The Consistency of Wavelet Density Estimator with Heteroscedastic Measurement Error

Xiaochen Zeng*, Jinru Wang
College of Applied Sciences, Beijing University of Technology, Pingle Yuan 100, Beijing, 100124, P. R. China.

* Corresponding author. Tel.: +86-13811551569; email: zengxiaochen@emails.bjut.edu.cn
Manuscript submitted April 17, 2017; accepted June 15, 2017.
doi: 10.17706/ijapm.2017.7.3.200-208

Abstract: The wavelet deconvolution estimation based on independent and identically distributed data has made a great progress in nonparametric statistics. However, heteroscedastic measurement errors occur in many natural processes. In this current paper, we construct a practical wavelet deconvolution estimator under severely ill-posed noise and show its mean consistency over L^p risk ($1 \le p < \infty$).

Key words: Consistency, density estimation, severely Ill-posed noise, wavelet.

1. Introduction

The density estimation with an additive noise plays an important role in both statistics and econometrics. Optimal convergence rate and consistency are two basic asymptotic criteria of the quality for an estimator. A lot of perfect achievements have been made for the wavelet estimation over $\mathrm{L_p}$ risk by Devroye [1]-[7].

However, in many real-life applications, the assumption of homoscedastic error is too restrictive. For instance, the data cannot be identically distributed which is called the heteroscedastic error. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathrm{Y_1, Y_2, \cdots, Y_n}$ be independent and un-identically distributed data of

$$Y_l = X_l + \varepsilon_l, \quad l = 1, 2, \cdots, n \tag{1.1}$$

where $\{X_l\}$ stands for independent and identically distributed (i.i.d.) real valued random variables with the unknown probability density f_X, $\{\varepsilon_l\}$ denotes independent random noise (error) with the probability density function f_{ε_l} $(l = 1, 2, \cdots, n)$, and X_l, ε_l are independent each other. It is well known that the probability density f_{Y_l} of Y_l equals to the convolution of f_X and f_{ε_l}.

To introduce the severely ill-posed noise, we need the Fourier transform of $f \in L^1(\mathbb{R})$, i.e.,

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

A standard method extends the definition to $L^2(\mathbb{R})$ functions. A random noise ε_l is said to be severely ill-posed if

$$\left| f_{\varepsilon_l}^{\mathrm{ft}}(t) \right| \gtrsim \exp\{-c_0 |\sigma_l t|^{\alpha}\}, \quad t \in \mathbb{R} \tag{1.2}$$

with $c_0, \alpha > 0$ and $l = 1, 2, \cdots, n$. For two variables A and B, $A \lesssim B$ denotes $A \leq cB$ for some constant $c > 0$ which is independent of A and B; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $B \lesssim A$. In particular, the model (1.1) reduces to the homoscedastic situation (classical deconvolution problem), when $\sigma_1 = \sigma_2 = \cdots = \sigma_n$ in (1.2).

Example 1.1. For $l = 1, 2, \cdots, n$, assume that $\sigma_l > 0$ are real numbers, $\varepsilon_l \backsim N(0, \sigma_l^2)$ which the density function can be represented by $f_{\varepsilon_l}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_l} \exp\left\{-\frac{x^2}{2\sigma_l^2}\right\}$. Then

$$f_{\varepsilon_l}^{\mathrm{ft}}(t) = \exp\left\{-\frac{1}{2}\sigma_l^2 t^2\right\}.$$

Thus (1.2) is satisfied with $\alpha = 2$ and $c_0 = \frac{1}{2}$.

In practical problem, normal distribution is the most commonly observed probability distribution, which is important and could be used in the natural and social sciences.

Example 1.2. Let $f_{\varepsilon_l}(x) = \frac{\lambda_l}{\pi(\lambda_l^2 + x^2)}$ $(l = 1, 2, \cdots, n)$ with $\lambda_l > 0$ which means the noise ε_l being Cauchy distribution. Then

$$f_{\varepsilon_l}^{\mathrm{ft}}(t) = \exp\{-\lambda_l |t|\}.$$

Hence, $\sigma_l = \lambda_l, \alpha = 1$ and $c_0 = 1$ in (1.2).

The Cauchy distribution arises widely in many application fields, which is also called Lorentzain distribution or Breit-Wigner distribution by physicists.

Under the heteroscedastic measurement error, Delaigle & Meister [8] studied the optimal convergence over L^2 risk by kernel method. Chesneau & Fadili [9] constructed a wavelet estimator of the density and investigated its MISE (L^2 risk) performance over Besov balls. The L^p risk $(1 \leq p < \infty)$ of wavelet deconvolution estimator was extended by Wang, Zhang & Kou [10]. However, we do not know whether the density function is smooth or not in some practical applications. Therefore, it is natural to consider the mean consistency of the wavelet estimator, which means that $E\left\|\hat{f}_n - f_X\right\|_p$ $(1 \leq p < \infty)$ converges to zero as the sample size n tends to infinity. As usual, for $1 \leq p < \infty$,

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p dx\right)^{\frac{1}{p}}.$$

This paper considers the mean consistency of a practical wavelet estimator with severely ill-posed noise for heteroscedastic model (1.1). More precisely, we define wavelet estimator for $f_X \in L^p(\mathbb{R})$ $(1 \leq p < \infty)$ by using Meyer's wavelet and study its mean L^p consistency.

2. Wavelet Estimator

This section is devoted to giving some useful concepts and lemmas. In order to introduce our estimator, we begin with a classical notation in wavelet analysis taken from Reference [11]. A multiresolution analysis (MRA) is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of the square integrable function space $L^2(\mathbb{R})$ satisfying the following properties:

1) $V_j \subseteq V_{j+1}$, $j \in \mathbb{Z}$;

2) $\overline{\cup_{j\in\mathbb{Z}}V_j} = L^2(\mathbb{R})$ (The space $\cup_{j\in\mathbb{Z}}V_j$ is dense in $L^2(\mathbb{R})$);
3) $f(2^j\cdot) \in V_j$ if and only if $f(\cdot) \in V_0$ for each $j \in \mathbb{Z}$;
4) There exists $\varphi \in L^2(\mathbb{R})$ (scaling function) such that $\{\varphi(\cdot -k), k \in \mathbb{Z}\}$ forms an orthonormal basis of $V_0 = \overline{span\{\varphi(\cdot -k), k \in \mathbb{Z}\}}$.

With the notation $h_{jk}(x) := 2^{j/2}h(2^jx-k)$ in wavelet analysis, we can show that $\{\varphi_{jk}(x), k \in \mathbb{Z}\}$ is an orthonormal basis of V_j. One of the important examples is Meyer's MRA (see [11]). The Fourier transform $\varphi^{\rm ft}$ of Meyer's scaling function φ is infinitely many times differentiable and their supports contained in the interval $[-a, a]$ with $a = 4\pi/3$.

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

$$P_jf(x) = \sum_{k\in\mathbb{Z}} \alpha_{jk}\varphi_{jk}(x),$$

where $\alpha_{jk} := \langle f, \varphi_{jk}\rangle$. If φ is Meyer's scaling function, then P_jf is well-defined for any $f \in L^p(\mathbb{R})$. All these claims can be found in [12].

Lemma 2.1(12). If φ is the Meyer's scaling function, then there exist $c_2 > c_1 > 0$ such that

$$c_1 2^{j\left(\frac{1}{2}-\frac{1}{p}\right)}\|\{\lambda_k\}\|_{l^p} \le \left\|\sum_k \lambda_k\varphi_{jk}(x)\right\|_p \le c_2 2^{j\left(\frac{1}{2}-\frac{1}{p}\right)}\|\{\lambda_k\}\|_{l^p},$$

where $\|\{\lambda_k\}\|_{l^p} := \left(\sum_k |\lambda_k|^p\right)^{\frac{1}{p}}$. Moreover, for some $f \in L^p(\mathbb{R})$ $(1 \le p < \infty)$,

$$\left\|P_jf - f\right\|_p \longrightarrow 0.$$

Since any density function $f_{\varepsilon_l} \in L^1(\mathbb{R})$, $f_{\varepsilon_l}^{\rm ft}$ is continuous. Therefore, if $f_{\varepsilon_l}^{\rm ft}(t) \neq 0$ and φ is Meyer's scaling function, we have

$$\int_{\mathbb{R}} \left|\frac{\varphi^{\rm ft}(t)}{f_{\varepsilon_l}^{\rm ft}(-2^jt)}\right| dt \lesssim \int_{-a}^{a} \left|\frac{1}{f_{\varepsilon_l}^{\rm ft}(-2^jt)}\right| dt < +\infty.$$

This together with $w_n := \sum_{l=1}^n \exp\{-c_0\sigma_l^{2\alpha}\}$, we know that

$$\hat{\alpha}_{jk} := \frac{2^{\frac{j}{2}}}{w_n}\sum_{l=1}^n \frac{\exp\{-c_0\sigma_l^{2\alpha}\}}{2\pi}\int_{\mathbb{R}} e^{it(2^jY_l-k)}\frac{\varphi^{\rm ft}(t)}{f_{\varepsilon_l}^{\rm ft}(-2^jt)}dt \tag{2.1}$$

is well-defined.

The classical linear wavelet estimator is defined by

$$\hat{f}_n(x) := \sum_{k\in\mathbb{Z}} \hat{\alpha}_{jk}\varphi_{jk}(x). \tag{2.2}$$

When $\sigma_1 = \sigma_2 = \cdots = \sigma_n$, $\hat{\alpha}_{jk}$ and $\hat{f}_n$ defined in (2.1)-(2.2) respectively reduce to the homoscedastic case automatically (see [2], [4]-[7]). In order to guarantee the estimator practical, we modify the above estimator (see (2.2)) as follows:

$$\hat{f}_{n,F}(x) := \sum_{|k|\leq K_n} \hat{\alpha}_{jk}\varphi_{jk}(x), \tag{2.3}$$

where the positive integer K_n will be specified later on.

The next lemma shows $\hat{\alpha}_{jk}$ defined by (2.1) is an unbiased estimation of $\alpha_{jk} := \int_{\mathbb{R}} f_X(x)\overline{\varphi_{jk}(x)}dx$.

Lemma 2.2. Let $\hat{\alpha}_{jk}$ be defined in (2.1), then $E\hat{\alpha}_{jk} = \alpha_{jk}$.

Proof. Since the intersection of $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ is dense in $L^1(\mathbb{R})$, there exists a sequence $\{f_n\}_{n\in\mathbb{N}} \subseteq L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ such that

$$\lim_{n\to\infty} \|f_n - f_X\|_1 = 0.$$

Note that Meyer's scaling function $\varphi \in L^\infty(\mathbb{R})$ and $\|\varphi_{jk}\|_\infty < +\infty$ for fixed j and k. Then

$$\left|\int_{\mathbb{R}} f_n(x)\overline{\varphi_{jk}(x)}\, dx - \int_{\mathbb{R}} f_X(x)\overline{\varphi_{jk}(x)}dx\right| \leq \int_{\mathbb{R}} |f_n(x) - f_X(x)|\,|\varphi_{jk}(x)|dx \leq \|f_n - f_X\|_1 \|\varphi_{jk}\|_\infty \longrightarrow 0, \tag{2.4}$$

as $n \to \infty$. On the other hand, because $\|f_n^{\text{ft}} - f_X^{\text{ft}}\|_\infty \leq \|f_n - f_X\|_1 \longrightarrow 0$ and $\varphi^{\text{ft}} \in L^1(\mathbb{R})$,

$$\left|\int_{\mathbb{R}} f_n^{\text{ft}}(t)\overline{(\varphi_{jk})^{\text{ft}}(t)}\, dt - \int_{\mathbb{R}} f_X^{\text{ft}}(t)\overline{(\varphi_{jk})^{\text{ft}}(t)}\, dt\right| \leq \|f_n^{\text{ft}} - f_X^{\text{ft}}\|_\infty \left\|(\varphi_{jk})^{\text{ft}}\right\|_1 \longrightarrow 0. \tag{2.5}$$

By the Plancherel formula, $\int_{\mathbb{R}} f_n(x)\overline{\varphi_{jk}(x)}\, dx = (2\pi)^{-1}\int_{\mathbb{R}} f_n^{\text{ft}}(t)\overline{(\varphi_{jk})^{\text{ft}}(t)}\, dt$. This with (2.4) and (2.5) leads to

$$\int_{\mathbb{R}} f_X(x)\overline{\varphi_{jk}(x)}dx = \frac{1}{2\pi}\int_{\mathbb{R}} f_X^{\text{ft}}(t)\overline{(\varphi_{jk})^{\text{ft}}(t)}\, dt. \tag{2.6}$$

Obviously, $Ee^{it2^jY_l} = \int_{\mathbb{R}} e^{it2^jy}f_{Y_l}(y)dy = f_{Y_l}^{\text{ft}}(-2^jt) = f_X^{\text{ft}}(-2^jt)f_{\varepsilon_l}^{\text{ft}}(-2^jt)$ thanks to X, ε_l independent each other and $f_{Y_l}^{\text{ft}} = f_X^{\text{ft}} \cdot f_{\varepsilon_l}^{\text{ft}}$. Hence,

$$E\hat{\alpha}_{jk} = \frac{2^{\frac{j}{2}}}{w_n}\sum_{l=1}^{n}\frac{\exp\{-c_0\sigma_l^{2\alpha}\}}{2\pi}\int_{\mathbb{R}} e^{-ikt}\varphi^{\text{ft}}(t)f_X^{\text{ft}}(-2^jt)\, dt = \frac{1}{2\pi}\int_{\mathbb{R}} f_X^{\text{ft}}(t)\overline{(\varphi_{jk})^{\text{ft}}(t)}\, dt.$$

It follows from (2.6) that $E\hat{\alpha}_{jk} = \int_{\mathbb{R}} f_X(x)\overline{\varphi_{jk}(x)}dx = \alpha_{jk}$. This completes the proof. □

In order to show Lemma 2.4, we state a classical inequality as follows.

Lemma 2.3 (Rosenthal's inequality, [12]). Let $X_1, X_2, \cdots, X_n$ be independent random variables such that $EX_l = 0$ and $E|X_l| < +\infty$ $(l = 1, 2, \cdots, n)$. Then

$$E\left|\sum_{l=1}^{n} X_l\right|^p \lesssim \begin{cases} \sum_{l=1}^{n} E|X_l|^p + \left(\sum_{l=1}^{n} EX_l^2\right)^{p/2}, & p > 2; \\ \left(\sum_{l=1}^{n} EX_l^2\right)^{p/2}, & 0 < p \leq 2. \end{cases}$$

Lemma 2.4. Let $1 \le p < \infty$, then

$$E\left|\hat{\alpha}_{jk} - \alpha_{jk}\right|^p \lesssim \left(\frac{2^j}{w_n}\right)^{\frac{p}{2}} \exp\left\{\frac{c_0\left(2^j a\right)^{2\alpha} p}{2}\right\},$$

where $a = \frac{4\pi}{3}$ is the support $\{t, |t| \le a\}$ of the Fourier transform of Meyer's scaling function.

Proof. Define

$$\xi_{l,j,k} := \frac{2^{\frac{j}{2}}}{w_n} \frac{\exp\{-c_0\sigma_l^{2\alpha}\}}{2\pi} \int_{\mathbb{R}} e^{it(2^j Y_l - k)} \frac{\varphi^{\mathrm{ft}}(t)}{f_{\varepsilon_l}^{\mathrm{ft}}(-2^j t)} dt$$

for $l = 1, \cdots, n$. By $\left|f_{\varepsilon_l}^{\mathrm{ft}}(t)\right| \gtrsim \exp\{-c_0|\sigma_l t|^\alpha\}$ and supp $\varphi^{\mathrm{ft}} \subseteq [-a, a]$, one obtains that

$$\left|\xi_{l,j,k}\right| \lesssim \frac{2^{\frac{j}{2}}}{w_n} \exp\{-c_0\sigma_l^{2\alpha}\} \exp\left\{c_0\left|\sigma_l 2^j a\right|^\alpha\right\} \le \frac{2^{\frac{j}{2}}}{w_n} \exp\{-c_0\sigma_l^{2\alpha}\} \exp\left\{\frac{c_0\left[\sigma_l^{2\alpha} + \left(2^j a\right)^{2\alpha}\right]}{2}\right\}$$

$$= \frac{2^{\frac{j}{2}}}{w_n} \exp\left\{-\frac{c_0\sigma_l^{2\alpha}}{2}\right\} \exp\left\{\frac{c_0\left(2^j a\right)^{2\alpha}}{2}\right\}. \tag{2.7}$$

Denote $\eta_{l,j,k} := \xi_{l,j,k} - E\xi_{l,j,k}$. Then $E\eta_{l,j,k} = 0$ obviously, and

$$\hat{\alpha}_{jk} - \alpha_{jk} = \hat{\alpha}_{jk} - E\hat{\alpha}_{jk} = \sum_{l=1}^{n}\left(\xi_{l,j,k} - E\xi_{l,j,k}\right) = \sum_{l=1}^{n} \eta_{l,j,k}$$

due to Lemma 2.2. This with Rosenthal's inequality (Lemma 2.3) tells that

$$E\left|\hat{\alpha}_{jk} - \alpha_{jk}\right|^p = E\left|\sum_{l=1}^n \eta_{l,j,k}\right|^{\mathrm{p}} \lesssim \begin{cases} \sum_{l=1}^n E\left|\eta_{l,j,k}\right|^p + \left(\sum_{l=1}^n E\eta_{l,j,k}^2\right)^{\frac{p}{2}}, & p > 2; \\ \left(\sum_{l=1}^n E\eta_{l,j,k}^2\right)^{\frac{p}{2}}, & 1 \le p \le 2. \end{cases} \tag{2.8}$$

On the other hand, (2.7) implies

$$E\left|\eta_{l,j,k}\right|^p \lesssim \left[\frac{2^{\frac{j}{2}}}{w_n} \exp\left\{-\frac{c_0\sigma_l^{2\alpha}}{2}\right\} \exp\left\{\frac{c_0\left(2^j a\right)^{2\alpha}}{2}\right\}\right]^p. \tag{2.9}$$

Hence, for $1 \le p \le 2$, it follows from (2.8)-(2.9) that

$$E\left|\alpha_{jk} - \alpha_{jk}\right|^p \lesssim \left[\sum_{l=1}^n \frac{2^j}{w_n^2} \exp\{-c_0\sigma_l^{2\alpha}\} \exp\left\{c_0\left(2^j a\right)^{2\alpha}\right\}\right]^{\frac{p}{2}} = \left[\frac{2^j}{w_n} \exp\left\{c_0\left(2^j a\right)^{2\alpha}\right\}\right]^{\frac{p}{2}}. \tag{2.10}$$

When $p > 2$, $\sum_{l=1}^n \exp\left\{-\frac{c_0\sigma_l^{2\alpha}p}{2}\right\} \le \sum_{l=1}^n \exp\{-c_0\sigma_l^{2\alpha}\} = w_n$. According to (2.9), one knows that

$$\sum_{l=1}^{n} E|\eta_{l,j,k}|^p \lesssim \sum_{l=1}^{n} 2^{\frac{jp}{2}} w_n^{-p} \exp\left\{-\frac{c_0\sigma_l^{2\alpha}p}{2}\right\} \exp\left\{\frac{c_0(2^j a)^{2\alpha}p}{2}\right\}$$
$$= 2^{\frac{jp}{2}} w_n^{-p} \sum_{l=1}^{n} \exp\left\{-\frac{c_0\sigma_l^{2\alpha}p}{2}\right\} \exp\left\{\frac{c_0(2^j a)^{2\alpha}p}{2}\right\}$$
$$\leq 2^{\frac{jp}{2}} w_n^{1-p} \exp\left\{\frac{c_0(2^j a)^{2\alpha}p}{2}\right\}.$$

Combining the above inequality with (2.8) and (2.10), one concludes that

$$E|\hat{\alpha}_{jk} - \alpha_{jk}|^p \lesssim 2^{\frac{jp}{2}} w_n^{1-p} \exp\left\{\frac{c_0(2^j a)^{2\alpha}p}{2}\right\} I_{\{p>2\}} + 2^{\frac{jp}{2}} w_n^{-\frac{p}{2}} \exp\left\{\frac{c_0(2^j a)^{2\alpha}p}{2}\right\}$$
$$\lesssim 2^{\frac{jp}{2}} w_n^{-\frac{p}{2}} \exp\left\{\frac{c_0(2^j a)^{2\alpha}p}{2}\right\}.$$

The proof is done.

3. Main Result

In this section, we state the main result as Theorem 3.1 and devote to give its proof.

Theorem 3.1. Let $f_X \in L^p(\mathbb{R})$ $(1 \leq p < \infty)$ satisfy $\|xf_X(x)\|_\infty \lesssim 1$ for $p > 1$ and $\|x^2 f_X(x)\|_\infty \lesssim 1$ for $p = 1$. Then for the estimator $\hat{f}_{n,F}$ defined in (2.3) with $j = \left\lfloor \frac{1}{2\alpha}\log_2(v\ln w_n)\right\rfloor$ ($\lfloor x \rfloor$ standing for the largest integer no more than x) and $K_n \sim \exp\{\ln^\theta w_n\}$ $\left(c_0 a^{2\alpha} v < 1,\ \theta \in (0,1)\right)$, one has

$$\lim_{n\to\infty} E\left\|\hat{f}_{n,F} - f_X\right\|_p = 0.$$

Proof. One need to estimate $E\left\|\hat{f}_{n,F} - f_X\right\|_p \leq I_1(n) + I_2(n) + I_3(n)$ with $I_1(n) := E\left\|\hat{f}_{n,F} - E\hat{f}_{n,F}\right\|_p$, $I_2(n) := \left\|E\hat{f}_{n,F} - P_j f_X\right\|_p$ and $I_3(n) := \left\|P_j f_X - f_X\right\|_p$. Obviously, $I_3(n) \longrightarrow 0$ due to $f_X \in L^p(\mathbb{R})$ and Lemma 2.1.

For $I_2(n)$, one considers the case $p > 1$ firstly, for which $\sum_{|k|>K_n}|k|^{-p} \sim \int_{K_n}^{+\infty} x^{-p}dx = \frac{1}{p-1}K_n^{1-p}$. Since $\alpha_{jk} := \int_{\mathbb{R}} f_X(x)\overline{\varphi_{jk}(x)}\,dx$ and φ is the Meyer's scaling function, $x\varphi(x) \in L^\infty(\mathbb{R})$ and

$$|k\alpha_{jk}| \leq \int_{\mathbb{R}} |k||\varphi_{jk}(x)|f_X(x)dx \leq \int_{\mathbb{R}} |2^j x - k||\varphi_{jk}(x)|f_X(x)dx + \int_{\mathbb{R}} |2^j x||\varphi_{jk}(x)|f_X(x)dx.$$

This with $\varphi_{jk}(x) := 2^{j/2}\varphi(2^j x - k)$ and the assumption $\|xf_X(x)\|_\infty \lesssim 1$ shows that

$$|k\alpha_{jk}| \lesssim 2^{\frac{j}{2}}\|x\varphi(x)\|_\infty + 2^j\|xf_X(x)\|_\infty 2^{-\frac{j}{2}}\|\varphi\|_1 \lesssim 2^{\frac{j}{2}}.$$

On the other hand, Lemma 2.2 tells $E\hat{\alpha}_{jk} = \alpha_{jk}$ and $E\hat{f}_{n,F} = \sum_{|k|\leq K_n}\alpha_{jk}\varphi_{jk}$. Therefore, it follows from $P_j f_X = \sum_{k\in\mathbb{Z}}\alpha_{jk}\varphi_{jk}$ and Lemma 2.1, one concludes that

$$I_2(n) = \left\| \sum_{|k|\le K_n} \alpha_{jk}\varphi_{jk} - \sum_{k\in\mathbb{Z}} \alpha_{jk}\varphi_{jk} \right\|_p = \left\| \sum_{|k|>K_n} \alpha_{jk}\varphi_{jk} \right\|_p \lesssim 2^{\frac{j}{2}-\frac{j}{p}} \left(\sum_{|k|>K_n} |\alpha_{jk}|^p \right)^{1/p}$$

$$\lesssim 2^{\frac{j}{2}-\frac{j}{p}} \left(\sum_{|k|>K_n} |k|^{-p} 2^{\frac{jp}{2}} \right)^{1/p} \lesssim 2^{j-\frac{j}{p}} K_n^{\frac{1}{p}-1} = \left(2^j K_n^{-1}\right)^{1-\frac{1}{p}}.$$

Moreover, $2^j K_n^{-1} \longrightarrow 0$ due to the choices of j and K_n. Then $I_2(n) \longrightarrow 0$ for $p > 1$.

When $p = 1$, it is assumed that $\|x^2 f_X(x)\|_\infty \lesssim 1$. Since $|\alpha_{jk}| \le \int_{\mathbb{R}} |\varphi_{jk}(x)| f_X(x)\, dx$, one finds that

$$|k^2\alpha_{jk}| \le \int_{\mathbb{R}} |2^j x - k|^2 |\varphi_{jk}(x)| f_X(x) dx + \int_{\mathbb{R}} |2^j x|^2 |\varphi_{jk}(x)| f_X(x) dx$$

$$\le 2^{\frac{j}{2}} \|x^2\varphi(x)\|_\infty + 2^{2j} \|x^2 f_X(x)\|_\infty 2^{-\frac{j}{2}} \|\varphi\|_1 \lesssim 2^{\frac{3j}{2}}.$$

Thus, $|\alpha_{jk}| \lesssim 2^{\frac{3j}{2}} |k|^{-2}$. According to Lemma 2.1, one has

$$I_2(n) = \left\| \sum_{|k|>K_n} \alpha_{jk}\varphi_{jk} \right\|_1 \lesssim 2^{\frac{j}{2}-j} \sum_{|k|>K_n} |\alpha_{jk}| \lesssim 2^{-\frac{j}{2}} \sum_{|k|>K_n} |k|^{-2} 2^{\frac{3j}{2}} \lesssim 2^j K_n^{-1}.$$

Hence, $I_2(n) \longrightarrow 0$ follows from $2^j K_n^{-1} \longrightarrow 0$.

It remains to estimate $I_1(n)$. By the definition of $\hat{f}_{n,F}$, $E\hat{f}_{n,F} = \sum_{|k|\le K_n} \alpha_{jk}\varphi_{jk}$ and Lemma 2.1,

$$I_1(n) \lesssim 2^{\frac{j}{2}-\frac{j}{p}} E\left(\sum_{|k|\le K_n} |\hat{\alpha}_{jk} - \alpha_{jk}|^p\right)^{\frac{1}{p}} \le 2^{\frac{j}{2}-\frac{j}{p}} \left(\sum_{|k|\le K_n} E|\hat{\alpha}_{jk} - \alpha_{jk}|^p\right)^{\frac{1}{p}}, \tag{3.1}$$

where Jensen's inequality is used in the second inequality of (3.1). Furthermore, combining (3.1) with Lemma 2.4, one obtains

$$I_1(n) \lesssim 2^{\frac{j}{2}-\frac{j}{p}} w_n^{-\frac{1}{2}} 2^{\frac{j}{2}} \exp\left\{\frac{c_0(2^j a)^{2\alpha}}{2}\right\} K_n^{\frac{1}{p}} \le w_n^{-\frac{1}{2}} 2^j K_n \exp\left\{\frac{c_0(2^j a)^{2\alpha}}{2}\right\}. \tag{3.2}$$

Choosing $j = \left\lfloor \frac{1}{2\alpha} \log_2(v \ln w_n) \right\rfloor$ and $K_n \sim \exp\{\ln^\theta w_n\}$, (3.2) reduces to

$$I_1(n) \lesssim w_n^{-\frac{1}{2}} \cdot (\ln w_n)^{\frac{1}{2\alpha}} \cdot \exp\{\ln^\theta w_n\} \cdot \exp\left\{\frac{c_0 a^{2\alpha} v \ln n}{2}\right\} = w_n^{-\frac{1}{2}} \cdot (\ln w_n)^{\frac{1}{2\alpha}} \cdot \exp\{\ln^\theta w_n\} \cdot w_n^{\frac{c_0 a^{2\alpha} v}{2}} \longrightarrow 0$$

as $n \longrightarrow \infty$, where $c_0 a^{2\alpha} v < 1$ and $\theta \in (0,1)$.

Therefore, the desired conclusion can be concluded by $I_k(n) \longrightarrow 0$ $(k = 1,2,3)$. The proof is completed. □

Remark 3.1. Note that the condition $x^2 f_X(x) \in L^\infty(\mathbb{R})$ is stronger than $x f_X(x) \in L^\infty(\mathbb{R})$, when $f_X \in L^\infty(\mathbb{R})$. Then Theorem 3.1 requires more for $p = 1$ than $p > 1$. This seems natural, because $f_X \in L^1 \cap L^\infty$ implies

$$f_X \in L^p \ (1 < p < \infty).$$

Remark 3.2. When $p = 1$ and $\sigma_1 = \sigma_2 = \cdots = \sigma_n$, the L^1 consistency under the Normal noise and Cauchy noise had been studied in [1] and [3] respectively.

4. Conclusion

We construct a practical wavelet density estimator $\hat{f}_{n,F}$ with severely ill-posed noise firstly. Then the mean L^p consistency of $\hat{f}_{n,F}$ is investigated under some mild condition on f_X, i.e., for $1 \le p < \infty$,

$$\lim_{n\to\infty} E \left\|\hat{f}_{n,F} - f_X\right\|_p = 0.$$

Our result can be seen as an extension of Devroye or Meister's work in some sense (see Reference [1] and [3] respectively).

Acknowledgment

The authors would like to thank Prof. Youming Liu for his many comments. This work was supported by the Beijing Nature Science Foundation (No. 1172001) and the National Natural Science Foundation of China (No. 11271038).

References

[1] Devroye, L. (1989). Consistent deconvolution in density estimation. *The Canadian Journal of Statistics*, *17(2)*, 235-239.

[2] Fan, J., & Koo, J. (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory*, *48(3)*, 734-747.

[3] Meister, A. (2009). Deconvolution problems in nonparametric statistics. *Lecture Notes in Statistics 193*. Berlin: Springer.

[4] Lounici, K., & Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Annals of Statistics, 39(1)*, 201-231.

[5] Li, R., & Liu, Y. (2013). Wavelet optimal estimations for density functions under severely ill-posed noises. *Abstract and Applied Analysis*.

[6] Li, R., & Liu, Y. (2014). Wavelet optimal estimations for a density with some additive noises. *Applied and Computational Harmonic Analysis*, *36(3)*, 416-433.

[7] Geng, Z., & Wang, J. (2015). The mean consistency of wavelet density estimators. *Journal of Inequalities and Applications*, 111.

[8] Delaigle, A., & Meister, A. (2008). Density estimation with heteroscedastic error. *Bernoulli*, *14(2)*, 562-579.

[9] Chesneau, C., & Fadili, J. (2013). Wavelet-based density estimation in a heteroscedastic convolution model. *Communications in Statistics - Theory and Methods*, *17(42)*, 3085-3099.

[10] Wang, J., Zhang, Q., & Kou, J. (2017). Wavelet estimators for the derivatives of the density function from data contaminated with heteroscedastic measurement errors. *Communications in Statistics - Theory and Methods*, *46(15)*, 7337-7354.

[11] Daubechies, I. (1992). *Ten Lectures on Wavelets.* Philadelphia: SIAM.

[12] Härdle, W., Kerkyacharian, G., Picard, D., & Tsybakov, A. (1998). Wavelets, approximation and statistical applications. *Lecture Notes in Statistics 129*. New York: Springer.

Xiaochen Zeng received the B.Sc. degree in mathematical sciences in 2012 from Beijing University of Technology, Beijing, P. R. China. Since 2012, he is an M.D.-Ph.D. based in College of Applied Sciences, Beijing University of Technology, whose advisor is Prof. Youming Liu. He has been pursuing the Ph.D. degree from 2014. His research interests include wavelet analysis and nonparametric density estimation in statistics.

Jinru Wang is associate professor, received a doctorate degree in probability theory and statistics from Beijing University of Technology in 2005. Her research interests are wavelet analysis and its applications in inverse problems.