Int. J. Nonlinear Anal. Appl. 14 (2023) 9, 367–377
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2023.28249.3821

Block thresholding in wavelet estimation of the regression function with errors in variables

Esmaeil Shirazi

Department of Statistics, Faculty of Science, Gonbad Kavous University, Gonbad Kavous 4971799151, Iran

(Communicated by Asadollah Aghajani)

Abstract

Errors-in-variables regression is the study of the association between covariates and responses where covariates are observed with errors. In this paper, we consider the estimation of regression functions when the independent variable is measured with error. We investigate the performances of an adaptive wavelet block thresholding estimator via the minimax approach under the L_p risk with $p \geq 1$ over Besov balls. We prove that it achieves the optimal rates of convergence.

Keywords: Adaptive estimation, Block Theresholding Method, Errors-in-variables, Minimax estimation, Nonparametric regression, Wavelets
2020 MSC: Primary 62G07; Secondary 62G20

1 Introduction and problem statement

Let $(Z_1, Y_1), (Z_2, Y_2), ..., (Z_n, Y_n)$ denote n independent pairs of random variables and consider the problem of estimating the regression function $g(z) = E(Y|Z = z)$. Due to the measuring mechanism or the nature of environment, the variable Z is measured with error and is not directly observable. Instead, Z is observed through $X = Z + U$, where U is a random noise. It is assumed that U has a known distribution and is independent of (X, Y). We aim to estimate g based on a given random sample $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ from the distribution of (X, Y).

The problem of estimating g for errors in variables model was originally investigated by several authors. For example: Nadaraya-Watson kernel type estimators, constructed as the ratio of two deconvolution kernel type estimators, see e.g. [1, 6, 7, 8, 11, 14]. One assumption usually done in all those works, is that the regularity of the regression function f and the regularity of the density g of the design are equal. In particular, when the regression function f and the density g admit kth-order derivatives, [7] give upper and lower bounds of the minimax risk for quadratic pointwise risk and for L_p risk on compact sets for ordinary and super smooth errors U.

In this article, we propose an extension of the wavelet estimator in [3] based on bivariate thresholding method and determine its convergence rate. We show that our estimator obtain the optimal rate of convergence under the mean integrated squared error (MISE) over Besov balls.

The paper is organized as follows. Assumptions on the model and some notations are introduced in Section 2. Section 3 briefly describes the periodized wavelet basis on $[0, 1]$ and the Besov balls. The estimators are presented in Section 4. The results are set in Section 5. The proofs are gathered in Section 6.

*Corresponding author
Email address: shirazi.esmaeel@gmail.com (Esmaeil Shirazi)

Received: March 2022 *Accepted:* September 2022

2 Wavelets, Besov balls and estimators

2.1 Some notations

Suppose that g belong to $L^2_{per}([0,1])$, the space of periodic functions of period one that are square-integrable on $[0,1]$:

$$L^2_{per}([0,1]) = \left\{ h: \ \|h\|_2 = \left(\int_0^1 h^2(x)dx \right)^{1/2} < \infty \right\}.$$

We assume that there exists a known constant $C_1 > 0$ such that

$$\|f\|_\infty = \sup_{x\in[0,1]} |f(x)| \le C_1 < \infty. \tag{2.1}$$

Any function $h \in L^2_{per}([0,1])$ can be represented by its Fourier series

$$h(t) = \sum_{k\in Z} \mathcal{F}(h)(k) e^{2i\pi kt}; \quad t \in [0,1],$$

where the equality is intended in mean-square convergence sense, and $\mathcal{F}(h)(k)$ denotes the Fourier coefficient given by

$$\mathcal{F}(h)(k) = \int_0^1 h(x) e^{2i\pi kx} dx; \quad k \in Z.$$

We consider the ordinary smooth case on g: there exist three constants, $c_g > 0, C_g > 0$ and $\delta > 1$, such that, for any $k \in Z$, the Fourier coefficient of g, i.e. $\mathcal{F}(g)(k)$, satisfies

$$\frac{c_g}{(1+k^2)^{\delta/2}} \le |\mathcal{F}(g)(k)| \le \frac{C_g}{(1+k^2)^{\delta/2}}. \tag{2.2}$$

This assumption controls the decay of the Fourier coefficients of g, and thus the smoothness of g. It is a standard hypothesis usually adopted in the field of nonparametric estimation for deconvolution problems. See e.g. [5], [12] and [16].

2.2 Meyer wavelets and Besov balls

In this part, for the purpose of this paper, we use the periodized Meyer wavelet bases on the unit interval. For any $x \in [0,1]$, any integer i and any $j \in \{0, ..., 2^i - 1\}$, let $\phi_{ij}(x) = 2^{i/2}\phi(2^i x - j)$ and $\psi_{ij}(x) = 2^{i/2}\psi(2^i x - j)$. We assume that the father and mother Meyer type wavelets, $\phi(x)$ and $\psi(x)$, are orthonormal and compactly supported over $[0,1]$. Also, we define

$$\phi^{per}_{ij} = \sum_{k\in Z} \phi_{ij}(x-k), \quad \psi^{per}_{ij} = \sum_{k\in Z} \psi_{ij}(x-k)$$

their periodized versions. There exists an integer i_0 such that the collection $\mathtt{B} = \{\phi^{per}_{i_0 j}, \ \ j = 1, ..., 2^{i_0} - 1, \ \ \psi^{per}_{ij}, \ \ i > i_0, \ \ j = 1, ..., 2^i - 1\}$ is an orthonormal basis $L^2_{per}([0,1])$. In what follows, the superscript "per" will be dropped to lighten the notation. Therefore, for all $g \in L^2_{per}([0,1])$, the wavelet expansion of g is

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

where

$$\alpha_{i_0 j} = \int_{[0,1]} g(x)\bar{\phi}_{i_0 j}(x)dx, \quad \beta_{ij} = \int_{[0,1]} g(x)\bar{\psi}_{i_0 j}(x))dx$$

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_{pq} , $s > 0$, $1 \le p, q \le \infty$. The parameter s measures the number of derivatives, where the existence of derivatives is required in an L^p-sense, whereas the parameter q 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. [19] and [4]).

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

$$|g|_{B^s_{pq}} = (\sum_j |\alpha_{i_0 j}|^p)^{1/p} + \{\sum_{i=i_0}^{\infty} [2^{i\sigma}(\sum_j |\beta_{ij}|^p)^{1/p}]^q\}^{1/q} \tag{2.4}$$

where $\sigma = s + 1/2 - 1/p$. [19] shows that the Besov function norm $\|g\|_{B^s_{pq}}$ is equivalent to the sequence norm $|g|_{B^s_{pq}}$ of the wavelet coefficients of g . Therefore we will use the sequence norm to calculate the Besov norm $\|g\|_{B^s_{pq}}$ in the sequel. We also consider a subset of Besov space B^s_{pq} such that $sp > 1, p, q \in [1, \infty]$. The spaces of regression functions that we consider in this paper are defined by

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

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

2.3 Block threshold estimator

The term-by-term hard thresholding procedure, estimates the function $g(x)$ by

$$\tilde{g}_H(x) = \sum_{j\in Z} \hat{\alpha}_{i_0 j}\phi_{i_0 j}(x) + \sum_{i\ge i_0}\sum_{j\in Z} \hat{\beta}_{ij} I(|\hat{\beta}_{ij}| > \kappa\lambda)\psi_{ij}(x), \tag{2.5}$$

where λ is a threshold and the empirical wavelet coefficients are defined by

$$\hat{\alpha}_{i_0 j} = \frac{1}{n}\sum_{\nu=1}^{n}\sum_{l\in C_i}(2i\pi l)\frac{\overline{F(\phi_{i_0 j})l}}{F(g)l} Y_\nu e^{-2i\pi l X_\nu} \tag{2.6}$$

$C_i = supp(F(\phi_{i,0})) = supp(F(\phi_{i,j}))$, and similarly,

$$\hat{\beta}_{ij} = \frac{1}{n}\sum_{\nu=1}^{n} G_\nu I(G_\nu \le \eta_i) \tag{2.7}$$

where

$$G_\nu = \sum_{l\in D_i}(2i\pi l)\frac{\overline{F(\psi_{ij})l}}{F(g)l} Y_\nu e^{-2i\pi l X_\nu}$$

$D_i = supp(F(\psi_{i,0})) = supp(F(\psi_{i,j}))$, and the threshold η_i is defined by

$$\eta_i = \theta 2^{\delta i}\sqrt{\frac{n}{\ln n}}. \tag{2.8}$$

The above term-by-term thresholded estimator (2.5) which is considered in [3], don't attain the optimal convergence rates of $n^{-2s/(2s+2\delta+1)}$, but do attain the rate $(n^{-1}\log_2 n)^{2s/(2s+2\delta+1)}$, which involves a logarithmic penalty. The reason is that a coefficient is more likely to contain a signal if neighboring coefficients do also. Therefore, incorporating information on neighboring coefficients will improve the estimation accuracy. But for a term-by-term thresholded estimator, other coefficients have no influence on the treatment of a particular coefficient.

A block thresholding estimator is to threshold empirical wavelet coefficients in groups rather than individually. It is constructed as follows. At each resolution level i, the integers j are divided among consecutive, nonoverlapping blocks of length l, say $\Gamma_{ik} = \{j : (k-1)l + 1 \le j \le kl,\ -\infty < k < \infty\}$. Within this block Γ_{ik}, the average estimated squared bias $\hat{A}_{ik} = l^{-1}\sum_{j\in B(k)}\hat{\beta}^p_{ij}$ will be compared to the threshold. Here, $B(k)$ refers to the set of indices j in block Γ_{ik}. If the average squared bias is larger than the threshold, all coefficients in the block will be kept. Otherwise, all coefficients will be discarded. For additional details, see [28].

Let $p \geq 2$ and $A_{ik} = l^{-1} \sum_{j\in B(k)} |\beta_{ij}|^p$ and estimating this with $\hat{A}_{ik}$, the block thresholding wavelet estimator of $g(x)$ becomes

$$\hat{g}(x) = \sum_{j\in Z} \hat{\alpha}_{i_0 j}\phi_{i_0 j}(x) + \sum_{i=i_0}^{R}\sum_{k}\sum_{j\in B(k)} \hat{\beta}_{ij}\psi_{ij}(x) I(\hat{A}_{ik} > cn^{-\frac{p}{2}}), \tag{2.9}$$

where the smoothing parameter R corresponds to the highest detail resolution level, parameter l is the block length and c is a threshold constant.

3 Asymptotic results

Our main theorem shows that the wavelet-based estimators, based on block thresholding of the empirical wavelet coefficients, attain optimal and nearly optimal convergence rates over a large range of Besov function classes.

Theorem 3.1. If the above conditions hold and if $\hat{g}(x)$ is as defined by (2.9), with the block length $l = \log n$ and $R = \lfloor \log_2(nl^{-2}) \rfloor$, then there exists a positive constant C such that for all $M > 0$ and $1/p < s < r$; $q \in [1, \infty]$:
1. if $p \in [2, \infty]$,

$$\sup_{g\in F^s_{p,q}(M)} E\int (\hat{g}(x) - g(x))^p \leq Cn^{-ps/(1+2s)},$$

2. if $p \in [1, 2)$,

$$\sup_{g\in F^s_{p,q}(M)} E\int (\hat{g}(x) - g(x))^p \leq C(\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{-ps/(1+2s)},$$

3.1 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 3.3 and 3.4 of [3] and we describe them below. They show that the estimators $\hat{\beta}_{jk}$ defined by (2.7) 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 3.2. 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 (2.7) satisfies

$$\begin{aligned} |\hat{\beta}_{ij} - \beta_{ij}| &\leq \left| \frac{1}{n}\sum_{\nu=1}^{n} \left(G_\nu I(G_\nu \leq \eta_i) - E\left(G_\nu I(G_\nu \leq \eta_i)\right)\right) + E\left(G_1 I(G_1 \leq \eta_i)\right) \right| \\ &\leq \left| \frac{1}{n}\sum_{\nu=1}^{n} \left(G_\nu I(G_\nu \leq \eta_i) - E\left(G_\nu I(G_\nu \leq \eta_i)\right)\right) \right| + E\left(\left|G_1\right| I(G_1 \leq \eta_i)\right). \end{aligned} \tag{3.1}$$

Now, we can show that,

$$E\left(\left|G_1\right| I(G_1 \leq \eta_i)\right) \leq \frac{E(G_1^2)}{\eta_i} \leq \frac{1}{\theta 2^{(\delta+d)i}}\sqrt{\frac{lnn}{n}}.$$

The inequality (3.1) holds for $\phi_{i,j}$ instead of $\psi_{i,j}$ and, a fortiori, $\hat{\alpha}_{i,j}$ instead of $\hat{\beta}_{i,j}$ and $\alpha_{i,j}$ instead of $\beta_{i,j}$. In addition to the inequality (3.1), the estimators $\hat{\beta}_{i,j}$ defined by (2.7) satisfy several specific probability inequalities. Two of them will be at the heart of the proof of the main result.

Proposition 3.3. 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 $j \geq j_0$, and n large enough, the estimator $\hat{\beta}_{jk}$, defined by (2.7) satisfies the following:

$$\mathbb{E}\left(|\hat{\beta}_{jk} - \beta_{jk}|^{2p}\right) \leq Cn^{-p} \tag{3.2}$$

The expression in (3.2) holds for $\hat{\alpha}_{jk}$ as well, replacing $\hat{\beta}_{jk}$ by $\hat{\alpha}_{jk}$ and β_{jk} by α_{jk}.

Proposition 3.4. Let $p \geq 2$. Under the assumptions of Theorem 3.1, there exists a constant $c > 0$ such that, for any $j \geq j_0$, and large enough n, the estimators $\hat{\beta}_{jk}$ defined by (2.7) satisfy the following concentration inequality:

$$\mathbb{P}\left(\left(\sum_{(i)} |\hat{\beta}_{jk} - \beta_{jk}|^p\right)^{1/p} \geq cn^{-p/2}\right) \leq Cn^{-p}, \tag{3.3}$$

for some constant $C > 0$.

4 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 Theorem 3.1: 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 $E\left(\|\hat{g} - g\|_p^p\right) \leq 4^{p-1}\left(T_1 + T_2 + T_3 + T_4\right)$, where

$$\begin{aligned}
T_1 &= E\|(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j})\phi_{i_0 j}(x)\|_p^p,\\
T_2 &= E\|\sum_{i=i_0}^{R}\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(x) I(\hat{A}_{ik} < cn^{-\frac{p}{2}})\|_p^p,\\
T_3 &= E\|\sum_{i=i_0}^{R}\sum_k \sum_{j\in B(k)} \hat{\theta}_{ij}\psi_{ij}(x) I(\hat{A}_{ik} \geq cn^{-\frac{p}{2}})\|_p^p,\\
T_4 &= E\|\sum_{i=R+1}^{\infty}\sum_{j=0}^{2^j-1} \beta_{ij}\|_p^p.
\end{aligned}$$

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

Lemma 4.1. 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 4.2. Using the L_p Minkowski inequality yields

$$\begin{aligned}
&1)\;\; I(\hat{A}_{ik} < cn^{-p/2}) I(A_{ik} \geq cn^{-p/2}) \leq I\left(|\hat{A}_{ik} - A_{ik}| \geq 2cn^{-p/2}\right)\\
&\qquad\qquad \leq I\left((l^{-1}\sum_{j\in B(k)} |\hat{\theta}_{ij}|^p) \geq 2cn^{-p/2}\right)\\
&2)\;\; I(\hat{A}_{ik} \geq cn^{-p/2}) I(A_{ik} < cn^{-p/2}) \leq I\left((l^{-1}\sum_{j\in B(k)} |\hat{\theta}_{ij}|^p) \geq 2cn^{-p/2}\right)
\end{aligned}$$

The upper bound for T_1: Using a L_p norm result on wavelet series (see [[9], Proposition 8.3]), the Cauchy-Schwarz inequality and Proposition 3.3, we obtain

$$\begin{aligned} T_1 = E\|(\hat{\alpha}_{i_0j}-\alpha_{i_0j})\phi_{i_0j}(x)\|_p^p &\leq C2^{i_0(\frac{p}{2}-1)}\sum_{j=0}^{2^{i_0}-1} E\left(\hat{\alpha}_{i_0j}-\alpha_{i_0j}\right)^p \\ &\leq C2^{i_0(\frac{p}{2}-1)}\sum_{j=0}^{2^{i_0}-1}\left(E\left(\hat{\alpha}_{i_0j}-\alpha_{i_0j}\right)^{2p}\right)^{\frac{1}{2}} \\ &\leq C2^{i_0(\frac{p}{2}-1)}2^{i_0}n^{-\frac{p}{2}} = C(2^{i_0}n^{-1})^{\frac{p}{2}}, \end{aligned}$$

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 4.1 and (2.4), 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 4.1, 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

$$\begin{aligned} T_{21} &= E\left(\|\textstyle\sum_{i=i_0}^{R}\sum_k\sum_{j\in B(k)}\beta_{ij}\psi_{ij}(x)I(\hat{A}_{ik} < cn^{-p/2})I(A_{ik} < cn^{-p/2})\|_p^p\right), \\ T_{22} &= E\left(\|\textstyle\sum_{i=i_0}^{R}\sum_k\sum_{j\in B(k)}\beta_{ij}\psi_{ij}(x)I(\hat{A}_{ik} < cn^{-p/2})I(A_{ik} \geq cn^{-p/2})\|_p^p\right). \end{aligned}$$

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

$$\begin{aligned} T_{211} = E&\left(\|\sum_{i=i_0}^{i_s}\sum_k\sum_{j\in B(k)}\beta_{ij}\psi_{ij}(x)I(\hat{A}_{ik} < cn^{-p/2})I(A_{ik} < cn^{-p/2})\|_p^p\right), \\ &\leq C\|\sum_{i=i_0}^{i_s}\sum_k\sum_{j\in B(k)}\beta_{ij}\psi_{ij}(x)I(A_{ik} < cn^{-p/2})\|_p^p \\ &\leq C\sum_{i=i_0}^{i_s}2^{i(p/2-1)}\left[\left(\sum_k\sum_{j\in B(k)}|\beta_{ij}|^p I(A_{ik} < cn^{-p/2})\right)^{1/p}\right]^p. \end{aligned}$$

Now, from the definition of A_{ik}, we have $\sum_{j\in B(k)}|\beta_{ij}|^p = lA_{ik}$. Since there are at most $l^{-1}2^i$ terms in $\sum_k$ for each i, we have

$$T_{211} \leq C\sum_{i=i_0}^{i_s}\left(2^{i(p/2-1)}2^i n^{-p/2}\right) \leq C\left(2^{i_s}n^{-1}\right)^{p/2}.$$

The upper bound for T_{212}: If $\nu \geq 2$, based on Lemma 4.1, for any $g \in B^s_{\nu,q}$, we have

$$T_{212} = E\left(\| \sum_{i=i_s+1}^{R} \sum_{k} \sum_{j \in B(k)} \beta_{ij}\psi_{ij}(x) I(\hat{A}_{ik} < cn^{-p/2}) I(A_{ik} < cn^{-p/2}) \|_p^p \right),$$
$$\leq C \sum_{i=i_s+1}^{R} \sum_{j} |\beta_{ij}|^p \leq C \sum_{i=i_s+1}^{R} 2^{-ips} \leq C 2^{-i_s ps}.$$

As to term T_{212}, for $\nu < 2$, nothing that $2cn^{-p/2}A_{ik}^{-1} > 1$, we have

$$T_{212} = E\left(\| \sum_{i=i_s+1}^{R} \sum_{k} \sum_{j \in B(k)} \beta_{ij}\psi_{ij}(x) I(\hat{A}_{ik} < cn^{-p/2}) I(A_{ik} < cn^{-p/2}) \|_p^p \right),$$
$$\leq C \sum_{i=i_s+1}^{R} \sum_{k} l A_{ik} 2^{i(p/2-1)} \|\psi\| \left(2cn^{-p/2} A_{ik}^{-1} \right)^{1-\frac{\nu}{p}}$$
$$\leq C l n^{-1+\frac{\nu}{2}} \sum_{i=i_s+1}^{R} 2^{i(p/2-1)} \sum_{k} A_{ik}^{\frac{\nu}{p}}$$
$$\leq C l^{1-\frac{\nu}{p}} n^{-\frac{p}{2}+\frac{\nu}{2}} \sum_{i=i_s+1}^{R} 2^{i(p/2-1)} 2^{-i\sigma\nu}$$
$$\leq C l^{1-\frac{\nu}{p}} n^{-\frac{p}{2}+\frac{\nu}{2}} 2^{-i_s\sigma\nu} 2^{i_s(p/2-1)}.$$

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

$$T_{21} \leq C \left(2^{i_s} n^{-p/2} + 2^{-i_s ps} \right). \tag{4.1}$$

The upper bound for T_{22}: From Lemma 4.2 and Proposition 3.4, we have

$$T_{22} = E\left(\| \sum_{i=i_0}^{R} \sum_{k} \sum_{j \in B(k)} \beta_{ij}\psi_{ij}(x) I(\hat{A}_{ik} < cn^{-p/2}) I(A_{ik} \geq cn^{-p/2}) \|_p^p \right)$$
$$\leq CE\left(\| (\sum_{i=i_0}^{R} \sum_{k} \sum_{j \in B(k)} |\beta_{ij}|^2 I(\hat{A}_{ik} < cn^{-p/2}) I(A_{ik} \geq cn^{-p/2}) |\psi_{ij}(x)|^2)^{\frac{1}{2}} \|_p^p \right)$$
$$\leq C\left(\| (\sum_{i=i_0}^{R} \sum_{k} \sum_{j \in B(k)} |\beta_{ij}|^2 \left[E(I(\hat{A}_{ik} < cn^{-p/2}) I(A_{ik} \geq cn^{-p/2})) \right]^{\frac{2}{p}} |\psi_{ij}(x)|^2)^{\frac{1}{2}} \|_p^p \right)$$
$$\leq C\left(\| (\sum_{i=i_0}^{R} \sum_{k} \sum_{j \in B(k)} |\beta_{ij}|^2 \left[P(l^{-1} \sum_{j \in B(k)} |\hat{\theta}_{ij}|^p \geq C2n^{-p/2}) \right]^{\frac{2}{p}} |\psi_{ij}(x)|^2)^{\frac{1}{2}} \|_p^p \right)$$
$$\leq Cn^{-\frac{p}{2}} \left(\| (\sum_{i=0}^{\infty} \sum_{k} \sum_{j \in B(k)} |\beta_{ij}|^2 |\psi_{ij}(x)|^2)^{\frac{1}{2}} \|_p^p \right) \leq C\|g\|_p^p n^{-\frac{p}{2}} \leq Cn^{-\frac{p}{2}} \tag{4.2}$$

Now, by using the results in Eq.(4.1) and (4.2), we have

$$T_2 \leq C \left(2^{i_s ps} + n^{-p/2} \right).$$

Now, if i_s satisfies $2^{i_s} \simeq n^{1/(1+2s)}$, then $T_2 \leq Cn^{-ps/(1+2s)}$.

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

$$T_{31} = E\|\sum_{i=i_0}^{R}\sum_{k}\sum_{j\in B(k)}\hat{\theta}_{ij}\psi_{ij}(x)I(\hat{A}_{ik}\geq cn^{-\frac{p}{2}})I(A_{ik}< cn^{-\frac{p}{2}})\|_p^p,$$
$$T_{32} = E\|\sum_{i=i_0}^{R}\sum_{k}\sum_{j\in B(k)}\hat{\theta}_{ij}\psi_{ij}(x)I(\hat{A}_{ik}\geq cn^{-\frac{p}{2}})I(A_{ik}\geq cn^{-\frac{p}{2}})\|_p^p.$$

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 4.2, the Cauchy-Schwarz inequality, and the propositions 3.3 and 3.4, we obtain

$$\begin{aligned} E\left(|\hat{\theta}_{ij}|^p I(\hat{A}_{ik}\geq cn^{-\frac{p}{2}})I(A_{ik}< cn^{-\frac{p}{2}})\right) &\leq E\left(|\hat{\theta}_{ij}|^p I(l^{-1}\sum_{j\in B(k)}|\hat{\theta}_{ij}|^p\geq C2n^{-p/2})\right)\\ &\leq \left[E(|\hat{\theta}_{ij}|)^{2p})\right]^{\frac{1}{2}}\left[P(l^{-1}\sum_{j\in B(k)}|\hat{\theta}_{ij}|^p\geq C2n^{-p/2})\right]^{\frac{1}{2}}\\ &\leq Cn^{-p}. \end{aligned} \tag{4.3}$$

From (4.3), and the fact that $\|\psi_{ij}\|_p^p = 2^{i(p/2-1)}\|\psi\|$, we have

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

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

$$\begin{aligned} T_{321} = &\quad E\left\|\textstyle\sum_{i=i_0}^{i_s}\sum_k\sum_{j\in B(k)}\hat{\theta}_{ij}\psi_{ij}(x)I(\hat{A}_{ik}\geq cn^{-\frac{p}{2}})I(A_{ik}\geq cn^{-\frac{p}{2}})\right\|_p^p\\ T_{322} = &\quad E\left\|\textstyle\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)}\hat{\theta}_{ij}\psi_{ij}(x)I(\hat{A}_{ik}\geq cn^{-\frac{p}{2}})I(A_{ik}\geq cn^{-\frac{p}{2}})\right\|_p^p. \end{aligned}$$

The upper bound for T_{321}**:** Using a L_p norm result on wavelet series (see ([9], Proposition 8.3)), Proposition

3.3 and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
T_{321} &\leq CE\left(\left\| \left(\sum_{i=i_0}^{i_s}\sum_k\sum_{j\in B(k)}|\hat{\theta}_{ij}|^2|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\right)\\
&\leq C\left\| \left(\sum_{i=i_0}^{i_s}\sum_k\sum_{j\in B(k)}\left[E(|\hat{\theta}_{ij}|^p)\right]^{\frac{2}{p}}|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq Cn^{-\frac{p}{2}}\sum_{i=i_0}^{i_s}\sum_k\sum_{j\in B(k)}\left\|\psi_{ij}(x)\right\|_p^p\\
&\leq Cn^{-\frac{p}{2}}\left\|\psi\right\|_p^p\sum_{i=i_0}^{i_s}2^i2^{i(p/2-1)}\leq C\left(2^{i_s}n^{-1}\right)^{\frac{p}{2}}. \qquad (4.4)
\end{aligned}$$

The upper bound for T_{322}: First, we find the upper bound for $\nu\geq 2$. Nothing $A_{ik}c^{-1}n^{\frac{p}{2}}\geq 1$ and from proposition 3.3, we have

$$\begin{aligned}
T_{322} &\leq CE\left(\left\| \left(\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)}|\hat{\theta}_{ij}|^2 I(A_{ik}\geq cn^{-\frac{p}{2}}))|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\right)\\
&\leq C\left\| \left(\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)}\left[E(|\hat{\theta}_{ij}|^p)\right]^{\frac{2}{p}}I(A_{ik}\geq cn^{-\frac{p}{2}}))|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq Cn^{-\frac{p}{2}}\left\| \left(\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)}A_{ik}n^{\frac{p}{2}}|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq C\sum_{i=i_s+1}^{R}\sum_j|\beta_{ij}|^p\leq C\sum_{i=i_s+1}^{R}2^{-ips}\leq C2^{-i_sps}. \qquad (4.5)
\end{aligned}$$

It follows from (4.4), (4.5) and the definition of $2^{i_s}\simeq n^{1/1+2s}$ that $T_{32}=O\left(n^{-ps/1+2s}\right)$. Also for $\nu<2$, we have

$$\begin{aligned}
T_{322} &\leq CE\left(\left\| \left(\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)}|\hat{\theta}_{ij}|^2 I(A_{ik}\geq cn^{-\frac{p}{2}}))|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\right)\\
&\leq C\left\| \left(\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)}\left[E(|\hat{\theta}_{ij}|^p)\right]^{\frac{2}{p}}I(A_{ik}\geq cn^{-\frac{p}{2}}))|\psi_{ij}(x)|^2\right)^{\frac{1}{2}}\right\|_p^p\\
&\leq C\sum_{i=i_s+1}^{R}\sum_k ln^{-\frac{p}{2}}\left(A_{ik}n^{\frac{p}{2}}\right)^{\nu/p}\left\|\psi_{ij}(x)\right\|_p^p\\
&\leq Cl^{1-\frac{\nu}{p}}n^{-\frac{p}{2}+\frac{\nu}{2}}\left\|\psi\right\|\sum_{i=i_s+1}^{R}2^{i(p/2-1)}\sum_k\left(\sum_{j\in B(k)}|\beta_{ij}|^p\right)^{\nu/p}\\
&\leq Cl^{1-\frac{\nu}{p}}n^{-\frac{p}{2}+\frac{\nu}{2}}2^{i_s(p/2-1)}\sum_{i=i_s+1}^{R}\sum_j|\beta_{ij}|^{\nu}\\
&\leq Cl^{1-\frac{\nu}{p}}n^{-\frac{p}{2}+\frac{\nu}{2}}2^{i_s(p/2-1)}2^{-i_s\sigma\nu}. \qquad (4.6)
\end{aligned}$$

Now, if $2^{i_s} \simeq (logn)^{\frac{(2-\nu)}{\nu(1+2s)}} n^{1/1+2s}$ and from (4.4), (4.6), then

$$T_{32} \leq C(\log_2 n)^{\frac{(2-\nu)+\nu^2(\sigma-2s/p)}{\nu(1+2s)}} n^{-ps/(1+2s)}.$$

Finally, by Combining these four bounds together, we complete the proof of Theorem 3.1. $\square$

References

[1] F. Comte and M.-L. Taupin, *Nonparametric estimation of the regression function in an errors-in-variables model*, Statistica Sinica **17** (2007), no. 3, 1065-1090.

[2] C.-L. Cheng and H. Schneeweiss, *Polynomial regression with errors in the variables*, J. Roy. Stat. Soc. Ser. B Stat. Methodol. **60** (1998), 189—199.

[3] C. Chesneau, *On adaptive wavelet estimation of the regression function and its derivatives in an errors-in-variables model*, Current Dev. Theory Appl. Wavelets **4** (2010), 185–208.

[4] D.L. Donoho and I.M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc. **90** (1995), no. 432, 1200–1224.

[5] J. Fan and J.Y. Koo, *Wavelet deconvolution*, IEEE Trans. Inf. Theory **48** (2002), 734—747.

[6] J. Fan and M. Masry, *Multivariate regression estimation with errors-in-variables: Asymptotic normality for mixing process*, J. Multivariate Anal. **43** (1992), 237–272.

[7] J. Fan and Y.K. Truong, *Nonparametric regression with errors-in-variables*, Ann. Statist. **21** (1993), no. 4, 1900–1925.

[8] J. Fan, Y.K. Truong and Y. Wang, *Nonparametric function estimation involving errors-in-variables*, Springer, 1991.

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

[10] P. Hall, G. Kerkyacharian D. Picard, *On the minimax optimality of block thresholded wavelet estimators*, Ann. Statist. **9** (1999), 33—50.

[11] D.A. Ioannides and P.D. Alevizos, *Nonparametric regression with errors in variables and applications*, Statist. Probab. Lett. **32** (1997), 35–43.

[12] I.M. Johnstone, G. Kerkyacharian, D. Picard and M. Raimondo, *Wavelet deconvolution in a periodic setting*, J. Roy. Statist. Soc. Ser. B **66** (2004), no. 3, 547—573.

[13] M.C. Jones, *Estimating densities, quantiles, quantile densities and density quantiles*, Ann. Inst. Statist. Math. **44** (1992), no. 4, 721–727.

[14] K. Kato and Y. Sasaki, *Uniform confidence bands for nonparametric errors-in-variables regression*, J. Econ. **213** (2019), no. 2, 516–555.

[15] J.Y. Koo and K.W. Lee, *B-splines estimations of regression functions with errors in variables*, Statist. Probab. Lett. **40** (1998), 57–66.

[16] M. Pensky and B. Vidakovic, *Adaptive wavelet estimator for nonparametric density deconvolution*, Ann. Statist. **27** (1999), 2033—2053.

[17] E. Masry, *Multivariate regression estimation with errors-in-variables for stationary processes*, Nonparamet. Statist. **3** (1993), 13–36.

[18] S. Mallat, *A wavelet tour of signal processing*, Elsevier/ Academic Press, Amsterdam, 2009.

[19] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, Cambridge, 1992

[20] N.U. Nair, P.G. Sankaran and B.V. Kumar, *Total time on test transforms of order n and its implications in reliability analysis*, J. Appl. Probab. **45** (2008), 1126–1139.

[21] N.U. Nair and P.G. Sankaran, *Quantile based reliability analysis*, Commun. Statist.: Theory Meth. **38** (2009), 222–232.

[22] E. Parzen, *Non parametric statistical data modeling*, J. Amer. Statist. Assoc. **74** (1979), 105–122.

[23] L. Peng and J.P. Fine, Nonparametric quantile inference with competing risks data, Biometrika **94** (2007), 735–744.

[24] N. Reid, *Estimating the median survival time*, Biometrika **68** (1981), 601–608.

[25] P.G. Sankaran and N.U. Nair, *Nonparametric estimation of hazard quantile function*, J. Nonparamet. Statist. **21** (2009), 757–767.

[26] P.G. Sankaran, I. Dewan and E.P. Sreedevi, *A non-parametric test for stochastic dominance using total time on test transform*, Amer. J. Math. Manag. Sci. **34** (2015), 162–183

[27] E. Shirazi, Y. Chaubey, H. Doosti, H.A. Nirumand, *Wavelet based estimation for the derivative of a density by block thresholding under random censorship*, J. Korean Statist. Soc. **41** (2012), 199—211.

[28] E. Shirazi, H. Doosti, H.A. Niroumand and N. Hosseinioun, *Nonparametric regression estimate with censored data based on block thresholding method*, J. Statist. Plan. Infer., **143** (2013), 1150–1165.

[29] E.V. Slud, D.P. Byar and S.B. Green, *A comparison of reflected versus test-based 355 confidence intervals for the median survival time, based on censored data*, Biometrics **40** (1984), 587–600.

[30] P. Soni, I. Dewan and K. Jain, *Nonparametric estimation of quantile density function*, Comput. Statist. Data Anal. **56** (2012), no. 12, 3876–3886.

[31] J.Q. Su and L.J. Wei, *Nonparametric estimation for the difference or ratio of median failure times*, Biometrics **49** (1993), 603—607.