Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 122 (2004) 65–78

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Robust signal extraction for on-line monitoring data

P.L. Davies[a], R. Fried[b,*], U. Gather[b]

[a]Fachbereich Mathematik, Universität Essen, D-45141 Essen, Germany
[b]Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany

Accepted 15 June 2003

Abstract

Data from the automatic monitoring of intensive care patients exhibits trends, outliers, and level changes as well as periods of relative constancy. All this is overlaid with a high level of noise and there are dependencies between the different items measured. Current monitoring systems tend to deliver too many false warnings which reduces their acceptability by medical staff. The challenge is to develop a method which allows a fast and reliable denoising of the data and which can separate artefacts from clinical relevant structural changes in the patients condition (Estadistica 53 (2001) 259). A simple median filter works well as long as there is no substantial trend in the data but improvements may be possible by approximating the data by a local linear trend. As a first step in this programme, the paper examines the relative merits of the L_1 regression, the repeated median (Biometrika 68 (1982) 242) and the least median of squares (Bull. Internat. Statist. Inst. 46 (1975) 375; J. Amer. Statist. Assoc. 79 (1984) 871). The question of dependency between different items is a topic for future research.

MSC: primary 62G07; secondary 62G35

Keywords: Linear regression; Signal extraction; Level change; Trend; Outliers; Small-sample efficiency

1. Introduction

On-line monitoring of intensive care patients poses an interesting challenge for statisticians. Clinical information systems register and save at least once a minute the physiological variables and device parameters for each patient. Reliable automatic monitoring systems are needed to process this large quantity of data in real time and to support

* Corresponding author. Fax: +49-231-755-5305.
E-mail address: fried@statistik.uni-dortmund.de (R. Fried).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.06.012

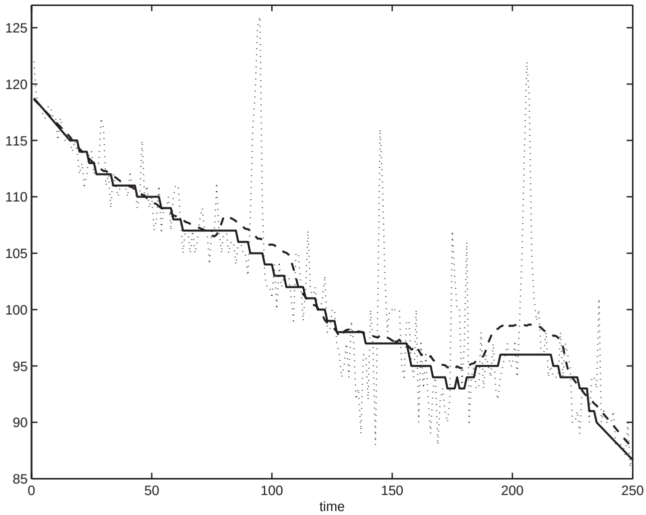

Fig. 1. Time series of the heart rate (dotted), as well as a running mean (dashed) and a running median (solid) with window width 31 both.

decision making at the bedside in critical situations where time is of utmost importance. Clinically relevant changes such as level shifts and trends need to be detected and to be distinguished from noise and clinically irrelevant outliers caused, for example, by the patient's movements or by measurement artefacts. If due care is not taken, the outliers may also adversely affect any subsequent data analysis. The reliable extraction of the underlying signal from a noisy time series with outliers is therefore an important basic step for further statistical off- or on-line analysis (Gather et al., 2001).

Fig. 1 shows a small excerpt from a series of measurements of the heart rate of a critically ill patient. An experienced physician analysed the data as being composed of a downward trend until time point 150 with noise and many clinically irrelevant outliers. Fig. 1 also shows the result of a running mean and of a running median (Tukey, 1977) with a time window of 31 observations. Both methods provide denoising, but the mean is clearly affected by the outliers although these are of moderate size only. In intensive care, the outliers in physiological time series may be much larger and affect the mean even more. The running median resists the outliers, even very large ones, much more successfully but it approximates the more or less linear trend by a step function.

The superiority of the median in resisting the clinically irrelevant outliers indicates the advantages of robust statistical functionals. It seems plausible that the difficulties of the median in adapting to local trends can be overcome by the use of robust regression functionals. As a first step in this programme we investigate the relative merits of the L_1 regression, the repeated median and the least median of squares which are defined below. Of particular interest are

- their ability to reproduce a linear trend in the presence of outliers,
- their ability to detect level changes,

- their ability to detect trend changes,
- the cost of computation.

Traditionally, the question of efficiency is also considered and we include some simulations for completeness. The important properties are however those listed above and these have little to do with efficiency (Davies and Gather, 1993). The situation we consider is a special one. The design points form a lattice and the sample size of about 20–30 observations is rather small but is necessitated by the requirement of being on-line. Clearly, a large time delay will allow a better statistical analysis but in very critical situations it may prove fatal for the patient.

2. Methods for robust linear regression

The robustification of even the simple linear regression model $y = \mu + \beta x + \varepsilon$ poses a considerable problem. One main weakness of all known high breakdown methods is their computational complexity. Huber (1995) has expressed this rather pointedly by saying that the high breakdown methods themselves break down because of their incomputability. The Hampel–Rousseeuw least median of squares (LMS) functional T_{LMS} (Hampel, 1975; Rousseeuw, 1984) is defined by

$$T_{\text{LMS}} = \operatorname{argmin}\{(\mu,\beta) : \operatorname{Median}(y_i - \mu - \beta x_i)^2\}. \tag{1}$$

As the design points lie on a lattice a breakdown can only be caused by outliers in the y variable. In this situation, the breakdown point of T_{LMS} in case of a sample of size n is $\lfloor n/2 \rfloor / n$. For the calculations, we use the algorithm by Stromberg (1993) which has a computational complexity of order n^4 for a sample of size n. This can be reduced if an approximate solution is calculated but as there are strict quality standards for on-line monitoring this is unlikely to be permitted. We are obliged to calculate the exact solution. As we are dealing with a single sample size of the order of 20 or 30 the complexity is no great problem. If, however, several hundred items have to be treated simultaneously then the computational complexity is a very important aspect. In principle, (1) may not have a unique solution but this has not been the case in all the data sets we have analysed so far.

Another high breakdown regression functional is Siegel's (1982) repeated median T_{RM} defined by

$$\tilde{\beta}_{\text{RM}} = \operatorname{med}_i \left(\operatorname{med}_{j \neq i} \frac{y_i - y_j}{x_i - x_j} \right),$$

$$\tilde{\mu}_{\text{RM}} = \operatorname{med}_i (y_i - \tilde{\beta}_{\text{RM}} x_i).$$

Its breakdown point is also $\lfloor n/2 \rfloor / n$ and the computational complexity of a straightforward implementation is of order n^2. It may therefore be preferred to T_{LMS} even if its small sample performance should turn out to be worse.

Finally, we also consider the L_1 regression T_{L1} defined by

$$T_{L1} = \operatorname{argmin}\left\{ (\mu,\beta) : \sum_{i=1}^{n} |y_i - \mu - \beta x_i| \right\}. \tag{2}$$

Table 1
Finite-sample replacement breakdown point q_m of T_{L1}

m	1	2	3	4	5	6	7	8	9	10	11	12
q_m	$\frac{1}{3}$	$\frac{2}{5}$	$\frac{2}{7}$	$\frac{3}{9}$	$\frac{4}{11}$	$\frac{4}{13}$	$\frac{5}{15}$	$\frac{5}{17}$	$\frac{6}{19}$	$\frac{7}{21}$	$\frac{7}{23}$	$\frac{7}{25}$
$\approx$	0.33	0.40	0.28	0.33	0.36	0.31	0.33	0.29	0.32	0.33	0.30	0.28

The L_1 regression functional is of the three methods the one most susceptible to outliers. We calculate T_{L1} using the descent technique as described in Sposito (1990). This is slightly faster than other methods for sample sizes $n \leqslant 30$ and considerably faster for larger sample sizes. The existence of multiple solutions is a noticeable problem for T_{L1} as we shall see below. The breakdown point of T_{L1} can be calculated from the results of He et al. (1990), Ellis and Morgenthaler (1992) and Mizera and Müller (1999). For design points on a lattice, $x_1 = -m, \ldots, x_n = m$, i.e. $n = 2m+1$, it reduces to

$$\min\left\{\frac{|I|}{2m+1} : \sum_{t\in I}|t| \geqslant \sum_{t\in I^c}|t|, I \subset \{-m,\ldots,m\}\right\}.$$

For large m, this is approximately $1 - 1/\sqrt{2} \approx 0.293$. Table 1 gives the exact values for small m. Clearly, T_{L1} is less robust than either T_{LMS} or T_{RM} but its speed of calculation may make it an interesting candidate and for this reason we include it in the comparison.

The T_{L1} belongs to the class of M-estimators. Huber-type M-estimators have robustness properties similar to the T_{L1} functional. This can be deduced from Mizera and Müller (1999). Redescending M-functionals are computationally difficult and require a good starting point (Rousseeuw and Leroy, 1987, p. 149). The three functionals investigated here do not need an initial solution and can if desired be used as robust starting values for a more efficient procedure as suggested by Rousseeuw and Leroy (1987, p. 129). However as we pointed out above efficiency is of little relevance when detecting outliers.

Observations from on-line monitoring are often correlated and it could be argued that this should be taken into account when using the regression functionals. The methods we use are conceptually simple and have proved to be more than adequate in their ability to detect outliers on-line in real data. Taking possible correlations into account leads to a considerable increase in complexity with the resulting increase in computing time as well as in a loss of robustness (Lucas, 1997; Meintanis and Donatos, 1999). Simplicity is of paramount importance in on-line monitoring.

3. Comparison of T_{L1}, T_{RM} and T_{LMS}

3.1. The basic simulation model

For a comparison of the finite-sample properties of the distinct regression methods 10 000 samples were simulated using the model

$$Y_t = \mu + \beta t + \varepsilon_t, \quad t = -m, \ldots, m$$

Table 2
Efficiencies relative to L_2 regression (in percent) measured by the simulated MSE for T_{L1} $(\tilde{\mu}_{L1}, \tilde{\beta}_{L1})$, T_{RM} $(\tilde{\mu}_{\mathrm{RM}}, \tilde{\beta}_{\mathrm{RM}})$ and T_{LMS} $(\tilde{\mu}_{\mathrm{LM}}, \tilde{\beta}_{\mathrm{LM}})$ for N(0, 1) errors

m	β	$\tilde{\mu}_{L1}$	$\tilde{\mu}_{\mathrm{RM}}$	$\tilde{\mu}_{\mathrm{LM}}$	$\tilde{\beta}_{L1}$	$\tilde{\beta}_{\mathrm{RM}}$	$\tilde{\beta}_{\mathrm{LM}}$
5	0.0	69.7	66.3	26.9	78.8	69.8	25.1
5	0.1	71.1	66.4	27.4	71.9	70.4	26.1
5	0.2	69.3	64.3	26.3	63.6	68.8	24.7
10	0.0	66.9	63.9	22.4	70.4	70.8	22.7
10	0.1	67.8	64.4	22.9	64.5	71.7	23.4
10	0.2	69.4	66.6	23.1	66.1	73.1	24.2
15	0.0	66.3	64.3	20.7	70.2	71.4	21.6
15	0.1	68.1	65.0	20.7	64.5	72.7	21.0
15	0.2	68.0	65.4	20.4	66.1	73.2	22.0

with $\mu = 0$ and for several different slopes β. The error ε was always taken to be Gaussian white noise with mean zero and unit variance. The estimated values of μ and β are used to provide a value of the signal at time $t = 0$. This represents a time delay of m. The value of m is determined by requiring on the one hand a certain stability (m large) and on the other hand the demands made by the on-line nature of the application (m small). In this paper, we restrict attention to the cases $m = 5, 10, 15$ which correspond to sample sizes $n = 2m + 1 = 11, 21, 31$.

3.2. Efficiency

As a first step, we give the relative efficiencies of the three functionals T_{L1}, T_{RM} and T_{LMS} with respect to the least-squares functional T_{L2} although as mentioned above efficiency is not an important consideration. The results are given in Table 2 for the slopes $\beta = 0$, 0.1 and 0.2. The results are as was to be expected apart perhaps from the slope component of T_{L1} where the relative efficiency is highest for $\beta = 0$. This may well be due to the non-uniqueness of the L_1 solution and the result of taking $\beta = 0$ as a starting point for the calculation of the solution. A similar phenomenon was noted by Terbeck (1996) in the case of the two-way table.

3.3. Outliers in the steady state

Data in intensive care medicine contain large isolated outliers as well as patches of outliers. We concentrate on outliers which have the same sign. Bias considerations imply that this is the most difficult case and this was confirmed by simulations. As positive outliers are more common in intensive care data we generated the outliers as follows where for the sake of brevity we restrict attention to a sample size $n = 21$. We replace an increasing number $1, 2, \ldots, 10$ of observations by additive outliers of sizes $\omega \in \{2, \ldots, 10\}$. This is done by generating the data as described in Section 3.1 and then adding ω to the y-value at points in the window which are chosen at random. For comparison, we also include the case of no outliers in the figures. The

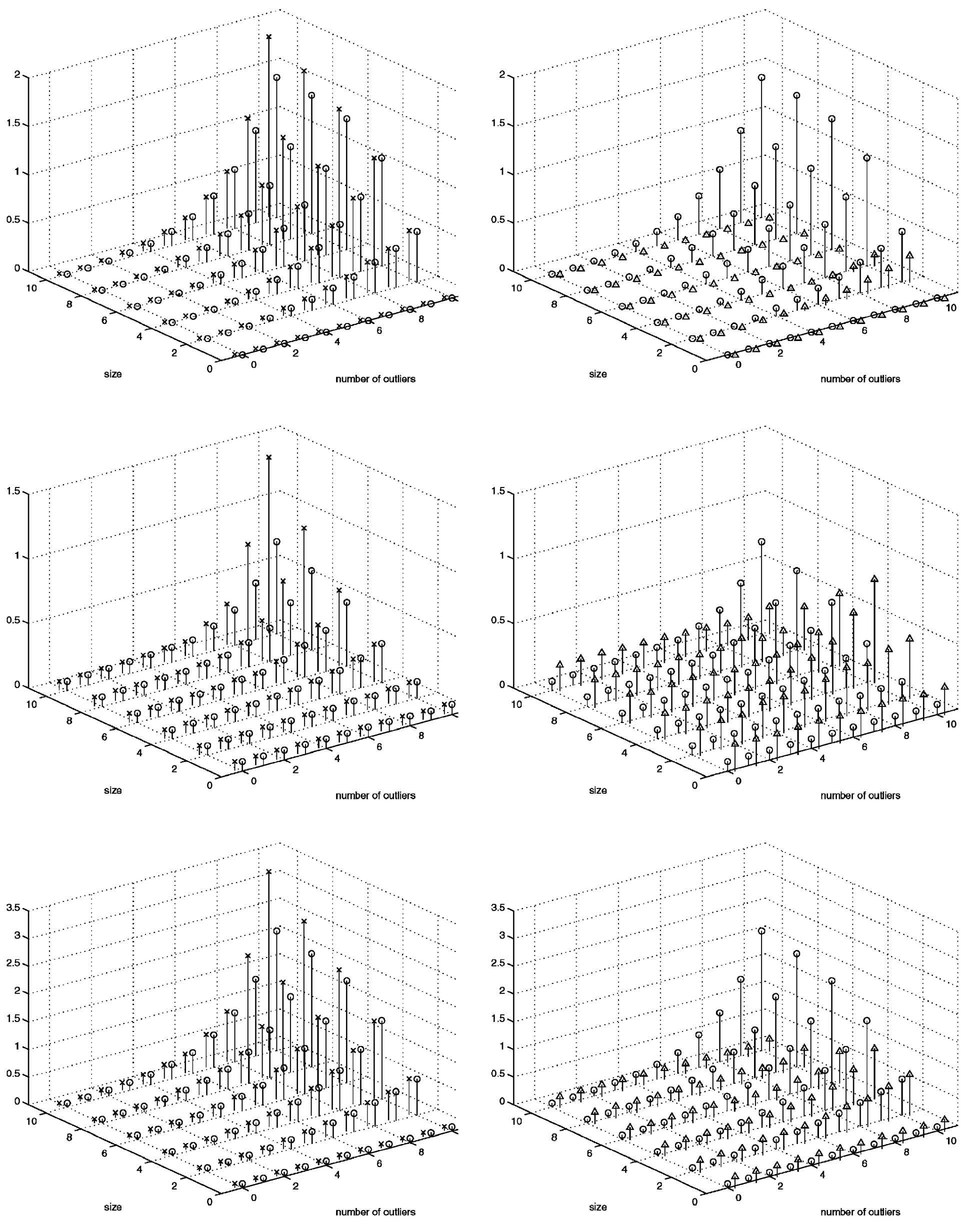

Fig. 2. Steady state: Simulated squared bias (top), variance (middle) and MSE (bottom) for the intercept. L_1 regression $\times$, repeated median $\circ$ and LMS $\triangle$.

simulations were performed with $\mu = \beta = 0$. Each of the 66 cases is simulated 2000 times and the squared bias, variance and mean square error (MSE) were calculated. The results are shown graphically in Fig. 2 for the intercept and in Fig. 3 for the

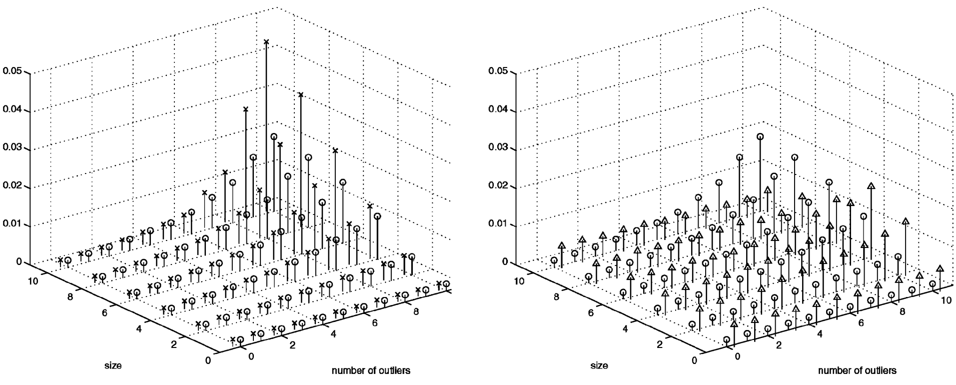

Fig. 3. Steady state: Simulated MSE for the slope. L_1 regression $\times$, repeated median $\circ$ and LMS $\triangle$.

slope. For the latter, only the MSE is shown as outliers occurring at positions chosen at random do not cause a bias for the slope. For 0–6 outliers or for outliers of size 0–4 there is little to choose between the methods. T_{L1} shows considerable bias in the intercept for seven or more outliers. This corresponds well with Table 1. T_{RM} performs similarly to T_{L1} for the intercept although it has the same breakdown point as T_{LMS}. Both T_{RM} and T_{L1} are dominated by T_{LMS} in the intercept for eight or more outliers of any size. With respect to the slope, T_{RM} has the smallest MSE amongst the three functionals in case of many small outliers whilst T_{LMS} is better for many large outliers.

3.4. Level shift and outliers

A situation that is particularly important in on-line monitoring is the occurrence of a level shift. In order to detect a level shift we need a reliable approximation for the level at time $t = 0$ when the most recent observations are at another level. Some definition of a level shift is required to distinguish it from a block of outliers. The definition we take is that the five most recent observations are of about the same size and sign and differ substantially from the preceding observations (cf. Imhoff et al., 1998; Gather et al., 2000). We report the results in detail for a level shift of size 10 although simulations were also carried out for level shifts of sizes 3 and 5. As outliers are again a problem, we generated positive ones at time points chosen at random. Again the squared bias, variance and MSE were calculated for each of the three regression functionals. The results for the level shift of 10 are shown in Figs. 4 (intercept) and 5 (slope). Five outliers occurring at the end of the time window cause T_{L1} to be biased for the intercept and the slope whilst T_{RM} is biased for the slope only. The superiority of T_{LMS} in this situation is apparent. It shows much less bias than the other functionals and can even accommodate up to seven outliers. The slope component of T_{LMS} shows considerable variability if a level shift combined with eight or more outliers occurs. When estimating the slope a moderate number of positive outliers and a negative shift can balance each other when T_{L1} and T_{RM} are

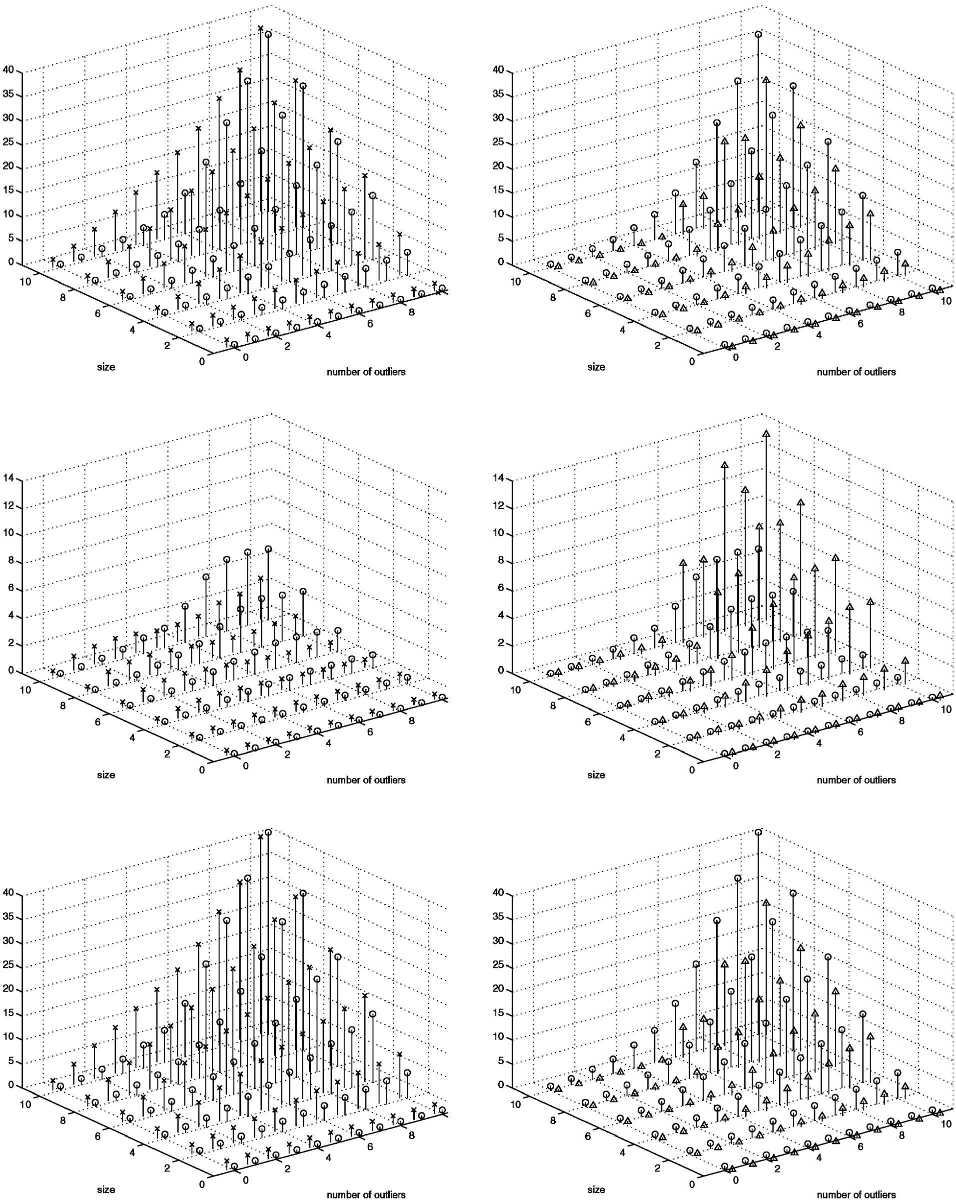

Fig. 4. Level shift of size 10: Simulated squared bias (top), variance (middle) and MSE (bottom) for the intercept. L_1 regression $\times$, repeated median $\circ$ and LMS $\triangle$.

used. This effect does not occur for T_{LMS}. In 3.3 as well as here, we also simulated a situation with positive and negative outliers. All methods showed a much smaller bias and MSE. While the MSE of T_{LMS} is the largest one in a steady state for up to five

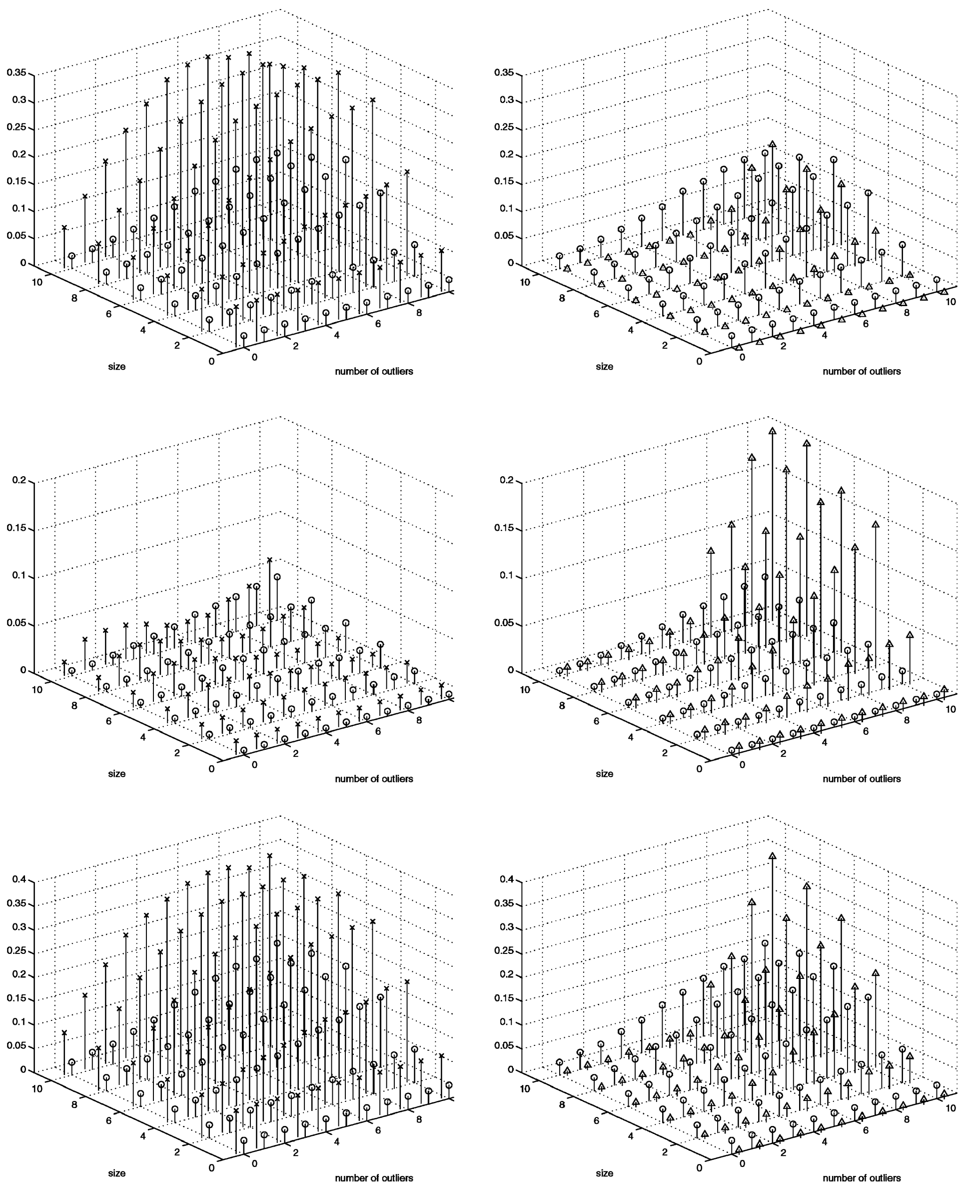

Fig. 5. Level shift of size 10: Simulated squared bias (top), variance (middle) and MSE (bottom) for the slope. L_1 regression $\times$, repeated median $\circ$ and LMS $\triangle$.

outliers, in case of a level shift we get almost the same results as for only positive outliers.

The results are of course less clear cut for the smaller level shifts of sizes 3 and 5 but again T_{LMS} is superior.

Table 3
Mean time of applying the functionals to 1000 samples of different sizes (in s) for the steady state $\beta=\mu=0$

n	T_{L2}	T_{L1}	T_{RM}	T_{LMS}
21	0.2	2.4	2.6	28.4
31	0.3	4.7	4.4	120.7

3.5. Computation times

As already noted above, the computation time needed is extremely important in on-line data when many variables are to be monitored and analysed simultaneously. Table 3 shows the mean times of applying the functionals to 1000 samples of sizes 21 and 31 using a self-written FORTRAN program on a Sun workstation ultra spark with 170 MHz and 320 MB Ram. We remark that for T_{L1} the time depends on the data as the number of iterations needed may vary. In case of a steep trend and an additional level shift the computation time increased to 3.5 ($n=21$) and 6.2 ($n=31$) s. The time needed for computation of T_{LMS} is much larger than that for the other functionals and increases rapidly with the sample size.

The results show that the computational complexity of T_{LMS} is such as to preclude its use in on-line monitoring except possibly in situations where the sampling rate and the number of variables to be analysed are both small. If in future fast algorithms for T_{LMS} become available then the situation would alter and preference may then be given to T_{LMS} because of its superior performance in the presence of outliers. In terms of computational complexity, there is at the moment little to choose between T_{L1} and T_{RM}.

3.6. Simulated time series

We now consider a simulated time series of length 250 with two time periods representing deterministic trends and two level shifts of size 5. It is shown in the upper panel of Fig. 6. The signal is overlaid with white noise generated from the standard normal distribution and 25 outliers of size 5 have been inserted. This corresponds to 10% of the observations being outliers which is reasonable for on-line monitoring. The outliers consist of seven single outliers, four patches of two, two patches of three and one patch of four. The time points of the outlier patches were selected randomly from a uniform distribution on all time points with the exception of the two outliers at times $t=195$ and 196 which were inserted to make the detection of the level shift at time $t=201$ more difficult. The functionals T_{L1}, T_{RM} and T_{LMS} were applied to this data set using a window width of $n=31$. We consider edge effects by extrapolating the trend estimated in the first and last time window, respectively.

In general, T_{LMS} shows more variability, but it is affected only by the outlier block at $t=112$. The most important difference between the methods in the clinical context

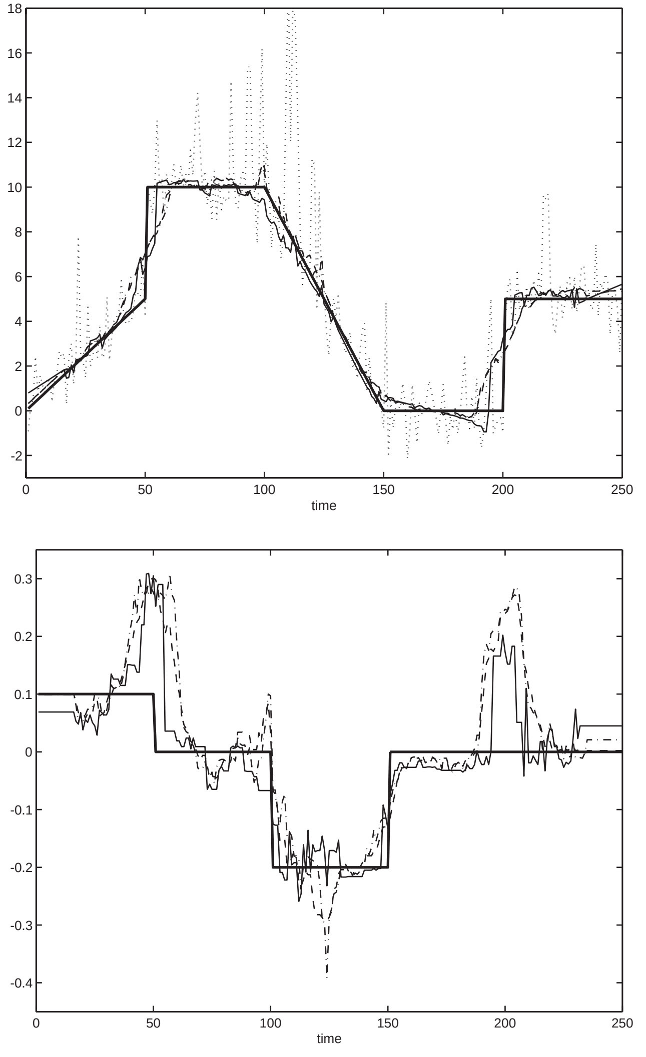

Fig. 6. Simulated time series. Top: Time series (dotted), underlying level (fat solid) and level estimates: T_{L1} (dashed–dotted), T_{RM} (dashed), T_{LMS} (solid). Bottom: Slope estimates (same styles). With respect to the level, T_{L1} and T_{RM} are almost identical.

can be seen at the time points of the level shifts. Both T_{L1} and T_{RM} are affected by the level shifts much earlier than T_{LMS} which is first affected only at times $t = 46$ and 194, respectively. At these time points there are, respectively, 10 and eight observations in the current time window which are affected by the level shift. With respect to the slope the differences between the methods are not very pronounced, but T_{LMS} preserves the slope changes better than the other functionals which tend to smooth the changes.

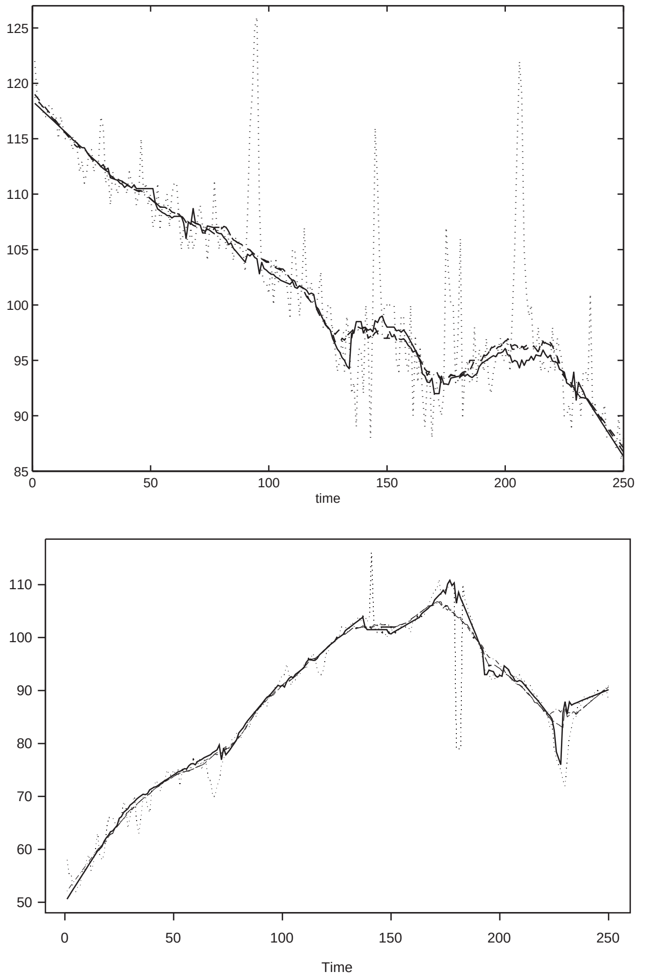

Fig. 7. Time series (dotted) representing heart rate (top) and arterial blood pressure (bottom) as well as some level approximates: T_{L1} (dashed–dotted), T_{RM} (dashed), T_{LMS} (solid). T_{L1} and T_{RM} are almost identical.

3.7. Real examples

Finally, we consider three real examples from the monitoring of intensive care patients. As such data often contain clinically irrelevant minor trends we use a time window of length $n = 31$ (see Fig. 7). The first example is the one shown in the introduction. All regression functionals perform well, however T_{L1} and T_{RM} are smoother than T_{LMS}.

The second time series represents the arterial blood pressure of another patient. Again there are outliers but only one section from $t = 225$ to 231 was judged to be clinically

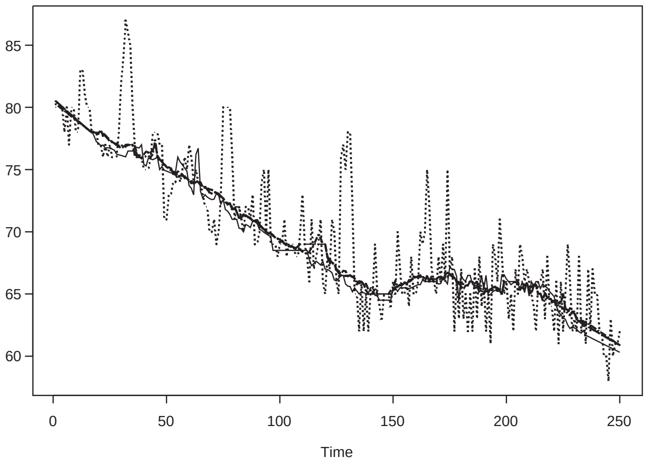

Fig. 8. Time series (dotted) representing heart rate as well as some level approximates: T_{L1} (dashed–dotted), T_{RM} (dashed), T_{LMS} (solid). T_{L1} and T_{RM} are almost identical.

relevant. T_{L1} and T_{RM} are both affected by the clinically irrelevant outliers at about $t = 166$ but miss the clinically relevant outliers at $t = 231$. T_{LMS} performs very well on this data set.

The final example depicted in Fig. 8 is the heart rate of a third patient, which was judged by a physician to show a downward trend with irrelevant outliers. Again, T_{L1} and T_{RM} are much smoother than T_{LMS} which exhibits a large spike at $t = 63$ due to a particular pattern in the data. T_{L1} and T_{RM} perform well but overestimate the signal between $t = 110$ and 140.

4. Discussion

Alarm systems in intensive care must be capable of on-line detection of clinically relevant patterns such as trends and level changes. The first step in the development of such systems is the on-line extraction of the signal which is corrupted by noise and outliers. In this paper, we have compared three robust methods of signal extraction namely T_{L1}, T_{RM} and T_{LMS}. The comparison used simulated and real data as they occur in the monitoring of intensive care patients. On the basis of the limited evidence presented here, our tentative conclusions are that T_{LMS} shows the smallest bias, but it is very variable and computationally expensive. The variability can possibly be reduced in a second smoothing step. T_{L1} and T_{RM} can withstand a large number of outliers, are computationally much less expensive and do not show such instabilities. Because of its higher breakdown point the present paper points to T_{RM} as being a prominent candidate for a first sweep over the data when computation time is critical.

Acknowledgements

We thank two anonymous referees for several suggestions which were helpful to improve the presentation of the paper. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Davies, P.L., Gather, U., 1993. The identification of multiple outliers. J. Amer. Statist. Assoc. 88, 782–793.

Ellis, S.P., Morgenthaler, S., 1992. Leverage and breakdown in L_1 regression. J. Amer. Statist. Assoc. 87, 143–148.

Gather, U., Fried R., Imhoff, M., 2000. Online classification of states in intensive care. In: Gaul, W., Opitz O., Schader, M. (Eds.), Data Analysis. Scientific Modeling and Practical Application. Springer, Berlin, pp. 413–428.

Gather, U., Fried, R., Lanius, V., Imhoff, M., 2001. Online-monitoring of high-dimensional physiological time series—a case-study. Estadistica 53, 259–298.

Hampel, F.R., 1975. Beyond location parameters: robust concepts and methods. Bull. Internat. Statist. Inst. 46, 375–382.

He, X., Jureckova, J., Koenker, R., Portnoy, S., 1990. Tail behavior of regression estimators and their breakdown points. Econometrica 58, 1195–1214.

Huber, P.J., 1995. Robustness: where are we now?. Student 1, 75–86.

Imhoff, M., Bauer, M., Gather, U., Löhlein, D., 1998. Statistical pattern detection in univariate time series of intensive care on-line monitoring data. Intensive Care Med. 24, 1305–1314.

Lucas, A., 1997. Asymptotic robustness of least median of squares for autoregressions with additive outliers. Commun. Statist. Theory Meth. 26, 2363–2380.

Meintanis, S.G., Donatos, G.S., 1999. Finite-sample performance of alternative estimators for autoregressive models in the presence of outliers. Comput. Statist. Data Anal. 31, 323–339.

Mizera, I., Müller, Ch., 1999. Breakdown points and variation exponents of robust M-estimators in linear models. Ann. Statist. 27, 1164–1177.

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

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

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

Sposito, V.A., 1990. Some properties of L_p-estimators. In: Lawrence, K.D., Arthur, J.L. (Eds.), Robust Regression. Analysis and Applications. Marcel Dekker, New York, pp. 23–58.

Stromberg, A.J., 1993. Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression. SIAM J. Sci. Comput. 14, 1289–1299.

Terbeck, W., 1996. Interaktionen in der Zwei-Faktoren-Varianzanalyse. Doctoral Thesis, Universität Essen, Essen, Germany.

Tukey, J.W., 1977. Exploratory Data Analysis [preliminary Ed. 1971]. Addison-Wesley, Reading, MA.