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

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

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

[2] *Fakultät Statistik, Technische Universität 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.

*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 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.

The scale estimates discussed in this paper are regression free, i.e. directly based on the observed data points without applying a local regression fit first. The advantage is that we do not have to make any choices for the estimation of

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 method we propose is an exploratory tool for online monitoring of the variability of a time series. It does not require any modeling of the underlying signal, and neither of the variability process. In this sense, the approach we take is not only regression-free but also *model-free*. It provides a complementary tool to procedures based on models which allow for prediction of future volatilities, as e.g. the conditional volatility models used in financial risk management, but which are usually neither robust nor work online for high frequency data.

The different candidate methods are described in Section 2. Their robustness properties are studied in Section 3 and their statistical efficiencies in Section 4. Section 5 presents a simulation study to assess the finite sample properties of the proposed methods in terms of mean bias and root mean squared error for various types of time series. Data applications can be found in Section 6. Finally, Section 7 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.

For deriving the influence function (Section 3.2) and the asymptotic variance (Section 4) of the estimators, we assume, for reasons of simplicity, that the noise components in (2.1) are independent. The model with independent errors serves as a testbed for a formal comparison of the different estimators. Most time series have dependent errors, however, and the procedures we propose are still applicable as descriptive measures of variability in this case. We conjecture that the relative performance of the different estimators carries over to models with dependent errors, as is confirmed by simulation experiments reported in Section 5.

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 $y_t - \hat{\mu}_t$, where $\hat{\mu}_t$ is an estimate of the level of the series. To estimate the latter, a robust regression filter as in Davies et al. (2004) and Gather et al. (2006) can be used. 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. Scale estimates based on that approach are considered in Fried and Gather (2003). However, in this case non-linearities can cause large bias. Alternatively, the signal could be estimated by an M-smoother with local linear fit, as in Rue et al. (2002), which can cope with jumps in the level (as in our approach). It is, however,

not clear how an online computable robust scale estimate should be defined here. A robust nonparametric regression scale curve estimator is proposed in Härdle and Tsybakov (1988). But again, besides the common problem of appropriate band-width selection, it is neither clear how to compute the estimator of Härdle and Tsybakov (1988) in an online setting. The scale estimators we propose in this paper are nonparametric in spirit and do not require preliminarily estimation of the signal μ_t which is the reason for referring to them as 'model-free'

2.1 Estimation methods

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. Note that this vertical height is different from the usual notion of height as a perpendicular length and corresponds to the non-zero residual of an L_1 fit to the three considered data points. Here, it is only assumed that 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

dependence 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 for the scale parameter 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

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

The functional form (2.4) is the probability limit of the estimator in (2.3). The vertical heights h_i are serially correlated, but under appropriate mixing conditions the empirical quantile will still converge to the associated population quantile. In particular, if the error terms in (2.1) are independent, then the vertical heights are only autocorrelated up to order two, and the estimator (2.3) will converge to (2.4).

Assuming normally distributed noise components, together with the assumption of having a linear trend within the considered window, 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 $c_q = 1.21$.

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). Note that for $\alpha = 1$, the estimator defined in (2.7) is not robust and coincides with the residual variance estimator in nonlinear regression proposed by Gasser et al. (1986). 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 $1/p$,

$$TM^{\alpha}_p : F \mapsto TM^{\alpha}_p(F) = E(X^p |\, X \leq H_F^{-1}(\alpha))^{1/p}\,, \tag{2.10}$$

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

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

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

with Q_N^α defined in (2.5), and $\varphi(z)$ the associated density of $\Phi(z)$. Details on how these expressions are obtained can be found in the first section of the Addendum[1], see Gelper et al. (2008). 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 Section 2 of the Addendum) 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_{adj}^α estimators, one could replace c_q by its finite sample counterpart c_q^n (obtainable by means of Monte-Carlo simulations). In the Addendum, a simple approximative formula for this finite sample factor c_q^n, with $\alpha = 0.5$, is derived:

$$c_q^n \approx 1.21\,\frac{n}{n + 0.44}. \tag{2.13}$$

When focussing on the estimation of the variance $E(\varepsilon_t^2)$, the robust measures of scale need to use the right correction factor, also if the distribution is not normal. The purpose of the paper is, however, to estimate variability in a broader sense than just the variance. Note that the variance as a measure of scale requires existence of the second order moment. For the pure descriptive purpose of monitoring scale, distributional assumptions on the noise component are not needed, and the estimators may be computed omitting the correction factors. The sequence of scale estimates S_t will, up to scalar multiplication, be the same.

[1]The Addendum, Gelper et al. (2008) is available on the following website: www.econ.kuleuven.be/sarah.gelper/public.

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. Finally, note that if one uses all possible triangles, formula (2.2) for the height of a triangle no longer applies, since most of these triangles are not based on equidistant observations.

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)):

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

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_{adj}^{α} estimator in a regression setup with random design; but we consider a fixed

	max. value of $\text{fsbp}(S, F_n)$	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 $\text{fsbp}(S, F_n)$ 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}.

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 the first section of the Addendum.

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.

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:

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

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 the first section of the Addendum. 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 results here; the mathematical derivations can be found in the Addendum. Let M denote one of the

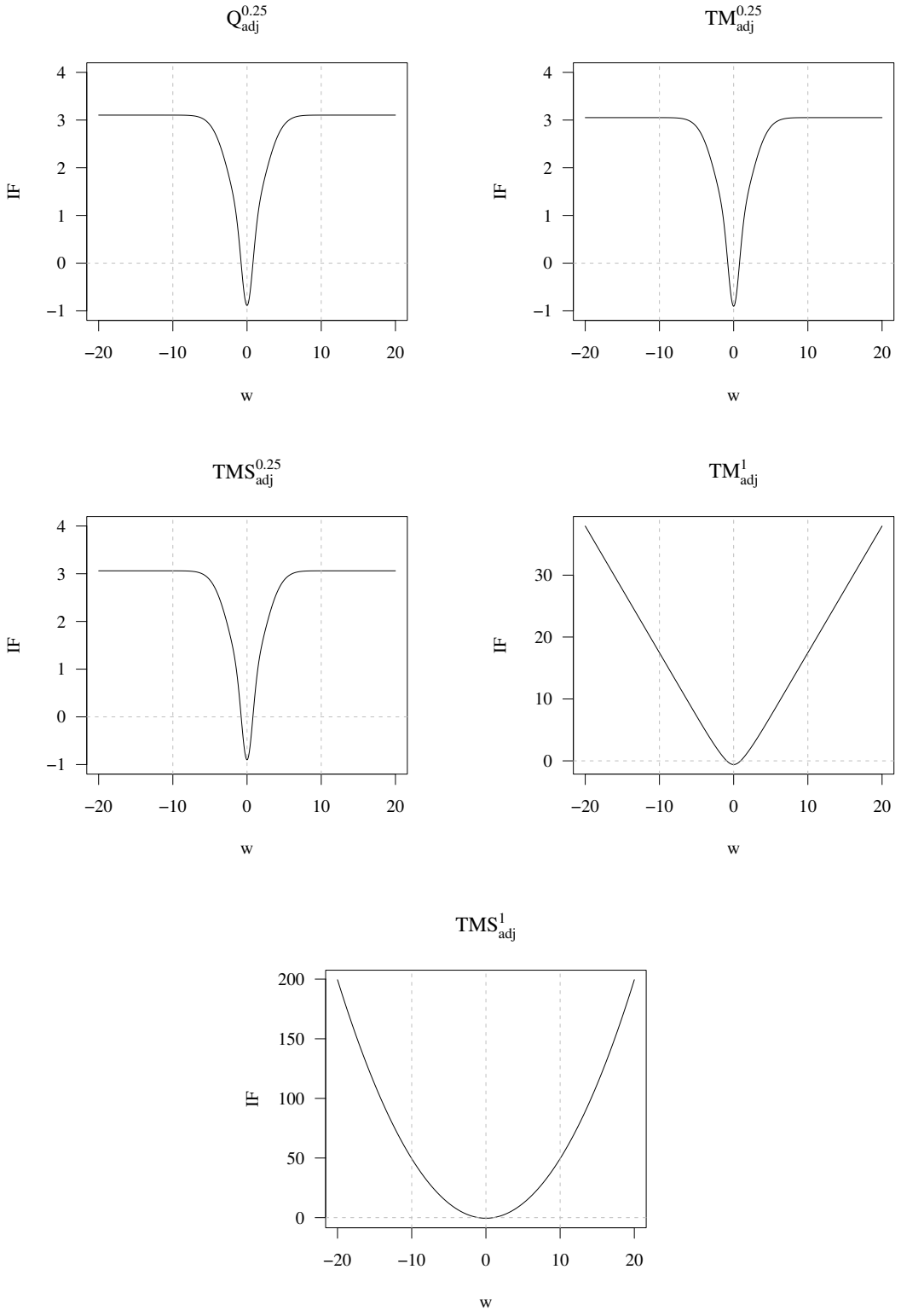

Figure 1: Influence functions for the Q^{α}_{adj}, TM^{α}_{adj}, and TMS^{α}_{adj} estimator, for $\alpha = 0.25$ and for $\alpha = 1$.

moment based functionals in equations (2.8) or (2.9), then the influence function at the standard normal $N(0,1)$ is given by

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

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 the Addendum). 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 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.

Finally, we also simulated empirical influence functions at finite samples to confirm the quite complicated expression for the theoretical IF (see the Addendum). 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 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

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

where Q_N^α 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_{adj}^1 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_{adj}^1 estimator. Naturally, the efficiencies are higher for higher values of α, except for the Q_{adj}^α where the efficiency decreases steeply when α is larger than 0.86. The TMS_{adj}^α estimator is slightly more efficient than the TM_{adj}^α estimator for every value of α. Surprisingly the most efficient scale estimator is the Q_{adj}^α, 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_{adj}^α 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_{adj}^α, 0.38% for the TM_{adj}^α and 0.43% for the TMS_{adj}^α 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 width $n = 20$ in the moving window approach:

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

where $\operatorname{Var}(S, F_n)$ is obtained by computing the scale estimate S 10000 times at a time series of length $n = 20$ with each F_n simulated from a model 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

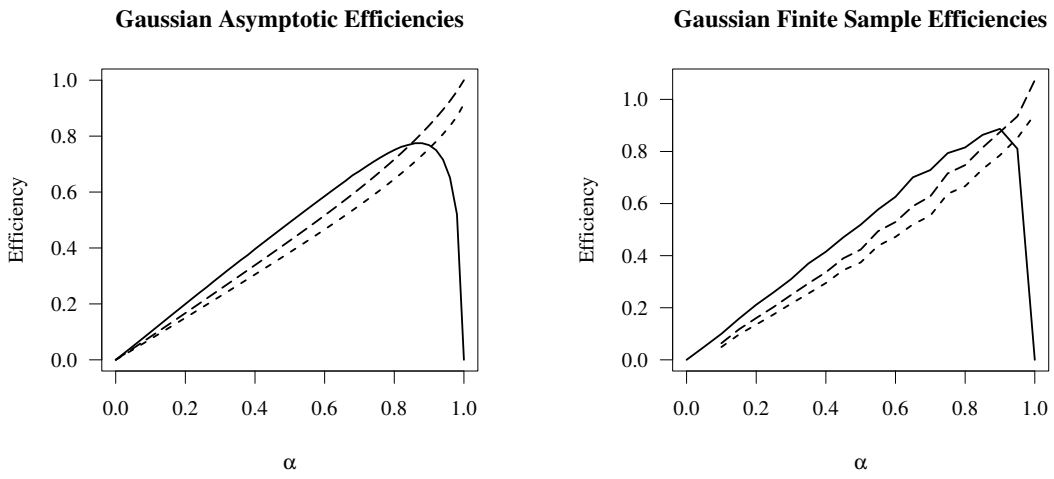

Figure 2: Asymptotic (left panel) and finite sample (window width 20, right panel) efficiencies for Q^{α}_{adj} (solid), TM^{α}_{adj} (short dash) and TMS^{α}_{adj} (long dash), for varying α.

finite sample setting.

5 Simulation Study

A simulation study is carried out to compare the finite sample performance of the estimators with respect to two criteria: the mean bias and root mean squared error. The Mean Bias (MB) of an estimator S is defined as the relative difference between the estimated and the true scale, averaged over each simulated time series:

$$\mathrm{MB}\,(S) = \frac{1}{T-n+1}\sum_{t=n}^{T}\frac{S_t-\sigma_t}{\sigma_t}\,, \tag{5.1}$$

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. Another summary measure is the Root Mean Squared Error (RMSE):

$$\mathrm{RMSE}\,(S) = \left(\frac{1}{T-n+1}\sum_{t=n}^{T}\frac{(S_t-\sigma_t)^2}{\sigma_t^2}\right)^{1/2}\,, \tag{5.2}$$

measuring the finite sample precision of the estimators. We simulate $N_s = 1000$ time series of length $T = 1000$, for several simulation settings. We take $n = 20$

and use the correction factors given by (2.13). For every simulated time series, we compute the mean bias and root mean squared error. We consider the robust estimators for $\alpha_{opt} = (n+1)/(4(n-2))$, 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 different simulation schemes can be found in Table 2.

In the first simulation setting, we consider time series of clean i.i.d. standard normal data. A boxplot of the $N_s = 1000$ values of the mean bias is presented in the top left panel of Figure 3. As expected, all scale estimators are nearly unbiased. The largest bias occurs for the estimates with $\alpha = \alpha_{opt}$. The average of the RMSE over the $N_s = 1000$ time series is presented in the first row of Table 3. 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 consider a heavy tailed distribution, namely a Student-t with three degrees of freedom. Again, appropriate finite sample correction factors are used. As can be seen from the top right panel of Figure 3, the mean bias is on average equal to zero. From Table 3 it follows that the smallest RMSE is now obtained by the non-robust TM^{1}_{adj} estimator, but the difference in RMSE with the robust estimators where α equals 0.5 is small.

The third and fourth simulation settings assess the behavior of the scale esti-

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 2: Simulation Schemes

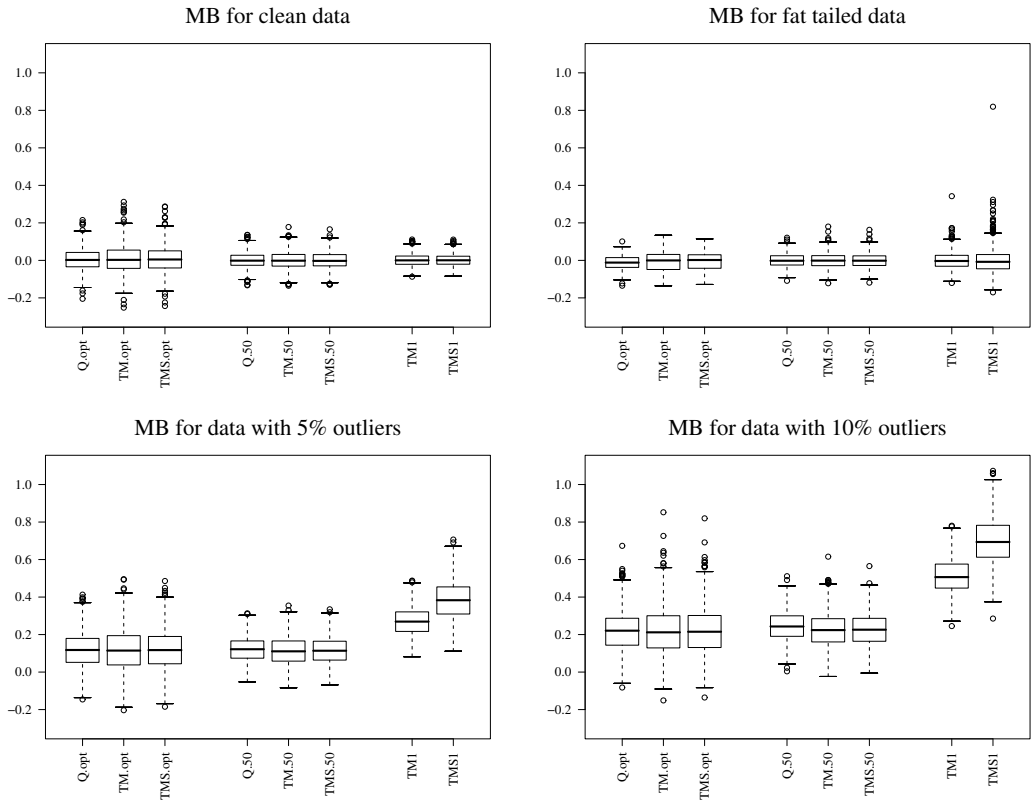

Figure 3: Boxplots of 1000 values of the Mean Bias (MB) for simulation schemes using clean data (top left), fat tailed data (top right), data with 5% outliers (bottom left) and 10 % outliers (bottom right). In every graph, the first three boxplots present the MB for the Q^{α}_{adj}, TM^{α}_{adj} and TMS^{α}_{adj} estimators with optimal fsbp; the middle three for α=0.5; the last two for the non-robust TM^{1}_{adj} and TMS^{1}_{adj}.

	$\alpha = \alpha_{opt} = \frac{n+1}{4(n-2)}$			$\alpha = 0.5$			$\alpha = 1$	
Setting	Q^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}	Q^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}	TM^{α}_{adj}	TMS^{α}_{adj}
1	0.44	0.54	0.51	0.29	0.34	0.32	0.22	0.21
2	0.43	0.48	0.47	0.36	0.39	0.37	0.33	0.41
3	0.52	0.62	0.59	0.38	0.41	0.40	0.51	0.70
4	0.60	0.70	0.67	0.50	0.52	0.51	0.75	1.00

Table 3: Average RMSE for clean data (1), fat tailed data (2), 5% outliers (3) and 10% outliers (4), using the window width $n = 20$, and 3 different values of α.

mation 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 3 and 10% in the bottom right panel. Under contamination, all procedures overestimate the scale, but the non-robust estimators TM^1_{adj} and TMS^1_{adj} perform particularly bad. The robust procedures also have some bias under contamination, but to a much smaller degree as the non-robust procedures. The difference in bias between the estimators based on $\alpha = 0.25$ and $\alpha = 0.5$ is small. The RMSE, as can be seen from Table 3, shows that $Q^{0.5}_{adj}$ has the smallest RMSE in presence of outliers. We thus suggest to use $Q^{0.5}_{adj}$ in practice, providing a good compromise between robustness and efficiency.

To complete the simulation study, we consider the case of dependent noise components. More specifically, we consider a GARCH setting, popular for analyzing financial time series. We simulate $N_s = 1000$ series of length $T = 1000$ from the GARCH(0,1) model

$$\begin{aligned} y_t &= z_t + I_t(d_{lo})w_{lo} \\ z_t &= \sigma_t \varepsilon_t + I_t(d_{vo})w_{vo} \\ \sigma_t^2 &= \alpha_0 + \alpha_1 z_{t-1}^2 \,, \end{aligned} \tag{5.3}$$

where ε_t is i.i.d. $N(0,1)$ such that σ_t measures the scale of the process y_t at time t. Here, $I_t(d) = 1$ if t belongs to a point in time included in the vector d, and zero if not. Following Hotta and Tsay (1998), we consider two kinds of outliers: level outliers (lo) and volatility outliers (vo). A level outlier is an isolated outlier in the level of the time series y_t, occuring at times d_{lo} having size w_{lo}. Volatility outliers occur at times d_{vo} and have a size of w_{vo}. A volatility outlier at t induces an increased level at period t and an increased volatility for the subsequent period $t+1$.

First, we consider clean data by setting both w_{lo} and w_{vo} equal to zero in

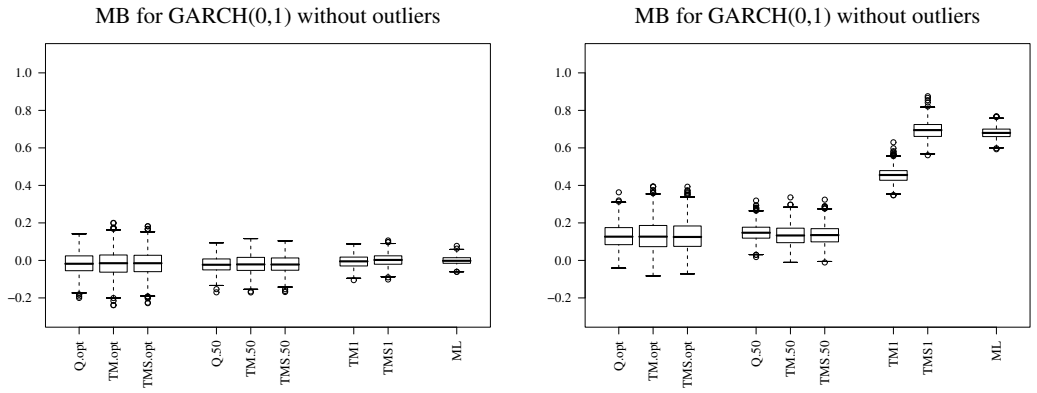

Figure 4: As Figure 3, but now for clean data (left) and 5% contaminated data (right) from a GARCH(0,1) model, and with the last boxplot corresponding to the ML estimator.

model (5.3). We then induce 5% contamination by creating alternating level and volatility outliers by setting $d_{lo} = (20, 60, 100, \ldots, 980)$, with $w_{lo} = 2$, and $d_{vo} = (40, 80, 120, \ldots, 1000)$, with $w_{vo} = 2$. The scale σ_t is then estimated using the proposed methods based on adjacent triangles, and using the standard Maximum Likelihood (ML) estimator in the GARCH model[2].

The boxplot of the $N_s = 1000$ values of the mean bias, computed again as in (5.1), is presented in Figure 4. The left panel corresponds to the clean GARCH(0,1) model. We see that the ML estimator is almost unbiased in this setting, and, as could be expected, it is the most precise estimator. The robust estimators, especially the Q^{α}_{adj} with $\alpha = 0.5$, still perform reasonably well. Note that there is a small downward bias in the methods based on adjacent triangles. This bias can be attributed to the use of correction factors for uncorrelated normal errors, while the unconditional distribution of y_t in model (5.3) is not normal. The right panel of Figure 4 presents the mean bias under contamination. Here, the non-robust estimators ML, TM^1_{adj}, and TMS^1_{adj}, result in a severe bias. From Figure 4 we see that also the robust estimators are somewhat biased under contamination, but to

[2]We use the standard GARCH routine provided in the *tseries* package of the R software.

a much smaller degree than the non-robust methods.

To show how the performance of the estimators changes over time, we compute the Mean Squared Error at every point in time,

$$\mathrm{MSE}_t(S) = \left(\frac{S_t - \sigma_t}{\sigma_t} \right)^2 .$$

Figure 5 shows the MSE_t for the subsequence of observations 440 to 560, containing three level outliers and two volatility outliers, averaged over $N_s = 1000$ simulations. (Similar plots are obtained for other subseries.) We compare the ML estimator with the advocated robust estimator $Q_\alpha^{0.5}$. In general, the MSE_t for the non-robust ML estimator is much larger than for the robust $Q_{adj}^{0.5}$ estimator. If a level outlier occurs, we observe an increased MSE_t of the ML estimator, while the $Q_{adj}^{0.5}$ estimator remains almost unaffected. The impact of a volatility outlier on the MSE of the ML estimator is also striking. Here one observes very small values of the MSE_t immediately after the occurrence of the volatility outlier, indicating that the ML estimator is completely dominated by these outliers and fits them well instead of the good data points. The $Q_{adj}^{0.5}$ is quite robust with respect to volatility outliers; the MSE_t only shows a small peak right after the occurrence of an outlier. We can conclude that (i) when using the robust estimator, the effect of level and volatility outliers is limited to the times when the outliers appear, (ii) when using the ML estimator, the effect of outliers is spread over the whole observation period.

6 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.

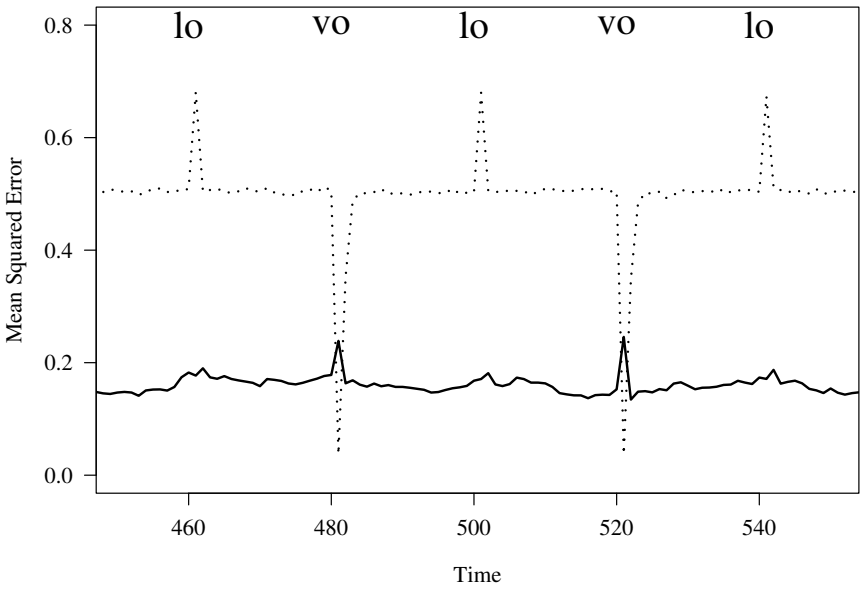

Figure 5: Average value of MSE_t evaluated from $N_s = 1000$ simulation runs for the time period 440 to 560 for the $Q_{adj}^{0.5}$ estimator (solid line) and the ML estimator of a GARCH(0,1) model (dotted line), in the presence of level outliers (lo) and volatility outliers (vo).

6.1 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 and equal to one for the first 300 observations, then it 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 $N(0, 10^2)$ distribution around the linear trend. The upper graph in Figure 6 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 6, the $Q_{adj}^{0.5}$ estimator performs quite well. 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 outliers in this time series. 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. Indeed, while the $Q_{adj}^{0.5}$ is robust with respect to additive outliers, level shifts, changing trends and a slowly changing scale, it is not robust with respect to a volatility shift.

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.

6.2 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 7 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 7 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 8 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 rounding error due to the limited 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.

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

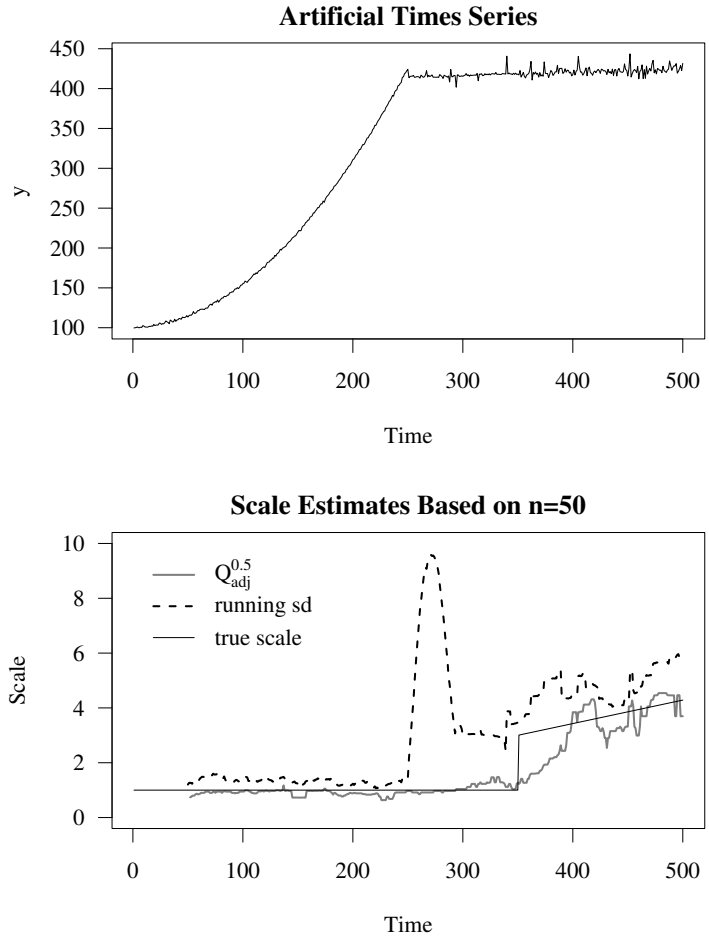

Figure 6: Artificial time series (top panel). The bottom panel presents the scale as estimated by the $Q^{0.5}_{adj}$ estimator and the residual standard deviation after an OLS-fit with $n = 50$. The true scale is represented by the thin solid line.

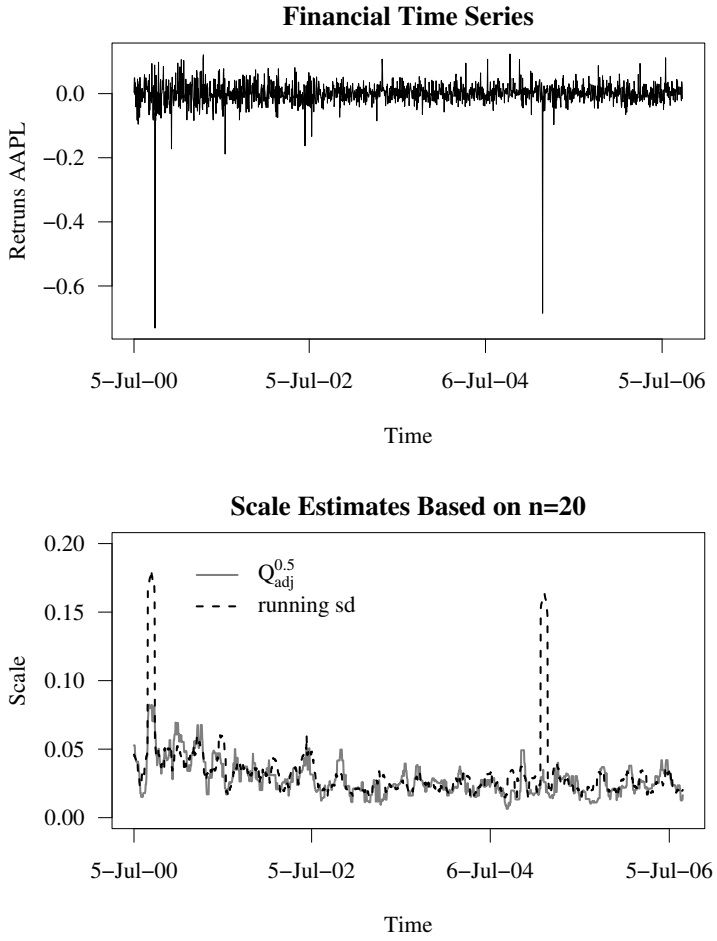

Figure 7: AAPL stock returns (top panel). The bottom panel presents the scale as estimated by the $Q_{adj}^{0.5}$ estimator and the residual standard deviation after an OLS-fit with $n = 20$.

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 8 demonstrates this: the estimations from the running standard deviation are larger than the Q^{α}_{adj} estimations, especially during the period of increased variability.

7 Conclusion

This paper studies model-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^{α}_{adj} estimator, based on the α-quantile of heights, turns out to have the best performance in terms of efficiency. Choosing α 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

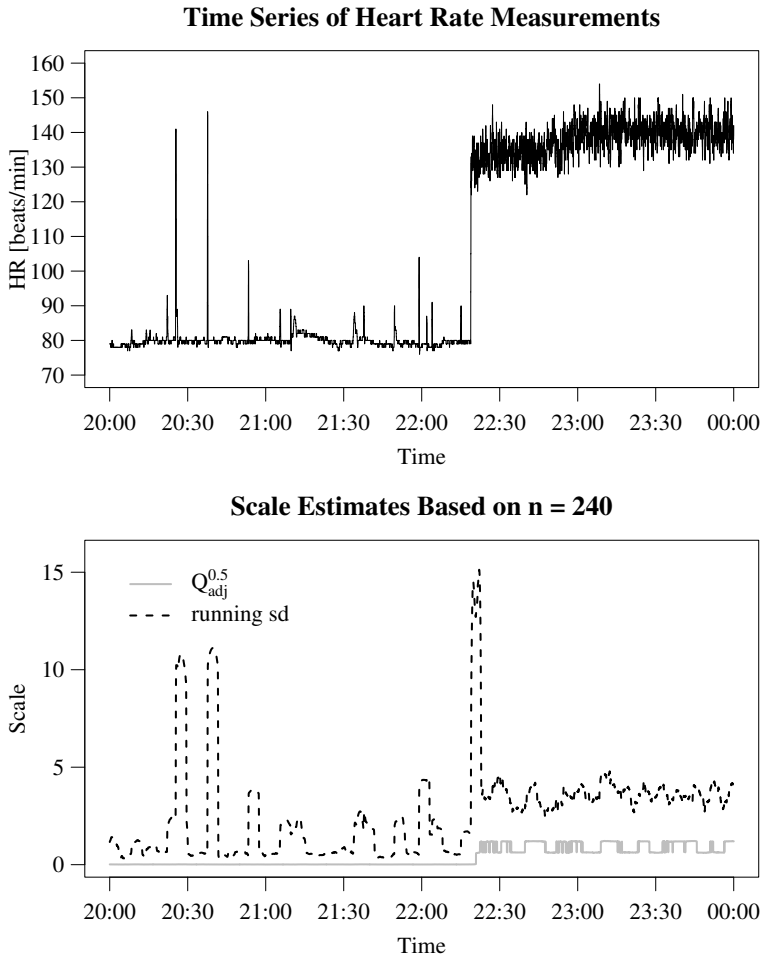

Figure 8: Physiological data (top panel). The bottom panel presents the scale as estimated by the $Q^{0.5}_{adj}$ estimator and the residual standard deviation after an OLS-fit with $n = 240$.

estimate S_t requires only $O(\log n)$ computing time, allowing for fast and easy online computation. Apart from that, 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.

The selection of the window width n is not treated in this paper. If n is chosen too large, then the assumption of a locally constant scale is not realistic anymore. Moreover, if there is a sudden jump in the variability of the series (not in the level, since the robust procedures proposed in this paper can cope with a level shift), the effect of this volatility shift will be more severe for longer window widths. On the other hand, if n is too small, the local scale estimate is based on too few observations. It then becomes less precise and more influenced by outliers. 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).

The proposed method can also be used as a tool for testing model assumptions on the volatility process σ_t. For example, one could estimate the process σ_t using a GARCH model and compare it with the proposed model-free estimates. If the difference is too large, then the GARCH model should be rejected. Similarly, a homoscedasticity assumption could be tested for. For deriving the theoretical influence functions and asymptotic efficiencies, we required local linearity and that the error terms within a single window are independent and normally distributed. A topic for future research is the study of the asymptotic properties of the estimators under less stringent conditions. Note, however, that the primary aim of the methods is to serve as a robust exploratory measure of volatility of a time series, which first of all is computable in an on-line setting. For obtaining the results on the breakdown point, we did not need any model assumptions.

Acknowledgements

We would like to thank the referees for their insightful comments and suggestions which have led to an improved version of this paper. We also gratefully acknowl-

edge the financial support of the German Science Foundation (DFG, SFB 475 "Reduction of Complexity for Multivariate Data Structures"), 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).

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 Mountains Mathematical Publications 26, 87–101.

Gasser, T., Sroka, L., Jennen-Steinmetz, C., 1986. Residual variance and residual pattern in nonlinear regression. Biometrika 73, 625–633.

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.

Gelper, S., Schettlinger, K., Croux, C., Gather, U., 2008. Robust online scale estimation in time series: A model-free approach. Technical Report, Katholieke Universiteit Leuven, www.econ.kuleuven.be/sarah.gelper/public.

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

Härdle, W., Tsybakov, A., 1988. Robust nonparameric regression with simultanious scale curve estimation. Annals of Statistics 16, 120–135.

Hotta, L., Tsay, R., 1998. Outliers in GARCH processes. Manuscript.

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.

Rue, H., Chu, C., Godtliebsen, F., Marron, J., 2002. M-smoother with local linear fit. Nonparametric Statistics 14, 155–168.

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