Fakultät Statistik

technische universität
dortmund

# Signal and Variability Extraction for Online Monitoring in Intensive Care

## Dissertation

zur Erlangung des akademischen Grades
'Doktor der Naturwissenschaften'
– Dr. rer. nat. –
der Technischen Universität Dortmund

Der Fakultät Statistik
der Technischen Universität Dortmund
vorgelegt von

**Karen Schettlinger**

geboren in
Neustadt am Rübenberge

Dortmund 2009

Promotionsausschuss

| | |
|---|---|
| 1. Gutachter: | Prof. Dr. Ursula Gather |
| 2. Gutachter: | Prof. Dr. Roland Fried |
| Vorsitzende der Prüfungskommission: | Prof. Dr. Katja Ickstadt |

Tag der mündlichen Prüfung

05. März 2009

## Abstract

This thesis proposes new methods for real-time signal and variability extraction, presents derivations of their robustness properties and discusses their value for practical applications to physiological time series. Although the proposed techniques are developed against the background of online monitoring in intensive care, they are also applicable to any other kind of time series.

For Repeated Median regression on an equidistant grid, the distribution of the position and number of zero residuals is investigated, and the correlation structure between the residual signs is examined.

For online signal extraction, an adaptive filter is proposed which essentially relies on a goodness-of-fit test based on residual signs from Repeated Median regression. After deriving suitable settings for this filter in the univariate case from a simulation study, the procedure is extended for application to multivariate time series.

For online variability extraction, three approaches to scale estimation are considered. The robustness properties of the newly proposed regression-free and model-free techniques are derived, and the different approaches are compared via an extensive simulation study.

# Table of Contents

# 1 Introduction

Monitoring systems in intensive care are important tools for diagnosing and judging the state of the critically ill. These systems assess and display measurements of physiological variables in real-time. As decision support for the medical staff, they also contain alarm systems and specific diagnosis tools. Currently used alarm systems for certain hemodynamic variables are essentially based on simple threshold rules where the violation of an upper or lower limit causes an alarm.

Physiological time series from an intensive care monitoring system often contain spikes due to technical problems (e.g. defective contacts or loose connections) or sudden movements of the patient (e.g. caused by coughing). Such outliers contain no information about the true state of the patient but may influence statistical analyses severely, see e.g. Charbonnier, Becq and Biot (2004). Furthermore, they cause many false alarms which increases the stress level for the clinical staff, see e.g. Lawless (1994).

Of course, it also may happen that outliers point at a relevant change of the patient's health: for example, arrhythmias may appear as spikes in the heart rate series. However, such events are controlled by additional monitoring rules, e.g. making use of ECG recordings.

Extracting the relevant information from the multitude of observations, e.g. by applying online filters to the time series and comparing the noise- and artefact-free level of the series to the alarm thresholds provides a possibility to reduce the number of false alarms (Kuhls, 2008). Further improvement can be expected when incorporating the variability of a time series into the alarm rules, e.g. by combining level and variability estimates. For certain variables, such as the heart rate, the variability itself might be a feature which is worth monitoring since it provides important information about the patient's health status and might even be used for early warnings.

In intensive care, methods are required which are computationally fast and also comprehensible and interpretable for the medical staff. Therefore, the techniques described in the following chapters are developed from an application-oriented background to guarantee their suitability for real-time application in intensive care. In particular, we focus on filtering techniques based on a moving time window. For extraction of the underlying level of a time series we focus on robust regression-based filters, in particular, on filters which use the Repeated Median regression. For online variability extraction we also consider regression-based techniques and compare them with methods which do not require this regression step.

Chapter 2 introduces the Repeated Median (or short: RM) regression defined by Siegel (1982) and states some important properties. For regression performed on an equidistant grid, some characteristics of the RM slope are investigated in Section 2.1, and Section 2.2 provides a detailed investigation of the RM residuals and their signs. The correlation structure of the RM residual signs is investigated in Section 2.3 and compared to the well-known correlation structure of Least Squares residuals. These attributes of the RM residuals are explored particularly in view of the adaptive signal extraction procedure, introduced in the following chapter.

Chapter 3 covers different approaches for signal extraction from high frequency time series. Section 3.1 provides an overview of some robust versions from the broad variety of univariate filtering methods that exhibit certain characteristics desirable in the online monitoring context in intensive care. This section particularly emphasises the advantages of filters based on Repeated Median regression, but the described filters are all based on moving time windows of fixed width. In Section 3.2 a univariate filtering procedure based on RM regression is described which chooses the window width adaptively, depending on the underlying data structure. Section 3.3 extends this approach to a filter for multivariate time series, taking into account the possible correlation structure between the different components of the series. All sections include exemplary applications to hemodynamic time series for proving their usefulness for real-time applications in intensive care.

Chapter 4 addresses the robust online extraction of the variability of the error term fluctuating around the true signal of a univariate time series. Here, three different types of scale estimators are considered: Section 4.1 describes a regression-based approach to scale estimation where the variability of the error term is judged by the residuals of a robust regression fit; the regression-free scale estimators introduced in Section 4.2 do not require a preceding regression step, but they rely on a local linear model; the approach described in Section 4.3 does not need this assumption and hence, we call the resulting scale estimators 'model-free'. We compare these three different approaches by means of an extensive simulation study in Section 4.4; Section 4.5 shows applications to a simulated and a physiological time series, and Section 4.6 concludes the results from this chapter.

Finally, Chapter 5 embraces the preceding chapters by summarising their most important findings. Furthermore, it provides a debate on the benefits of the investigated methods for current intensive care alarm systems and discusses possible combinations of the introduced online signal and variability extraction techniques.

## 2 Repeated Median Regression

Let $\mathbf{Y}_n = (Y_1, \ldots, Y_n)'$ denote a sample of size $n$ evaluated at equidistant design points $i = 1, \ldots, n$ and consider a simple linear model with noise terms $\varepsilon_i$:

$$Y_i \;=\; \mu + \beta \cdot i + \varepsilon_i \quad \text{for } i = 1, \ldots, n. \tag{2.1}$$

According to Siegel (1982), the Repeated Median (RM) regression estimates the slope and the (hierarchical) intercept at design point $i = 0$ by

$$\hat{\beta}^{RM} \;=\; \underset{i=1,\ldots,n}{\text{med}} \left\{ \underset{j \neq i}{\text{med}} \; \frac{Y_i - Y_j}{i - j} \right\} \tag{2.2}$$

$$\hat{\mu}^{RM} \;=\; \underset{i=1,\ldots,n}{\text{med}} \left\{ Y_i - \hat{\beta}^{RM} \cdot i \right\}. \tag{2.3}$$

Here, the median $\text{med}\{\cdot\}$ at an even sample size $n$ is defined as the arithmetic mean of the $(n/2)$th and $(n/2+1)$st order statistic. Of course, the regression estimators by Siegel (1982) are defined for arbitrary designs, but since here it is of interest to apply regression methods to (samples from) time series measured at equidistant time points, we will focus on an equidistant design in the following.

As an alternative to (2.3), Siegel (1982) proposes to estimate the RM intercept by

$$\hat{\mu}_1^{RM} \;=\; \underset{i=1,\ldots,n}{\text{med}} \left\{ \underset{j \neq i}{\text{med}} \; \frac{j\,Y_i - i\,Y_j}{j - i} \right\}. \tag{2.4}$$

Both estimators, (2.3) and (2.4), possess the same bounded influence function and asymptotic efficiency at standard normal data. However, the hierarchical intercept (2.3) has a lower computation time (Hössjer, Rousseeuw and Ruts, 1995) and thus, we will only consider this estimator. For definitions of the breakdown point, the influence function, efficiency, and equivariance properties see Appendix A.

In the simple bivariate linear model with fixed regressor variables and a symmetric error distribution, the RM regression estimator $(\hat{\mu}^{RM}, \hat{\beta}^{RM})'$ is unbiased and Fisher consistent (Siegel, 1982). It has a finite sample breakdown point of $\lfloor n/2 \rfloor / n$ if the data are in general position, meaning that the regression estimation still yields 'sensible' results when almost 50% of the data are contaminated by arbitrarily deviating values (Rousseeuw and Leroy, 1987). Since data from intensive care are very likely to be strongly contaminated by measurement artefacts, this *robustness* property is indeed required in this context.

Furthermore, the RM estimator is regression and scale equivariant. It is also equivariant w.r.t. affine transformations of the response variable but not w.r.t. affine transformations of the explanatory variable, see e.g. Section 2.7 of Schettlinger (2004). However, this is not a disadvantage here, since we only consider fixed equidistant times as design points.

A straightforward implementation of the RM estimator requires $\mathcal{O}(n^2)$ computation time (Siegel, 1982), but faster algorithms have been proposed by Rousseeuw, Netanyahu and Mount (1993) and Matoušek, Mount and Netanyahu (1998) with an expected computation time of $\mathcal{O}(n \log^2 n)$ or $\mathcal{O}(n \log n)$, respectively. For application to a moving time window, Bernholt and Fried (2003) have developed an update algorithm which only needs $\mathcal{O}(n)$ time. This is an important prerequisite for the real-time application to intensive care data.

Compared to least squares, the RM intercept estimator has a finite sample efficiency of almost 70% at standard normal data and even higher efficiencies at heavy tailed or skewed data (Gather, Schettlinger and Fried, 2006); the asymptotic efficiency at normal data amounts to 63.7% (Hössjer, Rousseeuw and Ruts, 1995). This means that it possesses a lower variance than many other robust estimators, causing the RM filter to result in smoother signal approximations than other robust regression filters for time series (see Section 3.1).

Hössjer, Rousseeuw and Croux (1994) show that the slope estimator $\hat{\beta}^{RM}$ is asymptotically normal and that its influence function is bounded if both, the distribution of the response variable and that of the explanatory variable, are continuous. Simulations with standard normal data show that this slope estimator has a finite sample efficiency of up to 61% while its asymptotic efficiency is only 40.5%. Thus, the RM slope estimator provides better estimations for small samples (in terms of a smaller variance) than the asymptotics suggest.

Furthermore, the RM estimator $(\hat{\mu}^{RM}, \hat{\beta}^{RM})'$ as defined by (2.3) and (2.2) possesses the *exact fit property*:

> When at least $n - \lfloor n/2 \rfloor + 1$ of the $n$ observations are collinear,
> the RM regression line runs exactly through these observations. (2.5)

Such a situation frequently occurs within the data we have in mind, because of the discrete measurement scale of the physiological time series extracted from an intensive care online monitoring system.

Hössjer, Rousseeuw and Ruts (1995), p. 62, point out that estimating the RM location hierarchically by the trend-corrected observations like in (2.3) causes the median of the residuals to be zero if both, the independent and the dependent variable, follow a continuous distribution. For even sample sizes $n$, this implies the 'balance of residual signs' which is important for the procedure described in Section 3.2. This means

$$\sum_{i=1}^{n} \operatorname{sign}(r_i) = 0\,, \tag{2.6}$$

where the RM residuals are given by

$$r_i = Y_i - \left(\hat{\mu}^{RM} + \hat{\beta}^{RM} i\right), \quad i = 1, \ldots, n, \tag{2.7}$$

and the sign function is defined as

$$\operatorname{sign}(r) = \begin{cases} -1, & \text{if} \quad r < 0 \\ 0, & \text{if} \quad r = 0 \\ 1, & \text{if} \quad r > 0 \end{cases}. \tag{2.8}$$

Equation (2.6) implies that the RM regression line separates a data cloud into halves: For an even sample size $n$, the RM line runs through either two observations or no observation at all, for an odd sample size $n$ the RM line leads through exactly one observation.
For odd sample sizes, suitable conditions are necessary for (2.6) to hold which will be addressed in the following. Furthermore, the condition of a continuous carrier distribution is not fulfilled for the data at hand because we consider equidistant points in time. A continuous distribution of the response generates a sample which is in general position with probability one, but this is not sufficient for (2.6) to be valid.
Since the balance of the residual signs (2.6) is a starting point for the procedure proposed in Section 3.2, it will be shown in the following that the probability for (2.6) not being fulfilled is negligible under certain conditions. The validity of this property is closely related with the number and position of residuals with a value of zero within the sample which will be investigated in Section 2.2. Previous to that, Section 2.1 explores some properties of the RM slope (2.2) which are essential for the remainder. Section 2.3 finishes with an investigation of the dependence structure of the signs of RM residuals providing possible explanations for certain facts found in Section 2.2.

## 2.1 The Repeated Median Slope

For the ease of notation, denote for any $i \neq j \in \{1, \ldots, n\}$ an *observational slope* by

$$\mathrm{b}(i, j) = \frac{Y_i - Y_j}{i - j}, \tag{2.9}$$

and denote the inner median of the RM slope by

$$\widetilde{\mathrm{b}}(i) := \underset{j \neq i}{\operatorname{med}} \; \{\mathrm{b}(i, j)\}, \tag{2.10}$$

such that the definition of the RM slope (2.2) can be expressed as

$$\hat{\beta}^{RM} = \underset{i=1,\ldots,n}{\operatorname{med}} \left\{ \underset{j \neq i}{\operatorname{med}} \; \mathrm{b}(i, j) \right\} = \underset{i=1,\ldots,n}{\operatorname{med}} \left\{ \widetilde{\mathrm{b}}(i) \right\}.$$

The RM regression line can only lead through more than one observation if the RM slope $\hat{\beta}^{RM}$ is equal to one of the slopes between a pair of observations $\mathrm{b}(i,j)$ (2.9). If $n$ is odd the inner median (2.10) is evaluated over an even number of slopes because every set $\{\mathrm{b}(i,j)\}$ with $j \neq i$ contains $n-1$ slopes; if $n$ is even, the outer median in (2.2) is calculated from the set $\{\widetilde{\mathrm{b}}(1),\ldots,\widetilde{\mathrm{b}}(n)\}$, containing $n$ elements. This means that for the RM slope at least one of the repeated medians is calculated from a set containing an even number of elements and hence, this median corresponds to a mean of two elements. However, it can still happen that $\hat{\beta}^{RM} = \mathrm{b}(i,j)$ for some $i,j \in \{1,\ldots,n\}$ – even for data in general position. An explanation for this fact is given below.
For the difference between $Y_i$ and $Y_k \in \mathbf{Y}_n = (Y_1,\ldots,Y_n)'$ let

$$\Delta_{i,k} := \Delta_{i,k}(\mathbf{Y}_n) = Y_k - Y_i \quad \text{for } i < k. \tag{2.11}$$

Because of the design with equidistant units of size one for the explanatory variable in the time series context here, the slope between two subsequent observations corresponds to their first difference, i.e.

$$\mathrm{b}(i,i+1) \;=\; \frac{Y_{i+1} - Y_i}{i+1-i} \;=\; \Delta_{i,i+1}\,.$$

Thus, every slope $\mathrm{b}(i,k)$ between $Y_i$ and $Y_k$ with $i < k \in \{1,\ldots,n\}$ can be calculated by an arithmetic mean of the first differences of intermediate observations:

$$\mathrm{b}(i,k) = \frac{1}{k-i}\sum_{h=i}^{k-1}\mathrm{b}(h,h+1) = \frac{1}{k-i}\sum_{h=i}^{k-1}\Delta_{h,h+1}\,.$$

In general, the slope $\mathrm{b}(i,k)$ can be rewritten as a weighted mean of slopes calculated from $Y_i$, $Y_k$ and $J \leq k-i-1$ intermediate observations with indices $i < j_1 < j_2 < \ldots < j_J < k$.

$$\begin{aligned}
\mathrm{b}(i,k) &= \frac{1}{k-i}\sum_{h=i}^{k-1}\Delta_{h,h+1} \\
&= \frac{j_1-i}{k-i}\cdot\frac{1}{j_1-i}\sum_{h=i}^{j_1-1}\Delta_{h,h+1} + \frac{j_2-j_1}{k-i}\cdot\frac{1}{j_2-j_1}\sum_{h=j_1}^{j_2-1}\Delta_{h,h+1} \\
&\quad + \ldots + \frac{k-j_J}{k-i}\cdot\frac{1}{k-j_J}\sum_{h=j_J}^{k-1}\Delta_{h,h+1} \\
&= \frac{j_1-i}{k-i}\cdot\mathrm{b}(i,j_1) + \frac{j_2-j_1}{k-i}\cdot\mathrm{b}(j_1,j_2) + \ldots + \frac{k-j_J}{k-i}\cdot\mathrm{b}(j_J,k)\,.
\end{aligned} \tag{2.12}$$

For this weighted mean, every weight has a value in $(0,1)$, and the sum of weights corresponds to $\frac{j_1-i}{k-i} + \frac{j_2-j_1}{k-i} + \ldots + \frac{k-j_J}{k-i} = 1$.

Accordingly, the slope between $Y_i$ and $Y_{i+2m}$ for $m \in \mathbb{N}$ and $i, i+2m \in \{1,\dots,n\}$, can be expressed as the median of the two observational slopes with the observation centred between $Y_i$ and $Y_{i+2m}$:

$$\begin{aligned} \mathrm{b}(i,i+2m) &= \frac{Y_{i+2m}-Y_i}{2m} = \frac{1}{2}\left[\frac{Y_{i+2m}-Y_{i+m}}{m} + \frac{Y_{i+m}-Y_i}{m}\right] \\ &= \frac{1}{2}\Big[\mathrm{b}(i,i+m) + \mathrm{b}(i+m,i+2m)\Big] \\ &= \mathrm{med}\Big\{\mathrm{b}(i,i+m),\ \mathrm{b}(i+m,i+2m)\Big\}. \end{aligned} \tag{2.13}$$

Thus, both – the inner and the outer median in (2.2) – can correspond to an observational slope (2.9), although at least one of these medians is evaluated on a set containing an even number of elements and thus, corresponds to an arithmetic mean of two slopes. In return, this allows for the RM slope to match an observational slope.
For example, consider a sample with even sample size $n$ where $\widetilde{\mathrm{b}}(i+m) = \mathrm{b}(i,i+m)$ and $\widetilde{\mathrm{b}}(i+2m) = \mathrm{b}(i+m,i+2m)$ for specific $i \in \{1,\dots,n-2m\}$ and $m \in \{1,\dots n/2-1\}$. Now, if $\widetilde{\mathrm{b}}(i+m)$ and $\widetilde{\mathrm{b}}(i+2m)$ are the two central values in the set of ordered $\widetilde{\mathrm{b}}(i)$, $i = 1,\dots,n$, then, according to (2.13), the RM slope corresponds to an observational slope:

$$\hat{\beta}^{RM} = \underset{i=1,\dots,n}{\mathrm{med}} \left\{ \widetilde{\mathrm{b}}(i) \right\} = \mathrm{med}\{\widetilde{\mathrm{b}}(i+m), \widetilde{\mathrm{b}}(i+2m)\} = \mathrm{b}(i,i+2m)\,.$$

Similarly, consider for a sample of odd size the median of the set $\{\widetilde{\mathrm{b}}(i)\ ;\ i = 1,\dots,n\}$ as $\hat{\beta}^{RM} = \widetilde{\mathrm{b}}(j+m)$, $j \in \{1,\dots,n-2m\}$ and $m \in \{1,\dots,(n-1)/2\}$. Now, if $\widetilde{\mathrm{b}}(j+m)$ corresponds to the median of the two observational slopes $\mathrm{b}(j,j+m)$ and $\mathrm{b}(j+m,j+2m)$, according to (2.13), the RM slope matches an observational slope:

$$\hat{\beta}^{RM} = \widetilde{\mathrm{b}}(j+m) = \mathrm{med}\{\mathrm{b}(j,j+m), \mathrm{b}(j+m,j+2m)\} = \mathrm{b}(j,j+2m)\,.$$

Furthermore, even for data in general position, it is possible that in the set $\{\widetilde{\mathrm{b}}(1),\dots,\widetilde{\mathrm{b}}(n)\}$ values appear repeatedly, such that for an even sample size $n$, the median of this set is possibly calculated by the mean of two observational slopes with the same value. Thus, depending on the intercept estimation, it is possible that the RM line leads through several observations.

## 2.2 Zero Residuals in Repeated Median Regression

In this section, the focus lies on residuals from an RM regression which equal a value of zero, or short: *zero residuals*. We consider independent standard normal variables such that we can assume the data to be in general position (with probability one). To evaluate the probability with which the balance of residual signs (2.6) is not true, we investigate the number and position of zero residuals within samples of size $n = 3$, $n = 4$ and $n = 5$, respectively.

Characterising a sample by the order of first differences, there are $(n-1)!$ different cases to investigate for a sample of size $n$. Half of the cases need not be investigated since the observations of these cases can be seen as a reflection of the observations from the other half of the cases along the $y$-axis. Thus, each case has a corresponding 'mirror case' for which the results can be achieved analogously. However, for a unique identification of the RM slope the order of first differences alone is often not sufficient and the cases need to be divided into subcases depending on further conditions. For a sample of size $n = 6$, $\frac{(n-1)!}{2} = 60$ cases possibly need to be divided into subcases, and in each (sub-)case the evaluation of RM slope and intercept have to take place separately. Therefore, we apply simulations for sample sizes larger than five.
An extreme case where (2.6) can not be true is given by an exact fit situation (2.5) where at least $k = n - \lfloor n/2 \rfloor + 1$ residuals equal the value zero and the signs of the remaining residuals are arbitrary – they can even be identical (Rousseeuw and Leroy, 1987).

### SAMPLE SIZE $n = 3$

For $n = 3$ observations in general position equation (2.6) is never fulfilled: for a sample $(Y_1, Y_2, Y_3)'$ the RM regression line always runs through the observations at $i = 1$ and $i = 3$, and the proof of this can be found in Appendix B. The sign of the central residual determines whether the sum of residual signs is equal to plus or minus one.

### SAMPLE SIZE $n = 4$

For a sample size of $n = 4$, there are twelve possibilities of data arrangements which uniquely define the RM slope and intercept, see Table 2.1. However, six of them correspond to a simple reflection of the data points along the $y$-axis and hence, result in expressions for the estimates analogous to the other six cases. The derivation of these cases and the calculations of the corresponding probabilities can be found in Appendix B.
Table 2.2 shows that for $n = 4$ observations in general position, only five different expressions are possible for the RM slope $\hat{\beta}^{RM}$. For the cases II(2), III(1), IV(2) and V(1) the slope estimate $\hat{\beta}^{RM}$ does not correspond to a slope between a pair of observations, i.e. the residuals of the corresponding regression line are all different from zero. In all other cases $\hat{\beta}^{RM}$ corresponds to an observational slope, namely either $\mathrm{b}(1,2)$, $\mathrm{b}(2,3)$ or $\mathrm{b}(3,4)$. For data situations where it is $\hat{\beta}^{RM} = \Delta_{2,3} = \mathrm{b}(2,3)$ (case I(1), I(2), VI(1) and VI(2)), the RM regression line runs through the observations at $i = 2$ and $i = 3$ only in case of an exact fit situation. This is the case if at least one of the remaining observational slopes $\mathrm{b}(1,2)$ or $\mathrm{b}(3,4)$ corresponds to $\mathrm{b}(2,3)$. Otherwise, the RM line separates the data cloud into the sets $\{Y_1, Y_4\}$ with corresponding residuals $r_1, r_4 > 0$ (or $r_1, r_4 < 0$, respectively) and $\{Y_2, Y_3\}$ with $r_2, r_3 < 0$ (or $r_2, r_3 > 0$, respectively), see Table 2.3.

| case | condition | additional condition | case differentiation |
|---|---|---|---|
| **I** | $(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4})$ | $\wedge\ (\Delta_{1,2}+\Delta_{3,4} \leq 2\Delta_{2,3})$ | **I (1)** |
| | | $\wedge\ (\Delta_{1,2}+\Delta_{3,4} \geq 2\Delta_{2,3})$ | **I (2)** |
| **II** | $(\Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{2,3})$ | $\wedge\ (\Delta_{2,3}+\Delta_{3,4} \leq 2\Delta_{1,2})$ | **II (1)** |
| | | $\wedge\ (\Delta_{2,3}+\Delta_{3,4} \geq 2\Delta_{1,2})$ | **II (2)** |
| **III** | $(\Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{1,2})$ | $\wedge\ (\Delta_{1,2}+\Delta_{2,3} \leq 2\Delta_{3,4})$ | **III (1)** |
| | | $\wedge\ (\Delta_{1,2}+\Delta_{2,3} \geq 2\Delta_{3,4})$ | **III (2)** |
| **IV** | $(\Delta_{1,2} \leq \Delta_{3,4} \leq \Delta_{2,3})$ | $\wedge\ (\Delta_{1,2}+\Delta_{2,3} \leq 2\Delta_{3,4})$ | **IV (1)** |
| | | $\wedge\ (\Delta_{1,2}+\Delta_{2,3} \geq 2\Delta_{3,4})$ | **IV (2)** |
| **V** | $(\Delta_{2,3} \leq \Delta_{1,2} \leq \Delta_{3,4})$ | $\wedge\ (\Delta_{2,3}+\Delta_{3,4} \leq 2\Delta_{1,2})$ | **V (1)** |
| | | $\wedge\ (\Delta_{2,3}+\Delta_{3,4} \geq 2\Delta_{1,2})$ | **V (2)** |
| **VI** | $(\Delta_{3,4} \leq \Delta_{2,3} \leq \Delta_{1,2})$ | $\wedge\ (\Delta_{1,2}+\Delta_{3,4} \leq 2\Delta_{2,3})$ | **VI (1)** |
| | | $\wedge\ (\Delta_{1,2}+\Delta_{3,4} \geq 2\Delta_{2,3})$ | **VI (2)** |

**Table 2.1**: Data situations for a sample of size $n = 4$ distinguished by the order of the first differences $\Delta_{i,i+1}$ for $i \in \{1, 2, 3\}$.

| case | $\hat{\boldsymbol{\beta}}^{RM}$ | $\hat{\boldsymbol{\mu}}^{RM}$ |
|---|---|---|
| **I(1) & VI(2)** | $\Delta_{2,3}$ | $1/2 \cdot (Y_4 - 6Y_3 + 7Y_2)$ |
| **I(2) & VI(1)** | $\Delta_{2,3}$ | $1/2 \cdot (-3Y_3 + 4Y_2 + Y_1)$ |
| **II(1) & V(2)** | $\Delta_{1,2}$ | $Y_1 - \hat{\beta}^{RM} = Y_2 - 2\hat{\beta}^{RM}$ |
| **II(2) & V(1)** | $\frac{1}{12}(2\Delta_{1,2} + 5\Delta_{2,3} + 5\Delta_{3,4})$ | no simple expression |
| **III(1) & IV(2)** | $\frac{1}{12}(5\Delta_{1,2} + 5\Delta_{2,3} + 2\Delta_{3,4})$ | no simple expression |
| **III(2) & IV(1)** | $\Delta_{3,4}$ | $Y_3 - 3\hat{\beta}^{RM} = Y_4 - 4\hat{\beta}^{RM}$ |

**Table 2.2**: Possible expressions of the coefficients resulting from a Repeated Median regression for a sample $(Y_1, Y_2, Y_3, Y_4)'$ observed at equidistant times $(1, 2, 3, 4)'$. The specification of the case IDs is given in Table 2.1.

| case | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| **(1)** | ( 1 -1 -1 1) | ( 0 0 1 -1) | (-1 1 -1 1) | ( 1 -1 0 0) | (-1 1 -1 1) | (-1 1 1 -1) |
| | 0.06693 | 0.08113 | 0.10194 | 0.08113 | 0.10194 | 0.06693 |
| **(2)** | ( 1 -1 -1 1) | ( 1 -1 1 -1) | (-1 1 0 0) | ( 1 -1 1 -1) | ( 0 0 -1 1) | (-1 1 1 -1) |
| | 0.06693 | 0.10194 | 0.08113 | 0.10194 | 0.08113 | 0.06693 |

**Table 2.3**: Order of the RM residual signs and the corresponding probability of occurrence for $n = 4$ independent observations from a standard normal distribution for the data situations specified by the cases defined in Table 2.1.

| **zero residuals** | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_5$ | $(r_1 \wedge r_3)$ | $(r_2 \wedge r_4)$ | $(r_3 \wedge r_5)$ | $(r_1 \wedge r_5)$ |
|---|---|---|---|---|---|---|---|---|---|
| **prob.** | 0.1706 | 0.0478 | 0.2970 | 0.0478 | 0.1706 | 0.0276 | 0.0444 | 0.0276 | 0.1666 |

**Table 2.4**: Probabilities for the position of the zero residuals in a sample of size $n = 5$ in case of independent standard normal data.

In the four cases II(1), V(2), III(2) and IV(1) the RM line results in two zero residuals each, where the line leads through $Y_1$ and $Y_2$ for the cases II(1) and V(2); in case III(2) and case IV(1) it runs through $Y_3$ and $Y_4$, cf. Table 2.3.

In Section 3.2 a procedure is described which is based on the assumption that for a good fit at the rightmost position in a sample, the balance of the residual signs (2.6) should also be true for a certain subset of the sample. The number and the position of zero residuals both have an impact on the sum of residual signs of such a subset, which is used as a test statistic in Section 3.2.

Table 2.3 contains the probabilities for the occurrence of a certain order of signs of residuals from an RM regression fit to $n = 4$ observations which are independently generated from a standard normal distribution. With a probability of approximately 0.675, the RM line does not cross any observation, and with a probability of about 0.268 the RM line does not have any zero residual but its slope corresponds to the slope between $Y_2$ and $Y_3$, i.e. $\hat{\beta}^{RM} = \mathrm{b}(2, 3)$ (cases I and VI). The probability for an RM line which runs through the first two of four observations (cases II(1) and V(2)) is approximately 0.162 which also corresponds to the probability for the RM line leading through $Y_3$ and $Y_4$ (cases III(2) and IV(1)). Thus, if there are zero residuals present, they either occur subsequently at the beginning or the end. However, the balance of residual signs (2.6) is fulfilled with probability one for $n = 4$ data generated according to a model with continuous error distribution.

### SAMPLE SIZE $n = 5$

Similar to the derivations for sample size $n = 4$, the probabilities for the number and location of zero residuals from an RM regression applied to five observations coming from independent standard normal random variables, can be evaluated by using the characterisation of different data situations via the first differences $\Delta_{i,i+1}$. The calculations can be found in Appendix B. Table B.9 lists all possible orders of residual signs in a sample of size $n = 5$ together with their probabilities.

Summarising the outcomes, Table 2.4 provides the probabilities for all possible positions of the zero residuals in a standard normal sample of size $n = 5$. Although the data will be in general position with probability one, the balance of residual signs (2.6) only occurs with probability 0.734 and it is

$$P\left(\sum_{i=1}^{5} \text{sign}(r_i) = -1\right) = P\left(\sum_{i=1}^{5} \text{sign}(r_i) = 1\right) = 0.1331\,.$$

Furthermore, it can be seen in Table 2.4 that a zero residual rather occurs at locations with odd position numbers, i.e. at the first, third, or fifth position in the sample.

### Sample Sizes Larger Than $n = 5$

The differentiation between odd and even sample sizes is intuitive, because for data in general position, the balance of residual signs (2.6) implies that for an odd sample size there will be one zero residual, and for an even sample size either no or two zero residuals can occur. However, the previous derivations for the sample sizes $n = 3$, $n = 4$ and $n = 5$ have shown that equation (2.6) is not necessarily always true. This appears in particular for odd sample sizes as will be shown in the following. All the results in this subsection are based on 100 000 simulated standard normal data sets for each of the considered sample sizes $n \in \{6, 7, \ldots, 121\}$.

#### Even Sample Sizes

For even sample sizes (2.6) is always true, i.e. it occurs for all simulated even sample sizes with a relative frequency of one. That means that there are either no or two zero residuals in the sample.

Assuming that there are two zero residuals in the sample, it is of interest at which positions they occur. Figure 2.1 illustrates the relative frequencies for the positions $(i, j)$ of the two zero residuals with $i, j \in \{1, \ldots, n\}$, conditional on the fact that there are exactly two residuals. The larger the square at (position1, position2), the higher the relative frequency for the two zero residuals appearing at these two positions in the sample.

It shows that some combinations of positions are much more frequent while others do not occur at all: for example, for a sample of size $n = 6$ (not shown here) the set of the zero residual position pairs appearing in 21 120 (out of 100 000) data sets, corresponds to $\{(1, 2), (1, 4), (1, 6), (3, 4), (3, 6), (5, 6)\}$. For larger sample sizes it can also be observed that the occurrence of both zero residuals at odd, or both at even positions, respectively, are very rare events (see Figure 2.1). Furthermore, Figure 2.1 shows that at least one of the two zero residuals is likely to occur at the edge of the sample; often both residuals appear at the boundaries, i.e. at positions $i = 1$ and $i = n$.

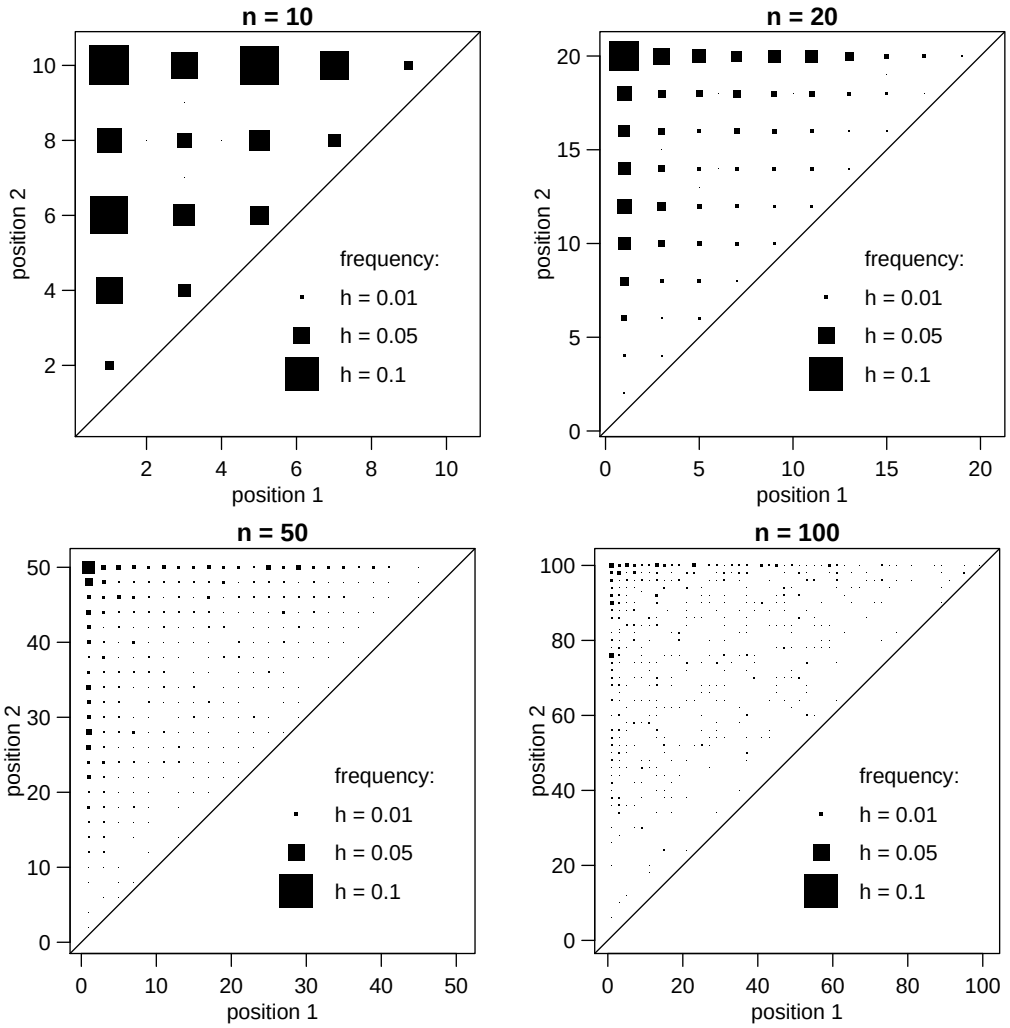


**Figure 2.1**: Relative frequencies for the positions of two zero RM residuals in samples of even size, conditional on the fact that there are exactly two residuals.

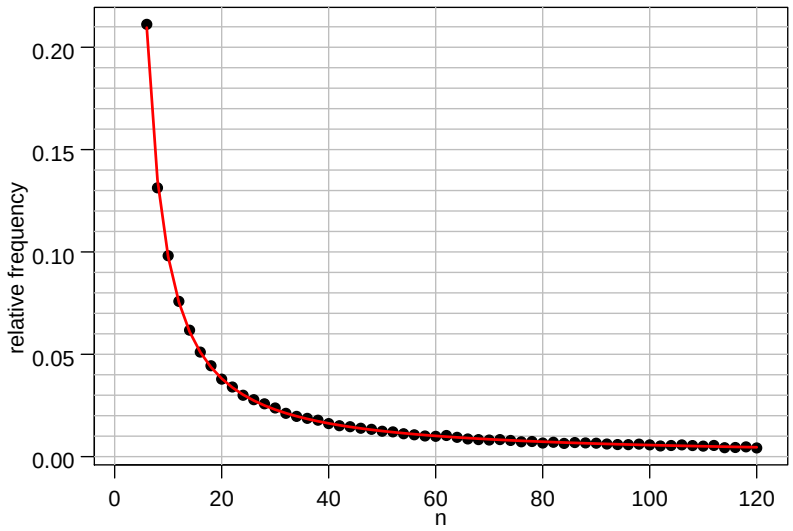


**Figure 2.2**: Relative frequencies of two zero RM residuals for even sample sizes $6 \leq n \leq 120$ with an approximation of the underlying probability (red line).

Figure 2.2 indicates that the relative frequency of zero residuals in a sample of even size decreases with an increasing sample size. Using the shown simulated relative frequencies, we can derive the following approximation of the probability of two zero RM residuals in even sample sizes:

$$P\Big(\#\{r_i = 0, i = 1, \ldots, n\} = 2 \,\Big|\, n \text{ even}\Big) \;\approx\; -0.0004 + \frac{0.5605}{n} + \frac{4.1951}{n^2}\,. \tag{2.14}$$

The red line in Figure 2.2 shows that this approximation fits the simulated relative frequencies nicely. Furthermore, it shows that for even sample sizes larger than $n = 20$ zero residuals appear in less than 4%, for $n$ larger than 40 in less than 2% of the cases; and for $n$ about 60 or larger just about one percent (or less) of the samples contain residuals with a value of zero.

Of course this approximation should be taken with care: The approximation (2.14) is achieved by means of simulations of 100 000 data sets for each sample size, and simulated values represent observations from a continuous distribution only to the extend of the precision of the computer used for the generation of the pseudo-random numbers. Moreover, the limited number of simulation runs also restricts the accuracy of the approximation.

For $n > 1408$ formula (2.14) returns a negative value, and for $n = 2$ it yields a value greater than one which both is not acceptable for an approximation of a probability. For $n = 2$ an RM line always leads through both observations and hence, the probability for two zero residuals is equal to one. For $n = 4$ the probability for two zero RM residuals under the normality assumption equals 0.32452 as can be derived from Table 2.3 and thus, an approximation of this probability is only required for $n \geq 6$.

For such sample sizes we can assume formula (2.14) to provide an approximation for the probability of two zero RM residuals under normality with sufficient precision for the procedure proposed in Section 3.2.

However, for sufficiently large, even sample size we assume the probability of zero residuals to be negligible, because for large $n$ the approximation yields values very close to zero: For $n = 100$ the probability for two zero residuals is approximatively 0.56% and for $n \geq 102$ the approximation yields values smaller than 0.5%. Hence, we propose to set the approximated probability to zero for $n \geq 100$.

### Odd Sample Sizes

For an odd number of observations in general position one can expect either one or two zero residuals resulting from RM regression: As could be seen from the calculations for standard normal data with $n = 5$, there is always at least one zero residual (cf. Table B.9), i.e.

$$P\big(\#\{r_i = 0, i = 1, \ldots, 5\} = 0\big) \;= 0\,.$$

Furthermore, our simulations revealed no case for any sample size $n \in \{7, 9, 11, \ldots, 121\}$ where an RM regression resulted in no zero residual. Thus, we assume the probability for no zero residual in a sample of odd size to be zero, and it is

$$\begin{aligned} P\Big(\#\{r_i = 0, i = 1, \ldots, n\} = 1 \,\Big|\, n \text{ odd}\Big) &= P\Big(\sum_{i=1}^{n} \text{sign}(r_i) = 0 \,\Big|\, n \text{ odd}\Big) \\ &= 1 - P\Big(\#\{r_i = 0, i = 1, \ldots, n\} = 2 \,\Big|\, n \text{ odd}\Big)\,. \end{aligned}$$

If there are two zero residuals, the balance of residual signs (2.6) is not true anymore. Figure 2.3 shows the distribution of the location of two zero residuals within the sample, conditioned on the fact that the sample size is odd.

Similar to the even sample sizes, in the majority of cases the zero residuals appear at particular positions, especially at locations with odd position number $i$, and the zero residuals are more likely to occur close to the boundaries. This is not very obvious for the larger sample sizes displayed in the bottom panels of Figure 2.3 which is due to the fact that these graphics show conditional distributions: they are based on those data sets (out of the 100 000 simulated ones) which result in exactly two zero residuals; for the sample size $n = 15$ there are 864 such data sets, for $n = 29$ there are still 244, for $n = 49$ it is 109, but the plot for $n = 99$ is only based on the positions of the two zero residuals within 25 data sets.

Figure 2.4 shows the relative frequencies for the occurrence of two zero RM residuals for an odd sample size; e.g. for $n = 99$ the frequency corresponds to 0.00025. The red line gives an approximation of the probability for two zero residuals in samples of odd size according to

$$P\Big(\#\{r_i = 0, i = 1, \ldots, n\} = 2 \,\Big|\, n \text{ odd}\Big) \approx -0.0015 + \frac{0.196}{n} - \frac{3.978}{n^2} + \frac{48.437}{n^3}\,. \quad (2.15)$$

Similar to the approximation (2.14) for even sample sizes, this approximation of the probability has to be taken with care. In particular, the approximation yields negative values for $n \geq 111$. However, analogous to the case of even sample sizes, we can assume the probability of two zero residuals to be negligible for sufficiently large $n$: Figure 2.4 shows that, even for small sample sizes like $n = 13$, the approximate probability is just about one percent (or less), and for increasing sample size it is fast decreasing and hence, we suggest to approximate this probability by zero for $n \geq 101$.

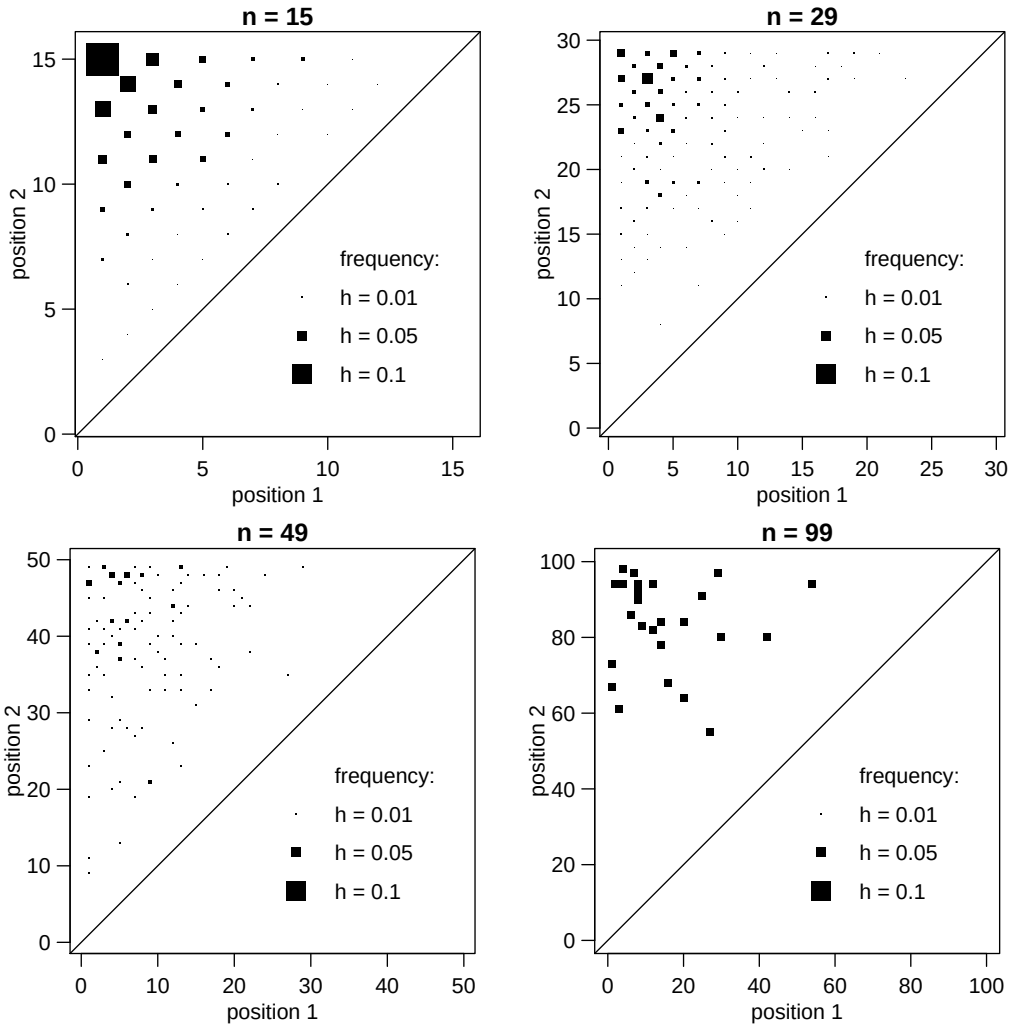


**Figure 2.3**: Relative frequencies for the positions of two zero RM residuals in samples of odd size, conditional on the fact that there are exactly two residuals.

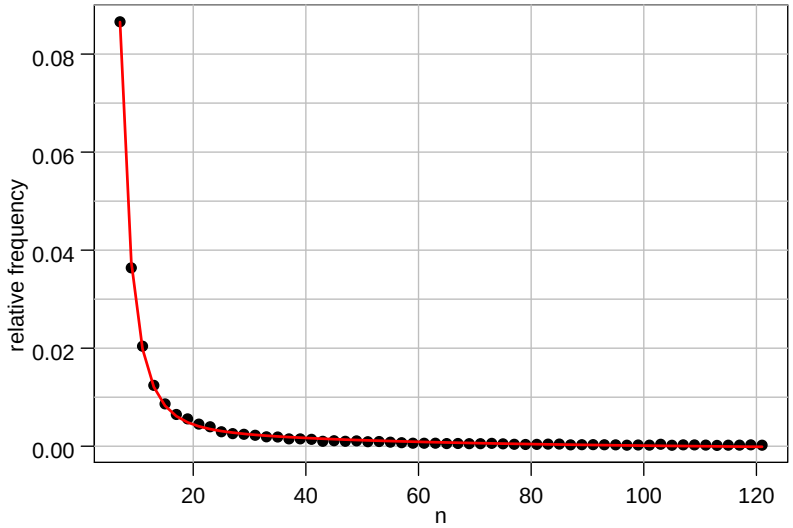


**Figure 2.4**: Relative frequencies of two zero RM residuals for odd sample sizes $7 \leq n \leq 121$ with an approximation of the underlying probability (red line).

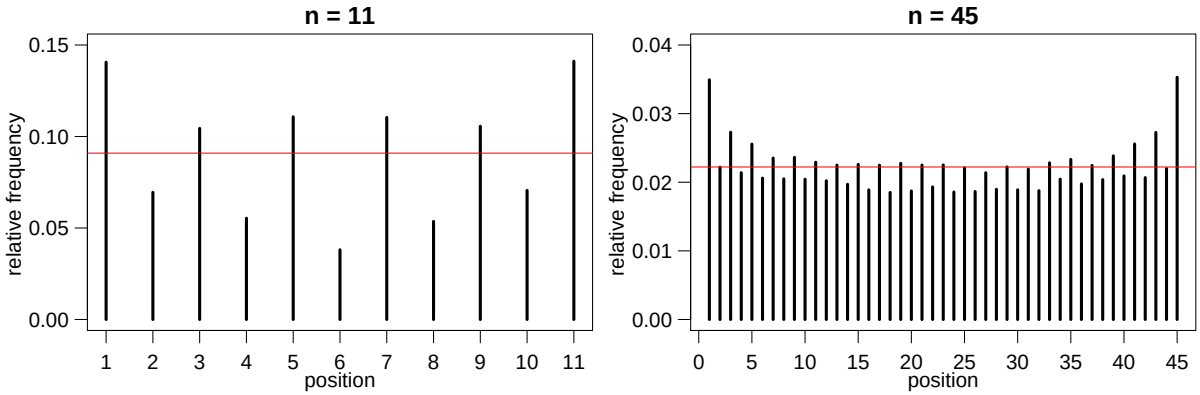


**Figure 2.5**: Empirical density for the position of one zero RM residual in a small and a medium sized sample with odd sample size, conditional on the fact that the sample contains exactly one zero residual. The red horizontal line displays the value $1/n$.

Observing two residuals with a value of zero means that the sum of signs of RM residuals is either minus one or plus one. In our simulations, both events occur with a similar frequency for each considered odd sample size. This is due to the symmetry of the error distribution, i.e. the normal distribution, because the distribution of $\varepsilon_i$ corresponds to the distribution of $-\varepsilon_i$ for $i = 1, \ldots, n$. Hence, it is

$$\begin{aligned} P\left(\sum_{i=1}^{n} \operatorname{sign}(r_i) = -1 \,\middle|\, n \text{ odd}\right) &= P\left(\sum_{i=1}^{n} \operatorname{sign}(r_i) = 1 \,\middle|\, n \text{ odd}\right) \\ &= \frac{1}{2}\left(1 - P\left(\sum_{i=1}^{n} \operatorname{sign}(r_i) = 0 \,\middle|\, n \text{ odd}\right)\right) \\ &= \frac{1}{2} P\Big(\#\{r_i = 0, i = 1, \ldots, n\} = 2 \,\Big|\, n \text{ odd}\Big)\,. \end{aligned}$$

In most samples there is exactly one zero residual resulting from an RM regression fit. If the residuals were independent the probability of the one zero residual occurring at position $i$ in the sample would be $1/n$, i.e. the zero residual could occur at each position with equal probability.

Figure 2.5 shows the empirical distribution of the position of a zero RM residual, conditioned on the fact that there is exactly one residual in the sample, derived from simulated standard normal data for a small sample size of $n = 11$ and a larger sample size of $n = 45$. The red horizontal line marks the value of $1/n$ for comparison of the empirical conditional density with the uniform density. The graphics for further odd sample sizes show the same characteristics, i.e. they look quite similar.

Both panels in Figure 2.5 show that the simulated relative frequencies lie close to the value of $1/n$, but that the occurrence of the zero residual at odd positions in the sample is more frequent than at even positions. Furthermore, it is more likely to observe a zero

residual at (or close to) the boundaries, in particular at the positions $i = 1$ or $i = n$. This behaviour does not fade for larger sample sizes (investigated for sizes up to $n = 121$) but shows about to the same extent for any sample size.
An explanation for this might be given by the fact that there are more combinatoric possibilities of data situations which result in a regression line leading through an observation at an odd position in the sample, especially at $i = 1$ or $i = n$. For $n = 5$ this is shown in Appendix B, but it also seems to be true for larger odd sample sizes.
Consequently, looking at residuals in the centre of a sample, the probability of one zero residual occurring at position $i$ could be approximated by $1/n$ for sufficiently large $n$, say $n \geq 15$. However, at the edges a value of $1/n$ would not provide a good approximation and simulated estimates for the requested probabilities might be more useful.

## 2.3 The Correlation Structure of the Residual Signs

The dependence structure of the RM residuals and their signs might provide further explanations for the positioning of zero RM residuals within the sample. Unfortunately, this structure is unknown so far. Intuitively, it is clear that some sort of correlation structure has to be present for the signs of the RM residuals. For example, it is easy to see that for an odd sample size $n$ no RM regression line could result in $\text{sign}(r_i) = -1$ for $i = 1, \ldots, (n-1)/2$; $\text{sign}(r_{(n-1)/2+1}) = 0$; and $\text{sign}(r_i) = 1$ for $i = (n+1)/2, \ldots, n$.
While for RM regression usually the residual signs sum up to zero (2.6), Least Squares (LS) regression is designed such that the residuals themselves sum up to zero. Therefore, we will investigate the correlations between residual signs from RM regression obtained from simulations with standard normal data and compare this with the behaviour of LS residuals for which the correlation structure is well known.
Rewriting model (2.1) using matrix notation, with $\mathbf{X}$ being the *design matrix* and $\boldsymbol{\theta}$ denoting the vector of coefficients, it is

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & n \end{pmatrix} \begin{pmatrix} \mu \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\Leftrightarrow \mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}\,.$$

The Least Squares (LS) estimator is then defined by

$$\widehat{\boldsymbol{\theta}}^{LS} = \begin{pmatrix} \hat{\mu}^{LS} \\ \hat{\beta}^{LS} \end{pmatrix} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\,\mathbf{Y} = \mathbf{H}\mathbf{Y}\,,$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ denotes the so called *hat matrix*.

| | $i=1$ | $i=2$ | $i=3$ | $i=4$ | $i=5$ |
|---|---|---|---|---|---|
| $i=1$ | 1.00 | −0.76 | −0.35 | 0.00 | 0.50 |
| $i=2$ | −0.76 | 1.00 | −0.27 | −0.14 | 0.00 |
| $i=3$ | −0.35 | −0.27 | 1.00 | −0.27 | −0.35 |
| $i=4$ | 0.00 | −0.14 | −0.27 | 1.00 | −0.76 |
| $i=5$ | 0.50 | 0.00 | −0.35 | −0.76 | 1.00 |

**Table 2.5**: Correlation matrix for Least Squares residuals for $n = 5$ observations from a standard normal distribution.

For an equidistant design as considered here, the hat matrix is given by $\mathbf{H} = (h_{ij})_{i,j\in\{1,\ldots,n\}}$ with

$$h_{ij} = \frac{1}{n(n-1)}\left(2(2n+1) - 6i + j\left(\frac{12}{n+1}i - 6\right)\right) .$$

This matrix determines the correlation structure for the LS residuals $r_i^{LS}$, because for $\mathbf{r}^{LS} = (r_1^{LS}, \ldots, r_n^{LS})'$, $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, and $\mathbf{I}_n$ denoting the $n$-dimensional identity matrix, it is

$$\begin{aligned} & \mathbf{Cov}(\mathbf{r}^{LS}) && = (\mathbf{I}_n - \mathbf{H})\sigma^2 \\ \Leftrightarrow \quad & \text{Cov}(r_i^{LS}, r_j^{LS}) && = (1 - h_{ij})\sigma^2 \quad \text{and} \quad \text{Var}(r_i^{LS}) = (1 - h_{ii})\sigma^2 \\ \Rightarrow \quad & \text{Corr}(r_i^{LS}, r_j^{LS}) && = \frac{1 - h_{ij}}{\sqrt{1 - h_{ii}}\sqrt{1 - h_{jj}}} . \end{aligned}$$

The values $h_{ii}$ on the diagonal of $\mathbf{H}$ are increasing towards the boundaries, i.e. towards $i = 1$ and $i = n$. They are called *leverages*, because they indicate stronger correlations of any residual with a residual close to the boundary and weaker correlations between residuals close to the centre. The larger the difference between $i$ and $j$, the stronger the correlation. Neighbouring LS residuals are always negatively correlated while those which are farthest apart are positively correlated. This can e.g. be seen in the correlation matrix for $n = 5$ given in Table 2.5.

The top left panel in Figure 2.6 provides a graphical representation of this correlation matrix; the other two top panels in Figure 2.6 represent the correlation matrices of LS residuals for the larger sample sizes $n = 10$ and $n = 15$ indicating that for increasing sample size the order of magnitude of all correlations decreases.

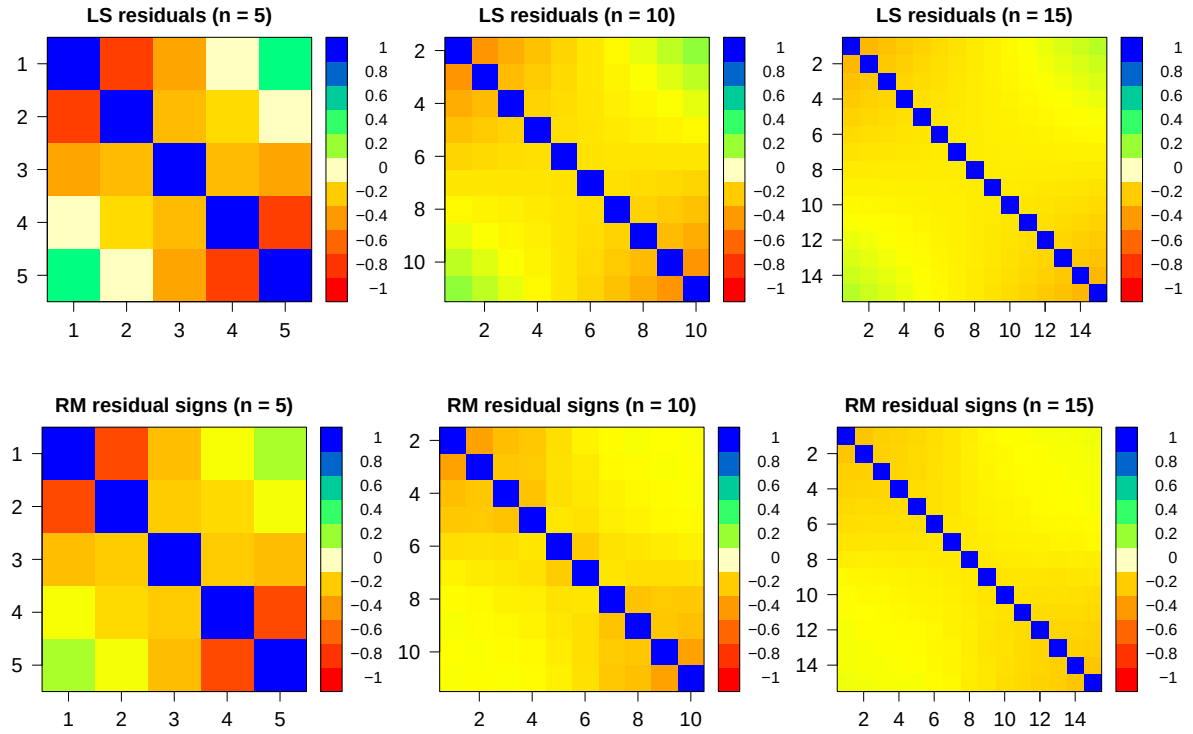


**Figure 2.6**: Graphical representation of the correlation matrices for residuals from least squares regression at an equidistant design (top panels) and for the signs of RM residuals (bottom panels) for some small sample sizes.

| | $i = 1$ | $i = 2$ | $i = 3$ | $i = 4$ | $i = 5$ |
|---|---|---|---|---|---|
| $i = 1$ | 1.00 | −0.71 | −0.26 | 0.02 | 0.17 |
| $i = 2$ | −0.71 | 1.00 | −0.21 | −0.15 | 0.01 |
| $i = 3$ | −0.26 | −0.21 | 1.00 | −0.20 | −0.25 |
| $i = 4$ | 0.02 | −0.15 | −0.20 | 1.00 | −0.72 |
| $i = 5$ | 0.17 | 0.01 | −0.25 | −0.72 | 1.00 |

**Table 2.6**: Simulated correlation matrix for the signs of RM residuals for $n = 5$ observations from a standard normal distribution.

For comparison, Table 2.6 shows the correlation matrix for the signs of RM residuals, calculated from 100 000 simulated standard normal data sets of size $n = 5$; and the bottom panels of Figure 2.6 display graphical representations of the correlation matrices for the signs of RM residuals for sample sizes $n = 5$, $n = 10$ and $n = 15$.

Generally, the correlations between RM residual signs are weaker than those between LS residuals, especially for design points $i$ which are far apart, but the signs of the pairwise correlations almost always agree (except for correlations close to zero). Often, even the order of magnitude is the same for correlations between LS residuals and for correlations between signs of RM residuals – especially for observations in the centre.
Although the distribution and the true underlying dependence structure of the signs of RM residuals is still unknown, the performed simulations suggest that, at least for a standard normal error distribution, they have a correlation structure similar to that of Least Squares residuals. Consequently, there also might be some sort of leverage effect for the signs of RM residuals towards the boundaries which could provide some explanation of the more frequent observations of the zero residuals at the boundaries (see Figure 2.5).

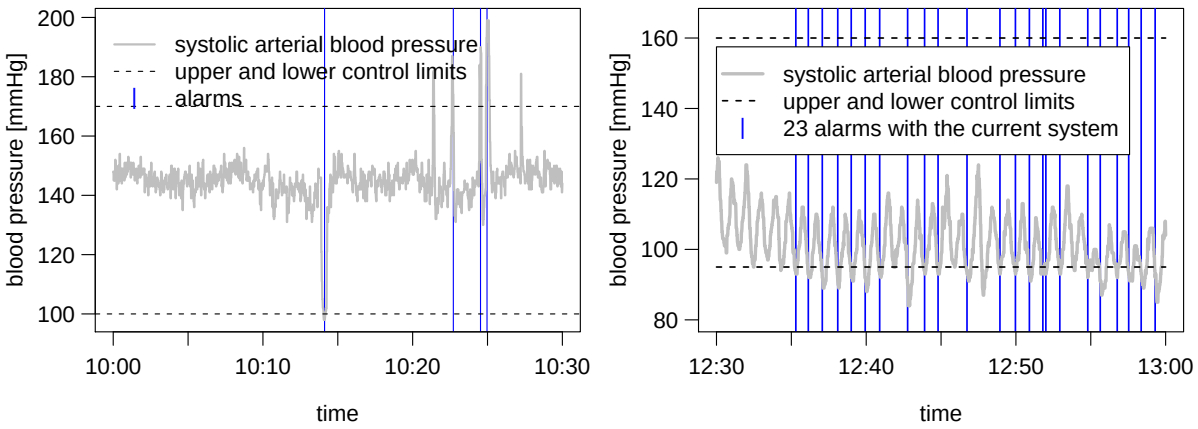


**Figure 3.1**: Shortcomings of current alarm systems. Measurement artefacts can trigger alarms – even if alarm delay rules are applied (left). At 10:21h and 10:27h the systolic arterial blood pressure does not exceed the upper limit for more than four seconds and thus, no alarm is set off (alarm delay rule). Furthermore, depending on the alarm settings, the overall alarm rate of current systems can be very high – even in the absence of outliers (right).

# 3 Online Signal Extraction

Currently, intensive care monitoring systems use alarm rules for variables such as heart rate, blood pressures or oxygen saturation which are essentially based on simple thresholds: violations of the upper or lower control limit activate an alarm – sometimes after a certain offset time. This mechanism produces many false alarms due to measurement artefacts, patient movements, or transient fluctuations around the alarm limits.

For example, the monitoring system we study here, triggers an alarm for the systolic arterial blood pressure if the observed values exceed the upper or lower control limit (set by the medical staff) for at least four seconds. This offset time provides a possibility for making the system robust against single measurement artefacts. However, experience with real data suggests that in practice such artefacts also occur in 'patches' of several consecutive values. Thus, even an offset time for the alarm does not completely avoid the occurrence of false alarms caused by artefacts as is illustrated in the left panel of Figure 3.1. Furthermore, the physiological time series may vary around a value very close to an alarm limit as can be seen in the right panel of Figure 3.1. This causes a sequence of alarms where one (possibly persistent) alarm would be sufficient to indicate that the physiological variable has reached a certain limit.

Pre-processing the input data for an alarm system by applying robust online filters and comparing the filter output with the preset alarm limits can reduce the number of false alarms (Kuhls, 2008). Such filters have to be able to deal with various levels of noise and many artefacts. Furthermore, they have to be suitable for time series with trends in the underlying level and, since a change in the patient's health status is often accompanied

by a trend change or a level shift, the applied filtering methods also need to be able to retain such patterns with no, or a very small time delay.

'Classical' time series techniques, dynamic linear models, Kalman filters, and methods in statistical process control often rely on an underlying parametric model. Their sensitivity to misspecification of the model parameters provides a source of error which is unacceptable for online monitoring in intensive care. This is possibly the reason why most of these approaches never got implemented in commercial products (Imhoff and Kuhls, 2006). Therefore, we focus on simple moving window approaches for robust online signal extraction and discuss their merits for preserving clinically relevant patterns such as trends, shifts, or extremes and for the removal of irrelevant outliers.

In this chapter we present procedures for univariate and multivariate signal extraction which are fast, efficient and robust, and can be used for discretely measured data with low variability as well as in situations with many outliers in the style of Gather and Schettlinger (2007).

## 3.1 UNIVARIATE SIGNAL EXTRACTION

As a simple and general working model for a univariate time series $(Y_t)_{t\in\mathbb{Z}}$, we assume the following:

$$Y_t = \mu_t + \epsilon_t \,, \tag{3.1}$$

where $\mu_t$ denotes the underlying signal at time $t$ which is assumed to run smoothly over time apart from some sudden trend changes or level shifts. The error term $\epsilon_t$ can further be decomposed into

$$\epsilon_t = \varepsilon_t + \nu_t \,,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an error process of random variables with $E(\varepsilon_t) = 0$ for all $t \in \mathbb{Z}$ and (possibly time-dependent) variance $Var(\varepsilon_t) = \sigma_t^2$; $(\nu_t)_{t\in\mathbb{Z}}$ represents an outlier-generating process which may be equal to zero for the majority of points in time.

Moving window techniques offer an intuitive and simple approach to extract the signal level $\mu_t$ from the noisy time series. Moving averages are popular since they trace trends and they are very efficient for Gaussian samples. However, sudden level shifts are not preserved as such and outliers can cause considerable bias (Figure 3.2).

Applying a local linear instead of a local constant approximation is known to improve the smoothing results, see e.g. Cleveland (1979). Local polynomial fits applied for smoothing date back to the end of the 19th / beginning of the 20th century (Woolhouse, 1870; Spencer, 1904). Actually, the least squares fit for the centre of a time window corresponds to the mean of the observations and hence, to the filter output of a moving average, i.e. these filters are strongly influenced by outliers.

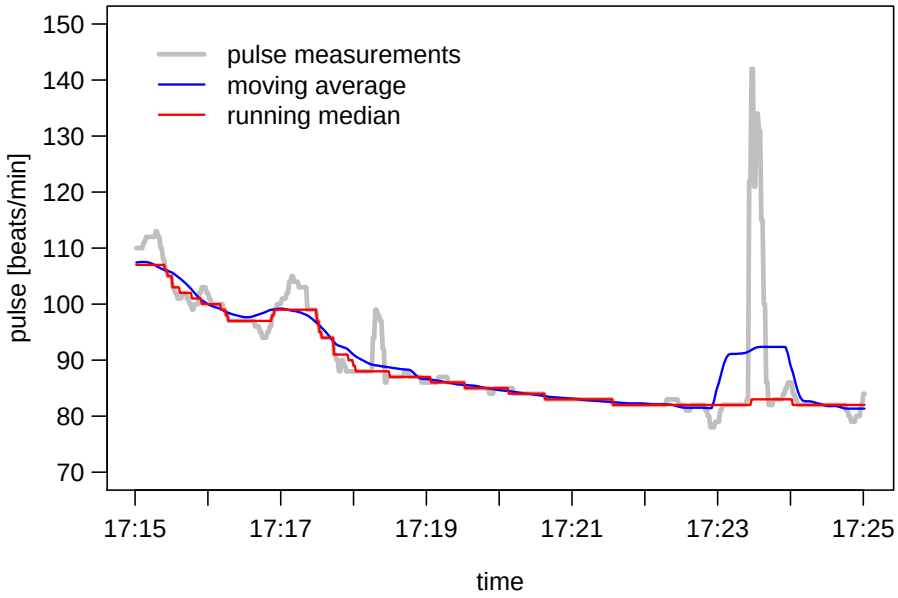


**Figure 3.2**: Moving average and running median applied to ten minutes of pulse measurements recorded on an intensive care unit with a frequency of once per second, both using a window with a width of $n = 61$ seconds.

Median filters, as suggested by Tukey (1977), are robust against artefacts, resist up to $\lfloor n/2 \rfloor$ outliers within one time window of length $n$, and they are capable of tracing level shifts. However, they deteriorate in trend periods (see Figure 3.2). An early approach to robust smoothing using local linear fits is proposed by Cleveland (1979) who assigns small weights to large residuals in an iterative linear regression procedure.

Using polynomials of orders higher than one for the local regression fit might yield less biased signal estimations, but at the same time it allows for the modeling of local extremes which might cause a higher variability of the filter output, see e.g. McDonald and Owen (1986). In addition, such filters generally require a high computation time.

On the basis of Schettlinger, Fried and Gather (2006), we review univariate filters which apply at most a local linear regression fit and thus, restrict the computation time; which are capable of tracing trends well and thus, guarantee a smooth output; and which are able to limit the bias of their output in the presence of outliers.

### 3.1.1 Simple Robust Regression Filters

In view of the weakness of running medians in trend periods, Davies, Fried and Gather (2004) apply robust regression methods to a moving time window of odd width $n = 2m + 1$, $m \in \mathbb{N}$, to approximate the signal level at the centre of each window. A real-time application of such a filter means a time lag of $m$ time units for the estimation of the current signal and hence, we call them *retrospective regression filters*. Gather, Schettlinger

and Fried (2006) investigate *online filters* based on the same principle, where the signal level is estimated at the most recent time by the last fitted value of a robust regression fit in each window. This allows a good and reliable approximation of the true signal without time delay – even in trend periods.

The local linearity assumption for the signal in a window of width $n \in \mathbb{N}$, can be expressed by

$$\mu_{t+i} \approx \mu_t + \beta_t\, i \tag{3.2}$$

$$\text{with } i = \begin{cases} -m, \ldots, m \text{ and } n = 2m+1 & \text{for retrospective filters and} \\ -n+1, \ldots, -1, 0 & \text{for online filters,} \end{cases} \tag{3.3}$$

where $\beta_t$ denotes the slope of the line in a window used to approximate the level at $t$.

For retrospective regression filters Davies, Fried and Gather (2004) compare the use of $L_1$ regression, Repeated Median (RM) regression (Siegel, 1982), and least median of squares (LMS) regression (Hampel, 1975; Rousseeuw, 1984). For regression on an equidistant grid, Davies, Fried and Gather (2004) show that $L_1$ only has a finite sample breakdown point of approximately 0.293 while LMS and RM regression achieve a maximum breakdown point of about 0.5, or exactly: $\lfloor n/2 \rfloor / n$, which is the maximum among all regression equivariant estimators. Furthermore, both methods possess the exact fit property (2.5), and LMS is regression-, scale- and affine equivariant while RM lacks equivariance w.r.t. affine transformations of the explanatory variable (see Section 2).

The LMS filter offers the highest robustness against many large outliers and is able to trace level shifts and trend changes well, but it tends to instabilities and has the highest computation time among the three compared methods. The RM filter slightly smoothes extremes and level shifts, but from the investigated methods, it is considered the best choice for signal extraction because it does not only offer considerable robustness against outliers, but it is also stable with regard to moderate variations in the data.

A filter based on the least quartile difference (LQD) estimator, introduced by Croux, Rousseeuw and Hössjer (1994), provides a compromise between the edge-preserving property of the LMS filter and the smoothness of the RM filter as can be seen in Figure 3.3. LQD has with $\lfloor n/2 \rfloor / n$ the same (maximum) finite sample breakdown point like RM and LMS, and is also regression-, scale and affine equivariant. However, this method requires high computational effort when using common algorithms.

For signal extraction without time delay, Schettlinger (2004) and Gather, Schettlinger and Fried (2006) examine online filters based on RM and LMS regression, because both show certain advantages in the retrospective situation (Davies, Fried and Gather, 2004). Furthermore, they consider least trimmed squares (LTS) regression (Rousseeuw, 1985),

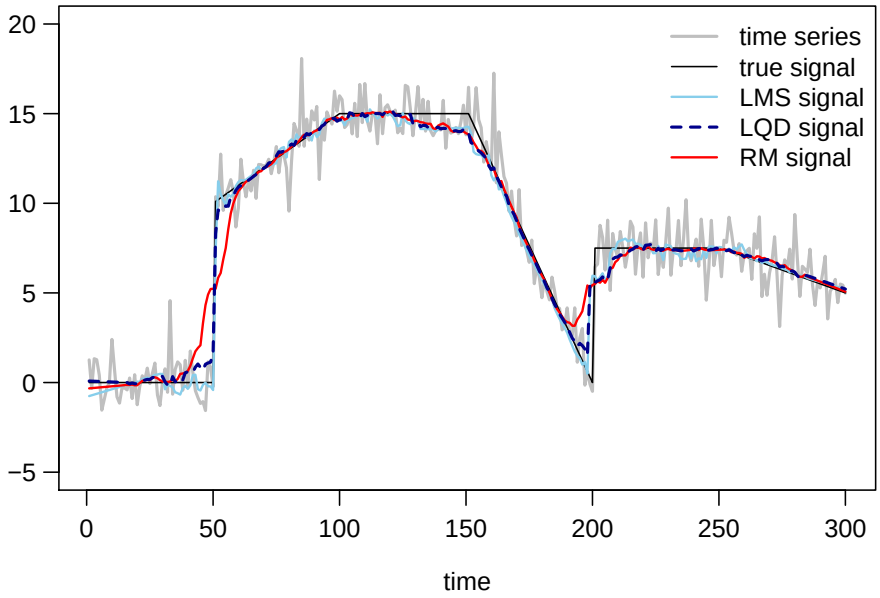


**Figure 3.3**: Comparison of retrospective RM, LQD and LMS regression filters applied to a simulated time series with standard normal noise, all using a window of width $n = 31$. The LQD filter provides a compromise between the smooth output of the RM filter and the shift preservation of the LMS filter.

since it is pointed out in Rousseeuw, Van Aelst and Hubert (1999) that LMS is outperformed by LTS regression because of its smoother objective function which results in a higher efficiency. Additionally, deepest regression (DR) (Rousseeuw and Hubert, 1999) is considered because it is expected to deal well with asymmetric and heteroscedastic errors. Both of these regression methods are regression-, scale- and affine-equivariant, but LTS achieves the maximum finite sample breakdown point of $\lfloor n/2 \rfloor / n$, while DR only has a breakdown point of about 1/3.

It turns out that the differences in the outcomes between the online LMS and LTS regression filters are negligible, and there is little difference between the RM and DR online filters as shown in Figure 3.4. In the online situation, LMS and LTS track shifts with a longer delay than their competitors and tend to overshoot shifts, while RM and DR show more stable results, see also Figure 3.4. A filter based on RM regression is considered the best competitor from the investigated online filters.

Robust regression methods generally need more computation time than non-robust techniques. To improve the applicability of robust regression filters in real-time, research has been carried out on the enhancement of the computational speed and on geometrical interpretations of the regression problem, see e.g. Edelsbrunner and Souvaine (1990) and Mount et al. (1997) for LMS, Rousseeuw and Van Driessen (2002) for LTS, and Langerman and Steiger (2003) for deepest regression.

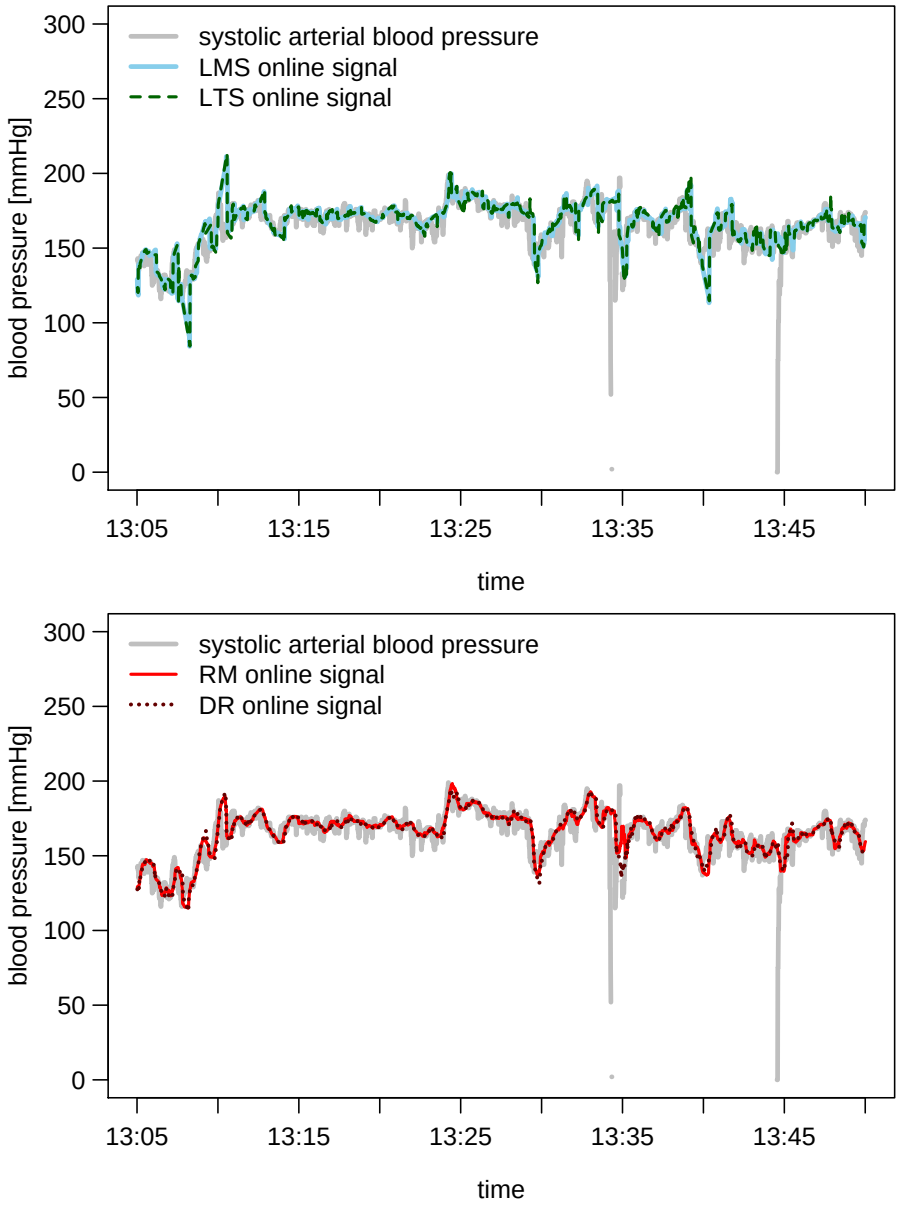


**Figure 3.4**: Online signals for a time series of systolic arterial blood pressure measurements extracted with four different robust regression filters using a moving window of width $n = 60$, i.e. the estimated signal is based on the last minute of measurements. The top panel shows the LMS and LTS online signals, the bottom panel shows the RM and DR online signals.

For LQD regression on $n$ bivariate data points Bernholt, Nunkesser and Schettlinger (2007) propose algorithms with expected running times of $\mathcal{O}(n^2 \log^2 n)$, or $\mathcal{O}(n^2 \log n)$, respectively, and a deterministic running time of $\mathcal{O}(n^2 \log^2 n)$. Therefor they apply the concept of geometric duality which Chazelle, Guibas and Lee (1985) use for solving geometrical problems. This provides a large improvement in computation time on currently available LQD algorithms: the exact algorithm by Croux, Rousseeuw and Hössjer (1994) needs $\mathcal{O}(n^5 \log n)$ time, and an adaptation of a fast LMS algorithm by Edelsbrunner and Souvaine (1990) still leads to a running time of $\mathcal{O}(n^4)$. Furthermore, the algorithms by

| | LQD | LMS | LTS | DR | RM |
|---|---|---|---|---|---|
| straightforward | $\mathcal{O}(n^5 \log n)$ | $\mathcal{O}(n^4)$ | $\mathcal{O}(n^3 \log^2 n)$ | $\mathcal{O}(n^3)$ | $\mathcal{O}(n^2)$ |
| update | $\mathcal{O}(n \log n)$ | $\mathcal{O}(n^2)$ | $\mathcal{O}(n^2)$ | $\mathcal{O}(n \log^2 n)$ | $\mathcal{O}(n)$ |

**Table 3.1**: Asymptotic computation times required for a regression fit to a sample of size $n$ for five robust regression methods. The first line contains the times when using a straightforward implementation, the second line shows the times required for an update of the regression estimates using improved algorithms.

Bernholt, Nunkesser and Schettlinger (2007) can be used to perform least quantile of squares (LQS) regression (Rousseeuw and Leroy, 1987; Rousseeuw and Hubert, 1997) and LMS regression through the origin in expected running time $\mathcal{O}(n \log n)$ or deterministic running time $\mathcal{O}(n \log^2 n)$.

Apart from the improved asymptotic computation times, Bernholt, Nunkesser and Schettlinger (2007) show in simulations that their proposed approximation algorithm is even faster for small sample sizes than the asymptotics suggest. Thus, these algorithms might increase the practical relevance of the LQD filter for online applications.

Another possibility to improve the computational efficiency of robust regression filters, is provided by the usage of update algorithms as suggested in McDonald and Owen (1986) for (non-robust) smoothing via local least squares fitting. Here the term 'update' means that estimation takes place using information stored from the last time window – only inserting the new information given by the most recent data point and deleting that of the oldest data point. Such update algorithms save a lot of computation time because the estimates do not have to be calculated for each window from scratch. For the RM regression line Bernholt and Fried (2003) present an algorithm which only needs linear time $\mathcal{O}(n)$ for an update. This update algorithm turns the RM filter into the one with the lowest computation time from investigated robust regression filters: Table 3.1 provides a comparison of the (asymptotic) computation times that the investigated robust regression filters need for estimating the signal at a particular point in time, and it clearly shows the predominance of the RM filter in terms of computation. Considering also the good performance of the RM filter for online signal extraction, it meets the most important requirements for online application in intensive care (Schettlinger and Imhoff, 2005).

### 3.1.2 Repeated Median Hybrid Filters

As pointed out above, a simple RM filter does not preserve sudden level shifts as such (Davies, Fried and Gather, 2004). Heinonen and Neuvo (1987, 1988) emphasise the advantages of linear median hybrid filters for preserving such signal edges. FIR median hybrid (FMH) filters are computationally even less demanding than running medians and

preserve shifts similarly well or even better than these. An FMH filter is defined as the median of several linear subfilters $\Phi_i$, $i = 1, \ldots, M$:

$$\hat{\mu}_t^{FMH} = \text{med}\{\Phi_1, \Phi_2, \ldots, \Phi_M\}\,.$$

For (retrospective) signal extraction from blood pressure measurements, Heinonen, Kalli, Turjanmaa and Neuvo (1985) use a simple FMH filter with $M = 3$ subfilters, consisting of two one-sided moving averages and the central observation as a central subfilter, i.e. for a sample $\mathbf{y}_t = (y_{t-m}, \ldots, y_t, \ldots, y_{t+m})'$:

$$\Phi_1(\mathbf{y}_t) = \frac{1}{m}\sum_{i=1}^{m} y_{t-i}\;, \quad \Phi_2(\mathbf{y}_t) = y_t\;, \quad \Phi_3(\mathbf{y}_t) = \frac{1}{m}\sum_{i=1}^{m} y_{t+i}$$

Similar to running medians, such simple FMH filters assume that the signal is locally constant. Predictive FMH (PFMH) filters use one-sided weighted averages instead of ordinary half-window averages for tracking linear trends (McDonald and Owen, 1986; Heinonen and Neuvo, 1988). Combined FMH filters combine the structures for a local constant and for a local linear signal. However, these filters can only remove single isolated outliers and hence they are not sufficiently robust for applications in intensive care.

Fried, Bernholt and Gather (2006) construct hybrid filters based on RM regression to combine the robustness of the repeated median with the better shift preservation of FMH filters. They investigate several filters, using either the central observation $y_t$ or the median of all observations in the window as a central subfilter. Instead of one-sided means they use one-sided medians

$$\hat{\mu}_t^F = \text{med}\{y_{t-m}, \ldots, y_{t-1}\} \quad \text{and} \quad \hat{\mu}_t^B = \text{med}\{y_{t+1}, \ldots, y_{t+m}\}\,,$$

and instead of the one-sided weighted averages they apply one-sided RM filters

$$\hat{\mu}_t^{RM,F} = \text{med}\{y_{t-m} + m\hat{\beta}_t^{RM,F}, \ldots, y_{t-1} + \hat{\beta}_t^{RM,F}\}\,,$$

where $\hat{\beta}_t^{RM,F}$ is the RM slope estimate based on the observations $y_{t-m}, \ldots, y_{t-1}$ and $\hat{\beta}_t^{RM,B}$ is defined analogously for the other half of the window. Since these subfilters make predictions for the central value, the procedures are called *predictive* (P), or *combined* (C) if both median and RM subfilters are used.

In general, RM-based filters are not affected by trends and attenuate both Gaussian and spiky noise well. The smoothest signal estimations are obtained using the ordinary RM filter, but on the other hand it also smoothes out shifts and trend changes. In contrast, the predictive RM hybrid filter

$$\hat{\mu}_t^{PRMH} = \text{med}\{\hat{\mu}_t^{RM,F},\, y_t,\, \hat{\mu}_t^{RM,B}\}$$

can preserve trend changes and level shifts almost exactly — even within trends — but it attenuates Gaussian noise less efficiently than simple RM filters. Like the other RM hybrid filters it is affected by many outliers. In addition, RM hybrid filters are designed for delayed signal extraction and hence, for online signal extraction, different subfilters have to be applied.

### 3.1.3 Nested Filters

An approach for combining the smoothness of the moving average with the robustness and shift preservation of the running median is given by modified trimmed means (MTM) (Lee and Kassam, 1985). The idea is to calculate the median of all observations in the window and then 'trim', i.e. discard, the observations that deviate by more than a specified multiple of a robust scale estimate, e.g. the median absolute deviation about the median (MAD) (Huber, 1981)

$$\hat{\sigma}_t^{MAD} = \underset{i=-m,\ldots,m}{\text{med}} \left\{ \left| y_{t+i} - \text{med}\{y_{t-m}, \ldots, y_{t+m}\} \right| \right\} \tag{3.4}$$

The arithmetic mean of the remaining observations is then taken as the signal estimate in the centre of the time window. These MTM estimates are robust against outliers and efficient for Gaussian noise. In addition, they can preserve large shifts in an otherwise constant signal level better than ordinary running medians (Himayat and Kassam, 1993). Since the location-based MTM deteriorates in trend periods, Gather and Fried (2004) extend this idea to the trimmed repeated median (TRM). Within each time window an RM regression line is fitted and the MAD is calculated from its residuals to estimate the local variability (Gather and Fried, 2003). Observations deviating by more than a multiple of the residual MAD from the fitted line are trimmed, and the final signal estimate is derived by a least squares fit to the remaining observations. This TRM filter is almost as robust as a variant applying another RM regression in the second step, but it is more efficient for Gaussian errors.

To further improve the preservation of shifts, Bernholt, Fried, Gather and Wegener (2006) use a smaller window width in the first step for the initial RM fit. Because of the nested design of the windows for the first and second regression step, the prefix 'double window' (DW) is added to the estimates, which results in DWMRM and DWTRM. An illustration of such a DWTRM fit to a single window containing $n = 31$ observations is provided by Figure 3.5.

Using this double-window technique considerably improves the performance of RM filters in terms of the preservation of shifts. In general, shifts that are large relative to the observational noise are traced more accurately than smaller shifts.

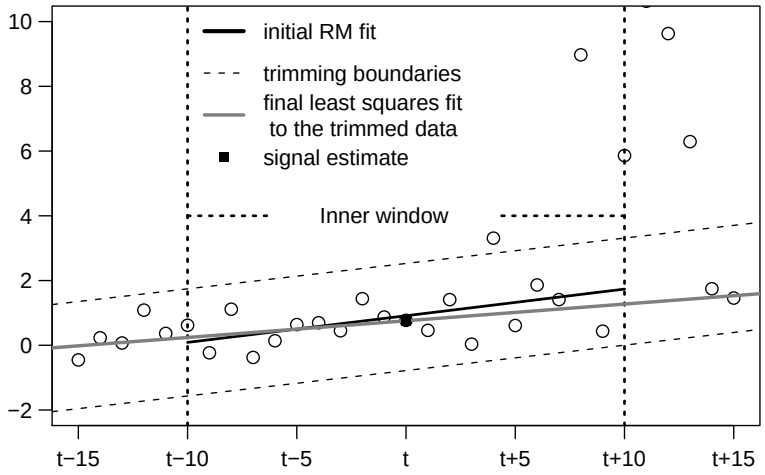


**Figure 3.5**: Fit of the DWTRM filter to a single time window of width $n = 31$. In the second step, only the observations within the trimming boundaries around the RM line are used to calculate the least squares fit.

If the application allows for a relatively large outer-window width, the signal estimation can also be improved by using a short inner window for the initial RM slope estimation and a larger outer window for the level estimation. First experiences show that this DWRM filter seems promising for signal extraction, bearing in mind the demands for robustness and the allowable time delay. However, these methods have not been investigated carefully in full online analysis yet.

### 3.1.4 Weighted Repeated Median Filters

In analogy to the popular weighted median (WM) filters, Fried, Einbeck and Gather (2007) construct weighted Repeated Median (WRM) filters. While WM filters are based on the idea that a constant level is more likely for close-by observations, WRM filters assume that the signal slope is more likely to be the same within short time lags. Suitable symmetric bell-shaped (in delayed or retrospective analysis) or monotonic (in full online analysis) weighting schemes allow the use of longer time windows than ordinary running medians or RM filters that correspond to uniform weights.

Consider a sample $\{(X_j, Y_j)\}$ of size $n$, where the observations of $X_j$ are not necessarily equidistant, and two sets of weights $w_j$, $\tilde{w}_j \in \mathbb{N}$, $j \in \{1, \ldots, n\}$. Analogous to the definitions of the RM slope (2.2) and RM intercept (2.3) the weighted Repeated Median (WRM) estimators for slope and intercept (at design point $x = 0$) are defined by:

$$\hat{\beta}^{WRM} = \underset{j=1,\ldots,n}{\text{med}} \left\{ \tilde{w}_j \diamond \left( \underset{k \neq j}{\text{med}} \; \tilde{w}_k \diamond \frac{Y_k - Y_j}{X_k - X_j} \right) \right\} \tag{3.5}$$

$$\hat{\mu}^{WRM} = \underset{j=1,\ldots,n}{\text{med}} \left\{ \tilde{w}_j \diamond \left( Y_j - \hat{\beta}^{WRM} \cdot X_j \right) \right\} . \tag{3.6}$$

The operator $\diamond$ in (3.5) and (3.6) symbolises replication, i.e. $w_j \diamond Y_j$ means that $Y_j$ is replicated $w_j$ times. In the time series context, the WRM level estimate at time $t$ results from a WRM fit to the observations in a time window of width $n$ corresponding to the sample $\{(j, y_j) : j = t + i\}$ with $i$ defined as in (3.3).

Fried, Einbeck and Gather (2007) compare the WRM filter to $L_1$ and weighted $L_1$ filters. Among other things, their study determines the minimal window width necessary for the investigated methods to resist a certain number $h$ of successive outliers, while taking such deviant values into account if their number is greater than $h$. The reason for this lies in the fact that when moving a time window through a series of measurements, at some point the time series contains $h$ subsequent outliers or 'spikes' (which are still regarded as a sequence of artefacts) while in the subsequent time window the presence of $h + 1$ successive outliers of the same size and sign may already indicate a shift (Imhoff, Bauer, Gather and Fried, 2002). In this way, window widths are determined that allow to track shifts lasting at least $h + 1$ observations while eliminating a smaller number of outliers.

For RM-based filters, weighting improves the adjustment to non-linear trends, allows for larger window widths, and increases the efficiency, while for the $L_1$ filter, weighting can increase robustness and efficiency.

For online signal extraction, the WRM filter tracks shifts better than the $L_1$ filter, which has some difficulties in distinguishing relevant from irrelevant patterns. The weighting reduces the bias of RM, implying that WRM also outperforms the standard RM filter in tracing shifts. In addition, the WRM filter generally shows the smoothest signal estimations, while the $L_1$ filter overshoots shifts and is wiggly. In conclusion, a suitably designed weighted RM filter can be recommended for online signal extraction.

In the retrospective situation, weighted $L_1$ filters provide even better results than the WRM filters. In particular, for moderate outliers, weighted $L_1$ filters show the least biased results, and they also trace large shifts with a smaller time delay. However, if several outlier patches occur close to each other and thus intrude into the same time window, the standard RM filter may still be the best choice because of its maximal breakdown point.

### 3.1.5 Extended Robust Regression Filters

In contrast to LMS filters, RM filters are more vulnerable to large outliers, while they accommodate small outliers well, see e.g. Gather and Fried (2003, 2004). Additionally, large outliers are usually easier to detect than small ones. Therefore, it is worthwhile to add automatic rules for outlier detection and replacement to RM filters to increase the robustness of the signal estimation (Fried, 2004). Likewise, it is possible to apply automatic rules for level shift detection to the RM filters investigated by Davies, Fried and Gather (2004) and Gather, Schettlinger and Fried (2006).

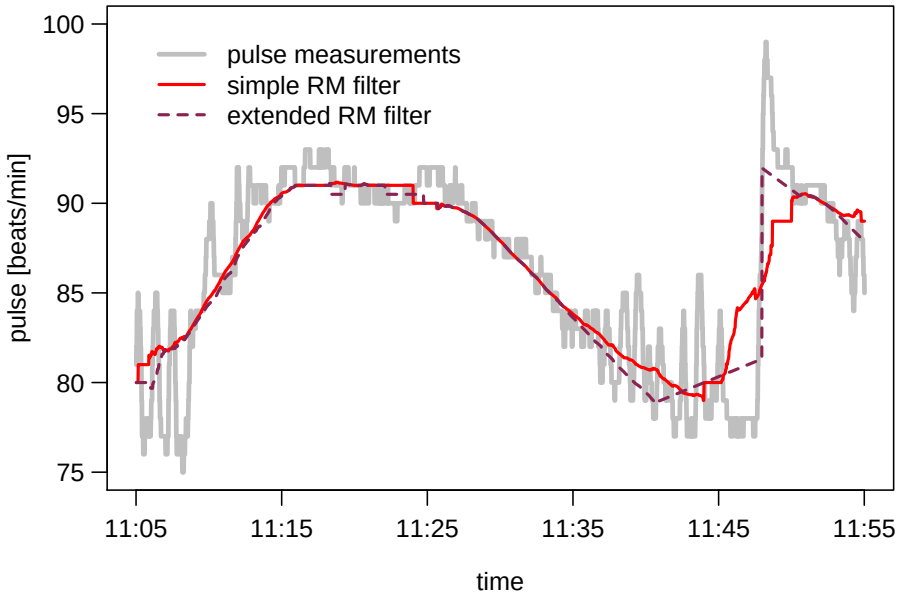


**Figure 3.6**: Comparison of the simple retrospective RM filter with its extended version including outlier and shift detection. Both filters use a moving window containing five minutes of observations.

Similarly to the nested filters approach in Section 3.1.3, an observation is regarded as an outlier if the corresponding absolute deviation from the current regression line is greater than a specified multiple $d$ of a robust scale estimation, i.e. if $|y_{t+i} - \hat{\mu}_{t+i}| > d \cdot \hat{\sigma}_t$. However, here only the next incoming observation is screened for outlier identification before entering the actualised time window by extrapolating the previous regression line. Detected outliers are replaced and no longer considered in the following analysis. In this way they lose their influence on the estimations. For certain 'worst case' scenarios, replacing outliers by the simple extrapolation of the regression line gives better results than other 'down-sizing' replacement strategies at a cost of reduced Gaussian efficiency.

For scale estimation Fried (2004) investigates the advantages of several robust estimators, including the MAD (3.4), the $S_n$ and $Q_n$ estimator of Rousseeuw and Croux (1993) and the 'length of the shortest half' (LSH), see Grübel (1988) or Rousseeuw and Leroy (1988). $Q_n$ and the LSH scale estimator give the best results in case of many large outliers of similar size, but $Q_n$ provides better efficiency, especially when identical measurements occur, e.g. due to rounding.

For shift detection, a simple majority rule is added to the filtering procedure. Considering the most recent $m$ observations in the time window, the number of observations with residuals greater than a certain limit and with same sign is counted. If this number exceeds $m/2$, this indicates a level shift and the procedure moves to the next window that does not overlap the current one. This rule enables the regression filters to detect and thus preserve shifts, and hence it overcomes the biggest disadvantage of the RM filter, see Figure 3.6. In addition, the delay in following shifts decreases, ideally to a minimal

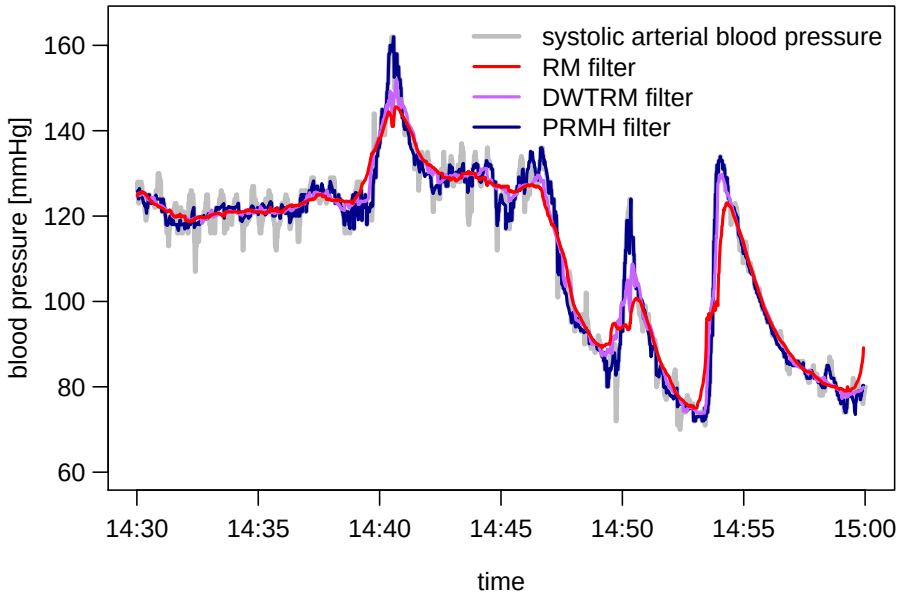


**Figure 3.7**: Comparison of a nested RM filter (DWTRM) and an RM hybrid filter (PRMH) with the simple RM filter in an retrospective application to half an hour of blood pressure measurements recorded once per second. For all filters the moving window contains two minutes of observations.

delay of $\lfloor m/2 \rfloor + 1$ time units. In this context, regression based filters with additional shift-detection rules seem preferable to other shift-preserving procedures such as LMS or FMH filters.

Furthermore, some simple rules can be added to overcome problems in the infrequent case that too many observations are identified as outliers and replaced by 'suitable' values (Fried, 2004). The rules for outlier treatment and shift detection can also be applied for online signal extraction. However, the minimal delay of shift detection cannot be reduced further because of the differentiation of shifts and outlier patches required.

### 3.1.6 Comparisons and Recommendations

The previous sections describe different approaches for filtering methods from which particularly the RM-based filters seem promising for application to online monitoring data from intensive care. However, the choice of the appropriate filter should depend on the characteristics of the underlying signal whenever known.

Summarising the outcomes described above, the following recommendations can be made: for retrospective signal extraction the predictive RM hybrid (PRMH) filter seems to be the best choice if the signal is assumed to contain many jumps and trend changes, while the simple RM filter yields better results if many outliers but no abrupt changes are expected. A compromise between these two methods is given by the DWTRM filter.

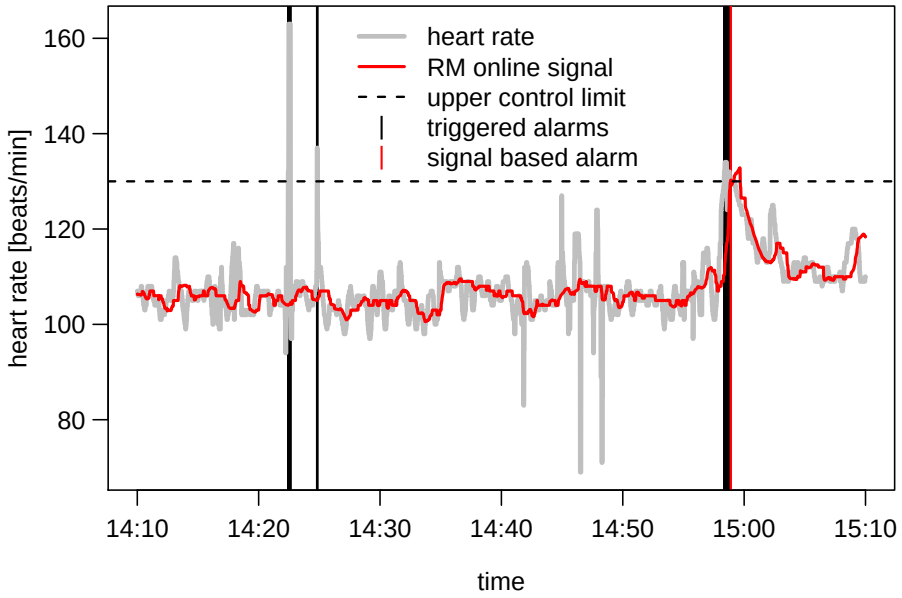


**Figure 3.8**: Alarm reduction by signal extraction. A filter based alarm system does not trigger alarms in case of measurement artefacts . The RM online signal based on a moving window of width $n = 151$ represents the level of the heart rate series well, it resists outliers, and it detects relevant changes.

A comparison of the RM filter with its extended version in Figure 3.6 shows how much a shift detection rule can improve the simple RM filter. However, local extremes, i.e. sudden trend changes, cannot be traced as well with this extended RM filter. In this case, application of the PRMH or the DWTRM filter is more suitable: Figure 3.7 shows that the predictive RM hybrid filter (PRMH) traces sudden shifts and local extremes very accurately. However, the PRMH signal shows the largest variability, especially in relatively constant periods e.g., from 14:30h to 14:40h. The simple RM filter output is the smoothest, but 'blurs out' sudden shifts and 'cuts' local extremes, see e.g. around 14:40h or 14:50h in Figure 3.7. As pointed out previously, the DWTRM signal provides a compromise between the RM and the PRMH filter output. It is smoother than the PRMH signal, but traces trend changes and shifts better than the RM filter.

Real-time applications of these filters imply a time delay of half the window width used for the signal extraction. Therefore, filters have been examined for their online application without any time delay. As shown in Figure 3.4, simple regression filters are suitable for this purpose, but even the online version of the RM filter still possesses some disadvantages, such as the slow reaction to level shifts. Weighted RM filters (WRM) can possibly improve upon simple online RM filters. Another promising and simple approach to overcome this problem is given by online filters which can choose the window width adaptively, depending on the underlying data structure (see Section 3.2).

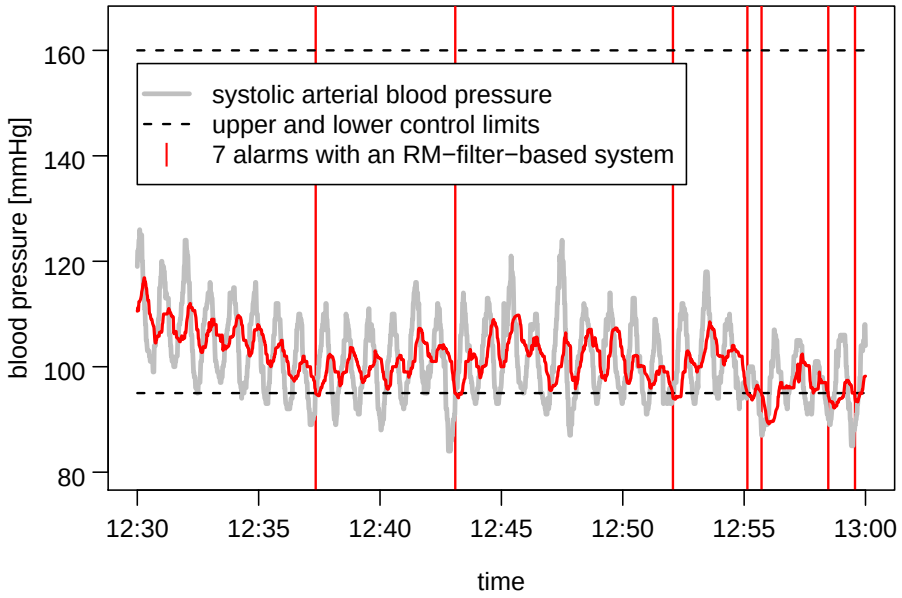


**Figure 3.9**: Alarm reduction by signal extraction. An RM-filter based alarm system can reduce the overall alarm rate drastically. Here, the RM online signal (red) is based on a moving window of width $n = 121$.

Figure 3.8 shows how the comparison of an extracted signal with preset alarms limits can prevent the triggering of false alarms because of (patches of) outliers while relevant alarms are preserved. Furthermore, Figure 3.9 shows that comparing an extracted online signal to the alarm limits can reduce the overall alarm rate.

### 3.1.7 Discussion

The methods recommended here for univariate signal extraction are based on a simple linear regression approach. The ordinary Repeated Median regression filter improves on running medians in trend periods but lacks the property of preserving sudden shifts as such. Arias-Castro and Donoho (2007) discuss that median filtering does not generally preserve edges better than linear filtering. However, for the signal-to-noise ratios we consider in this context, this statement still holds true. Different approaches to overcome the problems of the linear RM regression filter have been proposed and work well for particular situations, but there is no 'universal' procedure without any deficiencies. Double-window TRM filters are promising for delayed signal extraction, while weighted RM filters are possible candidates for online analysis.

Further investigations have shown that median-based filters are also robust against autocorrelations. Compared to procedures based on least squares, robust location or regression filters based on the median, trimmed means or RM show an increase in relative efficiency in the frequent case of positive correlations. In this case they also outperform filters incorporating autocorrelations explicitly into the analysis, see e.g. Fried (2007b) or Fried and

Gather (2005). In the infrequent situation of strong negative autocorrelations, a Prais-Winsten transformation of the data is worthwhile and improves the ordinary RM filter.
With the linear time RM update algorithm developed by Bernholt and Fried (2003) or its advancement to a linear storage algorithm (Fried, Bernholt and Gather, 2006), the RM filter is computationally feasible for high-frequency data. This update algorithm also implies a linear computation time for all the recommended filters. Outlier and shift detection as described above does not add further computation time when using, e.g. an $\mathcal{O}(\log n)$ MAD update algorithm (Bernholt, Fried, Gather and Wegener, 2006) for the scale estimation. Thus, an RM filter with such extensions represents an acceptable choice for signal extraction.
Similar to the remedian approach of Rousseeuw and Bassett, Jr. (1990), Fried and Gather (2002) propose to improve the computational speed by dividing the time window into $n_2$ disjoint segments, each of length $n_1$. The level within each segment is then estimated by an ordinary median or by repeated median regression. Then RM or another procedure can be applied to this pre-processed output window of width $n_2$. Hence, the computation time can be shrunk by a factor of $n_1$ when using a linear time algorithm and by $n_1^2$ when using an algorithm taking quadratic time. For retrospective analyses, computation times are not as crucial as for real-time applications, but they are still important because of the possible magnitude of the data sets.
All filters described in this section have been implemented in the `robfilter` package of the free statistics software R (Fried and Schettlinger, 2008) and are available on the CRAN server:

`http://cran.r-project.org/web/packages/robfilter/`.

For the filtering procedures described above, the window width $n$ has been assumed to be fixed throughout. The suitable choice of this width is no trivial task and depends on statistical demands as well as the user-oriented background. Larger window widths generally imply a smoother filter output, but they also increase the bias.
The following section proposes a filtering procedure for online signal extraction which is able to choose the window width data-adaptively and thus, does not require the specification of a 'suitable' window width prior to the analysis, i.e. when the data are still unknown. This is indeed a very important prerequisite for the real-time application of a filter, and especially for the application to online monitoring time series in intensive care.

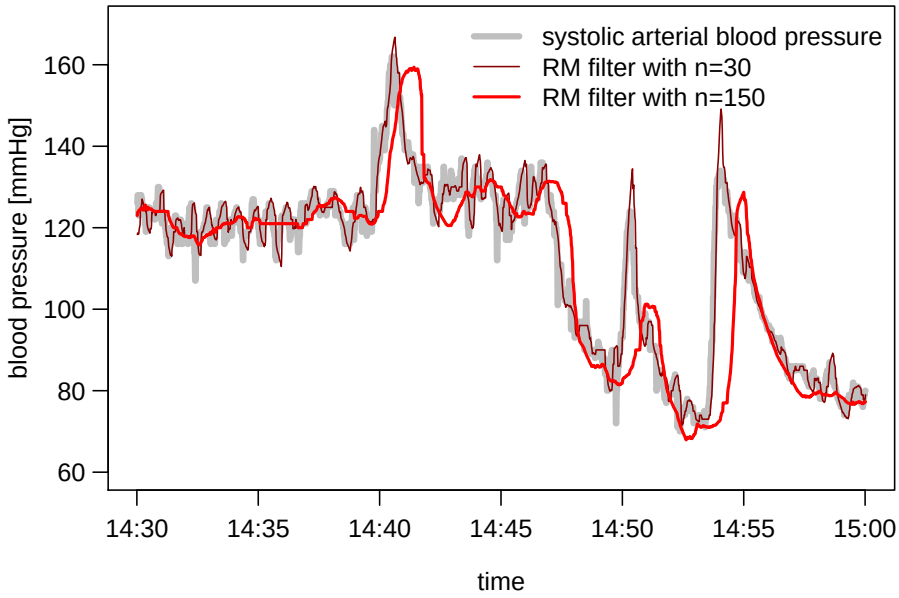


**Figure 3.10**: Influence of the window width on RM regression filters: an online signal extracted by a filter with a small window width of $n = 30$ yields estimations which trace sudden shifts and trend changes with small delay. An online RM filter using a large window width of $n = 150$ displays a much larger delay at such changes, but it achieves a much smoother signal estimation.

## 3.2 The Univariate Adaptive Online Repeated Median Filter

The choice of the window width $n$ for regression filters as described in the previous section can have a large impact on the extracted signal. Larger window widths assure robustness against a larger number of outlying values and result in smaller variability of the estimations and thus, yield a smoother extracted signal (see Figure 3.10). However, such filters also result in a larger bias for the estimated signal and larger computation times. A smaller window width leads to a smaller time delay in tracing sudden changes in the signal level but at the expense of higher variability (see also Figure 3.10).

The choice of the 'optimal' window width or bandwidth, respectively, and the corresponding bias-variance trade-off is a problem that also arises in kernel density estimation and non-parametric regression. It is well-known that methods based on local bandwidth selection can adapt better to the structure of the underlying function than methods using a fixed bandwidth, see e.g. Brockmann, Gasser and Herrmann (1993) or Fan, Hall, Martin and Patil (1996). Various *data-adaptive* choices of the 'optimal' window width for local polynomial fitting have been described by Friedman and Stützle (1981), Fan and Gijbels (1995), Fan and Gijbels (1996), Cleveland and Loader (1996), and Ruppert (1997), but none of these methods is robust against (patches of) outliers. The approach described in Goldenshluger and Nemirovski (1999) and Katkovnik (1999) can also be used for robust

filters (see Katkovnik, Egiazarian and Shmulevich (2001) and Chan and Zhang (2004)), but it is computer-intensive and hence not reasonable for online signal extraction in intensive care. Since the online RM filter outperforms other filters for intensive care applications (Gather, Schettlinger and Fried, 2006), we focus on a data-driven window width selection especially for this real-time filter, which can cope with trended data and is robust against outlying values.

In accordance with Schettlinger, Fried and Gather (2008) we will introduce a version of an online RM filter which is able to chose the window width $n$ adaptively, depending on the underlying data structure. Therefore, a goodness-of-fit test is applied to judge the appropriateness of the window width at each point in time. Several versions of this adaptive procedure are compared with a simulation study and the best settings are determined. We also propose some modifications of this adaptive online RM filter designed to meet practical demands. Furthermore, some examples illustrate the performance of the modified adaptive filters in applications to simulated and real time series from intensive care.

### 3.2.1 Sketch of the Adaptive Procedure

In Gather and Fried (2004) a procedure with a local window width choice for a delayed RM filter is proposed using the balance of residual signs (2.6) for assessing the model fit in each window: after fitting an RM regression to the data in the window, the total number of positive residuals $T^+$ and negative residuals $T^-$ within the first and the last quarter of the window is evaluated. The window width is reduced and the RM fit is repeated if either $T^+$ or $T^-$ is too small or too large compared to the expected number of residuals with the same sign in the described subset under a symmetric error distribution, which corresponds to a value of about $n/4$. However, this approach is based on a symmetric window, estimating the signal in the centre of the current time window and, if necessary, reducing the window by discarding both, the first and the last value in the window. This is not reasonable for online estimation because it may cause the elimination of the most recent observation.

The adaptive online filters presented here are based on the simple online RM filter estimating the signal in a sliding time window by (2.3) with the difference that the window width is not fixed but can vary over time. Therefore, we denote the window width for estimation at time $t$ by $n_t$.

The fundamental idea of the adaptive window width selection is as follows: if the current fit *is not* considered adequate, then the window width used to determine the fit is reduced, and the RM fit is re-estimated at the smaller sample. If the current fit *is* considered adequate, then the window width can be enlarged for the estimation at the next point in time. Figure 3.11 shows an outline of such an adaptive online RM filter.

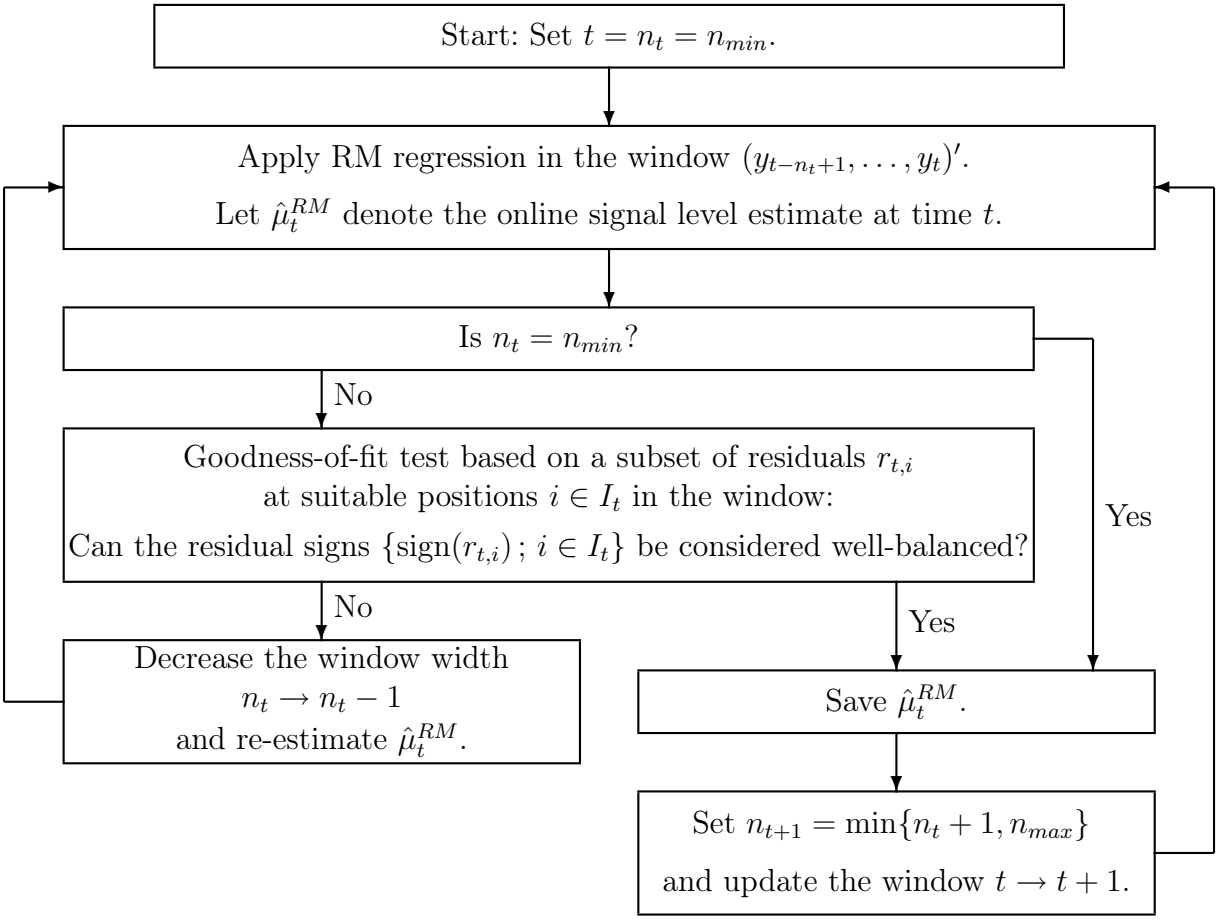


**Figure 3.11**: Flow chart of the adaptive online Repeated Median filter.

For application of this filter, the minimum and maximum window width have to be chosen by the user. The minimum window width $n_{min}$ controls how many outliers the filter can handle (within each window) and hence controls robustness. Furthermore, the smallest window should contain enough observations for a 'sensible' estimation. In intensive care, up to five subsequent outliers can occur without indicating a relevant change (Gather, Fried and Imhoff, 2000). To ensure a minimum robustness against such situations we choose $n_{min} = 11$ in the following.

The maximum window width $n_{max}$ should be chosen such that our working assumption (3.2) of local linearity of the signal is still appropriate. The computation time is also limited by the specification of $n_{max}$ because of the increasing computational demand with increasing window width. Moreover, considering that the proposed filter is intended for real-time application, the current time window should not contain information which seems 'too old' for the current estimation. With the background that a patient's health status can change within relatively short time, we choose $n_{max} = 121$ in the following, such that the largest window in applications to intensive care time series evaluated per second includes the current time and the preceding two minutes.

Here, the first signal estimation takes place when at least $n_{min}$ observations are present, i.e. the first time and the initial width for the first window is set to $t = n_t = n_{min}$. However, it is also possible to start with the online signal estimation at a later time using a (possibly) larger window width $t = n_{start} \in \{n_{min}, \ldots, n_{max}\}$.
For each time, the estimation of the current signal level is repeated, either until it cannot be rejected that the current estimate is appropriate, or until the window width cannot be reduced any more, i.e. $n_t = n_{min}$. If the estimation is considered suitable in the first place, no further iterations need to be performed resulting in a considerable gain of computation time compared to methods requiring more than one pass over the data, see e.g. Katkovnik, Egiazarian and Shmulevich (2001).
The core of this procedure consists of the goodness-of-fit test for checking the adequacy of the current signal estimate, which raises two main questions:

1. Which test should be used?
2. Which subset of residuals should be regarded for the test?

The first question will be addressed in the next section; the second question will be discussed thereafter, motivating the simulation study described in Section 3.2.6.

### 3.2.2 GOODNESS OF FIT TEST

Let $r_{t,i}$ for $i = 1, \ldots, n_t$ denote the residuals (2.7) in the window used for estimating the signal level at time $t$. Given the balance of the residual signs in the whole window (2.6), the proposed goodness-of-fit test is based on the idea that the balance of the signs of the residuals at certain points in time $i \in I_t \subset \{1, \ldots, n_t\}$ represents the adequacy of the fit at the last time point in the window: if either the negative or the positive residual signs prevail, the fit at the end of the window is not considered adequate and the window width has to be adjusted.
Testing the sign balance of the residuals $\{r_{t,i}\,;\, i \in I_t\}$ can be done using a sign test for the location of the error distribution in the selection. Here, we use $\tilde{\mu}_e^{I_t}$, the median of the distribution of the errors at points in time included in the selection $I_t$, as a measure of location, and we consider the test problem

$$H_0 : \tilde{\mu}_e^{I_t} = 0 \qquad \text{vs.} \qquad H_1 : \tilde{\mu}_e^{I_t} \neq 0\,. \tag{3.7}$$

In Section 2.2 it is proved that an RM residual can equal the exact value of zero with positive probability – not only for odd but also for even sample sizes or window widths, respectively. Thus, a test statistic consisting of the number of either all non-negative or non-positive residuals is not appropriate here, because it is nonspecific about the handling of residuals with value zero, and counting the zero as non-negative or as non-positive can make a relevant difference for the test decision.

Furthermore, the removal of zero residuals from the sample and thus the reduction of the sample size by the number of zero residuals does not provide a satisfying solution: in the extreme case of an exact fit (2.5), the number of zero residuals outweighs the number of residuals with positive or negative sign. In that case the regression line fits the majority of the data exactly, but by using a test statistic disregarding zero residuals, the regression fit would be judged solely by the residuals which do not lie on this line and hence, by the minority of the data. In intensive care an exact fit is not a rare event because the discreteness of the measurement scale can lead to collinear data.
Consequently, we choose a test statistic which takes into account that zero residuals support the null hypothesis, and we define the test statistic as

$$T = \sum_{i \in I_t} \operatorname{sign}(r_{t,i}) \tag{3.8}$$

where $r_{t,i}$ denote the RM residuals with $i \in I_t \subset \{1, \dots, n_t\}$, and the sign function is defined as in (2.8). If $|T|$ is large, either the positive or the negative residuals in the chosen subset predominate. Hence, we reject the null hypothesis that the current signal estimation is appropriate if

$$|T| > c(n_t, I_t)\,, \tag{3.9}$$

where $c(n_t, I_t)$ denotes a critical value depending on the window width $n_t$ and the subset selection $I_t$.
An intuitive choice for a critical value consists of a quantile of the distribution of the test statistic $T$ under the null. Therefore, the following section discusses the derivation of this unknown distribution. Particular critical values are derived in Section 3.2.5.

### 3.2.3 Distribution of the Test Statistic

Denote by

- $n_t$ the width of the current window,
- $k_t \in \{0, 1, 2\}$ the observed number of zero residuals in the current window,
- $I_t \subset \{1, \dots, n_t\}$ the index set indicating the subset of RM residuals whose signs are added up to evaluate the value of the test statistic $T$,
- $n_{I_t} < n_t$ the number of residual signs added for $T$, i.e. the cardinal number of $I_t$,
- $k_{I_t} \in \{0, \dots, k_t\}$ the number of zero residuals in the subset of the $n_{I_t}$ RM residuals at times $i \in I_t$.

The corresponding random variables are denoted by capital letters.

| | $P(K_t = 0)$ | $P(K_t = 1)$ | $P(K_t = 2)$ |
|---|---|---|---|
| $n_t = 3$ | 0 | 0 | 1 |
| $n_t = 4$ | 0.6755 | 0 | 0.3245 |
| $n_t = 5$ | 0 | 0.7338 | 0.2662 |

**Table 3.2**: Probabilities for no, one or two zero residuals resulting from an RM regression fit in a window containing $n_t \in \{3, 4, 5\}$ observations from a standard normal distribution.

In particular for small window widths, the values of $k_t$ and $k_{I_t}$ have a large impact on the distribution of $T$. Therefore, we decompose the density of this distribution into conditional probabilities, depending on the number of zero RM residuals. According to the total probability theorem, the probability for the test statistic $T$ taking on a value $z \in \{-n_{I_t}, \ldots, n_{I_t}\}$ can be written as

$$P\Big(T = z\Big) = \sum_{k_t=0}^{2} \Big( \sum_{k_{I_t}=0}^{k_t} P\big(T = z \big| k_{I_t}, k_t\big) \cdot P(K_{I_t} = k_{I_t} | k_t) \Big) \cdot P(K_t = k_t)\,. \tag{3.10}$$

Under the normality assumption, the term $P(K_t = k_t)$ can be derived using the results from Section 2.2. Table 3.2 lists the exact probabilities for $n_t \in \{3, 4, 5\}$. For $n_t > 5$, simulations are performed in Section 2.2 which lead for $P(K_t = 2)$ to approximation formulas (2.14) (for an even window width $n_t$) and (2.15) (for an odd window width). It is also shown there that for an even window width it is $P(K_t = 1) = 0$ and $P(K_t = 0) = 1 - P(K_t = 2)$, and for odd $n_t$ it is $P(K_t = 1) = 1 - P(K_t = 2)$ and $P(K_t = 0) = 0$. For increasing window width or sample size, respectively, Figures 2.2 and 2.4 in Section 2.2 show that $P(K_t = 2)$ tends to zero. Thus, for sufficiently large window widths ($n_t > 100$) we can take

- $P(K_t = 0) \approx 1$ for $n_t$ even, and
- $P(K_t = 1) \approx 1$ for $n_t$ odd.

(3.11)

In Section 2.2 it is also shown that, for standard normal data, zero residuals are more likely to arise from the first or last observations in a window (see e.g. Figure 2.5) meaning that zero residuals do not occur at each time with equal probability. In conclusion, the probability $P(K_{I_t} = k_{I_t} | k_t)$ for a certain number $k_{I_t}$ of zero residuals in the chosen subset strongly depends on the choice of the subset $I_t$.

As an example, Table 3.3 gives the probabilities $P(K_{I_t} = k_{I_t} | k_t)$ for two different subsets out of a window of width $n_t = 5$ under the normality assumption. For the first subset indicated by $I_t = \{1, 5\}$, i.e. the choice of the first and last RM residual in the window, it is very likely to observe two zero residuals in that subset while it is impossible when choosing the subset indicated by $I_t = \{3, 4\}$. Since for every window width $n_t$ there is a

| | $P(\cdot \mid K_t = 1)$ | | $P(\cdot \mid K_t = 2)$ | | |
|---|---|---|---|---|---|
| | $P(K_{I_t}=0 \mid 1)$ | $P(K_{I_t}=1 \mid 1)$ | $P(K_{I_t}=0 \mid 2)$ | $P(K_{I_t}=1 \mid 2)$ | $P(K_{I_t}=2 \mid 2)$ |
| $I_t = \{1, 5\}$ | 0.5350 | 0.4651 | 0.1668 | 0.2071 | 0.6260 |
| $I_t = \{3, 4\}$ | 0.5303 | 0.4698 | 0.6260 | 0.3738 | 0 |

**Table 3.3**: Probabilities for zero, one or two zero RM residuals in two different subsets of size $n_{I_t} = 2$ resulting from an RM regression fit to a window containing $n_t = 5$ observations from a standard normal distribution, conditioned on the total number $k_t \in \{1, 2\}$ of zero residuals in the whole window.

multitude of possible choices of a subset $I_t$ containing $n_{I_t}$ values, the following Section 3.2.4 discusses reasonable possibilities for $I_t$ while in Section 3.2.6 the best choice of $I_t$ and $n_{I_t}$ for the proposed adaptive procedure is deduced from a simulation study.
If $n_t$ is large enough for assumption (3.11) to hold, then $P(K_{I_t} = k_{I_t} | K_t = k_t)$ approximately corresponds to

$$P\big(K_{I_t} = k_{I_t} \,\big|\, K_t = 0\big) \;=\; P\big(K_{I_t} = 0 \,\big|\, K_t = 0\big) \;=\; 1 \quad \text{for } n_t \text{ even.} \tag{3.12}$$

Although it is always more likely to observe a zero residual close to the boundaries of a window (for any window width), the probability for a particular location of a zero residual is the closer to $1/n_t$, the larger the window width is (see e.g. Figure 2.5). Thus, for sufficiently large odd $n_t$ we can assume that the difference of a uniform distribution to the distribution of the location of one zero residual is negligible. In that case a zero residual can occur at any position in the subset $I_t$ approximately with equal probability and we can approximate $P(K_{I_t} = k_{I_t} | K_t = k_t)$ by

$$P\big(K_{I_t} = k_{I_t} \,\big|\, K_t = 1\big) \;\approx\; \begin{cases} \frac{n_{I_t}}{n_t} & \text{if } k_{I_t} = 1 \\ 1 - \frac{n_{I_t}}{n_t} & \text{if } k_{I_t} = 0 \end{cases} \quad \text{for } n_t \text{ odd.} \tag{3.13}$$

Section 2.3 also discusses the correlation structure between the signs of RM residuals which is similar to the one of least squares residuals: because of some sort of leverage effect, the correlation between the signs of RM residuals is the stronger, the closer the residuals lie to the boundaries of the window. However, it can also be seen that the correlations are the weaker the larger the sample size (see e.g. Figure 2.6). Thus, while this definitely does not hold for small samples, for sufficiently large window widths we can assume that the dependence between the signs of the RM residuals is weak.
Assuming independence for the residual signs, and assuming that a positive and negative sign are equally likely, we can also derive an approximation of $P(T = z | k_t, k_{I_t})$ for large window widths. Let

$$T^+ = \sum_{i \in I_t} \mathbb{1}_{\{1\}}(\text{sign}(r_{t,i})) \qquad \text{and} \qquad T^- = \sum_{i \in I_t} \mathbb{1}_{\{-1\}}(\text{sign}(r_{t,i}))$$

denote the number of strictly positive or strictly negative residuals, respectively, in the selection with indices $i \in I_t$. Here, $\mathbb{1}$ denotes the indicator function, i.e. for a set $B \subset \mathbb{R}$ and a real value $x \in \mathbb{R}$ it is

$$\mathbb{1}_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases} .$$

With the assumptions stated above, for large $n_t$ we can assume that the number of positive RM residuals $T^+$ follows approximately a hypergeometric distribution, i.e.

$$T^+ \quad \sim \quad \text{Hyper}(n_t - k_t, \frac{n_t - k_t}{2}, n_{I_t} - k_{I_t}) . \tag{3.14}$$

Furthermore, the number of non-zero residuals in the subset indicated by $I_t$ corresponds to $T^+ + T^- = n_{I_t} - k_{I_t}$, and we can use this relation to rewrite the test statistic (3.8) as

$$T \;=\; T^+ - T^- \;=\; 2T^+ - (n_{I_t} - k_{I_t}) .$$

Using this relation and the approximative hypergeometric distribution of $T^+$ (3.14), we derive an approximation for the density of $T$ via the total probability theorem (3.10). According to (3.11) and (3.12) for large and even window width $n_t$ $P(K_{I_t} = k_{I_t}|k_t)$ and $P(K_t = k_t)$ cancel out from formula (3.10), and only $P(T = z) = P(T = z|K_t = 0, K_{I_t} = 0)$ remains. Now let $S_{\text{even}}(n_{I_t}) := \{-n_{I_t}, -n_{I_t} + 2, \ldots, n_{I_t} - 2, n_{I_t}\}$. An approximation of the density of the test statistic $T$ is then given by

$$P\Big(T = z\Big) \approx \frac{\binom{\frac{n_t}{2}}{\frac{n_{I_t}+z}{2}} \binom{\frac{n_t}{2}}{\frac{n_{I_t}-z}{2}}}{\binom{n_t}{n_{I_t}}} \cdot \mathbb{1}_{S_{\text{even}}(n_{I_t})}(z) . \tag{3.15}$$

For large, odd window width $n_t$, using the approximations specified in (3.11) and (3.13), and letting $S_{\text{odd}}(n_{I_t}, k_{I_t}) := \{-(n_{I_t} - k_{I_t}), -(n_{I_t} - k_{I_t}) + 2, \ldots, (n_{I_t} - k_{I_t}) - 2, (n_{I_t} - k_{I_t})\}$, we get the following approximation for the density of $T$:

$$\begin{aligned} P\Big(T = z\Big) \quad \approx \quad & \frac{\binom{\frac{n_t-1}{2}}{\frac{n_{I_t}+z}{2}} \binom{\frac{n_t-1}{2}}{\frac{n_{I_t}-z}{2}}}{\binom{n_t - 1}{n_{I_t}}} \cdot \mathbb{1}_{S_{\text{odd}}(n_{I_t},0)}(z) \;\cdot\; \left(1 - \frac{n_{I_t}}{n_t}\right) \\ + \quad & \frac{\binom{\frac{n_t-1}{2}}{\frac{(n_{I_t}-1)+z}{2}} \binom{\frac{n_t-1}{2}}{\frac{(n_{I_t}-1)-z}{2}}}{\binom{n_t - 1}{n_{I_t} - 1}} \cdot \mathbb{1}_{S_{\text{odd}}(n_{I_t},1)}(z) \;\cdot\; \frac{n_{I_t}}{n_t} . \end{aligned} \tag{3.16}$$

The approximations (3.15) and (3.16) for the density of $T$ for large window widths only depend on the window width $n_t$ itself and the number of residual signs $n_{I_t}$ which enter $T$.

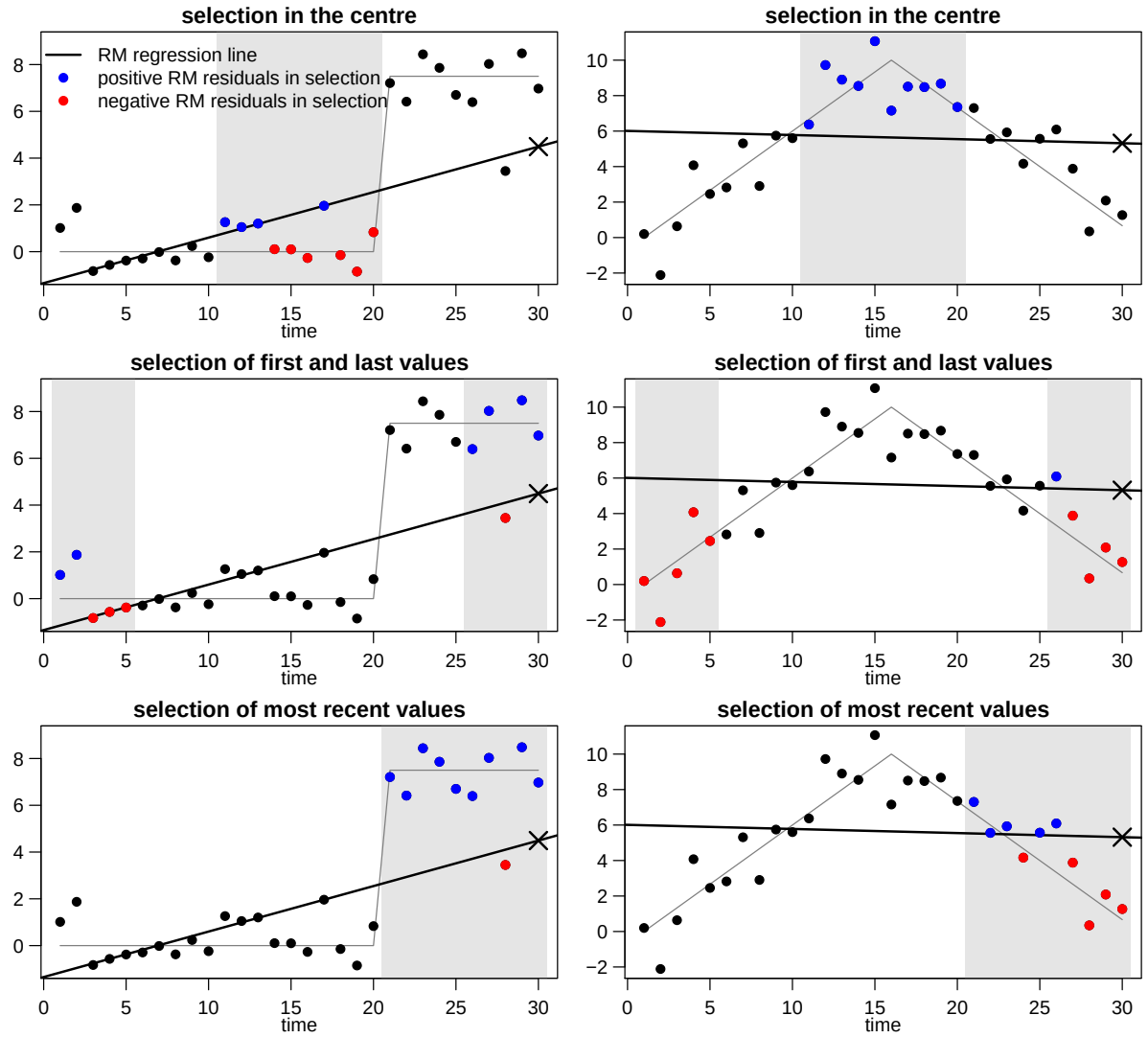


**Figure 3.12**: RM regression line with indication of the sign of the RM residuals in different subsets of size $n_{I_t} = 10$ in a window of width $n_t = 30$ containing a level shift (left) and a trend change (right). The RM online estimate ($\times$) deviates strongly from the true signal (grey line).

### 3.2.4 Subset Choice for the Test Statistic

The subset $I_t$ should not be chosen at random but as a representative selection for testing the goodness of the fit at the most recent time $t$. Therefore, we will discuss possible choices of $I_t$ and the consequences for the distribution of the test statistic.

The current window width needs to be adjusted particularly if the data in the window exhibit a pattern like a level shift or trend change. Therefore, the residual signs selected to calculate the test statistic should reflect such data structures. The left panels in Figure 3.12 show a window of width $n = 30$ containing a positive level shift at time $i = 21$; the right panels show a window with a change of the (linear) trend at time $i = 16$. In both cases, the online RM signal estimate deviates strongly from the true signal at the most recent time. Therefore, $T$ should be large in absolute value such that the null hypothesis (speaking for a 'good fit') can be rejected and the window width can be reduced.

Figure 3.12 exemplifies that the choice of the subset used for calculating the test statistic can have a large impact on the ability of the filter to adapt to certain data structures: For the level shift (left panels of Figure 3.12), choosing the most recent residual signs seems the best choice because in this subset there are mainly positive residuals resulting in a large test statistic, pointing at a bad fit at the most recent time. In the other subsets the prevalence of positive residuals might not be strong enough to indicate that the window width needs to be reduced. In case of a trend change (right panels of Figure 3.12), the residual signs in a subset containing the most recent values are almost balanced and thus would not lead to a window width reduction. In contrast, a subset containing residuals in the centre or from both ends of the window would yield a large absolute value of $T$ and hence would yield a detection of the misfit.

Seeing that different selections $I_t$ are advantageous in different data situations, we want to find the selection which results in an adaptive online RM filter yielding the least biased signal estimates after a sudden change by a simulation study in Section 3.2.6. In Section 3.2.3 it was already pointed out that because of the correlation structure of the signs of RM residuals the choice of the subset selection $I_t$ is crucial for the distribution of the test statistic $T$ (3.8). Of course, the number $n_{I_t}$ of residuals within the subset $I_t$ is also essential for the distribution of the test statistic and thus, the best choice for this number is also investigated in the simulation study in Section 3.2.6.

Simulations with independent standard normal data show that, if $I_t$ contains $n_{I_t}$ subsequent values, the distribution is the more concentrated around zero, the closer the subset is to the edge of the window (see Figure 3.13) which is due to the stronger correlations of the RM residual signs close to the boundaries of the window. For comparison Figure 3.13 also shows the large sample approximation of the density of $T$ according to equation (3.16), which – even for this small sample size of $n_t = 11$ – is reasonable if the subset $I_t$ indicates values close to the centre of the window.

Based on preliminary experiments, we only consider those subsets in the following which we think provide the most useful information about the fit at the most recent time. In particular, we consider the subsets such that the test statistic (3.8) consists either of the sum of the $n_{I_t}$ most recent residual signs; of residual signs from the beginning and the end of the time window; or of residual signs from the centre as e.g. illustrated in Figure 3.12. We define these subsets of times $I_t \subset \{1, \ldots, n_t\}$ by

- $I_t^{centre} = \left\{ \lfloor \frac{n_t - n_{I_t} + 1}{2} \rfloor + 1, \ldots, \lfloor \frac{n_t - n_{I_t} + 1}{2} \rfloor + n_{I_t} \right\}$,
- $I_t^{firstlast} = \left\{ 1, \ldots, \lfloor n_{I_t}/2 \rfloor, \; n_t - \lceil n_{I_t}/2 \rceil + 1, \ldots, n_t \right\}$ , and
- $I_t^{recent} = \left\{ n_t - n_{I_t} + 1, \ldots, n_t \right\}$ .

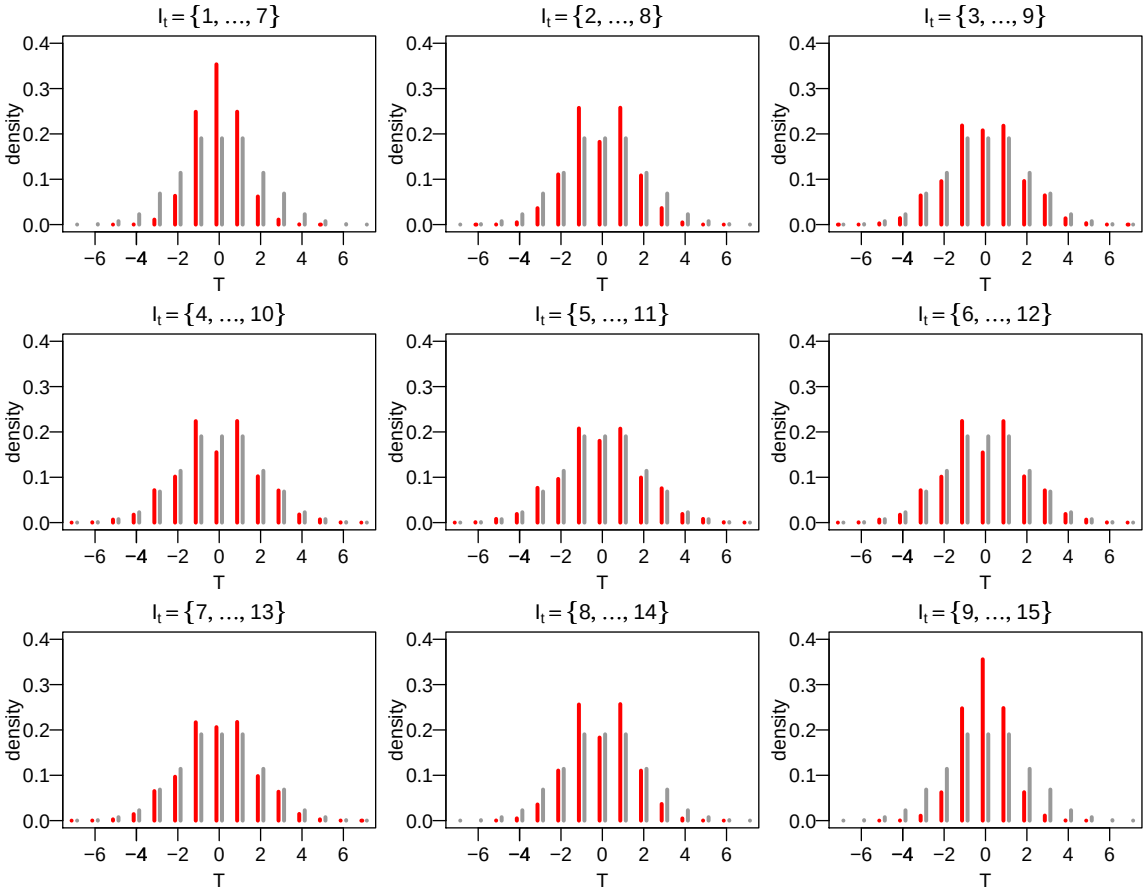


**Figure 3.13**: Empirical density of the distribution of the test statistic $T$ for all possible subsets $I_t$ indicating a set of $n_{I_t} = 7$ consecutive out of $n_t = 15$ RM residuals (red) for standard normal data. For reference, the (large sample) approximation for this density (3.16) appears in the background (grey).

### 3.2.5 Critical Values

To approximate the distribution of $T$ for the subset selections $I_t$ described in the previous section, we simulate 100 000 windows containing $n_t \in \{11, 12, \ldots, 121\}$ observations from a standard normal distribution. Generating the data from a distribution with heavier tails (like Cauchy) or a skewed distribution (like a lognormal) results in no significant difference in the distribution of $T$ as the test statistic only considers the signs of RM residuals. This similarity can be seen by comparing Figure 3.13 to Figures C.1 and C.2 in Appendix C which show the empirical density of $T$ derived from Cauchy and from lognormal data with zero mean and unit variance. Although there are slight differences in the empirical densities based on simulated data coming from different distributions, these differences are apparently very small – even for this small sample size.

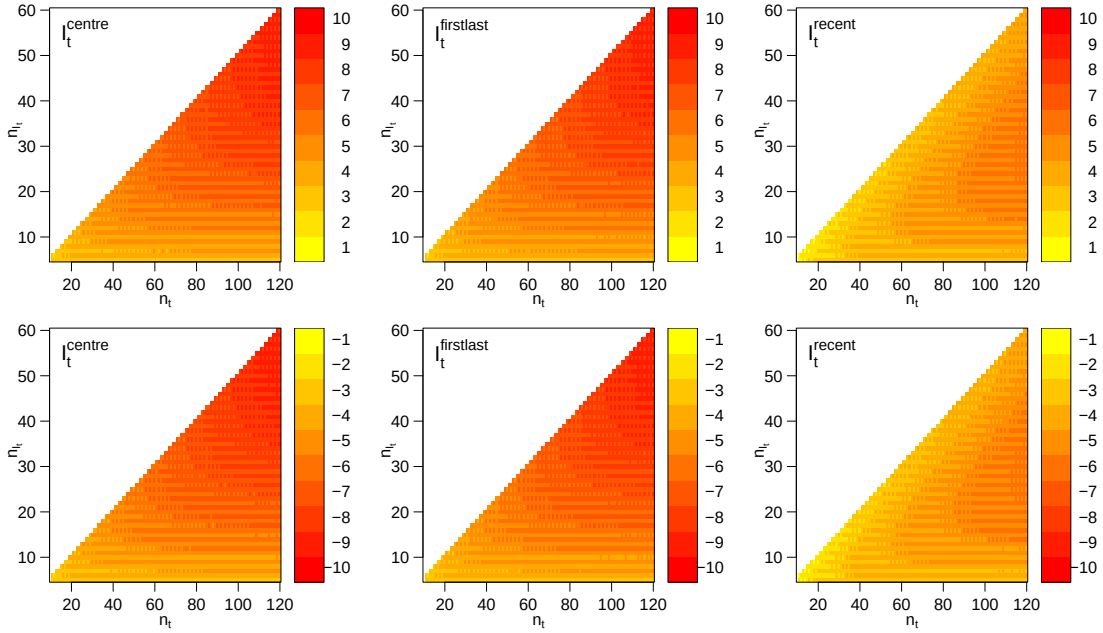


**Figure 3.14**: Empirical 0.95-quantiles (top) and 0.05-quantiles (bottom) of the simulated distribution of $T$ based on standard normal data for the subsets choosing RM residuals in the centre (left), in the first and last part of a window (middle) and subsets choosing the most recent RM residuals (right). These quantiles are neither monotone for increasing window width $n_t$ nor for increasing subset size $n_{I_t}$ and hence, they are not suitable as critical values.

As critical values $c(n_t, I_t)$ we investigate the empirical $p$-quantiles $q_p(n_t, I_t)$ of the simulated distributions of $T$ which depend on the size of the sample $n_t$, on the location of the chosen residuals, indicated by $I_t$, and the number of chosen residual signs $n_{I_t}$. Since further simulations suggest that choosing an approximate level of 0.1 for the test results in a good performance of the proposed adaptive online filter, we restrict the following descriptions to the 0.05- and 0.95-quantiles. The results achieved for the 0.025- and 0.975-quantiles are accordingly.

Figure 3.14 shows the 0.95-quantiles of the empirical distributions of $T$ resulting from the simulations for $n_t \in \{11, 12, \ldots, 121\}$ and $n_{I_t} \in \{5, \ldots, \lfloor n_t/2 \rfloor\}$. These empirical quantiles are neither monotonically increasing for increasing window width $n_t$ and a fixed subset size $n_{I_t}$; nor for increasing $n_{I_t}$ and a fixed window width; nor for both $n_t$ and $n_{I_t}$ increasing at the same time. Actually the montonicity is not even given when considering solely odd window widths or only even window widths, respectively. Thus, taking these simulated 0.95-quantiles as critical values for the test can lead to inconsistencies in the test decisions. Consider an exact fit situation in a window of size $n_t$ where only the last three observations at $t$, $t-1$ and $t-2$ deviate from the exact fit and produce positive residuals. For the test statistic applying $I_t^{recent}$ it is always $T = 3$, because all except the last three residual signs equal the value zero.

| $n_t$ | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
|---|---|---|---|---|---|---|---|---|
| $n_{I_t}$ | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $q_{0.95}(n_t, n_{I_t})$ | 3 | 2 | 3 | 2 | 3 | 2 | 2 | 2 |

**Table 3.4**: Non-monotonicity of the simulated 0.95-quantiles for the distribution of $T$ applying $I_t^{recent}$ for fixed $n_t$ and increasing $n_{I_t}$.

| $n_t$ | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 |
|---|---|---|---|---|---|---|---|---|
| $n_{I_t}$ | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| $q_{0.95}(n_t, n_{I_t})$ | 2 | 2 | 3 | 2 | 3 | 2 | 3 | 4 |

**Table 3.5**: Non-monotonicity of the simulated 0.95-quantiles for the distribution of $T$ applying $I_t^{recent}$ for increasing $n_t$ and fixed $n_{I_t}$.

Tables 3.4 and 3.5 provide some values of the 0.95 quantiles for the distribution of the test statistic using $I_t^{recent}$, i.e. adding up the most recent residual signs, shown in the top right panel of Figure 3.14. When adding the most recent $n_{I_t} = 5$ residual signs out of $n_t = 25$, Table 3.4 shows that the null hypothesis can not be rejected at the 0.1-level for such an exact fit situation, although three out of the five considered residual signs differ from zero, because the 0.95-quantile $q_{0.95}(25, 5)$ corresponds to three. However, because of $q_{0.95}(25, 12) = 2$, using a test statistic adding up $n_{I_t} = 12$ residual signs would lead to rejection even though the fraction of non-zero residuals in the considered subset is much smaller.

The same discrepancy arises for a fixed value of $n_{I_t}$ and increasing sample size $n_t$: when considering the most recent $n_{I_t} = 10$ residual signs for the test statistic, Table 3.5 shows that because of $q_{0.95}(29, 10) = 3$ the null hypothesis can not be rejected for three out of $n_t = 29$ non-zero residual signs while it would be rejected if the last three out of $n_t = 32$ signs are positive, since $q_{0.95}(32, 10) = 2$.

This lack of monotonicity is not due to the simulations but also appears for exact quantiles of discrete distributions which also causes the quantiles for the approximative distribution (3.15, 3.16) to lack this property. Let $n_{I_t} = \frac{n_t - k_t}{2}$ where $k_t = 0$ for $n_t$ even and $k_t = 1$ for $n_t$ odd. Table 3.6 contains the 0.95-quantiles $q_{0.95}^{approx}(n_t, n_{I_t})$ of the approximative distribution for $n_t \in \{20, \ldots, 30\}$. It shows that for increasing $n_t$ and $n_{I_t}$ these quantiles are not monotonically increasing. Figure 3.15 further displays that there is also no monotonic increase for either a fixed window width $n_t$, or a fixed subset size $n_{I_t}$ in the quantiles. However, the similarity of the structure of the approximative quantiles to the simulated quantiles (top panels of Figure 3.14) is apparent.

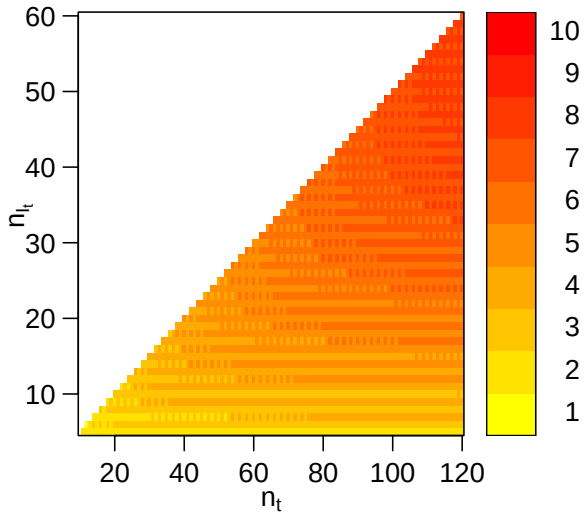


**Figure 3.15**: 0.95-quantiles of the approximative distribution defined by equations (3.15) and (3.16).

| $n_t$ | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_{I_t} = \frac{n_t - k_t}{2}$ | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 | 14 | 14 | 15 |
| $q_{0.95}^{approx}(n_t, n_{I_t})$ | 3 | 3 | 2 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 4 |

**Table 3.6**: Non-monotonicity of the 0.95-quantiles of the approximative distribution (3.15, 3.16).

To achieve logical consistency in the test decisions (3.9) for all possible settings of $n_t$ and $n_{I_t}$, we set the critical values at the level $\alpha$ to

$$c_\alpha(n_t, I_t) \;=\; \max_{m_1 \le n_t} \max_{m_2 \le n_{I_t}} \left\{ |q_{\alpha/2}(m_1, m_2)|,\; |q_{1-\alpha/2}(m_1, m_2)| \right\}. \tag{3.17}$$

For $\alpha = 0.1$ the resulting critical values are shown in Figure 3.16. They fulfill the following symmetry and monotonicity conditions for all $n_t$ and $I_t$:

$$\begin{aligned}
&1. \quad -c_\alpha(n_t, I_t) \;=\; c_{1-\alpha}(n_t, I_t) \\
&2. \quad c_\alpha(n_t, I_t) \;\le\; c_\alpha(n_s, I_t) \qquad \text{for} \quad n_t \;\le\; n_s \\
&3. \quad c_\alpha(n_t, I_t) \;\le\; c_\alpha(n_t, I_s) \qquad \text{for} \quad n_{I_t} \le\; n_{I_s}\,.
\end{aligned}$$

Examining the right panels of Figure 3.14, the 0.05- and 0.95-quantiles of $T$ applying $I_t^{recent}$ show a slightly different pattern to the quantiles for the other considered subsets. This is due to the fact that for fixed $n_t$ and increasing $n_{I_t}$ the range of the distribution increases. For the 0.95-quantiles this leads first to an increase but then causes a decrease because of the stronger concentration of the distribution around zero for large $n_{I_t}$. The right panel of Figure 3.16 shows that this structure is not maintained after the monotonisation of the critical values, but therefor logically consistent decisions are achieved by the test for increasing $n_t$ and $n_{I_t}$.

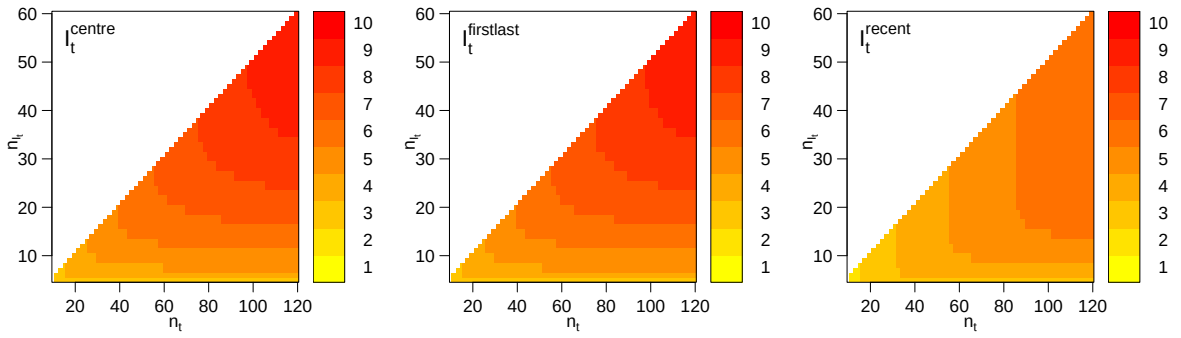


**Figure 3.16**: Critical values for the goodness of fit test at an approximate 0.1-level, derived from the modification of the simulated 0.05- and 0.95-quantiles. The modification ensures monotonicity of the critical values for increasing window width $n_t$ and increasing subset size $n_{I_t}$.

Figure 3.17 compares some specific simulated critical values with the 0.95-quantile of the distribution specified by (3.15) and (3.16). Regarding the colours in Figures 3.15 and 3.16 as heights like in a topographical map, the three panels in Figure 3.17 can be interpreted as cuts through the three-dimensional shape in Figures 3.15 and 3.16 along the lines $(n_t, \lfloor n_t/4 \rfloor)$ (top left), $(n_t, \lfloor n_t/3 \rfloor)$ (bottom left), and $(n_t, \lfloor n_t/2 \rfloor)$ (bottom right) showing the according profiles of the images. The semi-transparent colours in the background of Figure 3.17 indicate the corresponding 'height' colours used in the Figures 3.15 and 3.16.

For the displayed choices of $n_{I_t}$, the critical values for tests based on the test statistics applying $I_t^{centre}$ and $I_t^{firstlast}$ are identical and, due to the choice of the maximum value in (3.17), they are always larger than the approximate 0.95-quantiles, but the difference is never larger than two. This is also true for the critical values for $I_t^{recent}$ which are closer to the approximative quantiles, especially for small $n_{I_t}$. Concluding, these approximations can provide sensible critical values for tests in large windows for which no simulated critical values are available.

Using the described simulated critical values causes the true level of the test to be unknown. Definition (3.17) also means that the critical values are possibly chosen larger than the true critical value at a specified level such that the test may become conservative.

However, our procedure is meant as an exploratory tool for judging the appropriateness of the fit based on the current window width. Supporting the null hypothesis causes larger window widths to be in favour, but since this results in better stability, smaller variance, larger robustness and smoothness of the signal estimate, it is not a disadvantage for the performance of the proposed adaptive procedure.

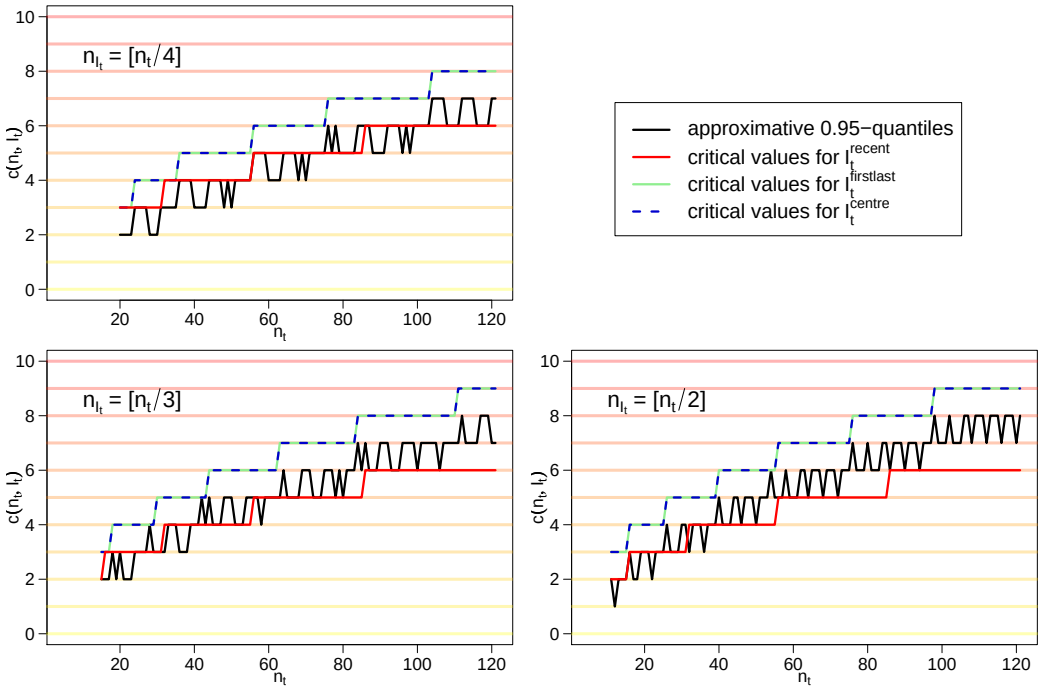


**Figure 3.17**: 0.95-quantiles of the large sample approximation (3.15, 3.16) in comparison with the simulated critical values for $n_{I_t}$ in dependence of $n_t$.

### 3.2.6 Simulation Study

To identify a filter which traces an abrupt change in the data with small time delay, we investigate the performance of the adaptive RM procedure (Figure 3.11) employing the three subsets $I_t^{centre}$, $I_t^{firstlast}$, and $I_t^{recent}$, specified at the end of Section 3.2.4. To examine the influence of the cardinality of $I_t$, we perform the simulations for $n_{I_t} \in \{\lfloor n_t/2 \rfloor,\ \lfloor n_t/3 \rfloor,\ \lfloor n_t/4 \rfloor\}$, i.e. with the cardinal number of $I_t$ depending on the window width $n_t$. If $n_{I_t}$ is smaller than five, we set $n_{I_t} = 5$ to guarantee a minimal number of observations in the subset $I_t$. Furthermore, we consider a subset containing a fixed number $n_{I_t}$ of most recent values, independent of the current window width $n_t$:

$$I_t^{fixed} = \{n_t - n_{I_t} + 1, \ldots, n_t\} \text{ with } n_{I_t} \in \{10, 15, 30\}.$$

For small window widths $n_t$, this subset may include a dominating part of the window, and it is even possible that $n_{I_t} > n_t$. Thus, if $n_{I_t}$ exceeds $n_t/2$, we set $n_{I_t} = \lfloor n_t/2 \rfloor$.
We also considered subsets $I_t^{thirds}$ which included $\lfloor n_{I_t}/3 \rfloor$ values in the first and last part of the window, and $n_{I_t} - 2\lfloor n_{I_t}/3 \rfloor$ in the centre. However, the filtering performance was worse than for the subsets described above in terms of time delay and bias and hence, the outcomes for such a filter are not shown here.

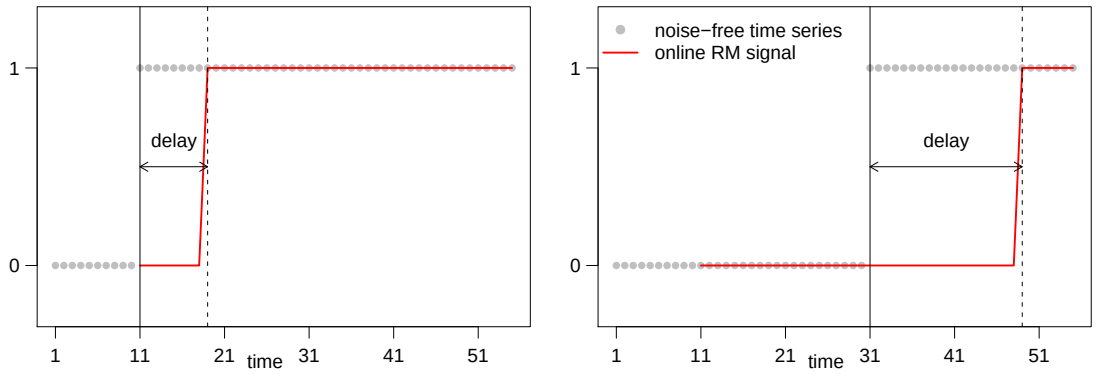


**Figure 3.18**: Noise-free time series with a level shift of size one at $t_0 = 11$ (left) and at $t_0 = 31$ (right), and online signal estimation by an adaptive RM filter using $I_t^{centre}$ and $n_{I_t} = \lfloor n_t/2 \rfloor$. The extracted RM signal traces the shift with shorter time delay if the constant period before the shift is shorter.

As simulation settings we look at level shifts and trend changes as e.g. shown in Figure 3.12. The length of the time period before the relevant change can have an impact on the signal approximation because the window width used by the adaptive filter at the time of the change can differ. Figure 3.18 illustrates that this transfers to the time delay for tracing such a pattern, here, a shift. Therefore, we consider different points in time for a shift and a trend change.

Furthermore, the signal-to-noise ratio may have an effect on the accuracy of the estimations around a change point. Therefore, different jump sizes for the level shift and different slopes of linear trends for the trend change are examined. In particular, we investigate time series with the true signal

$$\mu_t = \begin{cases} \mu_1(t) & \text{for } t = 1, \ldots, t_0 - 1 \\ \mu_2(t) & \text{for } t = t_0, \ldots, 2 \cdot t_0 \end{cases}, \tag{3.18}$$

where $t_0$ denotes the time of the shift or the trend change, respectively.

For a level shift we define

$$\mu_1(t) = 0 \quad \text{and} \quad \mu_2(t) = h\,, \tag{3.19}$$

and for a trend change we let

$$\mu_1(t) = \beta \cdot t \quad \text{and} \quad \mu_2(t) = 2\,\beta\,(t_0 - 1) - \beta \cdot t\,. \tag{3.20}$$

For the simulations, different settings for $h$, $\beta$ and $t_0$ are considered.

### NOISE-FREE LEVEL SHIFTS AND TREND CHANGES

As extreme settings, we consider the signal $\mu_t$ (3.18) as a time series without any additional noise or outliers. Because of the exact fit property and the regression equivariance of RM regression, the size of the shift $h$ in (3.19) or the size of the slope $\beta$ in (3.20), respectively, has no impact on the delay for tracing a change in the signal in noise free situations. Therefore, we examine

- a level shift of size $h = 1$ at $t_0 \in \{11, 21, \ldots, 131\}$ and
- a trend change with $\beta = 1$ at $t_0 \in \{11, 21, \ldots, 131\}$.

As a measure of performance, we take the delay from the time $t_0$ when the true change has taken place to the time when the filter first reacts to the change, illustrated in Figure 3.18. Let $\hat{\mu}_t$ denote the signal estimated by an adaptive online RM filter at time $t$, the delay is then defined by

$$\text{delay} = \min\left\{t \ : \ |\hat{\mu}_t - \mu_1(t)| > 0, \ t > t_0\right\} \ - \ t_0$$

with $\mu_1(t)$ according to (3.19) for a shift and according to (3.20) for a trend change. Using a different definition for the delay, e.g. the time difference until the estimated signal exceeds 50%, 75%, 90% or 95% of the difference between $\mu_1(t)$ and $\mu_2(t)$ does not change the findings presented in the following.
Figure 3.19 shows the delay of several adaptive RM filters needed to trace a level shift induced at different times $t_0$. A level shift after a large period of constancy (large $t_0$) induces a large window width $n_{t_0}$ at the time of the shift and hence, Figure 3.19 shows, as expected: the time delay increases with increasing $t_0$. Filters using a test statistic based on the subsets $I_t^{centre}$ or $I_t^{firstlast}$ show an almost linear increase for smaller $t_0$, while the delay using $I_t^{recent}$ or $I_t^{fixed}$ stays almost constant – regardless of $t_0$ or $n_{t_0}$, respectively. Furthermore, taking a smaller portion $n_{I_t}$ into account seems to increase the delay for filters applying a test statistic based on $I_t^{centre}$ or $I_t^{firstlast}$ whereas the delay with $I_t^{fixed}$ is a little lesser for some situations. When considering a certain number of residual signs depending on the current window width $n_t$ according to $I_t^{recent}$, no difference in delay can be seen for the simulated settings. The outcomes for a trend change are similar to the outcomes for a level shift (see Figure C.3 in Appendix C).
Concluding, adaptive RM filters based on a test statistic which takes into account the most recent residual signs by using $I_t^{recent}$ or $I_t^{fixed}$ result in the smallest time delay in tracing a shift or trend change if the data are noise-free. In these situations, the number of considered residual signs $n_{I_t}$ seems to have no impact on the delay.

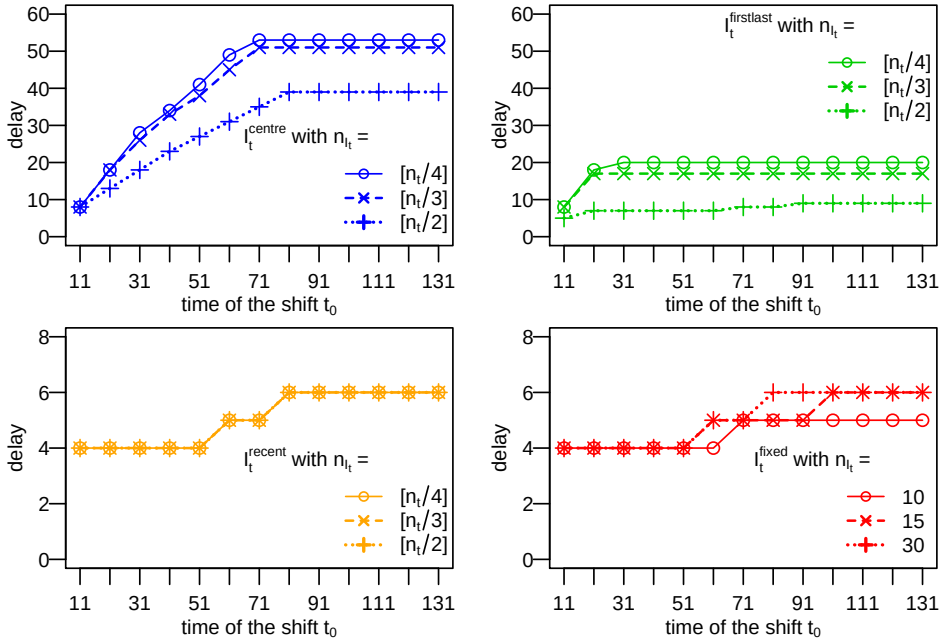


**Figure 3.19**: Time delay in tracing a level shift at time $t_0 \in \{11, 21, \ldots, 131\}$ produced by adaptive RM filters with different settings.

### Noisy Level Shifts and Trend Changes

The noisy data we consider are simulated from model (3.1), with the signal $\mu_t$ corresponding to (3.18) and either (3.19) for a level shift or (3.20) for a trend change. The errors are generated according to $\varepsilon_t \sim N(0, 1)$ and no outliers are included, i.e. $\nu_t = 0 \; \forall \; t$. The situations we consider consist of

- level shifts of size $h \in \{1, 2, 5, 10\}$ at $t_0 \in \{11, 61, 121\}$ and
- trend changes with slopes $\beta \in \{1, 2, 5, 10\}$ at $t_0 \in \{11, 61, 121\}$.

To judge the performance of the different filter settings we use the mean and median bias right after a shift or trend change, respectively, evaluated on 1000 simulation runs. The variance and mean squared error (MSE) of the signal estimates are also investigated and commented on but will not be displayed here for the sake of brevity.
Because of the robustness of the RM filters against patches of subsequent outliers, the estimated signal continues the linear trend defined by $\mu_1(t)$ for some time after $t_0$. For the level shift situations, this results in a bias of approximately $-h$ right after $t_0$, see e.g. Figure 3.20. In case of a trend change, the linear trend after $t_0$ points into the opposite direction than the trend before, causing a linear increase of about twice the size of $\beta$ for the bias as can be seen in Figure 3.21. A filter for which the bias curve tends more quickly towards zero after $t_0$ is said to be a method with shorter delay.

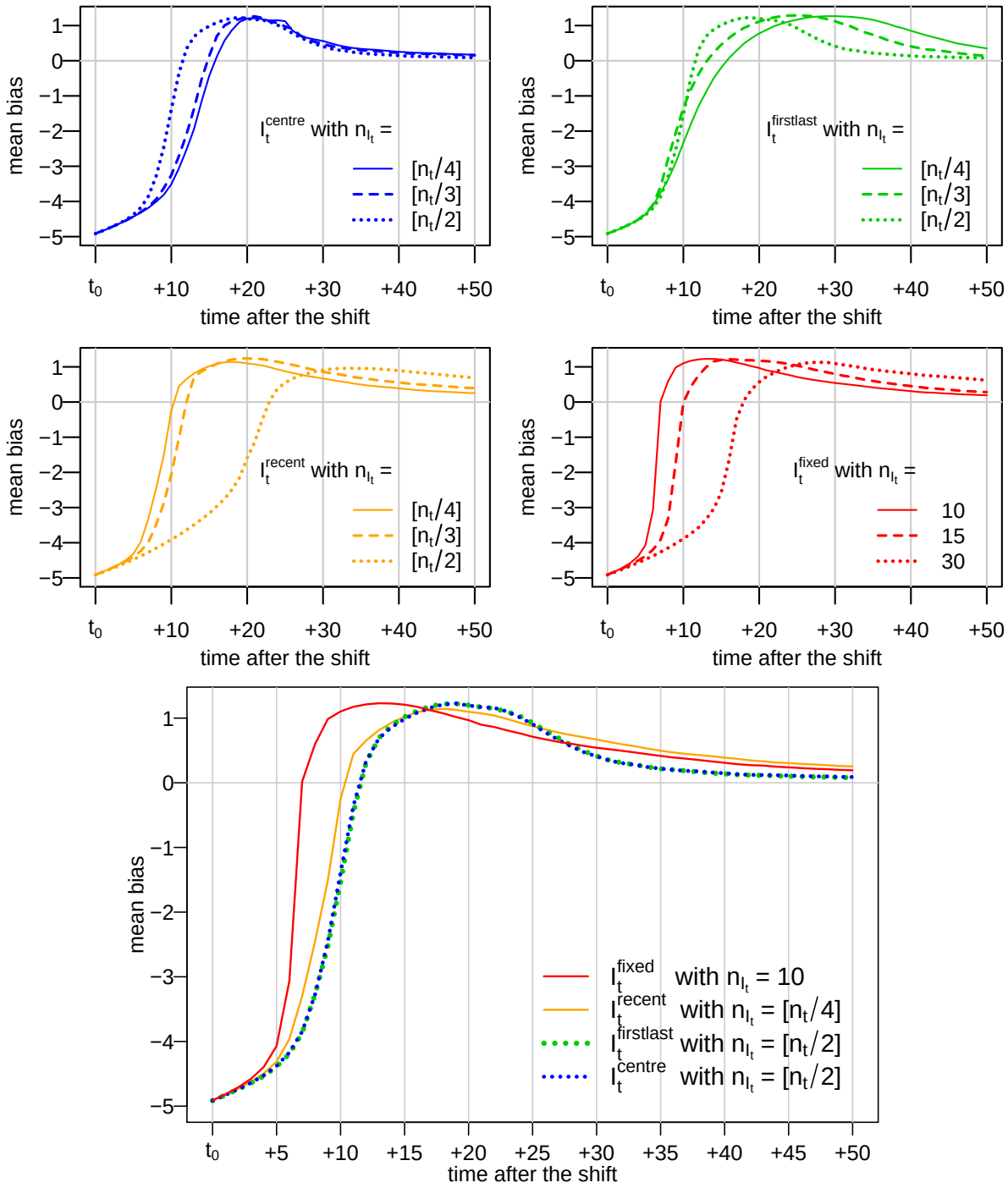


**Figure 3.20**: The top four panels show the mean bias after a level shift of size $h = 5$ at time $t_0 = 61$ for the adaptive RM filter with all considered subsets $I_t$ of size $n_{I_t}$. The bottom panel compares the 'best' mean bias curves from each of the plots on top, i.e. the ones which indicate the smallest time delay for tracing the shift, for each of the four subsets $I_t^{centre}$, $I_t^{firstlast}$, $I_t^{recent}$ and $I_t^{fixed}$.

### Level Shifts

Figure 3.20 displays the mean bias for all considered RM filters after a level shift of size $h = 5$ at $t_0 = 61$. The bottom panel compares the bias curves of the best performing filters for each of the four considered subsets $I_t$. It shows that filters based on a test statistic regarding a small number of the most recent residual signs outperform filters with other choices. In particular, the filter based on $I_t^{fixed}$ with $n_{I_t} = 10$ results in the smallest average time delay shown by the fact that the bias curve tends faster towards zero than for the other filters. This appears even more drastically with either, larger $t_0$ or larger shift size $h$.

For smaller shifts or shifts at an earlier time, the differences between the filters are not that obvious; for a level shift at $t_0 = 11$ almost no difference in mean or median bias is observed. Considering the median bias for large $t_0$ and large $h$, it is obvious that the estimated signal traces a sudden shift at some point in time, whereas the mean bias suggests a smooth and rather slow transition from the level before, to the level after the shift. However, since the difference between median and mean bias is generally small, the median bias is not shown here. All the considered filters have a decreasing time delay for an increasing shift size $h$, but increasing time delay for increasing $t_0$. However, this increase in delay is close to zero for a filter using $I_t^{fixed}$ with small $n_{I_t}$, whereas it appears prominently for all other filters.

The simulation study shows, as expected, that the variance of the signal estimates is the smaller, the larger the average window width $n_t$ at a certain time $t$. Since the window width increases until the time of the shift $t_0$, the variability decreases with time. However, because each filter has a certain delay in tracing the shift which also depends on the simulated data around that time, the variability shows a peak right after $t_0$. This peak is almost negligible for small $h$ and small $t_0$, but the peak is higher, the larger $h$ and $t_0$. Generally, this peak appears earlier for filters applying subsets $I_t$ regarding a lesser number of most recent residual signs and also the variance curve over time returns quicker to low values while the increased variability can be present for a longer time for $I_t^{centre}$ or $I_t^{firstlast}$. For the sake of brevity, the simulation results concerning the variance are not shown here.

Since the (squared) bias considerably dominates the MSE, this measure confirms the outcomes described above: The MSE curves for all filter settings show a large peak starting at $t_0$ and decrease thereafter, and the filter applying $I_t^{fixed}$ with $n_{I_t} = 10$ has the MSE curve that decreases the fastest. These curves are also excluded from the presentation here.

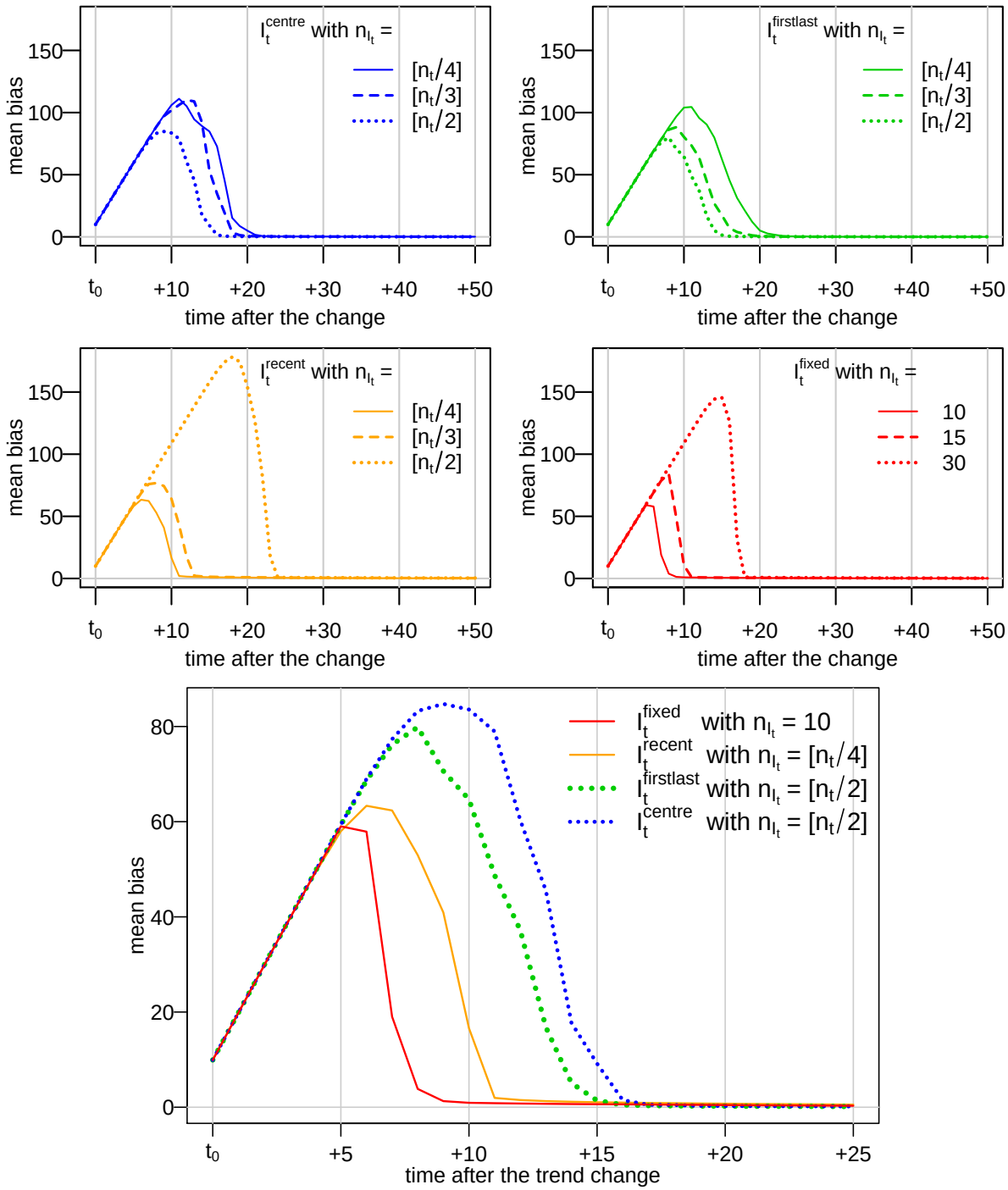


**Figure 3.21**: The top four panels show the mean bias after a trend change from $\beta = 5$ to $\beta = -5$ at time $t_0 = 61$ for the adaptive RM filter with all considered subsets $I_t$ of size $n_{I_t}$. The bottom panel compares the 'best' mean bias curves from each of the plots on top, i.e. the ones which indicate the smallest time delay for tracing the trend change, for each of the four subsets $I_t^{centre}$, $I_t^{firstlast}$, $I_t^{recent}$ and $I_t^{fixed}$.

### Trend Changes

Figure 3.21 shows the mean bias curves for all investigated filters for a trend change from $\beta = 5$ to $\beta = -5$ at $t_0 = 61$. Again, the filters using test statistics based on the subsets $I_t^{recent}$ and $I_t^{fixed}$ with small $n_{I_t}$ perform best w.r.t. bias and delay. For a trend change at $t_0 = 11$, a filter using $I_t^{fixed}$ can perform worse than other filters if the slope is not very large ($\beta = 1$ or $\beta = 2$); for all other settings of $\beta$ and $t_0$ the filter with $I_t^{fixed}$ clearly shows the best results. The larger $\beta$, the lesser the delay in tracing the change.
Analogous to the outcomes for the level shifts, all signal estimates show a peak in their variance curves after $t_0$. Here, it is even more obvious that those filters applying $I_t^{recent}$ or $I_t^{fixed}$ with low $n_{I_t}$ outperform the other filters because the variance is less, and the variance curve decreases the fastest after $t_0$. Again, the MSE is strongly dominated by the bias and thus, the MSE curves also imply using a filter based on $I_t^{fixed}$ with $n_{I_t} = 10$. Again, the outcomes for the variance and MSE are not shown here.

## 3.2.7 Modifications of the Adaptive Repeated Median Filter

In the last section it could be seen that the filter output has a strong bias after a sudden level or trend change. In extreme situations this may cause the signal estimate to exceed the range of the observations. This is not really acceptable in practical applications: for example, if a physiological time series in intensive care shows a drastic change within the alarm limits it is not the aim to trigger a threshold alarm. However, it is possible that the approximated signal may cause such an alarm because a continuation of a previous trend carries the estimation out of the alarm limits. Therefore, we apply the *restrict-to-range rule*, which restricts the estimated signal level $\hat{\mu}_t$ at time $t$ to a value within the range of a subset of the most recent observations in the current time window:

$$\hat{\mu}_t = \begin{cases} \min(\mathbf{y}_{t,n_t^{rtr}}), & \text{if } \hat{\mu}_t^{RM} < \min(\mathbf{y}_{t,n_t^{rtr}}) \\ \hat{\mu}_t^{RM}, & \text{if } \min(\mathbf{y}_{t,n_t^{rtr}}) \leq \hat{\mu}_t^{RM} \leq \max(\mathbf{y}_{t,n_t^{rtr}}) \\ \max(\mathbf{y}_{t,n_t^{rtr}}), & \text{if } \max(\mathbf{y}_{t,n_t^{rtr}}) < \hat{\mu}_t^{RM} \end{cases}, \tag{3.21}$$

where $\hat{\mu}_t^{RM}$ denotes the Repeated Median level estimate at time $t$; $\min(\mathbf{y}_{t,n_t^{rtr}})$ denotes the minimum, and $\max(\mathbf{y}_{t,n_t^{rtr}})$ denotes the maximum, evaluated at the sample $\mathbf{y}_{t,n_t^{rtr}} = (y_{t-n_t^{rtr}+1}, \ldots, y_t)'$. One possibility is to choose the subset size for the restrict-to-range rule equal to the window width $n_t^{rtr} = n_t$ to restrict the signal level estimate to the range of the observations in the window. However, to prevent that outliers that date back some time have an influence on the range of the current signal estimate we set $n_t^{rtr} = n_{I_t}$ in the following and thus, restrict the signal level to the range of the observations which are used to evaluate the appropriateness of the current fit when using $I_t^{recent}$ or $I_t^{fixed}$.
Another issue that needs to be addressed in practical applications is the treatment of

missing values: intensive care time series can contain missing values at single points in time due to short-term technical problems as well as long stretches of missing values, e.g. caused by disconnection of the measurement devices. To ensure reliability of the signal estimation, a certain number of (non-missing) observations should be available for the estimation. One possibility is to set the minimum required number of observations within one window equal to the value of the minimal window width $n_{min}$ to retain the robustness properties of the filter. However, for online estimation from data which can be expected to frequently show missing values, it is sensible to request a certain number of non-missings at the most recent times such that for the filter applying $I_t^{recent}$ or $I_t^{fixed}$ a minimum number of observations is available for the test procedure. In the following, we demand at least five observations at the most recent $n_{I_t}$ times; if less observations are present, the filter returns a missing value for the estimated signal level. In this way a signal estimation is guaranteed that is up-to-date, and it ensures a continuous signal estimation in case of short-run technical failures.

### 3.2.8 EXAMPLES

This section compares the adaptive RM filters with the best settings of $n_{I_t}$ for each of the four subsets $I_t^{centre}$, $I_t^{firstlast}$, $I_t^{recent}$, and $I_t^{fixed}$ determined in Section 3.2.6 also applying the modifications described in Section 3.2.7 to simulated and real time series. The comparisons also include filters based on fixed window widths.

First, we investigate some well-known examples from the regression context, typically used for evaluating the performance of non-parametric smoothers and filters. In particular, we take the Blocks and Doppler functions described by Donoho and Johnstone (1995) as time series signal $\mu_t$ at equidistant points $t$ according to (3.1). The observations are simulated by adding standard normal noise to the signal $\mu_t$, which is re-scaled to achieve a signal-to-noise ratio of five, plus 5% positive, additive outliers of size five at random times. Figures 3.22 and 3.23 show the signals of these time series extracted by four adaptive RM online filters. It can be seen that the restrict-to-range rule proposed in Section 3.2.7 does not avoid the over- or underestimation of the *true* signal after a level shift or trend change, but at least it prevents that the estimated signal falls outside the observational range. Such situations can for example be seen in Figure 3.22, e.g. after the negative level shifts around times 250 and 400, where the signal approximation is truncated by the minimum of the observations in the current window.

Furthermore, filters based on a local linear fit tend to be biased in regions of curvature. If the estimate is evaluated in the centre of a certain neighbourhood, such filters tend to 'trim the hills and fill the valleys' (cf. Hastie, Tibshirani and Friedman (2001), p. 171).

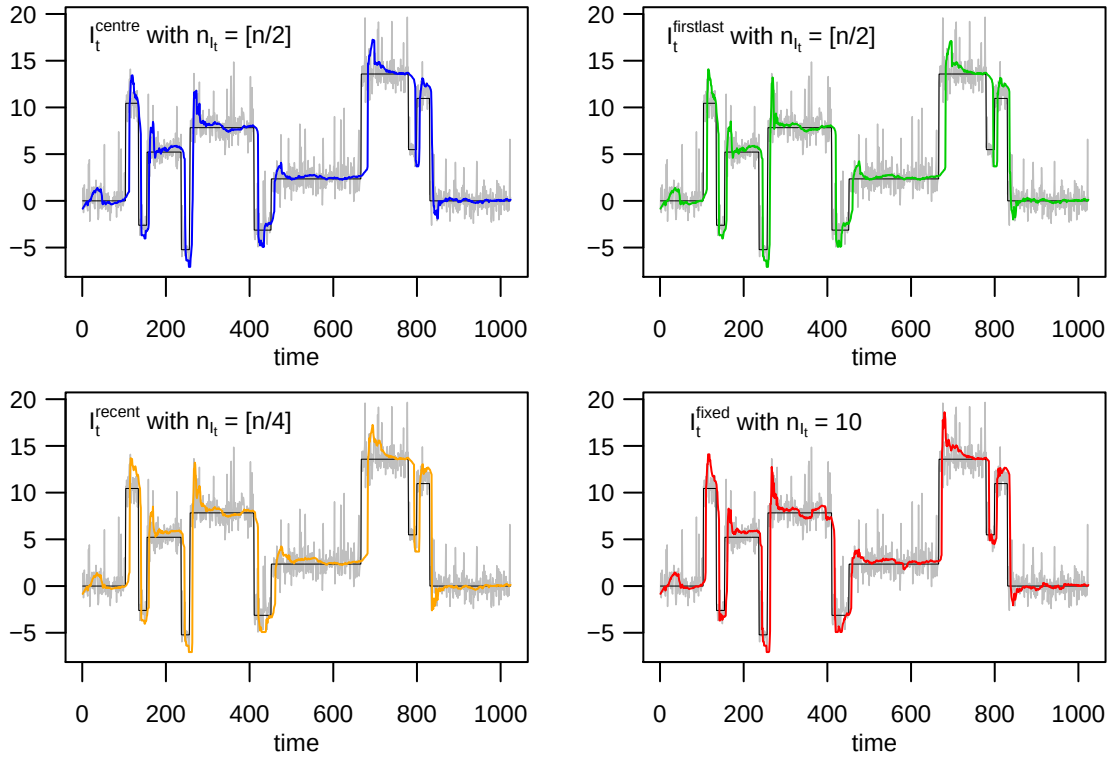


**Figure 3.22**: Signal of adaptive RM filters estimating the Blocks signal (black) from a time series (grey) with a signal-to-noise ratio of five and 5% positive additive outliers of size five.

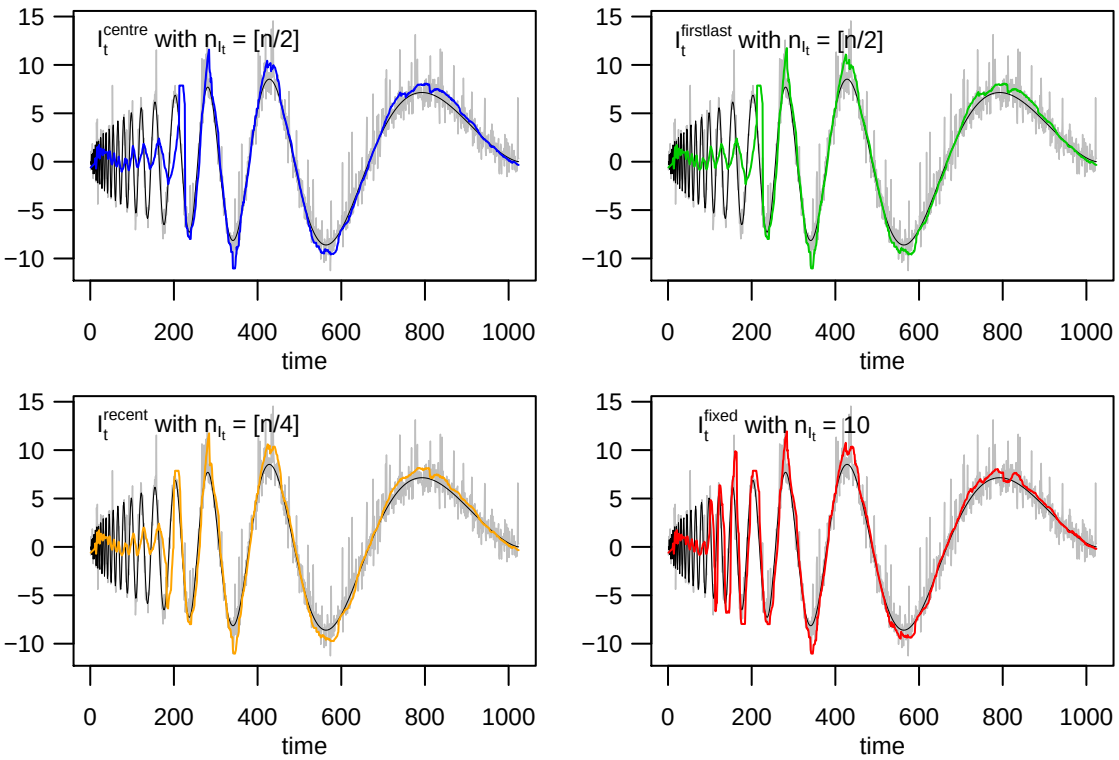


**Figure 3.23**: Signal of adaptive RM filters estimating the Doppler signal (black) from a time series (grey) with a signal-to-noise ratio of five and 5% positive additive outliers of size five.

For an online application, Figure 3.23 shows that the proposed online filters, based on an estimate at the end of each window, tend to 'augment the hills and deepen the valleys'.

In the first part of the Doppler signal (Figure 3.23), it is difficult to distinguish signal from noise, hence, all filters yield estimated signals around zero. The filter with $I_t^{fixed}$ and $n_{I_t} = 10$ traces the true signal close-by from about time $t = 100$ on while for the other filters the estimated and the true signal are close-by only from about time $t = 200$ on. Thus, the $I_t^{fixed}$-RM-filter has a better ability to trace patterns occurring in short time intervals and it also shows the smallest delay in tracing sudden shifts, see Figure 3.22.

Figure 3.24 shows another simulated time series with standard normal noise and 5% additive outliers of size five. The true signal contains level shifts of different heights as well as trends and trend changes. Again, it can be seen that using the test statistic applying $I_t^{fixed}$ causes the smallest delay in tracing sudden shifts.

For comparison Figure 3.25 shows the same time series but compares the estimated online signal resulting from the adaptive RM filter using the subset $I_t^{fixed}$ with the results from two online RM filters based on fixed window widths $n_t = 15$ and $n_t = 50$, for all $t$.

Figure 3.25 illustrates that all these filters are robust against artefacts, e.g. around time $t = 300$. The filter with large, fixed window width yields the smoothest signal estimations but also has the largest delay in tracing changes, most obvious for the shift at time $t = 250$. Furthermore, the estimated signal for this filter lies outside the observational range after the trend change around time $t = 75$, which could be prevented by applying the restrict-to-range rule proposed in Section 3.2.7. The RM filter based on $n_t = 15$ follows changes in the data with the smallest time delay, but exhibits much more variation around the true signal than the other filters in constant periods, e.g. from $t = 300$ to 350. Since the underlying data structure is not known in advance for intensive care time series, the example shown in Figure 3.25 illustrates the advantages of an RM filter with a data-driven choice of the window width: the signal approximation is smooth in times of constant or linear trend, but is also able to follow sudden changes with small delay.

As real data examples, we consider one hour of systolic arterial blood pressure and half an hour of heart rate measurements displayed in Figure 3.26. Both time series are measured and stored once per second. In addition to the measurements, the graphics display the upper and lower alarm limits set by the medical staff. Furthermore, the figure shows the online signal extracted by an adaptive RM filter with $I_t^{fixed}$. We used $n_{I_t} = 30$ for these two examples because application to several intensive care time series showed that a filter with these settings provides online signal estimates that are sufficiently smooth and have an acceptable time delay for the detection of relevant sudden changes. Choosing a smaller value for $n_{I_t}$ causes window width reduction for changes that may not be relevant from a medical perspective and hence, results in a more variable signal estimation.

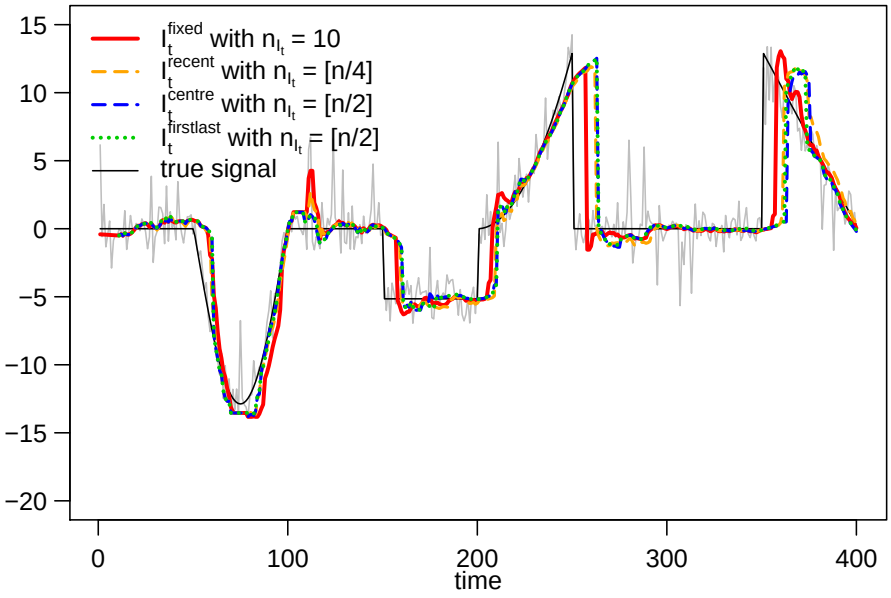


**Figure 3.24**: Comparison of four adaptive RM online filters at a simulated time series with standard normal noise, a signal-to-noise ratio of five, and 5% additive outliers of size five with random sign at random times.

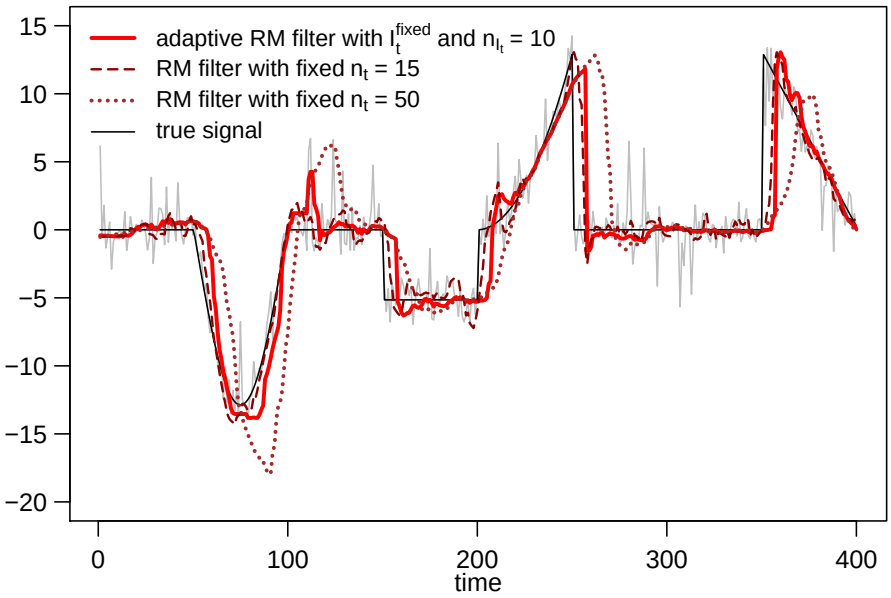


**Figure 3.25**: Comparison of the adaptive RM filter with $I_t^{fixed}$ and $n_{I_t} = 10$ with two RM online filters with fixed window widths $n_t = 15$ and $n_t = 50$ for all $t$ at a simulated time series with standard normal noise, a signal-to-noise ratio of five, and 5% additive outliers of size five with random sign at random times.

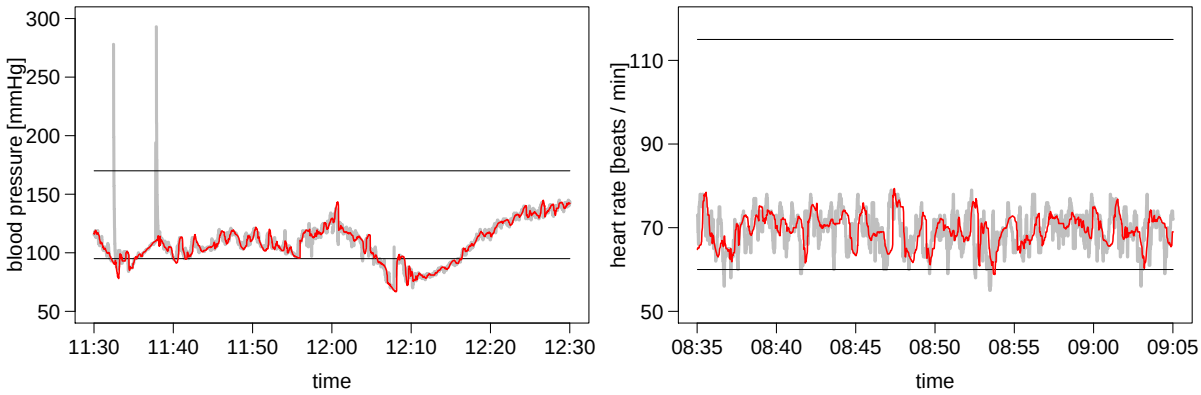


**Figure 3.26**: Systolic arterial blood pressure (left) and heart rate (right) with upper and lower alarm limits and a signal estimated by an adaptive RM filter using $I_t^{fixed}$ with $n_{I_t} = 30$.

For the blood pressure, the currently used alarm system triggers two artefact alarms caused by outliers, which violate the upper alarm limit, while the true blood pressure is close to the lower alarm limit. For the heart rate, a number of alarms are triggered because of lower limit violations while from a medical perspective it would rather be of interest that the heart rate is close to the lower limit with additional information about the variability. Using the signal extracted by the adaptive RM filter with $I_t^{fixed}$ as input to the threshold alarm system prevents the artefact alarms, displayed in the left panel of Figure 3.26, but still indicates relevant violations of the alarm limits. The right panel of Figure 3.26 illustrates that the total number of alarms can also be reduced because the estimated signal crosses the alarm threshold less often than the observed measurements but still transports the relevant information.

### 3.2.9 Search Algorithms

#### Linear Search

In Section 3.2.1 it is described that the adequate window width is determined at each time $t$ by a linear search algorithm: if the null hypothesis of the goodness of the current fit is rejected, then the current window width $n_t$ is reduced by one in each iteration step. The final window width is the first width for which the goodness of fit cannot be rejected. In that way a certain smoothness of the estimated signal is guaranteed because the algorithm chooses the largest window width possible and thus reduces the variability of the signal. However, the computing time of this algorithm might become quite large if the window width has to be reduced drastically because of a sudden change such as a trend change or level shift. In such a case the number of iterations might be as large as $n_{max} - n_{min}$. In the following, let $n_{t,\ell}$ denote the window width used to evaluate $\hat{\mu}_t$ in the $\ell$th iteration step where at each time $t$ we start with $n_{t,0} = \min\{n_{t-1} + 1, n_{max}\}$ and $n_t$ denotes the final window width chosen at time $t$.

### Binary Search

The binary search offers a faster approach to find an adequate window width because it uses nested intervals: if the goodness of fit is rejected in the first iteration for the width $n_{t,0}$, at the next step the adequateness of the fit based on the minimal window width $n_{min}$ is tested. If $H_0$ is rejected at that step, then $n_{t,1} = n_{min}$ is chosen as the final window width $n_t$ because the width cannot be chosen any smaller by definition. However, if $H_0$ is not rejected, the fit in a window with a width $n_{t,2} = \lceil (n_{min} + n_{t,0})/2 \rceil$ is evaluated. In that way the set of possible window widths, limited by $n_{min}$ and $n_{t,0}$, is divided into two halves and depending on the test decision for $n_{t,2}$ the set of possible widths is reduced either to $\{n_{min}, \ldots, n_{t,2}\}$ or to $\{n_{t,2}, \ldots, n_{t,0}\}$. Repeating this limitation to half intervals in each iteration step, a maximum of $\log(n_{max} - n_{min}) + 2$ iterations is possible.
The following pseudo code describes the procedure of finding an adequate window width and thus, a suitable signal level estimate at time $t$ with a binary search algorithm. The final level estimate, which is based on the final window width $n_t$, is denoted by $\hat{\mu}_t$, while an estimate in iteration $\ell$, based on the window width $n_{t,\ell}$, is denoted by $\hat{\mu}_{t,\ell}$.

1. (a) Perform RM regression in a window of width $n_{t,0} = \min\{n_{t-1} + 1, n_{max}\}$.
   (b) Test the goodness of fit of $\hat{\mu}_{t,0}$ according to the test problem specified in (3.7).
   If $H_0$ is not rejected: Save $\hat{\mu}_t = \hat{\mu}_{t,0}$, $n_t = n_{t,0}$ and **stop**.
   If $H_0$ is rejected: Set $n_{t,1} = n_{min}$ and $\ell = 1$.

2. (a) Perform RM regression in a window of width $n_{t,1} = n_{min}$.
   (b) Test the goodness of fit of $\hat{\mu}_{t,1}$.
   If $H_0$ is not rejected: Set $n_{low} = n_{min}$, $n_{up} = n_{t,0}$,
   $n_{t,2} = \lceil \frac{n_{low}+n_{up}}{2} \rceil$ and $\ell = 2$.
   If $H_0$ is rejected: Save $\hat{\mu}_t = \hat{\mu}_{t,1}$, $n_t = n_{t,1} = n_{min}$ and **stop**.

3. Repeat:
   (a) Perform RM regression in a window of width $n_{t,\ell}$.
   (b) Test the goodness of fit of $\hat{\mu}_{t,\ell}$.
   If $H_0$ is not rejected: Set $n_{low} = n_{t,\ell}$.
   If $H_0$ is rejected: Set $n_{up} = n_{t,\ell}$.
   (c) Set $n_{t,\ell+1} = \lceil \frac{n_{low}+n_{up}}{2} \rceil$ and $\ell = \ell + 1$.
   (d) If $n_{t,\ell+1} = n_{low}$ or $n_{t,\ell+1} = n_{up}$:
   Determine $\ell_m := \max\{j = 1, \ldots, \ell : H_0 \text{ not rejected}\}$,
   save $\hat{\mu}_t = \hat{\mu}_{t,\ell_m}$, $n_t = n_{t,\ell_m}$, $\ell_t = \ell$ and **stop**.

### GEOMETRIC SEARCH

The geometric search offers a possibility to combine fast computation with the fact that a window width is searched for which is adequate in the sense that the goodness of fit for the level estimate from this window cannot be rejected, and which is at the same time as large as possible, to reduce the variability of the estimated signal.

This search procedure starts by reducing the window width similar to the linear search algorithm, though here the window width is not reduced by one but by $2^\ell$. The window width in the $\ell$th iteration corresponds to $n_{t,\ell} = n_{t,0} - \sum_{i=1}^{\ell} 2^{i-1}$. If the goodness of fit for the level estimate $\hat{\mu}_{t,\ell}$ cannot be rejected at iteration $\ell$, a binary search within the limits $n_{t,\ell}$ and $n_{t,\ell-1}$ is performed.

Like the binary search procedure, the computation time for the geometric width search is of order $\mathcal{O}(\log(n_{max} - n_{min}))$. However, this search puts much more weight on large window widths because it starts with a reduction of $n_{0,t}$, the largest window width possible at time $t$, and reduces it first in small and then in exponentially increasing steps.

Depending on the underlying data, it is possible that the geometric and the binary search may not find the largest possible window width for which the corresponding fit is considered adequate. Therefore, a small comparison study will be conducted in the following to find out about the influence of the search algorithm on the final window width.

### NOISE-FREE LEVEL SHIFT

To compare the performance of the three different search algorithms, we first apply the adaptive online RM filter to a noise-free time series

$$y_t = \begin{cases} 0\,, & t = 1, \dots 50 \\ h\,, & t = 51, \dots 100 \end{cases},$$

with shift size $h = 5$, whereas the adequacy of the current fit is evaluated on basis of the five most recent residuals, i.e. we use $I_t^{fixed}$ with $n_{I_t} = 5$, and the estimated signal is restricted to the observational range of the five most recent observations, i.e. $n_t^{rtr} = n_{I_t} = 5$. The minimal window width is set to $n_{min} = 11$ and the maximum window width corresponds to $n_{max} = 100$. This procedure results in the online signal approximation

$$\hat{\mu}_t = \begin{cases} 0\,, & t = 1, \dots 54 \\ 5\,, & t = 55, \dots 100 \end{cases},$$

regardless of the algorithm used for finding the window width.

On a Linux Workstation with 3.06GHz CPU and 4GB RAM, the procession time for the analysis by the filter using the linear search corresponds to approximately one second while the filter applying the geometric search needs about 0.80 seconds; the analysis by the filter applying the binary search only takes about 0.64 seconds. As an explanation for

| $t$ | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|
| linear search | 1 | 1 | 1 | 44 | 2 | 1 | 2 | 1 | 1 | 1 |
| binary search | 1 | 1 | 1 | 2 | 2 | 1 | 3 | 1 | 1 | 1 |
| geometric search | 1 | 1 | 1 | 7 | 2 | 1 | 2 | 1 | 1 | 1 |

**Table 3.7**: Number of iterations $\ell_t$ required for finding the adequate window width by different search algorithms after a noise free level shift of size five at time $t_0 = 51$.

these computation times, Table 3.7 shows the number of iterations $\ell_t$ required by each algorithm at the times right after the level shift. At all other times ($t \leq 50$ and $t > 60$) the window width is simply increased by one and hence only one iteration takes place at each time. Since the filter output is the same regardless of the applied search algorithm, using a binary search would be the best option here, because it is the fastest.

### Level Shift with Standard Normal Noise

To compare the influence of the three search algorithms at noisy time series, we apply the adaptive RM filter to time series of length 100 with i.i.d. standard normal noise, a level shift at time $t_0 = 51$ and signal-to-noise ratios of zero, one, two and five. For each setting 1000 replications are performed. As performance measures, we consider the computation time, the window width $n_t$, and the number of iterations $\ell_t$ required by each algorithm at all times $t$.

For standard normal data without level shift (meaning a signal-to-noise-ratio of zero) the computation time is again the lowest for the binary search algorithm although the difference to the geometric search is not very large (cf. Figure C.4 in Appendix C). The maximum window width is the same for all search algorithms; the minimum, median and average window widths are slightly larger when using the linear search algorithm. However in terms of the maximum number of iterations at one time, it is very obvious that the binary and geometric search are at an advantage (cf. Figure C.5).

The three top panels in Figure 3.27 show the smallest, the mean, the median and the largest window width out of the 1000 simulation runs for the setting with signal-to-noise ratio five at each point in time. The bottom panels show the corresponding summary statistics for the number of iterations required by the three different search algorithms at each point in time.

For the simulated time series, an 'optimal' filter would have an increasing window width until the time of the level shift, a drastic drop to the smallest window width at the time of the shift and a width increase thereafter. Figure 3.27 shows that on average this is the case for all three search algorithms. Further, the maximum widths resulting from the different algorithms are quite similar for all times – except after the shift where the binary and the

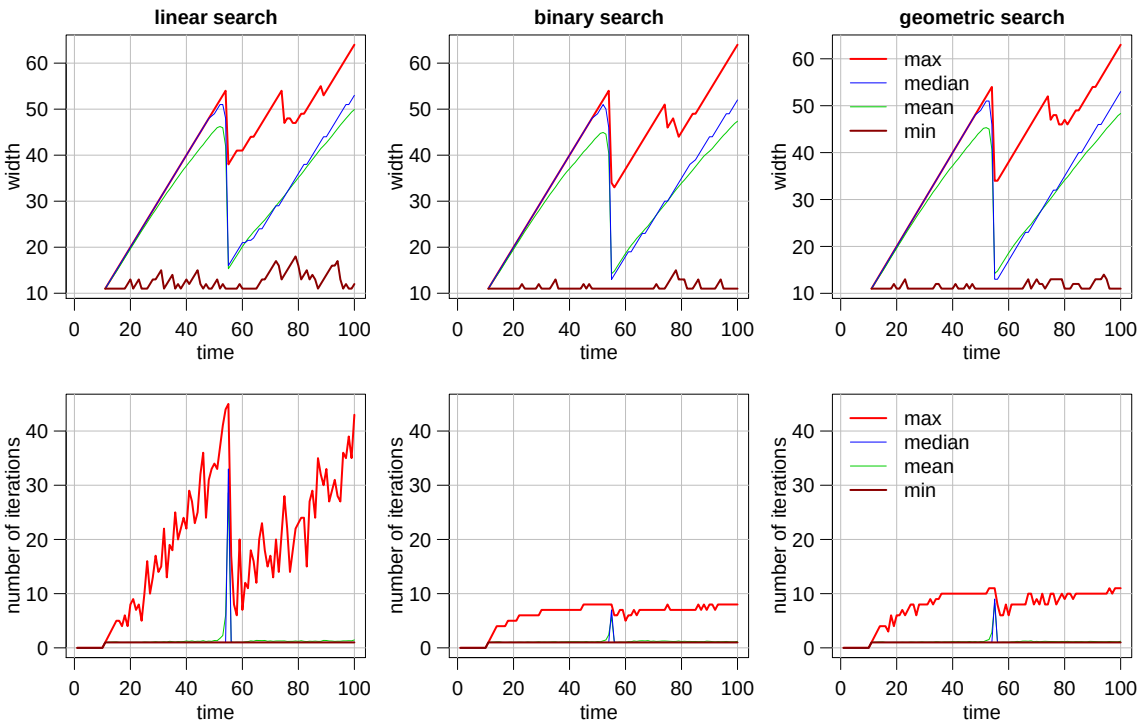


**Figure 3.27**: Summary statistics of the window widths $n_t$ and the number of iterations $\ell_t$ required by the adaptive online RM filter at time $t$ when using different search algorithms for the width. These summary statistics are based on 1000 simulated standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of five.

geometric search perform a little better because the width reduction is more drastic than for the linear search. However, the linear search results less often in $n_{min} = 11$ than the other search algorithms.

Looking at the number of iterations in the bottom panels of Figure 3.27, the advantages of the binary and geometric search are quite obvious: all three filter versions perform similar on average in stable time periods, but the maximum, the mean and even the median number of iterations around the level shift is much higher for the linear search algorithm. Indeed, the maximum number of iterations is much larger for the linear search over the whole period of time. Although the binary search requires less iterations than the geometric search, the difference is not large and the maximum number of iterations for the geometric search never exceeds eleven. The results for the signal-to-noise-ratios one and two lead to the same conclusions and are shown in Figures C.7 and C.9.

The higher number of iterations required by the linear search algorithm is reflected by a generally higher computation time, illustrated in Figure 3.28. This figure shows boxplots of the computation times needed for the analyses of the 1000 simulated time series. Again it shows that the binary search is the fastest option, but it also shows that the difference to the geometric search is not large. The same is true for the signal-to-noise ratios one and two as can be seen in Figures C.6 and C.8 in Appendix C.

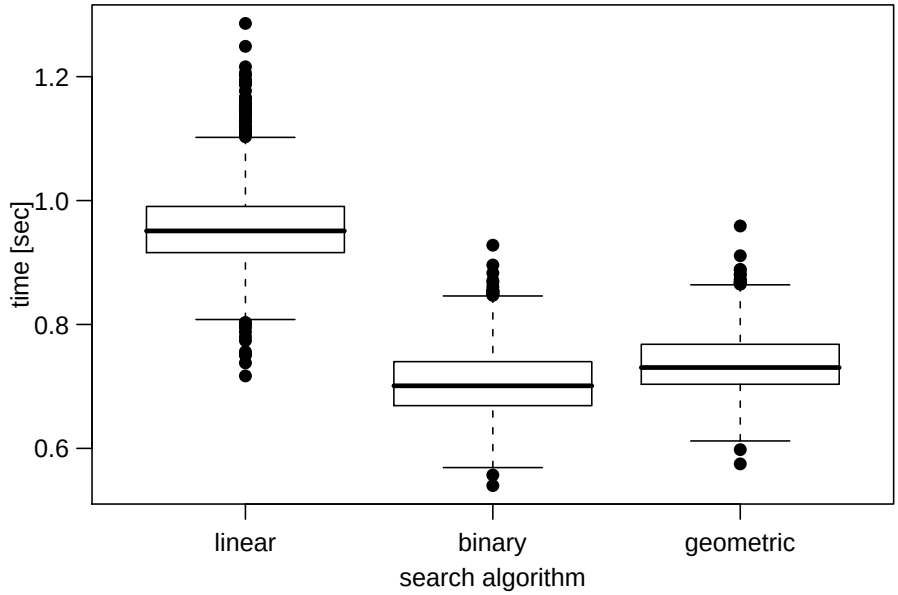


**Figure 3.28**: Computation times in seconds required by adaptive online RM filters applying different search algorithms for the window width, when analysing standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of five. The boxplots are based on 1000 simulation runs each.

Following from the definition, the linear search always results in the largest window width possible. However, the width found by the linear search is larger than the width found by a binary search for an average of only 27.7 out of 90 estimation times; for the geometric search it is just 26.1 times. For 49.8% of the time series the window widths found by all three search algorithms is the same at *at least* 49 out of 90 estimation times. In 5.9% the window widths found by the three algorithms are even identical at *all* time points. The similarity of the results of the binary and the geometric search is quite strong, but the difference to the widths found by the linear search is also small. While the binary search may need a few less iterations to find an adequate width at one point in time than the geometric search, the simulations indicate that the latter may result in a larger adequate window width which, in return, induces a smoother signal estimation.

Figure 3.29 shows an application of the adaptive online RM filter using the three different search algorithms to a standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of five. The minimal window width for this filter is $n_{min} = 11$, the number of most recent observations to test the goodness of fit of the signal is $n_{I_t} = 5$. Furthermore, the restrict-to-range rule is applied, restricting the online signal to the range of the most recent $n_{I_t}$ observations.

In the top left panel of Figure 3.29 the time series and the corresponding online signal are displayed. Since the procedure using the geometric search yields exactly the same window

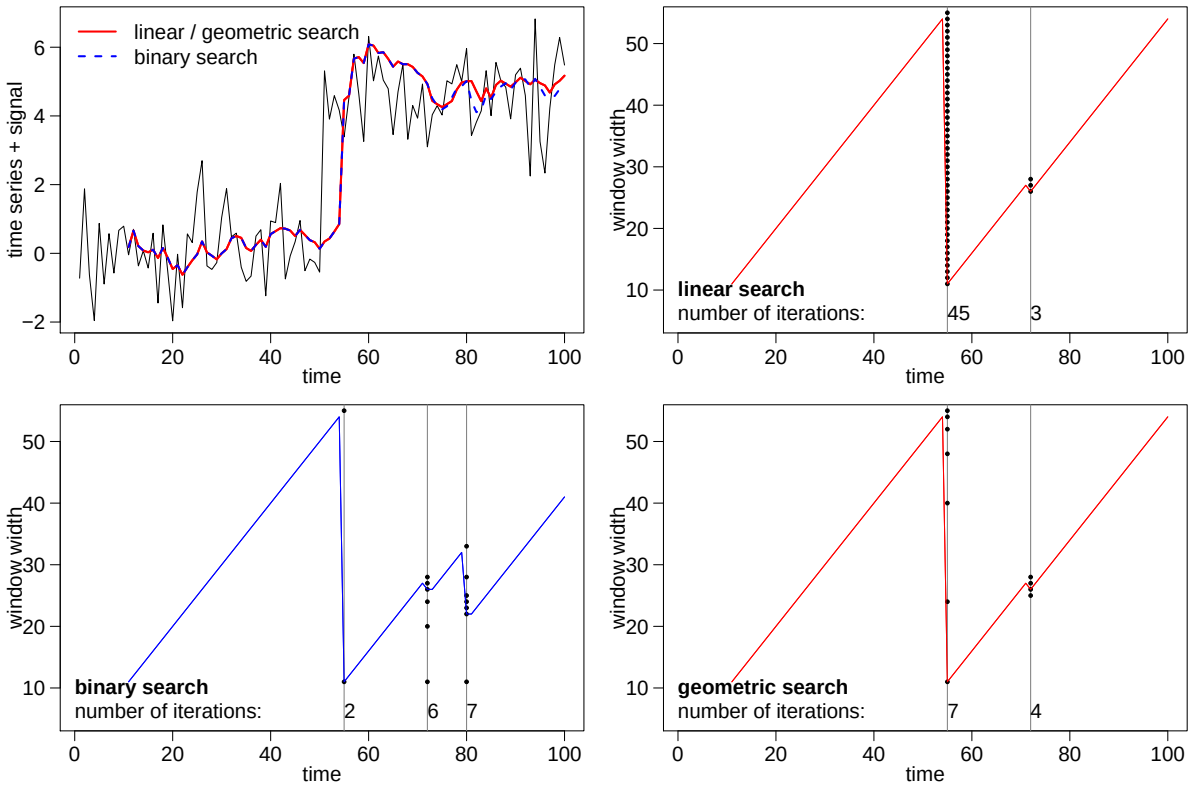


**Figure 3.29**: The top left panel shows a standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of five together with the online signal extracted by the adaptive online RM filter using different search algorithms. The remaining panels show the window width $n_t$ resulting from the specified search algorithm at each time $t$. Those times where more than one iteration was necessary for the final estimation are indicated by a grey vertical line; the black dots on the line mark the investigated window widths and the number below shows the number of iterations $\ell_t$ required at that time.

widths as the linear search, the corresponding signal estimations are identical. The other three panels show the window width $n_t$ determined by each of the three search algorithms and the number of iterations $\ell_t$ for those points in time where more than one iteration was performed (here: for $t \in \{52, 72, 80\}$).

It can be seen that although the binary search requires much less iterations if the window width has to be reduced drastically (here: after the detection of the level shift at $t = 55$), the geometric search is of advantage when the width is only reduced by a small amount, e.g. at time $t = 72$ because it then needs $\ell_t = 4$ instead of the $\ell_t = 6$ iterations required by the binary search. Furthermore, a smaller window width found by the binary search leads to another width reduction at $t = 80$ such that the binary search results in the smallest window widths and hence in a more variable signal, see e.g. the times $80 - 100$ in the top left panel of Figure 3.29. All these results imply that a geometric search algorithm seems to be the best choice for an online adaptive Repeated Median filter, because it is almost as fast as a binary search but may result in smoother signal estimations.

### 3.2.10 Conclusions

The proposed time series filter applying Repeated Median (RM) regression to a moving time window with a data-driven choice of the window width yields a smooth signal approximation; it is robust against isolated artefacts and small patches of outlying values. Furthermore, it is able to trace sudden changes with small time delay without applying different tests for artefact, trend change, and level shift detection. The filter does not require many parameter specifications, it does not put strong assumptions on the underlying data structure, and it is applicable to time series containing missing values.

For the best suitable choice of the window width in online applications, we recommend a filter which chooses the window width based on the sum of a fixed number of the most recent signs of RM residuals. This number of residual signs should be independent of the window width used for the signal estimation. For adaptation to sudden data changes with minimal time delay, we suggest to use a small number of say, fifteen, most recent residual signs. However, this number should be chosen based on the application-oriented background, e.g. on the maximally acceptable time delay and the frequency of measurement. For example, in applications to high-frequency data like time series from an online monitoring system in intensive care, one may take a larger number of residual signs, e.g. at the 30 most recent points in time, to ensure that medically irrelevant sudden changes are ignored while relevant ones are traced with acceptable time delay.

To guarantee the applicability of the filter in real-time, we propose to use a geometric search algorithm for the window width because it provides the best compromise solution between a small computation time and a smooth signal estimation.

Like the filters described in the previous section, the ***ad***aptive ***o***nline ***Re***peated Median ***filter*** has been implemented as the function `adore.filter` in the `robfilter` package of the free statistics software R (Fried and Schettlinger, 2008) and is available on the CRAN server:

`http://cran.r-project.org/web/packages/robfilter/`.

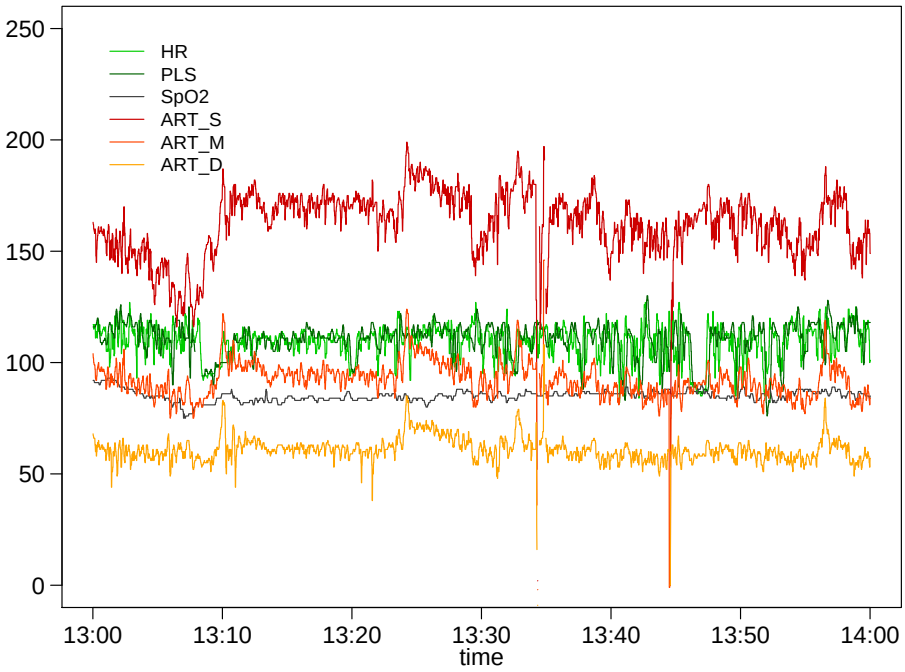


**Figure 3.30**: One hour of secondly measurements from a multivariate physiological time series including heart rate (HR) and pulse (PLS) in beats/min; oxygen saturation (SpO2) in per cent; and the systolic, mean and diastolic arterial blood pressures (ART_S, ART_M, ART_D) in mmHg.

## 3.3 The Multivariate Adaptive Online Repeated Median Filter

Current intensive care monitoring systems consider the observed time series for each physiological variable separately. However, regarding such a set of univariate time series as one multivariate time series allows for taking into account the correlation structure between the single components. Using the extracted multivariate signals instead of the raw measurements as input to a monitoring system seems a promising approach to reduce the false positive alarm rate of current alarm systems.

Figure 3.30 shows a part of a physiological multivariate time series measured at an intensive care unit at a frequency of once per second. Evidently, positive correlations are present – especially between heart rate and pulse, and between the three arterial blood pressures. Accounting for these multivariate features means incorporating possibly very important information into the monitoring and judgement process of the patient's health status.

Simply generalising univariate robust regression methods, as described in the previous sections, to the multivariate case does not result in affine equivariant procedures which causes reduced efficiency for the signal estimates, in particular if the components of the multivariate series are highly correlated (Chakraborty, 1999). On the other hand, multi-

variate affine equivariant regression methods with high breakdown point usually require the data to be in general position. This is often not the case for intensive care data since the measurements are recorded on a discrete scale.

Lanius (2005) and Lanius and Gather (2007) propose a multivariate signal extraction procedure for delayed signal extraction which provides a robust and reasonably efficient solution to this trade-off problem and is particularly designed for the application to intensive care data. Their proposal is based on a moving time window and it consists of two separate regression steps where first, univariate Repeated Median regression and second, multivariate Least Squares regression is performed. Similar to the (univariate) nested filters of Bernholt, Fried, Gather and Wegener (2006) described in Section 3.1.3, the sample is trimmed between these two regression steps in order to detect and replace (possibly multivariate) outliers where the outlyingness of an observation is judged w.r.t. the local covariance structure.

In this section a multivariate, robust regression-based filter for online signal extraction from multivariate high frequency time series is proposed following the lines of Borowski, Schettlinger and Gather (2009) and based on the approach of Lanius and Gather (2007). Unlike parallel univariate filters, the new procedure takes into account the local covariance structure between the single time series components. Since it is based on high-breakdown estimates, it is robust against (patches of) outliers in one or several of the components as well as against outliers with respect to the multivariate covariance structure. Moreover, this multivariate method makes use of the window width adaptation described in the previous section for univariate signal extraction and thus, overcomes the problem of the optimal choice of the window width. Hence, the new filter combines the advantages of the univariate adaptive online RM filter (Section 3.2) and the multivariate filter by Lanius and Gather (2007).

Analogous to the univariate case (3.1), we assume for a multivariate, $k$-dimensional time series $\boldsymbol{Y}_t = \big(Y_t(1), \ldots, Y_t(k)\big) \in \mathbb{R}^k$, $t \in \mathbb{Z}$, a simple additive working model:

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t + \boldsymbol{\nu}_t \,, \tag{3.22}$$

where $\boldsymbol{\mu}_t = \big(\mu_t(1), \ldots, \mu_t(k)\big)'$ denotes the $k$-dimensional, real signal vector at time $t$; $\boldsymbol{\epsilon}_t \in \mathbb{R}^k$ denotes an error term and $\boldsymbol{\nu}_t \in \mathbb{R}^k$ indicates an outlier generating process which most likely consists of the $k$-dimensional origin $\mathbf{0}_k$ most of the time, but may also indicate outliers in one, several, or all $k$ components.

Furthermore, our working assumption includes that each component of the multivariate signal is locally linear in a time window of width $n$, i.e. we assume (3.2) for each component $j = 1, \ldots, k$:

$$\mu_{t+i}(j) \;\approx\; \mu_t(j) \;+\; \beta_t(j) \cdot i \quad \text{for } i = -n+1, \ldots, -1, 0. \tag{3.23}$$

Due to these assumptions it is possible to evaluate the multivariate online signal estimate $\hat{\boldsymbol{\mu}}_t$ by fitting a multivariate regression model

$$\boldsymbol{Y}_{t+i} = \boldsymbol{\mu}_t + \boldsymbol{\beta}_t \cdot i + \boldsymbol{\varepsilon}_{t,i} \quad \text{for } i = -n+1, \ldots, -1, 0, \tag{3.24}$$

where $\boldsymbol{Y}_{t+i} \in \mathbb{R}^k$ describes the $n$ random variables in the current time window, $\boldsymbol{\mu}_t \in \mathbb{R}^k$ denotes the signal vector consisting of the $k$ components $\mu_t(j)$, $\boldsymbol{\beta}_t \in \mathbb{R}^k$ corresponds to the slope vector in the current window with components $\beta_t(j)$ and $\boldsymbol{\varepsilon}_{t,i} \in \mathbb{R}^k$ denotes an error term which is assumed to come from a symmetric distribution with zero median and (time-dependent) covariance matrix $\mathrm{Cov}(\boldsymbol{\varepsilon}_{t,i}) = \boldsymbol{\Sigma}_t \in \mathbb{R}^{k \times k}$ which may contain non-zero entries aside from the main diagonal.

To begin with, we describe the multivariate online filtering procedure based on a fixed window width, extending this to a procedure which adapts the window width at each time $t$. Corresponding to the description of the univariate adaptive RM filter, we discuss the treatment of missing values and some modifications of the multivariate filter. Finally, we exemplarily compare several online filters by an application to the physiological time series shown in Figure 3.30.

### 3.3.1 The Procedure

#### Multivariate Online Filtering With Fixed Window Width

The filtering procedure described by Lanius and Gather (2007) estimates the signal in a time window $(\mathbf{y}_{t+i};\, i = -m, \ldots, m)$ of fixed odd width $n = 2m+1$ centered around time $t$. Since this means a delay of $m$ time units for each estimation, this filter is more suitable for retrospective analyses. Here, we describe an online version of this filter which uses a moving time window $(\mathbf{y}_{t+i};\, i = -n+1, \ldots, 0)$ with a width $n \in \mathbb{N}$ which estimates the signal at the most recent time $t$.

To achieve a multivariate online signal estimate by the Trimmed Repeated Median-Least Squares procedure (TRM-LS), the following steps have to be performed within each time window $(\mathbf{y}_{t-n+1}, \ldots, \mathbf{y}_t)$ of length $n$ and dimension $k$:

1. Use univariate RM regression to find the signal estimate $\hat{\mu}_t(j)$ and the slope $\hat{\beta}_t(j)$ according to (2.3) and (2.2) for each component $j = 1, \ldots, k$, and combine these to the $k$-dimensional level and slope estimates

$$\begin{aligned} \hat{\boldsymbol{\mu}}_t &= \big(\hat{\mu}_t(1), \ldots, \hat{\mu}_t(k)\big)' \\ \text{and} \quad \hat{\boldsymbol{\beta}}_t &= \big(\hat{\beta}_t(1), \ldots, \hat{\beta}_t(k)\big)'. \end{aligned}$$

2. Calculate the residuals $\mathbf{r}_{t+i} = \mathbf{y}_{t+i} - (\hat{\boldsymbol{\mu}}_t + \hat{\boldsymbol{\beta}}_t\, i\,)$ for $i = -n+1, \ldots, 0$.

3. Use a robust method to estimate the local error covariance matrix $\mathbf{\Sigma}_t \in \mathbb{R}^{k \times k}$ based on the sample of residuals $\mathbf{r}_{t+i} \in \mathbb{R}^k,\ i = -n+1, \ldots, 0$.

4. Determine $S_t := \left\{ i = -n+1, \ldots, 0 \ : \ \mathbf{r}_{t+i}^\top \mathbf{\Sigma}_t^{-1} \mathbf{r}_{t+i} \leq d_n \right\}$

   the set of time points within the window at which the residuals $\mathbf{r}_{t+i}$ have a squared Mahalanobis distance w.r.t. the local covariance structure which is not larger than a specified value $d_n$.

5. Perform multivariate Least Squares regression on the trimmed sample

$$\{(i, \mathbf{y}_{t+i}) \ : \ i \in S_t\}$$

   to obtain the signal and slope estimates $\hat{\boldsymbol{\mu}}_t^{TRM-LS} \in \mathbb{R}^k$ and $\hat{\boldsymbol{\beta}}_t^{TRM-LS} \in \mathbb{R}^k$.

As a suitable estimator for the local error covariance matrix $\mathbf{\Sigma}_t$ in step 3, Lanius and Gather (2007) suggest to apply a slightly modified version of the fast computable orthogonalised Gnanadesikan-Kettenring estimator (OGK) by Maronna and Zamar (2002) to the residuals $\mathbf{r}_{t+i}$ in the window. This estimator is based on the fact that the covariance between two variables $X$ and $Y$ can be expressed as $\mathrm{Cov}(X, Y) = \left(\sigma(X+Y)^2 - \sigma(X-Y)^2\right)/4$ where $\sigma(\cdot)$ denotes the standard deviation. Since the multivariate OGK estimator inherits the explosion breakdown point of the univariate method, used for estimating the standard deviation, a high breakdown method should be applied to guarantee robustness against outliers. In a comparison study, Lanius and Gather (2007) find the $Q_n$ scale estimator (Rousseeuw and Croux, 1993) to be a suitable candidate which possesses a maximum asymptotic breakdown point of 50% if the data are in general position.
Due to the discrete measurement scale of the considered intensive care data, window samples might contain collinear data. This might cause the estimate $\hat{\sigma}_t^{Q_n}(\cdot)$ to be close to, or even equal to zero. To prevent the singularity of the estimated covariance matrix $\widehat{\mathbf{\Sigma}}_t$ (which is required for the inversion at step 4), Lanius and Gather (2007) propose to use a lower bound of $\vartheta = 0.02$ for the univariate estimates of the standard deviation, i.e.

$$\hat{\sigma}_t(\cdot) = \max\{\hat{\sigma}_t^{Q_n}(\cdot), \vartheta\}\,. \tag{3.25}$$

At step 4 an upper bound $d_n$ for the squared Mahalanobis distance of each $k$-dimensional residual vector determines whether an observation is regarded as an outlier or not. If the residuals were independently normal, the scaled squared distances had a $\chi^2$-distribution. Therefore, Lanius and Gather (2007) follow Maronna and Zamar (2002) in choosing a scaled $\chi^2$-quantile for 'hard' outlier rejection, i.e.

$$d_n \ = \ \frac{\chi^2_{k,\alpha} \cdot \mathrm{med}\left\{d_i;\ i = -n+1, \ldots, 0\right\}}{\chi^2_{k,0.5}},$$

where $\chi^2_{k,\alpha}$ is the $\alpha$-quantile of a $\chi^2$-distribution with $k$ degrees of freedom and the term $d_i = \mathbf{r}_{t+i}^\top \boldsymbol{\Sigma}_t^{-1} \mathbf{r}_{t+i}$, $i \in \{-n+1, \ldots, 0\}$, stands for the residual squared distances. Lanius and Gather (2007) choose this bound because for independent normal residuals the distribution of the squared distances asymptotically tends to a $\chi^2$-distribution.

### MULTIVARIATE ONLINE FILTERING WITH ADAPTIVE CHOICE OF THE WINDOW WIDTH

In the following let $n_{min}$ and $n_{max}$ specify the extreme values for the possible window widths, let $n_t(j)$ denote the window width used to evaluate the fit at time $t$ for component $j$, and let $n_{I_t}(j)$ denote the number of most recent residual signs which are considered for testing the adequacy of the current signal estimate in the $j$th component, analogous to the notation used in Section 3.2, with $j \in \{1, \ldots, k\}$.

Here, we propose a multivariate online filter with adaptive choice of the window width which replaces the univariate Repeated Median regression in the first step of the multivariate TRM-LS procedure, described in the previous section, by the *adaptive* univariate RM regression proposed in Section 3.2.

For the first step of this multivariate adaptive online Repeated Median filter the following has to be carried out:

1. (a) Use the adaptive univariate RM procedure to find the signal estimate $\hat{\mu}_t(j)$, the slope $\hat{\beta}_t(j)$ and the window width $n_t(j)$ for each component $j = 1, \ldots, k$, and use $n_{t,0} = \min\{N_{t-1} + 1, n_{max}\}$ as the starting window width for all $k$ components.

   (b) Set the overall window width for all components to

   $$N_t = \min\left\{n_t(j) : j = 1, \ldots, k\right\}.$$

   (c) Re-estimate $\hat{\mu}_t(j)$ and $\hat{\beta}_t(j)$ by univariate RM regression using the window width $N_t$ for all $j = 1, \ldots, k$.

   (d) Combine these values to the $k$-dimensional level and slope estimates

   $$\begin{aligned} \hat{\boldsymbol{\mu}}_t &= \left(\hat{\mu}_t(1), \ldots, \hat{\mu}_t(k)\right)' \\ \text{and} \quad \hat{\boldsymbol{\beta}}_t &= \left(\hat{\beta}_t(1), \ldots, \hat{\beta}_t(k)\right)'. \end{aligned}$$

The remaining steps of the procedure correspond to steps 2 to 5 of the TRM-LS procedure described above, with the difference that the fixed window width $n$ is replaced by the time-dependent overall window width $N_t$.

Similar to Section 3.2 the first online estimation takes place at time $t = n_{min}$ in a window of width $N_t = n_{min}$. For all other times $t$, the overall window width $N_t$ is set to the

minimum of the individual window widths $n_t(j)$, $j = 1, \ldots, k$, derived from the adaptive RM procedure in step 1.(a), to ensure that the local linearity assumption (3.23) holds for all components and thus, the remaining steps 2 to 5 can be applied.

### 3.3.2 Modifications and Treatment of Missing Values

For the proposed multivariate filter we also include the modifications proposed for the univariate case in Section 3.2.7. That means that for each component the signal value is restricted to the range of the most recent observations from this component, according to the restrict-to-range rule (3.21). Hence, outliers which date back some time have no influence on the range of the current signal estimate.

Furthermore, in Section 3.2 it was already stated that physiological time series from intensive care often contain missing values which might occur at single times; successively; in one, several or all components. Missing observations in the $j$th component lead to the missing of the $j$th component in the residual vector which in return impedes the estimation of the local covariance matrix $\mathbf{\Sigma}(t) \in \mathbb{R}^{k \times k}$ in step 3 of the procedure. Using only those residual vectors without missing components would mean an unnecessary loss of information for those components where the observations are present. Therefore, we suggest to replace missing observations – provided that 'enough' recent observations are present.

Generalising the proceeding described in Section 3.2 to the multivariate case, we request for each component $j \in \{1, \ldots, k\}$ at least $q$ non-missing observations at the recent $n_q$ points in time for the performance of an RM regression fit. Let $k_t \leq k$ denote the number of components which fulfil this requirement. For those $k_t$ components each missing observation in the window is replaced by the fitted value of the RM regression line. For the $k - k_t$ components with more than $q$ missing values at the most recent $n_q$ times, the signal is not estimated and the associated entries in the signal estimation vector are missing.

Obviously the residual components corresponding to the replaced observations equal the value zero. On the one hand this makes the estimation of $\mathbf{\Sigma}_t$ possible because the missing values are removed, on the other hand this causes a bias towards zero in the estimation. However, an implosion of $\widehat{\mathbf{\Sigma}}_t$ is prevented by the lower threshold (3.25) and the robustness against outliers is not influenced. This is indeed necessary since this matrix is used to detect outliers w.r.t. the local covariance structure.

According to the univariate case we propose to estimate the signal only for those components where at least $q = 5$ observations are present at the most recent $n_q = n_{I_t}$ time points, which are used to evaluate the goodness of fit.

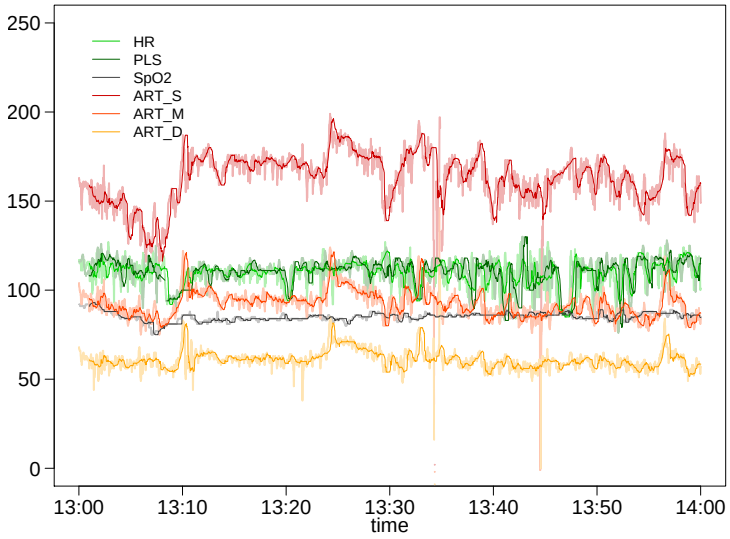


**Figure 3.31**: Multivariate physiological time series (semi-transparent colours) with componentwise online signal extraction by the univariate adaptive Repeated Median filter applying $I_t^{fixed}$ with $n_{I_t} = 30$, $n_{min} = 60$ and $n_{max} = 300$ seconds.

### 3.3.3 COMPARISONS

Here, the performance of the multivariate adaptive online Repeated Median filter is compared with its univariate counterpart as well as with the multivariate filter based on a fixed window width at an exemplary time series of hemodynamic variables (cf. Figure 3.30).

Figure 3.31 displays the extracted signals when separately applying a univariate adaptive RM filter to each of the components of the multivariate time series shown in Figure 3.30 with a minimal window width of $n_{min} = 60$ and a maximum window width of $n_{max} = 300$. Thus, any information older than five minutes will be disregarded, and the current state of the patient is judged only by recent observations.

Figure 3.32 shows the time-dependent window width $n_t$ used by the univariate RM filter for all variables. It shows that during this analysis for no component the maximum window width of five minutes is reached. Actually, the largest window width used (for the systolic arterial blood pressure) only amounts to $n_t = 167$ seconds.

The extracted online signal in Figure 3.31 runs relatively smooth, it is not influenced by outliers, e.g. around 13:35 or 13:45, and it reacts quickly to sudden changes, for example around times 13:25 or 13:30 for the arterial systolic blood pressure (ART_S). However, the parallel application of the univariate filter disregards possible dependences which might contain useful information for the signal extraction. For example, the sudden changes at the times given above, appear in all three systolic blood pressures at the same time and hence, they should evoke the same reaction for the window width chosen by the filter for all three components at the same time. This is not necessarily the case when regarding each

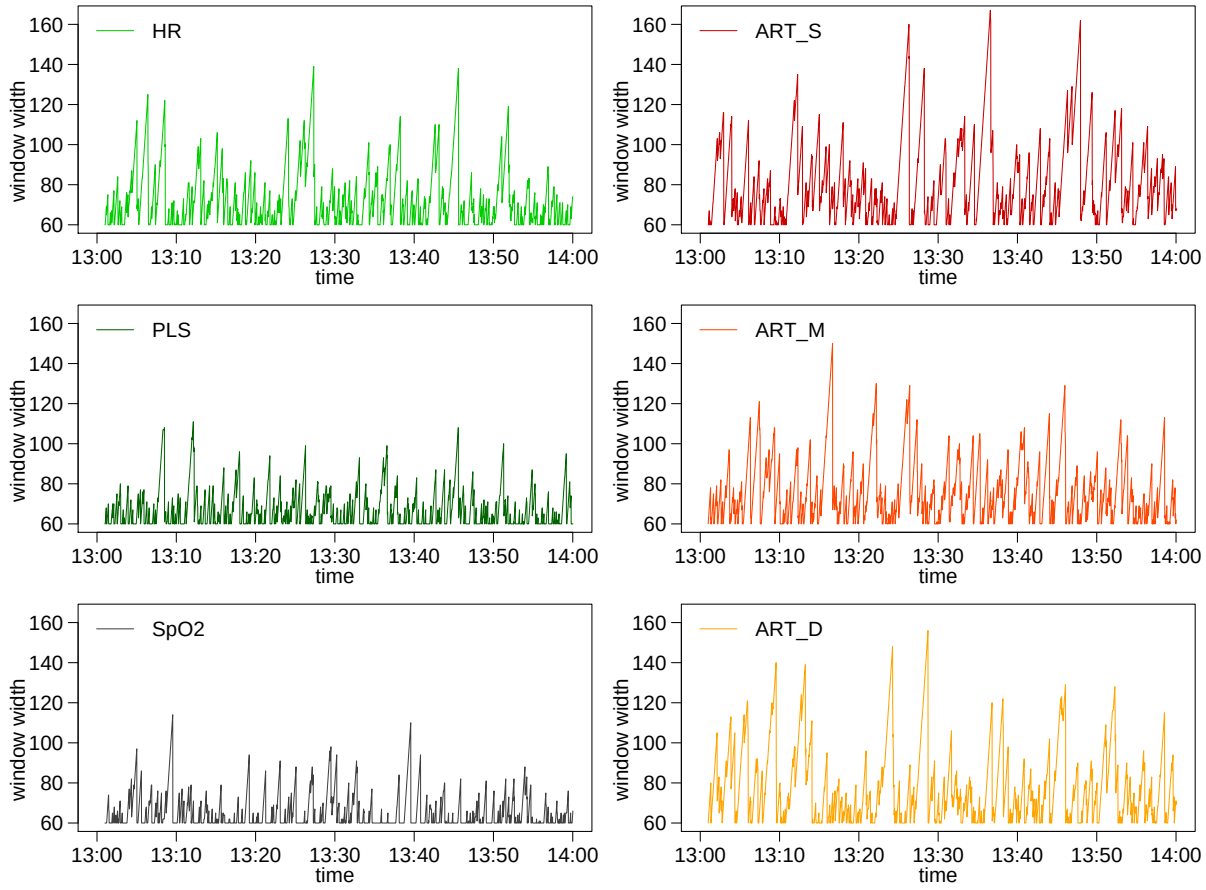


**Figure 3.32**: Window widths $n_t$ used by the univariate adaptive online Repeated Median filter for the single variables. The corresponding signals are shown in Figure 3.31.

component separately. Consider e.g. the window widths chosen for the three arterial blood pressures in the right hand side panels of Figure 3.32. One would expect the window width to increase, stay constant or decrease about at the same times at least for such highly correlated components, but the plots show obvious differences.

Figure 3.33 shows the signals extracted by a multivariate filter based on a fixed window width of 1.5 minutes, including the local correlation structure in the analysis. Comparing the window width of $n = 90$ seconds to the time-dependent window widths $n_t$ shown in Figure 3.32, it is no surprise that the extracted signal of this multivariate filter is generally smoother, because most of the time $n = 90$ is larger than the window widths used for the univariate component-wise RM filters. However, the delay in tracing sudden changes is larger for this multivariate filter, and possibly important short term trends are not traced properly, see e.g. the signal for the arterial systolic blood pressure around 13:10.

Applying the multivariate adaptive online RM filter proposed in this section, and using the same settings as for the univariate filter, yields an extracted signal which is shown in Figure 3.34. The effect of the restrict-to-range rule is here more obvious than for the univariate filter: some of the estimated signals look 'cut-off', e.g. the peaks of the signal for the arterial blood pressure around 13:10.

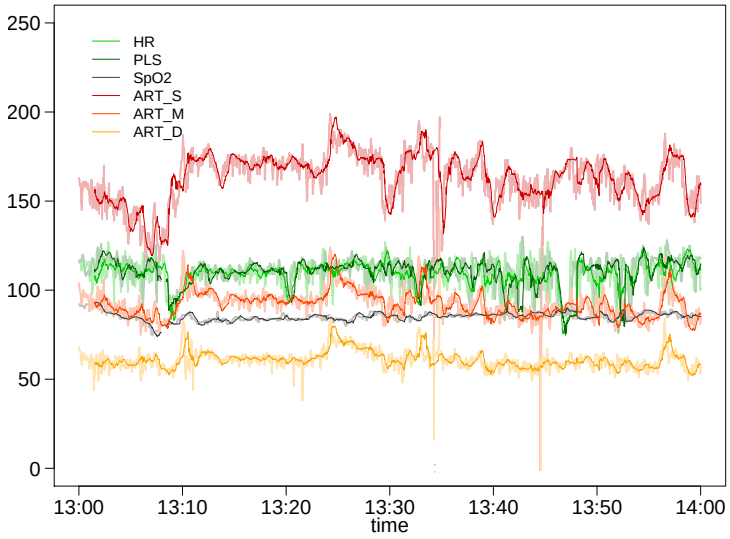


**Figure 3.33**: Multivariate physiological time series (semi-transparent colours) with signals extracted by the multivariate online Trimmed Repeated Median - Least Squares filter based on a fixed window width of 90 seconds.

| overall window width $N_t$ | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68-71 |
|---|---|---|---|---|---|---|---|---|---|
| relative frequency in % | 91.25 | 3.61 | 2.03 | 1.47 | 0.71 | 0.45 | 0.23 | 0.14 | 0.03 in each case |

**Table 3.8**: Relative frequencies in per cent for the overall window widths $N_t$ determined by the multivariate adaptive online Repeated Median filter with a window width range from $n_{min} = 60$ to $n_{max} = 300$ seconds.

Generally, the signal extracted by the multivariate adaptive filter reacts to sudden changes in the series as fast as the signal resulting from the univariate filter version, but unfortunately the former shows more variability. This is due to the fact that this filter generally chooses a quite small window width, because the overall window width for the multivariate Least Squares step is chosen as the minimum of the window widths determined for each component separately.

Table 3.8 lists the relative frequencies (in per cent) for the overall window width $N_t$ determined by the multivariate adaptive online Repeated Median filter in this application. For 91.25% of the analysed time points, the overall window width matches the minimal window width $n_{min} = 60$. Furthermore, the window width never exceeds a value of 71. Using such small window widths leads to a smaller efficiency of the adaptive multivariate RM filter compared to the other filters.

In order to increase the efficiency of this filter, it is sensible to use prior information on the correlation structure of the data and to apply the multivariate filter to blocks of highly correlated variables. Here, it is well known that heart rate and pulse are highly correlated as well as systolic, mean and diastolic blood pressures (Gather, Imhoff and Fried, 2002).

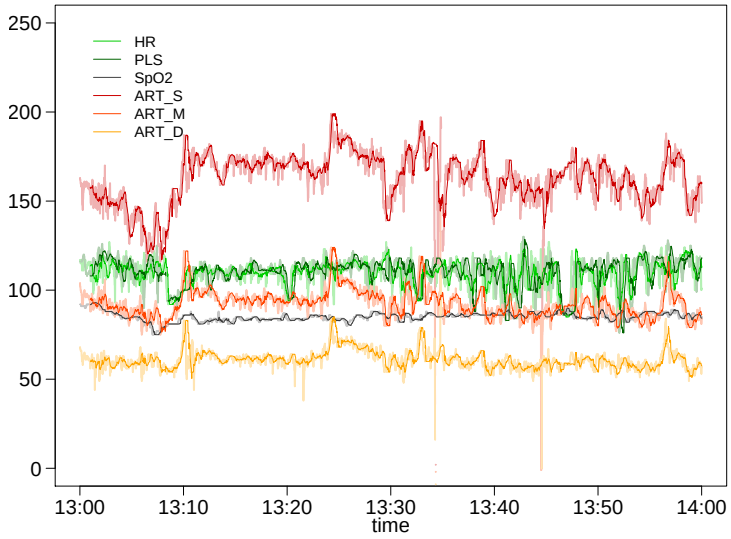


**Figure 3.34**: Multivariate physiological time series (semi-transparent colours) with signals extracted by the multivariate adaptive online Repeated Median filter with the window width (possibly) ranging from $n_{min} = 60$ to $n_{max} = 300$ seconds.

Therefore, we apply the multivariate adaptive filter including the restrict-to-range rule to these two 'blocks' of variables and use the univariate adaptive filter for the oxygen saturation. Figure 3.35 shows the signals resulting from this application; Figure 3.36 shows the corresponding window widths for the blood pressures as well as heart rate and pulse. The window widths for the oxygen saturation are displayed in Figure 3.32 (bottom left). The application of the multivariate filter to blocks of variables results in slightly smoother signal estimations than the simultaneous application of the multivariate filter to all components, see e.g. the time period 13:40-13:45 for the systolic arterial blood pressure in Figure 3.35 and Figure 3.34. This is due to the fact that larger window widths are chosen here (cf. Figure 3.36 and Table 3.8):
In about 43% of the time the overall window width $N_t$ chosen for heart rate and pulse is larger than that chosen by the multivariate filter applied to all components simultaneously; for the blood pressures this percentage amounts to approximately 54% . However, $N_t$ still matches $n_{min} = 60$ for 53.64% of the analysed time points for heart rate and pulse, and $N_t = n_{min}$ for 43.37% of the time for the blood pressures.

### 3.3.4 Summary

The multivariate adaptive online Repeated Median filter proposed in this section combines the advantages of the univariate adaptive online Repeated Median filter, introduced in Section 3.2, and the multivariate filtering procedure by Lanius and Gather (2007).
It separates relevant signals online from noise and, possibly multivariate, outliers; it is able to deal with missing values in one, several or possibly all components of the multivariate time series; it takes the correlation structure between the components into account; and it is able to choose the window width depending on the underlying data structure.

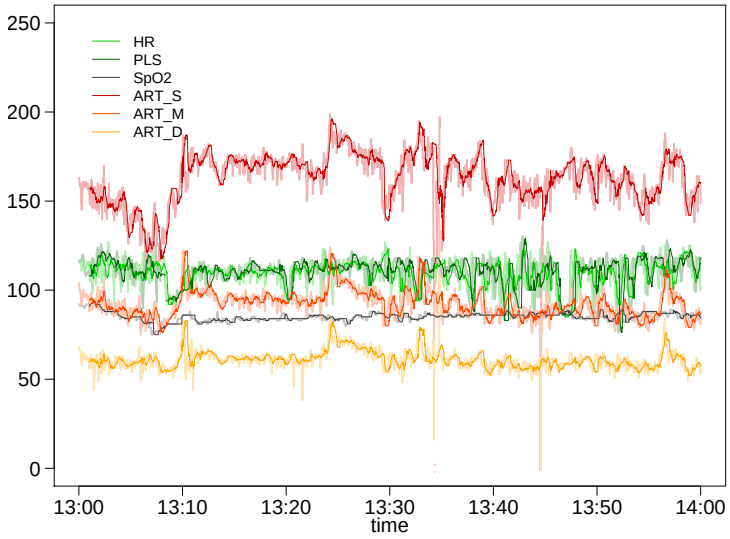


**Figure 3.35**: Physiological time series (semi-transparent colours) with signals from the multivariate adaptive online RM filter applied to the three arterial blood pressures and separately to heart rate and pulse. The signal for the oxygen saturation is extracted by the univariate adaptive RM filter. The minimum window width is set to 60, the maximum to 600 seconds.

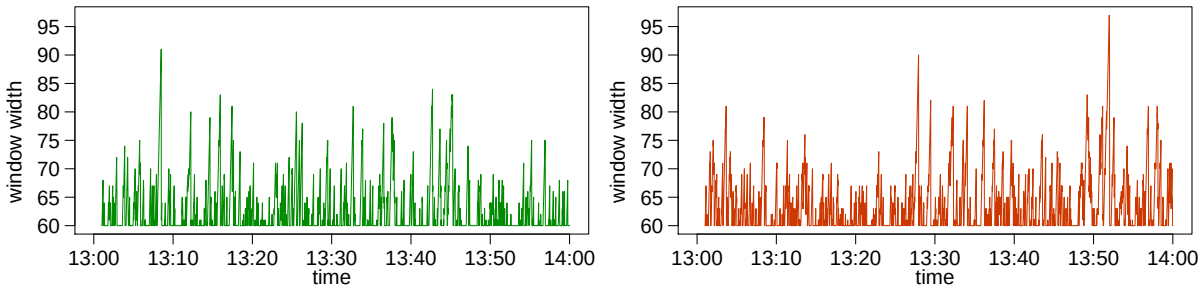


**Figure 3.36**: Window widths $n_t$ used by the multivariate adaptive online Repeated Median filter applied to the two-dimensional series of heart rate and pulse (left) and to the three arterial blood pressures (right). The corresponding signals are shown in Figure 3.35.

Application to a physiological time series from intensive care shows that this filter works quite well when applied to blocks of highly correlated variables. The extracted signals trace relevant changes in the time series with a very short time delay. However, this multivariate adaptive filter still lacks efficiency compared to the univariate adaptive RM filter. Several attempts to overcome this problem have not succeeded, for example choosing the overall window width equal to the mean or the median instead of the minimum of the component-wise window widths. Therefore, increasing the efficiency of this filter is still a challenge for future research. Furthermore, this filter is here only compared by means of an exemplary application. To find out more about its properties, extensive simulations and comparisons still have to be performed.

Analogous to the previous sections, an implementation of this ***m***ultivariate ***ad***aptive ***o***nline ***Re***peated Median ***filter*** is available as `madore.filter` function in the `robfilter` R-package (Fried and Schettlinger, 2008) on the CRAN server:

http://cran.r-project.org/web/packages/robfilter/.

# 4 Online Variability Extraction

The variability of a time series is an important feature which helps understanding, interpreting, and forecasting complex dynamic systems in various fields of application, especially when combined with information on other significant characteristics like e.g. the location of the signal level or the direction of a trend.

In intensive care the heart rate variability is for example used as a predictor for arrhythmias, for prediction of severity of illness and the mortality risk, or for choosing the right therapy for a patient. Changes in the variability of physiological variables contain useful information about the patient's state of health and thus, variability analysis is particularly useful in critical care units where a large number of variables is measured continuously and altered conditions have to be detected online since the patients are severely ill. An overview over several techniques to characterise the variability over time, especially for complex biological systems, can be found in Seely and Macklem (2004).

Since the time series considered here are measured at high frequencies, the methods applied for variability analysis should have a low computation time to deliver results with a minimal time delay. As could be seen in the previous chapter, such high frequency measurements typically lead to 'unclean' and noisy time series containing irrelevant outliers. Therefore, we focus on methods for variability extraction which possess certain robustness properties.

Physiological time series usually cannot be assumed to be generated by a mechanism following a 'clean' model: these series are generally not stationary because they can contain trends, sudden trend changes and level shifts and thus, classical time series techniques are not appropriate.

The scale estimators considered here, are able to deal with a certain amount of outliers, with trends or even shifts in the level, they are reasonably efficient, and they are computable online, i.e. they are able to present results in real-time. Since we assume the global structure of the observed time series to be unknown and quite complex, we focus on moving window techniques such that our assumptions are restricted to the local structure around a certain point in time.

In this chapter three different approaches for extracting the time-varying variability of a time series are compared: The first approach concerns scale estimators based on the residuals of a local linear regression; hence, we call it the *regression-based* approach. If the time series level is also of interest, the regression-based approach offers a reasonable and efficient possibility of estimating the variability around the underlying signal level (Gather and Fried, 2003). However, if the signal level is not estimated correctly this also affects the scale estimation.

The second approach includes methods which assume the time series to be locally linear but do not require the preceding regression step for the scale estimation. We call this the *regression-free* approach. Rousseeuw and Hubert (1996) propose such estimators which are based on the vertical heights of triangles formed by three observations from the sample. Here, two different regression-free scale estimators are investigated and a (yet unknown) finite sample breakdown is derived.

Third, an approach is considered where no assumptions on the underlying location, i.e. the time series level, are necessary and which is referred to as the *model-free* approach. This kind of scale estimator is particularly useful when the underlying level cannot be approximated by a locally linear trend. We derive the maximum finite sample breakdown point for the considered model-free scale estimators and discuss further properties.

In particular, the computational demands, the efficiency, and certain robustness properties like the breakdown point or the influence function of the estimators are discussed for each type of estimator. Definitions of these measures can be found in Appendix A.

For a systematic comparison of these three different types of robust online scale estimators with respect to their bias and mean squared error we conduct a simulation study where we compare (i) their finite sample efficiency, (ii) their behaviour in the presence of isolated and patchy outliers, (iii) their behaviour in the presence of level and scale shifts, (iv) their behaviour in the presence of a non-linear trend or volatility, and (v) their reaction towards temporal correlation. Furthermore, we compare their performance in an application to intensive care data and discuss their individual advantages.

Analogous to Section 3.1, we adopt the simple signal plus noise model (3.1) for a univariate time series $(Y_t)_{t\in\mathbb{Z}}$:

$$Y_t = \mu_t + \epsilon_t \,,$$

where $\mu_t$, the underlying signal at time $t$, is assumed to be relatively smooth apart from some sudden trend changes or level shifts, and

$$\epsilon_t = \varepsilon_t + \nu_t \,,$$

denotes an error term, where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a process coming from a symmetric distribution with median zero and (time-dependent) variance $Var(\varepsilon_t) = \sigma_t^2$, and $(\nu_t)_{t\in\mathbb{Z}}$ represents an outlier-generating process.

Our aim is to find adequate estimators for the variability of the errors $\varepsilon_t$ without time delay, reflecting the variability of the process at time $t$ without being influenced by outliers. Suppose that the scale $\sigma_t$ of the error term can be assumed to be approximately constant within a time window of width $n$, i.e.

$$\sigma_{t-n+i} \approx \sigma_t \,, \qquad i = 1, \ldots, n. \tag{4.1}$$

An online scale estimation can then be achieved by applying a scale functional to the observations within the most recent time window $(y_{t-n+1}, \ldots, y_t)'$.

Gather and Fried (2003) compare the finite-sample performance of several high-breakdown scale functionals, including the MAD, see equation (3.4) in Section 3.1.3, the length of the shortest half (Grübel, 1988) and the $S_n$ and $Q_n$ scale estimators (Rousseeuw and Croux, 1993). They find these estimators to be strongly biased if the underlying signal $\mu_t$ is locally not constant. This is due to the fact that for a non-constant signal, the application of a scale functional to the raw data does not result in estimation of the variability of the error term alone, but includes the estimation of the variability of the signal.
It is possible to remove the trend from the data prior to the scale estimation, e.g. by using robust regression estimators, resulting in regression-based scale estimates as described in the next section. This requires the assumption of a locally linear signal, cf. equation (3.2) in Section 3.1.1:

$$\mu_{t+i-n} \;\approx\; \mu_t \;+\; (i-n)\,\beta_t\,, \qquad i = 1, \ldots, n.$$

Another way of dealing with such trends is the application of scale functionals to trend-adjusted data, e.g. to a sequence of first differences. However, such techniques often show only small finite sample efficiencies as e.g. the median of the absolute first differences (Gather and Fried, 2003). Here, we consider online scale estimators which are unbiased at trended data, which are able to achieve reasonable efficiencies at standard normal data, and offer a certain robustness against outliers.
For the ease of notation, we will drop the time index $t$ in the following, such that the time window containing the $n$ observations considered for the estimation of the most recent scale $\sigma_t$ is denoted by $(y_1, \ldots, y_n)'$.

## 4.1 Regression-Based Scale Estimators

### 4.1.1 Definition

Gather and Fried (2003) propose to estimate the variability of the error term $\varepsilon$ by applying a robust scale functional to the residuals $\{r_i;\ i = 1, \ldots, n\}$ resulting from a robust linear regression fit to the current window. In particular, they propose to use residuals from Repeated Median regression (Siegel, 1982), cf. equations (2.2) and (2.3) in Chapter 2.
Let $r_i$ for $i = 1, \ldots, n$ denote the residuals from Repeated Median regression, defined in equation (2.7) and illustrated in Figure 4.1. The RM-based $Q_n$ scale estimator is then defined as an order statistic close to the quartile of the absolute differences of the RM residuals

$$Q_n = c_{Q_n}(n) \cdot \{\, |r_i - r_j| \;:\; 1 \le i < j \le n \}_{\binom{\lfloor n/2 \rfloor + 1}{2}} \tag{4.2}$$

where $c_{Q_n}(n)$ denotes a factor for achieving unbiasedness at samples of size $n$ coming from a model with specified error distribution.

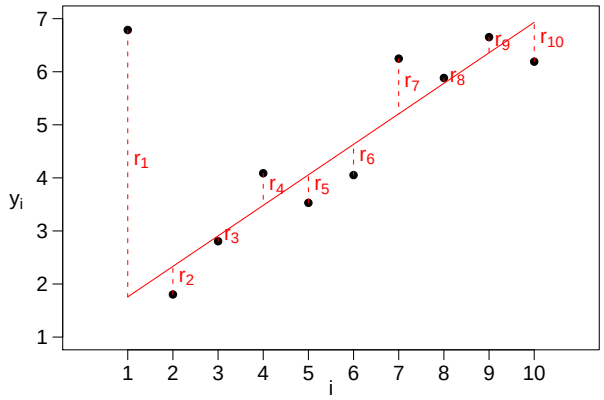


**Figure 4.1**: Repeated Median regression line in a window containing $n = 10$ observations. The variability of the error term can be estimated by applying the $Q_n$ estimator to the Repeated Median residuals $r_i$, $i = 1, \ldots, n$.

When focussing on the estimation of the variance $E(\varepsilon^2)$, or here: the standard deviation, the measure of scale requires the specification of an appropriate correction factor $c$ depending on the underlying error distribution. However, here we aim at estimating the variability of a time series which may also be represented by a multiple of the standard deviation of the true, but unknown, local error distribution. For the pure descriptive purpose of monitoring the time-dependent scale, distributional assumptions on the noise component (e.g. the existence of second order moments) are not required, and the scale estimates may be computed omitting the correction factors. The sequence of scale estimates will, up to a scalar factor, be the same.

### 4.1.2 PROPERTIES

For a scale estimator the breakdown point is determined by the minimal amount of contamination such that the estimated scale becomes either infinite (*explosion*) or zero (*implosion*), see Appendix A. For data in general position the finite sample breakdown point of $Q_n$ corresponds to $\lfloor n/2 \rfloor / n$, i.e. the estimate is bounded and stays away from zero even if almost 50% of the data are contaminated. Furthermore, it has a smooth and bounded influence function (Rousseeuw and Croux, 1993).

Gather and Fried (2003) compare several robust scale estimators for variability extraction from time series in a simulation study and find the $Q_n$ scale estimator to perform best. For independent Gaussian data, $Q_n$ is more stable than other high-breakdown scale estimators: it offers an asymptotic efficiency of 82% (relative to the empirical standard deviation) which is much larger than e.g. the asymptotic efficiency of the $MAD$, being only 36%.

In an online application to moving time windows of width $n$, the MAD can be updated in $\mathcal{O}(\log n)$ time (Bernholt, Fried, Gather and Wegener, 2006), while the application of the fastest offline algorithm for the $Q_n$ estimator needs $\mathcal{O}(n \log n)$ time (Croux and Rousseeuw, 1992).

Therefore, Nunkesser, Fried, Schettlinger and Gather (2008) propose an update algorithm for $Q_n$ which requires $\mathcal{O}(n \log n)$ time in a worst case scenario but runs much faster in practical applications. They compare the $Q_n$ estimator to the MAD, a 10%-trimmed standard deviation and a $\tau$-estimator of scale (Maronna and Zamar, 2002) in online applications to simulated time series with stationary errors and a real data example.

As compared to widely-used standard methods like the trimmed standard deviation, $Q_n$ leads to similarly good results in case of Gaussian data, but it provides a much higher resistance against large numbers of outliers.

Compared to other highly robust scale estimators, $Q_n$ yields less variable estimations than the MAD and performs similar to the $\tau$-estimator in case of GARCH(1,1) errors and for models with a piecewise constant variability. They find that, irrespective of the existence of autocorrelations, the proposed $Q_n$ update algorithm can directly be used in a nonparametric manner to estimate the marginal local volatility of a time series.

Moreover, due to its definition via pairwise differences, $Q_n$ does not require an estimate of the local mean and hence, it is less biased than its competitors, including the $\tau$-estimator, in the presence of outlier patches or level shifts. Fried (2007a) and Nunkesser, Schettlinger and Fried (2008) show that standardising a test statistic based on the medians of two time windows by $Q_n$ leads to robust tests for the detection of abrupt level shifts which are more powerful than tests using a standardisation by other highly robust and less efficient scale estimators (like the MAD).

However, the online scale extraction investigated by Fried (2007a), Nunkesser, Fried, Schettlinger and Gather (2008) and Nunkesser, Schettlinger and Fried (2008) is restricted to mean stationary time series and may become strongly biased for time series with a trend in the underlying signal.

Applying RM-based $Q_n$ (4.2) for online scale estimation avoids the bias for trended data. Using the update algorithm for $Q_n$ by Nunkesser, Fried, Schettlinger and Gather (2008) in combination with the linear time RM update algorithm by Bernholt and Fried (2003) an estimation can also be achieved in $\mathcal{O}(n \log n)$ time which makes the RM-based $Q_n$ scale estimator quite attractive for online application.

## 4.2 Regression-Free Scale Estimators

### 4.2.1 Definition

Rousseeuw and Hubert (1996) propose several scale estimators based on triangular heights. Here, we adapt their approach to data observed on an equidistant design space, namely discrete time. Within the recent time window, we consider the *vertical* heights of triangles

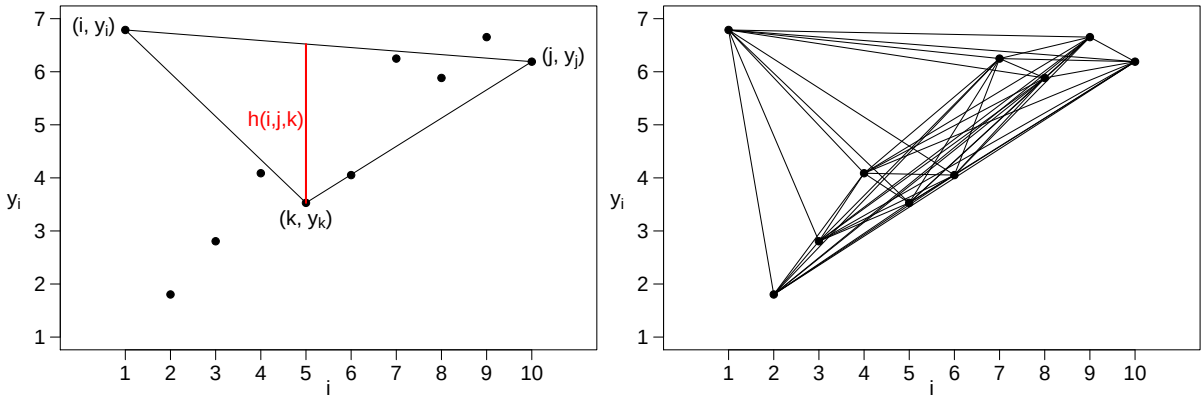


**Figure 4.2**: Illustration of a height $h(i, j, k)$ of a triangle as it is used for regression-free scale estimation (left panel). The $R$ and $Q^{\alpha}_{all}$ estimators both consider the heights of all possible $\binom{n}{3}$ triangles illustrated in the right panel.

with vertices given by three observations $y_i$, $y_j$ and $y_k$ with $i, j, k \in \{1, \ldots, n\}$:

$$h(i, j, k) = \left| y_j - y_i - \frac{(y_k - y_i)(j - i)}{(k - i)} \right| . \tag{4.3}$$

This vertical height does not match the 'height' defined as the perpendicular length of a triangle; the term 'height' as defined in (4.3) equals the non-zero residual of an $L_1$ regression fit to the three considered data points and is illustrated in the left panel of Figure 4.2.

Corresponding to the idea of the Repeated Median slope estimator (2.2), based on slopes between pairs of observations, Rousseeuw and Hubert (1996) propose a scale estimator based on a repeated median of the heights (4.3) formed by triples of observations:

$$R = c_R(n) \cdot \underset{i}{\text{med}} \left\{ \underset{j \neq i}{\text{med}} \left\{ \underset{k \neq i,j}{\text{med}} \, h(i, j, k) \right\} \right\} . \tag{4.4}$$

As alternative to this approach they also introduce an $\alpha$-quantile of the $\binom{n}{3}$ heights

$$Q^{\alpha}_{all} = c^{\alpha}_{Q_{all}}(n) \cdot \{h(i, j, k); 1 \leq i < j < k \leq n\}_{(\lfloor \alpha \binom{n}{3} \rfloor)} . \tag{4.5}$$

The terms, $c_R(n)$ and $c^{\alpha}_{Q_{all}}(n)$ denote factors which result in unbiasedness at a specified error distribution for samples of size $n$.

### 4.2.2 PROPERTIES

Let $\text{fsbp}^{+}(S, F_n)$ denote the finite sample explosion breakdown point of a scale estimator $S$ at an empirical distribution $F_n$ of a sample $\mathbf{y}_n$ of size $n$ according to definition (A.3) in Appendix A. Analogously, define $\text{fsbp}^{-}(S, F_n)$ the finite sample implosion breakdown point after (A.4).

According to (A.2) the finite sample breakdown point is then given by

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

Davies and Gather (2007) show that the upper bound for the finite sample breakdown point of affine equivariant scale functionals at samples in general position corresponds to $\lfloor n/2 \rfloor / n$, see Appendix A.

### Breakdown Point of $R$

Rousseeuw and Hubert (1996) prove for the $R$ estimator that for any sample in general position with empirical distribution function $F_n$ it is

$$\text{fsbp}^+(R, F_n) = \frac{\left\lfloor \frac{n-1}{2} \right\rfloor}{n} \quad \text{and} \quad \text{fsbp}^-(R, F_n) = \frac{\left\lfloor \frac{n}{2} \right\rfloor}{n} .$$

Thus, the breakdown point of $R$ is given by

$$\text{fsbp}(R, F_n) = \lfloor (n-1)/2 \rfloor / n \,, \tag{4.6}$$

which means that it almost attains the upper bound of $\lfloor n/2 \rfloor / n$.
Furthermore, Rousseeuw and Hubert (1996) point out that the finite sample breakdown point tends to a meaningful limit which they call *asymptotic breakdown point*. For the $R$ estimator this asymptotic breakdown point corresponds to $\text{abp}(R) = 50\%$.

### Breakdown Point of $Q_{all}^{\alpha}$

For the $Q_{all}^{\alpha}$ estimator Rousseeuw and Hubert (1996) state that the asymptotic explosion breakdown point corresponds to $\text{apb}^+(Q_{all}^{\alpha}) = 1 - \sqrt[3]{\alpha}$ while the asymptotic implosion breakdown point is given by

$$\text{apb}^-(Q_{all}^{\alpha}) = \begin{cases} \frac{1}{2} - \frac{1}{2}\cos(\varphi_\alpha) + \frac{\sqrt{3}}{2}\sin(\varphi_\alpha) & \text{if } 0 < \alpha < \frac{1}{2} \\ \frac{1}{2} & \text{if } \alpha = \frac{1}{2} \\ \frac{1}{2} + \frac{1}{2}\cos(\varphi_\alpha) + \frac{\sqrt{3}}{2}\sin(\varphi_\alpha) & \text{if } \frac{1}{2} < \alpha \leq 1 \end{cases}$$

where $\varphi_\alpha = \frac{1}{3}\arctan\left(\frac{\sqrt{\alpha(1-\alpha)}}{1/2-\alpha}\right)$. Thus, $Q_{all}^{\alpha}$ achieves its maximum asymptotic breakdown point of $\text{abp}_{max}(Q_{all}^{\alpha}) = 34.7\%$ when setting $\alpha = \alpha_{max}^{asymp} := 0.278$.
In Appendix D we derive the finite sample explosion and implosion breakdown points of $Q_{all}^{\alpha}$ at samples of size $n$ with empirical distribution function $F_n$. A graphical representation of $\text{fsbp}^+(Q_{all}^{\alpha}, F_n)$ and $\text{fsbp}^-(Q_{all}^{\alpha}, F_n)$ for different sample sizes $n$ is provided in Figure 4.3 which shows that the implosion breakdown point increases with increasing value of $\alpha$ while the explosion breakdown point decreases. For comparison, Figure 4.3 also shows the asymptotic explosion and implosion breakdown points.

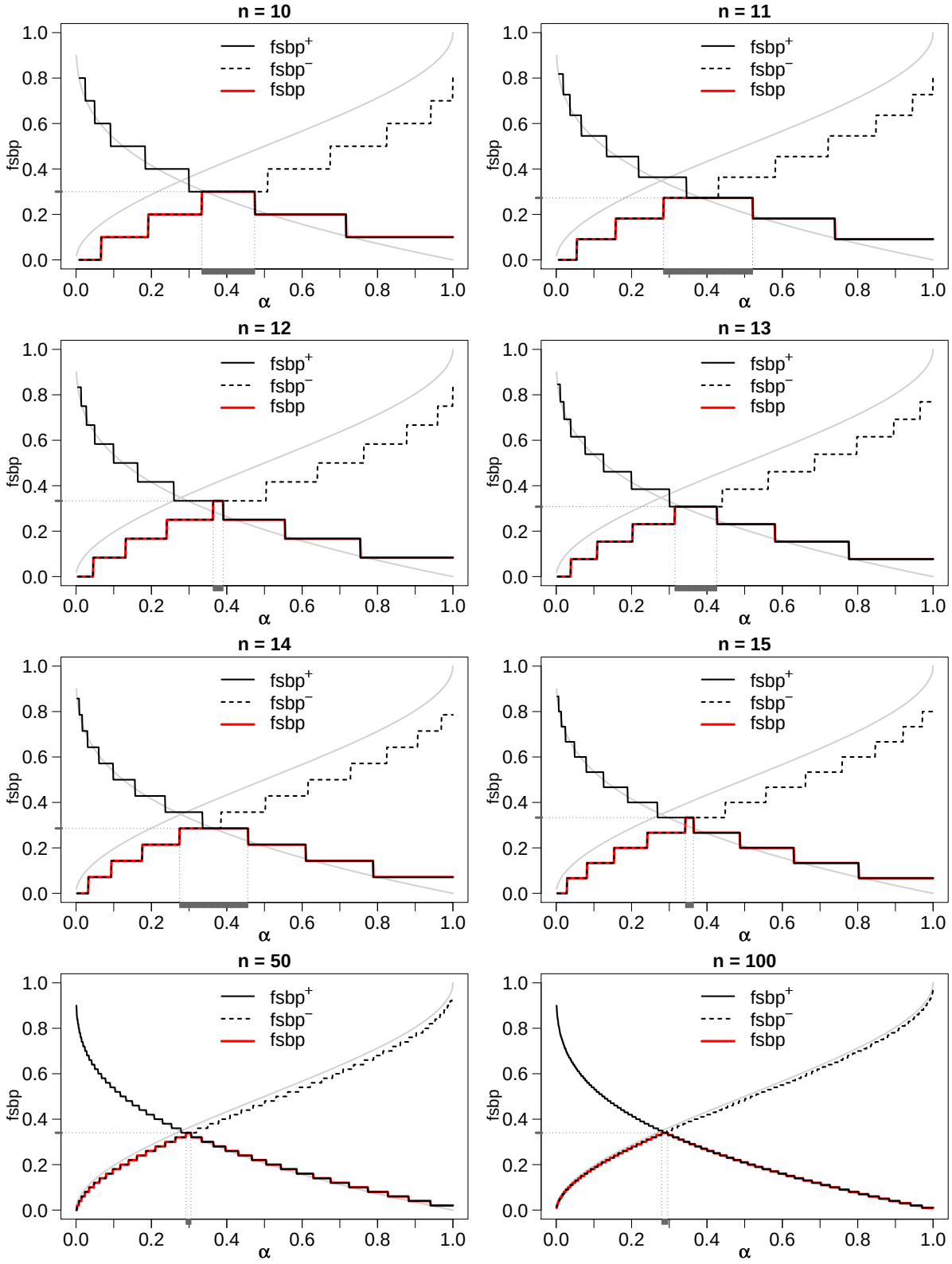


**Figure 4.3**: Finite sample explosion and implosion breakdown point of the regression-free $Q^{\alpha}_{all}$ estimator ($\text{fsbp}^{+}(Q^{\alpha}_{all})$ and $\text{fsbp}^{-}(Q^{\alpha}_{all})$) with resulting finite sample breakdown point $\text{fsbp}(Q^{\alpha}_{all})$ for all possible values of $\alpha \in [1/\binom{n}{3}, 1]$ and different sample sizes. The grey area at the x-axis marks the $\alpha$-interval for which $\text{fsbp}(Q^{\alpha}_{all})$ reaches its maximum value. The light grey lines in the background show the curves for the asymptotic explosion and implosion breakdown points $\text{abp}^{+}(Q^{\alpha}_{all})$ and $\text{abp}^{-}(Q^{\alpha}_{all})$.

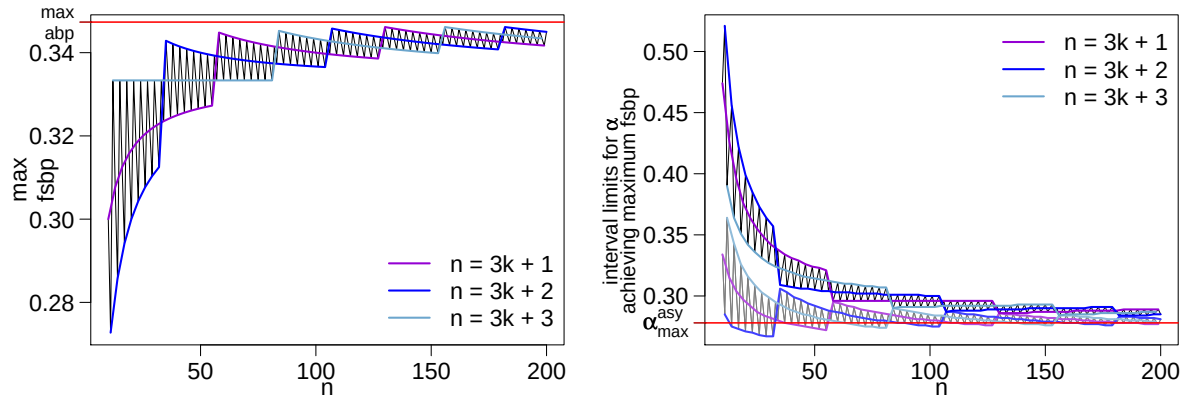


**Figure 4.4**: Maximum finite sample breakdown point $\text{fsbp}_{max}(Q^{\alpha}_{all})$ for sample sizes $n \in \{10, \ldots, 200\}$ in comparison with the maximum asymptotic breakdown point $\text{abp}_{max}(Q^{\alpha}_{all}) = 0.347$ (left), and corresponding interval limits for the values of $\alpha$ leading to a maximum fsbp, in comparison with the asymptotic value $\alpha^{asy}_{max}$ (right). The upper interval limits are plotted in full colours (black line), the lower interval limits are plotted in lighter colours (grey line).

Figure 4.3 shows that for each sample size $n$ the maximum finite sample breakdown point is reached for a range of values for $\alpha$. This interval is marked by a grey block on the x-axis in Figure 4.3. It can be seen that the length of this interval depends strongly on the sample size $n$. However, for increasing sample size, both interval limits tend to the asymptotic value of $\alpha^{asy}_{max} = 0.278$ and the maximum fsbp tends to the asymptotic breakdown point $\text{abp}_{max}(Q^{\alpha}_{all}) = 0.347$.

This shows even more obvious in Figure 4.4, displaying the maximum finite sample breakdown point for sample sizes $n \in \{10, \ldots, 200\}$ together with the corresponding interval limits for the values of $\alpha$ which lead to the maximum fsbp. The sawtooth pattern of the curves shown in both panels of Figure 4.4 can be explained by the fact that the estimator consists of a quantile which, for finite samples, corresponds to an order statistic with an integer order (here: $\lfloor \alpha \binom{n}{3} \rfloor$).

Additionally, Figure 4.4 shows that there is some sort of regular pattern in the curves of the breakdown point as well as the interval limits, depending on whether $n$, $n+1$ or $n+2$ is divisible by three. It is striking that – even when only considering every third value – both, the fsbp as well as the corresponding interval limits, do not tend monotonically to their asymptotic values but still in some sort of sawtooth pattern.

For example, for sample sizes $n \in \{3k + 1 : \ k \in \mathbb{N}\}$, where $n - 1$ is divisible by three, the maximum fsbp increases monotonically towards the asymptotic value for increasing $n$ up to $n = 55$, where at the same time not only the corresponding interval limits for $\alpha$ monotonically tend towards $\alpha^{asy}_{max}$, but also the length of the interval decreases. For $n = 58$, the maximum fsbp suddenly jumps to a value quite close to the maximum asymptotic breakdown point $\text{abp}_{max}$, but for sample sizes increasing from $n = 58$ to $n = 127$ the maximum fsbp *decreases* and thus increasingly departs from $\text{abp}_{max}$.

We also investigated the behaviour of the maximum fsbp curves at those sample sizes which are $k \in \{6, \ldots, 150\}$ steps apart, but did not find any monotonic behaviour. This finding is supported by the fact that the top peaks of the three coloured sawtooth curves in the left panel of Figure 4.4 appear for sample sizes from the set $\{12, 35, 58, 84, 107, 130, 156, 182\}$, and there is no obvious regularity in these values.

With the above results for $\text{fsbp}(Q^{\alpha}_{all}, F_n)$ it is possible to give a range of $\alpha$-values for which $Q^{\alpha}_{all}$ achieves its maximum finite sample breakdown point. However, the choice of $\alpha$ poses a trade-off problem between high efficiency and high robustness (in terms of high breakdown point), respectively. Therefore, for applications we propose to choose a value of $\alpha = 0.5$ as a compromise.

The estimator $Q^{0.5}_{all}$ still reaches an asymptotic breakdown point of $\text{abp}(Q^{0.5}_{all}) = 1 - \sqrt[3]{0.5} \approx 0.2063$ and achieves reasonable efficiency (cf. Section 4.4). For finite samples, the breakdown point might even be higher than the asymptotic value which will be shown below. This can also be seen, e.g. in the panel showing the fsbp for $n = 12$ in Figure 4.3, when comparing the finite sample breakdown point with the asymptotic value at $\alpha = 0.5$.

Although generally there is no simple expression for the implosion breakdown point $\text{fsbp}^{-}(Q^{\alpha}_{all}, F_n)$, it is possible to derive the following formula for quantiles close to the median (cf. Appendix D):

$$\text{fsbp}^{-}(Q^{\alpha}_{all}, F_n) \;=\; \frac{\lfloor (n-2)/2 \rfloor}{n} \qquad \text{for} \quad \alpha \in \left[ \frac{1}{2},\; \frac{1}{2} + \frac{1}{\binom{n}{3}} \right) . \tag{4.7}$$

For these values of $\alpha$, the explosion breakdown point $\text{fsbp}^{+}(Q^{\alpha}_{all}, F_n)$ corresponds to

$$\text{fsbp}^{+}(Q^{\alpha}_{all}, F_n) \;=\; \frac{1}{n} \lceil u + v + (n-1) \rceil \tag{4.8}$$

$$\text{with} \quad u \;:=\; \sqrt[3]{-3\left( \left\lfloor \alpha \binom{n}{3} \right\rfloor - 1 \right) + \sqrt{9\left( \left\lfloor \alpha \binom{n}{3} \right\rfloor - 1 \right)^2 - 1/27}}$$

$$\text{and} \quad v \;:=\; \sqrt[3]{-3\left( \left\lfloor \alpha \binom{n}{3} \right\rfloor - 1 \right) - \sqrt{9\left( \left\lfloor \alpha \binom{n}{3} \right\rfloor - 1 \right)^2 - 1/27}} .$$

Unfortunately $\text{fsbp}^{+}(Q^{\alpha}_{all}, F_n) \leq \text{fsbp}^{-}(Q^{\alpha}_{all}, F_n)$ for all $n \geq 6$, and thus the finite sample breakdown point of $Q^{\alpha}_{all}$ for $n \geq 6$ does not correspond to the short and simple expression (4.7) but to (4.8), i.e.

$$\text{fsbp}(Q^{\alpha}_{all}, F_n) \;=\; \text{fsbp}^{+}(Q^{\alpha}_{all}, F_n) \quad \text{for} \quad \frac{1}{2} \leq \alpha < \frac{1}{2} + \frac{1}{\binom{n}{3}} . \tag{4.9}$$

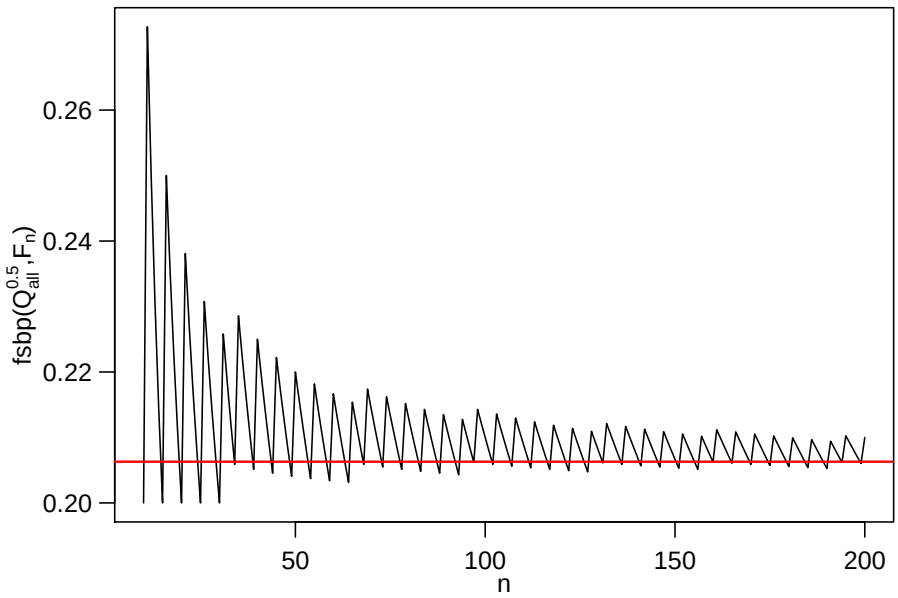


**Figure 4.5**: Finite sample breakdown point $\text{fsbp}(Q_{all}^{0.5}, F_n)$ of the $Q_{all}^{\alpha}$ estimator with $\alpha = 0.5$ for sample sizes $n \in \{10, \ldots, 200\}$ in comparison with its asymptotic breakdown point (red line).

Figure 4.5 displays the finite sample breakdown point of the $Q_{all}^{\alpha}$ estimator with $\alpha = 0.5$ for sample sizes ranging from ten to two hundred. It shows that for most sample sizes $\text{fsbp}(Q_{all}^{0.5}, F_n)$ is larger than the asymptotic value of about 20.6%, reaching values of up to 25% for $n = 10$ and ranging from 20.3% to 21.8% for sample sizes $n \in \{50, \ldots, 100\}$.

### Further Properties

Both estimators, $R$ and $Q_{all}^{\alpha}$ are regression invariant, i.e. a transformation $f(y_i) = y_i + i \cdot v$ of the observations $y_i$ with $i = 1, \ldots, n$ and $v \in \mathbb{R}$ does not change the outcomes of the scale estimation. However, a straightforward implementation of both regression-free estimators requires $\mathcal{O}(n^3)$ computation time. For online application, update algorithms should be used to reduce the computational complexity. Other properties of $R$ and $Q_{all}^{\alpha}$, particularly concerning their robustness, are to the best of our knowledge unknown so far.

The estimators (4.4) and (4.5) defined in this section are regression-free in the sense that they can estimate the variability of the error term around a locally linear signal without previously estimating the trend of data via a regression fit. However, they do require that the local linearity assumption (3.2) is valid. If this assumption is violated these methods also estimate the variability of the signal to some extent.

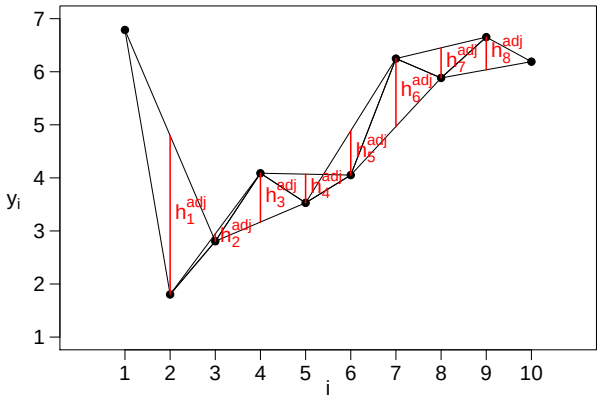


**Figure 4.6**: Illustration of the $n-2$ triangle heights $h_i^{adj}$, $i = 1, \ldots, n-2$, used for model-free scale estimation by the $Q_{adj}^{\alpha}$, $TM_{adj}^{\alpha}$ and $TMS_{adj}^{\alpha}$ estimators.

## 4.3 Model-Free Scale Estimators

Analogous to the methods described in the previous section, the scale estimators discussed in this section are regression-free, i.e. they do not depend on a local regression fit for the estimation of the variability of the error term.
Furthermore, they do not require any modeling of the underlying signal, and neither of the variability process. In this sense, the approach described here is not only regression-free but also *model-free*.

### 4.3.1 Definition

In this section we introduce an approach which, similarly to the $R$ and $Q_{all}^{\alpha}$ estimators (4.4) and (4.5), relies on triangle heights but here, the scale estimators are based on the $n-2$ vertical heights of triangles formed by three consecutive observations $y_i$, $y_{i+1}$ and $y_{i+2}$ with $i \in \{1, \ldots, n-2\}$, illustrated in Figure 4.6. For these heights formula (4.3) simplifies to

$$h_i^{adj} = \left| y_{i+1} - \frac{y_i + y_{i+2}}{2} \right|, \quad i = 1, \ldots, n-2. \tag{4.10}$$

Rousseeuw and Hubert (1996) propose a scale estimator based on an $\alpha$-quantile of these adjacent heights:

$$Q_{adj}^{\alpha} = c_{Q_{adj}}^{\alpha}(n) \cdot \{h_1^{adj}, \ldots, h_{n-2}^{adj}\}_{(\lfloor \alpha(n-2) \rfloor)}. \tag{4.11}$$

This estimator corresponds to the $\lfloor \alpha(n-2) \rfloor$-th value in the sequence of ordered heights $h_{(i)}^{adj}$, $i = 1, \ldots, n$, with $c_{Q_{adj}}^{\alpha}(n)$ a constant to achieve unbiasedness at a specified error distribution, depending on the window width $n$. Similar to the quantile of all triangle heights $Q_{all}^{\alpha}$ (4.5), the value of $\alpha$ controls the trade off between robustness and efficiency and will be addressed later on.

Gelper, Schettlinger, Croux and Gather (2009) further propose an $\alpha$-trimmed mean of adjacent heights

$$TM^{\alpha}_{adj} = c^{\alpha}_{TM_{adj}}(n) \cdot \frac{1}{\lfloor \alpha(n-2) \rfloor} \sum_{i=1}^{\lfloor \alpha(n-2) \rfloor} h^{adj}_{(i)} \tag{4.12}$$

and the square root of the $\alpha$ trimmed mean of squared adjacent heights

$$TMS^{\alpha}_{adj} = c^{\alpha}_{TMS_{adj}}(n) \cdot \sqrt{\frac{1}{\lfloor \alpha(n-2) \rfloor} \sum_{i=1}^{\lfloor \alpha(n-2) \rfloor} (h^{adj}_{(i)})^2}. \tag{4.13}$$

Again, the values $c^{\alpha}_{TM_{adj}}(n)$, and $c^{\alpha}_{TMS_{adj}}(n)$ are factors to achieve unbiasedness at a specified error distribution which also depend on the window width $n$ and the value of $\alpha$. The trimming proportion equals $(1-\alpha)$ where $\alpha$ can vary between zero and one. As for the $Q^{\alpha}_{adj}$ estimator, it controls the trade off between efficiency and robustness. Note that for $\alpha = 1$, the estimator defined in (4.13) is not robust and coincides with the residual variance estimator in nonlinear regression proposed by Gasser, Sroka and Jennen-Steinmetz (1986).

### Consistency Factors

For a normally distributed noise component in the local linear model (3.2), Gelper, Schettlinger, Croux and Gather (2008) derive the consistency factors for the three model-free scale estimators defined in (4.11), (4.12) and (4.13). In particular, they show that the asymptotic consistency factor for the $Q^{\alpha}_{adj}$ estimator is given by

$$c^{\alpha}_{Q_{adj}} = (Q^{\alpha}_{N})^{-1} \qquad \text{with} \qquad Q^{\alpha}_{N} := \sqrt{\frac{3}{2}}\, \Phi^{-1}\left(\frac{\alpha+1}{2}\right), \tag{4.14}$$

where $\Phi(z)$ is the standard normal cumulative distribution function at a value $z \in \mathbb{R}$, and the index $N$ refers to the assumption of normality.

Under the same assumptions, with $Q^{\alpha}_{N}$ as defined in (4.14), and $\varphi(z)$ the standard normal density associated with $\Phi(z)$, the asymptotic consistency factors of $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ can be written as

$$c^{\alpha}_{TM_{adj}} = \frac{\alpha}{\sqrt{6}\left[\varphi(0) - \varphi(\sqrt{2/3}\, Q^{\alpha}_{N})\right]}, \tag{4.15}$$

$$c^{\alpha}_{TMS_{adj}} = \frac{\sqrt{\alpha/3}}{\sqrt{\alpha/2 - \sqrt{2/3}\, Q^{\alpha}_{N}\, \varphi(\sqrt{2/3}\, Q^{\alpha}_{N})}}. \tag{4.16}$$

For example, for $\alpha = 0.5$ it is $c_{Q^{\alpha}_{adj}} = 1.21$, $c^{\alpha}_{TM_{adj}} = 2.51$, and $c^{\alpha}_{TMS_{adj}} = 2.16$.

The consistency factors (4.14), (4.15) and (4.16) are evaluated at the population level. However, extensive simulations showed that these values provide good approximations for finite samples of size $n = 20$ already (Gelper, Schettlinger, Croux and Gather, 2008).

Since the windows considered for online scale estimation are of limited size, the finite sample behaviour is indeed important here. To achieve unbiasedness at finite samples for a Gaussian distribution e.g. of the $Q^{\alpha}_{adj}$ estimator, the factor $c^{\alpha}_{Q_{adj}}$ can be replaced by its finite sample counterpart $c^{\alpha}_{Q_{adj}}(n)$ which can e.g. be obtained by means of Monte-Carlo simulations.

For the finite sample consistency factor $c^{0.5}_{Q_{adj}}(n)$ of the $Q^{\alpha}_{adj}$ estimator applying $\alpha = 0.5$ Gelper, Schettlinger, Croux and Gather (2008) derive the simple approximation formula

$$c^{0.5}_{Q_{adj}}(n) \approx 1.21 \frac{n}{n + 0.44}. \tag{4.17}$$

### 4.3.2 PROPERTIES

BREAKDOWN POINTS

Let $S^{\alpha}_{adj}$ denote any of the model-free scale estimators (4.11), (4.12) or (4.13) based on adjacent triangle heights. The maximum number of observations which may be replaced by arbitrary values within a sample of size $n$ with empirical distribution function $F_n$, such that the $S^{\alpha}_{adj}$ scale estimate remains positive, corresponds to

$$n\Delta(F_n) = \lfloor \alpha(n-2) \rfloor - 1 \,.$$

According to Davies and Gather (2005) an upper bound for the finite sample breakdown point of the considered scale estimators is then given by

$$\text{fsbp}(S^{\alpha}_{adj}, F_n) \leq \left\lfloor \frac{n - \lfloor \alpha(n-2) \rfloor}{2} \right\rfloor / n \,,$$

see definition(A.5) in Appendix A. Note that this upper bound is not reached for the model-free scale estimators $Q^{\alpha}_{adj}$, $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ as will be shown below.

For the $Q^{\alpha}_{adj}$ estimator Rousseeuw and Hubert (1996) derive the finite sample breakdown point in a regression setup with random design and come to the conclusion that the maximum asymptotic value for this estimator corresponds to $\text{abp}(Q^{\alpha}_{adj}) = 20\%$ when taking $\alpha = 0.4$.

However, here we consider a fixed design with equidistant time points which implies that possibly higher values can be achieved for the finite sample breakdown point. Thus, in the following we derive the breakdown point of $Q^{\alpha}_{adj}$ for an equidistant design and also derive expressions for the maximum finite sample breakdown point, depending on the sample size $n$. The results are also valid for the $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ estimators, since all of them only consider the $\lfloor \alpha(n-2) \rfloor$ smallest ordered adjacent triangle heights and thus, the finite sample breakdown point is the same for all three considered model-free scale estimators.

Let $\mathbf{y}_n$ denote a sample of size $n$ and $\mathbf{y}^k_n$ a sample where $k$ out of the $n$ values of the original sample $\mathbf{y}_n$ are replaced. Now, suppose that $\mathbf{y}_n$ is in general position and define

$$B := \lfloor \alpha(n-2) \rfloor \,.$$

If the replacement sample $\mathbf{y}_n^k$ 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 $B$th largest value of the ordered heights will be positive which implies for the finite sample implosion breakdown point for any of the model-free scale estimators $S_{adj}^{\alpha}$ that $\text{fsbp}^{-}(S_{adj}^{\alpha}, F_n) \geq B/n$. 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 $\text{fsbp}^{-}(S_{adj}^{\alpha}, F_n) \leq B/n$. We thus obtain for the finite sample implosion breakdown point as defined in (A.4)

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

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

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

Hence, the finite sample breakdown point (A.2) of any model free scale estimator $S_{adj}^{\alpha}$ defined by (4.11), (4.12) or (4.13) corresponds to

$$\text{fsbp}(S_{adj}^{\alpha}, F_n) = \frac{1}{n} \min \left\{ \left\lceil \frac{n-1-\lfloor \alpha(n-2) \rfloor}{3} \right\rceil, \lfloor \alpha(n-2) \rfloor \right\} . \tag{4.18}$$

Figure 4.7 shows that for increasing values of $\alpha \in [1/(n-2), 1]$, the explosion breakdown point is decreasing while the implosion breakdown point is increasing. However, the maximum value for $\text{fsbp}(S, F_n)$ depends not only on the choice of $\alpha$ but also on whether $n$ is divisible by four or not, see Figure 4.7 and Table 4.1. The derivations of the maximum finite sample breakdown point can be found in Appendix D.

Table 4.1 shows that, depending on $n$, more than one quantile might be chosen to achieve an estimate with maximum finite sample breakdown point with the order of the empirical quantile being $\lfloor \alpha(n-2) \rfloor \in \left\{ \left\lfloor \frac{n+1}{4} \right\rfloor, \ldots, n+1-3\left\lfloor \frac{n+1}{4} \right\rfloor \right\}$.

If collinear observations are rather expected than outliers, the best choice to prevent implosion is to set $\alpha$ to the maximal value within the range given in Table 4.1, i.e.

$$\alpha_{max}^{-} = \frac{n+1-3\left\lfloor \frac{n+1}{4} \right\rfloor}{(n-2)} \,.$$

However, if the aim is to prevent explosion it is recommendable to take the smallest empirical quantile possible, i.e. $\alpha$ should be set to

$$\alpha_{max}^{+} = \frac{n+1}{4(n-2)} \,. \tag{4.19}$$

| | maximum value $\text{fsbp}_{max}(S^{\alpha}_{adj}, F_n)$ | reached for $\alpha_{max} \in$ | 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 4.1**: Maximum values for the finite sample breakdown point $\text{fsbp}(S^{\alpha}_{adj}, F_n)$ with corresponding values of $\alpha$ and the rank $\lfloor \alpha(n-2) \rfloor$ of the adjacent triangle heights with $S^{\alpha}_{adj}$ representing one of the scale estimators $Q^{\alpha}_{adj}$, $TM^{\alpha}_{adj}$ or $TMS^{\alpha}_{adj}$.

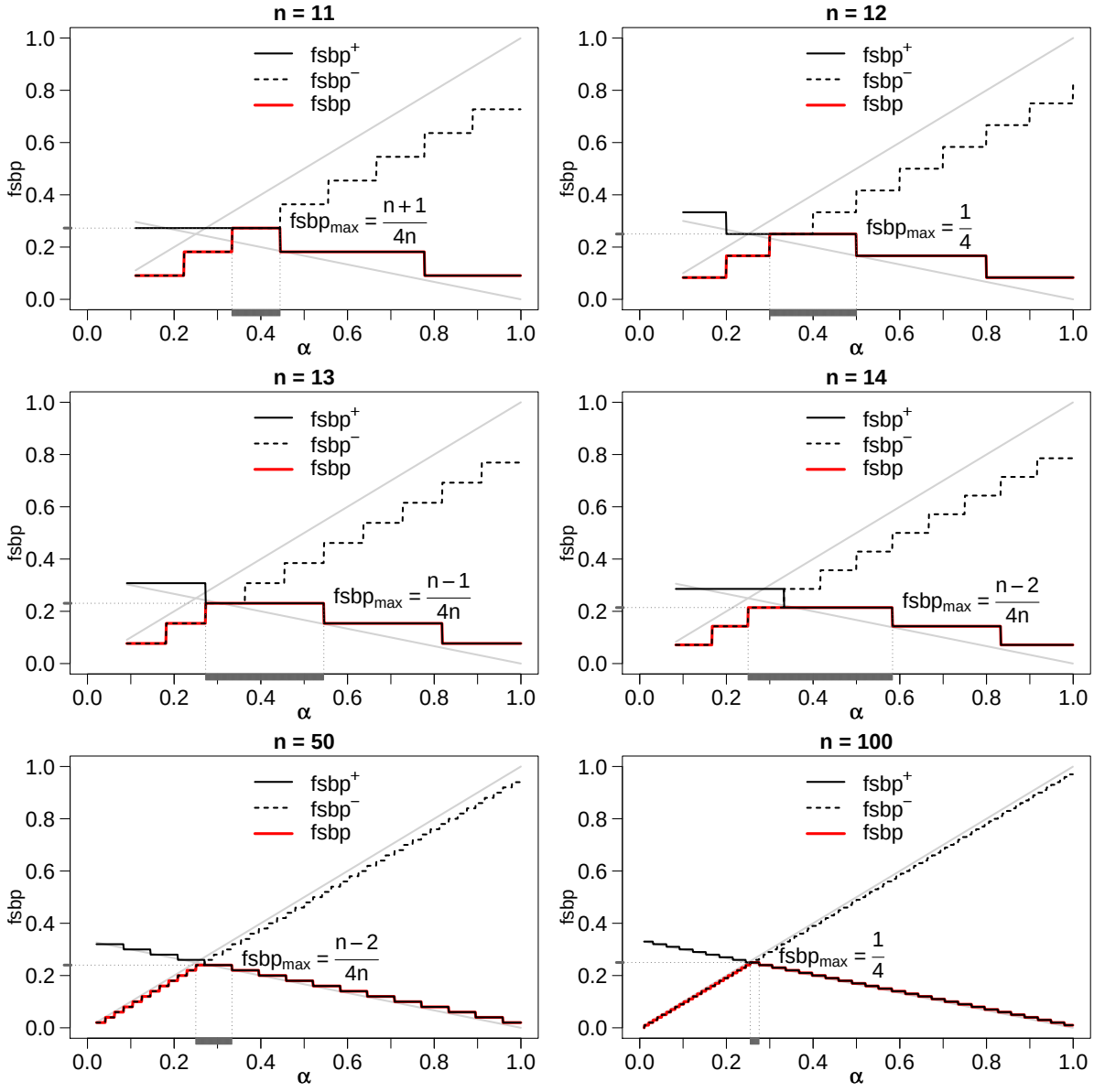


**Figure 4.7**: Finite sample breakdown points, explosion and implosion breakdown points of the model-free scale estimators $S^{\alpha}_{adj}$ for all possible values of $\alpha \in [1/(n-2), 1]$ and different sample sizes. The light grey lines in the background show the curves for the asymptotic explosion and implosion breakdown points $\text{abp}^{+}(Q^{\alpha}_{all})$ and $\text{abp}^{-}(Q^{\alpha}_{all})$.

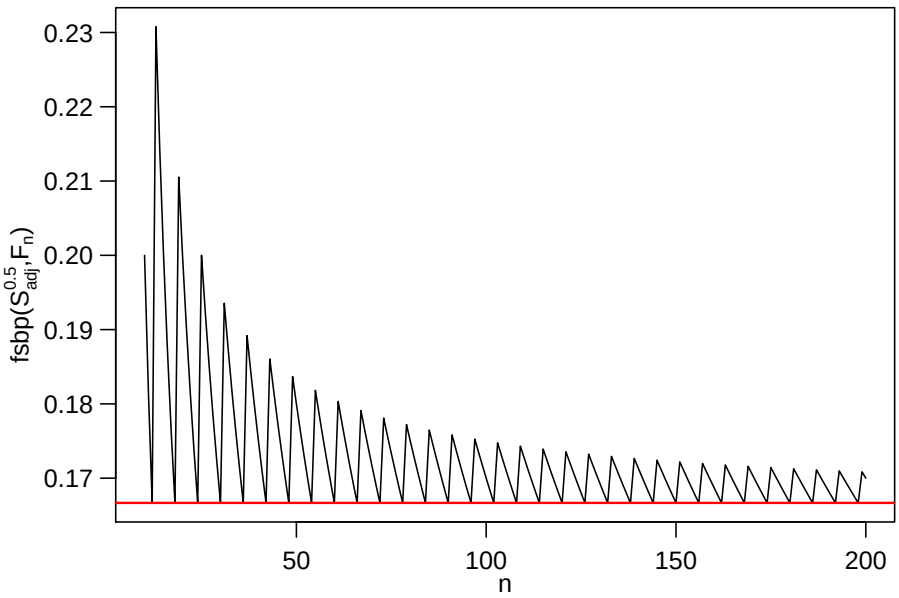


**Figure 4.8**: Finite sample breakdown point $\text{fsbp}(S^{0.5}_{adj}, F_n)$ of the model-free $S^{\alpha}_{adj}$ estimators with $\alpha = 0.5$ for sample sizes $n \in \{10, \ldots, 200\}$ in comparison with their asymptotic breakdown point (red line).

The bottom panels in Figure 4.7 illustrate that all interval limits for the $\alpha$ 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 $\text{abp}_{max}(S^{\alpha}_{adj}) = 0.25$ for $\alpha = 0.25$. For other values of $\alpha$, the asymptotic breakdown point equals $\text{abp}(S^{\alpha}_{adj}) = \min\{(1-\alpha)/3,\, \alpha\}$.

In particular, the finite sample breakdown point (A.1) for the model-free estimators applying $\alpha = 0.5$ corresponds to

$$\text{fsbp}^{+}(S^{0.5}_{adj}, F_n) \;=\; \left\lceil \frac{n - 1 - \lfloor (n-2)/2 \rfloor}{3} \right\rceil / n \,, \tag{4.20}$$

and the asymptotic breakdown point is given by

$$\text{abp}(S^{0.5}_{adj}) \;=\; 1/6 \,.$$

From Figure 4.7 it can be derived that the finite sample value might be larger than its asymptotic counterpart for some sample sizes, e.g. when considering the panel showing the results for $n = 13$. Figure 4.8 displays the finite sample breakdown point of the model-free scale estimators $S^{\alpha}_{adj}$ with $\alpha = 0.5$ for sample sizes $n \in 10, \ldots, 200$ in comparison with the asymptotic value of one sixth. It clearly shows that for all sample sizes the finite sample breakdown point is larger than or equal to 1/6, reaching values of up to 23% (for $n = 13$).

### INFLUENCE FUNCTIONS

The influence function (IF) quantifies the difference in estimated scale due to adding small amounts of outliers to the data (cf. definition (A.11) in Appendix A). Its derivation requires an appropriate definition of the considered functional. Therefore, we first introduce the functional forms of the considered model-free estimators before stating the results for the influence functions achieved by Gelper, Schettlinger, Croux and Gather (2008).

For the derivation of the influence functions, we consider the data to fulfill the assumptions of a local linear signal (3.2) and a locally constant scale within each time window (4.1) although these assumptions do not have to be fulfilled for application of the model-free estimators.

Because of the regression invariance of the considered estimators, we can assume without loss of generality that the underlying signal stays constantly at zero, i.e. the time series only consists of an error term and can locally (in a time window of length $n$) be represented by

$$y_i \;=\; \varepsilon_i \quad \text{for } i = 1, \ldots, n,$$

where for the error term we assume $\varepsilon_i \overset{\text{iid}}{\sim} F$ and $F$ denotes a probability distribution with mean zero and variance $\sigma^2$.

Considering observations sampled from a continuous distribution $F$, the corresponding triangle heights will also have a continuous distribution with distribution function $H_F$. In that case, the functional form of the $Q^{\alpha}_{adj}$ estimator (4.11) corresponds to

$$Q^{\alpha}_{adj}(F) \;=\; c^{\alpha}_{Q_{adj}} \cdot H_F^{-1}(\alpha) \tag{4.21}$$

with the consistency factor $c^{\alpha}_{Q_{adj}}$ defined according to (4.14).

The vertical heights $h_i^{adj}$ (4.10) 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 (3.1) are independent, then the vertical heights are only autocorrelated up to order two, and the estimator (4.11) will converge to (4.21).

Analogously, the functional forms of the estimators $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ can be derived: let $TM^{\alpha}_{p}$ denote a trimmed moment functional which is for a random variable $X \sim G$ defined as the $\alpha$-trimmed $p$th central moment to the power of $1/p$

$$TM^{\alpha}_{p} : F \mapsto TM^{\alpha}_{p}(F) = E(X^p \,|\, X \leq H_F^{-1}(\alpha))^{1/p}\,. \tag{4.22}$$

The functional form of the $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ estimators (4.12) and (4.13) is then given by

$$TM^{\alpha}_{adj}(F) \;=\; c^{\alpha}_{TM_{adj}} \cdot TM^{\alpha}_{1}(H_F) \tag{4.23}$$

and

$$TMS^{\alpha}_{adj}(F) \;=\; c^{\alpha}_{TMS_{adj}} \cdot TM^{\alpha}_{2}(H_F)\,. \tag{4.24}$$

The correction factors $c^{\alpha}_{TM_{adj}}$ and $c^{\alpha}_{TMS_{adj}}$ are defined according to (4.15) and (4.16). For the $Q^{\alpha}_{adj}$ functional (4.21) and a standard normal error distribution $F = N(0,1)$ the influence function is given by

$$\mathrm{IF}(x, Q^{\alpha}_{adj}, N(0,1)) = c^{\alpha}_{Q_{adj}} \frac{-G(Q^{\alpha}_N, x)}{2\sqrt{2/3}\ \varphi\left(\sqrt{2/3}\ Q^{\alpha}_N\right)}\,, \tag{4.25}$$

where $c^{\alpha}_{Q_{adj}}$ and $Q^{\alpha}_N$ are defined according to (4.14) and

$$\begin{aligned} G(Q^{\alpha}_N, x) \;=\; & -3\big(2\Phi(\sqrt{2/3}\,Q^{\alpha}_N) - 1\big) + \Phi(\sqrt{2}(Q^{\alpha}_N - x)) - \Phi(\sqrt{2}(-Q^{\alpha}_N - x)) \\ & + 2\big(\Phi(\sqrt{(4/5)}((x/2) + Q^{\alpha}_N)) - \Phi(\sqrt{(4/5)}((x/2) - Q^{\alpha}_N))\big). \end{aligned} \tag{4.26}$$

Let $M$ denote one of the moment based functionals defined in equations (4.23) and (4.24), then its influence function at the standard normal distribution $F = N(0,1)$ is given by

$$\begin{aligned} \mathrm{IF}(x, M, N(0,1)) \;=\; & \frac{c^p}{p\,\alpha}\Big[ -(Q^{\alpha}_N)^p\, G(Q^{\alpha}_N, x) - 3\frac{\alpha}{c^p} + \sqrt{2}\left(I^p_{\sqrt{2},\sqrt{2}} + I^p_{-\sqrt{2},\sqrt{2}}\right) \\ & + 2\sqrt{\frac{4}{5}}\left(I^p_{\sqrt{1/5},\sqrt{4/5}} + I^p_{\sqrt{1/5},-\sqrt{4/5}}\right)\Big]\,, \end{aligned} \tag{4.27}$$

with $p = 1$ and $c = c^{\alpha}_{TM_{adj}}$ (4.15) for $TM^{\alpha}_{adj}$ while $p = 2$ and $c = c^{\alpha}_{TMS_{adj}}$ (4.16) for the $TMS^{\alpha}_{adj}$ estimator, and the integral

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

can be computed numerically. For the derivation of the expressions of the influence functions (4.25) and (4.27) we refer to Gelper, Schettlinger, Croux and Gather (2008).

Figure 4.9 shows the influence functions of the the $Q^{\alpha}_{adj}$, $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ estimators for $\alpha = 0.25$ and $\alpha = 0.5$ for a standard normal distribution $F = N(0,1)$. All these influence curves exhibit three important properties: they are smooth, bounded and symmetric. Smoothness implies that a small change in one observation results in a small change of the estimated scale. Boundedness means that large outliers only have a limited impact on the scale estimator. For outlier sizes exceeding a certain threshold, all influence functions run (almost) parallel to the $x$-axis, i.e. the exact magnitude of the outlier is not of importance for the amount by which the bias of the scale estimation increases. Furthermore, all these influence functions are symmetric around zero, i.e. negative and positive outliers of the same size have the same effect on the scale estimation.

The influence functions of the three mode-free scale estimators with the same breakdown point, i.e. the same value of $\alpha$, are remarkably similar. Furthermore, Figure 4.9 shows the decreased robustness of the estimators for a larger value of $\alpha$: compared to the influence functions of the estimators based on $\alpha = 0.25$, the influence functions for $\alpha = 0.5$ converge at a slightly larger value for $x \to \pm\infty$.

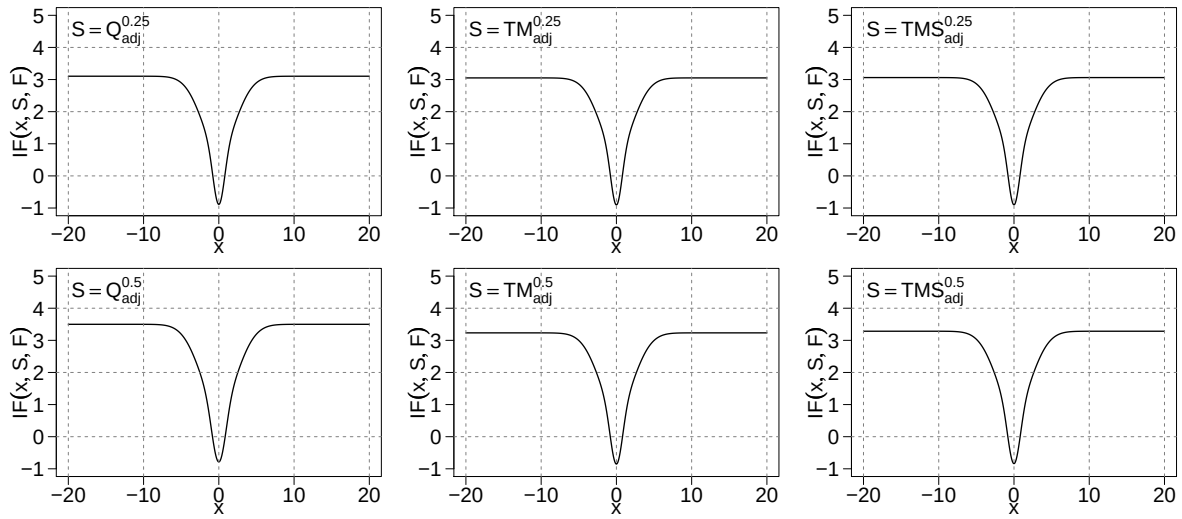


**Figure 4.9**: Influence functions of the $Q^{\alpha}_{adj}$, $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ estimators for $\alpha = 0.25$ (top) and $\alpha = 0.5$ (bottom) in case of an underlying standard normal distribution, i.e. $F = N(0, 1)$.

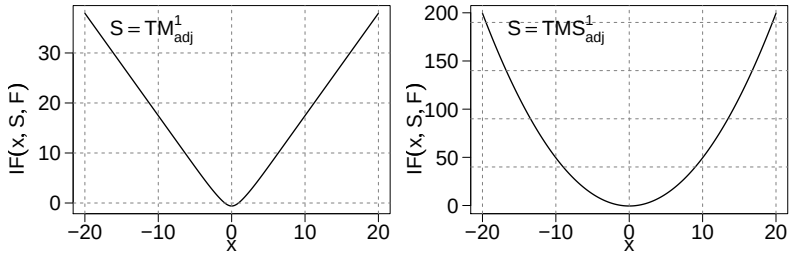


**Figure 4.10**: Influence functions of the non-robust $TM^{1}_{adj}$ and $TMS^{1}_{adj}$ estimators for an underlying standard normal distribution, i.e. $F = N(0, 1)$.

Figure 4.10 shows the influence functions of the non-robust estimators $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ with $\alpha = 1$. They are also smooth and symmetric around zero, but they are unbounded. As expected, the influence function of the $TMS^{1}_{adj}$ is quadratic, while the one of $TM^{1}_{adj}$ resembles the absolute value function. For decreasing values of $\alpha$, the difference between the influence functions of the two trimmed mean approaches becomes much less pronounced as can e.g. be seen in Figure 4.9.

### EFFICIENCY

The efficiency describes the precision of an estimator in relation to another estimator, see definitions (A.12) and (A.13) in Appendix A.

Gelper, Schettlinger, Croux and Gather (2009) discuss the efficiency of the robust model-free estimators $S^{\alpha}_{adj}$ relative to the non-robust mean of squared heights $TMS^{1}_{adj}$ shown in the left panel of Figure 4.11. Therefor they derive the asymptotic variances of these estimators for independent standard normal data and derive the asymptotic efficiencies (A.12) from that.

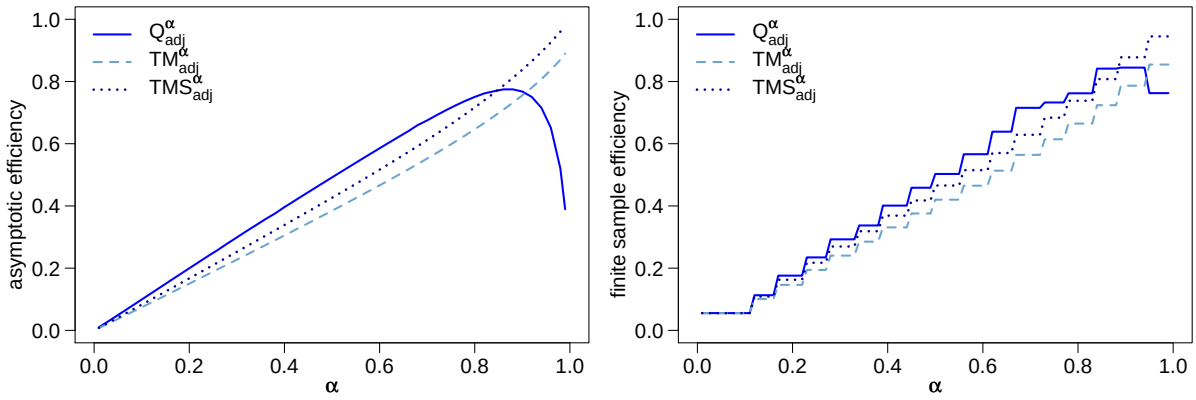


**Figure 4.11**: Asymptotic (left) and finite sample efficiency for a standard normal sample of size $n = 20$ (right) for the three model-free estimators $S^{\alpha}_{adj}$ for varying $\alpha \in (0, 1)$ in relation to the non-robust mean of squared heights estimator $TMS^{1}_{adj}$.

The $TMS^{\alpha}_{adj}$ estimator (4.13) with $\alpha = 1$ is equivalent to a scale estimator proposed by Gasser, Sroka and Jennen-Steinmetz (1986) who prove that for normally distributed data $(TMS^{1}_{adj})^2$ is asymptotically unbiased and strongly consistent for $\sigma^2$. These properties are also true for any other value of $\alpha$.

Figure 4.11 shows that, generally, the efficiency increases with increasing value of $\alpha$ for all considered estimators – asymptotically as well as for the finite sample case. The right panel of Figure 4.11 shows the finite sample efficiencies of the model-free estimators relative to $TMS^{1}_{adj}$ for a standard normal sample of size $n = 20$. It appears that the finite sample efficiency is reasonably close to the asymptotic efficiency, even for this small sample size.

For twenty observations there are $(n - 2) = 18$ triangle heights and thus, each estimator can only result in eighteen different estimates where each of the eighteen values can be achieved for $\alpha \in [k/(n-2), (k+1)/(n-2)]$, $k = 1, \ldots, n-3$. Hence, the curves of the finite sample efficiencies look like step functions.

Surprisingly, the quantile version of the model-free estimators is for $\alpha < 0.85$ more efficient than the trimmed mean of heights or even the trimmed mean of squared heights. Hence, replacing the quantile by a trimmed sum does not result in an increase of efficiency for a large range of values of $\alpha$. The $TMS^{\alpha}_{adj}$ estimator is slightly more efficient than the $TM^{\alpha}_{adj}$ estimator for all values of $\alpha$, but for small values of $\alpha$ the efficiencies of all estimators are quite similar. For values of $\alpha$ close to one, the efficiency of $Q^{\alpha}_{adj}$ is smaller than that of the other estimators, because the quantile tends to the maximum order statistic which is known to show a high variability.

For $\alpha = 0.25$ where the maximum asymptotic breakdown point of 25% is reached, $Q^{0.25}_{adj}$ only achieves an efficiency of 25%, and the trimmed mean estimators only achieve an efficiency of about 20%. Hence, the price paid for the maximal breakdown point is high.

Taking the median of the heights with $\alpha = 0.5$, results in an asymptotic efficiency of 49% for $Q^{\alpha}_{adj}$, in 38% for the $TM^{\alpha}_{adj}$ and in 43% for the $TMS^{\alpha}_{adj}$ estimator, while the asymptotic breakdown point equals 16.7%. The finite sample breakdown point for $n = 20$, however, is even higher because according to (4.20) it is $\text{fsbp}(S^{0.5}_{adj}) = \lfloor 10/3 \rfloor / 20 = 0.2$, allowing for four outliers in a window of size $n = 20$. Since this seems to provide a reasonable compromise between robustness and efficiency, we propose to choose $\alpha = 0.5$ for applications.

FURTHER PROPERTIES

One striking advantage of the model-free scale estimators introduced in this section is their low computation time: an update of any of the adjacent-type estimators can be achieved in only $\mathcal{O}(\log n)$ time. Furthermore, $Q^{\alpha}_{adj}$, $TM^{\alpha}_{adj}$ and $TMS^{\alpha}_{adj}$ do not rely on the local linearity assumption (3.2) but can also estimate the variability of the error term if the signal is non-linear in any sense, e.g. if it is quadratic or contains sudden level shifts or trend changes. In this sense, the adjacent-type estimators are model-free.

## 4.4 SIMULATION STUDY

In the following, we consider time series generated from the simple signal plus noise model $Y_t = \mu_t + \epsilon_t$ (3.1) with $t = 1, \ldots, T$ and the length of the time series corresponding to $T = 1000$. The simulation schemes we investigate, consist of time series with independent standard normal errors with and without contamination and of time series with autocorrelated errors. As contamination we consider isolated outliers, patches of outliers, and even level shifts. Furthermore, the effect of a sudden scale shift, the violation of the assumption of a locally constant scale and the violation of the linearity assumption for the underlying signal are examined. A description of this simulation study can also be found in Schettlinger, Gelper, Gather and Croux (2009).

For each setting $S = 1000$ time series are generated and the performance of the online scale estimators is judged at each point in time $t \in \{n, \ldots, T\}$ by the mean bias

$$\text{MB}_t = \frac{1}{S} \sum_{i=1}^{S} \frac{\widehat{\sigma}_t(i) - \sigma_t}{\sigma_t},$$

and the root mean squared error:

$$\text{RMSE}_t = \sqrt{\frac{1}{S} \sum_{i=1}^{S} \left( \frac{\widehat{\sigma}_t(i) - \sigma_t}{\sigma_t} \right)^2}.$$

| | $Q_{adj}^{0.5}$ | $TM_{adj}^{0.5}$ | $TMS_{adj}^{0.5}$ | $TM_{adj}^{1}$ | $TMS_{adj}^{1}$ | $R$ | $Q_{all}^{0.5}$ | $Q_n$ |
|---|---|---|---|---|---|---|---|---|
| $n = 20$ | 1.240 | 2.293 | 1.996 | 1.023 | 0.838 | 1.312 | 1.136 | 1.939 |
| $n = 50$ | 1.221 | 2.427 | 2.094 | 1.023 | 0.824 | 1.301 | 1.145 | 2.092 |

**Table 4.2**: Finite sample correction factors $c(n)$ for all scale estimators, achieving unbiasedness at normal samples of size $n \in \{20, 50\}$.

Here, $n$ denotes the window width, $\sigma_t$ is the true (uncontaminated) scale at that time, and $\widehat{\sigma}_t(i)$ is the online scale estimate at time $t \in \{n, \ldots, T\}$ for the $i$th simulated time series, $i \in \{1, \ldots, S\}$. The estimate $\widehat{\sigma}_t(i)$ is obtained by applying one of the considered scale estimators to the observations at times $(t - n + 1, \ldots, t)$ in time series $i$.

For the $Q_{all}^{\alpha}$ estimator (4.5) and the model-free scale estimators $Q_{adj}^{\alpha}$ (4.11), $TM_{adj}^{\alpha}$ (4.12) and $TMS_{adj}^{\alpha}$ (4.13), we consider the estimators with $\alpha = 0.5$ which offer a good compromise between a high breakdown point and high efficiency, see Sections 4.2 and 4.3. Furthermore, the non-robust alternatives $TM_{adj}^{1}$, i.e. the mean of all triangle heights, and $TMS_{adj}^{1}$, the square root of the mean of squared heights, are included for comparison.

The regression-based standard deviation

$$\text{sd} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (r_i^{LS})^2}\,, \tag{4.28}$$

calculated from least squares residuals $r_i^{LS}$ is included as non-robust alternative, and the $Q_n$ estimator based on the $RM$ residuals (4.2) is included as robust and regression-based reference method.

In the simulations we consider the two widths $n = 20$ and $n = 50$ for the windows which are moved over the whole length of the time series, resulting in $T - n + 1$ online scale estimates for each series. We use finite sample consistency factors $c(n)$ which are derived from simulations, where each of the estimators was applied to 10 000 standard normal samples of size $n \in \{20, 50\}$. The factors were calculated by inverting the mean of the 10000 scale estimates, to achieve an unbiased estimation of the standard deviation. They are given in Table 4.2. For an approximation of the correction factor for $Q_{adj}^{0.5}$ for large sample sizes the formula $c_{Q_{adj}}^{0.5}(n) \approx 1.21 \cdot n/(n + 0.44)$ could also be used, see Section 4.3 and Gelper, Schettlinger, Croux and Gather (2008).

A method performs best if it both, has a small mean bias $\text{MB}_t$ and is efficient, i.e. if it has a small $\text{RMSE}_t$, not only on average but preferably over the whole period of time.

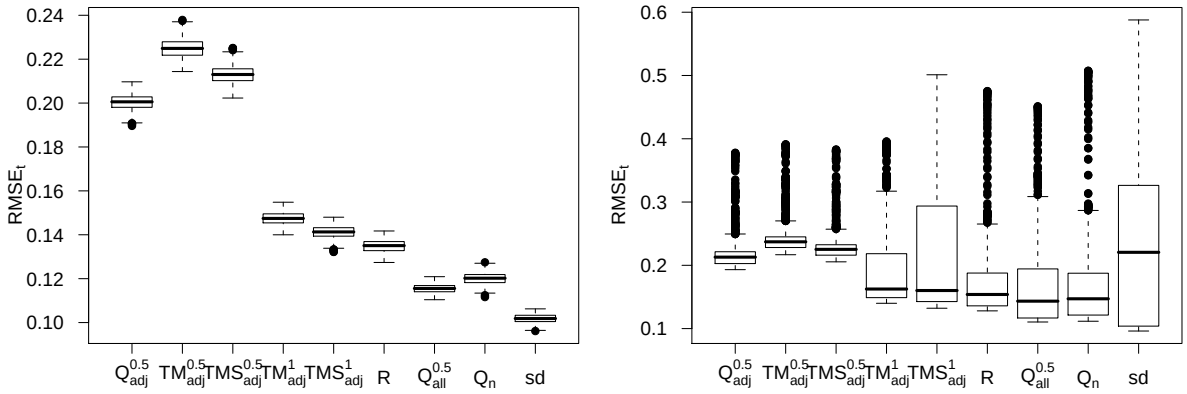


**Figure 4.12**: Boxplots of the $\text{RMSE}_t$ of online scale estimators using a window width of $n = 50$ for time series with independent non-contaminated (left) and contaminated $N(0,1)$ errors (right). The contaminated observations come from a $N(3, 1^2)$ distribution for $t \leq 500$ and from a $N(1, 3^2)$ distribution for $t > 500$.

| | $Q_{adj}^{0.5}$ | $TM_{adj}^{0.5}$ | $TMS_{adj}^{0.5}$ | $TM_{adj}^{1}$ | $TMS_{adj}^{1}$ | $R$ | $Q_{all}^{0.5}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | 0.325 | 0.360 | 0.341 | 0.240 | 0.229 | 0.230 | 0.194 | 0.210 | 0.167 |
| $n = 50$ | 0.200 | 0.225 | 0.213 | 0.147 | 0.141 | 0.135 | 0.115 | 0.120 | 0.102 |

**Table 4.3**: Average root mean squared error $\overline{RMSE}$ of the online scale estimators for time series consisting of Gaussian white noise.

### 4.4.1 Standard Normal Errors

The time series in this setting are generated according to model (3.1) with $\mu_t \equiv 0$ and $\varepsilon_t \overset{\text{iid}}{\sim} N(0,1)$, i.e. $y_t = \varepsilon_t$ for all $t = 1, \ldots, T$. All investigated methods yield unbiased online estimates of the scale. For the standard deviation a slight negative bias can be observed which is due to the fact that sd as defined in (4.28) *asymptotically* achieves unbiasedness at normal data while for all robust estimators finite sample correction factors are applied. In consequence, the standard deviation underestimates the true scale on average when using a small sample size. However, this effect diminishes for larger samples. The left panel of Figure 4.12 shows boxplots of the $\text{RMSE}_t$ of all estimators for $n = 50$; results for $n = 20$ look similar but are not presented here. Here, the standard deviation based on least squares residuals is, as expected, the most efficient method, showing the smallest $\text{RMSE}_t$. Furthermore, Table 4.3 contains the mean of the root mean squared error, averaged over the whole observed time period:

$$\overline{RMSE} = \frac{1}{T-n+1} \sum_{t=n}^{T} \text{RMSE}_t \,.$$

The non-robust $TM_{adj}^1$ and $TMS_{adj}^1$ are even less efficient than the robust $R$, $Q_{all}^{0.5}$ and $Q_n$. The three robust model-free estimators show the largest $\text{RMSE}_t$ with $Q_{adj}^{0.5}$ performing best among these three.

### 4.4.2 Standard Normal Errors with 5% Contamination

To investigate the performance of the online scale estimators in the presence of contamination, we consider standard normal time series where 5% gross-error outliers are induced at fixed time points. For the first half of the time period, the contamination distribution corresponds to a $N(d, 1)$-distribution leading to additive outliers of size $d$ in the level. In the second half, a $N(0, d^2)$-distribution is used to generate scale outliers. We consider values $d \in \{1, 2, 3, 4, 5\}$, generating small, moderately sized and large outliers. (For $d = 1$ only small additive outliers are generated in the first half of the time series.)
For both types of contamination five single outliers, two patches of two, two patches of three and one patch of five subsequent outliers is generated. In addition, for the additive outliers a stretch of ten successive level outliers is incorporated into the simulation scheme to investigate the behaviour of the scale estimators for a level shift.
In particular, the additive outliers are induced at $t \in W_1$ and the scale outliers appear at $t \in W_2$ with

$$\begin{aligned} W_1 &= \{51, 71, 81, 86, 91,\ \ 101, 102,\ \ 151, 152, \\ &\qquad 201, 202, 203,\ \ 241, 242, 243,\ \ 301, \ldots, 305,\ \ 401, \ldots, 410\} \\ \text{and}\quad W_2 &= \{551, 571, 581, 586, 591,\ \ 601, 602,\ \ 651, 652, \\ &\qquad 701, 702, 703,\ \ 741, 742, 743,\ \ 801, \ldots, 805\}. \end{aligned}$$

Figure 4.13 shows the mean bias averaged over time

$$\overline{\mathrm{MB}} = \frac{1}{T - n + 1} \sum_{t=n}^{T} \mathrm{MB}_t$$

as well as the root mean squared error averaged over time $\overline{\mathrm{RMSE}}$ for the considered methods applied in a moving window of width $n = 50$. Plots for $n = 20$ look similar, but are excluded from the presentation here. For all methods $\overline{\mathrm{MB}}$ and RMSE increase along with the magnitude of the outliers, determined by the value of $d$.
It shows clearly that the size of the outliers has a considerable influence on the bias of the non-robust scale estimators, while the mean bias does not increase as drastically with increasing $d$ for the robust methods. Particularly the $\overline{\mathrm{MB}}$ of the model-free estimators increases the least with $d$, and all three model-free estimators perform very similar – better than the other methods – for all values of $d$.
However, in terms of efficiency measured by the root mean squared error, the model-free estimators perform worse than the other robust estimators, and for small outliers, i.e. small $d$, they even have a higher $\overline{\mathrm{RMSE}}$ than the non-robust methods. All methods show some increase in $\overline{\mathrm{RMSE}}$ for increasing $d$, but this increase is much more pronounced for the non-robust methods.

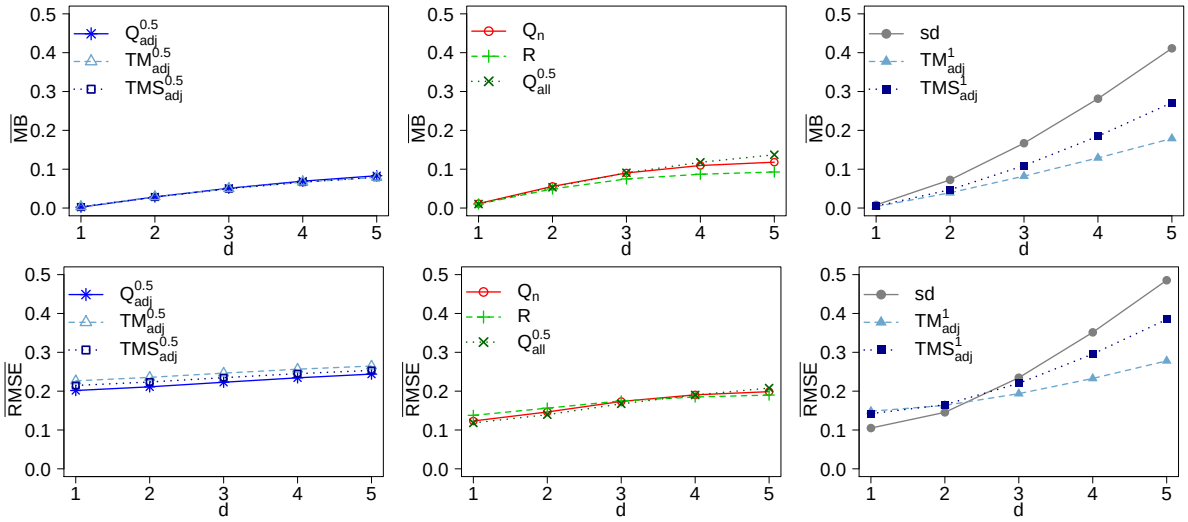


**Figure 4.13**: Average mean bias (top) and average RMSE (bottom) of online scale estimators applied to a moving window of width $n = 50$ for time series consisting of independent $N(0, 1)$ errors with 5% contamination. The contaminated observations come from a $N(d, 1^2)$ distribution for $t \leq 500$ and from a $N(1, d^2)$ distribution for $t > 500$ with the level of contamination $d \in \{1, 2, 3, 4, 5\}$.

For small to moderate sizes of outliers the non-robust $TM_{adj}^1$ and $TMS_{adj}^1$ scales even outperform its robust counterparts in terms of $\overline{\text{RMSE}}$, but for larger sizes of outliers the robust methods clearly perform better.

The right panel of Figure 4.12 shows boxplots of the $\text{RMSE}_t$ in the setting with $d = 3$ where the contamination comes either from a $N(3, 1^2)$ or a $N(0, 3^2)$ distribution. The average $\text{RMSE}_t$ for each of these boxplots corresponds to the corresponding $\overline{\text{RMSE}}$ value at $d = 3$ in the bottom panels of Figure 4.13. Although Figure 4.13 shows that the running standard deviation on average performs similar to the model-free non-robust methods for small outlier sizes, in Figure 4.12 it is evident that sd is heavily affected by the contamination. The $TMS_{adj}^1$ estimator is similarly affected, though not as bad, while $TM_{adj}^1$ still achieves reasonably efficient results for a non-robust estimator. These findings are confirmed for other simulation settings with $d \in \{1, 2, 4, 5\}$.

Figure 4.12 further shows that from the robust methods $R$, $Q_n$ and $Q_{all}^{0.5}$ perform best in terms of $\text{RMSE}_t$, with $R$ showing the best overall performance. From the robust model-free estimators $Q_{adj}^{0.5}$ shows the best performance again.

To evaluate the influence of the different types and number of consecutive outliers, Figure 4.14 shows the mean bias $\text{MB}_t$ over time for a normal series with contamination from a $N(3, 1^2)$ distribution for $t \leq 500$ and from a $N(0, 3^2)$ distribution for $t > 500$. For other values of $d$ the graphics look (up to a factor) very similar. Apparently, all robust methods show some sort of reaction towards outliers but not as drastic as the non-robust scale estimators. Furthermore, there is almost no difference between the three robust model-free scales.

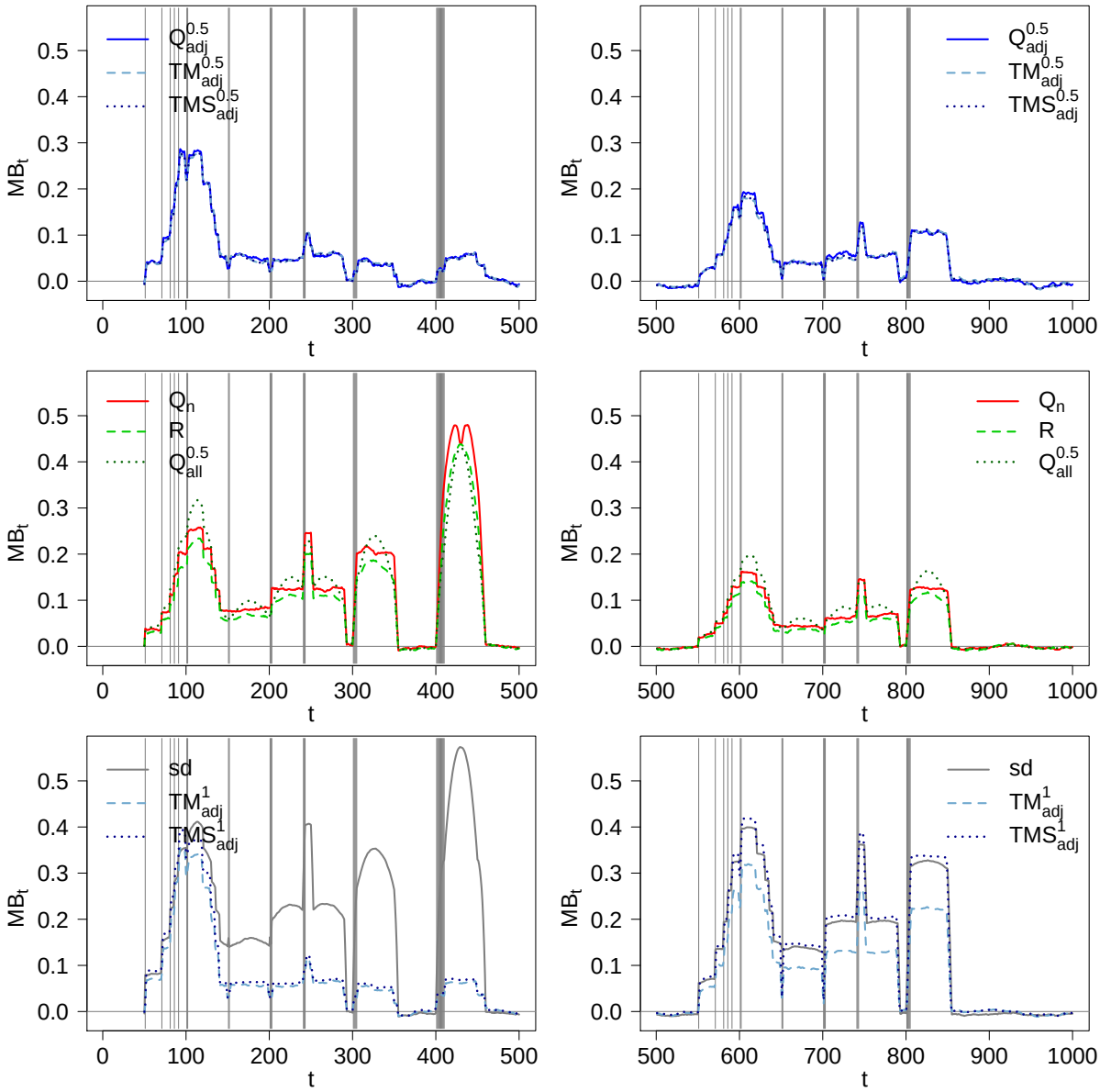


**Figure 4.14**: Mean bias $MB_t$ over time resulting from online scale estimation in a moving window of width $n = 50$ for time series with independent standard normal errors and 5% contamination. The contaminated observations come from a $N(3, 1^2)$ distribution for $t \leq 500$ (left) and from a $N(1, 3^2)$ distribution for $t > 500$ (right). The vertical grey lines mark the times where the outliers or outlier patches occur. The patch of ten outliers after $t = 400$ can also be interpreted as a temporary level shift.

Because all methods estimate the scale online, i.e. at the right endpoint of a time window, the increase in the bias appears after the time where the outlier(s) occur. All methods, regardless whether regression-based, regression-free or model-free, show an increasing $\text{MB}_t$ after the occurrence of outliers, a brief period of constantly increased mean bias and then a decrease; the time period until the decrease depends on the window width $n$.
Furthermore, it can be seen in Figure 4.14 that a set of single outliers results in a larger bias than consecutive ones for the model-free scale estimators and the number of subsequent outliers has no impact on the magnitude of the bias. In contrast, for the scale estimators $R$, $Q_{all}^{0.5}$ and $Q_n$, requiring the local linearity of the signal, it can be observed that $\text{MB}_t$ increases with the length of the outlier patch when additive outliers occur – most obvious for the level shift after $t = 400$.
Scale outliers seem to have a smaller influence on the $\text{MB}_t$ for $R$, $Q_{all}^{0.5}$ and $Q_n$, while for the robust model-free estimators their effect is similar to that of additive level outliers. For the non-robust model-free estimators, the influence of scale outliers is much worse than that of additive outliers, where the non-robust estimators perform even similar to their robust counterparts.
The biggest advantage of the model-free estimators, namely its independence of the local linearity assumption, appears most obviously for subsequent additive level outliers or level shifts, respectively, which cause a much smaller bias for this type of estimator compared to the other ones.

### 4.4.3 Scale Shift

In this setting, the scale of the data-generating normal distribution jumps from $\sigma_t = 1$ to $\sigma_t = 5$ at time $t = T/2+1 = 501$ while the level stays constant, $\mu_t \equiv 0$ for all $t = 1, \ldots, T$.
Because of the online estimation at the right end point of each time window, all methods trace the scale shift with some time delay. However, concerning the magnitude of $\text{MB}_t$ right after the shift and the duration until the bias returns back to zero, there is no real difference between the methods. The delay basically corresponds to the chosen window width.
In this setting it is $\text{MB}_t \approx -0.8$ right after the shift, regardless of the method or window width used (see the left panels of Figure 4.15).
Since all observations come from a normal distribution, the results for the root mean squared error basically correspond to the outcomes for the non-contaminated standard normal setting. Even the order of magnitude of $\overline{RMSE}$ is quite similar to this setting, cf. Tables 4.3 and 4.4, i.e. the standard deviation sd is the most efficient method, $Q_{all}^{0.5}$, $Q_n$ and $R$ perform not much worse, while the robust model-free estimators lack some efficiency.

| | $Q_{adj}^{0.5}$ | $TM_{adj}^{0.5}$ | $TMS_{adj}^{0.5}$ | $TM_{adj}^{1}$ | $TMS_{adj}^{1}$ | $R$ | $Q_{all}^{0.5}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | 0.331 | 0.365 | 0.347 | 0.244 | 0.233 | 0.237 | 0.200 | 0.217 | 0.171 |
| $n = 50$ | 0.219 | 0.244 | 0.232 | 0.162 | 0.153 | 0.155 | 0.135 | 0.139 | 0.115 |

**Table 4.4**: Average root mean squared error $\overline{RMSE}$ of the online scale estimators for time series consisting of normal errors with a scale shift from $\sigma_t = 1$ to $\sigma_t = 5$.

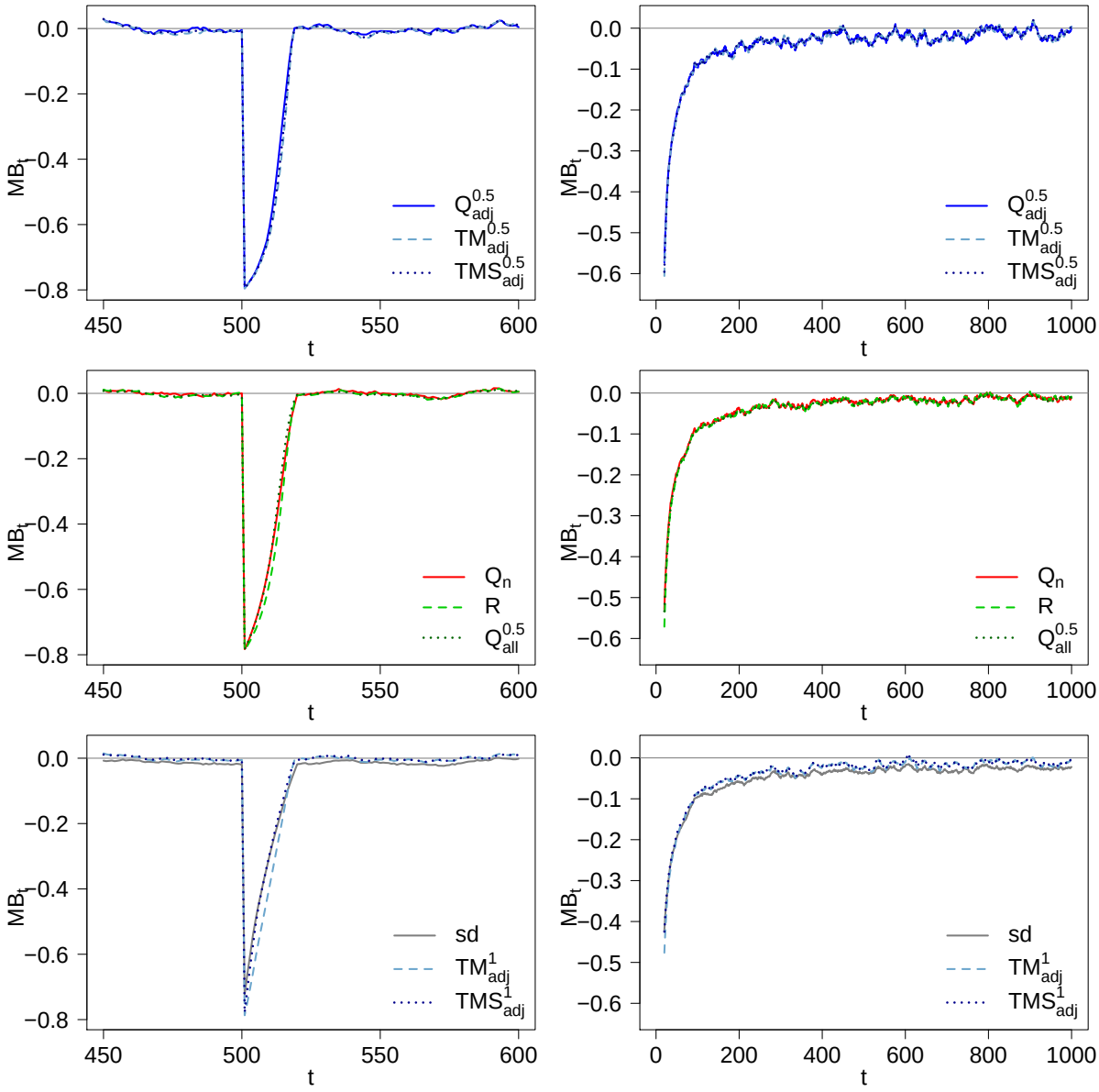


**Figure 4.15**: Mean bias $\mathrm{MB}_t$ for a time interval around the time $t_0 = 500$ where the scale of the independent error term shifts from $\sigma_t = 1$ to $\sigma_t = 5$ (left) and $\mathrm{MB}_t$ for time series with independent $N(0, t^2)$ errors (right) for the window width $n = 20$.

| | $Q^{0.5}_{adj}$ | $TM^{0.5}_{adj}$ | $TMS^{0.5}_{adj}$ | $TM^{1}_{adj}$ | $TMS^{1}_{adj}$ | $R$ | $Q^{0.5}_{all}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | 0.321 | 0.354 | 0.336 | 0.240 | 0.230 | 0.232 | 0.198 | 0.213 | 0.174 |
| $n = 50$ | 0.215 | 0.236 | 0.226 | 0.168 | 0.161 | 0.161 | 0.145 | 0.147 | 0.132 |

**Table 4.5**: Average root mean squared error $\overline{RMSE}$ of the online scale estimators for time series consisting of normal errors with a linearly changing scale $\sigma_t = t$.

| | $Q^{0.5}_{adj}$ | $TM^{0.5}_{adj}$ | $TMS^{0.5}_{adj}$ | $TM^{1}_{adj}$ | $TMS^{1}_{adj}$ | $R$ | $Q^{0.5}_{all}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | 0.422 | 0.434 | 0.419 | 0.361 | 0.360 | 19.7 | 18.7 | 28.9 | 30.2 |
| $n = 50$ | 0.385 | 0.400 | 0.390 | 0.333 | 0.318 | 114.3 | 109.1 | 165.1 | 189.0 |

**Table 4.6**: Average root mean squared error $\overline{RMSE}$ of the online scale estimators for normal time series with a quadratically increasing signal $\mu_t = t^2$.

### 4.4.4 Slow Scale Change

Here, the time series are generated from a model with constant level $\mu_t \equiv 0$ for all $t = 1, \ldots, T$ and independent errors from a $N(0, t^2)$ distribution, i.e. the assumption of a locally constant scale (4.1) is violated because the scale $\sigma_t = t$ changes linearly over time. Because all methods estimate the scale at the end of a time window, all previous observations $y_{t-n+1}, \ldots, y_{t-1}$ come from distributions with a smaller variance than $\sigma_t^2$. Therefore, all methods underestimate the true scale resulting in a negative $\text{MB}_t$ over the whole period of time, and on average the mean bias is quite similar for all methods.

Since $\text{MB}_t$ evaluates the bias relative to the true scale the deviation of the estimate from the true value is proportionally the largest for small values of $\sigma_t$. As the true scale $\sigma_t$ increases, the mean bias $\text{MB}_t$ tends to zero (see the right panels of Figure 4.15). For the larger window width of $n = 50$, $\text{MB}_t$ is further away from zero than for $n = 20$.

In terms of $\text{RMSE}_t$, the results are, even in order of magnitude, similar to the standard normal and the scale shift setting (cf. Tables 4.3 and 4.5); that means that all investigated methods perform well even if the scale is locally not constant, for the price of a small bias.

### 4.4.5 Quadratic Trend Change

To investigate the performance of the online scale estimators when the assumption of a local linear signal within each time window is violated, we generate time series from a model with independent $N(0, 1)$ errors $\varepsilon_t$ but a quadratically changing trend $\mu_t = t^2$.
In this setting $\text{MB}_t$ and $\text{RMSE}_t$ stay almost constant over the whole period of time and $\text{RMSE}_t$ strongly depends on the underlying bias which is quite large for all methods that rely on the assumption of a local linear signal (3.2), i.e. the regression-based and regression-free estimators.

| | $Q_{adj}^{0.5}$ | $TM_{adj}^{0.5}$ | $TMS_{adj}^{0.5}$ | $TM_{adj}^{1}$ | $TMS_{adj}^{1}$ | $R$ | $Q_{all}^{0.5}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | 0.360 | 0.377 | 0.368 | 0.322 | 0.316 | 0.238 | 0.206 | 0.225 | 0.192 |
| $n = 50$ | 0.312 | 0.321 | 0.317 | 0.296 | 0.293 | 0.146 | 0.128 | 0.134 | 0.119 |

**Table 4.7**: Average root mean squared error $\overline{RMSE}$ of the online scale estimators for time series with AR(1) errors.

| | $Q_{adj}^{0.5}$ | $TM_{adj}^{0.5}$ | $TMS_{adj}^{0.5}$ | $TM_{adj}^{1}$ | $TMS_{adj}^{1}$ | $R$ | $Q_{all}^{0.5}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | −0.284 | −0.287 | −0.286 | −0.279 | −0.276 | −0.070 | −0.095 | −0.061 | −0.080 |
| $n = 50$ | −0.281 | −0.282 | −0.282 | −0.279 | −0.278 | −0.029 | −0.043 | −0.026 | −0.033 |

**Table 4.8**: Average mean bias $\overline{MB}$ of the online scale estimators for time series with AR(1) errors.

Table 4.6 shows the $\overline{RMSE}$, the average of the $\text{RMSE}_t$ values over time, for all methods and both investigated window widths. The huge difference in $\overline{RMSE}$ between the model-free scale estimators and regression-based as well as regression-free estimators emphasise the fact that the model-free scale estimators achieve much better estimations if there are non-linearities in the underlying signal. The regression-free estimators $R$ and $Q_{all}^{0.5}$ perform better than the regression-based $Q_n$ and sd. $TMS_{adj}^{1}$ performs best here w.r.t. $\overline{RMSE}$, but the robust model-free estimators do not perform much worse.

### 4.4.6 AR(1) Errors

Departing from the assumption of independent errors, we consider model (3.1) with a constant level $\mu_t \equiv 0$ and autocorrelated errors. In particular, we generate the errors according to an AR(1) model, i.e.

$$\varepsilon_t = \varphi\varepsilon_{t-1} + \mathrm{e}_t\,,\ t \in \mathbb{Z}$$

with innovations $\mathrm{e}_t \sim N(0, \sigma_{\mathrm{e}}^2)$ for all $t = 1, \ldots, T$. The unconditional variance of $\varepsilon_t$ is then given by

$$\sigma_{\varepsilon_t}^2 \;=\; \mathrm{Var}(\varepsilon_t) \;=\; \frac{1}{1-\varphi^2}\,\sigma_{\mathrm{e}}^2\,.$$

For the simulations we use standard normal innovations and choose the parameter for the AR(1) model $\varphi = 0.4$ for moderate correlation between successive observations. This results in a marginal standard deviation of $\sigma_{\varepsilon_t} = 1.091$ for all $t$.

In case of AR(1) errors all methods loose some efficiency compared to the standard normal setting, in particular when using the larger window width (cf. Table 4.7). However, the results for the RMSE are still similar to the situation of independent $N(0, 1)$ errors, only $TM_{adj}^{1}$ and $TMS_{adj}^{1}$ are much less efficient here. This can also be seen when comparing the boxplots for the $\text{RMSE}_t$ values in case of AR(1) errors in the left panel of Figure 4.16 with those in the standard normal situation, shown in the left panel of Figure 4.12.

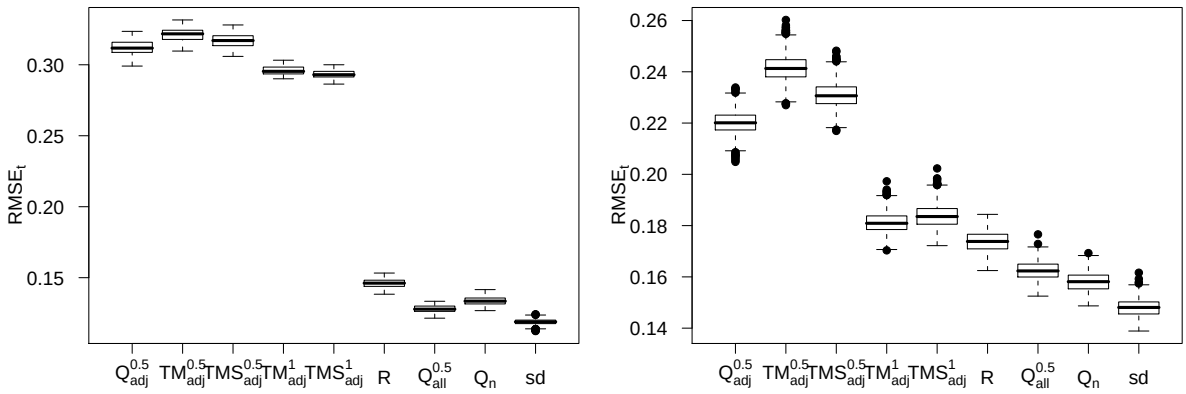


**Figure 4.16**: Boxplots of $\text{RMSE}_t$ for time series with AR(1) errors (left) and GARCH(1,1) errors (right) using the window width $n = 50$.

The increase in $\text{RMSE}_t$ is partly due to the fact that because of the positive autocorrelation all scale estimators show a negative bias (see Table 4.8). The regression-based and regression-free methods are only a little biased, and $Q_n$ is the least biased method. The model-free estimators have a larger bias which approximately has the same magnitude for all estimators of this type, independent of the window width $n = 20$ or $n = 50$ or of the fact whether they are robust or not. Thus, we do not recommend to use the model-free estimators on time series where autocorrelations can be expected.

### 4.4.7 GARCH(1,1) Errors

As a further deviation from our model assumptions we investigate time series with autocorrelated errors and a slowly varying scale, i.e. in model (3.1) we consider $\mu_t \equiv 0$ and $\varepsilon_t$ following a GARCH(1,1) model with parameters $\alpha_0, \alpha_1, \beta_1 \in \mathbb{R}$

$$\begin{aligned} \varepsilon_t &= \sigma_t \mathrm{e}_t\,,\ t \in \mathbb{Z} \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2\,, \end{aligned}$$

where $\mathrm{e}_t \sim N(0, \sigma_\mathrm{e}^2)$ for all $t = 1, \ldots, T$. For the simulations we use $\sigma_\mathrm{e}^2 = 1$, $\alpha_0 = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.7$ guaranteeing stationarity and moderate autocorrelation. The mean bias $\text{MB}_t$ and root mean squared error $\text{RMSE}_t$ are calculated w.r.t. the conditional $\sigma_t$ in the uncontaminated model.

The right panel of Figure 4.16 shows boxplots of the $\text{RMSE}_t$ values for each of the considered methods when applying the methods to a GARCH(1,1) process using the window width $n = 50$. Similar results are achieved for $n = 20$, only that $Q_n$ loses some efficiency while $TMS^1_{adj}$ gains some when using this smaller window width.

Figure 4.16 shows that the running standard deviation is the most efficient method for AR(1) and GARCH(1,1) errors, but the robust methods $Q_n$, $Q^{0.5}_{all}$ and $R$ are also quite good and more efficient than $TM^1_{adj}$ and $TMS^1_{adj}$. The three robust model-free estimators are least efficient for GARCH(1,1) errors whereas $Q^{0.5}_{adj}$ is the best one from these three.

| | $Q_{adj}^{0.5}$ | $TM_{adj}^{0.5}$ | $TMS_{adj}^{0.5}$ | $TM_{adj}^{1}$ | $TMS_{adj}^{1}$ | $R$ | $Q_{all}^{0.5}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | 0.326 | 0.358 | 0.341 | 0.249 | 0.244 | 0.243 | 0.210 | 0.222 | 0.183 |
| $n = 50$ | 0.220 | 0.241 | 0.231 | 0.181 | 0.184 | 0.174 | 0.162 | 0.158 | 0.148 |

**Table 4.9**: Average root mean squared error $\overline{RMSE}$ of the online scale estimators for time series with GARCH(1,1) errors.

| | $Q_{adj}^{0.5}$ | $TM_{adj}^{0.5}$ | $TMS_{adj}^{0.5}$ | $TM_{adj}^{1}$ | $TMS_{adj}^{1}$ | $R$ | $Q_{all}^{0.5}$ | $Q_n$ | sd |
|---|---|---|---|---|---|---|---|---|---|
| $n = 20$ | −0.023 | −0.025 | −0.024 | −0.014 | −0.008 | −0.006 | −0.012 | −0.004 | −0.021 |
| $n = 50$ | −0.023 | −0.025 | −0.025 | −0.005 | 0.006 | 0.001 | −0.002 | 0.003 | 0.001 |

**Table 4.10**: Average mean bias $\overline{MB}$ of the online scale estimators for time series with GARCH(1,1) errors.

Table 4.9 contains the corresponding $\overline{\text{RMSE}}$ values, showing that compared to the $N(0, 1)$ setting (cf. Table 4.3) there is a loss in efficiency for all methods although the results are quite similar in both situations.

Just like in the AR(1) setting, a negative bias can be observed for all considered methods where the absolute value of the $\overline{\text{MB}}$ is the largest for the robust model-free scale estimators. However, for GARCH(1,1) errors the magnitude of the bias is much smaller than in the AR(1) setting, and it is always close to zero, see Table 4.10. While a larger $n$ improves the bias for sd, $Q_n$, $Q_{all}^{0.5}$, and $R$, i.e. all estimates which require that the signal can be locally approximated by a line, it has no positive effect on the bias of the model-free estimators. All investigated methods provide sensible information on the slowly varying conditional scale although they estimate the local unconditional scale. However, the model-free estimators do not cope as well with autocorrelations as the other methods.

## 4.5 Applications

### 4.5.1 Simulated Time Series

An application to a simulated time series exhibiting various features can accentuate the advantages of the different scale estimation approaches. Figure 4.17 shows such a simulated series with its underlying true signal level $\mu_t$ where the errors $\varepsilon_t$ are independently generated from a normal distribution with mean zero and a standard deviation corresponding to the black line in the right hand side and bottom panels in Figure 4.17. This time series contains periods of a constant signal level, a level shift and periods with a linear or quadratic trend in the signal. Furthermore, the time-dependent scale of the error term also contains a shift, constant periods and periods with a linear and quadratic trend. The impact on the simulated observations can be seen in the top left panel of Figure 4.17.

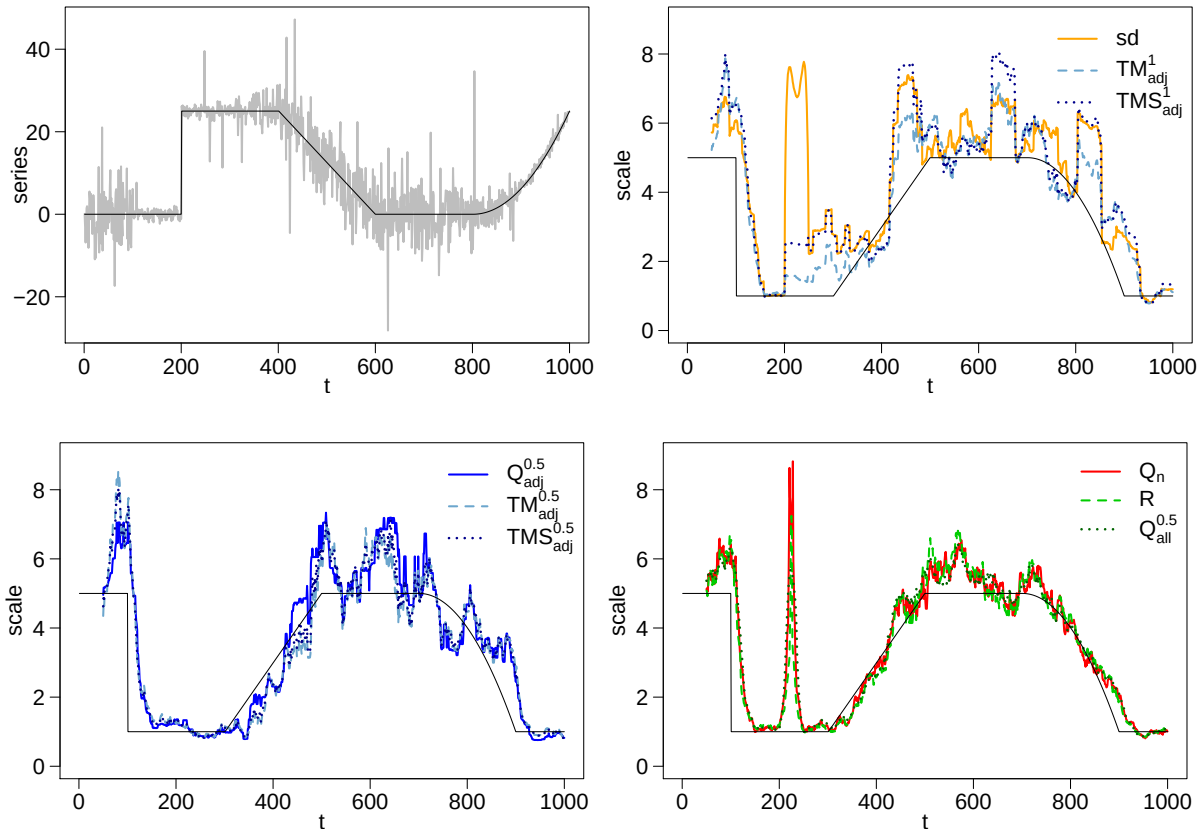


**Figure 4.17**: The top left panel shows a normal time series with varying signal level (black line), changing scale and 5% contamination. The remaining panels show the corresponding online scale estimates based on the window width $n = 50$ together with the underlying true scale (black line). The top right panel shows the non-robust scale estimates, the bottom panels display the robust scales.

Furthermore, at 50 time points (i.e. 5% of the time) outliers are generated by replacing the data with observations coming from a $N(0, 15^2)$ distribution.

It clearly shows that the non-robust scale estimators are not suitable if outliers are present because they drastically overestimate the true scale in all windows containing at least one outlier (see the top right panel of Figure 4.17).

All robust methods achieve online scale estimates which trace the course of the true scale quite nicely, only with a certain time delay because of the online estimation at the right endpoint of each time window. The three robust model-free estimators $Q^{0.5}_{adj}$, $TM^{0.5}_{adj}$ and $TMS^{0.5}_{adj}$, shown in the bottom left panel of Figure 4.17 perform very similar. Compared to the regression-based $Q_n$ and regression-free $Q^{0.5}_{all}$ and $R$ estimators (bottom right panel), the lower efficiency of the model-free estimators is reflected by a much larger variability in their estimations. However, they do not show the peak at $t = 200$ where a level shift in the signal of the series results in a drastic overestimation of the scale for all methods requiring

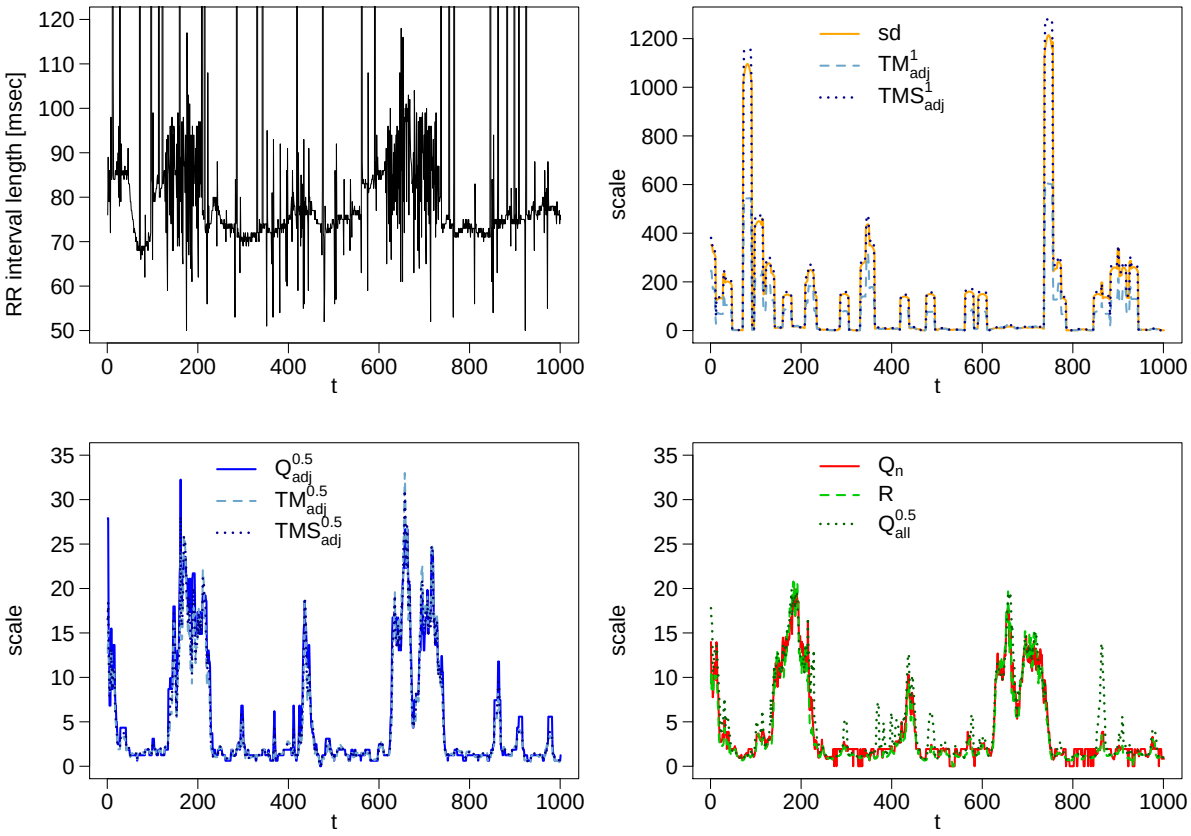


**Figure 4.18**: Sequence of 1000 RR interval lengths in msec (top left panel) with corresponding online scale estimates based on the window width $n = 20$. The top right panel shows the non-robust scale estimates, the bottom panels display the robust scales.

a local linear signal, i.e. $Q_n$, $R$ and $Q^{0.5}_{all}$. The difference in the online scale estimates from these three methods shown in the bottom right panel of Figure 4.17 is very small, but it appears that the $R$ estimator achieves the smoothest scale estimation.

### 4.5.2 Real Time Series

In intensive care, the heart rate in beats per minute is derived from the continuous electrocardiogram (ECG) signal. However, to prevent an intermediate calculation step, the heart rate variability is generally evaluated not from the heart rate measurements which are updated once a second but from the ECG itself: Therefore, the length of the time intervals between consecutive heart beats, conventionally named RR intervals, is measured in milliseconds and the heart rate variability is calculated from a sequence of such lengths. Technically, such a sequence is not a time series, because the sampling period is not equidistant. It would be easy to convert this sequence to a time series measured e.g. once per second, where an observation corresponds to the length of the last RR interval.

However, the difference to the sequence of true RR interval lengths would be minimal while there would be some loss in information and also in computation time.

The top left panel of Figure 4.18 shows a sequence of 1000 RR interval lengths in milliseconds from a patient taking part in the Cardiac Arrhythmia Suppression Trial (CAST). The data set is obtained from the Interbeat (RR) Interval Database on PhysioNet (Goldberger et al., 2000). It can be seen that the interval lengths do not permanently vary around a constant level but also show trends and trend changes (see e.g. the observations from 40 to 80 and the change thereafter).

For presentational reasons positive outliers with a size larger than 120msec are cut off. In this stretch of 1000 observations there are 25 such measurements with their values ranging between 653 and 5363, meaning a total percentage of 2.5% of very large outliers. However, some smaller values might also be considered as outliers because of the generally smaller variability of the observations at the time of observation, see e.g. the observations $x_{811}$ and $x_{812}$.

Some very large observations are given by $x_{72} = 4838$ and $x_{97} = 2032$, as well as by $x_{737} = 5363$ and $x_{755} = 1185$. For the windows including these observations, it is quite obvious that the non-robust scale estimators, shown in the top right panel of Figure 4.18, 'break down' because they reach huge values of more than 1000. It can also be seen that, according to the simulation results shown in the previous section, from the non-robust scale estimators the mean of adjacent triangle heights $TM^1_{adj}$ does not show such a strong reaction to outliers as $TMS^1_{adj}$ and the standard deviation. This appears even more obvious when using a larger window width like $n = 50$ (not shown here).

In contrast, the robust online scale estimators displayed in the bottom panels of Figure 4.18 perform well in the presence of outliers and – from looking at the data – also yield sensible estimations of the true scale. All robust methods indicate the increased variability from about 130 to 210 and from about 610 to 730 and also after 400.

The outcomes of the three robust scale estimators based on adjacent triangle heights are quite similar again, only that in times of increased variability $Q^{0.5}_{adj}$ tends to larger estimations than $TM^{0.5}_{adj}$ and $TMS^{0.5}_{adj}$. Furthermore, they all show some sort of reaction if several outliers occur within the window used for the estimation, especially if the window width is as small as, say, $n = 20$. This is due to their breakdown point of about 1/6.

From the regression-free estimators relying on the local linearity assumption (3.2) the $Q^{0.5}_{all}$ scale estimator performs similar to the adjacent-type estimators while $R$ is more robust towards the outliers and also more efficient because the $R$ estimations do not show such a large variability. The robust $Q_n$ scale estimator based on the $RM$ residuals performs similarly good as $R$ but shows larger variability.

## 4.6 Conclusions

The robust methods compared in this chapter provide useful tools for the online extraction of time varying variability. Because of the moving window approach all estimators have a computation time which allows for real-time application and they have given proof of their usefulness at a real data example. The methods are able to deal with contaminated data, trends and trend changes and work well even if the true scale cannot be assumed as locally constant within one time window, or if the errors are autocorrelated.

From the model-free adjacent-type scale estimators the two non-robust estimators $TM^1_{adj}$ and $TMS^1_{adj}$ are reasonably efficient at standard normal data and perform much better than the standard deviation in the presence of contamination. However, outliers cause considerable bias.

The three robust model-free estimators perform very similar in all settings but $Q^{0.5}_{adj}$ is slightly more efficient than $TM^{0.5}_{adj}$ and $TMS^{0.5}_{adj}$. Gelper, Schettlinger, Croux and Gather (2009) further investigate the behaviour of the model-free scale estimators with maximum breakdown point in a simulation study, but for their considered contamination models these estimators provide no advantage compared to the more efficient model-free estimators based on $\alpha = 0.5$, i.e. considering a half of the triangle heights.

The advantages of the model-free scale estimators consist in their very low update computation time of only $\mathcal{O}(\log n)$ and the fact that they work well in the presence of nonlinearities in the signal level such as quadratic trends or level shifts. In case of a linear change in the scale they also perform reasonably well, but for autocorrelated errors they show some bias and loss of efficiency.

Compared to the model-free estimators the investigated regression-free methods, $R$ and $Q^{0.5}_{all}$, have higher breakdown points and are much more efficient. Furthermore, their performance in the presence of autocorrelated errors is similar to the case of independent errors. If no contamination is present $Q^{0.5}_{all}$ is slightly more efficient than $R$, but in comparison $R$ shows a better overall performance because of its higher robustness against outliers.

The $Q_n$ estimator based on the residuals from a local Repeated Median regression is also highly robust and efficient (even at non-contaminated data) and it has a much lower computation time than the regression-free estimators. It also provides good estimations for autocorrelated errors. However, this scale estimator strongly relies on the underlying signal estimation which can cause a large bias for the scale if estimated wrongly.

If only a low percentage of additive or level outliers can be expected and the signal level of the time series is likely to contain many trend changes and level shifts and / or if the computing time is very limited, the model-free scale estimator $Q^{0.5}_{adj}$ should be chosen for

online scale estimation with a medium-sized window width, yielding a good compromise between robustness and efficiency. However, if temporal correlations are expected or if outliers are very likely to occur, a more robust and more efficient method like $R$ or $Q_n$ should be applied with a window width as large as possible but still ensuring the local linearity of the signal level within this window.

# 5 Discussion

In this thesis, several approaches for online signal and variability extraction from time series are introduced and compared with each other. Especially such methods are focussed on, which are able to deal appropriately with a certain amount of collinear observations and to resist or ignore a certain amount of outliers. For an appropriate identification of outliers and the exact times of their appearance, additional outlier detection rules should be applied to be able to specify and explain the underlying contamination component.

In particular, in Section 3.2 it is shown that a univariate online Repeated Median filter with adaptive choice of the window width works well for the removal of random noise and artefacts from time series which do not follow any known model. The basis for this filter is provided in Chapter 2, addressing especially those properties of Repeated Median (RM) regression which are important for the univariate adaptive RM filter.

The advantages of the univariate adaptive RM filter, particularly the high robustness and fast computation, also hold for the multivariate adaptive online RM filter, described in Section 3.3 which additionally takes into account possible correlations between the variables. However, since this filter often chooses a small window width when applied to intensive care time series, a high variability can be observed in the extracted signal. Thus, further research should be carried out to improve the efficiency of the multivariate filter.

From the three approaches for online variability extraction, introduced in Chapter 4, the regression-based and regression-free approaches offer robust and efficient scale extractions while they do not work well if there are non-linearities in the signal level of the time series. The first approach further relies on a previous regression fit, i.e. the scale estimate can be highly biased if the regression fit is not suitable. However, the introduced regression-free methods require a lot of computation time. The development of update algorithms for these estimators could improve their applicability for high frequency time series.

The introduced model-free scale estimators have a very low update time and hence, they are suitable for the application to ultra-high frequency time series. However, they are not as robust and efficient as the other approaches. Concluding, for online variability extraction from intensive care time series, a combination of a robust model-free scale estimator with a robust regression-based scale estimator could provide a solution which incorporates the processing of all relevant information.

For example, while e.g. a regression-based estimate could be used for monitoring the variability of a time series, the difference to a model-free estimate could be monitored alongside as a measure of correctness. For application to time series where the variability alone is of interest and the underlying signal is not extracted, a regression-free estimator could be used instead of a regression-based one.

An important issue which has not been raised yet, is the choice of the window width for the scale estimation. Of course, this is of crucial importance: while the model-free scale estimators seem to work better with small window widths, larger window widths improve the efficiency and robustness for the regression-free and regression-based scales. However, since we do not know the time series structure beforehand, it is not possible to determine an optimal window width for a real-time application and thus, the window width has to be determined from an application-oriented background or adaptively: either a technique for the data-driven choice of the window width has to be developed for the considered scale estimators, or they could be applied to windows where the window width is chosen by the adaptive RM signal filter.

Combining information about the signal level with information about the local variability offers a possibility for improved alarm systems, for example, by developing new alarm rules for early warnings. Of course, such rules can only be proposed based on expert knowledge from clinicians. The online extracted variability could also be displayed along with the online extracted signal on bedside patient monitors, e.g. illustrated by $\pm 3\sigma$ bounds around the signal. Even if the variability is not monitored, it could still provide essential information for diagnosis, treatment of the patient or simply for re-setting monitoring parameters like the alarm thresholds.

Another issue which has not been addressed appropriately so far, is the occurrence of collinear observations: for the physiological time series we have at hand, the secondly measurements from the monitoring system, e.g. the heart rate values displayed on the bedside monitor, are evaluated on a discrete scale. This discrete measurement scale and the fact that a patient often is in a steady state for a longer period of time, e.g. caused by sedation, results in collinear or even identical measurements for certain stretches of time. For the introduced scale extraction techniques, this increases the probability of implosion – which quite frequently occurs when applying the proposed methods to intensive care data. If an alarm rule were based on the extracted variability as well as the signal, then estimating a scale of zero could lead to an unwanted increase in the number of (false) alarms. Therefore, further research should be carried out to overcome this problem.

Finally, it should be stated that the spectrum of application for the introduced methods is not restricted to the clinical context. The usefulness of the online scale extraction methods for monitoring the volatility of financial time series of stock returns has been shown by Nunkesser, Fried, Schettlinger and Gather (2008) for the regression-based scale estimators, and for the model-free estimators by Gelper, Schettlinger, Croux and Gather (2009). However, a broad range of further possible fields of application is imaginable for the online signal and variability extraction techniques investigated in this thesis.

# A Supplementary Definitions

## Breakdown Point

The breakdown point is often used for representation of the global robustness of a functional or an estimator. A definition of the finite sample breakdown point can be found in Donoho and Huber (1983); Davies and Gather (2005, 2006, 2007) provide further discussions on the notion of breakdown point.

Let $\mathbf{y}_n = \{y_1, \ldots, y_n\}$ be a sample of size $n$ with empirical distribution function $F_n$, and let $T$ denote an estimator taking on values in the parameter space $\Theta$ which we consider equipped with a metric $D$ satisfying

$$\sup_{\theta_1, \theta_2 \in \Theta} D(\theta_1, \theta_2) = \infty \,.$$

Further, let $\mathbf{y}_n^k$ denote a sample obtained from the original sample $\mathbf{y}_n$ with $k \in \{1, \ldots, n\}$ observations replaced by 'contaminated' observations with corresponding (contaminated) empirical distribution function $F_n^k$.

The *finite sample breakdown point* (fsbp) of the estimator $T$ at the sample $\mathbf{y}_n$, or at the empirical distribution function $F_n$, respectively, is defined by

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

For a regression estimator as introduced in Chapter 2, the parameter space is given by $\Theta = \mathbb{R}^2$ and as a suitable metric $D$ the simple $L_2$ norm can be chosen. For such estimators contamination is generated by replacing values from the original sample $\mathbf{y}_n$ by points which are arbitrarily far away from that sample.

For scale estimators $S$, like those investigated in Chapter 4, the parameter space corresponds to $\Theta = (0, \infty)$. 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*). For evaluating the breakdown point of such scale estimators, the metric

$$D(\theta_1, \theta_2) = |\log(\theta_1/\theta_2)|\,, \quad \theta_1, \theta_2 \in (0, \infty),$$

seems a suitable choice as it yields $\infty$ in both cases, explosion and implosion. Here, contamination is generated either by replacing $k$ observations from the original sample $\mathbf{y}_n$ by observations which are arbitrarily far away from the original sample (causing the scale estimate to explode) or by collinear observations (causing implosion).

The finite sample breakdown point (A.1) of a scale estimator $S$ can also be expressed as

$$\text{fsbp}(S, F_n) = \min\left\{\text{fsbp}^+(S, F_n), \text{fsbp}^-(S, F_n)\right\}, \tag{A.2}$$

where

$$\text{fsbp}^+(S, F_n) = \min \frac{1}{n}\left\{k \in \{1, 2, \ldots, n\} : \sup_{F_n^k} S(F_n^k) = \infty\right\} \tag{A.3}$$

is called *explosion breakdown point*, and

$$\text{fsbp}^-(S, F_n) = \min \frac{1}{n}\left\{k \in \{1, 2, \ldots, n\} : \inf_{F_n^k} S(F_n^k) = 0\right\} \tag{A.4}$$

the *implosion breakdown point*.

Under suitable equivariance assumptions Davies and Gather (2005) derive an upper bound for the finite sample breakdown point (A.1) which is given by

$$\text{fsbp}(T, F_n) \quad \leq \quad \left\lfloor \frac{n - n\Delta(F_n) + 1}{2} \right\rfloor / n\,. \tag{A.5}$$

For a $p$-dimensional sample in general position, where no more than $p-1$ observations lie on a $(p-1)$-dimensional hyperplane, it is $n\Delta(F_n) = p - 1$ and thus the upper bound for a regression equivariant regression estimator $T : \mathbb{R}^{p\times n} \to \mathbb{R}^p$ is given by

$$\text{fsbp}(T, F_n) \quad \leq \quad (\lfloor \frac{n-p}{2} \rfloor + 1)/n \tag{A.6}$$

For a one-dimensional sample in general position it is $n\Delta(F_n) = 1$ and thus, the upper bound for the finite sample breakdown point of an affine equivariant scale estimator $S : \mathbb{R} \to (0, \infty)$ is

$$\text{fsbp}(S, F_n) \quad \leq \quad \left\lfloor \frac{n}{2} \right\rfloor / n\,, \tag{A.7}$$

cf. Davies and Gather (2007).

Apart from the definitions given above, there exist further definitions of the breakdown point, but throughout this thesis we only refer to the ones cited. Furthermore, different versions of breakdown points are often related: for example, Zuo (2001) shows that generating contamination via the replacement of observations (like described above) and generating contamination by adding observations to the sample which are generated from a contaminated distribution (leading to the so called *finite sample addition breakdown point*) asymptotically lead to the same result. Here, the limit of any finite sample breakdown point $\text{fsbp}(T, F_n)$ is referred to as the *asymptotic breakdown point* which can be characterised by

$$\text{abp}(T) \;=\; \lim_{n\to\infty} \text{fsbp}(T, F_n)\,.$$

However, this value should be handled with caution, because it is not always obvious to which limit the empirical distribution function $F_n$ converges.

### EQUIVARIANCE PROPERTIES

Let $\mathbf{X} := ((\mathbf{x}_1, y_1)', \ldots, (\mathbf{x}_n, y_n)')$ denote a sample of size $n$ where $\mathbf{x}_i \in \mathbb{R}^d$ is a $d$-dimensional row vector, describing the design, and $y_i \in \mathbb{R}$ is a one-dimensional response, $i = 1, \ldots, n$.

Equivariance towards certain transformations of a sample describes the ability of an estimator $T$ to change accordingly to the alteration of the data. The following definitions can be found in Rousseeuw and Leroy (1987).

An estimator $T : \mathbb{R}^{(d+1)\times n} \to \mathbb{R}^d$ is called *regression equivariant* if for the data set $\mathbf{X}$ and any column vector $\mathbf{v} \in \mathbb{R}$ it is

$$T\Big(\big\{(\mathbf{x}_i,\, y_i + \mathbf{x}_i\mathbf{v})'\big\}\Big) = T\Big(\big\{(\mathbf{x}_i,\, y_i)'\big\}\Big) + \mathbf{v} \quad \forall\, i = 1, \ldots, n\,. \tag{A.8}$$

$T$ is called *scale equivariant* if for any constant $c \in \mathbb{R}$ it is

$$T\Big(\big\{(\mathbf{x}_i,\, c\,y_i)'\big\}\Big) = c \cdot T\Big(\big\{(\mathbf{x}_i,\, y_i)'\big\}\Big) \quad \forall\, i = 1, \ldots, n\,. \tag{A.9}$$

If a linear transformation of $\mathbf{x}_i$ also transforms the estimate achieved by $T$ accordingly, the estimator $T$ is called *affine equivariant*. That means, for any nonsingular square matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$ it is

$$T\Big(\big\{(\mathbf{x}_i\,\mathbf{A},\, y_i)'\big\}\Big) = \mathbf{A}^{-1}\, T\Big(\big\{(\mathbf{x}_i,\, y_i)'\big\}\Big) \quad \forall\, i = 1, \ldots, n\,. \tag{A.10}$$

### INFLUENCE FUNCTION

The influence function, introduced by Hampel (1974), describes the effect of point mass contamination on the considered estimator. Illustratively, a smooth influence function means that a single outlier at one particular point has a similar effect on the estimator if it is located at a nearby point; bounded influence means that the effect on the estimator cannot get arbitrarily large.

Formally, let $T$ denote a functional which maps a distribution $F \in \mathcal{F}$ from the family of probability measures $\mathcal{F}$ onto the parameter space $\Theta$. Furthermore, define $\delta_{x_0}$ the point-mass at $x_0 \in \mathbb{R}$, and $\downarrow$ describes the 'limit from the right'. The *influence function* of the functional $T$ is then defined as

$$\mathrm{IF}(x_0, T, F) \;=\; \lim_{\varepsilon \downarrow 0} \frac{T\big((1-\varepsilon)F + \varepsilon\delta_{x_0}\big) - T\big(F\big)}{\varepsilon}\,. \tag{A.11}$$

### EFFICIENCY

The efficiency can be seen as a measure of precision of a particular estimator $T$ relative to another estimator $T_0$, usually the maximum likelihood estimator or an estimator attaining the Cramér-Rao bound.

Consider a random variable $X \sim F$ and a functional $T : \mathcal{F} \to \Theta$ mapping a distribution $F$ from the family of probability measures $\mathcal{F}$ onto the parameter space $\Theta$ with $T \sim G^F$, and $T_0 : \mathcal{F} \to \Theta$ with $T_0 \sim G_0^F$. The *asymptotic efficiency* of $T$ relative to $T_0$ at the distribution $F$ is then defined as

$$\begin{aligned} \text{eff}(T, T_0, F) &= \frac{\text{ASV}_{G_0^F}(T_0)}{\text{ASV}_{G^F}(T)} \\ &= \frac{E_{G_0^F}\left[(T_0 - E_{G_0^F}(T_0))^2\right]}{E_{G^F}\left[(T - E_{G^F}(T))^2\right]}, \end{aligned} \tag{A.12}$$

where ASV denotes the asymptotic variance, see e.g. Maronna, Martin and Yohai (2006), Section 2.2.2.

For unbiased estimators the *finite sample efficiency* is usually defined as the ratio of their population variances. However, these are often hard to derive and thus, the finite sample efficiency has to be estimated.

Consider the estimators $T$ and $T_0$ evaluated at $n$ random samples $X_1, \ldots, X_k \sim F$. Let $T_k(i)$ and $T_{0,k}(i)$ denote the estimates derived from applying $T$ or $T_0$ to the $i$th sample, $i = 1, \ldots, n$, and let $G_n^F$, $G_{0,n}^F$ denote their empirical distribution functions. Replacing the population variance by its finite sample counterpart, the empirical variance Var, we estimate the finite sample efficiency of the estimator $T$ relative to $T_0$ at the distribution $F$ by

$$\begin{aligned} \text{fs eff}(T, T_0, F) &= \frac{\text{Var}_{G_{0,k}^F}(T_0)}{\text{Var}_{G_k^F}(T)} \\ &= \frac{\frac{1}{n-1}\sum_{i=1}^n \left(T_{0,k}(i) - \frac{1}{n}\sum_{i=1}^n T_{0,k}(i)\right)^2}{\frac{1}{n-1}\sum_{i=1}^n \left(T_k(i) - \frac{1}{n}\sum_{i=1}^n T_k(i)\right)^2}. \end{aligned} \tag{A.13}$$

For biased estimators, the empirical variance is usually replaced by the mean squared error. Furthermore, $T_0$ does not have to be an estimator achieving the Cramér-Rao bound: the efficiency can also be used to measure the precision in relation to any estimator.

# B Zero Residuals in Repeated Median Regression – Proofs and Tables

In Section 2 it is described that Repeated Median (RM) regression for a simple linear model (2.1) generally leads to a separation of the data cloud into halves. This means that the number of positive residuals equals the number of negative residuals and hence, the sum of residual signs is zero (2.6). However, this is only true if the data fulfil certain requirements – the assumption of data in general position is not sufficient for all sample sizes.

In the following, we derive the RM regression coefficients for all possible combinations of observational arrangements for some small sample sizes. Furthermore, we calculate the probabilities for the location and the number of zero residuals in standard normal samples of size $n = 4$ and $n = 5$. For larger sample sizes these probabilities are approximated by simulations. According to (2.9), $\mathrm{b}(i,j)$ denotes the observational slope between $Y_i$ and $Y_j$, and the difference between $Y_i$ and $Y_k$ is denoted by $\Delta_{i,k}$ (2.11).

## Sample Size $n = 3$

According to equation (2.10), the inner medians for the RM slope calculation are defined as $\widetilde{\mathrm{b}}(i) := \underset{j \neq i}{\mathrm{med}} \ \{\mathrm{b}(i,j)\}$, (here) with $i = 1,2,3$ and $j \neq i$. Furthermore, it is

$$\begin{array}{rrcccccc}
 & \widetilde{\mathrm{b}}(1) & = & \mathrm{med}\{\mathrm{b}(1,2), \mathrm{b}(1,3)\} & = & \frac{1}{2}[\mathrm{b}(1,2) + \mathrm{b}(1,3)] & = & \frac{3}{4}\Delta_{1,2} + \frac{1}{4}\Delta_{2,3} \\
\wedge & \widetilde{\mathrm{b}}(2) & = & \mathrm{med}\{\mathrm{b}(1,2), \mathrm{b}(2,3)\} & = & \frac{1}{2}[\mathrm{b}(1,2) + \mathrm{b}(2,3)] & = & \frac{1}{2}\Delta_{1,2} + \frac{1}{2}\Delta_{2,3} \\
\wedge & \widetilde{\mathrm{b}}(3) & = & \mathrm{med}\{\mathrm{b}(1,3), \mathrm{b}(2,3)\} & = & \frac{1}{2}[\mathrm{b}(1,3) + \mathrm{b}(2,3)] & = & \frac{1}{4}\Delta_{1,2} + \frac{3}{4}\Delta_{2,3}.
\end{array}$$

From this, it follows for the RM slope (2.2) that

$$\begin{aligned}
\hat{\beta}^{RM} &= \mathrm{med}\big\{3/4\Delta_{1,2} + 1/4\Delta_{2,3},\ 1/2\Delta_{1,2} + 1/2\Delta_{2,3},\ 1/4\Delta_{1,2} + 3/4\Delta_{2,3}\big\} \\
&= 1/4(\Delta_{1,2} + \Delta_{2,3}) \ + \ 1/2 \cdot \mathrm{med}\big\{\Delta_{1,2},\ 1/2(\Delta_{1,2} + \Delta_{2,3}),\ \Delta_{2,3}\big\} \\
&\overset{(2.12)}{=} 1/4(\Delta_{1,2} + \Delta_{2,3}) \ + \ 1/2 \cdot 1/2(\Delta_{1,2} + \Delta_{2,3}) \quad = \ 1/2(\Delta_{1,2} + \Delta_{2,3}) \\
&= \mathrm{b}(1,3)\,,
\end{aligned}$$

and for the RM intercept (2.3) at design point $i = 0$ it is

$$\begin{aligned}
\hat{\mu}^{RM} &= \mathrm{med}\ \big\{\, Y_1 - \mathrm{b}(1,3),\ Y_2 - 2 \cdot \mathrm{b}(1,3),\ Y_3 - 3 \cdot \mathrm{b}(1,3) \,\big\} \\
&= \mathrm{med}\ \big\{\, 1/2 \cdot (3Y_1 - Y_3),\ Y_1 + Y_2 - Y_3,\ 1/2 \cdot (3Y_1 - Y_3) \,\big\} \\
&= 1/2 \cdot (3Y_1 - Y_3) \quad = \quad Y_1 - \mathrm{b}(1,3) \quad = \quad Y_3 - 3 \cdot \mathrm{b}(1,3)\,.
\end{aligned}$$

Thus, the RM regression line always runs through the observations at $i = 1$ and $i = 3$.

## SAMPLE SIZE $n = 4$

In a sample of size $n = 4$ there are the six different observational slopes

$$\mathrm{b}(1,2),\ \mathrm{b}(1,3),\ \mathrm{b}(1,4),\ \mathrm{b}(2,3),\ \mathrm{b}(2,4) \text{ and } \mathrm{b}(3,4)\,.$$

For a unique identification of the RM regression coefficients (2.2) and (2.3), the order of these observational slopes has to be known. This order is uniquely defined by the increasing order of observations which in turn can be characterised by the first differences $\Delta_{i,i+1}$, defined by (2.11).

The second column in Table 2.1, Section 2.2, lists the six possible permutations of the order of first differences. However, the sole specification of this order is not sufficient for a unique identification of the order of increasingly sorted observational slopes.

*Example:*

For case I, the condition $\Delta_{1,2} \le \Delta_{2,3} \le \Delta_{3,4}$ with $\mathrm{b}(i, i+1) = \Delta_{i,i+1}$ and equation (2.12) uniquely specifies the following order of slopes:

$$\mathrm{b}(1,2) \le \mathrm{b}(1,3) \le \mathrm{b}(2,3) \le \mathrm{b}(2,4) \le \mathrm{b}(3,4)$$

and with (2.12) it is $\mathrm{b}(1,3) \le \mathrm{b}(1,4) \le \mathrm{b}(3,4)$ and $\mathrm{b}(1,2) \le \mathrm{b}(1,4) \le \mathrm{b}(2,4)$

$$\Rightarrow \qquad \mathrm{b}(1,3) \le \mathrm{b}(1,4) \le \mathrm{b}(2,4)\,.$$

However, the order of $\mathrm{b}(1,4)$ and $\mathrm{b}(2,3)$ is not uniquely specified, and it depends on an additional condition on the first differences. Here, we have

$$\begin{aligned} & \mathrm{b}(1,4) && \le && \mathrm{b}(2,3) \\ \Leftrightarrow\ & 1/3 \cdot (\Delta_{1,2} + \Delta_{2,3} + \Delta_{3,4}) && \le && \Delta_{2,3} \\ \Leftrightarrow\ & \Delta_{1,2} + \Delta_{3,4} && \le && 2\Delta_{2,3}\,, \end{aligned}$$

and analogously $\Delta_{1,2} + \Delta_{3,4} \ge 2\Delta_{2,3} \quad \Leftrightarrow \quad \mathrm{b}(1,4) \ge \mathrm{b}(2,3)$.

Thus, we specify the sub-case I(1) by the conditions $(\Delta_{1,2} \le \Delta_{2,3} \le \Delta_{3,4})$ and $(\Delta_{1,2} + \Delta_{3,4} \le 2\Delta_{2,3})$ which uniquely determine the order

$$\mathrm{b}(1,2) \le \mathrm{b}(1,3) \le \mathrm{b}(1,4) \le \mathrm{b}(2,3) \le \mathrm{b}(2,4) \le \mathrm{b}(3,4)\,.$$

For both subcases of case I, cf. Table 2.1, the median slopes (2.10) correspond to

$$\begin{aligned} & \widetilde{\mathrm{b}}(1) && = && \mathrm{med}\{\mathrm{b}(1,2), \mathrm{b}(1,3), \mathrm{b}(1,4)\} && = && \mathrm{b}(1,3) \\ \wedge\ & \widetilde{\mathrm{b}}(2) && = && \mathrm{med}\{\mathrm{b}(1,2), \mathrm{b}(2,3), \mathrm{b}(2,4)\} && = && \mathrm{b}(2,3) \\ \wedge\ & \widetilde{\mathrm{b}}(3) && = && \mathrm{med}\{\mathrm{b}(1,3), \mathrm{b}(2,3), \mathrm{b}(3,4)\} && = && \mathrm{b}(2,3) \\ \wedge\ & \widetilde{\mathrm{b}}(4) && = && \mathrm{med}\{\mathrm{b}(1,4), \mathrm{b}(2,4), \mathrm{b}(3,4)\} && = && \mathrm{b}(2,4)\,, \end{aligned}$$

and because of $\mathrm{b}(1,3) \le \mathrm{b}(2,3) \le \mathrm{b}(2,4)$ the RM slope in case I corresponds to

$$\hat{\beta}^{RM} = \mathrm{med}\{\widetilde{\mathrm{b}}(1), \widetilde{\mathrm{b}}(2), \widetilde{\mathrm{b}}(3), \widetilde{\mathrm{b}}(4)\} = \mathrm{b}(2,3)\,.$$

Furthermore, it is for subcase I(1)

$$\left\{\begin{array}{rrcl} & \Delta_{1,2} & \leq & \Delta_{2,3} \\ \Leftrightarrow & Y_2 - Y_1 & \leq & Y_3 - Y_2 \\ \Leftrightarrow & Y_2 - 2(Y_3 - Y_2) & \leq & Y_1 - (Y_3 - Y_2) \end{array}\right. , \left\{\begin{array}{rrcl} & \Delta_{2,3} & \leq & \Delta_{3,4} \\ \Leftrightarrow & Y_3 - Y_2 & \leq & Y_4 - Y_3 \\ \Leftrightarrow & Y_2 - 2(Y_3 - Y_2) & \leq & Y_4 - 4(Y_3 - Y_2) \end{array}\right.$$

$$\text{and} \quad \left\{\begin{array}{rrcl} & \Delta_{1,2} + \Delta_{3,4} & \leq & 2\Delta_{2,3} \\ \Leftrightarrow & Y_4 - Y_3 + Y_2 - Y_1 & \leq & 2Y_3 - 2Y_2 \\ \Leftrightarrow & Y_4 - 4(Y_3 - Y_2) & \leq & Y_1 - (Y_3 - Y_2) \end{array}\right. ,$$

and thus, the RM intercept estimator in this subcase corresponds to

$$\begin{aligned} \hat{\mu}^{RM} &= \text{med}\left\{ Y_1 - \text{b}(2,3),\ Y_2 - 2\text{b}(2,3),\ Y_3 - 3\text{b}(2,3),\ Y_4 - 4\text{b}(2,3) \right\} \\ &= \text{med}\left\{ Y_1 - (Y_3 - Y_2),\ Y_2 - 2(Y_3 - Y_2),\ Y_3 - 3(Y_3 - Y_2),\ Y_4 - 4(Y_3 - Y_2) \right\} \\ &= \frac{1}{2}\big(Y_4 - 6Y_3 + 7Y_2\big) . \end{aligned}$$

This means: if additional to the underlying assumption of general position the conditions of case I(1) are satisfied, the RM regression line does not run through any observation. Analogously, this can be proved for case I(2).

The additional conditions, consisting of inequalities for the first differences and defining the sub-cases (1) and (2) for each of the cases II to VI, can be derived similarly. They specify the unique order of increasingly sorted slopes and hence, they also allow the determination of the coefficients $\hat{\mu}^{RM}$ and $\hat{\beta}^{RM}$. The last column of Table 2.1 in Section 2.2 contains the case IDs which uniquely define the order of the slopes given in Table B.1. While generally sub-case (1) and (2) result in different RM coefficient estimations, the RM estimates are identical for (1) and (2) if equality is fulfilled in the additional condition specified in the third column of Table 2.1. This also means that some observational slopes are identical. For example, if in case I not only $\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4}$ is fulfilled, but additionally it is $\Delta_{1,2} + \Delta_{3,4} = 2\Delta_{2,3}$, then the order of slopes is identical for sub-case (1) and (2) because it is $\text{b}(1,4) = \text{b}(2,3)$. The slopes which are identical, if the equality holds in the additional condition, are emphasised in bold blue in Table B.1.

For case I (1), the derivation of a closed form expression for the RM coefficients is shown above. This can be done analogously for all other cases. However, the derivations only have to be performed for half of the cases because the cases IV to VI describe the reversion of the cases I to III, i.e. the reflection of the data along the $y$-axis. Thus, the RM estimates can be expressed in the same way for two sub-cases each. Case VI(2), for example, can be seen as a reflection of case I(1) such that the observational slopes appear in the same order but reversed (cf. Table B.1). The obtainable expressions for the RM slope and intercept estimates for a sample of size $n = 4$ are given in Table 2.2 in Section 2.2.

| case | order of slopes | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I (1)** | b(1,2) | ≤ | b(1,3) | ≤ | **b(1,4)** | ≤ | **b(2,3)** | ≤ | b(2,4) | ≤ | b(3,4) | |
| **I (2)** | b(1,2) | ≤ | b(1,3) | ≤ | **b(2,3)** | ≤ | **b(1,4)** | ≤ | b(2,4) | ≤ | b(3,4) | |
| **II (1)** | b(3,4) | ≤ | **b(2,4)** | ≤ | **b(1,4)** | ≤ | **b(1,2)** | ≤ | b(1,3) | ≤ | b(2,3) | |
| **II (2)** | b(3,4) | ≤ | **b(1,2)** | ≤ | **b(1,4)** | ≤ | **b(2,4)** | ≤ | b(1,3) | ≤ | b(2,3) | |
| **III (1)** | b(2,3) | ≤ | b(2,4) | ≤ | **b(1,3)** | ≤ | **b(1,4)** | ≤ | **b(3,4)** | ≤ | b(1,2) | |
| **III (2)** | b(2,3) | ≤ | b(2,4) | ≤ | **b(3,4)** | ≤ | **b(1,4)** | ≤ | **b(1,3)** | ≤ | b(1,2) | |
| **IV (1)** | b(1,2) | ≤ | **b(1,3)** | ≤ | **b(1,4)** | ≤ | **b(3,4)** | ≤ | b(2,4) | ≤ | b(2,3) | |
| **IV (2)** | b(1,2) | ≤ | **b(3,4)** | ≤ | **b(1,4)** | ≤ | **b(1,3)** | ≤ | b(2,4) | ≤ | b(2,3) | |
| **V (1)** | b(2,3) | ≤ | b(1,3) | ≤ | **b(2,4)** | ≤ | **b(1,4)** | ≤ | **b(1,2)** | ≤ | b(3,4) | |
| **V (2)** | b(2,3) | ≤ | b(1,3) | ≤ | **b(1,2)** | ≤ | **b(1,4)** | ≤ | **b(2,4)** | ≤ | b(3,4) | |
| **VI (1)** | b(3,4) | ≤ | b(2,4) | ≤ | **b(1,4)** | ≤ | **b(2,3)** | ≤ | b(1,3) | ≤ | b(1,2) | |
| **VI (2)** | b(3,4) | ≤ | b(2,4) | ≤ | **b(2,3)** | ≤ | **b(1,4)** | ≤ | b(1,3) | ≤ | b(1,2) | |

**Table B.1**: Order of the observational slopes for a sample of size $n = 4$. If equality is fulfilled for the additional condition given in the third column of Table 2.1, then the equality also holds for the slopes marked in bold blue, and the sub-cases (1) and (2) do not differ.

### PROBABILITIES FOR POSSIBLE DATA SITUATIONS IN A SAMPLE OF SIZE $n = 4$

Under the assumption that $y_1, y_2, y_3, y_4$ are observations of independent, identically standard normally distributed random variables $Y_1, Y_2, Y_3, Y_4$, we will derive the probability of each of the cases listed in Table 2.1. For this we use the fact that for a linear transformation $\boldsymbol{Z}_1 = \boldsymbol{B}\boldsymbol{Z}_0$ of a multivariate, normally distributed random vector $\boldsymbol{Z}_0$ it is:

$$\boldsymbol{Z}_0 \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \Rightarrow \quad \boldsymbol{Z}_1 \sim N\left(\boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\top\right) . \tag{B.1}$$

To calculate the probability for the occurrence of a certain data situation characterised by the conditions given in Table 2.1, it is possible to find a linear transformation of the random vector $\boldsymbol{Y} = (Y_1, Y_2, Y_3, Y_4)'$ representing this case. Since $Y_1, Y_2, Y_3, Y_4$ are i.i.d. $N(0,1)$-distributed, $\boldsymbol{Y}$ has a multivariate standard normal distribution, i.e. with $\mathbf{0_4}$ being the four dimensional origin and $\boldsymbol{I_4}$ the $(4 \times 4)$ identity matrix, it is $\boldsymbol{Y} \sim N(\mathbf{0_4}, \boldsymbol{I_4})$.

#### Case I

The probability for case I(1) (cf. Table 2.1) is equal to

$$\begin{aligned} & P\big((\Delta_{1,2} \le \Delta_{2,3} \le \Delta_{3,4}) \wedge (\Delta_{1,2} + \Delta_{3,4} \le 2\Delta_{2,3})\big) \\ &= P\big((\Delta_{2,3} - \Delta_{1,2} \ge 0) \wedge (\Delta_{3,4} - \Delta_{2,3} \ge 0) \wedge (\Delta_{3,4} - \Delta_{2,3} \le \Delta_{2,3} - \Delta_{1,2})\big) , \end{aligned}$$

where $\Delta_{i,i+1} = Y_{i+1} - Y_i$ for $i \in \{1, 2, 3\}$.

Let $S_{I,1} = \Delta_{2,3} - \Delta_{1,2}$ and $S_{I,2} = \Delta_{3,4} - \Delta_{2,3}$. The probability of case I(1) occurring is then given by

$$P\big((S_{I,1} \geq 0) \wedge (S_{I,2} \geq 0) \wedge (S_{I,2} \leq S_{I,1})\big) \;=\; P\big((S_{I,1} \geq 0) \wedge (S_{I,2} \geq 0) \wedge (S_{I,2} - S_{I,1} \leq 0)\big)\,.$$

Furthermore, let $\boldsymbol{T}_I = \begin{pmatrix} S_{I,1} \\ S_{I,2} \\ S_{I,2} - S_{I,1} \end{pmatrix}$, i.e. $\boldsymbol{T}_I = \boldsymbol{B}_I \boldsymbol{Y}$ with $\boldsymbol{B}_I = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix}$, a three dimensional random variable, corresponding to the second and third differences of $\boldsymbol{Y}$. The probability we search for can then be evaluated via the distribution of $\boldsymbol{T}_I$. According to (B.1), it is $\boldsymbol{T}_I \sim N\big(\boldsymbol{0}_3, \boldsymbol{\Sigma}_{\boldsymbol{T}_I}\big)$ with $\boldsymbol{\Sigma}_{\boldsymbol{T}_I} = \boldsymbol{B}_I \boldsymbol{B}_I^\top = \begin{pmatrix} 6 & -4 & -10 \\ -4 & 6 & 10 \\ -10 & 10 & 20 \end{pmatrix}$. With $f_{\boldsymbol{T}_I}$ denoting the density of the distribution of $\boldsymbol{T}_I$ and $\boldsymbol{t} = (t_1,\, t_2,\, t_3)'$ an observation of this random variable, the probability of the occurrence of case I(1) corresponds to

$$\begin{aligned} & P\big((\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4}) \wedge (\Delta_{1,2} + \Delta_{3,4} \leq 2\Delta_{2,3})\big) \\ &= P\big((S_{I,1} \geq 0) \wedge (S_{I,2} \geq 0) \wedge (S_{I,2} - S_{I,1} \leq 0)\big) \\ &= \int_0^\infty \int_0^\infty \int_{-\infty}^0 f_{\boldsymbol{T}_I}(\boldsymbol{t})\; dt_3\, dt_2\, dt_1 \\ &= 0.06693\,. \end{aligned}$$

Analogously, the probability for case I(2) is given by

$$\begin{aligned} & P\big((\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4}) \wedge (\Delta_{1,2} + \Delta_{3,4} \geq 2\Delta_{2,3})\big) \\ &= P\big((S_{I,1} \geq 0) \wedge (S_{I,2} \geq 0) \wedge (S_{I,2} - S_{I,1} \geq 0)\big) \\ &= \int_0^\infty \int_0^\infty \int_0^\infty f_{\boldsymbol{T}_I}(\boldsymbol{t})\; dt_3\, dt_2\, dt_1 \\ &= 0.06693\,, \end{aligned}$$

where the results for these (and the following) integrals have been calculated numerically. Thus, the probability for case I corresponds to

$$P\big(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4}\big) \;=\; P\big((S_{I,1} \geq 0) \wedge (S_{I,2} \geq 0)\big) \;=\; 0.06693 + 0.06693 \;=\; 0.13386\,.$$

**Case II**

The probabilities for case II(1) and II(2) can be derived similarly to the first case, using the distribution of $\boldsymbol{T}_{II} = (S_{II,1}, S_{II,2}, S_{II,2} - S_{II,1})'$ with $S_{II,1} = \Delta_{1,2} - \Delta_{3,4}$ and $S_{II,2} = \Delta_{2,3} - \Delta_{1,2}$.

With $\boldsymbol{B}_{II} = \begin{pmatrix} -1 & 1 & 1 & -1 \\ 1 & -2 & 1 & 0 \\ 2 & -3 & 0 & 1 \end{pmatrix}$ and $\boldsymbol{\Sigma}_{\boldsymbol{T}_{II}} = \boldsymbol{B}_{II} \boldsymbol{B}_{II}^\top = \begin{pmatrix} 4 & -2 & -6 \\ -2 & 6 & 8 \\ -6 & 8 & 14 \end{pmatrix}$

it is $\boldsymbol{T}_{II} = \boldsymbol{B}_{II}\boldsymbol{Y} \sim N(\boldsymbol{0_3}, \boldsymbol{\Sigma}_{\boldsymbol{T}_{II}})$, and if $f_{\boldsymbol{T}_{II}}$ denotes the density of the corresponding normal distribution, the probability for case II(1) is given by

$$\begin{aligned} & P\big((\Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{2,3}) \wedge (\Delta_{2,3} + \Delta_{3,4} \leq 2\Delta_{1,2})\big) \\ &= P\big((S_{II,1} \geq 0) \wedge (S_{II,2} \geq 0) \wedge (S_{II,2} - S_{II,1} \leq 0)\big) \\ &= \int_0^\infty \int_0^\infty \int_{-\infty}^0 f_{\boldsymbol{T}_{II}}(\boldsymbol{t})\, dt_3\, dt_2\, dt_1 \\ &= 0.08113\,. \end{aligned}$$

The probability for case II(2) corresponds to

$$\begin{aligned} & P\big((\Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{2,3}) \wedge (\Delta_{2,3} + \Delta_{3,4} \geq 2\Delta_{1,2})\big) \\ &= \int_0^\infty \int_0^\infty \int_0^\infty f_{\boldsymbol{T}_{II}}(\boldsymbol{t})\, dt_3\, dt_2\, dt_1 \\ &= 0.10194\,, \end{aligned}$$

and hence, the probability of the occurrence of case II equals

$$P\big(\Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{2,3}\big) = 0.18307\,.$$

**Case III**

Analogously, the probabilities for case III(1) and III(2) can be derived via the distribution of $\boldsymbol{T}_{III} = (S_{III,1},\, S_{III,2},\, S_{III,2} - S_{III,1})'$, where $S_{III,1} = \Delta_{3,4} - \Delta_{2,3}$ and $S_{III,2} = \Delta_{1,2} - \Delta_{3,4}$.

With $\boldsymbol{B}_{III} = \begin{pmatrix} 0 & 1 & -2 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 0 & 3 & -2 \end{pmatrix}$ and $\boldsymbol{\Sigma}_{\boldsymbol{T}_{III}} = \boldsymbol{B}_{\boldsymbol{III}}\boldsymbol{B}_{III}^\top = \begin{pmatrix} 6 & -2 & -8 \\ -2 & 4 & 6 \\ -8 & 6 & 14 \end{pmatrix}$

it is $\boldsymbol{T}_{III} = \boldsymbol{B}_{III}\boldsymbol{Y} \sim N(\boldsymbol{0_3}, \boldsymbol{\Sigma}_{\boldsymbol{T}_{III}})$. For $f_{\boldsymbol{T}_{III}}$ denoting the density of the corresponding normal distribution, the probability for case III(1) is equal to

$$\begin{aligned} & P\big((\Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{1,2}) \wedge (\Delta_{1,2} + \Delta_{2,3} \leq 2\Delta_{3,4})\big) \\ &= P\big((S_{III,1} \geq 0) \wedge (S_{III,2} \geq 0) \wedge (S_{III,2} - S_{III,1} \leq 0)\big) \\ &= \int_0^\infty \int_0^\infty \int_{-\infty}^0 f_{\boldsymbol{T}_{III}}(\boldsymbol{t})\, dt_3\, dt_2\, dt_1 \\ &= 0.10194 \end{aligned}$$

and the probability for case III(2) corresponds to

$$\begin{aligned} & P\big((\Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{1,2}) \wedge (\Delta_{1,2} + \Delta_{2,3} \geq 2\Delta_{3,4})\big) \\ &= \int_0^\infty \int_0^\infty \int_0^\infty f_{\boldsymbol{T}_{III}}(\boldsymbol{t})\, dt_3\, dt_2\, dt_1 \\ &= 0.08113\,. \end{aligned}$$

Consequently, the probability of case III occurring is the same as for case II:

$$P\big(\Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{1,2}\big) = 0.18307\,.$$

**Cases IV, V and VI**

Because of the symmetry of the normal distribution, the probabilities of the remaining cases do not have to be calculated explicitly; they simply follow from the probabilities derived above: The probability for the occurrence of case IV(1) corresponds to the probability of case III(2) and similarly, the probability for case IV(2) is equal to the probability for case III(1) because $\boldsymbol{T_{IV}} = -\boldsymbol{T_{III}}$ with $\boldsymbol{B_{IV}} = -\boldsymbol{B_{III}}$ and $\boldsymbol{\Sigma_{B_{IV}}} = \boldsymbol{\Sigma_{B_{III}}}$. Likewise, it is $\boldsymbol{T}_V = -\boldsymbol{T}_{II}$ with $\boldsymbol{B}_V = -\boldsymbol{B}_{II}$ and $\boldsymbol{\Sigma_{B}}_V = \boldsymbol{\Sigma_{B}}_{II}$, as well as $\boldsymbol{T}_{VI} = -\boldsymbol{T}_I$ with $\boldsymbol{B}_{VI} = -\boldsymbol{B}_I$ and $\boldsymbol{\Sigma_{B}}_{VI} = \boldsymbol{\Sigma_{B}}_I$.

The probabilities resulting for all cases specified in Tables 2.1 and B.1 are given in Table 2.3 in Section 2.2.

| case | condition | prob. | case | condition (reverse to I - XII) |
|---|---|---|---|---|
| **I** | $(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{4,5})$ | 0.022 | **XXIV** | $(\Delta_{4,5} \leq \Delta_{3,4} \leq \Delta_{2,3} \leq \Delta_{1,2})$ |
| **II** | $(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{4,5} \leq \Delta_{3,4})$ | 0.026 | **XXIII** | $(\Delta_{3,4} \leq \Delta_{4,5} \leq \Delta_{2,3} \leq \Delta_{1,2})$ |
| **III** | $(\Delta_{1,2} \leq \Delta_{3,4} \leq \Delta_{2,3} \leq \Delta_{4,5})$ | 0.038 | **XXII** | $(\Delta_{4,5} \leq \Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{1,2})$ |
| **IV** | $(\Delta_{1,2} \leq \Delta_{3,4} \leq \Delta_{4,5} \leq \Delta_{2,3})$ | 0.060 | **XXI** | $(\Delta_{2,3} \leq \Delta_{4,5} \leq \Delta_{3,4} \leq \Delta_{1,2})$ |
| **V** | $(\Delta_{1,2} \leq \Delta_{4,5} \leq \Delta_{2,3} \leq \Delta_{3,4})$ | 0.038 | **XX** | $(\Delta_{3,4} \leq \Delta_{2,3} \leq \Delta_{4,5} \leq \Delta_{1,2})$ |
| **VI** | $(\Delta_{1,2} \leq \Delta_{4,5} \leq \Delta_{3,4} \leq \Delta_{2,3})$ | 0.048 | **XIX** | $(\Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{4,5} \leq \Delta_{1,2})$ |
| **VII** | $(\Delta_{2,3} \leq \Delta_{1,2} \leq \Delta_{3,4} \leq \Delta_{4,5})$ | 0.026 | **XVIII** | $(\Delta_{4,5} \leq \Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{2,3})$ |
| **VIII** | $(\Delta_{2,3} \leq \Delta_{1,2} \leq \Delta_{4,5} \leq \Delta_{3,4})$ | 0.033 | **XVII** | $(\Delta_{3,4} \leq \Delta_{4,5} \leq \Delta_{1,2} \leq \Delta_{2,3})$ |
| **IX** | $(\Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{4,5})$ | 0.038 | **XVI** | $(\Delta_{4,5} \leq \Delta_{1,2} \leq \Delta_{3,4} \leq \Delta_{2,3})$ |
| **X** | $(\Delta_{2,3} \leq \Delta_{4,5} \leq \Delta_{1,2} \leq \Delta_{3,4})$ | 0.065 | **XV** | $(\Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{4,5} \leq \Delta_{2,3})$ |
| **XI** | $(\Delta_{3,4} \leq \Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{4,5})$ | 0.059 | **XIV** | $(\Delta_{4,5} \leq \Delta_{2,3} \leq \Delta_{1,2} \leq \Delta_{3,4})$ |
| **XII** | $(\Delta_{3,4} \leq \Delta_{2,3} \leq \Delta_{1,2} \leq \Delta_{4,5})$ | 0.048 | **XIII** | $(\Delta_{4,5} \leq \Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4})$ |

**Table B.2**: Case differentiation for data situations in a sample of size $n = 5$ distinguished by the order of the first differences $\Delta_{i,i+1}$ and the corresponding probability of occurrence of each case for independent standard normal data.

## SAMPLE SIZE $n = 5$

For a sample size of $n = 5$, the probabilities for two, one or no zero residual will be derived in the same way as for sample size $n = 4$. Here, we have the $\binom{n}{2} = \frac{n(n-1)}{2} = 10$ observational slopes

$$\text{b}(1,2),\ \text{b}(1,3),\ \text{b}(1,4),\ \text{b}(1,5),\ \text{b}(2,3),\ \text{b}(2,4),\ \text{b}(2,5),\ \text{b}(3,4),\ \text{b}(3,5),\ \text{and}\ \text{b}(4,5).$$

The order of the increasingly sorted slopes is again determined by the arrangement or order of the observations which, as above, can be characterised by the first differences $\Delta_{i,i+1}$. The $(n-1)! = 24$ possible orders of the first differences, specifying the different cases, are given in Table B.2. Since the cases XIII to XXIV correspond to mirrored versions of the cases I to XII, they yield analogous expressions for the RM regression estimates. Because the calculations are similar for each case we will only show the derivations for the cases I to III.

The conditions on the first differences listed in Table B.2 are not sufficient to uniquely determine the order of increasingly sorted slopes. To be able to do so, for some cases additional conditions are necessary. These will be derived below, together with the corresponding order of observational slopes and RM coefficients. To calculate the probabilities for the occurrence of each of the cases for $n = 5$ independent observations coming from standard normally distributed random variables $Y_1, Y_2, Y_3, Y_4, Y_5$, we use equation (B.1). Therefore, let $\boldsymbol{Y} = (Y_1, Y_2, Y_3, Y_4, Y_5)' \sim N(\mathbf{0_5}, \boldsymbol{I_5})$ with $\mathbf{0_5} = (0, 0, 0, 0, 0)'$ and $\boldsymbol{I_5}$ the $(5 \times 5)$ identity matrix.

**CASE I** $(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{4,5})$

Let $S_1 = \Delta_{2,3} - \Delta_{1,2}$, $S_2 = \Delta_{3,4} - \Delta_{2,3}$ and $S_3 = \Delta_{4,5} - \Delta_{3,4}$, with $\Delta_{i,i+1}$ denoting the random variables of first differences of the $Y_i$. For the probability of case I we have

$$\begin{aligned} P\big(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{4,5}\big) &= P\big((\Delta_{1,2} \leq \Delta_{2,3}) \wedge (\Delta_{2,3} \leq \Delta_{3,4}) \wedge (\Delta_{3,4} \leq \Delta_{4,5})\big) \\ &= P\big((S_1 \geq 0) \wedge (S_2 \geq 0) \wedge (S_3 \geq 0)\big)\,. \end{aligned}$$

This can be evaluated via the joint distribution of $S_1$, $S_2$ and $S_3$. Let $\boldsymbol{T}_{\boldsymbol{I_0}} = (S_1, S_2, S_3)'$. According to equation (B.1) it is $\boldsymbol{T}_{\boldsymbol{I_0}} \sim N(\mathbf{0_3}, \boldsymbol{\Sigma}_{\boldsymbol{I_0}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{I_0}} = \begin{pmatrix} 6 & -4 & 1 \\ -4 & 6 & -4 \\ 1 & -4 & 6 \end{pmatrix}$. Numerical integration yields $P\big(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{3,4} \leq \Delta_{4,5}\big) = 0.02219$. The probabilities of all cases are listed in the third column of Table B.2.

Since we are interested in the number and position of zero residuals in the sample, we will now derive the RM coefficients for case I. With equation (2.12) we can deduce that

$$\left\{ \begin{array}{ll} & \mathrm{b}(1,2) \leq \mathrm{b}(1,3) \leq \mathrm{b}(1,4) \leq \mathrm{b}(1,5) \\ \wedge & \mathrm{b}(1,2) \leq \mathrm{b}(2,3) \leq \mathrm{b}(2,4) \leq \mathrm{b}(2,5) \\ \wedge & \mathrm{b}(1,3) \leq \mathrm{b}(2,3) \leq \mathrm{b}(3,4) \leq \mathrm{b}(3,5) \\ \wedge & \mathrm{b}(1,4) \leq \mathrm{b}(2,4) \leq \mathrm{b}(3,4) \leq \mathrm{b}(4,5) \\ \wedge & \mathrm{b}(1,5) \leq \mathrm{b}(2,5) \leq \mathrm{b}(3,5) \leq \mathrm{b}(4,5) \end{array} \right.$$

$$\Rightarrow \quad \left\{ \begin{array}{lll} & \widetilde{\mathrm{b}}(1) = \mathrm{med}\{\mathrm{b}(1,3), \mathrm{b}(1,4)\} = \frac{1}{12}\left[5\Delta_{1,2} + 5\Delta_{2,3} + 2\Delta_{3,4}\right] \\ \wedge & \widetilde{\mathrm{b}}(2) = \mathrm{med}\{\mathrm{b}(2,3), \mathrm{b}(2,4)\} = \frac{1}{4}\left[3\Delta_{2,3} + \Delta_{3,4}\right] \\ \wedge & \widetilde{\mathrm{b}}(3) = \mathrm{med}\{\mathrm{b}(2,3), \mathrm{b}(3,4)\} = \frac{1}{2}\left[\Delta_{2,3} + \Delta_{3,4}\right] = \mathrm{b}(2,4) \\ \wedge & \widetilde{\mathrm{b}}(4) = \mathrm{med}\{\mathrm{b}(2,4), \mathrm{b}(3,4)\} = \frac{1}{4}\left[\Delta_{2,3} + 3\Delta_{3,4}\right] \\ \wedge & \widetilde{\mathrm{b}}(5) = \mathrm{med}\{\mathrm{b}(2,5), \mathrm{b}(3,5)\} = \frac{1}{12}\left[2\Delta_{2,3} + 5\Delta_{3,4} + 5\Delta_{4,5}\right] \end{array} \right. .$$

In addition, we have $\widetilde{\mathrm{b}}(1) \leq \widetilde{\mathrm{b}}(2) \leq \widetilde{\mathrm{b}}(3) \leq \widetilde{\mathrm{b}}(4) \leq \widetilde{\mathrm{b}}(5)$ and thus

$$\hat{\beta}^{RM} = \mathrm{med}\{\widetilde{\mathrm{b}}(1), \widetilde{\mathrm{b}}(2), \widetilde{\mathrm{b}}(3), \widetilde{\mathrm{b}}(4), \widetilde{\mathrm{b}}(5)\} = \widetilde{\mathrm{b}}(3) = \mathrm{b}(2,4)\,.$$

Furthermore, it is $Y_3 - 3\hat{\beta}^{RM} \leq Y_2 - 2\hat{\beta}^{RM} = Y_4 - 4\hat{\beta}^{RM} \leq Y_5 - 5\hat{\beta}^{RM}$ and $Y_4 - 4\hat{\beta}^{RM} \leq Y_1 - \hat{\beta}^{RM}$. Hence, we get

$$\Rightarrow \quad \hat{\mu}^{RM} = Y_2 - 2\hat{\beta}^{RM} = Y_4 - 4\hat{\beta}^{RM}\,.$$

This means that the RM line runs through $Y_2$ and $Y_4$ for data fulfilling the conditions of case I. If the strict inequality is true in all of these conditions, we have for the corresponding residuals $r_1 > 0$, $r_2 = 0$, $r_3 < 0$, $r_4 = 0$ and $r_5 > 0$, i.e. $\sum_{i=1}^{n} \mathrm{sign}(r_i) = 1$. Hence, the balance of residual signs (2.6) is not fulfilled here. Nevertheless, the median of the residuals is zero.

| case II | condition | | | order of slopes |
|---|---|---|---|---|
| **(i)** | $\Delta_{1,2}+\Delta_{3,4}+\Delta_{4,5}$ | $\asymp$ | $3\Delta_{2,3}$ | $\mathrm{b}(1,5) \asymp \mathrm{b}(2,3)$ |
| **(ii)** | $\Delta_{1,2}+\Delta_{3,4}$ | $\asymp$ | $2\Delta_{2,3}$ | $\mathrm{b}(1,4) \asymp \mathrm{b}(2,3)$ |
| **(iii)** | $\Delta_{1,2}+\Delta_{2,3}+\Delta_{3,4}$ | $\asymp$ | $3\Delta_{4,5}$ | $\mathrm{b}(1,4) \asymp \mathrm{b}(1,5) \asymp \mathrm{b}(4,5)$ |

**Table B.3**: Additional conditions on the first differences and resulting order of observational slopes for case II in a sample of size $n = 5$. The operator $\asymp$ can be replaced by either $\leq$ or $\geq$ in each of the conditions.

| case | operator $\asymp$ in (i) | (ii) | (iii) | order of observational slopes |
|---|---|---|---|---|
| **II(1)** | $\leq$ | $\leq$ | $\leq$ | $\mathrm{b}(1,2)\leq\mathrm{b}(1,3)\leq\mathrm{b}(1,4)\leq\mathrm{b}(1,5)\leq\mathrm{b}(2,3)\leq\mathrm{b}(4,5)\leq\mathrm{b}(2,5)\leq\mathrm{b}(2,4)\leq\mathrm{b}(3,5)\leq\mathrm{b}(3,4)$ |
| **II(2)** | $\geq$ | $\leq$ | $\leq$ | $\mathrm{b}(1,2)\leq\mathrm{b}(1,3)\leq\mathrm{b}(1,4)\leq\mathrm{b}(2,3)\leq\mathrm{b}(1,5)\leq\mathrm{b}(4,5)\leq\mathrm{b}(2,5)\leq\mathrm{b}(2,4)\leq\mathrm{b}(3,5)\leq\mathrm{b}(3,4)$ |
| **II(3)** | $\geq$ | $\geq$ | $\leq$ | $\mathrm{b}(1,2)\leq\mathrm{b}(1,3)\leq\mathrm{b}(2,3)\leq\mathrm{b}(1,4)\leq\mathrm{b}(1,5)\leq\mathrm{b}(4,5)\leq\mathrm{b}(2,5)\leq\mathrm{b}(2,4)\leq\mathrm{b}(3,5)\leq\mathrm{b}(3,4)$ |
| **II(4)** | $\geq$ | $\geq$ | $\geq$ | $\mathrm{b}(1,2)\leq\mathrm{b}(1,3)\leq\mathrm{b}(2,3)\leq\mathrm{b}(4,5)\leq\mathrm{b}(1,5)\leq\mathrm{b}(1,4)\leq\mathrm{b}(2,5)\leq\mathrm{b}(2,4)\leq\mathrm{b}(3,5)\leq\mathrm{b}(3,4)$ |

**Table B.4**: Definition of sub-cases of case II for sample size $n = 5$ by specification of the inequality signs replacing the operator $\asymp$ in Table B.3. The last column gives the resulting order of observational slopes; under the condition $(\Delta_{1,2} \leq \Delta_{2,3} \leq \Delta_{4,5} \leq \Delta_{3,4})$ for case II no other order of observational slopes is possible.

### CASE II

The probability of case II occurring for $n = 5$ observations from standard normal i.i.d. variables can be derived similar to case I by using the joint distribution of the random variables $S_1 = \Delta_{2,3} - \Delta_{1,2}$, $S_2 = \Delta_{4,5} - \Delta_{2,3}$ and $S_3 = \Delta_{3,4} - \Delta_{4,5}$. However, further case differentiations are necessary here, to determine the cases in which zero residuals occur and at which positions within the sample they are located. The probabilities for such subcases can again be derived via the joint distribution of $S_1$, $S_2$, $S_3$ and suitable linear combinations of them.

To be able to uniquely sort the ten observational slopes, the appropriate inequality signs to replace the operator $\asymp$ in the additional conditions specified in Table B.3 have to be known. Below, we will use the notation $\mathbf{(i)}^{\geq}$ if the operator $\asymp$ is replaced by $\geq$ in the first row of Table B.3; and we use $\mathbf{(i)}_{\leq}$ if $\asymp$ is replaced by $\leq$. For example, $\mathbf{(iii)}_{\leq}$ corresponds to $\Delta_{1,2}+\Delta_{2,3}+\Delta_{3,4} \leq 3\Delta_{4,5}$ and thus it is $\mathrm{b}(1,4) \leq \mathrm{b}(1,5) \leq \mathrm{b}(4,5)$. In particular, we have for the three conditions from Table B.3 $\mathbf{(i)}_{\leq} \Rightarrow \mathbf{(ii)}_{\leq} \Rightarrow \mathbf{(iii)}_{\leq}$, and $\mathbf{(iii)}^{\geq} \Rightarrow \mathbf{(ii)}^{\geq} \Rightarrow \mathbf{(i)}^{\geq}$. In consideration of these implications, four possible sets of ordered observational slopes remain for case II, see Table B.4.

The unique order of observational slopes is not enough for a unique identification of expressions for the RM coefficients. Table B.5 gives further conditions for that and Table B.6 lists the corresponding results with their probabilities and the order of residual signs showing that (2.6) is fulfilled with probability one in case of continuous errors.

| cases | | additional conditions | | |
|---|---|---|---|---|
| **II(1), II(2), II(3)** | **(a)** | $(4\Delta_{2,3} + \Delta_{3,4} \leq 5\Delta_{4,5})$ | $\wedge$ | $(2\Delta_{2,3} + \Delta_{3,4} \leq 3\Delta_{4,5})$ |
| **II(1), II(2), II(3)** | **(b)** | $(4\Delta_{2,3} + \Delta_{3,4} \leq 5\Delta_{4,5})$ | $\wedge$ | $(2\Delta_{2,3} + \Delta_{3,4} \geq 3\Delta_{4,5})$ |
| **II(1), II(2), II(3)** | **(c)** | $(4\Delta_{2,3} + \Delta_{3,4} \geq 5\Delta_{4,5})$ | | |
| **II(4)** | **(a)** | $(4\Delta_{2,3} + \Delta_{3,4} \leq 5\Delta_{4,5})$ | | |
| **II(4)** | **(b)** | $(4\Delta_{2,3} + \Delta_{3,4} \geq 5\Delta_{4,5})$ | $\wedge$ | $(3\Delta_{1,2} + \Delta_{2,3} + \Delta_{3,4} \leq 5\Delta_{4,5})$ |
| **II(4)** | **(c)** | $(4\Delta_{2,3} + \Delta_{3,4} \geq 5\Delta_{4,5})$ | $\wedge$ | $(3\Delta_{1,2} + \Delta_{2,3} + \Delta_{3,4} \geq 5\Delta_{4,5})$ |

**Table B.5**: Further case differentiation of the cases II(1) to II(4) for $n = 5$ by specifying additional conditions for the unique determination of the RM regression coefficients.

| case | $\hat{\boldsymbol{\beta}}^{RM}$ | $\hat{\boldsymbol{\mu}}^{RM}$ | order of residual signs | probability |
|---|---|---|---|---|
| **II(1)(a)** | $\frac{1}{4}[2\Delta_{2,3} + \Delta_{3,4} + \Delta_{4,5}]$ | $Y_4 - 4\hat{\beta}^{RM}$ | (1 -1 -1 0 1) | 0.00232 |
| **II(1)(b)** | $\frac{1}{4}[2\Delta_{2,3} + \Delta_{3,4} + \Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00076 |
| **II(1)(c)** | $\frac{1}{6}[\Delta_{2,3} + \Delta_{3,4} + 4\Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00146 |
| **II(2)(a)** | $\frac{1}{4}[2\Delta_{2,3} + \Delta_{3,4} + \Delta_{4,5}]$ | $Y_4 - 4\hat{\beta}^{RM}$ | (1 -1 -1 0 1) | 0.00167 |
| **II(2)(b)** | $\frac{1}{4}[2\Delta_{2,3} + \Delta_{3,4} + \Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00024 |
| **II(2)(c)** | $\frac{1}{6}[\Delta_{2,3} + \Delta_{3,4} + 4\Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00017 |
| **II(3)(a)** | $\frac{1}{4}[2\Delta_{2,3} + \Delta_{3,4} + \Delta_{4,5}]$ | $Y_4 - 4\hat{\beta}^{RM}$ | (1 -1 -1 0 1) | 0.01036 |
| **II(3)(b)** | $\frac{1}{4}[2\Delta_{2,3} + \Delta_{3,4} + \Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00203 |
| **II(3)(c)** | $\frac{1}{6}[\Delta_{2,3} + \Delta_{3,4} + 4\Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00079 |
| **II(4)(a)** | $\frac{1}{4}[2\Delta_{2,3} + \Delta_{3,4} + \Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00122 |
| **II(4)(b)** | $\frac{1}{6}[\Delta_{2,3} + \Delta_{3,4} + 4\Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00103 |
| **II(4)(c)** | $\frac{1}{6}[\Delta_{2,3} + \Delta_{3,4} + 4\Delta_{4,5}]$ | $Y_5 - 5\hat{\beta}^{RM}$ | (1 -1 -1 1 0) | 0.00371 |

**Table B.6**: Expressions for the RM coefficients for all sub-cases of case II for $n = 5$. The fourth column gives the corresponding order of residual signs, the fifth column contains the probability of this event occurring for standard normal data.

### CASE III

Table B.7 shows five additional conditions for a unique identification of the RM coefficients. For these conditions we have $\mathbf{(i)}_{\leq} \Rightarrow \mathbf{(iii)}_{\leq} \Rightarrow \mathbf{(v)}_{\leq}$, $\mathbf{(i)}_{\leq} \Rightarrow \mathbf{(ii)}_{\leq}$, $\mathbf{(iv)}_{\leq} \Rightarrow \mathbf{(v)}_{\leq}$, $\mathbf{(v)}^{\geq} \Rightarrow \mathbf{(iii)}^{\geq} \Rightarrow \mathbf{(i)}^{\geq}$, $\mathbf{(ii)}^{\geq} \Rightarrow \mathbf{(i)}^{\geq}$, $\mathbf{(v)}^{\geq} \Rightarrow \mathbf{(iv)}^{\geq}$ as well as $(\mathbf{(ii)}_{\leq}$ & $\mathbf{(iii)}^{\geq}) \Rightarrow \mathbf{(iv)}^{\geq}$ and $(\mathbf{(ii)}^{\geq}$ & $\mathbf{(iii)}_{\leq}) \Rightarrow \mathbf{(iv)}_{\leq}$. Thus, there are ten possibilities for ordering the observational slopes. Table B.8 lists these cases together with the according RM regression coefficients and their probabilities at standard normal data. Here, all sub-cases result in the order (1, -1, 0, -1, 1) for the residual signs, i.e. (2.6) is fulfilled with probability one.

| | case III conditions | order of slopes |
|---|---|---|
| **(i)** | $\Delta_{1,2} + \Delta_{2,3} + \Delta_{4,5} \asymp 3\Delta_{3,4}$ | $\mathrm{b}(1,5) \asymp \mathrm{b}(3,4)$ |
| **(ii)** | $\Delta_{1,2} + \Delta_{2,3} \asymp 2\Delta_{3,4}$ | $\mathrm{b}(1,3) \asymp \mathrm{b}(1,4) \asymp \mathrm{b}(3,4)$ |
| **(iii)** | $\Delta_{1,2} + \Delta_{4,5} \asymp \Delta_{2,3} + \Delta_{3,4}$ | $\mathrm{b}(1,5) \asymp \mathrm{b}(2,4)$ |
| **(iv)** | $\Delta_{3,4} + \Delta_{4,5} \asymp 2\Delta_{2,3}$ | $\mathrm{b}(3,5) \asymp \mathrm{b}(2,5) \asymp \mathrm{b}(2,3)$ |
| **(v)** | $\Delta_{1,2} + \Delta_{3,4} + \Delta_{4,5} \asymp 3\Delta_{2,3}$ | $\mathrm{b}(1,5) \asymp \mathrm{b}(2,3)$ |

**Table B.7**: Additional conditions on the first differences and resulting order of observational slopes for case III in a sample of size $n = 5$ where the operator $\asymp$ can be replaced by $\leq$ or by $\geq$.

| **case** | **operator $\asymp$ in** (i) | (ii) | (iii) | (iv) | (v) | $\widehat{\boldsymbol{\beta}}^{RM}$ | $\widehat{\boldsymbol{\mu}}^{RM}$ | **probability** |
|---|---|---|---|---|---|---|---|---|
| **III(1)** | $\leq$ | $\leq$ | $\leq$ | $\leq$ | $\leq$ | $\frac{1}{4}[3\Delta_{3,4} + \Delta_{4,5}]$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00209 |
| **III(2)** | $\geq$ | $\leq$ | $\leq$ | $\leq$ | $\leq$ | $\frac{1}{4}[3\Delta_{3,4} + \Delta_{4,5}]$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00399 |
| **III(3)** | $\leq$ | $\leq$ | $\leq$ | $\geq$ | $\leq$ | $\mathrm{b}(2,4)$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00323 |
| **III(4)** | $\geq$ | $\leq$ | $\leq$ | $\geq$ | $\leq$ | $\mathrm{b}(2,4)$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00262 |
| **III(5)** | $\geq$ | $\leq$ | $\geq$ | $\geq$ | $\leq$ | $\mathrm{b}(2,4)$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00262 |
| **III(6)** | $\geq$ | $\leq$ | $\geq$ | $\geq$ | $\geq$ | $\mathrm{b}(2,4)$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00323 |
| **III(7)** | $\geq$ | $\geq$ | $\leq$ | $\leq$ | $\leq$ | $\mathrm{b}(1,5)$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00706 |
| **III(8)** | $\geq$ | $\geq$ | $\geq$ | $\leq$ | $\leq$ | $\mathrm{b}(1,5)$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00433 |
| **III(9)** | $\geq$ | $\geq$ | $\geq$ | $\geq$ | $\leq$ | $\frac{1}{4}[\Delta_{1,2} + 3\Delta_{2,3}]$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00399 |
| **III(10)** | $\geq$ | $\geq$ | $\geq$ | $\geq$ | $\geq$ | $\frac{1}{4}[\Delta_{1,2} + 3\Delta_{2,3}]$ | $Y_3 - 3\widehat{\beta}^{RM}$ | 0.00209 |

**Table B.8**: Expressions for the RM coefficients for all sub-cases of case III for $n = 5$. The last column contains the corresponding probability of occurrence for data from a standard normal distribution.

### CASES IV – XII

For the cases IV to XII the expressions for the RM coefficients, the resulting order of the residual signs and the corresponding probabilities for standard normal data can be derived analogously. Table B.9 summarises the outcomes by listing all possible orders of residual signs in a sample of size $n = 5$ together with their probabilities for standard normal data. A comprised interpretation of the results is given in Section 2.2.

| $\text{sign}(r_1)$ | $\text{sign}(r_2)$ | $\text{sign}(r_3)$ | $\text{sign}(r_4)$ | $\text{sign}(r_5)$ | **probability** |
|---|---|---|---|---|---|
| 0 | 1 | 1 | -1 | 0 | 0.02711 |
| 0 | -1 | -1 | 1 | 0 | 0.02711 |
| 0 | 1 | -1 | 1 | 0 | 0.02910 |
| 0 | -1 | 1 | -1 | 0 | 0.02910 |
| 0 | 1 | -1 | -1 | 0 | 0.02711 |
| 0 | -1 | 1 | 1 | 0 | 0.02711 |
| 0 | 1 | 0 | -1 | 1 | 0.00714 |
| 0 | -1 | 0 | 1 | -1 | 0.00714 |
| 0 | 1 | 0 | 1 | -1 | 0.00664 |
| 0 | -1 | 0 | -1 | 1 | 0.00664 |
| 1 | 0 | -1 | 0 | 1 | 0.02219 |
| -1 | 0 | 1 | 0 | -1 | 0.02219 |
| 1 | -1 | 0 | 1 | 0 | 0.00714 |
| -1 | 1 | 0 | -1 | 0 | 0.00714 |
| 1 | -1 | 0 | -1 | 0 | 0.00665 |
| -1 | 1 | 0 | 1 | 0 | 0.00665 |
| 0 | 1 | -1 | 1 | -1 | 0.03661 |
| 0 | -1 | 1 | -1 | 1 | 0.03661 |
| 0 | 1 | -1 | -1 | 1 | 0.04879 |
| 0 | -1 | 1 | 1 | -1 | 0.04879 |
| 1 | 0 | -1 | -1 | 1 | 0.02391 |
| -1 | 0 | 1 | 1 | -1 | 0.02391 |
| 1 | -1 | 0 | 1 | -1 | 0.08296 |
| -1 | 1 | 0 | -1 | 1 | 0.08296 |
| 1 | -1 | 0 | -1 | 1 | 0.06552 |
| -1 | 1 | 0 | 1 | -1 | 0.06552 |
| 1 | -1 | -1 | 0 | 1 | 0.02391 |
| -1 | 1 | 1 | 0 | -1 | 0.02391 |
| 1 | -1 | 1 | -1 | 0 | 0.03661 |
| -1 | 1 | -1 | 1 | 0 | 0.03661 |
| 1 | -1 | -1 | 1 | 0 | 0.04865 |
| -1 | 1 | 1 | -1 | 0 | 0.04865 |

**Table B.9**: All possible orders of the signs of RM residuals for a sample of size $n = 5$ with the corresponding probability of occurrence for independent standard normal data.

# C The Univariate Adaptive Repeated Median Filter – Supplementary Graphics

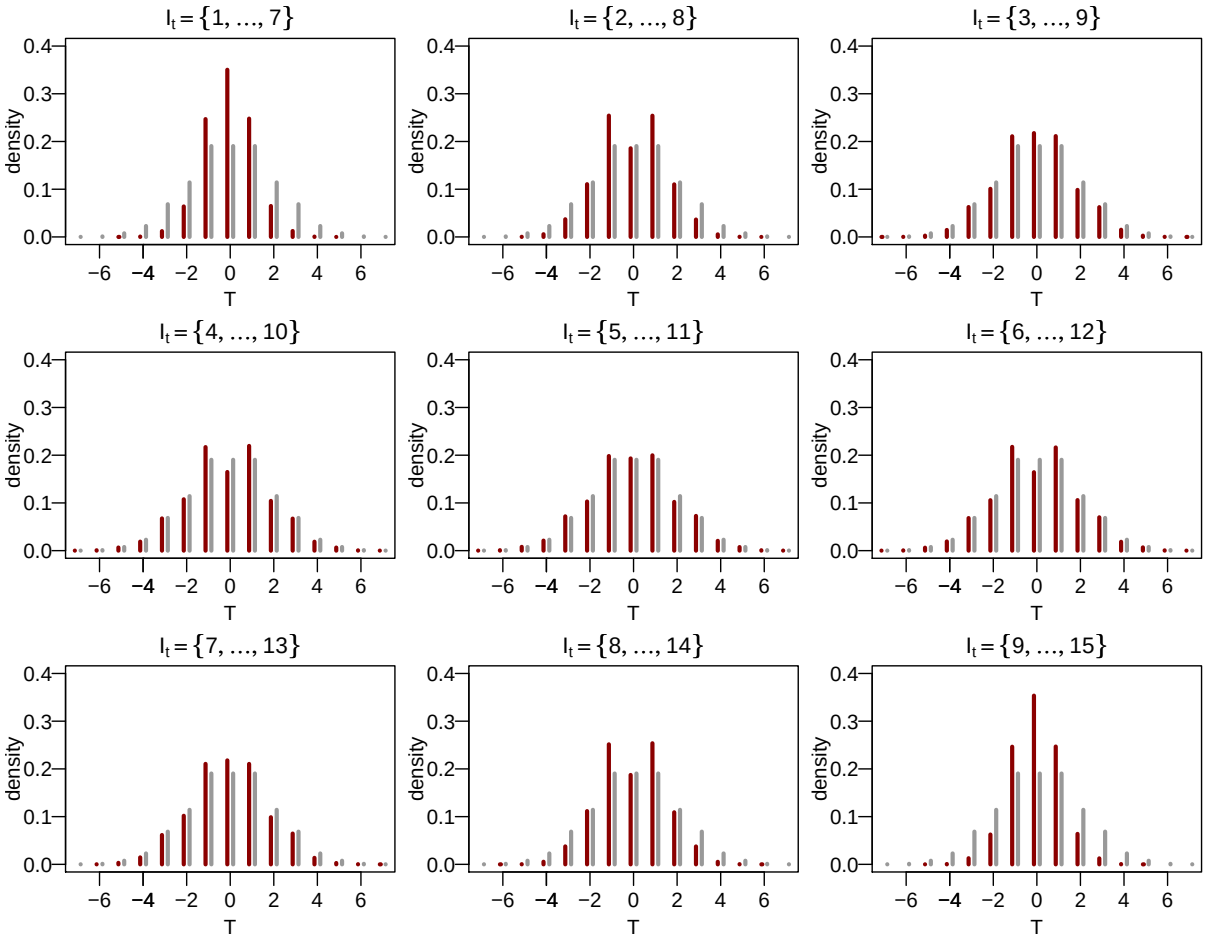


**Figure C.1**: Empirical density of the distribution of the test statistic $T$ for all possible subsets $I_t$ indicating a set of $n_{I_t} = 7$ consecutive out of $n_t = 15$ RM residuals (brown) for Cauchy data. For reference, the (large sample) approximation for this density (3.16) appears in the background (grey).

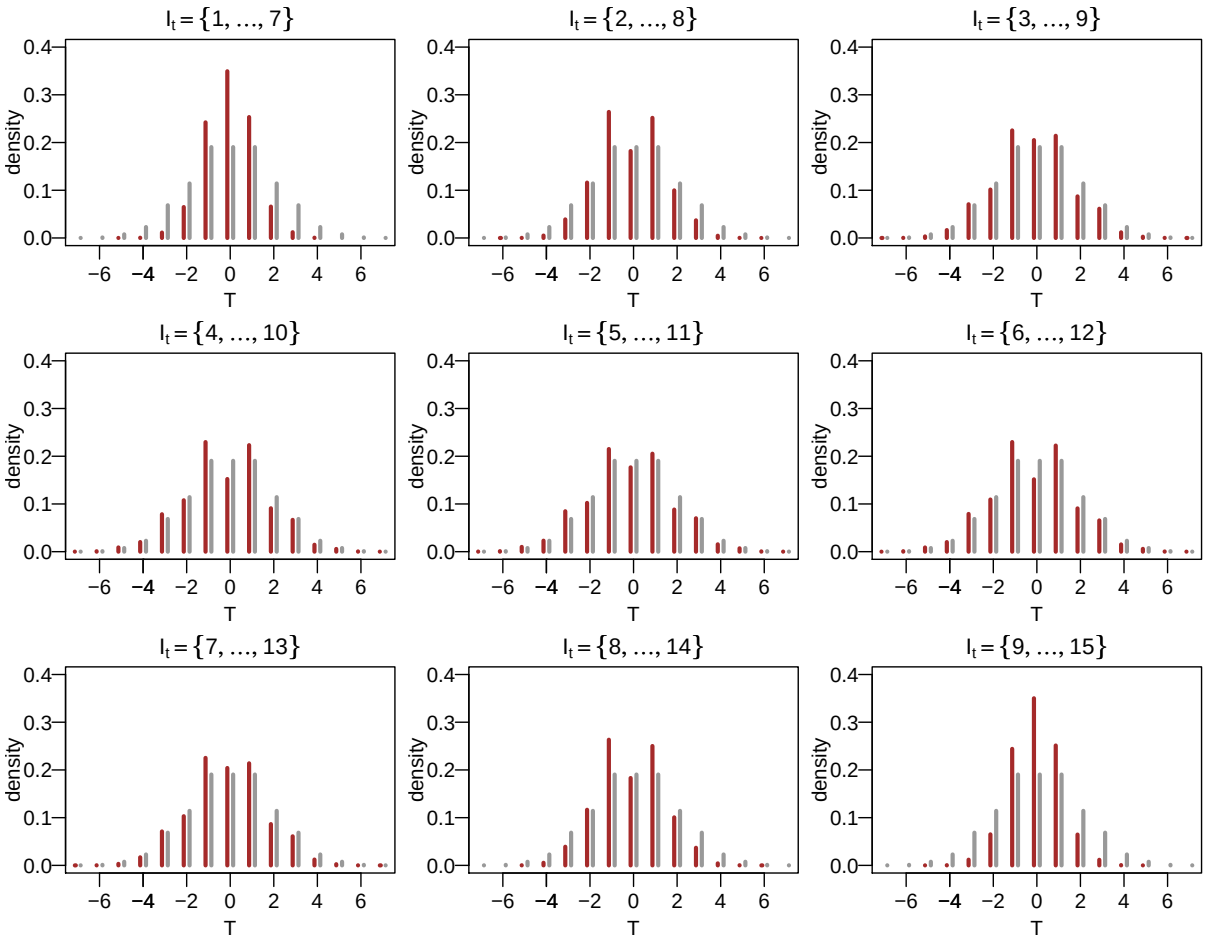


**Figure C.2**: Empirical density of the distribution of the test statistic $T$ for all possible subsets $I_t$ indicating a set of $n_{I_t} = 7$ consecutive out of $n_t = 15$ RM residuals (dark red) for data from a lognormal distribution with mean zero and unit variance. For reference, the (large sample) approximation for this density (3.16) appears in the background (grey).

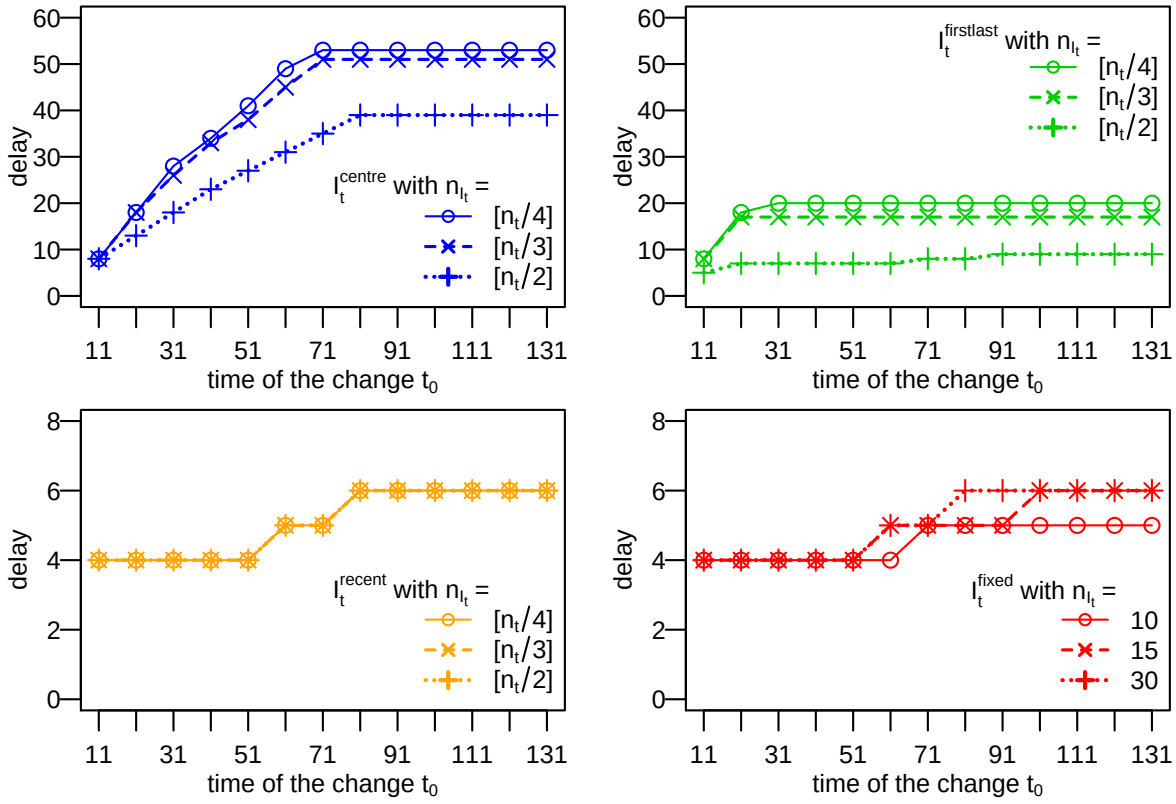


**Figure C.3**: Time delay in tracing a trend change at time $t_0 \in \{11, 21, \ldots, 131\}$ produced by adaptive RM filters with different settings.

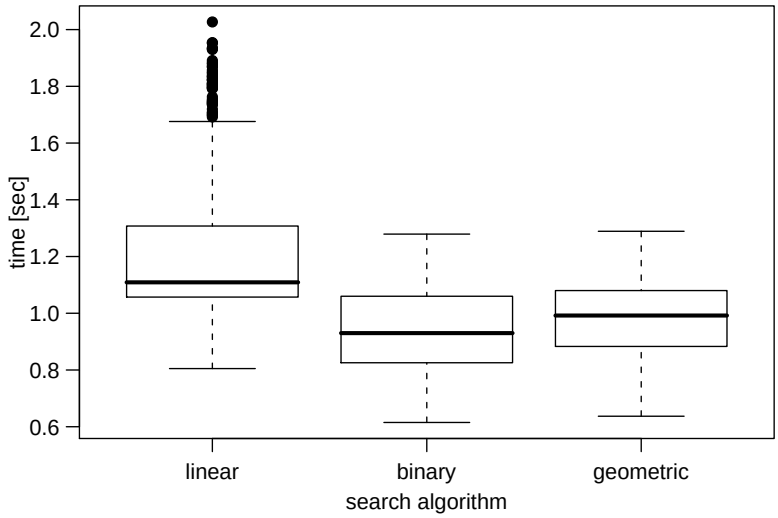


**Figure C.4**: Computation times in seconds required by adaptive online RM filters applying different search algorithms for the window width when analysing standard normal time series of length 100. The boxplots are based on 1000 simulation runs each.

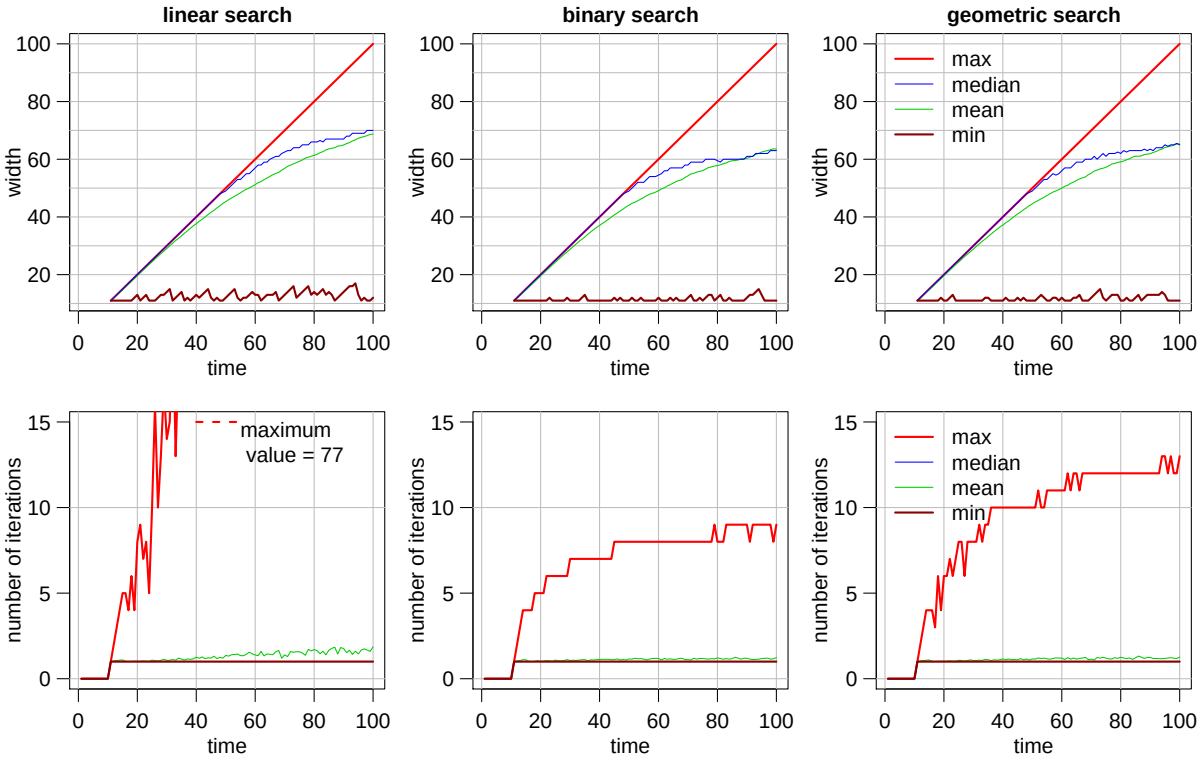


**Figure C.5**: Summary statistics of the window widths $n_t$ and the number of iterations $\ell_t$ required by the adaptive online RM filter at time $t$ when using different search algorithms for the width. These summary statistics are based on 1000 simulated standard normal time series of length 100.

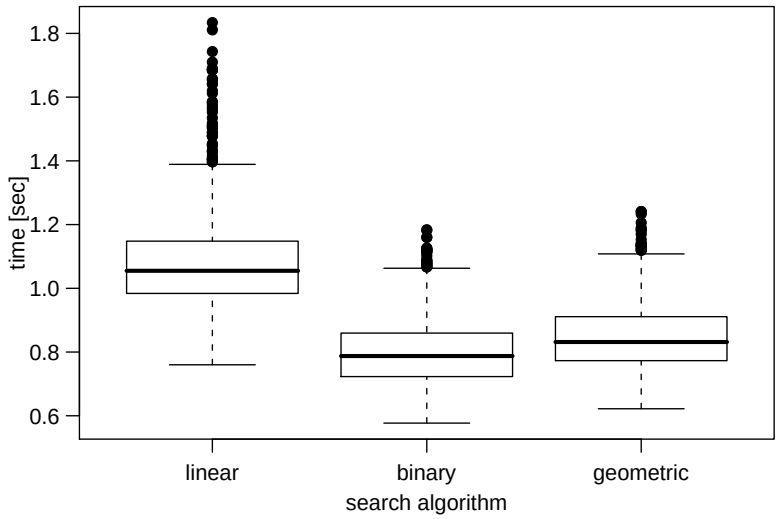


**Figure C.6**: Computation times in seconds required by adaptive online RM filters applying different search algorithms for the window width when analysing standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of one. The boxplots are based on 1000 simulation runs each.

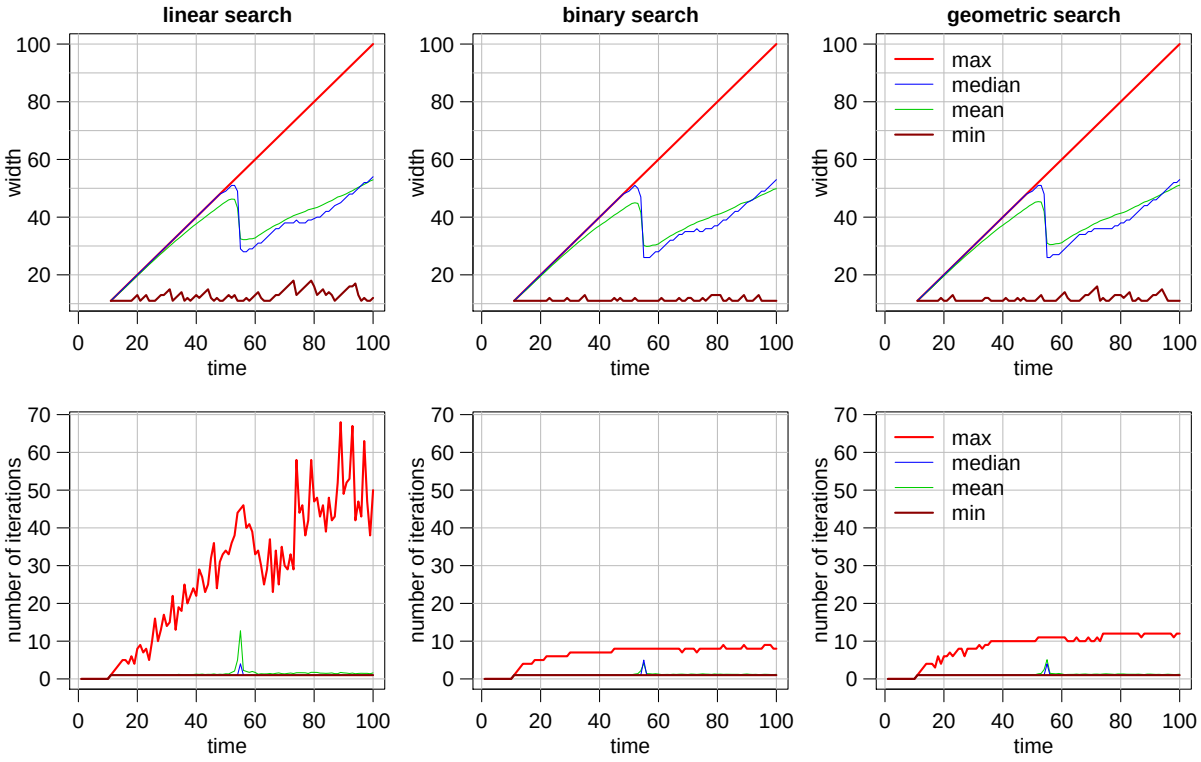


**Figure C.7**: Summary statistics of the window widths $n_t$ and the number of iterations $\ell_t$ required by the adaptive online RM filter at time $t$ when using different search algorithms for the width. These summary statistics are based on 1000 simulated standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of one.

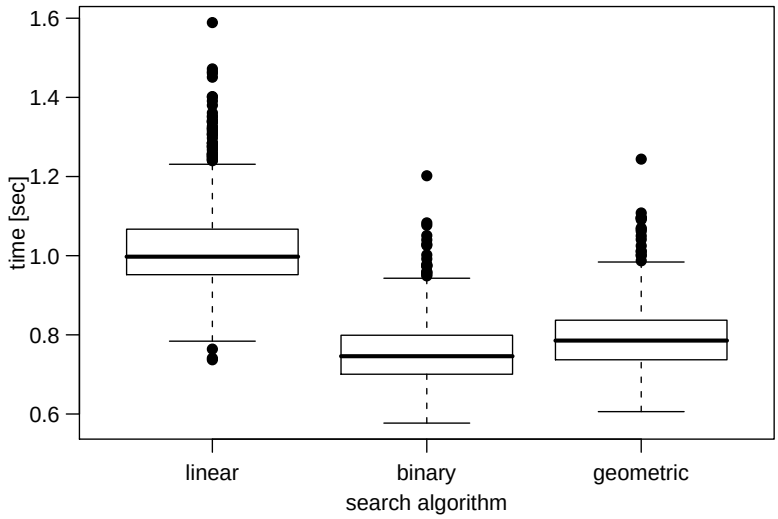


**Figure C.8**: Computation times in seconds required by adaptive online RM filters applying different search algorithms for the window width when analysing standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of two. The boxplots are based on 1000 simulation runs each.

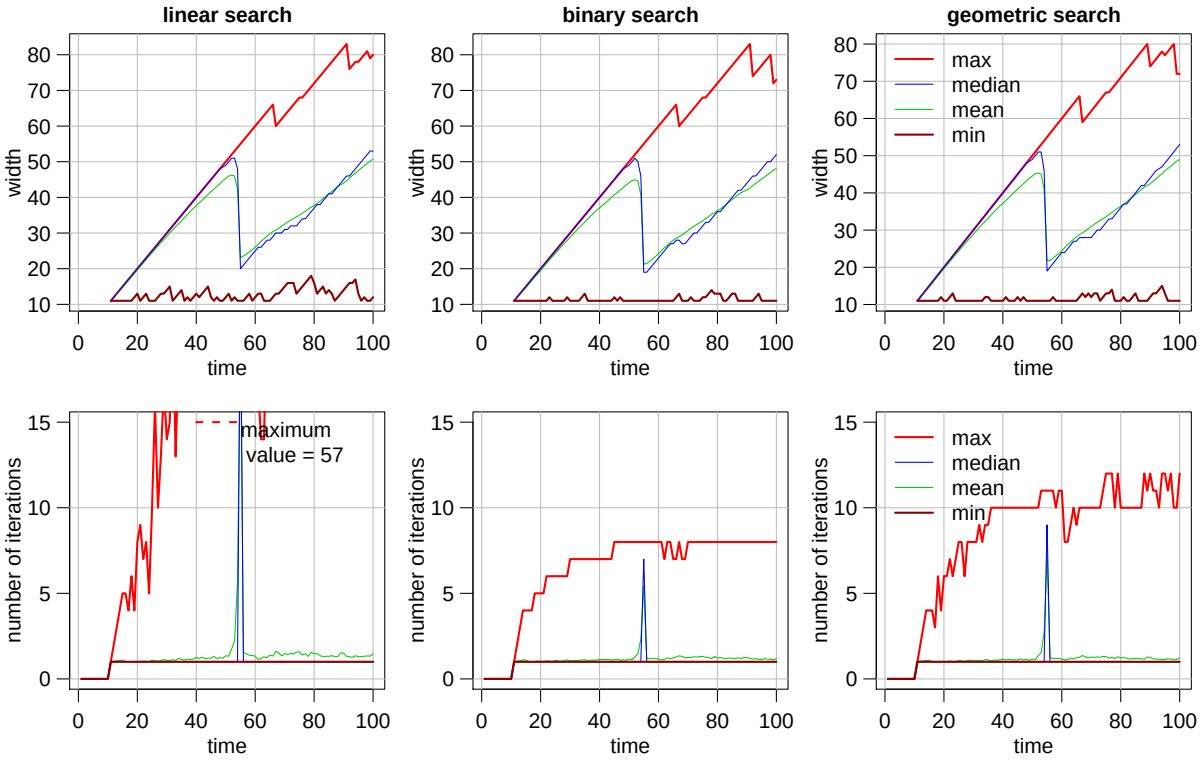


**Figure C.9**: Summary statistics of the window widths $n_t$ and the number of iterations $\ell_t$ required by the adaptive online RM filter at time $t$ when using different search algorithms for the width. These summary statistics are based on 1000 simulated standard normal time series of length 100 with a level shift at $t_0 = 51$ and a signal-to-noise ratio of two.

# D Breakdown Points of Robust Scale Estimators

## Finite Sample Breakdown Point of the $Q^{\alpha}_{all}$ Scale Estimator

### Explosion Breakdown Point of $Q^{\alpha}_{all}$

Let $\mathbf{y}_n^k$ denote a sample $\mathbf{y}_n$ with $k \in \{1, \ldots, n\}$ observations altered to arbitrary values such that only the $\binom{n-k}{3}$ heights as defined by (4.3) in Section 4.2 of triangles formed by three points from the original sample $\mathbf{y}_n$ are bounded. Rousseeuw and Hubert (1996) state that $Q^{\alpha}_{all}$ applied to the sample $\mathbf{y}_n^k$ explodes if and only if

$$\begin{aligned} \binom{n}{3} - \left\lfloor \alpha \binom{n}{3} \right\rfloor + 1 \quad &\leq \quad \binom{n}{3} - \binom{n-k}{3} \\ \Leftrightarrow \quad \binom{n-k}{3} \quad &< \quad \left\lfloor \alpha \binom{n}{3} \right\rfloor , \end{aligned} \tag{D.1}$$

i.e. $Q^{\alpha}_{all}$ explodes if the number of bounded heights is smaller than the order of the quantile $\left\lfloor \alpha \binom{n}{3} \right\rfloor$ defining the estimator.

To determine the exact finite sample explosion breakdown point of $Q^{\alpha}_{all}$ we have to find the smallest integer $k \in \{1, \ldots, n\}$ fulfilling (D.1), i.e.

$$\begin{aligned} \binom{n-k}{3} \quad &= \quad \left\lfloor \alpha \binom{n}{3} \right\rfloor - 1 \\ \Leftrightarrow \quad 0 \quad &= \quad \left\lfloor \alpha \binom{n}{3} \right\rfloor - 1 - \binom{n-k}{3} . \end{aligned} \tag{D.2}$$

For the ease of notation define

$$Q \;:=\; \left\lfloor \alpha \binom{n}{3} \right\rfloor . \tag{D.3}$$

The solution of equation (D.2) is given by $k^* := \lceil x^* \rceil$ where $x^* \in \mathbb{R}^+$ is a solution of the following transformation of this equation:

$$x^3 - (3n-3)x^2 + (3n^2 - 6n + 2)x - n^3 + 3n^2 - 2n + 6(Q-1) = 0 . \tag{D.4}$$

Substituting $y \;:=\; x - (n-1)$, equation (D.4) reads

$$y^3 - y + 6(Q-1) = 0 , \tag{D.5}$$

and a solution to this can be derived using Cardano's formula, see e.g. Lorenz and Lemmermeyer (2007), Chapter 15.

Setting $p := -1/3$ and $q := 3(Q-1)$ with $Q$ according to (D.3), equation (D.5) has exactly one real solution $y^*$ if the discriminant $D := q^2 + p^3 = 9(Q-1)^2 - 1/27$ is larger than zero. Since $D$ is always strictly positive for $Q \geq 2$, the real solution $y^*$ for all values $\alpha \geq 2/\binom{n}{3}$ is given by $y_1^*$, defined below. For a negative discriminant $D$, i.e. for $\alpha < 2/\binom{n}{3}$, there are three real solutions $y_1^*$, $y_2^*$ and $y_3^*$ to equation (D.5).
Define

$$u = \sqrt[3]{-q + \sqrt{D}} \quad \text{and} \quad v = \sqrt[3]{-q - \sqrt{D}}. \tag{D.6}$$

According to Cardano's formula the solutions for equation (D.5) are then given by

$$\begin{aligned} y_1^* &:= u + v \\ y_2^* &:= \frac{1}{2}(-1 + \sqrt{3}\mathrm{i}) \cdot u + \frac{1}{2}(-1 - \sqrt{3}\mathrm{i}) \cdot v = -\frac{1}{2}(u+v) + \frac{\sqrt{3}}{2}(u-v)\mathrm{i} \qquad \text{(D.7)} \\ y_3^* &:= \frac{1}{2}(-1 - \sqrt{3}\mathrm{i}) \cdot u + \frac{1}{2}(-1 + \sqrt{3}\mathrm{i}) \cdot v = -\frac{1}{2}(u+v) - \frac{\sqrt{3}}{2}(u-v)\mathrm{i}\,, \qquad \text{(D.8)} \end{aligned}$$

where $\mathrm{i} \in \mathbb{C}$ denotes the imaginary part of a complex number. Since we search for the smallest solution such that $x^*$ is positive, define

$$y^* := \min\Big\{\{y_1^*, y_2^*, y_3^*\} \cap \big(-(n-1), \infty\big)\Big\}.$$

The smallest number of replaced observations in the sample such that the number of bounded heights is smaller than $Q$ is then given by

$$k^* = \lceil x^* \rceil = \lceil y^* + (n-1) \rceil\,,$$

and the finite sample explosion breakdown point of $Q_{all}^{\alpha}$ at a sample with empirical distribution function $F_n$ corresponds to

$$\mathrm{fsbp}^+(Q_{all}^{\alpha}, F_n) = \frac{k^*}{n}\,. \tag{D.9}$$

Figure 4.3 in Section 4.2 provides a graphical representation of $\mathrm{fsbp}^+(Q_{all}^{\alpha}, F_n)$ for $\alpha \in [1/\binom{n}{3}, 1]$ and a comparison with its asymptotic counterpart $\mathrm{apb}^+(Q_{all}^{\alpha}) = 1 - \sqrt[3]{\alpha}$ for different sample sizes $n$.

### IMPLOSION BREAKDOWN POINT OF $Q_{all}^{\alpha}$

Here, let $\mathbf{y}_n^k$ denote a sample $\mathbf{y}_n$ (in general position) with $k \in \{1, \ldots, n\}$ observations replaced such that $k+1$ observations are identical. In that case, $\binom{k+1}{3} + \binom{k+1}{2}(n-k-1)$ of the triangle heights are zero. Rousseeuw and Hubert (1996) show that the estimator $Q_{all}^{\alpha}$ stays strictly positive when replacing $k$ observations in the described manner with

$$\binom{k+1}{3} + \binom{k+1}{2}(n-k-1) \quad \leq \quad \left\lfloor \alpha \binom{n}{3} \right\rfloor - 1\,.$$

Hence, we have to find the smallest value of $k \in \{1, \ldots, n\}$ such that

$$\binom{k+1}{3} + \binom{k+1}{2}(n-k-1) \quad \geq \quad \left\lfloor \alpha \binom{n}{3} \right\rfloor . \tag{D.10}$$

Denote the desired value by $k^*$. Setting $k^* := \lfloor x^* \rfloor$, we have to search for the smallest solution $x^* \in \mathbb{R}^+$ which solves the following equation, derived from a transformation of (D.10):

$$x^3 - \frac{3n-6}{2}x^2 - \frac{3n-4}{2} + 3Q \;=\; 0\,, \tag{D.11}$$

with $Q$ as defined in (D.3).

Substituting $y := x - \frac{n-2}{2}$, equation (D.11) reads

$$\begin{aligned} y^3 + 3py + 2q \;&=\; 0 \qquad\qquad (D.12)\\ \text{with} \quad p \;&=\; -\left(\frac{n-2}{2}\right)^2 - \frac{n-2}{2} - \frac{1}{3} \\ \text{and} \quad q \;&=\; -\left(\frac{n-2}{2}\right)^3 - \frac{3}{2}\left(\frac{n-2}{2}\right)^2 - \frac{1}{2}\left(\frac{n-2}{2}\right) + \frac{3}{2}Q\,. \end{aligned}$$

Analogous to the derivation of the explosion breakdown point, equation (D.12) can be solved using Cardano's formula, inserting $p$ and $q$ (as defined above) into equations (D.6) and (D.7).

Here, we have $D = q^2 + p^3 < 0$ for all values $n \in \mathbb{N}$ and $\alpha \in [1/\binom{n}{3}, 1]$, i.e. there always exist three real solutions to equation (D.12). However, since we seek a non-negative solution for (D.11), we choose

$$y^* \;:=\; \min\left\{ \{y_1^*, y_2^*, y_3^*\} \cap \left(-\frac{n-2}{2}, \infty\right) \right\}$$

and let

$$k^* = \lfloor x^* \rfloor = \left\lfloor y^* + \frac{n-2}{2} \right\rfloor .$$

The finite sample implosion breakdown point of $Q_{all}^{\alpha}$ at a sample of size $n$ with empirical distribution function $F_n$ is then given by

$$\text{fsbp}^-(Q_{all}^{\alpha}, F_n) \;=\; \frac{k^*}{n}\,. \tag{D.13}$$

Figure 4.3 in Section 4.2 provides a graphical representation of $\text{fsbp}^-(Q_{all}^{\alpha}, F_n)$ for $\alpha \in [1/\binom{n}{3}, 1]$ and a comparison with its asymptotic counterpart $\text{apb}^-(Q_{all}^{\alpha})$, defined in (4.7), for different sample sizes $n$.

For $\frac{1}{2} \leq \alpha < \frac{1}{2} + \frac{1}{\binom{n}{3}}$, we can also give a simple expression for the finite sample implosion breakdown point: if $\alpha$ lies in the specified interval, it is $q = 0$, leading to $y^* = 0$, and (D.13) corresponds to

$$\text{fsbp}^-(Q_{all}^{\alpha}, F_n) \;=\; \frac{\lfloor (n-2)/2 \rfloor}{n} \qquad \text{for} \quad \alpha \;\in\; \left[\frac{1}{2}, \; \frac{1}{2} + \frac{1}{\binom{n}{3}}\right) .$$

## MAXIMUM FINITE SAMPLE BREAKDOWN POINT OF THE MODEL-FREE SCALE ESTIMATORS

Let $S^{\alpha}_{adj}$ denote any of the model-free scale estimators $Q^{\alpha}_{adj}$ (4.11), $TM^{\alpha}_{adj}$ (4.12) or $TMS^{\alpha}_{adj}$ (4.13) based on adjacent triangle heights (4.10), described in Section 4.3. In the following, we determine the maximum value possible for the finite sample breakdown point $\text{fsbp}(S^{\alpha}_{adj}, F_n)$ at a sample $\mathbf{y}_n$ of size $n$ with empirical distribution function $F_n$. Furthermore, we derive the corresponding values of $\alpha$ for which the maximum breakdown point is achieved. Therefor consider the quantities

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

According to definition (A.2) and with the considerations from Section 4.3, the finite sample breakdown point corresponds to

$$\text{fsbp}(S^{\alpha}_{adj}, F_n) \;=\; \min\{A, B\}/n \,.$$

For increasing values of $\alpha \in [1/(n-2), 1]$, the quantity $A$ is decreasing while $B$ is increasing (see Figure 4.7 in Section 4.3). Hence, the maximum breakdown point is equal to $B/n$ for certain $\alpha \in [\alpha_{min}, \alpha_1)$, and it equals $A/n$ for certain $\alpha \in [\alpha_1, \alpha_{max})$. The maximum value for $\text{fsbp}(S^{\alpha}_{adj}, F_n)$ is reached for any value of $\alpha$ for which $A = B$, whenever this is possible. To determine the bounds $\alpha_{min}$ and $\alpha_{max}$, we distinguish between the four cases where either $n+1$, $n$, $n-1$ or $n+2$, respectively, is divisible by four.

### CASE I: $n \in \{4k-1, k \in \mathbb{N}\}$

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

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

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

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

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

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

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

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

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

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

### CASE II: $n \in \{4k, k \in \mathbb{N}\}$

Analogous to Case I we can show:

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

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

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

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

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

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

Here, we can show

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

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

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

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

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

**CASE IV:** $\boldsymbol{n \in \{4k+2, k \in \mathbb{N}\}}$

Here, it is

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

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

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

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

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

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

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

# List of Tables

## 2 Repeated Median Regression



## 3 Online Signal Extraction



## 4 Online Variability Extraction

**Appendix B: Zero Residuals in RM Regression**

# List of Figures

Simulation Study



**Appendix C: The Univariate Adaptive RM Filter**

# References


Arias-Castro, E., Donoho, D.L. (2007) Does Median Filtering Truly Preserve Edges Better Than Linear Filtering? *Technical Report* 2007-06, Department of Statistics, Stanford University.

Bernholt, T., 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., Wegener, I. (2006) Modified Repeated Median Filters. *Statistics and Computing* **16** (2), 177–192.

Bernholt, T., Nunkesser, R., Schettlinger, K. (2007) Computing the Least Quartile Difference Estimator in the Plane. *Computational Statistics and Data Analysis* **52** (2), 763–772.

Borowski, M., Schettlinger, K., Gather, U. (2009) Multivariate Real Time Signal Extraction by a Robust Adaptive Regression Filter. *Communications in Statistics - Simulation and Computation* **38** (2), 426–440.

Brockmann, M., Gasser, T., Herrmann, E. (1993) Locally Adaptive Bandwidth Choice for Kernel Regression Estimators. *Journal of the American Statistical Association* **88**, 1302–1309.

Chakraborty, X. (1999) On Multivariate Median Regression. *Bernoulli* **5**, 683–703.

Chan, S., Zhang, Z. (2004) Robust Local Polynomial Regression Using M-Estimator with Adaptive Bandwidth, in: *Proceedings of the 2004 International Symposium on Circuits and Systems (ISCAS 04), Circuits and Systems*, vol. 3, 333–336.

Charbonnier, S., Becq, G., Biot, L. (2004) On-Line Segmentation Algorithm for Continuously Monitored Data in Intensive Care Units. *IEEE Transaction on Biomedical Engineering* **51** (3), 484–492.

Chazelle, B., Guibas, L.J., Lee, D.T. (1985) The Power of Geometric Duality. *BIT* **25** (1), 76–90.

Cleveland, W.S. (1979) Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**, 829–836.

Cleveland, W.S., Loader, C.L. (1996) Smoothing by Local Regression: Principles and Methods, in: W. Härdle, M.G. Schimek (eds.), *Statistical Theory and Computational Aspects of Smoothing*, Springer, New York, 10–49.

Croux, C., Rousseeuw, P.J. (1992) Time-Efficient Algorithms for Two Highly Robust Estimators of Scale. *Computational Statistics* **1**, 411–428.

Croux, C., Rousseeuw, P.J., Hössjer, O. (1994) Generalized S-Estimators. *Journal of the American Statistical Association* **89** (428), 1271–1281.

Davies, P.L., Fried, R., Gather, U. (2004) Robust Signal Extraction for On-Line Monitoring Data. *Journal of Statistical Planning and Inference* **122**, 65–78.

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

Davies, P.L., Gather, U. (2006) Addendum to the Discussion of 'Breakdown and Groups'. *Annals of Statistics* **34** (3), 1577–1579.

Davies, P.L., Gather, U. (2007) The Breakdown Point – Examples and Counterexamples. *Revstat – Statistical Journal* **5**, 1–17.

Donoho, D.L., Huber, P.J. (1983) The Notion of Breakdown Point, in: P. Bickel, K. Doksum, J. Hodges, Jr. (eds.), *A Festschrift for Erich L. Lehmann*, Wadsworth, Belmont, California, 157–184.

Donoho, D.L., Johnstone, I.M. (1995) Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Association* **90** (432), 1200–1224.

Edelsbrunner, H., Souvaine, D.L. (1990) Computing Least Median of Squares Regression Lines and Guided Topological Sweep. *Journal of the American Statistical Association* **85** (409), 115–119.

Fan, J., Gijbels, I. (1995) Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation. *Journal of the Royal Statistical Society, Series B* **57** (2), 371–394.

Fan, J., Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

Fan, J., Hall, P., Martin, A., Patil, P. (1996) On Local Smoothing of Nonparametric Curve Estimators. *Journal of the American Statistical Association* **91**, 258–266.

Fried, R. (2004) Robust Filtering of Time Series with Trends. *Journal of Nonparametric Statistics* **16** (3), 313–328.

Fried, R. (2007a) On the Robust Detection of Edges in Time Series Filtering. *Computational Statistics and Data Analysis* **52** (2), 1063–1074.

Fried, R. (2007b) Robust Location Estimation under Dependence. *Journal of Statistical Computation and Simulation* **77** (2), 131 – 147.

Fried, R., Bernholt, T., Gather, U. (2006) Repeated Median and Hybrid Filters. *Computational Statistics and Data Analysis* **50** (9), 2313–2338.

Fried, R., Einbeck, J., Gather, U. (2007) Weighted Repeated Median Smoothing and Filtering. *Journal of the American Statistical Association* **102**, 1300–1308.

Fried, R., Gather, U. (2002) Fast and Robust Filtering of Time Series with Trends, in: W. Härdle, B. Rönz (eds.), *Proceedings in Computational Statistics (COMPSTAT 2002)*, Physica, Heidelberg, 367–372.

Fried, R., Gather, U. (2005) Robust Trend Estimation for AR(1) Disturbances. *Austrian Journal of Statistics* **34** (2), 139–151.

Fried, R., Schettlinger, K. (2008) robfilter: Robust Time Series Filters R Package Version 2.4. URL http://cran.r-project.org/web/packages/robfilter/

Friedman, J.H., Stützle, W. (1981) Projection Pursuit Regression. *Journal of the American Statistical Association* **76**, 817–823.

Gasser, T., Sroka, L., Jennen-Steinmetz, C. (1986) Residual Variance and Residual Pattern in Nonlinear Regression. *Biometrika* **73** (3), 625–633.

Gather, U., Fried, R. (2003) Robust Estimation of Scale for Local Linear Temporal Trends. *Tatra Mountains Mathematical Publications* **26**, 87–101.

Gather, U., Fried, R. (2004) Methods and Algorithms for Robust Filtering, in: J. Antoch (ed.), *Proceedings in Computational Statistics (COMPSTAT 2004)*, Physica, Heidelberg, 159–170.

Gather, U., Fried, R., Imhoff, M. (2000) Online Classification of States in Intensive Care, in: W. Gaul, O. Opitz, W. Schader (eds.), *Data Analysis. Scientific Modeling and Practical Application*, Springer, Berlin, 413–428.

Gather, U., Imhoff, M., Fried, R. (2002) Graphical Models for Multivariate Time Series from Intensive Care Monitoring. *Statistics in Medicine* **21** (18), 2685–2701.

Gather, U., Schettlinger, K. (2007) Robust and Adaptive Methods in Analysing Online Monitoring Data, in: *Proceedings of the 56th Session of the ISI*, Lisbon, Portugal, August 22-29, 2007.

Gather, U., Schettlinger, K., Fried, R. (2006) Online Signal Extraction by Robust Linear Regression. *Computational Statistics* **21** (1), 33–51.

Gelper, S., Schettlinger, K., Croux, C., Gather, U. (2008) Addendum to Robust Online Scale Estimation in Time Series: A Model-Free Approach. *Technical Report,* Faculty of Business and Economics, KU Leuven.
URL https://lirias.kuleuven.be/handle/123456789/201358

Gelper, S., Schettlinger, K., Croux, C., Gather, U. (2009) Robust Online Scale Estimation in Time Series: A Model-Free Approach. *Journal of Statistical Planning and Inference* **139**, 335 – 349.

Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E. (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101** (23), e215–e220.
URL http://circ.ahajournals.org/cgi/content/full/101/23/e215

Goldenshluger, A., Nemirovski, A. (1999) On Spatial Adaptive Estimation of Nonparametric Regression. *Mathematical Methods of Statistics* **6** (2), 135–170.

Grübel, R. (1988) The Length of the Shorth. *Annals of Statistics* **16** (2), 619–628.

Hampel, F.R. (1974) The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association* **69**, 383–393.

Hampel, F.R. (1975) Beyond Location Parameters: Robust Concepts and Methods. *Bulletin of the International Statistical Institute* **46**, 375–382.

Hastie, T., Tibshirani, R., Friedman, J. (2001) *The Elements of Statistical Learning*, 10 edn., Springer, New York.

Heinonen, P., Kalli, S., Turjanmaa, V., Neuvo, Y. (1985) Generalized Median Filters for Biological Signal Processing, in: *Proceedings of the 7th European Conference on Circuit Theory and Design*, 283–386.

Heinonen, P., Neuvo, Y. (1988) FIR-Median Hybrid Filters with Predictive FIR Substructures. *IEEE Transactions on Acoustics Speech and Signal Processing* **36** (6), 892–899.

Himayat, N., Kassam, S.A. (1993) Approximate Performance Analysis of Edge Preserving Filters. *IEEE Transactions on Signal Processing* **41** (9), 2764–2776.

Hössjer, O., Rousseeuw, P., Croux, C. (1994) Asymptotics of the Repeated Median Slope Estimator. *Annals of Statistics* **22** (3), 1478–1501.

Hössjer, O., Rousseeuw, P., Ruts, I. (1995) The Repeated Median Intercept Estimator: Influence Function and Asymptotic Normality. *Journal of Multivariate Analysis* **52** (1), 45–72.

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

Imhoff, M., Bauer, M., Gather, U., Fried, R. (2002) Pattern Detection in Intensive Care Monitoring Time Series with Autoregressive Models: Influence of the Model Order. *Biometrical Journal* **44** (6), 746–761.

Imhoff, M., Kuhls, S. (2006) Alarm Algorithms in Critical Care Monitoring. *Anesthesia and Analgesia* **102**, 1525–1537.

Katkovnik, V. (1999) A New Method for Varying Adaptive Bandwidth Selection. *IEEE Transactions on Signal Processing* **47** (9), 2567–2571.

Katkovnik, V., Egiazarian, K., Shmulevich, I. (2001) Adaptive Varying Window Size Filtering Based on Intersection of Confidence Intervals Rule, in: *Proceedings of the IEEE - EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP-01)*, June 3-6, Baltimore, Maryland, USA.

Kuhls, S. (2008) *Planung, Durchführung und Analyse einer klinischen Studie zur Bewertung und zum Vergleich von Alarmsystemen in der Intensivmedizin*, PhD Thesis, Fakultät Statistik, Technische Universität Dortmund.

Langerman, S., Steiger, W.L. (2003) The Complexity of Hyperplane Depth in the Plane. *Discrete and Computational Geometry* **30** (2), 299–309.

Lanius, V. (2005) *Statistische Extraktion relevanter Informationen aus multivariaten Online-Monitoring-Daten der Intensivmedizin*, PhD Thesis, Fachbereich Statistik, Universität Dortmund.

Lanius, V., Gather, U. (2007) Robust Online Signal Extraction from Multivariate Time Series. *Technical Report* 38/2007, SFB 475, Fakultät Statistik, Technische Universität Dortmund.

Lawless, S. (1994) Crying Wolf: False Alarms in a Pediatric Intensive Care Unit. *Critical Care Medicine* **22** (6), 981–985.

Lee, Y.H., Kassam, S.A. (1985) Generalized Median Filtering and Related Nonlinear Filtering Techniques. *IEEE Transactions on Acoustics Speech and Signal Processing* **33** (3), 672–683.

Lorenz, F., Lemmermeyer, F. (2007) *Algebra 1*, 4 edn., Spektrum Akademischer Verlag, Heidelberg.

Maronna, R.A., Martin, R.D., Yohai, V.J. (2006) *Robust Statistics, Theory and Methods*, Wiley.

Maronna, R.A., Zamar, R.H. (2002) Robust Estimates of Location and Dispersion for High-Dimensional Datasets. *Technometrics* **44** (4), 307–317.

Matoušek, J., Mount, D., Netanyahu, N. (1998) Efficient Randomized Algorithms for the Repeated Median Line Estimator. *Algorithmica* **20** (2), 136–150.

McDonald, J., Owen, A. (1986) Smoothing with Split Linear Fits. *Technometrics* **28** (3), 195–208.

Mount, D.M., Netanyahu, N.S., Romanik, K., Silverman, R., Wu, A.Y. (1997) A Practical Approximation Algorithm for the LMS Line Estimator, in: *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms*, 473–482.

Nunkesser, R., Fried, R., Schettlinger, K., Gather, U. (2008) Online Analysis of Time Series by the $Q_n$ Estimator. *Computational Statistics and Data Analysis* DOI:10.1016/j.csda.2008.02.027.

Nunkesser, R., Schettlinger, K., Fried, R. (2008) Applying the $Q_n$-Estimator Online, in: *Data Analysis, Machine Learning and Applications*, Springer, Berlin/Heidelberg, 277–284.

Rousseeuw, P.J. (1984) Least Median of Squares Regression. *Journal of the American Statistical Association* **79** (388), 871–880.

Rousseeuw, P.J. (1985) Multivariate Estimation with High Breakdown Point, in: W. Grossmann, G. Pflug, I. Vincze, W. Wertz (eds.), *Proceedings of the 4th Pannonian Symposium on Mathematical Statistics and Probability, Vol. B*, D. Reidel Publishing Company, Dordrecht.

Rousseeuw, P.J., Bassett, Jr., G.W. (1990) The Remedian: A Robust Averaging Method for Large Data Sets. *Journal of the American Statistical Association* **85**, 97–104.

Rousseeuw, P.J., Croux, C. (1993) Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association* **88**, 1273–1283.

Rousseeuw, P.J., Hubert, M. (1996) Regression-Free and Robust Estimation of Scale for Bivariate Data. *Computational Statistics and Data Analysis* **21** (1), 67–85.

Rousseeuw, P.J., Hubert, M. (1997) Recent Developments in PROGRESS, in: Y. Dodge (ed.), $L_1$-*Statistical Procedures and Related Topics*, *Lecture Notes-Monograph Series*, vol. 31, Institute of Mathematical Statistics, California, 201–214.

Rousseeuw, P.J., Hubert, M. (1999) Regression Depth. *Journal of the American Statistical Association* **94** (446), 388–402.

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

Rousseeuw, P.J., Leroy, A.M. (1988) A Robust Scale Estimator Based on the Shortest Half. *Statistica Neerlandica* **42**, 103–116.

Rousseeuw, P.J., Netanyahu, N.S., Mount, D.M. (1993) New Statistical and Computational Results on the Repeated Median Line, in: S. Morgenthaler, E. Ronchetti, W. Stahel (eds.), *New Directions in Data Analysis and Robustness*, Birkhäuser, Basel, 177–194.

Rousseeuw, P.J., Van Aelst, S., Hubert, M. (1999) Rejoinder to 'Regression Depth'. *Journal of the American Statistical Association* **94** (446), 419–433.

Rousseeuw, P.J., Van Driessen, K. (2002) Computing LTS Regression for Large Data Sets. *Estadística* **54**, 163–190.

Ruppert, D. (1997) Empirical-Bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation. *Journal of the American Statistical Association* **92**, 1049–1062.

Schettlinger, K. (2004) *Robust Methods for Signal Extraction from Time Series*, Diploma Thesis, Fachbereich Statistik, Universität Dortmund.

Schettlinger, K., Fried, R., Gather, U. (2006) Robust Filters for Intensive Care Monitoring: Beyond the Running Median. *Biomedical Engineering* **51** (2), 49–56.

Schettlinger, K., Fried, R., Gather, U. (2008) Real-Time Signal Processing by Adaptive Repeated Median Filters. *International Journal of Adaptive Control and Signal Processing* To appear.

Schettlinger, K., Gelper, S., Gather, U., Croux, C. (2009) Regression-Based, Rregression-Free and Model-Free Approaches for Robust Online Scale Estimation. *Journal of Statistical Computation and Simulation* Submitted.

Schettlinger, K., Imhoff, M. (2005) Signal Extraction from Time Series in Intensive Care Monitoring. *Biomedical Engineering* **50** (1), 489–490.

Seely, A.J.E., Macklem, P.T. (2004) Complex Systems and the Technology of Variability Analysis. *Critical Care* **8**, 367–384.

Siegel, A.F. (1982) Robust Regression Using Repeated Medians. *Biometrika* **69** (1), 242–244.

Spencer, J. (1904) On the Graduation of the Rates of Sickness and Mortality. *Journal of the Institute of Actuaries* **38**, 334–347.

Tukey, J.W. (1977) *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

Woolhouse, W.S.B. (1870) Explanation of a New Method of Adjusting Mortality Tables With Some Observations upon Mr. Makeham's Modification of Gompertz's Theory. *Journal of the Institute of Actuaries* **15**, 389–410.

Zuo, Y. (2001) Some Quantitative Relationships Between Two Types of Finite Sample Breakdown Point. *Statistics and Probability Letters* **51** (4), 369–375.