Online Signal Extraction by Robust Linear Regression

Ursula Gather[1], Karen Schettlinger[1] and Roland Fried[2]

[1] Department of Statistics, University of Dortmund, 44221 Dortmund, Germany
[2] Department of Statistics, University Carlos III, 28903 Getafe (Madrid), Spain

Summary

In intensive care, time series of vital parameters have to be analysed online, i.e. without any time delay, since there may be serious consequences for the patient otherwise. Such time series show trends, slope changes and sudden level shifts, and they are overlaid by strong noise and many measurement artefacts. The development of update algorithms and the resulting increase in computational speed allows to apply robust regression techniques to moving time windows for online signal extraction. By simulations and applications we compare the performance of *least median of squares*, *least trimmed squares*, *repeated median* and *deepest regression* for online signal extraction.

Keywords: Robust filtering, least median of squares, least trimmed squares, repeated median, deepest regression, breakdown point.

1 Introduction

The online analysis of vital parameters in intensive care requires fast and reliable methods as a small fault can yield life-threatening consequences for the patient. Methods need to be able to deal with a high level of noise and measurement artefacts and to provide robustness against outliers. The variables in question include for example heart rate, temperature and different blood pressures.

Davies, Fried and Gather (2004) apply robust regression techniques to moving time windows to extract a signal containing constant periods, monotonic trends with time-varying slopes, and sudden level shifts. Since there is no commonly accepted standard for robust regression, they compare the popular L_1, *repeated median* (RM), and *least median of squares* (LMS) *regression* in this context. They report that repeated median regression is preferable to L_1 in most respects; as opposed to these methods, LMS regression tends to instabilities and it is slower, but it traces level shifts better and it is less biased in the presence of many large outliers.

These findings concern signal approximation in the centre of each time window, i.e. with some time delay. Since fast reaction is of utmost importance in intensive care, we exploit online versions of such procedures. Resuming the work from Davies et al. (2004) we analyse RM and LMS in this modified scenario without delay, and we investigate two further regression methods since their possible benefits have caused considerable attention in the statistical literature:

In Rousseeuw, Van Aelst and Hubert (1999, p. 425), Rousseeuw points out that he considers LMS to be outperformed by *least trimmed squares* (LTS) *regression* because of its smoother objective function, which results in a higher efficiency; the only advantage of LMS would be its minimax bias among all residual-based estimators. One concern here is whether LTS regression may outmatch LMS with respect to stability. Additionally, we investigate *deepest regression* (DR), which is expected to deal well with asymmetric and heteroscedastic errors (Rousseeuw and Hubert 1999) and compare it to RM regression which showed best performance for delayed signal extraction. Based on the results obtained here, the discussion at the end points at further promising regression candidates.

Section 2 introduces the regression methods and discusses some of their properties. Section 3 reports a simulation study carried out to investigate the performance of the methods in different data situations. Section 4 describes applications to some time series from intensive care, and finally, Section 5 closes with a discussion of the results.

2 Procedures for Online Signal Extraction

In the following, we consider a real valued time series $(\mathbf{y}_t)_{t\in\mathbb{Z}}$ observed at time points $t = 1, \ldots, N$. For the applicability of robust regression methods, we assume the data to be locally well approximated by a linear trend. This means, within time windows of fixed length $n = 2m + 1$ we assume a model

$$y_{t+i} = \mu_t + \beta_t i + \varepsilon_{t,i}, \quad i = -m, \ldots, m, \tag{1}$$

where μ_t denotes the underlying level (the signal) and β_t the slope at time t; the $\varepsilon_{t,i}$ are independent error terms with zero median. We consider different distributional assumptions for $\varepsilon_{t,i}$ below.

Regarding only one time window, we may drop the index t for simplicity. Hence, for a time window centred at time t we write $y_i = \mu + \beta i + \varepsilon_i$ for $i = -m, \ldots, m$. The window width n is chosen based on statistical and medical arguments as explained in Section 3.

2.1 Methods for Robust Regression

Let now $\mathbf{y} = (y_{-m}, \ldots, y_m)'$ denote a time window of width n from $(\mathbf{y}_t)_{t\in\mathbb{Z}}$, and let $r_i = y_i - (\tilde{\mu} + \tilde{\beta} i)$, $i = -m, \ldots, m$, denote the corresponding residuals. For estimation of the level μ and the slope β we consider the following robust regression functionals $T : \mathbb{R}^n \to \mathbb{R}^2$:

1. *Least Median of Squares* (Rousseeuw 1984)
$$T_{LMS}(\mathbf{y}) = (\tilde{\mu}_{LMS}, \tilde{\beta}_{LMS}) = \arg\min_{\tilde{\mu},\tilde{\beta}} \left\{ med \left(r_i^2 \, ; \; i = -m, \ldots, m \right) \right\} .$$

2. *Least Trimmed Squares* (Rousseeuw 1983)
$$T_{LTS}(\mathbf{y}) = (\tilde{\mu}_{LTS}, \tilde{\beta}_{LTS}) = \arg\min_{\tilde{\mu},\tilde{\beta}} \sum_{k=1}^{h} (r^2)_{k:n} \, ,$$
where $(r^2)_{k:n}$ denotes the kth ordered squared residual for the current time window, i.e. $(r^2)_{1:n} \le \ldots \le (r^2)_{k:n} \le \ldots \le (r^2)_{n:n}$ for any $k \in \{1, \ldots, n\}$, and h is a trimming proportion. We take $h = \lfloor n/2 \rfloor + 1$.

3. *Repeated Median* (Siegel 1982)
$$\begin{aligned} T_{RM}(\mathbf{y}) &= (\tilde{\mu}_{RM}, \tilde{\beta}_{RM}) \\ \text{with} \quad \tilde{\beta}_{RM} &= med_i \big(med_{j \ne i} \frac{y_i - y_j}{i - j} \; ; \; i, j = -m, \ldots, m \big) \\ \text{and} \quad \tilde{\mu}_{RM} &= med_i \big(\; y_i - \tilde{\beta}_{RM} i \, ; \qquad i = -m, \ldots, m \big), \end{aligned}$$
where the median for an even sample size is defined as the mean of the two midmost observations.

4. *Deepest Regression* (Rousseeuw and Hubert 1999)

$$T_{DR}(\mathbf{y}) = (\tilde{\mu}_{DR}, \tilde{\beta}_{DR}) = \arg\max_{\tilde{\mu},\tilde{\beta}} \Big\{ rdepth\Big((\tilde{\mu}, \tilde{\beta}), \mathbf{y} \Big) \Big\},$$

where the *regression depth* of a fit $(\tilde{\mu}, \tilde{\beta})$ to a sample $\mathbf{y}$ is defined as

$$rdepth\Big((\tilde{\mu}, \tilde{\beta}), \mathbf{y} \Big) = \min_{-m \le i \le m} \Big\{ \min\{L^+(i) + R^-(i), R^+(i) + L^-(i)\} \Big\}$$

$$\begin{aligned} \text{with} \quad L^+(i) = L^+_{\tilde{\mu},\tilde{\beta}}(i) = & \quad \#\Big\{ j \in \{ -m, \ldots, i \} \; : \; r_j(\tilde{\mu}, \tilde{\beta}) \ge 0 \Big\} \\ \text{and} \quad R^-(i) = R^-_{\tilde{\mu},\tilde{\beta}}(i) = & \quad \#\Big\{ j \in \{ i+1, \ldots, m \} \; : \; r_j(\tilde{\mu}, \tilde{\beta}) < 0 \Big\}. \end{aligned}$$

$L^-(i)$ and $R^+(i)$ are defined analogously.

We note that in the definition of LMS we could choose another quantile than the median, leading to the class of LQS regression estimates. Using the median maximizes the robustness of the method.

Applying such regression functionals, we estimate the level and the slope in the window centre, as in Davies, Fried and Gather (2004). This implies a delay of m time units for the current estimation. As we are rather interested in the level at the most recent time point, which is at the end of the window, we investigate the behaviour of the *online estimates* defined as $\tilde{\mu}^{online} = \tilde{\mu} + \tilde{\beta}m$.

In the context considered here, finite sample properties in case of an equidistant design and fast computation are of particular interest, and thus, the following sections investigate the performance of the methods with respect to such criteria. For more general information on LMS and LTS see Rousseeuw and Leroy (1987). Further properties of the LMS estimator are given by Davies (1993) and by Sheather, McKean and Hettmansperger (1997). The asymptotics of the RM estimator are investigated by Hössjer, Rousseeuw and Croux (1994), and by Hössjer, Rousseeuw and Ruts (1995). The asymptotic distribution of the DR estimator can be derived from Bai and He (1999), while van Aelst, Rousseeuw, Hubert and Struyf (2002) give a general overview over interesting DR properties.

2.2 Algorithms and Computational Speed

We use exact algorithms for all estimates (Bernholt 2004). For RM we have an update algorithm available. This prevents calculating the new value for each time window from scratch and thus enhances the computational speed.

	LMS	LTS	RM	DR
time	$O(n^2)$	$O(n^2)$	$O(n)$	$O(n \log^2 n)$
memory space	$O(n^2)$	$O(n^2)$	$O(n^2)$	$O(n)$

Table 1: Computational complexity of the considered algorithms.

	LMS	LTS	RM	DR
$n = 21$	0.161	0.161	0.035	0.747
$n = 31$	0.323	0.324	0.049	0.956

Table 2: Mean computation time of 10000 updates in msec.

The algorithms for LMS and LTS regression are based on the results of Edelsbrunner and Souvaine (1990). The repeated median algorithm is described in detail by Bernholt and Fried (2003), and the deepest regression estimates are computed by an algorithm based on results from van Kreveld, Mitchell, Rousseeuw, Sharir, Snoeyink and Speckmann (1999). This algorithm does not take the average over all deepest regression fits, if there are several, but chooses one of the deepest fits at random which increases the speed of computation but might lead to some loss of efficiency.

Table 1 shows the computational complexities of the resulting algorithms. However, these values only reflect asymptotic behaviour. Therefore, Table 2 shows the mean time needed for an update in milliseconds for small sample sizes, measured on a PC with *Pentium IV* processor with 2.4 GHz and 512 MB memory.

It turns out that, when using these algorithms, the repeated median is by far the fastest method for the considered sample sizes. In contrast to its low asymptotic computation time, an update of the DR estimate takes about 20 times longer than that of the repeated median. The algorithms for LMS and LTS are faster than that for DR for the small sample sizes considered here; the smaller asymptotic computation time of the latter seems to need considerable sample sizes to become dominant.

2.3 Breakdown and Exact Fit

In case of normal errors, least squares is the most efficient regression method. However, least squares regression can be strongly influenced by a single outlier, resulting in a finite sample replacement breakdown point of $1/n$. Since data from critical care can contain several outliers within short time spans, we prefer robust methods which show stable results and small bias even for a high percentage of contamination, preferably combined with satisfactory efficiency in periods without measurement problems and artefacts.

LMS, RM, and LTS (with $h = \lfloor n/2 \rfloor + 1$) possess a finite sample replacement breakdown point of $\lfloor n/2 \rfloor / n \approx 50\%$. This is the highest possible value for a regression equivariant functional (Rousseeuw and Leroy 1987). Rousseeuw and Hubert (1999) show that deepest regression has a breakdown point of at least about one third in any case. This raises the question if its breakdown is larger than this in case of an equidistant fixed design as treated here. For example, the L_1 breakdown point is $1/n$ if contamination in the explanatory variable is allowed, while it is about 25% in case of an equally-spaced design. However, below we will provide evidence that even in this case deepest regression only guarantees protection against up to one third contaminated observations in the sample.

To accomplish this, we first regard the exact fit property: data from intensive care often contain repeated values as the measurements are on a discrete scale, and the patient's physiological parameters can stay steady for some time. In such situations the exact fit property is informative.

A regression functional $T : \mathbb{R}^n \to \mathbb{R}^2$ possesses the exact fit property if for some fit $(\tilde{\mu}, \tilde{\beta})$ and $k \in \{0, 1, \ldots, \lceil n/2 \rceil - 1\}$ the following is satisfied: Whenever $y_i = \tilde{\mu} + \tilde{\beta} i$ fits at least $n - k$ of the n observations exactly, then $T = (\tilde{\mu}, \tilde{\beta})$ whatever the other k observations are. Roughly spoken: if the majority of the data lies on a straight line, the solution of the functional T will be exactly this line (Rousseeuw and Leroy 1987, p. 122).

The smallest possible fraction of contamination which can cause a regression functional T to deviate from $(\tilde{\mu}, \tilde{\beta})$ is called the *exact fit point*: consider a sample $\mathbf{y}_n$ of size n such that $y_i = \tilde{\mu} + \tilde{\beta} i$ for all i, and let $\mathbf{y}_{k,n}$ be a sample where k out of the n observations of $\mathbf{y}_n$ are replaced by arbitrary values. Then, the exact fit point of T is defined as

$$\delta_n^*(T, \mathbf{y}_n) = \min_k \left\{ \frac{k}{n} \;\middle|\; \text{there exists a sample } \mathbf{y}_{k,n} \text{ such that } T(\mathbf{y}_{k,n}) \neq (\tilde{\mu}, \tilde{\beta}) \right\}.$$

For regression and scale equivariant functionals as considered here, this value gives an upper bound for the finite sample replacement breakdown point ε_n^* (Rousseeuw and Leroy 1987, pp. 122-124), i.e.

$$\varepsilon_n^*(T, \mathbf{y}_n) \;\leq\; \delta_n^*(T, \mathbf{y}_n).$$

The exact fit point for LMS and LTS is $\frac{\lceil n/2 \rceil}{n}$ (Rousseeuw and Leroy 1987, Section 3.4). For RM one less observation is needed to pull the fit away from the line in case of a sample of odd size, because its slope component is calculated by taking sets of two observations. Hence, its exact fit point $\frac{\lfloor n/2 \rfloor}{n}$ is equal to its breakdown point.

For deepest regression an upper bound for the exact fit point can be derived as follows: consider a sample $\mathbf{y}_{n,k}$ of size n where $n - k$ observations lie

n	5	7	9	11	13	15	17	19	21	23	25	27
k	2	2	3	4	4	5	6	6	7	8	8	9
n	29	31	33	35	37	39	41	43	45	47	49	51
k	10	10	11	12	12	13	14	14	15	16	16	17

Table 3: Upper bound for the exact fit point k/n of the deepest regression functional for selected sample sizes n.

on a straight line $l_0 : y_j = \mu_0 + \beta_0 j$, $j = -m, \ldots, m$. The exact fit point δ_n^* equals the smallest fraction k/n of values not lying on l_0 such that the deepest regression fit departs from the line l_0. This means we are searching for a number k with $T_{DR}(\mathbf{y}_{n,k-1}) = (\mu_0, \beta_0)$ and $T_{DR}(\mathbf{y}_{n,k}) \neq (\mu_0, \beta_0)$.

W.l.o.g. we assume $\mu_0 = 0$ and $\beta_0 = 0$. Furthermore, we take the first $n-k$ observations to lie on the line l_0, i.e. we have $y_j = 0$ for $j = -m, \ldots, m-k$, and we put the remaining k observations on another line $l_1 : y_j = \mu_1 + \beta_1 j$ for $j = m-k+1, \ldots, m$, with $\mu_1 = -\frac{n+1}{2}$ and $\beta_1 = 1 \neq \beta_0$. This guarantees that l_1 has a regression depth of at least k, because at least k observations lie on l_1. Also, the residuals of these observations have the same (positive) sign with respect to l_0. In this way, the fit of l_0 to the full sample $\mathbf{y}_{n,k}$ is worsened with increasing k. Table 3 gives the smallest number k of non-zero observations which, in this configuration, forces the deepest regression estimate away from $(0,0)$ for small to moderate sample sizes.

In this particular data situation and for the considered sample sizes, we see that the departure of $\lfloor \frac{n+1}{3} \rfloor$ observations from l_0 can cause the deepest regression fit to do so, too.

Hence, we can conclude that the *smallest* k with $T_{DR}(\mathbf{y}_{n,k}) \neq (\mu, \beta)$ is at most $\lfloor \frac{n+1}{3} \rfloor$ and thus

$$\delta_n^*(T_{DR}, \mathbf{y}) \quad \leq \quad \frac{1}{n} \cdot \left\lfloor \frac{n+1}{3} \right\rfloor .$$

Rousseeuw and Hubert (1999) show that the breakdown point of the T_{DR} at any data set is *at least* one third:

$$\varepsilon_n^*(T_{DR}, \mathbf{y}) \geq \frac{1}{n} \left(\left\lceil \frac{n}{3} \right\rceil - 1 \right) \approx \frac{1}{3}.$$

Thus,

$$\frac{1}{n} \left(\left\lceil \frac{n}{3} \right\rceil - 1 \right) \leq \varepsilon_n^*(T_{DR}, \mathbf{y}) \leq \delta_n^*(T_{DR}, \mathbf{y}) \leq \frac{1}{n} \cdot \left\lfloor \frac{n+1}{3} \right\rfloor .$$

This leads to the claim that, even in case of an equally-spaced design, the breakdown point of the DR functional equals 1/3.

3 Monte Carlo Study

In the following, we compare the performance of the *online estimates* $\tilde{\mu}^{online} = \tilde{\mu} + \tilde{\beta}m$ in different data situations. In particular, we consider scenarios which are of importance in the online monitoring context. The performance of the estimates will be judged by their standard deviation, bias and root mean squared error. For comparison, we also include results for least squares (LS) regression. Data are generated from the simple linear model

$$Y_i = \mu + \beta i + \varepsilon_i,\ i = -m, \dots, m,$$

- normal errors,
- heavy tailed errors,
- skewed errors,
- normal errors with additive outliers at random time points,
- normal errors with subsequent additive outliers.

We set $\mu = \beta = 0$ w.l.o.g. since all methods considered here are regression equivariant, and set the error variance to one w.l.o.g. because of the scale equivariance. In each case $S = 10000$ independent samples are generated.

On the one hand, the assumption of a linear trend within each time window becomes less reliable if a large window width is chosen: even a small bias in the estimation for the window centre can cause a considerable bias of the online estimates because these are based on linear extrapolation. On the other hand, a large window width stands for smaller variability and produces smoother estimates. As a compromise, a choice of $m = 10$ or $m = 15$ is considered acceptable for the physiological data we have in mind, leading to window widths of $n = 21$ or $n = 31$ respectively, with the time units being minutes.

3.1 Standard Normal Errors

In the ideal situation of normal errors all methods yield unbiased results, due to the symmetry of the underlying error distribution.

Repeated median and deepest regression do not perform much worse than least squares (LS) regression whilst the LMS and LTS estimates spread much further (cf. Table 4). The similar behaviour of LMS and LTS can be explained by the fact that both pick about 50% of the observations which can be optimally described by a straight line, without restrictions for symmetry, while RM and DR seek for a balanced fit.

		LMS	LTS	RM	DR	LS
standard normal	$n = 21$	0.875	0.887	0.500	0.533	0.420
errors	$n = 31$	0.767	0.785	0.422	0.450	0.352
heavy tailed	$n = 21$	0.544	0.551	0.345	0.354	0.413
errors	$n = 31$	0.450	0.455	0.279	0.287	0.342
skewed	$n = 21$	0.489	0.495	0.353	0.384	0.429
errors	$n = 31$	0.389	0.399	0.285	0.317	0.350

Table 4: Standard deviations for the estimates at standard normal, re-scaled t_3 distributed and re-scaled lognormal data.

As a result, the LMS and LTS online estimates provide only about 20% efficiency as compared to LS, while for DR we have about 61%, and for RM even approximately 70% efficiency. This is consistent with previous research, and the results here reflect the fact that for small samples LMS regression is slightly more efficient than LTS regression (Rousseeuw and Leroy 1987, p.183).

3.2 Heavy Tails and Skewness

As real data sets may contain large aberrant values, the normal distribution is often not appropriate to model the error term. Therefore, we examine errors from a re-scaled t-distribution with three degrees of freedom and unit variance as well as errors from a shifted lognormal distribution with zero median and unit variance.

At the t_3-distribution, all methods yield unbiased results because of symmetry. Again, the results for LMS and LTS regression are similar, like those for RM and DR. The standard deviations (cf. Table 4) show that compared to the standard normal situation the variability has decreased for all robust methods, while for least squares it remains about the same since its standard deviation only depends on the error variance.

A larger window width causes less variability, but the proportions of the outcomes from the different methods stay approximately the same for different window widths. The LMS and LTS standard deviations are about 60% the size of their values in the standard normal case, but nevertheless they are still outperformed by LS. This is not true for RM and DR, having standard deviations about 66% of their former size, with repeated median regression showing the smallest variability here.

Figure 1 shows boxplots of the results for the online estimates at lognormal errors with a window width of $n = 31$. The black line in the box denotes the median, the grey line the arithmetic mean.

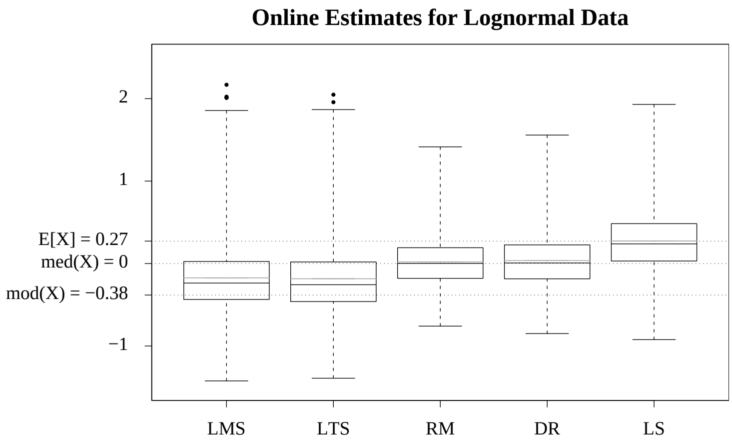

Figure 1: Boxplots of the simulation results for the window width $n = 31$.

The figure clearly shows systematic differences among the considered methods. Rousseeuw, Van Aelst and Hubert (1999) point out that LMS and LTS are 'mode-seeking' in contrast to the 'median-like' behaviour of deepest regression and, as we want to add, the repeated median. Indeed, the LMS and LTS yield estimates mainly between the mode and the median of the underlying error distribution, while RM and DR yield results centred at the median and LS at the expectation.

Since the methods apparently estimate different quantities, an examination of bias is not sensible. Thus, we will only regard variability (cf. Table 4).

The RM and DR standard deviations are only about 70% that for the standard normal situation, and for LMS and LTS they are only about half as large. Comparing the results of the robust methods to least squares we see that the RM standard deviation is only slightly more than 80% as large as the corresponding least squares value, while the DR standard deviation is approximately 90% as large. LMS and LTS on the other hand perform again worse than least squares where LTS shows a little more variability than LMS. Hence, again the repeated median provides the best results w.r.t. variability.

Also, in the context of online monitoring of vital parameters it might be more sensible to estimate the median instead of the mode as the signal can rather be assumed to lie within the centre of the data. Due to the discreteness of the measurements for the considered data it may happen that the mode equals the maximum (or the minimum) while the rest of the data cover a wide range of values. Because of this and the previous results on variability, RM and DR might be preferable to LMS and LTS in case of skewed errors.

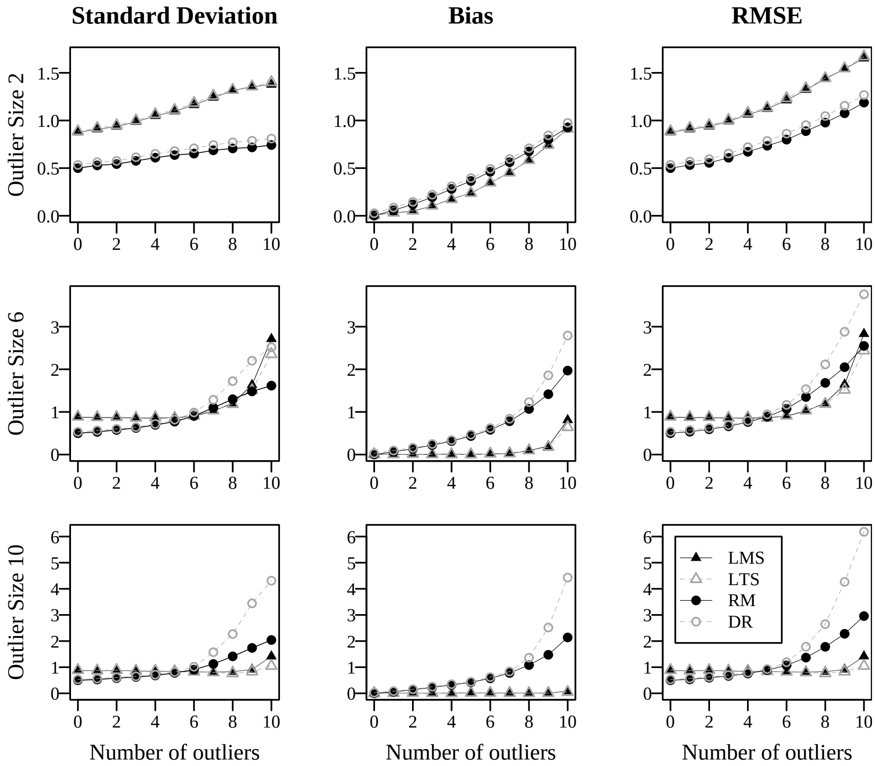

Figure 2: Standard deviation, bias and root mean squared error ($RMSE$) for the online estimates at standard normal data with additive outliers at random time points.

3.3 Additive Outliers

In intensive care, data suffer from a broad variety of perturbations, either caused by medical reasons or by external sources such as a loose cable. As these disturbances often produce similar deviations at several time points, we investigate the influence of additive outliers with same sign and size.

We generate samples from a standard normal model and add a value $a \in \{2, 4, 6, 8, 10\}$ to an increasing number $k \in \{1, \ldots, 10\}$ of observations chosen at random from the sample. Negative additive outliers would yield analogous results. For the sake of brevity we only consider the sample size $n = 21$.

We note that positive outliers at random time points do not cause a bias for the slope, but for the level estimation. The unbiasedness for the slope is due to symmetry since the effects of outliers to the right and to the left of the window centre cancel out on average. In contrast, positive outliers may systematically raise the level estimate, causing a positive bias for the level in the centre, which also affects the online estimates. Figure 2 shows standard deviation, bias and root mean squared error ($RMSE$) of the online estimates for outliers of size 2, 6 and 10. Results for outlier sizes of 4 and 8 lie in between.

Again, the similarity of the RM and DR outcomes shows up clearly, and the differences in the results between LMS and LTS regression are negligible. LTS is only slightly less variable than LMS for a large number of $9-10$ outliers. We also see that LMS and LTS are more heavily affected by smaller outliers than by larger ones.

Comparing repeated median and deepest regression, RM is preferable here as it yields a smaller standard deviation and bias for all considered numbers and sizes of outliers. However, this advantage is only significant in case of seven or more outliers in accordance with the lower breakdown point of DR.

Overall, LMS and LTS perform best in terms of bias although with respect to the $RMSE$ they only outperform the other methods in case of many large outliers. For small outliers or a small to moderate number of outliers the repeated median should be preferred as it has the smallest $RMSE$.

3.4 Outlying Sequences

For online monitoring it is of special importance to track sudden jumps in the signal because this may point at an abrupt change of the patient's state. Looking at single time windows such a *level shift* is indicated by a patch of outlying values of the same size and sign at the end of a time window.

We simulate such situations by generating positive additive outliers of the same size as in the previous subsection - only that now the value $a \in \{2, 4, 6, 8, 10\}$ is added to $k \in \{1, \ldots, 10\}$ subsequent values at the end of the time window. Again, only the case $n = 21$ is investigated.

As the online estimates approximate the level at the end of the window, a small bias w.r.t. level in the centre of the time window is not necessarily what we aim at: for the monitoring data at hand, as a medical rule of thumb a sequence of five or more largely deviating values is assumed to indicate a shift whereas a smaller number is typically regarded as series of outliers (Imhoff, Bauer, Gather and Fried 2003). Hence, a method performs well if it maintains the central level in case of a few subsequent outliers but jumps up to the level of these largely deviant observations when their number is five or more - to estimate the new (higher) level rather than the former (lower) one in the centre of the window.

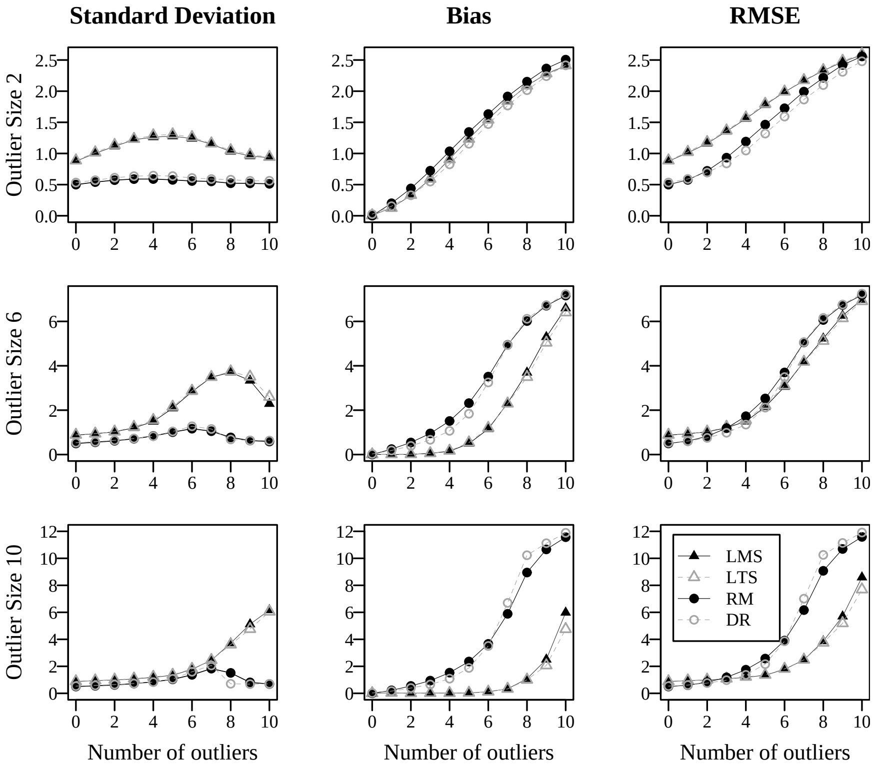

Figure 3: Standard deviation, bias and root mean squared error ($RMSE$) for the online estimates at standard normal data with additive outliers, occurring subsequently at the end of the time window.

Again, Figure 3 shows standard deviation, bias and $RMSE$ only for outlier sizes 2, 6 and 10 as the results for the sizes 4 and 8 lie in between.

No method shows exactly the bias behaviour described above, although for medium-sized to large outliers the LMS and LTS bias curves remain constantly low for a smaller number of outliers and then show a sudden drastic increase. However, the number of outliers which is necessary to make the LMS or LTS bias increase is the larger, the larger the size of the outliers is. In other words: the LMS and LTS online estimates follow a large level shift with a considerable delay, in contrast to the estimates obtained by these methods in the window centre, see Davies et al. (2004). On the other hand, the RM and DR estimates typically smear a moderately large shift.

Figure 3 further reveals that the standard deviations of RM and DR are always smaller than those of LMS and LTS regression, and also that they stay almost constant. Again, the difference between LMS and LTS, and between RM and DR is negligible, both with respect to bias and variability, in spite of the different breakdown points of the latter.

The time delay of the LMS and LTS online estimates w.r.t. the tracking of level shifts could be reduced by choosing a higher quantile instead of the median for LQS regression, or a higher proportion h for LTS regression. Inspection of bias curves shows that such methods smear shifts, according to their smaller breakdown point, although - depending on the quantile or the h chosen - the smearing might be less than for RM or DR. However, using such LQS or LTS methods with smaller breakdown point would lower the robustness against outliers in general, representing a disadvantage as compared to RM regression with its 50% breakdown point. Also, the strict separation of 'old' and 'new' estimated level by the LMS and LTS methods with 50% breakdown point is advantageous when applying automatic rules for shift detection like those proposed by Fried (2004).

4 Application to Time Series

In this section, we analyse the stability of the estimates as well as their ability to track trends, slope changes and sudden level shifts by applying them to a simulated and to a real time series. In both cases we use a window width of $n = 21$ observations.

The simulated time series is 250 time units long and consists of a signal containing constant as well as trend periods and a level shift, plus additive standard normal noise. 10% of the observations are replaced by positive additive outliers of size 6, which are bundled in patches of four subsequent outliers (twice), three outliers (twice), two outliers (three times), and single outliers (five times). The starting point of each sequence is chosen at random.

Figure 4 shows the online estimates and the underlying signal for the simulated times series. All methods trace the trends and the slope changes. Also, the similarity of the results from LMS and LTS regression as well as from repeated median and deepest regression shows up clearly. RM and DR yield more stable results than LMS and LTS, and they are less influenced by values deviating moderately from the underlying signal, e.g. see the results around time points $50 - 60$ and around time point 150.

As the online estimates are based on a linear extrapolation of the level estimates for the centre of the time window, the LMS and LTS estimates continue the pre-level-shift trend until some time points after the level shift. This is due to their small bias with respect to the 'old' level before the shift.

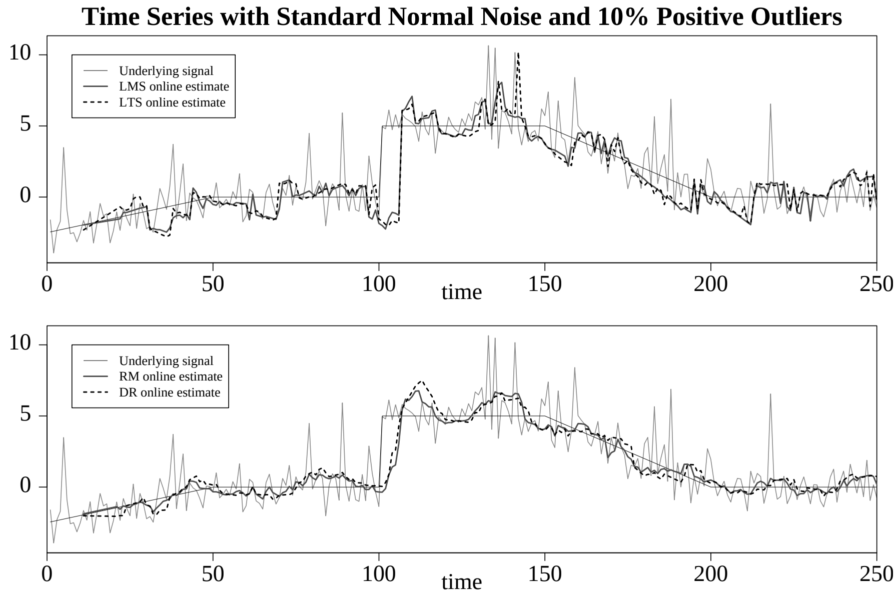

Figure 4: Online estimates based on windows of size $n = 21$.

Repeated median and deepest regression trace the level shift with a shorter delay than LMS or LTS, but they do not capture the abruptness of the jump. Also, the RM and DR estimates are closer to the signal around the times of a slope change - especially around the times 150 and 200. After the transition to the 'new' level, after the shift, all methods overestimate the signal, due to the strongly positive slope estimates around the shift. This phenomenon is well-known when using local linear fits, see e.g. Einbeck and Kauermann (2003).

Finally, we apply the methods to a medical time series of length 250, representing the mean pulmonary artery blood pressure of an intensive care patient. Figure 5 shows that RM and DR yield much more stable results while LMS and LTS are affected by moderate variation in the data.

The RM and DR method trace the level shift around time point 70 better than LMS or LTS regression. Also, LMS and LTS overestimate the level right after the shift by far more drastically. However, they capture the abruptness of the shifts better (e.g. around the times 150 and 175) while RM and DR smear them. Here, the analyst must decide whether it is better to get a 'smeared' transition from one level to the other, or to catch the suddenness of the jump with some time delay.

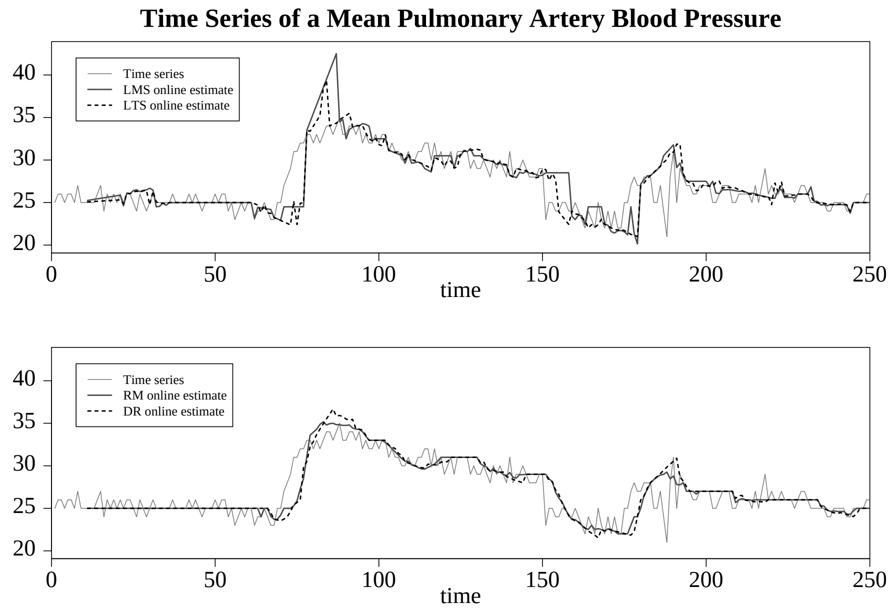

Figure 5: Online estimates based on windows of size $n = 21$.

Both examples show the superiority of repeated median and deepest regression in terms of stability. Also, the repeated median does not overestimate the signal after a shift as much as deepest regression.

5 Discussion

All of the considered methods follow trends and slope changes and trace level shifts quite well. The differences in the outcomes from least median of squares and least trimmed squares regression are negligible while repeated median and deepest regression also show very similar results.

For symmetric, unimodal errors all methods provide unbiased estimates of the median and the mode, which are identical in this case; in case of unimodal, but skewed errors, the LMS and LTS estimates lie somewhere in between the median and the mode while RM and DR estimate the median.

LMS and LTS are less biased than RM and DR in the presence of many large outliers. However, as explained in Section 3.4, in case of a level shift a small bias does not mean better performance of the online estimates. Although RM and DR smear a shift somewhat, these methods still might be preferred because LMS and LTS follow a shift with a longer delay - especially if the shift is large. In spite of the claim that deepest regression is particularly appropriate for skewed errors due to its construction, the repeated median performed even better for lognormal errors.

Further, repeated median and deepest regression yield a more stable signal extraction; and the LMS and LTS estimates are stronger influenced by small or medium-sized outliers.

Summarising, repeated median and deepest regression outperform LMS and LTS regression w.r.t. online signal extraction without delay. Repeated median regression yields the best results in most respects: among these robust methods, RM is the least variable in most of the considered situations; it gives stable estimations in the applications and also, it is computationally the fastest.

Other methods such as the *least quartile difference* (LQD) *estimator* (Croux, Rousseeuw and Hössjer 1994) or the *least trimmed differences* (LTD) *estimator* (Stromberg, Hössjer and Hawkins 2000) might also be worth considering in the online monitoring context. Both methods are based on pairwise absolute differences between the observations. They estimate the slope by searching an optimal subset of these differences for which different error criteria are minimised. This principle is similar to LMS and LTS regression. The level can then be estimated by the median of the trend-corrected observations - just like for RM. Therefore, these methods possibly mean a compromise between LMS/LTS and RM. Accordingly, LQD and LTD are considered to be more stable and better suitable for asymmetric errors than LMS or LTS regression. However, as Stromberg et al. (2000) point out, both methods still show instabilities towards moderate variation in the data - in contrast to RM.

A possible problem is the time needed for exact computation of LQD and LTD in a high risk environment like intensive care: present exact algorithms for LQD and LTD are based on LMS and LTS fits to the pairwise differences (Stromberg et al. 2000). Faster algorithms for both methods are currently under research.

The regression methods investigated in this paper serve for an initial approximation of the signal value. More refined methods for signal extraction can be constructed by subsequent application of another method. Reweighted least squares based on an initial LMS fit is a popular approach to robust regression but can also show instabilities in an automatic application due to the initial use of the LMS. Instead, Bernholt, Fried, Gather and Wegener (2004) suggest reweighted least squares based on an initial RM fit. Application of rank regression (Chang, McKean, Naranjo and Sheather 1999) instead of least squares seems also promising because of its nonparametric flavour and may give an impulse for further research.

Acknowledgement

We thank an anonymous referee for several insightful suggestions, particularly, for pointing out LQD and highly robust rank regression. Financial support of the *Deutsche Forschungsgemeinschaft* (SFB 475: 'Reduction of Complexity for Multivariate Data Structures') is gratefully acknowledged.

References

Bai, Z.D. and He, X. (1999). Asymptotic Distributions of the Maximal Depth Estimators for Regression and Multivariate Location, *Ann. Stat.* ***27*** *(5)*, 1616-1637.

Bernholt, T. (2004). Exact Algorithms for the Repeated Median, LMS, LTS and Deepest Regression, *Personal Communication*.

Bernholt, T. and Fried, R. (2003). Computing the Update of the Repeated Median Regression Line in Linear Time, *Inf. Process. Lett.* ***88*** *(1)*, 111-117.

Bernholt, T., Fried, R., Gather, U. and Wegener I. (2004). Modified Repeated Median Filters, *Technical Report* ***46****, SFB 475, University of Dortmund, Germany.*

Chang, W.H., McKean, J.W., Naranjo, J.D. and Sheather, S.J. (1999). High-Breakdown Rank Regression, *J. Am. Stat. Assoc.* ***94****, No. 445*, 205-219.

Croux, C., Rousseeuw, P.J. and Hössjer, O. (1994). Generalized S-Estimators, *J. Am. Stat. Assoc.* ***89****, No. 428*, 1271-1281.

Davies, P.L. (1993). Aspects of Robust Linear Regression, *Ann. Stat.* ***21*** *(4)*, 1843-1899.

Davies, P.L., Fried, R. and Gather, U. (2004). Robust Signal Extraction for On-line Monitoring Data, *J. Stat. Plann. Inference* ***122*** *(1-2)*, 65-78.

Edelsbrunner, H. and Souvaine, D.L. (1990). Computing Least Median of Squares Regression Lines and Guided Topological Sweep, *J. Am. Stat. Assoc.* ***85****, No. 409*, 115-119.

Einbeck, J. and Kauermann, G. (2003). Online Monitoring with Local Smoothing Methods and Adaptive Ridging, *J. Statist. Comput. Simul.* ***73***, 913-929.

Fried, R. (2004). Robust Filtering of Time Series with Trends, *J. Nonparametric Statistics* ***16***, 313-328.

Hössjer, O., Rousseeuw, P.J. and Ruts, I. (1995). The Repeated Median Intercept Estimator: Influence Function and Asymptotic Normality, *J. Multivariate Anal.* ***52***, 45-72.

Hössjer, O., Rousseeuw, P.J. and Croux, C. (1994). Asymptotics of the Repeated Median Slope Estimator, *Ann. Stat.* ***22*** *(3)*, 1478-1501.

Imhoff, M., Bauer, M., Gather, U. and Fried, R. (2002). Pattern Detection in Intensive Care Monitoring Time Series with Autoregressive Models: Influence of the AR-Model Order, *Biom. J.* ***44***, 746-761.

Rousseeuw, P.J. (1983). Multivariate Estimation with High Breakdown Point, in W. Grossmann, G. Pflug, I. Vincze, W. Wertz (eds.) *Proceedings of the 4th Pannonian Symposium on Mathematical Statistics and Probability, Vol. B*, D. Reidel Publishing Company, Dordrecht (The Netherlands).

Rousseeuw, P.J. (1984). Least Median of Squares Regression, *J. Am. Stat. Assoc.* ***79****, No. 388*, 871-880.

Rousseeuw, P.J. and Hubert, M. (1999). Regression Depth, *J. Am. Stat. Assoc.* ***94****, No. 446*, 388-402.

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

Rousseeuw, P.J., Van Aelst, S. and Hubert, M. (1999). Rejoinder to 'Regression Depth', *J. Am. Stat. Assoc.* ***94****, No. 446*, 419-433.

Sheather, S.J., McKean, J.W. and Hettmansperger, T.P. (1997). Finite Sample Stability Properties of the Least Median of Squares Estimator, *J. Stat. Comput. Simul.* ***58***, 371-383.

Siegel, A.F. (1982). Robust Regression Using Repeated Medians, *Biometrika* ***69***, 242-244.

Stromberg, A.J., Hössjer, O., Hawkins, D.M. (2000). The Least Trimmed Differences Regression Estimator and Alternatives, *J. Am. Stat. Assoc.* ***95****, No. 451*, 853-864.

Van Aelst, S., Rousseeuw, P.J., Hubert, M. and Struyf, A. (2002). The Deepest Regression Method, *J. Multivariate Anal.* ***81***, 138-166.

Van Kreveld, M., Mitchell, J.S.B., Rousseeuw, P.J., Sharir, M., Snoeyink, J. and Speckmann, B. (1999). Efficient Algorithms for Maximum Regression Depth, *Proceedings of the 15th Annual ACM Symposium of Computational Geometry*, ACM Press, New York (NJ), 31-40.