Statistica Sinica **19** (2009), 1371-1387

SMOOTHING NON-EQUISPACED HEAVY NOISY DATA WITH WAVELETS

Anestis Antoniadis, Irène Gijbels and Jean-Michel Poggi

Université Joseph Fourier, Katholieke Universiteit Leuven and Université Paris-Sud

Abstract: We consider a nonparametric noisy data model $Y_k = f(x_k) + \epsilon_k$, $k = 1, \ldots, n$, where the unknown signal $f: [0,1] \to \mathbb{R}$ is assumed to belong to a wide range of function classes, including discontinuous functions, and the $\epsilon_k' s$ are independent identically distributed noises with zero median. The distribution of the noise is assumed to be unknown and to satisfy some weak conditions. Possible noise distributions may have heavy tails, so that, for example, no moments of the noise exist. The design points are assumed to be deterministic points, not necessarily equispaced within the interval $[0,1]$. Since the functions can be nonsmooth and the noise may have heavy tails, traditional estimation methods (for example, kernel methods) cannot be applied directly in this situation. As in Brown, Cai, and Zhou (2008), our approach first uses local medians to construct certain variables Z_k structured as a Gaussian nonparametric regression but, unlike in this paper, the resulting data being not equispaced, we apply a wavelet block penalizing procedure adapted to non-equidistant designs to construct an estimator of the regression function. Under mild assumptions on the design it is shown that our estimator simultaneously attains the optimal rate of convergence over a wide range of Besov classes, without prior knowledge of the smoothness of the underlying functions or prior knowledge of the error distribution. The performance of our procedures is evaluated on simulated data sets covering a broad variety of settings and on some data examples, and are compared with other proposals made in the literature for treating similar problems.

Key words and phrases: Median, non equispaced design, penalization, robust regression, wavelets.

1. Introduction and Set-up

Suppose we want to recover a signal $f(\cdot)$ on the basis of independent observations

$$Y_k = f(x_k) + \epsilon_k, \quad k = 1, \ldots, n, \tag{1.1}$$

where $\{x_k\}_{k=1,\ldots,n} = \{x_{k,n}\}_{k=1,\ldots,n} \subset [0,1] = I$ is a design to be specified later, f belongs to a broad nonparametric class of functions, a Besov ball $B^{\alpha}_{p,q}([0,1])$ (for a definition see for example Amato, Antoniadis, and Pensky (2006)), and where the ϵ_k's are assumed to be i.i.d., each ϵ_k having a distribution that is absolutely continuous with respect to the Lebesgue measure with density, say, h. We will assume in particular that each ϵ_k has a zero median. Note that the noise distribution with density h may have heavy tails (e.g., Cauchy or Laplace) so that, for instance, the expectation of noise does not exist. Therefore, in this paper, the function f in (1.1) has a meaning of conditional median and the model will be called *nonparametric median regression* hereafter.

Inspired by the work of Brown, Cai, and Zhou (2008), our method may be summarized as a blockwise penalized wavelet kernel estimation on a signal built from the medians of suitably binned data. However, we focus here on non-equally spaced design, while in the paper just cited the authors have focused their work on equally spaced design using standard blockwise wavelet thresholding techniques for denoising. For a non-equally spaced design, we bin the sample so that each bin contains the same number of observations, and then take the median of each bin. The method therefore produces unequally spaced medians that are homoscedastic since the number of observations in the bins are the same. A wavelet kernel procedure for unequally spaced observations with homoscedastic noise, developed by Amato, Antoniadis, and Pensky (2006), applied to the local medians, together with a bias correction, leads to the desired estimator of f. In particular, we show that the optimality results obtained by Brown, Cai, and Zhou (2008) for the equidistant design model can be carried over to the non-equispaced design case.

Another way to handle the non-equidistant design case would be to bin the data into equal length bins, but this time the resulting medians would be approximated by Gaussian data that are heteroscedastic. Moreover the variance of each of the medians depends on how the x_i values are distributed within the bin, and the quality of the Gaussian approximation therefore depends on the "density" of the design points that usually is unknown. Even if this could be estimated, it is not at all clear that the results could be extended to that context.

In Section 2 we introduce the proposed method. Theoretical foundations and properties of the resulting estimator are established in Section 3. The finite-sample performance of the estimation procedure is investigated via a simulation study in Section 4, including also comparisons with other available methods. The method is also applied to two data examples.

2. The Proposed Procedure

We now give a detailed description of our procedure for robust estimation. Let $0 < m < n$ be an integer (its choice is discussed later) and partition the data into bins so that each bin contains the same number of observations m (i.e., each bin, say I_j is given by $I_j = [x_{(j-1)m+1}, x_{jm}]$ with $j = 1, \ldots, T = \lfloor n/m \rfloor$). Denote by X_j the median of the observations Y_k within bin I_j, $j = 1, \ldots, T$, and let

$$b_m = \mathbb{E}[\text{median}\{\epsilon_1, \ldots, \epsilon_m\}]. \tag{2.1}$$

Treating the X_j as data from a common Gaussian error regression model with mean $f(x_{jm}) + b_m$ (see Proposition 3.1 in the next section), we apply on the resulting non-equidistant data the wavelet kernel penalized estimation procedure of Amato, Antoniadis, and Pensky (2006) to obtain an estimate for $f(x_{jm}) + b_m$. An estimator of f at the points $\{x_{jm}\}_{j=1,\ldots,T}$ is obtained by subtracting an appropriate estimator of b_m as in Brown, Cai, and Zhou (2008). More precisely, divide again each bin I_j into two sub-bins with the first bin of the size $\lfloor m/2 \rfloor$ and denote by X_j^* the median of the observations in the first sub-bin. Then

$$\hat{b}_m = \frac{1}{T} \sum_{j=1}^{T} (X_j^* - X_j). \tag{2.2}$$

To apply the above procedure in practice one needs to choose the number of observations per bin, m, and the smoothing parameter, λ, in the wavelet kernel penalized procedure. The latter parameter is chosen by K-fold generalized cross-validation, while for choosing the number of observations per bin we have adopted the median cross-validation criterion proposed by Zheng and Yang (1998). Note that by using the wavelet denoising procedure of Amato, Antoniadis, and Pensky (2006) we do not need to estimate $h^2(0)$, as is necessary in the regular design case when using Cai's method (see Cai and Brown (1998)).

Remark 2.1. In the above procedure f is evaluated, for simplicity (to avoid notational complication), at the extremal points x_{jm} of the interval I_j. Instead, any point in the bin I_j could be chosen, such as for example the point at which the median of the Y observations (in the bin) is attained.

3. Theoretical Properties

Before stating the theoretical properties of our estimator, we introduce some notations and assumptions.

Assumptions on the noise distribution

As in Brown, Cai, and Zhou (2008) set

$$\mathcal{H}_{c_1,c_2} = \Big\{ h \text{ density } : \int_{-\infty}^{0} h(x)dx = 0.5,\ c_1 \le h(0) \le \frac{1}{c_1}, \text{ and } |h(x) - h(0)| \le \frac{|x|}{c_1} \text{ for all } |x| < c_2 \Big\}.$$

(A1) The ϵ_k's are i.i.d., each ϵ_k having a distribution that is absolutely continuous with respect to the Lebesgue measure with density h. There exists $0 < c_1 < 1$, $c_2 > 0$, such that $h \in \mathcal{H}_{c_1,c_2}$.

(A2) $\int |x|^\gamma h(x)dx < \infty$ for some $\gamma > 0$.

(A3) There exists $0 < c_1 < 1$, $c_2, c_3, c_4 > 0$ such that $h \in \mathcal{H}_{c_1,c_2,c_3,c_4}$, where

$$\mathcal{H}_{c_1,c_2,c_3,c_4} = \Big\{ h : h \in \mathcal{H}_{c_1,c_2},\ |h^{(3)}(x)| \leq c_4 \text{ for } |x| \leq c_3 \text{ and } \int |x|^{c_3} h(x)dx < c_4 \Big\}.$$

Assumption on the design

We assume that the design points are already ordered and that no ties occur. The theoretical results are established under the follow assumption on the design.

(B) There exists a sequence of integers $m = m_n$ with $0 < m_n < n$ such that, if the design points are binned into bins so that each bin contains the same number of observations m_n, and a $\delta \geq 3/4$, such that

$$\rho_n = \max_{1\leq j\leq T} \max_{(j-1)m_n+1\leq i\leq jm_n} |x_{jm_n,n} - x_{i,n}| = O(n^{-\delta}).$$

Remark 3.1. It is worth noting that the condition

(B') There exists a Lipschitz function $\kappa(\cdot)$ such that

$$\max_{1\leq j\leq T} \left| x_{jm_n,n} - x_{(j-1)m_n+1,n} - \frac{\kappa(x_{jm_n,n})}{n} \right| = o(n^{-1}).$$

is a sufficient condition for Condition (B) above. Indeed, with $\kappa(\cdot) = 1$, (B') implies (B) with $\delta > 1$, as discussed in Antoniadis, Grégoire, and McKeague (1994). Condition (B') is a standard assumption for the fixed design model, and is somewhat weaker than the "asymptotic equidistance" assumption of Gasser and Müller (1979) in which $\kappa(t) = 1$. Furthermore (B') implies the following standard condition of fixed design

(B") The design is such that $\max_i |x_{i,n} - x_{i+1,n}| = O(n^{-1})$,

which is also sufficient for Condition (B).

In case of random design, Condition (B) is satisfied in case the design density is continuous and strictly positive on $]0,1[$ (and hence bounded below and above on each closed subset of $]0,1[$).

Denote by X_j the median of the observations Y_k within bin I_j, $j = 1, \ldots, T$, and let $b_m = \mathbb{E}[\text{median}\{\epsilon_1, \ldots, \epsilon_{m_n}\}]$. Assume that f in model (1.1) belongs to the Besov ball $B^{\alpha}_{p,q}$ with $\alpha > 1/p + 1/2$ which, in particular, implies that f is Hölder with smoothness index $d = \min(\alpha - 1/p, 1) > 0$ and constant $C_f > 0$ (see Meyer (1992)). Then, similarly to Brown, Cai, and Zhou (2008) we can prove the following proposition.

Proposition 3.1. *Assume that the design satisfies* (B) *and that the noise distribution satisfies* (A1) *and* (A2). *The random variables* X_j *can be written as*

$$\sqrt{m_n} X_j = \sqrt{m_n}\{f(x_{jm_n,n}) + b_m\} + e_j + Z_j + \xi_j, \quad j = 1, \ldots, T, \tag{3.1}$$

where

(i) $Z_j \overset{i.i.d.}{\sim} N(0, [1/(4h^2(0))])$.

(ii) e_j *are deterministic constants with* $|e_j| \le C_f \sqrt{m} \rho_n^d$.

(iii) *there exists a constant* $C > 0$ *such that the random variables* ξ_j *are mean-zero independent variables with, for any* $\ell > 0$,

$$\mathbb{E}|\xi_j|^{\ell} \le C^{\ell} m^{-\ell/2} + C^{\ell} m^{\ell/2} (\max\{\rho_n^d, |b_m|\})^{\ell},$$

and for any $a > 0$,

$$\mathbb{P}\{|\xi_j| > a\} \le C^{\ell} (a^2 m)^{-\ell/2} + C^{\ell} \left(\frac{a^2 (\max\{\rho_n^d, |b_m|\})^{-2}}{m} \right)^{-\ell/2}.$$

A proof is given in the Appendix.

So far the approach outlined above, and the results concerning the median observations, is along the lines of Brown, Cai, and Zhou (2008) with minor complications due to the non-equispaced design, and with a more careful handling of the constants involved in the upper bounds. The main challenge remains the estimation of the signal by wavelet denoising since here, by the adopted binning approach, the binned data to be denoised are homoscedastic with Gaussian errors but with a non-equispaced design. One therefore needs to apply a wavelet denoising procedure that leads to estimates with the same optimal rates as in the equidistant design, this is a difficult task. Indeed, several wavelet denoising procedures for unequally spaced observations with homoscedastic Gaussian noise exist in the literature such as those developed by, for example, Antoniadis, Grégoire, and Vial (1997), Cai and Brown (1998), Kovac and Silverman (2000), Antoniadis and Fan (2001), and Kerkyacharian and Picard (2004), among others. However none of these procedures have allowed derivation of optimal rates of convergence for the estimators, similar to those for the equispaced setup. When applying the wavelet kernel penalized procedure developed by Amato, Antoniadis, and Pensky

(2006), we show below that the resulting estimators achieve the optimal rates of convergence.

Note first that by (3.1) the binned data are written as

$$\sqrt{m}X_j = \underbrace{\sqrt{m}g(x_{jm}) + e_j}_{\text{perturbated true signal}} + \underbrace{Z_j + \xi_j,}_{\text{perturbated Gaussian error}} \tag{3.2}$$

where $g(x_{jm}) = f(x_{jm}) + b_m$. Adopting the same notations as in Amato, Antoniadis, and Pensky (2006), the estimator of f at the points x_{jm} is obtained by minimizing

$$A(f) := \frac{1}{T}\sum_{i=1}^{T}\left\{\sqrt{m}X_i - \sqrt{m}f(x_{im})\right\}^2 + \lambda^2 R_J(\sqrt{m}f), \tag{3.3}$$

with $R_J(f) = \sum_{j=0}^{J}\sum_r \|P_{jr}f\|_{\mathcal{H}_{\Gamma,j,r}}$, $T = 2^J$, and where the penalty term in (3.3) is a sum of wavelet-based Reproducing Kernel Hilbert Space (RKHS) norms and a pseudo-norm. The main issue is to show that the perturbations in (3.2) can be controlled in such a way that they do not affect the final optimal rate of convergence. This result is stated in the following theorem.

Theorem 3.1. *Consider the nonparametric median regression model* $Y_i = f_0(x_i)+\epsilon_i$, $i = 1,\ldots,n$ *where* x_i*'s are given deterministic points in* $[0,1)$ *satisfying* (B), *and the* ϵ_i*'s are independent identically distributed noise variables satisfying* (A3). *Assume that* $\alpha \geq \max(1/q-1/2, 1/p+1/2)$, $p,q \geq 1$, *and let* $J = \delta(\log_2 n)$ *and* $s = 2/(2\alpha+1)$. *Assume that* δ *in* (B) *satisfies* $\max(3/4, 1/(2d+1)) \leq \delta < 1$ *and also that* f_0 *belongs to the unit ball of* $B^{\alpha}_{p,q}([0,1])$. *In the penalized kernel wavelet procedure, assume equal block sizes of length* $L_{jr} \approx (\log T)^{1+\nu}$ *with* $\nu \geq 0$ *and take* $\Gamma_{jr} = \Gamma_j = 2^{\mu j}$ *with* $1 < \mu < 2(\alpha - 1/p)$. *Consider the estimator* $\hat{\mathbf{f}}$ *of the values of the unknown regression function at the points defined by the binning process which is obtained by minimizing* (3.3).

Then (i) *if* f_0 *is not a constant and* $\lambda = \lambda_n$ *satisfies* $\lambda_n^{-1} = O_P(n^{(2-\rho)/4}) R_J^{(1-s)/2}(f_0)$, *we have* $(1/n)\|\hat{\mathbf{f}} - \mathbf{f}_0\|_n = O_P(\lambda_n)R_J^{1/2}(f_0)$; (ii) *if* f_0 *is constant, we have* $(1/n)\|\hat{\mathbf{f}} - \mathbf{f}_0\|_n = O_P(\max\{(n\lambda_n)^{-2/3}, n^{-1/2}\})$.

An analog of Theorem 3.3 of Amato, Antoniadis, and Pensky (2006), establishing integrated mean squared error consistency of the estimator, can also be stated and proved in similar fashion.

In the above cited paper, in order to get the right optimal consistency rate, the authors rely on Theorem 10.2 of van de Geer (2000) which involves conditions on the entropy of Besov balls as well as a sub-gaussianity condition on the noise. To prove Theorem 3.1 we show that the entropy bounds still hold despite the perturbation of the deterministic part in (3.2), and that the perturbated error is indeed sub-gaussian (see Lemma 3.2). The perturbation in the deterministic part can be handled via the result in Lemma 3.1.

Lemma 3.1. *Let* b_m *and* $\hat{b}_m$ *be defined as in* (2.1) *and* (2.2). *Under* (B) *and* (A3) *we have*

$$\sup_{\mathcal{H}_{c_1,c_2,c_3,c_4}} \left| b_m + \frac{h'(0)}{8h^3(0)m} \right| = O(m^{-2})$$

$$\text{and} \quad \sup_{\mathcal{H}_{c_1,c_2,c_3,c_4}} \sup_{f \in \text{unit ball of } B^{\alpha}_{p,q}([0,1])} \mathbb{E}(\hat{b}_m - b_m)^2 = O(\max(\rho_n^{2d}, m^{-4})).$$

For a proof of this lemma see Lemma 5 of Brown, Cai, and Zhou (2008).

Lemma 3.2. *Let the assumptions of Proposition* 3.1 *hold with* $\max(3/4, 1/(2d+1)) \leq \delta \leq 1$. *Then the perturbated error* $U_j = Z_j + \xi_j$*, with* Z_j *and* ξ_j *as in Proposition* 3.1*, is subgaussian.*

A proof of this lemma is given in the Appendix.

4. Experimental Results

This section reports results from a simulation study and some data examples that were conducted to evaluate the practical performance of our median based procedure for robust wavelet regression (RWR for short). All computations were performed using either Matlab or R. The Matlab codes for implementation of the procedure can be obtained from the authors upon request.

Our estimator is compared to competitors whose codes are available in the literature. Comparisons are done for both, equispaced and non-equispaced design.

4.1. Description of the algorithms and parameters choices

Algorithms. The algorithm proceeds in two major steps. The first uses local medians to construct new variables structured as a Gaussian nonparametric regression with non-equispaced design. The second step applies a wavelet block penalizing procedure adapted to non-equidistant designs to construct an estimator of the regression function. The computational algorithm and implementations issues are described in detail and discussed in Amato, Antoniadis, and Pensky (2006). The performance of the penalized estimator depends on some parameters, more precisely a regularization parameter, the chosen maximum resolution, and a block size, when thresholding is performed by blocks. All the default values provided by the waveker MATLAB code (see Amato, Antoniadis, and Pensky (2006)) are used here. In order to explore features of the data arising on different scales, scales from 1 to 6 are considered. The regularization parameter is chosen by 4-fold generalized cross-validation. The default wavelet is the near symmetric Symmlet 6 Daubechies wavelet (see Daubechies (1992)).

Selecting the number of data points in a bin. We propose to use the median cross-validation criterion of Zheng and Yang (1998) to select the number of nearest neighbors for estimating the regression function by local sample medians. Their procedure naturally deals with outliers and this data-driven selection method is used in the analysis of the data sets, even if the consistency results are only available in the equally spaced fixed design points regression model. For the simulated data we prefer to use $m_n = n^{1/4}$, in order to get a fixed number of observations for a given n, remaining the same across the samples. Let us mention that this strategy seems sometimes to select a slightly too small value for m.

4.2. Simulation study

The experimental setup was essentially the same as in Sardy, Tseng and Bruce (2001). Two test functions previously considered by Amato, Antoniadis, and Pensky (2006) were used: Heavisine and Corner. Alltogether a sample size $n = 1,024$, with two different types of noise considered — a Gaussian noise with a variance σ^2, and a Cauchy noise with scale σ — and two signal-to-noise ratios SNR= 3 and 7. The computation of the signal-to-noise ratio in the Cauchy noise case is a little tricky. For equispaced data, using a number of observations within each bin, say m (equal to $n^{1/4}$ in the simulations), and using the asymptotic results of our previous sections, the variance of the median (Gaussian) binned data X_j is approximately $(\sigma^2\pi^2)/(4m)$ for Cauchy noise with scale parameter σ. Denoting by $\sigma^2(f)$ the "variance" of the signal, we chose the scale parameter σ such that the ratio of the variance of the signal and the variance of the "noise" leads to a given signal-to-noise ratio (SNR), i.e., $\mathrm{SNR} = (\sqrt{4m}\sigma(f))/(\sigma\pi)$ or, equivalently, $\sigma = \sqrt{4m}\sigma(f)/(\mathrm{SNR} \times \pi)$. We adopted the same signal-to-noise ratio for the simulations involving a non-equispaced design. Figure 4.1 displays each of the signals together with a simulated (equidistant) sample from Cauchy data having a SNR= 7.

For each combination of test function and noise, 100 samples were generated. For the random design case non-equispaced placement of the sample points was done by distributing the points on the interval $[0, 1]$ according to a Beta$(1/2, 1/2)$ random variable.

For equidistant design, our method (RWR) is compared to three others.

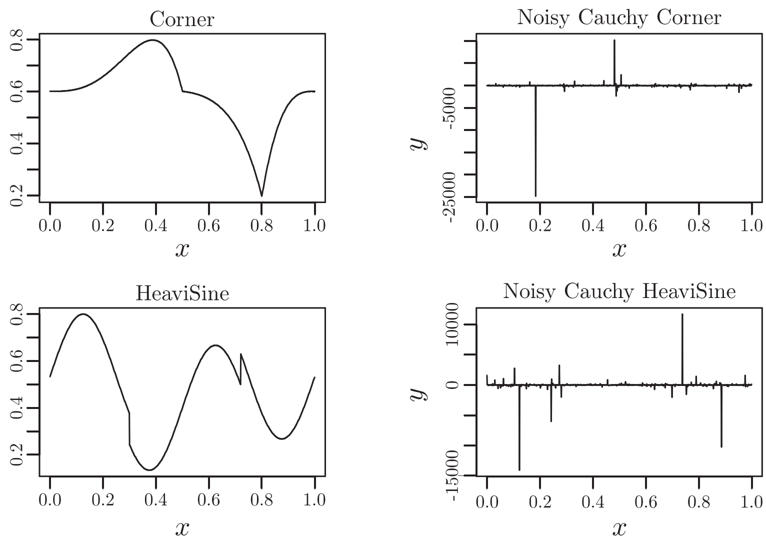

Figure 4.1. Each of the signals used in the simulation study, together with a typical simulated sample obtained by adding a Cauchy noise having a scale such that the resulting SNR is 7.

- RWS: the robust smoother-cleaner of Bruce and Gao (1994), with the default parameters and universal thresholding; software available using S+Wavelets (similar to the one developed by Sardy, Tseng and Bruce (2001) for robust wavelet denoising, or by Oh, Nychka, and Lee (2007) for robust smoothing).
- NPRQ: a local linear quantile regression estimator with a cross-validation bandwidth, see Yu and Jones (1998); available in the R package `quantreg` by Koenker (2007).
- BCZ: the wavelet median smoother of Brown, Cai, and Zhou (2008), where an estimate of the pseudo-data noise variance is obtained using a difference-type estimator (as suggested by these authors).

For our RWR estimators, whenever an estimate of the pseudo-data noise variance was needed, we used the estimate based on the median absolute deviation of finer detail wavelet coefficients (see, e.g., Donoho and Johnstone (1994)).

For each simulation setup (regular or irregular design), the mean squared error $MSE = (1/n)\sum(\hat{f}(x_i)-f(x_i))^2$ was calculated for each regression estimate $\hat{f}$ evaluated on an equally spaced grid of x's. This is important, especially in the case of irregularly spaced design points, to fairly appreciate the effect of the irregularity of the design on the estimation error. We also computed a standard nonrobust wavelet smoother (WAV) obtained by wavelet universal thresholding (see Donoho and Johnstone (1994)), but we do not show the results here since its

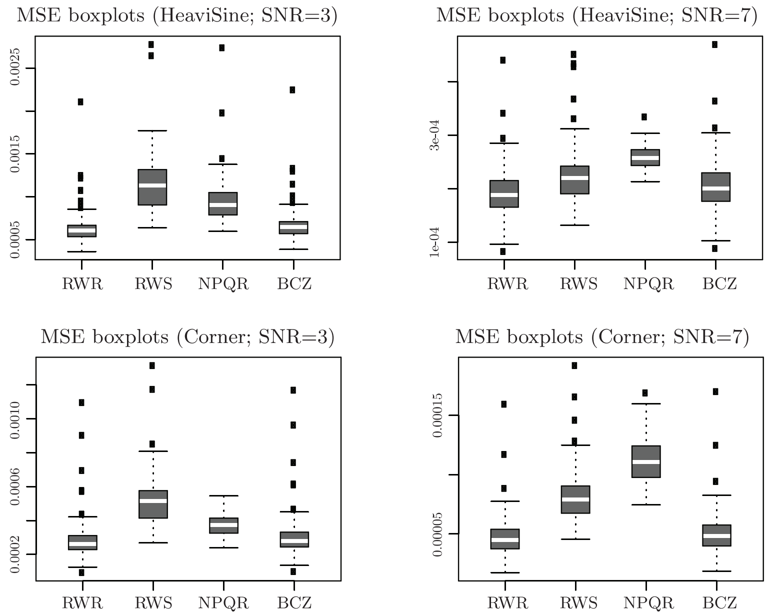

Figure 4.2. Boxplots of the MSE values of the Heavisine and Corner curve estimates from Cauchy noisy data for the four robust regression methods (equispaced design) and for the low and high SNR's in the simulation study.

MSE's are far too large when compared with those obtained with the methods mentioned above.

For each method, each function and each SNR, boxplots of these MSE's are displayed in Figure 4.2.

For non-equispaced design we compared our method with the NPQR method, which is the only available algorithm for random design robust smoothing. For the 100 simulations involving a non-equispaced placement of the sample points according to a Beta$(1/2, 1/2)$ random variable, some of the results are presented in Figure 4.3, depicting boxplots of the mean square error of the RWR method and NPQR.

For the two irregular functions that we used in our simulations, we also compared our robust smoother with a standard wavelet smoother (universal thresholding) on the Gaussian data sets in order to measure the loss in efficiency due to the binning. The resulting boxplots for each function and each SNR are displayed in Figure 4.4

The following major observations can be made. First, our proposed procedure RWR outperformed the others in most cases. Secondly, the standard wavelet denoising procedure (WAV) was better than the robust method for signals contaminated by Gaussian noise, but the loss of efficiency was not very large.

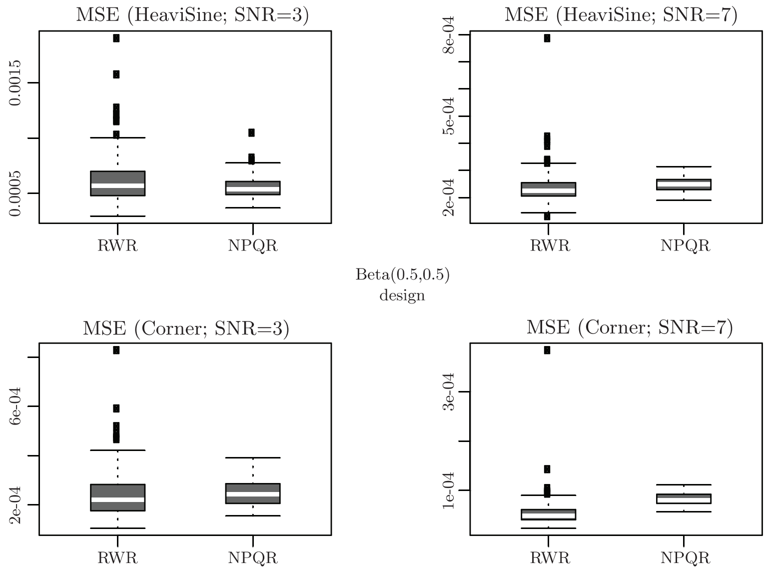

Figure 4.3. Boxplots of the MSE values of the Heavisine and Corner curve estimates from non-equispaced Cauchy noisy data for our robust regression method and Koenker's local linear quantile regression smoother.

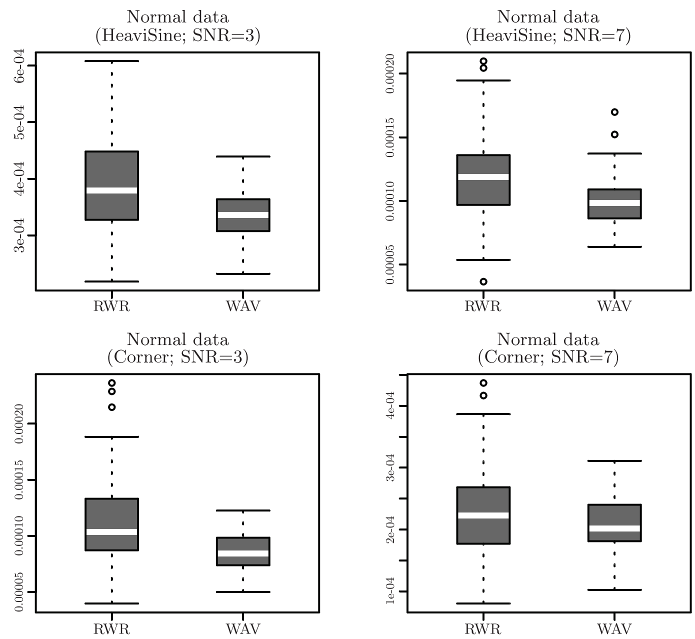

Figure 4.4. Boxplots of the MSE values of the Heavisine and Corner curve estimates from equispaced Gaussian noisy data for our robust regression method and a standard denoising wavelet based procedure.

A likely explanation is the fact that these procedures do not really rely upon the same number of points since RWR involves a preliminary binning. Lastly, the good average performance of NPRQ on the Heavisine signal is due to the fact that the tall small jump is hard to distinguish when the SNR is not large enough. Note also a small loss in efficiency when using the difference-based variance estimator of Brown, Cai, and Zhou (2008) in the case of equispaced data. In the non-equidistant case, see Figure 4.3, our RWR procedure performed better than the NPQR estimator in most cases, but was more variable. We believe that this is due to the way binning pre-processes the data.

4.3. Analysis of data

We apply our robust wavelet smoother to two data sets, the bone mineral density dataset with data observed on a non-equispaced design, and the radar glint noise data with equispaced design.

The glint data set. This data set is available in the module "Wavelets" of Splus, and has been previously examined by Bruce and Gao (1994) and Sardy, Tseng and Bruce (2001). The data are radar glint spikes observations from a target captured at $n = 512$ angles. Figure 4.5 compares robust denoising using our procedure with symmlets 6 and m, the number of data points in a bin, chosen by the median cross-validation criterion of Zheng and Yang (1998), to denoising with wavelet shrinkage combined with a clean and repeat procedure as it is implemented in the robust wavelet smoother-cleaner of Splus.

It can be seen that RWR is resistant to the adverse effects of outliers, notably at target angles near 5, 90, 140, 200, 320 (corresponding to abscissas near 0, 0.17, 0.27, 0.39, 0.62), and the range from 420 to 470 (resp. 0.82 to 0.92), while the smoother-cleaner procedure (RWS) of Bruce and Gao (1994) is still somewhat sensitive to the glint spikes.

Bone mineral density data. This example is based on measurements of bone mineral density (BMD) on 261 adolescents. The data were originally reported and analyzed by Bachrach, Hastie, Wang, Narasimhan, and Marcus (1999). The response is the relative change in spinal BMD and the covariate is the age of the adolescent. The data are also available in the R-package `ElemStatlearn`.

Figure 4.6 shows a robust regression analysis for the variable BMD conditional on gender. The response in the vertical axis is relative change in spinal BMD and the covariate on the horizontal axis is the age of the adolescent. The top panel provides the robust smoothing of RWR and of NPQR on the male subpopulation, while the bottom one is devoted to the female subpopulation. Whatever estimation method is used one clearly identifies different patterns of growth in the early ages, and especially the delay of roughly two years for males with respect to females.

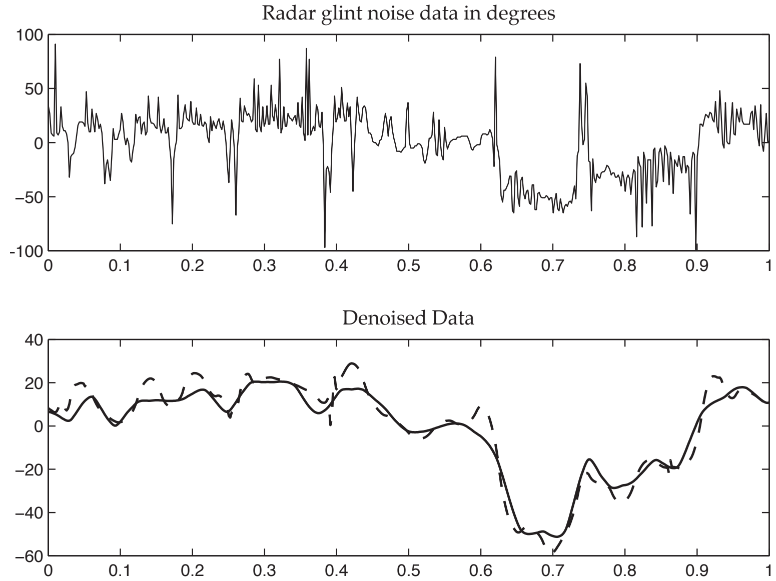

Figure 4.5. Top: radar glint noise data in degrees; Bottom: denoising by waveshrink combined with the robust clean and repeat wavelet procedure RWS (dashed line) compared to our robust wavelet smoothing procedure RWR (solid line).

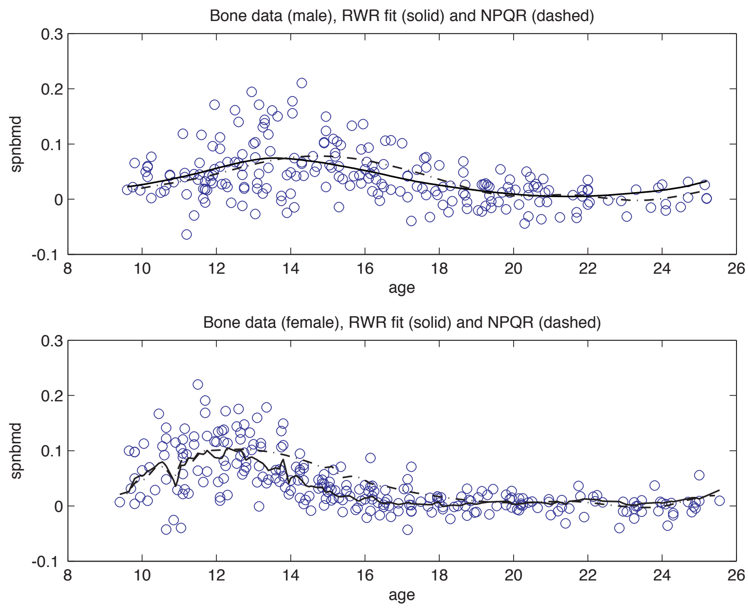

Figure 4.6. Relative change in Spinal BMD versus age. Data ('o') and corresponding RWR fit (solid curve) and NPQR fit (dashed curve) for the male population (Top) and the female population (Bottom).

Acknowledgement

Support from the IAP research network P6/03, Federal Science Policy, Belgium, is acknowledged. The second author also gratefully acknowledges financial support by the GOA/07/04-project of the Research Fund KULeuven. The authors would like to thank the Editor, an AE, and two anonymous referees for their useful suggestions and comments on improvements to this paper.

Appendix

A.1. Proof of Proposition 3.1

Let $\eta_j = \text{median}\{\epsilon_{(j-1)m+1}, \ldots, \epsilon_{jm}\}$. We define $Z_j = 2\Phi^{-1}(G(\eta_j))/h(0)$, where G is the cumulative distribution function of η_j and Φ is the cumulative distribution function of a standard normal variable. It is obvious that the random variables X_j, $j = 1, \ldots, T$, can be written as (with m_n denoted as m)

$$\sqrt{m}X_j = \sqrt{m}\{f(x_{jm}) + b_m\} + e_j + Z_j + \xi_j,$$

where

$$e_j = \sqrt{m}\mathbb{E}\{X_j - f(x_{jm}) - \eta_j\} \quad \text{and} \quad \xi_j = \sqrt{m}\{X_j - f(x_{jm}) - b_m\} - e_j - Z_j.$$

From Theorem 1 and Corollary 1 of Brown, Cai, and Zhou (2008), it follows that there exists a zero-mean Gaussian variable Z with variance $1/(4h^2(0))$ such that Z_j and Z have the same distribution, and moreover there exists a constant D such that

$$|\sqrt{m}\eta_j - Z_j| \leq Dm^{-1/2}\Big(1 + 4Z_j^2\Big) \tag{A.1}$$

when $|Z_j| \leq \varepsilon\sqrt{m}$, for some $\varepsilon > 0$.

Since X_j is the median of the observations in bin I_j, it is obvious that

$$\min_{(j-1)m+1\leq i\leq jm}\{f(x_i)\} \leq X_j - \eta_j \leq \max_{(j-1)m+1\leq i\leq jm}\{f(x_i)\}, \tag{A.2}$$

and therefore

$$|e_j| \leq \sqrt{m}\mathbb{E}\{|X_j - f(x_{jm}) - \eta_j|\} \leq \sqrt{m} \max_{(j-1)m+1\leq i\leq jm} |f(x_i) - f(x_{jm})| \leq C_f\sqrt{m}\rho_n^d. \tag{A.3}$$

The random term ξ_j has a zero mean and we split it into two terms $\xi_j^{(1)}$ and $\xi_j^{(2)}$, where

$$\xi_j^{(1)} = \sqrt{m}\{X_j - f(x_{jm}) - \eta_j\} - e_j - \sqrt{m}b_m \quad \text{and} \quad \xi_j^{(2)} = \sqrt{m}\eta_j - Z_j.$$

From (A.2) and (A.3) we have $|\xi_j^{(1)}| \leq \max(2C_f, 1)\sqrt{m}\max\{\rho_n^d, |b_m|\}$ and therefore, for any $\ell > 0$,

$$\mathbb{E}|\xi_j^{(1)}|^\ell \leq C^\ell m^{\ell/2}(\max\{\rho_n^d, |b_m|\})^\ell \tag{A.4}$$

with $C = \max(2C_f, 1)$.
By (A.1) we also have

$$\mathbb{E}|\xi_j^{(2)}|^\ell \leq C^\ell m^{-\ell/2} + (\mathbb{E}|\xi_j^{(2)}|^{2\ell})^{1/2}\mathbb{P}\{|Z_j| > \varepsilon\sqrt{m}\},$$

where C is a generic constant proportional to D. Since Z_j has a Gaussian distribution, assertion (iii) follows along the same lines as in the proof of Proposition 1 of Brown, Cai, and Zhou (2008).

A.2. Proof of Theorem 3.1 and Proof of Lemma 3.2

Proof of Theorem 3.1.

As in the proof of Theorem 3.2 of Amato, Antoniadis, and Pensky (2006), the conditions on the unknown regression function f_0 in Theorem 3.1 are only active for its wavelet coefficients and do not include the V_0 scaling coefficient of f_0. However, one has to control also the difference between the discretized values of $f_0(x_{jm})$ and $g_0(x_{jm}) + e_j/\sqrt{m} = f_0(x_{jm}) + b_m + e_j/\sqrt{m}$. For any $f \in \mathcal{H}_{J,\Gamma}$, write $f = a + f_1$ where $a \in V_0$ and $f_1 \in \mathcal{V}_{J,\Gamma}$. The conditions of Theorem 3.1 are equivalent to the fact that the function f_0 is such that $f_{01} \in \mathcal{V}_{J,\Gamma}$. Now, by Lemma 3.1, the constant b_m is uniformly bounded by a term of order m^{-1} and by Proposition 3.1 - (ii), $|e_j|/\sqrt{m} = O(n^{-\delta d})$. Hence, $(g_{01} + e_{0j}/\sqrt{m}) \in \mathcal{V}_{J,\Gamma}$, where $g_{01}+e_{0j}/\sqrt{m}$ is the corresponding projection of the perturbed deterministic signal onto the detail space $\mathcal{V}_{J,\Gamma}$. Moreover, the fact that b_m is uniformly bounded by a term of order m^{-1} leads to the same lower and upper bounds for the entropy of unit Besov balls as at (30) of Amato, Antoniadis, and Pensky (2006). By definition of the penalty R_J we have $R_J(\sqrt{m}f) = \sqrt{m}R_J(f)$. Using the above remark and the fact that the entropy condition in Theorem 10.2 of van de Geer (2000) involves the normalized quantity $(f - f_0)/(R_J(f) + R_J(f_0))$, invariant with respect to scalar multiplication of f and f_0, one sees easily that the factor $\sqrt{m}$ does not affect the entropy conditions. With the sub-gaussianity of the perturbated error term established in Lemma 3.2, the proof is complete.

Proof of Lemma 3.2.

Since both random variables Z_j and ξ_j have zero mean, subgaussianity of U_j is equivalent to the fact that there exists a $K > 0$ such that

$$\mathbb{E}\Bigg(\exp\Big(\frac{U_j^2}{K^2}\Big)\Bigg) < +\infty. \tag{A.5}$$

This assertion is true due to Theorem 2 of Pollard (2005). We have

$$\mathbb{E}\Bigg(\exp\Big(\frac{(Z_j+\xi_j)^2}{K^2}\Big)\Bigg) \leq \mathbb{E}\Bigg(\exp\Big(\frac{2Z_j^2}{K^2}\Big)\Bigg)\Bigg(\exp\Big(\frac{2\xi_j^2}{K^2}\Big)\Bigg).$$

From the Cauchy-Schwarz inequality it now follows that

$$\mathbb{E}\Big(\exp\Big(\frac{(Z_j+\xi_j)^2}{K^2}\Big)\Big) \le \mathbb{E}^{1/2}\Big(\exp\Big(\frac{4Z_j^2}{K^2}\Big)\Big)\mathbb{E}^{1/2}\Big(\exp\Big(\frac{4\xi_j^2}{K^2}\Big)\Big).$$

Since Z_j is zero-mean Gaussian with variance $1/(4h^2(0))$, the first factor on the right side of the last expression is finite. For the second factor we use Proposition 3.1 to get

$$\begin{aligned}\mathbb{E}\Big(\exp\frac{4\xi_j^2}{K^2}\Big) &= \sum_{\ell=0}^{\infty}\frac{1}{\ell!}\Big(\frac{4}{K^2}\Big)^{\ell}\mathbb{E}(\xi_j^{2\ell}) \\ &\le \sum_{\ell=0}^{\infty}\frac{1}{\ell!}\Big(\frac{4}{K^2}\Big)^{\ell}\Big\{C^{2\ell}m^{\ell}+C^{2\ell}m^{\ell}(\max\{\rho_n^d,|b_m|\})^{2\ell}\Big\} \\ &= \exp\Big(\frac{4C^2m}{K^2}\Big)+\exp\Big(\frac{4C^2m(\max\{\rho_n^d,|b_m|\})^2}{K^2}\Big).\end{aligned}$$

For K large enough and $\max(3/4, 1/(2d+1)) \le \delta \le 1$, both terms are finite and the result follows.

References

Amato, U., Antoniadis, A. and Pensky, M. (2006). Wavelet kernel penalized estimation for non-equispaced design regression. *Statist. Comput.* **16**, 37-56.

Antoniadis, A. and Fan, J. (2001). Regularization by wavelet approximations. *J. Amer. Statist. Assoc.* **96**, 939-967.

Antoniadis, A., Grégoire, G. and McKeague, I. (1994). Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.* **89**, 1340-1353.

Antoniadis, A., Grégoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statist. Probab. Lett.* **35**, 225-232.

Bachrach, L. K., Hastie, T., Wang, M.-C., Narasimhan, B. and Marcus, R. (1999). Bone mineral acquisition in healthy asian, hispanic, black and caucasian youth. A longitudinal study. *The Journal of Clinical Endocrinology & Metabolism* **84**, 4702-4712.

Brown, L. D., Cai, T. and Zhou, H. (2008). Robust nonparametric estimation via wavelet median regression. *Ann. Statist.* **36**, 2055-2084.

Bruce, A. and Gao, H. Y. (1994). *S+ Wavelets, Users Manual.* StatSci, Seatlle, WA.

Cai, T. and Brown, L. D. (1998). Wavelet shrinkage for nonequispaced samples. *Ann. Statist.* **26**, 1783-1799.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425-455.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation.* Lecture Notes in Mathematics **757**, 23-68.

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli* **10**, 1053-1105.

Koenker, R. (2007). Quantile Regression In R: A Vignette. Package quantreg can be freely downloaded from CRAN: `http://cran.r-project.org`.

Kovac, A. and Silverman, B. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Amer. Statist. Assoc.* **95**, 172-183.

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

Oh, H.-L., Nychka, D. W. and Lee, T. C. M. (2007). The role of pseudo data for robust smoothing with application to wavelet regression. *Biometrika* **94**, 893-904.

Pollard, D. (2005). *Asymptopia.* Book in Progress, Yale University (see `http://www.stat.yale.edu/~pollard/607.spring05/handouts/finite-max.pdf`).

Sardy, S., Tseng, P. and Bruce, A. (2001). Robust wavelet denoising. *IEEE Trans. Signal Process.* **49**, 1146-1152.

van de Geer, S. A. (2000). *Empirical Processes in M-estimation.* Cambridge University Press, Cambridge.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228-237.

Zheng, Z. and Yang, Y. (1998). Cross-validation and median criterion. *Statist. Sinica* **8**, 907-921.

Laboratoire Jean Kuntzmann, Department de Statistique, Université Joseph Fourier, Tour IRMA, B.P.53, 38041 Grenoble CEDEX 9, France.

E-mail: Anestis.Antoniadis@imag.fr

Department of Mathematics and Leuven Statistics Research Centre (LStat), Celestijnenlaan 200B, Box 2400, B-3001 Leuven (Heverlee), Belgium.

E-mail: Irene.Gijbels@wis.kuleuven.be

Laboratoire de Mathématiques, Equipe de Probabilités, Statistique et Modélisation, Université Paris-Sud, Bâtiment 425, 91405 Orsay Cedex, France.

E-mail: jean-michel.poggi@math.u-psud.fr

(Received April 2008; accepted December 2008)