Numerical performances of a warped wavelet estimation procedure for regression in random design

Christophe Chesneau, Thomas Willer

Numerical performances of a warped wavelet estimation procedure for regression in random design

Chesneau Christophe, Willer Thomas

Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6-Paris 7, 175 rue de Chevaleret, F-75013 Paris, France.

Abstract

The purpose of this paper is to investigate the numerical performances of the hard thresholding procedure introduced by Kerkyacharian and Picard [17] for the non-parametric regression model with random design. That construction adopts a new approach by using a wavelet basis warped with a function depending on the design, which enables to estimate regression functions under mild assumptions on the design. We compare our numerical properties to those obtained for other constructions based on hard wavelet thresholding. The performances are evaluated on numerous simulated data sets covering a broad variety of settings including known and unknown design density models, and also on real data sets.

Keywords and phrases: Nonparametric regression, random design, warped wavelet basis, wavelet thresholding.

AMS: Primary: 62G05, Secondary: 62G20.

1 Motivation

Suppose we observe data $(Y_1, X_1), ..., (Y_n, X_n)$ where $(Y_i)_{i=1,...,n}$ is characterized by the following equation:

$$Y_i = f(X_i) + s\epsilon_i, \qquad i = 1, ..., n, \tag{1}$$

the ϵ_i's are i.i.d centered standard normal variables, s is a fixed noise level, and the X_i's are i.i.d random variables representing the design points with density g. The aim is to recover the unknown function f from the observations. To reach this aim, we propose to focus our attention on reconstruction methods using wavelet analysis.

Wavelet thresholding algorithms are popular methods in curve estimation, as well as in nonparametric estimation problems in general. Their advantages with respect to linear procedures are well known: they provide adaptive estimators enjoying near minimax properties in a wide variety of settings, whereas linear estimators can be far from optimal in many situations as established in a series of paper by Donoho and Johnstone ([8], [9]), and Donoho, Johnstone, Picard and Kerkyacharian ([10], [11], [12] and [13]).

We will concentrate here on the regression problem with non equispaced samples defined above. There exist numerous methods in this setting. Among them let us cite Hall and Turlach [14], Antoniadis, Gregoire and Vial [2], Cai and Brown [4], and Maxim [20]. Compared to standard algorithms, the thresholding is notably more complicated because it has to incorporate the variations of the density of the design. As for general minimax results, a study over various function spaces and for different risks can be found in the books [18] and [26] for non-parametric estimation problems, including regression with random design.

Recently a quite different algorithm was developed by Picard and Kerkyacharian in [17]. The procedure stays very close to the equispaced Donoho and Johnstone's Visushrink procedure, and thus is very simple in its form (preliminary estimators are no longer needed) and in its implementation (the standard uniform threshold suffices). On the other side, the projection is done on an unusual non-orthonormal basis, called warped wavelet basis, so their analytic properties need to be studied to derive the performances of the estimator. Such a basis can be defined as a usual wavelet basis composed with the repartition function $G(y) = \int_0^y g(t)dt$. Some theoretical results, including maxiset properties (see [6] and [16]) were established in their paper. Another important advantage of the warped basis estimator is that it is near optimal in the minimax sense over a large class of function spaces for a wide variety of design densities, not necessarily bounded above and below as generally required by other wavelet estimators. Basically, the condition on the design refers to the Muckenhoupt weights theory introduced in Muckenhoupt [23].

The purpose of this paper is to provide numerical performances for the warped basis procedure and to compare these results to those obtained for other wavelet procedures based on the hard thresholding rules. In many constructions, the first step consists in determining a function $Y(x)$ of the form:

$$Y(x) = \sum_m w_m(x) Y_m$$

where $w_m(x)$ is a sequence of functions suitably chosen. For instance, in Hall and Turlach [14] the w_m's correspond to a polynomial which depends to the variable $(X_i)_{i=1,\dots,n}$. In Cai and Brown [4] (and in Maxim [20]), the w_m's corresponds to scale wavelets warped with G. In Antoniadis, Grégoire and Vial [2], the random design is transformed into equispaced data via a binning method and the w_m's are defined by scale wavelets. In a second step, the function Y is expanded on a standard wavelet basis and a hard thresholding algorithm is performed. In all the techniques described above, the thresholds have similar forms and depend on the quantity $\sup_t \frac{1}{g(t)}$, which corresponds to an upper bound for the variance of the estimated wavelet coefficients. For the sake of conciseness, only the construction developed by Cai and Brown [4] will be considered for the simulations since it appears to be relatively representative of these kinds of methods.

In all the estimation procedures, the nature of the density g plays an important role. First we consider the usual context of known and bounded design density g and we investigate which characteristics of the design mainly affect the behaviour of each estimator. Second we consider the context of unbalanced designs, and we propose an adaptation of Cai and Brown's procedure to allow vanishing densities (the minimax properties of such a procedure are not established then, here we only provide a numerical comparison). Third, we examine the case of unknown densities.

The paper is organized as follows. Section 2 introduces basics on wavelets and the main procedures. Section 3 gives a qualitative comparison of the estimators described below, and section 4 presents the results of the simulation study. Lastly in Section 5 the two procedures are applied to real data sets.

2 Estimation procedures

Let us briefly summarize the basics on wavelets that will be needed in the later sections. Let ϕ and ψ be respectively a scaling function and a wavelet associated to a multiresolution analysis on $\mathbb{R}$. With an appropriate treatment at the boundaries of these functions, any f of $\mathbb{L}^2([0,1])$ can be expanded into a wavelet series as:

$$f(x) = \sum_{j,k\in\Lambda} \beta_{j,k}\psi_{j,k}(x), \qquad \beta_{j,k} = \int_0^1 f(t)\psi_{j,k}(t)dt, \qquad x \in [0,1],$$

where $\Lambda = \{(j,k)|\ -1 \leq j \leq \infty, 0 \leq k \leq 2^j - 1\}$, $\psi_{j,k}(.) = 2^{\frac{j}{2}}\psi(2^j. - k)$, $\phi_{j,k}(.) = 2^{\frac{j}{2}}\phi(2^j. - k)$, and for convenience we have set $\psi_{-1,k}(.) = \phi_{0,k}(.)$. See Cohen, Daubechies and Vial [5] and Meyer [22] for further details on wavelet bases on the unit interval $[0,1]$. For wavelets on the line we refer the reader to Daubechies [7], the books of Meyer [21] and Mallat [19].

2.1 The procedure of Kerkyacharian and Picard

Let us consider the regression problem described in (1). Let us recall that the function G is defined by:

$$G(x) = \mathbb{P}(X_1 \leq x) = \int_0^x g(t)dt, \qquad x \in [0,1].$$

Kerkyacharian and Picard [17] propose a construction where the unknown function is expanded on a warped basis instead of a regular wavelet basis. Proceeding in such a way, the estimates of the coefficients become more natural. Let us briefly describe the construction of this procedure.

In the case of known g, we consider the following estimator:

$$\hat{\beta}^{\dagger}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i\psi_{j,k}(G(X_i)),$$

for (j,k) in the set:

$$\Lambda_n = \{(j,k)| -1 \leq j \leq j_1(n),\ 0 \leq k \leq 2^j - 1\},$$

where $j_1(n)$ is an integer such that $2^{j_1(n)}$ is of the order $\sqrt{\frac{n}{\ln(n)}}$, and we perform a hard thresholding algorithm:

$$\hat{f}^{\dagger}(t) = \sum_{j,k\in\Lambda_n} \hat{\beta}^{\dagger}_{j,k} 1_{\left\{|\hat{\beta}^{\dagger}_{j,k}| \geq \kappa s\sqrt{\frac{2\ln(n)}{n}}\right\}} \psi_{j,k}(G(t)), \quad t \in [0,1],$$

where κ is a large enough constant.

In the sequel, we will refer to this estimator as **estimator** 1.

2.2 The construction of Cai and Brown

Estimator 1 will be compared to a wavelet thresholding procedure based on the construction of Cai and Brown ([4]) in the case where the density of the design g is known and bounded from below. The construction consists in the following three steps:

1. Compute a preliminary estimator $\hat{f}$ as if the data were equispaced by using the scaling function at high resolution level $j_2(n) = \log_2(n)$:
$$\hat{f}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \phi_{j_2,i}(t).$$

2. Warp $\hat{f}$ by the function G and compute the wavelet coefficients of the resulting function:
$$\hat{\beta}^*_{j,k} = \int_0^1 \hat{f}(G(t))\psi_{j,k}(t)dt,$$
for (j,k) in the set:
$$\Omega_n = \{(j,k) | -1 \leq j \leq j_2,\ 0 \leq k \leq 2^j - 1\}.$$

3. Perform a hard thresholding algorithm:
$$\hat{f}^*(t) = \sum_{j,k \in \Omega_n} \hat{\beta}^*_{j,k} 1_{\left\{|\hat{\beta}^*_{j,k}| \geq \kappa s \lambda^*_{j,k} \sqrt{\frac{2\ln(n)}{n}}\right\}} \psi_{j,k}(t), \qquad t \in [0,1], \tag{2}$$
where
$$\lambda^*_{j,k} = \sqrt{\max_{t \in S_{j,k}} \frac{1}{g(t)}},$$
κ is a positive constant, $S_{j,k} = [2^{-j}k, 2^{-j}(k+N)]$ and N is the length of the filter associated to the wavelet.

In the sequel, we will refer to this estimator as **estimator** 2.

Comments. The choice of the threshold is linked to the variance of the $\hat{\beta}^*_{j,k}$. Following the Subsection 3.3 of Cai and Brown [4], we have:
$$nVar(\hat{\beta}^*_{j,k}) = \sum_{i=1}^{n} \left(\int_0^1 \phi_{j_2,k}(G(x))\psi_{j,k}(x)dx \right)^2 \leq \int_{S_{j,k}} \psi^2_{j,k}(t) \frac{1}{g(t)} dt = u^2_{j,k}$$
and the threshold $\lambda^*_{j,k}$ is chosen in such a way that:
$$u^2_{j,k} \leq (\lambda^*_{j,k})^2.$$
In practice $u_{j,k}$ is difficult to compute, that is why $\lambda^*_{j,k}$ is used instead. In a second part of the study we propose a slight modification of estimator 2, by adapting the thresholds so as to allow constructions for vanishing densities too. Let us suppose that $\frac{1}{g}$ belongs to $\mathbb{L}^1([0,1])$ and consider the thresholding procedure (2) in which we remplace $\lambda^*_{j,k}$ by another bound of $u_{j,k}$ defined as:
$$\tilde{\lambda}_{j,k} = \sqrt{2^j \left(\int_{S_{j,k}} \frac{1}{g(t)} dt \right)}.$$
In the sequel, we will refer to this estimator as **estimator** $2'$.

Remark 2.1. *In the simulation study we will consider the case of unknown noise level s. The thresholds are thus modified by replacing s by an estimator:*

$$\hat{s} = \sqrt{\frac{1}{2(n-2)} \sum_{i=2}^{n} (Y_{(i)} - Y_{(i-1)})^2},$$

where each $Y_{(i)}$ refers to the value Y_k such that X_k is the i-th higher coordinate of the vector $(X_j)_{1 \leq j \leq n}$.

3 Preliminary comparison of the two estimators

3.1 Implementation

3.1.1 Thresholding algorithms

Obviously, estimator 2 (and $2'$) is more difficult to implement than estimator 1. It requires each coefficient to be compared to a specific value recalculated for each scale and shift parameter. On the contrary, the thresholds of estimator 1 are simple, and approximations of the coefficients can be computed very easily. Indeed if G is replaced by the empirical distribution function $\hat{G}_n$ in the expression of $\hat{\beta}^{\dagger}_{j,k}$ one obtains the following coefficients:

$$\hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} Y_i \psi_{j,k}(\frac{i}{n}),$$

which can be obtained directly by performing a wavelet decomposition of the vector Y. These coefficients were used in the sequel for estimator 1 instead of coefficients $\hat{\beta}^{\dagger}_{\lambda}$.

The differences are illustrated on figure 1: as a toy example, the wavelet coefficients and the thresholds for a 'Sine' regression function and a 'Sine' design density (see their representations further in this paper) were computed with 2^6 observations and using the Haar basis. The figure represents the estimated detail coefficients and their thresholds from the coarsest ($j = 0$) to the highest ($j = 5$) resolution level, for estimator 1 (top) and estimator 2 (bottom). Estimator 1 needs constant thresholds, whatever the scale and the shift parameter, whereas for estimator 2 the thresholds vary with respect to the density g: the lower g on the interval $[\frac{k}{2^j}, \frac{k+1}{2^j}]$, the higher the threshold.

3.1.2 The case of unknown densities

In the case where g is unknown, we replace G wherever it appears in the construction of estimators 1 and 2 or $2'$ by the empirical distribution function of the X_i's:

$$\hat{G}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}.$$

Then adapting estimator 1 is quite easy: we only need to replace G in the warped basis. In the Cai and Brown procedure: first we have to warp $\hat{f}_{j_2}$ with G_n instead of G in step 1 and secondly we have to replace the thresholds in step 3 by estimators, thus a density estimator of g is also needed. Theorically many techniques are available, including log spline (see Maxim ([20])), Kernel or wavelet methods (see the book of Härdle, Picard, Kerkyacharian and Tsybakov ([15])). In the simulation study we used a method based on binning and wavelet thresholding which will be detailed further.

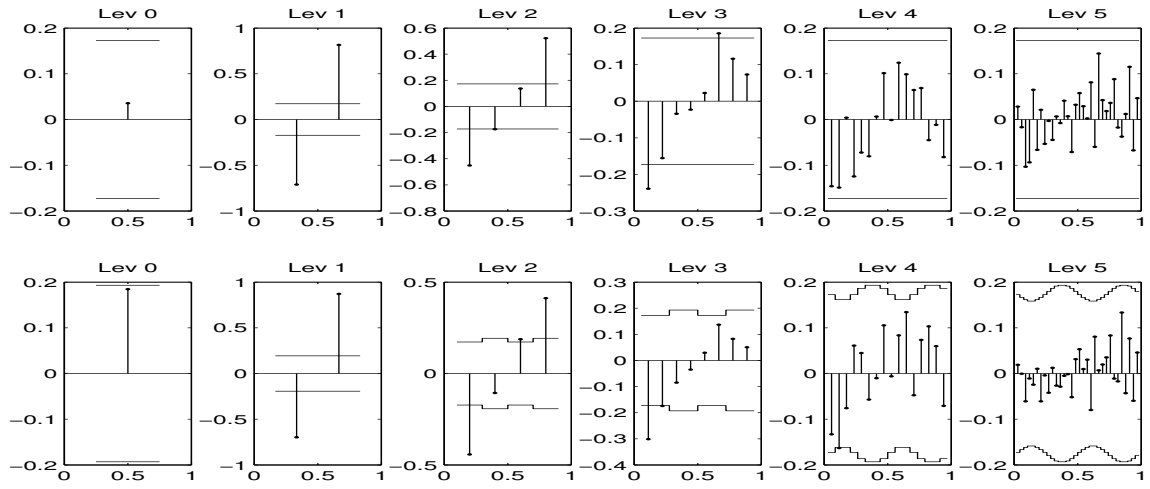

Figure 1: Coefficients (stems) and their corresponding thresholds (stairs) for estimator 1 (top), estimator 2 (bottom) and each level of resolution (from left to right)

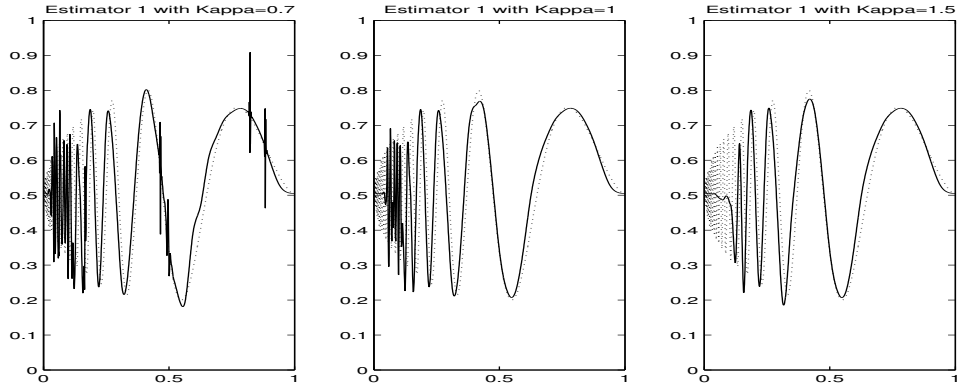

Figure 2: Estimator 1 for the 'Doppler' target (in dots) for three different values of κ (0.7, 1, 1.5)

3.2 Some examples of settings

Before a thorough study in a wide variety of settings, we investigate the behaviours of the estimators in several particular models to highlight their main differences. In this subsection we take:

- number of observations: $n = 1024$,
- root signal noise ratio: $rsnr = 3$,
- wavelet basis: Symlet of order 8.

First we investigate the choice of the constant κ for the thresholds of estimator 1. There is no optimal constant suited to any setting, however choosing $\kappa = 1$ proves efficient in general as can be seen on figure 2: for smaller κ there remains unfiltered noise and for larger κ the first oscillations of the Doppler function are not recovered.

Let us now compare the two estimators. Predictably in most settings close to optimal conditions, i.e for smooth densities close to the uniform, the two estimators behave similarly. For example when the density is a sine function with relatively small amplitude, both estimators behave well as can be seen on figure 3. Secondly, estimator 1

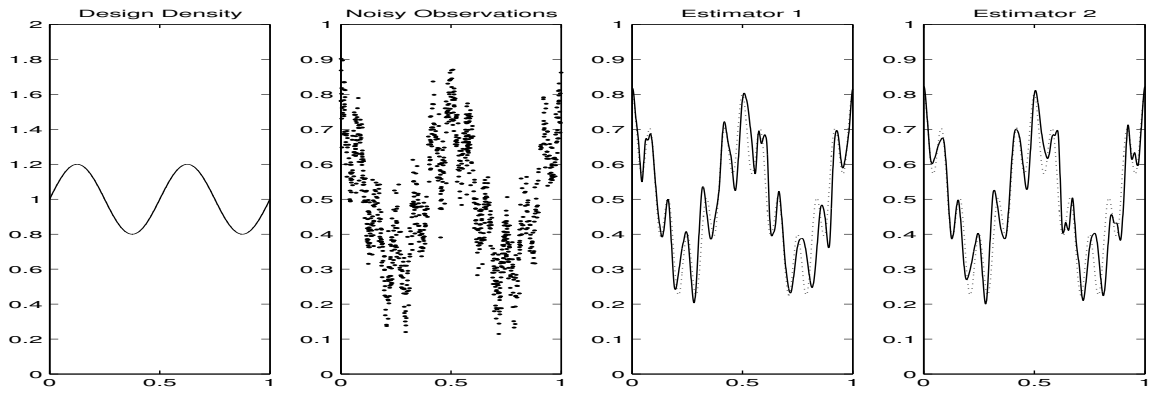

Figure 3: The two estimators of the 'Wave' target (in dots)

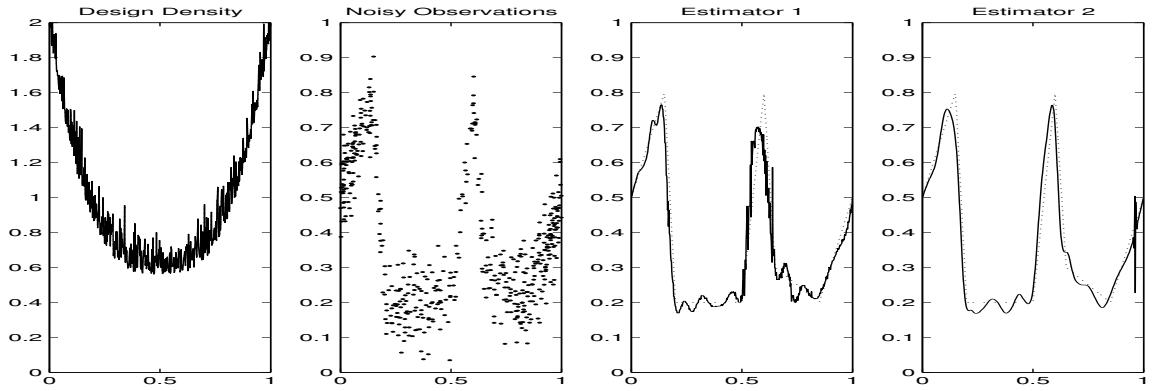

Figure 4: The two estimators of the 'Angles' target (in dots)

presents some defaults when the target function is much smoother than the density of the design. Indeed in this case the warping deteriorates the regularity of the estimator which is visually less pleasant than estimator 2, see for instance figure 4.

Beside that smoothness effect, interesting differences appear when the design is far from uniform, i.e when the distribution of the design points in $[0, 1]$ is very unbalanced. Two deteriorations can then be noticed for estimator 2. First it does not capture as many details of the target function as estimator 1 in the zones where the observations are sparse (figure 5). Secondly it presents artifacts in the high density zones (figure 6). These two problems have the same origin: the wide variations of g provoke disproportionate thresholds, leaving some noise unfiltered when g is too high and on the contrary erasing useful details when g is too low. This can be seen respectively at level 5 in figure 7 and at level 9 in figure 8 where the thresholds associated to the two previous settings are represented.

The simulation study presented in the next section enables to analyse these differences more thoroughly.

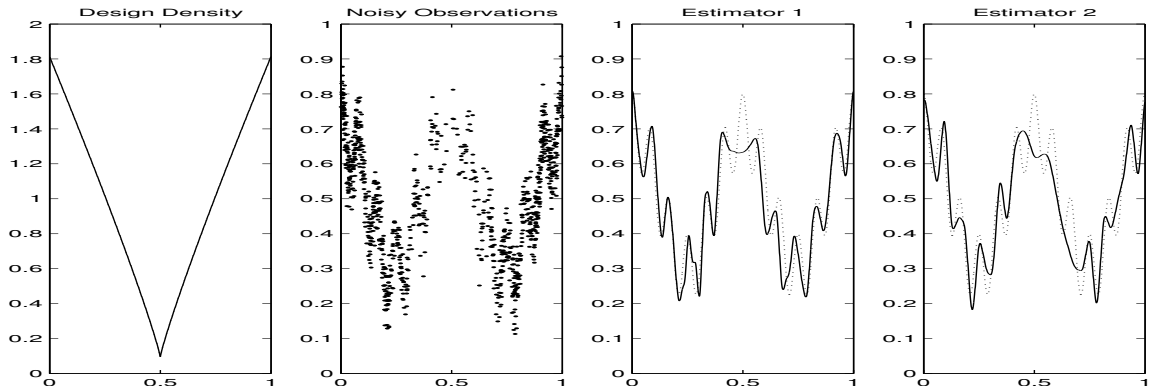

Figure 5: The two estimators of the 'Wave' target (in dots)

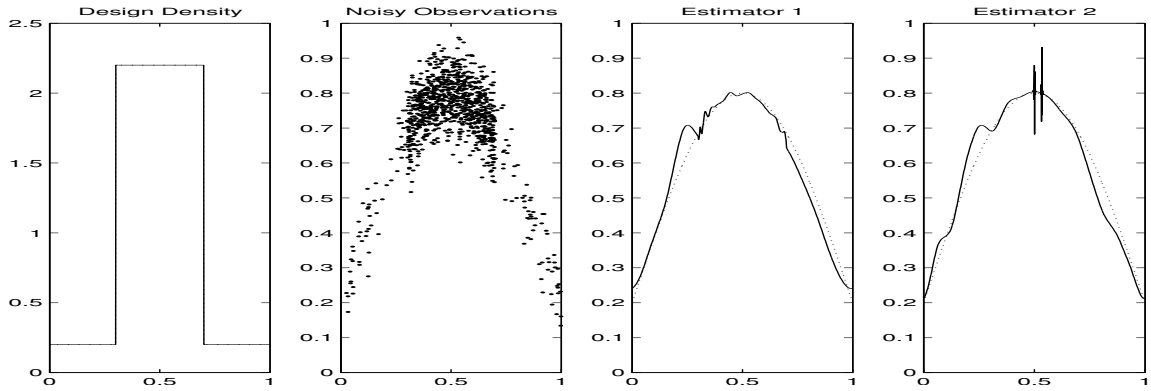

Figure 6: The two estimators of the 'Sine' target (in dots)

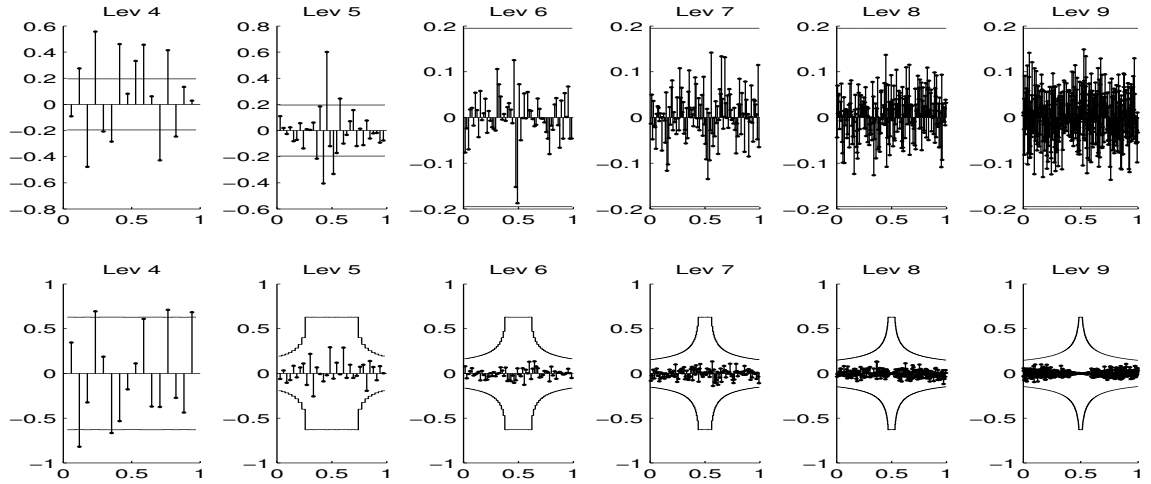

Figure 7: Coefficients (stems) and their corresponding thresholds (stairs) associated to figure 5 for estimator 1 (top), estimator 2 (bottom) and each level of resolution (from left to right)

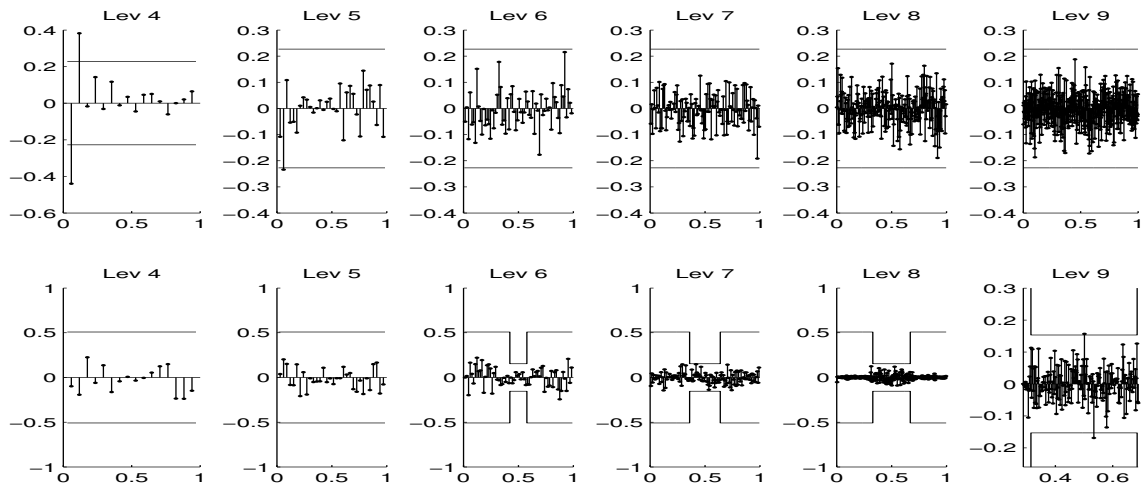

Figure 8: Coefficients (stems) and their corresponding thresholds (stairs) associated to figure 6 for estimator 1 (top), estimator 2 (bottom) and each level of resolution (from left to right)

4 Simulation study

4.1 Description of the simulation

We compare the behaviour of the two estimators for different regression functions and different densities of the design. For each one of these two factors, we used the functions represented on figures 9 and 10. Most of the target functions are borrowed from Antoniadis et al. ([1]), where they are used to highlight differences between linear and non linear estimators. We refer to their paper for the mathematical expressions. As for the densities, there are two groups: the first and second ones are uniform or slightly varying, whereas the next four ones are used to test if the estimators behave well in case of one or numerous holes in the density, i.e of zones where one has hardly any observation of the unknown function. The mathematical expressions are given in the appendix. Notice that all these functions are bounded from below.

In addition the effects of n, of the root signal to noise ratio (denoted by $rsnr$) and of the choice of the wavelet basis are examined. A series of results are given in the appendix for samples with $n \in \{2^9, 2^{10}, 2^{11}, 2^{12}\}$, $rsnr = 1$ (high noise) or $rsnr = 7$ (small noise), and the wavelet basis is the Symlet of order 8 or the Coiflet of order 3 (as in [1]).

The quality of each estimator was evaluated by computing approximations of the mean integrated error (L1), the root mean integrated square error (RMSE) and the maximum deviation (MXDV). The criterium $MXDV$ reflects the amplitude of localized errors in the estimation, whereas $L1$ reflects the mean quality of the estimation along the whole domain of definition of the target function. These quantities were estimated in the following way:

- L1 is computed as the average over 100 runs of $\frac{1}{n}\sum_{i=1}^{n} |f(\frac{i}{n}) - \hat{f}(\frac{i}{n})|$.
- RMSE is computed as the average over 100 runs of $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(\frac{i}{n}) - \hat{f}(\frac{i}{n}))^2}$.

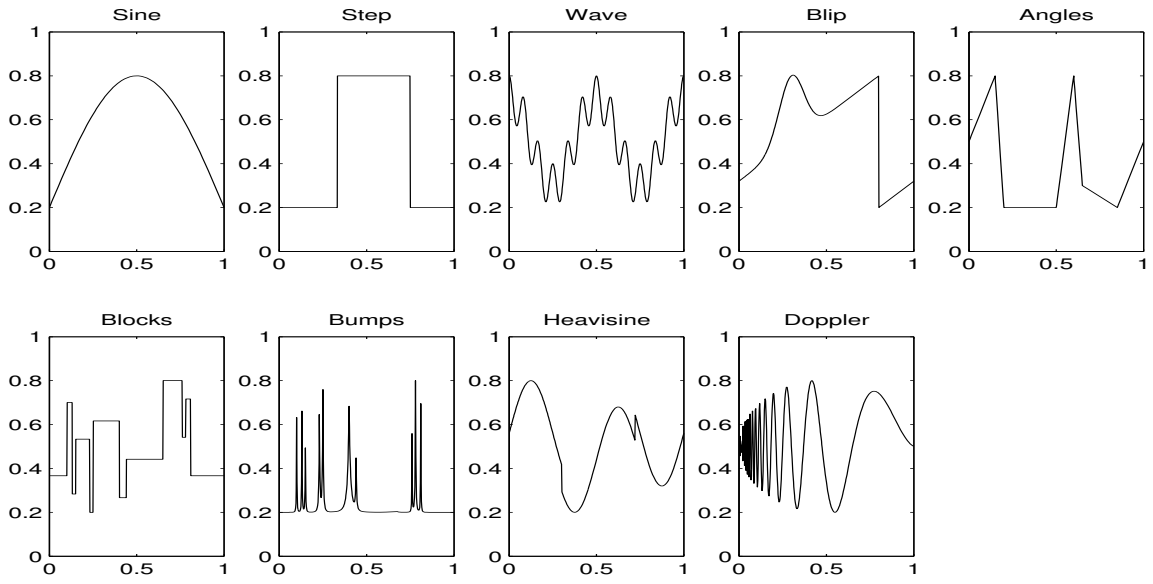

Figure 9: Target functions

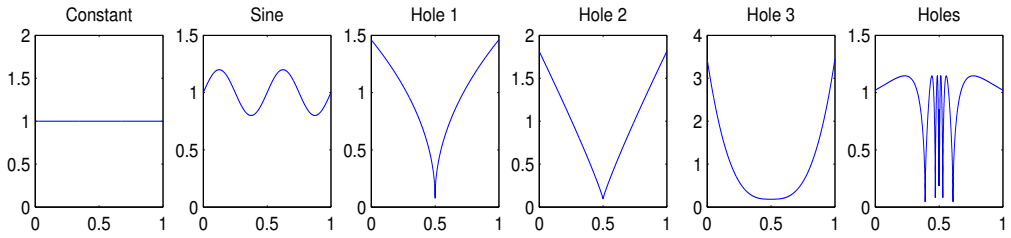

Figure 10: Densities of the design

- MXDV is computed as the average over 100 runs of $\max_{1\leq i\leq n} |f(\frac{i}{n}) - \hat{f}(\frac{i}{n})|$.

In each run, the random variables X and ϵ were simulated independently of their values in the other runs.

4.2 Results for known and bounded densities

Let us first focus on a setting with small sample $n = 512$, high noise $rsnr = 1$ and a Symlet basis (figure 15). The performance of estimator 1 relatively to estimator 2 for the three criteria is given in the first column of the figure. For densities close to the uniform, the two estimators have similar performances, except that estimator 2 has a better MXDV for the Sine and Heavisine target function and the Sine density design.

On the contrary differences appear for other densities. Especially considering the L_1 loss, estimator 1 is better for moderately complicated targets, such as Wave or Blip, because it recovers details that estimator 2 ignores. This advantage is all the more significant as the the hole is wide ('Hole2' and 'Hole3' densities). For more complicated targets (Doppler or Blocks) neither estimator captures the details very well in high noise, so their performances are equivalent.

For larger samples (figure 17) the advantage of estimator 1 in case of holes in the design density grows more and more obvious. When $n = 4096$ estimator 2 is outperformed whatever the rsnr, the design and the target (except Heavisine). This is particularly true for 'Hole1' and 'Holes' densities, where the thresholding rule of estimator 2 is probably inadapted.

In the small noise settings, the previous comparison remains valid but the advantage of estimator 1 is generally less significant. Similarly replacing the Symlet by the Coiflet basis does not change the advantage of estimator 1 over estimator 2, but this one tends to be reduced in most settings. That may come from the fact that the Coiflet scaling functions have better approximation properties than the Symlets, and thus the detail coefficients and the thresholds play a minor role in both methods.

4.3 Results for known and vanishing densities

In this part the simulation study is performed with the same model parameters as before, except for three densities, namely 'Hole1' 'Hole2' and 'Holes', which are allowed to vanish (the new expressions are given in the appendix). The constant κ of estimator $2'$ was fitted in practice such that the estimator behaves well whenever the design density is uniform, as it was done for estimator 1.

The results are the following. For the first two densities of figure 10, estimator $2'$ behaves in a similar way as estimator 1 and estimator 2. However for vanishing densities, estimator $2'$ corrects the main oversmoothing default of estimator 2. Indeed the thresholding method is less rough than the one used in the previous section, so the estimator behaves better now in some of the settings investigated earlier as can be seen in figure 11.

Nevertheless the values of the quality criteria (see figures 18) show that estimator 1 generally remains better than estimator $2'$, even if its advantage is much smaller for

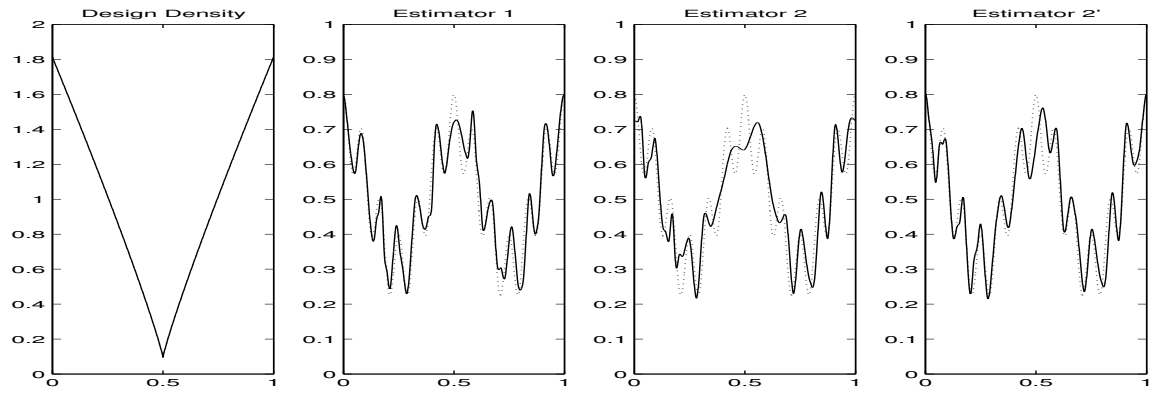

Figure 11: Estimators 1, 2 and 2′ for the Wave target (in dots), $n = 1024$ and $rsnr = 3$

densities with narrow holes such as 'Hole1' and 'Holes' in the case of high noise (first column of figures 18 versus first column of figure 17). So even with thresholds sharper than before, estimator 2′ does not manage to recover the details of the target functions as well as estimator 1, especially when the observations are sparse in a wide range of abscisses (for 'Hole2' and 'Hole3' densities).

4.4 Results for unknown densities

In this context, a density estimator is necessary to implement estimator 2′. In Wavelab a procedure is available, which consists first in computing an histogram H_n of the data:

$$H_n(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}(t_{j+1} - t_j)^{-1}1_{T_j}(x)1_{T_j}(X_i),$$

where $t_1, ..., t_m$ is an equispaced sample of $[0, 1]$, $T_i = [t_i, t_{i+1}[$ and m is a large integer. Then a wavelet thresholding algorithm is performed on H_n.

A common choice for m is $m = [n/l]$, where the width l of the steps can be seen as a bandwidth parameter chosen according to the smoothness of the underlying density. However a fixed $l = 16$ leads to reasonable estimators so this choice was adopted in the sequel.

Predictably, the estimation errors in the density cause both estimators to deteriorate. The warping makes estimator 1 unsmooth in domains where the observations are sparse, and the thresholding generally leaves some noise unfiltered for estimator 2. For example in figure 12 artifacts appear even though the density is relatively well estimated. Some of them could have been erased, had the true thresholds been taken into account.

Analysing the three quality criteria for high noise (first column of figure 16), estimator 1 is now clearly stronger than estimator 2 whatever the density, especially for the maximum deviation. However for small noise the performances are close. Estimator 1 is more robust with respect to the lack of knowledge of the design density.

Conclusion. The two procedures have similar performances for smooth and homogenous design densities. For vanishing densities, estimator 2′ is better than estimator 2 but both are generally outperformed by estimator 1, and this ranking is even clearer if unknown densities are taken into account.

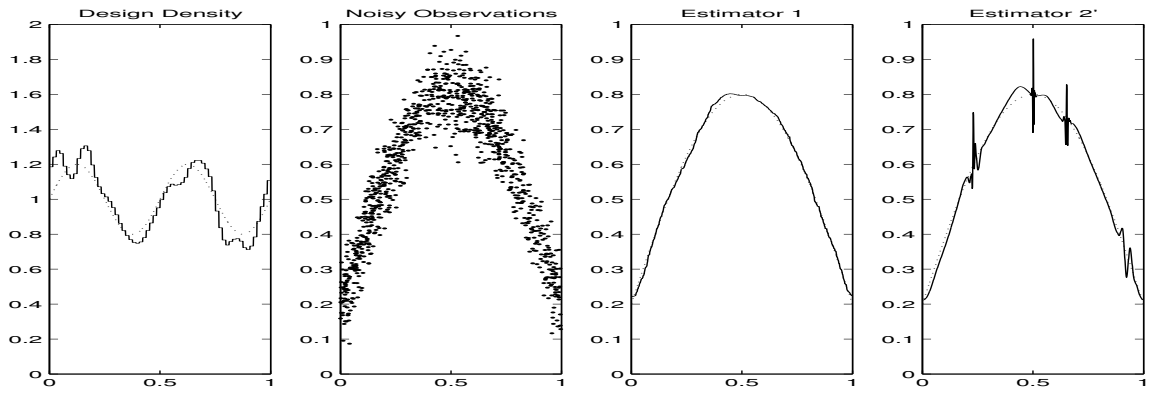

Figure 12: Density estimator, noisy data and the two estimators for the Sine target (in dots)

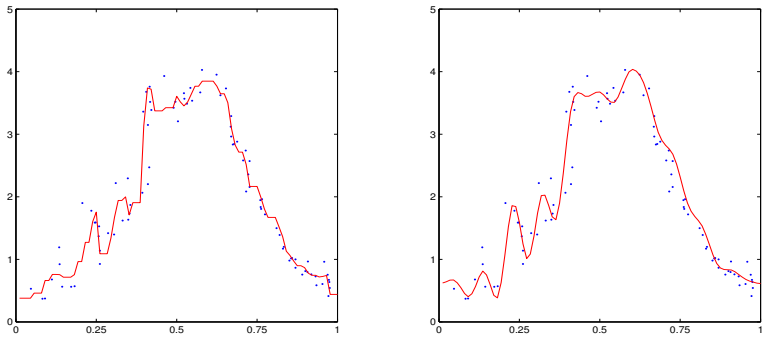

Figure 13: Ethanol data (dots) with estimator 1 (left) and estimator 2 (right)

5 Applications to real data sets

5.1 Ethanol data

We investigate the performance of the two wavelet thresholding procedures when applied to the ethanol data introduced by Brinkman [3]. The data consists of 88 measurements from an experiment where ethanol was burned in a one-cylinder automobile engine. The concentration of the total amount of nitric oxide and nitrogen dionide (y-axis) is related to the "equivalence ratio" (x-axis), a measure of the richness of the air ethanol mixture. To fit the data to our model, the range of the x-axis variable is linearly shifted to $[0, 1]$.

The two procedures considered here yield satisfactory results, compared to the numerous other estimators applied to this dataset (eg [2]). As can be seen on figure 13, estimator 1 seems slightly better than estimator 2 for $x \in [0.7, 1]$, but globally both capture the variations in the data quite precisely. We can remark that estimator 1 is a bit unsmooth because of the warping with $\hat{G}_n$. If we wish to obtain a visually more pleasant result, an alternative is to use a smoother estimator of G.

So as to quantify the performances of the the two procedures, we use a criterium developed by Nason ([24]) adapted to the regression model with random design. This approach consists in evaluating the following estimator of the mean square error:

$$\hat{M} = n^{-1} \sum_{i=1}^{n} (\hat{f}^{-i}(X_i) - Y_i)^2$$

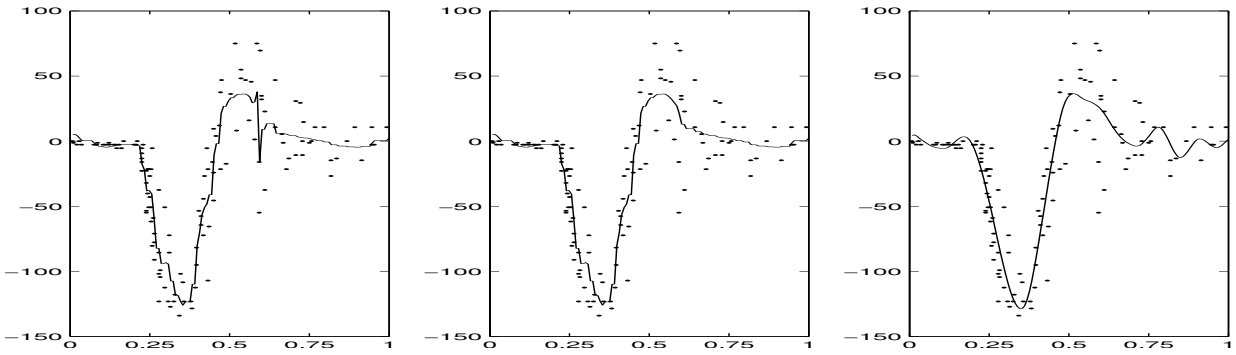

Figure 14: Motorcycle data (dots) with estimator 1 with $\kappa = 1$ (left), $\kappa = 1.06$ (middle) and estimator 2 (right)

where $\hat{f}^{-i}$ is each of our procedures contructed from all the data except the i-th observation (Y_i, X_i). As in Nason [24], we compute the values of $\hat{M}$ for various choices of the wavelet basis and of the coarsest level j_0. The results are similar to the ones obtained in his paper. For example the performances of estimator 1 are given in table 1, the best result being 103.9 versus 98 in Nason [24].

	Sym7	**Sym8**	**Sym9**	**Sym10**
$j_0 = 4$	156.8	145.1	119.0	127.8
$j_0 = 5$	167.5	151.4	125.6	136.8
$j_0 = 6$	140.8	139.0	103.9	126.8

Table 1: Values of $\hat{M}$ ($\times 1000$) of estimator 1 for various choices of the wavelet basis and of the coarsest level j_0

5.2 Motorcycle acceleration data

Lastly we apply our procedures to the motorcycle acceleration data considered in Silverman [25]. These 133 observations are taken from a crash test and show the acceleration of a motorcyclist's head. The explanatory variable is time (rescaled to the unit interval) and the dependent variable is the head acceleration (in g).

As can be seen on Figure 14, the data are heteroscedastic with an increasing variance with respect to the time. If we apply blindly the two procedures, estimator 1 exhibits a high frequency feature due to a large variance. This can be corrected by slightly increasing the level of the threshold (i.e by setting $\kappa = 1.06$ instead of $\kappa = 1$). We can notice that estimator 2 seems to be less precise than estimator 1.

6 Appendix

6.1 Target functions and densities

Target functions. The target functions in figure 9 have the following expressions:

- Sine: $f(x) = 0.2 + 0.6\sin(\pi x)$,
- Step: $f(x) = 0.2 + 0.6I\{1/3 < x < 3/4\}$,

- Wave: $f(x) = 0.5 + 0.2\cos(4\pi x) + 0.1\cos(24\pi x)$,
- Blip: $f(x) = (0.32 + 0.6x + 0.3\exp(-100(x-0.3)^2)I\{0 \leq x \leq 0.8\} + (-0.28 + 0.6x + 0.3\exp(-100(x-1.3)^2)I\{0.8 < x \leq 1\}$,
- Angles: $f(x) = (2x+0.5)I\{0 \leq x \leq 0.15\} + (-12(x-0.15)+0.8)I\{0.15 < x \leq 0.2\} + 0.2I\{0.2 < x \leq 0.5\} + (6(x-0.5)+0.2)I\{0.5 < x \leq 0.6\} + (-10(x-0.6)+0.8)I\{0.6 < x \leq 0.65\} + (-0.5(x-0.65)+0.3)I\{0.65 < x \leq 0.85\} + (2(x-0.85)+0.2)I\{0.85 < x \leq 1\}$,
- Blocks, Bumps, Heavisine and Doppler are Donoho and Johnstone's functions (used for example in [8]) vertically rescaled to $[0.2, 0.8]$.

Densities. The design densities in figure 10 have the following expressions, up to a normalisation constant:

- Constant: $g(x) = 1$,
- Sine: $g(x) = 1 + 0.2\sin(4\pi x)$,
- Hole 1: $g(x) \asymp |x-0.5|^{0.5} + 0.04$ in section 3 and $g(x) \asymp |x-0.5|^{0.5}$ in section 4,
- Hole 2: $g(x) \asymp |x-0.5|^{0.9} + 0.03$ in section 3 and $g(x) \asymp |x-0.5|^{0.9}$ in section 4,
- Hole 3: $g(x) \asymp |x-0.5|^3 + 0.007$,
- Holes: $g(x) \asymp |\sin(\frac{1}{2|x-0.5|+0.1})|^{0.5} + 0.02$ in section 3 and $g(x) \asymp |\sin(\frac{1}{2|x-0.5|+0.1})|^{0.5}$ in section 4.

6.2 Simulation results

Some of the simulation results are summarized here graphically. Each graph provides, for a given quality criterium, the ratio of the value of the criterium for estimator 1 and the sum of the two values of the criterium for estimator 1 and estimator 2 (or 2′). Thus estimator 1 is better than estimator 2 whenever the ratio is below the value 0.5.

Each one of the six groups of nine successive columns refers to the nine target functions and to one of the six densities in the following order: 'Constant', 'Sine', 'Hole1', 'Hole2', 'Hole3' and 'Holes'.

Acknowledgments. The authors would like to thank Professor Dominique Picard and Professor Gérard Kerkyacharian for their help.

References

[1] A. Antoniadis and J. Bigot. Wavelet estimators in nonparametric regression : A comparative simulation study. *Journal of Statistical Software*, 6, 2001.

[2] A. Antoniadis, G. Grégoire, and P. Vial. Random design wavelet curve smoothing. *Statistics and Probability Letters*, 35, 1997.

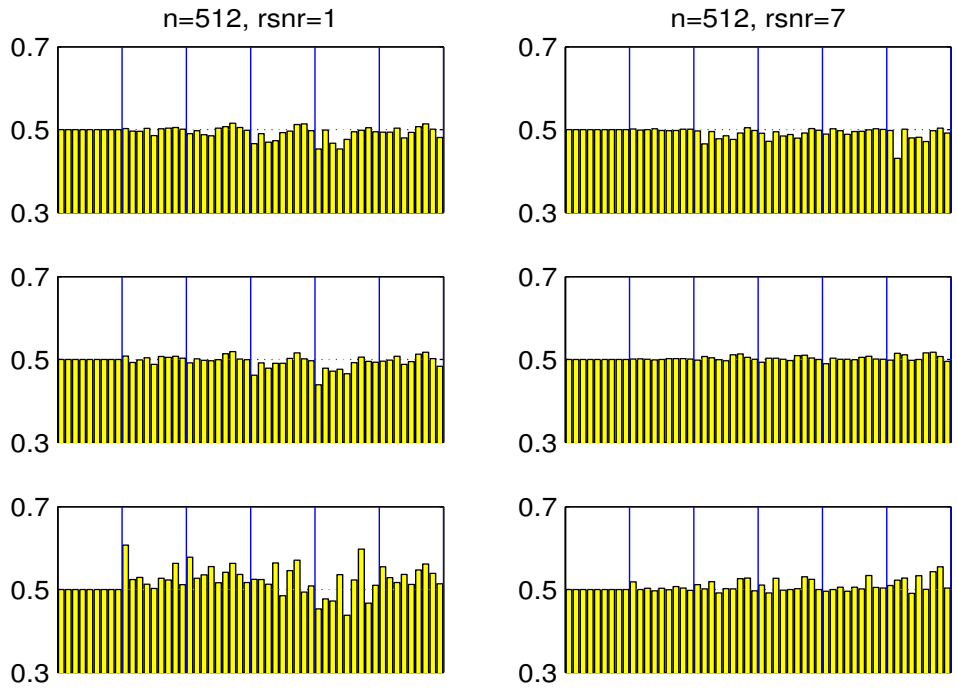

Figure 15: Estimator 1 versus estimator 2: ratios of L1 (top), RMSE (middle) and MXDV (bottom) in the classical setting

[3] N.D. Brinkman. Ethanol- a single-cylinder engine study of efficiency and exhaust emissions. *SAE Transactions*, 90:1410–1424, 1981.

[4] T. Cai and L.D. Brown. Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, 26:1783–1799, 1998.

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

[6] A. Cohen, R. De Vore, G. Kerkyacharian, and D. Picard. Maximal q-paces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11:167–191, 2000.

[7] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia., 1992.

[8] D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 1994.

[9] D. Donoho and I. Johnstone. Adaptating to unknown smoothness via wavelet shrinkage. *J . Am Stat. Assoc*, 90(432):1200–1224, 1995.

[10] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage by optimal recovery. *Lucien LeCam Festschrift D. Pollard and G.Yang (eds)*, 1994.

[11] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptotia. *J. Roy. Statist. Soc. Ser.*, B(57):301–369, 1995.

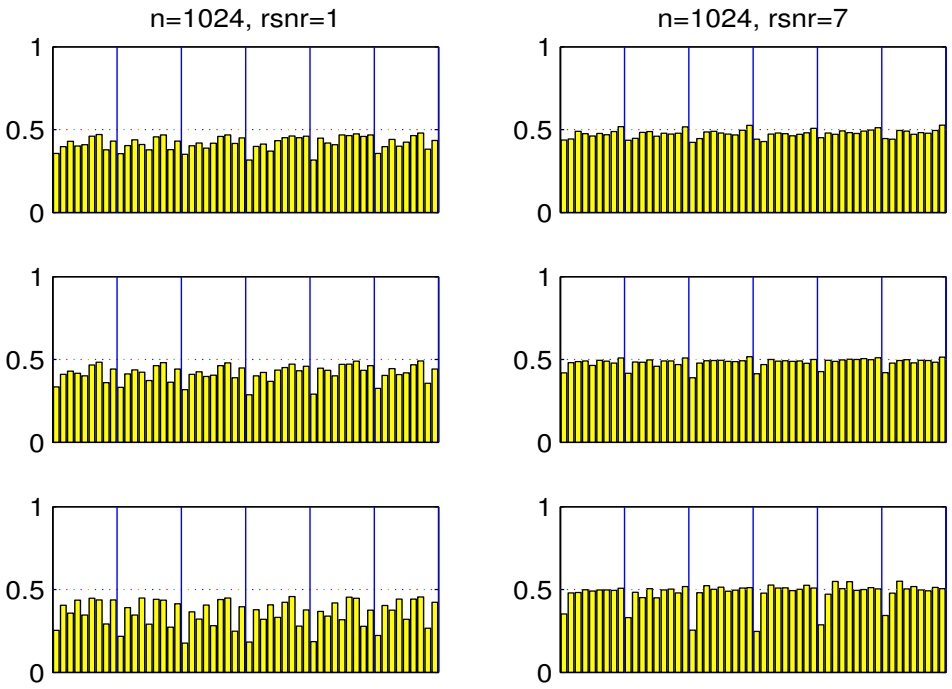

Figure 16: Estimator 1 versus estimator 2: ratios of L1 for unknown densities

[12] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24:508–539, 1996.

[13] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Universal near minimaxity of wavelet shrinkage. *Festschrift for Lucien Le Cam*, pages 183–218, 1997.

[14] P. Hall and B.A. Turlach. Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *The Annals of Statistics*, 25, 1997.

[15] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation and Statistical Applications*. Springer-Verlag, 1998.

[16] G. Kerkyacharian and D. Picard. Minimax or maxisets? *Bernoulli*, 8(2):219–253, 2002.

[17] G. Kerkyacharian and D. Picard. Regression in random design and warped wavelets. *Bernoulli*, 10(6):1053–1105, 2004.

[18] V. Korostelev and A. Tsybakov. *Minimax theory of image reconstruction*. Springer-Verlag, 1993.

[19] S. Mallat. *A wavelet tour of signal processing (2nd edition)*. Academic Press Inc., San Diego, CA, 1998.

[20] V. Maxim. Restauration de signaux bruités observés sur des plans d'expérience aléatoires. *Thèse de l'Université Joseph Fourier, Grenoble 1*, 2003.

[21] Y. Meyer. *Ondelettes et Opérateurs-I*. Hermann, 1990.

[22] Y. Meyer. Ondelettes sur l'intervalle. *Rev. Mat. Iberoamericana*, 7:115–133, 1990.

[23] B. Muckenhoupt. Weighted norm inequalities for the hardy maximal function. *Trans. Amer. Math. Hencec*, 165, 1972.

[24] G.P. Nason. Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage. *Statistics and Computing*, 12:219–227, 2002.

[25] B.W. Silverman. Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J. Roy. Statist. Soc.*, 47:1–52, 1985.

[26] A. Tsybakov. *Introduction à l'estimation Non-paramétrique*. Springer, 2004.

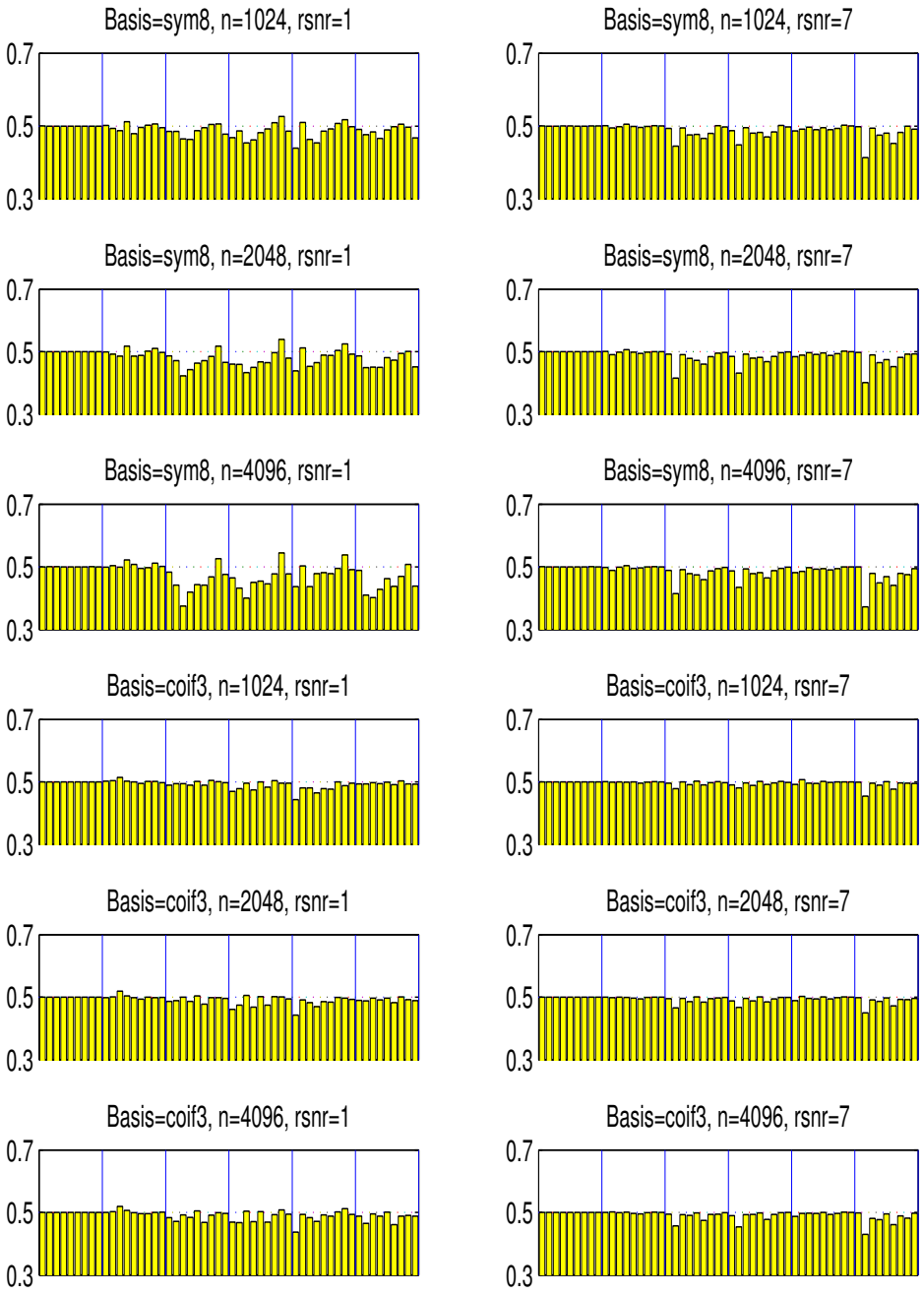

Figure 17: Estimator 1 versus estimator 2: ratios of L1 in the classical setting

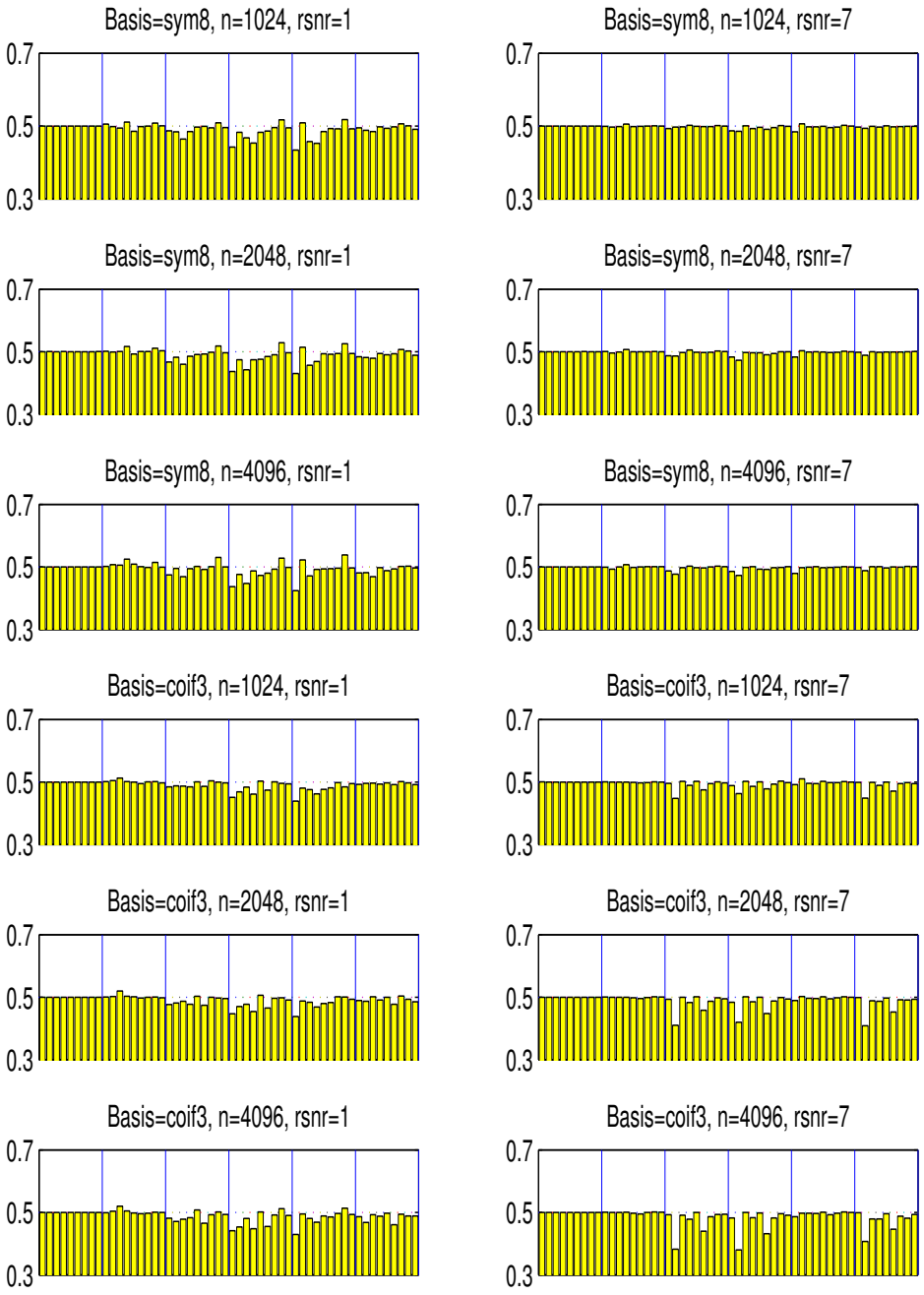

Figure 18: Estimator 1 versus estimator 2′: ratios of L1 for known vanishing densities