RESEARCH

Open Access

CrossMark

Wavelet optimal estimations for a two-dimensional continuous-discrete density function over L^p risk

Lin Hu[1], Xiaochen Zeng[2] and Jinru Wang[2*]

[*]Correspondence:
815150929@qq.com
[2]Department of Applied
Mathematics, Beijing University of
Technology, Beijing, P.R. China
Full list of author information is
available at the end of the article

Abstract

The mixed continuous-discrete density model plays an important role in reliability, finance, biostatistics, and economics. Using wavelets methods, Chesneau, Dewan, and Doosti provide upper bounds of wavelet estimations on L^2 risk for a two-dimensional continuous-discrete density function over Besov spaces $B_{r,q}^s$. This paper deals with L^p ($1 \le p < \infty$) risk estimations over Besov space, which generalizes Chesneau–Dewan–Doosti's theorems. In addition, we firstly provide a lower bound of L^p risk. It turns out that the linear wavelet estimator attains the optimal convergence rate for $r \ge p$, and the nonlinear one offers optimal estimation up to a logarithmic factor.

Keywords: Wavelets; Density estimation; Continuous-discrete density; Optimality

1 Introduction
1.1 Introduction

The density estimation plays an important role in both statistics and econometrics. This paper considers a two-dimensional density estimation model defined over mixed continuous and discrete variables [2]. More precisely, let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be independent and identically distributed (i.i.d.) observations of a bivariate random variable (X, Y), where X is a continuous random variable, and Y is a discrete one. The joint density function of (X, Y) is given by

$$f(x, v) = \frac{\partial}{\partial x} F(x, v)$$

with $F(x, v) = P(X \le x, Y = v)$ being the distribution function of (X, Y). We are interested in estimating $f(x, v)$ from $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. This continuous-discrete density model also arises in survival analysis, economics, and social sciences. For example, consider a series system with m components, which fails as soon as one of the components fails. Let X be the failure time of the system, and let Y be the component whose failure resulted in the failure of the system. Then (X, Y) is a bivariate continuous-discrete random variable. For more examples, see [1] and [4].

The conventional kernel method gives a nice estimation for the continuous-discrete density function [1, 10, 14]. However, it is hard to provide the optimal estimation for the

densities in Besov spaces. In addition, the complexity of bandwidth selection increases the difficulty of the kernel method.

Recently, wavelet methods have made the remarkable achievements in density estimation [7, 8, 11, 12, 15] due to their time and frequency localization, multiscale decomposition, and fast algorithm in numerical computations. In fact, wavelet estimation attains optimality for densities in Besov spaces, which avoids the disadvantage of kernel methods. Using the wavelet method, Chesneau et al. [2] constructed linear and nonlinear wavelet estimators for a two-dimensional continuous-discrete density function and derived their mean integrated squared errors performance over Besov balls.

This paper addresses L^p ($1 \leq p < \infty$) risk estimations on Besov balls by using wavelet bases, which generalizes Chesneau–Dewan–Doosti's theorems. It should be pointed out that a lower bound for L^p risk of all estimators is derived firstly. It turns out that the linear wavelet estimator is optimal for $r \geq p$ and the nonlinear one attains optimal estimation up to a logarithmic factor.

1.2 Notations and definitions

In this paper, we use the tensor product method to construct an orthonormal wavelet basis for $L^2(\mathbb{R}^2)$, which will be used in later discussions. With a one-dimensional Daubechies scaling function D_{2N} and a wavelet function ψ_{2N} (ψ_{2N} can be constituted from the scaling function D_{2N}), we construct two-dimensional tensor product wavelets φ, ψ^1, ψ^2, and ψ^3 as follows:

$$\varphi(x, y) := D_{2N}(x)D_{2N}(y), \qquad \psi^1(x, y) := D_{2N}(x)\psi_{2N}(y),$$
$$\psi^2(x, y) := \psi_{2N}(x)D_{2N}(y), \qquad \psi^3(x, y) := \psi_{2N}(x)\psi_{2N}(y).$$

Then φ and ψ^i ($i = 1, 2, 3$) are compactly supported in time domain, because Daubechies' wavelet D_{2N} and ψ_{2N} are [5, 8].

Denote

$$\varphi_{j,k}(x, y) := 2^j \varphi\left(2^j x - k_1, 2^j y - k_2\right), \qquad \psi^i_{j,k}(x, y) := 2^j \psi^i\left(2^j x - k_1, 2^j y - k_2\right)$$

for $k = (k_1, k_2) \in \mathbb{Z}^2$ and $i = 1, 2, 3$. Then for each $f \in L^2(\mathbb{R}^2)$,

$$f = \sum_{k \in \mathbb{Z}^2} \alpha_{j_0,k} \varphi_{j_0,k} + \sum_{j=j_0}^{\infty} \sum_{i=1}^{3} \sum_{k \in \mathbb{Z}^2} \beta^i_{j,k} \psi^i_{j,k}$$

holds in L^2 sense, where $\alpha_{j,k} := \langle f, \varphi_{j,k} \rangle$, $\beta^i_{j,k} := \langle f, \psi^i_{j,k} \rangle$. As usual, let P_j be the orthogonal projection operator defined by

$$P_j f := \sum_{k \in \mathbb{Z}^2} \langle f, \varphi_{j,k} \rangle \varphi_{j,k}.$$

Details on wavelet basis can be found in [5, 8]. A scaling function φ is called m-regular if $\varphi \in C^m(\mathbb{R}^2)$ and $|D^\alpha \varphi(x)| \leq C(1 + |x|^2)^{-l}$ for each $l \in \mathbb{Z}$ ($|\alpha| = 0, 1, \ldots, m$). By the definition of tensor product wavelets we find that the scaling function φ is m-regular, since Daubechies' function D_{2N} is smooth enough for large N.

One of advantages of wavelet bases is that they can characterize Besov spaces, which contain Hölder spaces and L^2-Sobolev spaces as particular examples. Throughout the paper, we work within a Besov space on a compact subset of $\mathbb{R}^2$. The following lemma shows equivalent definitions for those spaces, which are fundamental in our discussions.

Lemma 1.1 ([13]) *Let φ be an m-regular orthonormal scaling function with the corresponding wavelets ψ^i ($i = 1, 2, 3$). If $f \in L^r(\mathbb{R}^2)$, $\alpha_{j,k} = \langle f, \varphi_{j,k} \rangle$ $\beta_{j,k}^i = \langle f, \psi_{j,k}^i \rangle$, and $1 \leq r, q \leq \infty$, $0 < s < m$. Then following assertions are equivalent*:

   (i) $f \in B_{r,q}^s(\mathbb{R}^2)$;

   (ii) $\{2^{js} \| P_{j+1}f - P_j f \|_r\}_{j \geq 0} \in l^q$;

   (iii) $\| \{2^{j(s+1-\frac{2}{p})} \| \beta_{j,\cdot} \|_r\}_{j \geq 0} \|_q < \infty$.

The Besov norm of f can be defined by

$$\|f\|_{B_{r,q}^s} := \|\alpha_{j_0,\cdot}\|_r + \left\| \left\{ 2^{j(s+1-\frac{2}{p})} \|\beta_{j,\cdot}\|_r \right\}_{j \geq j_0} \right\|_q,$$

where $\|\alpha_{j_0,\cdot}\|_r^r := \sum_{k \in \mathbb{Z}^2} |\alpha_{j_0,k}|^r$ and $\|\beta_{j,\cdot}\|_r^r := \sum_{i=1}^3 \sum_{k \in \mathbb{Z}^2} |\beta_{j,k}^i|^r$.

Here and further, $A \lesssim B$ means that $A \leq CB$ for some constant $C > 0$ independent of A and B, $A \gtrsim B$ means $B \lesssim A$, and $A \sim B$ stands for both $A \lesssim B$ and $A \gtrsim B$.

Remark 1.1 By (i) and (ii) of Lemma 1.1 we observe that

$$\|P_j f - f\|_r = \left\| \sum_{l=j}^{\infty} (P_{l+1}f - P_l f) \right\|_r \leq \sum_{l=j}^{\infty} \|P_{l+1}f - P_l f\|_r \lesssim \sum_{l=j}^{\infty} 2^{-ls} \lesssim 2^{-js}$$

for $f \in B_{r,q}^s(\mathbb{R}^2)$. Hence

$$\|P_j f - f\|_r \lesssim 2^{-js}. \tag{1.1}$$

Remark 1.2 When $r \leq p$, Lemma 1.1(i) and (iii) imply that, for $s' - \frac{2}{p} = s - \frac{2}{r} > 0$,

$$B_{r,q}^s(\mathbb{R}^2) \hookrightarrow B_{p,q}^{s'}(\mathbb{R}^2),$$

where $A \hookrightarrow B$ stands for a Banach space A continuously embedded in another Banach space B. More precisely, $\|u\|_B \leq C\|u\|_A$ ($u \in A$) for some constant $C > 0$.

Lemma 1.2 ([13]) *Let $\varphi \in L^2(\mathbb{R}^2)$ be a scaling function or a wavelet with $\sup_{k \in \mathbb{Z}^2} |\varphi(x-k)| < \infty$. Then, for $\lambda = \{\lambda_k\} \in l^p(\mathbb{Z}^2)$ and $1 \leq p \leq \infty$*,

$$\left\| \sum_{k \in \mathbb{Z}^2} \lambda_k \varphi_{j,k} \right\|_p \sim 2^{j(1-2/p)} \|\lambda\|_p.$$

Here $\|\lambda\|_p$ is the $l^p(\mathbb{Z}^2)$ norm of $\lambda \in l^p(\mathbb{Z}^2)$:

$$\|\lambda\|_p := \begin{cases} (\sum_{k \in \mathbb{Z}^2} |\lambda_k|^p)^{1/p} & \text{if } p < \infty, \\ \sup_{k \in \mathbb{Z}^2} |\lambda_k| & \text{if } q = \infty. \end{cases}$$

1.3 Main results

In this subsection, we state our main results and discuss relations to some other work. To do that, we propose a new bivariate function $f_*(x, y)$, which is an improved one of that in [2]. Define

$$f_*(x, y) := \sum_{v=1}^{m} u(y, v) P(Y = v) f(x | Y = v)$$

with

$$u(y, v) = \begin{cases} \dfrac{1}{1 + e^{\frac{1}{y-v} + \frac{1}{y-v+1}}} 1_{(v-1, v)}(y) + \dfrac{e^{\frac{1}{y-v} + \frac{1}{y-v-1}}}{1 + e^{\frac{1}{y-v} + \frac{1}{y-v-1}}} 1_{(v, v+1)}(y), & y \neq v, \\ 1, & y = v, \end{cases}$$

where 1_D is the indicator function of a set D.

The construction of f_* follows the idea proposed by Chesneau [2] but is different from [2]. The weight $u(y, v)$ equals to characteristic function $1_{\{v - \frac{1}{2} \leq y < v + \frac{1}{2}\}}$ in [2]. By a careful verification our weight $u(y, v)$ is differentiable with respect to y for each $v \in \{1, 2, \dots, m\}$. The modification of $u(y, v)$ from the characteristic function to the smooth one makes f_* continuous in y. It is easy to see that, for any $y = v \in \{1, 2, \dots, m\}$,

$$f_*(x, y) = f(x, v).$$

Hence, the problem is converted to construct an estimator of f_*. As in [2], we assume that f_* belongs to the space $B_{r,q}^s(H, Q)$ or, equivalently, f_* belongs to the Besov ball

$$B_{r,q}^s(H) := \left\{ f, f \in B_{r,q}^s(\mathbb{R}^2) \text{ and } \|f\|_{B_{r,q}^s} \leq H \right\}$$

and that the support of $f_*(x, \cdot)$ is contained in $[-Q, Q]$ for fixed v ($Q > 0$, $v = 1, 2, \dots, m$).

To introduce the wavelet estimator, we need the wavelet coefficient estimators of $\alpha_{j,k}$ and $\beta_{j,k}^i$:

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{l=1}^{n} \int_{\mathbb{R}} \varphi_{j,k}(X_l, y) u(y, Y_l) \, dy, \qquad \hat{\beta}_{j,k}^i = \frac{1}{n} \sum_{l=1}^{n} \int_{\mathbb{R}} \psi_{j,k}^i(X_l, y) u(y, Y_l) \, dy. \tag{1.2}$$

Define $\wedge_{j_0} := \{k \in \mathbb{Z}^2, \operatorname{supp} f_* \cap \operatorname{supp} \varphi_{j_0, k} \neq \emptyset\}$. When f_* and φ have compact supports, the cardinality of $\wedge_j$ satisfies $\sharp \wedge_j \lesssim 2^{2j}$. Then the linear wavelet estimator of f_* is given as follows:

$$\hat{f}_n^{\text{lin}}(x, y) := \sum_{k \in \wedge_{j_0}} \hat{\alpha}_{j_0, k} \varphi_{j_0, k}(x, y), \tag{1.3}$$

where j_0 is chosen such that $2^{j_0} \sim n^{\frac{1}{2s'+1}}$, $s' := s - (\frac{2}{r} - \frac{2}{p})_+$, and $x_+ := \max\{x, 0\}$.

To obtain a nonlinear estimator, we take j_0 and j_1 such that $2^{j_1} \sim \frac{n}{\ln n}$ and $2^{j_0} \sim n^{\frac{1}{2m+1}}$ with $m > s$. Define $\wedge_j := \{k \in \mathbb{Z}^2, \operatorname{supp} f_* \cap \operatorname{supp} \psi_{j,k}^i \neq \emptyset\}$ and $\lambda_j := \frac{T}{2} 2^{-\frac{j}{2}} \sqrt{\frac{\ln n}{n}}$ (T is the constant

described as Lemma 2.3). Then the nonlinear estimator is given by

$$\hat{f}_n^{\text{non}}(x,y) := \sum_{k \in \wedge_{j_0}} \hat{\alpha}_{j_0,k} \varphi_{j_0,k}(x,y) + \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} \sum_{k \in \wedge_j} \hat{\beta}_{j,k}^i 1_{\{|\hat{\beta}_{j,k}^i| > \lambda_j\}} \psi_{j,k}^i(x,y). \tag{1.4}$$

From the definition of $\hat{f}_n^{\text{non}}$ we find that the nonlinear estimator has the advantage to be adaptive, since it does not depend on the indices s, r, q and H in its construction.

The following theorem gives a lower bound estimation for L^p risk.

Theorem 1.1 *Let $\hat{f}$ be an estimator of $f_* \in B_{r,q}^s(H)$ with $s > \frac{2}{r}$ and $r, q \geq 1$. Then there exists $C > 0$ such that, for $1 \leq p < \infty$,*

$$\sup_{f_* \in B_{r,q}^s(H)} E\|\hat{f}_n - f_*\|_p^p \geq C \max\left\{ n^{-\frac{sp}{2s+1}}, \left(\frac{\ln n}{n}\right)^{-\frac{(s-\frac{2}{r}+\frac{2}{p})p}{2(s-\frac{2}{r})+1}} \right\}.$$

The upper bounds of the linear and nonlinear wavelet estimators are provided by Theorems 1.2 and 1.3, respectively.

Theorem 1.2 *Let $\hat{f}_n^{\text{lin}}$ be the estimator of $f_* \in B_{r,q}^s(H,Q)$ defined by (1.3) with $1 \leq r, q < \infty$, $s > 0$. If the density of X is bounded, then for $r \geq p \geq 1$ or $r \leq p < \infty$ and $s > \frac{2}{r}$,*

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\|\hat{f}_n^{\text{lin}} - f_*\|_p^p \lesssim n^{-\frac{ps'}{2s'+1}}$$

with $s' = s - (\frac{2}{r} - \frac{2}{p})_+$ and $x_+ := \max(x,0)$.

Remark 1.3 If $r \geq 2$, $p = 2$ and $s > 0$, $s' = s$, then Theorem 1.2 reduces to Theorem 4.1 in [2]. In addition, Theorem 1.2 does not make any restriction on Q, and so the assumptions are weaker than in [2]. Theorem 1.2 extends the corresponding theorem of [2] from $p = 2$ to $p \in [1, \infty)$.

When $r \geq p$, $s' = s$ and the linear wavelet estimator $\hat{f}_n^{\text{lin}}$ attains optimality thanks to Theorems 1.1 and 1.2. However, the linear estimator does not offer optimal estimation for $r < p$, because of $s' < s$ and $\frac{s'}{2s'+1} < \frac{s}{2s+1}$ in this case.

To give a suboptimal estimation for $r < p$, we need the nonlinear wavelet estimators defined by (1.4).

Theorem 1.3 *Let $\hat{f}_n^{\text{non}}$ be the estimator of $f_* \in B_{r,q}^s(H,Q)$ defined by (1.4) with $1 \leq r, q < \infty$, $s > 0$. If the density of X is bounded, then for $r \geq p \geq 1$ or $r \leq p < \infty$ and $s > \frac{2}{r}$,*

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\|\hat{f}_n^{\text{non}} - f_*\|_p^p \lesssim (\ln n)^p \left(\frac{\ln n}{n}\right)^{\alpha p}$$

with $\alpha := \min\{\frac{s}{2s+1}, \frac{s-\frac{2}{r}+\frac{2}{p}}{2(s-\frac{2}{r})+1}\}$.

Remark 1.4 Theorems 1.1 and 1.3 tell us that the nonlinear estimator is suboptimal up to a logarithmic factor. Moreover, if $p = 2$ and $\{r \geq 2, s > 0\}$ or $\{1 \leq r < 2, s > \frac{2}{r}\}$, then $\alpha = \frac{s}{2s+1}$, and Theorem 1.3 is the same as Theorem 4.2 in [2] up to a logarithmic factor. Hence Theorem 1.3 can be considered as an extension of Theorem 4.2 in [2] from $p = 2$ to $p \in [1, \infty)$.

In particular, we can extend the theorems to the multidimensional case as in [3] by using the technique developed by [9]. It is a challenging problem to study the estimation of a multivariate continuous-discrete conditional density. We refer to [3] for further details.

2 Some lemmas

We shall show several lemmas in this section, which are needed for proofs of our main theorems.

Lemma 2.1 *Let $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ be defined by* (1.2). *Then*

$$E(\hat{\alpha}_{j,k}) = \alpha_{j,k} \quad and \quad E(\hat{\beta}^i_{j,k}) = \beta^i_{j,k}$$

for $j \geq j_0$, $k \in \mathbb{Z}^2$, and $i = 1, 2, 3$.

Proof Denote $c_{j,k}(v) = \int \phi_{j,k_2}(y) u(y, v)\, dy$. Then

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \int \varphi_{j,k}(X_i, y) u(y, Y_i)\, dy = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,k_1}(X_i) c_{j,k}(Y_i).$$

Since $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are independent and identically distributed, we have

$$E(\hat{\alpha}_{j,k}) = E\big(\phi_{j,k_1}(X_1) c_{j,k}(Y_1)\big) = E\big(E\big(\phi_{j,k_1}(X_1) c_{j,k}(Y_1)|Y_1\big)\big)$$

$$= E\big(c_{j,k}(Y_1) E\big(\phi_{j,k_1}(X_1)|Y_1\big)\big) = E\left(c_{j,k}(Y_1) \int \phi_{j,k_1}(x) f(x|Y_1)\, dx\right)$$

$$= \sum_{v=1}^{m} P(Y_1 = v) c_{j,k}(v) \int \phi_{j,k_1}(x) f(x|Y_1 = v)\, dx$$

$$= \int \int \left(\sum_{v=1}^{m} P(Y_1 = v) u(y, v) f(x|Y_1 = v)\right) \phi_{j,k_1}(x) \phi_{j,k_2}(y)\, dx\, dy$$

$$= \int \int f_*(x, y) \varphi_{j,k}(x, y)\, dx\, dy = \alpha_{j,k}.$$

Similarly to the previous arguments, $E(\hat{\beta}^i_{j,k}) = \beta^i_{j,k}$. The proof of Lemma 2.1 is done. □

To show Lemma 2.2, we introduce Rosenthal's inequality.

Rosenthal's inequality ([8]) Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $EX_l = 0$ and $E|X_l|^p < \infty$ ($l = 1, 2, \ldots, n$). Then, with $C_p > 0$,

$$E\left|\sum_{l=1}^{n} X_l\right|^p \leq \begin{cases} C_p[\sum_{l=1}^{n} E|X_l|^p + (\sum_{l=1}^{n} E|X_l|^2)^{p/2}], & p \geq 2, \\ C_p(\sum_{l=1}^{n} E|X_l|^2)^{p/2}, & 0 < p \leq 2. \end{cases}$$

Lemma 2.2 *Let $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ be defined by* (1.2). *If the density of X is bounded, then there exists a constant $C > 0$ such that*

$$E|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p \leq 2^{-\frac{p}{2}j}n^{-\frac{p}{2}} \quad and \quad E|\hat{\beta}_{j,k} - \beta_{j,k}|^p \leq 2^{-\frac{p}{2}j}n^{-\frac{p}{2}}$$

for $1 \leq p < \infty$ and $2^j \leq n$.

Proof We only prove the first inequality, since the second one is similar. By the definition of $\hat{\alpha}_{j,k}$,

$$\hat{\alpha}_{j,k} = \frac{1}{n}\sum_{l=1}^{n}\int_{\mathbb{R}}\varphi_{j,k}(X_l, y)u(y, Y_l)\,dy = \frac{1}{n}\sum_{l=1}^{n}\phi_{j,k_1}(X_l)c_{j,k_2}(Y_l),$$

where $c_{j,k_2}(Y_l) := \int_{\mathbb{R}}\phi_{j,k_2}(y)u(y, Y_l)\,dy$, and ϕ is a one-dimensional Daubechies scaling function D_{2N}. Since $|u(y, v)| \leq 2$, we obtain that

$$\left|c_{j,k_2}(Y_l)\right| \leq \int_{\mathbb{R}}\left|\phi_{j,k_2}(y)\right|\left|u(y, Y_l)\right|\,dy \leq 2^{-\frac{j}{2}}\|\phi\|_1 \tag{2.1}$$

and

$$E\left|\phi_{j,k_1}(X_l)c_{j,k_2}(Y_l)\right|^p \lesssim 2^{-\frac{p}{2}j}E\left|\phi_{j,k_1}(X_l)\right|^p \lesssim 2^{-\frac{p}{2}j}\int_{\mathbb{R}}\left|\phi_{j,k_1}(x)\right|^p f_X(x)\,dx \lesssim 2^{-j} \tag{2.2}$$

due to the boundedness of f_X. Define $\xi_l := \phi_{j,k_1}(X_l)c_{j,k_2}(Y_l) - \alpha_{j,k}$. Then

$$E|\xi_l|^p = E\left|\phi_{j,k_1}(X_l)c_{j,k_2}(Y_l) - \alpha_{j,k}\right|^p \lesssim E\left|\phi_{j,k_1}(X_l)c_{j,k_2}(Y_l)\right|^p + E|\alpha_{j,k}|^p. \tag{2.3}$$

It follows from Lemma 2.1 and Jensen's inequality that

$$E|\alpha_{j,k}|^p = \left|E\left[\phi_{j,k_1}(X_l)c_{j,k_2}(Y_l)\right]\right|^p \leq E\left|\phi_{j,k_1}(X_l)c_{j,k_2}(Y_l)\right|^p.$$

Hence (2.3) reduces to

$$E|\xi_l|^p \lesssim E\left|\phi_{j,k_1}(X_l)c_{j,k_2}(Y_l)\right|^p \lesssim 2^{-j} \tag{2.4}$$

thanks to (2.2). By the definition of $\hat{\alpha}_{j,k}$ and ξ_l, $\hat{\alpha}_{j,k} - \alpha_{j,k} = \frac{1}{n}\sum_{l=1}^{n}\xi_l$, where $\xi_1, \xi_2, \ldots, \xi_n$ are independent because $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ also are. On the other hand, Lemma 2.1 implies $E(\xi_l) = 0$. Then Rosenthal inequality leads to

$$E|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p = E\left|\frac{1}{n}\sum_{l=1}^{n}\xi_l\right|^p \lesssim \begin{cases} n^{-p}[\sum_{l=1}^{n}E|\xi_l|^p + (\sum_{l=1}^{n}E|\xi_l|^2)^{\frac{p}{2}}], & p \geq 2, \\ n^{-p}(\sum_{l=1}^{n}E|\xi_l|^2)^{\frac{p}{2}}, & 1 \leq p \leq 2. \end{cases} \tag{2.5}$$

By (2.4) we know that

$$n^{-p}\left(\sum_{l=1}^{n}E|\xi_l|^2\right)^{\frac{p}{2}} \lesssim n^{-p}(n2^{-j})^{\frac{p}{2}} \lesssim n^{-\frac{p}{2}}2^{-\frac{p}{2}j}$$

for $1 \le p < 2$ and

$$n^{-p}\left[\sum_{l=1}^{n} E|\xi_l|^p + \left(\sum_{l=1}^{n} E|\xi_l|^2\right)^{\frac{p}{2}}\right] \lesssim n^{-p}\left(n2^{-j} + n^{\frac{p}{2}}2^{-\frac{p}{2}j}\right) \lesssim n^{-\frac{p}{2}}2^{-\frac{p}{2}j}$$

for $p \ge 2$ thanks to the assumption $2^j \le n$. Combining these with (2.5), we receive the desired conclusion

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

This completes the proof. □

To prove Lemma 2.3, we need the well-known Bernstein inequality.

Bernstein's inequality ([8]) Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $E(X_i) = 0$ and $\|X_i\|_\infty \le M$. Then, for each $\gamma > 0$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \gamma\right\} \le 2\exp\left(-\frac{n\gamma^2}{2[E(X_i^2) + \|X\|_\infty\gamma/3]}\right).$$

The next lemma is an extension of Proposition 4.2 in [2].

Lemma 2.3 *Let $2^j \le \frac{n}{\ln n}$, $\hat{\beta}_{j,k}^i$ $(i = 1, 2, 3)$ be defined in (1.2). If the density of X is bounded, then for each $\varepsilon > 0$, there exists $T > 0$ such that, for $j \ge 0$ and $k \in \mathbb{Z}^2$,*

$$P\left\{|\hat{\beta}_{j,k}^i - \beta_{j,k}^i| > \frac{T}{2}2^{-\frac{1}{2}j}\sqrt{\frac{\ln n}{n}}\right\} \lesssim 2^{-\varepsilon j}. \tag{2.6}$$

Proof We only show (2.6) for $i = 1$. By the definition of $\hat{\beta}_{j,k}^1$, $\hat{\beta}_{j,k}^1 = \frac{1}{n}\sum_{l=1}^{n}\int_{\mathbb{R}}\psi_{j,k}^1(X_l, y)u(y, Y_l)\,dy$, and

$$\hat{\beta}_{j,k}^1 - \beta_{j,k}^1 = \frac{1}{n}\sum_{l=1}^{n}[\phi_{j,k_1}(X_l)d_{j,k_2}(Y_l) - \beta_{j,k}^1],$$

where $d_{j,k_2}(Y_l) := \int_{\mathbb{R}}\psi_{j,k_2}(y)u(y, Y_l)\,dy$ (ϕ, ψ stand for the one-dimensional Daubechies scaling function and wavelet function, respectively). Define $\eta_l := \phi_{j,k_1}(X_l)d_{j,k_2}(Y_l) - \beta_{j,k}^1$. Then $\hat{\beta}_{j,k}^1 - \beta_{j,k}^1 = \frac{1}{n}\sum_{l=1}^{n}\eta_l$ and $E(\eta_l) = 0$ because of $\beta_{j,k}^l = E(\hat{\beta}_{j,k}^l) = E[\phi_{j,k_1}(X_i)d_{j,k_2}(Y_i)]$.

Using (2.1) with ψ instead of ϕ, we get $|d_{j,k_2}(Y_l)| \lesssim 2^{-\frac{j}{2}}$. Note that $|\phi_{j,k_1}(X_l)| := 2^{\frac{j}{2}}|\phi(2^j X_l - k_1)| \le 2^{\frac{j}{2}}\|\phi\|_\infty$. Then $|\phi_{j,k_1}(X_l)d_{j,k_2}(Y_l)| \lesssim 1$ and $|\beta_{j,k}^1| = |E[\phi_{j,k_1}(X_l)d_{j,k_2}(Y_l)]| \lesssim 1$. Hence

$$|\eta_l| \le |\phi_{j,k_1}(X_l)d_{j,k_2}(Y_l) - \beta_{j,k}^1| \lesssim 1. \tag{2.7}$$

By replacing c_{j,k_2} and $\alpha_{j,k}$ with d_{j,k_2} and $\beta_{j,k}^1$, respectively, arguments similar to (2.1)–(2.4) show that

$$E|\eta_l|^p \lesssim 2^{-j}. \tag{2.8}$$

Because $\eta_1, \eta_2, \ldots, \eta_n$ are i.i.d. and $E(\eta_l) = 0$ ($l = 1, 2, \ldots, n$), Bernstein's inequality tells us that

$$P\left\{ \left| \hat{\beta}_{j,k}^l - \beta_{j,k}^l \right| = \left| \frac{1}{n} \sum_{l=1}^{n} \eta_l \right| > \frac{T}{2} 2^{-\frac{1}{2}j} \sqrt{\frac{\ln n}{n}} \right\} \leq 2\exp\left(-\frac{n\lambda_j^2}{2[E(\eta_l^2) + \frac{\lambda_j}{3}\|\eta\|_\infty]} \right) \qquad (2.9)$$

with $\lambda_j = \frac{T}{2} 2^{-\frac{1}{2}j} \sqrt{\frac{\ln n}{n}}$. This with (2.7)–(2.8) implies

$$\frac{n\lambda_j^2}{2[E(\eta_l^2) + \frac{\lambda_j}{3}\|\eta\|_\infty]} \geq \frac{T^2 \ln n}{8(C_1 + \frac{C_2}{6} T 2^{\frac{j}{2}}\sqrt{\frac{\ln n}{n}})} \geq \frac{T^2 \ln n}{8(C_1 + \frac{C_2}{6} T)}$$

because $2^{\frac{j}{2}}\sqrt{\frac{\ln n}{n}} \leq 1$ by the assumption $2^j \leq \frac{n}{\ln n}$. Note that $\ln n > j \ln 2$ due to $n \geq 2^j \ln n > 2^j$. Hence

$$\frac{n\lambda_j^2}{2[E(\eta_l^2) + \frac{\lambda_j}{3}\|\eta\|_\infty]} \geq \frac{T^2 \ln 2}{8(C_1 + \frac{C_2}{6} T)} j > \varepsilon j$$

by choosing $T > 0$ such that $\frac{T^2 \ln 2}{8(C_1 + \frac{C_2}{6} T)} > \varepsilon$. Then (2.9) reduces to

$$P\left\{ \left| \hat{\beta}_{j,k}^1 - \beta_{j,k}^1 \right| > \frac{T}{2} 2^{-\frac{1}{2}j} \sqrt{\frac{\ln n}{n}} \right\} \leq 2^{-\varepsilon j},$$

which concludes (2.6) with $i = 1$. Similarly, the conclusions with $i = 2, 3$ hold. This completes the proof. $\qquad\square$

At the end of this section, we introduce two classical lemmas, which are needed for the proof of lower bound.

Lemma 2.4 (Varshamov–Gilbert lemma, [11]) *Let $\Theta := \{\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m), \varepsilon_i \in \{0, 1\}\}$. Then there exists a subset $(\varepsilon^0, \varepsilon^1, \ldots, \varepsilon^T)$ of Θ with $\varepsilon^0 = (0, 0, \ldots, 0)$ such that $T \geq 2^{\frac{m}{8}}$ and*

$$\sum_{k=1}^{m} \left| \varepsilon_k^i - \varepsilon_k^j \right| \geq \frac{m}{8} \quad (0 \leq i \neq j \leq T).$$

To state Fano's lemma, we introduce a concept: When P is absolutely continuous with respect to Q (denoted by $P \ll Q$), the Kullback divergence of P and Q between two measures P and Q is defined by

$$K(P, Q) := \int p(x) \ln \frac{p(x)}{q(x)} \, dx,$$

where $p(x)$ and $q(x)$ are the density functions of P and Q, respectively.

Lemma 2.5 (Fano's lemma, [6]) *Let $(\Omega, \mathcal{F}, P_k)$ be a probability spaces, and let $A_k \in \mathcal{F}$, $k = 0, 1, \ldots, m$. If $A_k \cap A_{k'} = \emptyset$ for $k \neq k'$, then with A^C standing for the complement of A and*

$\mathcal{K}_m := \inf_{0 \le k \le m} \frac{1}{m} \sum_{k \ne k'} K(P_k, P_{k'}),$

$$\sup_{0 \le k \le m} P_k(A_k^C) \ge \min \left\{ \frac{1}{2}, \sqrt{m} \exp(-3e^{-1} - \mathcal{K}_m) \right\},$$

where $K(P_k, P_{k'})$ is the Kullback distance of P_k and $P_{k'}$ $(k = 0, 1, \ldots, m)$.

3 Proofs of lower bounds

We rewrite Theorem 1.1 as follows before giving its proof.

Theorem 3.1 *Let $\hat{f}_n$ be an estimator of $f_* \in B_{r,q}^s(H)$ with $s > \frac{2}{r}$ and $1 \le r, q \le \infty$. Then, for $1 \le p < \infty$,*

$$\sup_{f_* \in B_{r,q}^s(H)} E\|\hat{f}_n - f_*\|_p^p \gtrsim \max \left\{ n^{-\frac{sp}{2s+1}}, \left(\frac{\ln n}{n} \right)^{\frac{(s - \frac{2}{r} + \frac{2}{p})p}{2(s - \frac{2}{r}) + 1}} \right\}.$$

Proof As in Sect. 1, we take the two-dimensional tensor product wavelet

$$\psi^1(x, y) := D_{2N}(x) \psi_{2N}(y),$$

where $D_{2N}(\cdot)$ and $\psi_{2N}(\cdot)$ are the one-dimensional Daubechies scaling function and wavelet function, respectively. Then ψ^1 is m-regular $(m > s)$ for large N, and

$$\operatorname{supp} \psi^1 \subseteq [0, 2N - 1] \times [-N + 1, N]$$

due to $\operatorname{supp} D_{2N} \subseteq [0, 2N - 1]$ and $\operatorname{supp} \psi_{2N} \subseteq [-N + 1, N]$. Then there exists a compactly supported density function g_0 such that $\int_{\mathbb{R}^2} g_0(x)\, dx = 1$, $g_0(x)|_{[0,2N-1] \times [-N+1,N]} = c_0$, and $g_0 \in B_{r,q}^s(H)$. Define $\Delta_j := \Delta_j^1 \times \Delta_j^2$ with

$$\Delta_j^1 := \{0, 2N, 4N, \ldots, 2(2^j - 1)N\}, \qquad \Delta_j^2 := \{0, \pm 2N, \pm 4N, \ldots, \pm 2(2^{j-1} - 1)N\}.$$

Then $\sharp \Delta_j = 2^j(2^j - 1) \sim 2^{2j}$ ($\sharp \Delta_j$ denotes the cardinality of Δ_j). Denote $a_j := 2^{-(2s+1)j}$ and

$$\wedge := \left\{ g_\varepsilon(x, y) = g_0(x, y) + a_j \sum_{k \in \Delta_j} \varepsilon_k \psi_{j,k}^1(x, y), \varepsilon_k \in \{0, 1\} \right\}.$$

Obviously, the supports of $\psi_{j,k}^1$ and $\psi_{j,k'}^1$ are disjoint for $k \ne k' \in \Delta_j$ and $\operatorname{supp} \psi_{j,k}^1 \subseteq \operatorname{supp} g_0$. When $(x, y) \in [0, 2N - 1] \times [-N + 1, N]$,

$$g_\varepsilon \ge c_0 - a_j \|\psi_{j,k}^1\|_\infty \ge c_0 - 2^{-2sj} \|\psi^1\|_\infty > 0$$

for large j. On the other hand,

$$\int_{\mathbb{R}^2} g_\varepsilon(x, y)\, dx = \int_{\mathbb{R}^2} g_0(x, y)\, dx = 1.$$

Hence g_ε is a bivariate density function for $\varepsilon = (\varepsilon_k)_{k \in \Delta_j}$.

Moreover, $g_\varepsilon \in B^s_{r,q}(H)$. In fact, for $\varepsilon_k \in \{0,1\}$, $\sum_{k \in \Delta_j} |\varepsilon_k|^r \leq 2^{2j}$ and

$$2^{j(s+1-\frac{2}{r})} a_j \left(\sum_{k \in \Delta_j} |\varepsilon_k|^r \right)^{\frac{1}{r}} \leq 1.$$

By Lemma 1.1, $\|a_j \sum_{k \in \Delta_j} \varepsilon_k \psi^1_{j,k}\|_{B^s_{r,q}} \leq H$. This with $g_0 \in B^s_{r,q}(H)$ implies $g_\varepsilon \in B^s_{r,q}(H)$.

According to Lemma 2.4 (Varshamov–Gilbert theorem), for $\Omega = \{\varepsilon = (\varepsilon_k)_{k \in \Delta_j}, \varepsilon_k \in \{0,1\}\}$, there exists a subset $\{\varepsilon^{(0)}, \varepsilon^{(1)}, \ldots, \varepsilon^{(M)}\}$ of Ω such that $M \geq 2^{\frac{2^{2j}}{8}}$, $\varepsilon^{(0)} = (0,0,\ldots,0)$, and for $m, n = 0, 1, \ldots, M$, $m \neq n$,

$$\sum_{k \in \Delta_j} \left| \varepsilon^{(m)}_k - \varepsilon^{(n)}_k \right| \geq \frac{2^{2j}}{8}. \tag{3.1}$$

Denote $\wedge' := \{g_{\varepsilon^{(0)}}, g_{\varepsilon^{(1)}}, \ldots, g_{\varepsilon^{(M)}}\}$. Then $\wedge' \subseteq \wedge$, and for $g_{\varepsilon^{(m)}}, g_{\varepsilon^{(n)}} \in \wedge'$,

$$\|g_{\varepsilon^{(m)}} - g_{\varepsilon^{(n)}}\|^p_p = a^p_j \sum_{k \in \Delta_j} \left| \varepsilon^m_k - \varepsilon^n_k \right|^p \|\psi^1_{j,k}\|^p_p = 2^{-2(sp+1)j} \|\psi^1\|^p_p \sum_{k \in \Delta_j} \left| \varepsilon^m_k - \varepsilon^n_k \right|^p,$$

since the supports of $\psi^1_{j,k}$ ($k \in \Delta_j$) are mutually disjoint. This with (3.1) leads to

$$\|g_{\varepsilon^{(m)}} - g_{\varepsilon^{(n)}}\|^p_p \geq C_1 2^{-2psj} := \delta^p_j.$$

Define

$$A_{\varepsilon^{(i)}} := \left\{ \|\hat{f}_n - g_{\varepsilon^{(i)}}\|_p < \frac{\delta_j}{2} \right\},$$

$i = 0, 1, 2, \ldots, M$. Then $A_{\varepsilon^{(m)}} \cap A_{\varepsilon^{(n)}} = \emptyset$ for $m \neq n$. Denote by P^n_f the probability measure with the density $f^n(x,y) := \prod^n_{i=1} f(x_i, y_i)$. By the construction of $g_{\varepsilon^{(i)}}$, $P^n_{g_{\varepsilon^{(i)}}} \ll P^n_{g_0}$. Then it follows from Lemma 2.5 (Fano's lemma) that

$$\sup_{0 \leq i \leq M} P^n_{g_{\varepsilon^{(i)}}} \left(\|\hat{f}_n - g_{\varepsilon^{(i)}}\|_p \geq \frac{\delta_j}{2} \right) \geq \sup_{0 \leq i \leq M} P^n_{g_{\varepsilon^{(i)}}} \left(A^c_{\varepsilon^{(i)}} \right) \geq \min \left\{ \frac{1}{2}, \sqrt{M} e^{-\frac{3}{e}} e^{-\mathcal{K}_M} \right\}.$$

Furthermore,

$$E\|\hat{f}_n - g_{\varepsilon^{(i)}}\|^p_p \geq \frac{\delta^p_j}{2^p} P^n_{g_{\varepsilon^{(i)}}} \left(\|\hat{f}_n - g_{\varepsilon^{(i)}}\|_p \geq \frac{\delta_j}{2} \right) \geq 2^{-2psj} P^n_{g_{\varepsilon^{(i)}}} \left(A^c_{\varepsilon^{(i)}} \right).$$

Taking $2^j \sim n^{\frac{1}{2(2s+1)}}$, we obtain that

$$\sup_{0 \leq i \leq M} E\|\hat{f}_n - g_{\varepsilon^{(i)}}\|^p_p \geq 2^{-2psj} \sup_{0 \leq i \leq M} P^n_{g_{\varepsilon^{(i)}}} \left(A^c_{\varepsilon^{(i)}} \right) \geq n^{-\frac{ps}{2s+1}} \min \left\{ \frac{1}{2}, \sqrt{M} e^{-\frac{3}{e}} e^{-\mathcal{K}_M} \right\} \tag{3.2}$$

with $\mathcal{K}_M := \inf_{0 \leq v \leq M} \frac{1}{M} \sum_{i \neq v} K(P^n_{g_{\varepsilon(i)}}, P^n_{g_{\varepsilon(v)}})$. By the definition of Kullback divergence,

$$
\begin{aligned}
K\left(P^n_{g_{\varepsilon(i)}}, P^n_{g_0}\right) &= \int_{\mathbb{R}^{2n}} \left[\ln \frac{\prod_{i=1}^n g_{\varepsilon(i)}(x_i, y_i)}{\prod_{i=1}^n g_0(x_i, y_i)} \right] \prod_{i=1}^n g_{\varepsilon(i)}(x_i, y_i) \, dx_1 \, dy_1 \, dx_2 \, dy_2 \cdots dx_n \, dy_n \\
&= n \int_{\mathbb{R}^2} g_{\varepsilon(i)}(x_1, y_1) \ln \frac{g_{\varepsilon(i)}(x_1, y_1)}{g_0(x_1, y_1)} \, dx_1 \, dy_1 \\
&\leq n \int_{\mathbb{R}^2} g_{\varepsilon(i)}(x_1, y_1) \left[\frac{g_{\varepsilon(i)}(x_1, y_1)}{g_0(x_1, y_1)} - 1 \right] dx_1 \, dy_1,
\end{aligned}
\tag{3.3}
$$

where we applied the inequality $\ln u \leq u - 1$ for $u > 0$ in the last inequality. Note that

$$
\begin{aligned}
&\int_{\mathbb{R}^2} g_{\varepsilon(i)}(x_1, y_1) \left[\frac{g_{\varepsilon(i)}(x_1, y_1)}{g_0(x_1, y_1)} - 1 \right] dx_1 \, dy_1 \\
&= \int_{\mathbb{R}^2} \left[g_0(x_1, y_1) \right]^{-1} \left[g_{\varepsilon(i)}(x_1, y_1) - g_0(x_1, y_1) \right]^2 dx_1 \, dy_1
\end{aligned}
$$

and $g_0(x_1, y_1) = c_0$ for $(x_1, y_1) \in [0, 2N-1] \times [-N+1, N]$. Combining this with the Parseval identity, we reduce (3.3) to

$$
K\left(P^n_{g_{\varepsilon(i)}}, P^n_{g_0}\right) \leq n c_0^{-1} a_j^2 \left\| \sum_{k \in \Delta_j} \varepsilon_k^i \psi_{j,k}^1(x, y) \right\|_2^2 = n c_0^{-1} a_j^2 \sum_{k \in \Delta_j} \left| \varepsilon_k^i \right|^2 \leq n c_0^{-1} a_j^2 2^{2j}.
\tag{3.4}
$$

Hence

$$
\mathcal{K}_M \leq \frac{1}{M} \sum_{i=1}^M K\left(P^n_{g_{\varepsilon(i)}}, P^n_{g_0}\right) \leq c_0^{-1} n a_j^2 2^{2j}.
$$

On the other hand, $2^j \sim n^{\frac{1}{2(2s+1)}}$ implies $n a_j^2 \leq C$. Then it follows from $M \geq 2^{\frac{2^{2j}}{8}} \geq e^{2^{2j} \frac{\ln 2}{8}}$ that

$$
\sqrt{M} e^{-\mathcal{K}_M} \geq e^{2^{2j} \frac{\ln 2}{16} - c_0^{-1} C 2^{2j}} \geq 1
$$

by choosing $C > 0$ such that $C < \frac{\ln 2}{16} c_0$. This with (3.2) leads to

$$
\sup_{0 \leq i \leq M} E \|\hat{f}_n - g_{\varepsilon(i)}\|_p^p \geq n^{-\frac{ps}{2s+1}} \min \left\{ \frac{1}{2}, \sqrt{M} e^{-\frac{3}{e}} e^{-\mathcal{K}_M} \right\} \gtrsim n^{-\frac{ps}{2s+1}}.
\tag{3.5}
$$

Now, it remains to show that

$$
\sup_{f_* \in B^s_{r,q}(H)} E \|\hat{f}_n - f_*\|_p^p \geq C \left(\frac{\ln n}{n} \right)^{\frac{(s - \frac{2}{r} + \frac{2}{p}) p}{2(s - \frac{2}{r}) + 1}}.
\tag{3.6}
$$

Similarly to the proof of (3.5), we construct the family of density functions $\{g_k, k \in \Delta_j\}$ as follows:

$$
g_k(x, y) := g_0(x, y) + a_j \psi_{j,k}^1(x, y), \quad k \in \Delta_j,
$$

where $a_j := 2^{-j(s+1-\frac{2}{r})}$. Obviously, $\int_{\mathbb{R}^2} g_k(x,y)\,dx\,dy = \int_{\mathbb{R}^2} g_0(x,y)\,dx\,dy = 1$, and

$$g_k(x,y)|_{[0,2N-1]\times[-N+1,N]} \geq c_0 - 2^{-j(s-\frac{2}{r})}\|\psi^1\|_\infty > 0$$

for large j since $s > \frac{2}{r}$. Then g_k is a bivariate density function for fixed $k \in \Delta_j$. From the proof of (3.5) we know that $g_0 \in B_{r,q}^s(H)$. This with

$$\|a_j\psi_{j,k}^1\|_{B_{r,q}^s} \sim a_j 2^{j(s+1-\frac{2}{r})} \leq 1$$

implies $g_k \in B_{r,q}^s(H)$ for $k \in \Delta_j$.

To prove (3.6), we need to show that

$$\sup_{k\in\Delta_j} E\|\hat{f}_n - g_k\|_p^p \geq C\left(\frac{\ln n}{n}\right)^{\frac{(s-\frac{2}{r}+\frac{2}{p})p}{2(s-\frac{2}{r})+1}}. \tag{3.7}$$

When $k \neq k' \in \Delta_j$, $\operatorname{supp}\psi_{j,k}^1 \cap \operatorname{supp}\psi_{j,k'}^1 = \emptyset$ and

$$\|g_k - g_{k'}\|_p^p = a_j^p\|\psi_{j,k}^1 - \psi_{j,k'}^1\|_p^p = 2a_j^p 2^{j(p-2)}\|\psi^1\|_p^p = 2 \cdot 2^{-j(s-\frac{2}{r}+\frac{2}{p})p}\|\psi^1\|_p^p.$$

Moreover,

$$\|g_k - g_{k'}\|_p = 2^{\frac{1}{p}}\|\psi^1\|_p 2^{-j(s-\frac{2}{r}+\frac{2}{p})} := \delta_j.$$

Define $B_k := \{\|\hat{f}_n - g_k\|_p < \frac{\delta_j}{2}\}$. Then $B_k \cap B_{k'} = \emptyset$ $(k \neq k')$. According to Lemma 2.5 (Fano's lemma), we find that

$$\sup_{k\in\Delta_j} P_{g_k}^n\left(\|\hat{f}_n - g_k\|_p \geq \frac{\delta_j}{2}\right) \geq \min\left\{\frac{1}{2}, \sqrt{M}e^{-3e^{-1}}e^{-\mathcal{K}_M}\right\}, \tag{3.8}$$

where $M = \sharp\Delta_j$ and $\mathcal{K}_M := \inf_{0\leq\nu\leq M}\frac{1}{M}\sum_{k\neq\nu}K(P_{g_k}^n, P_{g_\nu}^n) \leq \frac{1}{M}\sum_{k\neq 0}K(P_{g_k}^n, P_{g_0}^n)$. Similar to (3.3)–(3.4), we conclude that

$$K\left(P_{g_k}^n, P_{g_\nu}^n\right) \leq n\int_{\mathbb{R}^2}\left[g_0(x,y)\right]^{-1}\left[g_k(x,y) - g_0(x,y)\right]^2 dx\,dy \leq c_0^{-1}C_1 n a_j^2.$$

Hence $\mathcal{K}_M \leq c_0^{-1}C_1 n a_j^2$. By taking $2^j \sim (\frac{n}{\ln n})^{\frac{1}{2(s-\frac{2}{r})+1}}$ we obtain that $\ln 2^j \geq C'\ln n$ and $e^{-\mathcal{K}_M} \geq e^{-c_0^{-1}C_1 n a_j^2} \geq e^{-c_0^{-1}C\ln n}$, thanks to $na_j^2 \leq C_2\ln n$ $(C = C_1 C_2)$. Moreover, choosing C_1 and C' such that $C' > c_0^{-1}C$, we have

$$\sqrt{M}e^{-3e^{-1}}e^{-\mathcal{K}_M} \gtrsim e^{\ln 2^j}e^{-3e^{-1}}e^{-\mathcal{K}_M} \geq e^{C'\ln n - c_0^{-1}C\ln n - 3e^{-1}} \gtrsim 1$$

due to $M \sim 2^{2j}$. This with (3.8) implies $\sup_{k\in\Delta_j} P_{g_k}^n(\|\hat{f}_n - g_k\|_p \geq \frac{\delta_j}{2}) \gtrsim 1$. Furthermore,

$$\sup_{k\in\Delta_j} E\|\hat{f}_n - g_k\|_p^p \geq \frac{\delta_j^p}{2^p}P_{g_k}\left(\|\hat{f}_n - g_k\|_p \geq \frac{\delta_j}{2}\right)e^{-3e^{-1}} \gtrsim \delta_j^p.$$

Then the desired conclusion (3.7) follows from $\delta_j := 2^{\frac{1}{p}} \|\psi^1\|_p 2^{-j(s-\frac{2}{r}+\frac{2}{p})}$ and the choice of $2^j \sim (\frac{n}{\ln n})^{\frac{1}{2(s-\frac{2}{r})+1}}$. This completes the proof. □

4 Proofs of upper bounds

In this section, we prove the upper bounds of wavelet estimators. The result of the linear one is derived firstly. We restate and prove Theorem 1.2 as Theorem 4.1.

Theorem 4.1 *Let $\hat{f}_n^{\lin}$ be the linear estimator of $f_* \in B_{r,q}^s(H,Q)$ defined in (1.3) with $1 \le r, q < \infty, s > 0$. If the density of X is bounded, then for $\{r \ge p \ge 1\}$ or $\{r \le p < \infty$ and $s > \frac{2}{r}\}$,*

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\|\hat{f}_n^{\lin} - f_*\|_p^p \lesssim n^{-\frac{ps'}{2s'+1}}$$

with $s' = s - (\frac{2}{r} - \frac{2}{p})_+$ and $x_+ := \max\{x, 0\}$.

Proof When $r \le p$, $s' := s - (\frac{2}{r} - \frac{2}{p})_+ = s - \frac{2}{r} + \frac{2}{p}$ and $B_{r,q}^s(\mathbb{R}^2) \hookrightarrow B_{p,q}^{s'}(\mathbb{R}^2)$ thanks to Remark 1.2. Then

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\|\hat{f}_n^{\lin} - f_*\|_p^p \lesssim \sup_{f_* \in B_{r,q}^{s'}(H,Q)} E\|\hat{f}_n^{\lin} - f_*\|_p^p.$$

When $r > p$ and f_* has a compact support, then $\hat{f}_n^{\lin}$ does due to φ having the same property. By the Hölder inequality,

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\|\hat{f}_n^{\lin} - f_*\|_p^p \lesssim \sup_{f_* \in B_{r,q}^s(H,Q)} \left(E\|\hat{f}_n^{\lin} - f_*\|_r^r\right)^{\frac{p}{r}}.$$

Because $s' = s$ in that case, it is sufficient to prove that

$$\sup_{f_* \in B_{r,q}^{s'}(H,Q)} E\|\hat{f}_n^{\lin} - f_*\|_p^p \lesssim n^{-\frac{ps'}{2s'+1}} \tag{4.1}$$

for the conclusion of Theorem 4.1.

Recall that $\hat{f}_n^{\lin} := \sum_{k \in \wedge_{j_0}} \hat{\alpha}_{j_0,k} \phi_{j_0,k}$. Then by Lemma 2.1 we conclude that

$$E\|\hat{f}_n^{\lin} - E(\hat{f}_n^{\lin})\|_p^p = E\left\|\sum_{k \in \wedge_{j_0}} (\hat{\alpha}_{j_0,k} - \alpha_{j_0,k})\varphi_{j_0,k}\right\|_p^p \lesssim 2^{j_0(p-2)} \sum_{k \in \wedge_{j_0}} E|\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^p$$

due to Lemma 1.2. It follows from Lemma 2.2 and $\sharp \wedge_{j_0} \lesssim 2^{2j_0}$ that

$$E\|\hat{f}_n^{\lin} - E(\hat{f}_n^{\lin})\|_p^p \lesssim 2^{j_0(p-2)} 2^{2j_0} 2^{-\frac{pj_0}{2}} n^{-\frac{p}{2}} \lesssim 2^{\frac{pj_0}{2}} n^{-\frac{p}{2}} \lesssim n^{-\frac{ps'}{2s'+1}} \tag{4.2}$$

thanks to the choice of $2^{j_0} \sim n^{\frac{1}{2s'+1}}$.

On the other hand, by Lemma 2.1, $E(\hat{f}_n^{\lin}) = \sum_{k \in \wedge_{j_0}} \alpha_{j_0,k} \varphi_{j_0,k} = P_{j_0} f_*$. Combining this with $f_* \in B_{p,q}^{s'}(\mathbb{R}^2)$ and Remark 1.1, we get that

$$\|E(\hat{f}_n^{\lin}) - f_*\|_p^p = \|P_{j_0} f_* - f_*\|_p^p \lesssim 2^{-j_0 ps'}.$$

Taking $2^{j_0} \sim n^{\frac{1}{2s'+1}}$, it is easy to show

$$\left\| E\big(\hat{f}_n^{\mathrm{lin}}\big) - f_* \right\|_p^p \lesssim n^{-\frac{ps'}{2s'+1}}. \tag{4.3}$$

Hence, by (4.2)–(4.3),

$$\sup_{f_* \in B_{r,q}^{s'}(H,Q)} E\big\| \hat{f}_n^{\mathrm{lin}} - f_* \big\|_p^p \lesssim \sup_{f_* \in B_{r,q}^{s'}(H,Q)} E\big\| \hat{f}_n^{\mathrm{lin}} - E\big(\hat{f}_n^{\mathrm{lin}}\big) \big\|_p^p + \sup_{f_* \in B_{r,q}^{s'}(H,Q)} \big\| E\big(\hat{f}_n^{\mathrm{lin}}\big) - f_* \big\|_p^p$$

$$\lesssim n^{-\frac{ps'}{2s'+1}},$$

which means that (4.1) holds. The proof is done. $\qquad\square$

Next, we are in a position to prove the conclusion of the nonlinear one.

Theorem 4.2 *Let $\hat{f}_n^{\mathrm{non}}$ be the nonlinear estimator of $f_* \in B_{r,q}^s(H,Q)$ defined in (1.4) with $1 \le r, q < \infty$, $s > 0$. If the density of X is bounded, then for $\{r \ge p \ge 1\}$ or $\{r \le p < \infty$ and $s > \frac{2}{r}\}$,*

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\big\| \hat{f}_n^{\mathrm{non}} - f_* \big\|_p^p \lesssim (\ln n)^p \left(\frac{\ln n}{n} \right)^{\alpha p}$$

with $\alpha := \min\{ \frac{s}{2s+1}, \frac{s - \frac{2}{r} + \frac{2}{p}}{2(s-\frac{2}{r})+1} \}$.

Proof We only need to prove the case $r \le p$. In fact, when $r > p$, $\hat{f}_n^{\mathrm{non}}$ has a compact support because of φ, ψ, and f_* have the same property. By the Hölder inequality,

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\big\| \hat{f}_n^{\mathrm{non}} - f_* \big\|_p^p \lesssim \sup_{f_* \in B_{r,q}^s(H,Q)} \big(E\big\| \hat{f}_n^{\mathrm{non}} - f_* \big\|_r^r \big)^{\frac{p}{r}}.$$

Using Theorem 4.2 for the case $r = p$, we find that $\sup_{f_* \in B_{r,q}^s(H,Q)} E\| \hat{f}_n^{\mathrm{non}} - f_* \|_r^r \lesssim (\ln n)^r (\frac{\ln n}{n})^{\alpha r}$, and therefore

$$\sup_{f_* \in B_{r,q}^s(H,Q)} E\big\| \hat{f}_n^{\mathrm{non}} - f_* \big\|_p^p \lesssim (\ln n)^p \left(\frac{\ln n}{n} \right)^{\alpha p}.$$

It remains to estimate the case $r \le p$. Recall that

$$\hat{f}_n^{\mathrm{non}} - f_* = \big(\hat{f}_n^{lin} - P_{j_0} f_* \big) + \big(P_{j_1+1} f_* - f_* \big) + \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} \sum_{k \in \wedge_j} \big(\hat{\beta}_{j,k}^i 1_{\{|\hat{\beta}_{j,k}^i| > \lambda_j\}} - \beta_{j,k}^i \big) \psi_{j,k}^i$$

with $\lambda_j = T 2^{-\frac{j}{2}} \sqrt{\frac{\ln n}{n}}$. Denote $f_{j_0,j_1} := \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} \sum_{k \in \wedge_j} (\hat{\beta}_{j,k}^i 1_{\{|\hat{\beta}_{j,k}^i| > \lambda_j\}} - \beta_{j,k}^i) \psi_{j,k}^i$. Then

$$E\big\| \hat{f}_n^{\mathrm{non}} - f_* \big\|_p^p \lesssim E\big\| \hat{f}_n^{\mathrm{lin}} - P_{j_0} f_* \big\|_p^p + \| P_{j_1+1} f_* - f_* \|_p^p + E\| f_{j_0,j_1} \|_p^p. \tag{4.4}$$

From the proof of Theorem 4.1 we obtain that

$$E\big\|\hat{f}_n^{\mathrm{lin}} - P_{j_0}f_*\big\|_p^p \lesssim 2^{\frac{j_0 p}{2}} n^{-\frac{p}{2}} \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p}$$

and

$$\|P_{j_1+1}f_* - f_*\|_p^p \lesssim 2^{-j_1 p s'} \lesssim \left(\frac{\ln n}{n}\right)^{\alpha p} \tag{4.5}$$

due to $2^{j_0} \sim n^{\frac{1}{2m+1}}$, $2^{j_1} \sim \frac{n}{\ln n}$ and $\alpha = \min\{\frac{s}{2s+1}, \frac{s-\frac{2}{r}+\frac{2}{p}}{2(s-\frac{2}{r})+1}\}$.

By $f_{j_0,j_1} := \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} \sum_{k\in\wedge_j} (\hat{\beta}_{j,k}^i 1_{\{|\hat{\beta}_{j,k}^i|>\lambda_j\}} - \beta_{j,k}^i)\psi_{j,k}^i$ and Lemma 1.2,

$$E\|f_{j_0,j_1}\|_p^p \lesssim (j_1 - j_0 + 1)^{p-1} \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k\in\wedge_j} E\big|\hat{\beta}_{j,k}^i 1_{\{|\hat{\beta}_{j,k}^i|>\lambda_j\}} - \beta_{j,k}^i\big|^p.$$

On the other hand, it is easy to see that

$$\big|\hat{\beta}_{j,k}^i 1_{\{|\hat{\beta}_{j,k}^i|>\lambda_j\}} - \beta_{j,k}^i\big| = \big|\hat{\beta}_{j,k}^i - \beta_{j,k}^i\big|\big(1_{\{|\hat{\beta}_{j,k}^i|\geq\lambda_j,|\beta_{j,k}^i|<\lambda_j/2\}} + 1_{\{|\hat{\beta}_{j,k}^i|\geq\lambda_j,|\beta_{j,k}^i|\geq\lambda_j/2\}}\big)$$

$$+ \big|\beta_{j,k}^i\big|\big(1_{\{|\hat{\beta}_{j,k}^i|<\lambda_j,|\beta_{j,k}^i|>2\lambda_j\}} + 1_{\{|\hat{\beta}_{j,k}^i|<\lambda_j,|\beta_{j,k}^i|\leq2\lambda_j\}}\big)$$

and $1_{\{|\hat{\beta}_{j,k}^i|\geq\lambda_j,|\beta_{j,k}^i|<\lambda_j/2\}} \leq 1_{\{|\hat{\beta}_{j,k}^i-\beta_{j,k}^i|>\lambda_j/2\}}$. Then

$$E\|f_{j_0,j_1}\|_p^p \lesssim T_1 + T_2 + T_3 + T_4 \tag{4.6}$$

with

$$T_1 := (\ln n)^{p-1} \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k\in\wedge_j} E\big[\big|\hat{\beta}_{j,k}^i - \beta_{j,k}^i\big|^p 1_{\{|\hat{\beta}_{j,k}^i-\beta_{j,k}^i|>\lambda_j/2\}}\big],$$

$$T_2 := (\ln n)^{p-1} \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k\in\wedge_j} E\big[\big|\hat{\beta}_{j,k}^i - \beta_{j,k}^i\big|^p 1_{\{|\hat{\beta}_{j,k}^i|\geq\lambda_j,|\beta_{j,k}^i|\geq\lambda_j/2\}}\big],$$

$$T_3 := (\ln n)^{p-1} \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k\in\wedge_j} E\big[\big|\beta_{j,k}^i\big|^p 1_{\{|\hat{\beta}_{j,k}^i|<\lambda_j,|\beta_{j,k}^i|\leq2\lambda_j\}}\big],$$

$$T_4 := (\ln n)^{p-1} \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k\in\wedge_j} E\big[\big|\beta_{j,k}^i\big|^p 1_{\{|\hat{\beta}_{j,k}^i|<\lambda_j,|\beta_{j,k}^i|>2\lambda_j\}}\big].$$

When $|\hat{\beta}_{j,k}^i| < \lambda_j$ and $|\beta_{j,k}^i| > 2\lambda_j$, $|\hat{\beta}_{j,k}^i - \beta_{j,k}^i| \geq |\beta_{j,k}^i| - |\hat{\beta}_{j,k}^i| > |\hat{\beta}_{j,k}^i|/2$. Hence

$$\big|\beta_{j,k}^i\big|^p 1_{\{|\hat{\beta}_{j,k}^i|<\lambda_j,|\beta_{j,k}^i|>2\lambda_j\}} \lesssim \big|\hat{\beta}_{j,k}^i - \beta_{j,k}^i\big|^p 1_{\{|\hat{\beta}_{j,k}^i-\beta_{j,k}^i|>\lambda_j/2\}}.$$

Then (4.6) reduces to

$$E\|f_{j_0,j_1}\|_p^p \lesssim T_1 + T_2 + T_3. \tag{4.7}$$

By (4.4)–(4.5) and (4.7) it is sufficient to show

$$T_\ell \lesssim (\ln n)^p \left(\frac{\ln n}{n}\right)^{\alpha p}, \quad \ell = 1, 2, 3, \tag{4.8}$$

for the conclusion of Theorem 4.2.

To estimate T_1, using the Hölder inequality, we find that

$$T_1 \lesssim (\ln n)^{p-1} \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k \in \wedge_j} \left(E\big|\hat{\beta}_{j,k}^i - \beta_{j,k}^i\big|^{2p}\right)^{\frac{1}{2}} \left[E(1_{\{|\hat{\beta}_{j,k}^i - \beta_{j,k}^i| \ge \lambda_j/2\}})\right]^{\frac{1}{2}}.$$

Note that $E(1_{\{|\hat{\beta}_{j,k}^i - \beta_{j,k}^i| \ge \lambda_j/2\}}) = P(|\hat{\beta}_{j,k}^i - \beta_{j,k}^i| \ge \frac{\lambda_j}{2}) \le 2^{-\varepsilon j}$ due to Lemma 2.3. Taking ε such that $\varepsilon > p$, we conclude that

$$T_1 \lesssim (\ln n)^{p-1} n^{-\frac{p}{2}} \sum_{j=j_0}^{j_1} \sum_{i=1}^{3} 2^{\frac{p-\varepsilon}{2}j} \lesssim (\ln n)^{p-1} n^{-\frac{p}{2}} 2^{\frac{p}{2}j_0} \lesssim (\ln n)^{p-1} n^{-\frac{ps}{2s+1}}$$

thanks to Lemma 2.2, $\sharp \wedge_j \lesssim 2^{2j}$ and the choice of j_0. Hence (4.8) with $\ell = 1$ holds since $\alpha \le \frac{s}{2s+1}$.

To estimate T_2 and T_3, define

$$2^{j_0^*} \sim \left(\frac{n}{\ln n}\right)^{1-2\alpha}, \qquad 2^{j_1^*} \sim \left(\frac{n}{\ln n}\right)^{\frac{\alpha}{s - \frac{2}{r} + \frac{2}{p}}}.$$

Recall that $2^{j_0} \sim n^{\frac{1}{2m+1}}$, $2^{j_1} \sim \frac{n}{\ln n}$ and $\alpha := \min\{\frac{s}{2s+1}, \frac{s - \frac{2}{r} + \frac{2}{p}}{2(s - \frac{2}{r}) + 1}\}$. Then

$$1 - 2\alpha \ge \frac{1}{2s+1} > \frac{1}{2m+1} \quad \text{and} \quad \frac{\alpha}{s - \frac{2}{r} + \frac{2}{p}} \le \frac{1}{2(s - \frac{2}{r}) + 1} \le 1.$$

Hence $2^{j_0} \le 2^{j_0^*}$ and $2^{j_1^*} \le 2^{j_1}$. Moreover, a simple computation shows that $1 - 2\alpha \le \frac{\alpha}{s - \frac{2}{r} + \frac{2}{p}}$, which implies $2^{j_0^*} \le 2^{j_1^*}$.

Now, we estimate T_2 by dividing T_2 into

$$T_2 = (\ln n)^{p-1} \left(\sum_{j=j_0}^{j_0^*} + \sum_{j=j_0^*+1}^{j_1}\right) \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k \in \wedge_j} E\big[\big|\hat{\beta}_{j,k}^i - \beta_{j,k}^i\big|^p 1_{\{|\hat{\beta}_{j,k}^i| \ge \lambda_j, |\beta_{j,k}^i| \ge \lambda_j/2\}}\big]$$

$$:= t_1 + t_2. \tag{4.9}$$

Since $1_{\{|\hat{\beta}_{j,k}^i| \ge \lambda_j, |\beta_{j,k}^i| \ge \lambda_j/2\}} \le 1$, by Lemma 2.2 we know that

$$t_1 \lesssim (\ln n)^{p-1} \sum_{j=j_0}^{j_0^*} \sum_{i=1}^{3} 2^{\frac{pj}{2}} n^{-\frac{p}{2}} \lesssim (\ln n)^{p-1} n^{-\frac{p}{2}} 2^{\frac{j_0^*}{2}p} \lesssim (\ln n)^p \left(\frac{\ln n}{n}\right)^{\alpha p} \tag{4.10}$$

thanks to $\sharp \wedge_j \lesssim 2^{2j}$ and the choice of j_0^*. To estimate t_2, we observe that

$$1_{\{|\hat{\beta}_{j,k}^i| \ge \lambda_j, |\beta_{j,k}^i| \ge \frac{\lambda_j}{2}\}} \le 1_{\{|\beta_{j,k}^i| \ge \frac{\lambda_j}{2}\}} \lesssim \left(\frac{|\beta_{j,k}^i|}{\lambda_j}\right)^r.$$

This with Lemma 2.2 leads to

$$t_2 \lesssim (\ln n)^{p-1} \sum_{j=j_0^*+1}^{j_1} \sum_{i=1}^{3} 2^{j(\frac{p}{2}-2)} n^{-\frac{p}{2}} \sum_{k \in \wedge_j} \left(\frac{|\beta_{j,k}^i|}{\lambda_j} \right)^r. \tag{4.11}$$

Note that $\|\beta_{j,\cdot}\|_r \lesssim 2^{-j(s+1-\frac{2}{r})}$ because of $f_* \in B_{r,q}^s$ and Lemma 1.1. Then (4.11) reduces to

$$t_2 \lesssim (\ln n)^{p-\frac{r}{2}-1} n^{\frac{r-p}{2}} \sum_{j=j_0^*+1}^{j_1} 2^{-j(sr+\frac{r}{2}-\frac{p}{2})} \tag{4.12}$$

thanks to $\lambda_j = \frac{T}{2} 2^{-\frac{j}{2}} \sqrt{\frac{\ln n}{n}}$. Denote $\theta := sr + \frac{r}{2} - \frac{p}{2}$. When $\theta > 0$, $r > \frac{p}{2s+1}$ and

$$t_2 \lesssim (\ln n)^{p-\frac{r}{2}-1} n^{\frac{r-p}{2}} 2^{-j_0^*(sr+\frac{r}{2}-\frac{p}{2})} \lesssim (\ln n)^p \left(\frac{\ln n}{n} \right)^{\alpha p} \tag{4.13}$$

due to the choice of j_0^*. In (4.13), we use the fact $\alpha = \frac{s}{2s+1}$ in the case $r > \frac{p}{2s+1}$.

To show (4.13) for $\theta \leq 0$, define $r_1 := (1-2\alpha)p > 0$. Then $\alpha = \frac{s-\frac{2}{r}+\frac{2}{p}}{2(s-\frac{2}{r})+1} \leq \frac{s}{2s+1}$ and $r \leq \frac{p}{2s+1} \leq (1-2\alpha)p = r_1$ because $\theta \leq 0$. The same arguments as (4.11) show that

$$t_2 \lesssim (\ln n)^{p-1} \sum_{j=j_0^*+1}^{j_1} \sum_{i=1}^{3} 2^{j(\frac{p}{2}-2)} n^{-\frac{p}{2}} \sum_{k \in \wedge_j} \left(\frac{|\beta_{j,k}^i|}{\lambda_j} \right)^{r_1}.$$

It follows from $f_* \in B_{r,q}^s$ and Lemma 1.1 that

$$\|\beta_{j,\cdot}\|_{r_1} \leq \|\beta_{j,\cdot}\|_r \leq 2^{-j(s+1-\frac{2}{r})}$$

due to $r \leq r_1$. Therefore, similarly to (4.12), we get that

$$t_2 \lesssim (\ln n)^{p-\frac{r_1}{2}-1} n^{\frac{r_1-p}{2}} \sum_{j=j_0^*+1}^{j_1} 2^{j[\frac{p-2}{2}-(s-\frac{2}{r}-\frac{1}{2})r_1]}.$$

Note that $\frac{p}{2} - 2 - (s - \frac{2}{r} + \frac{1}{2})r_1 = 0$ because of $r_1 = (1-2\alpha)p$ and $\alpha = \frac{s-\frac{2}{r}+\frac{2}{p}}{2(s-\frac{2}{r})+1}$. Then

$$t_2 \lesssim (\ln n)^{p-\frac{r_1}{2}-1} n^{\frac{r_1-p}{2}} \lesssim (\ln n)^p \left(\frac{\ln n}{n} \right)^{\alpha p}, \tag{4.14}$$

which implies that (4.13) holds for $\theta \leq 0$. The desired conclusion (4.8) with $\ell = 2$ follows from (4.9)–(4.10) and (4.13)–(4.14).

Finally, by splitting T_3 into

$$T_3 = (\ln n)^{p-1} \left(\sum_{j=j_0}^{j_0^*} + \sum_{j=j_0^*+1}^{j_1} \right) \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k \in \wedge_j} E\left[|\beta_{j,k}^i|^p 1_{\{|\hat{\beta}_{j,k}^i| < \lambda_j, |\beta_{j,k}^i| \leq 2\lambda_j\}} \right]$$

$$:= e_1 + e_2 \tag{4.15}$$

we obtain that

$$e_1 \lesssim (\ln n)^{p-1} \sum_{j=j_0}^{j_0^*} \sum_{i=1}^{3} 2^{jp}|\lambda_j|^p \lesssim (\ln n)^{\frac{3}{2}p-1} n^{-\frac{p}{2}} 2^{\frac{j_0^* p}{2}} \lesssim (\ln n)^p \left(\frac{\ln n}{n}\right)^{\alpha p} \tag{4.16}$$

thanks to $\sharp \wedge_j \lesssim 2^{2j}$ and the choice of λ_j and j_0^*.

To estimate e_2, we use the fact $1_{\{|\hat{\beta}_{j,k}^i| \le \lambda_j, |\beta_{j,k}^i| \le 2\lambda_j\}} \le (\frac{2\lambda_j}{|\beta_{j,k}^i|})^{p-r}$ because of $r \le p$. Similarly to (4.11)–(4.13),

$$e_2 \lesssim (\ln n)^p \left(\frac{\ln n}{n}\right)^{\alpha p} \tag{4.17}$$

for $\theta > 0$, where $\theta := sr + \frac{r}{2} - \frac{p}{2}$. When $\theta \le 0$, we rewrite e_2 as follows:

$$e_2 = (\ln n)^{p-1} \left(\sum_{j=j_0^*+1}^{j_1^*} + \sum_{j=j_1^*+1}^{j_1}\right) \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k \in \wedge_j} E\big[|\beta_{j,k}^i|^p 1_{\{|\hat{\beta}_{j,k}^i| < \lambda_j, |\beta_{j,k}^i| \le 2\lambda_j\}}\big]$$

$$:= e_1^* + e_2^*. \tag{4.18}$$

Proceeding as in (4.11) and (4.12), we find that

$$e_1^* \lesssim (\ln n)^{p-1} \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2}} \sum_{j=j_0^*+1}^{j_1^*} 2^{-j(sr+\frac{r-p}{2})} \lesssim (\ln n)^{p-1} \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2}} 2^{-j_1^*(sr+\frac{r-p}{2})}.$$

This with the choice of $2^{j_1^*} \sim (\frac{n}{\ln n})^{\frac{\alpha}{s-\frac{2}{r}+\frac{2}{p}}}$ leads to

$$e_1^* \lesssim (\ln n)^p \left(\frac{\ln n}{n}\right)^{\alpha p} \tag{4.19}$$

due to $\alpha = \frac{s-\frac{2}{r}+\frac{2}{p}}{2(s-\frac{2}{r})+1}$ for $\theta \le 0$. When $r \le p$,

$$\|\beta_{j,\cdot}\|_p \le \|\beta_{j,\cdot}\|_r \lesssim 2^{-j(s+1-\frac{2}{r})}$$

thanks to $f_* \in B_{r,q}^s$ and Lemma 1.1. Therefore

$$e_2^* \lesssim (\ln n)^{p-1} \sum_{j=j_1^*+1}^{j_1} \sum_{i=1}^{3} 2^{j(p-2)} \sum_{k \in \wedge_j} |\beta_{j,k}^i|^p \lesssim (\ln n)^{p-1} \sum_{j=j_1^*+1}^{j_1} 2^{-j(sp-\frac{2p}{r}+2)}.$$

Combining this with the choice of $2^{j_1^*} \sim (\frac{n}{\ln n})^{\frac{\alpha}{s-\frac{2}{r}+\frac{2}{p}}}$, we observe that

$$e_2^* \lesssim (\ln n)^{p-1} 2^{-j_1^*(sp-\frac{2p}{r}+2)} \lesssim (\ln n)^p \left(\frac{\ln n}{n}\right)^{\alpha p}.$$

This with (4.19) implies that (4.17) holds for $\theta \leq 0$. Hence

$$T_3 \lesssim (\ln n)^{p-1} \left(\frac{\ln n}{n} \right)^{\alpha p}$$

follows from (4.15)–(4.17).

Therefore, the desired conclusion can be concluded by (4.4)–(4.8) with $\ell = 1, 2, 3$, which completes the proof. $\square$

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors finish this work together. All authors read and approved the final manuscript.

Author details

[1]Department of Basic Courses, Beijing Union University, Beijing, P.R. China. [2]Department of Applied Mathematics, Beijing University of Technology, Beijing, P.R. China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ahmad, I.A., Cerrito, P.B.: Nonparametric estimation of joint discrete-continuous probability densities with applications. J. Stat. Plan. Inference **41**, 349–364 (1994)
2. Chesneau, C., Dewan, I., Doosti, H.: Nonparametric estimation of a two dimensional continuous-discrete density function. Stat. Methodol. **18**, 64–78 (2014)
3. Chesneau, C., Doosti, H.: A note on the adaptive estimation of a conditional continuous-discrete multivariate density by wavelet methods. Chin. J. Math. **2016**, 6204874 (2016)
4. Crowder, M.J.: Classical Competing Risks. Chapman & Hall, London (2001)
5. Daubechies, I.: Ten Lectures on Wavelets. SIAM, Philadelphia (1992)
6. Devore, R., Kerkyacharian, G., Piard, D., Temlyakov, V.: Approximation methods for supervised learning. Found. Comput. Math. **6**, 3–58 (2006)
7. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D.: Density estimation by wavelet thresholding. Ann. Stat. **24**, 508–539 (1996)
8. Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.: Wavelets, Approximation and Statistical Applications. Springer, New York (1998)
9. Kou, J.K., Liu, Y.M.: An extension of Chesneau's theorem. Stat. Probab. Lett. **108**, 23–32 (2016)
10. Li, Q., Maasoumi, E., Racine, J.S.: A nonparametric test for equality of distributions with mixed categorical and continuous data. J. Econom. **148**, 186–200 (2009)
11. Li, R., Liu, Y.M.: Wavelet optimal estimations for a density with some additive noises. Appl. Comput. Harmon. Anal. **32**, 416–433 (2014)
12. Liu, Y.M., Wang, H.Y.: Convergence order of wavelet thresholding estimator for differential operators on Besov spaces. Appl. Comput. Harmon. Anal. **32**, 342–356 (2012)
13. Meyer, Y.: Wavelets and Operators. Cambridge University Press, Cambridge (1992)
14. Ouyang, D., Li, Q., Racine, J.: Cross-validation and the estimation of probability distributions with categorical data. J. Nonparametr. Stat. **18**, 69–100 (2006)
15. Zeng, X.C.: A note on wavelet deconvolution density estimation. Int. J. Wavelets Multiresolut. Inf. Process. **15**(6), 1750055 (2017)