Robust Online Scale Estimation in Time Series: A Regression-Free Approach

Sarah Gelper[1,*], Karen Schettlinger[2],
Christophe Croux[1], and Ursula Gather[2]

May 3, 2007

[1] *Faculty of Economics and Management, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium.*

[2] *Department of Statistics, University of Dortmund, 44221 Dortmund, Germany.*

Abstract: This paper presents variance extraction procedures for univariate time series. The volatility of a times series is monitored allowing for non-linearities, jumps and outliers in the level. The volatility is measured using the height of triangles formed by consecutive observations of the time series. This idea was proposed by Rousseeuw and Hubert (1996, Regression-free and robust estimation of scale for bivariate data, *Computational Statistics and Data Analysis*, 21, 67–85) in the bivariate setting. This paper extends their procedure to apply for online scale estimation in time series analysis. The statistical properties of the new methods are derived and finite sample properties are given. A financial and a medical application illustrate the use of the procedures.

Keywords: Breakdown Point, Influence function, Online monitoring, Outliers, Robust scale estimation.

*Corresponding author. E-mail: sarah.gelper@econ.kuleuven.be, Tel: 0032/16326928, Fax: 0032/1632.67.32

1 Introduction

In this paper we propose a method to monitor variability in univariate time series. The procedure allows one to get insight in the evolution of the variability of the series over time. Moreover, it (i) can cope with highly non-linear signals, (ii) is suitable for online applications and (iii) is robust with respect to outliers and level shifts. This is achieved by making use of the vertical height of triangles formed by consecutive data points. The method is explorative; it does not require an explicit modeling of the time series. This technique is of interest in various applied fields. In finance for instance, variability of returns is associated with risk and thus directly relates to portfolio management and option pricing. In intensive care, measurement of variables like heart rate and blood pressure need to be constantly monitored since changes in these variables and their variability contain crucial information on the well-being of the patient.

For both the financial and the intensive care applications, the data are recorded with high frequency, e.g. every minute or every second. For these applications, it is important to monitor the variability instantaneously. For this reason, the proposed methods are designed to work online: for every new incoming observation, the variability is easily determined by a fast updating step. The scale estimate at the present time point is obtained by using a finite number of the most recent observations, making it a local approach.

High frequency measurements typically lead to 'unclean' and noisy series containing irrelevant outliers. Hence, we focus on robust techniques. For every method, the robustness with respect to outliers is studied in detail by computing breakdown points and influence functions. Statistical efficiencies are also derived. These are accompanied by simulation results which provide insight into the finite sample properties of the different methods (Appendix B).

The scale estimates discussed in this paper are regression free, i.e. directly based on the observed data points without applying a regression fit first. The advantage is that we do not have to bother about estimating the main signal in the series

before estimating the variability. Regression free scale estimation methods have already been studied by Rousseeuw and Hubert (1996) in the general bivariate setting. Here, we are especially interested in time series scale estimation, and adapt the estimators proposed by Rousseeuw and Hubert (1996) to be applicable to time series with non-linear trends, trend changes and jumps. In this more special setting of univariate times series, we are able to derive theoretical properties of these estimators as well.

The different candidate methods are described in Section 2. Their robustness properties are studied in Section 3 and their statistical efficiencies in Section 4. Data applications can be found in Section 5. Finally, Section 6 briefly summarizes the results and gives concluding remarks.

2 Description of the methods

We define a simple time series model, where the time series $(y_t)_{t\in\mathbb{Z}}$ is decomposed into a level component μ_t and a random noise component e_t

$$y_t \;=\; \mu_t + e_t\,. \tag{2.1}$$

The noise component e_t is assumed to have zero mean and time varying scale σ_t. The focus in this study lies on estimating and monitoring σ_t, which reflects the variability of the process around its underlying level μ_t. The level or *signal* μ_t can vary smoothly over time but can also contain sudden jumps or trend changes. The variability of the process y_t is then captured by the scale of the e_t, where the latter may contain outliers.

We make use of a moving window approach for the estimation of σ_t. To obtain a scale estimate of the time series at time point t, denoted by S_t, we only use information contained in the time window formed by the n time points $t-n+1$ to t. As the window moves along the series, we obtain a scale estimate S_t for every time point $t = n, \ldots, T$. As such, a *running scale* approach is obtained, suitable

for online application. An example would be a running standard deviation, which would of course not be robust with respect to outliers nor be suitable for time series containing a trend.

One possibility for online estimation of σ_t is to apply a scale estimate to the residuals of a robust regression fit within a time window, as studied in Fried and Gather (2003). This procedure is based on the fact that the local level μ_t can be estimated well by robust regression filters (see e.g. Davies et al. (2004) and Gather et al. (2006)). In that case it is assumed that, within a time window of length n, the underlying signal μ_t of the series y_t can be reasonably well approximated by a linear trend. The approach presented in this paper allows for stronger non-linearities in the time series.

2.1 Estimation Methods

The methods under consideration are regression free, i.e. a scale estimate for e_t in (2.1) is obtained without fitting a regression line within the time window. Following the approach of Rousseeuw and Hubert (1996), the scale estimates are constructed using the vertical heights of triangles formed by triples of successive data points. These heights correspond to the non-zero residual of an L_1 fit to these three data points. Here it is assumed that only within each triple of consecutive observations, the series can well be approximated by a linear trend.

Consider any three successive observations y_i, y_{i+1} and y_{i+2}. Assuming the series to be observed at equidistant time points, the height of the triangle formed by these observations is given by the simple formula

$$h_i = \left| y_{i+1} - \frac{y_i + y_{i+2}}{2} \right|. \tag{2.2}$$

The more variation there is in the time series, the larger the h_i will be. Within a window of length n, the heights of the $n-2$ *adjacent* triangles are used in the construction of the scale estimators studied here. Note that the heights h_i in (2.2) are invariant with respect to adding a linear trend to the time series, having the

beneficial consequence that linear trend changes do not affect the scale monitoring procedure.

Suppose we want to estimate the variability at time t using the observations in the time window $t-n+1,\ldots,t$ of length n. For ease of notation, we drop the dependency on t and denote these observations by y_1 to y_n, and the associated heights as defined in (2.2) by h_i, for $i = 1,\ldots,n-2$. The first estimator we consider is proposed in Rousseeuw and Hubert (1996) and is defined via the α-quantile of the heights obtained from adjacent triangles, with $0 < \alpha < 1$. Let $h_{(i)}$ be the i-th value of the ordered sequence of all heights in the current window. The scale estimate Q^{α}_{adj} is then given by

$$Q^{\alpha}_{adj}(y_1,\ldots,y_n) \;=\; c_q \cdot h_{(\lfloor \alpha(n-2) \rfloor)}\,, \tag{2.3}$$

which is the $\lfloor \alpha(n-2) \rfloor$-th value in the sequence of ordered heights, with c_q a constant to achieve Fisher consistency at a specified error distribution, referred to as the *consistency factor*. The value of α regulates the trade off between robustness and efficiency, as will be discussed in detail in sections 3 and 4.

Considering observations sampled from a continuous distribution F, the corresponding triangle heights will also have a continuous distribution, denoted by H_F. In that case the functional form of the estimator (2.3) corresponds to

$$Q^{\alpha}_{adj}(F) \;=\; c_q \cdot H_F^{-1}(\alpha). \tag{2.4}$$

Assuming normality of the noise component, it is not difficult to show that one needs to select

$$c_q = (Q^{\alpha}_N)^{-1} \text{ with } Q^{\alpha}_N := \sqrt{\frac{3}{2}}\,\Phi^{-1}\left(\frac{\alpha+1}{2}\right). \tag{2.5}$$

Here $\Phi(z)$ is the standard normal cumulative distribution function at z, and the index N refers to the assumption of normality. For example, for $\alpha = 0.5$ we have $c_q = 2.51$.

We present two alternatives to the Q^{α}_{adj} estimator making use of averages instead of the quantile. The first alternative is constructed as the Trimmed Mean

(TM) of the adjacent triangle heights and is defined by

$$TM^{\alpha}_{adj}(y_1,\ldots,y_n) \;=\; c_m \cdot \frac{1}{\lfloor \alpha\,(n-2)\rfloor} \sum_{i=1}^{\lfloor \alpha\,(n-2)\rfloor} h_{(i)}\,. \tag{2.6}$$

The second alternative is the square root of a Trimmed Mean of Squares (TMS) of the adjacent triangle heights:

$$TMS^{\alpha}_{adj}(y_1,\ldots,y_n) \;=\; c_s \cdot \sqrt{\frac{1}{\lfloor \alpha\,(n-2)\rfloor} \sum_{i=1}^{\lfloor \alpha\,(n-2)\rfloor} h^2_{(i)}}\,. \tag{2.7}$$

The trimming proportion equals $(1-\alpha)$ where α can vary between zero and one. As for the Q^{α}_{adj} estimator, it regulates the trade off between efficiency (high α) and robustness (low α). The functional form of these estimators is given by

$$TM^{\alpha}_{adj}(F) \;=\; c_m \cdot TM^{\alpha}_{1}(H_F) \tag{2.8}$$

and

$$TMS^{\alpha}_{adj}(F) \;=\; c_s \cdot TM^{\alpha}_{2}(H_F)\,. \tag{2.9}$$

Here, we use a trimmed moment functional TM^{α}_{p} which is defined as the α-trimmed pth central moment to the power of $1/p$,

$$TM^{\alpha}_{p} : G \mapsto TM^{\alpha}_{p}(G) = E(X^p |\, X \leq Q^{\alpha}(G))^{1/p}\,, \tag{2.10}$$

with $X \sim G$. The consistency factors c_m and c_s can be derived for Gaussian noise:

$$c_m \;=\; \frac{\alpha}{\sqrt{6}\,\left[\varphi(0) - \varphi(\sqrt{2/3}\,Q^{\alpha}_N)\right]}\,, \tag{2.11}$$

$$c_s \;=\; \frac{\sqrt{\alpha/3}}{\sqrt{\alpha/2 - \sqrt{2/3}\,Q^{\alpha}_N\,\varphi(\sqrt{2/3}\,Q^{\alpha}_N)}}\,, \tag{2.12}$$

with Q^{α}_N defined in (2.5), and $\varphi(z)$ the associated density of $\Phi(z)$. Details on how these expressions are obtained can be found in Appendix A. For example, for $\alpha = 0.5$, one has $c_m = 2.51$ and $c_s = 2.16$. The consistency factors c_q, c_m and c_s have been derived at the population level. However, extensive simulations (reported in Appendix B) have shown that they yield very good approximations, already

for samples of size $n = 20$. We stress that the finite sample case is not without importance in this setting since the scale estimates are computed within windows of limited size. To achieve unbiasedness at finite samples for a Gaussian distribution of, for example, the Q^{α}_{adj} estimators, one could replace c_q by its finite sample counterpart c^n_q (obtainable by means of Monte-Carlo simulations). In Appendix B, a simple approximative formula for this finite sample factor c^n_q, with $\alpha = 0.5$, is derived:

$$c^n_q \approx 1.21 \frac{n}{n + 0.44}. \tag{2.13}$$

We only consider scale estimators based on heights of adjacent triangles. Alternatively, one could use the heights of triangles formed by all triples of data points within the window, or any other subset of them. Several such possibilities are described in Rousseeuw and Hubert (1996). However, for online monitoring of high frequency time series, the use of adjacent triangles is natural and appealing. The adjacent based methods are fast to compute and the update of the estimate for a new incoming observation is quick. The fastest algorithm to insert a new observation in an ordered series takes only $O(\log n)$ time and hence, so does the update of the adjacent based estimators. Moreover, using all possible triangles in one window requires the local linearity assumption to hold in the entire window and not only for triples of consecutive observations. As such, methods based on adjacent triangles are more suitable when the underlying signal has strong nonlinearities.

3 Robustness Properties

To evaluate the robustness of the estimators with respect to outlying observations, we look at their breakdown points and influence functions.

3.1 Breakdown Points

Loosely speaking, the breakdown point of a scale estimator is the minimal amount of contamination such that the estimated scale becomes either infinite (explosion)

or zero (implosion).

Let $\mathbf{y}_n = \{y_1, \ldots, y_n\}$ be a sample of size n with empirical distribution function F_n. Let S denote one of the investigated scale functionals (2.4), (2.8) or (2.9) taking values in the parameter space $(0, \infty)$ which we consider equipped with a metric D satisfying $\sup_{s_1,s_2\in(0,\infty)} D(s_1, s_2) = \infty$. For evaluating the breakdown point of scale functionals, the metric $D(s_1, s_2) = |\log(s_1/s_2)|$ seems a suitable choice as it yields ∞ in both cases, explosion and implosion.

Further, let $\mathbf{y}_n^k$ be a sample obtained from $\mathbf{y}_n$ but with a proportion of k/n observations altered to arbitrary values ($k \in \{1, \ldots, n\}$), and let F_n^k denote the the empirical distribution of $\mathbf{y}_n^k$. We define the *finite sample breakdown point* (fsbp) of S at the sample $\mathbf{y}_n$, or at F_n, by

$$\text{fsbp}(S, F_n, D) = \min \frac{1}{n} \left\{ k \in \{1,\, 2,\, \ldots,\, n\} :\, \sup_{F_n^k} D\left(S(F_n), S(F_n^k)\right) = \infty \right\},$$

which is equal to

$$\text{fsbp}(S, F_n) = \min \left\{ \text{fsbp}^+(S, F_n), \text{fsbp}^-(S, F_n) \right\}, \tag{3.1}$$

where

$$\text{fsbp}^+(S, F_n) = \min \frac{1}{n} \left\{ k \in \{1,\, 2,\, \ldots,\, n\} :\, \sup_{F_n^k} S(F_n^k) = \infty \right\} \tag{3.2}$$

is the explosion breakdown point, and

$$\text{fsbp}^-(S, F_n) = \min \frac{1}{n} \left\{ k \in \{1,\, 2,\, \ldots,\, n\} :\, \inf_{F_n^k} S(F_n^k) = 0 \right\} \tag{3.3}$$

the implosion breakdown point.

It is possible to give an upper bound for the finite sample breakdown point for affine equivariant scale estimates S (Davies and Gather (2005)):

$$\text{fsbp}(S, F_n) \le \left\lfloor \frac{n - n\Delta(F_n) + 1}{2} \right\rfloor / n\,, \tag{3.4}$$

where $n\Delta(F_n)$ is the maximal number of observations which might be replaced within the sample, such that the scale estimate remains positive. For scale estimates based on adjacent triangle heights $n\Delta(F_n)$ is equal to $\lfloor\alpha(n-2)\rfloor - 1$. Note that the bound (3.4) is not obtained for the scale estimates S considered here.

Rousseeuw and Hubert (1996) calculated the finite sample breakdown point of the Q^{α}_{adj} estimator in a regression setup with random design; but we consider a fixed design with equidistant time points, yielding higher values for the finite sample breakdown point: Suppose that $\mathbf{y}_n$ is in general position and define $B := \lfloor\alpha(n-2)\rfloor$. If the replacement sample $\mathbf{y}^k_n$ is chosen with $k = B-1$ such that $B+1$ observations are collinear, then this results in $B-1$ zero triangle heights and $n-B-1$ heights larger than zero. Hence, the Bth largest value of the ordered heights will be positive which implies $\text{fsbp}^-(S, F_n) \geq B/n$. On the other hand, replacing B observations such that $B+2$ observations are collinear implies that at least B heights will be zero and therefore $\text{fsbp}^-(S, F_n) \leq B/n$. We thus obtain

$$\text{fsbp}^-(S, F_n) = \lfloor\alpha(n-2)\rfloor/n\,.$$

For the explosion breakdown point, we follow the proof of Theorem 3 in Rousseeuw and Hubert (1996) and obtain

$$\text{fsbp}^+(S, F_n) = \left\lceil \frac{n-1-\lfloor\alpha(n-2)\rfloor}{3} \right\rceil /n\,.$$

Hence, the finite sample breakdown point corresponds to

$$\text{fsbp}(S, F_n) = \frac{1}{n}\min\left\{ \left\lceil \frac{n-1-\lfloor\alpha(n-2)\rfloor}{3} \right\rceil, \lfloor\alpha(n-2)\rfloor \right\}. \tag{3.5}$$

The maximum value for $\text{fsbp}(S, F_n)$ depends not only on the choice of α but also on whether n is divisible by four or not (see Table 1). A proof can be found in Appendix A.

Table 1 shows that, depending on n, more than one quantile might be chosen to achieve an estimate with maximum fsbp, with the order of the empirical quantile being $\lfloor\alpha(n-2)\rfloor \in \left\{ \left\lfloor \frac{n+1}{4} \right\rfloor, \ldots, n+1-3\left\lfloor \frac{n+1}{4} \right\rfloor \right\}$. This is due to the fact that both implosion and explosion of the estimator are regarded as breakdown.

	max. value of fsbp(S, F_n)	reached for $\alpha \in$	corresponding $\lfloor \alpha(n-2) \rfloor \in$
$n \in \{4k-1, k \in \mathbb{N}\}$:	$\frac{n+1}{4n}$	$\left[\frac{n+1}{4(n-2)}, \frac{n+5}{4(n-2)}\right)$	$\left\{\frac{n+1}{4}\right\}$
$n \in \{4k, k \in \mathbb{N}\}$:	$\frac{1}{4}$	$\left[\frac{n}{4(n-2)}, \frac{n+8}{4(n-2)}\right)$	$\left\{\frac{n}{4}, \frac{n+4}{4}\right\}$
$n \in \{4k+1, k \in \mathbb{N}\}$:	$\frac{n-1}{4n}$	$\left[\frac{n-1}{4(n-2)}, \frac{n+11}{4(n-2)}\right)$	$\left\{\frac{n-1}{4}, \frac{n+3}{4}, \frac{n+7}{4}\right\}$
$n \in \{4k+2, k \in \mathbb{N}\}$:	$\frac{n-2}{4n}$	$\left[\frac{n-2}{4(n-2)}, \frac{n+14}{4(n-2)}\right)$	$\left\{\frac{n-2}{4}, \frac{n+2}{4}, \frac{n+6}{4}, \frac{n+10}{4}\right\}$

Table 1: Maximum values for the finite sample breakdown point fsbp(S, F_n) with corresponding values of α and the rank $\lfloor \alpha(n-2) \rfloor$ of the triangle heights with S representing one of the scale estimates Q^{α}_{adj}, TM^{α}_{adj} or TMS^{α}_{adj}.

If collinear observations rather than outliers are expected in the sample, the best choice is to set α to the maximal value within the range given in Table 1, i.e. $\alpha = \frac{n+1-3\left\lfloor \frac{n+1}{4} \right\rfloor}{(n-2)}$. However, if the aim is to prevent explosion, then setting $\alpha = \frac{n+1}{4(n-2)}$, and hence taking the smallest empirical quantile, is recommendable. Since we only consider data in general position, preventing explosion is more important here. Thus, in the remainder of this paper, we choose α to be equal to

$$\alpha_{opt} = \frac{n+1}{4(n-2)}\,. \tag{3.6}$$

As Rousseeuw and Hubert (1996) point out, the finite sample breakdown point tends to a meaningful limit which they call *asymptotic breakdown point*. Here, all interval limits for the α attaining the maximum fsbp tend to 0.25 as n goes to infinity. So, the maximal asymptotic breakdown point for the considered scale estimates is 0.25 for $\alpha = 0.25$. For other values of α, the asymptotic breakdown point equals $\min\{(1-\alpha)/3, \alpha\}$.

3.2 Influence Functions

The Influence Function (IF) quantifies the difference in estimated scale due to adding small amounts of outliers to the data. The uncontaminated time series is

denoted by y_t and, for deriving the IF, we assume local linearity and a constant scale within the time window considered. Hence,

$$y_i = a + bi + \epsilon_i \sigma \tag{3.7}$$

for $i = 1, \ldots, n$, where $\epsilon_i \overset{iid}{\sim} F_0$. Typically, F_0 will be taken as the standard normal $N(0,1)$. Since all our estimation methods are regression invariant, we assume that $a = b = 0$ in equation (3.7) without loss of generality. As defined by Hampel (1974), the *influence function* of a scale functional S at the model distribution F is given by

$$\mathrm{IF}(w, S, F) \;=\; \lim_{\varepsilon \downarrow 0} \frac{S((1-\varepsilon)F + \varepsilon \Delta_w) - S(F)}{\varepsilon}\,, \tag{3.8}$$

where Δ_w denotes the point mass distribution at w for every $w \in \mathbb{R}$. For each possible value w, $\mathrm{IF}(w, S, F)$ quantifies the change in estimated scale when a very small proportion of all observations is set equal to the value w. Applying definition (3.8) to the Q^{α}_{adj} functional (2.4), and taking the standard normal distribution $N(0,1)$ for F, we obtain the following expression for the influence function:

$$\mathrm{IF}(w, Q^{\alpha}_{adj}, N(0,1)) = c_q \, \frac{-G(Q^{\alpha}_N, w)}{2\sqrt{2/3}\;\varphi\left(\sqrt{2/3}\;Q^{\alpha}_N\right)}\,, \tag{3.9}$$

where c_q and Q^{α}_N are defined according to (2.5) and

$$\begin{aligned} G(Q^{\alpha}_N, w) \;=\; & -3\big(2\Phi(\sqrt{2/3}\,Q^{\alpha}_N) - 1\big) + \Phi(\sqrt{2}(Q^{\alpha}_N - w)) - \Phi(\sqrt{2}(-Q^{\alpha}_N - w)) \\ & + 2\big(\Phi(\sqrt{(4/5)}((w/2) + Q^{\alpha}_N)) - \Phi(\sqrt{(4/5)}((w/2) - Q^{\alpha}_N))\big). \end{aligned} \tag{3.10}$$

The analytical derivation of this expression can be found in Appendix A. The IF of the Q^{α}_{adj} estimator for $\alpha = 0.25$ is depicted in the upper left panel of Figure 1. We notice three important properties: the IF is smooth, bounded and symmetric. Smoothness implies that a small change in one observation results in a small change of the estimated scale. Because the influence function is bounded, large outliers only have a limited impact on the estimated scale. As soon as the value of an outlier exceeds a certain level (approximately 9), the IF is flat and the exact magnitude

of the outlier is of no importance anymore for the amount by which the estimated scale increases. Finally, we note that the influence function is symmetric around zero, i.e. a negative and positive outlier of the same size have an equal effect on the estimated scale.

Influence functions have also been computed for the estimators based upon trimmed sums of (squared) heights. We only present the outcomes here; the mathematical derivations can be found in Appendix A. Let M denote one of the moment based functionals in equations (2.8) or (2.9), then the influence function at the standard normal $N(0,1)$ is given by

$$\begin{aligned} \mathrm{IF}(w, M, N(0,1)) &= \frac{c^p}{p\,\alpha}\Big[-(Q_N^\alpha)^p\, G(Q_N^\alpha) - 3\frac{\alpha}{c^p} + \sqrt{2}\left(I_{\sqrt{2},\sqrt{2}} + I_{-\sqrt{2},\sqrt{2}}\right) \\ &\quad +2\sqrt{\frac{4}{5}}\left(I^p_{\sqrt{1/5},\sqrt{4/5}} + I^p_{-\sqrt{1/5},\sqrt{4/5}}\right)\Big], \end{aligned} \tag{3.11}$$

with $p = 1$ and $c = c_m$ for TM^α_{adj} while $p = 2$ and $c = c_s$ for the TMS^α_{adj} estimator. In the above expression, we also need the integral

$$I^p_{a,b} = \int_0^{Q_N^\alpha} h^p \varphi(aw + bh) dh\,,$$

which can be computed analytically (see Appendix A). The upper right panel of Figure 1 shows the IF for TM^α_{adj}, where α equals 0.25. It shows the same properties as the influence function of Q^α_{adj} – it is smooth, bounded and symmetric. In the middle left panel we see the corresponding IF of the TMS estimator, which is remarkably close to that of TM. When comparing the influence function of the three robust estimators, sharing the same breakdown point, we can see that they are very similar.

In the middle right and lower panel of Figure 1, the IF of the non robust estimators, TM^α_{adj} and TMS^α_{adj} with $\alpha = 1$, are plotted. The influence functions are smooth and symmetric but unbounded. As expected, the IF of the TMS-method is clearly quadratic, while the IF of the TM-approach resembles the absolute value function. For smaller values of α, the difference between the IFs of the two trimmed mean approaches becomes much less pronounced.

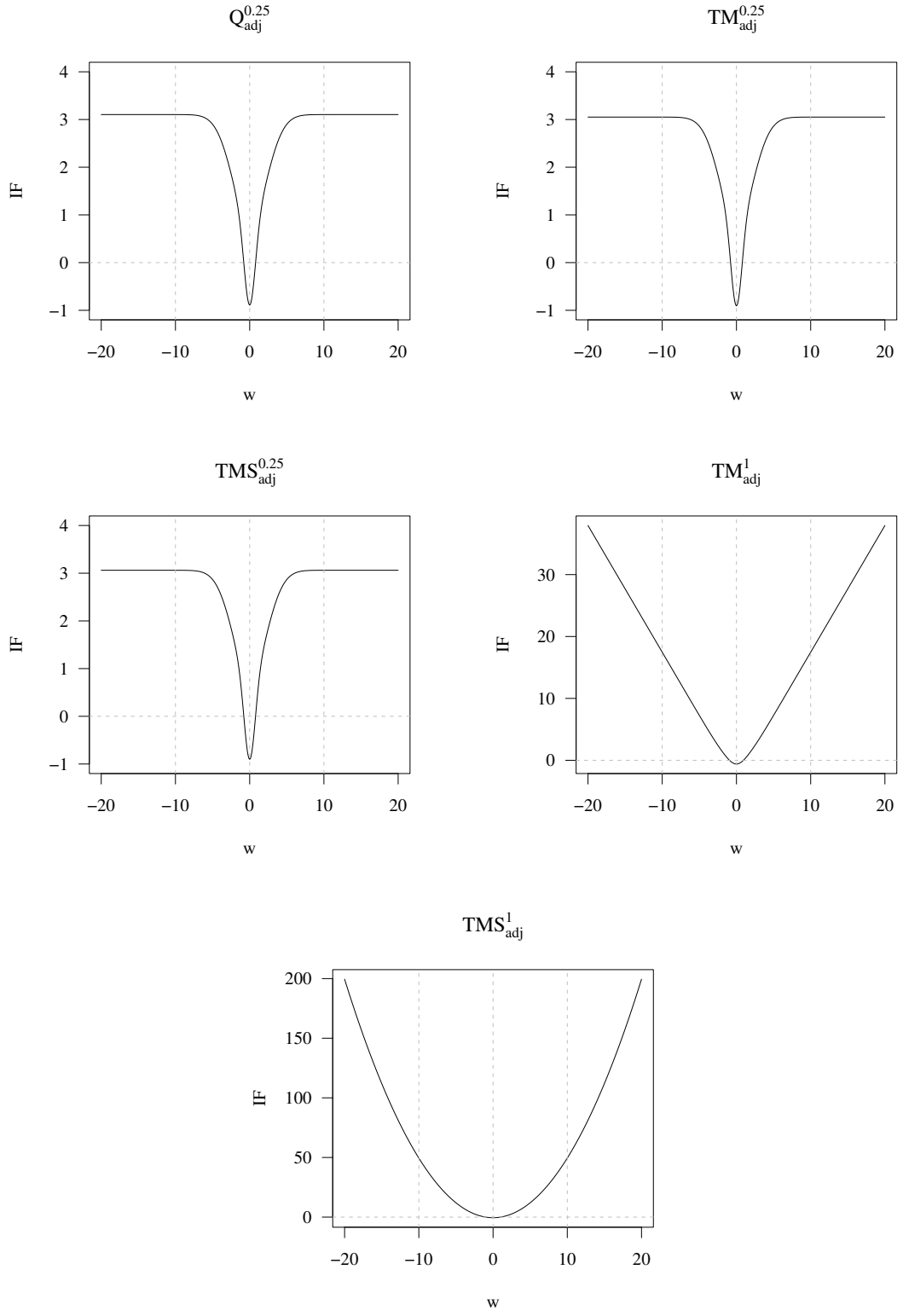

Figure 1: Influence functions for the Q^{α}_{adj}, TM^{α}_{adj}, and TMS^{α}_{adj} estimator, for $\alpha = 0.25$ and for $\alpha = 1$.

Finally, we also simulated empirical influence functions at finite samples to confirm the quite complicated expression for the theoretical IF (see Appendix B). It can be observed that already for $n = 20$, the empirical IF is very close to its theoretical counterpart.

4 Statistical Efficiencies

The efficiency of an estimator measures its relative precision and is related to its asymptotic variance (ASV). Here, we study the efficiency of an estimator S relative to the non-robust TMS^1_{adj} estimator:

$$\text{Efficiency}(S, F) = \frac{\text{ASV}(TMS^1_{adj}, F)}{\text{ASV}(S, F)}\,.$$

We maintain the local linearity assumption (3.7) and let F indicate the distribution of the error terms, supposed to be independent. Computing the asymptotic variance of the scale estimators requires caution because the estimators are based on heights of triangles, and these heights are autocorrelated. Similar to Portnoy (1977), we can write the asymptotic variance of an estimator based on the heights h_i as

$$\text{ASV}(S, F) = \sum_{l=-\infty}^{+\infty} E\big(\psi(h_i, S, H_F)\,\psi(h_{i+l}, S, H_F)\big), \tag{4.1}$$

where $\psi(h_i, S, H_F)$ is the influence function of the estimator S as a function of the heights h_i, which follow distribution H_F determined by F. Note that $\psi(h_i, S, H_F)$ is different from the influence function as described in Section 3, where we examine the effect of an outlying *observation*, while here we need the influence function of the *heights*, as these are the elements in the construction of the estimators. If the error terms in equation (3.7) are independently distributed, the heights are auto-correlated up to two lags, and equation (4.1) reduces to

$$\begin{aligned}\text{ASV}(S, F) \;&=\; E\big(\psi^2(h_i, S, H_F)\big) + 2\,E\big(\psi(h_i, S, H_F)\,\psi(h_{i+1}, S, H_F)\big)\\ &\quad + 2\,E\big(\psi(h_i, S, H_F)\,\psi(h_{i+2}, S, H_F)\big).\end{aligned}$$

As in Jureckova and Sen (1996), when F is a standard normal distribution, the ψ-functions for our estimators are given by

$$\begin{aligned}
\psi(h, Q^{\alpha}_{adj}, H_N) &= c_q \left(\frac{\alpha - I(h < Q^{\alpha}_N)}{2\sqrt{2/3}\,\psi\left(\sqrt{2/3}\,Q^{\alpha}_N\right)} \right) \\
\psi(h, TM^{\alpha}_{adj}, H_N) &= \tfrac{c_m}{\alpha}\big(hI(h < Q^{\alpha}_N) + Q^{\alpha}_N(\alpha - I(h < Q^{\alpha}_N))\big) - 1 \\
\psi(h, TMS^{\alpha}_{adj}, H_N) &= \tfrac{c_s^2}{2\alpha}\big(h^2 I(h < Q^{\alpha}_N) + (Q^{\alpha}_N)^2(\alpha - I(h < Q^{\alpha}_N))\big) - \tfrac{1}{2},
\end{aligned}$$

where Q^{α}_N is the α-quantile of the distribution of the heights under the standard normal distribution (see equation (2.5)), N is an index referring to the assumption of normality and I is the indicator function. The exact value of the ASV for the non-trimmed mean-squared-heights estimator TMS^1_{adj} equals 35/36. For the other estimators, the ASV is obtained by numerical integration. The left panel of Figure 2 evaluates the ASV of the estimators relative to the ASV of the TMS^1_{adj} estimator. Naturally, the efficiencies are higher for higher values of α, except for the Q^{α}_{adj} where the efficiency decreases steeply when α is larger than 0.86. The TMS^{α}_{adj} estimator is slightly more efficient than the TM^{α}_{adj} estimator for every value of α. Surprisingly the most efficient scale estimator is the Q^{α}_{adj}, at least for α smaller than 0.85. Hence, replacing the quantile by a trimmed sum does not result in an increase of efficiency for a large range of values of α.

At the optimal breakdown point of 25%, where α equals 0.25, we obtain an efficiency of only 25% for the Q^{α}_{adj} estimator and of around 20% for both trimmed mean estimators. Hence the price paid for the maximal breakdown point is very high. Taking the median of the heights, $\alpha = 0.5$, results in an efficiency of 49% for the Q^{α}_{adj}, 0.38% for the TM^{α}_{adj} and 0.43% for the TMS^{α}_{adj} estimator. These efficiencies are more reasonable and hence $\alpha = 0.5$ is recommended. Then, the asymptotic breakdown point is 16.6% and the finite sample breakdown point (see (3.5)) allows for three outliers in a window of 20 observations.

To compare the asymptotic and finite sample behavior of the estimators, the right panel of Figure 2 presents a simulated approximation of the ASV for window

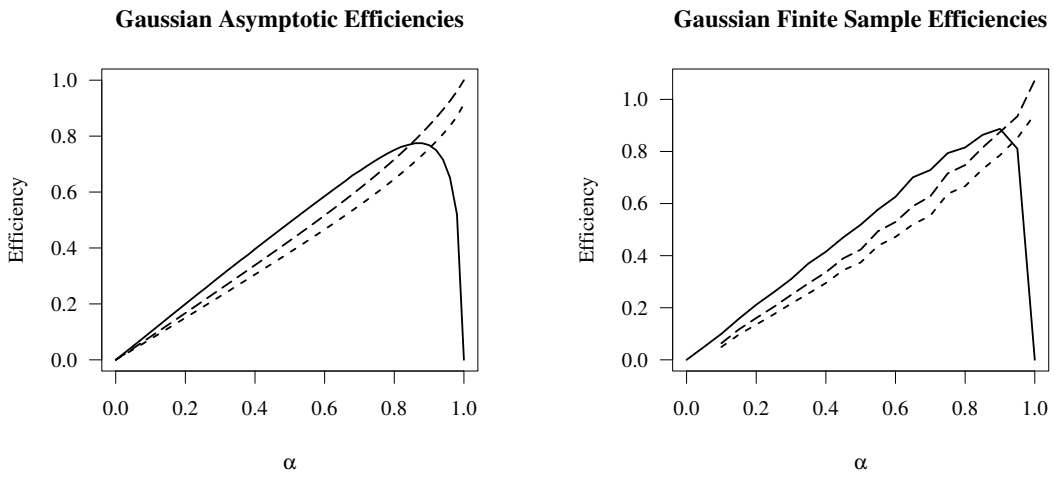

Figure 2: Asymptotic (left panel) and finite sample (window width 20, right panel) efficiencies for Q_{adj}^{α} (solid), TM_{adj}^{α} (short dash) and TMS_{adj}^{α} (long dash), for varying α.

width $n = 20$ in the moving window approach:

$$ASV(S, F) \approx n \operatorname{Var}(S_n, F)\,,$$

where $\operatorname{Var}(S_n, F)$ is obtained by computing the scale estimate S_n 10000 times for a simulated time series of length n with i.i.d. standard normal noise. Comparing the right and left panel of Figure 2 indicates that a window width of 20 already provides a good approximation of the asymptotic variance and that the ordering of the scale estimates remains unchanged in the finite sample setting. Moreover, a much more elaborated simulation study, summarized in Appendix B, has been undertaken, where times series where generated from different sampling schemes, including outlier generating ones. A conclusion is that for values of α not too close to one, the three different robust procedures remain quite close to each other under a large variety of sampling scheme. The simulation results confirm that taking α too small results in too large a loss of efficiency. We thus suggest using $Q_{adj}^{0.5}$ in practice: This estimator yields a good compromise between robustness and efficiency.

5 Applications

In this section we present an artificial data example to illustrate the online scale estimation methods and two real data applications, one application in finance and one in medicine.

5.1 Artificial data

The running scale approach is illustrated using a simulated time series of length 500, starting with a quadratic trend. After a trend change (at observation 250) and a level shift (at observation 250), the trend becomes linear. The true scale σ_t is constant at one for the first 300 observations, then jumps to three and grows linearly thereafter. Contamination is only included in the subseries with linear trend, i.e. starting from observation 251 on. We include 5% replacement outliers from a $N(0, 10^2)$ around the linear trend. The upper graph in Figure 3 plots the time series, while the bottom graph shows the estimated scales using either the $Q_{adj}^{0.5}$, or the non-robust standard deviation computed from an OLS-fit within each window considered. The latter estimation approach is called here a *running sd*. The true scale function σ_t, which is known here, is also presented.

As can be seen from Figure 3, the $Q_{adj}^{0.5}$ estimator performs quite well. The shift in the magnitude of the scale (after observation 300) is detected with some delay since for the first observation after this shift, most observations included in the window for scale estimation are still from the period before the scale shift. Furthermore, the $Q_{adj}^{0.5}$ estimator can easily cope with the non-linearities in the signal of the times series and with the presence of the outliers in this time series.

Comparing this with the scale estimates which use the running sd approach, one can first notice that during the period of the quadratic trend, when no outliers are present, the true scale is systematically overestimated. The reason for this is that the running sd method relies on the local linearity assumption to be true within each window. The latter assumption is clearly violated in the first part of

the series. As expected, the running sd approach is not robust w.r.t. the trend and level shift in the signal at $t = 250$, resulting in a high spike. Finally, in the last part of the series, the running sd is again substantially overestimating the true scale, now caused by the presence of outliers in the second part of this time series.

5.2 Real data applications

To illustrate the use of the online scale estimation methods for financial data, we look at Apple Computer, Inc. stock returns (AAPL). The more volatile the returns of a stock are, the more risky it seems to invest in it. The upper panel of Figure 4 plots the returns of the AAPL stock from July 5th 2000 until September 27th 2006. These returns are based on daily closing prices. There are a few large negative outliers, which indicate that the stock price during that particular day decreased steeply. The lower panel of Figure 4 presents the scale, estimated using both the $Q_{adj}^{0.5}$ and the running sd-estimator, here for $n = 20$. Note that the negative outliers strongly influence the running sd-estimates during certain time periods. This is undesirable since we do not want a single isolated observation to potentially result in extremely high scale estimates for several periods. If we are not in the neighborhood of outliers, then the robust and non-robust approaches give similar results. During the period we consider, the volatility of the stock return has decreased. From the beginning of the period until the beginning of 2003, the AAPL stock has become less risky. From then on, the volatility has stabilized.

The second application concerns heart rate measurements recorded at an intensive care unit once per second. The top panel in Figure 5 shows a time series of such heart rate measurements plus $N(0, 0.01^2)$-noise, added to prevent the scale estimates from imploding due to measurement accuracy. The first part of the time series seems rather stable with a few positive outlying values while at around 22:27h not only does the heart rate of the patient suddenly increase but also its variability.

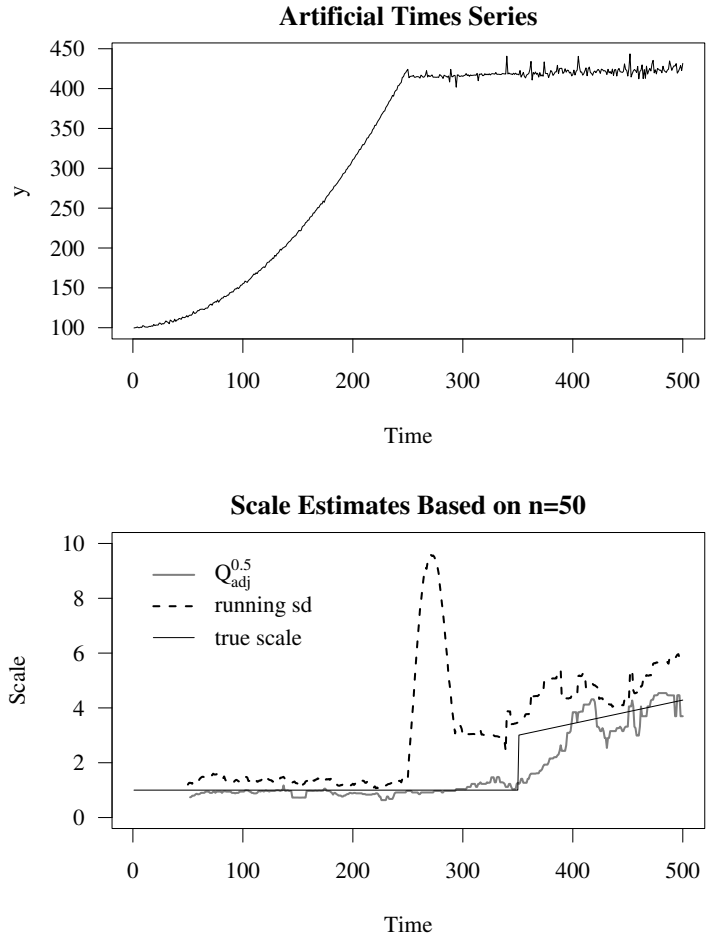

Figure 3: Artificial time series (top panel). The bottom panel presents the scale as estimated by the $Q^{0.5}_{adj}$ estimator and the residual standard deviation after an OLS-fit with $n = 50$. The true scale is represented by the thin solid line.

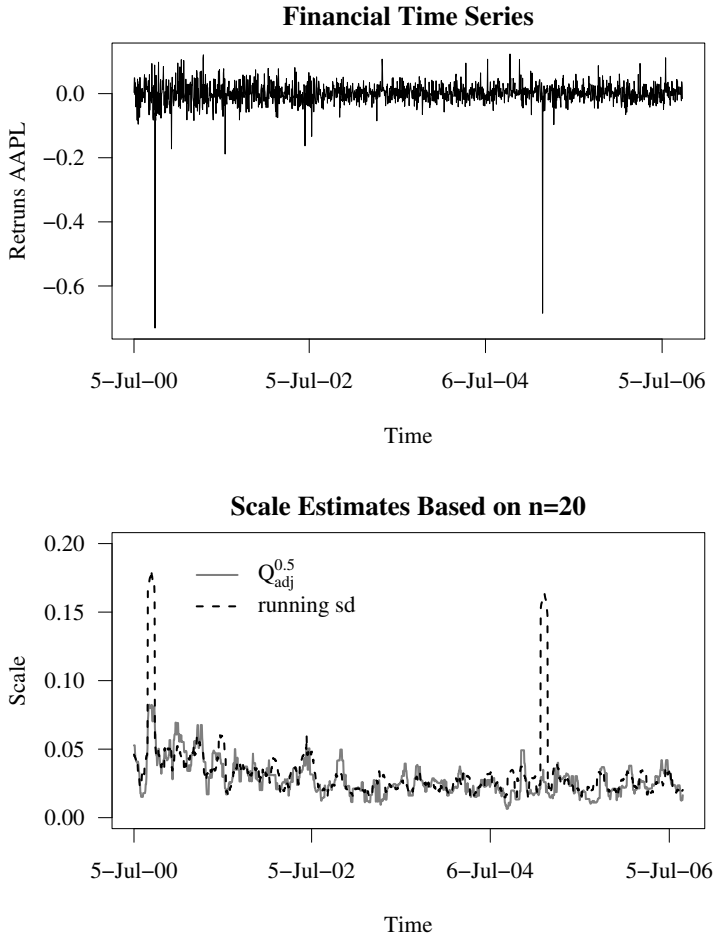

Figure 4: AAPL stock returns (top panel). The bottom panel presents the scale as estimated by the $Q^{0.5}_{adj}$ estimator and the residual standard deviation after an OLS-fit with $n = 20$.

The bottom panel presents again the $Q_{adj}^{0.5}$ and the running sd estimator using a window width of $n = 240$ seconds. Both methods detect the sudden increase in variability. However, the effect of the outliers on the running sd clearly motivates the need for robust methods in this application. Similar to the artificial example, the running sd estimates outstanding large variability around the level shift which does not reflect the data. This results from the preceding regression step where the level around the jump is not estimated correctly and thus, the residuals indicate a large variability. This problem occurs for all regression-based scale estimates, including robust approaches such as a running Q_n scale estimate (Rousseeuw and Croux (1993)) based on the residuals from a repeated median regression (Siegel (1982)), as described in Fried and Gather (2003).

Additionally, regression-based methods estimate the variability around a line within the whole window while the proposed adjacent-type methods are only based on a linear approximation for three consecutive data points and hence rather estimate short-term variability. Figure 5 demonstrates this: the estimations from the running standard deviation are larger than the Q_{adj}^{α} estimations, especially during the period of increased variability.

6 Conclusion

This paper studies regression-free scale estimation procedures for online application in time series. The estimators are based on the heights of adjacent triangles which makes them suitable for time series with non-linearities. Moreover, it is shown that the presented methods perform well for time series with trend changes, level changes, time varying scale and outliers. This is confirmed by theoretical and simulation based evidence, as well as by real data examples including a financial and a physiological application. The estimators achieve a maximal asymptotic breakdown point of 25% while the Q_{adj}^{α} estimator, based on the α-quantile of heights, turns out to have the best performance in terms of efficiency. Choosing α

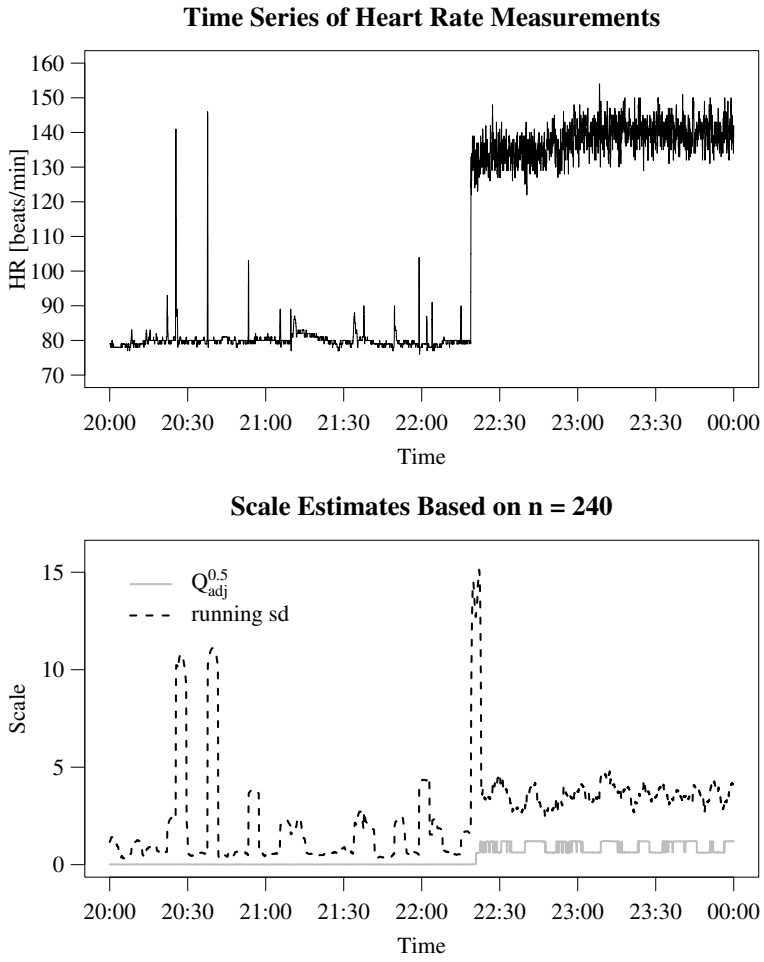

Figure 5: Physiological data (top panel). The bottom panel presents the scale as estimated by the $Q^{0.5}_{adj}$ estimator and the residual standard deviation after an OLS-fit with $n = 240$.

to be equal to 0.5 provides both, reasonable robustness and efficiency.

The proposed online scale monitoring procedure is easy to implement since all scale estimates are defined explicitly. For every new observation, the associated estimate S_t requires only $O(\log n)$ computing time, allowing for fast online computation. The selection of the window length n is not treated in this paper. We assume here that the practitioner provides some subject matter information for a reasonable choice for n. Alternatively, an automatic adaptive selection procedure for n could be developed, similar as in Gather and Fried (2004).

Besides allowing for fast and easy online computation, the estimates based on adjacent triangles are more robust in the face of non-linearities than other existing robust scale estimation procedures in the time series context. For deriving the theoretical influence function and asymptotic efficiencies, we have required local linearity and that the error terms within a single window be independent. However, these assumptions are required only to maintain analytical tractability of the theoretical results. When used as an explorative tool in an applied time series context, the latter assumptions are by no means required.

Acknowledgements

We gratefully acknowledge the financial support of the German Science Foundation (DFG, SFB 475 "Reduction of Complexity for Multivariate Data Structures"), the the European Science Foundation (ESF-network SACD "Statistical Analysis of Complex Data with Robust and Related Statistical Methods") and the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (Contract number G.0594.05).

Appendix A: Proofs

Consistency Factors

To obtain the asymptotic consistency factor for the Q^{α}_{adj} estimator, we rely on the local linearity assumption with normally distributed i.i.d. error terms as in equation (3.7). If $\epsilon_i \overset{idd}{\sim} N(0,\sigma^2) = F$, then we need to select c_q such that $Q^{\alpha}_{adj}(F) = \sigma$. The heights for triangles formed by equidistant and consecutive time points are defined by (2.2). Now under the local linearity assumption, substitute (3.7) in (2.2) to obtain

$$h_i = \left|\left(\epsilon_{i+1} - \frac{\epsilon_i + \epsilon_{i+2}}{2}\right)\sigma\right| = \left|v_i\sqrt{\frac{3}{2}}\,\sigma\right|. \tag{A.1}$$

with $v_i \sim N(0,1)$.

Since Q^{α}_{adj} is defined by the correction constant c_q times the α quantile of all heights, it follows that

$$P\left(c_q\,\sqrt{3/2}\,\sigma\,|v_i| \le Q^{\alpha}_{adj}(F)\right) = 2\,P\left(v_i \le \frac{Q^{\alpha}_{adj}(F)}{c_q\,\sqrt{3/2}\,\sigma}\right) - 1 = \alpha$$

$$\Rightarrow \Phi\left(\frac{Q^{\alpha}_{adj}(F)}{c_q\,\sqrt{3/2}\,\sigma}\right) = \frac{\alpha+1}{2} \Rightarrow Q^{\alpha}_{adj}(F) = \Phi^{-1}((\alpha+1)/2)\,c_q\,\sqrt{3/2}\,\sigma\,.$$

Hence, in order to get Fisher consistency, we need to select

$$c_q = \sqrt{2/3}\;\frac{1}{\Phi^{-1}((\alpha+1)/2)}\,.$$

Note that $c_q = (Q^{\alpha}_N)^{-1}$.

Using definition (2.8) and equation (A.1), we get

$$\begin{aligned} TM^{\alpha}_{adj}(F) &= c_m\,E\left(\sigma\sqrt{3/2}|v| \;\middle|\; \sigma\sqrt{3/2}|v| \le Q^{\alpha}_N\;\sigma\right) \\ &= c_m\frac{\sigma\,\sqrt{3/2}}{\alpha}\left(2\int_0^{\sqrt{\frac{2}{3}}Q^{\alpha}_N} v\,\varphi(v)dv\right). \\ &= c_m\,\frac{\sigma\,\sqrt{6}}{\alpha}\left[\varphi(0) - \varphi(\sqrt{2/3}\,Q^{\alpha}_N)\right]. \end{aligned}$$

Hence, to get $TM^{\alpha}_{adj}(F) = \sigma$, we take

$$c_m = \frac{\alpha}{\sqrt{6}\,\left[\varphi(0) - \varphi(\sqrt{\frac{2}{3}}Q^{\alpha}_N)\right]}\,.$$

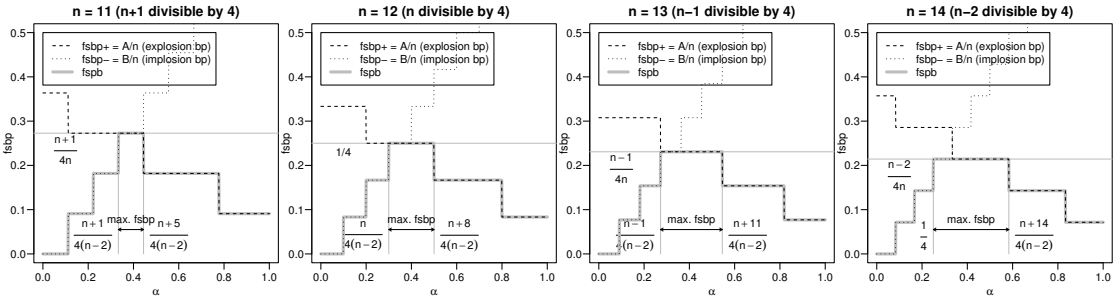

Figure 6: Finite sample breakdown points, explosion and implosion breakdown points for varying values of α for exemplary n.

For the TMS^{α}_{adj} defined by (2.9), we get

$$\begin{aligned} TMS^{\alpha}_{adj}(F) &= c_s\, E\left(\sigma^2 \frac{3}{2} v^2 \Big| |v| \le \sqrt{\frac{2}{3}} Q^{\alpha}_N\right)^{1/2} \\ &= c_s \sqrt{\frac{3}{2}} \sqrt{\frac{2}{\alpha}}\, \sigma \left(\int_0^{\sqrt{\frac{2}{3}} Q^{\alpha}_N} v^2 \varphi(v) dv\right)^{1/2} \\ &= c_s \sqrt{\frac{3}{\alpha}}\, \sigma \left(-\sqrt{2/3}\, Q^{\alpha}_N\, \varphi(\sqrt{2/3}\, Q^{\alpha}_N) + \frac{\alpha+1}{2} - \frac{1}{2}\right)^{1/2}. \end{aligned}$$

Fisher consistency is ensured by taking

$$c_s = \frac{\sqrt{\alpha/3}}{\sqrt{\alpha/2 - \sqrt{2/3}\, Q^{\alpha}_N\, \varphi\left(\sqrt{2/3}\, Q^{\alpha}_N\right)}}.$$

The Finite Sample Breakdown Point

In the following, the maximum value possible for the finite sample breakdown point $\mathrm{fsbp}(S, F_n)$, given by (3.5), of any of the considered scale estimates S ((2.3), (2.6) or (2.7)) at a sample $\mathbf{y}_n$ of size n with empirical distribution function F_n is determined. Further, the corresponding values of α for which the maximum breakdown is achieved are derived. Therefore, consider the quantities

$$A := \left\lceil \frac{n - 1 - \lfloor \alpha(n-2) \rfloor}{3} \right\rceil \qquad \text{and} \qquad B := \lfloor \alpha(n-2) \rfloor\,.$$

By (3.5), the finite sample breakdown point $\text{fsbp}(S, F_n)$ equals $\min\{A, B\}/n$.

For increasing values of $\alpha \in [0, 1]$, the quantity A is decreasing while B is increasing (see Figure 6). Hence, the maximum breakdown point will be equal to B/n for certain $\alpha \in [\alpha_{min}, \alpha_1)$, and it will equal A/n for certain $\alpha \in [\alpha_1, \alpha_{max})$. The maximum value for $\text{fsbp}(S, F_n)$ is reached for any value of α for which $A = B$ whenever this is possible. To determine the bounds α_{min} and α_{max}, we distinguish between the four cases where either $n+1$, n, $n-1$ or $n+2$ respectively is divisible by four.

Case I: $n \in \{4k - 1, k \in \mathbb{N}\}$.

(a) Let $\alpha = \frac{n+1}{4(n-2)} - \varepsilon$ with arbitrary small $\varepsilon > 0$. Then

$$B = \left\lfloor \left(\frac{n+1}{4(n-2)} - \varepsilon \right) (n-2) \right\rfloor = \left\lfloor \frac{n+1}{4} - \varepsilon(n-2) \right\rfloor < \frac{n+1}{4}\,.$$

Hence $B/n < \frac{n+1}{4n}$ for $\alpha < \alpha_{min} := \frac{n+1}{4(n-2)}$.

(b) Consider $x \in [n+1, n+5)$ and $\alpha = \frac{x}{4(n-2)}$. Then $B = \lfloor x/4 \rfloor = (n+1)/4$,

$$A = \left\lceil \frac{n-1-B}{3} \right\rceil = \left\lceil \frac{n-1-(n+1)/4}{3} \right\rceil = \left\lceil \frac{n+1}{4} - \frac{2}{3} \right\rceil = \frac{n+1}{4} = B$$

and hence $A/n = B/n = \frac{n+1}{4n}$ for $\alpha \in \left[\frac{n+1}{4(n-2)}, \frac{n+5}{4(n-2)} \right)$.

(c) For $\alpha = \frac{n+5}{4(n-2)}$ it is $B = \lfloor (n+5)/4 \rfloor = (n+1)/4 + 1$ and

$$A = \left\lceil \frac{n-1-(n+1)/4-1}{3} \right\rceil = \left\lceil \frac{n+1}{4} - 1 \right\rceil = \frac{n+1}{4} - 1\,.$$

Thus, $A/n < \frac{n+1}{4n}$ for all $\alpha \geq \alpha_{max} := \frac{n+5}{4(n-2)}$.

From (a), (b), (c) and as A and B are decreasing and increasing in α, respectively, the maximum value for $\text{fsbp}(S, F_n) = \frac{n+1}{4n}$ is reached for $\alpha \in \left[\frac{n+1}{4(n-2)}, \frac{n+5}{4(n-2)} \right)$.

Case II: $n \in \{4k, k \in \mathbb{N}\}$. Analogously to Case I, one can show:

(a) For $\alpha = \frac{n}{4(n-2)} - \varepsilon$, $\varepsilon > 0$, it is $B = \left\lfloor \frac{n}{4} - \varepsilon(n-2) \right\rfloor < \frac{n}{4}$ and hence $\frac{1}{n}B < \frac{1}{4}$ for $\alpha < \alpha_{min} := \frac{n}{4(n-2)}$.

(b) **(i)** Consider $x \in [n, n+4)$ and $\alpha = \frac{x}{4(n-2)}$. Then $B = \lfloor x/4 \rfloor = n/4$ and $A = \left\lceil \frac{n-1-n/4}{3} \right\rceil = \left\lceil \frac{n}{4} - \frac{1}{3} \right\rceil = \frac{n}{4}$, and thus $A/n = B/n = 1/4$ for $\alpha \in \left[\frac{n}{4(n-2)}, \frac{n+4}{4(n-2)} \right)$.

(ii) For $x \in [n+4, n+8)$ and $\alpha = \frac{x}{4(n-2)}$, it is $B = \lfloor x/4 \rfloor = n/4+1$ and $A = \left\lceil \frac{n}{4} - \frac{2}{3} \right\rceil = \frac{n}{4}$. Following it is $A/n = 1/4 < B/n$ for $\alpha \in \left[\frac{n+4}{4(n-2)}, \frac{n+8}{4(n-2)} \right)$.

(c) For $\alpha = \frac{n+8}{4(n-2)}$ it is $B = \lfloor (n+8)/4 \rfloor = n/4 + 2$ and $A = \left\lceil \frac{n}{4} - 1 \right\rceil = \frac{n}{4} - 1$. Hence, it is $A/n < \frac{1}{4} \; \forall \; \alpha \geq \alpha_{max} := \frac{n+8}{4(n-2)}$.

Thus, the maximum value $1/4$ for the finite sample breakdown point $\text{fsbp}(S, F_n)$ is reached for $\alpha \in \left[\frac{n}{4(n-2)}, \frac{n+8}{4(n-2)} \right)$.

Case III: $n \in \{4k+1, k \in \mathbb{N}\}$

Analogously one can show that

(a) $\frac{1}{n}B < \frac{n-1}{4n}$ for $\alpha < \alpha_{min} := \frac{n-1}{4(n-2)}$.

(b) **(i)** $\frac{1}{n}A = \frac{1}{n}B = \frac{n-1}{4n}$ for $\alpha \in \left[\frac{n-1}{4(n-2)}, \frac{n+3}{4(n-2)} \right)$.

(ii) $\frac{1}{n}A = \frac{n-1}{4n} < \frac{1}{n}B$ for $\alpha \in \left[\frac{n+7}{4(n-2)}, \frac{n+11}{4(n-2)} \right)$

(c) $\frac{1}{n}A < \frac{n-1}{4n} \; \forall \; \alpha \geq \alpha_{max} := \frac{n+11}{4(n-2)}$.

Thus, the maximum value $\frac{n-1}{4n}$ for $\text{fsbp}(S, F_n)$ is reached for $\alpha \in \left[\frac{n-1}{4(n-2)}, \frac{n+11}{4(n-2)} \right)$.

Case IV: $n \in \{4k+2, k \in \mathbb{N}\}$

Again, analogously one can show that

(a) $\frac{1}{n}B < (n-2)/4n$ for $\alpha < 1/4$.

(b) $\frac{1}{n}B = \frac{n-2}{4n}$ for $\alpha \in \left[\frac{1}{4}, \frac{n+2}{4(n-2)} \right)$.

(c) $\frac{1}{n}A = \frac{n-2}{4n}$ for $\alpha \in \left[\frac{n+2}{4(n-2)}, \frac{n+14}{4(n-2)} \right)$.

(d) $\frac{1}{n}A \leq \frac{n-2}{4n} - \frac{1}{n}$ for any $\alpha \geq \frac{n+14}{4(n-2)}$.

From (a), (b), (c) and (d) we can conclude that the maximum breakdown point $\text{fsbp}(S, F_n) = \frac{n-2}{4n}$ is reached for $\alpha \in \left[\frac{1}{4}, \frac{n+14}{4(n-2)}\right)$. Note that in this case, the equation $A = B$ does not hold for any value of $\alpha \in [0, 1]$ because:

$$\begin{aligned} & A = B \\ \Leftrightarrow \quad & \left\lceil \frac{n-1-\lfloor \alpha(n-2) \rfloor}{3} \right\rceil = \lfloor \alpha(n-2) \rfloor \\ \Leftrightarrow \quad & \lfloor \alpha(n-2) \rfloor - 1 < \tfrac{n-1-\lfloor \alpha(n-2) \rfloor}{3} \leq \lfloor \alpha(n-2) \rfloor \\ \Leftrightarrow \quad & \frac{n-1}{4} \leq \lfloor \alpha(n-2) \rfloor < \frac{n+2}{4}\,. \end{aligned}$$

By definition $\lfloor \alpha(n-2) \rfloor \in \mathbb{N}$. However, for $n \in \{4k+2, k \in \mathbb{N}\}$ there is no integer in the interval $[\frac{n-1}{4}, \frac{n+2}{4})$.

Influence Functions

For the derivation of equation (3.9), recall that the functional of the Q^{α}_{adj} estimator is defined by (2.4)

$$Q^{\alpha}_{adj}(F) \;=\; c_q\, H^{-1}_F(\alpha),$$

for any distribution F of the error terms ϵ_i. Here,

$$H_F(u) = P\big(\big|\epsilon_{i+1} - \frac{\epsilon_i + \epsilon_{i+2}}{2}\big| \leq u\big),$$

for all $u > 0$. The IF as defined in (3.8) then equals

$$\text{IF}(w, Q^{\alpha}_{adj}, F) = c_q\, \frac{\partial}{\partial \varepsilon}\; H^{-1}_{F_\varepsilon}(\alpha)\big|_{\varepsilon\,=\,0} = c_q\, \frac{-\frac{\partial H_{F_\varepsilon}}{\partial \varepsilon}\left(H^{-1}_F(\alpha)\right)\big|_{\varepsilon\,=\,0}}{H'_F\left(H^{-1}_F(\alpha)\right)}\,. \tag{A.2}$$

Here, assuming $F = N(0, 1)$, we have

$$H^{-1}_F(\alpha) = H^{-1}_N(\alpha) = Q^{\alpha}_N = \sqrt{\frac{3}{2}}\; \Phi^{-1}\left(\frac{\alpha+1}{2}\right)\,,$$

referring to the assumption of normality by the index N.

We now compute numerator and denominator of (A.2) separately. For the numerator, we need the distribution function of the heights from the contaminated

series y_t^ε, which can be written as

$$\begin{aligned} H_{F_\varepsilon}(u) &= (1-\varepsilon)^3 H_F(u) + \varepsilon(1-\varepsilon)^2 P\big(\big|w - \frac{y_{t-1}+y_{t+1}}{2}\big| \le u\big) \\ &\quad +2\varepsilon(1-\varepsilon)^2 P\big(\big|y_t - \frac{w+y_{t+1}}{2}\big| \le u\big) + O(\varepsilon^2)\,, \end{aligned}$$

for any $u > 0$. Computing the derivative of this expression with respect to ε and evaluating in $\varepsilon = 0$ yields

$$\frac{\partial H_{F_\varepsilon}(u)}{\partial \varepsilon}\Big|_{\varepsilon\,=\,0} = -3\,H_F(u) + P\big(\big|w - \frac{y_{t-1}+y_{t+1}}{2}\big| \le u\big) + P\big(\big|y_t - \frac{w+y_{t+1}}{2}\big| \le u\big).$$

Since $F = N(0,1)$, standard calculations give

$$\begin{aligned} \frac{\partial H_{N_\varepsilon}}{\partial \varepsilon}(h)\Big|_{\varepsilon\,=\,0} &= -3\big(2\Phi(\sqrt{2/3}\,h) - 1\big) + \Phi(\sqrt{2}(h-w)) - \Phi(\sqrt{2}(-h-w)) \\ &\quad +2\big(\Phi(\sqrt{(4/5)}((w/2)+h)) - \Phi(\sqrt{(4/5)}((w/2)-h))\big) \\ &:= G(h,w), \end{aligned} \tag{A.3}$$

and we note that $G(Q_N^\alpha)$ equals expression (3.10).

For the denominator of expression (A.2), we note that

$$H_F(u) = P(h \le u) = P\big(\big|y_{t+1} - \frac{y_t + y_{t+2}}{2}\big| \le u\big) = \Phi(\sqrt{2/3}u) - \Phi(-\sqrt{2/3}u)$$

from which it follows that

$$\frac{\partial H_F(u)}{\partial u}\Big|_{u\,=\,Q_N^\alpha} = 2\sqrt{2/3}\,\varphi(\sqrt{2/3}\;Q_N^\alpha). \tag{A.4}$$

The expression of the $IF(Q_{adj}^\alpha, N(0,1), w)$ for normal distributed error terms follows from equations (A.2), (A.3) and (A.4).

The functionals TM_{adj}^α and TMS_{adj}^α both take the same form

$$M : F \mapsto M(F) = c\,E(h^p|\,h \le Q_F^\alpha)^{1/p} = \frac{c}{\alpha}\left(\int_0^{Q_F^\alpha} h^p\,dH_F(h)\right)^{1/p}, \tag{A.5}$$

where $p = 1$ for the TM_{adj}^α and $p = 2$ for the TMS_{adj}^α, c is either c_m (equation (2.11)) or c_s (equation (2.12)) and $Q_F^\alpha = H_F^{-1}(\alpha)$. The influence function, as defined in equation (3.8), is given by

$$\mathrm{IF}(w, M, F) = \frac{\partial}{\partial \varepsilon}\frac{c}{\alpha}\left(\int_0^{Q_{F_\varepsilon}^\alpha} h^p\,dH_{F_\varepsilon}(h)\right)^{1/p}\Big|_{\varepsilon\,=\,0},$$

and the chain rule delivers

$$\mathrm{IF}(w,M,F) = \frac{c}{\alpha}\frac{1}{p}E(h^q|\, h \le Q_F^\alpha)^{(1/p)-1}\,\frac{\partial}{\partial\varepsilon}\int_0^{Q^\alpha_{F_\varepsilon}} h^p dH_{F_\varepsilon}(h)\big|_{\varepsilon=0}.$$

We know that $E(h^p|\, h \le Q_F^\alpha) = 1/c^p$, so

$$\begin{aligned}\mathrm{IF}(w,M,F) &= \frac{c^p}{p\,\alpha}\left[\frac{\partial}{\partial p}\int_0^p h^p dH_F(h)\big|_{p=Q_F^\alpha}\frac{\partial}{\partial\varepsilon}Q^\alpha_{F_\varepsilon}\big|_{\varepsilon=0} + \frac{\partial}{\partial\varepsilon}\int_0^{Q_F^\alpha} h^p dH_{F_\varepsilon}(h)\big|_{\varepsilon=0}\right]\\ &:= \frac{c^p}{p\,\alpha}(T_1+T_2)\,.\end{aligned}$$

We will now separately examine the first and second term T_1 and T_2.

Applying the Leibnitz rule to the first term yields

$$T_1 = (Q_F^\alpha)^p H_F'(Q_F^\alpha)\,\frac{\partial}{\partial\varepsilon}Q^\alpha_{F_\varepsilon}\big|_{\varepsilon=0}.$$

Recall now that the influence function of Q^α_{adj} can be rewritten as

$$\mathrm{IF}(w,Q_{adj},N(0,1)) = c_q\,\frac{\partial}{\partial\varepsilon}Q^\alpha_{N_\varepsilon}\big|_{\varepsilon=0} = c_q\,\frac{-G(Q_N^\alpha,w)}{H_N'(Q_N^\alpha)}\,,$$

hence for $F = N(0,1)$

$$T_1 = -(Q_N^\alpha)^p\,G(Q_N^\alpha,w)\,.$$

To obtain the second term T_2, we use the expression for G in (A.3) and obtain

$$\begin{aligned}T_2 &= \int_0^{Q_N^\alpha} h^p dG(h)\\ &= -3\int_0^{Q_N^\alpha} h^p dH_F(h) + \int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{2}(h-w)) - \int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{2}(-h-w))\\ &\quad + 2\int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{(4/5)}((w/2)+h) - 2c/\alpha) - 2\int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{(4/5)}((w/2)-h))\\ &= -3\frac{\alpha}{c^p} + \int_0^{Q_N^\alpha} h^p\Big[\sqrt{2}\big(\varphi(\sqrt{2}(w+h)+\varphi(\sqrt{2}(w-h)\big)\\ &\quad + 2\sqrt{4/5}\big(\varphi(\sqrt{4/5}(w/2+h)) + \varphi(\sqrt{4/5}(w/2-h))\big)\Big]dh\,.\end{aligned}$$

Defining

$$I^p_{a,b} = \int_0^{Q_N^\alpha} h^p\varphi(aw+bh)dh\,, \tag{A.6}$$

results in expression (3.11) for $\mathrm{IF}(w,M,N(0,1))$

In expression (3.11) of the influence function, there still appear integrals which have to be computed. In practice, we can do this by numerical integration techniques; or we can derive an analytic expression for $I^p_{a,b}$. Using the substitution method for solving integrals, define $u = aw + bh$ and rewrite equation (A.6) as

$$I^p_{a,b} = \int_{aw}^{aw+bQ} \left(\frac{u-aw}{b}\right)^p \varphi(u)\frac{1}{b}\,du\,.$$

We now distinguish between $p = 1$ and $p = 2$. For $p = 1$ one can verify that

$$I^1_{a,b} = \frac{\varphi(aw) - \varphi(aw+bQ)}{b^2} + \frac{aw}{b^2}\left(\Phi(aw) - \Phi(aw+bQ)\right)\,,$$

and for $p = 2$, after some calculations, we get

$$I^2_{a,b} = \frac{1}{b^3}\big[(aw-bQ)\varphi(aw+bQ) - aw\varphi(aw) + (1+a^2w^2)\Phi(aw+bQ) - (1+a^2w^2)\Phi(aw)\big]$$

In the special case where $\alpha = 1$, we obtain simplified expressions for the mean over all heights and the mean over all squared heights:

$$\begin{aligned}\mathrm{IF}(w, TM^1_{adj}, N(0,1)) \;&=\; c_m\Big[-3\sqrt{6}\,\varphi(0) + 2w\,\Phi(\sqrt{2}w) + \sqrt{2}\,\varphi(\sqrt{2}\,w) - w \\ &\quad + 2\left(\sqrt{5}\,\varphi(\sqrt{1/5}\,w) + w\,\Phi(\sqrt{1/5}\,w) - w/2\right)\Big]\,,\end{aligned}$$

and

$$\mathrm{IF}(w, TMS^1_{adj}, N(0,1)) = \frac{1}{2}\,(w^2 - 1).$$

Appendix B: Simulations

Correction Factors

In definitions (2.3), (2.6) and (2.7), consistency factors are used to achieve Fisher consistency. However, these estimators may still be biased for finite samples; replacing the consistency factors by correction factors then ensures unbiasedness of the scale estimators. Those correction factors c^n_q, c^n_m and c^n_s depend on the

	$\alpha = \frac{n+1}{4(n-2)}$			$\alpha = 0.5$			$\alpha = 1$	
n	Q^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}	Q^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}
10	2.18	4.40	4.07	1.27	2.61	2.32	1.02	0.85
15	2.43	4.90	4.42	1.34	2.76	2.42	1.02	0.84
20	2.61	5.27	4.70	1.24	2.57	2.24	1.02	0.84
50	2.46	4.98	4.34	1.22	2.53	2.18	1.02	0.82
100	2.57	5.17	4.49	1.22	2.52	2.18	1.02	0.82
200	2.57	5.18	4.49	1.21	2.52	2.17	1.02	0.82
∞	2.56	5.17	4.47	1.21	2.51	2.16	1.02	0.82

Table 2: Correction factors obtained by 10000 simulation runs for several values of n and α, together with the consistency factors ($n \to \infty$), under the assumption of Gaussian noise.

window width n and differ slightly from the consistency factors. To obtain the correction factors, we assume linearity and a constant scale σ within the time window considered, as in (3.7). Here, F_0 will be taken to be the standard normal $N(0,1)$. Table 2 gives values of the correction factors under the assumption of i.i.d. Gaussian error terms and based on 10000 simulation runs for different window widths and different values of α. In the first column, we choose α such that the optimal finite sample breakdown point is achieved, as is explained in Section 3.1. From Table 2 it is clear that as the window width increases, the correction factors converge to their asymptotic values. These asymptotic values are the consistency factors in the definition of the functionals in equations (2.3), (2.6) and (2.7).

In practice, we suggest using Q^{α}_{adj} with $\alpha = 0.5$ because this yields a good compromise between robustness and efficiency. For application to real data, we advise using the finite sample correction factor for this estimator. Though this can be obtained for any sample size by simulation, it might be rather time consuming. Therefore, we propose a simple approximative formula for the finite sample cor-

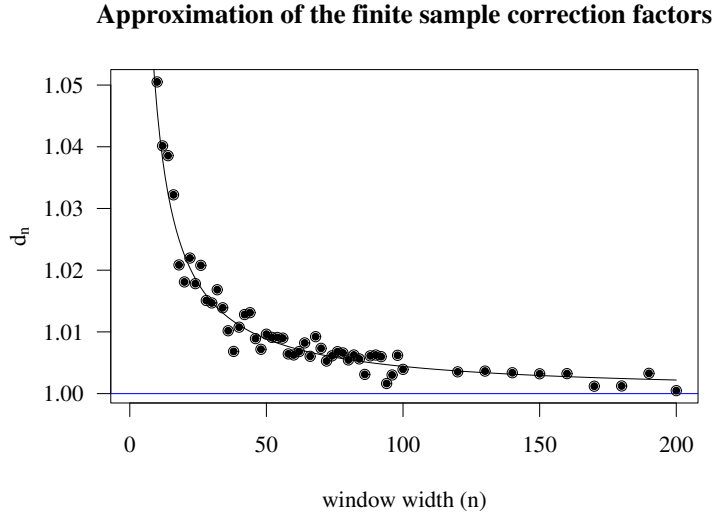

Figure 7: Approximation of the finite sample correction factors for the Q^{α}_{adj} for α equal to 0.5. The dots represent the simulated value of d_n for 10000 simulations and the solid line is the approximation using formula (B.7).

rection factor of the $Q^{0.5}_{adj}$ under the assumption of i.i.d. normal error terms. The correction factor c^n_q can be written as a multiple of its asymptotic value:

$$c^n_q = d_n \, c_q \, .$$

As c^n_q approaches c_q for n tending to infinity, we know that d_n attains one. We run 10000 simulations for varying window widths, up to $n = 200$ observations. A plot of the simulated values for d_n can be found in Figure 7. The shape of the plot is well approximated by a function of the form

$$d_n = \frac{n}{n-a},$$

where the constant a can be estimated by an OLS regression of the following model:

$$n(d_n - 1) = a \, d_n + u_t$$

which yields $a = 0.44$. Since c_q equals 1.21 for $\alpha = 0.5$, we get

$$d_n \approx \frac{n}{n-0.44} \Rightarrow c^n_q \approx 1.21 \, \frac{n}{n+0.44}. \tag{B.7}$$

This approximation gives values close to the large scale simulated values of d_n, as can be seen by the solid line in Figure 7, and is easy to use in practice.

For the pure descriptive purpose of monitoring scale, one will often not be willing to make distributional assumptions on the noise component. If this is the case, then the estimators may be computed omitting the correction factors. The estimated sequence S_t will of course be the same, up to scalar multiplication.

Empirical Influence Functions

The theoretical influence function in equation (3.8) is an asymptotic concept. In practice however, we work with finite sample sizes or, for the considered application, window widths. A finite sample equivalent of the influence function defined in (3.8) is given by the Empirical Influence Function (EIF). It is a simulation based curve and defined by

$$\mathrm{EIF}(w; S, F, n, N_s) = \frac{n}{N_s} \sum_{s=1}^{N_s} (S(\tilde{\mathbf{y}}^s) - S(\mathbf{y}^s)) \,, \tag{B.8}$$

where $n = 20$ the window width, and $N_s = 10000$ is the number of simulations. For every simulated sample $\mathbf{y}^s$ of size n from a $N(0, 1)$ distribution the scale estimate is computed. This sample is then contaminated by replacing one randomly selected observation by the value w, resulting in a contaminated series $\mathbf{y}^s$ from which again the scale estimator S is computed. Figure 8 shows the EIF of the different adjacent type estimators, similar to Figure 1, for $\alpha = 1$ and $\alpha = \alpha_{opt} \approx 0.29$ as defined in equation (3.6). It is striking how close the theoretical and empirical influence functions are already for $n = 20$.

Simulated bias and efficiency under contamination

In this section, a simulation study is carried out to compare the finite sample performance of the estimators in the moving window approach. The estimation methods are compared with respect to two criteria: their mean bias and root

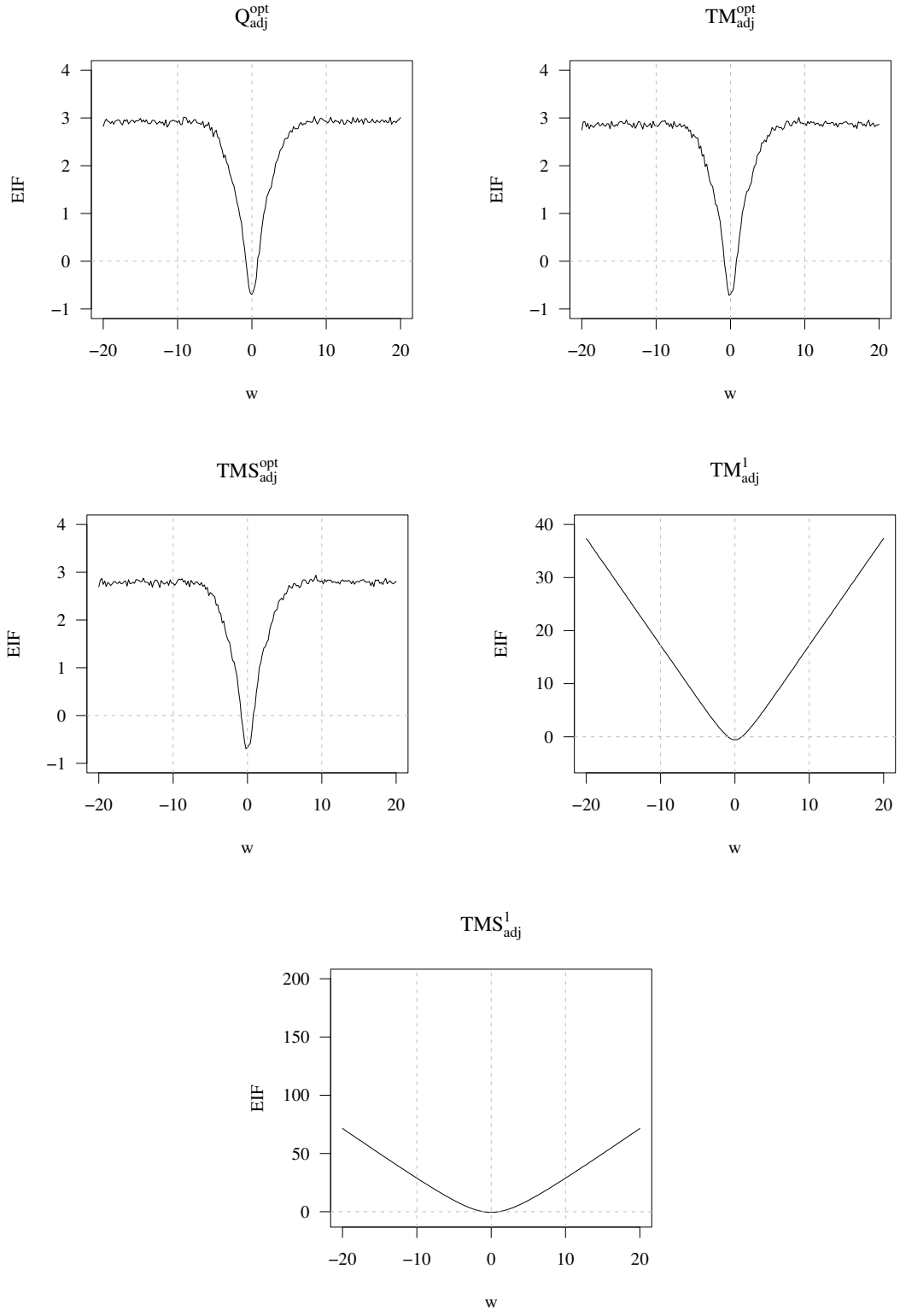

Figure 8: Empirical Influence functions (EIF) for the Q^{α}_{adj}, TM^{α}_{adj}, and TMS^{α}_{adj} estimators, for α corresponding to the optimal finite sample breakdown point and for $\alpha = 1$.

mean squared error. The Mean Bias (MB) of an estimator S is defined as the mean relative difference between the estimated and the true scale over all scale estimates resulting from the moving window:

$$\text{MB}\,(S) = \frac{1}{T-n+1}\sum_{t=n}^{T}\frac{S_t-\sigma_t}{\sigma_t}\,, \tag{B.9}$$

where n denotes the window width, T the length of the time series, σ_t the true scale at time t and S_t the estimated scale. Using the simulated finite sample correction factors, we expect the bias to be zero on average for all proposed estimators. It gives insight in the deviation of the estimated scale from the true scale. Another summary measure is the Mean Root Squared Error (MRSE):

$$\text{MRSE}\,(S) = \frac{1}{T-n+1}\left(\sum_{t=n}^{T}\frac{(S_t-\sigma_t)^2}{\sigma_t^2}\right)^{1/2}. \tag{B.10}$$

It measures the finite sample precision of the estimators.

We consider four simulation settings for which we simulate $N_s = 1000$ time series of length $T = 1000$. We choose the window width n equal to 20 and make use of the correction factors as presented above. For every simulated time series, we compute the mean bias and mean root squared error. The estimators are evaluated for α equal to $\alpha_{opt} = (n+1)/(4(n-2))$, where the optimal finite sample breakdown point is achieved, and α equal to 0.5. We also consider the non-robust versions of the TM^{α}_{adj} and TMS^{α}_{adj} estimators, where α equals 1. An overview of the simulation schemes can be found in Table 6.

Setting		Description
1	Clean data:	$y_t \overset{iid}{\sim} N(0,1)$
2	Fat tailed data:	$y_t \overset{iid}{\sim} t_3$
3	5% outliers:	$y_t \overset{iid}{\sim} 0.95N(0,1) + 0.05N(0,5)$
4	10% outliers:	$y_t \overset{iid}{\sim} 0.90N(0,1) + 0.10N(0,5)$

Table 3: Simulation Settings

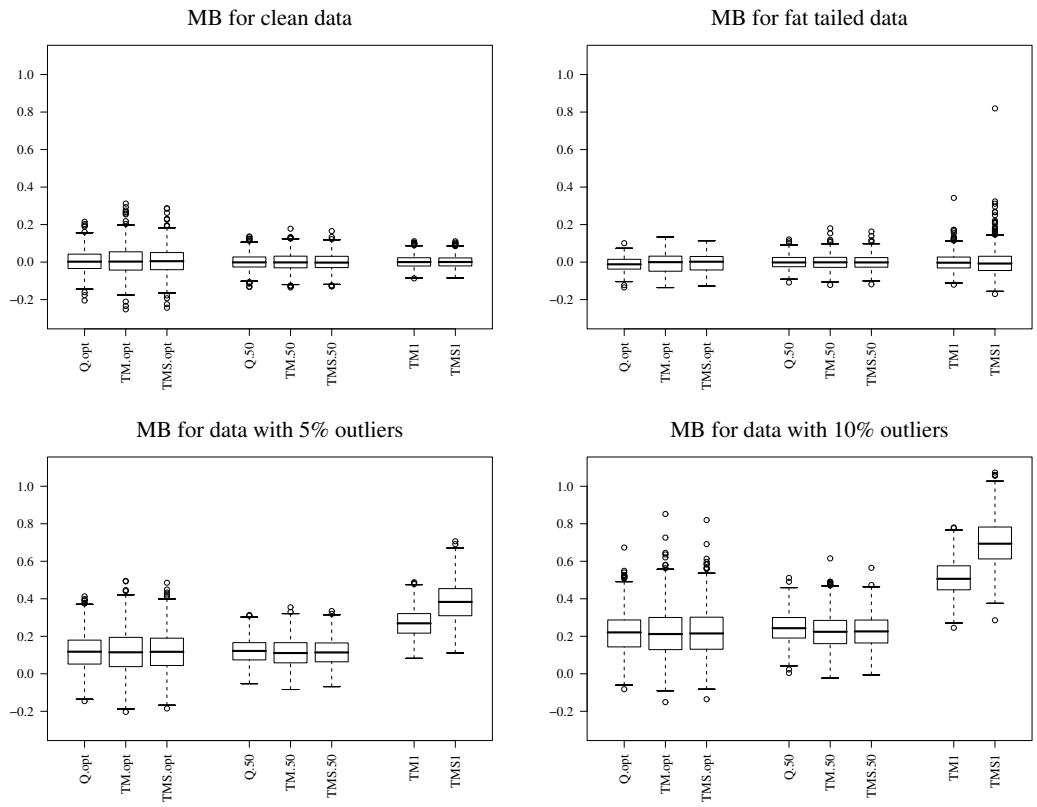

Figure 9: Mean bias for clean data (top left), fat tailed data (top right), data with 5% outliers (bottom left) and 10 % outliers (bottom right). In every graph, the first three boxplots present the MB for the Q^{α}_{adj}, TM^{α}_{adj} and TMS^{α}_{adj} estimators so that the optimal fsbp is achieved. The middle three boxplots correspond to the three estimators for α=0.5 and the last two boxplots represent the MB for the non-robust TM^{1}_{adj} and TMS^{1}_{adj} estimators.

	$\alpha = \alpha_{opt} = \frac{n+1}{4(n-2)}$			$\alpha = 0.5$			$\alpha = 1$	
	Q^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}	Q^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}
Clean data	0.44	0.54	0.51	0.29	0.34	0.32	0.22	0.21
Fat tailed data	0.43	0.48	0.47	0.36	0.39	0.37	0.33	0.41
5% outliers	0.52	0.62	0.59	0.38	0.41	0.40	0.51	0.70
10% outliers	0.60	0.70	0.67	0.50	0.52	0.51	0.75	1.00

Table 4: Average MRSE for the four simulation schemes, using window width n 20, and 3 different values of α: $\alpha = \frac{n+1}{4(n-2)}$ for obtaining the optimal fsbp, $\alpha = 0.5$ and $\alpha = 1$ for the non-robust estimators.

In the first simulation setting, we consider a time series of clean i.i.d. standard normal data. The mean bias is presented in the top left panel of Figure 9. As expected, all scale estimators are unbiased, i.e. the mean bias is on average zero. The largest biases occur for α_{opt}. The average of the MRSE over the 1000 time series is presented in the first row of Table 4. The non-robust procedures, where α equals 1, have the smallest variation of the estimated scale around the true scale. This is in line with the findings presented in Figure 2, where it is shown that the efficiency of both moment based estimators is higher for larger values of α.

In the second setting, we look at a heavy tailed distribution, namely a Student-t with three degrees of freedom. Again, we use finite sample correction factors to obtain unbiasedness. The mean bias is on average equal to zero, as illustrated in the top right panel of Figure 9. Table 4 indicates that the smallest MRSE is obtained by the non-robust TM^{1}_{adj} estimator, the difference in MRSE with the robust estimators where α equals 0.5 is small.

The third and fourth simulation settings assess the behavior of the scale estimation procedures for contaminated data. We induce respectively 5% and 10% outliers. The outliers come from a replacement outlier generating process with a proportion ε of the observations coming from a normal distribution with standard deviation 5. We consider 5% outliers in the bottom left panel of Figure 9 and 10%

in the bottom right panel. We use the finite sample correction factors obtained from simulations based on non-contaminated i.i.d. standard normal data. Under contamination, all procedures overestimate the scale, but the non-robust estimators TM^1_{adj} and TMS^1_{adj} perform particularly bad. The difference in bias between the estimators based on $\alpha = 0.25$ or 0.5 is small in both settings. Among the robust estimators, the $Q^{0.5}_{adj}$ has the smallest MRSE.

References

Davies, P., Fried, R., Gather, U., 2004. Robust signal extraction for on-line monitoring data. Journal of Statistical Planning and Inference 122, 65–78.

Davies, P., Gather, U., 2005. Breakdown and groups. The Annals of Statistics 33, 977–1035.

Fried, R., Gather, U., 2003. Robust estimation of scale for local linear temporal trends. Tatra Mountans Mathematical Publications 26, 87–101.

Gather, U., Fried, R., 2004. Methods and algorithms for robust filtering. In: Antoch, J. (Ed.), invited paper in COMPSTAT 2004, Proceedings in Computational Statistics. pp. 159–170.

Gather, U., Schettlinger, K., Fried, R., 2006. Online signal extraction by robust linear regression. Computational Statistics 21, 33–51.

Hampel, F., 1974. The influence curve and its role in robust estimation. Annals of statistics 69, 383–393.

Jureckova, J., Sen, P., 1996. Robust Statistical Procedures: Asymptotics and Interrelations. Wiley, New York.

Portnoy, S., 1977. Robust estimation in dependent situations. The Annals of Statistics 5, 22–43.

Rousseeuw, P., Croux, C., 1993. Alternatives to the median absolute deviation. Journal of the American Statistical Association 88, 1273–1283.

Rousseeuw, P., Hubert, M., 1996. Regression-free and robust estimation of scale for bivariate data. Computational Statistics and Data Analysis 21, 67–85.

Siegel, A., 1982. Robust regression using repeated medians. Biometrika 69, 242–244.