
Volume 41, Issue 2, June 2012
ISSN 1226-3192
ELSEVIER
JOURNAL OF
THE KOREAN
STATISTICAL
SOCIETY
Available online at www.sciencedirect.com
SciVerse ScienceDirect
THE KOREAN STATISTICAL SOCIETY founded 1971

Journal of the Korean Statistical Society 41 (2012) 199–211

Contents lists available at SciVerse ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Wavelet based estimation for the derivative of a density by block thresholding under random censorship

Esmaeil Shirazi[a,*], Yogendra P. Chaubey[b], Hassan Doosti[c], Hossein Ali Nirumand[a]

[a] Department of Statistics, Ferdowsi University, Mashhad, 91755-1159, Iran
[b] Department of Mathematics and Statistics, Concordia University, Montréal, QC H3G 1M8, Canada
[c] Department of Mathematics, Tarbiat Moallem University, Tehran, Iran

ARTICLE INFO

Article history:
Received 7 March 2011
Accepted 19 August 2011
Available online 15 September 2011

AMS 2000 subject classifications:
primary 62Gxx
62Nxx
secondary 62G05
62N02

Keywords:
Adaptive estimation
Block thresholding
Censored data
Nonparametric estimator of derivative of a density
Rates of convergence

ABSTRACT

We consider wavelet based method for estimating derivatives of a density via block thresholding when the data obtained are randomly right censored. The proposed method is analogous to that of Hall and Patil (1995) for density estimation in the complete data case that has been extended recently by Li (2003, 2008). We find bounds for the L_2-loss over a large range of Besov function classes for the resulting estimators. The results of Hall and Patil (1995), Prakasa Rao (1996) and Li (2003, 2008) are obtained as special cases and the performance of the proposed estimator is investigated by a numerical study.

1. Introduction

In various applications, estimation of a density and its derivatives may be required, for example, for the evaluation of modes and inflection points, for the estimation of the derivatives of a regression function and for the evaluation of the density scores. There are many excellent texts that are now available on the subject dealing with theoretical aspects and applications. In this context the reader may be referred to the texts by Devroye (1987), Silverman (1986) and Wand and Jones (1995), concentrating more on kernel methods (Parzen, 1962 and Rosenblatt, 1956) and its variants. For a more theoretical treatment of the general topic of nonparametric functional estimation, one may consult for Prakasa Rao (1983). There are two excellent texts originating from the discipline of econometrics that provide an excellent review of various nonparametric methods for the estimation of a density and its derivative (see Li & Racine, 2006 and Pagan & Ullah, 1999). In most of these applications the kernel method is more popular and as a result many software packages include this method as a member of their toolbox, for example, one may see Wand and Ripley (2009) for the details of the `KernSmooth` package in R that is becoming a powerful tool for statisticians lately.

* Corresponding author.
E-mail addresses: esm.shirazi@stu-mail.um.ac.ir, shirazi.esmaeel@gmail.com (E. Shirazi), chaubey@alcor.concordia.ca (Y.P. Chaubey), hassandoosti1353@yahoo.com (H. Doosti).

1226-3192/$ – see front matter © 2011 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jkss.2011.08.004

The estimation of derivatives of densities was featured in Bhattacharya (1967) in estimating Fisher's information from a sample of *i.i.d.* observations and this theme was carried out in Singh (1977) highlighting many other applications using the estimators of density derivatives. Härdle, Marron, and Wand (1990) discussed the issue of the band width choice, specifically for density derivatives.

Initially, after the advent of the kernel method most of the investigation were with complete data until the paper by McNichols and Padgett (1984) that presented a modified kernel estimator for the density based on the right censored data. Padgett and McNichols (1984) provided the status of density estimation for censored data until 1984. Further, Diehl and Stute (1988) considered hazard rate estimation using the kernel method. For a recent review on smooth functional estimation under censored data the reader may be referred to Chaubey, Sen, and Sen (2007).

This is the subject matter of the present paper where the data typically represent the survival times that are incomplete in some way due to the presence of a number of events which potentially censor the event of interest. Withdrawals from a clinical trial, deaths unrelated to the disease under study, individuals still alive at the end of a follow-up study, and so on, are examples of censoring issues.

Currently, wavelet based methods for the estimation of the density and its derivatives are becoming increasingly popular as an alternative to the usual kernel method. These methods offer fast computations and easy updating in addition to being easily adapted to the design (Delouille, Simoens, & von Sachs, 2001), and specific smoothness (Donoho & Johnstone, 1995). Donoho et al. [DJKP] (1995) proposed wavelet based methods for density estimation for the *i.i.d.* data which were subsequently adapted to the estimation of derivatives by Prakasa Rao (1996). Hall and Patil (1995) used a modification of DJKP estimator adapting to a specific smoothness parameter and obtained an asymptotic formula for the mean integrated squared error of the corresponding density estimator. Further work in this context may be referred, specifically for density estimation by Antoniadis, Grégoire, and Nason (1999), Cai (1999), Hall, Kerkyacharian, and Picard (1998, 1999) and Li (2003) and for nonparametric regression function estimation by Cai (2002) and Dabrowska (1995). Recently, Li (2002, 2003) has extended Hall and Patil's (1995) analysis to hazard rate and density estimation under the right random censorship model, respectively. We may mention here that Patil (1997) also considered the estimation of hazard rate based on orthogonal wavelets, however, under uncensored data. In this paper, we follow a similar plan as in Li (2003, 2008) for analyzing the block thresholding wavelet estimator for the d-th ($d \geq 0$) derivative of a density using censored data. Our analysis parallels to that of Li (2008) dealing with the density estimation in providing an upper bound on L^2-loss for the resulting estimator. Naturally, for $d = 0$, we get the result in Li (2008) that optimal convergence rates for density estimation are achieved over a large range of Besov function classes.

The organization of the paper is as follows. In Section 2, we describe some preliminary notions about the wavelet system and Besov spaces. In Section 3, the form of block thresholding wavelet estimator for the d-th derivative of a density is given along with the main result that gives an upper bound on the L_2-loss for the estimator. In Section 4, we study the performance of various wavelet threshold estimators. Details of the steps in the sketch of the proof are relegated to Appendix.

2. Preliminaries on wavelets and Besov spaces

Here we provide a brief introduction to the wavelet system and Besov spaces that have become essential to the statistical literature. For the details of the theory and applications of wavelets, the reader may refer to the excellent text by Vidakovic (1999) or to the excellent survey by Antoniadis (2007). For the properties of the Besov spaces, the reader is referred to Meyer (1992) and Tribel (1992) (cf. Härdle, Kerkycharian, & Tsybakov, 1998; Leblanc, 1996).

2.1. Wavelet system

A wavelet system is composed of an infinite collection of functions that are obtained by dilation and translation of two basic functions $\phi(.)$ and $\psi(.)$ called the *scaling function* and *mother wavelet*, respectively. The function ϕ is assumed to satisfy

$$\int_{-\infty}^{\infty} \phi(x) \mathrm{d}x = 1 \tag{2.1}$$

and is obtained as the solution from the equation

$$\phi(x) = \sum_{k=-\infty}^{\infty} C_k \phi(2x - k), \tag{2.2}$$

for a given sequence of constants $\{C_k\}$, and the function $\psi(x)$ is given by

$$\psi(x) = \sum_{k=-\infty}^{\infty} (-1)^k C_{-k+1} \phi(2x - k). \tag{2.3}$$

Define

$$\phi_{i,j}(x) = 2^{i/2} \phi(2^i x - j), \quad -\infty < i, j < \infty \tag{2.4}$$

and

$$\psi_{i,j}(x) = 2^{j/2}\phi(2^i x - j), \quad -\infty < i, j < \infty. \tag{2.5}$$

Suppose that the coefficients $\{C_k\}$ satisfy

$$\sum_{k=-\infty}^{\infty} C_k C_{k+2l} = 2 \quad \text{if } l = 0 \tag{2.6}$$

$$= 0 \quad \text{if } l \neq 0. \tag{2.7}$$

It is known (cf. Daubechies, 1992) that under some additional conditions on $\phi(.)$, the collection $\{\psi_{i,j}; -\infty < j < \infty, -\infty < k < \infty\}$ forms an orthonormal basis for $L^2(\mathbb{R})$, and $\{\phi_{i_0,j}; -\infty < j < \infty\}$ constitute an orthonormal basis for $L^2(\mathbb{R})$, for every fixed $i_0 \in \mathbb{Z}$, as well.

Definition 2.1. The scaling function ϕ is said to be r-regular for an integer $r \geq 1$, if for every nonnegative integer $\ell \leq r$, the ℓ-th derivative of $\phi(.)$, $\phi^\ell(.)$ is such that for any integer $k \geq 1$,

$$|\phi^\ell(x)| \leq c_k(1 + |x|)^{-k}, \quad -\infty < x < \infty \tag{2.8}$$

for some $c_k \geq 0$ depending only on k.

Definition 2.2. A *multiresolution analysis* of $L^2(\mathbb{R})$ consists of an increasing sequence of closed spaces $\{V_j\}$ of $L^2(\mathbb{R})$ such that

(i) $\bigcap_{j=-\infty}^{\infty} V_j = \{0\}$;
(ii) $\bigcup_{j=-\infty}^{\infty} V_j = L^2(\mathbb{R})$;
(iii) there is a scaling function $\phi \in V_0$ such that $\{\phi(x - k), -\infty < k < \infty\}$ forms an orthonormal basis for V_0;
(iv) for all $h(.) \in L^2(\mathbb{R})$, $-\infty < k < \infty$, $h(x) \in V_0 \Longrightarrow h(x - k) \in V_0$; and
(v) $h(x) \in V_j \Longrightarrow h(2x) \in V_{j+1}$.

Mallat (1989) has connected the multiresolution analysis to the wavelet theory by showing that given any multiresolution analysis, it is possible to construct a function $\psi(.)$, (called the mother wavelet), such that for any fixed j, $-\infty < j < \infty$, the family $\{\psi_{j,k}, j, -\infty < k < \infty\}$ constitutes a basis of the orthogonal complement W_j of V_j in V_{j+1} so that $\{\psi_{j,k}, -\infty < j, k < \infty\}$ is an orthonormal basis of $L^2(\mathbb{R})$ (cf. Daubechies, 1992). The corresponding multiresolution analysis is said to be r-regular if the scaling function $\phi(.)$ is so.

Suppose that both the functions ϕ and ψ belong to the space of functions with r continuous derivatives denoted by $\mathbf{C}^r$, for some $r \geq 1$, and have compact supports included in $[-L, L]$, for some $L > 0$. It follows, from Corollary 5.5.2 in Daubechies (1988), that the mother wavelet ψ is orthogonal to polynomials of degree $\leq r$, *i.e.*

$$\int \psi(x)x^l \mathrm{d}x = 0, \quad \forall l = 0, 1, \ldots, r.$$

Any function $f \in L^2(\mathbb{R})$ can be expanded in the form (cf. Daubechies, 1992):

$$f(x) = \sum_{j\in Z} a_{i_0,j}\phi_{i_0,j}(x) + \sum_{i\geq i_0}\sum_{j\in Z} b_{i,j}\psi_{i,j}(x)$$

$$= \mathcal{P}_{i_0}f(x) + \sum_{i\geq i_0} \mathcal{D}_i f(x) \tag{2.9}$$

for any integer $i_0 \in \mathbb{Z}$. The so-called wavelet coefficients $a_{i_0,j}$ and $b_{i,j}$ are given by

$$a_{i_0,j} = \int f(x)\phi_{i_0,j}(x)\mathrm{d}x \tag{2.10}$$

and

$$b_{i,j} = \int f(x)\psi_{i,j}(x)\mathrm{d}x \tag{2.11}$$

respectively.

For the later analysis, we choose $i_0 = 0$ and use the notation

$$a_j \equiv a_{i_0,j} \quad \text{and} \quad \phi_j \equiv \phi_{i_0,j}. \tag{2.12}$$

2.2. Besov spaces

Besov spaces are normed spaces defined for *weakly-differentiable* functions belonging to $L^2(\mathbb{R})$. We present the following definition of a weakly differentiable function f from Härdle et al. (1998).

Definition 2.3. Let $f \in L^2(\mathbb{R})$ be an integrable function on every bounded interval. It is said to be weakly differentiable if there exists a function g defined on the real line which is integrable on every bounded interval such that

$$\int_x^y g(u)\mathrm{d}u = f(y) - f(x).$$

The function g is defined almost everywhere and is called the *weak derivative* of f.

Then it is known that for any $\phi \in D(\mathbb{R})$,

$$\int_{-\infty}^{\infty} f(u)\phi'(u)\mathrm{d}u = -\int_{-\infty}^{\infty} g(u)\phi(u)\mathrm{d}u$$

where $D(\mathbb{R})$ denotes the space of infinitely differentiable functions, on the real line, with compact support.

Definition 2.4. Let $1 \le p \le \infty$ and $m \ge 0$ be an integer. A function $f \in L_p(\mathbb{R})$ belongs to the *Sobolev space* $W_p^m(\mathbb{R})$, if it is m-times weakly-differentiable and the m-th weak derivative $f^{(m)} \in L_P(\mathbb{R})$. The space $W_p^m(\mathbb{R})$ is equipped with the norm $\|f\|_{W_p^m}$, called the Sobolev norm, where $\|f\|_p$ denotes the norm for $L_p(\mathbb{R})$.

Let $f \in L_p(\mathbb{R})$ for some $1 \le p \le \infty$. Let $\Delta_h f(x) = f(x+h) - f(x)$ and define $\Delta_h^2 f = \Delta_h(\Delta_h f)$. For $t \ge 0$, let

$$\omega_p^1(f, t) = \sup_{|h|\le t} \|\Delta_h f\|_p$$

and

$$\omega_p^2(f, t) = \sup_{|h|\le t} \|\Delta_h^2 f\|_p.$$

Let $1 \le q \le \infty$. Suppose there exists a function $\epsilon(t)$ on $[0, \infty)$ such that $\|\epsilon\|_p^* < \infty$ where

$$\begin{aligned} \|\epsilon\|_q^* &= \left(\int_0^\infty t^{-1}|\epsilon(t)|^q \mathrm{d}t\right)^{1/q}, \quad \text{if } 1 \le q < \infty, \\ &= \operatorname{ess\,sup}_t |\epsilon(t)|, \quad \text{if } q = \infty. \end{aligned} \tag{2.13}$$

Definition 2.5. Let $1 \le p, q \le \infty$ and $s = m + \alpha$ where $m \ge 0$ is an integer and $0 < \alpha \le 1$. The *Besov space* $B_{p,q}^s$ is the space of all functions f such that $f \in W_p^m(\mathbb{R})$ and $\omega_p^2(f^{(m)}, t) = \epsilon(t)t^\alpha$ where $\|\epsilon\|_q^* < \infty$.

The norm on the Besov space is induced by the corresponding Sobolev space that can be written as

$$\|f\|_{B_{p,q}^s} = \|\mathcal{P}_0 f\|_p + \left(\sum_{i\ge 0} (\|\mathcal{D}_i f\|_p 2^{is})^q\right)^{1/q},$$

where

$$\mathcal{P}_0 f = \sum_{j\in Z} a_j \phi_j$$

represents the orthogonal projection of f on the space spanned by functions $\phi_j,\ j \ge 0$ and

$$\mathcal{D}_i f = \sum_{j\in Z} b_{i,j}\psi_{i,j}$$

represents that on the space spanned by functions $\psi_{i,j}, j \in \mathbb{Z}$. Suppose that f belongs to the Besov class (see Meyer, 1992, Section VI.10),

$$F_{p,q}^s(M, L) = \{f \in B_{p,q}^s, \|f\|_{B_{p,q}^s} \le M, \operatorname{supp} f \subset [-L, L]\}$$

for some $0 \le s \le r+1$, $p \ge 1$ and $q \ge 1$. Then in view of the representation (2.9), it can be shown (cf. Härdle et al., 1998) that the function say $f \in B_{p,q}^s$ if and only if

$$\|a_{i_0,.}\|_{\ell_p} < \infty, \quad \text{and} \quad \left(\sum_{i\ge i_0} (\|b_{i,.}\|_{\ell_p} 2^{i\sigma})^q\right)^{1/q} < \infty, \tag{2.14}$$

where $\sigma = s + (1/2) - (1/p)$, and $\|\gamma_j, .\|_{\ell_p}$ represents the following norm for a double sequence $\{\gamma_{j,k}\}$:

$$\|\gamma_{i,.}\|_{\ell_p} = \left(\sum_{k\in Z} \gamma_{i,k}^p\right)^{1/p}.$$

Let $\phi(.)$ be the scaling function as defined earlier. Define

$$\theta_\phi(x) = \sum_{k=-\infty}^{\infty} |\phi(x-k)|.$$

Suppose the following conditions hold:

(C1) The ess $\sup_x \theta_\phi(x) < \infty$ where

$$\text{ess}\sup_x g(x) = \inf\{y : \lambda(|x : g(x) > y) = 0\}$$

and λ is the Lebesgue measure on the real line.

(C2) There exists a bounded nondecreasing function $\Phi(.)$ such that $|\Phi(u)| \le \Phi(|u|)$ almost every where and

$$\int_0^\infty |u|^r \Phi(|u|)\mathrm{d}u < \infty$$

for some integer $r \ge 0$.

Then the Besov norm can be written (here we take $i_0 = 0$) (cf. Härdle et al., 1998, p. 123) in terms of the wavelet coefficients:

$$\|f\|_{B^s_{p,q}} = \|a_{0.}\|_p + \left(\sum_{j\ge 0}(2^{j\sigma}\|b_{j,.}\|_p)^q\right)^{1/q}. \tag{2.15}$$

3. Block thresholding estimator of the *d*-th derivative of a probability density function

Let $\{X_1, X_2, \ldots\}$ and $\{Y_1, Y_2, \ldots\}$ be two sequences of random variables. The first sequence may be regarded as representing survival times (or failure times), having a common unknown distribution function $F(.)$ and density function $f(.)$ and the second sequence as the censoring times as a result of random right censoring. The censoring times are assumed to be distributed according to a common distribution function $G(.)$. In this set up we may only observe

$$Z_i = \min(Y_i, X_i) := Y_i \wedge X_i \quad \text{and} \quad \delta_i = I(X_i \le Y_i), \tag{3.1}$$

where $I(.)$ denotes the indicator function. In this random censorship model, we assume that the survival times $\{X_i\}$ are independent of the censoring times $\{Y_i\}$. Following the convention in the survival analysis literature, we assume that both X_i and Y_i are nonnegative random variables. In contrast to statistics for complete data, that are based on the sequence $\{X_1, X_2, \ldots\}$, the estimation for the censored data depends on the pairs $(Z_1, \delta_1), (Z_2, \delta_2), \ldots, (Z_n, \delta_n)$. For example the Kaplan–Meier estimators of the distribution functions F and G are defined, respectively, by

$$\hat{F}_n(x) = 1 - \prod_{i=1}^n \left[1 - \frac{\delta_{(i)}}{n-i+1}\right]^{I(Z_{(i)} \le x)}, \tag{3.2}$$

$$\hat{G}_n(x) = 1 - \prod_{i=1}^n \left[1 - \frac{1-\delta_{(i)}}{n-i+1}\right]^{I(Z_{(i)} \le x)}, \tag{3.3}$$

where $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ denote the order statistics of $Z_1, Z_2, \ldots, Z_n$, and is the concomitant of $Z_{(i)}$, i.e., $\delta_{(m)} = \delta_k$ if $Z_{(m)} = Z_k$. Note that $\delta_k/n(1-\hat{G}(Z_m^-))$ is the jump of the Kaplan–Meier estimator $\hat{F}_n$ at Z_m.

We are interested in estimating $f^{(d)}$, the d-th ($d \ge 0$) derivative of f based on (Z_i, δ_i), $i = 1, 2, \ldots, n$. Note that the case $d = 0$ refers to the estimation of the density itself. Let $T < \tau_H$ be a fixed constant, where $\tau_H = \inf\{x : H(x) = 1\} \le \infty$ is the least upper bound for the support of H, the distribution function of Z_1 and $f_1(x) = f(x)I(x \le T)$. We would like to estimate the d-th ($d \ge 0$) derivative of f over the interval $x \in (-\infty, T)$, or equivalently we estimate $f_1^{(d)}(x)$, for $x \in (-\infty, T)$. To motivate the estimator, we consider the wavelet expansion of f_1 as in Hall and Patil (1995) (see their Section 2.2 with $s = 0$) (cf. Li, 2008),

$$f_1(x) = \sum_{j\in Z} a_j\phi_j(x) + \sum_{i\ge 0}\sum_{j\in Z} b_{ij}\psi_{ij}(x),$$

$$a_j = \int f_1(x)\phi_j(x)\mathrm{d}x, \qquad b_{ij} = \int f_1(x)\psi_{ij}(x)\mathrm{d}x.$$

The non-linear wavelet estimator of f_1, as given below (see Li, 2008, Eq. (2.5)) may be motivated by the plug-in method resulting in:

$$\hat{f}_1(x) = \sum_{j\in Z}\hat{a}_j\phi_j(x) + \sum_{i=0}^\infty\sum_{j\in Z}\hat{b}_{ij}I(|\hat{b}_{ij}| > \eta)\psi_{ij}(x), \tag{3.4}$$

where $\eta > 0$ is a "threshold" parameter; the constants $\hat{a}_j$ and $\hat{b}_{ij}$ are defined as

$$\hat{a}_j = \int I(x \le T)\phi_j(x)\mathrm{d}\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i I(Z_i \le T)\phi_j(Z_i)}{1-\hat{G}(Z_i^-)},$$

$$\hat{b}_{ij} = \int I(x \le T)\psi_{ij}(x)\mathrm{d}\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i I(Z_i \le T)\psi_{ij}(Z_i)}{1-\hat{G}(Z_i^-)}.$$

For estimators of the derivatives we use the generalized Fourier coefficients of $f_1^{(d)}$ (see Section 2.5 of Hall & Patil, 1995), namely,

$$\alpha_j^{(d)} = \int f_1^{(d)}\phi_j = (-1)^d \int f(x)\phi^{(d)}(x-j)\mathrm{d}x,$$
$$\beta_{ij}^{(d)} = \int f_1^{(d)}\psi_{ij} = (-1)^d 2^{i(d+(1/2))} \int f(x)\psi^{(d)}(2^i x - j)\mathrm{d}x.$$

Using plug-in estimators of $\alpha_j^{(d)}$ and $\beta_{ij}^{(d)}$, the nonlinear wavelet thresholding estimator of $f_1^{(d)}$ is given by

$$\tilde{f}_1^{(d)}(x) = \sum_{j\in Z} \hat{\alpha}_j\phi_j(x) + \sum_{i=0}^{\infty}\sum_{j\in Z} \hat{\beta}_{ij} I(|\hat{\beta}_{ij}| > \lambda)\psi_{ij}(x), \tag{3.5}$$

where we have suppressed the superfix (d) from the estimators of the wavelet coefficients, i.e.

$$\begin{aligned}\hat{\alpha}_j &= (-1)^d \int \phi_j^{(d)}(x) I(x\le T)\mathrm{d}\hat{F}_n(x)\\ &= (-1)^d \sum_{m=1}^{n} \frac{\delta_m I(Z_m\le T)\phi_j^{(d)}(Z_m)}{n(1-\hat{G}_n(Z_m^-))},\end{aligned} \tag{3.6}$$

$$\begin{aligned}\hat{\beta}_{ij} &= (-1)^d \int \psi_{ij}^{(d)}(x) I(x\le T)\mathrm{d}\hat{F}_n(x)\\ &= (-1)^d \sum_{m=1}^{n} \frac{\delta_m I(Z_m\le T)\psi_{ij}^{(d)}(Z_m)}{n(1-\hat{G}_n(Z_m^-))}.\end{aligned} \tag{3.7}$$

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 estimator in Eq. (3.5) is known as a 'hard-thresholding' wavelet estimator and may not achieve the optimal convergence rate (see Li, 2008) for wavelet density estimation. Hence a block thresholding method (see Cai, 1999, 2002 and Chicken & Cai, 2005 for details) as described in Li (2008) may be employed. This method provides the following form of the wavelet estimator of $f^{(d)}(x)$,

$$\hat{f}_1^{(d)}(x) = \sum_{j\in Z} \hat{\alpha}_j\phi_j(x) + \sum_{i=0}^{R}\sum_{k\in\mathbb{Z}}\sum_{j\in\Gamma_{ik}} \hat{\beta}_{ij}\psi_{ij}(x) I(\hat{B}_{ik} > Cn^{-1}), \tag{3.8}$$

where R is a smoothing parameter, $\Gamma_{ik} = \{j : (k-1)l+1 \le j \le kl\}$, $-\infty < k < \infty$ represents consecutive, nonoverlapping blocks of length l for each resolution i and $\hat{B}_{ik}$ $\left(=l^{-1}\sum_{j\in\Gamma_{ik}}\hat{\beta}_{ij}^2\right)$ represents the average estimated squared bias for the block Γ_{ik}. This may be presented in a simplified form as

$$\hat{f}_1^{(d)}(x) = \sum_{j\in Z} \hat{\alpha}_j\phi_j(x) + \sum_{i=0}^{R}\sum_{k\in\mathbb{Z}} \hat{d}_{ik}(x) I(x\in J_{ik}) I(\hat{B}_{ik} > Cn^{-1}), \tag{3.9}$$

where $\hat{d}_{ik}(x) = \sum_{j\in\Gamma_{ik}}\hat{\beta}_{ij}\psi_{ij}(x)$ and

$$J_{ik} = \bigcup_{j\in\Gamma_{ik}}\{x : \psi_{ij}\ne 0\} = \bigcup_{j\in\Gamma_{ik}}\{\mathrm{supp}\ \psi_{ij}\}.$$

Our basic aim is to extend the results of Li (2008) that deals with the density estimation to the derivatives of compactly supported densities. For completeness we reproduce some of his notations that will be useful for further analysis. Define the reproducing wavelet kernel $K(x,y) = (-1)^d\sum_j \phi_j(x)\phi_j^{(d)}(y)$ (see Müller & Gasser, 1979). By self-similarity of multi-resolution subspaces, $K_i(x,y) = 2^{i(s+1)}K(2^i x, 2^i y)$, $i = 0, 1, 2, \ldots,$ is a reproducing kernel on the space spanned by functions ϕ_j, $j\in\mathbb{Z}$. Now, we may define

$$\mathcal{K}_i f^{(d)}(x) = \int K_i(x,y) f(y)\mathrm{d}y, \tag{3.10}$$

$$\mathcal{D}_i f^{(d)}(x) = \int D_i(x,y) f(y)\mathrm{d}y \tag{3.11}$$

where $D_i(x,y) = (-1)^d\sum_j \psi_{ij}(x)\psi_{ij}^{(d)}(y)$. In terms of the above notation wavelet expansion of $f_1^{(d)} \in L^2$, may be written as

$$f_1^{(d)}(x) = \mathcal{K}_0 f_1^{(d)}(x) + \sum_{i=0}^{\infty}\mathcal{D}_i f_1^{(d)}(x) \tag{3.12}$$

where $\mathcal{K}_0 f^{(d)}$ represents the orthogonal projection of $f^{(d)}$ on the space spanned by functions ϕ_j, $j\in\mathbb{Z}$, and $\mathcal{D}_i f^{(d)}$ represents that on the space spanned by functions ψ_{ij}, $j\in\mathbb{Z}$.

Noting that $D_i(x, y) = K_{i+1}(x, y) - K_i(x, y)$ and there exists an integrable function Q such that $|K(x, y)| \le Q(x - y)$ for all x, y, which implies that for all integers i and all $1 \le p \le \infty$,

$$\|\mathcal{K}_i f^{(d)}\|_p \le \|Q\|_1 \|f\|_p \quad \text{and} \quad \|\mathcal{D}_i f^{(d)}\|_p \le \|Q\|_1 \|f\|_p.$$

We will also use the following estimators for $\mathcal{K}_i f^{(d)}(x)$ and $\mathcal{D}_i f^{(d)}(x)$ as suggested by Eqs. (3.10) and (3.11):

$$\hat{\mathcal{K}}_i^{(d)}(x) = \frac{1}{n}\sum_{m=1}^{n} K_i(x, X_m), \tag{3.13}$$

$$\hat{\mathcal{D}}_i^{(d)}(x) = \frac{1}{n}\sum_{m=1}^{n} D_i(x, X_m). \tag{3.14}$$

In what follows we consider wavelets ϕ and ψ as those in Cohen, Daubechies, and Vial (1993), *i.e.* we assume that ϕ_j and ψ_{ij} are compactly supported in $[-L, T]$ and form a complete orthonormal basis of $L^2(-L, T)$. Li (2008) considered a subset of densities in Besov class $B^s_{p,q}$ for $s > 1/p;\ p, q \in [1, \infty)$. Our choice of the subset of the Besov class is different in that it depends on the order of the derivative to be estimated. For example, for $p \in [2, \infty)$, we assume that $s > \max(d, 1/p)$. This obviously gives the same subset as in Li (2008) for $d = 0$. Since, $F^s_{p,q}(M, L)$ is a subset of the space of bounded functions, we study the wavelet estimator for densities belonging to $F^s_{p,q}(M, L) \bigcap B_\infty(A)$ where $B_\infty(A)$ is the space of all densities f such that $\|f\|_\infty \le A$.

The following theorem extends Theorem 3.1 of Li (2008) for the wavelet estimators of the derivatives of densities over certain Besov spaces.

Theorem 3.1. *Assume that the wavelets ϕ and ψ are $(r + d)$-regular and have d bounded derivatives. Let $\hat{f}_1^{(d)}$ be the block thresholding wavelet estimator (3.8) where the block length $l = \log n$, and $R = \lfloor \log_2(nl^{-2}) \rfloor$ and the thresholding constant C is given as*

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

with C_1 and C_2 being universal constants as given in Talagrand(1994) (see also Li, 2008, pp 1535). Then there exists a constant C_0 such that for all $M, L \in (0, \infty)$ and $q \in [1, \infty]$:

(i) *For $p \in [2, \infty]$ and $s \ge \max(d, 1/p)$,*

$$\sup_{f^{(d)} \in F^s_{p,q}(M,L) \bigcap B_\infty(A)} E\int (\hat{f}_1^{(d)} - f_1^{(d)})^2 \le C_0 n^{-2(s-d)/(1+2s)}.$$

(ii) *For $p \in [1, 2)$ and $s \ge \max\left(\frac{(2d+1)+(2-p)/2}{p}, 1/p\right)$,*

$$\sup_{f^{(d)} \in F^s_{p,q}(M,L) \bigcap B_\infty(A)} E\int (\hat{f}_1^{(d)} - f_1^{(d)})^2 \le C_0(\log_2 n)^{\frac{(2-p)(2d+1)}{p(1+2s)}} n^{-2(s-d)/(1+2s)}.$$

Remark 3.1. This study can be considered as an extension from complete data to randomly right censored data. If we assume that there is no censoring, i.e., $G \equiv 0$ on $(-\infty, \infty)$, then $\delta_i \equiv 1$, for all $i = 1, 2, \ldots, n$ and we get the same result in Chaubey, Doosti, and Prakasa Rao (2006, 2008), Hall and Patil (1995) and Prakasa Rao (1996). Also, assuming $d = 0$, we get the same result as in Li (2008) based on censored data.

Remark 3.2. Since the choice of block size l and thresholding constant C largely determines the performance of the resulting estimator, it is important to study in detail the effect of l and C on the properties of the estimator and derive the optimal l and C if such values exist. Between the two parameters l and C, the block size l is more important and plays a similar role as the bandwidth in the traditional kernel estimation. Here we can consider the block thresholding estimator in (3.8) with general block size $l = (\log n)^s$ with some $s \ge 0$, but based on Cai (2002), we know that to achieve the optimal global adaptivity, the block size must be at least of the order $O(\log n)$. On the other hand, to achieve the optimal local adaptivity, the block size must be no more than $O(\log n)$. Therefore no block thresholding estimator can achieve simultaneously the optimal global and local adaptivity if the block size is not of order $O(\log n)$.

4. A simulation study

In this section, we study the performance of the proposed estimator (3.9) and compare it with various other threshold wavelet estimators as available in the *WaveLab* package developed by Buckheit, Chen, Donoho, Johnstone, and Scargle (1995) at Stanford University. Also it may be better to refer to Antoniadis, Bigot, and Sapatinas (2001) for a detailed description of the procedures. In the following two examples, we generate two sets of random variables and find the average norm criterion

Table 1
Computed values for ANorm and simulated standard errors for Example 1 for various sample sizes; standard errors are in parentheses.

Estimation methods	ANorm and (simulated standard error)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Block thresholding	33.378 (6.670)	24.103 (4.318)	19.186 (2.651)	16.627 (1.994)
NeighBlock	33.558 (6.716)	24.149 (4.312)	19.160 (2.674)	16.578 (1.998)
Minimax	32.891 (5.499)	24.046 (3.507)	19.180 (2.398)	16.731 (1.878)
SureShrink	39.544 (5.204)	28.454 (3.893)	21.895 (2.579)	18.378 (1.893)
Hard threshold	39.858 (5.609)	27.842 (4.158)	21.374 (2.687)	17.769 (2.051)
Linear	45.061 (4.684)	32.446 (2.934)	24.824 (2.381)	20.485 (1.803)

Table 2
Computed values for ANorm and simulated standard errors for Example 2 for various sample sizes; standard errors are in parentheses.

Estimation methods	ANorm and (simulated standard error)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Block thresholding	22.786 (10.100)	16.043 (5.525)	11.455 (4.221)	9.814 (2.206)
NeighBlock	23.117 (10.193)	16.202 (5.658)	11.611 (4.241)	9.928 (2.181)
Minimax	25.873 (8.366)	18.419 (4.483)	13.193 (3.502)	10.863 (1.100)
SureShrink	30.733 (11.608)	21.032 (6.865)	13.893 (5.087)	11.873 (3.249)
Hard threshold	31.721 (10.529)	21.557 (6.290)	14.3871 (4.900)	11.555 (2.936)
Linear	44.839 (7.547)	32.707 (4.126)	23.687 (3.458)	17.871 (2.100)

(ANorm) calculated for several thresholding estimators, including (3.9) for different sample sizes n. We show in detail the results for (ANorm) and the total number of replications $N = 100$ and $d = 1$; the ANorm criterion is defined as

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

where $\hat{f}_l^{(d)}(.)$ denotes an estimator of $f^{(d)}(.)$ at the lth replication. In this simulation study, we used Daubechies's compactly supported wavelet Symmlet 4 (see Daubechies, 1992, p. 198) and Coiflet 2 (see Daubechies, 1992, p. 258), and primary resolution level $j_0 = 3$. The code was written in MATLAB environment using the WaveLab software. Lower values of ANorm is believed to indicate better performance. We also list the corresponding standard errors in these tables.

Example 1. Here we generate the random samples X_i, $1 \le i \le n$ from a Beta distribution with parameters $\alpha = 2$, $\beta = 5$ along with the independent random sample Y_i, $1 \le i \le n$ from the uniform distribution on the interval $(0, 1)$. Table 1, shows the values of ANorm for several wavelet estimators of the derivative of the beta density function in different sample sizes. Block thresholding estimator has better performance compared to other wavelet estimators. When sample size increases the performance of other estimators come closer to the performance of our proposed estimator. Fig. 1 shows the original derivative of *pdf* of X_i with black line, Block thresholding estimator with red line, hard thresholding estimator with blue line and linear estimator with dotted line, respectively.

Example 2. For this example, we consider the simulated data from an exponential family. We generate our sample, X_i, $1 \le i \le n$ from exponential distribution with mean 1. The sample, Y_i, $1 \le i \le n$ is generated from exponential density with mean 5. Table 2 and Fig. 2 give similar results as in Example 1.

In both of these examples, the simulation results show that the block thresholding estimator performs better or close to the NeighBlock estimator that seems to outperform all other competing estimators. It may be noted here that NeighBlock estimator is a modified form of block thresholding estimators that borrows information from neighboring blocks (see Cai & Silverman, 2001).

Remark 4.1. Similar numerical study for density function estimation was carried out in Li (2008) where ANorm values are smaller than those in our study. However, we should note that the values in Li (2008) are for the estimation of the probability density functions whereas in our case these are for the estimation of its derivative. Note that the rate of convergence for estimating the *pdf* is not the same as the rate of convergence for estimating its derivative.

Acknowledgments

The authors are thankful to an anonymous referee for a careful revision of the manuscript and some constructive suggestions. Y.P. Chaubey acknowledges NSERC of Canada for the partial support of this research through his discovery grant.

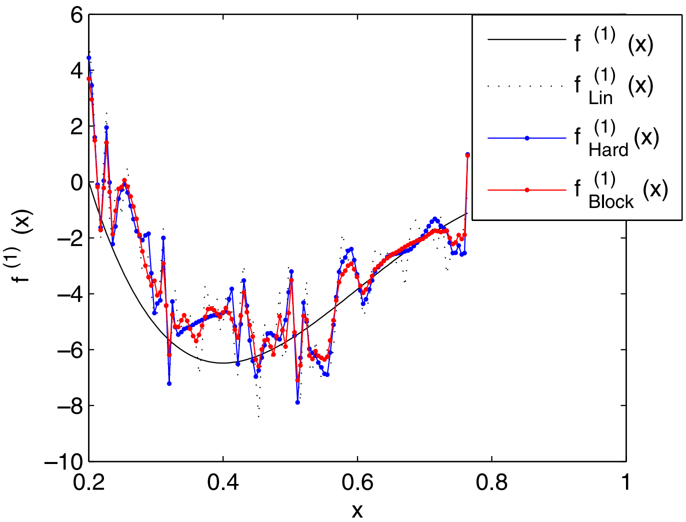

Fig. 1. Derivative estimation for beta density. The original derivative of density function (*solid line*), block thresholding estimator (*red line*), hard thresholding estimator (*blue line*) and linear estimator (*dotted line*) respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

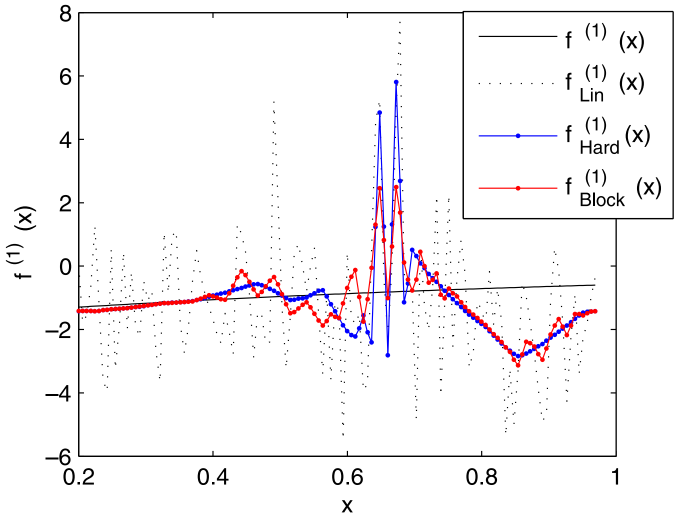

Fig. 2. Derivative estimation for exponential density. The original derivative of density function (*solid line*), block thresholding estimator (*red line*), hard thresholding estimator (*blue line*) and linear estimator (*dotted line*) respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

E. Shirazi was supported by Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad. The third author was supported by a grant from Ferdowsi University of Mashhad, No. MS88056DST.

Appendix. Proof of Theorem 3.1

First we sketch the basic steps of the proof that follows along the lines of Li (2008). First we state a version of Lemma 5.1 of Li (2008) for wavelet density derivatives, that compares the estimators of the wavelet coefficients to those for the complete data case. Its proof can be provided basically by mimicking the steps of the proof of Li (2008), hence it is omitted.

Lemma. *Let $\hat{\alpha}_j$ and $\hat{\beta}_{ij}$ be defined as in Eqs. (3.6) and (3.7). Also, let*

$$\varphi_j^{(d)}(x) = \phi_j^{(d)}(x) I(x \le T) \quad j = 0, \pm 1, \pm 2, \ldots \tag{A.1}$$

$$\varphi_{ij}^{(d)}(x) = \psi_{ij}^{(d)}(x) I(x \le T) \quad i = 0, 1, \ldots, R;\ j = 0, \pm 1, \pm 2, \ldots, \tag{A.2}$$

$$\bar{\alpha}_j = \frac{(-1)^d}{n} \sum_{m=1}^{n} \frac{\delta_m \varphi_j^{(d)}(Z_m)}{1 - \hat{G}(Z_m)}, \quad j = 0, \pm 1, \pm 2, \ldots, \tag{A.3}$$

$$\bar{\beta}_{ij} = \frac{(-1)^d}{n} \sum_{m=1}^{n} \frac{\delta_m \varphi_{ij}^{(d)}(Z_m)}{1 - \hat{G}(Z_m)}, \quad i = 0, 1, \ldots, R;\ j = 0, \pm 1, \pm 2, \ldots. \tag{A.4}$$

Then the following equations hold.

$$\hat{\alpha}_j = \bar{\alpha}_j + \bar{W}_j + R_{n,j}, \qquad E(R_{n,j}^2) = O\left(\frac{1}{n^2}\right)\int (\varphi_j^{(d)})^2 \mathrm{d}F. \tag{A.5}$$

$$\hat{\beta}_{ij} = \bar{\beta}_{ij} + \bar{W}_{ij} + R_{n,ij}, \qquad E(R_{n,ij}^2) = O\left(\frac{1}{n^2}\right)\int (\varphi_{ij}^{(d)})^2 \mathrm{d}F \tag{A.6}$$

where

$$W_j(Z_m) = U_j(Z_m) - V_j(Z_m), \qquad W_{ij}(Z_m) = U_{ij}(Z_m) - V_{ij}(Z_m),$$

$$\bar{W}_j = n^{-1}\sum_{m=1}^{n} W_j(Z_m), \qquad \bar{W}_{ij} = n^{-1}\sum_{m=1}^{n} W_{ij}(Z_m), \tag{A.7}$$

$$U_j(Z_m) = \frac{(-1)^d(1-\delta_{(m)})}{1-H(Z_m)}\int_{Z_m}^{\tau_H} \varphi_j^{(d)}(\omega)F(\mathrm{d}\omega)$$

$$U_{ij}(Z_m) = \frac{(-1)^d(1-\delta_{(m)})}{1-H(Z_m)}\int_{Z_m}^{\tau_H} \varphi_{ij}^{(d)}(\omega)F(\mathrm{d}\omega)$$

$$V_j(Z_m) = (-1)^d\int_{-L}^{\tau_H}\int_{-L}^{\tau_H} \frac{\varphi_j^{(d)}(\omega)I(\nu < Z_m \wedge \omega)}{(1-H(\nu))(1-G(\nu))}G(\mathrm{d}\nu)F(\mathrm{d}\omega)$$

$$V_{ij}(Z_m) = (-1)^d\int_{-L}^{\tau_H}\int_{-L}^{\tau_H} \frac{\varphi_{ij}^{(d)}(\omega)I(\nu < Z_m \wedge \omega)}{(1-H(\nu))(1-G(\nu))}G(\mathrm{d}\nu)F(\mathrm{d}\omega). \tag{A.8}$$

As in Li (2008), we decompose $E\|\hat{f}_1^{(d)} - f_1^{(d)}\|_2^2$ into several parts, in view of (3.9) and (3.12):

$$E\|\hat{f}_1^{(d)} - f_1^{(d)}\|_2^2 \le 4(I_1 + I_2 + I_3 + I_4) \tag{A.9}$$

where

$$I_1 = E\|\hat{\mathcal{K}}_0 - \mathcal{K}_0 f^{(d)}\|_2^2,$$

$$I_2 = E\left\|\sum_{i=0}^{i_s}\left[\sum_k \hat{d}_{ik}I(J_{ik})I(\hat{B}_{ik} > Cn^{-1}) - \mathcal{D}_i f^{(d)}\right]\right\|_2^2,$$

$$I_3 = E\left\|\sum_{i=i_s+1}^{R}\left[\sum_k \hat{d}_{ik}I(J_{ik})I(\hat{B}_{ik} > Cn^{-1}) - \mathcal{D}_i f^{(d)}\right]\right\|_2^2,$$

$$I_4 = E\left\|\sum_{i=R+1}^{\infty} \mathcal{D}_i f^{(d)}\right\|_2^2. \tag{A.10}$$

Next, we obtain the upper bounds for I_1, I_2, I_3 and I_4, as follows:

$$I_1 = O(n^{-1});$$
$$I_2 \le C(\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{-2(s-d)/(1+2s)};$$
$$I_3 \le Cn^{-2(s-d)/(1+2s)};$$
$$I_4 = o(n^{-2(s-d)/(1+2s)}),$$

proofs of which are detailed subsequently.

Using the above bounds in Eq. (A.10) the proof is completed. □

Bound for I_1: Using orthogonality of wavelets ϕ_j, we have

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

From the lemma given earlier and the fact that

$$3(a^2 + b^2 + c^2) - (a + b + c)^2 \ge 0,$$

we have

$$I_1 \le 3\left\{\sum_j E(\bar{\alpha}_j-\alpha_j)^2+\sum_j E\bar{W}_j^2+\sum_j ER_{n,j}^2\right\}$$
$$=: 3(I_{11}+I_{12}+I_{13}). \tag{A.11}$$

Noting that

$$E(\bar{\alpha}_j-\alpha_j)^2 = E\left\{(-1)^d 2^{i_0(d+1/2)}n^{-1}\sum_{m=1}^{n}\frac{\delta_m\varphi^{(d)}(2^{i_0}Z_m-j)}{1-G(Z_m)}\right\}^2-n^{-1}\alpha_j^2$$
$$= (-1)^{2d}2^{i_0(2d+1)}n^{-1}\int(\varphi^{(d)})^2(y)\frac{f_1^{(d)}((y+j)/2^{i_0})}{1-G((y+j)/2^{i_0})}\mathrm{d}y-n^{-1}\alpha_j^2,$$

we obtain

$$\sum_j E(\bar{\alpha}_j-\alpha_j)^2 = 2^{i_0(2d+1)}n^{-1}\int(\varphi^{(d)})^2(y)\sum_j 2^{-i_0}\frac{f_1((y+j)/2^{i_0})}{1-G((y+j)/2^{i_0})}\mathrm{d}y-n^{-1}\sum_j\alpha_j^2.$$

Since

$$\sum_j 2^{-i_0}f_1((y+j)/2^{i_0})/(1-G((y+j)/2^{i_0}))\to\int f_1/(1-G)$$

and

$$\sum_j\alpha_j^2 = O\left(\int f_1^2\right)\int(\varphi_j^{(d)})^2,$$

we get $E\sum_j(\bar{\alpha}_j-\alpha_j)^2 = 2^{i_0(2d+1)}n^{-1}\int f_1/(1-G)\int(\varphi^{(d)})^2+o(2^{i_0(2d+1)}n^{-1})$, therefore

$$I_{11} = O(2^{i_0(2d+1)}n^{-1}) = O(n^{-1}).$$

Further from (A.7)

$$I_{12}\le n^{-1}\sum_j EW_j^2(Z_1)\le 2n^{-1}\sum_j(EU_j^2(Z_1)+EV_j^2(Z_1)).$$

In view of (A.8), applying the Cauchy–Schwarz inequality and using the compact support of ϕ, we finally obtain

$$EU_j^2(Z_1)\le\frac{1}{[(1-H(T))][(1-G(T))]}2^{i_0(2d+1)}\int(\varphi^{(d)})^2(y)f_1^2((y+j)/2^{i_0})\mathrm{d}y.$$

Hence,

$$n^{-1}\sum_j EU_j^2(Z_1) = O\left(2^{i_0(2d+1)}n^{-1}\int(\varphi^{(d)})^2(y)\sum_j 2^{-i_0}f_1^2((y+j)/2^{i_0})\mathrm{d}y\right) = o(2^{i_0(2d+1)}n^{-1}).$$

Similarly, we obtain

$$EV_j^2(Z_1)\le\frac{1}{[(1-H(T))]^2[(1-G(T))]^2}2^{i_0(2d+1)}\int(\varphi^{(d)})^2(y)f_1^2((y+j)/2^{i_0})\mathrm{d}y.$$

Thus, $n^{-1}\sum_j EV_j^2(Z_1) = o(2^{i_0(2d+1)}n^{-1})$ and we get

$$I_{12} = o(2^{i_0(2d+1)}n^{-1}) = o(n^{-1}).$$

Using the property of $E(R_{n,ij}^2)$ from (A.5), we have

$$I_{13} = O(n^{-2})\sum_j\int(\varphi_j^{(d)})^2\mathrm{d}F = O(2^{i_0(2d+1)}n^{-2}) = o(2^{i_0(2d+1)}n^{-2}) = O(n^{-2}).$$

Now, using the bounds obtained for I_{11}, I_{12} and I_{13} in (A.11), we have $I_1 = O(n^{-1})$. □

Bound for I_2: From similar methods as for the proof of Lemma 5.6 in Li (2008), we get

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

Writing $\mathcal{D}_i f^{(d)}(x) = \sum_j \beta_{ij}\psi_{ij}(x) =: \sum_k \sum_{j\in \Gamma_{ik}} \beta_{ij}\psi_{ij}(x) =: \sum_k d_{ik}f^{(d)}$, we have for the term in the square brackets

$$\begin{aligned}
&E\int\left(\sum_k \hat{d}_{ik}(x)I(J_{ik})I(\hat{B}_{ik} > Cn^{-1}) - \mathcal{D}_i f^{(d)}(x)\right)^2 \mathrm{d}x \\
&\le 3\left\{E\int\left[\sum_k(\hat{d}_{ik}(x) - d_{ik}f^{(d)}(x))I(J_{ik})I(\hat{B}_{ik} > Cn^{-1})\right]^2 \mathrm{d}x + E\int_{J_{ik}}\left[\sum_k d_{ik}f^{(d)}(x)I(B_{ik} < 2Cn^{-1})\right.\right. \\
&\quad\left.\left.\times I(\hat{B}_{ik} \le Cn^{-1})\right]^2 \mathrm{d}x + E\int_{J_{ik}}\left[\sum_k d_{ik}f^{(d)}(x)I(B_{ik} > 2Cn^{-1})I(\hat{B}_{ik} \le Cn^{-1})\right]^2 \mathrm{d}x\right\} \\
&\le 3\left\{E\int(\hat{d}_{ik}(x) - d_{ik}f^{(d)}(x))^2\mathrm{d}x\right\} + \left\{E\sum_k\int_{J_{ik}}(d_{ik}f^{(d)}(x))^2\mathrm{d}xI(B_{ik} \le 2Cn^{-1})\right\} \\
&\quad + \left\{E\sum_k\int_{J_{ik}}(d_{ik}f^{(d)}(x))^2\mathrm{d}xI(B_{ik} > 2Cn^{-1})I(\hat{B}_{ik} \le Cn^{-1})\right\} \\
&=: 3(I_{21} + I_{22} + I_{23}),
\end{aligned}$$

where $I_{22} \le C2^i n^{-1}$ and $I_{23} \le Cn^{-1}$ for all i. Also

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

Applying the same arguments as for the term I_1, we get

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

Thus, $I_{21} \le C2^{i(2d+1)}n^{-1} = O\left(n^{-\frac{2(s-d)}{1+2s}}\right)$ and we get

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

Now, under the assumptions of Theorem 3.1, if $p \ge 2$ and i_s satisfies $2^{i_s} \simeq n^{1/(2s+1)}$, then $I_2 \le Cn^{-2(s-d)/(2s+1)}$. If $1 \le p < 2$ and i_s satisfies $2^{i_s} \simeq (\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{1/(1+2s)}$, then

$$I_2 \le C(\log_2 n)^{\frac{2-p}{p(1+2s)}} n^{-2(s-d)/(1+2s)}. \quad \square$$

Bound for I_3: The arguments to obtain the desired bound for I_3, are similar to the proof of Lemma 5.7 in Li (2008), hence the details are omitted. $\square$

Bound for I_4: The arguments for the term I_4 are also very similar to that for I_4 in Li (2008), except that the smoothing index s is different. It leads to the result that for $p < 2$, $I_4 \le M^2 2^{-2i\sigma}$. On the basis of the choice of R with $2^R \simeq n(\log_2 n)^{-2}$ and $2\sigma = 1 + 2(s - 1/p) > 2(s - d)/(1 + 2s)$, we have $I_4 = o(n^{-2(s-d)/(1+2s)})$. For $p \ge 2$ proof goes along the similar lines. $\square$

References

Antoniadis, A. (2007). Wavelet methods in statistics: some recent dvelopments and their applications. *Statistics Surveys*, *1*, 16–55.

Antoniadis, A., Bigot, J., & Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, *6*, See http://www.jstatsoft.org/v06/i06.

Antoniadis, A., Grégoire, G., & Nason, G. (1999). Density and hazard rate estimation for right-censored data by using wavelet methods. *Journal of the Royal Statistical Society*, *61*, 63–84.

Bhattacharya, P. K. (1967). Estimation of a probability density function and its derivatives. *Sankhyā A*, *29*, 373–382.

Buckheit, J. B., Chen, S., Donoho, D. L., Johnstone, I. M., & Scargle, J. (1995). About waveLab. *Technical report*. USA: Department of Statistics, Stanford University. Available at: http://www-stat.stanford.edu/~wavelab.

Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Annals of Statistics*, *27*, 898–924.

Cai, T. (2002). On block thresholding in wavelet regression: adaptivity, block size, and threshold level. *Statistica Sinica*, *12*, 1241–1273.

Cai, T. T., & Silverman, B. W. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhyā B*, *63*, 127–148.

Chaubey, Y. P., Doosti, H., & Prakasa Rao, B. L. S. (2006). Wavelet based estimation of the derivatives of a density with associated variables. *International Journal of Pure & Applied Mathematics*, *27*, 97–106.

Chaubey, Y. P., Doosti, H., & Prakasa Rao, B. L. S. (2008). Wavelet based estimation of derivative of a density for negatively associated process. *Journal of Statistical Theory and Practice*, *2*, 453–463.

Chaubey, Y. P., Sen, A., & Sen, P. K. (2007). Smoothed funcional estimation for censored data. In *Encyclopedia of statistics in quality and reliability*. Chichester, UK: John Wiley and Sons.

Chicken, E., & Cai, T. (2005). Block thresholding for density estimation: local and global adaptivity. *Journal of Multivariate Analysis*, *95*, 76–106.

Cohen, A., Daubechies, I., & Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, *27*, 97–106.

Dabrowska, D. M. (1995). Nonparametric regression with censored covariates. *Journal of Multivariate Analysis*, *54*, 253–283.

Daubechies, I. (1988). Orthogonal bases of compactly supported wavelets. *Statistica Sinica*, *41*, 909–996.

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

Delouille, V., Simoens, J., & von Sachs, R. (2001). Smooth design-adapted wavelets for nonparametric stochastic regression. *Discussion paper 0117*. Belgium: Institut de Statistique, UCL.

Devroye, L. (1987). *A course in density estimation*. MA, USA: Birkhaüser.

Diehl, S., & Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *Journal of Multivariate Analysis*, *25*, 299–310.

Donoho, D. L., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinking. *Journal of the American Statistical Association*, *90*, 1200–1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., & Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *Journal of the Royal Statistical Society: Series B*, *57*, 301–369.

Hall, P., Kerkyacharian, G., & Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet method. *Annals of Statistics*, *26*, 922–942.

Hall, P., Kerkyacharian, G., & Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Annals of Statistics*, *9*, 33–50.

Hall, P., & Patil, P. (1995). Formulae for mean integated squared error of non-linear wavelet-based density estimators. *Annals of Statistics*, *23*, 905–928.

Härdle, W., Kerkycharian, , Picard, D., & Tsybakov, T. (1998). *Lecture notes in statistics*: *Vol. 129*. *Wavelets, approximations, and statistical applications*. New York: Springer.

Härdle, W., Marron, J. S., & Wand, M. P. (1990). Bandwidth choice for density derivatives. *Journal of the Royal Statistical Society: Series B*, *52*, 223–232.

Leblanc, F. (1996). Wavelet linear density estimator for a discrete-time stochastic process: L_p-losses. *Statistics & Probability Letters*, *27*, 71–84.

Li, L. (2002). Hazard rate estimation for censored data by wavelet methods. *Communications in Statistics—Theory and Methods*, *31*, 943–960.

Li, L. (2003). Non-linear wavelet-based density estimators under random censorship. *Journal of Statistical Planning and Inference*, *117*, 35–58.

Li, L. (2008). On the block thresholding wavelet estimators with censored data. *Journal of Multivariate Analysis*, *99*, 1518–1543.

Li, Q., & Racine, J. S. (2006). *Nonparametric econometrics: theory and practice*. New Jersey: Princeton University Press.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 674–693.

McNichols, D. T., & Padgett, W. J. (1984). Nonparametric estimation from accelerated life tests with random censorship. In *Notes rep. comput. sci. appl. math.*: *Vol. 10*. *Reliability theory and models (Charlotte, NC, 1983)* (pp. 155–167). Orlando, FL: Academic Press.

Meyer, Y. (1992). *Wavelets and operators*. Cambridge: Cambridge University Press.

Müller, H., & Gasser, T. (1979). Optimal convergence properties of kernel estimates of derivatives of a density function. In *Lecture notes in math.*: *Vol. 757*. *Smoothing techniques for curve estimation (Proc. workshop, Heidelberg, 1979)* (pp. 144–154). Berlin: Springer.

Padgett, W. J., & McNichols, D. T. (1984). Nonparametric density estimation from censored data. *Communications in Statistics—Theory and Methods*, *13*, 1581–1611.

Pagan, A., & Ullah, A. (1999). *Nonparametric econometrics*. Cambridge: Cambridge University Press.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, *33*, 1065–1076.

Patil, P. (1997). Nonparametric hazard rate estimation by orthogonal wavelet methods. *Journal of Statistical Planning and Inference*, *60*, 153–168.

Prakasa Rao, B. L. S. (1983). *Nonparametric functional estimation*. New York: Academic Press.

Prakasa Rao, B. L. S. (1996). Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bulletin of Informatics and Cybernetics*, *28*, 91–100.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, *27*, 832–837.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York, NY: Chapman and Hall.

Singh, R. S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems. *Journal of the Royal Statistical Society*, *B39*, 357–363.

Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Annals of Probability*, *22*(1), 28–76.

Tribel, H. (1992). *Theory of function space II*. Berlin: Birkhaüser Verlag.

Vidakovic, A. (1999). *Statistical modelling by wavelets*. New York: Springer.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman and Hall.

Wand, M. P., & Ripley, B. D. (2009). KernSmooth: functions for kernel smoothing for Wand and Jones (1995). R package version 2. 22-19. URL: http://CRAN.R-project.org/.