SFB 823

Robust repeated median regression in moving windows with data-adaptive width selection

Discussion Paper

Matthias Borowski, Roland Fried

Nr. 28/2011

Robust Repeated Median regression in moving windows with data-adaptive width selection

SCARM – Slope Comparing Adaptive Repeated Median

Matthias Borowski
Faculty of Statistics
TU Dortmund University
borowski@statistik.tu-dortmund.de

Roland Fried
Faculty of Statistics
TU Dortmund University
fried@statistik.tu-dortmund.de

August 29, 2011

Abstract

Online (also 'real-time' or 'sequential') signal extraction from noisy and outlier-interfered data streams is a basic but challenging goal. Fitting a robust Repeated Median (Siegel, 1982) regression line in a moving time window has turned out to be a promising approach (Davies et al., 2004; Gather et al., 2006; Schettlinger et al., 2006). The level of the regression line at the rightmost window position, which equates to the current time point in an online application, is then used as signal extraction. However, the choice of the window width has large impact on the signal extraction, and it is impossible to predetermine an optimal fixed window width for data streams which exhibit signal changes like level shifts and sudden trend changes. We therefore propose a robust test procedure for the online detection of such signal changes. An algorithm including the test allows for online window width adaption, meaning that the window width is chosen w.r.t. the current data situation at each time point. Comparison studies show that our new procedure outperforms an existing Repeated Median filter with automatic window width selection (Schettlinger et al., 2010).

1 Introduction

In many fields, such as intensive care online-monitoring, industrial process control or financial markets, data are measured at high sampling rates of e.g. one observation per second. The resulting time series are typically non-stationary and often exhibit a large amount of noise and outliers. The separation of the unknown underlying signal (which carries the relevant information) from noise and outliers *online*, i.e. for every new incoming observation, is a basic but challenging goal. However, standard methods for *signal extraction* or *filtering* like running means or medians are not appropriate due to the difficult data situation. A running mean is not robust against outliers and blurs level shifts (also called step changes, edges or jumps), whereas a running median improperly depicts linear trends by steps and with a delay of half a window width. For the difficult data that we are concerned with, Davies et al. (2004) and Gather et al. (2006) recommend signal

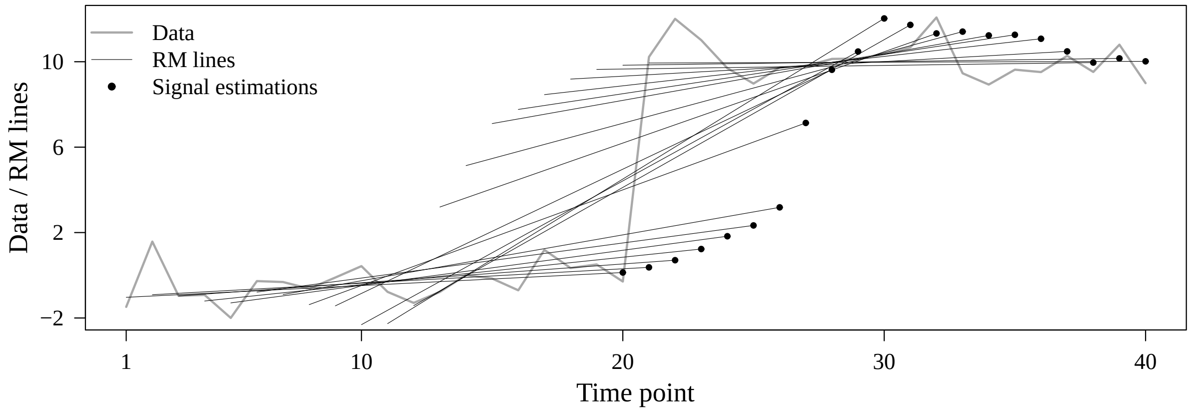

Figure 1: Principle of signal extraction by moving window RM regression: the right-end level of the regression line is used as signal estimation for the corresponding time point; here the time window contains $n = 20$ observation

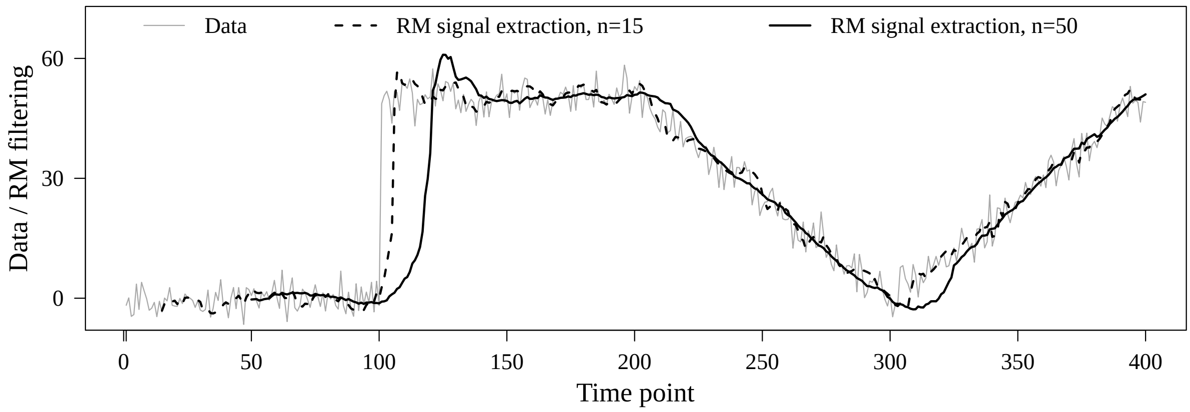

Figure 2: Trade-off for the choice of the window width: data (grey) and RM signal extraction with $n = 15$ (black, dashed) and $n = 50$ (black, solid)

filtering by robust Repeated Median (RM) regression (Siegel, 1982) in a moving time window, taking the regression level either at the central or at the rightmost window position as signal estimation. Taking the rightmost position yields signal extractions without time delay, which is preferable in most online applications, and also allows for choosing even window widths. This approach is outlined in Figure 1. The choice of the window width n has large impact on the resulting signal extraction. A large value of n extracts a smooth signal with little variability whereas a small n gives a signal close to the observations, cf. Figure 2. A predetermined fixed window width n cannot be suitable at all time points t. As long as the data show a stable trend (that can also be zero), a large n is required to obtain smooth signal estimation time series. In contrast, since level shifts and suddenly changing trends yield crucial information, the signal extraction should trace such signal changes as exactly as possible – meaning that a small n is required in these situations.
Following the idea of the *adaptive online RM* filter (aoRM, Schettlinger et al., 2010), we propose an online signal extraction procedure which adapts the window width to the current data situation at each time point t. The procedure is based on a test for detecting level shifts and sudden trend changes. Since the test compares RM slopes which are

estimated in separate time windows, we term our filtering procedure *Slope Comparing Adaptive Repeated Median*.
In the next Section we introduce the RM regression and its statistical properties and give a short outline of existent modifications and extensions of the RM, including the aoRM. In Section 3 the SCARM filter is explained in detail. The test for the detection of signal changes, an approach for the online estimation of the variance of RM slope differences, and the final filtering algorithm are presented. In Section 4 we compare the tests used by the SCARM and the aoRM w.r.t. the detection of level shifts and trend changes. Applications to simulated and real data are given in Section 5. Section 6 gives a summary and an outlook.

2 Repeated median regression for online signal extraction

We formalize the extraction of a signal (μ_t) from time series (x_t) using the components model

$$\begin{aligned} X_t &= \mu_t + \xi_t, \\ \xi_t &= \sigma_t\, \varepsilon_t + \eta_t, \quad t = 1, 2, \ldots . \end{aligned} \tag{1}$$

The data x_t are realizations of the real-valued random variables X_t, and μ_t is the underlying signal which is assumed to be smoothly varying most of the time but can exhibit sudden trend changes and level shifts. The outlier process η_t generates impulsive spiky noise which is zero most of the time but occasionally takes large absolute values. The noise process $\sigma_t\, \varepsilon_t$ has zero mean and variance σ_t^2. The noise variance σ_t^2 may change slowly over time, so that we can treat it locally as (almost) constant.

Following Schettlinger et al. (2010), we assume that the underlying signal μ_t can be locally well approximated by a regression line within a short time window $\{t-n+1, \ldots, t\}$, $t \geq n$:

$$\mu_{t-n+i} \approx \mu_t + \beta_t \cdot (i-n), \quad i = 1, \ldots, n, \tag{2}$$

where μ_{t-n+i} is the level (of the regression line) at the window position (i.e. to the time point) $t-n+i$ and β_t the slope of the regression line. Davies et al. (2004) and Gather et al. (2006) compare several regression estimators w.r.t. robustness, efficiency, and computing time, and find RM regression to provide best compromise results. Given a window sample $\mathbf{x}_t = (x_{t-n+1}, \ldots, x_t)$ of size n, the RM estimate of the slope β_t and the level μ_t in (2) is

$$\begin{aligned} \hat{\beta}_t :=& \hat{\beta}(\mathbf{x}_t) = \underset{i \in \{1,\ldots,n\}}{\text{med}} \left\{ \underset{i' \neq i,\, i' \in \{1,\ldots,n\}}{\text{med}} \frac{x_{t-n+i} - x_{t-n+i'}}{i - i'} \right\}, \\ \hat{\mu}_t :=& \hat{\mu}(\mathbf{x}_t) = \underset{i \in \{1,\ldots,n\}}{\text{med}} \left\{ x_{t-n+i} - \hat{\beta}_t \cdot (i-n) \right\}. \end{aligned} \tag{3}$$

The RM has a finite sample replacement breakdown point of $\lfloor n/2 \rfloor / n \approx 50\%$ (Rousseeuw and Leroy, 1987) which is the maximum possible value for a regression equivariant regression estimator (Davies and Gather, 2005). It yields good efficiency at non-contaminated Gaussian samples (Gather et al., 2006) and is fast to compute with an update algorithm for moving time windows that only needs linear time (Bernholt and Fried, 2003). Moreover, for data being drawn from

$$X_{t-n+i} = \mu_t + \beta_t \cdot (i-n) + \xi_{t-n+i}, \quad i = 1, \ldots, n,$$

with symmetric errors $(\xi_{t-n+1}, \dots, \xi_t)$ that have the same distribution as $(-\xi_{t-n+1}, \dots, -\xi_t)$, the RM slope $\hat{\beta}_t$ is an unbiased estimate of the true slope β_t (Siegel, 1982).
Like for any other localized signal extraction technique, there is the trade-off problem for the choice of the window width when using the RM. This trade-off problem is mainly based on the fact that signal changes, e.g. level shifts, may occur at unknown time points. Several extensions of the RM have been proposed to overcome the problem of tracing level shifts (Fried, 2004; Gather and Fried, 2004; Fried et al., 2006; Bernholt et al., 2006; Fried et al., 2007), but all apply a fixed predetermined window width n. Schettlinger et al. (2010) tackle the trade-off problem by the *adaptive online RM* (aoRM) filter which uses a residual-sign-test to adapt the window width to the current data situation at each time t. The test used by the aoRM is based on the fact that an RM regression results in an equal number of positive and negative residuals. The aoRM first fits an RM line in a time window $\{t-n+1, \dots, t\}$. Then the null hypothesis, that the median of the distribution of the $r \leq \lfloor n/2 \rfloor$ rightmost RM residual signs is zero, is tested against the alternative, that the median is not zero.
Let the residuals of an RM fit in a time window $\{t-n+1, \dots, t\}$ be denoted as

$$R_{t,i} = X_{t-n+i} - \Big(\hat{\mu}(\mathbf{X}_t) - \hat{\beta}(\mathbf{X}_t) \cdot (i-n)\Big), \quad i = 1, \dots, n,$$

where $\mathbf{X}_t = (X_{t-n+1}, \dots, X_t)$. The aoRM test statistic is the absolute sum of the $r \leq \lfloor n/2 \rfloor$ rightmost RM residual signs:

$$T_{\text{aoRM}} = \left| \sum_{i \in I} \text{sign}(R_{t,i}) \right|, \quad I = \{n-r+1, \dots, n\}, \quad \text{sign}(\omega) = \begin{cases} -1, & \text{if } \omega < 0 \\ 0, & \text{if } \omega = 0 \\ 1, & \text{if } \omega > 0 \end{cases}.$$

The test rejects the null hypothesis if $T_{\text{aoRM}} > c_\alpha(n, r)$ where the critical value $c_\alpha(n, r)$ depends on n, r and the level of significance α. For small n and r, the critical values are obtained by Monte Carlo simulations; for large n and r, Schettlinger et al. use quantiles of a hypergeometric distribution.
At each time t, the test is used to find an adequate window width n: If H_0 is rejected, n is set to $n-1$ by removing the leftmost/oldest observation from the sample, and it is tested again. This loop ends as soon as H_0 is no longer rejected. The RM regression is then performed on the window sample of the adapted width n. Our proposed Slope Comparing Adaptive Repeated Median procedure adopts the idea of the aoRM, however yielding a more powerful and faster detection of level shifts and trend changes, cf. Section 4.

3 The Slope Comparing Adaptive Repeated Median

Firstly we introduce the test used by the Slope Comparing Adaptive Repeated Median (SCARM) for the detection of level shifts and trend changes, simply speaking of signal changes in the following. The test statistic is the difference of RM slopes relative to the variance of this difference. Hence, we also propose an approach for the online estimation of the variance of RM slope differences. The entire SCARM algorithm is presented at the end of this Section.

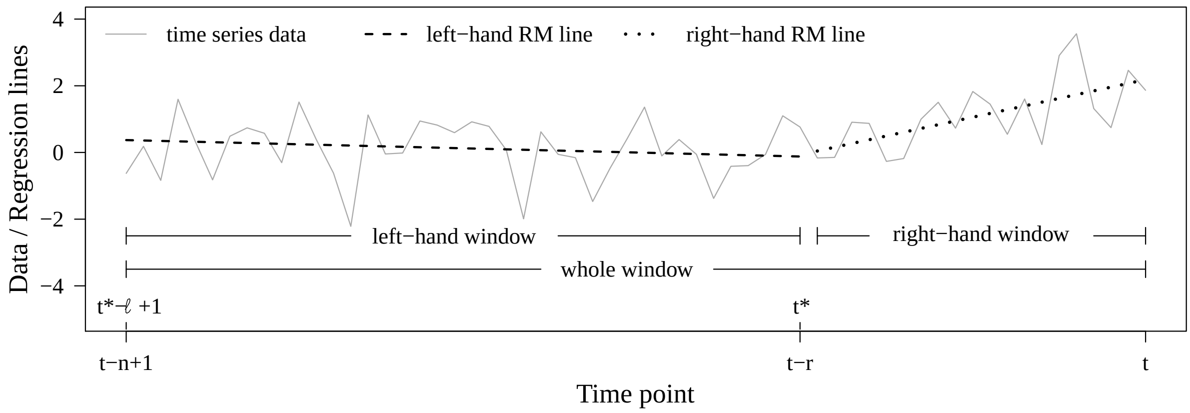

Figure 3: Separation of a window sample into a left-hand and a right-hand sample with left-hand (black dashed) and right-hand (black dotted) RM line

3.1 A test for the online-detection of signal changes

The test used by the SCARM is based on the assumption (1) that the data come from $X_t = \mu_t + \xi_t$ where μ_t is the underlying relevant signal and $\xi_t = \sigma_t\,\varepsilon_t + \eta_t$ the error process. In the theoretical development of our test we assume that $\sigma_t\,\varepsilon_t \sim N(0, \sigma_t^2)$ and $\eta_t = 0$. In Appendix B we inspect the SCARM test statistic distribution given skewed, heavy-tailed and contaminated errors by means of Monte Carlo simulations.
Following the change-point regression approach of Chen et al. (2011), we test the null hypothesis that the signal in a time window $\{t-n+1, \ldots, t\}$ is linear against the alternative that a signal change in the form of a level shift and/or trend change is present after some time point t_0:

$$\mathrm{H}_0: \ \mu_{t-n+i} = \mu + \beta \cdot (i-n), \quad i = 1, \ldots, n,$$
$$\mathrm{H}_1: \ \mu_{t-n+i} = \begin{cases} \mu_{\text{old}} + \beta_{\text{old}} \cdot (i-n), & i = 1, \ldots, t_0 \\ \mu_{\text{new}} + \beta_{\text{new}} \cdot (i-n), & i = t_0+1, \ldots, n \end{cases},$$

where $t_0 \in \{1, \ldots, n-1\}$ and $\mu_{\text{old}} \neq \mu_{\text{new}}$ and/or $\beta_{\text{old}} \neq \beta_{\text{new}}$. Due to (1), H_0 can also be written as

$$X_{t-n+i} = \mu_t + \beta_t \cdot (i-n) + \sigma_t\,\varepsilon_{t-n+i}. \tag{4}$$

We divide the *whole window* $\{t-n+1, \ldots, t\}$ of length n into two separate parts, a *right-hand window* $\{t-r+1, \ldots, t\}$ of width r and a *left-hand window* $\{t^*-\ell+1, \ldots, t^*\}$ of width ℓ where $t^* = t-r$, $t^*-\ell+1 = t-n+1$, and $n = \ell + r$ (cf. Figure 3). Denote by

$$\mathbf{x}_t = (x_{t-n+1}, \ldots, x_t) \in \mathbb{R}^n, \quad \mathbf{x}_t^{\text{right}} = (x_{t-r+1}, \ldots, x_t) \in \mathbb{R}^r,$$
$$\mathbf{x}_{t^*}^{\text{left}} = (x_{t^*-\ell+1}, \ldots, x_{t^*}) = (x_{t-n+1}, \ldots, x_{t-n+\ell}) = (x_{t-n+1}, \ldots, x_{t-r}) \in \mathbb{R}^\ell$$

the *whole*, *right-hand*, and *left-hand window sample*, and let $\mathbf{X}_t$, $\mathbf{X}_t^{\text{right}}$, and $\mathbf{X}_{t^*}^{\text{left}}$ be the corresponding random vectors. Furthermore, $\hat{\mu}(\mathbf{x}_{t^*}^{\text{left}})$ and $\hat{\beta}(\mathbf{x}_{t^*}^{\text{left}})$ are the RM level and slope which are estimated from the left-hand sample and which specify the left-hand RM line (the black dashed line in Figure 3):

$$\hat{\mu}_{t^*-\ell+j}(\mathbf{x}_{t^*}^{\text{left}}) = \hat{\mu}(\mathbf{x}_{t^*}^{\text{left}}) + \hat{\beta}(\mathbf{x}_{t^*}^{\text{left}})(j-\ell), \tag{5}$$

where $j = 1, \ldots, \ell$. The RM level and slope in the right-hand window are denoted by $\hat{\mu}(\mathbf{x}_t^{\text{right}})$ and $\hat{\beta}(\mathbf{x}_t^{\text{right}})$. They specify the right-hand RM line (the black dotted line in

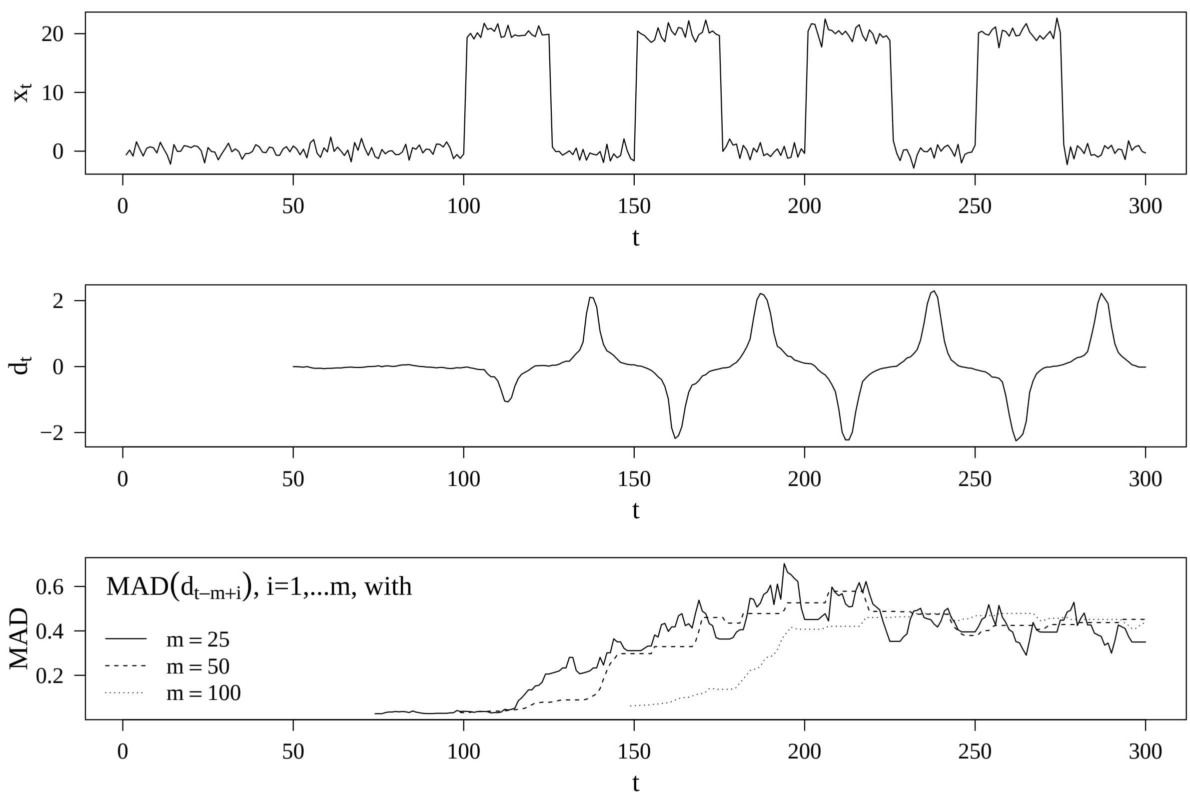

Figure 4: Top: simulated time series (x_t); center: time series (d_t) of RM slope differences with $\ell, r = 25$; bottom: MAD of the recent m slope differences d_{t-m+i}, $i = 1, \ldots, m$

Figure 3):

$$\hat{\mu}_{t-r+k}(\mathbf{x}_t^{\text{right}}) = \hat{\mu}(\mathbf{x}_t^{\text{right}}) + \hat{\beta}(\mathbf{x}_t^{\text{right}})(k - r), \tag{6}$$

where $k = 1, \ldots, r$.

The RM slope is unbiased for symmetric error distributions (Siegel, 1982). Hence, the RM slopes within the whole, left-hand and right-hand window are unbiased under this assumption if H_0 is true:

$$\text{E}(\hat{\beta}(\mathbf{X}_t)) = \text{E}(\hat{\beta}(\mathbf{X}_{t^*}^{\text{left}})) = \text{E}(\hat{\beta}(\mathbf{X}_t^{\text{right}})) = \beta_t,$$

implying that the expectation of $D_t := \hat{\beta}(\mathbf{X}_{t^*}^{\text{left}}) - \hat{\beta}(\mathbf{X}_t^{\text{right}})$ is zero. Our proposed test statistic is

$$T_t := \frac{D_t}{\sqrt{\widehat{\text{Var}}(D_t)}}, \tag{7}$$

where $\widehat{\text{Var}}(D_t)$ is an estimate of $\text{Var}(D_t)$. If $|T_t|$ is 'too large', the whole window sample $\mathbf{x}_t := (x_{t-n+1}, \ldots, x_t)$ is assumed not to come from (4), and H_0 is rejected. Since the RM reacts to trend changes *and* level shifts, the test is sensitive to both kinds of signal changes.

An intuitive approach to estimate $\text{Var}(D_t)$ would be to apply a robust scale estimator like the *Median Absolute Deviation* (MAD) to the latest m realizations of RM slope differences $d_{t-m+1}, \ldots, d_t$. However, this approach is problematic since signal changes induce large absolute values d_t, as can be seen in Figure 4. The top graph shows a time series (x_t) with several level shifts; the central graph shows the RM slope differences (d_t) with

$\ell, r = 25$; and the lower graph shows the MAD at time t, computed on the m slope differences $d_{t-m+1}, \ldots, d_t$ where $m = 25$ (solid line), $m = 50$ (dashed line) and $m = 100$ (dotted line). As can be seen, shortly after an upward (downward) shift in the time series (x_t), the time series (d_t) of RM slope differences show a downward (upward) peak. Thus, when several level shifts in the data occur close to each other, there are several peaks in the time series (d_t), meaning a large amount of variability. Hence, even a robust scale estimator like the MAD yields large estimates of $\text{Var}(D_t)$, see graph (c) of Figure 4. That is, since level shifts induce large estimates of $\text{Var}(D_t)$, they can 'mask' subsequent level shifts, meaning reduced power of the test. We therefore propose an alternative approach to estimate $\text{Var}(D_t)$ in the following subsection. Using this approach we obtain $\widehat{\text{Var}}(D_t)$ and reject H_0 if

$$T_t = \frac{D_t}{\sqrt{\widehat{\text{Var}}(D_t)}} < q_{\text{low}} \quad \text{or} \quad T_t = \frac{D_t}{\sqrt{\widehat{\text{Var}}(D_t)}} > q_{\text{up}},$$

where the critical values q_{low} and q_{up} are the $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of T_t. Monte Carlo simulations in Appendix B indicate that the distribution of T_t under H_0 can be well approximated by a t_f-distribution – even for skewed, heavy-tailed, and contaminated errors. Hence, we reject H_0 if $|T_t| > t_{f,1-\alpha/2}$. The degrees of freedom $f = f(\ell, r)$ depend on ℓ and r. Appendix B describes how we determine $f(\ell, r)$.

3.2 The estimation of $\text{Var}(D_t)$

We propose a method to estimate $\text{Var}(D_t)$ under H_0 with $\xi_{t-n+i} = \sigma_t \varepsilon_{t-n+i} \sim N(0, \sigma_t^2)$ i.i.d. within the whole window $\{t-n+1, \ldots, t\}$. The effects of violations of these assumptions because of outliers, for instance, are discussed below. If H_0 is true, we have

$$\begin{aligned} \text{Var}(D_t) &= \text{Var}\left(\hat{\beta}_{t^*}(\mathbf{X}_{t^*}^{\text{left}}) - \hat{\beta}_t(\mathbf{X}_t^{\text{right}})\right) \\ &= \text{Var}\left(\hat{\beta}_{t^*}(\mathbf{X}_{t^*}^{\text{left}})\right) + \text{Var}\left(\hat{\beta}_t(\mathbf{X}_t^{\text{right}})\right), \end{aligned} \tag{8}$$

since (i) we assume (ε_t) to form independent Gaussian noise, (ii) the left-hand and the right-hand samples are separate and (iii) $\text{Var}\left(\hat{\beta}_{t^*}(\mathbf{X}_{t^*}^{\text{left}})\right)$ and $\text{Var}\left(\hat{\beta}_t(\mathbf{X}_t^{\text{right}})\right)$ do not depend on the true β_t:

Theorem 1. *If H_0 is true, the variance of $\hat{\beta}(\mathbf{X}_t)$ does not depend on the true β_t. It is equivariant to the noise variance σ_t^2 and depends on the window width n.*

Proof. The RM slope estimate $\hat{\beta}(\mathbf{X}_t)$ can be written as

$$\begin{aligned} \hat{\beta}(\mathbf{X}_t) &= \underset{i \in \{1,\ldots,n\}}{\text{med}} \left\{ \underset{i' \neq i,\, i' \in \{1,\ldots,n\}}{\text{med}} \frac{X_{t-n+i} - X_{t-n+i'}}{i - i'} \right\} \\ &= \beta_t + \underset{i \in \{1,\ldots,n\}}{\text{med}} \left\{ \underset{i' \neq i,\, i' \in \{1,\ldots,n\}}{\text{med}} \frac{\sigma_t \varepsilon_{t-n+i} - \sigma_t \varepsilon_{t-n+i'}}{i - i'} \right\} \end{aligned}$$

since

$$X_{t-n+i} = \mu_t + \beta_t(i-n) + \xi_{t-n+i} \quad \text{and} \quad X_{t-n+i'} = \mu_t + \beta_t(i'-n) + \xi_{t-n+i'}.$$

Hence, it is

$$\mathrm{Var}(\hat{\beta}(\mathbf{X}_t)) = \sigma_t^2 \cdot \mathrm{Var}\left(\underset{i \in \{1,\ldots,n\}}{\mathrm{med}} \left\{ \underset{i' \neq i,\, i' \in \{1,\ldots,n\}}{\mathrm{med}} \frac{\varepsilon_{t-n+i} - \varepsilon_{t-n+i'}}{i - i'} \right\} \right), \tag{9}$$

where $\mathrm{Var}(\varepsilon_{t-n+i}) = 1$. □

Furthermore, the data within the left-hand and the right-hand window have the same linear structure if H_0 is true. From this it follows:

Corollary 1. *If H_0 is true,* $\mathrm{Var}(\hat{\beta}(\mathbf{X}_{t^*}^{\mathrm{left}}))$ *and* $\mathrm{Var}(\hat{\beta}(\mathbf{X}_t^{\mathrm{right}}))$ *do not depend on the true slope β_t. They are equivariant to the noise variance σ_t^2 and depend on the respective window width ℓ and r.*

Therefore, we regard the variance of the left-hand and the right-hand RM slope as a function $\mathrm{V} : \mathbb{N} \times \mathbb{R}^+ \longmapsto \mathbb{R}^+$ with

$$\mathrm{V}(\ell, \sigma_t) := \mathrm{Var}\left(\hat{\beta}_{t^*}(\mathbf{X}_{t^*}^{\mathrm{left}})\right) \quad \text{and} \quad \mathrm{V}(r, \sigma_t) := \mathrm{Var}\left(\hat{\beta}_t(\mathbf{X}_t^{\mathrm{right}})\right). \tag{10}$$

In order to approximate $\mathrm{V}(\ell, \sigma_t)$ and $\mathrm{V}(r, \sigma_t)$ for any $\ell, r \geq 5$ and any noise variance $\sigma_t^2 > 0$, we use the fact that

$$\mathrm{V}(\ell, \sigma_t) = \mathrm{V}(\ell, 1) \cdot \sigma_t^2 \quad \text{and} \quad \mathrm{V}(r, \sigma_t) = \mathrm{V}(r, 1) \cdot \sigma_t^2, \tag{11}$$

cf. (9). Combining (8) – (11) yields, under H_0,

$$\mathrm{Var}(D_t) = \sigma_t^2 \cdot (\mathrm{V}(\ell, 1) + \mathrm{V}(r, 1)). \tag{12}$$

By means of Monte Carlo simulations we obtain approximations $\hat{\mathrm{V}}(\ell, 1) =: v_\ell$ and $\hat{\mathrm{V}}(r, 1) =: v_r$ for $\ell, r \geq 5$, see Appendix A. The estimation of σ_t^2 could be carried out on the residuals of an RM regression on the whole window sample. However, violations of the null hypothesis due to signal changes lead to large absolute residuals. Hence, even robust scale estimators could deliver inadequately large estimations of σ_t^2, meaning that signal changes can 'mask' themselves or subsequent signal changes, as explained above.
In order to prevent such masking effects, we estimate σ_t directly on the data using a scale estimator proposed by Rousseeuw and Hubert (1996) and Gelper et al. (2009). It estimates σ_t by the heights of adjacent triangles, formed by triples of consecutive observations. Thus, it does not depend on any previous estimate of the regression parameters μ_t and β_t, is invariant to linear trends, and performs well for level shifts, time varying scale and outliers. Given a sample vector $\mathbf{x} := (x_1, \ldots, x_n)$, this scale estimator is

$$Q_\delta(\mathbf{x}) = c_n \cdot h_{(\lfloor \delta(n-2) \rfloor)},$$

where $h_{(\lfloor \delta(n-2) \rfloor)}$ is the δ-quantile of the $n-2$ adjacent triangle heights

$$h_w = \left| x_{w+1} - \frac{x_w + x_{w+2}}{2} \right|, \quad w = 1, \ldots, n-2,$$

which are formed by all triples of successive observations x_w, x_{w+1}, x_{w+2}. The constant c_n denotes a factor to achieve unbiasedness for samples of size n drawn from a normal distribution. The choice of δ is complicated by a trade-off problem between robustness and efficiency. We act on the suggestion of Gelper et al. (2009) and set $\delta = 0.5$ to achieve reasonable robustness and efficiency. Using the $Q_{0.5}$ scale estimator and the approximations v_ℓ and v_r, we estimate $\mathrm{Var}(D_t)$ according to (12), i.e. $\widehat{\mathrm{Var}}(d_t) = Q_{0.5}(\mathbf{x}_t)^2 \cdot (v_\ell + v_r)$. Note that even though the assumption of an outlier-free error process ($\eta_{t-n+i} = 0$) is made in the theoretical development of our approach for the estimation of $\mathrm{Var}(D_t)$, the robustness of the Q_δ scale estimator ensures adequate estimates when outliers are present. Furthermore, the Q_δ – and thus our proposed approach for the estimation of $\mathrm{Var}(D_t)$ – is robust against signal changes in the time window (i.e. H_1 is true). We therefore prevent signal changes from masking themselves or following breaks. However, the Q scale estimator may become zero if ties are present in the data. We therefore propose to set a positive lower bound for the Q scale estimation to prevent that the realization of the SCARM test statistic is infinite due to ties in the data:

$$\widehat{\mathrm{Var}}(d_t) = \max\{b,\, Q_{0.5}(\mathbf{x}_t)^2 \cdot (v_\ell + v_r)\}. \tag{13}$$

The choice of the bound b depends on the application and can have large influence on the resulting SCARM signal extraction, cf. Section 5.

3.3 The filtering algorithm

Next, we propose a moving window algorithm that uses the proposed test to find an adequate window width at each time point t. Hence, we denote the time-dependent window width as n_t instead of n in the following. We modify the algorithm of Schettlinger et al. (2010), particularly by replacing the test with our proposed test procedure. Our algorithm requires the following input arguments which must be chosen by the analyst in consideration of the given data:

- the level of significance α of the test,
- the fixed width of the right-hand window r,
- a minimum bound $\ell_{\min}$ for the left-hand window width ℓ_t which ensures that $\ell_t \geq \ell_{\min}$ (the left-hand window width is not fixed but varies over time since $\ell_t := n_t - r$), and
- a minimum and maximum window width $n_{\min}$ and $n_{\max}$, so that $n_t \in \{n_{\min}, \ldots, n_{\max}\}$ and $n_{\min} \leq \ell_{\min} + r$.

The choice of the input arguments is discussed at the end of this Section.
The flow chart in Figure 5 illustrates the SCARM algorithm. The core of the algorithm is the test of whether n_t is adequate for an RM regression. If H_0 cannot be rejected, the signal is estimated on the window sample of size n_t. However, if the test rejects H_0, n_t is set to its minimum value $n_{\min}$, since we assume a signal change within the time window. (This adaption principle is different from that of the aoRM. If the aoRM rejects the null hypothesis, n_t is set to $n_t - 1$ and it is tested again. This is repeated until n_t is adequate.) Ideally, $n_{\min}$ is such that the time window $\{t - n_{\min} + 1, \ldots, t\}$ solely includes observations after the signal change; sensible choices for $n_{\min}$ are discussed below. After the signal is estimated on the sample $(x_{t-n_t+1}, \ldots, x_t)$, the window is updated for the next time point $t + 1$ by incorporating the new incoming observation x_{t+1} into the window sample. Provided that H_0 is not rejected, the window grows gradually with each incoming new observation, until n_t equals the maximum possible value $n_{\max}$. In this case

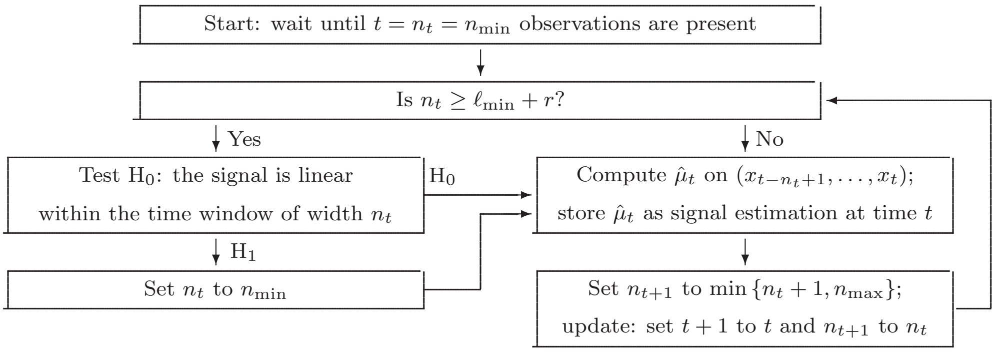

Figure 5: A flow chart of the SCARM algorithm.

the window does not grow anymore, but moves instead to the right.
It is important to understand that (i) the window width can increase only 'bit by bit' but decreases at once to $n_{\min}$ when the test detects a signal change, (ii) the window width can only reach its maximum value $n_{\max}$ if the test does not reject H_0 for an accordant period of time, and (iii) when the algorithm sets the window width n_t to $n_{\min}$ and $n_{\min}$ is chosen smaller than $\ell_{\min} + r$, the window width n_t cannot be decreased before it has grown again up to $n_t = \ell_{\min} + r$. This rule ensures that there are enough observations for a meaningful testing.

The level of significance α of the test defines the expected number of falsely detected signal changes. For example, $\alpha = 0.1\%$ means that one expects one false detection in 1000 time points.
The choice of the right-hand width r is crucial. It defines the cutoff between a patch of outliers and a signal change. For instance, if less than 10 subsequent aberrant observations are treated as a patch of outliers, r should be greater than 20 because of the 50% breakdown point of the RM. However, actually the RM estimation is already considerably biased if the outlier proportion is around 1/3. This can be seen in Figure 1, where the RM line (window width $n = 20$) is affected by the level shift when the time window contains around six or seven level-shifted observations. Hence, we recommend to choose r three times larger than the length of the largest expected outlier patch.
On the one hand r determines the maximum length of outlier-patches the filtering procedure can resist, on the other hand r determines the time needed to detect a signal change, which is approximately $r/3$. That is, if the test detects a signal change, it is justifiable to assume that this break happened around time $t - r/3$. Hence, we recommend to choose $n_{\min} = r/3$.
The minimum bound $\ell_{\min}$ for the left-hand window width ℓ_t ensures that the test is only performed if the left-hand and right-hand window contain 'enough' observations: After n_t has been decreased to the value $n_{\min}$, it must grow again for the following time points until $n_t = \ell_{\min} + r$. Similar to the choice of r, $\ell_{\min}$ must be chosen w.r.t. the requested amount of robustness: in order that the left-hand window can resist as many outliers as the right-hand window, we recommend to choose $\ell_{\min} = r$. Finally, the maximum width $n_{\max}$ ensures that $n_t \leq n_{\max}$. It limits the time needed for the computation of the RM.

4 A comparison of the aoRM and SCARM test

We compare the tests used by the aoRM and SCARM w.r.t. their power to detect level shifts and trend changes as well as the mean time needed for the detection of such signal changes.

4.1 Detection of level shifts and trend changes

The comparison is done by simulations using several different input values, in particular $(n, r) = (40, 20)$, $(60, 20)$, $(80, 40)$, $(120, 40)$ and the significance levels $\alpha = 0.001$, 0.005, 0.01 and 0.1. The critical values for the SCARM and aoRM test statistics are obtained by simulations, applying the tests to Gaussian noise.
For each (n, r)-combination we generate time series (x_t) and (y_t), $t = 1, \ldots, n + r$, where a level shift and trend change, respectively, is present at time $t = n + 1$. The time series (x_t) are given by

$$x_t = \begin{cases} \varepsilon_t, & t = 1, \ldots, n \\ a + \varepsilon_t, & t = n+1, \ldots, n+r \end{cases},$$

where $\varepsilon_t \sim N(0, 1)$ and a is the height of the level shift. The time series (y_t) are given by

$$y_t = \begin{cases} \varepsilon_t, & t = 1, \ldots, n \\ b \cdot (t - n) + \varepsilon_t, & t = n+1, \ldots, n+r \end{cases},$$

where b is the slope of the trend. We generate 1000 time series (x_t) and (y_t) for each (n, r)-combination and for each $a = 1, \ldots, 4$ and $b = 0.1, \ldots, 0.4$. We then apply the aoRM and SCARM test with fixed n and r in a moving time window of width n. Thus, both tests deliver test decisions for the time points $t = n, \ldots, n + r$. If at least one of these $r + 1$ tests decides to reject H_0, the signal change is considered as detected by the test procedure. Figure 6 shows the resulting rates of detected level shifts (top) and trend changes (bottom) for $(n, r) = (40, 20)$; the results for $(n, r) = (120, 40)$ are shown in Figure 7. In Figure 6, the SCARM test offers distinctly higher detection rates than the aoRM test. For small levels of significance α, the aoRM test has small power to detect signal changes, in particular trend changes. Although the performance of the aoRM becomes better for higher levels of significance, it is worse than the SCARM test for each α-value. However, in an online application, a high significance level induces a high rate of false detections, meaning that the window width is often decreased unnecessarily. These results are obtained using $(n, r) = (40, 20)$, i.e. $\ell = r = 20$ since $\ell = n - r$. Further simulations using other (n, r)-values show that the SCARM offers a considerably higher power than the aoRM when r is small and/or when the difference between ℓ and r is small. For larger (ℓ, r)-values, as in Figure 7, the difference between the detection rates of the SCARM and aoRM becomes less distinct. However, the SCARM still yields higher power when the significance level and the signal change are small.
In contrast to the aoRM test, the test of the SCARM offers high detection rates for small ℓ- and r-values. This is a crucial advantage of the SCARM over the aoRM, because a smaller r induces that signal changes are detected earlier. Since detection time is crucial, we have also allocated the time spans from the signal change at time point $t = n + 1$ to the time points when the changes have been detected.
Figure 8 shows the *mean detection times*, i.e. the mean of the allocated time spans, of the aoRM and SCARM test for level shifts (top) and trend changes (bottom). Since the

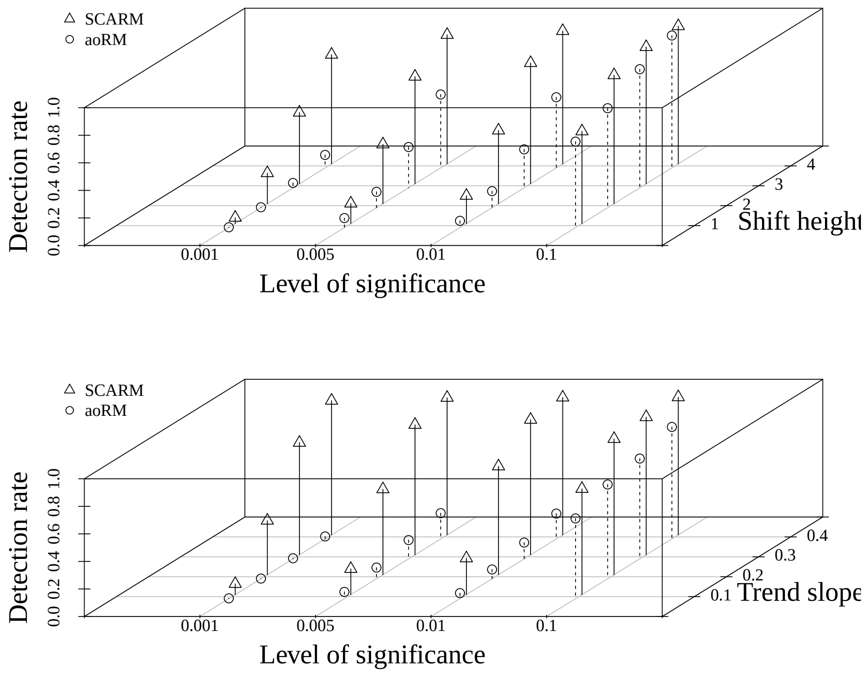

Figure 6: Rate of detected level shifts (top) and trend changes (bottom) with $(n, r) = (40, 20)$ or $(\ell, r) = (20, 20)$, respectively

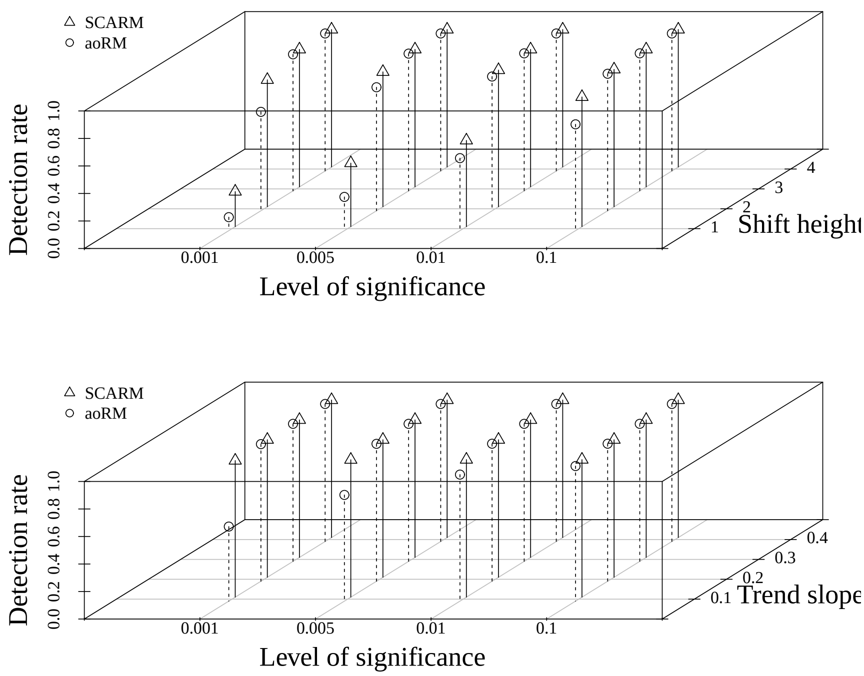

Figure 7: Rate of detected level shifts (top) and trend changes (bottom) with $(n, r) = (120, 40)$ or $(\ell, r) = (80, 40)$, respectively

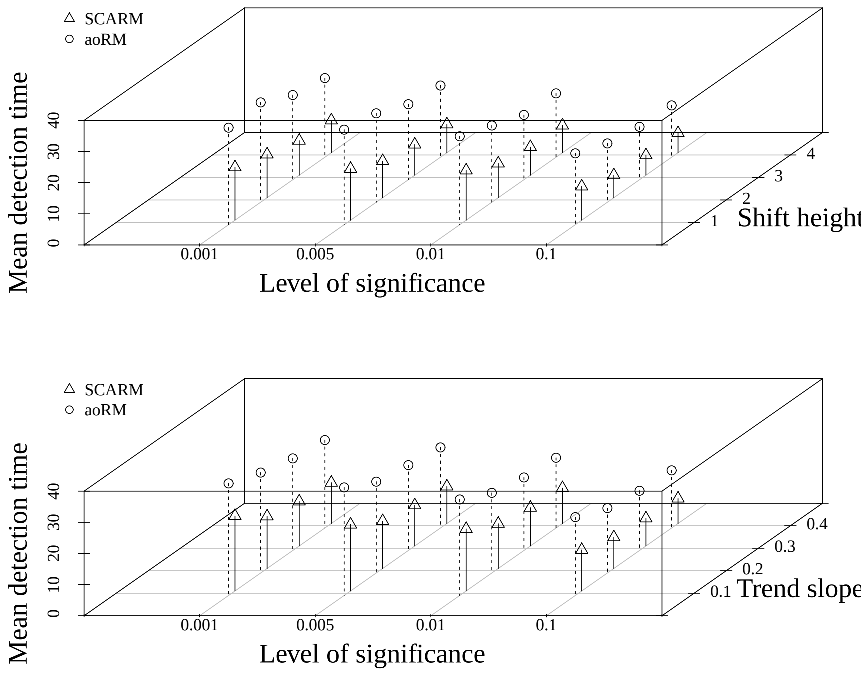

Figure 8: Mean detection times of the aoRM and SCARM for level shifts (top) and trend changes (bottom) with $(n, r) = (120, 40)$ or $(\ell, r) = (80, 40)$, respectively

results are similar for all considered (n, r)-values, we only show the detection times for $(n, r) = (120, 40)$ or $(\ell, r) = (80, 40)$, respectively. Both procedures tend to react faster when the level of significance and the magnitude of the signal change are large. However, the SCARM yields faster reaction times than the aoRM in any of the situations considered here. Roughly spoken, the SCARM offers a mean detection time of approximately $\frac{1}{4}r$ to $\frac{1}{2}r$, whereas the aoRM needs approximately $\frac{1}{2}r$ to $\frac{2}{3}r$ time points on average to detect signal changes.

5 Application

The SCARM and aoRM are provided in the R package *robfilter* 3.0 (Fried et al., 2011), which offers several time series filtering methods based on robust concepts. The functions are termed *scarm.filter* and *adore.filter* (*ad*aptive *o*nline *re*peated median). The input arguments *right.width*, *min.left.width*, *min.width*, and *max.width* of the *scarm.filter* function match the input values r, $\ell_{\min}$, $n_{\min}$, and $n_{\max}$. The main input arguments of the *adore.filter* are the number of RM residuals for the test statistic, cf. Section 2, and a minimum and maximum window width, matching the input arguments *p.test*, *min.width* and *max.width*. We apply the *scarm.filter* and *adore.filter* function retrospectively to simulated data and to real data from intensive care online-monitoring.

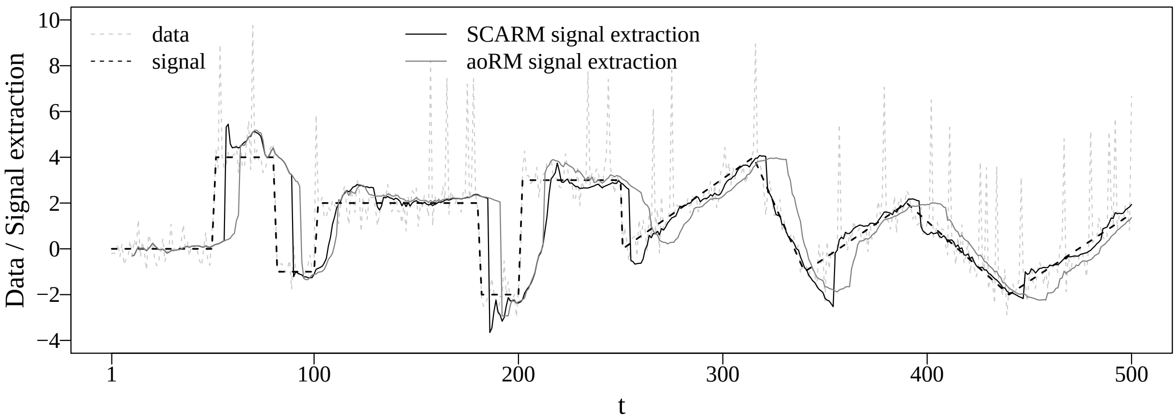

Figure 9: Simulated time series (grey, dashed) with underlying signal (black, dashed) and signal extraction by SCARM (black, solid) and aoRM (grey, solid)

5.1 Simulated data

We create a signal which exhibits several level shifts and trend changes and add normal noise and 5% outliers of value 5 at randomly chosen positions. For both the *scarm.filter* and *adore.filter* we choose the input arguments $min.width = 11$ and $max.width = 120$. Furthermore, we choose $p.test = right.width = 22$, so that the *scarm.filter* and the *adore.filter* use an equal number of observations for the test decisions. Since the *adore.filter* function uses the fixed significance level $\alpha = 0.1$, we also choose this relatively high α-value for the *scarm.filter* function. Figure 9 shows the simulated data with the underlying signal and the signal extraction time series.
The SCARM signal extractions are apparently closer to the underlying signal. Both filters trace the signal changes with a certain time delay, since the signal is estimated robustly at the right end of each time window. However, the SCARM delivers signal extractions that react faster to changes. These impressions support the findings as to the mean detection times from Section 4. However, the SCARM and aoRM window width adaption principles are different. The aoRM searches for a window width n_t which is suitable *and as large as possible*, whereas the SCARM sets the window width down to its minimum value when it detects a signal change. Hence, the SCARM traces signal changes more accurately than the aoRM.

5.2 Online-monitoring data from intensive care

The time series from intensive care online-monitoring are systolic blood pressure measurements of a patient, measured at a frequency of one observation per second. After consulting an experienced intensive care physician, aberrant data patches that are shorter than 10 observations can be regarded as clinically irrelevant outlier patches. Hence, for the *scarm.filter* we choose the input arguments $right.width = min.left.width = 30$ and $min.width = right.width/3 = 10$, and we set $max.width = 180$. Furthermore, we choose the significance level $sign.level = 0.001$. Since the blood pressure measurements are integer valued, ties occur frequently. Therefore, H_0 would be often rejected due to zero noise scale estimations and thus infinite values of the test statistic. The input argument *bound.noise.sd* of the *scarm.filter* function matches the input argument b, the lower bound for the noise scale estimation in (13). This bound avoids rejections of H_0 due to ties in

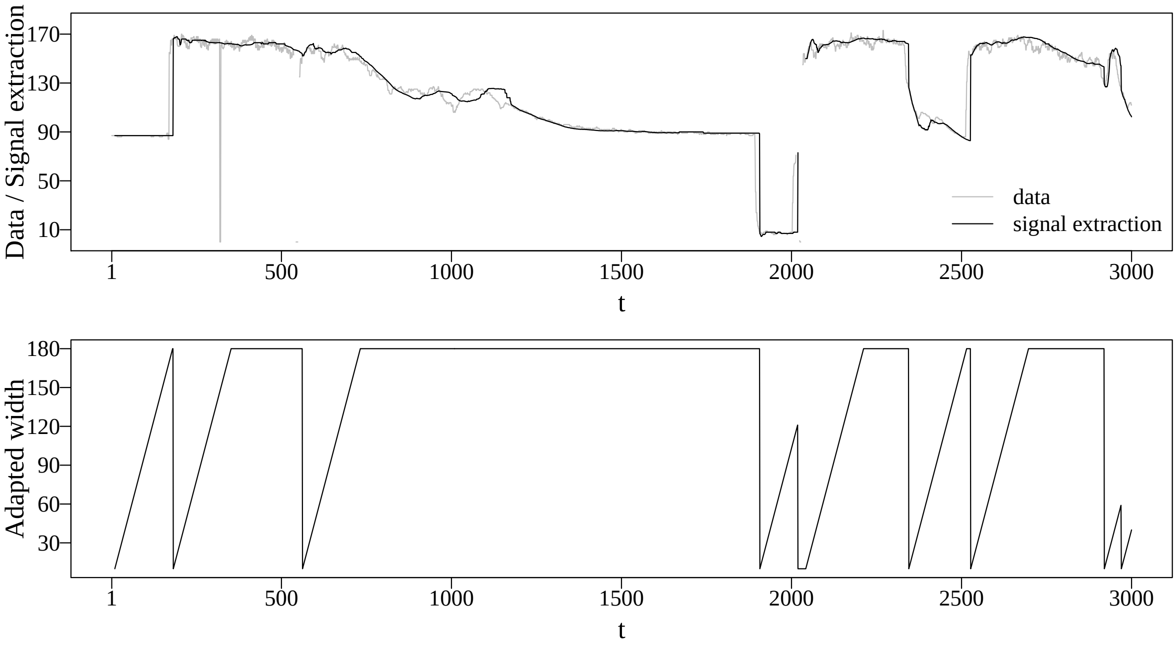

Figure 10: Top: time series of systolic artery blood pressure and SCARM signal extraction; bottom: adapted window widths

the data. Moreover, the blood pressure time series are strongly positively autocorrelated. Hence, ordinary fluctuations very often cause falsely detected signal changes. If a smooth signal extraction is requested, the user can simply choose a large value of *bound.noise.sd* to prevent false detections. We applied the *scarm.filter* to the real data using several different values of *bound.noise.sd*, aiming at a signal extraction which is as smooth as possible without missing any signal change, and obtained good results for *bound.noise.sd* = 10, cf. Figure 10 (a). The signal extraction is smooth in phases of a stable trend, whereas level shifts and changing trends are traced fast and exactly. Although we choose a large value of *bound.noise.sd* here, the SCARM test has still enough power to detect the conspicuous signal changes. This example shows that the input argument *bound.noise.sd* can also be used as a tuning parameter that has large influence on the signal extraction. It is also visible how the *scarm.filter* handles the missing values around time $t = 2000$. In order to obtain reliable signal estimations, the *scarm.filter* function requires at least $\lceil right.width/2 \rceil$ non-missing observations in the right-hand and $\lceil min.left.width/2 \rceil$ in the left-hand window. Otherwise, the signal estimation output is a missing value. That is, when subsequent observations are missing, the *scarm.filter* first continues the signal extraction, but stops when too many values are missing. The signal estimation outputs are then missing values. When enough new non-missing observations are given again, the *scarm.filter* re-starts the signal extraction.
The time series of adapted widths n_t in Figure 10 (b) give information about the test decisions taken by the SCARM test for any time point t. When H_0 is rejected, i.e. when the SCARM detects a signal change, the width n_t is set to its minimum. Hence, we can evaluate whether the test decisions are correct or not by comparing the time series (n_t) with the data and signal extraction time series. In our opinion all rejections excepting the second rejection of H_0 (around time $t = 550$) are necessary and therefore correct.

6 Summary and outlook

Robust RM regression in a moving time window is a promising approach for online signal extraction from noisy and outlier-contaminated non-stationary data stream time series. However, it is impossible to predetermine an optimal window width as the data structure changes over time. The SCARM tackles this problem by following the idea of the aoRM (Schettlinger et al., 2010) of choosing the window width automatically at each time point t w.r.t. the current data situation. Both procedures use a test to decide whether or not the width must be adapted. The test statistic of the SCARM is the absolute difference of RM slopes, computed in separate time windows, divided by the variance of the slope difference. In order to avoid masking effects, we propose a sophisticated approach to estimate the variance. The test procedure including the variance estimation is the core of the SCARM signal extraction algorithm.

Since the signal extractions strongly depend on the chosen window width, the performance of the test is crucial. In a simulation study, we find that the SCARM outperforms the aoRM w.r.t. the power and the mean time needed for detecting level shifts and sudden trend changes. Applications to simulated data, using the R-functions in the package *robfilter*, further illustrate the advantages of the SCARM over the aoRM, and applications to real data from intensive care demonstrate the capability of the SCARM in practice. However, these applications also show that the test of the SCARM has a distinctly increased type I error rate in the case of positive autocorrelations. Choosing a bound for the noise variance estimation (13) may fix this problem, but at the cost of decreased power for detecting signal changes. There is apparently still some need for research as to the effect of autocorrelations on the SCARM test.

The SCARM and aoRM can also be used to extract signals from multivariate time series, simply by applying the filters to each univariate component of the multivariate time series. However, the dependence structure of the data is not regarded then. Borowski et al. (2009) develop a multivariate extension of the aoRM that takes cross-correlations into account in order to improve the efficiency of the signal estimation and the robustness against multivariate outliers. A similar multivariate extension of the SCARM would probably be beneficial.

Acknowledgement

This work has been supported by the Collaborative Research Center 'Statistical modelling of nonlinear dynamic processes' (SFB 823, Projects C1 and C3) of the German Research Foundation (DFG).

References

Bernholt, T. and Fried, R. (2003). Computing the update of the repeated median regression line in linear time. *Information Processing Letters*, 88(3):111–117.

Bernholt, T., Fried, R., Gather, U., and Wegener, I. (2006). Modified repeated median filters. *Statistics and Computing*, 16:177–192.

Borowski, M., Schettlinger, K., and Gather, U. (2009). Multivariate real time signal processing by a robust adaptive regression filter. *Communications in Statistics – Simulation and Computation*, 38(2):426–440.

Chen, C., Chan, J., Gerlach, R., and Hsieh, W. (2011). A comparison of estimators for regression models with change points. *Statistics and Computing*, 21(3):395–414.

Davies, P., Fried, R., and Gather, U. (2004). Robust signal extraction for on-line monitoring data. *Journal of Statistical Planning and Inference, Special Issue: Contemporary Data Analysis: Theory and Methods in Honor of John W. Tukey*, 122:65–78.

Davies, P. and Gather, U. (2005). Breakdown and groups. *The Annals of Statistics*, 33(3):977–1035.

Fried, R. (2004). Robust filtering of time series with trends. *Nonparametric Statistics*, 16(3-4):313–328.

Fried, R., Bernholt, T., and Gather, U. (2006). Repeated median and hybrid filters. *Computational Statistics & Data Analysis*, 50(9):2313 – 2338.

Fried, R., Einbeck, J., and Gather, U. (2007). Weighted repeated median smoothing and filtering. *Journal of the American Statistical Association*, 102:1300–1308.

Fried, R., Schettlinger, K., and Borowski, M. (2011). *robfilter: Robust Time Series Filters.* R package version 3.0.

Gather, U. and Fried, R. (2004). Methods and algorithms for robust filtering. In *Proceedings in Computational Statistics (COMPSTAT 2004)*, pages 159–170. Physica, Heidelberg.

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

Gelper, S., Schettlinger, K., Croux, C., and Gather, U. (2009). Robust online scale estimation in time series: A model-free approach. *Journal of Statistical Planning and Inference*, 139(2):335 – 349.

Rousseeuw, P. and Hubert, M. (1996). Regression-free and robust estimation of scale for bivariate data. *Computational Statistics & Data Analysis*, 21(1):67–85.

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

Schettlinger, K., Fried, R., and Gather, U. (2006). Robust filters for intensive care monitoring: beyond the running median. *Biomedical Engineering*, 51:49–56.

Schettlinger, K., Fried, R., and Gather, U. (2010). Real time signal processing by adaptive repeated median filters. *International Journal of Adaptive Control and Signal Processing*, 24:346–362.

Siegel, A. (1982). Robust regression using repeated medians. *Biometrika*, 69:242–244.

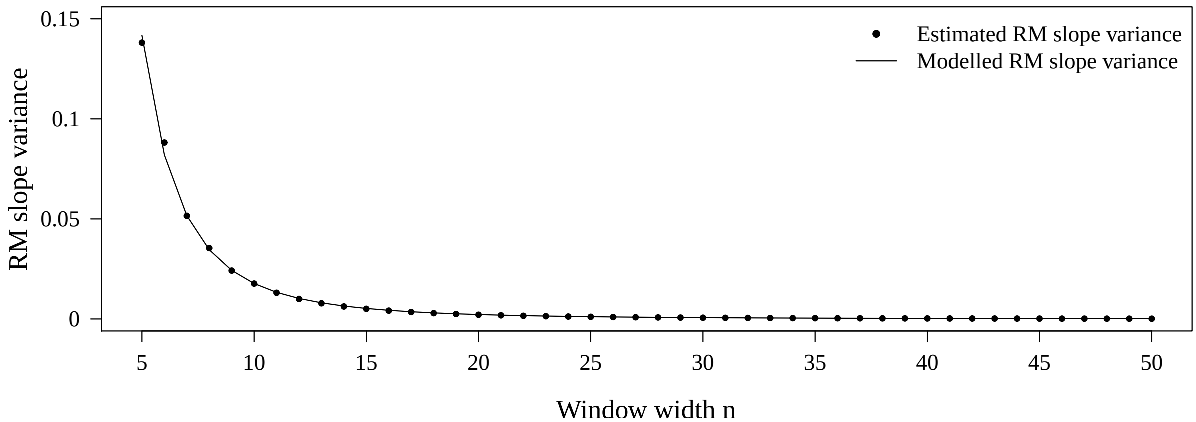

Figure 11: Empirical variance of 100000 RM slopes estimated on standard normal samples of size $n = 5, \ldots, 50$

A Monte Carlo approximations of $\hat{\mathbf{V}}(\ell, 1)$ and $\hat{\mathbf{V}}(r, 1)$

We approximate $v_n := \hat{\mathrm{V}}(n, 1)$ for $n = 5, \ldots, 300$ by the empirical variance of RM slopes which are estimated on samples coming from

$$X_{t-n+i} = \mu_t + \beta_t \cdot (i - n) + \varepsilon_{t-n+i}, \qquad i = 1, \ldots, n,$$

with standard normal i.i.d. errors $\varepsilon_{t-n+i} \sim N(0, 1)$. W.l.o.g. we set $\mu_t = \beta_t = 0$ because of the regression equivariance of the RM slope (Rousseeuw and Leroy, 1987), i.e. $X_{t-n+i} = \varepsilon_{t-n+i}$.

We generate time series (x_t) consisting of $100000 + 300 - 1 = 100299$ observations. Then for each $n = 5, \ldots, 300$ we move a time window $\{t - n + 1 \ldots, t\}$ over the time series, starting at time point $t = 300$. Hence, for each n we obtain 100000 RM slopes, and v_n is the empirical variance computed on these 100000 RM slopes. Due to this time series design the RM slopes are autocorrelated, as they are in practice. However, in another simulation study we also approximated the RM slope variance for independent samples. These estimates are comparable to those obtained by the time series design. As was to be expected, the variance of the RM slope decreases monotonically with increasing window size n, see Figure 11. In order to obtain approximations v_n for $n > 300$, we model the relationship between n and v_n and find that the function

$$v(n) = 4.77 \cdot 10^{-7} + 17.71 \cdot n^{-3}$$

is an appropriate model with standard error 0.0004 and coefficient of determination 0.9983.

B The empirical distribution of the SCARM test statistic

This Monte Carlo study analyzes the distribution of the SCARM test statistic T_t under the null hypothesis. That is, we compute T_t on samples $\mathbf{x}_t$ that come from model (4):

$$\begin{aligned} X_{t-n+i} &= \mu_t + \beta_t \cdot (i - n) + \xi_{t-n+i}, \\ \xi_{t-n+i} &= \sigma_t \, \varepsilon_{t-n+i} + \eta_{t-n+i}, \quad i = 1, \ldots, n, \end{aligned}$$

left-hand width ℓ	right-hand width r: 5	10	15	20	25	30	35	40	45	50
5	3.3	–	–	–	–	–	–	–	–	–
10	4.7	6.2	–	–	–	–	–	–	–	–
15	6.9	7.7	10.9	–	–	–	–	–	–	–
20	8.0	9.1	12.3	14.8	–	–	–	–	–	–
25	10.2	14.2	15.8	16.5	19.1	–	–	–	–	–
30	11.8	12.6	16.1	20.1	20.5	20.7	–	–	–	–
35	12.0	18.2	18.7	18.1	29.4	27.3	24.8	–	–	–
40	14.8	15.6	16.4	23.7	22.3	24.9	31.9	21.7	–	–
45	14.7	16.7	23.6	25.8	21.2	38.1	26.9	25.1	38.2	–
50	20.5	26.7	19.9	20.0	31.9	28.5	24.8	51.1	30.6	41.9

Table 1: Approximation of the SCARM test statistic distribution by a t-distribution: suitable degrees of freedom $f(\ell, r)$

where we set $\mu_t = \beta_t = 0$ w.l.o.g. due to the regression equivariance of the RM slope. First of all we consider standard normal errors, as assumed in the theoretical development of the SCARM test, i.e. $\eta_{t-n+i} = 0$ and $\sigma_t \, \varepsilon_{t-n+i} \sim N(0, \sigma_t^2)$, and w.l.o.g. we set $\sigma_t = 1$. We generated 10000 samples of length $n = \ell + r$ for $r \in \{5, 10, \ldots, 100\}$ and $\ell \in \{r, r + 5, \ldots, 100\}$. Thus, for any combination ℓ, r we obtained 10000 realizations of the SCARM test statistic T_t.

We find that the distribution of the SCARM test statistic can be well approximated by a t-distribution with f degrees of freedom, where f depends on ℓ and r. For each combination (ℓ, r), we compare the empirical α- and $(1 - \alpha)$-quantiles, $\alpha = 0.01, 0.02, \ldots, 0.05$, of the SCARM test statistic to the corresponding theoretical quantiles of a t-distribution with degrees of freedom $f = 0.1, 0.2, \ldots, 100$, in order to find a suitable f for each combination (ℓ, r). For each (ℓ, r)-combination, we choose that f that minimizes the mean absolute difference between the empirical and the theoretical quantiles. Table 1 lists the suitable degrees of freedom $f(\ell, r)$ for $r \in \{5, 10, \ldots, 50\}$ and $\ell \in \{r, r + 5, \ldots, 50\}$. The degrees of freedom f, and thus the quantiles $t_{f,\alpha/2}$ and $t_{f,1-\alpha/2}$, are expected to be monotonically increasing in ℓ and r. However, this is not true for the approximations of $f(\ell, r)$ in Table 1. Therefore, for $r \in \{5, \ldots, 100\}$ and $\ell \in \{r, \ldots, 100\}$ we set

$$f(\ell, r) = \min_{\ell' \geq \ell} \min_{r' \geq r} \{f(\ell', r')\}, \tag{14}$$

with $r' \in \{5, 10, \ldots, 50\}$ and $\ell' \in \{r', r' + 5, \ldots, 50\}$ to achieve monotonic degrees of freedom f and thus monotonic critical values $t_{f,\alpha/2}$ and $t_{f,1-\alpha/2}$. By taking the minimum in (14) we decide for larger absolute critical values in order that the test keeps the chosen level of significance α. If ℓ or r is larger than 100, we use standard normal quantiles as critical values.

B.1 Other error types

We further investigate the distribution of T_t for heavy-tailed, skewed, and contaminated errors ξ_{t-n+i}, in particular:

- Noise type 1: heavy-tailed errors from a standardized t-distribution with three degrees of freedom;

ℓ	r	0.995-	0.9975- quantile	0.9995-	quantile type
		3.66	4.25	5.83	$t_{f(10,10)=6.2}$
		3.12	3.21	3.29	empirical, noise type 1
10	10	3.91	4.88	6.36	empirical, noise type 2
		2.81	3.00	3.99	empirical, noise type 3
		3.03	3.13	5.10	empirical, noise type 4
		2.84	3.14	3.83	$t_{f(50,50)=20.7}$
		2.43	2.63	3.28	empirical, noise type 1
50	50	2.68	2.78	3.31	empirical, noise type 2
		2.17	2.22	2.85	empirical, noise type 3
		2.52	2.75	3.17	empirical, noise type 4
		2.75	3.02	3.64	$t_{f(100,20)=30.7}$
		2.43	2.74	3.28	empirical, noise type 1
100	20	2.80	2.92	3.05	empirical, noise type 2
		2.36	2.69	3.53	empirical, noise type 3
		2.32	2.42	2.81	empirical, noise type 4

Table 2: Quantiles of a $t_{f(\ell,r)}$ distribution and empirical quantiles of the SCARM test statistic computed on different types of noise

- Noise type 2: skewed errors from a standardized Weibull distribution with scale and shape parameter two and one;
- Noise type 3: standard normal errors with 10% contamination from $N(10, 1)$;
- Noise type 4: standard normal errors with 10% contamination from $N(0, \sigma_t^2 = 100)$.

Table 2 gives the empirical $(1 - \alpha/2)$-quantiles, $\alpha = 0.01, 0.005, 0.001$, of the computed SCARM test statistics for the four noise types and for different combinations (ℓ, r). Furthermore, the table lists the quantiles of the $t_{f(\ell,r)}$-distribution which are used as critical values for test decision. The empirical quantiles are generally lower than the t_f-quantiles that are used for test decision, except for $(\ell, r) = (10, 10)$ and given the skewed noise type 2. That is, the test keeps the chosen level of significance, even if the noise is heavy-tailed or contaminated. However, if the noise is skewed, ℓ and r should both not be too small.