Weighted Repeated Median Smoothing and Filtering

Roland FRIED, Jochen EINBECK, and Ursula GATHER

We propose weighted repeated median filters and smoothers for robust nonparametric regression in general and for robust online signal extraction from time series in particular. The new methods allow us to remove outlying sequences and to preserve discontinuities (shifts) in the underlying regression function (the signal) in the presence of local linear trends. Suitable weighting of the observations according to their distances in the design space reduces the bias arising from nonlinearities and improves the efficiency using larger bandwidths, while still distinguishing long-term shifts from outlier sequences. Other localized robust regression techniques like S, M, and MM estimators as well as weighted L_1 regression, are examined for comparison.

KEY WORDS: Breakdown point; Outlier; Robust regression; Signal extraction.

1. INTRODUCTION

Online analysis of a variable observed in short time lags is a common task. A basic objective is the extraction of the time-varying level (the signal) underlying a noisy time series. Relevant signal details, such as monotonic trends and abrupt shifts, need to be preserved, whereas irrelevant spikes due to measurement errors should be eliminated. Robust filtering procedures also should be fast and simple.

Standard median filters (Tukey 1977) remove spikes and preserve shifts, but they have difficulties if the signal is not almost constant within each window (e.g., Davies, Fried, and Gather 2004). For some improvement, we can weight the observations according to their temporal distances to the current target point, calculating a weighted median (WM). Whereas the median of observations $y_1, \ldots, y_n$ minimizes the L_1 distance, the WM $\hat{\mu}$ of $y_1, \ldots, y_n$ with positive weights $w_1, \ldots, w_n$ (which dates back to Edgeworth 1887) minimizes the weighted L_1 distance,

$$\hat{\mu} = \operatorname*{argmin}_{\mu} \sum_{i=1}^{n} w_i \cdot |y_i - \mu|. \tag{1}$$

In time series filtering with data $y_1, \ldots, y_n$ measured at fixed design points $x_1, \ldots, x_n$, we choose w_i depending on the distance between x_i and the target point x, $w_i = w(|x - x_i|)$, using a monotonically decreasing weight function w. WM filters are popular because of their flexibility. For a given minimal length $\ell + 1$ of signal details to be preserved, a WM filter with window width larger than the $2\ell + 1$ needed for a standard median can be selected. This allows more efficient noise suppression (Yang, Yin, Gabbouj, Astola, and Neuvo 1995). Time series filtering is a special case of nonparametric smoothing with a fixed design. Generally, locally weighted median smoothing, studied initially by Härdle and Gasser (1984), allows robust nonparametric estimation of the conditional median $\mu = g(x)$ of a response Y given a covariate x.

Local linear fits are usually preferable to local constant fits (Fan, Hu, and Truong 1994). The advantages are well known in case of L_2 regression (Fan 1992; Hastie and Loader

1993); however, robustness is needed in the presence of outliers. Davies et al. (2004) suggested the repeated median (RM) (Siegel 1982) for the extraction of monotonic trends from time series. The RM has the same optimal asymptotic 50% breakdown point as the standard median but relies on a constant slope within each window instead of a constant level.

We develop weighted repeated medians (WRMs) for robust nonparametric smoothing in the presence of trends. WRMs allow for application of longer windows than "standard" RM filters without being severely biased when the signal slope varies over time. For a full online analysis, we approximate the signal at the current time point without any time delay, giving the largest weights to the most recent observations.

We note that there are locally weighted versions of other robust regression techniques; equal weighting results in the highest efficiency of weighted Theil–Sen estimators and the highest asymptotic breakdown point of 29.3% among all efficiency-optimal weighting schemes in the case of an equally spaced design (Scholz 1978). Simpson and Yohai (1998) discussed the stability of one-step GM estimators (including weighted L_1 regression) in approximately linear regression with a random design.

The article is organized as follows. Section 2 reviews WMs and introduces WRMs and weighted L_1 regression. Section 3 derives analytical properties of these methods. Section 4 reports results from simulations. Section 5 exemplifies the methods on some time series, followed by some conclusions. Proofs of the analytical results are given in an Appendix.

2. ROBUST SMOOTHING AND FILTERING

We start with alternative derivations of WMs. WM filters downweight remote observations, which reduces problems due to trends but does not overcome them completely. For further improvement, we apply regression techniques with weighting according to the temporal distances. We review weighted L_1 regression before introducing WRMs.

2.1 Alternative Derivations of Weighted Medians

For nonnegative integer valued weights $w_1, \ldots, w_n$, a simple representation of the WM of real numbers $y_1, \ldots, y_n$ is given by

$$\hat{\mu} = \operatorname{med}\{w_1 \diamond y_1, \ldots, w_n \diamond y_n\} \tag{2}$$

where $w \diamond y$ denotes replication of y to obtain w identical copies of it.

Notation (2) can be used in an extended way for positive real weights as well. Let $y_{(1)} \leq \cdots \leq y_{(n)}$ denote the ordered observations and let $w_{(1)}, \ldots, w_{(n)}$ denote the corresponding positive weights. Then the WM of $y_1, \ldots, y_n$ is $\hat{\mu} = y_{(k)}$, where

$$k = \max \left\{ h : \sum_{i=h}^{n} w_{(i)} \geq \frac{1}{2} \sum_{i=1}^{n} w_i \right\}. \quad (3)$$

For example, the WM of 1, 2, 3 and 7 with weights .1, 1.6, 1.4, and .5 is $y_{(3)} = 3$, because $.5 + 1.4 \geq 3.6/2$. Generally, (3) and (1) yield the same results. However, the entire interval $[y_{(k-1)}, \ldots, y_{(k)}]$ solves (1) if $\sum_{i=k}^{n} w_{(i)} = \frac{1}{2} \sum_{i=1}^{n} w_i$. The solution $y_{(k-1)}$ would be obtained in (3) by summing from the bottom instead of from the top. This ambiguity can be solved as usual by choosing the midpoint of the interval.

Two WMs with respective weights $w_1, \ldots, w_n$ and $w'_1, \ldots, w'_n$ are equivalent iff they always give the same result. This is the case iff for every index set $I \subset \{1, \ldots, n\}$,

$$\sum_{i \in I} w_i \geq .5 \sum_{i=1}^{n} w_i \quad \Longleftrightarrow \quad \sum_{i \in I} w'_i \geq .5 \sum_{i=1}^{n} w'_i.$$

For $n = 3$, the WM with weights $(w_1, w_2, w_3) = (2, 4, 3)$ is equivalent to the standard median. For this, the weights must be balanced, such that no subset of less than $\lfloor (n+1)/2 \rfloor$ weights sums to at least half of the total mass. The WM is an order statistic with rank depending on the observations and the weights.

2.2 Weighted Median Smoothing and Filtering

Let $y_1, \ldots, y_N$ be observed at fixed design points $x_1 < \cdots < x_N$ under the model

$$Y_i = g(x_i) + u_i + v_i, \qquad i = 1, \ldots, N, \quad (4)$$

where u_i is symmetric observational noise with mean 0 and finite variance σ^2 and v_i is spiky noise from an outlier generating mechanism. The goal is to approximate the signal $g(x)$ for $x \in [x_1, x_N]$, representing the level of Y as a function of x. To distinguish signal and noise, we assume that $\mu = g(x)$ is smooth with infrequent shifts. The observational noise is assumed to be rough and the number of subsequent spikes small compared with the durations between the shifts.

Fan and Hall (1994) and Wang and Scott (1994) proposed local constant weighted L_1 estimates $\hat{g}(x)$ based on (1), using weights $w_1(x), \ldots, w_N(x)$. In time series filtering, the design is usually equidistant, $x_i = i, i = 1, \ldots, N$. In retrospective applications, when some delay is possible, we usually approximate the level in the window center choosing bell-shaped weights that are symmetric to the center and monotonically decreasing to both sides of it. When focusing on online analysis, where the target point x at which we estimate the signal is at the end of the window, we apply monotonically increasing weights (e.g., Einbeck and Kauermann 2003).

2.3 Weighted L_1 Regression

The theoretical properties of local linear mean estimators carry over to local linear median estimators based on L_1 regression (Fan et al. 1994). For the local linear median at x, $\hat{\mu}$, we fit a straight line to the data using a weight function,

$$(\hat{\mu}, \hat{\beta}) = \underset{\mu, \beta}{\mathrm{argmin}} \sum_{i=1}^{N} w_i(x) |y_i - \mu - \beta(x_i - x)|. \quad (5)$$

The solution of weighted L_1 regression (WL_1) generally is not unique. In case of a fixed design, $w_1(x), \ldots, w_N(x)$ are fixed, and WL_1 regression minimizes the residuals with respect to a norm. Thus the set of minimizing values is at least convex.

Several algorithms have been developed for L_1 regression and for quantile regression in general (Portnoy and Koenker 1997; Koenker 2005). These can be adapted to weighted L_1 regression because the ordinary L_1 solution of the modified problem

$$\mathrm{min!} \sum_{i=1}^{N} |w_i(x) \cdot y_i - w_i(x) \cdot \mu - \beta \cdot w_i(x) \cdot (x_i - x)| \quad (6)$$

with data $(w_i(x), w_i(x) \cdot x_i, w_i(x) \cdot y_i)$ is the same as the original WL_1 solution. We use an approximative L_1 procedure for simplicity and increased robustness. Starting from the standard RM, the algorithm iterates a finite number of steps between minimization of the objective function with respect to μ given the current β and vice versa.

2.4 Weighted Repeated Medians

Davies et al. (2004) investigated robust regression techniques like the standard RM and L_1 regression for delayed signal extraction from time series. Online versions of such procedures have been compared by Gather, Schettlinger, and Fried (2006). The RM is found to be preferable to the alternatives in both situations. The resulting (standard) RM filters fit a linear trend to the data in each window, replacing the assumption of a locally constant signal underlying the median by a trend with locally constant slope. This motivates us to generalize the RM, permitting localization by weighting.

Consider a window of width n with observations $(x_1, y_1), \ldots, (x_n, y_n)$, where $x_1 < \cdots < x_n$. We define the WRM with two possibly different sets of weights $w_i, \tilde{w}_i, i = 1, \ldots, n$, as

$$\tilde{\beta}^{WRM}(x) = \underset{j=1,\ldots,n}{\mathrm{med}} \ \tilde{w}_j \diamond \left(\underset{i \neq j}{\mathrm{med}} \ \tilde{w}_i \diamond \frac{y_i - y_j}{x_i - x_j} \right) \quad (7)$$

and

$$\tilde{\mu}^{WRM}(x) = \mathrm{med} \left(w_1 \diamond (y_1 - (x_1 - x)\tilde{\beta}^{WRM}(x)), \ldots, \right.$$
$$\left. w_n \diamond (y_n - (x_n - x)\tilde{\beta}^{WRM}(x)) \right); \quad (8)$$

that is, we weight the pairwise slopes in the inner median depending on the position of x_i, and in the outer median on the position of x_j when estimating the slope $\beta(x)$. We choose both sets of weights w_i and $\tilde{w}_i$ to be monotonic for online application.

We call two WRMs with weights $w_1, \ldots, w_n, \tilde{w}_1, \ldots, \tilde{w}_n$ and $w'_1, \ldots, w'_n, \tilde{w}'_1, \ldots, \tilde{w}'_n$ equivalent if the slope and the level estimate are always identical. A necessary condition for this is the equivalence of the WMs corresponding to $w_1, \ldots, w_n$

and $w'_1, \ldots, w'_n$: if the slope estimates are identical, then there are samples in which the WMs of the slope-corrected observations are different otherwise. The following additional condition for $\tilde{w}_1, \ldots, \tilde{w}_n$ and $\tilde{w}'_1, \ldots, \tilde{w}'_n$ guarantees the equivalence of WRMs:

> The WMs corresponding to $\tilde{w}_1, \ldots, \tilde{w}_n$ and to $\tilde{w}'_1, \ldots, \tilde{w}'_n$ are equivalent, and for each $i \in \{1, \ldots, n\}$, the WMs corresponding to $\tilde{w}_1, \ldots, \tilde{w}_{i-1}, \tilde{w}_{i+1}, \ldots, \tilde{w}_n$ and to $\tilde{w}'_1, \ldots, \tilde{w}'_{i-1}, \tilde{w}'_{i+1}, \ldots, \tilde{w}'_n$ also are equivalent.

3. ANALYTICAL PROPERTIES

As usual, we assume that for every target point x, the subset of design points with nonzero weights forms a window of subsequent points. We discuss analytical properties of the foregoing smoothers for a single window of width n, under the following condition:

(C) $y_1, \ldots, y_n$ are values of a response observed at fixed $x_1 < \cdots < x_n$. $w_1, \ldots, w_n$ and $\tilde{w}_1, \ldots, \tilde{w}_n$ are the corresponding sets of strictly positive weights. (We suppress the dependence on x because we treat a single target point x.)

3.1 Equivariances

Equivariances guarantee that an estimate reacts as expected to systematic changes in the data. Location equivariance means that adding a constant c changes the estimate by c. Scale equivariance means that multiplying all of the $y_1, \ldots, y_n$ by c changes the estimate by the same factor. The level estimates obtained from WMs and WRMs have both of these properties.

We also require that the quality of the smoothing does not depend on linear trends. This can be guaranteed by applying regression equivariant estimators. When regressing a variable y on a variate $\mathbf{z} \in \mathbb{R}^d$, regression equivariance means that adding a multiple $\mathbf{c}'\mathbf{z}$ to y for a $\mathbf{c} \in \mathbb{R}^d$ changes the estimate by this vector $\mathbf{c}$. (Weighted) RMs as defined here are equivariant with respect to adding a vector multiple $(a, b)\mathbf{z}_i = a + bx_i$ of $\mathbf{z}_i = (1, x_i)'$ to y_i, $i = 1, \ldots, n$. A procedure for (weighted) L_1 regression fulfills this equivariance if the initial estimator (e.g., the RM) fulfills it because we just act on the residuals thereafter. The performance of the WMs depends on trends, because they do not make use of the covariate values $x_1, \ldots, x_n$, but rather regress on a constant level only.

3.2 Removal of Spiky Noise

The removal of irrelevant spikes and the preservation of relevant signal details, particularly of long-term shifts, are essential properties of robust smoothers. The performance of moving-window techniques can be measured by two related quantities, the breakdown point and the exact fit point of the underlying functional.

The finite-sample replacement breakdown point measures the minimal fraction of data that can drive an estimate beyond all bounds when set to arbitrary values (Ellis and Morgenthaler 1992). In the context of nonparametric smoothing by moving-window techniques, this corresponds to the smallest fraction of contamination within a window that can cause an arbitrarily large spike in the output. A single outlier can already do so in local (weighted) least squares fits. (See Davies and Gather 2005 for a discussion of breakdown points.)

Another popular quantity in signal extraction is the number of spikes that a procedure can remove completely from a prototype signal in noise-free conditions, where the variance σ^2 of the observational noise equals 0. When applying a regression functional to a moving window assuming a locally linear trend, this number of spikes corresponds to the exact fit point of the functional. This is the smallest fraction of observations that can cause an estimated regression hyperplane to deviate from another hyperplane even though all of the other data points lie on the latter (Rousseeuw and Leroy 1987, sec. 3.4). For regression and scale-equivariant functionals, the exact fit point is not smaller than the finite-sample breakdown point. Let $\lfloor a \rfloor$ be the largest integer not larger than a. The standard median fits a constant exactly if less than $\lfloor (n+1)/2 \rfloor$ out of n observations deviate from it, which equals its breakdown point. Up to $\lfloor (n-1)/2 \rfloor$ subsequent spikes are removed completely from a constant signal. In retrospective application with a symmetric window, a shift from one constant to another is preserved exactly when applying an odd $n = 2m + 1$. In online applications, the shift is delayed by m time points.

Within a trend period, a standard median cannot exactly preserve a shift into the opposite direction, and a single spike causes smearing. An advantage of regression techniques is that the removal of outliers and preservation of shifts do not depend on linear trends, because the WRM and WL_1 regression are equivariant to them. The breakdown and the exact fit point of the standard RM for fitting a straight line both equal $\lfloor n/2 \rfloor/n$; thus it can remove $\lfloor n/2 \rfloor - 1$ subsequent spikes from a linear trend, which is only slightly less than for the standard median when the signal is constant.

For the derivation of breakdown and exact fit points of robust weighted regression methods, let $\mathbf{z}_i \in \mathbb{R}^d$ be fixed regressors, $\boldsymbol{\gamma} \in \mathbb{R}^d$ be the parameter to be estimated, and

$$y_i = \mathbf{z}'_i \cdot \boldsymbol{\gamma} + u_i, \qquad i = 1, \ldots, n.$$

We transfer results for standard L_1 regression to the weighted case using the modified problem (6). Based on work of He, Jureckova, Koenker, and Portnoy (1990, thm. 5.3), Ellis and Morgenthaler (1992, thm. 2.3), and Mizera and Müller (1999, thm. 2) we conclude that both the breakdown and the exact fit point of WL_1 regression equal k/n, where $k = \min |I|$, $I \subset \{1, \ldots, n\}$, for which $0 \neq \tilde{\boldsymbol{\gamma}} \in \mathbb{R}^d$ exists such that

$$\sum_{i \in I} w_i \cdot |\mathbf{z}'_i \cdot \tilde{\boldsymbol{\gamma}}| \geq \sum_{i \notin I} w_i \cdot |\mathbf{z}'_i \cdot \tilde{\boldsymbol{\gamma}}|. \tag{9}$$

Because a WM regresses on a constant, $\mathbf{z}_i \equiv 1$, its breakdown and exact fit point is the minimal fraction of weights that sum to at least $.5 \sum_{i=1}^n w_i$. It is straightforward that a WM that is not equivalent to the standard median has a breakdown point smaller than the optimal value $\lfloor (n+1)/2 \rfloor/n$ of the latter. The more variable the weights, the greater the loss in robustness due to weighting.

Calculating the numerical value of the breakdown and exact fit points of (weighted) L_1 regression is more difficult for $d \geq 2$, because then more directions must be considered. (For an algorithmic solution, see Giloni and Padberg 2004.)

Simple upper bounds for simple linear regression, $y_i = \mu + \beta(x_i - x)$, result from choosing the coordinate axis as directions $\tilde{\boldsymbol{\gamma}}$ in (9). The breakdown point of WL_1-regression with

weights $w_1, \ldots, w_n$ is not larger than $\min\{k_l, k_s\}/n$, where k_l is the minimal cardinality of $I \subset \{1, \ldots, n\}$ such that

$$\sum_{i \in I} w_i \geq \sum_{i \notin I} w_i,$$

and k_s is the minimal cardinality of $I \subset \{1, \ldots, n\}$ such that

$$\sum_{i \in I} w_i |x_i - x| \geq \sum_{i \notin I} w_i |x_i - x|.$$

This upper bound is generally not strict, because it considers only two directions. For standard L_1 regression and an equidistant, centered design, the upper bound is $1 - 1/\sqrt{2} = 29.3\%$ asymptotically, whereas the true value is at most 25% (Ellis and Morgenthaler 1992, prop. 4.1). Nevertheless, the upper bound is attained by the approximative weighted L_1 algorithm outlined in Section 2.3.

Next, we address the breakdown and exact fit of WRMs.

Proposition 1. Let a WRM $(\tilde{\mu}, \tilde{\beta})$ weighted by $w_1, \ldots, w_n$ and $\tilde{w}_1, \ldots, \tilde{w}_n$ fulfill (C). Then the following results hold:

a. A lower bound for the breakdown and the exact fit point of $(\tilde{\mu}, \tilde{\beta})$ is $\min\{k_s, k_l\}/n$, where k_s is the minimal number for which $\sum_{i=1}^{k_s} \tilde{w}_{[i]} \geq \sum_{i=k_s+2}^{n} \tilde{w}_{[i]}$, with $\tilde{w}_{[1]} \geq \tilde{w}_{[2]} \geq \cdots \geq \tilde{w}_{[n]}$ denoting the ordered weights, and k_l is the minimal number of weights $w_{[1]} \geq w_{[2]} \geq \cdots \geq w_{[n]}$ for which $\sum_{i=1}^{k_l} w_{[i]} \geq \sum_{i=k_l+1}^{n} w_{[i]}$.
b. An upper bound for the breakdown and exact fit points of $(\tilde{\mu}, \tilde{\beta})$ is $\min\{k_s' - 1, k_l\}/n$, where k_l is as in part a and k_s' is minimal with $\sum_{i=2}^{k_s'} \tilde{w}_{[i]} \geq \sum_{i=k_s'+1}^{n} \tilde{w}_{[i]}$.
c. The breakdown and the exact fit points of $(\tilde{\mu}, \tilde{\beta})$ do not exceed the $\lfloor n/2 \rfloor / n$ value of the standard repeated median.

The lower and the upper bounds given in parts a and b are not always identical; consider $n = 5$ and $(w_1, \ldots, w_5) = (\tilde{w}_1, \ldots, \tilde{w}_5) = (1, 1, 1, 3, 2)$, for which $k_s = 1$, but $k_s' = 3$. The lower bound is attained in the most relevant cases.

Proposition 2. Under (C), the breakdown and the exact fit points of a WRM with symmetric bell-shaped or monotonic weights equal $\min\{k_s, k_l\}/n$.

There are also WRMs that attain their respective upper bound, for example, the one for $n = 5$ mentioned earlier. The previous results allow us to determine weighted L_1 and WRM filters that remove outlier patches up to a given length completely while exactly preserving longer shifts under idealized conditions ($\sigma^2 = 0$).

In the simulations, we consider full online applications using the target point $x = x_n$ and triangular weighting schemes with $w_i(x) = \tilde{w}_i(x) = i$, $i = 1, \ldots, n$. Table 1 gives the minimal widths n necessary to remove outlier patches of different

lengths for weighted and standard L_1 and RM filtering. An equidistant design is assumed for L_1, whereas in case of the WRM, the results hold for any fixed design. n increases for the WRM compared with the standard RM, whereas weighting allows us to decrease n for L_1 filtering because of increased robustness. Nevertheless, the WL_1 does not achieve the optimal robustness of the standard RM and needs a somewhat larger n.

3.3 Continuity

(Lipschitz) continuity guarantees local stability to small changes in the data due to observational noise or rounding. Every WM is Lipschitz continuous with constant 1, because changing every observation by less than δ changes any order statistic by at most δ, and a WM always corresponds to one of these. For fixed design, the slope estimate of a WRM changes by at most $2\delta / \min(x_i - x_{i-1})$, so that the WRM level is Lipschitz continuous with constant $2 \max\{|x_1 - x|, |x_n - x|\} / \min(x_i - x_{i-1})$, because none of the slope-corrected observations changes more.

4. MONTE CARLO STUDY

Robust filters should preserve long-term shifts as discussed in Section 3.2, while removing irrelevant outlier sequences. We compare the online filters through simulations, concentrating on equidistant designs as in time series filtering. Data are generated from model (4) with standard Gaussian white noise u_i. The signal is a sine function, $g(x_i) = \nu \cdot .5 \cdot \sin(i \cdot \pi/100)$, $i = 1, \ldots, 100$, where $\nu \in \{0, 1, \ldots, 20\}$ determines the degree of nonlinearity. We treat a single window with target point $x = 50$.

In intensive care, five subsequent strongly deviant observations in hemodynamic series point to a relevant shift, whereas shorter sequences are typically irrelevant (Imhoff, Bauer, Gather, and Fried 2002). Accordingly, we fix widths in preliminary experiments with the aim of closely tracking large shifts from the fifth deviant observation on. Choosing the maximal width under this restriction optimizes both efficiency and robustness.

For the standard RM and L_1 regressions, we select $n = 11$. Using triangular weights, for the WRM, we choose the maximal width, leading to elimination of at most $\ell = 4$ observations ($n = 18$ according to Table 1), whereas we choose $n = 16$ for WL_1. We also include the fast S estimator from R with $n = 10$ (command `fast.s`); the least median of squares (LMS) with $n = 9$; the least trimmed squares (LTS) with $n = 10$, and the S estimator with $n = 10$ from the R package MASS (command `lqs`); the reweighted LTS (RLTS; command `ltsReg`) with $n = 11$ from the R package RRCOV; the M estimators with the Huber, Hampel, and Tukey's bi-squared functions, all with $n = 14$ and the MM estimator with $n = 10$ also from MASS (command `rlm`); and, finally, the MM estimator from

Table 1. The minimal window width n necessary to remove outlier patches of length ℓ in online application, weighted L_1 and RM regression

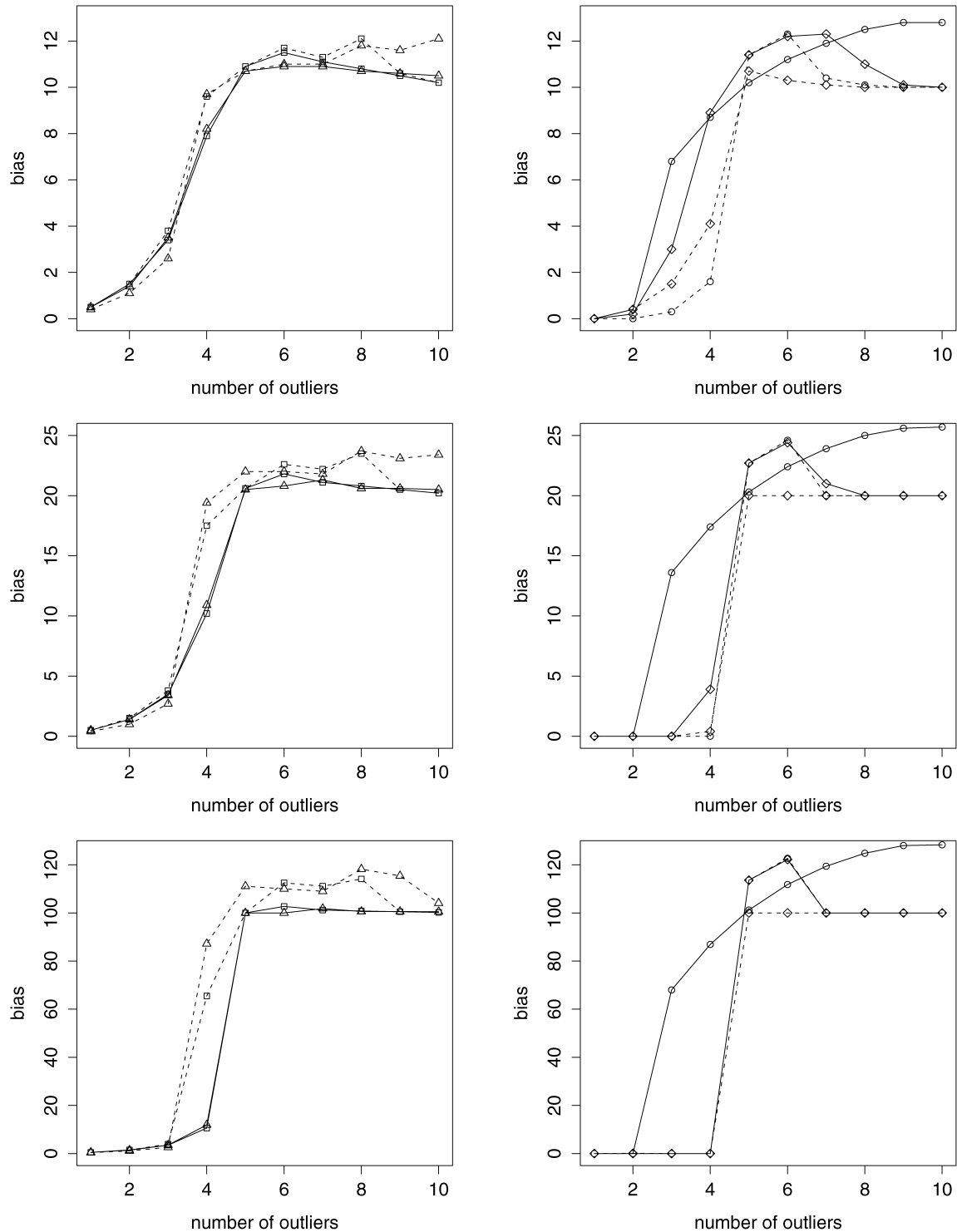| ℓ	L_1						RM					
	1	2	3	4	5	6	1	2	3	4	5	6
Standard L_1/RM	5	8	11	15	18	22	4	6	8	10	12	14
Triangular, $w_i(x) = i$	4	7	10	14	17	21	5	9	12	15	19	22

Figure 1. Bias for the level due to an increasing number of observations shifted by $c = 10$ (top), $c = 20$ (center), and $c = 100$ (bottom) at the end of the window. Left, solid line: RM (□), WRM (△); left, dashed line: L_1 (□), WL_1 (△); right, solid line: M with the bi-square (○), MM (◇); right, dashed line: LMS (◇), RLTS (○).

the ROBLM package with $n = 11$. We find that the latter outperforms the other MM estimator in our context. Similarly, the LTS and the S estimators demonstrated little advantage over the LMS and RLTS, and the M estimators with the Huber or the Hampel function over the Tukey biweight function.

Comparing the ability of the procedures to distinguish relevant from irrelevant patterns, we generate data resembling the

intrusion of a shift into the window, adding the same constant c to an increasing number of observations at the window end. In accordance with the foregoing demands, up to four deviant observations are considered outliers and should not affect the estimation, whereas the shift should be reproduced from five deviant observations on up. Figure 1 compares the bias for the signal value before the shift caused by $\ell = 1, 2, \ldots, 10$ shifted

observations at times $t = 50, 49, \ldots, 41$, calculated from 2,000 windows each. An ideal curve stays at 0 up to $\ell = 4$ and then increases abruptly to the added c representing the new level.

The RMs are less biased than the L_1 estimates in case of four outliers, with small differences between the weighted and the unweighted versions. In addition, the RMs closely reproduce the shift from the fifth observation on, whereas the L_1 estimates overshoot the new signal value for some observations. The LMS and the RLTS are even less biased than the RMs in case of two–four outliers but the RLTS overshoots shifts. The MM resists a few outliers very well, but becomes biased in case of three or four outliers and also overshoots shifts. The M estimator, finally, would strongly smooth shifts. The LMS has been proposed repeatedly for image analysis because of its excellent edge preservation (Meer, Mintz, Rosenfeld, and Kim 1991; Müller 1999; Rousseeuw and Van Aelst 1999). It is followed by the RLTS or the WRMs in this exercise, depending on whether one considers blurring of edges to be worse than overshooting or vice versa. We obtain similar results for the estimation of the slopes (not shown here). The LMS is generally the least biased, followed by the RLTS and the WRM. All of these results have been confirmed for other shift sizes and for window widths chosen to preserve shifts from the fourth observation on.

Figure 2 compares the efficiencies for Gaussian noise as a function of the nonlinearity ν in the absence of outliers and shifts. Because of a bias for $\nu \neq 0$, we measure the efficiency by the percentage mean squared error (MSE) compared with the standard RM, obtained from 10,000 runs for each $\nu = 0, \ldots, 20$. Weighting provides to gain a considerable amount of efficiency both for the RM and L_1 regression due to the longer window widths possible then. This is even more true for the slope. Only the M estimator is more efficient than these, but as we have seen before, it does not reproduce shifts. The RLTS is somewhat less efficient than the unweighted RM and L_1, whereas the LMS is much less efficient.

5. APPLICATION

For further comparison, we apply the filters to some time series. The simulated data depicted in Figure 3 are generated by overlaying a senoidal signal of length $N = 250$ with a shift by standard Gaussian white noise. A temporary shift of duration 7 is inserted at $x_i = 70$ to investigate the preservation of relevant patterns.

The procedures are challenged by inserting irrelevant sequences of up to three outliers of size 10. We choose the widths cited in Section 4 for tracking shifts after $\ell = 4$ observations; therefore, the filters resist the irrelevant outliers while delaying the shifts by four observations. We present the results only for the WRM, the LMS, and the RLTS because they outperformed the other methods in the simulation study. As was expected, the RLTS (like the MM and the L_1) overshoots the shifts. The LMS, as well as the LTS and the S estimators, provide wiggly outcomes according to its large variability. The lack of stability of the LMS has been noted previously in time series filtering (Davies et al. 2004). Like the MM, it reacts to irrelevant patterns in the data (e.g., at $t = 200$). The WRMs provide stable outcomes and track the shifts well, although they are somewhat affected by long outlier sequences.

We also consider real data representing the arterial pressure of a patient in intensive care (see Fig. 3). The filters are applied using the same widths as before. The MM, RLTS, and L_1 filters again overshoot the shifts. The WRM provides the best results because it tracks the shifts well, like the LMS and LTS, while being less variable.

6. CONCLUSION

We have investigated WRM and weighted L_1 filters for robust, detail-preserving online smoothing of noisy data with underlying trends. In case of the repeated median, weighting the observations according to their distance in the design space improves the local adaption to nonlinear regression functions, allows the use of longer windows, and increases the efficiency
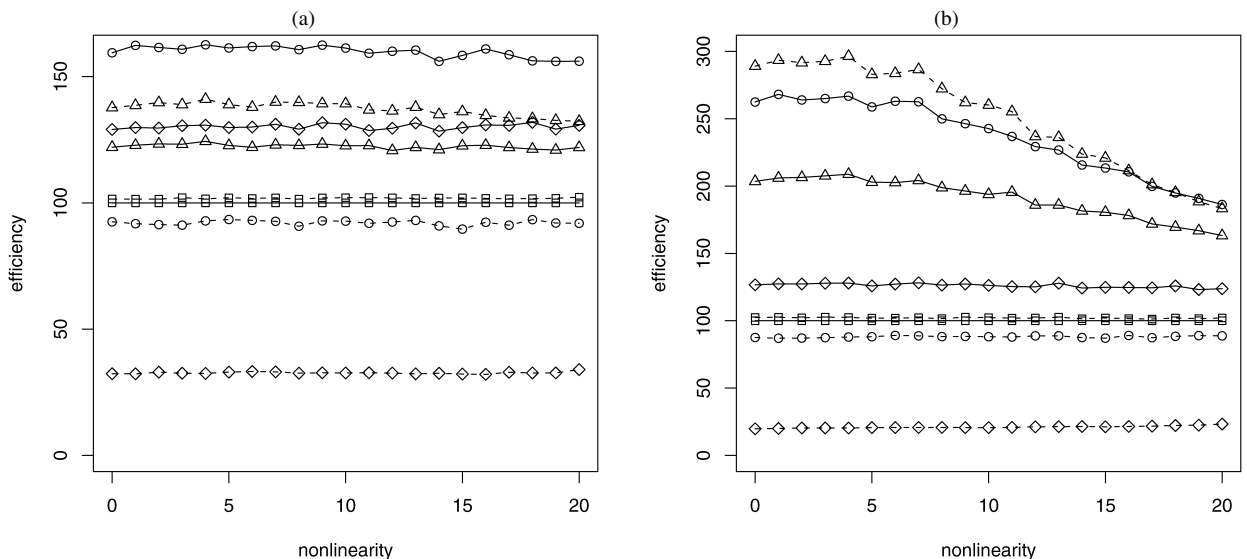

Figure 2. Efficiency for the level (a) and the slope (b) due to an increasing nonlinearity. Solid line: RM ($\square$), WRM ($\triangle$), M bi-squared ($\circ$), MM ($\diamond$); dashed line: L_1 regression ($\square$), WL_1 ($\triangle$), LMS ($\diamond$), RLTS ($\circ$).
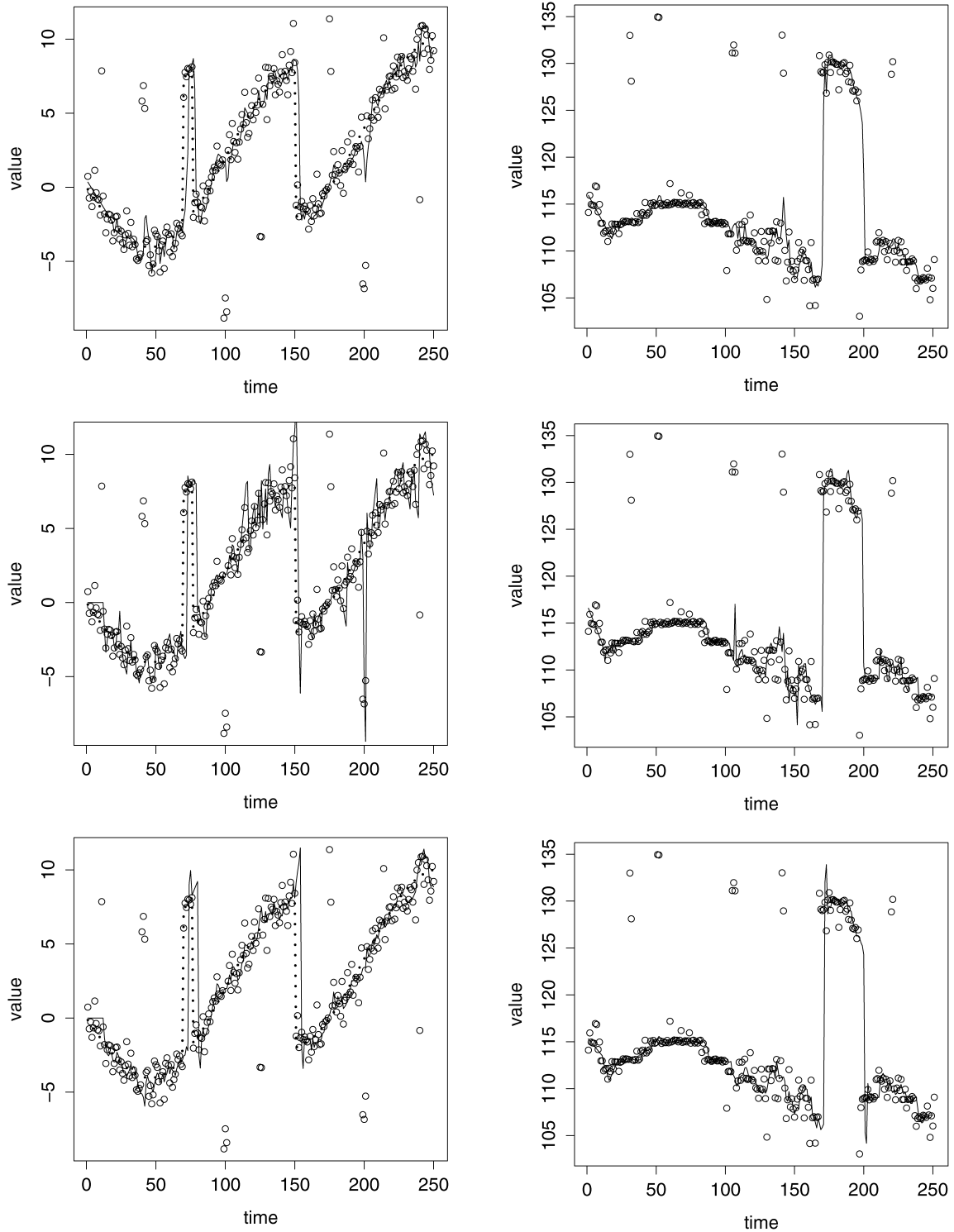
Figure 3. Simulated (left) and real time series (right), underlying signal (···), and filter outputs (—): for WRM (top), LMS (center), and RLTS (bottom).

compared with the standard version, retaining the suppression of outlying spikes and the preservation of relevant shifts. In the case of L_1 regression, in contrast, weighting can increase the robustness and the discrimination between sequences of relevant and irrelevant length.

We have compared the methods with several competitors also based on robust regression, namely LMS, LTS, S, M, and MM estimators. The LMS and LTS can provide even greater robust-

ness and better edge preservation than WRMs, but this advantage is outweighed by a much greater variability and poor stability of the outcome, making these methods less reliable in automatic application. M or MM estimators do not preserve shifts well. WRMs combine the advantages of stability, good edge preservation, high robustness, and considerable Gaussian efficiency. In addition, WRMs with triangular weights allow application of linear time algorithms based on updating the output

from the previous time point (Bernholt and Fried 2003). Experiments with different error distributions resulted in an average computation time for an update of about $2n \cdot 10^{-6}$ seconds on a 2-GHz Intel Core Duo with 512 MB DDR2/667 using a window of width n, whereas exact computation of the LMS and related methods needs $O(n^2)$ time. Thus WRMs with triangular weight functions are much faster than these competitors and can be used to analyze high-frequency data in real time. An implementation of the WRM filter is available in the R-package ROBFILTER, which can be found at *http://cran.r-project.org*.

A question not addressed here is the automatic identification of level shifts, for which many different rules have been suggested. A comparison of the different possibilities arising in combination with WRMs is beyond the scope of this article. Because we can tune a WRM to track level shifts with a prescribed delay of, say, ℓ observations, the following approach seems natural. Future signal values can be predicted by extrapolation of the regression line fitted to the most recent window. The fact that an intruding shift starts to influence the filter output when $\ell + 1$ observations are shifted and the output reaches the new level suggests comparing the current filter output at time t with its prediction calculated at time $t - \ell - 1$. A shift is detected if this difference is large relative to its standard deviation. The variance can be estimated by exponential smoothing of the squared differences even in the case of time-varying variability. We tested this approach on some examples using percentiles of the standard Gaussian distribution as thresholds and found that it worked well, although it sometimes was more sensitive to small changes than desired. This problem can be overcome by using thresholds corresponding to relevant changes.

In the simulations we have concentrated on the typical equidistant designs arising in time series filtering. Nonequidistant designs are found in, for example, option pricing. The analytical results for the WRMs presented here remain valid then. Based on our limited experience, we can say that the foregoing comparisons with respect to variability, robustness, and shift preservation carry over to more general situations, assuming that there are no outliers in the design space. Therefore, we tentatively recommend WRMs also for online applications with nonequidistant designs as well, although more investigations are needed with respect to the suitable choice of the window width.

All of these results rely on outlier patches being well separated. When such patches occur close to each other, using a standard RM with a reasonable width still may be the best decision, because it can deal with the largest fraction of outliers.

APPENDIX: PROOFS

Proof of Proposition 1

Because for regression and scale-equivariant functionals like WRMs, the exact fit point (EFP) is at least as large as the finite-sample breakdown point (BP), it suffices to prove part a for the BP and part b for the EFP.

a. Fewer than $k = \min\{k_s, k_l\}$ modifications have bounded effects on the level and the slope. When excluding an unmodified, "clean" observation y_j, the sum of the weights is still larger for the clean observations than for the modified observations. Thus, for every clean y_j, the inner median in the slope corresponds to a clean pair and is bounded. The WRM slope is bounded by the same quantity. The weighted majority of the slope-corrected y_j, and thus the WRM level, are then also bounded.

b. Because of regression equivariance, we may assume that all observations are 0 and need to find $k = \min\{k'_s - 1, k_l\}$ substitutions causing the fit to deviate from the horizontal axis. If $k = k'_s - 1$, then let the positions $I = \{i_1, \ldots, i_{k+1}\}$ correspond to the largest weights $\tilde{w}_{[1]} \geq \cdots \geq \tilde{w}_{[k+1]}$. Set the rightmost k of these observations (i.e., that the with largest x) on an increasing line with slope $b > 0$ through the leftmost of them. For each observation in I, the total weight of the other observations in I is at least the total weight of the unmodified zero observations. Thus the corresponding inner medians and the WRM slope is at least $b/2$. If $k = k_l$, then set the k observations with largest w_i to an arbitrary value M, obtaining a WRM level of at least $M/2$.

c. The standard RM has maximal BP among regression equivariant methods, including WRMs. Its EFP is maximal because it equals its upper bound, $k_s = k'_s - 1$.

Proof of Proposition 2

It is sufficient to prove that the EFP equals its lower bound. We assume without loss of generalization that all n observations equal 0 and show that $k = \min\{k_l, k_s\}$ modifications can make the WRM line deviate from the horizontal axis. Symmetric bell-shaped weights and monotonic weights can be treated in the same way. The k largest weights $\tilde{w}_j$ are at subsequent positions $x_{i-k+1} < \cdots < x_i$.

If $k_s \leq k_l$, then proceed as follows. If $\tilde{w}_1 + \cdots + \tilde{w}_{i-k} \geq \tilde{w}_{i+1} + \cdots + \tilde{w}_n$, then set the k observations at $x_{i-k+1}, \ldots, x_i$ to an increasing line with slope 1 through $(x_{i-k}, 0)$. Then $\tilde{w}_{i-k}$ is the $(k+1)$th largest $\tilde{w}_j$. The pairwise slope is 1 if both design points are selected from $x_{i-k}, \ldots, x_i$; strictly positive if one is from $x_1, \ldots, x_{i-k-1}$ and the other from $x_{i-k+1}, \ldots, x_i$; and it is 0 if both are from $x_1, \ldots, x_{i-k}, x_{i+1}, \ldots, x_n$. The inner median corresponding to x_{i-k} is strictly positive, because the total weight of the modified is at least that of the unmodified observations. This also holds for those at $x_{i-k+1}, \ldots, x_i$ because the pairwise slopes through $x_1, \ldots, x_i$ are larger than 0. Because the total weight at $x_{i-k}, \ldots, x_i$ is larger than the rest, the WRM slope is larger than 0 and the WRM line deviates from the horizontal axis.

If $\tilde{w}_1 + \cdots + \tilde{w}_{i-k} < \tilde{w}_{i+1} + \cdots + \tilde{w}_n$, then set the observations at $x_{i-k+1}, \ldots, x_i$ to an increasing line through $(x_{i+1}, 0)$ and use the same arguments as before, interchanging the role of $x_1, \ldots, x_{i-k}$ and $x_{i+1}, \ldots, x_n$.

If $k_l < k_s$, then set the k observations with largest w_i to 1. From the proof of Proposition 1, part a, it follows that the slope estimate is 0, but the level estimate is at least .5.

[Received August 2005. Revised July 2007.]

REFERENCES

Bernholt, T., and Fried, R. (2003), "Computing the Update of the Repeated Median Regression Line in Linear Time," *Information Processing Letters*, 88, 111–117.

Davies, P. L., and Gather, U. (2005), "Breakdown and Groups," *The Annals of Statistics*, 33, 977–1035.

Davies, P. L., Fried, R., and Gather, U. (2004), "Robust Signal Extraction for On-Line Monitoring Data," *Journal of Statistical Planning and Inference*, 122, 65–78.

Edgeworth, F. Y. (1887), "A New Method of Reducing Observations Relating to Several Questions," *Philosophical Magazine*, 24, 184–191.

Einbeck, J., and Kauermann, G. (2003), "Online Monitoring With Local Smoothing Methods and Adaptive Ridging," *Journal of Statistical Computation and Simulation*, 73, 913–929.

Ellis, S. P., and Morgenthaler, S. (1992), "Leverage and Breakdown in L1 Regression," *Journal of the American Statistical Association*, 87, 143–148.

Fan, J. (1992), "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

Fan, J., and Hall, P. (1994), "On Curve Estimation by Minimizing Mean Absolute Deviation and Its Implications," *The Annals of Statistics*, 22, 867–885.

Fan, J., Hu, T.-C., and Truong, Y. K. (1994), "Robust Nonparametric Function Estimation," *Scandinavian Journal of Statistics*, 21, 433–446.

Gather, U., Schettlinger, K., and Fried, R. (2006), "Online Signal Extraction by Robust Linear Regression," *Computational Statistics*, 21, 33–51.

Giloni, A., and Padberg, M. (2004), "The Finite Sample Breakdown Point of L1-Regression," *SIAM Journal of Optimization*, 14, 1028–1042.

Härdle, W., and Gasser, T. (1984), "Robust Non-Parametric Function Fitting," *Journal of the Royal Statistical Society*, Ser. B, 46, 42–51.

Hastie, T., and Loader, C. (1993), "Local Regression: Automatic Kernel Carpentry," *Statistical Science*, 8, 120–129.

He, X., Jureckova, J., Koenker, R., and Portnoy, S. (1990), "Tail Behavior of Regression Estimators and Their Breakdown Points," *Econometrica*, 58, 1195–1214.

Imhoff, M., Bauer, M., Gather, U., and Fried, R. (2002), "Pattern Detection in Intensive Care Monitoring Time Series With Autoregressive Models: Influence of the Model Order," *Biometrical Journal*, 44, 746–761.

Koenker, R. (2005), *Quantile Regression*, Economic Society Monograph Series, Cambridge, U.K.: Cambridge University Press.

Meer, P., Mintz, D., Rosenfeld, A., and Kim, D. Y. (1991), "Robust Regression Methods in Computer Vision: A Review," *International Journal of Computer Vision*, 6, 59–70.

Mizera, I., and Müller, C. H. (1999), "Breakdown Points and Variation Exponents of Robust M–Estimators in Linear Models," *The Annals of Statistics*, 27, 1164–1177.

Müller, C. H. (1999), "On the Use of High Breakdown Point Estimators in the Image Analysis," *Tatra Mountains Mathematical Publications*, 17, 283–293.

Portnoy, S., and Koenker, R. (1997), "The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators," *Statistical Science*, 12, 279–300.

Rousseeuw, P. J., and Leroy, A. (1987), *Robust Regression and Outlier Detection*, New York: Wiley.

Rousseeuw, P. J., and Van Aelst, S. (1999), "Positive-Breakdown Robust Methods in Computer Vision," in *Computing Science and Statistics*, Vol. 31, eds. K. Berk and M. Pourahmadi, Interface Foundation of North America, pp. 451–460.

Scholz, F.-W. (1978), "Weighted Median Regression Estimates," *The Annals of Statistics*, 6, 603–609.

Siegel, A. F. (1982), "Robust Regression Using Repeated Medians," *Biometrika*, 68, 242–244.

Simpson, D. G., and Yohai, V. J. (1998), "Functional Stability of One-Step GM-Estimators in Approximately Linear Regression," *The Annals of Statistics*, 26, 1147–1169.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Wang, F. T., and Scott, D. W. (1994), "The L_1 Method for Robust Nonparametric Regression," *Journal of the American Statistical Association*, 89, 65–76.

Yang, R., Yin, L., Gabbouj, M., Astola, J., and Neuvo, Y. (1995), "Optimal Weighted Median Filtering Under Structural Constraints," *IEEE Transactions on Signal Processing*, 43, 591–604.