Robust Estimation of Scale for Local Linear Temporal Trends

Roland Fried and Ursula Gather

Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany

{gather, fried}@statistik.uni-dortmund.de

Abstract. Online monitoring data measured in intensive care exhibit periods of relative constancy corresponding to a steady state and interruptions representing slow monotonic trends and level shifts. These signals are overlaid with a high level of noise and many measurement artifacts. Methods are needed that allow a fast and reliable denoising of the data and separate artifacts from information of clinical relevant changes in the patient's condition. The underlying signal can be approximated by fitting a local linear trend within a moving time window using e.g. the repeated median (Davies, Fried and Gather, 2002). We compare the finite-sample performance of robust functionals for scale estimation in trend periods as these can be used to detect outliers and level shifts.

Keywords: Signal extraction, linear regression, level shift, trend, outliers, small-sample efficiency.

Mathematical Subject Classification: primary 62G07; secondary 62G35

1 Introduction

In intensive care physiological variables like the heart rate are recorded at least every minute. For suitable bedside decision support in time-critical situations, systematic changes in the variables have to be detected quickly and distinguished from clinical irrelevant short term fluctuations and measurement artifacts. Therefore, methods working in real-time are needed which decompose the observed time series into a smooth underlying signal and rough noise terms (Wernecke et al., 1988, Gather et al., 2002).

Tukey (1977) suggests median filtering to prefilter a noisy time series, to clean the data from artifacts and to extract an underlying signal. Atypical observations (outliers) and sudden changes (level shifts) in the data generating mechanism can be detected if we additionally apply a reliable estimator of scale like the Median Absolute Deviation about the Median (MAD). However, in view of high sampling frequencies most changes occur gradually, and in trend periods most scale estimators are strongly biased. Davies, Fried and Gather (2002) fit a local linear trend to the data using e.g. the repeated median (Siegel, 1982) because of its high robustness and its computational speed. This provides information on the occurrence of trends and it may improve the estimation of scale.

In this paper, we compare the finite-sample performance of functionals for scale estimation in trend periods. Important aspects for online monitoring are the unique existence of the estimate in any data situation, low computation time, high robustness and satisfactory finite sample efficiency. Functionals which do not have a unique solution in any data situation are useless in an automatic online application, just like functionals which afford too much computation time. Therefore we restrict to functionals for which explicit formulae in terms of e.g. order statistics exist to guarantee unique existence and to avoid iterations. This excludes most M- and S-functionals for instance.

W.r.t. robustness we take a global perspective considering the finite-sample replacement breakdown point and bias curves. The breakdown point should ideally be 50%, and (patches of) outliers should not cause a large bias. Satisfactory finite-sample efficiency in a Gaussian sample is preferable but it is not an overriding concern here.

We proceed as follows. In Section 2 we review some robust functionals for scale estimation. In Section 3 we perform a simulation study to compare the finite-sample properties of these methods when applied to a single time window with a linear trend. In Section 4 applications to real and simulated time series are presented before we give some conclusions.

2 Methods

Let $x_1, \ldots, x_N$ be real valued data measured at time points $t = 1, \ldots, N$. We assume that there is an underlying signal μ_t, $t = 1, \ldots, N$, that is overlaid by additive random noise. In order to separate signal and noise we assume that the signal is simple, i.e. smooth with possibly a few sudden changes, while the noise $E_1, \ldots, E_N$ consists of independently distributed random variables with mean zero. Hence, we consider $x_1, \ldots, x_N$ to be a realization of random variables $X_1, \ldots, X_N$, that are decomposed as

$$X_t = \mu_t + E_t, t = 1, \ldots, N.$$

Our main concern is the estimation of $\sigma_t = \sqrt{Var(E_t)}$.

Data in intensive care monitoring cannot be assumed to be stationary as there are slow monotonic changes and the variance σ_t^2 may vary over time as well (Gather et al., 2002). In order to accommodate such changes we approximate μ_t and σ_t using a moving time window of small or moderate length $n = 2m + 1$. For simplicity we renumber the observations in the current time window by $x_{-m}, \ldots, x_m$ in the following.

2.1 Robust scale estimators

Median filtering is a highly robust method for extraction of an underlying signal which preserves sudden changes in the data. It implicitly assumes that the level is almost constant in the current time window, i.e. $\mu_t = \mu$ and $\sigma_t = \sigma$, $t = -m, \ldots, m$. Rousseeuw and Croux (1992) discuss a couple of explicit robust scale estimators in this location-scale model. They consider the influence function, the gross error sensitivity and the breakdown point, and they compare the finite sample efficiencies by simulation. We select some promising scale estimators based on their and other author's results for further comparison. All these estimators can be calculated in $O(n \log n)$ time using suitable algorithms.

The perhaps most commonly used robust scale estimator in the location-scale model is the median absolute deviation about the median $\tilde{\mu}$

$$MAD = c \cdot med_t\{|x_t - \tilde{\mu}|\}.$$

Here, c is a small sample correction factor that can be chosen depending on the sample size to achieve unbiasedness. The breakdown point of the MAD is asymptotically 50% and its gross-error sensitivity is the smallest possible for a Fisher-consistent scale estimator, namely 1.17. It is the minimax asymptotically biased Huber M-estimator in a model with symmetric (strongly) unimodal noise distribution (Hampel et al., 1986, Martin and Zamar, 1993). The asymptotic efficiency of the MAD is 36.7%, which is rather small. Collins (1999) points out that the discontinuity of the influence function of the MAD causes its asymptotic variance to increase in an $\epsilon-$contamination neighbourhood with arbitrarily small ϵ resulting in an even smaller efficiency of 14.5% then.

While the MAD is a location-based estimator as it measures the deviations of the observations from a (robust) location estimate, the interquartile range

$$IQR = c\left(x_{(n-\lfloor n/4\rfloor)} - x_{(\lfloor n/4\rfloor+1)}\right),$$

where $x_{(1)}, \dots, x_{(n)}$ is the ordered sample, is a location-free estimator. Location-free estimators have the advantage that they do not implicitly rely on a symmetric noise distribution. At symmetric distributions, IQR has the same influence function as the MAD. However, its breakdown point is only 25%.

A location-free estimator with breakdown point 50% can be derived using the nested scale functional

$$SN = c \cdot med_i med_{j\neq i}|x_i - x_j|\,.$$

Its asymptotic efficiency is 58.2%, which is larger than those of similar functionals with the outer median replaced by other quantiles. Replacing the outer median by a minimum for instance results in an estimator with the same influence function and the same asymptotic efficiency like the MAD. The influence function of SN is also discontinuous, and its gross error sensitivity is 1.625. An algorithm for computation of SN in $O(n \log n)$ time is described by Croux and Rousseeuw (1992).

Another location-free 50% breakdown estimator is obtained using a trimmed mean of median deviations

$$TMM = c \cdot \sum_{k=1}^{m} \{med_{j\neq i}|x_i - x_j| : i = -m, \dots, m\}_{(k)}.$$

TMM has a continuous influence function and a gross error sensitivity of 1.458 but smaller asymptotic efficiency than SN, namely 52%. Replacing absolute by squared deviations results in an estimator with a slightly larger asymptotic efficiency (53%) and slightly larger gross error sensitivity (1.467), that is not considered further on here.

Rousseeuw and Croux (1993) propose a location-free estimator with 50% breakdown point and higher efficiency based on an order statistic of all pairwise differences

$$QN = c \cdot \{|x_i - x_j| : -m \leq i < j \leq m\}_{(h)}\,,\; h = \binom{m+1}{2}\,.$$

Its asymptotic efficiency is 82.3%, which is larger than that of estimators based on weighted sums of the smallest $\binom{m+1}{2}$ absolute or squared differences. Its influence function is smooth and its gross error sensitivity is 2.069. It can also be computed in $O(n \log n)$ time (Croux and Rousseeuw, 1992).

The length of the shortest half (Grübel, 1988, Rousseeuw and Leroy, 1988)

$$LSH = c \cdot min\{|x_{(i+m)} - x_{(i)}|; i = -m, \dots, 0\}$$

is another location-free scale estimator with 50% breakdown point. It has the same influence function and thus the same asymptotic efficiency as the MAD. For small samples, however, its efficiency is larger than that of the MAD, and its maximum bias in case of many outliers is smaller than that of the MAD. The reason is that the MAD is centered by the median, that itself is influenced by a large percentage of outliers, while a location estimator is not necessary for the LSH. Indeed, the shorth is the minimax bias robust scale estimator with general location (Martin and Zamar, 1993).

Replacing the range by the standard deviation we get a least trimmed squares estimator

$$LTS = c \cdot min_i sdv\{x_{(i)}, \ldots, x_{(i+m)}, i = -m, \ldots, 0\}\,.$$

Its asymptotic efficiency is smaller than that of the LSH, namely 30.7%, but its influence function is continuous. Croux and Haesbroeck (2001) find LTS to resist outliers better than estimators based on subranges like the LSH, while LSH performs better for inliers.

While for LTS the standard deviation needs to be calculated for all subranges, this is not necessary if we use location-based trimming like

$$SMAD = \sqrt{\frac{1}{m+1} \sum_{k=1}^{m+1} \{(x_i - \tilde{\mu})^2; i = -m, \ldots, m\}_{(k)}},$$

Rousseeuw and Croux (1992) call this estimator a smoothed MAD as it is similarly defined like the MAD and has a continuous influence function. For $SMAD$ we use asymmetric trimming based on absolute values to achieve a breakdown point of 50%. The breakdown point of the 50%-trimmed standard deviation

$$TS = \sqrt{\frac{1}{m+1} \sum_{k=\lfloor n/4 \rfloor + 1}^{n - \lfloor n/4 \rfloor} \{(x_i - \tilde{\mu})^2; i = -m, \ldots, m\}_{(k)}}$$

is 25% only as it trims symmetrically w.r.t. the location estimate. It has the same influence function as LTS and $SMAD$.

All these scale estimators are strongly biased in trend periods. An estimator which avoids this bias can be based on pairs of subsequent observations (x_i, x_{i+1}) instead of all pairs (x_i, x_j). The maximal breakdown point is 25% then obtained by using

$$MAS = c \cdot med\{|x_{i+1} - x_i|, i = -m, \ldots, m-1\}\,.$$

While for QN there are $n(n-1)/2$ pairs, there are only $n-1$ pairs of subsequent observations. Hence, the efficiency of MAS is much smaller than that of QN in the steady state, but it is much less biased in trend periods.

We do not consider M-estimators or S-estimators here which need to be calculated iteratively. Optimal B-robust M-estimators e.g. are appealing but they cannot attain an efficiency of more than 50.6% if we insist on a 50% breakdown point (Hampel et al., 1986). Highly efficient and high-breakdown M-estimators suffer from a strongly increasing (finite) bias for contamination much smaller than 50% (Croux, 1994), whereas k-step M-estimators which also combine 50% breakdown point and high efficiency typically have discontinuous influence function and large gross-error sensitivity, and their bias curve is close to that of the exact M estimator in spite of the higher breakdown point (Rousseeuw and Croux, 1994). The above functionals can be used for fast preprocessing of online monitoring data as well as for initialization of algorithms which compute robust M-estimators.

2.2 Robust filtering with trends

For better estimation of monotonic trends we can approximate not only the level μ_0 in the center but all levels μ_t in the current time window $t = -m, \ldots, m$. We assume a trend to be locally linear, i.e.

$$X_t = \mu_0 + t\beta_0 + E_t, \quad t = -m, \ldots, m, \tag{1}$$

where $\mu_t \approx \mu_0 + t\beta_0, t = -m, \ldots, m$, is the level at the t-th time point in the current window and β_0 is the slope in the center (as the slope is assumed to be approximately constant in the window). Thus, we need to approximate the level μ_0, the slope β_0 and the scale σ_0.

Davies et al. (2002) compare the small sample performance of L_1-regression, the repeated median (Siegel, 1982) and the least median of squares (Hampel, 1975, Rousseeuw, 1984) for extraction of a local linear trend. The first two methods have higher efficiency, while the least median of squares has the smallest bias and MSE if there are between 20% and 40% outlying observations in the time window. The repeated median performs somewhat better than L_1-regression in accordance to its higher breakdown point of 50%. Since analysis of real and simulated data shows that the least median of squares can be seriously mislead by certain patterns in the time window, we use the repeated median

$$\begin{aligned} \tilde{\beta}_0 &= med_i \left(med_{j \neq i} \frac{x_i - x_j}{i - j} \right) \\ \tilde{\mu}_0 &= med_i (x_i - \tilde{\beta}_0 i) \, , \end{aligned}$$

in the following to approximate μ_0 and β_0. We can then estimate σ_0 from the trend corrected observations $r_t = x_t - t\tilde{\beta}_0$, $t = -m, \ldots, m$. Any of the estimators mentioned above can be used for this purpose. To achieve unbiasedness we derive finite sample correction factors for all methods by simulation.

3 Comparison within a single time window

Most of the properties of the above scale estimators are based on large sample asymptotics. For an application to online monitoring data, however, we have to choose small to medium window widths $n = 2m + 1$ as we assume the underlying approximations to hold only locally and since the time delay m must not be too large. The preliminary robust linear regression step may also influence the properties of the scale estimators. Therefore, we compare the finite-sample performance of the methods for a single time window in the following. We simulate data from the model

$$Y_t = \mu + \beta \cdot t + E_t, t = -m, \ldots, m,$$

with $\mu = 0$ and for several different slopes β. The error E is always Gaussian white noise with zero mean and unit variance.

3.1 Efficiency

First we calculate the small sample efficiencies measured by the mean square error MSE of the distinct methods for various sample sizes, cf. Table 1. For each setting 20000 samples are generated. All estimators are unbiased when using finite sample correction factors with the only exception MAS, which shows an increasing positive bias with increasing slope. We find this bias to be small for a realistic slope $|\beta| \leq 0.5$ and all sample sizes considered here. Thus, we concentrate on $\beta = 0$. Obviously, the asymptotic efficiencies are good approximations for most of the estimators if $n \geq 20$. Only QN has a much smaller finite sample efficiency which converges slowly. Nevertheless, its finite sample efficiency is the largest among these robust methods, and the differences increase with n. The ordering w.r.t. finite sample efficiency is

$$QN > SN > TMM > LSH > IQR > LTS \approx MAD > SMAD \approx TS \approx MAS$$

for $n \geq 30$. We also consider a trend-corrected version $MASC$ of MAS but do not find a noteworthy improvement over MAS for realistic slopes here and in the following, and thus neglect it.

3.2 Small contamination

In online monitoring, rules for the replacement of detected outliers can be formulated to improve the results. However, we cannot completely rely on such rules since especially outliers that are not very large may remain undetected. Therefore, the robustness of the methods against some small to moderate outliers is important. Since the repeated median is regression equivariant we restrict to $\mu = \beta = 0$, and we consider the window width $n = 31$ only. We replace an increasing number of 0(1)7 observations by additive outliers of increasing size 2(1)5 with random sign at random time points in the window. This corresponds to a percentage of contaminated observations between 0% and 22.5%. Each of the 32 cases is simulated 10000 times and the squared bias, variance and MSE are calculated.

Table 1: Finite sample efficiencies relatively to the MLE measured by the simulated MSE for N(0,1) errors.

m	β	MAS	MAD	SN	TMM	QN	LSH	LTS	$SMAD$	TS	$MASC$	IQR
5	0.0	41.7	32.2	47.8	44.7	49.8	34.6	33.9	29.9	40.0	40.6	42.7
10	0.0	35.6	34.6	51.6	47.1	61.3	38.6	35.8	30.6	31.0	35.5	35.2
15	0.0	33.5	35.0	54.4	48.3	66.4	39.5	35.7	31.1	34.1	33.6	38.6
20	0.0	32.9	35.5	55.4	48.9	69.7	40.2	35.5	30.7	30.7	32.9	35.5
25	0.0	33.0	35.4	56.0	49.2	72.0	40.1	34.9	30.1	32.0	33.0	37.5
30	0.0	31.7	36.4	57.1	50.4	73.6	40.9	35.2	31.1	30.7	31.7	36.4
50	0.0	31.2	35.6	57.6	50.9	76.9	40.9	34.4	30.4	30.4	31.2	35.9

Figure 1: Moderate number of small to medium-sized outliers: Simulated MSE for MAS x, MAD $\circ$, SN $\triangle$, TMM $*$, QN $\bigtriangledown$ (left), and LSH x, LTS $\circ$, $SMAD$ $\triangle$, TS $*$, IQR $\bigtriangledown$ (right).

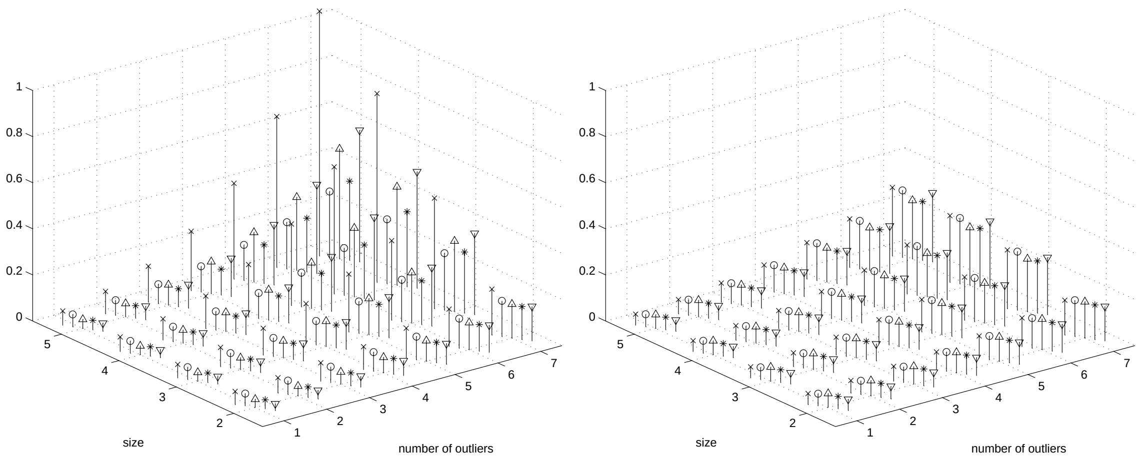


Figure 1 depicts the results for outliers with random sign. Obviously, most of the estimators resist a few small to medium-sized outliers rather well. Only MAS shows poor performance. Closer analysis shows that both its bias and its variance strongly increase with the number and the size of the outliers, while the variance of the other methods is not much affected. QN looses its largest relative efficiency in the presence of three or more moderate outliers because of its increasing bias, and the same is true to a smaller extent for SN. All the other methods show similar performance here.

When considering only one-sided positive outliers (Figure 2), IQR becomes more strongly affected as well, while SN and QN are slightly inferior to the other methods for more than 4 positive outliers. Only TS is worse than these for 7 outliers. For 4 or more outliers, LSH and LTS show the best performance, closely followed by $SMAD$ and TMM and then the MAD.

3.3 Implosion

Variables measured in intensive care, e.g. the central venous pressure, sometimes have little variability relative to the measurement scale. This may result in identical measurements causing scale estimators to become negatively biased and possibly even in zero estimates ("implosion"). Therefore we investigate the effect of identical observations on the scale estimators. For $\mu = \beta = 0$ and $n = 31$ we replace an increasing number 0(1)15 of observations by zero values. Each of the 15 cases is simulated 10000 times and the squared bias, variance and MSE are calculated. Since the variances of all estimators are slightly decreasing with increasing number of zero measurements without important differences we restrict to the bias in Figure 3.

QN shows the best performance here, followed by MAD, TMM and SN, while $SMAD$ and LTS are inferior to the other methods. This may be due to the positive weights given to small deviations. We also considered other settings. In case of an alternation between two values, e.g., the MAD shows a strongly increasing variance and worsens somewhat, while the ordering of the

Figure 2: Moderate number of small to medium-sized positive outliers: Simulated MSE for MAS x, MAD $\circ$, SN $\triangle$, TMM $*$, QN $\bigtriangledown$ (left), and LSH x, LTS $\circ$, $SMAD$ $\triangle$, TS $*$, IQR $\bigtriangledown$ (right).

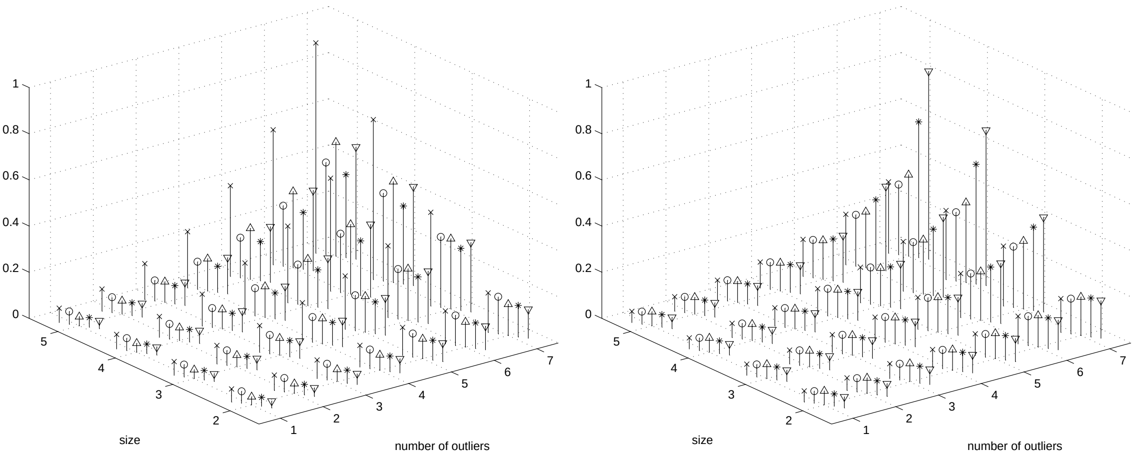
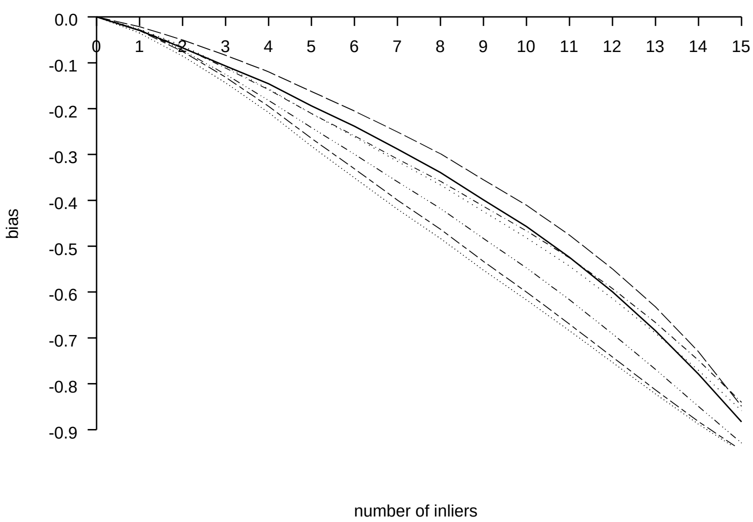


Figure 3: Inliers: Simulated bias for MAD (solid), SN (wide dots), TMM (dash-dot), QN (dashed), LSH (dash-dot-dot-dot), LTS (short dashed) and $SMAD$ (dotted).



other methods remains the same. However, the situation shown here is much worse. MAS, IQR and TS are not considered any longer because of their vulnerability to outliers found above.

3.4 Level shift and explosion

Estimators with identical breakdown points can nevertheless be very differently influenced by many outliers. Berrendero and Zamar (1999) compare the maximal asymptotic bias of MAD, SN, QN, and LSH for almost 50% contamination and find LSH to have the best worst-case behaviour as the maximal bias goes to infinity twice as fast for the MAD as for the LSH, while this rate is 1.6 and 2.12 for SN and QN respectively. Therefore, we analyze the relative performance of the estimators for a large percentage of contamination. Again we consider a window width $n = 31$ and replace an increasing number 0(1)15 of observations by additive outliers. For the outlier generating mechanism we use:

1. Additive outliers of fixed, increasing size 0(2)10.
2. Additive outliers generated from an $N(0, \{2 + j\}^2)$-distribution, $j = 1, \ldots, 4$.
3. Additive outliers generated from an $N(3j, j^2)$-distribution, $j = 1, \ldots, 4$.

Additionally, we insert level shifts of several sizes into the time window as the detection of level shifts is important in online monitoring. Since in intensive care changes that last at least five minutes can often be regarded as clinical relevant, we define a level shift by five outliers of the same size at the end of the time window. For each case 4000 samples are generated and the squared bias, variance and MSE are calculated.

Figure 4 depicts the results for the first setting and a level shift of size 10. LSH, LTS and TMM perform very similar with LTS having the smallest bias and MSE. QN performs worse than these methods but better than MAD and SN in case of many outliers of the same size. The reason is that many outliers of the same size and possibly a level shift mean that we actually sample from several distributions with the same variance and different means, and this influences a method like QN which is based on an α-quantile of the pairwise differences with $\alpha < 0.5$ less than location-based methods like the MAD. The $SMAD$ not shown here performs similar to the MAD. As the $SMAD$ does not seem to provide significant improvements on the MAD in our context we neglect it in the following.

Although outliers of similar size are a typical problem in intensive care monitoring because of e.g. the drawing of blood samples, we also consider outliers with random size and random sign. Figure 5 depicts the results for setting 2. Only the MSE is shown as there are only minor differences w.r.t. the variance, that is much smaller than the squared bias. Here, QN is worse than MAD and SN, while LSH, LTS and TMM are again superior with LTS being slightly better than LSH and TMM.

As one-sided outliers (setting 1) affect the estimators more strongly than outliers with random sign (setting 2), we also considered another setting 3 with positive outliers generated from a normal distribution with increasing mean and variance. Here, we do not find large differences and thus the results are not shown here.

Figure 4: Outliers of fixed positive size and level shift of size 10. Simulated squared bias (top), variance (middle) and MSE (bottom) for LSH x, LTS $\circ$, TMM $\triangle$ (left), and MAD x, SN $\circ$, QN $\triangle$ (right).

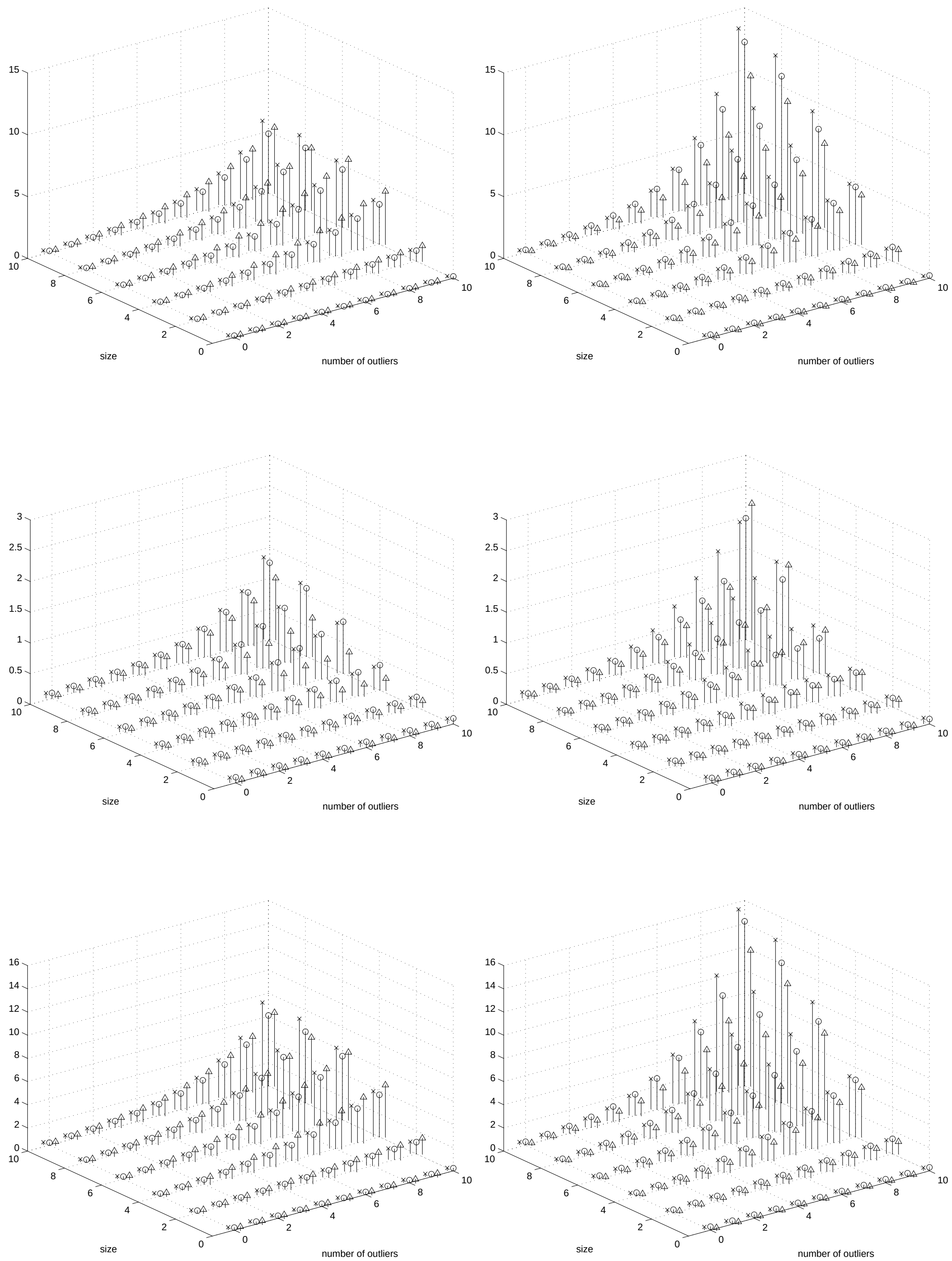
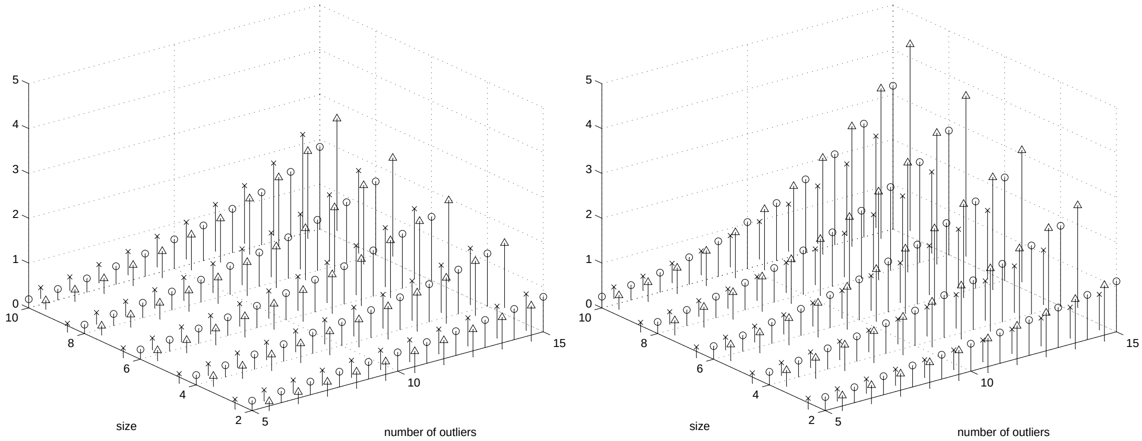


Figure 5: Outliers of random size and random sign, level shift of size 10, MSE for LSH x, LTS $\circ$, TMM $\triangle$ (left), and MAD x, SN $\circ$, QN $\triangle$ (right).



4 Application

Now we check the performance of the distinct methods for time series filtering. We consider two real examples from intensive care monitoring as well as a simulated time series. The performance of the repeated median w.r.t. the real time series is discussed in Davies et al. (2002). Therefore we concentrate on the scale estimators here.

The first time series shown in Figure 6 represents the heart rate of a critically ill patient. There are some patches of clinically irrelevant positive outliers, that cause all estimates to increase. The QN is the most affected method by these outliers. The MAD shows some large fluctuations, while the LSH is considerably smaller than the other methods at about $t = 145$. This might be due to some subsequent observations which are almost identical.

The second time series also shown in Figure 6 represents the arterial blood pressure of another patient. The methods perform rather similarly here with the QN and the MAD being occasionally slightly more affected by outliers than the other methods.

Finally, we consider a simulated time series of length 500. We have inserted linear trend periods between $t = 100$ and $t = 200$ as well as between $t = 300$ and $t = 400$, and σ increases at $t = 300$ from 1 to 2. 10% of the observations have been replaced by additive (patchy) $N(0, 6^2)$-distributed outliers, and another 10% have been replaced by additive outliers of fixed size 6. To explore the effect of identical measurements, we have replaced another 10% of the observations by the preceding values. The results are shown in Figure 7. Again, QN is most affected by the outliers, but it also shows the best performance when identical measurements occur. Between the other methods there is not a large difference.

Figure 6: Top: Time series (dotted) representing heart rate (left) and arterial blood pressure (right) and level approximates (solid). Bottom: Some scale approximates: MAD (solid), TMM (dashed-dotted), QN (dotted), LSH (dashed).

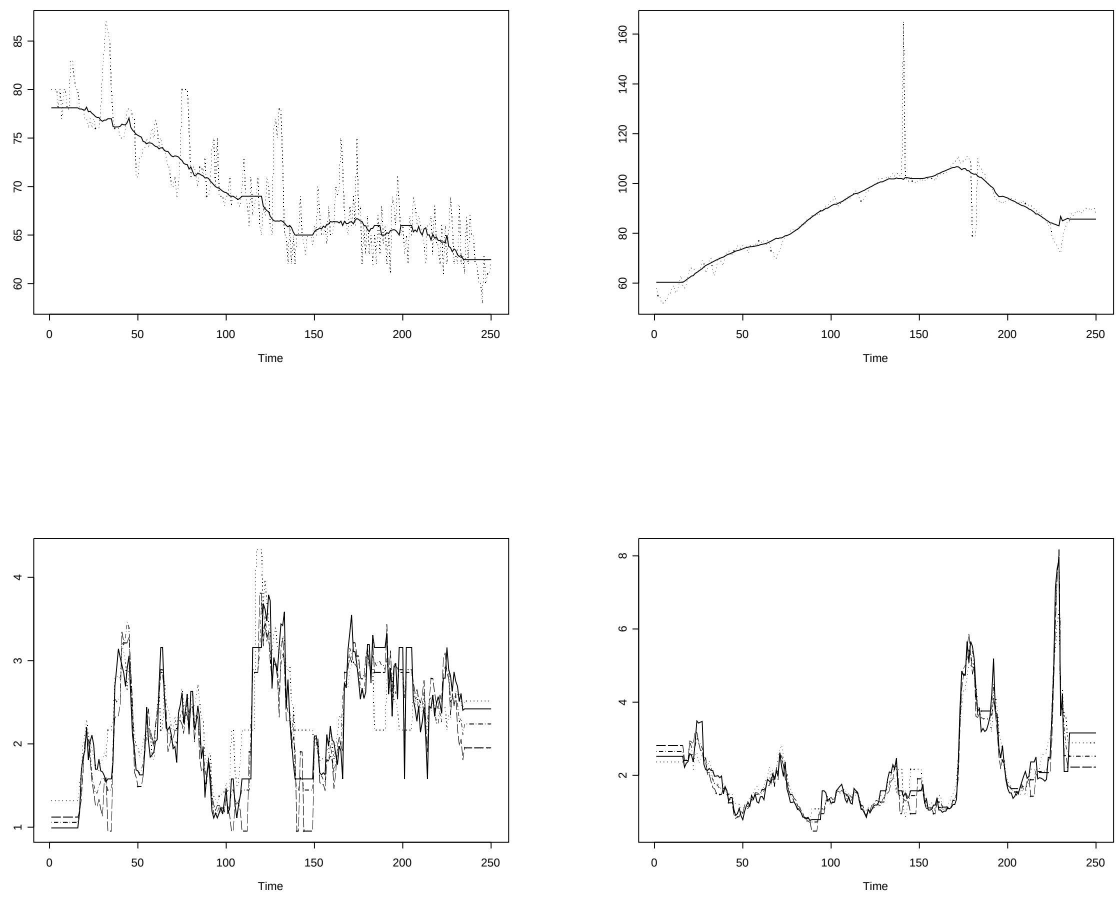
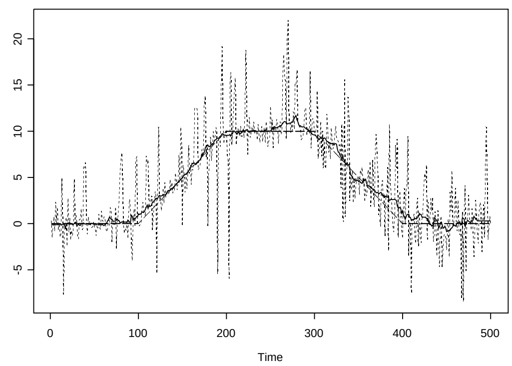
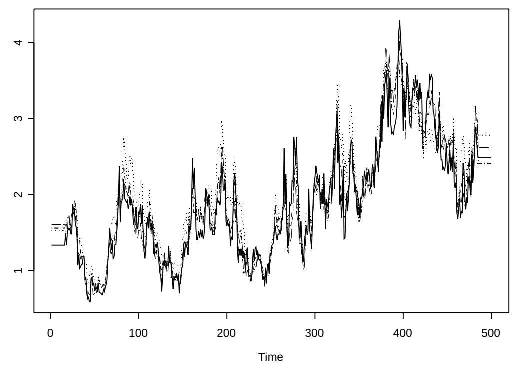


Figure 7: Left hand side: Simulated time series with 20% outliers and 10% inliers (dotted), underlying level (solid) and level approximate (dashed). Right hand side: Scale approximates: MAD (solid), TMM (dashed-dotted), QN (dotted), LSH (dashed). The true scale increases from $\sigma = 1$ to $\sigma = 2$ at $t = 300$.





5 Conclusion

Reliable algorithms for artefact detection should be applied before a very noisy time series can be analyzed finally. We have investigated explicit robust scale estimators for filtering with locally linear time trends. High breakdown point methods have to be recommended as some standard methods like IQR with 25% breakdown point turned out to be strongly influenced already by more than 10% contamination of moderate size. An additional analysis not presented here shows that IQR and TS can be useful if there are many inliers and a few outliers only. Among the high breakdown point methods, none of them has been found to be uniformly superior in all situations. QN and SN have large efficiency in a Gaussian sample, but they loose this advantage if there are about 10% outliers of moderate size. QN performs rather well for level shifts and outliers of the same size as well as for inliers. The MAD provides robustness against outliers without being much affected by some inliers, but it is outperformed by LSH, LTS and TMM in terms of efficiency in a Gaussian sample and in case of many large outliers. LTS and $SMAD$ cannot be recommended if one has to protect against inliers. Altogether, if a large percentage of outliers is more a problem than inliers, LSH and TMM can be recommended for robust scale estimation in time series filtering with TMM providing stronger protection against inliers.

Acknowledgments: We thank Laurie Davies for helpful discussions. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Berrendero, J.R. and Zamar, R.H. (1999). Global robustness of location and dispersion estimates. *Statistics & Probability Letters* 44, 63-72

Collins, J.R. (1999). Robust M-estimators of scale: Minimax bias versus maximal variance. *Canadian J. of Statistics* 27, 81-96.

Croux, Ch. (1994). Efficient high-breakdown M-estimators of scale. *Statistics & Probability Letters* 19, 371-379.

Croux, Ch. and Haesbrock, G. (2001). Maxbias curves of robust scale estimators based on subranges. *Metrika* 53, 101-122.

Croux, C. and Rousseeuw, P.J. (1992). Time-efficient algorithms for two highly robust estimators of scale. In Y. Dodge and J. Whittaker (eds.): *Computational Statistics, Volume 1.* Heidelberg: Physica-Verlag.

Davies, P.L., Fried, R. and Gather, U. (2002). Robust signal extraction from on-line monitoring data. Technical Report 2/2002, SFB 475, University of Dortmund, Germany.

Gather, U., Fried, R., Lanius, V. and Imhoff, M. (2002). Online-monitoring of high-dimensional physiological time series - a case-study. *Estadistica*, in press.

Grübel, R. (1988). The length of the shorth. *The Annals of Statistics* 16, 619-628.

Hampel, F.R. (1975). Beyond location parameters: Robust concepts and methods. Bulletin of the Int. Statist. Inst. 46, 375–382.

Hampel, F.R., Ronchetti, E.M, Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The approach based on influence functions.* New York: Wiley.

Huber, P.J. (1981). *Robust Statistics.* New York: Wiley.

Martin, R.D. and Zamar, R.H. (1993). Bias robust estimation of scale. *The Annals of Statistics* 21, 991-1017.

Rousseeuw, P.J. (1984). Least median of squares regression. J. Amer. Statist. Assoc. 79, 871–880.

Rousseeuw, P.J. and Croux, C. (1992). Explicit scale estimators with high breakdown point. In Y. Dodge (ed.) *L_1-Statistical Analysis and Related Methods.* Amsterdam: North Holland. 77-92.

Rousseuw, P.J. and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Americ. Statist. Assoc.* 88, 1273-1283.

Rousseeuw, P.J. and Croux, Ch. (1994). The bias of k-step M-estimators. *Statistics & Probability Letters* 20, 411-420.

Rousseeuw, P.J. and Leroy, A.M. (1988). A robust scale estimator based on the shortest half. *Statistica Neerlandica* 42, 103-116.

Siegel, A.F. (1982). Robust regression using repeated medians. *Biometrika* 68, 242–244.

Tukey, J.W. (1977). *Exploratory Data Analysis.* Reading, Mass.: Addison-Wesley (preliminary ed. 1971).

Wernecke, K.-D., Kalb, G. and Stuerzenbecher, E. (1988). On classification strategies in medical diagnostics (with special reference to mixed models). In H.H. Bock (ed.): *Classification and Related Methods of Data Analysis.* Amsterdam: North-Holland.