mathematics

MDPI

Article

Adaptive Wavelet Estimations in the Convolution Structure Density Model

Kaikai Cao [1,*] and Xiaochen Zeng [2]

[1] School of Mathematics and Information Science, Weifang University, Weifang 261061, China
[2] School of Mathematics, Faculty of Science, Beijing University of Technology, Beijing 100124, China; zengxiaochen@bjut.edu.cn
* Correspondence: caokaikai@wfu.edu.cn

Received: 23 June 2020; Accepted:17 August 2020; Published: 19 August 2020

Abstract: Using kernel methods, Lepski and Willer study a convolution structure density model and establish adaptive and optimal L^p risk estimations over an anisotropic Nikol'skii space (Lepski, O.; Willer, T. Oracle inequalities and adaptive estimation in the convolution structure density model. *Ann. Stat.* **2019**, *47*, 233–287). Motivated by their work, we consider the same problem over Besov balls by wavelets in this paper and first provide a linear wavelet estimate. Subsequently, a non-linear wavelet estimator is introduced for adaptivity, which attains nearly-optimal convergence rates in some cases.

Keywords: generalized deconvolution; adaptive density estimation; wavelet; Besov space

1. Introduction

The estimation of a probability density from independent and identically distributed (i.i.d.) random observations $X_1, X_2, \cdots, X_n$ of X is a classical problem in statistics. The representative work is Donoho et al. [1], they established an adaptive and nearly-optimal estimate (up to a logarithmic factor) over Besov spaces using wavelets.

However, the observed data are always polluted by noises in many real-life applications. One of the important problems is the density estimation with an additive noise. Let $Z_1, Z_2, \cdots, Z_n$ be i.i.d. random variables and have the same distribution as

$$Z = X + Y, \tag{1}$$

where X denotes a real-valued random variable with unknown probability density function f and Y stands for an independent random noise (error) with a known probability density g. The problem is to estimate f by $Z_1, Z_2, \cdots, Z_n$ in some sense. Moreover, it is also called a denconvolution problem (model), because the density h of Z equals the convolution of f and g. Fan and Koo [2] studied the MISE performance (L^2-risk) of linear wavelet deconvolution estimator over a Besov ball. The L^∞ risk optimal wavelet estimations were investigated by Lounici and Nickl [3]. Furthermore, Li and Liu [4] provided L^p $(1 \le p \le \infty)$ risk optimal deconvolution estimations using wavelet bases.

In this paper, we consider a generalized deconvolution model introduced by Lepski & Willer [5,6]. More precisely, let $(\Omega, \mathcal{F}, P)$ be a probability space and $Z_1, Z_2, \cdots, Z_n$ be i.i.d. random variables having the same distribution as

$$Z = X + \varepsilon Y, \tag{2}$$

where the symbols X and Y are same as model (1), f and g are the corresponding densities respectively. Moreover, the biggest difference with model (1) is that a Bernoulli random variable $\varepsilon \in \{0, 1\}$ with

$P\{\varepsilon = 1\} = \alpha$ is added in (2), and $\alpha \in [0,1]$ is known. The problem is also to estimate f by the observed data $Z_1, Z_2, \cdots, Z_n$ in some sense.

When $\alpha = 1$, model (2) reduces to the deconvolution one (see [2–4,7,8] et al.), while $\alpha = 0$ corresponds to the classical density model with no errors (see [1,9–11] et al.). Clearly, the density function h of Z in (2) satisfies

$$h = (1-\alpha)f + \alpha f * g.$$

Here, $f * g$ stands for the convolution of f and g. Furthermore, when the function $G_\alpha(t) := 1 - \alpha + \alpha g^{ft}(t) \neq 0$ for $t \in \mathbb{R}$, we have

$$f^{ft}(t) = [(1-\alpha) + \alpha g^{ft}(t)]^{-1} h^{ft}(t) = (G_\alpha(t))^{-1} h^{ft}(t),$$

where g, α are known and f^{ft} is the Fourier transform of $f \in L^1(\mathbb{R})$ given by

$$f^{ft}(t) := \int_{\mathbb{R}} f(x) e^{-itx} dx.$$

Based on the model (2) with some mild assumptions on G_α, Lepski and Willer [5] provided a lower bound estimation over L^p risk on an anistropic Nikol'skii space. Moreover, they investigated an adaptive and optimal L^p estimate by using kernel method in Ref. [6]. Recently, Wu et al. [12] established a pointwise lower bound estimation for model (2) under the local Hölder condition.

When compared with the classical kernel estimation of density functions, the wavelet estimations provide more local information and fast algorithm [13]. We will consider the L^p $(1 \leq p < \infty)$ risk estimations under the model (2) over Besov balls by using wavelets and expect to obtain the corresponding convergence rates.

The same as Assumption 4 in [6], we also need the following condition on Y,

$$|G_\alpha(t)| \gtrsim (1 + |t|^2)^{-\frac{\beta(\alpha)}{2}} \tag{3}$$

with $\beta(\alpha) = \beta \geq 0$ for $\alpha = 1$ and $\beta(\alpha) = 0$ for others. It is reasonable, because it holds automatically for $\alpha = 0$, while the same condition for $\alpha = 1$ is necessary for the deconvolution estimations [4,7]. In addition, when $0 < \alpha < \frac{1}{2}$, $\beta(\alpha) = 0$ and $|G_\alpha(t)| \geq 1 - \alpha - \alpha|g^{ft}(t)| \gtrsim 1$ thanks to $\|g^{ft}\|_\infty \leq 1$. In fact, the condition (3) is necessary to prove Lemmas 2 and 3 in Section 2. Here and after, $A \lesssim B$ denotes $A \leq cB$ for a fixed constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $A \gtrsim B$.

It is well-known that the wavelet estimation depends on an orthonormal wavelet expansion in $L^2(\mathbb{R})$, even in $L^p(\mathbb{R})$. Let $\{V_j : j \in \mathbb{N}\}$ be a classical Multiresolution Analysis of $L^2(\mathbb{R})$ with scaling function φ and ψ being the corresponding wavelet. Subsequently, for $f \in L^2(\mathbb{R})$,

$$f = \sum_{k\in\mathbb{Z}} \alpha_{j_0 k}\varphi_{j_0 k} + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}, \tag{4}$$

where $\alpha_{jk} := \langle f, \varphi_{jk}\rangle$, $\beta_{jk} := \langle f, \psi_{jk}\rangle$ and $\vartheta_{jk}(\cdot) := 2^{\frac{j}{2}}\vartheta(2^j \cdot -k)$ $(\vartheta = \varphi$ or $\psi)$. A scaling function φ is called m-regular $(m \in \mathbb{N})$, if $\varphi \in \mathbf{C}^m(\mathbb{R})$ and $|\varphi^{(k)}(x)| \lesssim (1 + |x|^2)^{-l}$ for each $l \in \mathbb{N}$ $(k = 0, 1, \cdots, m)$. Clearly, the m-regularity of φ implies that of the corresponding ψ, and $|\varphi^{ft}(t)| \lesssim (1 + |t|^2)^{-\frac{m}{2}}$ due to the integration by parts. An important example is Daubechies' function D_{2N} with N large enough.

As usual, let P_j be the orthogonal projection from $L^2(\mathbb{R})$ onto V_j,

$$P_j f := \sum_{k\in\mathbb{Z}} \alpha_{jk}\varphi_{jk}. \tag{5}$$

If φ is m-regular, then $P_j f$ is well-defined for $f \in L^p(\mathbb{R})$. Moreover, the identity (4) holds in $L^p(\mathbb{R})$ for $1 \leq p \leq \infty$.

The following lemma is needed for later discussions.

Lemma 1 ([13]). *Let ϑ be an orthogonal scaling function or a wavelet satisfying m-regularity. Subsequently, there exist $C_2 \geq C_1 > 0$, such that, for $\lambda = \{\lambda_k\} \in l^p(\mathbb{Z})$ and $1 \leq p \leq \infty$,*

$$C_1 2^{j(\frac{1}{2}-\frac{1}{p})} \|\lambda\|_{l_p} \leq \| \sum_{k\in\mathbb{Z}} \lambda_k \vartheta_{jk} \|_p \leq C_2 2^{j(\frac{1}{2}-\frac{1}{p})} \|\lambda\|_{l_p}.$$

One of the advantages of wavelet bases is that they can characterize Besov spaces, which contain the L^2-Sobolev spaces and Hölder spaces as special examples.

Proposition 1 ([13]). *Let scaling function φ be m-regular with $m > s > 0$ and ψ be the corresponding wavelet. Afterwards, for $r, q \in [1, \infty]$ and $f \in L^r(\mathbb{R})$, the following conditions are equivalent:*

(i). $f \in B^s_{r,q}(\mathbb{R})$;
(ii). $\{2^{js}\|P_j f - f\|_r\}_{j\geq 0} \in l_q$; *and,*
(iii). $\|\alpha_{0\cdot}\|_{l_r} + \|(2^{j(s+\frac{1}{2}-\frac{1}{r})}\|\beta_{j\cdot}\|_{l_r})\|_{l_q} < \infty$.

The Besov norm can be defined by

$$\|f\|_{B^s_{r,q}} := \|\alpha_{j_0\cdot}\|_{l_r} + \|(2^{j(s+\frac{1}{2}-\frac{1}{r})}\|\beta_{j\cdot}\|_{l_r})\|_{l_q}.$$

When $s > 0$ and $1 \leq r, p, q \leq \infty$, it is well-known that

(1) $B^s_{r,q} \hookrightarrow B^s_{r,\infty} \hookrightarrow B^{s-\frac{1}{r}}_{\infty,\infty}$ for $s > \frac{1}{r}$;
(2) $B^s_{r,q} \hookrightarrow B^{s'}_{p,q}$ for $r \leq p$ and $s - \frac{1}{r} = s' - \frac{1}{p}$,

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

In this paper, we use the notation $B^s_{r,q}(L, M)$ with some constants $L, M > 0$ to stand for a Besov ball, i.e.,

$$B^s_{r,q}(L, M) := \{f \in L^r(\mathbb{R}),\ f \text{ is a density}, \|f\|_{B^s_{r,q}} \leq L \text{ and supp } f \subseteq [-M, M]\}.$$

Next, we will estimate f with L^p risk by constructing wavelet estimators from the observed data $Z_1, Z_2, \cdots, Z_n$. To introduce wavelet estimators, we take φ having compact support and m-regularity with $m \geq \beta(\alpha) + 2$ in this paper. Moreover, denote

$$\hat{\alpha}_{jk} := \frac{1}{n} \sum_{l=1}^{n} (K\varphi_{jk})(Z_l),$$

where

$$(K\varphi_{jk})(x) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \overline{(\varphi_{jk})^{ft}(t)} G_\alpha^{-1}(t) dt \tag{6}$$

and $K\psi_{jk}$ is defined by the way. Clearly, $E\hat{\alpha}_{jk} = \alpha_{jk}$ due to the Plancherel formula. Subsequently, the linear wavelet estimator is given by

$$\hat{f}_n^{lin} := \sum_{k\in\Lambda_{j_0}} \hat{\alpha}_{j_0 k} \varphi_{j_0 k}, \tag{7}$$

where $\Lambda_j := \{k \in \mathbb{Z},\ \text{supp}\, f \cap \text{supp}\, \varphi_{jk} \neq \emptyset\} \cup \{k \in \mathbb{Z},\ \text{supp}\, f \cap \text{supp}\, \psi_{jk} \neq \emptyset\}$. In particular, the cardinality of Λ_j satisfies that $|\Lambda_j| \sim 2^j$, when f and φ have compact supports.

Now, we are in a position to state the first result of this paper.

Theorem 1. *For $r, p \in [1, +\infty)$, $q \in [1, +\infty]$ and $\frac{1}{r} < s < m$, the estimator $\hat{f}_n^{lin}$ in (7) with $2^{j_0} \sim n^{\frac{1}{2s'+2\beta(\alpha)+1}}$ satisfies*

$$\sup_{f \in B^s_{r,q}(L,\, M)} E\|\hat{f}_n^{lin} - f\|_p^p \lesssim n^{-\frac{s'p}{2s'+2\beta(\alpha)+1}},$$

where $s' = s - (\frac{1}{r} - \frac{1}{p})_+$ and $a_+ = \max\{a, 0\}$.

Remark 1. *When $\alpha = 1$, $\beta(\alpha) = \beta$, the conclusion of Theorem 1 reduces to Theorem 3 of Li & Liu [4].*

Note that the estimator $\hat{f}_n^{lin}$ is non-adaptive, because the choice of j_0 depends on the unknown parameter s. To obtain an adaptive estimate, define

$$\hat{\beta}_{jk} := \frac{1}{n}\sum_{l=1}^{n}(K\psi_{jk})(Z_l) \quad \text{and} \quad \tilde{\beta}_{jk} := \hat{\beta}_{jk} I_{\{|\hat{\beta}_{jk}| > \tau_{j,n}\}}. \tag{8}$$

Here, $\tau_{j,n} = c\gamma 2^{j\beta(\alpha)}\sqrt{\frac{j}{n}}$ and the constants c, γ will be determined later on. Subsequently, the non-linear wavelet estimator is defined by

$$\hat{f}_n^{non} := \sum_{k \in \Lambda_{j_0}} \hat{\alpha}_{j_0 k}\varphi_{j_0 k} + \sum_{j=j_0}^{j_1}\sum_{k \in \Lambda_j} \tilde{\beta}_{jk}\psi_{jk}, \tag{9}$$

where j_0, j_1 are positive integers satisfying $2^{j_0} \sim n^{\frac{1}{2m+2\beta(\alpha)+1}}$ and $2^{j_1} \sim \frac{n}{\ln n}$ respectively. Clearly, j_0 and j_1 do not depend on the unknown parameters s, r, q, which means that the estimator $\hat{f}_n^{non}$ in (9) is adaptive.

Theorem 2. *Let $r, p \in [1, +\infty)$, $q \in [1, +\infty]$ and $\frac{1}{r} < s < m$. Then the estimator $\hat{f}_n^{non}$ in (9) satisfies*

$$\sup_{f \in B^s_{r,q}(L,M)} E\|\hat{f}_n^{non} - f\|_p^p \lesssim (\ln n)^p \Big(\frac{\ln n}{n}\Big)^{\theta p},$$

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

Remark 2. *When $\alpha = 0$, $\beta(\alpha) = 0$ and $\theta = \min\{\frac{s}{2s+1}, \frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-\frac{1}{r})+1}\}$, the convergence rate of Theorem 2 coincides with that of Theorem 3 in Donoho et al. [1]. On the other hand, $\beta(\alpha) = \beta$ and $\theta = \min\{\frac{s}{2s+2\beta+1}, \frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-\frac{1}{r})+2\beta+1}\}$ for the case $\alpha = 1$, while the conclusion of Theorem 4 in Li & Liu [4] can follow directly from this theorem.*

Remark 3. *When comparing the result of Theorem 2 with Theorem 1, we find easily that for the case $r \leq p$, the convergence rate of non-linear estimator is better than that of the linear one with $n^{-\frac{s'p}{2s'+2\beta(\alpha)+1}}$ and $s' = s - \frac{1}{r} + \frac{1}{p}$.*

Remark 4. *The convergence rates of Theorem 2 with the cases $\alpha = 0$ and $\alpha = 1$ are nearly-optimal (up to a logarithmic factor) by Donoho et al. [1] and Li & Liu [4] respectively. However, it is not clear whether the estimation in Theorem 2 is optimal (nearly-optimal) or not for $\alpha \in (0, 1)$. Therefore, one of our future work is to*

determine a low bound estimate for model (2) with $\alpha \in (0,1)$. This problem may be much more complicated than the cases of $\alpha = 0$ and $\alpha = 1$.

2. Preliminaries

This section is devoted to introduce some useful lemmas. The following inequality is necessary in the proof of Lemma 2.

Rosenthal's inequality ([13]). *Let $p > 0$ and $X_1, X_2, \cdots, X_n$ be the independent random variables such that $EX_l = 0$ and $E|X_l|^p < \infty$ $(l = 1, 2, \cdots, n)$. Subsequently, there exists $C(p) > 0$, such that*

$$E\left|\sum_{l=1}^{n} X_l\right|^p \le C(p)\left\{\sum_{l=1}^{n} E|X_l|^p I_{\{p>2\}} + \left(\sum_{l=1}^{n} EX_l^2\right)^{\frac{p}{2}}\right\}.$$

Lemma 2. *Let $1 \le 2^j \le n$ and $\|f\|_\infty \lesssim 1$. Then for $p \in [1, +\infty)$,*

$$E|\hat{\alpha}_{jk} - \alpha_{jk}|^p \lesssim n^{-\frac{p}{2}} 2^{j\beta(\alpha)p} \quad \text{and} \quad E|\hat{\beta}_{jk} - \beta_{jk}|^p \lesssim n^{-\frac{p}{2}} 2^{j\beta(\alpha)p}.$$

Proof. Obviously, one only needs to prove the first inequality and the second one is similar. Define $\xi_l := (K\varphi_{jk})(Z_l) - E(K\varphi_{jk})(Z_l)$ $(l = 1, \cdots, n)$. Subsequently, $\{\xi_l\}_{l=1}^n$ are i.i.d. samples and $E\xi_l = 0$ $(l = 1, \cdots, n)$. By the definitions of α_{jk} and $\hat{\alpha}_{jk}$, $E\hat{\alpha}_{jk} = \alpha_{jk}$ and

$$E|\hat{\alpha}_{jk} - \alpha_{jk}|^p = \frac{1}{n^p} E\left|\sum_{l=1}^{n}[(K\varphi_{jk})(Z_l) - E(K\varphi_{jk})(Z_l)]\right|^p = \frac{1}{n^p} E\left|\sum_{l=1}^{n} \xi_l\right|^p. \tag{10}$$

According to (6), one obtains that

$$|(K\varphi_{jk})(Z_l)| \lesssim \int_{\mathbb{R}} |(\varphi_{jk})^{ft}(t) G_\alpha^{-1}(t)| dt = 2^{-\frac{j}{2}} \int_{\mathbb{R}} |\varphi^{ft}(2^{-j}t) G_\alpha^{-1}(t)| dt.$$

This with (3), m $(m \ge \beta(\alpha) + 2)$ regularity of φ and $2^j \ge 1$ shows

$$|(K\varphi_{jk})(Z_l)| \lesssim 2^{\frac{j}{2}} \int_{\mathbb{R}} (1 + |t|^2)^{-\frac{m}{2}} (1 + |2^j t|^2)^{\frac{\beta(\alpha)}{2}} dt \lesssim 2^{j(\frac{1}{2} + \beta(\alpha))}. \tag{11}$$

Hence, for $l = 1, 2, \cdots, n$,

$$\|\xi_l\|_\infty \lesssim 2^{j(\frac{1}{2} + \beta(\alpha))}. \tag{12}$$

On the other hand, $\|h\|_\infty = \|(1 - \alpha)f + \alpha(f * g)\|_\infty \lesssim 1$ follows from $\|f\|_\infty \lesssim 1$. Afterwards,

$$E|\xi_l|^2 \le E(K\varphi_{jk})^2(Z_l) = \int_{\mathbb{R}} |K\varphi_{jk}(z)|^2 h(z) dz \lesssim \int_{\mathbb{R}} |K\varphi_{jk}(z)|^2 dz.$$

Furthermore, $E|\xi_l|^2 \lesssim \int_{\mathbb{R}} |(\varphi_{jk})^{ft}(z) G_\alpha^{-1}(z)|^2 dz$ due to the Plancherel formula. The same arguments as (11) imply that

$$E|\xi_l|^2 \lesssim \int_{\mathbb{R}} (1 + |t|^2)^{-m} (1 + |2^j t|^2)^{\beta(\alpha)} dt \lesssim 2^{2j\beta(\alpha)}. \tag{13}$$

By Rosenthal's inequality, $E|\sum_{l=1}^n \xi_l|^p \lesssim \sum_{l=1}^n E|\xi_l|^p I_{\{p>2\}} + (\sum_{l=1}^n E\xi_l^2)^{\frac{p}{2}}$. This with (12) and (13) and $2^j \le n$ shows

$$E\left|\sum_{l=1}^{n} \xi_l\right|^p \lesssim n^{\frac{p}{2}} 2^{j\beta(\alpha)p} [(n^{-1}2^j)^{\frac{p}{2}-1} I_{\{p>2\}} + 1] \lesssim n^{\frac{p}{2}} 2^{j\beta(\alpha)p}. \tag{14}$$

Finally, the desired conclusion is concluded from (10) and (14). □

We state another classical inequality, before giving the proof of Lemma 3.

Bernstein's inequality ([13]). *Let $X_1, X_2, \cdots, X_n$ be independent random variables with $EX_l = 0, EX_l^2 \le M$ and $|X_l| \le \|X\|_\infty$ $(l = 1, 2, \cdots, n)$. Then for each $\epsilon > 0$,*

$$P\left\{\left|\frac{1}{n}\sum_{l=1}^{n} X_l\right| > \epsilon\right\} \le 2\exp\left\{-\frac{n\epsilon^2}{2\left(M + \|X\|_\infty \epsilon/3\right)}\right\}.$$

Lemma 3. *If $j2^j \le n$, $2^j \ge 1$ and $\|f\|_\infty \lesssim 1$, then there exists some constat $c > 0$ such that for any $\gamma > 0$,*

$$P\left\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\tau_{j,n}}{2}\right\} \lesssim 2^{-\gamma j},$$

where $\tau_{j,n} = c\gamma 2^{j\beta(\alpha)}\sqrt{\frac{j}{n}}$.

Proof. Denote $\eta_l := (K\psi_{jk})(Z_l) - E(K\psi_{jk})(Z_l)$, $l = 1, \cdots, n$. Then $\{\eta_l\}_{l=1}^n$ are i.i.d., $E\eta_l = 0$ and $|\hat{\beta}_{jk} - \beta_{jk}| = |\frac{1}{n}\sum_{l=1}^n \eta_l|$ by (8). Moreover, the same arguments as (12) and (13) show

$$\|\eta_l\|_\infty \le H_1 2^{j(\frac{1}{2}+\beta(\alpha))} \quad \text{and} \quad E\eta_l^2 \le H_2 2^{2j\beta(\alpha)}$$

with some positive constants H_1, H_2.

According to Bernstein's inequality,

$$P\left\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\tau_{j,n}}{2}\right\} \le 2\exp\left\{-\frac{n\tau_{j,n}^2}{8[H_2 2^{2j\beta(\alpha)} + H_1 2^{j(\frac{1}{2}+\beta(\alpha))}\tau_{j,n}/6]}\right\}. \tag{15}$$

Note that for $j2^j \le n$,

$$8[H_2 2^{2j\beta(\alpha)} + H_1 2^{j(\frac{1}{2}+\beta(\alpha))}\tau_{j,n}/6] \le H2^{2j\beta(\alpha)}(1 + c\gamma\sqrt{\frac{j2^j}{n}}) \le (1 + c\gamma)H2^{2j\beta(\alpha)}$$

with some positive constant H. Hence,

$$\frac{n\tau_{j,n}^2}{8[H_2 2^{2j\beta(\alpha)} + H_1 2^{j(\frac{1}{2}+\beta(\alpha))}\tau_{j,n}/6]} \ge \frac{nc^2\gamma^2 jn^{-1}2^{2j\beta(\alpha)}}{(1+c\gamma)H2^{2j\beta(\alpha)}} = \frac{c^2\gamma^2 j}{(1+c\gamma)H}.$$

Furthermore, (15) reduces to

$$P\left\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\tau_{j,n}}{2}\right\} \le 2e^{-\frac{c^2\gamma^2 j}{(1+c\gamma)H}} \lesssim 2^{-\gamma j}$$

by choosing $c \ge \max\{2H\ln 2, \sqrt{2H\gamma^{-1}\ln 2}\}$. This completes the proof. □

3. Proofs of Main Results

We shall show the proofs of Theorems 1 and 2 in this section.

Proof of Theorem 1. It is sufficient to prove the case for $r \le p$. In fact, when $r > p$ and f has a compact support, $\hat{f}_n^{lin}$ does because of φ having the same property. Subsequently, it follows from Hölder inequality and Jensen's inequality that

$$\sup_{f\in B^s_{r,q}(L,M)} E\|\hat{f}_n^{lin}-f\|_p^p \lesssim \sup_{f\in B^s_{r,q}(L,M)} (E\|\hat{f}_n^{lin}-f\|_r^r)^{\frac{p}{r}} \lesssim n^{-\frac{s'p}{2s'+2\beta(\alpha)+1}}.$$

According to (5) and (7), one easily finds that

$$E\|\hat{f}_n^{lin}-f\|_p^p \lesssim \|P_{j_0}f-f\|_p^p + E\|\hat{f}_n^{lin}-E\hat{f}_n^{lin}\|_p^p. \tag{16}$$

Clearly, by Proposition 1, $\|P_{j_0}f-f\|_p^p \lesssim 2^{-j_0s'p}$ thanks to the well-known embedding theorem $B^s_{r,q}\hookrightarrow B^{s'}_{p,q}$ for $r\le p$.

On the other hand, $s>\frac{1}{r}$ implies $\|f\|_\infty \lesssim 1$ and Lemma 2 tells that $E|\hat{\alpha}_{jk}-\alpha_{jk}|^p \lesssim 2^{j\beta(\alpha)p}n^{-\frac{p}{2}}$. This with Lemma 1 and $|\Lambda_{j_0}|\sim 2^{j_0}$ shows that

$$E\|\hat{f}_n^{lin}-E\hat{f}_n^{lin}\|_p^p \lesssim 2^{j_0(\frac{p}{2}-1)}\sum_{k\in\Lambda_{j_0}} E|\hat{\alpha}_{j_0k}-\alpha_{j_0k}|^p \lesssim 2^{j_0(\beta(\alpha)+\frac{1}{2})p}n^{-\frac{p}{2}}. \tag{17}$$

Finally, (16) reduces to

$$E\|\hat{f}_n^{lin}-f\|_p^p \lesssim 2^{-j_0s'p}+2^{j_0(\beta(\alpha)+\frac{1}{2})p}n^{-\frac{p}{2}} \lesssim n^{-\frac{s'p}{2s'+2\beta(\alpha)+1}}$$

because of $2^{j_0}\sim n^{\frac{1}{2s'+2\beta(\alpha)+1}}$. The proof is done. □

Now, we give a proof of Theorem 2, which is the most important result.

Proof of Theorem 2. The same arguments as the proof of Theorem 1, one only needs to prove the case for $r\le p$.

According to (4), (5) and (9), $E\|\hat{f}_n^{non}-f\|_p^p \lesssim A_n+B_n+C_n$, where

$$A_n=E\|\sum_{k\in\Lambda_{j_0}}(\hat{\alpha}_{j_0k}-\alpha_{j_0k})\varphi_{j_0k}\|_p^p,\ \ B_n=E\|\sum_{j=j_0}^{j_1}\sum_{k\in\Lambda_j}(\tilde{\beta}_{jk}-\beta_{jk})\psi_{jk}\|_p^p$$

and $C_n=\|f-P_{j_1}f\|_p^p$.

It is known that $B^s_{r,q}\hookrightarrow B^{s'}_{p,q}$ for $r\le p$ with $s'=s-\frac{1}{r}+\frac{1}{p}$. Hence, $C_n=\|f-P_{j_1}f\|_p^p \lesssim 2^{-j_1s'p}$ due to Proposition 1. Moreover, it follows from the choice of $2^{j_1}\sim\frac{n}{\ln n}$, $s>\frac{1}{r}$ and $\theta:=\min\{\frac{s}{2s+2\beta(\alpha)+1},\frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-\frac{1}{r})+2\beta(\alpha)+1}\}$ that

$$C_n\lesssim\Big(\frac{\ln n}{n}\Big)^{s'p}\lesssim\Big(\frac{\ln n}{n}\Big)^{\theta p}.$$

Similar to (17), one obtains

$$A_n=E\|\sum_{k\in\Lambda_{j_0}}(\hat{\alpha}_{j_0k}-\alpha_{j_0k})\varphi_{j_0k}\|_p^p \lesssim 2^{j_0(\beta(\alpha)+\frac{1}{2})p}n^{-\frac{p}{2}}\lesssim\Big(\frac{\ln n}{n}\Big)^{\theta p}$$

by $2^{j_0}\sim n^{\frac{1}{2m+2\beta(\alpha)+1}}$, $s<m$ and the definition of θ.

Next, the main work of this proof is to estimate B_n. By Lemma 1,

$$B_n=E\|\sum_{j=j_0}^{j_1}\sum_{k\in\Lambda_j}(\tilde{\beta}_{jk}-\beta_{jk})\psi_{jk}\|_p^p \lesssim j_1^{p-1}\sum_{j=j_0}^{j_1}2^{j(\frac{p}{2}-1)}\sum_{k\in\Lambda_j}E|\tilde{\beta}_{jk}-\beta_{jk}|^p.$$

Define $\hat{B}_j := \{k : |\hat{\beta}_{jk}| > \tau_{j,n}\}$, $B_j := \{k : |\beta_{jk}| > \frac{\tau_{j,n}}{2}\}$, $C_j := \{k : |\beta_{jk}| > 2\tau_{j,n}\}$ and $\xi_{jk} := \hat{\beta}_{jk} - \beta_{jk}$. Subsequently, $B_n \lesssim (\ln n)^{p-1} \sum_{l=1}^{4} Ee_l$, where

$$\begin{aligned} e_1 &:= \sum_{j=j_0}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\xi_{jk}|^p I_{\{k\in\hat{B}_j\cap B_j^c\}}, \\ e_2 &:= \sum_{j=j_0}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\xi_{jk}|^p I_{\{k\in\hat{B}_j\cap B_j\}}, \\ e_3 &:= \sum_{j=j_0}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\beta_{jk}|^p I_{\{k\in\hat{B}_j^c\cap C_j\}}, \\ e_4 &:= \sum_{j=j_0}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\beta_{jk}|^p I_{\{k\in\hat{B}_j^c\cap C_j^c\}} \end{aligned}$$

with A^c denoting the complement of A in $\mathbb{Z}$. Hence, it suffices to prove $Ee_l \lesssim \ln n(\frac{\ln n}{n})^{\theta p}$ $(l = 1, 2, 3, 4)$.

To estimate Ee_1, note that $|\hat{\beta}_{jk} - \beta_{jk}| \geq \frac{\tau_{j,n}}{2}$ follows from $k \in \hat{B}_j \cap B_j^c$. Moreover, $E|\xi_{jk}|^p I_{\{k\in\hat{B}_j\cap B_j^c\}} \leq (E|\xi_{jk}|^{2p})^{\frac{1}{2}} P^{\frac{1}{2}}(|\xi_{jk}| > \frac{\tau_{j,n}}{2})$ by the Hölder inequality. Afterwards, for large γ,

$$Ee_1 \lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} n^{-\frac{p}{2}} 2^{j\beta(\alpha)p} 2^{-\frac{\gamma j}{2}} \lesssim 2^{j_0(\frac{p}{2}+\beta(\alpha)p-\frac{\gamma}{2})} n^{-\frac{p}{2}} \lesssim \Big(\frac{\ln n}{n}\Big)^{\theta p}$$

thanks to $|\Lambda_j| \sim 2^j$, Lemmas 2 and 3.

For e_3, $|\hat{\beta}_{jk} - \beta_{jk}| \geq \frac{\tau_{j,n}}{2}$ due to $k \in \hat{B}_j^c \cap C_j$. Combining this with Lemma 3, one knows $EI_{\{k\in\hat{B}_j^c\cap C_j\}} \leq P(|\xi_{jk}| \geq \frac{\tau_{j,n}}{2}) \lesssim 2^{-\gamma j}$. Therefore, by $|\beta_{jk}| \lesssim 2^{\frac{j}{2}}$, $|\Lambda_j| \sim 2^j$ and large γ,

$$Ee_3 \lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} 2^{\frac{pj}{2}} 2^{-\gamma j} \lesssim 2^{-(\gamma-p)j_0} \lesssim \Big(\frac{\ln n}{n}\Big)^{\theta p}.$$

In order to estimate e_2 and e_4, one defines $j_0^*, j_1^* \in \mathbb{Z}$ satisfying

$$2^{j_0^*} \sim \Big(\frac{n}{\ln n}\Big)^{\frac{1-2\theta}{2\beta(\alpha)+1}} \text{ and } 2^{j_1^*} \sim \Big(\frac{n}{\ln n}\Big)^{\frac{\theta}{s-\frac{1}{r}+\frac{1}{p}}}.$$

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

To estimate Ee_2, one divides e_2 into

$$e_2 = \Big(\sum_{j=j_0}^{j_0^*} + \sum_{j=j_0^*+1}^{j_1}\Big) 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\xi_{jk}|^p I_{\{k\in\hat{B}_j\cap B_j\}} := e_{21} + e_{22}.$$

Similar to (17), according to $2^{j_0^*} \sim (\frac{n}{\ln n})^{\frac{1-2\theta}{2\beta(\alpha)+1}}$,

$$Ee_{21} \lesssim 2^{j_0^*(\beta(\alpha)+\frac{1}{2})p} n^{-\frac{p}{2}} \lesssim \Big(\frac{\ln n}{n}\Big)^{\theta p}.$$

Define

$$\omega := \Big(s + \beta(\alpha) + \frac{1}{2}\Big)r - \Big(\beta(\alpha) + \frac{1}{2}\Big)p.$$

When $\omega \geq 0$, $\theta := \min\{\frac{s}{2s+2\beta(\alpha)+1}, \frac{s-1/r+1/p}{2(s-1/r)+2\beta(\alpha)+1}\} = \frac{s}{2s+2\beta(\alpha)+1}$. It is easy to see that $\frac{|\beta_{jk}|}{\tau_{j,n}/2} > 1$ for $k \in B_j$. Hence,

$$Ee_{22} \lesssim \sum_{j=j_0^*+1}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} E|\xi_{jk}|^p \Big(\frac{|\beta_{jk}|}{\tau_{j,n}/2}\Big)^r. \tag{18}$$

In addition, $E|\xi_{jk}|^p \lesssim 2^{j\beta(\alpha)p} n^{-\frac{p}{2}}$ by Lemma 2 and $\sum_k |\beta_{jk}|^r \lesssim 2^{-j(sr+\frac{r}{2}-1)}$ due to Proposition 1 and $f \in B^s_{r,q}$. These with (18), $\tau_{j,n} \sim 2^{j\beta(\alpha)}\sqrt{\frac{j}{n}}$ and $2^{j_0^*} \sim (\frac{n}{\ln n})^{\frac{1-2\theta}{2\beta(\alpha)+1}}$ lead to

$$Ee_{22} \lesssim \sum_{j=j_0^*+1}^{j_1} \Big(\frac{\ln n}{n}\Big)^{\frac{p-r}{2}} 2^{-j\omega} \lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{p-r}{2}} 2^{-j_0^*\omega} \lesssim \Big(\frac{\ln n}{n}\Big)^{\theta p}.$$

For the case $\omega < 0$, $\theta := \min\{\frac{s}{2s+2\beta(\alpha)+1}, \frac{s-1/r+1/p}{2(s-1/r)+2\beta(\alpha)+1}\} = \frac{s-1/r+1/p}{2(s-1/r)+2\beta(\alpha)+1}$. Denote $r_1 := (1-2\theta)p$. Then $r \leq r_1 \leq p$. Hence, the same arguments as (18) show that

$$Ee_{22} \lesssim \sum_{j=j_0^*+1}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} E|\xi_{jk}|^p \Big(\frac{|\beta_{jk}|}{\tau_{j,n}/2}\Big)^{r_1}. \tag{19}$$

On the other hand, $\sum_k |\beta_{jk}|^{r_1} \lesssim (\sum_k |\beta_{jk}|^r)^{\frac{r_1}{r}} \lesssim 2^{-jr_1(s+\frac{1}{2}-\frac{1}{r})}$ follows from $f \in B^s_{r,q}$ and Proposition 1. Furthermore, $E|\xi_{jk}|^p \lesssim 2^{j\beta(\alpha)p} n^{-\frac{p}{2}}$ thanks to Lemma 2. These with (19) and $\tau_{j,n} \sim 2^{j\beta(\alpha)}\sqrt{\frac{j}{n}}$ imply that

$$Ee_{22} \lesssim \sum_{j=j_0^*+1}^{j_1} 2^{j[\frac{p}{2}-1+\beta(\alpha)p-\beta(\alpha)r_1-r_1(s+\frac{1}{2}-\frac{1}{r})]} \Big(\frac{\ln n}{n}\Big)^{\frac{p-r_1}{2}} \lesssim \ln n \Big(\frac{\ln n}{n}\Big)^{\theta p},$$

because $\frac{p-r_1}{2} = \theta p$ and $\frac{p}{2} - 1 + \beta(\alpha)p - \beta(\alpha)r_1 - r_1(s+\frac{1}{2}-\frac{1}{r}) = 0$.

Finally, it remains to estimate Ee_4. Obviously, $\theta = \frac{s}{2s+2\beta(\alpha)+1}$ for $\omega \geq 0$. Thus,

$$e_4 = (\sum_{j=j_0}^{j_0^*} + \sum_{j=j_0^*+1}^{j_1}) 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\beta_{jk}|^p I_{\{k\in\hat{B}_j^c\cap C_j^c\}} := e_{41} + e_{42}. \tag{20}$$

Because $k \in C_j^c$ and $\tau_{j,n} \sim 2^{j\beta(\alpha)}\sqrt{\frac{j}{n}}$, $|\beta_{jk}| < 2\tau_{j,n} \sim 2^{j\beta(\alpha)}\sqrt{\frac{j}{n}}$. This with $|\Lambda_j| \sim 2^j$ implies that

$$Ee_{41} \lesssim \sum_{j=j_0}^{j_0^*} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} \tau_{j,n}^p \lesssim 2^{j_0^* p(\beta(\alpha)+\frac{1}{2})} \Big(\frac{\ln n}{n}\Big)^{\frac{p}{2}}. \tag{21}$$

Note that $\frac{2\tau_{j,n}}{|\beta_{jk}|} > 1$ by $k \in C_j^c$, and $\sum_k |\beta_{jk}|^r \lesssim 2^{-j(sr+\frac{r}{2}-1)}$ due to $f \in B^s_{r,q}$. Subsequently, with $r \leq p$ and $\tau_{j,n} \sim 2^{j\beta(\alpha)}\sqrt{\frac{j}{n}}$,

$$Ee_{42} \lesssim \sum_{j=j_0^*+1}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \Big(\frac{2\tau_{j,n}}{|\beta_{jk}|}\Big)^{p-r} \lesssim 2^{-j_0^*\omega} \Big(\frac{\ln n}{n}\Big)^{\frac{p-r}{2}}. \tag{22}$$

Combining (20)–(22) with $2^{j_0^*} \sim (\frac{n}{\ln n})^{\frac{1-2\theta}{2\beta(\alpha)+1}}$ and $\theta = \frac{s}{2s+2\beta(\alpha)+1}$ ($\omega \geq 0$), one obtains that

$$Ee_4 = Ee_{41} + Ee_{42} \lesssim \left(\frac{\ln n}{n}\right)^{\theta p}$$

holds for $\omega \geq 0$.

At last, one shows $Ee_4 \lesssim (\frac{\ln n}{n})^{\theta p}$ for the case $\omega < 0$. Subsequently,

$$e_4 = (\sum_{j=j_0}^{j_1^*} + \sum_{j=j_1^*+1}^{j_1}) 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\beta_{jk}|^p I_{\{k\in\hat{B}_j^c \cap C_j^c\}} := e'_{41} + e'_{42}. \tag{23}$$

The same arguments as (22), one finds

$$Ee'_{41} \lesssim 2^{-j_1^*\omega} \left(\frac{\ln n}{n}\right)^{\frac{p-r}{2}}. \tag{24}$$

On the other hand, $\sum_k |\beta_{jk}|^p \lesssim (\sum_k |\beta_{jk}|^r)^{\frac{p}{r}} \lesssim 2^{-jp(s+\frac{1}{2}-\frac{1}{r})}$ by $f \in B^s_{r,q}$ $(r \leq p)$ and Proposition 1. Furthermore,

$$Ee'_{42} \leq \sum_{j=j_1^*+1}^{j_1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\Lambda_j} |\beta_{jk}|^p \lesssim 2^{-j_1^* p(s-\frac{1}{r}+\frac{1}{p})}$$

because of $s > \frac{1}{r}$. This with (23) and (24) and $2^{j_1^*} \sim (\frac{n}{\ln n})^{\frac{\theta}{s-\frac{1}{r}+\frac{1}{p}}}$ leads to

$$Ee_4 = Ee'_{41} + Ee'_{42} \lesssim \left(\frac{\ln n}{n}\right)^{\theta p}$$

thanks to $\theta = \frac{s-1/r+1/p}{2(s-1/r)+2\beta(\alpha)+1}$ for $\omega < 0$. The proof is completed. □

4. Conclusions

This current paper shows L^p $(1 \leq p < \infty)$ risk estimations of both linear and non-linear wavelet estimators under a convolution structure density model over Besov balls. The corresponding conclusions are introduced by Theorems 1 and 2 in this paper, which can be seen as an extension of the works of Donoho et al. [1] and Li & Liu [4].

It should be pointed out that the non-linear wavelet estimator is adaptive, and the convergence rate of non-linear estimator are better than that of linear one for the case $r \leq p$. However, it is not clear whether the estimations are optimal (nearly-optimal) or not for $\alpha \in (0, 1)$. Therefore, one of our future work is to determine a low bound estimate for model (2) with $\alpha \in (0, 1)$.

Author Contributions: Conceptualization, K.C. and X.Z.; methodology, K.C. and X.Z.; validation, X.Z.; formal analysis, K.C.; writing–original draft preparation, K.C.; writing–review and editing, K.C. and X.Z.; supervision, X.Z.; funding acquisition, X.Z.; project administration, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by the National Natural Science Foundation of China (No. 11901019) and the Science and Technology Program of Beijing Municipal Commission of Education (No. KM202010005025).

Acknowledgments: The authors would like to thank the referees for their very helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Donoho, D.L.; Johnstone, I.M.; Kerkyacharian, G.; Picard, D. Density estimation by wavelet thresholding. *Ann. Stat.* **1996**, *24*, 508–539. [CrossRef]
2. Fan, J.; Koo, J.-Y. Wavelet deconvolution. *IEEE Trans. Inform. Theory* **2002**, *48*, 734–747.

3. Lounici, K.; Nickl, R. Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Stat.* **2011**, *39*, 201–231. [CrossRef]
4. Li, R.; Liu, Y.M. Wavelet optimal estimations for a density with some additive noises. *Appl. Comput. Harmon. Anal.* **2014**, *36*, 416–433. [CrossRef]
5. Lepski, O.; Willer, T. Lower bounds in the convolution structure density model. *Bernoulli* **2017**, *23*, 884–926. [CrossRef]
6. Lepski, O.; Willer, T. Oracle inequalities and adaptive estimation in the convolution structure density model. *Ann. Stat.* **2019**, *47*, 233–287. [CrossRef]
7. Liu, Y.M.; Zeng, X.C. Asymptotic normality for wavelet deconvolution density estimators. *Appl. Comput. Harmon. Anal.* **2020**, *48*, 321–342. [CrossRef]
8. Pensky, M.; Vidakovic, B. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Stat.* **1999**, *27*, 2033–2053.
9. Cao, K.K.; Liu, Y.M. On the Reynaud-Bouret–Rivoiard–Tuleau-Malot problem. *Int. J. Wavelets Multiresolut. Inf. Process.* **2018**, *16*, 1850038. [CrossRef]
10. Kerkyacharian, G.; Picard, D. Density estimation in Besov spaces. *Stat. Probab. Lett.* **1992**, *13*, 15–24. [CrossRef]
11. Liu, Y.M.; Wang, H.Y. Convergence order of wavelet thresholding estimator for differential operators on Besov spaces. *Appl. Comput. Harmon. Anal.* **2012**, *32*, 342–356. [CrossRef]
12. Wu, C.; Wang, J.R.; Zeng, X.C. A pointwise lower bound for generalized deconvolution density estimation. *Appl. Anal.* **2019**. [CrossRef]
13. Härdle, W.; Kerkyacharian, G.; Picard, D.; Tsybakov, A. *Wavelets, Approximation and Statistical Applications*; Springer: New York, NY, USA, 1998.