WAVELET BLOCK THRESHOLDING FOR DENSITY ESTIMATION IN THE PRESENCE OF BIAS

Christophe Chesneau

WAVELET BLOCK THRESHOLDING FOR DENSITY ESTIMATION IN THE PRESENCE OF BIAS

Christophe Chesneau

Laboratoire de Mathématiques Nicolas Oresme,
Université de Caen Basse-Normandie,
Campus II, Science 3,
14032 Caen, France.

http://www.chesneau-stat.com *christophe.chesneau@gmail.com*

Abstract

We consider the density estimation problem from i.i.d. biased observations. We investigate the performances of an adaptive wavelet block thresholding estimator via the minimax approach under the $\mathbb{L}^p$ risk with $p \geq 1$ over Besov balls. We prove that it achieves near optimal rates of convergence.

Key words: Adaptive estimation, Biased data, Wavelets, Block thresholding.

1991 MSC: Primary: 62G07, Secondary: 62N01.

---

1 MOTIVATIONS

The classic density estimation problem can be formulated as follows. Let $X_1, ..., X_n$ be an i.i.d. sample from a distribution with density function f. The objective is to estimate the density function f based on the sample. Various estimation techniques have been proposed in the statistical literature. The most popular of them are presented in the books of Devroye and Györfi [8], Silverman [17], Efromovich [9], Härdle et al. [11] and Tsybakov [19].

In this paper, we consider the problem of estimating f without observe directly the sample $X_1, ..., X_n$. We record i.i.d. observations $Z_1, ..., Z_n$ from a biased distribution with the following density function

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

where w denotes a positive function and μ is the real number defined by $\mu = \int w(y)f(y)dy$. Here, w is known. The density function f and the real number μ are unknown. The objective is to estimate the density function f from the observations $Z_1, ..., Z_n$. In what follows, it is always assumed that, without loss of generality, the functions f and w are defined on the unit interval $[0, 1]$. Moreover, we suppose that there exist two constants w_1 and w_2 such that, for any $x \in [0, 1]$,

$$0 < w_1 \leq w(x) \leq w_2 < \infty.$$

The concept of biased data has several applications in various domains such as biology, see Buckland et al. [2], industry, see Cox [7], and economics, see Heckman [12]. We may equally refer to the survey by Patil and Rao [14] on several practical examples of biased distributions.

The density estimation problem for biased data has been considered in several papers. Among them, let us briefly present some recent results established by Efromovich [10] and Brunel et al. [1]. In the article of Efromovich [9], the Efromovich-Pinsker adaptive estimator based on a blockwise shrinkage algorithm is developed. It achieves the exact minimax rate of convergence under the $\mathbb{L}^2$ risk over a Sobolev class, $B^s_{2,2}$. In the article of Brunel et al. [1], a penalized projection density estimator is proposed. It achieves the exact minimax rate of convergence under the $\mathbb{L}^2$ risk over a particular Besov class, $B^s_{2,\infty}$. The main differences between these two estimators are discussed in Remark 3.2 of Brunel et al. [1].

In the present paper, we adopt the minimax point of view under the $\mathbb{L}^p$ risk with $p \geq 1$, not only for $p = 2$, and we consider the Besov classes, $B^s_{\pi,r}$, with no particular restriction on the parameters π and r. We estimate the unknown density function f by using an adaptive wavelet block thresholding estimator suitably adapted to our minimax framework. It can be viewed as a $\mathbb{L}^p$ version of the BlockShrink algorithm initially developed by Cai [3] for the $\mathbb{L}^2$ risk and the regression model with equispaced deterministic sample. We prove that the considered estimator achieves optimal or near optimal rates of convergence according to the values of the parameters s, π, r of the Besov classes, and the parameter p of the $\mathbb{L}^p$ risk. This result follows from a general theorem proved by Chesneau [4] and some technical probability inequalities. Thus, the main difficulties of this work are the proofs of a moment inequality and a specific concentration inequality. Finally, let us mention that, if we restrict our study to the $\mathbb{L}^2$ risk and the Sobolev class $B^s_{2,2}$ or the Besov class $B^s_{2,\infty}$, then our estimator has similar asymptotic minimax performances to the Efromovich-Pinsker adaptive estimator and the penalized projection density estimator.

The paper is organized as follows. In Section 2, we present wavelets and Besov balls. The considered adaptive wavelet block thresholding estimator is defined in Section 3. Section 4 is devoted to the main result. The proofs are postponed in Section 5.

2 WAVELETS AND BESOV BALLS

We consider an orthonormal wavelet basis generated by dilation and translation of a compactly supported "father" wavelet ϕ and a compactly supported "mother" wavelet ψ. For the purposes of this paper, we use the periodized wavelet bases on the unit interval. For any $x \in [0,1]$, any integer j and any $k \in \{0, \ldots, 2^j - 1\}$, let

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \qquad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

be the elements of the wavelet basis and

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

there periodized version. There exists an integer τ such that the collection ζ defined by $\zeta = \{\phi^{per}_{\tau,k}, k = 0, ..., 2^\tau - 1;\ \psi^{per}_{j,k}, \quad j = \tau, ..., \infty,\ k = 0, ..., 2^j - 1\}$ constitutes an orthonormal basis of $\mathbb{L}^2([0,1])$. In what follows, the superscript "*per*" will be suppressed from the notations for convenience. For any integer $l \geq \tau$, a function $f \in \mathbb{L}^2([0,1])$ can be expanded into a wavelet series as

$$f(x) = \sum_{k=0}^{2^l - 1} \alpha_{l,k}\phi_{l,k}(x) + \sum_{j=l}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x), \quad x \in [0,1],$$

where $\alpha_{j,k} = \int_0^1 f(t)\phi_{j,k}(t)dt$ and $\beta_{j,k} = \int_0^1 f(t)\psi_{j,k}(t)dt$. For further details about wavelet bases on the unit interval, we refer to Cohen et al. [6].

Now, let us define the Besov balls. Let $M \in (0,\infty)$, $s \in (0,\infty)$, $\pi \in [1,\infty]$ and $r \in [1,\infty]$. Let us set $\beta_{\tau-1,k} = \alpha_{\tau,k}$. We say that a function f belongs to the Besov balls $B^s_{\pi,r}(M)$ if and only if there exists a constant $M^* > 0$ such that the associated wavelet coefficients satisfy

$$\left(\sum_{j=\tau-1}^{\infty}\left(2^{j(s+1/2-1/\pi)}\left(\sum_{k=0}^{2^j-1}|\beta_{j,k}|^\pi\right)^{1/\pi}\right)^r\right)^{1/r} \leq M^*.$$

For a particular choice of parameters s, π and r, these sets contain the Hölder and Sobolev balls. See Meyer [13].

3 ESTIMATOR

Now, let us describe the main estimator of the study. As mentioned in Section 1, it can be viewed as $\mathbb{L}^p$ version of the BlockShrink estimator initially developed by Cai [3] for the $\mathbb{L}^2$ risk for the regression model with equispaced deterministic sample. The considered $\mathbb{L}^p$ version has been introduced by Picard and Tribouley [16] for general statistical problems in the context of the adaptive confidence interval for pointwise curve estimation.

Let $p \in [1, \infty)$ and $d \in (0, \infty)$. Let j_1 and j_2 be the integers defined by

$$j_1 = \lfloor ((p \vee 2)/2) \log_2(\log n) \rfloor \quad and \quad j_2 = \lfloor \log_2(n/\ln n) \rfloor.$$

Here, $p \vee 2 = \max(p, 2)$ and the quantity $\lfloor a \rfloor$ denotes the whole number part of a.

For any $j \in \{j_1, ..., j_2\}$, let us set $L = \lfloor (\log n)^{(p\vee 2)/2} \rfloor$ and $A_j = \{1, ..., \lfloor 2^j L^{-1} \rfloor\}$. For any $K \in A_j$, we consider the set

$$U_{j,K} = \{k \in \{0, ..., 2^j - 1\};\ (K-1)L \leq k \leq KL - 1\}.$$

We define the ($\mathbb{L}^p$ version of the) BlockShrink estimator by

$$\hat{f}_n(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k}\phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2} \sum_{K \in A_j} \sum_{k \in U_{j,K}} \hat{\beta}_{j,k} 1_{\left\{\hat{b}_{j,K} \geq d n^{-1/2}\right\}} \psi_{j,k}(x), \quad x \in [0,1], \tag{3.1}$$

where $\hat{b}_{j,K} = (L^{-1} \sum_{k \in U_{j,K}} |\hat{\beta}_{j,k}|^p)^{1/p}$,

$$\hat{\alpha}_{j_1,k} = \hat{\mu} n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\phi_{j_1,k}(Z_i), \qquad \hat{\beta}_{j,k} = \hat{\mu} n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i), \tag{3.2}$$

with

$$\hat{\mu} = \left(n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i) \right)^{-1}.$$

Let us mention that $\hat{f}_n$ does not require any a priori knowledge on f (and, a fortiori, μ) in its construction. It is adaptive.

The sets $\mathcal{A}_j$ and $U_{j,K}$ are chosen such that $\cup_{K \in \mathcal{A}_j} U_{j,K} = \{0, ..., 2^j - 1\}$, $U_{j,K} \cap U_{j,K'} = \emptyset$ for any $K \neq K'$ with $K, K' \in \mathcal{A}_j$, and $|U_{j,K}| = L = \lfloor (\log n)^{(p\vee 2)/2} \rfloor$.

The estimators $\hat{\alpha}_{j_1,k}$ and $\hat{\beta}_{j,k}$ are wavelet versions of those proposed by Brunel et al. [1].

In fact, the penalized projection density estimator developed by Brunel et al. [1] can be constructed from a wide variety of bases (trigonometric, polynomial, wavelets ...). However, the choice of the basis has no influence on the minimax performances of this estimator. In our study, due to the complexity of our minimax approach, we really use the multiresolution nature of the wavelet bases to obtain the desired result.

Lemma 3.1 below determines an upper bound for $\left|\hat{\beta}_{j,k} - \beta_{j,k}\right|$.

Lemma 3.1 *Let us consider the biased density model described in the second paragraph of Section* 1*. Suppose that there exists a known constant* f_2 *such that, for any* $x \in [0,1]$,

$$f(x) \leq f_2.$$

We recall that there exist two constants w_1 *and* w_2 *such that, for any* $x \in [0,1]$, *we have* $0 < w_1 \leq w(x) \leq w_2 < \infty$. *Then, for any* $j \in \{j_1, ..., j_2\}$ *and any* $k \in \{0, ..., 2^j - 1\}$, *the estimator* $\hat{\beta}_{j,k}$ *defined by (3.2) satisfies*

$$\begin{aligned}\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| \leq\, & w_2 w_1^{-1} \left|\mu n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k}\right| \\ & + w_2 \left|\beta_{j,k}\right| \left|n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i) - \mu^{-1}\right|. \end{aligned} \tag{3.3}$$

The inequality (3.3) holds for $\phi_{j,k}$ instead of $\psi_{j,k}$ and, a fortiori, $\hat{\alpha}_{j,k}$ instead of $\hat{\beta}_{j,k}$ and $\alpha_{j,k}$ instead of $\beta_{j,k}$.

In addition to the inequality (3.3), the estimators $(\hat{\beta}_{j,k})_{j,k}$ defined by (3.2) satisfy several specific probability inequalities. Two of them will be at the heart of the proof of the main result. Further details are given in Section 4 below.

4 MAIN RESULT

Theorem 4.1 below determines the rates of convergence achieved by the ($\mathbb{L}^p$ version of the) BlockShrink estimator under the $\mathbb{L}^p$ risk over Besov balls.

Theorem 4.1 *Let us consider the biased density model described in the second paragraph of Section* 1*. Suppose that there exists a known constant* f_2 *such that, for any* $x \in [0,1]$,

$$f(x) \leq f_2.$$

We recall that there exist two constants w_1 *and* w_2 *such that, for any* $x \in [0,1]$, *we have* $0 < w_1 \leq w(x) \leq w_2 < \infty$. *Let* $p \in [1, \infty[$. *Let us consider the BlockShrink estimator* $\hat{f}_n$ *defined by (3.1) with a large enough threshold* *constant d. Then there exists a constant $C > 0$ such that, for any $\pi \in [1, \infty]$, $r \in [1, \infty]$, $s \in (1/\pi, \infty)$ and n large enough, we have*

$$\sup_{f \in B^s_{\pi,r}(M)} \mathbb{E}\left(\int_0^1 \left| \hat{f}_n(x) - f(x) \right|^p dx \right) \leq C\varphi_n,$$

where φ_n is defined by

$$\varphi_n = \begin{cases} n^{-\alpha_1 p} (\log n)^{\alpha_1 p 1_{\{p > \pi\}}}, & \textit{when} \quad \epsilon > 0, \\ (n^{-1} \log n)^{\alpha_2 p} (\log n)^{(p - \pi/r) + 1_{\{\epsilon = 0\}}}, & \textit{when} \quad \epsilon \leq 0, \end{cases}$$

with $\alpha_1 = s/(2s+1)$, $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$ and $\epsilon = \pi s + 2^{-1}(\pi - p)$.

Now, let us discuss the optimality of the rates φ_n described in Theorem 4.1 above. Using standard lower bound techniques, we can prove that there exists a constant $c > 0$ such that

$$\inf_{\tilde{f}} \sup_{f \in B^s_{\pi,r}(M)} \mathbb{E}\left(\int_0^1 \left| \tilde{f}(x) - f(x) \right|^p dx \right) \geq c\varphi_n^*,$$

where $\inf_{\tilde{f}}$ denotes the infimum over all the possible estimators $\tilde{f}$ of f and φ_n^* is defined by

$$\varphi_n^* = \begin{cases} n^{-\alpha_1 p}, & \text{when} \quad \epsilon > 0, \\ (n^{-1} \log n)^{\alpha_2 p} & \text{when} \quad \epsilon \leq 0, \end{cases}$$

with $\alpha_1 = s/(2s+1)$, $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$ and $\epsilon = \pi s + 2^{-1}(\pi - p)$. Thanks to the assumptions of boundedness made on w, the proof is similar to the proof of the lower bound for the standard density estimation problem (i.e., when we observe the i.i.d. sample $X_1, ..., X_n$ from a distribution with unknown density function f). Further details can be found in the book of Härdle et al. [11] (Section 10.4).

Thus, the rates of convergence presented in Theorem 4.1 above are minimax except in the cases $\{p > \pi\} \cap \{\epsilon > 0\}$ and $\epsilon = 0$ where there is an extra logarithmic term. Moreover, they are better than those achieved by the conventional term-by-term thresholding estimators (hard, soft,...). The main difference is for the case $\{\pi \geq p\}$ where there is no extra logarithmic term.

As mentioned in Section 1, if we adopt the same minimax framework than Efromovich [10] and Brunel et al. [1], i.e. $p = 2$ and $\pi = r = 2$ or $\pi = 2$, $r = \infty$, then our estimator has similar asymptotic minimax performances to the Efromovich-Pinsker adaptive estimator developed by Efromovich [10] and the penalized projection density estimator proposed by Brunel et al. [1].

Notice that, for the classic case $w(x) = 1$ (and, a fortiori, $\hat{\mu} = \mu = 1$) and $p = 2$, Theorem 4.1 has been established by Chicken and Cai [5].

Thanks to Theorem 4.2 of Chesneau [4], the proof of Theorem 4.1 is an immediate consequence of Propositions 1 and 2 below. These propositions show that the estimators $(\hat{\beta}_{j,k})_{j,k}$ defined by (3.2) satisfy a standard moment inequality and a specific concentration inequality.

Proposition 1 *Let $p \geq 2$. Suppose that the assumptions of Theorem 4.1 are satisfied. Then there exists a constant $C > 0$ such that, for any $j \in \{j_1, ..., j_2\}$, any $k \in \{0, ..., 2^j - 1\}$ and n large enough, the estimator $\hat{\beta}_{j,k}$ defined by (3.2) satisfies the following moment inequality*

$$\mathbb{E}\left(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^{2p}\right) \leq Cn^{-p}. \tag{4.1}$$

The inequality (4.1) holds for $\hat{\alpha}_{j,k}$ instead of $\hat{\beta}_{j,k}$ and $\alpha_{j,k}$ instead of $\beta_{j,k}$.

The proof of Proposition 1 uses Lemma 3.1 and some moment inequalities as the Rosenthal inequality (see Petrov [15]).

Proposition 2 *Let $p \geq 2$. Suppose that the assumptions of Theorem 4.1 are satisfied. Then there exists a constant $d_* > 0$ such that, for any $j \in \{j_1, ..., j_2\}$, any $K \in \mathcal{A}_j$ and n large enough, the estimators $(\hat{\beta}_{j,k})_{k \in U_{j,K}}$ defined by (3.2) satisfy the following concentration inequality*

$$\mathbb{P}\left(\left(\sum_{k \in U_{j,K}} \left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^{p}\right)^{1/p} \geq d_* 2^{-1} n^{-1/2} (\log n)^{1/2}\right) \leq n^{-p}.$$

The proof of Proposition 2 uses Lemma 3.1 and some concentration inequalities as the Talagrand inequality (see Talagrand [18]) and the Bernstein inequality (see Petrov [15]).

In Propositions 1 and 2 above, we have only considered the case $p \geq 2$. Indeed, if the moment inequality and the concentration inequality are satisfied with $p \geq 2$, then they are satisfied for $p \in [1, 2]$. This is a consequence of the Hölder inequality for the moment inequality. For the concentration inequality, this is a consequence of the following inequality of the l_p norm: for any sequence $(a_i)_{i \in \mathbb{N}^*}$, any $m \in \mathbb{N}^*$ and any $p \in [1, 2]$, we have $\left(m^{-1} \sum_{i=1}^{m} |a_i|^p\right)^{1/p} \leq \left(m^{-1} \sum_{i=1}^{m} (a_i)^2\right)^{1/2}$.

Now, let us discuss the choice of the thresholding constant d. From a theoretical point of view, it is difficult to determine the exact minimum value of d such that $\hat{f}_n$ achieves the rates of convergence exhibited in Theorem 4.1. In fact, Theorem 4.1 holds for $d \geq d_*$ where d_* refers to the constant of Proposition 2 above.

5 PROOFS

In this section, $(C_i)_{i=1,...,12}$ denote positive constants. They take different values in each proof. They are independent of f, μ and n.

PROOF OF LEMMA 3.1. For any $j \in \{j_1, ..., j_2\}$ and any $k \in \{0, ..., 2^j - 1\}$, we have the following decomposition

$$\begin{aligned}\hat{\beta}_{j,k} - \beta_{j,k} &= \hat{\mu} n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k} \\ &= \hat{\mu}\mu^{-1} \left(\mu n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k} \right) + \hat{\mu}\beta_{j,k}(\mu^{-1} - \hat{\mu}^{-1}).\end{aligned}$$

It follows from the triangular inequality that

$$\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| \leq \left|\hat{\mu}\mu^{-1}\right| \left|\mu n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k}\right| + |\hat{\mu}||\beta_{j,k}| \left|\hat{\mu}^{-1} - \mu^{-1}\right|.$$

Since $w(x) \leq w_2$ for any $x \in [0,1]$, we have $\hat{\mu} \leq w_2$. Since $w_1 \leq w(x)$ for any $x \in [0,1]$, we have $\mu = \int_0^1 f(y)w(y)dy \geq w_1 \int_0^1 f(y)dy = w_1$. Therefore

$$\begin{aligned}\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| &\leq w_2 w_1^{-1} \left|\mu n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k}\right| \\ &+ w_2 \left|\beta_{j,k}\right| \left|n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i) - \mu^{-1}\right|.\end{aligned}$$

Lemma 3.1 is proved.

PROOF OF PROPOSITION 1. Let $p \geq 2$. Using Lemma 3.1 and the elementary inequality

$$(|x+y|)^a \leq 2^{a-1}(|x|^a + |y|^a), \qquad x, y \in \mathbb{R}, \qquad a \geq 1, \tag{5.1}$$

for any $j \in \{j_1, ..., j_2\}$ and any $k \in \{0, ..., 2^j - 1\}$, we obtain

$$\mathbb{E}\left(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}\right) \leq c_{j,k}\left(S_{j,k} + T_{j,k}\right),$$

where $c_{j,k} = 2^{p-1}\left(\max\left(w_2 w_1^{-1},\ w_2|\beta_{j,k}|\right)\right)^{2p}$,

$$S_{j,k} = \mathbb{E}\left(\left|\mu n^{-1} \sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k}\right|^{2p}\right)$$

and

$$T_{j,k} = \mathbb{E}\left(\left|n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i) - \mu^{-1}\right|^{2p}\right).$$

Let us investigate the upper bounds for $c_{j,k}$, $S_{j,k}$ and $T_{j,k}$, in turn.

The upper bound for $c_{j,k}$. Using the Cauchy-Schwarz inequality and the fact that $f(x) \le f_2$ for any $x \in [0,1]$ and $\int_0^1 (\psi_{j,k}(x))^2 dx = 1$, we obtain

$$|\beta_{j,k}| \le \left(\int_0^1 (f(x))^2 dx\right)^{1/2} \left(\int_0^1 (\psi_{j,k}(x))^2 dx\right)^{1/2} \le f_2. \tag{5.2}$$

It follows that

$$c_{j,k} \le 2^{p-1}\left(\max\left(w_2 w_1^{-1},\ w_2 f_2\right)\right)^{2p}. \tag{5.3}$$

The obtained upper bound does not depend on n, f, and the parameters j and k.

The upper bound for $S_{j,k}$. For any $i \in \{1, ..., n\}$, let us set $D_i = \mu w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k}$. Clearly, $D_1, ..., D_n$ are i.i.d. random variables such that

$$\mathbb{E}(D_1) = \int_0^1 \mu w^{-1}(x)\psi_{j,k}(x)\left(f(x)w(x)\mu^{-1}\right) dx - \beta_{j,k} = 0.$$

Now, in order to apply the Rosenthal inequality (see Petrov [15]), let us investigate the upper bound for the moments of $|D_1|$. Using the elementary inequality (5.1) and the fact that $|\beta_{j,k}| \le f_2$ (see (5.2)), for any $a \ge 2$, we have

$$\begin{aligned}\mathbb{E}(|D_1|^a) &\le 2^{a-1}\left(\mathbb{E}\left(\left|\mu w^{-1}(Z_1)\psi_{j,k}(Z_1)\right|^a\right) + |\beta_{j,k}|^a\right)\\ &\le 2^{a-1}\left(\mathbb{E}\left(\left|\mu w^{-1}(Z_1)\psi_{j,k}(Z_1)\right|^a\right) + f_2^a\right).\end{aligned} \tag{5.4}$$

Using the inequalities $w_1 \le w(x)$ for any $x \in [0,1]$, $\mu \le w_2$, $|\beta_{j,k}| \le f_2$ (see (5.2)) and $|\psi_{j,k}(Z_i)| \le \sup_{x\in[0,1]} |\psi_{j,k}(x)| \le 2^{j/2} \sup_{x\in[0,1]} |\psi(x)|$, it comes

$$\begin{aligned}&\mathbb{E}\left(\left|\mu w^{-1}(Z_1)\psi_{j,k}(Z_1)\right|^a\right)\\ &\le w_2^{a-1} w_1^{-(a-1)} 2^{j(a-2)/2}\left(\sup_{x\in[0,1]} |\psi(x)|\right)^{a-2} \mathbb{E}\left(\mu w^{-1}(Z_1)\left(\psi_{j,k}(Z_1)\right)^2\right).\end{aligned} \tag{5.5}$$

Since $f(x) \le f_2$ for any $x \in [0,1]$, we have

$$\begin{aligned}\mathbb{E}\left(\mu w^{-1}(Z_1)\left(\psi_{j,k}(Z_1)\right)^2\right) &= \int_0^1 \mu w^{-1}(x)(\psi_{j,k}(x))^2\left(\mu^{-1}w(x)f(x)\right)dx \\ &= \int_0^1 (\psi_{j,k}(x))^2 f(x)dx \\ &\leq f_2\int_0^1 (\psi_{j,k}(x))^2 dx = f_2. \qquad (5.6)\end{aligned}$$

By putting the inequalities (5.5) and (5.6) together and using the definition of the integer j_2, for any $j \in \{j_1, ..., j_2\}$, we have

$$\begin{aligned}\mathbb{E}\left(\left|\mu w^{-1}(Z_1)\psi_{j,k}(Z_1)\right|^a\right) &\leq w_2^{a-1}w_1^{-(a-1)}\left(\sup_{x\in[0,1]}|\psi(x)|\right)^{a-2} f_2 2^{j_2(a-2)/2} \\ &\leq C_1 n^{a/2-1}.\end{aligned}$$

This, with the inequality (5.4), yields the existence of a constant C_2 such that

$$\mathbb{E}(|D_1|^a) \leq 2^{a-1}\left(C_1 n^{a/2-1} + f_2^a\right) \leq C_2 n^{a/2-1}. \qquad (5.7)$$

The Rosenthal inequality and the inequality (5.7) taken with the values $a = 2p$ and $a = 2$ imply

$$\begin{aligned}S_{j,k} &= \mathbb{E}\left(\left|n^{-1}\sum_{i=1}^n D_i\right|^{2p}\right) \leq C_3\left(n^{1-2p}\mathbb{E}\left(|D_1|^{2p}\right) + n^{-p}\left(\mathbb{E}\left((D_1)^2\right)\right)^p\right) \\ &\leq C_4\left(n^{1-2p}n^{p-1} + n^{-p}\right) \leq C_5 n^{-p}. \qquad (5.8)\end{aligned}$$

The upper bound for $T_{j,k}$. For any $i \in \{1, ..., n\}$, let us set $G_i = w^{-1}(Z_i) - \mu^{-1}$. Clearly, $G_1, ..., G_n$ are i.i.d. such that

$$\mathbb{E}(G_1) = \int_0^1 w^{-1}(x)\left(f(x)w(x)\mu^{-1}\right)dx - \mu^{-1} = 0.$$

Now, in order to apply the Rosenthal inequality (see Petrov [15]), let us investigate the upper bound for the moments of $|G_1|$. Using the elementary inequality (5.1), the inequalities $w_1 \leq w(x)$ for any $x \in [0,1]$ and $\mu^{-1} \leq w_1^{-1}$, for any $a \geq 2$, we have

$$\mathbb{E}(|G_1|^a) \leq 2^{a-1}\left(\mathbb{E}\left(\left(w^{-1}(Z_1)\right)^a\right) + \mu^{-a}\right) \leq 2^a w_1^{-a}. \qquad (5.9)$$

The Rosenthal inequality and the inequality (5.9) taken with the values $a = 2p$ and $a = 2$ imply

$$T_{j,k} = \mathbb{E}\left(\left|n^{-1}\sum_{i=1}^{n} G_i\right|^{2p}\right) \le C_6\left(n^{1-2p}\mathbb{E}\left(|G_1|^{2p}\right) + n^{-p}\left(\mathbb{E}\left((G_1)^2\right)\right)^p\right)$$
$$\le C_7 n^{-p}. \tag{5.10}$$

Combining the upper bounds (5.3), (5.8) and (5.10), we prove the existence of a constant C_8 such that

$$\mathbb{E}\left(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}\right) \le c_{j,k}(S_{j,k} + T_{j,k}) \le C_8 n^{-p}.$$

This proved Proposition 1.

PROOF OF PROPOSITION 2. Let $p \ge 2$. Using Lemma 3.1 and the Minkowski inequality for the l_p norm, for any $j \in \{j_1, ..., j_2\}$ and any $K \in A_j$, we have

$$\left(\sum_{k\in U_{j,K}} |\hat{\beta}_{j,k} - \beta_{j,k}|^p\right)^{1/p} \le w_2 w_1^{-1}\left(\sum_{k\in U_{j,K}}\left|\mu n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k}\right|^p\right)^{1/p}$$
$$+ w_2\left|n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i) - \mu^{-1}\right|\left(\sum_{k\in U_{j,K}} |\beta_{j,k}|^p\right)^{1/p}.$$

Using an elementary inequality of the l_p norm with the fact that $p \ge 2$, the orthonormality of the wavelet basis and the fact that $f(x) \le f_2$ for any $x \in [0,1]$, we have

$$\left(\sum_{k\in U_{j,K}} |\beta_{j,k}|^p\right)^{1/p} \le \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^2\right)^{1/2} \le \left(\int_0^1 (f(x))^2\, dx\right)^{1/2} \le f_2.$$

It follows from these inequalities that, for any $\lambda > 0$, we have

$$\mathbb{P}\left(\left(\sum_{k\in U_{j,K}} |\hat{\beta}_{j,k} - \beta_{j,k}|^p\right)^{1/p} \ge \lambda n^{-1/2}(\log n)^{1/2}\right) \le F_{j,K} + Q_{j,K}, \tag{5.11}$$

where

$$F_{j,K} = \mathbb{P}\Bigg(w_2 w_1^{-1}\left(\sum_{k\in U_{j,K}}\left|\mu n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k}\right|^p\right)^{1/p} \ge$$
$$2^{-1}\lambda n^{-1/2}(\log n)^{1/2}\Bigg)$$

and

$$Q_{j,K} = \mathbb{P}\left(w_2 f_2 \left| n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i) - \mu^{-1} \right| \geq 2^{-1}\lambda n^{-1/2}(\log n)^{1/2} \right).$$

Now, let us analyze the upper bounds for $F_{j,K}$ and $Q_{j,K}$, in turn.

The upper bound for $F_{j,K}$. First of all, let us present the Talagrand inequality in Lemma 5.1 below.

Lemma 5.1 (Talagrand [18]) *Let $V_1, ..., V_n$ be i.i.d. random variables and $\epsilon_1, ..., \epsilon_n$ be independent Rademacher variables, also independent of $V_1, ..., V_n$. Let $\mathcal{F}$ be a class of functions uniformly bounded by T. Let $r_n : \mathcal{F} \to R$ be the operator defined by*

$$r_n(h) = n^{-1}\sum_{i=1}^{n} h(V_i) - \mathbb{E}(h(V_1)).$$

Suppose that

$$\sup_{h\in\mathcal{F}} \mathbb{V}\left(h(V_1)\right) \leq v \quad \textit{and} \quad \mathbb{E}\left(\sup_{h\in\mathcal{F}}\sum_{i=1}^{n}\epsilon_i h(V_i)\right) \leq nH.$$

Then, there exist two absolute constants $C_1^* > 0$ and $C_2^* > 0$ such that, for any $t > 0$, we have

$$\mathbb{P}\left(\sup_{h\in\mathcal{F}} r_n(h) \geq t + C_2^* H\right) \leq \exp\left(-nC_1^*\left(t^2 v^{-1} \wedge tT^{-1}\right)\right).$$

In order to apply the Talagrand inequality, let us consider the set $\mathcal{C}_q$ defined by

$$\mathcal{C}_q = \left\{ a = (a_{j,k}) \in \mathbb{Z}^*; \sum_{k\in U_{j,K}} |a_{j,k}|^q \leq 1 \right\}$$

and the functions class $\mathcal{F}$ defined by

$$\mathcal{F} = \left\{ h; \quad h(x) = \mu w^{-1}(x)\sum_{k\in U_{j,K}} a_{j,k}\psi_{j,k}(x), \quad a \in \mathcal{C}_q \right\}.$$

By an argument of duality, we have

$$\begin{aligned}
&\left(\sum_{k\in U_{j,K}} \left| \mu n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k} \right|^p \right)^{1/p} \\
&= \sup_{a\in\mathcal{C}_q}\sum_{k\in U_{j,K}} a_{j,k}\left(\mu n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i) - \beta_{j,k} \right) = \sup_{h\in\mathcal{F}} r_n(h),
\end{aligned}$$

where r_n denotes the function defined in Lemma 5.1. Now, let us evaluate the parameters T, H and v of the Talagrand inequality.

The value of T. Let h be a function in $\mathcal{F}$. Using the inequalities $w_1 \le w(x)$ for any $x \in [0,1]$, $\mu \le w_2$ and the fact that ψ is compactly supported, we obtain

$$|h(x)| \le \mu w^{-1}(x) \sum_{k \in U_{j,K}} |\psi_{j,k}(x)| \le w_2 w_1^{-1} \sum_{k=0}^{2^j-1} |\psi_{j,k}(x)| \le C_2 2^{j/2},$$

where C_2 denotes a constant depending on w_1, w_2, $\sup_{x\in[0,1]} |\psi(x)|$ and the length of the support of ψ.

Hence $T = C_2 2^{j/2}$.

The value of H. Let $\epsilon_1, ..., \epsilon_n$ be independent Rademacher variables independent of $\mathbf{Z} = (Z_1, ..., Z_n)$.

The Hölder inequality for the l_p norm, the Hölder inequality and the definition of $\mathcal{C}_q$ imply

$$\begin{aligned}
&\mathbb{E}\left(\sup_{a\in\mathcal{C}_q} \sum_{i=1}^{n} \sum_{k\in U_{j,K}} a_{j,k}\epsilon_i \mu w^{-1}(Z_i)\psi_{j,k}(Z_i)\right) \\
&\le \sup_{a\in\mathcal{C}_q}\left(\sum_{k\in U_{j,K}} |a_{j,k}|^q\right)^{1/q} \mathbb{E}\left(\left(\sum_{k\in U_{j,K}}\left|\sum_{i=1}^{n}\epsilon_i \mu w^{-1}(Z_i)\psi_{j,k}(Z_i)\right|^p\right)^{1/p}\right) \\
&\le \left(\sum_{k\in U_{j,K}} \mathbb{E}\left(\left|\sum_{i=1}^{n}\epsilon_i \mu w^{-1}(Z_i)\psi_{j,k}(Z_i)\right|^p\right)\right)^{1/p}. \qquad (5.12)
\end{aligned}$$

Since $\epsilon_1, ..., \epsilon_n$ are independent Rademacher variables, also independent of $\mathbf{Z} = (Z_1, ..., Z_n)$, the Khintchine inequality implies the existence of a constant C_3 such that

$$\begin{aligned}
&\mathbb{E}\left(\left|\sum_{i=1}^{n}\epsilon_i \mu w^{-1}(Z_i)\psi_{j,k}(Z_i)\right|^p\right) = \mathbb{E}\left(\mathbb{E}\left(\left|\sum_{i=1}^{n}\epsilon_i \mu w^{-1}(Z_i)\psi_{j,k}(Z_i)\right|^p \;\middle|\; \mathbf{Z}\right)\right) \\
&\le C_3 \mathbb{E}\left(\mathbb{E}\left(\left|\sum_{i=1}^{n}\mu^2 w^{-2}(Z_i)\left(\psi_{j,k}(Z_i)\right)^2\right|^{p/2} \;\middle|\; \mathbf{Z}\right)\right) = C_3 M_{j,k}, \qquad (5.13)
\end{aligned}$$

where

$$M_{j,k} = \mathbb{E}\left(\left|\sum_{i=1}^{n}\mu^2 w^{-2}(Z_i)\left(\psi_{j,k}(Z_i)\right)^2\right|^{p/2}\right).$$

Now, for any $i \in \{1, ..., n\}$, let us set $N_i = \mu^2 w^{-2}(Z_i)\left(\psi_{j,k}(Z_i)\right)^2$. Clearly, the variables $N_1, ..., N_n$ are i.i.d. . The triangular inequality and an elementary inequality of convexity give

$$M_{j,k} \leq \mathbb{E}\left(\left(\left|\sum_{i=1}^{n}(N_i - \mathbb{E}(N_1))\right| + n|\mathbb{E}(N_1)|\right)^{p/2}\right) \leq 2^{p/2-1}(I_{j,k} + J_{j,k}),$$

where

$$I_{j,k} = \mathbb{E}\left(\left|\sum_{i=1}^{n}(N_i - \mathbb{E}(N_1))\right|^{p/2}\right) \quad and \quad J_{j,k} = n^{p/2}\left(\mathbb{E}(N_1)\right)^{p/2}.$$

Let us analyze the upper bounds for $I_{j,k}$ and $J_{j,k}$, in turn.

The upper bound for $I_{j,k}$. The Rosenthal inequality applied to the i.i.d. centered random variables $N_1 - \mathbb{E}(N_1), ..., N_n - \mathbb{E}(N_n)$ and the Hölder inequality imply the existence of two constants C_4 and C_5 such that

$$\begin{aligned} I_{j,k} &\leq C_4\left(n\mathbb{E}(|N_1 - \mathbb{E}(N_1)|^{p/2}) + \left(n\mathbb{E}(|N_1 - \mathbb{E}(N_1)|^2)\right)^{p/4}\right) \\ &\leq C_5\left(n\mathbb{E}(|N_1|^{p/2}) + \left(n\mathbb{E}(|N_1|^2)\right)^{p/4}\right). \end{aligned}$$

Using the inequalities $w_1 \leq w(x)$, $f(x) \leq f_2$ and $|\psi_{j,k}(x)| \leq 2^{j/2}\sup_{x\in[0,1]}|\psi(x)|$ for any $x \in [0,1]$, $\mu \leq w_2$, and the definition of the integer j_2, for any $a \geq 1$ and any $j \in \{j_1, ..., j_2\}$, we have

$$\begin{aligned} \mathbb{E}(|N_1|^a) &= \int_0^1 \mu^{2a}w^{-2a}(x)\left(\psi_{j,k}(x)\right)^{2a}\left(\mu^{-1}f(x)w(x)\right)dx \\ &= \int_0^1 \mu^{2a-1}w^{-2a+1}(x)\left(\psi_{j,k}(x)\right)^{2a}f(x)dx \\ &\leq w_2^{2a-1}w_1^{-2a+1}f_2\left(\sup_{x\in[0,1]}|\psi(x)|\right)^{2a-2}2^{j(a-1)}\int_0^1\left(\psi_{j,k}(x)\right)^2dx \\ &\leq w_2^{2a-1}w_1^{-2a+1}f_2\left(\sup_{x\in[0,1]}|\psi(x)|\right)^{2a-2}2^{j_2(a-1)} \leq C_6n^{a-1}. \end{aligned}$$

Therefore, if we consider the previous inequality with the values $a = p/2$ and $a = 2$, we obtain $I_{j,k} \leq C_7n^{p/2}$.

The upper bound for $J_{j,k}$. Since $\mathbb{E}(N_1) \leq C_6$, we have $J_{j,k} \leq C_6n^{p/2}$.

Combining the obtained upper bounds for $I_{j,k}$ and $J_{j,k}$, we have

$$M_{j,k} \leq 2^{p/2-1}(I_{j,k} + J_{j,k}) \leq C_8n^{p/2}. \tag{5.14}$$

Putting (5.12), (5.13) and (5.14) together, we obtain

$$\mathbb{E}\left(\sup_{a\in\mathcal{C}_q}\sum_{i=1}^{n}\sum_{k\in U_{j,K}} a_{j,k}\epsilon_i\mu w^{-1}(Z_i)\psi_{j,k}(Z_i)\right) \leq C_3^{1/p}\left(\sum_{k\in U_{j,K}} M_{j,k}\right)^{1/p}$$
$$\leq C_9 n^{1/2}L^{1/p}.$$

Hence $H = C_9 n^{-1/2}L^{1/p}$.

The value of v. Using the inequalities $w_1 \leq w(x)$ and $f(x) \leq f_2$ for any $x \in [0,1]$, $\mu \leq w_2$, the fact that the wavelet basis is orthonormal, an inequality of l_p norm combined with the inequality $q = 1+(p-1)^{-1} \leq 2$ and the definition of $\mathcal{C}_q$, we have

$$\begin{aligned}
&\sup_{h\in\mathcal{F}}\mathbb{V}\left(h(Z_1)\right)\\
&\leq \sup_{a\in\mathcal{C}_q}\mathbb{E}\left(\mu^2 w^{-2}(Z_1)\left(\sum_{k\in U_{j,K}} a_{j,k}\psi_{j,k}(Z_1)\right)^2\right)\\
&= \sup_{a\in\mathcal{C}_q}\left(\int_0^1 \mu^2 w^{-2}(x)\left(\sum_{k\in U_{j,K}} a_{j,k}\psi_{j,k}(x)\right)^2\left(\mu^{-1}f(x)w(x)\right)dx\right)\\
&\leq w_2 w_1^{-1} f_2 \sup_{a\in\mathcal{C}_q}\left(\int_0^1\left(\sum_{k\in U_{j,K}} a_{j,k}\psi_{j,k}(x)\right)^2 dx\right)\\
&= w_2 w_1^{-1} f_2 \sup_{a\in\mathcal{C}_q}\left(\sum_{k\in U_{j,K}}(a_{j,k})^2\right) \leq w_2 w_1^{-1} f_2 \sup_{a\in\mathcal{C}_q}\left(\sum_{k\in U_{j,K}}(a_{j,k})^q\right)^{2/q}\\
&\leq w_2 w_1^{-1} f_2.
\end{aligned}$$

Hence $v = w_2 w_1^{-1} f_2$.

Now, let us notice that, for any $j \in \{j_1, ..., j_2\}$, we have $n2^j \leq n2^{j_2} \leq 2n^2(\log n)^{-1}$. Therefore, if we chose $t = 4^{-1}\lambda^* n^{-1/2}(\log n)^{1/2}$ with $\lambda^* = \lambda w_2^{-1}w_1$, then we have

$$\left(t^2v^{-1}\wedge tT^{-1}\right) \geq C_{10}\left(\lambda^2(n^{-1}\log n)\wedge\lambda n^{-1/2}2^{-j/2}(\log n)^{1/2}\right) \geq C_{11}\lambda^2 n^{-1}\log n.$$

Since $(\log n)^{1/2} \leq L^{1/p} < 2^{1/p}(\log n)^{1/2}$, we have $H < C_9 2^{1/p}n^{-1/2}(\log n)^{1/2}$. Therefore, for λ large enough and $t = 4^{-1}\lambda^* n^{-1/2}(\log n)^{1/2}$ with $\lambda^* = \lambda w_2^{-1}w_1$, the Talagrand inequality described in Lemma 5.1 yields

$$
\begin{aligned}
&F_{j,K}\\
&= \mathbb{P}\left(\left(\sum_{k\in U_{j,K}}\left|\mu n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i)-\beta_{j,k}\right|^p\right)^{1/p} \geq 2^{-1}\lambda^* n^{-1/2}(\log n)^{1/2}\right)\\
&\leq \mathbb{P}\Bigg(\left(\sum_{k\in U_{j,K}}\left|\mu n^{-1}\sum_{i=1}^{n} w^{-1}(Z_i)\psi_{j,k}(Z_i)-\beta_{j,k}\right|^p\right)^{1/p} \geq 4^{-1}\lambda^* n^{-1/2}(\log n)^{1/2} +\\
&\quad C_2^* C_9 2^{1/p} n^{-1/2}(\log n)^{1/2}\Bigg)\\
&\leq \mathbb{P}\left(\sup_{h\in\mathcal{F}} r_n(h) \geq t + C_2^* H\right) \leq \exp\left(-nC_1^*\left(t^2 v^{-1}\wedge tT^{-1}\right)\right)\\
&\leq \exp\left(-nC_1^* C_{11}\lambda^2(\log n/n)\right) \leq 2^{-1}n^{-p}. \qquad (5.15)
\end{aligned}
$$

We obtain the desired upper bound for $F_{j,K}$.

The upper bound for $Q_{j,K}$. Let us set $W_i = w^{-1}(Z_i) - \mu^{-1}$. Clearly, $W_1, ..., W_n$ are i.i.d. with

$$
\begin{aligned}
\mathbb{E}(W_1) &= \mathbb{E}\left(w^{-1}(Z_1)\right) - \mu^{-1} = \int_0^1 w^{-1}(x)\left(\mu^{-1}f(x)w(x)\right)dx - \mu\\
&= \mu^{-1}\int_0^1 f(x)dx - \mu = 0.
\end{aligned}
$$

Moreover, since $w_1 \leq w(x)$ for any $x \in [0,1]$ and $w_1 \leq \mu$, the triangular inequality yields, for any $i \in \{1, ..., n\}$,

$$
|W_i| \leq w^{-1}(Z_i) + \mu^{-1} \leq 2w_1^{-1}.
$$

The Bernstein inequality (see Petrov [15]) gives us, for any $l > 0$,

$$
\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^{n} W_i\right| \geq l\right) \leq 2\exp\left(-nl^2\left(2(4w_1^{-2} + 3^{-1}lw_1^{-1})\right)^{-1}\right).
$$

If we apply this inequality with $l = l_* = 2^{-1}w_2^{-1}f_2^{-1}\lambda n^{-1/2}(\log n)^{1/2}$, we prove the existence of a constant C_{12} such that, for λ large enough,

$$
Q_{j,K} = \mathbb{P}\left(\left|n^{-1}\sum_{i=1}^{n} W_i\right| \geq l_*\right) \leq 2\exp\left(-C_{12}\lambda^2(\log n)\right) \leq 2^{-1}n^{-p}. \qquad (5.16)
$$

We have the desired upper bound for $Q_{j,K}$.

It follows from the inequalities (5.11), (5.15) and (5.16) that

$$
\mathbb{P}\left(\left(\sum_{k\in U_{j,K}}|\hat{\beta}_{j,k} - \beta_{j,k}|^p\right)^{1/p} \geq \lambda n^{-1/2}(\log n)^{1/2}\right) \leq F_{j,K} + Q_{j,K} \leq n^{-p}.
$$

Therefore, there exists a constant $d^* > 0$ such that

$$\mathbb{P}\left(\left(\sum_{k\in U_{j,K}} |\hat{\beta}_{j,k} - \beta_{j,k}|^p\right)^{1/p} \geq 2^{-1}d_* n^{-1/2}(\log n)^{1/2}\right) \leq n^{-p}.$$

This proved Proposition 2.

References

[1] Brunel, E., Comte, F., and Guilloux, A. (2007). Nonparametric density estimation in presence of bias and censoring. *TEST (to appear)*.

[2] Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (1993). *Distance Sampling: Estimating Abundance of Biological Populations*. Chapman and Hall, London.

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

[4] Chesneau, C. (2007). Wavelet estimation via block thresholding: A minimax study under the lp risk. *Statistica Sinica (to appear)*.

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

[6] Cohen, A., Daubechies, I., Jawerth, B., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24(1):54–81.

[7] Cox, D. (1969). *Some sampling problems in technology. In New Developments in Survey Sampling (N. L. Johnson and H. Smith, Jr., eds.)*. Wiley, New York, 506-527.

[8] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L^1 View*. Wiley, New York.

[9] Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications*. Springer, New York.

[10] Efromovich, S. (2004). Density estimation for biased data. *J. Statist. Plann. Inference*, 32(3):1137–1161.

[11] Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*. vol. 129 of Lecture Notes in Statistics. Springer-Verlag, New York.

[12] Heckman, J. (1985). *Selection bias and self-selection. In The new Palgrave : A Dictionary of Economics, 287-296*. MacMillan Press, Stockton, New York.

[13] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge.

[14] PATIL, G. P. and RAO, C. R. (1977). *The weighted distributions: A survey of their applications. In Applications of Statistics (P. R. Krishnaiah, ed.) 383-405*. North-Holland, Amsterdam.

[15] PETROV, V. V. (1995). *Limit Theorems of Probability Theory*. Clarendon Press, Oxford.

[16] PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann.Statist.*, 28(1):298–335.

[17] SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[18] TALAGRAND, M. (1994). Sharper bounds for gaussian and empirical processes. *Annals of Probability*, 22:28–76.

[19] TSYBAKOV, A. (2004). *Introduction à l'estimation nonparamétrique*. Springer Verlag.