can easily get $n \delta_j^2 \sim n \cdot n^{-\frac{2(s+\frac{d}{2})}{2s+d}} \sim 1$. By the Varshamov-Gilbert Lemma, there exists a sufficiently small positive constant c_1 such that

$$\sqrt{D} e^{-K_D} \geq e^{2^{jd-3}} e^{-n \delta_j^2 2^{jd}} \geq \sqrt{D} e^{-K_D} \geq e^{2^{jd-3}} e^{-c_1 2^{jd}} \geq 1.$$

According to (10), we have $\sup_{0 \leq \mu \leq D} P_{g_{\lambda^\mu}}^n(G_{\lambda^\mu}^c) \gtrsim 1$. Hence, by using Markov's inequality,

$$\begin{aligned} \sup_{0 \leq \mu \leq D} \mathbb{E} \left[\|\hat{f}_n(\mathbf{x}) - f_{\lambda^\mu}(\mathbf{x})\|_p \right] &\geq \sup_{0 \leq \mu \leq D} \frac{\sigma_j}{2} P_{g_{\lambda^\mu}}^n \left(\|\hat{f}_n - f_{\lambda^\mu}\|_p \geq \frac{\sigma_j}{2} \right) \\ &\gtrsim \sigma_j \sim n^{-\frac{s}{2s+d}}, \end{aligned}$$

which is the desired conclusion.

The proof of the second part. For the second part, it is necessary to construct a series of density functions $f_k(\mathbf{x})$ such that $f_k(\mathbf{x}) \in B_{\eta,q}^s(\Omega, T)$ and

$$\sup_k \mathbb{E} \left[\|\hat{f}_n(\mathbf{x}) - f_k(\mathbf{x})\|_p \right] \gtrsim \left(\frac{\ln n}{n} \right)^{\frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d}},$$

with $\hat{f}_n(\mathbf{x})$ being any possible estimator of $f(\mathbf{x})$.

Now, we define $\delta_j := 2^{-j(s - \frac{d}{\eta} + \frac{d}{2})}$ and $f_k(\mathbf{x}) := y(\mathbf{x}) + \delta_j \Psi_{j,k,u}(\mathbf{x})$ with $k \in \Lambda_j$. Then, we need to prove that functions $f_k(\mathbf{x})$ are density functions and $f_k(\mathbf{x}) \in B_{\eta,q}^s(\Omega, T)$ in the discussions that follow.

Note that

$$\int_{\Omega} f_k(\mathbf{x}) d\mathbf{x} = \int_{\Omega} \left[y(\mathbf{x}) + \delta_j \Psi_{j,k,u}(\mathbf{x}) \right] d\mathbf{x} = 1.$$

Due to $\min_{\mathbf{x} \in \Omega} y(\mathbf{x}) = c_0 > 0$ and $\|\delta_j \Psi_{j,k,u}(\mathbf{x})\|_{\infty} = 2^{-j(s - \frac{d}{\eta})} \|\Psi_u\|_{\infty}$, we can choose a large parameter j such that

$$f_k(\mathbf{x}) \geq c_0 - \delta_j 2^{-j(s - \frac{d}{\eta})} \|\Psi_u\|_{\infty} \geq 0.$$

Obviously, $\delta_j 2^{j(s + \frac{d}{2} - \frac{d}{\eta})} = 1$ and $\|\delta_j \Psi_{j,k,u}\|_{B_{\eta,q}^s} \sim \delta_j 2^{j(s + \frac{d}{2} - \frac{d}{\eta})} \|\Psi_{j,k,u}\|_{\eta} \sim 1$. This, together with the condition $y(\mathbf{x}) \in B_{\eta,q}^s(\Omega, T)$, shows that $f_k(\mathbf{x}) \in B_{\eta,q}^s(\Omega, T)$.

For any $k, k' \in \Lambda_j$, and $k \neq k'$, we can obtain $\text{supp } \Psi_{j,k,u} \cap \text{supp } \Psi_{j,k',u} = \emptyset$ and $\text{supp } f_k \cap \text{supp } f_{k'} = \emptyset$. Furthermore,

$$\|f_k(\mathbf{x}) - f_{k'}(\mathbf{x})\|_p = \|\delta_j \Psi_{j,k,u} - \delta_j \Psi_{j,k',u}\|_p = 2^{-j(s + \frac{d}{p} - \frac{d}{\eta})} \|\Psi_u\|_p := \sigma_j.$$

In addition, take $G_k := \{\|\hat{f}_n(\mathbf{x}) - f_k(\mathbf{x})\|_p < \frac{\sigma_j}{2}\}$. Then, $G_k \cap G_{k'} = \emptyset$ holds when $k \neq k'$. Using Fano's lemma, one has

$$\sup_{k \in \Lambda_j} P_{g_k}^n(G_k^c) \geq \min \left\{ \sqrt{2^{jd}} \exp(-K_{2^{jd}}), \frac{1}{2} \right\}, \quad (11)$$

where their common distribution is characterized by probability measure $P_{g_k}^n$ with corresponding joint density function $g^n(\mathbf{x}) := \prod_{i=1}^n g(\mathbf{x}_i)$.

By Markov's inequality, we get

$$\mathbb{E}[\|\hat{f}_n(\mathbf{x}) - f_k(\mathbf{x})\|_p] \geq \frac{\sigma_j}{2} P_{g_k}^n \left(\|\hat{f}_n - f_k\|_p \geq \frac{\sigma_j}{2} \right) = \frac{\sigma_j}{2} P_{g_k}^n(G_k^c). \quad (12)$$

Combining (11) and (12),

$$\sup_{k \in \Lambda_j} \mathbb{E}[\|\hat{f}_n(\mathbf{x}) - f_k(\mathbf{x})\|_p] \geq \frac{\sigma_j}{2} \min \left\{ \sqrt{2^{jd}} \exp(-K_{2^{jd}}), \frac{1}{2} \right\}. \quad (13)$$

Using Lemma 3 and the definition of $K_{2^{jd}}$,

$$K_{2^{jd}} = \inf_{k' \in \Lambda_j} 2^{-jd} \sum_{k \neq k'} \mathbb{K}(P_{g_k}^n, P_{g_{k'}}^n) \leq n\delta_j^2.$$

Taking $2^j \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2(s-\frac{d}{\eta}+\frac{d}{2})}}$, we have $n\delta_j^2 = n \cdot 2^{-2j(s-\frac{d}{\eta}+\frac{d}{2})} \sim n \left(\frac{n}{\ln n}\right)^{-\frac{2(s-\frac{d}{\eta}+\frac{d}{2})}{2(s-\frac{d}{\eta}+\frac{d}{2})+d}} = \ln n$. In addition, there exists a small enough positive constant c_4 such that $n\delta_j^2 \leq c_4 \ln n$ and $\left[4(s-\frac{d}{\eta})+2d\right]^{-1} > c_4$. Hence,

$$\sqrt{2^{jd}} e^{-K_{2^{jd}}} \geq \sqrt{2^{jd}} e^{-c_4 \ln n} \succ \left(\frac{n}{\ln n}\right)^{\left[4(s-\frac{d}{\eta})+2d\right]^{-1}} n^{-c_4} \geq 1.$$

This, together with (13), shows that

$$\sup_{k \in \Lambda_j} \mathbb{E}[\|\hat{f}_n(\mathbf{x}) - f_k(\mathbf{x})\|_p] \gtrsim \sigma_j \gtrsim 2^{-j(s-\frac{d}{\eta}+\frac{d}{p})} \sim \left(\frac{\ln n}{n}\right)^{\frac{s-\frac{d}{\eta}+\frac{d}{p}}{2(s-\frac{d}{\eta}+\frac{d}{2})+d}},$$

which is the desired conclusion. \square

4. Upper bound estimations

In this section, we will construct two density estimators by using the wavelet method. The upper bounds over L^p ($1 \leq p < +\infty$) risk of those two wavelet estimators are considered with some mild conditions.

4.1. Linear estimator

In this section, we develop a linear wavelet estimator using wavelet theory, and discuss its rate of convergence under L^p risk ($1 \leq p < \infty$) in Besov spaces.

For the estimation model (1), a linear wavelet estimator is constructed as follows:

$$\hat{f}_n^{\text{lin}}(\mathbf{x}) := \sum_{k \in \Lambda_{j_n}} \hat{\alpha}_{j_n, k} \Phi_{j_n, k}(\mathbf{x}), \mathbf{x} \in \Omega.$$

$\hat{\alpha}_{j_n, k}$ is defined as follows, but it can be obtained from Lemma 4 as an unbiased estimator:

$$\hat{\alpha}_{j_n, k} := \frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j_n, k}(X_i) - \mu_{j_n, k}, \quad (14)$$

with $\mu_{j_n, k} = \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Phi_{j_n, k}(\mathbf{x}) d\mathbf{x}$. Now, in order to prove important theorems for the linear estimator, we firstly give some lemmas.

Lemma 4. *Assuming that $\hat{\alpha}_{j_n, k}$ is defined as (14), then*

$$\mathbb{E}[\hat{\alpha}_{j_n, k}] = \alpha_{j_n, k}.$$

Proof. According to $\hat{\alpha}_{j^*,k} = \frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i) - \mu_{j^*,k}$ and $\mu_{j^*,k} = \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x}$, we can easily get

$$\begin{aligned} \mathbb{E}[\hat{\alpha}_{j^*,k}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i) - \mu_{j^*,k}\right] \\ &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i) - \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x}\right] \\ &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i)\right] - \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (15)$$

Since the random variables X_1, X_2, \dots, X_n are i.i.d. and due to its corresponding common density function $g(\mathbf{x})$,

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i)\right] = \mathbb{E}\left[\frac{1}{1-\theta} \Phi_{j^*,k}(X_1)\right] = \int_{\Omega} \frac{1}{1-\theta} g(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x}. \quad (16)$$

By (15), (16), and the model in equation (1), which has $g(\mathbf{x}) = \theta h(\mathbf{x}) + (1-\theta)f(\mathbf{x})$, we know that

$$\begin{aligned} \mathbb{E}[\hat{\alpha}_{j^*,k}] &= \int_{\Omega} \frac{1}{1-\theta} g(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x} - \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\Omega} \frac{1}{1-\theta} [\theta h(\mathbf{x}) + (1-\theta)f(\mathbf{x})] \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x} - \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\Omega} f(\mathbf{x}) \Phi_{j^*,k}(\mathbf{x}) d\mathbf{x} = \alpha_{j^*,k}. \end{aligned}$$

The proof of Lemma 4 is finished. \square

Rosenthal's inequality [12] Consider the independent random variables X_1, \dots, X_n with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[|X_i|^p] < \infty$,

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

Lemma 5. Let $2^{j^*d} \leq n$, $0 < \theta \leq c_3 < 1$, and $\hat{\alpha}_{j^*,k}$ be defined as in (14). Then, for any $1 \leq p < \infty$, we can derive

$$\mathbb{E}[|\hat{\alpha}_{j^*,k} - \alpha_{j^*,k}|^p] \lesssim n^{-\frac{p}{2}}.$$

Proof. By the definition of $\hat{\alpha}_{j^*,k}$ and Lemma 4, we obtain

$$\begin{aligned} |\hat{\alpha}_{j^*,k} - \alpha_{j^*,k}| &= \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i) - \mu_{j^*,k} - \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i) - \mu_{j^*,k}\right] \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i) - \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Phi_{j^*,k}(X_i)\right] \right| \\ &= \frac{1}{n} \left| \sum_{i=1}^n \frac{1}{1-\theta} (\Phi_{j^*,k}(X_i) - \mathbb{E}[\Phi_{j^*,k}(X_i)]) \right| \\ &= \frac{1}{n} \left| \sum_{i=1}^n D_i \right|, \end{aligned}$$

where $D_i := \frac{1}{1-\theta} (\Phi_{j_s, k}(X_i) - \mathbb{E}[\Phi_{j_s, k}(X_i)])$. Obviously, $\mathbb{E}[D_i] = 0$ and

$$\mathbb{E}[|\hat{\alpha}_{j_s, k} - \alpha_{j_s, k}|^p] = \mathbb{E}\left[\left|\frac{1}{n} \sum_{i=1}^n D_i\right|^p\right] = \frac{1}{n^p} \mathbb{E}\left[\left|\sum_{i=1}^n D_i\right|^p\right].$$

Using the assumption that $0 < \theta \leq c_3 < 1$, we can easily get

$$\begin{aligned} \mathbb{E}[|D_i|^p] &= \mathbb{E}\left[\left|\frac{1}{1-\theta} (\Phi_{j_s, k}(X_i) - \mathbb{E}[\Phi_{j_s, k}(X_i)])\right|^p\right] \\ &= \frac{1}{(1-\theta)^p} \mathbb{E}[|\Phi_{j_s, k}(X_i) - \mathbb{E}[\Phi_{j_s, k}(X_i)]|^p] \\ &\lesssim \mathbb{E}[|\Phi_{j_s, k}(X_i) - \mathbb{E}[\Phi_{j_s, k}(X_i)]|^p] \\ &\lesssim \mathbb{E}[|\Phi_{j_s, k}(X_i)|^p]. \end{aligned} \tag{17}$$

Due to $g(\mathbf{x}) = \theta h(\mathbf{x}) + (1-\theta)f(\mathbf{x})$, and the density functions $h(\mathbf{x})$ and $f(\mathbf{x})$ bounded in the defining interval Ω , it is easy to see that

$$\begin{aligned} \mathbb{E}[|\Phi_{j_s, k}(X_i)|^p] &= \int_{\Omega} g(\mathbf{x}) |\Phi_{j_s, k}(\mathbf{x})|^p d\mathbf{x} \\ &= \int_{\Omega} \theta h(\mathbf{x}) |\Phi_{j_s, k}(\mathbf{x})|^p d\mathbf{x} + \int_{\Omega} (1-\theta)f(\mathbf{x}) |\Phi_{j_s, k}(\mathbf{x})|^p d\mathbf{x} \\ &\lesssim \int_{\Omega} |\Phi_{j_s, k}(\mathbf{x})|^p d\mathbf{x}. \end{aligned}$$

This, together with Lemma 1 and (17), shows that

$$\mathbb{E}[|D_i|^p] \lesssim \int_{\Omega} |\Phi_{j_s, k}(\mathbf{x})|^p d\mathbf{x} \lesssim 2^{j_s(\frac{d}{2} - \frac{d}{p})p}.$$

In particular, when $p = 2$, $\mathbb{E}[|D_i|^2] \lesssim 1$.

By Rosenthal's inequality and $2^{j_s d} \leq n$, when $p > 2$, we have

$$\begin{aligned} \mathbb{E}[|\hat{\alpha}_{j_s, k} - \alpha_{j_s, k}|^p] &= \frac{1}{n^p} \mathbb{E}\left[\left|\sum_{i=1}^n D_i\right|^p\right] \\ &\lesssim \frac{1}{n^p} \left\{ \sum_{i=1}^n \mathbb{E}[|D_i|^p] + \left(\sum_{i=1}^n \mathbb{E}[D_i^2]\right)^{\frac{p}{2}} \right\} \\ &\lesssim n^{-p} \left(n 2^{j_s(\frac{d}{2} - \frac{d}{p})p} + n^{\frac{p}{2}} \right) \\ &\lesssim n^{-\frac{p}{2}}. \end{aligned}$$

When $1 \leq p \leq 2$, we get

$$\begin{aligned} \mathbb{E}[|\hat{\alpha}_{j_s, k} - \alpha_{j_s, k}|^p] &= \frac{1}{n^p} \mathbb{E}\left[\left|\sum_{i=1}^n D_i\right|^p\right] \\ &\lesssim \frac{1}{n^p} \left(\sum_{i=1}^n \mathbb{E}[D_i^2]\right)^{\frac{p}{2}} \\ &\lesssim n^{-\frac{p}{2}}. \end{aligned}$$

In conclusion, when $1 \leq p < \infty$, the following result holds:

$$\mathbb{E}[|\hat{\alpha}_{j_s, k} - \alpha_{j_s, k}|^p] \lesssim n^{-\frac{p}{2}}.$$

□

In this position, the convergence rate of the linear wavelet estimator is stated in the following theorem. In the sequel, the following symbol is introduced by $x_+ = \max\{x, 0\}$.

Theorem 2. For the model (1), $f(\mathbf{x}) \in B_{\eta,q}^s(\Omega)$, with $\eta, q \in [1, \infty)$, $s > 0$, and $1 \leq p < \infty$. If $\{1 \leq p < \eta\}$ or $\{1 \leq \eta \leq p, s > \frac{d}{\eta}\}$, the linear estimator $\hat{f}_n^{\text{lin}}(\mathbf{x})$ with $2^{j_*} \sim n^{\frac{1}{2s'+d}}$ and $s' = s - d(\frac{1}{\eta} - \frac{1}{p})_+$ satisfies

$$\mathbb{E}[\|\hat{f}_n^{\text{lin}}(\mathbf{x}) - f(\mathbf{x})\|_p^p] \lesssim n^{-\frac{s'p}{2s'+d}}.$$

Remark 2. It is easy to see from Theorem 2 that the convergence rates of this linear wavelet estimator are different under distinct cases. According to the lower bound estimations in Theorem 1, this linear wavelet estimator can attain the optimal convergence rate $n^{-\frac{s'p}{2s'+d}}$ in the case of $1 \leq p < \eta$. On the other hand, when $p = 2$, our results reduce to the conclusion of Liang and Kou [17].

Remark 3. Note that the convergence rate of this linear wavelet estimator is $n^{-\frac{(s-d)(\eta+d/p)p}{2(s-d)(\eta+d/p)+d}}$ when $1 \leq \eta \leq p$. Compared with the results of Theorem 1, the linear wavelet estimator gets worse with $1 \leq \eta \leq p$. On the other hand, the linear wavelet estimator relies on the smoothness parameter s of the unknown density function $f(\mathbf{x})$. Hence, this linear estimator is not adaptive. In order to overcome those shortages, a nonlinear wavelet estimator will be proposed by the hard thresholding method in the following section.

Proof of Theorem 2. Based on the definitions of $\hat{f}_n^{\text{lin}}(\mathbf{x})$ and orthogonal projection operators, we obtain

$$\begin{aligned} \mathbb{E}[\|\hat{f}_n^{\text{lin}}(\mathbf{x}) - f(\mathbf{x})\|_p^p] &= \mathbb{E}[\|\hat{f}_n^{\text{lin}}(\mathbf{x}) - \mathbb{P}_{V_{j_*}} f(\mathbf{x}) + \mathbb{P}_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_p^p] \\ &\lesssim \mathbb{E}[\|\hat{f}_n^{\text{lin}}(\mathbf{x}) - \mathbb{P}_{V_{j_*}} f(\mathbf{x})\|_p^p] + \|\mathbb{P}_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_p^p. \end{aligned} \quad (18)$$

By Lemma 1, it is easy to obtain

$$\begin{aligned} \mathbb{E}[\|\hat{f}_n^{\text{lin}}(\mathbf{x}) - \mathbb{P}_{V_{j_*}} f(\mathbf{x})\|_p^p] &= \mathbb{E} \left[\left\| \sum_{k \in \Lambda_{j_*}} (\hat{\alpha}_{j_*,k} - \alpha_{j_*,k}) \Phi_{j_*,k}(\mathbf{x}) \right\|_p^p \right] \\ &\lesssim 2^{j_* (\frac{d}{2} - \frac{d}{p}) p} \sum_{k \in \Lambda_{j_*}} \mathbb{E} [|\hat{\alpha}_{j_*,k} - \alpha_{j_*,k}|^p]. \end{aligned}$$

Furthermore, using Lemma 5, $|\Lambda_{j_*}| \sim 2^{j_* d}$, and $2^{j_*} \sim n^{\frac{1}{2s'+d}}$,

$$\mathbb{E}[\|\hat{f}_n^{\text{lin}}(\mathbf{x}) - \mathbb{P}_{V_{j_*}} f(\mathbf{x})\|_p^p] \lesssim 2^{j_* (\frac{d}{2} - \frac{d}{p}) p} \sum_{k \in \Lambda_{j_*}} n^{-\frac{p}{2}} \lesssim 2^{j_* (\frac{d}{2} - \frac{d}{p}) p} \cdot 2^{j_* d} \cdot n^{-\frac{p}{2}} \sim n^{-\frac{s'p}{2s'+d}}. \quad (19)$$

When $1 \leq p < \eta$, $s' = s - d(\frac{1}{\eta} - \frac{1}{p})_+ = s$. By the Hölder inequality and $f(\mathbf{x}) \in B_{\eta,q}^s(\Omega)$, there are

$$\begin{aligned} \|\mathbb{P}_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_p^p &= \int_{\Omega} |\mathbb{P}_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})|^p \cdot 1 \, d\mathbf{x} \\ &\leq \left(\int_{\Omega} |\mathbb{P}_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})|^{p \cdot \frac{\eta}{p}} \, d\mathbf{x} \right)^{\frac{p}{\eta}} \left(\int_{\Omega} 1^{\frac{\eta}{\eta-p}} \, d\mathbf{x} \right)^{1-\frac{p}{\eta}} \\ &\lesssim \left(\int_{\Omega} |\mathbb{P}_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})|^{\eta} \, d\mathbf{x} \right)^{\frac{p}{\eta}} \\ &= \|\mathbb{P}_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_{\eta}^p. \end{aligned}$$

Due to Lemma 2, the definition of $P_{V_{j_*}} f(\mathbf{x})$, and $f(\mathbf{x}) \in B_{\eta,q}^s(\Omega)$, it is easy to see that

$$\begin{aligned} \|P_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_{\eta} &= \left\| \sum_{j=j_*}^{\infty} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_{j_*}} \beta_{j,k,u} \Psi_{j,k,u}(\mathbf{x}) \right\|_{\eta} \\ &\leq \sum_{j=j_*}^{\infty} \|P_{V_{j+1}} f(\mathbf{x}) - P_{V_j} f(\mathbf{x})\|_{\eta} \\ &\lesssim \sum_{j=j_*}^{\infty} 2^{-js}. \end{aligned} \quad (20)$$

Therefore, according to (20), there are

$$\|P_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_p^p \lesssim \sum_{j=j_*}^{\infty} 2^{-js^p} \lesssim 2^{-j_* s^p} \sim n^{-\frac{sp}{2s^p+d}}. \quad (21)$$

When $1 \leq \eta \leq p$ and $s > \frac{d}{\eta}$, $s' = s - d(\frac{1}{\eta} - \frac{1}{p})_+$. The Besov spaces embedding theorem gives $B_{\eta,q}^s(\Omega) \subseteq B_{p,q}^{s-\frac{d}{\eta}+\frac{d}{p}}(\Omega)$ and

$$\|P_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_p^p \lesssim \sum_{j=j_*}^{\infty} 2^{-js'p} \lesssim 2^{-j_* s'p} \sim n^{-\frac{s'p}{2s'^p+d}}. \quad (22)$$

Hence, according to (21) and (22), for each $1 \leq p < \infty$,

$$\|P_{V_{j_*}} f(\mathbf{x}) - f(\mathbf{x})\|_p^p \lesssim n^{-\frac{s'p}{2s'^p+d}}. \quad (23)$$

Finally, owing to (18), (19), (23), we prove that

$$\mathbb{E}[\|\hat{f}_n^{\text{lin}}(\mathbf{x}) - f(\mathbf{x})\|_p^p] \lesssim n^{-\frac{s'p}{2s'^p+d}}.$$

□

4.2. Nonlinear wavelet estimator

This section constructs a nonlinear wavelet estimator based on the thresholding function, and discusses the convergence rate under L^p ($1 \leq p < \infty$) risk in Besov spaces. Firstly, we define the nonlinear wavelet estimator by

$$\hat{f}_n^{\text{non}}(\mathbf{x}) := \sum_{k \in \Lambda_{j_*}} \alpha_{j_*,k} \Phi_{j_*,k}(\mathbf{x}) + \sum_{j=j_*}^{j_2} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} \Psi_{j,k,u}(\mathbf{x}), \mathbf{x} \in \Omega,$$

with $t_n := \sqrt{\frac{\ln n}{n}}$. In addition, we develop an estimator for the coefficients $\beta_{j,k,u}$

$$\hat{\beta}_{j,k,u} := \frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Psi_{j,k,u}(X_i) - \gamma_{j,k,u}, \quad (24)$$

where $\gamma_{j,k,u} = \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Psi_{j,k,u}(\mathbf{x}) d\mathbf{x}$. This is the unbiased estimate of $\beta_{j,k,u}$ that converges to $\beta_{j,k,u}$ in $L^p(\Omega)$, which follows from Lemma 6. To facilitate the proof of Theorem 3, we first introduce several lemmas.

Lemma 6. *Assuming that $\hat{\beta}_{j,k,u}$ is defined as in (24), we obtain*

$$\mathbb{E}[\hat{\beta}_{j,k,u}] = \beta_{j,k,u}.$$

Proof. Since $\hat{\beta}_{j,k,u} = \frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Psi_{j,k,u}(X_i) - \gamma_{j,k,u}$ and $\gamma_{j,k,u} = \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Psi_{j,k,u}(\mathbf{x}) d\mathbf{x}$, by the properties of expectation, one has

$$\begin{aligned} \mathbb{E}[\hat{\beta}_{j,k,u}] &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Psi_{j,k,u}(X_i) - \gamma_{j,k,u} \right] \\ &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Psi_{j,k,u}(X_i) \right] - \gamma_{j,k,u}. \end{aligned}$$

Since it is known that X_1, X_2, \dots, X_n are i.i.d. random variables with a common density function $g(\mathbf{x})$, we can obtain

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Psi_{j,k,u}(X_i) \right] = \mathbb{E} \left[\frac{1}{1-\theta} \Psi_{j,k,u}(X_1) \right] = \int_{\Omega} \frac{1}{1-\theta} g(\mathbf{x}) \Psi_{j,k,u}(\mathbf{x}) d\mathbf{x}.$$

Combining the above arguments and the mixture distribution $g(\mathbf{x}) = \theta h(\mathbf{x}) + (1-\theta)f(\mathbf{x})$, we can have

$$\begin{aligned} \mathbb{E}[\hat{\beta}_{j,k,u}] &= \int_{\Omega} \frac{1}{1-\theta} [\theta h(\mathbf{x}) + (1-\theta)f(\mathbf{x})] \Psi_{j,k,u}(\mathbf{x}) d\mathbf{x} - \int_{\Omega} \frac{\theta}{1-\theta} h(\mathbf{x}) \Psi_{j,k,u}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\Omega} f(\mathbf{x}) \Psi_{j,k,u}(\mathbf{x}) d\mathbf{x} = \beta_{j,k,u}. \end{aligned}$$

Lemma 6 is proved. \square

Lemma 7. Let $2^{jd} \leq n$, $0 < \theta \leq c_3 < 1$, and $\hat{\beta}_{j,k,u}$ be defined as in (24). Then, for any $1 \leq p < \infty$, we can derive

$$\mathbb{E}[|\hat{\beta}_{j,k,u} - \beta_{j,k,u}|^p] \lesssim n^{-\frac{p}{2}}.$$

Proof. By Lemma 6, we have $\mathbb{E}[\hat{\beta}_{j,k,u}] = \beta_{j,k,u}$. Therefore,

$$\begin{aligned} |\hat{\beta}_{j,k,u} - \beta_{j,k,u}| &= \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Psi_{j,k,u}(X_i) - \gamma_{j,k,u} - \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{1-\theta} \Psi_{j,k,u}(X_i) - \gamma_{j,k,u} \right] \right| \\ &= \frac{1}{n} \left| \sum_{i=1}^n \frac{1}{1-\theta} (\Psi_{j,k,u}(X_i) - \mathbb{E}[\Psi_{j,k,u}(X_i)]) \right| \\ &= \frac{1}{n} \left| \sum_{i=1}^n G_i \right|, \end{aligned} \tag{25}$$

with $G_i := \frac{1}{1-\theta} (\Psi_{j,k,u}(X_i) - \mathbb{E}[\Psi_{j,k,u}(X_i)])$. Obviously, $\mathbb{E}[G_i] = 0$ and

$$\mathbb{E}[|\hat{\beta}_{j,k,u} - \beta_{j,k,u}|^p] = \mathbb{E} \left[\left| \frac{1}{n} \sum_{i=1}^n G_i \right|^p \right] = \frac{1}{n^p} \mathbb{E} \left[\left| \sum_{i=1}^n G_i \right|^p \right].$$

From the condition $0 < \theta \leq c_3 < 1$, it is known that

$$\begin{aligned} \mathbb{E}[|G_i|^p] &= \mathbb{E} \left[\left| \frac{1}{1-\theta} (\Psi_{j,k,u}(X_i) - \mathbb{E}[\Psi_{j,k,u}(X_i)]) \right|^p \right] \\ &= \frac{1}{(1-\theta)^p} \mathbb{E} [|\Psi_{j,k,u}(X_i) - \mathbb{E}[\Psi_{j,k,u}(X_i)]|^p] \\ &\lesssim \mathbb{E} [|\Psi_{j,k,u}(X_i) - \mathbb{E}[\Psi_{j,k,u}(X_i)]|^p] \\ &\lesssim \mathbb{E} [|\Psi_{j,k,u}(X_i)|^p]. \end{aligned}$$

The mixture model $g(\mathbf{x}) = \theta h(\mathbf{x}) + (1 - \theta)f(\mathbf{x})$ in (1), combined with $\Psi(\mathbf{x})$'s characteristics, yields

$$\begin{aligned} \mathbb{E} [|\Psi_{j,k,u}(X_i)|^p] &= \int_{\Omega} g(\mathbf{x}) |\Psi_{j,k,u}(\mathbf{x})|^p d\mathbf{x} \\ &= \int_{\Omega} \theta h(\mathbf{x}) |\Psi_{j,k,u}(\mathbf{x})|^p d\mathbf{x} + \int_{\Omega} (1 - \theta)f(\mathbf{x}) |\Psi_{j,k,u}(\mathbf{x})|^p d\mathbf{x} \\ &\lesssim \int_{\Omega} |\Psi_{j,k,u}(\mathbf{x})|^p d\mathbf{x}. \end{aligned}$$

Then we can obtain $\mathbb{E}[|G_i|^p] \lesssim \int_{\Omega} |\Psi_{j,k,u}(\mathbf{x})|^p d\mathbf{x}$ and

$$\mathbb{E}[|G_i|^p] \lesssim \int_{\Omega} |\Psi_{j,k,u}(\mathbf{x})|^p d\mathbf{x} \lesssim 2^{j(\frac{d}{2} - \frac{d}{p})p}.$$

Especially, when $p = 2$, $\mathbb{E}[|G_i|^2] \lesssim 1$.

Applying Rosenthal's inequality and given that $2^{jd} \leq n$, for $p > 2$, we have

$$\begin{aligned} \mathbb{E}[|\hat{\beta}_{j,k,u} - \beta_{j,k,u}|^p] &= \frac{1}{n^p} \mathbb{E} \left[\left| \sum_{i=1}^n G_i \right|^p \right] \\ &\lesssim \frac{1}{n^p} \left\{ \sum_{i=1}^n \mathbb{E}[|G_i|^p] + \left(\sum_{i=1}^n \mathbb{E}[G_i^2] \right)^{\frac{p}{2}} \right\} \\ &\lesssim n^{-p} \left(n 2^{j(\frac{d}{2} - \frac{d}{p})p} + n^{\frac{p}{2}} \right) \\ &\lesssim n^{-\frac{p}{2}}. \end{aligned}$$

When $1 \leq p \leq 2$, we get

$$\begin{aligned} \mathbb{E}[|\hat{\beta}_{j,k,u} - \beta_{j,k,u}|^p] &= \frac{1}{n^p} \mathbb{E} \left[\left| \sum_{i=1}^n G_i \right|^p \right] \\ &\lesssim \frac{1}{n^p} \left(\sum_{i=1}^n \mathbb{E}[G_i^2] \right)^{\frac{p}{2}} \\ &\lesssim n^{-\frac{p}{2}}. \end{aligned}$$

Hence, we can obtain

$$\mathbb{E}[|\hat{\beta}_{j,k,u} - \beta_{j,k,u}|^p] \lesssim n^{-\frac{p}{2}}.$$

□

Bernstein's inequality[14] Let X_1, \dots, X_n be independent random variables. We assume that $\mathbb{E}[X_i] = 0$, $|X_i| < c$, and $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \mathbf{var}(X_i)$. Under these conditions, for each $\epsilon > 0$,

$$\Pr \left(\frac{1}{n} \left| \sum_{i=1}^n X_i \right| \geq \epsilon \right) \leq 2 \exp \left\{ -\frac{n\epsilon^2}{2(\sigma^2 + c\epsilon/3)} \right\}.$$

Lemma 8. Let $2^{jd} \lesssim \frac{n}{\ln n}$, $0 < \theta \leq c_3 < 1$, and $p \in [1, \infty)$. The estimator $\hat{\beta}_{j,k,u}$ is given by (24). For some constant $\kappa > 1$, we have

$$\Pr(|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| \geq \kappa t_n) \lesssim n^{-2p}.$$

Proof. According to (25), note that $|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| = \frac{1}{n} \left| \sum_{i=1}^n G_i \right|$ and

$$\left\{ |\hat{\beta}_{j,k,u} - \beta_{j,k,u}| \geq \kappa t_n \right\} \subseteq \left\{ \frac{1}{n} \left| \sum_{i=1}^n G_i \right| \geq \kappa t_n \right\}. \quad (26)$$

It is known that $G_i := \frac{1}{1-\theta} (\Psi_{j,k,u}(X_i) - \mathbb{E}[\Psi_{j,k,u}(X_i)])$, which gives $\mathbb{E}[G_i] = 0$. Because $h(\mathbf{x})$, $f(\mathbf{x})$, and $\Phi(\mathbf{x})$ are all bounded functions,

$$\begin{aligned} |G_i| &\leq |\Psi_{j,k,u}(X_i)| + |\mathbb{E}[\Psi_{j,k,u}(X_i)]| \\ &= \left| 2^{\frac{jd}{2}} \Psi_u(2^j X_i - \mathbf{k}) \right| + \left| \int_{\Omega} g(\mathbf{x}) 2^{\frac{jd}{2}} \Psi_u(2^j \mathbf{x} - \mathbf{k}) d\mathbf{x} \right| \\ &= \left| 2^{\frac{jd}{2}} \Psi_u(2^j X_i - \mathbf{k}) \right| + \left| \int_{\Omega} \theta h(\mathbf{x}) 2^{\frac{jd}{2}} \Psi_u(2^j \mathbf{x} - \mathbf{k}) d\mathbf{x} \right| \\ &\quad + \left| \int_{\Omega} (1-\theta) f(\mathbf{x}) 2^{\frac{jd}{2}} \Psi_u(2^j \mathbf{x} - \mathbf{k}) d\mathbf{x} \right| \lesssim 2^{\frac{jd}{2}}. \end{aligned}$$

By the condition $2^{jd} \lesssim \frac{n}{\ln n}$,

$$|G_i| \lesssim \sqrt{\frac{n}{\ln n}}. \quad (27)$$

Due to the proof of Lemma 7,

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \mathbf{var}[G_i] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[G_i^2] \lesssim 1. \quad (28)$$

According to (26)–(28) and Bernstein's inequality, we have

$$\begin{aligned} \Pr \left(\frac{1}{n} \left| \sum_{i=1}^n G_i \right| \geq \kappa t_n \right) &\lesssim \exp \left\{ -\frac{n \kappa^2 t_n^2}{2(\sigma^2 + (\kappa t_n \sqrt{\frac{n}{\ln n}})/3)} \right\} \\ &\lesssim \exp \left\{ -\frac{(\ln n) \kappa^2}{2(\sigma^2 + \kappa/3)} \right\} \\ &\lesssim n^{-\frac{\kappa^2}{2(\sigma^2 + \kappa/3)}}. \end{aligned}$$

Then, a sufficiently large κ can be selected such that

$$\Pr(|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| \geq \kappa t_n) \lesssim n^{-\frac{\kappa^2}{2(\sigma^2 + \kappa/3)}} \lesssim n^{-2p}.$$

□

Under the above conditions, the convergence rate of the nonlinear wavelet estimator is characterized by the following theorem.

Theorem 3. For the model (1), $f(\mathbf{x}) \in B_{\eta,q}^s(\Omega)$ with $\eta, q \in [1, \infty)$, $s > 0$ and $1 \leq p < \infty$. If $\{1 \leq p < \eta\}$ or $\{1 \leq \eta \leq p, s > \frac{d}{\eta}\}$, the nonlinear estimator $\hat{f}_n^{\text{non}}(\mathbf{x})$ with $2^{j^*} \sim (n/\ln n)^{\frac{1}{d}}$ and $2^{j^*} \sim n^{\frac{1}{2m+d}}$ ($m > s$) has

$$\mathbb{E}[\|\hat{f}_n^{\text{non}}(\mathbf{x}) - f(\mathbf{x})\|_p^p] \lesssim (\ln n)^{\frac{3p}{2}} n^{-\delta p},$$

where

$$\delta = \min \left\{ \frac{s}{2s+d}, \frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d} \right\} = \begin{cases} \frac{s}{2s+d}, & \eta > \frac{pd}{2s+d}, \\ \frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d}, & \eta \leq \frac{pd}{2s+d}. \end{cases}$$

Remark 4. Taking into account the lower bound estimations in Theorem 1, note that the convergence rate of this nonlinear wavelet estimator matches the optimal convergence rate up to an $(\ln n)^{\frac{3p}{2}}$ factor. In addition, the results of this theorem are consistent with the convergence rate of Liang and Kou [17], in the case of $p = 2$.

Remark 5. While both linear and nonlinear wavelet estimators achieve optimal convergence rates when $\eta > p$, the nonlinear estimator demonstrates superior convergence performance for $p \leq \eta$. Notably, the nonlinear wavelet estimator exhibits adaptivity as its construction does not require prior knowledge of the smoothness parameters of the unknown density function.

Remark 6. Note that the mixture density model (1) reduces to the nonparametric estimation problem of Kou and Chen [16], when the density function $h(\mathbf{x}) \equiv 1$, which means that the function $h(\mathbf{x})$ denotes the density function of uniform distribution. Then, those upper bound estimations of two wavelet estimators in Theorems 2 and 3 are the same as the results of Kou and Chen [16]. It should be pointed out that Kou and Chen [16] only focus on the upper bound estimations of the wavelet method, they do not study the lower bound estimations. In this paper, in order to discuss the optimality of wavelet estimators, the lower bound estimations of wavelet estimators are investigated in Section 3.

Proof. By the definitions of $\hat{f}_n^{\text{non}}(\mathbf{x})$, $\hat{f}_n^{\text{lin}}(\mathbf{x})$, $\mathbb{P}_{V_j} f(\mathbf{x})$, and (2),

$$\begin{aligned} \hat{f}_n^{\text{non}}(\mathbf{x}) - f(\mathbf{x}) &= \sum_{k \in \Lambda_{j_*}} (\hat{\alpha}_{j_*,k} - \alpha_{j_*,k}) \Phi_{j_*,k}(\mathbf{x}) - \sum_{j=j_2+1}^{\infty} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \beta_{j,k,u} \Psi_{j,k,u}(\mathbf{x}) \\ &\quad + \sum_{j=j_*}^{j_2} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} \Psi_{j,k,u}(\mathbf{x}) - \sum_{j=j_*}^{j_2} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \beta_{j,k,u} \Psi_{j,k,u}(\mathbf{x}) \\ &= (\hat{f}_n^{\text{lin}}(\mathbf{x}) - \mathbb{P}_{V_{j_*}} f(\mathbf{x})) - (f(\mathbf{x}) - \mathbb{P}_{V_{j_2+1}} f(\mathbf{x})) \\ &\quad + \sum_{j=j_*}^{j_2} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} (\hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} - \beta_{j,k,u}) \Psi_{j,k,u}(\mathbf{x}). \end{aligned}$$

Hence,

$$\mathbb{E} \left[\left\| \hat{f}_n^{\text{non}}(\mathbf{x}) - f(\mathbf{x}) \right\|_p^p \right] \lesssim A + B + C. \quad (29)$$

In the above equality,

$$\begin{aligned} A &:= \mathbb{E} \left[\left\| \hat{f}_n^{\text{lin}}(\mathbf{x}) - \mathbb{P}_{V_{j_*}} f(\mathbf{x}) \right\|_p^p \right], \\ B &:= \left\| f(\mathbf{x}) - \mathbb{P}_{V_{j_2+1}} f(\mathbf{x}) \right\|_p^p, \\ C &:= \mathbb{E} \left[\left\| \sum_{j=j_*}^{j_2} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} (\hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} - \beta_{j,k,u}) \Psi_{j,k,u}(\mathbf{x}) \right\|_p^p \right]. \end{aligned}$$

For A, it follows from Lemma 1 that

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{f}_n^{\text{lin}}(\mathbf{x}) - \mathbb{P}_{V_{j_*}} f(\mathbf{x}) \right\|_p^p \right] &= \mathbb{E} \left[\left\| \sum_{k \in \Lambda_{j_*}} (\hat{\alpha}_{j_*,k} - \alpha_{j_*,k}) \Phi_{j_*,k}(\mathbf{x}) \right\|_p^p \right] \\ &\lesssim 2^{j_* (\frac{d}{2} - \frac{d}{p}) p} \sum_{k \in \Lambda_{j_*}} \mathbb{E} \left[|\hat{\alpha}_{j_*,k} - \alpha_{j_*,k}|^p \right]. \end{aligned}$$

By Lemma 5, we have $|\Lambda_{j_*}| \sim 2^{j_* d}$, with $2^{j_*} \sim n^{\frac{1}{2m+d}}$ ($m > s$),

$$A \lesssim 2^{j_* (\frac{d}{2} - \frac{d}{p}) p} \sum_{k \in \Lambda_{j_*}} n^{-\frac{p}{2}} \lesssim 2^{\frac{j_* d}{2}} \cdot n^{-\frac{p}{2}} \sim n^{-\frac{mp}{2m+d}} \leq n^{-\frac{sp}{2s+d}} \leq n^{-\delta p}. \quad (30)$$

For B , when $1 \leq p < \eta$, $s' = s - d(\frac{1}{\eta} - \frac{1}{p})_+ = s$. By the Hölder inequality and $f(\mathbf{x}) \in B_{\eta,q}^s(\Omega)$,

$$\begin{aligned} \|f(\mathbf{x}) - \mathbb{P}_{V_{j_2+1}} f(\mathbf{x})\|_p^p &= \int_{\Omega} |f(\mathbf{x}) - \mathbb{P}_{V_{j_2+1}} f(\mathbf{x})|^p \cdot 1 d\mathbf{x} \\ &\leq \left(\int_{\Omega} |f(\mathbf{x}) - \mathbb{P}_{V_{j_2+1}} f(\mathbf{x})|^{p \cdot \frac{\eta}{p}} d\mathbf{x} \right)^{\frac{p}{\eta}} \left(\int_{\Omega} 1^{1 \cdot \frac{p}{p-\eta}} d\mathbf{x} \right)^{1-\frac{p}{\eta}} \\ &\leq \left(\int_{\Omega} |f(\mathbf{x}) - \mathbb{P}_{V_{j_2+1}} f(\mathbf{x})|^{\eta} d\mathbf{x} \right)^{\frac{p}{\eta}} \\ &= \|f(\mathbf{x}) - \mathbb{P}_{V_{j_2+1}} f(\mathbf{x})\|_{\eta}^p. \end{aligned}$$

Then, the same method is used as in the proof of (20), and we can obtain

$$B \lesssim \sum_{j=j_2+1}^{\infty} 2^{-jsp} \lesssim 2^{-j_2 sp} \sim \left(\frac{\ln n}{n} \right)^{\frac{sp}{d}} \leq \left(\frac{\ln n}{n} \right)^{\frac{sp}{2s+d}} \leq \left(\frac{\ln n}{n} \right)^{\delta p}. \quad (31)$$

Furthermore, when $1 \leq \eta \leq p$ and $s > \frac{d}{\eta}$, $s' = s - \frac{d}{\eta} + \frac{d}{p}$. The Besov spaces embedding theorem gives $B_{\eta,q}^s(\Omega) \subseteq B_{p,q}^{s-\frac{d}{\eta}+\frac{d}{p}}(\Omega)$ and

$$B \lesssim \sum_{j=j_2+1}^{\infty} 2^{-js'p} \lesssim 2^{-j_2(s-\frac{d}{\eta}+\frac{d}{p})p} \sim \left(\frac{\ln n}{n} \right)^{\frac{(s-\frac{d}{\eta}+\frac{d}{p})p}{d}} \leq \left(\frac{\ln n}{n} \right)^{\frac{(s-\frac{d}{\eta}+\frac{d}{p})p}{2(s-\frac{d}{\eta})+d}} \leq \left(\frac{\ln n}{n} \right)^{\delta p}. \quad (32)$$

In conclusion, by (31) and (32), when $1 \leq p < \infty$,

$$B \lesssim \left(\frac{\ln n}{n} \right)^{\delta p}. \quad (33)$$

By Lemma 1 and the Hölder inequality, C can be written as

$$\begin{aligned} C &:= \mathbb{E} \left[\left\| \sum_{j=j_*}^{j_2} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \left(\hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} - \beta_{j,k,u} \right) \Psi_{j,k,u}(\mathbf{x}) \right\|_p^p \right] \\ &\lesssim (j_2 - j_* + 1)^{p-1} \sum_{j=j_*}^{j_2} \mathbb{E} \left[\left\| \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \left(\hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} - \beta_{j,k,u} \right) \Psi_{j,k,u}(\mathbf{x}) \right\|_p^p \right] \\ &\lesssim (j_2 - j_* + 1)^{p-1} \sum_{j=j_*}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} - \beta_{j,k,u} \right|^p \right]. \end{aligned}$$

Note that

$$\begin{aligned} \left| \hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} - \beta_{j,k,u} \right| &\lesssim \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} + \left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n\}} \\ &= \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}, |\beta_{j,k,u}| < \frac{\kappa t_n}{2}\}} \\ &\quad + \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}, |\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \\ &\quad + \left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| > 2\kappa t_n\}} \\ &\quad + \left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| \leq 2\kappa t_n\}}. \end{aligned}$$

We will discuss the following four parts:

$$(1) \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n, |\beta_{j,k,u}| < \frac{\kappa t_n}{2}\}},$$

$$(2) \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n, |\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}},$$

$$(3) \left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| > 2\kappa t_n\}},$$

$$(4) \left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| \leq 2\kappa t_n\}}.$$

For (1), the conditions $|\hat{\beta}_{j,k,u}| \geq \kappa t_n$ and $|\beta_{j,k,u}| < \frac{\kappa t_n}{2}$ imply that $|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| \geq |\hat{\beta}_{j,k,u}| - |\beta_{j,k,u}| > \frac{\kappa t_n}{2}$ and

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

Therefore, we can get

$$\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n, |\beta_{j,k,u}| < \frac{\kappa t_n}{2}\}} \leq \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \frac{\kappa t_n}{2}\}}.$$

For (2), since $\mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n, |\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \leq \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}}$, we can get

$$\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n, |\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \leq \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}}.$$

For (4), due to $\mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| \leq 2\kappa t_n\}} \leq \mathbb{I}_{\{|\beta_{j,k,u}| \leq 2\kappa t_n\}}$, it is easy to see that

$$\left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| \leq 2\kappa t_n\}} \leq \left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\beta_{j,k,u}| \leq 2\kappa t_n\}}.$$

For (3), note that $|\hat{\beta}_{j,k,u}| < \kappa t_n$ and $|\beta_{j,k,u}| > 2\kappa t_n$; we see that $|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > |\beta_{j,k,u}| - |\hat{\beta}_{j,k,u}| > \kappa t_n$,

$$\begin{aligned} \left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| > 2\kappa t_n\}} &\leq (|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| + |\hat{\beta}_{j,k,u}|) \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| > 2\kappa t_n\}} \\ &< (|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| + |\hat{\beta}_{j,k,u}|) \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \kappa t_n\}} \\ &< (|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| + |2\kappa t_n|) \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \kappa t_n\}} \\ &\lesssim \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \frac{\kappa t_n}{2}\}}. \end{aligned}$$

Therefore, we know that

$$\left| \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| < \kappa t_n, |\beta_{j,k,u}| > 2\kappa t_n\}} \lesssim \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \frac{\kappa t_n}{2}\}}.$$

In line with these discussions, we obtain

$$\begin{aligned} \left| \hat{\beta}_{j,k,u} \mathbb{I}_{\{|\hat{\beta}_{j,k,u}| \geq \kappa t_n\}} - \beta_{j,k,u} \right|^p &\lesssim \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \frac{\kappa t_n}{2}\}} \\ &\quad + \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} + \left| \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\beta_{j,k,u}| \leq 2\kappa t_n\}}. \end{aligned}$$

Therefore, C can be decomposed into three parts:

$$C \lesssim (j_2 - j_* + 1)^{p-1} (C_1 + C_2 + C_3), \quad (34)$$

where

$$\begin{aligned} C_1 &= \sum_{j=j_*}^{j_2} 2^{j(\frac{d}{2} - \frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \frac{\kappa t_n}{2}\}} \right], \\ C_2 &= \sum_{j=j_*}^{j_2} 2^{j(\frac{d}{2} - \frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \right], \\ C_3 &= \sum_{j=j_*}^{j_2} 2^{j(\frac{d}{2} - \frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \left| \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\beta_{j,k,u}| \leq 2\kappa t_n\}}. \end{aligned}$$

For C_1 , it follows from the Hölder inequality that

$$\begin{aligned} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \frac{\kappa t_n}{2}\}} \right] &\leq \left\{ \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^{2p} \right] \right\}^{\frac{1}{2}} \left\{ \mathbb{E} \left[\mathbb{I}_{\{|\hat{\beta}_{j,k,u} - \beta_{j,k,u}| > \frac{\kappa t_n}{2}\}} \right] \right\}^{\frac{1}{2}} \\ &\leq \left\{ \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^{2p} \right] \right\}^{\frac{1}{2}} \left\{ \Pr \left(\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right| > \frac{\kappa t_n}{2} \right) \right\}^{\frac{1}{2}}. \end{aligned}$$

Based on Lemma 7, Lemma 8, and $|\hat{\beta}_{j,k,u} - \beta_{j,k,u}|^p \lesssim \left(\frac{n}{\ln n}\right)^{\frac{p}{2}}$, we get

$$\begin{aligned} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^{2p} \right] &\leq \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \right] \\ &\lesssim \left(\frac{n}{\ln n}\right)^{\frac{p}{2}} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \right] \\ &\lesssim \left(\frac{n}{\ln n}\right)^{\frac{p}{2}} n^{-\frac{p}{2}} \lesssim (\ln n)^{-\frac{p}{2}}. \end{aligned}$$

Furthermore, since $|\Lambda_{j^*}| \sim 2^{j^*d}$ and $2^{j^2} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{d}}$,

$$\begin{aligned} C_1 &\lesssim \sum_{j=j^*}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} (\ln n)^{-\frac{p}{4}} n^{-p} \lesssim \sum_{j=j^*}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{jd} (\ln n)^{-\frac{p}{4}} n^{-p} \\ &\lesssim 2^{j_2(\frac{d}{2}-\frac{d}{p})p+j_2d} (\ln n)^{-\frac{p}{4}} n^{-p} \lesssim (\ln n)^{-\frac{3p}{4}} n^{-\frac{p}{2}} \\ &\leq \left(\frac{\ln n}{n}\right)^{\frac{p}{2}} \leq \left(\frac{\ln n}{n}\right)^{\delta p}. \end{aligned} \tag{35}$$

For C_2 , one defines $2^{j'} \sim n^{\frac{1}{2s+d}}$, we obtain $2^{j^*} \sim n^{\frac{1}{2m+d}}$ ($m > s$) $\leq 2^{j'} \sim n^{\frac{1}{2s+d}} \leq 2^{j_2} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{d}}$, and C_2 can be defined as

$$\begin{aligned} C_2 &= \left(\sum_{j=j^*}^{j'} + \sum_{j=j'+1}^{j_2} \right) 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \right] \\ &:= C_{21} + C_{22}. \end{aligned} \tag{36}$$

From Lemma 7 and $2^{j'} \sim n^{\frac{1}{2s+d}}$,

$$\begin{aligned} C_{21} &:= \sum_{j=j^*}^{j'} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \right] \\ &\lesssim \sum_{j=j^*}^{j'} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} n^{-\frac{p}{2}} \lesssim \sum_{j=j^*}^{j'} 2^{j(\frac{d}{2}-\frac{d}{p})p} \cdot 2^{jd} \cdot n^{-\frac{p}{2}} \\ &\lesssim 2^{\frac{j'pd}{2}} \cdot n^{-\frac{p}{2}} \sim n^{-\frac{sp}{2s+d}} \leq n^{-\delta p}. \end{aligned} \tag{37}$$

For C_{22} , when $p \geq 1$, using Lemma 7, one obtains

$$\begin{aligned} C_{22} &:= \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{E} \left[\left| \hat{\beta}_{j,k,u} - \beta_{j,k,u} \right|^p \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \right] \\ &\lesssim n^{-\frac{p}{2}} \sum_{j=j^*}^{j'} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{I}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}}. \end{aligned}$$

When $1 \leq p < \eta$, by the Hölder inequality, Lemma 7, $t_n = \sqrt{\frac{\ln n}{n}}$, and $2^{j'} \sim n^{\frac{1}{2s+d}}$, we get

$$\begin{aligned}
 C_{22} &\lesssim n^{-\frac{p}{2}} \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \left(\frac{|\beta_{j,k,u}|}{\kappa t_n/2} \right)^p \lesssim (\ln n)^{-\frac{p}{2}} \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \|\beta_{j,k,u}\|_p^p \\
 &\leq \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{jd(1-\frac{d}{\eta})p} \|\beta_{j,k,u}\|_\eta^p \leq \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{jd(1-\frac{d}{\eta})p} 2^{-j(s-\frac{d}{\eta}+\frac{d}{2})p} \\
 &= \sum_{j=j'+1}^{j_2} 2^{-jps} \lesssim 2^{-j'ps} \sim n^{-\frac{sp}{2s+d}} \leq n^{-\delta p}.
 \end{aligned} \tag{38}$$

When $1 \leq \eta < p$ and $s > \frac{d}{\eta}$,

$$\begin{aligned}
 C_{22} &\lesssim n^{-\frac{p}{2}} \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \left(\frac{|\beta_{j,k,u}|}{\kappa t_n/2} \right)^\eta \\
 &\lesssim (\ln n)^{-\frac{\eta}{2}} n^{\frac{\eta-p}{2}} \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \|\beta_{j,k,u}\|_\eta^\eta \\
 &\leq (\ln n)^{-\frac{\eta}{2}} n^{\frac{\eta-p}{2}} \sum_{j=j'+1}^{j_2} 2^{-j(s\eta-\frac{d}{2}(p-\eta))}.
 \end{aligned}$$

Let $\varepsilon := s\eta - \frac{d}{2}(p - \eta)$. Then C_{22} can be written as

$$C_{22} \lesssim (\ln n)^{-\frac{\eta}{2}} n^{\frac{\eta-p}{2}} \sum_{j=j'+1}^{j_2} 2^{-j\varepsilon}.$$

When $\varepsilon > 0$, if and only if $\eta \leq \frac{dp}{2s+d}$, then $\delta = \frac{s}{2s+d}$ and

$$C_{22} \lesssim (\ln n)^{-\frac{\eta}{2}} n^{\frac{\eta-p}{2}} 2^{-j'\varepsilon} \lesssim (\ln n)^{-\frac{\eta}{2}} n^{-\frac{sp}{2s+d}} \leq n^{-\frac{sp}{2s+d}} \leq n^{-\delta p}. \tag{39}$$

When $\varepsilon \leq 0$, if and only if $\eta \leq \frac{dp}{2s+d}$, then we have $\delta = \frac{s-\frac{d}{\eta}+\frac{d}{p}}{2(s-\frac{d}{\eta})+d}$. We take $2^{j_1} \sim \frac{1}{2(s-\frac{d}{\eta})+d}$. Obviously, $2^{j'} \sim n^{\frac{1}{2s+d}} < 2^{j_1} \sim \frac{1}{2(s-\frac{d}{\eta})+d} < 2^{j_2} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{d}}$, C_{22} can be written as

$$\begin{aligned}
 C_{22} &:= \left(\sum_{j=j'+1}^{j_1} + \sum_{j=j_1+1}^{j_2} \right) 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \mathbb{E} \left[|\beta_{j,k,u} - \beta_{j,k,u}|^p \mathbb{1}_{\{|\beta_{j,k,u}| \geq \frac{\kappa t_n}{2}\}} \right] \\
 &:= C_{221} + C_{222}.
 \end{aligned}$$

For C_{221} , note that $\frac{\eta-p}{2} - \frac{\delta\varepsilon}{s-\frac{d}{\eta}+\frac{d}{p}} = -\delta p$ and

$$\begin{aligned}
 C_{221} &\lesssim (\ln n)^{-\frac{\eta}{2}} n^{\frac{\eta-p}{2}} \sum_{j=j'+1}^{j_1} 2^{-j\varepsilon} \lesssim (\ln n)^{-\frac{\eta}{2}} n^{\frac{\eta-p}{2}} 2^{-j_1\varepsilon} \\
 &\lesssim (\ln n)^{-\frac{\eta}{2}} n^{\frac{\eta-p}{2}} n^{-\frac{\delta\varepsilon}{s-\frac{d}{\eta}+\frac{d}{p}}} = (\ln n)^{-\frac{\eta}{2}} n^{-\delta p} \leq n^{-\delta p}.
 \end{aligned} \tag{40}$$

For C_{222} , when $1 \leq \eta \leq p$ and $s > \frac{d}{\eta}$, $B_{\eta,q}^s(\Omega) \subseteq B_{p,q}^{s-\frac{d}{\eta}+\frac{d}{p}}(\Omega)$. Furthermore, we can get

$$\begin{aligned} C_{222} &\lesssim n^{-\frac{p}{2}} \sum_{j=j_1+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} \left(\frac{|\beta_{j,k,u}|}{\kappa t_n/2} \right)^p \lesssim (\ln n)^{-\frac{p}{2}} \sum_{j=j_1+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \|\beta_{j,k,u}\|_p^p \\ &\leq (\ln n)^{-\frac{p}{2}} \sum_{j=j_1+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{-j(s-\frac{d}{\eta}+\frac{d}{2})p} \lesssim (\ln n)^{-\frac{p}{2}} 2^{-j_1(s-\frac{d}{\eta}+\frac{d}{2})p} \\ &\sim (\ln n)^{-\frac{p}{2}} n^{-\frac{(s-\frac{d}{\eta}+\frac{d}{2})p}{2(s-\frac{d}{\eta})+d}} \leq (\ln n)^{-\frac{p}{2}} n^{-\delta p} \leq n^{-\delta p}. \end{aligned} \quad (41)$$

By combining (38), (39), (40), and (41), we get

$$C_{22} \lesssim n^{-\delta p}. \quad (42)$$

This, together with (36), (37), and (42), shows that

$$C_2 \lesssim n^{-\delta p}. \quad (43)$$

Finally, C_3 could be rewritten as

$$\begin{aligned} C_3 &= \left(\sum_{j=j_*}^{j'} + \sum_{j=j'+1}^{j_2} \right) 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} |\beta_{j,k,u}|^p \mathbb{I}_{\{|\beta_{j,k,u}| \leq 2\kappa t_n\}} \\ &:= C_{31} + C_{32}. \end{aligned}$$

For C_{31} ,

$$\begin{aligned} C_{31} &= \sum_{j=j_*}^{j'} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} |\beta_{j,k,u}|^p \mathbb{I}_{\{|\beta_{j,k,u}| \leq 2\kappa t_n\}} \leq \sum_{j=j_*}^{j'} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} |2\kappa t_n|^p \\ &\leq \sum_{j=j_*}^{j'} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{jd} \left(\frac{\ln n}{n} \right)^{\frac{p}{2}} \lesssim \left(\frac{\ln n}{n} \right)^{\frac{p}{2}} 2^{\frac{j'pd}{2}} \\ &\lesssim (\ln n)^{\frac{p}{2}} n^{-\frac{sp}{2s+d}} \leq (\ln n)^{\frac{p}{2}} n^{-\delta p}. \end{aligned}$$

For C_{32} , when $1 \leq p < \eta$, by the Hölder inequality, Lemma 2, $t_n = \sqrt{\frac{\ln n}{n}}$, and $2^{j'} \sim n^{\frac{1}{2s+d}}$,

$$\begin{aligned} C_{32} &\leq \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} |\beta_{j,k,u}|^p = \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \|\beta_{j,k,u}\|_p^p \\ &\leq \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{jd(1-\frac{d}{\eta})p} \|\beta_{j,k,u}\|_\eta^p \leq \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{jd(1-\frac{d}{\eta})p} 2^{-j(s-\frac{d}{\eta}+\frac{d}{2})p} \\ &\lesssim \sum_{j=j'+1}^{j_2} 2^{-jps} \lesssim 2^{-j'ps} \sim n^{-\frac{sp}{2s+d}} \leq n^{-\delta p}. \end{aligned}$$

When $1 \leq \eta \leq p$, one has

$$\begin{aligned} C_{32} &\leq \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} |\beta_{j,k,u}|^p \left(\frac{2\kappa t_n}{|\beta_{j,k,u}|} \right)^{p-\eta} \lesssim \left(\frac{\ln n}{n} \right)^{\frac{p-\eta}{2}} \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} \|\beta_{j,k,u}\|_\eta^\eta \\ &\lesssim \left(\frac{\ln n}{n} \right)^{\frac{p-\eta}{2}} \sum_{j=j'+1}^{j_2} 2^{j(\frac{d}{2}-\frac{d}{p})p} 2^{-j\eta(s-\frac{d}{\eta}+\frac{d}{2})} \leq \left(\frac{\ln n}{n} \right)^{\frac{p-\eta}{2}} \sum_{j=j'+1}^{j_2} 2^{-j(s\eta-\frac{d}{2}(p-\eta))} \\ &= \left(\frac{\ln n}{n} \right)^{\frac{p-\eta}{2}} \sum_{j=j'+1}^{j_2} 2^{-j\varepsilon}. \end{aligned}$$

When $\varepsilon > 0$, $\eta \leq \frac{dp}{2s+d}$ and $\delta = \frac{s}{2s+d}$. Then

$$\begin{aligned} C_{32} &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\eta}{2}} 2^{-j'\varepsilon} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\eta}{2}} 2^{-j'(s\eta - \frac{d}{2}(p-\eta))} \\ &\lesssim (\ln n)^{\frac{p-\eta}{2}} n^{\frac{\eta-p}{2}} n^{-\frac{s\eta - \frac{d}{2}(p-\eta)}{2s+d}} \leq (\ln n)^{\frac{p}{2}} n^{-\frac{sp}{2s+d}} \leq (\ln n)^{\frac{p}{2}} n^{-\delta p}. \end{aligned}$$

When $\varepsilon \leq 0$, $\eta \leq \frac{dp}{2s+d}$ and $\delta = \frac{s - \frac{d}{\eta} + \frac{d}{p}}{2(s - \frac{d}{\eta}) + d}$. One defines $2^{j_1} \sim \frac{1}{2(s - \frac{d}{\eta}) + d}$; obviously, $2^{j'} \sim n^{\frac{1}{2s+d}} < 2^{j_1} \sim \frac{1}{2(s - \frac{d}{\eta}) + d} < 2^{j_2} \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{d}}$, and C_{32} can be written as

$$\begin{aligned} C_{32} &= \left(\sum_{j=j'+1}^{j_1} + \sum_{j=j_1+1}^{j_2} \right) 2^{j(\frac{d}{2} - \frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} |\beta_{j,k,u}|^p \mathbb{I}_{\{|\beta_{j,k,u}| \leq 2\kappa t_n\}} \\ &:= C_{321} + C_{322}. \end{aligned}$$

For C_{321} ,

$$C_{321} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\eta}{2}} \sum_{j=j'+1}^{j_1} 2^{-j\varepsilon} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\eta}{2}} 2^{-j_1\varepsilon} \sim (\ln n)^{\frac{p}{2}} n^{-\delta p}.$$

For C_{322} , when $1 \leq \eta \leq p$ and $s > \frac{d}{\eta}$, $B_{\eta,q}^s(\Omega) \subseteq B_{p,q}^{s - \frac{d}{\eta} + \frac{d}{p}}(\Omega)$, and

$$\begin{aligned} C_{322} &\lesssim \sum_{j=j_1+1}^{j_2} 2^{j(\frac{d}{2} - \frac{d}{p})p} \sum_{u=1}^{2^d-1} \sum_{k \in \Lambda_j} |\beta_{j,k,u}|^p \lesssim \sum_{j=j_1+1}^{j_2} 2^{j(\frac{d}{2} - \frac{d}{p})p} \|\beta_{j,k,u}\|_p^p \\ &\lesssim \sum_{j=j_1+1}^{j_2} 2^{j(\frac{d}{2} - \frac{d}{p})p} 2^{-j(s - \frac{d}{\eta} + \frac{d}{2})p} \leq \sum_{j=j_1+1}^{j_2} 2^{-jp(s - \frac{d}{\eta} + \frac{d}{p})} \\ &\lesssim 2^{-j_1 p(s - \frac{d}{\eta} + \frac{d}{p})} \leq n^{-\delta p}. \end{aligned}$$

Therefore, in all cases,

$$C_3 \lesssim (\ln n)^{\frac{p}{2}} n^{-\delta p}. \quad (44)$$

Combining the above results (34), (35), (43), and (44), we have

$$C \lesssim (j_2 - j_* + 1)^{p-1} (C_1 + C_2 + C_3) \lesssim (\ln n)^{\frac{3p}{2}} n^{-\delta p}.$$

This, together with (29), (30), and (33), shows that

$$\mathbb{E}[\|\hat{f}_n^{\text{non}}(\mathbf{x}) - f(\mathbf{x})\|_p^p] \lesssim (\ln n)^{\frac{3p}{2}} n^{-\delta p},$$

which is the desired conclusion. \square

5. Numerical simulations

To assess the empirical performance of the proposed wavelet-based estimators, we conduct a series of numerical simulations. Specifically, we consider the estimation problem defined in (1), where the goal is to reconstruct the unknown probability density function $f(x)$ from a set of observed data points $\{X_i\}_{i=1}^n$. In order to evaluate the performance of the estimators, we employed the $\text{MSE}(f(X_i), \hat{f}(X_i)) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - \hat{f}(X_i))^2$ as the assessment criterion. For the linear estimator $\hat{f}_n^{\text{lin}}(\mathbf{x})$, the wavelet scale

parameter j_* is selected from the discrete set $j_* \in \{0, 1, \dots, \log_2(n) - 1\}$. The optimal choice of j_* is obtained by minimizing the mean squared error ($\text{MSE}(f(X_i), \hat{f}_n^{\text{lin}}(X_i))$). For the nonlinear wavelet estimator, the scale parameter j_2 is fixed at the maximum level of wavelet decomposition $j_2 = \log_2(n) - 1$. Similar to the choice of j_* , the other optimal threshold parameter $\lambda = \kappa t_n$ is selected by minimizing the MSE ($\text{MSE}(f(X_i), \hat{f}_n^{\text{non}}(X_i))$). The simulation experiments are done using RStudio software. In addition, the Daubechies compactly supported wavelet with 8 vanishing moments is used in the following studies.

In the simulation study of model (1), we choose $n = 4096$ and the proportionality coefficient $\theta = 0.1$. Six functions are selected as $f(x)$. As shown in Figures 1–6, both linear and nonlinear wavelet estimators have excellent performance. Table 1 presents the MSE results for both linear and nonlinear wavelet estimators across various experimental settings. The simulation study indicates that these estimators are capable of effectively approximating the unknown density function $f(x)$. Notably, the nonlinear wavelet estimator exhibits superior performance compared to its linear counterpart.

Example 1. In model (1), we consider two density functions: $f_1(x) = 2.38x^2e^{-x^2}$ and $h(x) = 0.15x^2 + 0.6 \cos 2x + 0.15$ defined on the interval $x \in [0, 2]$. As depicted in Figure 1(a), the optimal scale parameter is determined to be $j_* = 5$, and then the corresponding $\text{MSE}(\hat{f}_n^{\text{lin}}, f) = 0.01949769$. Based on Figure 1(c), we find the optimal threshold parameter to be $\lambda = 0.0478260870$. The performance of both linear and nonlinear wavelet estimators is visualized in Figures 1(b) and 1(d), respectively. These results clearly demonstrate the effectiveness of the proposed wavelet-based estimators in approximating the target density function.

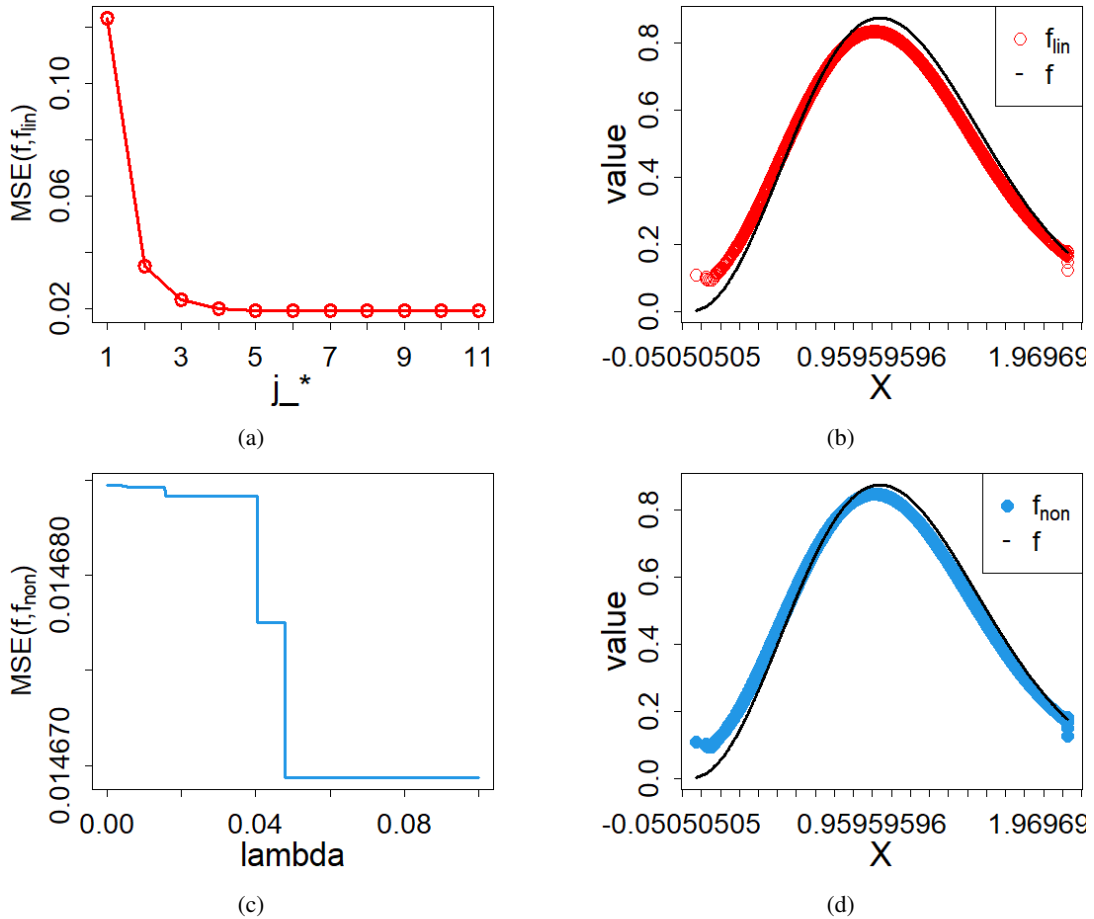


Figure 1: The estimates of $f_1(x)$: (a) $\text{MSE}(\hat{f}_n^{\text{lin}}, f)$ with a different scale parameter j_* ; (b) the linear wavelet estimate $\hat{f}_n^{\text{lin}}(x)$; (c) $\text{MSE}(\hat{f}_n^{\text{non}}, f)$ with a different thresholding parameter λ ; (d) the nonlinear wavelet estimate $\hat{f}_n^{\text{non}}(x)$.

Example 2. We choose the density functions $f_2(x) = (x - \mathbb{I}_{\{x \leq 0.3\}})^2 + 0.9$, $h(x) = 0.5x^2 + 0.5 \cos 2x + 0.53$, $x \in [0, 1]$. The optimal scale parameter $j_* = 10$ can be obtained from Figure 2(a). Figure 2(c) identifies the optimal threshold parameter as $\lambda = 0.11391304$. The estimation results for the linear wavelet estimator are depicted in Figure 2(b). The corresponding results for the nonlinear wavelet estimator are shown in Figure 2(d). Comparative analysis reveals that, when compared to the linear estimator, the nonlinear wavelet estimator demonstrates superior performance, particularly at discontinuity points.

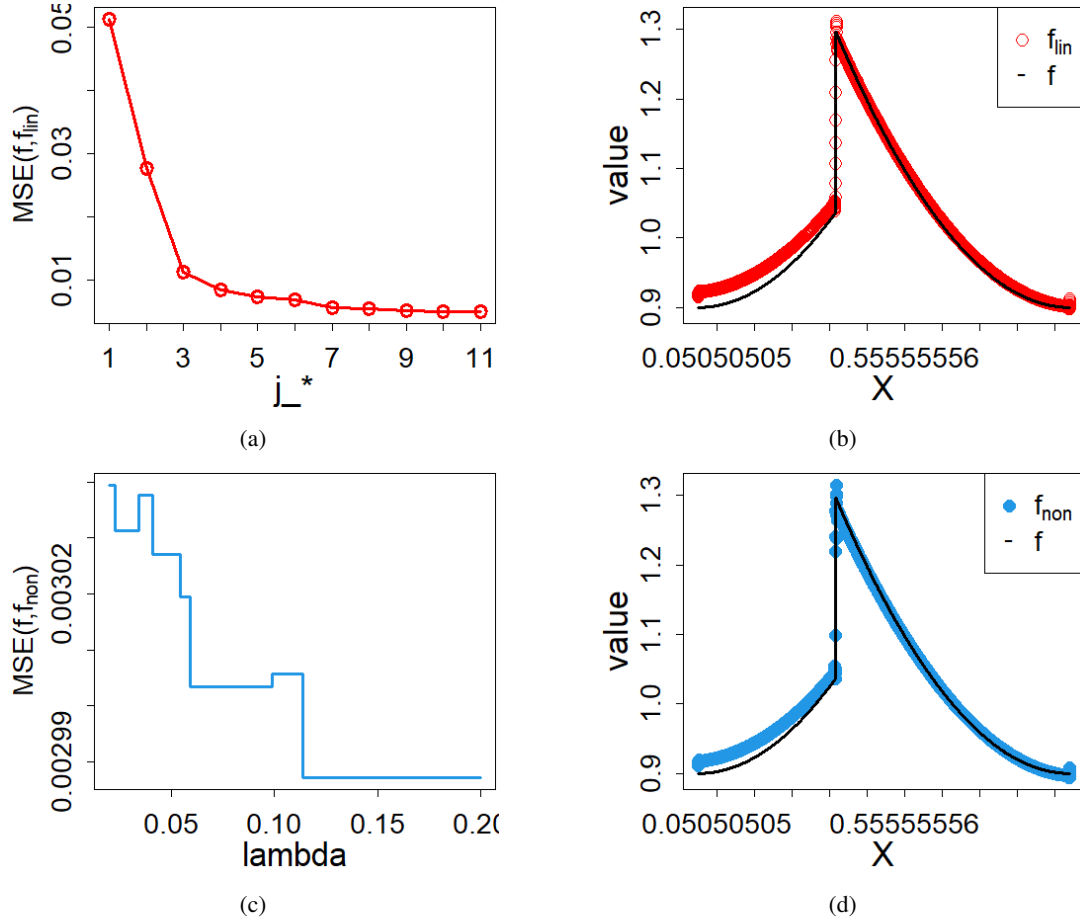


Figure 2: The estimates of $f_2(x)$.

Example 3. In the context of model (1), we consider two density functions $f_3(x) = 1.6e^{-200(x-0.75)^2} + 2e^{-50(x-0.33)^2}$ and $h(x) = 0.9x^2 + 0.35 \cos 2x + 0.55$, with $x \in [0, 1]$. The optimal scale parameter $j_* = 6$ can be obtained from Figure 3(a). Furthermore, the optimal threshold parameter is determined to be $\lambda = 0.0030100334$, as shown in Figure 3(c). Figures 3(b) and 3(d) show that the wavelet estimators employed are capable of effectively approximating the unknown density function.

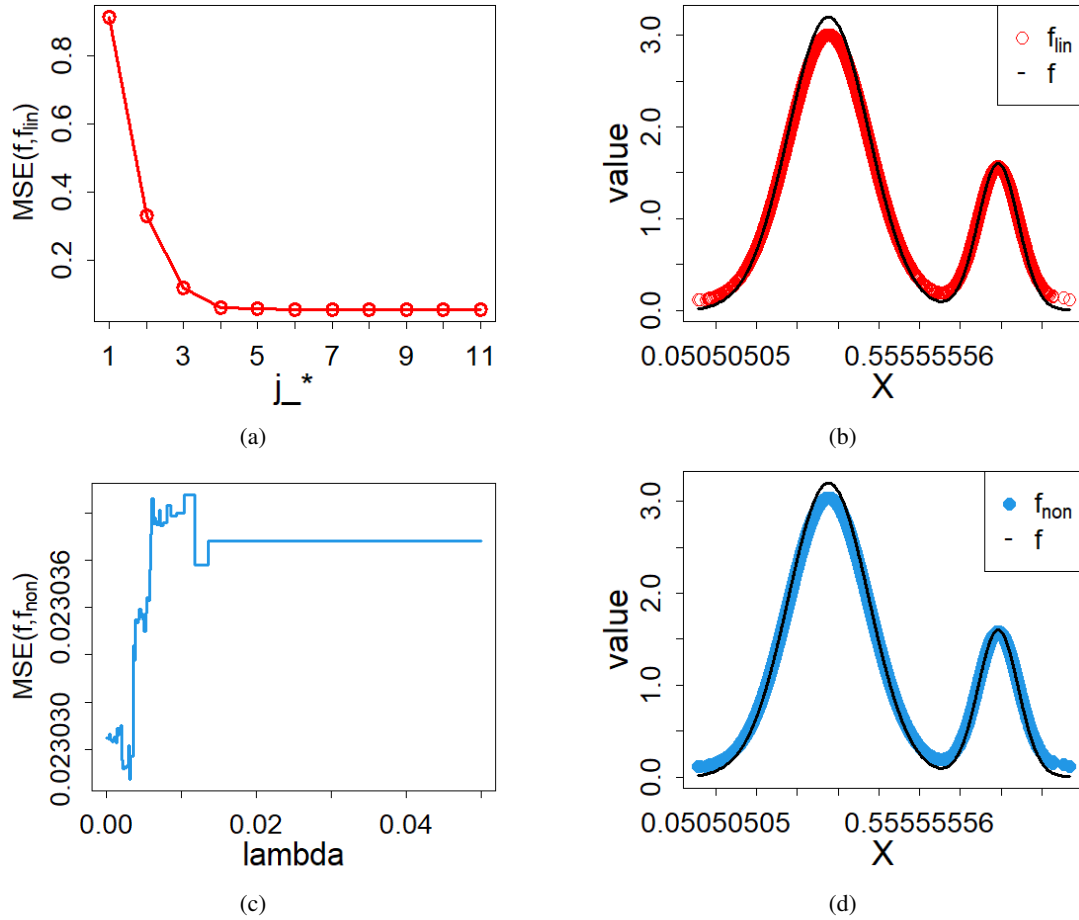


Figure 3: The estimates of $f_3(x)$.

Example 4. We choose the density functions $f_4(x) = e^{-x^2} \sin^2 x$ and $h(x) = 0.1x^2 + 0.6 \cos 2x + 0.11$, $x \in [-2.5, 2.5]$. The optimal scale parameter $j_* = 5$ can be obtained from Figure 4(a). The optimal threshold parameter, denoted as λ , is determined to be 0.00301003344, as depicted in Figure 4(c). The corresponding estimation results are displayed in Figure 4(b) for the linear wavelet estimator, and in Figure 4(d) for the nonlinear wavelet estimator.

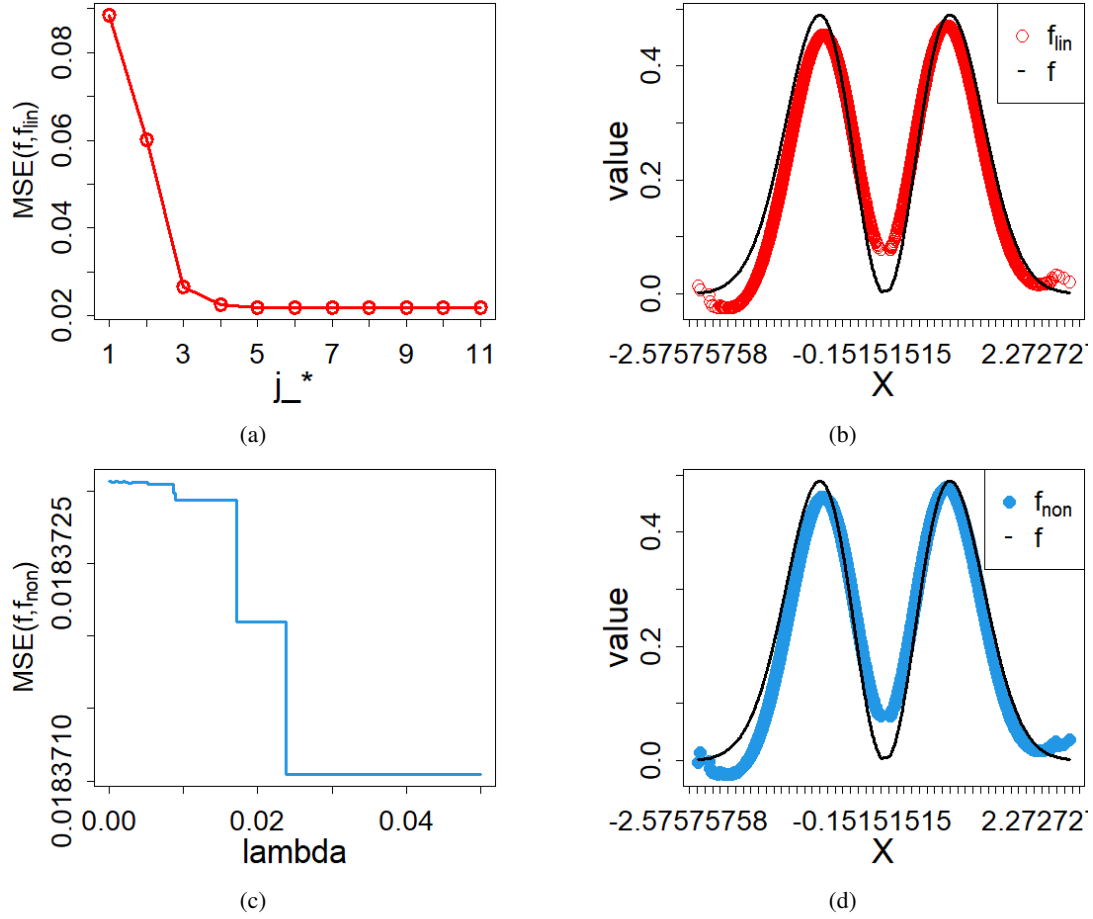


Figure 4: The estimates of $f_4(x)$.

Example 5. In the context of model (1), we choose the density functions $f_5(x) = (4 \sin 4\pi x - \text{sign}(x + 0.12) - \text{sign}(0.22 - x))/5 + 1.2$ and $h(x) = 0.6x^2 + 0.5 \cos 2x + 0.53$, $x \in [-0.5, 0.5]$. Figures 5(a) and 5(c) demonstrate that the optimal parameters are determined to be $j_* = 8$ for the scale parameter, and $\lambda = 0.0237458194$ for the threshold parameter. Under those optimal parameters, two wavelet estimators have good performance.

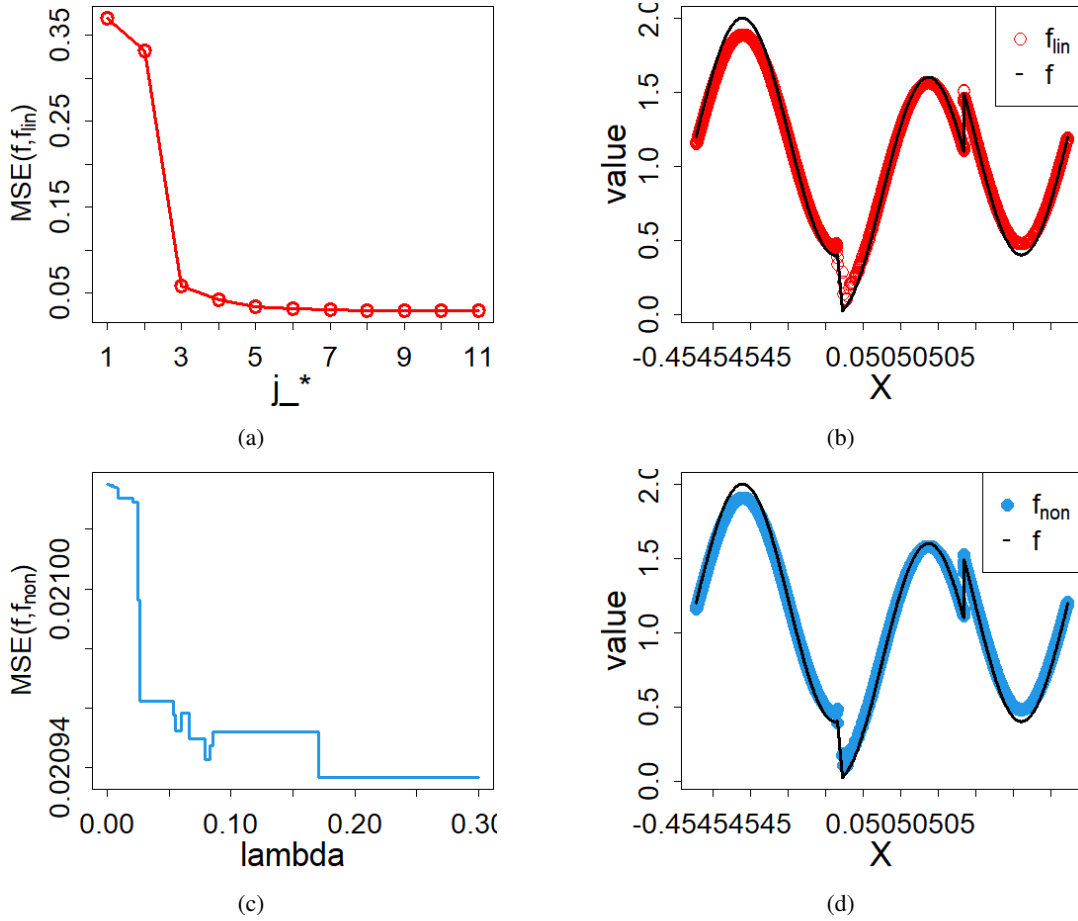


Figure 5: The estimates of $f_5(x)$.

Example 6. We choose the density functions $f_6(x) = 8(2-x)e^{-4(2-x)^2}$, $h(x) = 0.2x^2 + 0.35 \cos 2x + 0.51$, and $x \in [0.5, 2]$. The optimal scale parameter $j_* = 6$ can be obtained from Figure 6(a). The optimal threshold parameter is $\lambda = 0.0284280936$, as depicted in Figure 6(c). Figures 6(b) and 6(d) show that both wavelet estimators can effectively approximate the target density function.

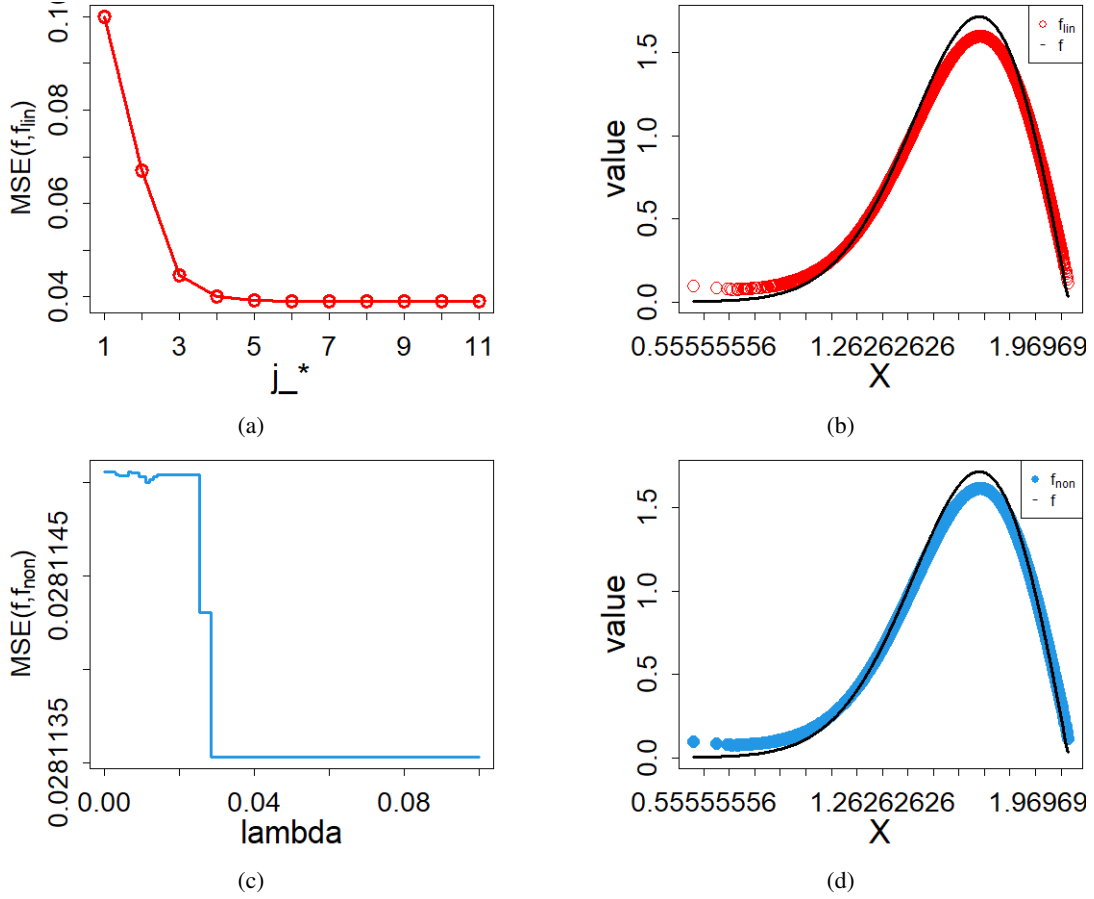


Figure 6: The estimates of $f_6(x)$.

The following Table 1 shows the optimal scale parameters, the optimal threshold parameters, and the MSE of two wavelet estimators in different examples. For more simulation studies, we choose a different sample size n to observe the performance of two wavelet estimators. According to Table 2, it is easy to see that the MSE of two wavelet estimators becomes smaller when the sample size n is larger. In addition, the nonlinear wavelet estimators \tilde{f}_n perform better than the linear wavelet estimators \hat{f}_n .

	f_1	f_2	f_3
j_*	5	10	6
λ	0.0478260870	0.11391304	0.0030100334
$MSE(\hat{f}_n^{\text{lin}}, f)$	0.01949769	0.00514771	0.03986530
$MSE(\hat{f}_n^{\text{non}}, f)$	0.01466935	0.00298714	0.02302871
	f_4	f_5	f_6
\tilde{j}_*	5	8	6
$\tilde{\lambda}$	0.0237458194	0.170568562	0.0284280936
$MSE(\tilde{f}_n^{\text{lin}}, f)$	0.02185451	0.02971702	0.03902024
$MSE(\tilde{f}_n^{\text{non}}, f)$	0.01837104	0.02093666	0.02811353

Table 1: The MSE of the wavelet estimator.

n	128	256	512	1024	2048	4096	8192	16384	
1	$MSE(\hat{f}_n, f)$	0.02004337	0.01969370	0.01967366	0.01965352	0.01963825	0.01949769	0.01938861	0.01934151
	$MSE(\tilde{f}_n, f)$	0.01531049	0.017234	0.014139	0.012880	0.01479997	0.01466935	0.01466550	0.01464125
2	$MSE(\hat{f}_n, f)$	0.00560728	0.00517845	0.00516828	0.00516463	0.00515295	0.00514771	0.00501177	0.00500330
	$MSE(\tilde{f}_n, f)$	0.00329194	0.00326762	0.00310774	0.00309442	0.00301939	0.00298714	0.00297252	0.00295634
3	$MSE(\hat{f}_n, f)$	0.04774499	0.04085642	0.03986695	0.03974946	0.03953559	0.03949926	0.03949290	0.03936086
	$MSE(\tilde{f}_n, f)$	0.02323339	0.02308795	0.02308036	0.02307001	0.02306140	0.02302871	0.022824	0.02281976
4	$MSE(\hat{f}_n, f)$	0.02285294	0.02189756	0.02189058	0.02188800	0.02188289	0.02185451	0.02162162	0.02151109
	$MSE(\tilde{f}_n, f)$	0.01921598	0.01861709	0.01860857	0.01850824	0.01839866	0.01837104	0.01827369	0.01809206
5	$MSE(\hat{f}_n, f)$	0.036478	0.03209222	0.03113655	0.03061721	0.02962667	0.02971702	0.02965899	0.02965899
	$MSE(\tilde{f}_n, f)$	0.02417582	0.02267460	0.02257770	0.02212211	0.02151516	0.02093666	0.02088335	0.02086120
6	$MSE(\hat{f}_n, f)$	0.03968358	0.03950713	0.03935347	0.03927934	0.03927852	0.03902024	0.038165	0.03792818
	$MSE(\tilde{f}_n, f)$	0.02851510	0.02843545	0.028391	0.02832763	0.02821251	0.02811353	0.02774405	0.02690849

Table 2: Estimation results of two wavelet estimators with a different sample size n .

Acknowledgements

The authors would like to thank the reviewers for their helpful and constructive comments that contributed to improving the original version of this paper. This paper is supported by the National Natural Science Foundation of China (No. 12361016), Guangxi Natural Science Foundation (No. 2024GXNSFBA010379), Center for Applied Mathematics of Guangxi (GUET), Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation.

References

- [1] F. ABRAMOVICH, T. C. BAILEY, and T. SAPATINAS, [Wavelet analysis and its statistical applications](#), *J. Roy. Statist. Soc. Ser. D.* **49** (2000), no. 1, 1–29.
- [2] P. BALDI, G. KERKYACHARIAN, D. PICARD, and A. B. TSYBAKOV, [Adaptive density estimation for directional data using needlets](#), *Ann. Statist.* **37** (2009), no. 6A, 3362–3395.
- [3] F. BERTOLINO, S. CABRAS, E. M. CASTELLANOS, and W. RACUGNO, [Unscaled Bayes factors for multiple hypothesis testing in microarray experiments](#), *Stat. Methods Med. Res.* **24** (2015), no. 6, 1030–1043.

- [4] S. H. BODHA, G. JITI, M. J. SILVAPULLE, and P. SILVAPULLE, [Forecasting a nonstationary time series using a mixture of stationary and nonstationary factors as predictors](#), *J. Bus. Econom. Statist.* **42** (2024), no. 1, 122–134.
- [5] K. K. CAO and X. C. ZENG, [Data-driven wavelet estimations for density derivatives](#), *Bull. Malays. Math. Sci. Soc.* **47** (2024), no. 6, 169–187.
- [6] G. CHAGNY, A. CHANNAROND, V. H. HOANG, and A. ROCHE, [Adaptive nonparametric estimation of a component density in a two-class mixture model](#), *J. Statist. Plann. Inference.* **216** (2022), 51–69.
- [7] M. CHEN, C. GAO, and Z. REN, [A general decision theory for Huber’s \$\epsilon\$ -contamination model](#), *Electron. J. Stat.* **10** (2016), no. 2, 3752–3774.
- [8] C. CHESNEAU, I. DEWAN, and H. DOOSTI, [Nonparametric estimation of a two dimensional continuous–discrete density function by wavelets](#), *Statist. Methodol.* **18** (2014), 64–78.
- [9] C. CHESNEAU, S. E. KOLEI, J. K. KOU, and F. NAVARRO, [Nonparametric estimation in a regression model with additive and multiplicative noise](#), *J. Comput. Appl. Math.* **380** (2020), 112971.
- [10] D. L. DONOHO, I. M. JOHNSTONE, G. KERKYACHARIAN, and D. PICARD, [Density estimation by wavelet thresholding](#), *Ann. Statist.* **24** (1996), no. 2, 508–539.
- [11] B. EFRON, R. TIBSHIRANI, J. D. STOREY, and V. TUSHER, [Empirical Bayes analysis of a microarray experiment](#), *J. Am. Stat. Assoc.* **96** (2001), no. 456, 1151–1160.
- [12] Z. GENG and J. WANG, [The mean consistency of wavelet density estimators](#), *J. Inequal. Appl.* **2015** (2015), no. 1, 1–14.
- [13] P. J. HUBER, [A robust version of the probability ratio test](#), *Ann. Math. Stat.* **36** (1965), no. 6, 1753–1758.
- [14] W. HÄRDLE, G. KERKYACHARIAN, D. PICARD, and A. TSYBAKOV, [Wavelets, Approximation, and Statistical Applications](#), Springer, New York, 1998.
- [15] A. JUDITSKY and S. LAMBERT-LACROIX, [On minimax density estimation on \$\mathbb{R}\$](#) , *Bernoulli.* **10** (2004), no. 2, 187–220.
- [16] J. K. KOU and X. M. CHEN, [Wavelet estimations of a density function in two-class mixture model](#), *AIMS Math.* **9** (2024), no. 8, 20588–20611.
- [17] D. LIANG and J. K. KOU, [Nonparametric density estimation in a mixed model using wavelet](#), *Axioms.* **14** (2025), 741.
- [18] H. LIU and C. GAO, [Density estimation with contamination: minimax rates and theory of adaptation](#), *IMS-BS.* **13** (2019), no. 2, 3613–3653.
- [19] S. Y. LIU and T. YU, [Kernel density estimation in mixture models with known mixture proportions](#), *Stat. Med.* **40** (2021), no. 28, 6360–6372.
- [20] H. Q. MA and J. M. JIANG, [Pseudo-Bayesian classified mixed model prediction](#), *J. Amer. Statist. Assoc.* **118** (2023), no. 543, 1747–1759.
- [21] R. MAIBORODA and O. SUGAKOVA, [Nonparametric density estimation for symmetric distributions by contaminated data](#), *Metrika.* **75** (2012), no. 1, 109–126.
- [22] E. C. J. MARIA, I. SALAZAR, L. SANZ, and M. GÓMEZ-VILLEGAS, [Using copula to model dependence when testing multiple hypotheses in DNA microarray experiments: A Bayesian approximation](#), *Mathematics.* **8** (2020), no. 9, 1514–1514.
- [23] D. M. MARZIO and C. C. TAYLOR, [Boosting kernel density estimates: A bias reduction technique?](#), *Biometrika.* **91** (2004), no. 1, 226–233.
- [24] G. McLACHLAN and D. PEEL, [Finite Mixture Models](#), Wiley, New York, 2000.
- [25] G. A. MOSTOFI and K. M. KOPAEI, [Nonparametric Bayesian inferences on the skewed data using a Dirichlet process mixture model](#), *Statist. Papers.* **66** (2025), no. 1, 28–52.
- [26] V. H. NGUYEN and C. MATIAS, [Nonparametric estimation of the density of the alternative hypothesis in a multiple testing setup. Application to local false discovery rate estimation](#), *ESAIM: Probab. Stat.* **18** (2014), no. 1, 584–612.
- [27] R. K. PATRA and B. SEN, [Estimation of a two-component mixture model with applications to multiple testing](#), *J. R. Stat. Soc.* **78** (2016), no. 4, 869–893.
- [28] S. ROBIN, A. BAR-HEN, J. J. DAUDIN, and L. PIERRE, [A semi-parametric approach for mixture models: Application to local false discovery rate estimation](#), *Comput. Stat. Data Anal.* **51** (2007), no. 12, 5483–5493.
- [29] A. T. SHEN, X. LI, Y. J. ZHANG, and X. J. ZHANG, [Consistency and asymptotic normality of wavelet estimator in a nonparametric regression model](#), *Stochastics.* **93** (2021), no. 6, 868–885.

- [30] H. SHU, B. NAN, and R. KOEPPE, [Multiple testing for neuroimaging via hidden Markov random field](#), *Biometrics*. **71** (2015), no. 3, 741–750.
- [31] A. B. TSYBAKOV, *Introduction to Nonparametric Estimation*, Springer, New York, 2008.
- [32] H. Y. WANG, [Convergence rates of density estimation in Besov spaces](#), *Appl. Math.* **02** (2011), no. 10, 1258–1262.
- [33] A. M. WINKLER, M. A. WEBSTER, J. C. BROOKS, I. TRACEY, S. SMITH, and T. NICHOLS, [Non-parametric combination and related permutation tests for neuroimaging](#), *Hum. Brain Mapp.* **37** (2016), no. 4, 1486–1511.
- [34] L. YANG, W. FAN, and N. BOUGUILA, [Clustering analysis via deep generative models with mixture models](#), *IEEE Trans. Neural Netw. Learn. Syst.* **33** (2022), no. 1, 340–350.
- [35] J. ZHANG, B. Q. LIN, and Y. ZHOU, [Kernel density estimation for partial linear multivariate responses models](#), *J. Multivariate Anal.* **185** (2021), 104768.