Journal of Data Science and Modeling, Vol. 1, No. 1, 143-158, December 2022

Research Manuscript

Nonparametric Wavelet Quantile Density Estimation Based on the Biased Data

Esmaeil Shirazi*

Assistant professor of statistics, Department of Statistics, Faculty of Science,

Gonbad Kavous University, Gonbad Kavous 4971799151, Iran.

Received: 16/7/2018 Accepted: 28/10/2018

Abstract: The estimation of a quantile density function in the biased nonparametric regression model is investigated. We propose and develop a new wavelet-based methodology for this problem. In particular, an adaptive hard thresholding wavelet estimator is constructed. Under mild assumptions on the model, we prove that it enjoys powerful mean integrated squared error properties over Besov balls. The performance of the proposed estimator is investigated by a numerical study.

Keywords: Adaptivity, Biased data, Quantile density estimation, Wavelets.
Mathematics Subject Classification (2010): Primary 62G07; secondary 62G20.

*Corresponding author: shirazi.esmaeel@gmail.com

1. Introduction

Estimation of a quantile density function from biased data is a frequent problem in industrial life testing experiments and medical studies. Let X be a continuous random variable with cumulative density function $F(x)$, density function $f(x)$ and hazard function $r(x)$. The quantile function of X is defined as

$$Q(x) = F^{-1}(x) = \inf\{y \in \mathbb{R};\ F(y) > x\} \tag{1.1}$$

It satisfies F (Q(x)) = x. Parzen (1979) and Jones (1992) defined the quantile density function as the derivative of $Q(x)$, that is, $q(x) = Q'(x)$. Note that the sum of two quantile density functions is again a quantile density function.
Differentiating (1.1), we get

$$q(x) = \frac{1}{f(Q(x))}, \quad x \in [0, 1] \tag{1.2}$$

Nair and Sankaran (2009) defined the hazard quantile function as follows:

$$R(x) = r(Q(x)) = \frac{f(Q(x))}{1 - F(Q(x))} = \frac{1}{(1-x)q(x)}, \quad x \in (0, 1).$$

Hence $q(x)$ appears in the expression for hazard quantile function and it would be useful to study nonparametric estimators of this unknown quantile density function.

In this paper, we consider the problem of estimating $q(x)$ without observing directly the sample $X_1, ..., X_n$. We observe an *i.i.d.* sample $Y_1, ..., Y_n$ from a biased distribution with the following density function

$$g(y) = \mu^{-1} w(y) f(y)$$

where w denotes a positive function and μ is the real number defined by $\mu = \int w(y)f(y)dy$. Here, w is known. The density function f and the real number μ are unknown. The objective is to estimate the quantile density function q from $Y_1, ..., Y_n$.

The concept of the quantile density function estimation has been considered in several papers and some smooth quantile function estimators have been proposed for complete samples. For examples, the kernel method were used by Jones (1992) and Soni and *et al.* (2012) for studying nonparametric estimators of quantile density estimation. They proposed some smooth estimators and investigated their asymptotic properties. Chesneau and *et al.* (2015) discussed the nonparametric wavelet estimators of the quantile density function and proposed two kinds of the projection estimators in linear and nonlinear form.

In this study, we develop two types of wavelet estimators for the quantile density function when data comes from a biased distribution function. Our wavelet hard thresholding estimator which is introduced as a nonlinear estimator, has the feature to be adaptive according to $q(x)$. We show that these estimators attain optimal and nearly optimal rates of convergence over a wide range of Besov function classes.

The rest of this paper is organized as follows. In the next section, we present our wavelet estimators. The main theoretical results are described in Section 3. In Section 4, we investigate the performance of proposed estimator. The proofs of the technical results appear in Sections 5.

2. Notations and estimators

We start this section by introducing the concept of Multiresolution Analysis (MRA) on $\mathbb{R}$ as described in Meyer (1992). Let ϕ be a scale function and ψ its associated wavelet basis of $L^p[0,1]$, and define $\phi_{i_0 j}(x) = 2^{i_0/2}\phi(2^{i_0}x - j)$ and $\psi_{ij}(x) = 2^{i/2}\psi(2^i x - j)$. We assume that the father and mother wavelets, $\phi(x)$ and $\psi(x)$, are bounded and compactly supported over $[0,1]$, that $\int \phi = 1$ and that the wavelets are *r-regular*. We call a wavelet ψ *r-regular* if ψ has r vanishing moments and r continuous derivatives. An empirical wavelet expansion for all $q \in L^p[0,1]$ is given by

$$q(x) = \sum_{j \in Z} \alpha_{i_0 j}\phi_{i_0 j}(x) + \sum_{i \geq i_0}\sum_{j \in Z} \beta_{ij}\psi_{ij}(x), \tag{2.3}$$

where,

$$\alpha_{i_0 j} = \int_{[0,1]} g(x)\phi_{i_0 j}(x)dx = \int_{[0,1]} \frac{1}{f(F^{-1}(x))}\phi_{i_0 j}(x)dx = \int_{[0,1]} \phi_{i_0 j}(F(x))dx$$

Similarly,

$$\beta_{ij} = \int_{[0,1]} \psi_{i_0 j}(F(x))dx$$

Since F is unknown, we estimate it by the empirical estimator based on a sample from biased data:

$$\hat{F}_n(x) = \frac{\hat{\mu}}{n}\sum_{i=1}^{n} \frac{I(Y_i < x)}{w(x)}, \qquad x \in [0,1] \tag{2.4}$$

This leads the following integral estimator for $\alpha_{i_0 j}$ and β_{ij}

$$\hat{\alpha}_{i_0 j} = \int_{[0,1]} \phi_{i_0 j}(\hat{F}(x))dx, \qquad \hat{\beta}_{ij} = \int_{[0,1]} \psi_{i_0 j}(\hat{F}(x))dx \tag{2.5}$$

Clearly, $\hat{\alpha}_{i_0 j}$ and $\hat{\beta}_{ij}$ are not unbiased estimators for $\alpha_{i_0 j}$ and β_{ij}. However, using the dominated convergence theorem, one can prove that they are asymptotically unbiased.

Based on $\hat{\alpha}_{i_0 j}$ and $\hat{\beta}_{ij}$, we consider two kinds of wavelet estimators for $q(x)$: a linear wavelet estimator $q_L(x)$ and a hard thresholding wavelet estimator $q_H(x)$, both defined below:

Linear wavelet estimator. We define the linear wavelet estimator $q_L(x)$ by

$$\tilde{q}_L(x) = \sum_{j=1}^{2^{i_0}-1} \hat{\alpha}_{i_0 j} \phi_{i_0 j}(x) \tag{2.6}$$

Hard thresholding wavelet estimator. We define the hard thresholding wavelet estimator $q_H(x)$ by

$$\hat{q}_H(x) = \sum_{j \in Z} \hat{\alpha}_{i_0 j} \phi_{i_0 j}(x) + \sum_{i \geq i_0}^{R} \sum_{j \in Z} \hat{\beta}_{ij} I(|\hat{\beta}_{ij}| > \kappa \lambda_i) \psi_{ij}(x), \tag{2.7}$$

where $\hat{\alpha}_{i_0 j}$ and $\hat{\beta}_{ij}$ are defined by (2.5), the smoothing parameters i_0 and R satisfying $2^{i_0} \simeq n^{1/1+2s}$ and $2^R \simeq n(\log_2 n)^{-2}$, κ is a large enough constant and λ_i represents a threshold. Both R and λ_i will be chosen a posteriori (see Theorem 3.2).

The construction of $\tilde{q}_H(x)$ exploits the sparse nature of the wavelet decomposition of $q(x)$: only the wavelet coefficients with large magnitude contain the main information (in terms of details) of q. Hence $\tilde{q}_H(x)$ aims to only estimate the larger coefficients, and to remove the other (or estimate it by 0). Further aspects and explanation related to this selection techniques can be found in Hardel and *et al.* (1998) and Vidakovic (1999, p.171-173).

As is done in the wavelet literature, we investigate wavelet-based estimators asymptotic convergence rates over a large range of Besov function classes $B^s_{\nu,\pi}$, $s > 0$, $1 \leq \nu, \pi \leq \infty$. The parameter s measures the number of derivatives, where the existence of derivatives is required in an L^p-sense, whereas the parameter π provides a further finer gradation.

The Besov spaces include, in particular, the well-known Sobolev and Hölder spaces of smooth functions H^m and C^s and (B^m_{22} and $B^s_{\infty,\infty}$ respectively), but in addition less traditional spaces, like the spaces of functions of bounded variation, sandwiched between $B^1_{1,1}$ and $B^1_{1,\infty}$. The latter functions are of statistical interest because they allow for better models of spatial of inhomogeneity (e.g. Meyer (1992) and Donoho and Johnstone (1995)).

For a given *r-regular* mother wavelet ψ with $r > s$, define the sequence norm of the wavelet coefficients of a quantile density function $g \in B^s_{\nu,q}$ by

$$|q|_{B^s_{\nu,\pi}} = (\sum_j |\alpha_{i_0 j}|^\nu)^{1/\nu} + \{\sum_{i=i_0}^{\infty} [2^{i\sigma} (\sum_j |\beta_{ij}|^\nu)^{1/\nu}]^\pi\}^{1/\pi} \tag{2.8}$$

Where $\sigma = s + 1/2 - 1/\nu$. Meyer (1992) shows that the Besov function norm $\|q\|_{B^s_{\nu,\pi}}$ is equivalent to the sequence norm $|q|_{B^s_{\nu,\pi}}$ of the wavelet coefficients of q . Therefore we will use the sequence norm to calculate the Besov norm $\|q\|_{B^s_{\nu,\pi}}$ in the sequel. We also consider a subset of Besov space $B^s_{\nu,\pi}$ such that $s\nu > 1$ and $\nu, \pi \in [1, \infty]$. The spaces of the unknown function $q(x)$ that we consider in this paper are defined by

$$F^s_{\nu,\pi}(M) = \{q : q \in B^s_{\nu,\pi}, \|q\|_{B^s_{\nu,\pi}} \le M, supp\ q \subseteq [0, 1]\},$$

i.e., $F^s_{\nu,\pi}(M)$ is a subset of functions with fixed compact support and bounded in the norm of one of the Besov spaces $B^s_{\nu,\pi}$. Moreover, $s\nu > 1$ implies that $F^s_{\nu,\pi}(M)$ is a subset of the space of bounded continuous functions.

3. Asymptotic results

3.1 Assumptions

Before describing our results, we formulate the following assumptions:

- A.1: There exist two constants ω_1 and ω_2 such that, for any $x \in [0, 1]$,

$$0 < \omega_1 < w(x) < \omega_2 < \infty$$

- A.2: There exist a known constant f_2 such that, for any $x \in [0, 1]$,

$$f(x) \le f_2$$

Also, In what follows, it is always assumed that, without loss of generality, the functions f and w are defined on the unit interval $[0, 1]$.

3.2 Main results

In following theorems, we consider the rate of convergence of wavelet estimators $\hat{q}_L(x)$ and $\hat{q}_H(x)$ under L^P risk function. They attain optimal and nearly optimal rates of convergence over a wide range of Besov space classes. Moreover,C denotes any constant that does not depend on i, j and n. Its value may change from one term to another and may depends on ϕ or ψ.

Theorem 3.1. *Let $p \geq 2$. Assume that the assumptions A.1 and A.2 hold and $q \in F^s_{\nu,\pi}(M)$ with $s > 1/r$, $r \geq 1$ and $\pi \geq 1$. Let $\hat{q}_L(x)$ be as in (2.6) with i_0 being the integer such that $2^{i_0} \simeq n^{1/1+2s}$. Then there exists a constant $C > 0$ such that*

$$E\left(\int_{[0,1]} |\hat{q}_L(x) - q(x)|^p\right) \leq Cn^{-\frac{ps}{2s+1}}$$

In following, theorem 3.2 explores the rates of convergence of $\hat{q}_H(x)$ under the L^p risk over Besov balls.

Theorem 3.2. *Under assumptions A.1 and A.2, when $p \geq 2$ and $\hat{q}_H(x)$ be as in (2.7) with i_s being the integer satisfying*

$$2^{i_s} \simeq \left(\frac{n}{\log_2 n}\right)^{1/1+2s}$$

and λ_n being the threshold:

$$\lambda_n = \kappa\sqrt{\frac{\log n}{n}}$$

Assume that $q \in F^s_{\nu,\pi}(M)$ with $s > 1/r$, $r \geq 1$ and $\pi \geq 1$. Then there exists a constant $C > 0$ such that

$$E\left(\int_{[0,1]} |\hat{q}_H(x) - q(x)|^p\right) \leq C\left(\frac{\log_2 n}{n}\right)^{-\frac{ps}{2s+1}}$$

If we do a global comparison between the results of Theorems 3.1 and 3.2, the rates of convergence achieved by $\hat{q}_H(x)$ are better than the one achieved by $\hat{q}_L(x)$. Moreover, let us recall that $\hat{q}_H(x)$ is adaptive while $\hat{q}_L(x)$ is not adaptive due to its dependence on s in its construction.

4. A Simulation Study

In this section we study the performance of our quantile density estimators based on (2.7) using the preliminary estimators for cumulative density function F which mentioned in (2.4). In the following example, we simulate a biased random sample from Beta distribution for different sample sizes and calculate the average norm criterion (ANorm) based on aforementioned competitors, where the ANorm criterion of the estimator $\hat{g}$ is defined as

$$ANorm = \frac{1}{N}\sum_{b=1}^{N}\left(\sum_{i=1}^{n}(\hat{f}_b(x_i) - f(x_i))^2\right)^{1/2}.$$

with $\hat{f}_b(.)$ being defined as an estimator of $f(.)$ at the b^{th} replication. The results in this simulation study are obtained using Daubechies's compactly supported wavelet Symmlet 4 (see Daubechies (1992), p. 198) and Coiflet 2 (see Daubechies (1992), p. 258), and primary resolution level $j_0 = 6$ based on $N = 100$ replications. The code was written in MATLAB environment using the WaveLab software. Lower values of ANorm are indicative of better performance. We also list the corresponding standard errors. Recall that when the parameters in Beta distribution are chosen from (0, 1), the corresponding quantile density function satisfies all the conditions required to prove the results..

Example. Here we generate the random samples $X_i,\ 1 \leq i \leq n$ from a Beta distribution with parameters $\alpha = 0.5,\ \beta = 0.6$ along with the following non-negative, bounded discontinuous biasing function,

$$w(x) = \begin{cases} 1, & for \quad x < 0.3, \\ x, & for \quad x \geq 0.3. \end{cases}$$

Figure 1 shows the original density function *pdf* with black line along with two versions of the new wavelet estimator of q, namely (i) the hard thresholding estimator ($\hat{q}_H$) with blue line, (ii) the linear wavelet estimator ($\tilde{q}_L$) with dotted green line and (iii) the smoothed version of the hard thresholding estimator ($\hat{q}_{SH}$) with dotted red line respectively.

Table 1 shows the values of ANorm and simulated standard errors for each one of the three different quantile wavelet estimators, using hard threshold estimator, a smoothed version of a hard threshold estimator, and the linear wavelet estimator for the beta density mentioned above for different sample sizes. The new smoothed version of hard threshold estimator has better performance compared to the hard threshold estimator and the linear wavelet estimator. Also, the hard threshold estimator is almost better than the linear wavelet estimator.

Table 1: Computed values of ANorm and simulated standard errors for various sample sizes; ANorms are located in the first row and standard errors in the second row.

Estimation Methods	ANorm and Simulated Standard Error				
	$n = 128$	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
The smoothed hard threshold estimator $\hat{g}_{SH}$	1.776	2.492	3.197	3.874	8.325
	0.157	0.156	0.141	0.121	0.184
The hard threshold estimator $\hat{g}_H$	1.885	2.643	3.583	4.90	8.859
	0.166	0.165	0.158	0.140	0.196
The linear wavelet estimator $\hat{g}_L$	2.484	5.145	3.969	5.435	8.448
	0.220	0.322	0.175	0.170	0.187

4.1 Auxiliary results

In the following section we provide some asymptotic results that are of importance in proving the theorem. The proof of Theorem 3.1 is a consequence of Propositions 4.2 and 4.3 of Chesneau and *et al.* (2015) and we describe them below. They show that the estimators $\hat{\beta}_{jk}$ defined by (2.5) satisfy a standard moment inequality and a specific concentration inequality. Before presenting these inequalities, the following lemma determines an upper bound for $|\hat{\beta}_{ij} - \beta_{ij}|$.

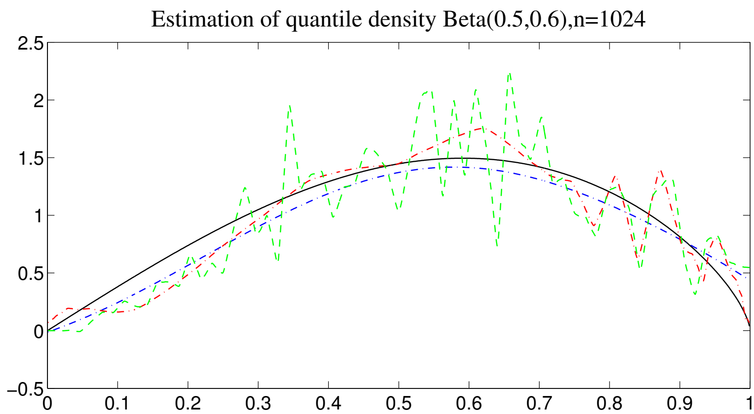
Lemma 4.1. *Suppose that the assumptions of Theorem 3.1 are satisfied. Then, for any $i \in \{i_0 + 1, ..., R\}$ and any $j \in \{0, ..., 2^i - 1\}$, the estimator $\hat{\beta}_{ij}$ defined by* (2.5) *satisfies*

$$\begin{aligned} |\hat{\beta}_{ij} - \beta_{ij}| &\leq K 2^{3i/2} \int_{[0,1]} |\hat{F}(x) - F(x)| dx \\ &\leq K 2^{3i/2} \sup_{[0,1]} |\hat{F}(x) - F(x)|. \end{aligned}$$

with $K = \sup_{[0,1]} |\psi'(x)|$ and $\psi'_{ij}(x) = 2^{3i/2} \psi(2^i x - j)$.

Proposition 4.2. *Let $p \geq 2$. Suppose that the assumptions of Theorem 3.1 are satisfied, then there exists a constant $C > 0$ such that, for any $i \geq i_0$, and n large enough, the estimator $\hat{\beta}_{ij}$, defined by (2.5) satisfies the following:*

$$\mathbb{E}\left(|\hat{\beta}_{ij} - \beta_{ij}|^{2p}\right) \leq C n^{-p} \tag{4.9}$$



]

Figure 1: Quantile biased density estimators via hard thresholding method; the true density function (*solid line*), the smoothed version of hard thresholding estimator (*blue line*), the hard thresholding estimator (*dotted red line*) and the linear wavelet estimator(*dotted green line*).

The expression in proposition (4.2) holds for $\hat{\alpha}_{ij}$ as well, replacing $\hat{\beta}_{ij}$ by $\hat{\alpha}_{ij}$ and β_{ij} by α_{ij}.

Proposition 4.3. *Let $p \geq 2$. Under the assumptions of Theorem 3.1, there exists a constant $c > 0$ such that, for any $i \geq i_0$, and large enough n, the estimators $\hat{\beta}_{ij}$ defined by (2.5) satisfy the following concentration inequality:*

$$\mathbb{P}\left(\left(|\hat{\beta}_{ij} - \beta_{ij}|\right) \geq \lambda_n\right) \leq 2\left(\frac{\log n}{n}\right)^p, \tag{4.10}$$

for some constant $C > 0$.

5. Proof

In this section, C represents a constant which may differ from one term to another. We suppose that n is large enough.

Proof of the propositions 4.2 and 4.3: Let us observe that

$$\begin{aligned}\hat{\beta}_{ij} - \beta_{ij} &= \int_{[0,1]} \left(\psi_{i_0 j}(\hat{F}(x)) - \psi_{i_0 j}(F(x))\right) dx \\ &= \int_{[0,1]} \left(\psi_{i_0 j}(\hat{U}(x)) - \psi_{i_0 j}(x)\right) q(x) dx\end{aligned}$$

with

$$\hat{U}(x) = \frac{\hat{\mu}}{n} \sum_{i=1}^{n} \frac{I(Y_i < x)}{w(x)}, \qquad U_i = F(X_i)$$

Then the proofs of propositions 4.2 and 4.3 follow from the technical part of (Kerkyacharian and Picard (2004), Subsection 9.2.2. pages 1093 - 1098) with q instead of $f(G^{-1})$ Let us mention that for the validity of results we need to suppose (A) and a restriction on i considered in our study, i.e., $2^i \leq \sqrt{n \ \log n}$.

Proof of the Theorem 3.1: Based on empirical wavelet expansion in Eq.(2.3), we can write

$$E\left(\int_{[0,1]} |\hat{q}_L(x) - q(x)|^p\right) \leq 2^{p-1}\left(T_1 + T_2\right) \tag{5.11}$$

where

$$T_1 = E\left(\int_{[0,1]} |\sum_{j=0}^{2^{i_0}-1} \left(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j}\right) \phi_{i_0 j}(x)|^p dx\right)$$

and

$$T_2 = \int_{[0,1]} \left|\sum_{i=i_0}^{\infty} \sum_{j=1}^{2^i-1} \beta_{ij}\psi_{ij}(x)\right|^p dx$$

Using Proposition 4.2 and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} T_1 &\leq C 2^{i_0(p/2-1)} \sum_{j=0}^{2^{i_0}-1} E\left(\left|\hat{\alpha}_{i_0 j} - \alpha_{i_0 j}\right|^p\right) \\ &\leq C 2^{i_0(p/2-1)} \sum_{j=0}^{2^{i_0}-1} \left(E\left(\left(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j}\right)^{2p}\right)\right)^{1/2} \\ &\leq C 2^{i_0(p/2-1)} 2^{i_0} n^{-p/2} = C(2^{i_0} n^{-1})^{p/2} \end{aligned} \tag{5.12}$$

On the other hand, by noting that $q \in F^s_{\nu,\pi}(M)$ and proceeding as in (Donoho and *et al.* (1996), eq. (24)), we have immediately

$$T_2 \leq C 2^{-i_0 sp} \tag{5.13}$$

It follows from (5.11), (5.12), (5.13) and the definition of i_0 that

$$E\left(\int_{[0,1]} |\hat{q}_L(x) - q(x)|^p\right) \leq C n^{-\frac{ps}{2s+1}} \tag{5.14}$$

Proof of the Theorem 3.2: For the sake of simplicity, we set $\hat{\theta}_{ij} = \hat{\beta}_{ij} - \beta_{ij}$. Applying the Minkowski inequality and an elementary inequality of convexity, we have $E\left(\|\hat{q}_H - q\|_p^p\right) \leq 4^{p-1}\left(T_1 + T_2 + T_3 + T_4\right)$, where

$$T_1 = E\|(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j})\phi_{i_0 j}(x)\|_p^p,$$

$$T_2 = E\|\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}\beta_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}| < \lambda_n)\|_p^p,$$

$$T_3 = E\|\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}\hat{\theta}_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}| \geq \lambda_n)\|_p^p,$$

$$T_4 = E\|\sum_{i=R+1}^{\infty}\sum_{j=0}^{2^i-1}\beta_{ij}\|_p^p,$$

In order to prove the above theorem, it suffices to bound each term T_1, T_2, T_3 and T_4 separately.

Lemma 5.1. *Assume $u \in \mathbb{R}^n$ and $\|u\|_p = (\sum_i |u_i|^p)^{1/p}$, for $0 < p_1 \leq p_2 \leq \infty$. Then the following inequalities hold:*

$$\|u\|_{p_2} \leq \|u\|_{p_1} \leq n^{\frac{1}{p_1}-\frac{1}{p_2}}\|u\|_{p_2}.$$

Lemma 5.2. *Using the L_p Minkowski inequality yields*

- $\{|\hat{\beta}_{ij}| < \kappa\lambda_n,\ |\beta_{ij}| \geq 2\kappa\lambda_n\} \subseteq \{|\hat{\beta}_{ij} - \beta_{ij}| \geq \frac{\kappa\lambda_n}{2}\}$
- $\{|\hat{\beta}_{ij}| \geq \kappa\lambda_n,\ |\beta_{ij}| < \frac{\kappa\lambda_n}{2}\} \subseteq \{|\hat{\beta}_{ij} - \beta_{ij}| \geq \frac{\kappa\lambda_n}{2}\}$
- $\{|\hat{\beta}_{ij}| < \kappa\lambda_n,\ |\beta_{ij}| \geq 2\kappa\lambda_n\} \subseteq \{|\beta_{ij}| \leq |\hat{\beta}_{ij} - \beta_{ij}|\}$

The upper bound for T_1: Using a L_p norm result on wavelet series (see [Hardel and *et al.* (1998), Proposition 8.3]), the Cauchy-Schwarz inequality and Proposition 4.2, we obtain

$$\begin{aligned}T_1 = E\|(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j})\phi_{i_0 j}(x)\|_p^p &\leq C2^{i_0(\frac{p}{2}-1)}\sum_{j=0}^{2^{i_0}-1}E\left(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j}\right)^p\\ &\leq C2^{i_0(\frac{p}{2}-1)}\sum_{j=0}^{2^{i_0}-1}\left(E\left(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j}\right)^{2p}\right)^{\frac{1}{2}}\\ &\leq C2^{i_0(\frac{p}{2}-1)}2^{i_0}n^{-\frac{p}{2}} = C(2^{i_0}n^{-1})^{\frac{p}{2}},\end{aligned}$$

Based on our choice of $i_0 = 0$, we have $T_1 = O(n^{-p/2})$.

The upper bound for T_4: First, let's consider $\nu < p$. From Lemma 5.1 and (2.8), we have $\|\beta_{i.}\|_p \leq \|\beta_{i.}\|_\nu \leq M2^{-i\sigma}$. Thus $\sum_j |\beta_{ij}|^p \leq M^p 2^{-ip\sigma}$. Since $s\nu > 1$ and $\sigma > 1/2$, we have

$$T_4 \leq C\left(\sum_{i=R+1}^{\infty} 2^{-i\sigma}\right)^p \leq C2^{-R\sigma p}$$

On the basis of our choice R with $2^R \simeq n(\log_2 n)^{-2}$ and $p\sigma > ps/(1+2s)$, we obtain $T_4 = O(n^{-ps/(1+2s)})$.
For $\nu \geq p$ which $p \geq 2$, from Lemma 5.1, we have $\|\beta_{i.}\|_p \leq (C2^i)^{\frac{1}{p}-\frac{1}{\nu}}\|\beta_{i.}\|_\nu \leq M2^{-is}$. However, we can show that

$$T_4 \leq C\left[\sum_{i=R+1}^{\infty}\left(\sum_{j=0}^{2^i-1}|\beta_{ij}|^p\right)^{\frac{1}{p}}\right]^p \leq C\left(\sum_{i=R+1}^{\infty} 2^{-is}\right)^p \leq C2^{-Rsp}$$

Again, on the basis of our choice R with $2^R \simeq n(\log_2 n)^{-2}$, we obtain $T_4 = O(n^{-ps/(1+2s)})$.
The upper bound for T_2: Applying the Minkowski inequality and an elementary inequality of convexity, we have $T_2 \leq 2^{p-1}(T_{21} + T_{22})$, where

$$\begin{aligned} T_{21} &= E\left(\| \textstyle\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1} \beta_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}| < \kappa\lambda_n)I(|\beta_{ij}| < 2\kappa\lambda_n)\|_p^p\right), \\ T_{22} &= E\left(\| \textstyle\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1} \beta_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}| < \kappa\lambda_n)I(|\beta_{ij}| \geq 2\kappa\lambda_n)\|_p^p\right) \end{aligned}$$

The upper bound for T_{21}: For the first term T_{21}, we have $T_{21} \leq 2^{p-1}(T_{211} + T_{212})$, where

$$\begin{aligned} T_{211} &= E\left(\|\sum_{i=i_0}^{i_s}\sum_{j=0}^{2^i-1} \beta_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}| < \kappa\lambda_n)I(|\beta_{ij}| < 2\kappa\lambda_n)\|_p^p\right), \\ &\leq C\|\sum_{i=i_0}^{i_s}\sum_{j=0}^{2^i-1} \beta_{ij}\psi_{ij}(x)I(|\beta_{ij}| < 2\kappa\lambda_n)\|_p^p \\ &\leq C\sum_{i=i_0}^{i_s} 2^{i(p/2-1)}\left[\left(\sum_{j=0}^{2^i-1}|\beta_{ij}|^p I(|\beta_{ij}| < 2\kappa\lambda_n)\right)^{1/p}\right]^p \end{aligned}$$

Now, from the definition of λ_n and by considering $2^{i_s} \simeq \left(n(\log_2 n)^{-1}\right)^{1/1+2s}$, we have

$$T_{211} \leq C \sum_{i=i_0}^{i_s} \left(2^{i(p/2-1)} 2^i (\lambda_n)^p\right) \leq C \left(2^{i_s} \lambda_n^2\right)^{p/2} \leq C \left(\frac{n}{\log_2 n}\right)^{-ps/1+2s}$$

The upper bound for T_{212}: For $\nu \geq 2$, based on Lemma 5.1, for any $g \in B^s_{\nu,q}$, we have

$$\begin{aligned} T_{212} &= E\left(\| \sum_{i=i_s+1}^{R} \sum_{j=0}^{2^i-1} \beta_{ij}\psi_{ij}(x) I(|\hat{\beta}_{ij}| < \kappa\lambda_n) I(|\beta_{ij}| < 2\kappa\lambda_n)\|_p^p\right), \\ &\leq C \sum_{i=i_s+1}^{R} \sum_{j} |\beta_{ij}|^p \leq C \sum_{i=i_s+1}^{R} 2^{-ips} \leq C 2^{-i_s ps} \leq C \left(\frac{n}{\log_2 n}\right)^{-ps/1+2s} \end{aligned}$$

Putting the upper bounds of T_{211} and T_{212} together, we conclude that

$$T_{21} \leq C \left(\frac{n}{\log_2 n}\right)^{-ps/1+2s} \tag{5.15}$$

The upper bound for T_{22}: By the Cauchy-Schwarz inequality and from Lemma 5.2, Proposition 4.2 and 4.3, we have

$$\begin{aligned} T_{22} &= E\left(\| \sum_{i=i_0}^{R} \sum_{j=0}^{2^i-1} \beta_{ij}\psi_{ij}(x) I(|\hat{\beta}_{ij}| < \kappa\lambda_n) I(|\beta_{ij}| \geq 2\kappa\lambda_n)\|_p^p\right) \\ &\leq C \sum_{i=i_0}^{R} \sum_{j=0}^{2^i-1} \left[E\left(|\hat{\beta}_{ij} - \beta_{ij}|^{2p}\right)\right]^{1/2} \left[P\left(|\hat{\beta}_{ij} - \beta_{ij}| > \frac{\kappa\lambda_n}{2}\right)\right]^{1/2} \\ &\leq C \sum_{i=i_0}^{R} 2^i n^{-p/2} \left(\frac{\log_2 n}{n}\right)^{p/2} \leq C 2^R n^{-p/2} \left(\frac{\log_2 n}{n}\right)^{p/2} \\ &\leq C \left(\frac{\log_2 n}{n}\right)^{p/2} \leq C \left(\frac{\log_2 n}{n}\right)^{ps/1+2s} \end{aligned} \tag{5.16}$$

Now, by using the results in Eq.(5.15) and (5.16), we have

$$T_2 \leq C \left(\frac{n}{\log_2 n}\right)^{-ps/1+2s}.$$

The upper bound for T_3: By the Minkowski inequality and an elementary inequality of convexity, we have $T_3 \leq 2^{p-1}(T_{31}+T_{32})$, where

$$T_{31} = E\|\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}\hat{\theta}_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}|\geq \kappa\lambda_n)I(|\beta_{ij}|<\frac{\kappa\lambda_n}{2})\|_p^p,$$
$$T_{32} = E\|\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}\hat{\theta}_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}|\geq \kappa\lambda_n)I(|\beta_{ij}|\geq\frac{\kappa\lambda_n}{2})\|_p^p$$

Applying the same argument as in T_2, to find an upper bound for T_{31} and T_{32}.
The upper bound for T_{31}: Using Lemma 5.2, the Cauchy-Schwarz inequality, and the propositions 4.2 and 4.3, we obtain

$$E\left(|\hat{\theta}_{ij}|^p I(|\hat{\beta}_{ij}|\geq\kappa\lambda_n)I(|\beta_{ij}|<\frac{\kappa\lambda_n}{2})\right)\leq\left[E(|\hat{\theta}_{ij}|)^{2p})\right]^{\frac{1}{2}}\left[P\left(|\hat{\beta}_{ij}-\beta_{ij}|>\frac{\kappa\lambda_n}{2}\right)\right]^{1/2}$$
$$\leq Cn^{-p} \tag{5.17}$$

From (5.17), and the fact that $\|\psi_{ij}\|_p^p = 2^{i(p/2-1)}\|\psi\|$, we have

$$\begin{aligned}
T_{31} &\leq CE\left(\left\|\left(\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}|\hat{\theta}_{ij}|^2 I(|\hat{\beta}_{ij}|\geq\kappa\lambda_n)I(|\beta_{ij}|<\frac{\kappa\lambda_n}{2})|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\right)\\
&\leq C\left\|\left(\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}\left[E(|\hat{\theta}_{ij}|^p I(|\hat{\beta}_{ij}|\geq\kappa\lambda_n)I(|\beta_{ij}|<\frac{\kappa\lambda_n}{2})\right]^{\frac{2}{p}}|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq Cn^{-p}\left\|\left(\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p \leq Cn^{-p}\sum_{i=i_0}^{R}\sum_{j=0}^{2^i-1}\|\psi_{ij}(x)\|_p^p\\
&\leq Cn^{-p}2^{R(p/2-1)}\leq Cn^{-p/2}
\end{aligned}$$

Where the last inequality arises from this fact $2^R\leq n$.
The upper bound for T_{32}: By the Minkowski inequality and an elementary inequality of convexity, we have $T_{32}\leq 2^{p-1}(T_{321}+T_{322})$, where

$$T_{321} = E\left\|\sum_{i=i_0}^{i_s}\sum_{j=0}^{2^i-1}\hat{\theta}_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}|\geq\kappa\lambda_n)I(|\beta_{ij}|\geq\tfrac{\kappa\lambda_n}{2})\right\|_p^p$$
$$T_{322} = E\left\|\sum_{i=i_s+1}^{R}\sum_{j=0}^{2^i-1}\hat{\theta}_{ij}\psi_{ij}(x)I(|\hat{\beta}_{ij}|\geq\kappa\lambda_n)I(|\beta_{ij}|\geq\tfrac{\kappa\lambda_n}{2})\right\|_p^p$$

The upper bound for T_{321}: Using a L_p norm result on wavelet series (see (Hardel and *et al.* (1998), Proposition 8.3)), Proposition 4.2 and the Cauchy-

Schwarz inequality, we obtain

$$\begin{aligned}
T_{321} &\leq CE\left(\left\|\left(\sum_{i=i_0}^{i_s}\sum_{j=0}^{2^i-1}|\hat{\theta}_{ij}|^2|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\right)\\
&\leq C\left\|\left(\sum_{i=i_0}^{i_s}\sum_{j=0}^{2^i-1}\left[E(|\hat{\theta}_{ij}|^p)\right]^{\frac{2}{p}}|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq Cn^{-\frac{p}{2}}\sum_{i=i_0}^{i_s}\sum_{j=0}^{2^i-1}\|\psi_{ij}(x)\|_p^p \leq Cn^{-\frac{p}{2}}\|\psi\|_p^p\sum_{i=i_0}^{i_s}2^i2^{i(p/2-1)}\\
&\leq C\left(2^{i_s}n^{-1}\right)^{\frac{p}{2}} \qquad (5.18)
\end{aligned}$$

The upper bound for T_{322}: First, we find the upper bound for $\nu \geq 2$. Nothing $2|\beta_{ij}|(\kappa\lambda_n)^{-1} \geq 1$ and from proposition 4.2, we have

$$\begin{aligned}
T_{322} &\leq CE\left(\left\|\left(\sum_{i=i_s+1}^{R}\sum_{j=0}^{2^i-1}|\hat{\theta}_{ij}|^2 I(|\beta_{ij}| \geq \frac{\kappa\lambda_n}{2})|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\right)\\
&\leq C\left\|\left(\sum_{i=i_s+1}^{R}\sum_{j=0}^{2^i-1}\left[E(|\hat{\theta}_{ij}|^p)\right]^{\frac{2}{p}} I(|\beta_{ij}| \geq \frac{\kappa\lambda_n}{2})|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq Cn^{-\frac{p}{2}}\left\|\left(\sum_{i=i_s+1}^{R}\sum_{j=0}^{2^i-1}|\beta_{ij}|\lambda_n^{-1}|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq Cn^{-\frac{p}{2}}\lambda_n^{-p/2}\sum_{i=i_s+1}^{R}\sum_{j}|\beta_{ij}|^p \leq C\sum_{i=i_s+1}^{R}2^{-ips} \leq C2^{-i_sps} \qquad (5.19)
\end{aligned}$$

It follows from (5.18), (5.19) and the definition of $2^{i_s} \simeq \left(\frac{\log n}{n}\right)^{1/1+2s}$ that $T_{32} = O\left(\left(\frac{\log n}{n}\right)^{-ps/1+2s}\right)$.

Finally, by Combining these four bounds together, we complete the proof of Theorem 3.2. □

References

Chesneau, C., Dewan, I and Doosti, H. (2015). Nonparametric estimation of a quantile density function by wavelet methods. *Computational Statistics and Data Analysis,* **94**, 161-174.

Daubechies, I. (1992). *Ten Lectures on Wavelets,* CBMS-NSF regional conferences series in applied mathematics. SIAM, Philadelphia.

Donoho, D. L. and Jonestone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. **90**, 1200-1224.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Annals of Statistics*, **24**, 508-539.

Hardle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). Wavelet, Approximation and Statistical Applications, Lectures Notes in Statistics, Springer Verlag. New York.

Jones, M.C. (1992). Estimating densities, quantiles, quantile densities and density quantiles, Annals Institute Statistical Mathematics 44, **4**, 721-727.

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets, *Bernoulli,* **10**, 1053-1105.

Meyer, Y. (1992). Wavelets and Operators, Cambridge University Press, Cambridge.

Parzen, E. (1979). Non parametric statistical data modeling, Journal of American Statistical Association, **74**, 105-122.

Parzen, E. (2004). Quantile Probability and Statistical Data Modeling, Statistical Science, **19**, 652-662.

Soni, P., Dewan, I. and Jain, K. (2012). Nonparametric estimation of quantile density function, Computational Statistics and Data Analysis, **56**, 12, 3876-3886.

Vidakovic, B. (1999). Statistical Modeling by Wavelets, John Wiley and Sons, Inc., New York.