Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 1518–1543
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

On the block thresholding wavelet estimators with censored data

Linyuan Li

Department of Mathematics and Statistics, University of New Hampshire, USA

Received 2 December 2005
Available online 24 January 2008

Abstract

We consider block thresholding wavelet-based density estimators with randomly right-censored data and investigate their asymptotic convergence rates. Unlike for the complete data case, the empirical wavelet coefficients are constructed through the Kaplan–Meier estimators of the distribution functions in the censored data case. On the basis of a result of Stute [W. Stute, The central limit theorem under random censorship, Ann. Statist. 23 (1995) 422–439] that approximates the Kaplan–Meier integrals as averages of i.i.d. random variables with a certain rate in probability, we can show that these wavelet empirical coefficients can be approximated by averages of i.i.d. random variables with a certain error rate in L^2. Therefore we can show that these estimators, based on block thresholding of empirical wavelet coefficients, achieve optimal convergence rates over a large range of Besov function classes $B^s_{p,q}, s > 1/p, p \geq 2, q \geq 1$ and nearly optimal convergence rates when $1 \leq p < 2$. We also show that these estimators achieve optimal convergence rates over a large class of functions that involve many irregularities of a wide variety of types, including chirp and Doppler functions, and jump discontinuities. Therefore, in the presence of random censoring, wavelet estimators still provide extensive adaptivity to many irregularities of large function classes. The performance of the estimators is tested via a modest simulation study.

MSC: primary 62G07; secondary 62G20

Keywords: Adaptive estimation; Besov spaces; Block thresholding; Censored data; Density estimation; Minimax estimation; Nonlinear wavelet-based estimator; Rates of convergence

1. Introduction

It is well known that in medical follow-up research, industrial life-testing and other studies, the observation on the survival time of a patient or a testing subject is often incomplete due to

E-mail address: linyuan@cisunix.unh.edu.

doi:10.1016/j.jmva.2008.01.001

right censoring. Classical examples of the causes of this type of censoring are that the patient was alive at the termination of the study, that the patient withdrew alive during the study, or that the patient died from causes other than those under study. In those cases only part of the observations are real death times. Our goal is to estimate nonparametrically the density functions of the survival times from censored data and investigate their asymptotic properties.

Formally, let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) survival times with a common distribution function F and a density function f. Also let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. censoring times with a common distribution function G. It is assumed that X_m is independent of Y_m for every m. Rather than observing $X_1, X_2, \ldots, X_n$, the variables of interest, in the randomly right-censored models, one observes $Z_m = \min(X_m, Y_m) = X_m \wedge Y_m$ and $\delta_m = I(X_m \le Y_m)$, $m = 1, 2, \ldots, n$, where $I(A)$ denotes the indicator function of the set A. We are interested in estimating f based on bivariate data (Z_m, δ_m), $m = 1, 2, \ldots, n$.

There is a huge literature on how to estimate density function or hazard rate function based on censored data; for example, nonparametric kernel estimators are typically used. However it is usually assumed that the underlying density function is a fixed smooth function. In this paper, we consider that the underlying density functions to be estimated belong to a large function class. We propose block thresholding wavelet estimators with censored data and investigate their asymptotic convergence rates over that large class of functions.

Nonparametric regression estimation by wavelets is developed in a series of works of Donoho and Johnstone [7–9] and Donoho et al. [10]. The recent monograph by Härdle et al. [15] and the book by Vidakovic [23] provide excellent systematic discussions on wavelets and their applications in statistics. Because wavelets are localized in both time and frequency and have remarkable approximation properties, wavelet estimators automatically adapt to these varying degrees of regularity (discontinuities, cusps, sharp spikes, etc.) of the underlying curves. Therefore, wavelet estimators typically achieve optimal convergence rates over a large class of functions with unknown degree of smoothness.

Most of nonlinear wavelet estimators are constructed through term-by-term thresholding of the empirical wavelet coefficients. These estimators usually achieve nearly optimal convergence rates within a logarithmic term. Hall et al. [13,14] introduce block thresholding wavelet estimators for density and regression respectively by shrinking wavelet coefficients in groups rather than individually. They show that block thresholded estimators achieve optimal global convergence rates without a logarithm term. Cai [2,3] study block thresholding via the approach of ideal adaptation with an oracle. On the basis of an oracle inequality, he investigates the asymptotic global and local rates of convergence and numerical properties of a class of block thresholding estimators for regression functions. It is shown that these estimators have excellent performance over a large range of Besov classes. However, all above estimators are constructed for the complete data case.

For randomly right-censored data, Antoniadis, Grégoire and Nason [1] describe a wavelet method for estimation of a single density and a single hazard rate function. They obtain the estimator's asymptotic normality and asymptotic mean integrated squared error (MISE). Li [16] considers a nonlinear wavelet estimator of a single density function with randomly censored data and derives its mean integrated squared error.

The objective of this paper is to propose block thresholding wavelet estimators with censored data for the density functions which belong to a large function class and investigate their asymptotic convergence rates. We show that these estimators attain optimal and nearly optimal rates of convergence over a wide range of Besov function classes. These results are analogous to those in [13] and [4] for density estimation in the complete data case.

In the next section, we give the elements of Besov spaces and wavelet transform, and provide block thresholding wavelet density estimators. The main results are described in Section 3. Section 4 contains a modest simulation study. The proofs of main results appear in Section 5 and Appendix.

2. Block thresholding wavelet estimators

This section contains some facts about wavelets that will be used in the sequel. Let $\phi(x)$ and $\psi(x)$ be father and mother wavelets, having the properties: ϕ and ψ are bounded and compactly supported, and $\int \phi = 1$. We call a wavelet ψ *r-regular* if ψ has r vanishing moments and r continuous derivatives. Let

$$\phi_{i_0 j}(x) = 2^{i_0/2}\phi(2^{i_0}x - j), \qquad \psi_{ij}(x) = 2^{i/2}\psi(2^i x - j), \qquad x \in \mathbb{R},\ i_0,\ i \in \mathbb{Z};$$

then, the collection $\{\phi_{i_0 j}, \psi_{ij}, i \ge i_0, j \in \mathbb{Z}\}$ is an orthonormal basis (ONB) of $L^2(\mathbb{R})$. Therefore, for all $f \in L^2(\mathbb{R})$,

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

where

$$\alpha_{i_0 j} = \int f(x)\phi_{i_0 j}(x)\mathrm{d}x, \qquad \beta_{ij} = \int f(x)\psi_{ij}(x)\,\mathrm{d}x.$$

The orthogonality properties of ϕ and ψ imply

$$\int \phi_{i_0 j_1}\phi_{i_0 j_2} = \delta_{j_1 j_2}, \qquad \int \psi_{i_1 j_1}\psi_{i_2 j_2} = \delta_{i_1 i_2}\delta_{j_1 j_2}, \qquad \int \phi_{i_0 j_1}\psi_{i j_2} = 0, \quad \forall i_0 \le i,$$

where δ_{ij} denotes the Kronecker delta, i.e., $\delta_{ij} = 1$, if $i = j$; and $\delta_{ij} = 0$, otherwise. For more on the wavelets see [6].

As is done in the wavelet literature, we investigate wavelet-based estimators' asymptotic convergence rates over a large range of Besov function classes $B^s_{p,q}$, $s > 0$, $1 \le p,\ q \le \infty$. Parameter s is an index of regularity or smoothness and parameters p and q are used to specify the type of norm. Besov spaces contain many traditional function spaces, in particular, the well-known Sobolev and Hölder spaces of smooth functions H^m and C^s ($B^m_{2,2}$ and $B^s_{\infty,\infty}$ respectively). Besov spaces also include significant spatial inhomogeneity function classes, such as the bump algebra and bounded variations classes. For a more detailed study we refer the reader to [22].

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

$$|f|_{B^s_{p,q}} = \left(\sum_j |\alpha_{i_0 j}|^p\right)^{1/p} + \left\{\sum_{i=i_0}^{\infty}\left[2^{i\sigma}\left(\sum_j |\beta_{ij}|^p\right)^{1/p}\right]^q\right\}^{1/q}, \tag{2.1}$$

where $\sigma = s + 1/2 - 1/p$. [19] shows that the Besov function norm $\|f\|_{B^s_{p,q}}$ is equivalent to the sequence norm $|f|_{B^s_{p,q}}$ of the wavelet coefficients of f. Therefore we will use the sequence norm to calculate the Besov norm $\|f\|_{B^s_{p,q}}$ in the sequel. We consider a subset of Besov space

$B^s_{p,q}$ such that $s > 1/p,\ p,\ q \in [1, \infty]$. The spaces of densities that we consider in this paper are defined by

$$F^s_{p,q}(M, L) = \left\{ f : \int f = 1,\ f \geq 0,\ f \in B^s_{p,q},\ \|f\|_{B^s_{p,q}} \leq M,\ \mathrm{supp} f \subset [-L,\ L] \right\},$$

i.e., $F^s_{p,q}(M, L)$ is a subset of densities with fixed compact support and bounded in the norm of one of the Besov spaces $B^s_{p,q}$. Moreover, $s > 1/p$ implies that $F^s_{p,q}(M, L)$ is a subset of the space of bounded continuous functions. Hence, we shall consider the intersection of this set $F^s_{p,q}(M, L)$ with $B_\infty(A)$, where $B_\infty(A)$ is the space of all functions f such that $\|f\|_\infty \leq A$.

In order to demonstrate the optimality of block thresholding wavelet estimators, following the notation of [13], let's consider another function space $\tilde{V}_{s_1}(F^s_{2,\infty}(M, L))$. This function space basically includes functions which may be written as a sum of a regular function in $F^s_{2,\infty}(M, L)$ and an irregular function in $F^{s_1}_{\tau,\infty}(M, L)$ with $\tau = (s + 1/2)^{-1}$. From embedding properties of Besov spaces ([15], p.124), we have $F^{s_1}_{\tau,\infty} \subseteq B^{s_1 - s}_{2,\infty}$. But whenever $s_1 < s + \frac{1}{2}$ (i.e., $s_1 - s < \frac{1}{2}$), space $F^{s_1}_{\tau,\infty}$ can be a very large function space which can include discontinues functions. For more discussion on this function space, see Remark 3.4 in the next section.

In our random censorship model, we observe $Z_m = \min(X_m, Y_m)$, and $\delta_m = I(X_m \leq Y_m)$, $m = 1, 2, \ldots, n$. Let $T < \tau_H$ be a fixed constant, where $\tau_H = \inf\{x : H(x) = 1\} \leq \infty$ is the least upper bound for the support of H, the distribution function of Z_1. We estimate $f_T(x)$, i.e., the density function $f(x)$ for $x \in [-L, T]$ (for the reason, see the following Remark 3.4). We can select wavelets ϕ and ψ as those in [5] such that they form an orthonormal basis of $L^2[-L,\ T]$. Hence, in the following, we also assume that $\phi_{i_0 j}$ and ψ_{ij} are compactly supported in $[-L,\ T]$ and form a complete orthonormal basis of $L^2[-L,\ T]$.

The wavelet expansion of $f_T(x)$ is

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

where

$$\alpha_{i_0 j} = \int f_T(x) \phi_{i_0 j}(x)\,\mathrm{d}x, \qquad \beta_{ij} = \int f_T(x) \psi_{ij}(x)\,\mathrm{d}x.$$

We will use the same notation as in [13], consider $i_0 = 0$ and write ϕ_j for ϕ_{0j}, α_j for α_{0j}, etc., we have

$$f_T(x) = \sum_j \alpha_j \phi_j(x) + \sum_{i=0}^{\infty} \sum_j \beta_{ij} \psi_{ij}(x).$$

Let $K(x, y) = \sum_j \phi(x - j)\phi(y - j)$ be the wavelet projection kernel. Then there exists a compactly supported $Q \in L^2$ such that $|K(x,\ y)| \leq Q(x - y)$ for all x and y. Similarly, define $K_i(x,\ y) = 2^i K(2^i x, 2^i y),\ i = 0, 1, 2, \ldots,$ and $K_i f(x) = \int K_i(x, y) f(y)\mathrm{d}y$. Then $D_i f(x) = \int D_i(x, y) f(y)\mathrm{d}y$, where $D_i(x, y) = \sum_j \psi_{ij}(x)\psi_{ij}(y) = K_{i+1}(x, y) - K_i(x, y)$. These $D_i(x, y)$ are called innovation kernels. In terms of the above notation, we have

$$f_T(x) = K_0 f(x) + \sum_{i=0}^{\infty} D_i f(x). \tag{2.2}$$

Without loss of generality, we can assume that there is a common compactly supported Q, such that

$$|K(x, y)| \le Q(x-y) \quad \text{and} \quad |D_0(x, y)| \le Q(x-y) \quad \text{for all } x \text{ and } y. \tag{2.3}$$

[6] shows that above conditions are met for certain compactly supported wavelets.

The above representation suggests the following estimators for $K_i f(x)$ and $D_i f(x)$:

$$\hat{K}_i(x) = \frac{1}{n}\sum_{m=1}^{n} K_i(x, X_m), \qquad \hat{D}_i(x) = \frac{1}{n}\sum_{m=1}^{n} D_i(x, X_m). \tag{2.4}$$

The term-by-term hard thresholded wavelet estimator of f_T (see [16, p.37]) is

$$\tilde{f}_T(x) = \sum_j \hat{\alpha}_j \phi_j(x) + \sum_{i=0}^{q}\sum_j \hat{\beta}_{ij} I(|\hat{\beta}_{ij}| > \lambda)\psi_{ij}(x), \tag{2.5}$$

where q is a smoothing parameter, λ is a threshold and the empirical wavelet coefficients are

$$\begin{aligned} \hat{\alpha}_j &= \int \phi_j(x) I(x \le T)\,\mathrm{d}\hat{F}_n(x) = \frac{1}{n}\sum_{m=1}^{n} \frac{\delta_m I(Z_m \le T)\phi_j(Z_m)}{1-\hat{G}_n(Z_m-)}, \\ \hat{\beta}_{ij} &= \int \psi_{ij}(x) I(x \le T)\,\mathrm{d}\hat{F}_n(x) = \frac{1}{n}\sum_{m=1}^{n} \frac{\delta_i I(Z_m \le T)\psi_{ij}(Z_m)}{1-\hat{G}_n(Z_m-)}. \end{aligned} \tag{2.6}$$

Here $\hat{F}_n$ and $\hat{G}_n$ denote the Kaplan–Meier estimators of distribution functions F and G, respectively, i.e.,

$$\begin{aligned} \hat{F}_n(x) &= 1 - \prod_{m=1}^{n}\left[1 - \frac{\delta_{(m)}}{n-m+1}\right]^{I(Z_{(m)} \le x)}, \\ \hat{G}_n(x) &= 1 - \prod_{m=1}^{n}\left[1 - \frac{1-\delta_{(m)}}{n-m+1}\right]^{I(Z_{(m)} \le x)}, \end{aligned}$$

where $Z_{(m)}$ is the m-th ordered Z-value and $\delta_{(m)}$ is the concomitant of the m-th-order Z statistic, i.e., $\delta_{(m)} = \delta_k$ if $Z_{(m)} = Z_k$. Note that $\delta_m/n(1-\hat{G}_n(Z_m-))$ is the jump of the Kaplan–Meier estimator $\hat{F}_n$ at Z_m.

The above term-by-term thresholded estimators (2.5) which are also considered in [17] don't attain the optimal convergence rates of $n^{-2s/(1+2s)}$, but do attain the rate $(n^{-1}\log_2 n)^{2s/(1+2s)}$, 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 $l^{-1}\sum_{j\in B(k)} \hat{\beta}_{ij}^2 (=: \hat{B}_{ik})$ 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 [2–4].

Let $B_{ik} = l^{-1} \sum_{j \in B(k)} \beta_{ij}^2$ and estimating this with $\hat{B}_{ik}(= l^{-1} \sum_{j \in B(k)} \hat{\beta}_{ij}^2)$, the block thresholding wavelet estimator of f_T becomes

$$\hat{f}_T(x) = \sum_j \hat{\alpha}_j \phi_j(x) + \sum_{i=0}^{R} \sum_k \sum_{j \in B(k)} \hat{\beta}_{ij} \psi_{ij}(x)\, I(\hat{B}_{ik} > C_0 n^{-1})$$

$$=: \hat{K}_0(x) + \sum_{i=0}^{R} \sum_k \hat{D}_{ik}(x) I(x \in J_{ik})\, I(\hat{B}_{ik} > C_0 n^{-1}), \tag{2.7}$$

where the last line defines $\hat{K}_0(x)$ and $\hat{D}_{ik}(x)$. The smoothing parameter R corresponds to the highest detail resolution level, parameter l is the block length and C_0 is a threshold constant. Notice that $\hat{D}_{ik}(x)$ is an estimate of $D_{ik} f(x) = \sum_{j \in B(k)} \beta_{ij} \psi_{ij}(x)$, and

$$J_{ik} = \bigcup_{j \in B(k)} \left\{ x : \psi_{ij}(x) \neq 0 \right\} = \bigcup_{j \in B(k)} \left\{ \operatorname{supp} \psi_{ij} \right\}.$$

Note that if the support of ψ is of length v, then the length of J_{ik} is $(l + v - 1)/2^i \leq 2l/2^i$, and these intervals overlap each other at either end by $(v - 1)2^{-i}$. Therefore, without loss of generality, we assume that the length of the support of ψ is 1; then these intervals J_{ik} are nonoverlapping.

Remark 2.1. We can define the wavelet estimator of $f(x)$ (i.e., on the whole support $[-L, L]$ of f without truncation at T), say $\hat{f}(x)$, instead of $f_T(x)$, similarly to (2.5) and (2.6) in this paper. However, in this case, the finiteness of the MISE, i.e. $E \int (\hat{f} - f)^2 < \infty$, cannot be ensured, because of the endpoint effect. This treatment is analogous to that of the MISE with kernel estimation, which includes a nonnegative bounded and compactly supported weight function w. Its role for w is to eliminate the endpoint effect [18, p.1523]. Thus we typically consider $\int (\hat{f}_T - f_T)^2$ to eliminate the endpoint effect. See also [16, p.38].

Remark 2.2. Although here we consider survival time setting, the random variables X and Y need not necessarily be positive. Suppose that there is no censoring, i.e., $G \equiv 0$ on $(-\infty, \infty)$. Then $\delta_m \equiv 1$, for all $m = 1, 2, \ldots, n$, and upon taking $T = \tau_H = \tau_F$, we see that $f_T \equiv f$ and the above estimator $\hat{f}_T$ is analogous to those of [13].

3. Main results

The following theorems show 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 and behave themselves as if they know in advance in which class the functions lie.

Theorem 3.1. *Let $\hat{f}_T$ be the block thresholding wavelet density estimator (2.7) with the block length $l = \log n$, $R = \lfloor \log_2(nl^{-2}) \rfloor$ and the threshold constant C_0 given as*

$$C_0 = 112.5\, A\, [1 - H(T)]^{-2} [1 - G(T)]^{-2} (C_2 \|Q\|_2 + \sqrt{2} \|Q\|_1 C_1^{-1/2})^2,$$

where C_1 and C_2 are the universal constants from Talagrand (1994) given in the Appendix. Suppose that the wavelets ϕ and ψ are r-regular. Then, there exists a constant C such that for all $M,\ L \in (0, \infty),\ 1/p < s < r,\ q \in [1, \infty]$:

1. *If* $p \in [2, \infty]$,

$$\sup_{f \in F^s_{p,q}(M,L) \cap B_\infty(A)} E \int \left(\hat{f}_T - f_T\right)^2 \le C\, n^{-2s/(1+2s)}.$$

2. *If* $p \in [1, 2)$,

$$\sup_{f \in F^s_{p,q}(M,L) \cap B_\infty(A)} E \int \left(\hat{f}_T - f_T\right)^2 \le C\, \left(\log_2 n\right)^{\frac{2-p}{p(1+2s)}}\, n^{-2s/(1+2s)}.$$

Remark 3.1. The above block thresholding wavelet estimators defined as in (2.7) are adaptive in the sense that they don't depend on unknown parameters s, p and q. When $p \ge 2$, minimax theory indicates that the best convergence rates over $F^s_{p,q}(M, L)$ are at $n^{-2s/(1+2s)}$. Thus, the above estimators achieve exact optimal convergence rates, without one knowing the smoothness parameters. This result is analogous to that for the complete data case. In the case $1 \le p < 2$, the above convergence rates are the same as those of the *BlockJS* estimators given in [2]. From [8,9], the traditional linear estimators (including the kernel estimator with a fixed bandwidth) cannot achieve the rates stated in Theorem 3.1. Hence using our above block thresholding estimators has advantages over the traditional linear method.

Remark 3.2. Our block length $l = \log n$ is different from that, $l = (\log n)^2$, in [13]. From [2,3], this block length $l = \log n$ is optimal both for the global error measure and estimating functions at a point simultaneously. Our threshold constant C_0 with censored data is analogous to that in the complete data case, but it depends on the censoring distribution G also.

Theorem 3.2. *Let $\hat{f}_T$ be the block thresholding wavelet density estimator* (2.7) *with the block length $l = \log n$, $R = \lfloor \log_2(nl^{-2}) \rfloor$ and the threshold constant C_0 given as*

$$C_0 = 112.5\, A\, [1 - H(T)]^{-2} [1 - G(T)]^{-2} (C_2 \|Q\|_2 + \sqrt{2} \|Q\|_1 C_1^{-1/2})^2,$$

where C_1 and C_2 are the universal constants from [24] *given in the Appendix. Suppose that the wavelet ψ is r-regular with $r > s_1$ and $s/(1+2s) < s_1 - s$. Then, there exists a constant C such that for all $M, L \in (0, \infty)$, $1/2 < s$,*

$$\sup_{f \in \tilde{V}_{s_1}(F^s_{2,\infty}(M,L)) \cap B_\infty(A)} E \int \left(\hat{f}_T - f_T\right)^2 \le C\, n^{-2s/(1+2s)}.$$

Remark 3.3. The above result, in Theorem 3.2, is analogous to those in [13,4] for the complete data case.

Remark 3.4. From the characterization of Besov spaces, it can be verified that a function which has a finite number of jumps and the "regularity" ν elsewhere belongs to class $F^\nu_{\nu^{-1},\infty}$ (see [15, p. 114]). Now from embedding properties of Besov spaces [15, p. 124], it can be shown that $F^{s+1/2}_{\tau,\infty}$ ($= F^\nu_{\nu^{-1},\infty}$ with $\nu = s + 1/2$) is included in $F^{s_1}_{\tau,\infty}$ from $s + 1/2 > s_1$. Therefore $F^{s_1}_{\tau,\infty}$ can be a much larger space than the regular space $F^s_{2,\infty}$, since it can contain discontinuous functions whenever $s_1 < s + 1/2$ or $s_1 < \tau^{-1}$, where $\tau = (s + 1/2)^{-1}$. In Theorem 3.2, we only require $s/(1 + 2s) < s_1 - s$, which may include the case $s/(1 + 2s) < s_1 - s < 1/2$ or $s_1 < s + 1/2$. Therefore, function space $F^{s_1}_{\tau,\infty}$ can contain discontinuous functions. For the kernel estimators with fixed bandwidth, the convergence rates could not achieve rates at $n^{-2s/(1+2s)}$ over $\tilde{V}_{s_1}(F^s_{2,\infty}(M, L))$ if the underlying functions to be estimated are discontinuous. Minimax theory indicates that the best convergence rates over $F^s_{2,\infty}$ are at $n^{-2s/(1+2s)}$. Thus, Theorem 3.2 shows that wavelet estimators attain optimal convergence rates over a large class of functions which include discontinuous functions.

4. Simulation results

To investigate the performance of the proposed wavelet estimator, we present a modest simulation study. However, the proposed estimators in Theorems 3.1 and 3.2 are of purely theoretical interest; they reduce the mean integrated squared error based on variance–bias tradeoff. They are not practical for implementation, since the threshold constant C_0 depends on unknown constants A, C_1, C_2 and unknown distributions F and G. In this simulation study, we determine the threshold C_0 by *cross-validation*, which minimizes the prediction error generated by comparing a prediction, based on half of the data, to the other half of the data. For details on cross-validation, see [20]. In order to compare our wavelet estimators to the existing competitors, we choose the [12] local linear regression smoother for comparison. Although the estimators are devised for nonparametric regression settings, one can use a proper definition of binned values to convert a density problem into a standard regression problem (we are aware that there is a certain approximating error from the effect of binning). Therefore, for convenience of comparison, we consider the regression settings in the simulation study and use the same function as in [12], *i.e.,*

$$Y_i = 4.5 - 64X_i^2(1 - X_i)^2 - 16(X_i - 0.5)^2 + 0.25\epsilon_i, \quad \epsilon_i \sim \text{i.i.d. } N(0, 1).$$

For convenience of the discrete wavelet transform, we let $X_i = i/n, i = 1, 2, \dots, n$, where n is the sample size. We consider three different sample sizes: $n = 256,\ 512$ and 1024. The censoring time T_i is conditionally independent of the survival time Y_i given X_i and is distributed as $(T_i|X_i = x) \sim \exp(t(x))$, where $t(x)$ is the mean conditional censoring time given by

$$t(x) = \begin{cases} 3(1.25 - |4x - 1|), & \text{if} \quad 0 \le x \le 0.5; \\ 3(1.25 - |4x - 3|), & \text{if} \quad 0.5 < x \le 1. \end{cases}$$

For the above censoring variable, approximately 40% of the data are censored. We also consider another censoring variable $(T_i|X_i = x) \sim \exp(2.2 * t(x))$ which results in approximately 20% of data are censored. Fan and Gijbels [12] use the explicit local average transformation to replace the censored observation Z_i with Y_i^*, which is a weighted average of all uncensored responses which are larger than Z_i within a small neighborhood of X_i, i.e.,

$$Y_i^* = \frac{\sum\limits_{j:Z_j>Z_i} Z_j K\left(\frac{X_i - X_j}{(X_{i+k} - X_{i-k})/2}\right)\delta_j}{\sum\limits_{j:Z_j>Z_i} K\left(\frac{X_i - X_j}{(X_{i+k} - X_{i-k})/2}\right)\delta_j},$$

where K is a nonnegative kernel function and k plays the role of the bandwidth. The value of k can be determined by *cross-validation*. For numerical comparisons we consider the average norm (ANorm) of the estimators at the sample points

$$\text{ANorm} = \frac{1}{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{n}\left(\hat{f}_l(x_i) - f(x_i)\right)^2\right)^{1/2},$$

Table 1
Average norm from $N = 100$ replications

	20% censored			40% censored		
	$n = 256$	$n = 512$	$n = 1024$	$n = 256$	$n = 512$	$n = 1024$
Local linear	3.844	5.549	7.886	3.869	5.497	7.860
Wavelet	3.867	5.519	7.774	3.850	5.463	7.780

Table 2
Average norm from $N = 100$ replications

	20% censored		40% censored	
	$n = 256$	$n = 512$	$n = 256$	$n = 512$
Local linear	7.691	9.199	8.540	10.721
Wavelet	7.332	8.941	9.421	11.851

where $\hat{f}_l$ is the estimate of f in the l-th replication and N is the total number of replications. Since different wavelets yield very similar results, we only use Daubechies' compactly supported wavelet *Symmlet* 8. We find that our wavelet estimators based on cross-validation are very close to the Stein unbiased risk estimator [8] in terms of the mean integrated squared error. The simulation results for different sample sizes and different censoring proportion are summarized in Table 1. On the basis of these results, we see that our wavelet estimator has a very similar average norm which is usually slightly smaller than that of the local linear regression smoother.

The second example that we considered is the following model: $Y_i = g(X_i) + \epsilon_i$, where $\epsilon_i \sim$ i.i.d. $N(0, 1)$, $i = 1, \ldots, n$ ($n = 256$ and 512), and $g(x)$ is a piecewise HeaviSine function:

$$g(x) = \begin{cases} 4\sin(4\pi x) + 20, & \text{if} \quad 0 \le x < 0.3; \\ 4\sin(4\pi x) + 18, & \text{if} \quad 0.3 \le x < 0.7; \\ 4\sin(4\pi x) + 20, & \text{if} \quad 0.7 \le x \le 1. \end{cases}$$

We also consider the censoring time T_i to be conditionally independent of the survival time Y_i given X_i and to be distributed as $(T_i | X_i = x) \sim \exp(t(x))$, where $t(x) = 4g(x)$ results in approximately 40% censoring and $t(x) = 2g(x)$ results in approximately 20% censoring. The average norms for two estimators for different sample sizes and censorings are summarized in Table 2. On the basis of the above simulation results, we found that for light censoring (approximately 20%) and moderate sample sizes, our wavelet estimator does slightly better than Fan and Gijbels' local linear estimator. However, for heavier censoring, this advantage is lost.

Remark 4.1. It is well known that wavelet methods are very effective in estimating functions which have locally varying (heterogeneous) degree of smoothing (inhomogeneity), i.e., these functions are quite smooth on one part of the domain but much less regular on another part, whereas kernel estimators with fixed bandwidth are an appropriate tool for estimating functions with homogeneous degree of smoothness (regular functions). Typically, kernel methods ask for a certain degree of smoothness for the underlying functions, while wavelet methods can deal with functions which belong to a quite general function space. Therefore, wavelet methods do very well in estimating the peaks and valleys of the underlying curves, but they are not very well satisfied over the smooth portion. Wavelet methods are also very suitable in estimating the sudden changes, such as discontinuities. It is evident that, because of the adaptability of the wavelet estimators to many different types of nonsmoothness, a price will be paid on the estimation of a truly smooth curve. Of course, wavelet methods can't replace the other smoothing methods, but they do complement each other. From this small simulation study, we see that wavelet estimators are very comparable to the [12] local linear regression smoother, which involves a variable bandwidth.

5. Proofs

The overall proofs for the above theorems follow along the lines of [13] and [11] for density estimation in the complete data case. But moving from complete data to censored data involves a significant change in complexity. For complete data, the empirical wavelet coefficients would be defined as averages of i.i.d. random variables. In this case, it is relatively easy to investigate the large deviation behavior of the empirical wavelet coefficients. For the censored data case, the empirical wavelet coefficients are constructed through the Kaplan–Meier estimators of the distribution functions as in (2.6). Hence they are no longer sums of i.i.d. random variables. Li [17] considers term-by-term thresholded estimators with censored data which don't attain the optimal convergence rates as in this paper. The proof for the block thresholding estimators is significantly different from that for the term-by-term thresholded estimator in [17]. For the block thresholding estimators, we will use a result from Talagrand (1994) to deal with large deviation behavior of the empirical wavelet coefficients, instead of the standard Bernstein inequality.

The key part of the proof is approximating the empirical wavelet coefficients with averages of i.i.d. random variables with a sufficiently small error rate. In the complete data case, we may write empirical wavelet coefficients as integrals with respect to an empirical distribution function. Naturally, in the random censored data case, we may write empirical wavelet coefficients as integrals with respect to the Kaplan–Meier estimator. In this case, they are no longer sums of i.i.d. random variables. Stute [21] approximates the Kaplan–Meier integrals as averages of i.i.d. random variables with a certain rate in probability. Nevertheless we are able to show that these empirical wavelet coefficients can be approximated by averages of i.i.d. random variables with a certain error rate in L^2 also, since the MISE considers L^2 error (for details see the following Lemma 5.1).

The proof of the above theorems can be broken into several parts. In view of (2.2) and (2.7), we have

$$E\|\hat{f}_T - f_T\|_2^2 \le 4(I_1 + I_2 + I_3 + I_4),$$

where

$$\begin{aligned}
I_1 &= E\left\|\hat{K}_0 - K_0 f\right\|_2^2,\\
I_2 &= E\left\|\sum_{i=0}^{i_s}\left[\sum_k \hat{D}_{ik} I(J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f\right]\right\|_2^2,\\
I_3 &= E\left\|\sum_{i=i_s+1}^{R}\left[\sum_k \hat{D}_{ik} I(J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f\right]\right\|_2^2,\\
I_4 &= \left\|\sum_{i=R+1}^{\infty} D_i f\right\|_2^2,
\end{aligned} \tag{5.1}$$

where i_s be the integer such that $2^{i_s} \simeq n^{1/(2s+1)}$ (i.e., $2^{i_s} \le n^{1/(2s+1)} < 2^{i_s+1}$) for $p \ge 2$ and $2^{i_s} \simeq (\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{1/(2s+1)}$ for $1 \le p < 2$. In order to prove the above theorems, it suffices to bound each term $I_1,\ I_2,\ I_3$ and I_4 separately, which is done in the following Lemmas 5.4–5.7 respectively.

In order to prove Lemmas 5.4–5.7, we need some preparations. We begin with some lemmas. The following first lemma is completely analogous to Lemma 4.1 of [16] with different notation ($\hat{\alpha}_j, \bar{\alpha}_j, \hat{\beta}_{ij}$ and $\bar{\beta}_{ij}$ here play the roles of $\hat{b}_j, \tilde{b}_j, \hat{b}_{ij}$ and $\tilde{b}_{ij}$ there with $p = 1$, and similarly for U, V and W). The second and third lemmas concern inequalities which will be used in the sequel. All proofs of these lemmas are omitted. For proofs, see [16,13,2].

Lemma 5.1. *Let $\hat{\alpha}_j$ and $\hat{\beta}_{ij}$ be defined as in Eq. (2.6). Also, let*

$$\begin{aligned}
&\varphi_j(x) = \phi_j(x) I(x \le T), \quad j = 0, \pm 1, \pm 2, \ldots,\\
&\varphi_{ij}(x) = \psi_{ij}(x) I(x \le T),\ i = 0, 1, \ldots, R; \quad j = 0, \pm 1, \pm 2, \ldots,\\
&\bar{\alpha}_j = \frac{1}{n}\sum_{m=1}^{n} \frac{\delta_m \varphi_j(Z_m)}{1 - G(Z_m)}, \quad j = 0, \pm 1, \pm 2, \ldots,\\
&\bar{\beta}_{ij} = \frac{1}{n}\sum_{m=1}^{n} \frac{\delta_m \varphi_{ij}(Z_m)}{1 - G(Z_m)}, \quad i = 0, 1, \ldots, R;\ j = 0, \pm 1, \pm 2, \ldots.
\end{aligned}$$

Then the following equations hold:

$$\begin{aligned}
&\hat{\alpha}_j = \bar{\alpha}_j + \overline{W}_j + R_{n,j}, \qquad E(R_{n,j}^2) = O\left(\frac{1}{n^2}\right) \int \varphi_j^2 \, \mathrm{d}F,\\
&\hat{\beta}_{ij} = \bar{\beta}_{ij} + \overline{W}_{ij} + R_{n,ij}, \qquad E(R_{n,ij}^2) = O\left(\frac{1}{n^2}\right) \int \varphi_{ij}^2 \, \mathrm{d}F,
\end{aligned}$$

where

$$\begin{aligned}
&W_j(Z_m) = U_j(Z_m) - V_j(Z_m), \qquad W_{ij}(Z_m) = U_{ij}(Z_m) - V_{ij}(Z_m),\\
&\overline{W}_j = \frac{1}{n}\sum_{m=1}^{n} W_j(Z_m), \qquad \overline{W}_{ij} = \frac{1}{n}\sum_{m=1}^{n} W_{ij}(Z_m),
\end{aligned}$$

and

$$\begin{aligned}
U_j(Z_m) &= \frac{1 - \delta_m}{1 - H(Z_m)} \int_{Z_m}^{\tau_H} \varphi_j(w)\, F(\mathrm{d}w),\\
U_{ij}(Z_m) &= \frac{1 - \delta_m}{1 - H(Z_m)} \int_{Z_m}^{\tau_H} \varphi_{ij}(w)\, F(\mathrm{d}w),\\
V_j(Z_m) &= \int_{-L}^{\tau_H} \int_{-L}^{\tau_H} \frac{\varphi_{ij}(w) I(v < Z_m \wedge w)}{[1 - H(v)][1 - G(v)]} G(\mathrm{d}v) F(\mathrm{d}w),\\
V_{ij}(Z_m) &= \int_{-L}^{\tau_H} \int_{-L}^{\tau_H} \frac{\varphi_{ij}(w) I(v < Z_m \wedge w)}{[1 - H(v)][1 - G(v)]} G(\mathrm{d}v) F(\mathrm{d}w).
\end{aligned}$$

Remark 5.1. This lemma shows that empirical wavelet coefficients $\hat{\beta}_{ij}$ (or $\hat{\alpha}_j$) can be approximated by averages of i.i.d. random variables $\bar{\beta}_{ij}$ (or $\bar{\alpha}_j$) and $\overline{W}_{ij}$ (or $\overline{W}_j$) with a sufficiently small term $R_{n,ij}$ (or $R_{n,j}$). Through the detailed calculation in [16], we are able to show that $\overline{W}_{ij}$ (or $\overline{W}_j$) is negligible compared to the main term $\bar{\beta}_{ij}$ (or $\bar{\alpha}_j$). Hence, asymptotically, empirical wavelet coefficients $\hat{\beta}_{ij}$ (or $\hat{\alpha}_j$) with censored data are equivalent to empirical wavelet coefficients $\bar{\beta}_{ij}$ (or $\bar{\alpha}_j$) for the complete data case.

The following lemma is very useful when one calculates the sequence norm of a function for different values of p.

Lemma 5.2. *Let $u \in \mathbb{R}^n$, $\|u\|_p = (\sum_i |u_i|^p)^{1/p}$, and $0 < p_1 \le p_2 \le \infty$. Then the following inequalities hold:*

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

Lemma 5.3. *Let $D_i f$ and $\hat{D}_i$ be defined as in (2.2) and (2.4). Then*

$$E\left\|\sum_{i=I}^{J}\left(\hat{D}_i - D_i f\right)\right\|_2^2 \le \left\{\sum_{i=I}^{J}\left[E\int\left(\hat{D}_i - D_i f\right)^2\right]^{1/2}\right\}^2.$$

Lemma 5.4. *Suppose that the assumptions of Theorem 3.1 hold. Then*

$$I_1 = E\left\|\hat{K}_0 - K_0 f\right\|_2^2 = o\left(n^{-2s/(1+2s)}\right).$$

Proof. On the basis of the orthogonality of wavelets ϕ, we have

$$I_1 = \sum_j E\left(\hat{\alpha}_j - \alpha_j\right)^2.$$

From Lemma 5.1,

$$\begin{aligned} I_1 \le\ & 3\left\{\sum_j E(\bar{\alpha}_j - \alpha_j)^2 + \sum_j E\overline{W}_{j_0 k}^2 + \sum_j E R_{n,j}^2\right\} \\ =:\ & 3(I_{11} + I_{12} + I_{13}). \end{aligned}$$

Apply the same arguments as in Lemma 4.2 in [16, pp. 42, 43] and notice that when $p = 1$ in [16], we have

$$I_{11} = O(n^{-1}p) = O(n^{-1}), \quad I_{12} = o(n^{-1}p) = o(n^{-1}), \quad I_{13} = O(n^{-2}p) = O(n^{-2}).$$

Hence, Lemma 5.4 is proved. □

In the proofs below, C represents a generic finite constant, the concrete value of which may change from line to line in the sequel.

Lemma 5.5. *Suppose that the assumptions of Theorem 3.1 hold. Then*

$$I_4 = \left\|\sum_{i=R+1}^{\infty} D_i f\right\|_2^2 = o\left(n^{-2s/(1+2s)}\right).$$

Proof. On the basis of the orthogonality of wavelets ψ, we have

$$I_4 = \sum_{i=R+1}^{\infty}\sum_j \beta_{ij}^2.$$

From the wavelet expansion of f_T in (2.2), the wavelet coefficients $\beta_{ij} = \int f_T(x)\psi_{ij}(x)\mathrm{d}x$. Because of $\operatorname{supp} f_T \subset [-L, T]$ and $\operatorname{supp}\psi \subset [-v, v]$, we have, for any level i, that there are at most $C2^i$ nonzero coefficients β_{ij}.

First, let's consider $p < 2$. From Lemma 5.2 and (2.1), we have $\|\beta_{i.}\|_2 \le \|\beta_{i.}\|_p \le M2^{-i\sigma}$. Thus $\sum_j \beta_{ij}^2 \le M^2 2^{-2i\sigma}$. Since $sp > 1$ and $\sigma > 1/2$, we have $I_4 \le \sum_{i=R+1}^{\infty} M^2 2^{-2i\sigma} = M^2 2^{-2R\sigma} 2^{-2\sigma}(1-2^{-2\sigma})^{-1} \le M^2 2^{-2R\sigma}$. On the basis of our choice R with $2^R \simeq n(\log_2 n)^{-2}$ and $2\sigma = 1 + 2(s - 1/p) > 2s/(2s+1)$, we obtain $I_4 = o\left(n^{-2s/(1+2s)}\right)$.

For $p \ge 2$, from Lemma 5.2, we have $\|\beta_{i.}\|_2 \le (C2^i)^{\frac{1}{2}-\frac{1}{p}} \|\beta_{i.}\|_p \le C2^{-is}$. Thus, we have

$$I_4 \le C \sum_{R+1}^{\infty} 2^{-2is} = C2^{-2Rs}2^{-2s}(1-2^{-2s})^{-1} \le C2^{-2Rs}.$$

Again, on the basis of our choice of R and $s > 0$, we have $I_4 = o\left(n^{-2s/(1+2s)}\right)$.

Taken together with $p < 2$, this completes the proof of the lemma. $\square$

Lemma 5.6. *Suppose that the assumptions of Theorem 3.1 hold with $p \ge 2$ and $2^{i_s} \simeq n^{1/(2s+1)}$. Then*

$$I_2 = E\left\| \sum_{i=0}^{i_s} \left[\sum_k \hat{D}_{ik} I(J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f \right] \right\|_2^2 \le Cn^{-2s/(1+2s)}.$$

Suppose that the assumptions of Theorem 3.2 hold with $1 \le p < 2$ and $2^{i_s} \simeq (\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{1/(2s+1)}$. Then

$$I_2 = E\left\| \sum_{i=0}^{i_s} \left[\sum_k \hat{D}_{ik} I(J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f \right] \right\|_2^2 \le C\left(\log_2 n\right)^{\frac{2-p}{p(1+2s)}} n^{-2s/(1+2s)}.$$

Proof. From Lemma 5.3, we have

$$I_2 \le \left\{ \sum_{i=0}^{i_s} \left[E\int \left(\sum_k \hat{D}_{ik}(x) I(x \in J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f(x) \right)^2 \mathrm{d}x \right]^{1/2} \right\}^2.$$

Writing $D_i f(x) = \sum_j \beta_{ij}\psi_{ij}(x) = \sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(x) =: \sum_k D_{ik} f(x)$, we have for the term in brackets

$$\begin{aligned}
&E\int \left(\sum_k \hat{D}_{ik}(x) I(x \in J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f(x) \right)^2 \mathrm{d}x \\
&\le 3\left\{ E\int \left[\sum_k \left(\hat{D}_{ik}(x) - D_{ik} f(x) \right) I(x \in J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) \right]^2 \mathrm{d}x \right. \\
&\quad + E\int \left[\sum_k D_{ik} f(x) I(\hat{B}_{ik} \le C_0 n^{-1}) I(B_{ik} \le 2C_0 n^{-1}) \right]^2 \mathrm{d}x \\
&\quad \left. + E\int \left[\sum_k D_{ik} f(x) I(\hat{B}_{ik} \le C_0 n^{-1}) I(B_{ik} > 2C_0 n^{-1}) \right]^2 \mathrm{d}x \right\}
\end{aligned}$$

$$\le 3\left\{E\int\left(\hat{D}_i(x)-D_i f(x)\right)^2 \mathrm{d}x\right.$$
$$+E\sum_k\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x I(B_{ik}\le 2C_0n^{-1})$$
$$\left.+E\sum_k\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x I(\hat{B}_{ik}\le C_0n^{-1})I(B_{ik}>2C_0n^{-1})\right\}$$
$$=: 3\left(I_{21}+I_{22}+I_{23}\right),$$

where the last inequality follows from the orthogonality of ψ and $\hat{D}_i(x) := \sum_j \hat{\beta}_{ij}\psi_{ij}(x)$.

As to the first term I_{21}, from Lemma 5.1, we have

$$\begin{aligned} I_{21} &= \sum_j E(\hat{\beta}_{ij}-\beta_{ij})^2 \\ &\le 3\left\{\sum_j E(\bar{\beta}_{ij}-\beta_{ij})^2+\sum_j E\overline{W}_{ij}^2+\sum_j ER_{n,ij}^2\right\} \\ &=: 3(I_{211}+I_{212}+I_{213}). \end{aligned}$$

Through direct calculation as in term I_1 (see also (4.15) and (4.16) in [16, p. 43]), we can get

$$I_{211}=O(n^{-1}2^i), \qquad I_{212}=o(n^{-1}2^i), \qquad I_{213}=O(n^{-2}2^i).$$

Thus, we have $I_{21}\le C2^in^{-1}$.

As to the term I_{22}, from the definition of B_{ik}, we have $\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x = \sum_{j\in B(k)}\beta_{ij}^2 = lB_{ik}$. Since there are at most $l^{-1}2^i$ terms in $\sum_k$ for each i, we have $I_{22} = \sum_k l2C_0n^{-1} \le C2^in^{-1}$.

The term I_{23} involves large deviation behavior and will be considered separately in the next section. Assume that $I_{23}\le Cn^{-1}$ for all i for the time being; we obtain

$$\begin{aligned} I_2 &\le \left\{\sum_{i=0}^{i_s}\left[C2^in^{-1}+C2^in^{-1}+Cn^{-1}\right]^{1/2}\right\}^2 \\ &\le C\left\{\sum_{i=0}^{i_s}\left[(2^in^{-1})^{1/2}+n^{-1/2}\right]\right\}^2 \\ &\le C\left(2^{i_s}n^{-1}+i_s^2n^{-1}\right). \end{aligned}$$

Now, if i_s satisfies $2^{i_s}\simeq n^{1/(2s+1)}$, then $I_2\le Cn^{-2s/(1+2s)}$. If i_s satisfies $2^{i_s}\simeq (\log_2 n)^{\frac{2-p}{p(1+2s)}}n^{1/(2s+1)}$, then $I_2\le C\left(\log_2 n\right)^{\frac{2-p}{p(1+2s)}}n^{-2s/(1+2s)}$, which completes the proof of the lemma. □

Lemma 5.7. *Suppose that the assumptions of Theorem 3.1 hold with $p\ge 2$ and $2^{i_s}\simeq n^{1/(2s+1)}$. Then*

$$I_3 = E\left\|\sum_{i=i_s+1}^{R}\left[\sum_k \hat{D}_{ik}I(J_{ik})I(\hat{B}_{ik}>C_0n^{-1})-D_if\right]\right\|_2^2 \le C\,n^{-2s/(1+2s)}.$$

Suppose that the assumptions of Theorem 3.2 hold with $1 \le p < 2$ *and* $2^{i_s} \simeq (\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{1/(2s+1)}$. *Then*

$$I_3 = E \left\| \sum_{i=i_s+1}^{R} \left[\sum_k \hat{D}_{ik} I(J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f \right] \right\|_2^2 \le C \left(\log_2 n\right)^{\frac{2-p}{p(1+2s)}} n^{-2s/(1+2s)}.$$

Proof. From Lemma 5.3, we have

$$I_3 \le \left\{ \sum_{i=i_s+1}^{R} \left[E \int \left(\sum_k \hat{D}_{ik} I(J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f(x) \right)^2 \mathrm{d}x \right]^{1/2} \right\}^2.$$

Applying the same argument as in I_2, we can write the term in brackets as

$$\begin{aligned} &E \int \left(\sum_k \hat{D}_{ik} I(J_{ik}) I(\hat{B}_{ik} > C_0 n^{-1}) - D_i f(x) \right)^2 \mathrm{d}x \\ &\le E \sum_k \int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik} f(x) \right)^2 \mathrm{d}x \, I(\hat{B}_{ik} > C_0 n^{-1}) I(B_{ik} > C_0 (2n)^{-1}) \\ &\quad + E \sum_k \int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik} f(x) \right)^2 \mathrm{d}x \, I(\hat{B}_{ik} > C_0 n^{-1}) I(B_{ik} \le C_0 (2n)^{-1}) \\ &\quad + E \sum_k \int_{J_{ik}} (D_{ik} f(x))^2 \, \mathrm{d}x \, I(\hat{B}_{ik} \le C_0 n^{-1}) I(B_{ik} \le 2C_0 n^{-1}) \\ &\quad + E \sum_k \int_{J_{ik}} (D_{ik} f(x))^2 \, \mathrm{d}x \, I(\hat{B}_{ik} \le C_0 n^{-1}) I(B_{ik} > 2C_0 n^{-1}) \\ &=: I_{31} + I_{32} + I_{33} + I_{34}. \end{aligned}$$

The terms I_{32} and I_{34} involve large deviation behavior and will be considered separately in the next section. Assume that $I_{32} \le C n^{-1}$ and $I_{34} \le C n^{-1}$ for all i for the time being.

Let's consider the first part of the lemma with $p \ge 2$ and $2^{i_s} \simeq n^{1/(2s+1)}$. For the first term I_{31}, noticing $B_{ik} C_0^{-1} 2n \ge 1$ and $E(\hat{\beta}_{ij} - \beta_{ij})^2 \le C n^{-1}$ for all i and j, we have

$$\begin{aligned} I_{31} &\le \sum_k E \int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik} f(x) \right)^2 \mathrm{d}x \cdot B_{ik} C_0^{-1} 2n \\ &\le C \sum_k \sum_{j \in B(k)} E(\hat{\beta}_{ij} - \beta_{ij})^2 \cdot l^{-1} \sum_{j \in B(k)} \beta_{ij}^2 \, n \\ &\le C \sum_k l \, n^{-1} \cdot l^{-1} \sum_{j \in B(k)} \beta_{ij}^2 \, n = C \sum_j \beta_{ij}^2. \end{aligned}$$

When $p \ge 2$, as in Lemma 5.5, we have $\sum_j \beta_{ij}^2 \le C 2^{-2si}$. Hence $I_{31} \le C 2^{-2si}$.

Similarly, when $p \ge 2$,

$$I_{33} \le \sum_k \int_{J_{ik}} (D_{ik} f(x))^2 \, \mathrm{d}x \cdot I(B_{ik} \le 2C_0 n^{-1})$$

$$\leq C \sum_j \beta_{ij}^2 \leq C 2^{-2si}.$$

Combine these four terms together; we have, for $p \geq 2$,

$$\begin{aligned}
I_3 &\leq \left[\sum_{i=i_s+1}^{R} \left(C2^{-2si} + Cn^{-1} + C2^{-2si} + Cn^{-1} \right)^{1/2} \right]^2 \\
&\leq C \left[\sum_{i=i_s+1}^{R} \left(2^{-si} + n^{-1/2} \right) \right]^2 \\
&\leq C \left(2^{-2si_s} + R^2 n^{-1} \right) \\
&\leq C n^{-2s/(1+2s)}.
\end{aligned}$$

Now let's consider the second part of the lemma with $1 \leq p < 2$ and $2^{i_s} \simeq (\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{1/(2s+1)}$. Since $C_0^{-1} 2nB_{ik} > 1$ in I_{31}, we have

$$\begin{aligned}
I_{31} &\leq E \sum_k \int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik} f(x) \right)^2 \mathrm{d}x \cdot I(B_{ik} > C_0 (2n)^{-1}) \\
&\leq C \sum_k l n^{-1} \cdot n^{\frac{p}{2}} B_{ik}^{\frac{p}{2}} \\
&\leq C l n^{-1+\frac{p}{2}} \sum_k \left(l^{-1} \sum_{j \in B(k)} \beta_{ij}^2 \right)^{\frac{p}{2}} \\
&\leq C l^{1-\frac{p}{2}} n^{-1+\frac{p}{2}} \sum_k \left(\sum_{j \in B(k)} \beta_{ij}^2 \right)^{\frac{p}{2}}.
\end{aligned}$$

From Lemma 5.2, when $p < 2$, we have $\|\beta_{i\cdot}\|_2 \leq \|\beta_{i\cdot}\|_p$, thus $\left(\sum_{j \in B(k)} \beta_{ij}^2 \right)^{p/2} \leq \sum_{j \in B(k)} \beta_{ij}^p$. Hence,

$$I_{31} \leq C l^{1-\frac{p}{2}} n^{-1+\frac{p}{2}} \sum_j \beta_{ij}^p \leq C l^{1-\frac{p}{2}} n^{-1+\frac{p}{2}} M^p 2^{-i\sigma p},$$

and the last inequality follows from Lemma 5.5 when $p < 2$.

As to the term I_{33}, for $p < 2$, noticing that $2C_0 n^{-1} B_{ik}^{-1} \geq 1$, we have

$$\begin{aligned}
I_{33} &\leq \sum_k \int_{J_{ik}} (D_{ik} f(x))^2 \,\mathrm{d}x \cdot I(B_{ik} \leq 2C_0 n^{-1}) \\
&\leq \sum_k l B_{ik} \cdot \left(2C_0 n^{-1} B_{ik}^{-1} \right)^{1-\frac{p}{2}} \\
&= C l n^{-1+\frac{p}{2}} \sum_k B_{ik}^{\frac{p}{2}} \\
&\leq C l^{1-\frac{p}{2}} n^{-1+\frac{p}{2}} M^p 2^{-i\sigma p},
\end{aligned}$$

and the last step follows from that of I_{31}.

Therefore, when $1 \leq p < 2$, we have

$$
\begin{aligned}
I_3 &\leq \left\{ \sum_{i=i_s+1}^{R} \left[C(n^{-1}l)^{1-\frac{p}{2}} 2^{-i\sigma p} + Cn^{-1} + C(n^{-1}l)^{1-\frac{p}{2}} 2^{-i\sigma p} + Cn^{-1} \right]^{1/2} \right\}^2 \\
&\leq C \left\{ \sum_{i=i_s+1}^{R} \left[(n^{-1}l)^{\frac{1}{2}-\frac{p}{4}} 2^{-i\sigma p/2} + Cn^{-\frac{1}{2}} \right] \right\}^2 \\
&\leq C \left((n^{-1}l)^{1-\frac{p}{2}} 2^{-i_s \sigma p} + R^2 n^{-1} \right) \\
&\leq C \left(\log_2 n \right)^{\frac{2-p}{p(1+2s)}} n^{-2s/(1+2s)},
\end{aligned}
$$

which proves the second part of the lemma. $\square$

We are now in a position to give the proofs of Theorems 3.1 and 3.2.

Proof of Theorem 3.1. The proof follows from Lemmas 5.4–5.7 and the fact that

$$E\|\hat{f}_T - f_T\|_2^2 \leq 4(I_1 + I_2 + I_3 + I_4). \quad \square$$

Proof of Theorem 3.2. The proof is similar to that of Theorem 3.1 and is simpler, since we only need to consider the $p \geq 2$ case. Notice that the large deviation results for empirical wavelet coefficients and the proofs for term I_1 in Lemma 5.4 and term I_2 in Lemma 5.6 still hold for the new function space $\tilde{V}_{s_1}(F^s_{2,\infty}(M, L))$, since these results do not depend on the smoothness parameter s. Hence in order to prove Theorem 3.2, it suffices to prove Lemmas 5.5 and 5.7.

Let's consider Lemma 5.5 first. In the sequel, for the sake of convenience, we write f as f_T, suppressing T. Recall that for all $f \in \tilde{V}_{s_1}(F^s_{2,\infty}(M, L))$ f may be written as $f = f_1 + f_2$ where $f_1 \in F^s_{2,\infty}(M, L)$ and $f_2 \in F^{s_1}_{\tau,\infty}(M, L)$. Then, from Lemma 5.5, we have $I_4 \leq 2(I_{4,1} + I_{4,2})$, where $I_{4,1} = \sum_{i=R+1}^{\infty} \sum_j \beta^2_{ij,1}$, $I_{4,2} = \sum_{i=R+1}^{\infty} \sum_j \beta^2_{ij,2}$, $\beta_{ij,1} = \int f_1 \psi_{ij}$ and $\beta_{ij,2} = \int f_2 \psi_{ij}$. Since $f_1 \in F^s_{2,\infty}(M, L)$, from Lemma 5.5 we have $I_{4,1} = o\left(n^{-2s/(1+2s)}\right)$. As to $I_{4,2}$, using the inclusion properties of Besov spaces, we have $f_2 \in F^{s_1}_{\tau,\infty}(M, L) \subseteq B^{s_1-s}_{2,\infty}(M, L)$. Applying the norm equivalence between the Besov norm of a function and the sequence norm of wavelet coefficients of a function in (2.1), we have $\sum_j \beta^2_{ij,2} = \|\beta_{i.,2}\|^2 \leq M 2^{-2i(s_1-s)}$. Hence, $I_{4,2} \leq M \sum_{i=R+1}^{\infty} 2^{-2i(s_1-s)} = M\left(n^{-1}(\log_2 n)^2\right)^{2(s_1-s)}$. From the additional condition $s/(1+2s) < s_1 - s$ stated in Theorem 3.2, we finally obtain $I_{4,2} = o(n^{-2s/(1+2s)})$. Combining this with $I_{4,1}$, we complete the proof of Lemma 5.5.

For the term I_3, the proof can be derived in the same spirit as in I_4. We can treat $I_{3,1}$ and $I_{3,2}$ separately. The proof of $I_{3,1}$ is the same as that in Lemma 5.7, while the proof for $I_{3,2}$ can be derived by Besov space embedding properties. Since the proof is similar to that in [13, p. 940–941] and the step by step details are long, we omit them here. $\square$

Acknowledgments

The author thanks two referees and one associate editor for their careful reading of an earlier version of the manuscript and for their valuable suggestions and insightful comments. This greatly improved the presentation of the paper. The author also wishes to thank Professors Jianqing Fan, Iréne Gijbels and Brenda MacGibbon for making their program code used for the local linear regression smoother available for this research. Finally, the author expresses his deep gratitude to Christopher Valenta for his contribution on simulation study. This research was supported in part by the NSF grant DMS-0604499.

Appendix

In order to obtain the bounds on the terms I_{23}, I_{32} and I_{34} in Lemmas 5.6 and 5.7, we need the following theorem from Talagrand (1994) as stated in [13, p. 937].

Theorem A.1. *Let $U_1, U_2, \ldots, U_n$ be independent and identically distributed random variables. Let $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ be independent Rademacher random variables that are also independent of the U_i. Let $\mathcal{F}$ be a class of functions uniformly bounded by M. If there exists v, $H > 0$ such that for all n,*

$$\sup_{h\in\mathcal{F}} \operatorname{var} h(U) \le v, \quad E \sup_{h\in\mathcal{F}} \sum_{m=1}^{n} \epsilon_m h(U_m) \le nH,$$

then there exist universal constants C_1 and C_2 such that for all $\lambda > 0$,

$$P\left\{\sup_{h\in\mathcal{F}}\left[\frac{1}{n}\sum_{m=1}^{n} h(U_m) - Eh(U)\right] \ge \lambda + C_2 H\right\} \le \mathrm{e}^{-nC_1\left[\frac{\lambda^2}{v}\wedge\frac{\lambda}{M}\right]}.$$

The following lemma will also be needed in the sequel. For the proof, see [13, p. 939].

Lemma A.1. *If $\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x \le 2^{-1}lC_0n^{-1}$, then*

$$\left\{\int_{J_{ik}}(\hat{D}_{ik}(x))^2\mathrm{d}x \ge lC_0n^{-1}\right\} \subseteq \left\{\int_{J_{ik}}(\hat{D}_{ik}(x) - D_{ik}f(x))^2\mathrm{d}x \ge 0.08lC_0n^{-1}\right\}.$$

If $\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x > 2lC_0n^{-1}$, then

$$\left\{\int_{J_{ik}}(\hat{D}_{ik}(x))^2\mathrm{d}x \le lC_0n^{-1}\right\} \subseteq \left\{\int_{J_{ik}}(\hat{D}_{ik}(x) - D_{ik}f(x))^2\mathrm{d}x \ge 0.16lC_0n^{-1}\right\}.$$

Lemma A.2. *Suppose that the assumptions of Theorem 3.1 hold. Then*

$$I_{23} = E\sum_k \int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x\, I(\hat{B}_{ik} \le C_0n^{-1})I(B_{ik} > 2C_0n^{-1}) = O\left(n^{-1}\right).$$

Proof. From Lemma A.1, we have

$$I_{23} \le E\sum_k \int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x \cdot I\left(\int_{J_{ik}}\left(\hat{D}_{ik}(x) - D_{ik}f(x)\right)^2 \mathrm{d}x \ge 0.16lC_0n^{-1}\right).$$

From Lemma 5.1, we may write

$$\begin{aligned}\hat{D}_{ik}(x) &= \sum_{j\in B(k)} \hat{\beta}_{ij}\psi_{ij}(x)\\ &= \sum_{j\in B(k)} (\bar{\beta}_{ij} + \overline{W}_{ij})\psi_{ij}(x) + \sum_{j\in B(k)} R_{n,ij}\psi_{ij}(x)\\ &=: \overline{D}_{ik}(x) + R_{n,ik}(x),\end{aligned}$$

where

$$\begin{aligned}
\overline{D}_{ik}(x) = \ & \frac{1}{n}\sum_{m=1}^{n}\Bigg[\frac{\delta_m I(Z_m \le T)}{1-G(Z_m)}\sum_{j\in B(k)}\psi_{ij}(Z_m)\psi_{ij}(x) \\
& + \frac{1-\delta_m}{1-H(Z_m)}\int_{Z_m}^{\tau_H}\sum_{j\in B(k)}\psi_{ij}(w)\psi_{ij}(x)I(w\le T)\,F(\mathrm{d}w) \\
& - \int_{-L}^{\tau_H}\int_{-L}^{\tau_H}\frac{I(v<Z_m\wedge w)}{[1-H(v)][1-G(v)]}\sum_{j\in B(k)}\psi_{ij}(w)\psi_{ij}(x)\,G(\mathrm{d}v)F(\mathrm{d}w)\Bigg] \\
=: \ & \frac{1}{n}\sum_{m=1}^{n}\Bigg[\frac{\delta_m I(Z_m\le T)}{1-G(Z_m)}D_{ik}(x,Z_m) \\
& + \frac{1-\delta_m}{1-H(Z_m)}\int_{Z_m}^{\tau_H}D_{ik}(x,w)I(w\le T)\,F(\mathrm{d}w) \\
& - \int_{-L}^{\tau_H}\int_{-L}^{\tau_H}\frac{I(v<Z_m\wedge w)}{[1-H(v)][1-G(v)]}D_{ik}(x,w)\,G(\mathrm{d}v)F(\mathrm{d}w)\Bigg] \\
=: \ & \frac{1}{n}\sum_{m=1}^{n}T_{ik}(x,Z_m).
\end{aligned}$$

The second equality defines $D_{ik}(x,y) = \sum_{j\in B(k)}\psi_{ij}(x)\psi_{ij}(y)$ and the last equality defines $T_{ik}(x,Z_m)$.

From Lemma 5.1 and applying the triangle inequality, we have

$$\begin{aligned}
I_{23} \le \ & E\sum_k\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x\cdot I\left(\int_{J_{ik}}\left(\overline{D}_{ik}(x)-D_{ik}f(x)\right)^2\,\mathrm{d}x \ge 0.08\alpha l C_0 n^{-1}\right) \\
& + E\sum_k\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x\cdot I\left(\int_{J_{ik}}\left(R_{n,ik}(x)\right)^2\,\mathrm{d}x \ge 0.08\beta l C_0 n^{-1}\right) \\
=: \ & I_{231}+I_{232},
\end{aligned}$$

where α and β are any positive numbers such that $\alpha+\beta=1$.

Let's consider the term I_{232} first. From the orthogonality of ψ and applying the Markov inequality, we have

$$\begin{aligned}
I_{232} & \le \sum_k\int_{J_{ik}}(D_{ik}f(x))^2\mathrm{d}x\cdot\sum_{j\in B(k)}E(R_{n,ij}^2)(0.08\beta l C_0 n^{-1})^{-1} \\
& \le \sum_k\sum_{j\in B(k)}\beta_{ij}^2\cdot C(ln)^{-1}\sum_{j\in B(k)}\int\psi_{ij}^2\mathrm{d}F \\
& \le Cn^{-1}\sum_j\beta_{ij}^2,
\end{aligned}$$

where the second inequality follows from Lemma 5.1 and the last inequality follows from there being l terms in the sums $\sum_{j\in B(k)}$ and $\int\psi_{ij}^2\mathrm{d}F \le \|f\|_\infty$ for all i and j. As in Lemma 5.5, $\sum_j\beta_{ij}^2 \le M^2 2^{-2i\sigma}$ when $p<2$ and $\sum_j\beta_{ij}^2\le C2^{-is}$ when $p\ge 2$. Hence $I_{232}\le Cn^{-1}$ for any value of p.

As to the first term I_{231}, we can apply the above Talagrand Theorem A.1. Write

$$\begin{aligned}
&\left\{\int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \,\mathrm{d}x \geq 0.08\alpha l C_0 n^{-1}\right\} \\
&= \left\{\left[\int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \,\mathrm{d}x\right]^{1/2} \geq \sqrt{0.08\alpha l C_0 n^{-1}}\right\} \\
&= \left\{\sup_{\|g\|_2 \leq 1} \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right) g(x) \,\mathrm{d}x \geq \sqrt{0.08\alpha l C_0 n^{-1}}\right\} \\
&= \left\{\sup_{\|g\|_2 \leq 1} \left[\frac{1}{n}\sum_{m=1}^{n} \int_{J_{ik}} T_{ik}(x, Z_m) g(x) \,\mathrm{d}x - E\int_{J_{ik}} T_{ik}(x, Z_m) g(x) \mathrm{d}x\right]\right. \\
&\left.\geq \sqrt{0.08\alpha l C_0 n^{-1}}\right\} \\
&=: \left\{\sup_{h\in\mathcal{F}} \left[\frac{1}{n}\sum_{m=1}^{n} h(Z_m) - Eh(Z)\right] \geq \sqrt{0.08\alpha l C_0 n^{-1}}\right\},
\end{aligned}$$

where $\mathcal{F} = \{\int_{J_{ik}} T_{ik}(x, \cdot) I(j \in B(k)) g(x) \mathrm{d}x : \|g\|_2 \leq 1\}$ and the third equality follows from $ET_{ik}(x, Z) = D_{ik} f(x)$.

In order to apply the Talagrand Theorem, we need to compute these constants $M,\ v,\ H$ and λ.

$$M = \sup_y \left|\int_{J_{ik}} g(x) T_{ik}(x, y) \mathrm{d}x\right| \leq \sup_y \|g\|_2^2 \left\{\int_{J_{ik}} T_{ik}^2(x, y) \mathrm{d}x\right\}^{1/2}.$$

Recall that

$$\begin{aligned}
T_{ik}(x, Z_m) = {} & \frac{\delta_m I(Z_m \leq T)}{1 - G(Z_m)} D_{ik}(x, Z_m) \\
& + \frac{1-\delta_m}{1 - H(Z_m)} \int_{Z_m}^{\tau_H} D_{ik}(x, w) I(w \leq T)\, F(\mathrm{d}w) \\
& - \int_{-L}^{\tau_H}\int_{-L}^{\tau_H} \frac{I(v < Z_m \wedge w)}{[1 - H(v)][1 - G(v)]} D_{ik}(x, w)\, G(\mathrm{d}v) F(\mathrm{d}w). \\
=: {} & T_{ik,1}(x, Z_m) + T_{ik,2}(x, Z_m) + T_{ik,3}(x, Z_m).
\end{aligned}$$

Thus

$$M \leq \sup_y \left\{3\int_{J_{ik}} T_{ik,1}^2(x, y) \mathrm{d}x + 3\int_{J_{ik}} T_{ik,2}^2(x, y) \mathrm{d}x + 3\int_{J_{ik}} T_{ik,3}^2(x, y) \mathrm{d}x\right\}^{1/2}.$$

Since

$$\begin{aligned}
&\int_{J_{ik}} T_{ik,1}^2(x, y) \mathrm{d}x \leq \frac{1}{[1 - G(T)]^2} \int_{J_{ik}} D_{ik}^2(x, y) \mathrm{d}x \\
&\leq \frac{1}{[1 - G(T)]^2} \int_{J_{ik}} 2^{2i} Q^2(2^i(x - y)) \mathrm{d}x,
\end{aligned}$$

we have

$$\sup_y \int_{J_{ik}} T_{ik,1}^2(x,y)\mathrm{d}x \le \frac{2^i \|Q\|_2^2}{[1-G(T)]^2}.$$

Similarly, we can have

$$\sup_y \int_{J_{ik}} T_{ik,2}^2(x,y)\mathrm{d}x \le \frac{2^i \|Q\|_2^2}{[1-H(T)]^2},$$

$$\sup_y \int_{J_{ik}} T_{ik,3}^2(x,y)\mathrm{d}x \le \frac{2^i \|Q\|_2^2}{[1-G(T)]^2[1-H(T)]^2}.$$

Thus, we have

$$\begin{aligned} M &\le \left\{\frac{3}{[1-G(T)]^2} + \frac{3}{[1-H(T)]^2} + \frac{3}{[1-G(T)]^2[1-H(T)]^2}\right\}^{1/2} 2^{i/2}\|Q\|_2 \\ &\le \frac{3}{[1-G(T)][1-H(T)]} 2^{i/2}\|Q\|_2. \end{aligned}$$

Through similar direct calculation, we can obtain that

$$\begin{aligned} v &\le \sup_{\|g\|_2\le 1} E\left\{\int_{J_{ik}} T_{ik}(x,Z)g(x)\mathrm{d}x\right\}^2 \\ &\le 3\sup_{\|g\|_2\le 1} E\left\{\int_{J_{ik}} T_{ik,1}(x,Z)g(x)\mathrm{d}x\right\}^2 + 3\sup_{\|g\|_2\le 1} E\left\{\int_{J_{ik}} T_{ik,2}(x,Z)g(x)\mathrm{d}x\right\}^2 \\ &\quad + 3\sup_{\|g\|_2\le 1} E\left\{\int_{J_{ik}} T_{ik,3}(x,Z)g(x)\mathrm{d}x\right\}^2. \end{aligned}$$

As to the first term, we have

$$\begin{aligned} E\left\{\int_{J_{ik}} T_{ik,1}(t,Z)g(t)\mathrm{d}t\right\}^2 &= \iint_{x<y} \left\{\int_{J_{ik}} \frac{I(x\le T)}{1-G(x)} D_{ik}(t,x)g(t)\mathrm{d}t\right\}^2 \mathrm{d}F(x)\mathrm{d}G(y) \\ &\le \int \left\{\int_{J_{ik}} D_{ik}(t,x)g(t)\mathrm{d}t\right\}^2 \frac{I(x\le T)}{1-G(x)}\mathrm{d}F(x) \\ &\le \frac{1}{1-G(T)} \int \left\{\int_{J_{ik}} D_{ik}(t,x)g(t)\mathrm{d}t\right\}^2 f(x)\mathrm{d}x. \end{aligned}$$

From the exact same argument as in [13, p. 938], we can get

$$\sup_{\|g\|_2\le 1} E\left\{\int_{J_{ik}} T_{ik,1}(x,Z)g(x)\mathrm{d}x\right\}^2 \le \frac{1}{1-G(T)}\|f\|_\infty \|Q\|_1^2.$$

Similarly, we can obtain

$$\sup_{\|g\|_2\le 1} E\left\{\int_{J_{ik}} T_{ik,2}(x,Z)g(x)\mathrm{d}x\right\}^2 \le \frac{1}{[1-G(T)][1-H(T)]}\|f\|_\infty \|Q\|_1^2,$$

$$\sup_{\|g\|_2\le 1} E\left\{\int_{J_{ik}} T_{ik,3}(x,Z)g(x)\mathrm{d}x\right\}^2 \le \frac{1}{[1-G(T)]^2[1-H(T)]^2}\|f\|_\infty \|Q\|_1^2.$$

Combining above terms together, we finally obtain

$$v \leq \frac{9}{[1-G(T)]^2[1-H(T)]^2} \|f\|_\infty \|Q\|_1^2.$$

At last, from [13, p. 939], we have

$$\begin{aligned} nH &= E\left\{\sup_{\|g\|_2 \leq 1} \sum_{m=1}^{n} \int_{J_{ik}} T_{ik}(x, Z_m) g(x) \mathrm{d}x \cdot \varepsilon_m \right\} \\ &\leq n^{1/2} \left\{ \int_{J_{ik}} E\, T_{ik}^2(x, Z) \mathrm{d}x \right\}^{1/2}. \end{aligned}$$

Through direct calculation, we have

$$\begin{aligned} ET_{ik,1}^2(t, Z) &= \iint_{x<y} \frac{I(x \leq T)}{[1-G(x)]^2} D_{ik}^2(t, x) \mathrm{d}F(x) \mathrm{d}G(y) \\ &\leq \frac{1}{1-G(T)} \|f\|_\infty 2^i \|Q\|_2^2, \end{aligned}$$

for all the values of t. Since the length of J_{ik} is $l2^{-i}$, we have

$$\int_{J_{ik}} E\, T_{ik,1}^2(x, Z) \mathrm{d}x \leq \frac{1}{1-G(T)} l \|f\|_\infty \|Q\|_2^2.$$

Like in the above calculation, we obtain

$$\begin{aligned} \int_{J_{ik}} E\, T_{ik,2}^2(x, Z) \mathrm{d}x &\leq \frac{1}{[1-H(T)]^2} l \|f\|_\infty \|Q\|_2^2, \\ \int_{J_{ik}} E\, T_{ik,3}^2(x, Z) \mathrm{d}x &\leq \frac{1}{[1-G(T)]^2[1-H(T)]^2} l \|f\|_\infty \|Q\|_2^2. \end{aligned}$$

Thus, we obtain

$$nH \leq \frac{3}{[1-G(T)][1-H(T)]} \left(nl \|f\|_\infty \|Q\|_2^2 \right)^{1/2}.$$

Now applying Theorem A.1 with the above constants M, v, H and

$$\lambda = \sqrt{0.08\alpha C_0 l n^{-1}} - 3C_2[1-H(T)]^{-1}[1-G(T)]^{-1} \sqrt{\|f\|_\infty \|Q\|_2^2 l n^{-1}},$$

we have

$$P\left(\int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x) \right)^2 \geq 0.08\alpha l C_0 n^{-1} \right) \leq \exp\left\{ -nC_1 \left(\frac{\lambda^2}{v} \wedge \frac{\lambda}{M} \right) \right\}.$$

On the basis of our choice of R such that $2^R \simeq n(\log n)^{-2}$, we have $\lambda^2 v^{-1} \leq \lambda M^{-1}$. Thus we have

$$\begin{aligned} &P\left(\int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x) \right)^2 \geq 0.08\alpha l C_0 n^{-1} \right) \\ &\quad \leq \exp\left\{ -nC_1 \left(\frac{\lambda^2}{9[1-H(T)]^{-2}[1-G(T)]^{-2} \|f\|_\infty \|Q\|_1^2} \right) \right\}. \end{aligned}$$

In order to make

$$P\left(\int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \geq 0.08\alpha l C_0 n^{-1}\right) \leq n^{-1}, \tag{A.1}$$

through direct calculation we only need that the threshold constant C_0 in λ satisfies

$$C_0 \geq \frac{9\|f\|_\infty}{0.08\alpha[1-H(T)]^2[1-G(T)]^2}(C_2\|Q\|_2)^2.$$

But, based on our choice C_0 as in Theorem 3.1, let α be close to 1; C_0 is greater than the above constant.

Thus, $I_{231} \leq n^{-1}\sum_j \beta_{ij}^2 \leq Cn^{-1}$. Combining another term, I_{232}, we prove the lemma. □

Lemma A.3. *Suppose that the assumptions of Theorem 3.1 hold. Then*

$$\begin{aligned} I_{32} &= E\sum_k \int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x\, I(\hat{B}_{ik} > C_0 n^{-1}) I(B_{ik} \leq C_0(2n)^{-1}) \\ &= O\left(n^{-1}\right). \end{aligned}$$

Proof. Applying the triangle inequality and Lemma 5.1, we have

$$\begin{aligned} I_{32} \leq\ & 2E\sum_k \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x\, I(\hat{B}_{ik} > C_0 n^{-1}) I(B_{ik} \leq C_0(2n)^{-1}) \\ & + 2E\sum_k \int_{J_{ik}} R_{n,ik}^2(x)\mathrm{d}x \\ =:\ & 2(I_{321} + I_{322}). \end{aligned}$$

As to the second term I_{322}, using Lemma 5.1, we have

$$I_{322} \leq E\sum_k \sum_{j\in B(k)} R_{n,ij}^2 = O\left(\frac{1}{n^2}\right)\sum_j \int \psi_{ij}^2 \mathrm{d}F \leq \frac{C}{n^2} 2^i \|f\|_\infty \leq Cn^{-1},$$

where the last two inequalities follow from there being at most $C2^i$ terms in the sum $\sum_j$ and $2^i \leq n$ for all i. As to the first term, write

$$\begin{aligned} I_{321} \leq\ & E\sum_k \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x \cdot I \\ & \times \left(\int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x \geq 0.16 l C_0 n^{-1}\right) \\ \leq\ & E\sum_k \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x \cdot I \\ & \times \left(\int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x \geq 0.08 l C_0 n^{-1}\right) \\ & + E\sum_k \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x \cdot I\left(\int_{J_{ik}} R_{n,ik}^2(x)\mathrm{d}x \geq 0.08 l C_0 n^{-1}\right) \\ =:\ & I_{3211} + I_{3212}. \end{aligned}$$

As to the first term I_{3211}, we need the following identity:

$$E[Y^2 I(Y > a)] = a^2 P(Y > a) + \int_a^\infty 2yP(Y > y)\mathrm{d}y,$$

with $Y^2 = \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik} f(x)\right)^2 \mathrm{d}x$ and $a = \sqrt{0.08lC_0 n^{-1}}$. Therefore

$$I_{3211} = \sum_k a^2 P(Y > a) + \sum_k \int_a^\infty 2yP(Y > y)\mathrm{d}y =: L_1 + L_2. \tag{A.2}$$

From (A.1), we have

$$L_1 \le \sum_k (0.08)C_0 l n^{-1} n^{-1} \le \sum_k Cln^{-2} \le Cn^{-1},$$

and the last inequality follows from there being at most $Cl^{-1}2^i$ terms in the sum $\sum_k$ and $2^i \le 2^R \le n$ for all i.

As to the probability $P(Y > y)$ in the term L_2, applying the Talagrand Theorem A.1 with $\lambda = y - C_2 H$, we have

$$\begin{aligned} P(Y > y) &= P\left(Y > (y - C_2H) + C_2H\right) \\ &\le \exp\left\{-nC_1\left(\frac{(y - C_2H)^2}{9[1 - H(T)]^{-2}[1 - G(T)]^{-2}\|f\|_\infty \|Q\|_1^2} \wedge \right.\right. \\ &\qquad \left.\left. \wedge \frac{y - C_2H}{3[1 - H(T)]^{-1}[1 - G(T)]^{-1}2^{i/2}\|Q\|_2}\right)\right\}. \end{aligned}$$

Let y_0 satisfy

$$\frac{(y - C_2H)^2}{9[1 - H(T)]^{-2}[1 - G(T)]^{-2}\|f\|_\infty \|Q\|_1^2} = \frac{y - C_2H}{3[1 - H(T)]^{-1}[1 - G(T)]^{-1}2^{i/2}\|Q\|_2},$$

which results in $y_0 = \frac{3}{[1-H(T)][1-G(T)]} \frac{\|f\|_\infty \|Q\|_1^2}{2^{i/2}\|Q\|_2} - C_2H > 0$. Therefore, we have

$$\begin{aligned} \int_a^\infty 2yP(Y > y)\mathrm{d}y &= \int_a^{y_0} 2y \exp\left\{-nC_3(y - C_2H)^2\right\} \mathrm{d}y \\ &\quad + \int_{y_0}^\infty 2y \exp\left\{-nC_4(y - C_2H)\right\} \mathrm{d}y \\ &=: L_{21} + L_{22}, \end{aligned}$$

where $C_3 = 9^{-1}C_1^{-1}[1 - H(T)]^2[1 - G(T)]^2\|f\|_\infty^{-1}\|Q\|_1^{-2}$ and $C_4 = 3^{-1}C_1^{-1}[1 - H(T)][1 - G(T)]2^{-i/2}\|Q\|_2^{-1}$. Through the change of variable $u = y - C_2H$, we have

$$L_{21} \le \int_{a-C_2H}^\infty 2u\mathrm{e}^{-C_3nu^2}\mathrm{d}u + 2C_2H \int_{a-C_2H}^\infty \mathrm{e}^{-C_3nu^2}\mathrm{d}u =: L_{211} + L_{212}.$$

Through direct calculation, we get $L_{211} \le (C_3n)^{-1}\mathrm{e}^{-C_3n\sqrt{a-C_2H}}$. In order to make $L_{211} \le Cn^{-2}$, it suffices to require $\mathrm{e}^{-C_3n\sqrt{a-C_2H}} \le n^{-1}$, which is equivalent to

$$C_0 \ge \frac{9\|f\|_\infty}{0.08[1 - H(T)]^2[1 - G(T)]^2}(C_2\|Q\|_2 + \|Q\|_1 C_1^{-1/2})^2,$$

which is satisfied from our choice of C_0 as in Theorem 3.1.

Apply inequality $\int_x^\infty e^{-t^2/2}dt \le x^{-1}e^{-x^2/2}$; through simple algebra, we have

$$L_{212} = \frac{2C_2H}{\sqrt{2C_3n}(a - C_2H)} e^{-C_3n(a-C_2H)^2}.$$

In order to make $L_{212} \le Cn^{-2}$, we only need $e^{-C_3n(a-C_2H)^2} \le Cn^{-2}$, which is equivalent to asking that

$$C_0 \ge \frac{9\|f\|_\infty}{0.08[1 - H(T)]^2[1 - G(T)]^2}(C_2\|Q\|_2 + \sqrt{2}\|Q\|_1 C_1^{-1/2})^2.$$

Thus, we have $L_{21} \le Cn^{-2}$.

Let's turn back to the second term L_{22}. Through integration by parts, we have

$$\begin{aligned} L_{22} &= \frac{2y_0}{C_4n2^{i/2}} e^{-C_4n2^{-i/2}(y_0-C_2H)} + \frac{2\,2^{i/2}}{C_4n}\int_{y_0}^\infty e^{-C_4n2^{-i/2}(y-C_2H)}\,dy \\ &=: L_{221} + L_{222}. \end{aligned}$$

Through direct calculation, we can obtain that $L_{221} \le C2^{-i}n^{-1}$ and $L_{222} \le C2^i n^{-3}$. Thus

$$L_{22} \le \frac{C}{n2^i} + \frac{C2^i}{n^3}.$$

Now, combining this with the term $L_{21} \le Cn^{-2}$ and noticing that there are at most $Cl^{-1}2^i$ terms in the sum $\sum_k$ and $2^i \le 2^R \le n$ for all i, we have $L_2 \le \sum_k L_{21} + \sum_k L_{22} \le Cn^{-1}$ for all i. Combining this with $L_1 \le Cn^{-1}$, we prove $I_{3211} \le Cn^{-1}$.

In order to prove the lemma, we only need to show that $I_{3212} \le Cn^{-1}$ also. Let $E = \{\int_{J_{ik}}(\overline{D}_{ik}(x) - D_{ik}f(x))^2dx \ge 0.08lC_0n^{-1}\}$; thus

$$\begin{aligned} I_{3212} &= E\sum_k \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik}f(x)\right)^2 dx \cdot I\left(\int_{J_{ik}} R_{n,ik}^2(x)dx \ge 0.08lC_0n^{-1}\right) I_{E^c} \\ &\quad + E\sum_k \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik}f(x)\right)^2 dx \cdot I\left(\int_{J_{ik}} R_{n,ik}^2(x)dx \ge 0.08lC_0n^{-1}\right) I_E \\ &\le \sum_k \left(0.08lC_0n^{-1}\right) P\left(\int_{J_{ik}} R_{n,ik}^2(x)dx \ge 0.08lC_0n^{-1}\right) \\ &\quad + E\sum_k \int_{J_{ik}} \left(\overline{D}_{ik}(x) - D_{ik}f(x)\right)^2 dx \cdot I_E \\ &= \sum_k \sum_{j\in B(k)} ER_{n,ij}^2 + I_{3211}. \end{aligned}$$

However, from the same argument as for term I_{322}, we that see $\sum_k \sum_{j\in B(k)} ER_{n,ij}^2 \le Cn^{-1}$. Combining this with the term $I_{3211} \le Cn^{-1}$, we complete the proof. □

Lemma A.4. *Suppose that the assumptions of Theorem 3.1 hold. Then*

$$I_{34} = E\sum_k \int_{J_{ik}} (D_{ik}f(x))^2\,dx\, I(\hat{B}_{ik} \le C_0n^{-1})I(B_{ik} > 2C_0n^{-1}) = O\left(n^{-1}\right).$$

Proof. Compare the term I_{23} in Lemma A.2 with the above term I_{34}; we notice that the only difference between them is the different ranges of i. Since the proof of Lemma A.2 holds for all $i \leq 2^R$, we have $I_{34} = O\left(n^{-1}\right)$ also. $\square$

References

[1] A. Antoniadis, G. Grégoire, G. Nason, Density and hazard rate estimation for right-censored data by using wavelet methods, J. Roy. Statist. Soc. B 61 (1999) 63–84.
[2] T. Cai, Adaptive wavelet estimation: A block thresholding and oracle inequality approach, Ann. Statist. 27 (1999) 898–924.
[3] T. Cai, On block thresholding in wavelet regression: Adaptivity, block size, and threshold level, Statistic Sinica 12 (2002) 1241–1273.
[4] E. Chicken, T. Cai, Block thresholding for density estimation: Local and global adaptivity, J. Multivariate Anal. 95 (2005) 76–106.
[5] A. Cohen, I. Daubechies, P. Vial, Wavelets on the interval and fast wavelet transforms, Appl. Comput. Harm. Anal. 1 (1993) 54–82.
[6] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.
[7] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1994) 425–455.
[8] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinking, J. Am. Statist. Assoc. 90 (1995) 1200–1224.
[9] D.L. Donoho, I.M. Johnstone, Minimax estimation via wavelet shrinkage, Ann. Statist. 26 (1998) 879–921.
[10] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Wavelet shrinkage: asymptopia? (with discussion), J. Roy. Statist. Soc. B 57 (1995) 301–369.
[11] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, Ann. Statist. 24 (1996) 508–539.
[12] J. Fan, I. Gijbels, Censored regression: Local linear approximations and their applications, J. Am. Statist. Assoc. 89 (1994) 560–570.
[13] P. Hall, G. Kerkyacharian, D. Picard, Block threshold rules for curve estimation using kernel and wavelet method, Ann. Statist. 26 (1998) 922–942.
[14] P. Hall, G. Kerkyacharian, D. Picard, On the minimax optimality of block thresholded wavelet estimators, Statistica Sinica 9 (1999) 33–50.
[15] W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, Wavelets, Approximation and Statistical Applications, in: Lecture Notes in Statistics, vol. 129, 1998.
[16] L. Li, Non-linear wavelet-based density estimators under random censorship, J. Statist. Planning Inference 117 (2003) 35–58.
[17] L. Li, On the minimax optimality of wavelet estimators with censored data, J. Statist. Planning Inference 137 (2007) 1138–1150.
[18] J.S. Marron, W.J. Padgett, Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples, Ann. Statist. 15 (1987) 1520–1535.
[19] Y. Meyer, Wavelets and Operators, Cambridge University Press, Cambridge, 1992.
[20] G.P. Nason, Wavelet shrinkage using cross-validation, J. Roy. Statist. Soc. B 58 (1996) 463–479.
[21] W. Stute, The central limit theorem under random censorship, Ann. Statist. 23 (1995) 422–439.
[22] H. Triebel, Theory of Function Spaces II, Birkhäuser, Basel, 1992.
[23] B. Vidakovic, Statistical Modeling by Wavelets, John Wiley & Sons, Inc., 1999.
[24] M. Talagrand, Sharper bounds for Gaussian and empirical processes, Ann. Probab. 22 (1) (1994) 28–76.