AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(9): 28628–28645.
DOI: 10.3934/math.20261139
Received: 18 July 2026
Revised: 24 August 2026
Accepted: 01 September 2026
Published: 07 September 2026

Research article

Oracle inequalities and adaptive wavelet estimations for a mixture density model

Kaikai Cao *

School of Mathematics and Statistics, Weifang University, Weifang 261061, China

* **Correspondence:** Email: caokaikaiwfu@163.com.

Abstract: This paper proposes a data-driven wavelet estimator for a mixture density model, in the spirit of the Goldenshluger and Lepski method. In addition, we investigate the fully adaptive estimations for multivariate density functions. First, a pointwise oracle inequality is given, which does not require any assumptions on the underlying function. Moreover, the pointwise estimation under the local Hölder condition and L^p risk ($1 \leq p < \infty$) estimations on Besov spaces are investigated, respectively.

Keywords: wavelet; nonparametric estimation; mixture model; density function; data-driven
Mathematics Subject Classification: 42C40, 62G07, 62G20

1. Introduction

The estimation of a probability density from independent and identically distributed (i.i.d.) random observations $X_1, X_2, \ldots, X_n$ is a classical problem in statistics. Numerous studies have been devoted to the problem of nonparametric density estimation under various statistical models. In this paper, we consider a mixture density model [14,26]. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with a common density $g(x)$ defined by

$$g(x) = \theta h(x) + (1-\theta) f(x), \;\; x \in \mathbb{R}^d. \tag{1.1}$$

In the above equation, $h(x)$ and $f(x)$ are both bounded density functions, and the function $h(x)$ is known. The parameter $\theta \in (0, 1)$ is a known mixing ratio. The problem is how to estimate the unknown density $f(x)$ from the observations $X_1, X_2, \ldots, X_n$ in some sense.

In particular, when h is the density of the uniform distribution on [0,1], the model (1.1) reduces to

$$g(x) = \theta + (1-\theta) f(x), \;\; x \in [0, 1], \tag{1.2}$$

which is known as a two-class mixture model; see [5, 13].

As a powerful statistical tool, the above mixture model plays a significant role in statistics, economics, big data processing, and other fields; see [10, 19, 25]. In the context of dealing with contamination problems [23], model (1.1) describes the case where the density function $f(x)$ of an unknown distribution is contaminated by an arbitrary distribution $h(x)$ with a proportion θ, which constitutes a particular instance of the pollution model [12]. In addition, Maiboroda and Sugakova [20], Chen et al. [6], and Liu and Gao [16] 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; see [1, 22, 27].

The kernel estimation serves as a prevalent tool for nonparametric mixture density inference. Focusing on the extended version of Model (1.2), where the mixing parameter θ is treated as an unknown quantity, Nguyen and Matias [26] provided a randomly weighted kernel estimator and established an upper bound for its pointwise quadratic risk on a class of Hölder density. Furthermore, Chagny et al. [5] investigated the same problem by constructing a data-driven kernel estimator.

Compared with kernel estimators, wavelet estimators provide more local information, which is effective for the estimation of density function with cusps, as they have the properties of time-frequency localization and multiresolution analysis (see [7, 8, 21]). In 2024, Kou and Chen [13] constructed linear and adaptive nonlinear wavelet estimators for Model (1.2) and established upper bounds for their $L^p(1 \le p < \infty)$ risks over Besov spaces. For model (1.1), Liang and Kou [15] employed wavelet techniques to develop similar estimators and investigated their convergence rates in terms of mean integrated squared error in 2025. One year later, Kou and Liang [14] established the minimax lower bound over $L^p(1 \le p < \infty)$ loss and derived corresponding upper bounds.

In contrast to the traditional adaptive estimations, Goldenshluger and Lepski [9] developed a data-driven kernel estimator for the classical density framework in 2014. The parameters of their estimator are determined by a fully adaptive selection rule. In 2019, Liu and Wu [18] introduced a data-driven wavelet estimator and considered pointwise density estimations under an anisotropic local Hölder condition. In 2021, Cao and Zeng [4] investigated L^p risk $(1 \le p < \infty)$ estimations under the independence hypothesis by using the data-driven wavelet estimator based on the classical density model. In 2024, Cao and Zeng [2] proposed a data-driven wavelet estimation method for estimating the derivatives of the density function. Furter, more relevant studies with data-driven estimation can be found in [3, 17].

The objective of this paper is to construct a data-driven wavelet estimator for the mixture density model (1.1) and analyze its performance under pointwise and L^p risks. The main contributions of this paper are as follows. We construct a fully adaptive data-driven wavelet estimator and establish a pointwise oracle inequality that imposes no regularity assumptions on the underlying function f (except for the restrictions ensuring the existence of the model and the risk). Moreover, using this inequality, we derive the pointwise estimations under a local Hölder condition and L^p risk $(1 \le p < \infty)$ estimations on Besov spaces, respectively.

This paper is organized as follows. Section 2 introduces the required function spaces and constructs a data-driven wavelet estimator. Section 3 establishes the corresponding oracle inequality. Section 4 presents two auxiliary propositions. Section 5 deals with adaptive estimation. Section 6 provides numerical simulations for the proposed estimator. Finally, Section 7 draws conclusions and discusses possible directions for further research.

2. Function spaces and data-driven estimator

In this section, we introduce the function spaces required in the subsequent analysis and propose a fully adaptive wavelet estimator based on data-driven methods.

2.1. Wavelets and function spaces

We begin with a classical concept in wavelet analysis. A multiresolution analysis (MRA, [24]) is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of the square integrable function space $L^2(\mathbb{R}^d)$ satisfying the following properties:

(i). $V_j \subset V_{j+1},\ j \in \mathbb{Z}$ and $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}^d)$;

(ii). $f(2\cdot) \in V_{j+1}$ if and only if $f(\cdot) \in V_j$ for each $j \in \mathbb{Z}$;

(iii). There exists $\Phi \in L^2(\mathbb{R}^d)$ (scaling function) such that $\{\Phi(\cdot - k),\ k \in \mathbb{Z}^d\}$ forms an orthonormal basis of $V_0 = \overline{\mathrm{span}\{\Phi(\cdot - k)\}}$.

When $d = 1$, a wavelet function Ψ can be constructed from the scaling function Φ in a simple way such that $\{2^{j/2}\Psi(2^j \cdot -k),\ j, k \in \mathbb{Z}\}$ constitutes an orthonormal basis (wavelet basis) of $L^2(\mathbb{R})$. Examples include Daubechies wavelets [11], which have compact supports in time domain. For $d \geq 2$, the tensor product method gives an MRA $\{V_j\}$ of $L^2(\mathbb{R}^d)$ from one-dimensional MRA. In fact, with a scaling function Φ of tensor products, we find $2^d - 1$ wavelet functions Ψ^ℓ $(\ell = 1, 2, \cdots, 2^d - 1)$ such that

$$\{2^{jd/2}\Psi^\ell(2^j \cdot -k),\ \ j \in \mathbb{Z},\ k \in \mathbb{Z}^d,\ \ell = 1, 2, \cdots, 2^d - 1\}$$

constitutes an orthonormal basis (wavelet basis) of $L^2(\mathbb{R}^d)$.

A scaling function Φ is said to be t regular if $\Phi \in C^t(\mathbb{R}^d)$, and

$$|D^\alpha\Phi(x)| \leq c_l(1 + |x|^2)^{-\frac{l}{2}},\ \ \alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d) \in \mathbb{N}^d \tag{2.1}$$

for any $l \in \mathbb{N}$, $|\alpha| = 0, 1, \cdots, t$ and some independent positive constants c_l. Here and after, $|\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_d$, and

$$D^\alpha f := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} f.$$

Then, the corresponding wavelet Ψ^ℓ $(\ell = 1, 2, \cdots, 2^d - 1)$ also satisfies (2.1). The Daubechies scaling function $\underbrace{D_{2N} \times \cdots \times D_{2N}}_{d \text{ times}}$ with $N > t + d$ is an example, and the tensor product of D_{2N} with large N is used in the whole paper.

As usual, let P_j be the orthogonal projection operator from $L^2(\mathbb{R}^d)$ onto the scaling space V_j with the orthonormal basis $\{\Phi_{jk}(\cdot) = 2^{jd/2}\Phi(2^j \cdot -k),\ k \in \mathbb{Z}^d\}$. Then, for each $f \in L^2(\mathbb{R}^d)$,

$$P_j f = \sum_{k\in\mathbb{Z}^d} \alpha_{jk}\Phi_{jk} \tag{2.2}$$

with $\alpha_{jk} := \langle f, \Phi_{jk}\rangle$. Specially, when a scaling function Φ is t regular, the identity (2.2) holds in $L^p(\mathbb{R}^d)$ for $p \geq 1$ ([11]).

One of advantages of wavelet bases is that they can characterize Besov spaces, which are needed in order to establish the L^p risk estimations.

Lemma 2.1 ([11,24]). *Let Φ be t regular, Ψ^ℓ ($\ell = 1, 2, \cdots, 2^d - 1$) be the corresponding wavelets, and $f \in L^r(\mathbb{R}^d)$. If $\alpha_{jk} := \langle f, \Phi_{jk}\rangle$, $\beta^\ell_{jk} := \langle f, \Psi^\ell_{jk}\rangle$, $r, q \in [1, \infty]$, and $0 < s < t$, then the following assertions are equivalent:*

(i). $f \in B^s_{r,q}(\mathbb{R}^d)$;

(ii). $\{2^{js}\|P_j f - f\|_r\} \in l_q$;

(iii). $\{2^{j(s-\frac{d}{r}+\frac{d}{2})}\|\beta_{j\cdot}\|_{l_r}\} \in l_q$.

The Besov norm of f can be defined by

$$\|f\|_{B^s_{r,q}} := \|\alpha_{j_0\cdot}\|_{l_r} + \|(2^{j(s-\frac{d}{r}+\frac{d}{2})}\|\beta_{j\cdot}\|_{l_r})_{j\geq j_0}\|_{l_q},$$

where $\|\alpha_{j_0\cdot}\|^r_{l_r} := \sum\limits_{k\in\mathbb{Z}^d} |\alpha_{j_0k}|^r$ *and* $\|\beta_{j\cdot}\|^r_{l_r} = \sum\limits_{\ell=1}^{2^d-1} \sum\limits_{k\in\mathbb{Z}^d} |\beta^\ell_{jk}|^r$.

Moreover, Lemma 2.1 (i) and (ii) shows that $\|P_j f - f\|_r \lesssim 2^{-js}$ holds for $f \in B^s_{r,q}(\mathbb{R}^d)$. Here and throughout, the notation $A \lesssim B$ denotes $A \leq cB$ with some fixed and independent constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $A \gtrsim B$.

Note that l_r is continuously embedded to l_p for $r \leq p$. Then, Lemma 2.1 (i) and (iii) imply that

$$B^s_{r,q}(\mathbb{R}^d) \hookrightarrow B^{s'}_{p,q}(\mathbb{R}^d) \text{ for } r \leq p, \text{ and } s' - \frac{d}{p} = s - \frac{d}{r} > 0,$$

where $A \hookrightarrow B$ stands for a Banach space A continuously embedded in another Banach space B.

Subsequently, we introduce a local Hölder condition, which plays a key role in investigating the pointwise estimations.

As in [17, 18], for a univariate function f, the local Hölder condition of order $0 < s \leq 1$ at the point $x_0 \in \mathbb{R}^d$ means that for a fixed constant $L > 0$ and each $x, y \in \Omega_{x_0}$ (a neighborhood of the point x_0),

$$|f(y) - f(x)| \leq L|y - x|^s.$$

We use $H^s(\Omega_{x_0})$ to denote all those functions.

For $s = N + \epsilon$ with $\epsilon \in (0, 1]$ and $N \in \mathbb{N}$, one defines

$$H^s(\Omega_{x_0}) := \{f, D^\alpha f \in H^\epsilon(\Omega_{x_0}),\ |\alpha| = N\}.$$

The following lemma is necessary for the pointwise estimations.

Lemma 2.2 ([17, 28]). *Let $\Phi \in L^2(\mathbb{R}^d)$ be a compactly supported and t regular scaling function and $\Psi^\ell(\ell = 1, 2, \cdots, 2^d - 1)$ be the corresponding wavelet. If $f \in H^s(\Omega_{x_0}) \cap L^2(\mathbb{R}^d)$ with $s > 0$ and $t \geq [s]$, then for $x \in \Omega_{x_0}$ and sufficiently large j,*

(i). $f(x) = \sum\limits_{k\in\mathbb{Z}^d} \alpha_{j_0k}\Phi_{j_0k}(x) + \sum\limits_{j=j_0}^{\infty} \sum\limits_{\ell=1}^{2^{d-1}} \sum\limits_{k\in\mathbb{Z}^d} \beta^\ell_{jk}\Psi^\ell_{jk}(x)$ *holds pointwisely;*

(ii). $\sup\limits_{x\in\Omega_{x_0}} \sup\limits_{f\in H^s(\Omega_{x_0})\cap L^2(\mathbb{R}^d)} |f(x) - P_j f(x)| \lesssim 2^{-js}$.

In the whole paper, the notations $H^s(\Omega_{x_0}, M)$ and $B^s_{r,q}(M, T)$ stand for

$$H^s(\Omega_{x_0}, M) := \{f \in H^s(\Omega_{x_0}), f \text{ is a density function, and } \|f\|_\infty \leq M\}.$$

Further,

$$B^s_{r,q}(M, T) := \{f \in B^s_{r,q}(M),\ \|f\|_\infty \leq M, \text{ and } supp\ f \subseteq [-T, T]^d\}, \tag{2.3}$$

where $B^s_{r,q}(M) := \{f \in B^s_{r,q}(\mathbb{R}^d), f$ is a density function, and $\|f\|_{B^s_{r,q}} \leq M\}$, and $M, T > 0$ are some positive constants.

2.2. Data-driven estimator

As in [14], the linear wavelet estimator for model (1.1) is given by

$$\widehat{f_j}(x) := \sum_{k\in\mathbb{Z}^d} \widehat{\alpha}_{jk}\Phi_{jk}(x), \tag{2.4}$$

where $\widehat{\alpha}_{jk} := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\theta}\Phi_{jk}(X_i) - \mu_{jk}$, and $\mu_{jk} = \int_{\mathbb{R}^d}\frac{\theta}{1-\theta}h(x)\Phi_{jk}(x)dx$. Clearly, $E\widehat{\alpha}_{jk} = \alpha_{jk} := \langle f, \Phi_{jk}\rangle$, and $E\widehat{f_j} = P_j f$; see [14].

Motivated by [3, 9, 17], we introduce a selection rule for parameter j defined in (2.4), which relies exclusively on the observed samples $X_1, \cdots, X_n$.

Let $\mathcal{H} := \left\{0, 1, \cdots, \lfloor\frac{1}{d}\log_2\frac{n}{\ln n}\rfloor\right\}$ with $\lfloor a\rfloor$ denoting the integer part of a. Thus, $j_0 \in \mathcal{H}$ is determined by the following rule:

$$\widehat{R}_j(x) := \sup_{j'\in\mathcal{H}}\left[\left|\widehat{f_{j\wedge j'}}(x) - \widehat{f_{j'}}(x)\right| - \tau_n(j\wedge j') - \tau_n(j')\right]_+, \tag{2.5}$$

$$j_0 = j_0(x) = \operatorname*{arginf}_{j\in\mathcal{H}}\left[\widehat{R}_j(x) + 2\tau_n(j)\right].$$

Here and throughout, $a\wedge b := \min\{a,\ b\}$, $a_+ := \max\{a,\ 0\}$, and

$$\tau_n(j) := \left(\frac{\lambda 2^{jd}\ln n}{n}\right)^{\frac{1}{2}}, \tag{2.6}$$

where $\lambda > 0$ is a constant determined later on. Thus, the data-driven wavelet estimator is obtained by

$$\widehat{f_n}(x) := \widehat{f_{j_0}}(x) = \sum_{k\in\mathbb{Z}^d}\widehat{\alpha}_{j_0k}\Phi_{j_0k}(x). \tag{2.7}$$

3. Oracle inequality

In this section, we provide a pointwise oracle inequality for the data-driven estimator in (2.7), which plays the key roles in the proofs of Theorems 5.1 and 5.2.

We first introduce two fundamental quantities for the wavelet estimator $\widehat{f_j}$. Define

$$B_j(x,f) := |P_jf(x) - f(x)| \quad\text{and}\quad S_n(x,j) := \widehat{f_j}(x) - E\widehat{f_j}(x), \tag{3.1}$$

which correspond to the bias and stochastic error of $\widehat{f_j}$, respectively. Furthermore, we construct two auxiliary supremum functionals based on the above quantities:

$$B_j^*(x,f) := \sup_{j'\in\mathcal{H},\ j'\geq j} B_{j'}(x,f) \quad\text{and}\quad \aleph_n(x) := \sup_{j\in\mathcal{H}}\Big[|S_n(x,j)| - \tau_n(j)\Big]_+, \tag{3.2}$$

where $\tau_n(j)$ is given by (2.6).

Then, the following pointwise oracle inequality holds.

Theorem 3.1. *For any $x\in\mathbb{R}^d$, the estimator $\widehat{f_n}(x)$ defined in (2.7) satisfies that*

$$|\widehat{f_n}(x) - f(x)| \leq \inf_{j\in\mathcal{H}}\left\{5B_j^*(x,f) + 5\tau_n(j)\right\} + 5\aleph_n(x),$$

where $\tau_n(j)$ is given by (2.6), and $B_j^*(x,f)$, $\aleph_n(x)$ are determined by (3.2).

Proof. By the triangle inequality,

$$\left|\widehat{f_{j_0}}(x) - f(x)\right| \le \left|\widehat{f_{j_0\wedge j}}(x) - \widehat{f_{j_0}}(x)\right| + \left|\widehat{f_{j_0\wedge j}}(x) - \widehat{f_j}(x)\right| + \left|\widehat{f_j}(x) - f(x)\right| \tag{3.3}$$

holds for any $x \in \mathbb{R}^d$ and $j \in \mathcal{H}$.

According to (2.5), one obtains that

$$\left|\widehat{f_{j_0\wedge j}}(x) - \widehat{f_j}(x)\right| \le \widehat{R}_{j_0}(x) + \tau_n(j_0 \wedge j) + \tau_n(j), \tag{3.4}$$

and

$$\left|\widehat{f_{j_0\wedge j}}(x) - \widehat{f_{j_0}}(x)\right| = \left|\widehat{f_{j\wedge j_0}}(x) - \widehat{f_{j_0}}(x)\right| \le \widehat{R}_j(x) + \tau_n(j \wedge j_0) + \tau_n(j_0) \tag{3.5}$$

because of the identity $\widehat{f_{j_0\wedge j}} = \widehat{f_{j\wedge j_0}}$.

Note that $\tau_n(j \wedge l) = \min\{\tau_n(j), \tau_n(l)\}$ by (2.6). This with (3.4) and (3.5) concludes

$$\begin{aligned}
&\left|\widehat{f_{j_0\wedge j}}(x) - \widehat{f_{j_0}}(x)\right| + \left|\widehat{f_{j_0\wedge j}}(x) - \widehat{f_j}(x)\right| \\
&\le \widehat{R}_j(x) + 2\tau_n(j_0) + \widehat{R}_{j_0}(x) + 2\tau_n(j) \\
&\qquad \le 2\widehat{R}_j(x) + 4\tau_n(j)
\end{aligned} \tag{3.6}$$

by the selection of j_0 in (2.6).

On the other hand, it follows from (3.1) and (3.2) that

$$\left|\widehat{f_j}(x) - f(x)\right| \le B_j(x, f) + |S_n(x, j)| \le B_j^*(x, f) + \aleph_n(x) + \tau_n(j), \tag{3.7}$$

because $|S_n(x, j)| \le [|S_n(x, j)| - \tau_n(j)]_+ + \tau_n(j) \le \aleph_n(x) + \tau_n(j)$ by (3.2).

Combining (3.3), (3.6), and (3.7) yields the bound

$$\left|\widehat{f_{j_0}}(x) - f(x)\right| \le 2\widehat{R}_j(x) + B_j^*(x, f) + \aleph_n(x) + 5\tau_n(j). \tag{3.8}$$

Next, it suffices to estimate $\widehat{R}_j(x)$. By using (2.5) and (3.1),

$$\begin{aligned}
\widehat{R}_j(x) &= \sup_{j'\in\mathcal{H}} \Big[|\widehat{f_{j\wedge j'}}(x) - \widehat{f_{j'}}(x)| - \tau_n(j \wedge j') - \tau_n(j')\Big]_+ \\
&\le \sup_{j'\in\mathcal{H}} \Big[|E\widehat{f_{j\wedge j'}}(x) - E\widehat{f_{j'}}(x)| + |S_n(x, j \wedge j')| - \tau_n(j \wedge j') + |S_n(x, j')| - \tau_n(j')\Big]_+.
\end{aligned}$$

Clearly, $\sup_{j'\in\mathcal{H}} |E\widehat{f_{j\wedge j'}}(x) - E\widehat{f_{j'}}(x)| \le \sup_{\{j'\in\mathcal{H},\ j'\ge j\}}\{B_{j\wedge j'}(x, f) + B_{j'}(x, f)\}$. Hence,

$$\widehat{R}_j(x) \le 2B_j^*(x, f) + 2\aleph_n(x) \tag{3.9}$$

thanks to (3.2).

Substituting Ineq (3.9) into (3.8), one deduces that

$$\left|\widehat{f_{j_0}}(x) - f(x)\right| \le 5B_j^*(x, f) + 5\aleph_n(x) + 5\tau_n(j)$$

holds for each $j \in \mathcal{H}$ and $x \in \mathbb{R}^d$. Furthermore,

$$\left|\widehat{f_n}(x) - f(x)\right| = \left|\widehat{f_{j_0}}(x) - f(x)\right| \le \inf_{j\in\mathcal{H}} \left\{5B_j^*(x, f) + 5\tau_n(j)\right\} + 5\aleph_n(x)$$

thanks to $\widehat{f_n}(x) = \widehat{f_{j_0}}(x)$ in (2.7). The proof is done. □

Remark 3.1. *As can be seen from the proof, the pointwise oracle inequality of Theorem 3.1 does not require any assumptions on the underlying function. In addition, Theorem 3.1 is the key technical tool for bounding pointwise and L^p risk estimations of the data-driven estimator in (2.7).*

4. Two propositions

This section presents two auxiliary propositions which are needed for proving the main results. To begin, the following lemma and a well-known inequality are needed to prove Proposition 4.1.

Lemma 4.1. *Let $K_j(v,x) := \sum_{k\in\mathbb{Z}^d}[\frac{1}{1-\theta}\Phi_{jk}(v) - \mu_{jk}]\Phi_{jk}(x)$ and Φ be t regular with $t > 0$. Then, for $f \in L^\infty(M)$,*

$$\max\{|K_j(v,x)|,\ E|K_j(X_1,x)|^2\} \le M_1 2^{jd},$$

where $M_1 \ge 1$ is some constant.

Proof. From the regularity property of Φ, it is known that there exists a function $\Upsilon \in L^1 \cap L^\infty$ satisfying $|\sum_{k\in\mathbb{Z}^d}\Phi(x-k)\Phi(y-k)| \le \Upsilon(x-y)$. Moreover, according to the definition of $K_j(v,x)$, one obtains

$$\begin{aligned}
|K_j(v,x)| &\le \Big|\sum_{k\in\mathbb{Z}^d}\frac{1}{1-\theta}\Phi_{jk}(v)\Phi_{jk}(x)\Big| + \Big|\sum_{k\in\mathbb{Z}^d}\mu_{jk}\Phi_{jk}(x)\Big| \\
&\le \frac{1}{1-\theta}2^{jd}\Big|\sum_{k\in\mathbb{Z}^d}\Phi(2^jv-k)\Phi(2^jx-k)\Big| + 2^{jd}\Big|\sum_{k\in\mathbb{Z}^d}\int_{\mathbb{R}^d}\frac{\theta}{1-\theta}h(t)\Phi(2^jt-k)dt\Phi(2^jx-k)\Big| \\
&\le \frac{1}{1-\theta}2^{jd}\Big|\sum_{k\in\mathbb{Z}^d}\Phi(2^jv-k)\Phi(2^jx-k)\Big| + \frac{\theta}{1-\theta}2^{jd}\int_{\mathbb{R}^d}h(t)\Big|\sum_{k\in\mathbb{Z}^d}\Phi(2^jt-k)\Phi(2^jx-k)\Big|dt \\
&\le \frac{1}{1-\theta}\|\Upsilon\|_\infty 2^{jd} \qquad (4.1)
\end{aligned}$$

thanks to $0<\theta<1$.

On the other hand, $\|g\|_\infty \le \theta\|h\|_\infty + (1-\theta)\|f\|_\infty \le \max\{\|h\|_\infty, M\}$ by model (1.1) and $f \in L^\infty(M)$. Hence,

$$\begin{aligned}
&E|K_j(X_1,x)| = \int_{\mathbb{R}^d}|K_j(v,x)|g(v)dv \\
\le\ & \frac{1}{1-\theta}\int_{\mathbb{R}^d}\Big|\sum_{k\in\mathbb{Z}^d}\Phi_{jk}(v)\Phi_{jk}(x)\Big|g(v)dv + \Big|\sum_{k\in\mathbb{Z}^d}\mu_{jk}\Phi_{jk}(x)\Big| \\
\le\ & \frac{2^{jd}}{1-\theta}\int_{\mathbb{R}^d}\Big|\sum_{k\in\mathbb{Z}^d}\Phi(2^jv-k)\Phi(2^jx-k)\Big|g(v)dv + \frac{\theta 2^{jd}}{1-\theta}\int_{\mathbb{R}^d}h(t)\Big|\sum_{k\in\mathbb{Z}^d}\Phi(2^jt-k)\Phi(2^jx-k)\Big|dt \\
\le\ & \frac{1}{1-\theta}\|\Upsilon\|_1(\max\{\|h\|_\infty, M\} + \|h\|_\infty)
\end{aligned}$$

by $0<\theta<1$. This with (4.1) obtains

$$E|K_j(X_1,x)|^2 \le \frac{1}{(1-\theta)^2}\|\Upsilon\|_\infty\|\Upsilon\|_1(\max\{\|h\|_\infty, M\} + \|h\|_\infty)2^{jd}. \qquad (4.2)$$

Choosing $M_1 := \max\{\frac{1}{1-\theta}\|\Upsilon\|_\infty,\ \frac{1}{(1-\theta)^2}\|\Upsilon\|_\infty\|\Upsilon\|_1(\max\{\|h\|_\infty, M\} + \|h\|_\infty),\ 1\}$, the desired conclusion follows from (4.1) and (4.2). The proof is completed. □

Bernstein's inequality ([11]). *Let $Y_1, \cdots, Y_n$ be i.i.d. random variables, $EY_i = 0$, and $|Y_i| \le M$ $(i = 1, 2, \cdots, n)$. Then, for any $\gamma > 0$,*

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| \ge \gamma\right\} \le 2\exp\Big\{-\frac{n\gamma^2}{2(EY_i^2 + M\gamma/3)}\Big\}.$$

Now, we introduce the first proposition.

Proposition 4.1. *Let $f \in L^\infty(M)$ and Φ be t regular with $t > 0$. Then, for each $x \in \mathbb{R}^d$ and $p \in [1, \infty)$, there exists $\lambda > 6M_1^2 p^2$ such that*

$$E[\aleph_n(x)]^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}},$$

where $\aleph_n(x)$ is given by (3.2), and $M_1 \ge 1$ is the constant in Lemma 4.1.

Proof. For each $j \in \mathcal{H}$, one introduces the quantity

$$\overline{\tau_n(j)} := \left(\frac{6M_1^2 p 2^{jd}\lambda_j}{n}\right)^{\frac{1}{2}}, \tag{4.3}$$

where $\lambda_j := \max\{dpj\ln 2,\ 1\}$. Clearly, the inequality $\lambda \ln n \ge 6M_1^2 p\lambda_j$ holds for large n, as $\lambda > 6M_1^2 p^2$, and $j \in \mathcal{H}$. Then, $\overline{\tau_n(j)} \le \tau_n(j)$ due to (2.6) and (4.3). Furthermore,

$$\Big[|S_n(x, j)| - \tau_n(j)\Big]_+ \le \Big[|S_n(x, j)| - \overline{\tau_n(j)}\Big]_+. \tag{4.4}$$

For any $t \ge 0$,

$$P\Big\{[|S_n(x, j)| - \overline{\tau_n(j)}\ ]_+ > t\Big\} = P\Big\{|S_n(x, j)| - \overline{\tau_n(j)} > t\Big\}.$$

Therefore,

$$\begin{aligned}
E\Big[|S_n(x, j)| - \overline{\tau_n(j)}\Big]_+^p &= p\int_0^\infty t^{p-1} P\Big\{|S_n(x, j)| - \overline{\tau_n(j)} > t\Big\}\,dt \\
&= p\int_0^\infty t^{p-1} P\Big\{|S_n(x, j)| > t + \overline{\tau_n(j)}\Big\}\,dt \\
&= p[\overline{\tau_n(j)}]^p \int_0^\infty t^{p-1} P\Big\{|S_n(x, j)| > \overline{\tau_n(j)}(t+1)\Big\}\,dt.
\end{aligned} \tag{4.5}$$

According to the definition of $\widehat{f_j}(x)$ and (3.1), one obtains

$$S_n(x, j) = \frac{1}{n}\sum_{i=1}^{n}[K_j(X_i, x) - EK_j(X_i, x)],$$

where $K_j(v, x) = \sum_{k\in\mathbb{Z}^d}[\frac{1}{1-\theta}\Phi_{jk}(v) - \mu_{jk}]\Phi_{jk}(x)$ is given by Lemma 4.1. On the other hand, Lemma 4.1 implies that

$$\max\{|K_j(X_i, x)|,\ E|K_j(X_i, x)|^2\} \le M_1 2^{jd}.$$

Using Bernstein's inequality, one knows

$$P\left\{|S_n(x,j)| > \overline{\tau_n(j)}(t+1)\right\} \leq 2\exp\left\{-\frac{n[\overline{\tau_n(j)}]^2(t+1)^2}{2[M_1 2^{jd} + 2M_1 2^{jd}\overline{\tau_n(j)}(t+1)/3]}\right\}. \tag{4.6}$$

Note that $\overline{\tau_n(j)} = \left(\frac{6M_1^2 p 2^{jd}\lambda_j}{n}\right)^{\frac{1}{2}} \leq 3M_1 p$ for $j \in \mathcal{H}$ and large n. Thus,

$$2[M_1 2^{jd} + 2M_1 2^{jd}\overline{\tau_n(j)}(t+1)/3] \leq 6M_1^2 p 2^{jd}(t+1)$$

due to $M_1, p \geq 1$, and $t > 0$. Substituting the above estimate into (4.6), one obtains

$$P\left\{|S_n(x,j)| > \overline{\tau_n(j)}(t+1)\right\} \leq 2e^{-\lambda_j(t+1)} \leq 2e^{-t}e^{-\lambda_j}$$

because of $\lambda_j \geq 1$.

This with (4.5) and $\overline{\tau_n(j)} = \left(\frac{6M_1^2 p 2^{jd}\lambda_j}{n}\right)^{\frac{1}{2}}$ concludes that

$$E\Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+^p \lesssim [\overline{\tau_n(j)}]^p e^{-\lambda_j}\int_0^\infty t^{p-1}e^{-t}dt \lesssim \left(\frac{6M_1^2 p 2^{jd}\lambda_j}{n}\right)^{\frac{p}{2}} e^{-\lambda_j}.$$

Hence, according to $e^{-\lambda_j} \leq 2^{-dpj}$ and $\lambda_j \lesssim \ln n$ for large n, one knows

$$E[\aleph_n(x)]^p \lesssim \sum_{j\in\mathcal{H}} E\Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+^p \lesssim \sum_{j\in\mathcal{H}}\left(\frac{\ln n}{n}\right)^{\frac{p}{2}} 2^{-\frac{dpj}{2}} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}}$$

thanks to (3.2) and (4.4). This finishes the proof of Proposition 4.1. □

Before giving another proposition, we need the following notations. Denote

$$\mathfrak{M}(x,f) := \inf_{j\in\mathcal{H}}\{B_j^*(x,f) + \tau_n(j)\}, \tag{4.7}$$

$$\Lambda_m := \{x \in [-T,T]^d,\ 2^m\delta_n < \mathfrak{M}(x,f) \leq 2^{m+1}\delta_n\}, \tag{4.8}$$

where $\delta_n = (\frac{C\ln n}{n})^{\frac{s}{2s+d}}$, $C > 1$ is some constant, and $T > 0$ is given by (2.3).

Note that $\mathfrak{M}(x,f) \leq c_0 := \sup_x \mathfrak{M}(x,f)$ if Φ is t regular and if $\|f\|_\infty \lesssim 1$. Then, there exists

$$m_2 := \min\{m \in \mathbb{Z},\ 2^m\delta_n \geq c_0\} \tag{4.9}$$

such that $\Lambda_m = \emptyset$ for each $m > m_2$. Clearly, $m_2 > 0$ for large n.

Proposition 4.2. *Let $f \in B_{r,q}^s(M)$ and Φ be t regular with $t > 0$. Then, for $m \in \mathbb{Z}$ satisfying $0 \leq m \leq m_2$ and each $p \in [1,\infty)$,*

$$\int_{\Lambda_m}[\mathfrak{M}(x,f)]^p dx \lesssim 2^{m(p-r-\frac{2sr}{d})}\delta_n^p.$$

Moreover, if $s > \frac{d}{r}$ and if $r \leq p$, then with $s' := s - \frac{d}{r} + \frac{d}{p}$,

$$\int_{\Lambda_m}[\mathfrak{M}(x,f)]^p dx \lesssim 2^{-\frac{2ms'p}{d}}\delta_n^{\frac{s'}{s}p},$$

where $\mathfrak{M}(x,f)$ and Λ_m are defined in (4.7) and (4.8), respectively.

Proof. One chooses $j_1 \in \mathbb{Z}$ satisfying

$$c_1 2^{2m}\delta_n^{-\frac{d}{s}} \le 2^{j_1 d} \le c_2 2^{2m}\delta_n^{-\frac{d}{s}},$$

where two positive constants c_1, c_2 satisfy

$$(2M)^{\frac{d}{s}} I_{\{r=\infty\}} < c_1 < c_2 < \min\left\{\frac{C}{4c_0^2},\ \frac{C}{4\lambda}\right\}.$$

It is easy to see that $2^{m_2} \le 2c_0\delta_n^{-1}$ from (4.9). Then,

$$1 < c_1\delta_n^{-\frac{d}{s}} \le 2^{j_1 d} \le c_2 2^{2m_2}\delta_n^{-\frac{d}{s}} \le 4c_2c_0^2\delta_n^{-(\frac{d}{s}+2)} < \frac{n}{\ln n}$$

due to $0 \le m \le m_2$ and $c_2 < \frac{C}{4c_0^2}$. Hence, $j_1 \in \mathcal{H}$.

On the other hand, according to $\delta_n = (\frac{C\ln n}{n})^{\frac{s}{2s+d}}$ and $c_2 < \frac{C}{4\lambda}$,

$$\begin{aligned} \tau_n(j_1) &= \left(\frac{\lambda 2^{j_1 d}\ln n}{n}\right)^{\frac{1}{2}} \le \left(\frac{\lambda c_2 2^{2m}\delta_n^{-\frac{d}{s}}\ln n}{n}\right)^{\frac{1}{2}} = \left(\frac{\lambda c_2 2^{2m}\delta_n^{-\frac{d}{s}}\delta_n^{2+\frac{d}{s}}}{C}\right)^{\frac{1}{2}} \\ &= (\lambda c_2 C^{-1})^{\frac{1}{2}} 2^m\delta_n \le 2^{m-1}\delta_n. \end{aligned} \tag{4.10}$$

Recall that $\Lambda_m = \{x \in [-T,T]^d,\ 2^m\delta_n < \mathfrak{M}(x,f) \le 2^{m+1}\delta_n\}$. Then,

$$\int_{\Lambda_m} [\mathfrak{M}(x,f)]^p dx \le (2^{m+1}\delta_n)^p|\Lambda_m|, \tag{4.11}$$

where $|\Lambda_m|$ stands for the Lebesgue measure of the set Λ_m. Furthermore, according to the definition of $\mathfrak{M}(x,f)$, $j_1 \in \mathcal{H}$, and (4.10), one obtains that

$$\begin{aligned} |\Lambda_m| &\le |\{x \in [-T,T]^d,\ B_{j_1}^*(x,f) > 2^{m-1}\delta_n\}| + |\{x \in [-T,T]^d,\ \tau_n(j_1) > 2^{m-1}\delta_n\}| \\ &= |\{x \in [-T,T]^d,\ B_{j_1}^*(x,f) > 2^{m-1}\delta_n\}|. \end{aligned} \tag{4.12}$$

When $r = \infty$, it follows from $f \in B_{r,q}^s(M)$ and $m \ge 0$ that

$$B_{j_1}^*(x,f) = \sup_{j' \ge j_1} B_{j'}(x,f) \le M2^{-j_1 s} \le Mc_1^{-\frac{s}{d}}2^{-\frac{2ms}{d}}\delta_n \le 2^{m-1}\delta_n$$

thanks to the choice of $2^{j_1 d} \ge c_1 2^{2m}\delta_n^{-\frac{d}{s}}$ with $c_1 > (2M)^{\frac{d}{s}}$. Hence, $|\Lambda_m| = 0$ because of (4.12). Therefore,

$$\int_{\Lambda_m} [\mathfrak{M}(x,f)]^p dx \le (2^{m+1}\delta_n)^p|\Lambda_m| = 0$$

by (4.11).

For the case $1 \le r < \infty$, by (3.2), (4.12), Chebyshev's inequality, and $f \in B_{r,q}^s(M)$,

$$|\Lambda_m| \le \sum_{j \in \mathcal{H}, j \ge j_1} |\{x \in [-T,T]^d,\ B_j(x,f) > 2^{m-1}\delta_n\}|$$

$$\leq \sum_{j\in\mathcal{H}, j\geq j_1} \frac{\|B_j(\cdot,f)\|_r^r}{(2^{m-1}\delta_n)^r} \lesssim 2^{-mr}\delta_n^{-r}2^{-j_1 sr}. \tag{4.13}$$

This with (4.11) leads to

$$\int_{\Lambda_m} [\mathfrak{M}(x,f)]^p dx \lesssim (2^{m+1}\delta_n)^p 2^{-mr}\delta_n^{-r}2^{-j_1 sr} \lesssim 2^{m(p-r)}\delta_n^{p-r}2^{-j_1 sr} \lesssim 2^{m(p-r-\frac{2sr}{d})}\delta_n^p$$

due to $2^{j_1} \sim 2^{\frac{2m}{d}}\delta_n^{-\frac{1}{s}}$.

Finally, one discusses the case of $s > \frac{d}{r}$ and $r \leq p$. Note that $f \in B^s_{r,q} \hookrightarrow B^{s'}_{p,q}$ with $s' = s - \frac{d}{r} + \frac{d}{p}$. The same arguments with (4.13) yield

$$|\Lambda_m| \leq \sum_{j\in\mathcal{H}, j\geq j_1} \frac{\|B_j(\cdot,f)\|_p^p}{(2^{m-1}\delta_n)^p} \lesssim 2^{-mp}\delta_n^{-p}2^{-j_1 s' p}.$$

This with (4.11) and $2^{j_1} \sim 2^{\frac{2m}{d}}\delta_n^{-\frac{1}{s}}$ implies that

$$\int_{\Lambda_m} [\mathfrak{M}(x,f)]^p dx \lesssim (2^{m+1}\delta_n)^p 2^{-mp}\delta_n^{-p}2^{-j_1 s' p} \lesssim 2^{-j_1 s' p} \lesssim 2^{-\frac{2ms'p}{d}}\delta_n^{\frac{s'}{s}p}.$$

The proof is thus complete. □

5. Adaptive estimations

This section investigates fully adaptive pointwise and L^p risk ($1 \leq p < \infty$) estimations by the data-driven estimator defined in (2.7). First, we establish and prove the pointwise estimation.

Theorem 5.1. *Let Φ be t regular with $t > 0$. Then, for $0 < s < t$, the data-driven estimator $\widehat{f_n}$ in (2.7) satisfies*

$$\sup_{x\in\Omega_{x_0}} \sup_{f\in H^s(\Omega_{x_0},M)} E|\widehat{f_n}(x) - f(x)|^p \lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{sp}{2s+d}}.$$

Remark 5.1. *When $d = 1$, $p = 2$, and h denotes the density of the uniform distribution on [0, 1], the convergence rate derived in Theorem 5.1 coincides with the results in Nguyen and Matias [26] and Chagny et al. [5]. It is worth pointing out that our analysis relies on the local Hölder condition $H^s(\Omega_x)$ rather than the global Hölder space $H^s([0,1])$. In fact, a function $f \in H^s([0,1])$ must be contained in $H^s(\Omega_x)$ for each $x \in [0,1]$, whereas the converse does not hold in general; see [18].*

Proof. According to Theorem 3.1,

$$\begin{aligned} E\Big|\widehat{f_n}(x) - f(x)\Big|^p &\leq \inf_{j\in\mathcal{H}}\Big\{5B_j^*(x,f) + 5\tau_n(j)\Big\} + 5\aleph_n(x) \\ &\lesssim [B_{j_2}^*(x,f)]^p + [\tau_n(j_2)]^p + E[\aleph_n(x)]^p \end{aligned} \tag{5.1}$$

holds for any $x \in \Omega_{x_0}$, where $j_2 \in \mathcal{H}$ is determined later on.

By (2.6) and Proposition 4.1,

$$[\tau_n(j_2)]^p + E[\aleph_n(x)]^p \lesssim \Big(\frac{2^{j_2 d}\ln n}{n}\Big)^{\frac{p}{2}} + \Big(\frac{\ln n}{n}\Big)^{\frac{p}{2}}. \tag{5.2}$$

On the other hand, it follows from (3.1), (3.2), and Lemma 2.2 that the inequality

$$[B^*_{j_2}(x,f)]^p \quad = \quad \left[\sup_{j'\in\mathcal{H},\ j'\geq j_2} |P_{j'}f(x)-f(x)|\right]^p \lesssim 2^{-j_2 sp} \tag{5.3}$$

holds for any $x \in \Omega_{x_0}$ and $f \in H^s(\Omega_{x_0}, M)$.

Take j_2 satisfying $2^{j_2} \sim (\frac{n}{\ln n})^{\frac{1}{2s+d}}$ by balancing the two terms $\left(\frac{2^{j_2 d}\ln n}{n}\right)^{\frac{p}{2}}$ and $2^{-j_2 sp}$. Then, $j_2 \in \mathcal{H}$ by large n and $s > 0$. Furthermore,

$$\sup_{x\in\Omega_{x_0}} \sup_{f\in H^s(\Omega_{x_0},M)} E|\widehat{f_n}(x)-f(x)|^p \lesssim \Big(\frac{2^{j_2 d}\ln n}{n}\Big)^{\frac{p}{2}} + 2^{-j_2 sp} + \Big(\frac{\ln n}{n}\Big)^{\frac{p}{2}} \lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{sp}{2s+d}}$$

due to (5.1)–(5.3). The proof is finished. □

Next, the $L^p(1 \leq p < \infty)$ risk estimation is given.

Theorem 5.2. *Let Φ be t regular with $t > 0$. Then, for $0 < s < t$, $r, q \in [1,\infty]$, and $p \in [1,\infty)$, the data-driven estimator $\widehat{f_n}$ in (2.7) satisfies*

$$\sup_{f\in B^s_{r,q}(M,T)} E\|\widehat{f_n}I_{[-T,T]^d} - f\|_p^p \lesssim \Big(\frac{\ln n}{n}\Big)^{\theta p},$$

where

$$\theta := \begin{cases} \frac{s}{2s+d}, & 1 \leq p < \frac{2sr}{d} + r; \\ \frac{sr}{dp}, & p \geq \frac{2sr}{d} + r,\ s \leq \frac{d}{r}; \\ \frac{s-\frac{d}{r}+\frac{d}{p}}{2(s-\frac{d}{r})+d}, & p \geq \frac{2sr}{d} + r,\ s > \frac{d}{r}. \end{cases}$$

Remark 5.2. *When $s > \frac{d}{r}$, the convergence rates in Theorem 5.2 are optimal (up to a logarithmic factor) by Theorem 1 in [14]. However, the optimal convergence behavior remains unclear for $s \leq \frac{d}{r}$. Fortunately, the convergence rate $\frac{sr}{dp}$ matches the optimal convergence rate of standard nonparametric wavelet density estimation problems; see [4, 9].*

Remark 5.3. *When $s > \frac{d}{r}$, the convergence rate*

$$\theta = \min\left\{\frac{s}{2s+d},\ \frac{s-\frac{d}{r}+\frac{d}{p}}{2(s-\frac{d}{r})+d}\right\}$$

coincides with the work of Kou and Liang in [14]. More importantly, the estimation for the case $s \leq \frac{d}{r}$ is considered in Theorem 5.2. On the other hand, Theorem 5.2 with $p = 2$ can be reduced to Theorem 2 of Liang and Kou [15]. It should be pointed out that their estimators belong to classical nonlinear wavelet estimators, whereas ours are fully adaptive data-driven wavelet estimators.

Proof. It follows from Theorem 3.1 and (4.7) that

$$|\widehat{f_n}(x) - f(x)| \lesssim \mathfrak{M}(x,f) + \aleph_n(x)$$

holds for any $x \in [-T, T]^d$. This with $supp\ f \subset [-T, T]^d$ obtains

$$E\|\widehat{f_n} I_{[-T,T]^d} - f\|_p^p = E\int_{[-T,T]^d} \left|\widehat{f_n}(x) - f(x)\right|^p dx \qquad (5.4)$$
$$\lesssim \int_{[-T,T]^d} [\mathfrak{M}(x, f)]^p dx + \int_{[-T,T]^d} E[\aleph_n(x)]^p dx.$$

Recall that

$$\Lambda_m = \{x \in [-T, T]^d,\ 2^m \delta_n < \mathfrak{M}(x, f) \le 2^{m+1}\delta_n\}$$

in (4.8). Define $\Lambda_0^- := \{x \in [-T, T]^d,\ \mathfrak{M}(x, f) \le \delta_n\}$ with $\delta_n = (\frac{C \ln n}{n})^{\frac{s}{2s+d}}$. Then, for any $p \in [1, \infty)$,

$$\int_{[-T,T]^d} [\mathfrak{M}(x, f)]^p dx \lesssim \int_{\Lambda_0^-} [\mathfrak{M}(x, f)]^p dx + \sum_{m=0}^{m_2} \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx \qquad (5.5)$$
$$\lesssim \delta_n^p + \sum_{m=0}^{m_2} \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx,$$

where $m_2 \in \mathbb{Z}$ is given by (4.9).

Substituting (5.5) into (5.4), one obtains

$$\begin{aligned} E\|\widehat{f_n} I_{[-T,T]^d} - f\|_p^p &\lesssim \delta_n^p + \sum_{m=0}^{m_2} \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx + \int_{[-T,T]^d} E[\aleph_n(x)]^p dx \\ &\lesssim \delta_n^p + \sum_{m=0}^{m_2} \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx + \Big(\frac{\ln n}{n}\Big)^{\frac{p}{2}} \\ &\lesssim \delta_n^p + \sum_{m=0}^{m_2} \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx \qquad (5.6) \end{aligned}$$

thanks to Proposition 4.1 and $\delta_n \sim (\frac{\ln n}{n})^{\frac{s}{2s+d}}$.

From (4.8) and (4.9), one has $2^{m_2} \sim \delta_n^{-1}$ and $\delta_n \sim (\frac{\ln n}{n})^{\frac{s}{2s+d}}$. To finish the proof, we split Ineq (5.6) into three separate cases and establish the corresponding bounds by Proposition 4.2.

(i). For $1 \le p < \frac{2sr}{d} + r$,

$$\delta_n^p + \sum_{m=0}^{m_2} \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx \lesssim \delta_n^p \lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{sp}{2s+d}}. \qquad (5.7)$$

(ii). For $p \ge \frac{2sr}{d} + r$,

$$\delta_n^p + \sum_{m=0}^{m_2} \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx \lesssim \delta_n^p + 2^{m_2(p-r-\frac{2sr}{d})}\delta_n^p \lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{sr}{d}}. \qquad (5.8)$$

(iii). For the case $p \ge \frac{2sr}{d} + r$ and $s > \frac{d}{r}$, let $m_1 \in \mathbb{Z}$ satisfy

$$2^{m_1} \sim \delta_n^{\frac{s' p(\frac{1}{s} - \frac{1}{s'})}{(\frac{2s'}{d}+1)p - \frac{2sr}{d} - r}}$$

with $s' = s - \frac{d}{r} + \frac{d}{p}$. It is straightforward to verify $0 \le m_1 \le m_2$ by $r \le p,\ p \ge \frac{2sr}{d} + r$, and $s > \frac{d}{r}$. Therefore,

$$\begin{aligned}\delta_n^p + \sum_{m=0}^{m_2}\int_{\Lambda_m}[\mathfrak{M}(x,f)]^p dx &\le \delta_n^p + \sum_{m=0}^{m_1}\int_{\Lambda_m}[\mathfrak{M}(x,f)]^p dx + \sum_{m=m_1}^{m_2}\int_{\Lambda_m}[\mathfrak{M}(x,f)]^p dx\\ &\lesssim \delta_n^p + 2^{m_1(p-r-\frac{2sr}{d})}\delta_n^p + 2^{-\frac{2m_1 s' p}{d}}\delta_n^{\frac{s'}{s}p}.\end{aligned}$$

Then, due to the choice of 2^{m_1}, $\delta_n \sim (\frac{\ln n}{n})^{\frac{s}{2s+d}}$, and $s' = s - \frac{d}{r} + \frac{d}{p}$, the above inequality leads to

$$\delta_n^p + \sum_{m=0}^{m_2}\int_{\Lambda_m}[\mathfrak{M}(x,f)]^p dx \lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{s'p}{2(s-\frac{d}{r})+d}}. \tag{5.9}$$

The proof of Theorem 5.2 is completed thanks to (5.6)–(5.9). □

6. Numerical simulations

In this section, two numerical simulations are conducted to examine the empirical performance of the proposed wavelet estimator with dimension $d = 1$. Specifically, we focus on the estimation problem defined in (1.1), whose objective is to reconstruct the unknown probability density function $f(x)$ using the observed data points $X_1, X_2, , \cdots, X_n$. The performance of the wavelet estimator $\widehat{f_n}$ is measured by the mean square error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{f_n}(X_i) - f(X_i))^2,$$

where the corresponding integer j_0 is determined by our selection rule given in the preceding section.

In the simulation study, we consider sample size $n = 4000$ and the mixing ratio parameter $\theta = 0.2$. The codes are written in MATLAB software. In addition, the Daubechies wavelet D6 is employed throughout the simulations. As shown in Figures 1 and 2, the proposed wavelet estimator has excellent performance. Table 1 presents the MSE results under different sample sizes. The simulation study shows that the proposed wavelet estimator provides an effective approximation to the unknown density function $f(x)$.

Example 6.1. *We consider two density functions*

$$f_1(x) = 6(x - x^2)I_{[0,1]}(x) \ \ \text{and} \ \ h_1(x) = I_{[0,1]}(x).$$

The present example is compatible with model (1.1) and model (1.2) simultaneously. The performance of the proposed wavelet estimator is visualized in Figure 1. This result clearly demonstrates the effectiveness of the proposed wavelet estimator in approximating the target density function.

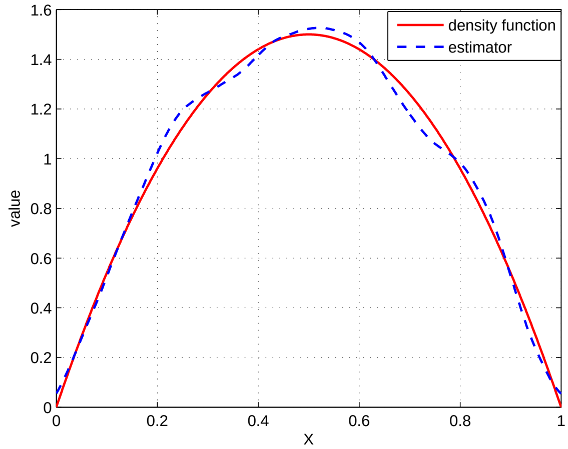

Figure 1. Example 6.1.

Example 6.2. *We choose the density functions*

$$f_2(x) = \frac{3}{4}\sqrt{1-|x|}I_{[-1,1]}(x) \;\; and \;\; h_2(x) = (0.6x^2 + 0.4cos2x + 0.1181405)I_{[-1,1]}(x).$$

The present example is compatible with model (1.1), and f_2 has a cusp at $x = 0$. We plot the performance of our wavelet estimator in Figure 2. This result shows that the proposed wavelet estimator can still achieve effective approximation even when the target density possesses a cusp.

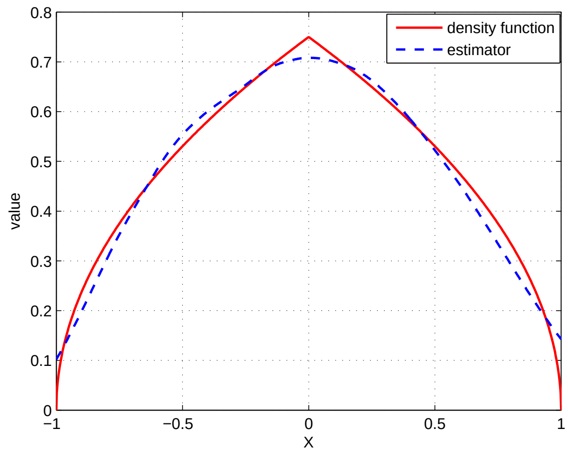

Figure 2. Example 6.2.

The following Table 1 presents the MSE values of the proposed wavelet estimator under different sample sizes n for the two numerical examples. According to Table 1, the MSE values decrease with increasing sample size n.

Table 1. The MSE values for different sample sizes n.

n	500	1000	2000	4000	8000	16000
Ex.6.1	0.0251512	0.0148637	0.0065220	0.0037389	0.0017663	0.0009376
Ex.6.2	0.0034964	0.0008705	0.0007434	0.0005568	0.0005469	0.0005348

7. Conclusions and future research

We study the mixture density model

$$g(x) = \theta h(x) + (1-\theta) f(x),\ \ x \in \mathbb{R}^d.$$

Subsequently, we construct a data-driven wavelet estimator and establish its pointwise and L^p risk estimates.

For the pointwise estimates, when $d = 1,\ p = 2$, and h is the density of the uniform distribution on $[0,\ 1]$, the convergence rate derived in Theorem 5.1 coincides with the results in Nguyen and Matias [26] and Chagny et al. [5]. It is worth pointing out that our analysis relies on the local Höder condition $H^s(\Omega_x)$, rather than the global Hölder space $H^s([0,1])$. In fact, a function $f \in H^s([0,1])$ must be contained in $H^s(\Omega_x)$ for each $x \in [0,1]$, whereas the converse does not hold in general.

For the L^p risk estimates, when $s > \frac{d}{r}$, the convergence rates in Theorem 5.2 are optimal (up to a logarithmic factor) thanks to Theorem 1 of Kou and Liang [14]. Moreover, Theorem 3 of Kou and Liang [14] and Theorem 2 of Liang and Kou [15] can follow directly from Theorem 5.2. On the other hand, the estimation for the case $s \le \frac{d}{r}$ is considered in Theorem 5.2, whereas it is none for that with traditional adaptive wavelet estimators. However, the optimal convergence behavior remains unclear for $s \le \frac{d}{r}$.

It should be pointed out that our current work assumes that the mixing ratio parameter θ is known in model (1.1), which introduces an evident limitation. In practical applications, the exact value of θ remains unknown in most cases. Hence, the preliminary parameter estimation is required to construct the estimator.

Finally, future work may concentrate on the construction of data-driven wavelet estimators for density estimation when the mixing ratio parameter θ is unknown, together with the associated error analysis. Besides, the optimality of the proposed estimators for the case $s \le \frac{d}{r}$ remains a problem for further investigation.

Use of Generative-AI tools declaration

The author declares that he has not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (No. 12101459).

Conflict of interest

The author declares no conflict of interest.

References

1. F. Bertolino, S. Cabras, E. M. Castellanos, W. Racugno, Unscaled Bayes factors for multiple hypothesis testing in microarray experiments, *Stat. Meth. Med. Res.*, **24** (2015), 1030–1043. https://doi.org/10.1177/0962280212437827
2. K. K. Cao, X. C. Zeng, Data-driven wavelet estimations for density derivatives, *Bull. Malays. Math. Sci. Soc.*, **47** (2024), 169–187. https://doi.org/10.1007/s40840-024-01766-5
3. K. K. Cao, X. C. Zeng, A data-driven wavelet estimator for deconvolution density estimations, *Results Math.*, **78** (2023), 156. https://doi.org/10.1007/s00025-023-01928-0
4. K. K. Cao, X. C. Zeng, Adaptive wavelet density estimation under independence hypothesis, *Results Math.*, **76** (2021), 196. https://doi.org/10.1007/s00025-021-01506-2
5. G. Chagny, A. Channarond, V. H. Hoang, A. Roche, Adaptive nonparametric estimation of a component density in a two-class mixture model, *J. Statist. Plann. Infer.*, **216** (2022), 51–69. https://doi.org/10.1016/j.jspi.2021.05.004
6. M. Chen, C. Gao, Z. Ren, A general decision theory for Huber's ϵ-contamination model, *Electron. J. Stat.*, **10** (2016), 3752–3774. https://doi.org/10.1214/16-EJS1216
7. I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.
8. I. Daubechies. The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inform. Theory*, **36** (1990), 961–1005. https://doi.org/10.1109/18.57199
9. A. Goldenshluger, O. Lepski, On adaptive minimax density estimation on $\mathbb{R}^d$, *Probab. Theory Relat. Fields*, **159** (2014), 479–543. https://doi.org/10.1007/s00440-013-0512-1
10. S. B. Hannadige, J. T. Gao, M. J. Silvapulle, P. Silvapulle, Forecasting a nonstationary time series using a mixture of stationary and nonstationary factors as predictors, *J. Bus. Econom. Statist.*, **42** (2024), 122–134. https://doi.org/10.1080/07350015.2023.2166048
11. W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, *Wavelets, Approximation and Statistical Applications*, New York: Springer-Verlag, 1998.
12. P. J. Huber, A robust version of the probability ratio test, *Ann. Math. Stat.*, **36** (1965), 1753–1758. https://doi.org/10.1214/aoms/1177699803
13. J. K. Kou, X. M. Chen, Wavelet estimations of a density function in two-class mixture model, *AIMS Math.*, **9** (2024), 20588–20611. https://doi.org/10.3934/math.20241000
14. J. K. Kou, D. Liang, Optimal convergence rates of wavelet estimators for a hidden density in a mixture model, *Math. Commun.*, **31** (2026), 83–113. https://doi.org/10.64785/mc.31.1.8
15. D. Liang, J. K. Kou, Nonparametric density estimation in a mixed model using wavelets, *Axioms*, **14** (2025), 741. https://doi.org/10.3390/axioms14100741
16. H. Y. Liu, C. Gao, Density estimation with contamination: minimax rates and theory of adaptation, *Election J. Stat.*, **13** (2019), 3613–3653. https://doi.org/10.1214/19-EJS1617

17. Y. M. Liu, C. Wu, Point-wise wavelet estimation in the convolution structure density model, *J. Fourier Anal. Appl.*, **26** (2020), 81. https://doi.org/10.1007/s00041-020-09794-y

18. Y. M. Liu, C. Wu, Point-wise estimation for anisotropic densities, *J. Multivar. Anal.*, **171** (2019), 112–125. https://doi.org/10.1016/j.jmva.2018.11.014

19. H. Q. Ma, J. M. Jiang, Pseudo-Bayesian classified mixed model prediction, *J. Amer. Statist. Assoc.*, **118** (2023), 1747–1759. https://doi.org/10.1080/01621459.2021.2008944

20. R. Maiboroda, O. Sugakova, Nonparametric density estimation for symmetric distributions by contaminated data, *Metrika*, **75** (2012), 109–126. https://doi.org/10.1007/s00184-010-0317-5

21. S. Mallat, A theorey for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **11** (1989), 674–693. https://doi.org/10.1109/34.192463

22. E. C. J. Maria, I. Salazar, L. Sanz, M. Gmez-Villegas, Using copula to model dependence when testing multiple hypotheses in DNA microarray experiments: A Bayesian approximation, *Mathematics*, **8** (2020), 1514. https://doi.org/10.3390/math8091514

23. G. McLachlan, D. Peel, *Finite Mixture Models*, New York: Wiley, 2000.

24. Y. Meyer, *Wavelets and Operators*, Cambridge: Cambridge University Press, 1992.

25. A. G. Mostofi, M. Kharrati-Kopaei, Nonparametric Bayesian inferences on the skewed data using a Dirichlet process mixture model, *Statist. Papers*, **66** (2025), 28–52. https://doi.org/10.1007/s00362-024-01658-7

26. V. H. Nguyen, 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), 584–612. https://doi.org/10.1051/ps/2013041

27. H. Shu, B. Nan, R. Koeppe, Multiple testing for neuroimaging via hidden Markov random field, *Biometrics*, **71** (2015), 741–750. https://doi.org/10.1111/biom.12329

28. C. Wu, X. C. Zeng, N. Mi, Adaptive and optimal pointwise deconvolution density estimations by wavelets, *Adv. Comput. Math.*, **47** (2021), 14. https://doi.org/10.1007/s10444-021-09844-z

AIMS Press