technische universität
dortmund

Robust and non-parametric control charts for time series with a time-varying trend

Dissertation

in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften

presented to the

Faculty of Statistics
TU Dortmund University

by

Sermad Abbas

Submitted:	Dortmund, March 2019
Advisor and primary referee:	Prof. Dr. Roland Fried
Secondary referee:	Prof. Dr. Christine Müller
Chairman of the commission:	JProf. Dr. Andreas Groll
Assessor:	Dr. Leo Geppert
Day of the oral examination:	19th June 2019

Yoda: “Always with you what cannot be done. Hear you nothing that I say?”

Luke: “Master, moving stones around is one thing, but this is... totally different!”

Yoda: “No! No different! Only different in your mind. You must unlearn what you have learned.”

Luke: “All right, I’ll give it a try.”

Yoda: “No! Try not. Do... or do not. There is no try.”

The Empire Strikes Back

“Kept you waiting, huh?”

Solid Snake in Metal Gear Solid 2: Sons of Liberty

Contents

1 Introduction **1**

2 Methods for real-time filtering and pattern detection **5**

2.1 Introduction 5
2.2 Data sets 6
2.2.1 Online monitoring of vital parameters in intensive care units 7
2.2.2 Detection of nano-size objects 8
2.2.3 Monitoring of crack widths in concrete bridges 10
2.3 Robust time-series filtering 11
2.3.1 Location-based filtering 12
2.3.2 Regression-based filtering 15
2.4 Online detection of shifts and volatility changes 21
2.4.1 Detection of location shifts 21
2.4.2 Detection of volatility changes 25
2.5 Discussion 26

3 Robust control charts for the mean based on repeated two-sample location tests **27**

3.1 Introduction 27
3.2 Model 29
3.3 Methods 30
3.3.1 Measuring the performance of a control chart 30
3.3.2 Selected two-sample tests for the location problem 33
3.4 Simulations 39
3.4.1 In-control comparison under the $\mathcal{N}(0,1)$-distribution 40
3.4.2 ARL_0 under non-normal distributions 47
3.4.3 Out-of-control analysis 51
3.4.4 Influence of outliers in the out-of-control setting 58
3.5 Applications 59
3.5.1 Setting up the control charts for applications 60
3.5.2 PAMONO data 62
3.5.3 Time series of heart-rate measurements 63
3.6 Discussion 64

4 Robust control charts for the mean of locally linear time series **67**

4.1 Introduction 67

4.2 Model . . . 70
4.3 Methods . . . 71
4.3.1 Removing trends by local regression . . . 71
4.3.2 Comparison of repeated median and ordinary least squares . . . 72
4.3.3 Improving the computation time of the repeated median . . . 75
4.3.4 Modification of the simplified randomisation . . . 75
4.3.5 Using one-sample tests on the forecast errors . . . 77
4.4 Simulations . . . 79
4.4.1 In-control comparison under the $\mathcal{N}(0,1)$-distribution . . . 79
4.4.2 ARL_0 under non-normal distributions . . . 83
4.4.3 Out-of-control analysis . . . 86
4.4.4 Local AR(1)-models . . . 92
4.4.5 In-control performance for AR(1)-models . . . 94
4.4.6 Out-of-control performance for AR(1)-models . . . 95
4.5 Applications . . . 98
4.5.1 PAMONO data . . . 98
4.5.2 Time series with sine-wave signal . . . 100
4.5.3 Time series of heart-rate measurements . . . 101
4.5.4 Time series of crack-width measurements . . . 102
4.6 Summary . . . 105

5 Summary and Outlook **107**
5.1 Summary . . . 107
5.2 Outlook . . . 108

A Proofs **111**

B Figures **115**

C Tables **129**

D Supplementary Definitions **143**
D.1 Asymptotic relative efficiency . . . 143
D.2 Breakdown point . . . 143
D.3 Pitman asymptotic relative efficiency . . . 144
D.4 Selected equivariance and invariance properties . . . 144

List of Figures **147**

List of Tables **149**

References **151**

1 Introduction

Change is an inevitable part of many dynamic systems. Even though it is often caused by the natural evolution, a change can also point at a serious malfunction of the system. This thesis is devoted to the detection of sudden changes in univariate time series which exhibit a slowly time-varying trend. The signal is assumed to be covered by additive noise and occasional unusually large or small values, called outliers, may be present. Many procedures for the detection of change points have been proposed throughout the years. An extensive discussion on this topic can be found in the book "Detection of Abrupt Changes: Theory and Application" by Basseville and Nikiforov (1993).

In this work, we focus on the detection of change points in online-monitoring applications. Our goal is to develop methods that are able to identify structural breaks in the process with a short delay. At the same time, we restrict the number of false alarms by controlling the run length, which we define as the duration between two alarms. Common methods of choice under these premises are *control charts*, which have their origin in *Statistical Process Control*. A control chart is a tool to decide whether a process follows its natural behaviour. Based on a sample of consecutive observations, a control statistic is computed and compared to predefined control limits.

In this thesis, we study two-sample location tests in a moving time window to identify sudden changes in the mean function of a time series with a slowly varying signal. The time window consists of the most recent observations in the process and is split into two subwindows to test for a location shift between them. The main contributions of this thesis are extensive comparisons between control charts based on different test statistics under consideration of the run length.

The studied principle has some advantages over ordinary control charts. As only the newest observations in the process are used to check for a change point, no large historical data sets have to be acquired to estimate the process parameters. This relaxes the assumptions on the signal and allows for slow natural changes in the parameters. As the past of the process is disregarded by the new control charts, they adapt to the local signal course. This reduces the risk of falsely identifying natural changes as relevant ones and thus helps to keep the false alarm frequency small. Moreover, using appropriate test statistics makes the charts robust against a predefined number of consecutive outliers, avoiding confusion between outlier patches and level shifts.

In what follows, we provide an overview of the main themes and results of the Chapters 2 - 4 in this thesis. A global summary and some ideas for possible future research can be found in Chapter 5.

Methods for real-time filtering and pattern detection

Chapter 2 introduces the basic terms and problems of online monitoring, which are relevant to this thesis. It is based on the publication "Online analysis of medical time series" by Fried et al. (2017). The article has been modified and rephrased to fit more into the context of this thesis. Information on the changes can be found at the beginning of Chapter 2.

The chapter introduces three time series from real-world applications, which will appear throughout this thesis as illustrative examples for the developed control charts to demonstrate their versatility and wide applicability.

The first time series consists of heart-rate measurements obtained from an online-monitoring of a patient in an intensive care unit. The underlying signal of the time series in such an application is known to vary over time. It is important to have outlier-resistant procedures that are able to distinguish between unavoidable variation and sudden, clinically relevant changes that may threaten the patient's health.

The second time series represents greyscale values. They stem from the biosensor technology *Plasmon-Assisted Microscopy of Nano-Size Objects* (PAMONO) (Siedhoff et al., 2014b). The signal is piecewise constant, making the time series a good testing ground for our procedures to illustrate their detection qualities in a rather undisturbed setting.

As a third example, we consider measurements of crack widths from a bridge monitoring. These data are used as an example to show how the control charts, originally designed for detecting changes in the mean, can be modified to find structural breaks in the variance.

Parts of Chapter 2 are based on a review article and it mainly serves the purpose to summarise prior research. It focuses on two fields which are combined in the later chapters of this thesis. We first give an overview of robust local filtering procedures. They are separated into the groups of filters based on location estimators and filters based on regression. Their goal is to approximate the signal of a time series so that they preserve the relevant signal characteristics without being influenced by outliers. Regression estimators are especially suited for signal extraction in trend-affected time series. From the summarised literature, the repeated median (Siegel, 1982) turns out to be convincing from the viewpoint of efficiency and robustness against outliers.

In addition, we discuss deficiencies of ordinary control schemes for the applications we have in mind and motivate the use of the control charts studied in the following chapters.

Robust control charts for the mean based on repeated two-sample location tests

Chapter 3 is based on the article "Control charts for the mean based on robust two-sample tests" by Abbas and Fried (2017). Some extensions have been made compared to the original article. They are described at the beginning of the chapter.

We introduce control charts based on sequentially applied two-sample tests to identify sudden changes in an otherwise locally constant signal.

In extensive simulation studies, we compare several two-sample tests with respect to their performance in the absence of relevant changes, that is when the process operates

in control, and under sudden location shifts, so-called out-of-control settings. In the first case, we aim for charts which ensure the same, or at least a similar, run-length behaviour over a wide range of noise distributions. This is a desirable property because the true distribution is often unknown in practice. Moreover, the charts should detect structural breaks fast and with a high probability, so that we study the detection speed and rate, again over a broad spectrum of distributions.

We show that control charts that are based on rank tests have a distribution-free in-control run-length distribution. Furthermore, we conclude that a control chart based on the two-sample Hodges-Lehmann estimator for shift (Hodges and Lehmann, 1963) provides a good compromise regarding the requested properties. For an in-control process, it is nearly distribution free. Moreover, it is robust against outliers and fast in detecting location shifts.

Robust control charts for the mean of a locally linear time series

Although the moving time window is able to deal with slow trends in the signal, strong trends can easily be confused with location shifts. In Chapter 4, we modify the control-chart principle from Chapter 3 in order to find abrupt level and trend changes in time series where the assumption of a locally constant signal is violated.

Assuming a locally linear signal, we construct residual control charts which apply the tests to one-step-ahead forecast errors of a fitted local regression model. Residual charts have been studied extensively in the control-chart literature; see for example Knoth and Schmid (2004). However, our herein developed charts have again the advantage that they do not require much knowledge on the global process behaviour.

We study two cases. In the first one, we use the repeated median to compute the forecast errors, assuming a simple linear model within a moving time window. If this assumption is exactly fulfilled, the forecast errors vary regularly around a constant. However, the one-step-ahead forecast errors are correlated. This causes the control charts which are based on rank tests to lose the exact distribution independence of the in-control run length.

Simulation studies indicate that the main properties of the control charts, which are observed in Chapter 3, also apply in this setting. Again, the two-sample Hodges-Lehmann estimator provides a good basis for a control chart by retaining its detection and robustness properties from the locally constant setting. Especially when the forecast errors are computed from a large window, it performs similarly well as for the locally constant scenario. Moreover, the two-sample testing approach is able to deal with model misspecification. Even if the local linearity assumption is not exactly fulfilled, the charts can deal with the resulting slow trends in the series of forecast errors.

In the second scenario, we briefly consider a local autoregression of order one. Here, all charts lose their distribution independence. For small to moderate autocorrelations, we can still expect reasonably good detection properties.

Computations and Software

The programming part of this thesis is performed with the statistical software package `R` (R Core Team, 2017), version 3.4.1. Most simulations are conducted on the Linux HPC cluster LiDOng at TU Dortmund University. Each node consists of a 3.00 GHz Intel Xeon E5450 machine with 15 GB RAM. We use the `R` package `batchtools` (Lang et al., 2017) to carry out the simulations in parallel on the cluster. For simulations on a local machine, we use `future.apply` (Bengtsson, 2018). The graphics are created with the packages `ggplot2` (Wickham, 2009), `gridExtra` (Auguie, 2016), `scales` (Wickham, 2017), and `tikzDevice` (Sharpsteen and Bracken, 2018). Many tables are generated with the help of `xtable` (Dahl, 2016). Data manipulation is achieved with the packages `data.table` (Dowle and Srinivasan, 2017), `plyr` (Wickham, 2011), `dplyr` (Wickham et al., 2017), and `tidyr` (Wickham and Henry, 2017).

Acknowledgements

Parts of this work were supported by the Collaborative Research Centre "Providing Information by Resource-Constrained Data Analysis" (SFB 876) and the Collaborative Research Centre "Statistical Modelling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG), which is gratefully acknowledged.

Thanks also go to the following recipients for providing and pre-processing the real-world data used in this thesis:

- Project C4 of the Collaborative Research Centre "The Reduction of Complexity in Multivariate Data Structures" (SFB 475) for providing the time series of heart-rate measurements.
- Project B2 of the SFB 876 for providing the PAMONO data.
- Reinhard Maurer, Jens Heinrich, and Melanie Horn of project B5 of the SFB 823 as well as KHP Dortmund and the Civil Engineering Department Bochum for collecting, processing, and providing the crack-width measurements.

The provision of the HPC cluster LiDOng by the ITMC of TU Dortmund University is also gratefully acknowledged.

2 Methods for real-time filtering and pattern detection

This chapter is based on the article "Online analysis of medical time series", published in the *Annual Review of Statistics and Its Application* by Fried et al. (2017). The article is joint work with Roland Fried (TU Dortmund University), Matthias Borowski (University of Münster), and Michael Imhoff (Ruhr-University Bochum).

To better fit into the context of this thesis, the article has been revised. The emphasis is put on procedures for online signal extraction and shift detection in univariate time series.

Major changes are the following:

- The sections on multivariate techniques, clinical applications, outlier and trend detection, and online segmentation procedures have been removed or shortened.
- The original article deals with medical applications only. Here, the procedures are looked at from a more general point of view.
- Three time series from real-world applications are introduced. These are not part of the original article.
- A brief introduction to control charts is added to motivate the development of alternatives in the Chapters 3 and 4. This is not part of the original article.
- This chapter serves as an introduction to the basic problems considered in this thesis. This required to rewrite Introduction and Conclusion for a better transition to the following chapters.

2.1 Introduction

The analysis of quantities that are measured continuously over time is a frequent task in a multitude of applications like health care, biology, or engineering. A frequent objective is to detect unusual patterns like monotonic trends or sudden changes of the level, the trend, or the variability. These can indicate severe disturbances in the monitored characteristic. For example, when monitoring vital parameters of a patient in an intensive care unit, a sudden increase or decrease of the measurements can point at critical changes in the well-being of the patient.

We consider monitoring tasks in which the analysis of new incoming observations is executed automatically and in real-time. Relevant changes have to be detected as soon as possible, while false alarms should only occur rarely. In general, an alarm indicates that some kind of intervention is required to correct the process. Triggering false alarms too

often can make the responsible staff insensitive to alarms. This implies the danger that important situations are ignored (Borowski et al., 2011). Moreover, a closer inspection can be costly, for example when a production machine has to be stopped for a certain amount of time (Basseville and Nikiforov, 1993, p. 6ff.),

Several aspects can complicate the real-time analysis. Due to technical reasons, the measurements are often noisy, so that essential information on the quantity of interest can be hidden. Furthermore, the data are often collected with a high frequency. In addition, measurement artefacts can lead to outliers. Such observations with an unusually large absolute value can corrupt the analysis and lead to false decisions by mimicking or masking relevant patterns. To deal with these difficulties, procedures are required that can extract the important information of a time series, are computationally efficient, separate relevant from irrelevant structures, and identify unusual situations with only a short time delay.

This chapter summarises research that contributes to the outlined problems. In Section 2.2, data sets from three different real-world applications are introduced. They serve as examples throughout this thesis and contain several of the structures described before. In Section 2.3 we give an overview of robust filtering procedures. They aim at approximating the true signal underlying a noisy and outlier-contaminated time series. Section 2.4 serves as a motivation for the shift-detection methods discussed in the Chapters 3 and 4. We outline some existing techniques and describe their drawbacks with respect to our intended applications. The procedures aim at giving real-time information on sudden shift or variability changes. In Section 2.5, we give some remarks on how the methods described in this chapter are connected to the procedures studied in the other chapters of this thesis.

2.2 Data sets

The data sets described in the following serve as application examples for the methods discussed in the Chapters 3 and 4. In this chapter, we use them merely as illustrative and motivational examples.

The research questions behind these data are very different but can be boiled down to the detection of sudden changes in specific process parameters. However, our goal is not to solve the underlying problems, since solutions of an acceptable quality often require several steps which are adjusted to the data structures to be analysed. Development and discussion of such procedures are beyond the scope of this thesis. Instead, we motivate how the herein discussed procedures can be used for a single step on which further analysis can be build up.

In Subsection 2.2.1, we describe a data set from a monitoring application in an intensive care unit where the measurements are affected by a time-varying trend and outliers. We discuss a time series with a piecewise-constant signal in Subsection 2.2.2. It stems from a biosensor technology which can be used to detect very small particles in sample fluids. The goal on both data sets is to detect abrupt location changes. The measurements shown

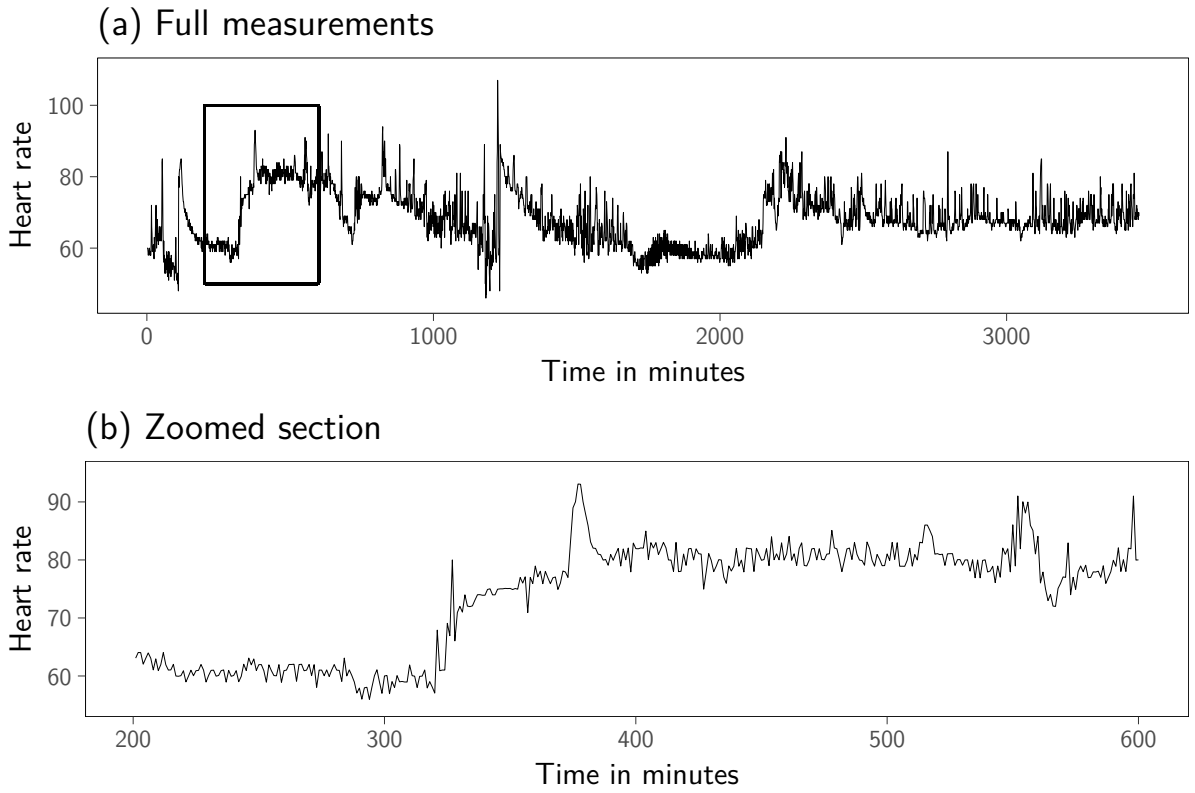

Figure 2.1: Heart-rate measurements of a patient in an intensive care unit from an online monitoring. Panel (a) shows the complete time series, panel (b) the time series restricted to the box in panel (a).

in Subsection 2.2.3 originate from a monitoring of crack widths in a concrete bridge where the signal of the time series is non-linear. Here, the objective is to find variability changes.

2.2.1 Online monitoring of vital parameters in intensive care units

In intensive care units, online systems assess the physical state of a critically ill patient by continuously measuring vital signs like heart rate, blood pressure, oxygen saturation, or respiration rate with a high sampling frequency. The resulting time series are often very noisy and contaminated by outliers caused by, for example, patient movements, manipulation, nursing, or therapeutic interventions (Imhoff et al., 2009).

A fundamental task in the online analysis of such time series is the extraction of the underlying time-varying level of the measurements. It contains important information on the patient's health. Clinically irrelevant short-term fluctuations have to be eliminated and outliers need to be resisted while preserving clinically relevant information, such as changing trends or sudden shifts (Schettlinger et al., 2006).

The use of threshold alarms is a common approach for the quick detection of such relevant events. The observations are compared to critical values which are manually specified by health-care professionals. However, in several studies, it was found that about 90% of the detected alarms are irrelevant and instead related to artefacts or minor

fluctuations around the threshold (Imhoff et al., 2009).

Figure 2.1(a) shows minutely observed heart-rate measurements over a duration of 3462 minutes. Many of the aforementioned structures can be found in the time series. The signal varies over time and is affected by several sharp location or trend changes. The measurements are corrupted by outliers. A global parametric modelling of the signal course is, in general, not possible. Thus, the data are typically analysed under local assumptions; see for example Imhoff et al. (2002) and Schettlinger et al. (2006).

In the remainder of this thesis, we focus on the enlarged section of length 400 shown in Figure 2.1(b). The restriction to a shorter time range improves the presentation of the results later in this thesis. A small location shift can be seen around the time point $t = 290$. After $t = 320$, a large, persistent level shift occurs, followed by a large peak near $t = 375$. Behind the peak, the level seems to be slightly larger than before. The signal bends at roughly $t = 430$. A smaller peak can be seen at $t = 515$ and a larger one at $t = 550$. Here, the level falls below the one prior to the peak. At the end of the time range, the signal increases slowly.

Figure 2.1 demonstrates the difficulty of finding an adequate global parametric model for the course of the time series. Moreover, as the future development is unknown, global models seem to be inappropriate due to their lack of flexibility.

Local procedures use the most recent observations to analyse the data. This is especially helpful for the detection of relevant structural breaks, since the procedures are less prone to confuse the natural long-term fluctuations of the signal with abrupt changes. This idea is the basis for the shift-detection procedures studied in the Chapters 3 and 4. Most of the procedures presented in this chapter were studied for the online monitoring of data from intensive care units.

2.2.2 Detection of nano-size objects

The *Plasmon-Assisted Microscopy of Nano-Size Objects* (PAMONO) is a biosensor technology for the indirect identification of particles with a size on the nano-scale in a sample fluid (Zybin et al., 2010). A detailed description of the construction and functioning of the sensor can be found in Weichert et al. (2010) or Siedhoff et al. (2011). These references also describe an analysis pipeline for this application.

The sensor generates a sequence of greyscale images which show reflections of a laser beam that is pointed on the sensor. If a particle adheres to the surface, a persistent bright spot is visible in the image sequence, caused by increased greyscale values. Such spots affect several adjacent pixel coordinates in the images. Due to physical effects, they are typically surrounded by a dark circle.

The research goal is to automatically analyse, preferably in real time, these data and get an indication on the particle concentration in the sample. One possibility is to extract the greyscale values for each pixel coordinate and apply procedures for step detection to

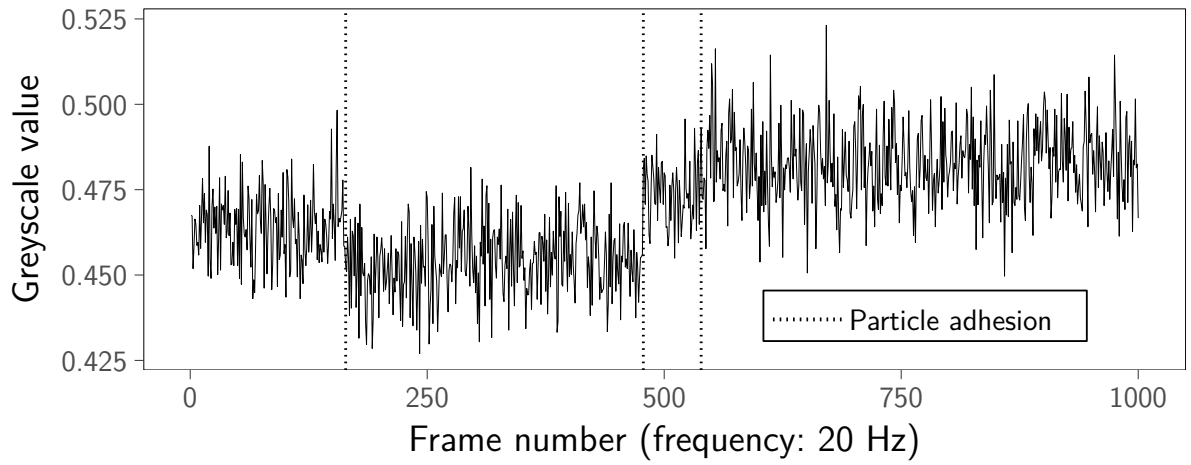

Figure 2.2: Pixel time series of greyscale values obtained by the PAMONO biosensor. Sudden location shifts can be attributed to particles adhering to the sensor surface.

the resulting time series. This is because the bright spots cause sudden location shifts in the time series of the affected coordinates. The bright spots lead to positive shifts and the dark circles to negative ones. It is possible that multiple particles adhere to the same or neighbouring positions on the sensor surface. Thus, a time series may be affected by multiple upward and downward shifts.

The height of a shift depends on the pixel coordinates associated with the time series. Typically, the largest increase of the greyscale values is in the centre of a spot. The shifts become smaller with increasing distance of the pixel coordinates to the centre. To our knowledge, typical values for the jump height do not exist.

Figure 2.2 shows a PAMONO time series of length 1 000 for a single pair of pixel coordinates. The measurements were taken with a frequency of 20 Hz. The data set, which contains the corresponding greyscale images, is freely available for public use and stems from project B2 of the Collaborative Research Centre *SFB 876* [1] (Siedhoff et al., 2014b). We use synthetically generated data, which has the advantage that we know the true time points at which particles adhere to the sensor surface. These data consist of real particle signals that have been placed manually on images which contain only real background noise. A detailed explanation of the synthesising procedure can be found in Siedhoff et al. (2014a). For the extraction of the pixel time series from the image data, we use `R` code generated for the Master's thesis of Abbas (2013).

The depicted time series is affected by three particle adhesions, occurring at the time points $t = 164$, $t = 478$, and $t = 539$. Between the resulting location shifts, the level of the time series is nearly constant. It does not contain any visible trends or outliers. Compared

[1] As of March 2019, the data and a detailed description are available on the project's website `http://sfb876.tu-dortmund.de/SPP/sfb876-b2.html`. In our example, we use the time series with pixel coordinates $(497, 53)$ which can be found in the folder `synth/pcount100/nstt/training` of the data set `PAMONO Sensor Data 200nm_10Apr13`.

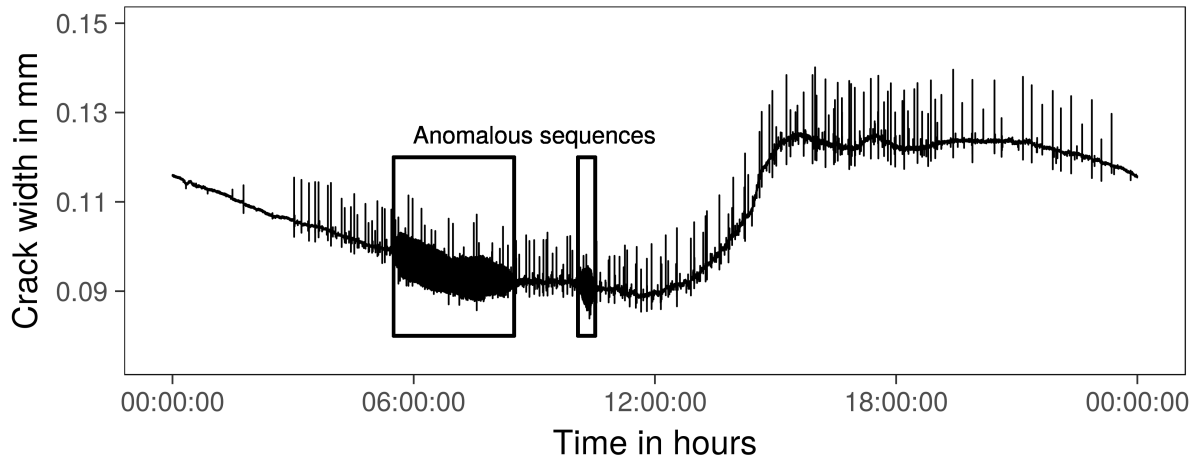

Figure 2.3: Time series of crack-width measurements from a bridge monitoring. The black boxes highlight sequences of increased variability.

to the heart-rate measurements, the process is very well-behaved. This provides a good opportunity to apply our step-detection procedures in a rather clean scenario.

2.2.3 Monitoring of crack widths in concrete bridges

Sudden increases of crack widths in a concrete bridge can point at the breaking of tension wires inside the bridge. Such events give valuable information on the remaining lifetime of a bridge. Hence, their timely detection is important to perform appropriate actions like, for example, closing the bridge.

A monitoring of crack widths has been started on a bridge in Bochum in June 2016. Several monitoring devices were installed over existing cracks on the bridge. The measurements are taken every two seconds. A detailed description of the data and the underlying research goals can be found in Abbas et al. (2018).

For this thesis, we consider the time series depicted in Figure 2.3 as an example. It shows 43 200 measurements taken on 5th August 2016. The signal is strongly non-linear. This is because of the temperature, which is also measured but not shown in the figure to put an emphasis on the crack widths. The black boxes highlight sequences, called anomalous sequences, with comparatively large variability. Such increases occur abruptly and are caused by unknown physical effects. They can complicate the analysis of these data. Other challenges are frequently occurring isolated peaks, which are caused by heavy vehicles or the tram passing over the bridge.

In Chapter 4, we will compare methods for shift detection in time series with a locally linear signal. The crack-width measurements will serve as an example to show how the discussed procedures can be adjusted to detect variability changes.

2.3 Robust time-series filtering

In this section, we summarise procedures, called *time-series filters* in the following, which aim at extracting the signal of a time series in real time. Motivated by the idea that the signal often cannot be modelled globally, they are based on local model assumptions.

The real-valued time series $(Y_t\colon\ t \in \mathbb{N})$ is assumed to be decomposable according to the additive components model

$$Y_t = \mu_t + \varepsilon_t + \eta_t,\ t \in \mathbb{N}. \tag{2.1}$$

The smooth sequence $(\mu_t\colon\ t \in \mathbb{N})$ denotes the true and unknown deterministic signal, which exhibits changing trends and occasional level shifts, but otherwise varies slowly over time. By $(\varepsilon_t\colon\ t \in \mathbb{N})$ we denote additive random noise, which we consider to be independent with expectation $\mathrm{E}(\varepsilon_t) = 0$ and time-varying variance $\mathrm{Var}(\varepsilon_t) = \sigma_t^2$. The outlier-generating mechanism $(\eta_t)_{t\in\mathbb{N}}$ consists of random variables that are zero most of the time, but sometimes take large absolute values.

We consider moving-window techniques which adapt to the local signal course. The methods presented in Chapters 3 and 4 are also based on this idea. A moving window contains the most recent observations of the time series and can be used for the approximation of local signal characteristics like μ_t or σ_t. It avoids making, often unrealistic, global parametric assumptions on the model in Equation (2.1).

This section contains a review of several time-series filters, which are based on the assumption that the signal can be approximated by a constant or a linear function in the window. Implementations for the described techniques can be found in the `R` package `robfilter` (Fried et al., 2014), which is available for free on the Comprehensive `R` Archive Network (CRAN).

Two possible ways of estimating the signal at time t using a moving time window are the *delayed* and the *full online estimation*. For the delayed estimation, the window is of odd width $n = 2 \cdot h + 1$, $h \in \mathbb{N}$, and centred at the target time t. Then, the sample for the estimation is $\boldsymbol{Y}_t^{(n)} = (Y_{t-h}, \ldots, Y_t, \ldots, Y_{t+h})$. This leads to a time delay of h time points before the signal can be estimated. An estimate for the location at the rightmost time point in the window can be obtained by extrapolation. In the full online estimation, we use only observations prior to the current time point t to estimate the signal. By using samples of the form $\boldsymbol{Y}_t^{(n)} = (Y_{t-n+1}, \ldots, Y_t)$, we estimate the signal without a time delay at the time point t in windows of width n, which can be an even or odd number.

The size of the window width depends on several aspects. The validity of a locally approximately constant or linear signal in the window is crucial. Thus, n should not be chosen too large if the signal exhibits many trend changes or level shifts. For the delayed approach, a large window also increases the delay until μ_t can be estimated. On the other hand, a small window width can increase the variability of the estimator and decrease its robustness against outliers (Gather et al., 2006).

We focus on the delayed estimation when describing the procedures in the sequel. They

can be straightforwardly adjusted to the full online estimation by replacing the indices accordingly. The signal is estimated at the time points $t = h+1, h+2, \ldots$ for the delayed, and $t = n, n+1, \ldots$ for the full online approach.

In Subsection 2.3.1, we describe filters for locally constant signals. Locally linear signals are treated in Subsection 2.3.2.

2.3.1 Location-based filtering

When the signal within the window is nearly constant, so that $\mu_{t+i} \approx \mu_t$ for $i = -h, \ldots, h$, it is reasonable to use a location estimator $\hat{\mu}_t$ to approximate μ_t. An obvious choice is the *moving average*

$$\hat{\mu}_t^{(\mathrm{MEAN})} = \frac{1}{n} \sum_{i=-h}^{h} Y_{t+i}.$$

Drawbacks are its sensitivity to outliers and that it smooths level shifts. A robust procedure with better shift preservation is given by Tukey's *running median* (Tukey, 1977, p. 120ff., Mäkivirta et al., 1991), defined as

$$\hat{\mu}_t^{(\mathrm{MED})} = \underset{i=-h,\ldots,h}{\mathrm{median}} \left(Y_{t+i} \right).$$

The robustness of an estimator can be quantified by the *finite-sample breakdown point* (Donoho and Huber, 1983), which is the number of observations in the sample that need to be altered so that the estimate deviates arbitrarily strong from the value for the original sample. A definition can be found in Appendix D.2. The finite-sample breakdown point of the median is $\lceil n/2 \rceil / n$, so that it can resist up to $\lceil n/2 - 1 \rceil$ outliers. Here, $\lceil . \rceil$ means rounding up to the nearest integer. In the absence of noise and trends, the running median can preserve permanent and temporary level shifts exactly when they last at least $h+1$ observations. The running median tends to deteriorate in trend periods as the assumption of an approximately locally constant signal is only legitimate for very short time windows (Gather et al., 2006). Moreover, under normality, the sample median is substantially less efficient than the mean. As measured by the *asymptotic relative efficiency* (ARE), which is the ratio of the asymptotic variances, the median has an ARE of $2/\pi$ compared to the mean (Serfling, 2011). This means that the asymptotic variance of $\hat{\mu}_t^{(\mathrm{MEAN})}$ is about 37% smaller than the one of $\hat{\mu}_t^{(\mathrm{MED})}$. However, $\hat{\mu}_t^{(\mathrm{MED})}$ can be more efficient than $\hat{\mu}_t^{(\mathrm{MEAN})}$ under heavy-tailed distributions. Nevertheless, the low efficiency under normality points at a small one for distributions with a shape similar to the one of the normal distribution (Lehmann, 1997, p. 360).

The behaviour during trend periods can be improved by *modified trimmed means* (MTM) and *double-window modified trimmed means* (DWMTM) (Lee and Kassam, 1985). These filters remove normal noise more efficiently than the running median and are robust against outliers. In a first step, the level μ_t and variability σ_t within $\boldsymbol{Y}_t^{(n)}$ are calculated robustly by using the median $\hat{\mu}_t^{(\mathrm{MED})}$ and the *median absolute deviation from the median* (MAD).

The latter is given by

$$\hat{\sigma}_t^{(\mathrm{MAD})} = c_n \cdot \underset{i=-h,\ldots,h}{\mathrm{median}} \left(|Y_{t+i} - \hat{\mu}_t^{(\mathrm{MED})}| \right). \tag{2.2}$$

The factor $c_n \in \mathbb{R}$ depends on the sample size and assures an asymptotically unbiased estimation at a specified error distribution. For example, at normal noise and for a large window width, we use $c_n = 1.483$ (Maronna et al., 2006, p. 33). The second step consists of trimming all observations that deviate more than a specific multiple $d_t = d \cdot \hat{\sigma}_t^{(\mathrm{MAD})}$, $d \geq 0$, from $\hat{\mu}_t^{(\mathrm{MED})}$. The remaining observations are averaged, so that the resulting signal estimator is given by

$$\begin{aligned} \hat{\mu}_t^{(\mathrm{MTM})} &= \frac{1}{|J_t|} \sum_{i=-h}^{h} Y_{t+i} \cdot I_{J_t}(i), \\ J_t &= \left\{ i = -h, \ldots, h \colon \ |Y_{t+i} - \hat{\mu}_t^{(\mathrm{MED})}| \leq d_t \right\}, \end{aligned}$$

where $I_A(x)$ denotes the indicator function of x on the set A and $|J_t|$ represents the cardinality of the set J_t. Lee and Kassam (1985) suggest $d = 2$ as a compromise between the robustness of the running median ($d = 0$) and the efficiency of the moving average ($d = \infty$) under normality.

The DWMTM is a refinement of the MTM. A short window of width $2 \cdot m + 1$, $m < h$, is used in the first step. The final estimates are calculated from a long window of width $2 \cdot h + 1$. This combination allows for an efficient noise reduction even close to level shifts and during trends. When these patterns appear, the initially calculated estimates are still reliable. Thus, the final estimates are calculated from nearby observations with values close the target level. In total, the estimator is given by

$$\begin{aligned} \hat{\mu}_t^{(\mathrm{DWMTM})} &= \frac{1}{|J_t|} \sum_{i=-h}^{h} Y_{t+i} \cdot I_{J_t}(i), \\ J_t &= \left\{ i = -h, \ldots, h \colon \ |Y_{t+i} - \hat{\mu}_t^{(\mathrm{MED},m)}| \leq d_t \right\}, \\ \hat{\mu}_t^{(\mathrm{MED},m)} &= \underset{i=-m,\ldots,m}{\mathrm{median}} \left(Y_{t+i} \right). \end{aligned}$$

Linear median hybrid filters with finite impulse response (Heinonen and Neuvo, 1987; Heinonen and Neuvo, 1988; Astola et al., 1989; Wichman et al., 1990) start with applying a couple of linear subfilters to the sample. These are selected to trace specific types of polynomial trends well. The median of the outcome of these subfilters is used as the estimator for μ_t. We describe two filters which follow this idea.

The *finite impulse responsive median hybrid* (FMH) filter uses the central observation

in the sample and the averages of its left and right part as subfilters, so that

$$\hat{\mu}_t^{(\mathrm{FMH})} = \mathrm{median}\left\{\Phi_1(t), \Phi_2(t), \Phi_3(t)\right\},$$
$$\Phi_1(t) = \frac{1}{h}\sum_{i=1}^{h} Y_{t-i},\ \Phi_2(t) = Y_t,\ \Phi_3(t) = \frac{1}{h}\sum_{i=1}^{h} Y_{t+i}.$$

Compared to the running median with the same window width, this filter preserves shifts better. However, it is considerably less robust against outliers. Similar to the running median, the FMH filter deteriorates in trend periods.

Replacing the simple averages with weighted averages can improve the performance. The weighting scheme is chosen so that the mean square error of the prediction of a polynomial trend in the centre of the window is minimised. The *predictive linear median hybrid* (PFMH) filter results from the assumption of a linear trend. It uses the weights $w_i = {}^{(4\cdot h-6\cdot i+2)}/_{\left(h^2-h\right)}$, $i = 1, \ldots, h, h > 1$, and the subfilters

$$\Phi_1(t) = \sum_{i=1}^{h} w_i \cdot Y_{t-i} \quad \text{and} \quad \Phi_3(t) = \sum_{i=1}^{h} w_i \cdot Y_{t+i}.$$

The PFMH filter can remove normal noise and preserves shifts during a linear trend, which is a benefit compared to the running median. However, it is prone to impulsive spiky noise.

To illustrate the described filters, we estimate the level of a simulated time series generated from standard normal noise. In Figure 2.4, the original time series and the filter outputs are depicted. Using an artificial instead of a time series from a real application, for example as introduced in Section 2.2, has the advantage that we can control the trends, outliers, and level shifts. The time series in Figure 2.4 contains two level shifts at $t = 50$ and $t = 90$. From $t = 141$ to $t = 180$ there is a positive linear trend. A negative one reaches from $t = 251$ to $t = 300$ and is followed by a location shift. In addition, 7.5% of the observations are outliers at random positions. Furthermore, there are two patches of consecutive outliers from $t = 196$ to $t = 200$ and from $t = 311$ to $t = 315$. We apply the filters with $h = 21$. For the DWMTM, we choose $m = 11$ for the inner window.

Both hybrid filters lead to rather wiggly estimates of the level (Figure 2.4(c)). Linear trends and location shifts are reproduced accurately. The PFMH filter is severely affected by the outlier patches. MTM (Figure 2.4(b)) and running median (Figure 2.4(a)) smooth the level shifts to some extent, particularly visible at $t = 301$ after the declining trend. The running median furthermore shows a step-like behaviour during the trend periods. The DWMTM traces shifts and trends more precisely. In addition, it is not affected by the outliers. The signal estimate is stable over time, so that it shows a slightly better overall performance than the other considered filters.

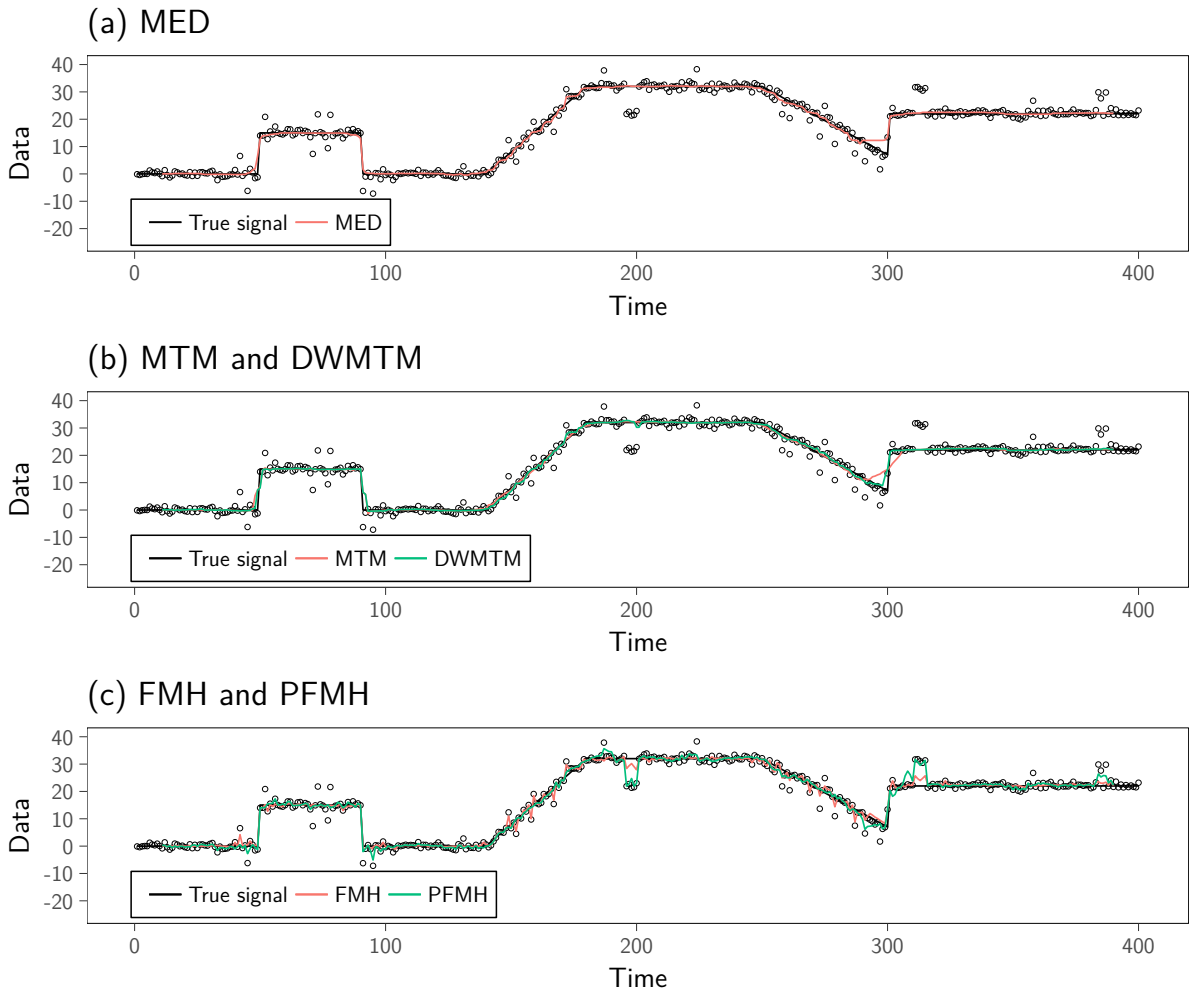

Figure 2.4: Level estimates of selected location-based filters with window widths $h = 21$ and $m = 11$ (for DWMTM) on a simulated time series with standard normal noise and 7.5% additive outliers at random time points as well as two outlier patches.

2.3.2 Regression-based filtering

Robust regression techniques can help to construct filters with similarly good properties as the running median for locally approximately constant signals and a better performance during trend periods. We assume that the signal is linear in the time window, so that

$$\mu_{t+i} \approx \mu_t + \beta_t \cdot i, \; i = -h, \ldots, h. \tag{2.3}$$

Then, μ_t and β_t are level and slope in the window centred at time point t. Let

$$\hat{e}_{t+i} = Y_{t+i} - \left(\hat{\mu}_t + \hat{\beta}_t \cdot i\right), \; i = -h, \ldots, h,$$

be the regression residuals for estimators $\hat{\mu}_t$ and $\hat{\beta}_t$ for μ_t and β_t. In the following, we summarise several possible choices for $\hat{\mu}_t$ and $\hat{\beta}_t$.

The most popular technique for linear regression is *ordinary least squares* (OLS) regres-

sion, for which the estimators are defined as

$$\hat{\beta}_t^{(\mathrm{OLS})} = \frac{\sum_{i=-h}^{h} \left(i - \bar{i}_t\right) \cdot \left(Y_{t+i} - \overline{Y}_t\right)}{\sum_{i=-h}^{h} \left(i - \bar{i}_t\right)^2},$$
$$\hat{\mu}_t^{(\mathrm{OLS})} = \overline{Y}_t - \hat{\beta}_t \cdot \bar{i}_t,$$
$$\overline{Y}_t = \frac{1}{n} \sum_{i=-h}^{h} Y_{t+i}, \quad \bar{i}_t = \frac{1}{n} \sum_{i=-h}^{h} i.$$

Similar to the moving average, the OLS estimators can be easily influenced by outliers (Rousseeuw and Leroy, 1987, p. 23f.).

Motivated by the median being the L_1 estimate of location, L_1-*regression* is a reasonable alternative to OLS regression (Rousseeuw and Leroy, 1987, p. 10). The regression estimators are given by

$$\left(\hat{\mu}_t^{(L_1)}, \hat{\beta}_t^{(L_1)}\right) = \arg \min_{\hat{\mu}_t, \hat{\beta}_t} \sum_{i=-h}^{h} |\hat{e}_{t+i}|.$$

The finite-sample replacement breakdown point for L_1-regression is slightly less than 30% in large samples with a fixed equidistant design $i = -h, \ldots, h$ (Davies et al., 2004).

Rousseeuw and Hubert (1999) propose *deepest regression* (DR), which is a generalisation of the median to the regression context. It has similar properties as L_1-regression. The finite-sample breakdown point is 1/3 for fixed equidistant designs and therefore slightly larger than for L_1-regression in large samples (Gather et al., 2006). The idea is to compute estimators that maximise the *regression depth*, given by

$$\left(\hat{\mu}_t^{(\mathrm{DR})}, \hat{\beta}_t^{(\mathrm{DR})}\right) = \arg \max_{\hat{\mu}_t, \hat{\beta}_t} \left\{\mathrm{rdepth}\left(\left(\hat{\mu}_t, \hat{\beta}_t\right), \boldsymbol{Y}_t^{(n)}\right)\right\},$$

where

$$\mathrm{rdepth}\left(\left(\hat{\mu}_t, \hat{\beta}_t\right), \boldsymbol{Y}_t^{(n)}\right) = \min_{i=-h,\ldots,h} \left\{\min\left\{L^+(i) + R^-(i),\ R^+(i) + L^-(i)\right\}\right\},$$

with

$$L^+(i) = |\{j = -h, \ldots, i\colon\ \hat{e}_{t+j} \geq 0\}|, \quad R^-(i) = |\{j = i+1, \ldots, h\colon\ \hat{e}_{t+j} < 0\}|,$$

and $L^-(i)$ and $R^+(i)$ defined analogously. Hence, the regression depth quantifies how deeply a regression function with parameters $\hat{\mu}_t$ and $\hat{\beta}_t$ lies in the data.

Least median of squares (LMS) regression minimises the median of the squared residuals, so that

$$\left(\hat{\mu}_t^{(\mathrm{LMS})}, \hat{\beta}_t^{(\mathrm{LMS})}\right) = \arg \min_{\hat{\mu}_t, \hat{\beta}_t} \left\{\underset{i=-h,\ldots,h}{\mathrm{median}}\left(\hat{e}_{t+i}^2\right)\right\}$$

(Hampel, 1975; Rousseeuw, 1984). Its finite-sample breakdown point equals $\lfloor n/2 \rfloor / n$, so that it can resist almost 50% of observations in the window being affected by outliers or a level shift (Davies and Gather, 2005). Hence, it provides greater robustness than L_1-regression and DR. The LMS is computationally expensive and rather inefficient under normally distributed noise, so that the filter output can be quite wiggly (Gather et al., 2006).

Similar remarks apply to *least trimmed squares* (LTS) regression, which is asymptotically more efficient than LMS regression (Rousseeuw, 1984). It minimises the sum of the $r \in \mathbb{N}$, $r < n$, smallest squared residuals. The estimators are given by

$$\left(\hat{\mu}_t^{(\mathrm{LTS})}, \hat{\beta}_t^{(\mathrm{LTS})}\right) = \arg \min_{\hat{\mu}_t, \hat{\beta}_t} \sum_{i=1}^{r} \left(\hat{e}_t^2\right)_{i:n},$$

where $\left(\hat{e}_t^2\right)_{i:n}$ is the i-th ordered squared residual in the window at time point t.

Repeated median (RM) regression by Siegel (1982) provides a good compromise between robustness and efficiency under normal noise (Davies et al., 2004; Gather et al., 2006). The estimators can be defined as

$$\begin{aligned} \hat{\beta}_t^{(\mathrm{RM})} &= \underset{i=-h,\ldots,h}{\mathrm{median}} \left(\underset{\substack{j=-h,\ldots,h \\ j \neq i}}{\mathrm{median}} \left(\frac{Y_{t+i} - Y_{t+j}}{i - j} \right) \right), \\ \hat{\mu}_t^{(\mathrm{RM})} &= \underset{i=-h,\ldots,h}{\mathrm{median}} \left(Y_{t+i} - \hat{\beta}_t^{(\mathrm{RM})} \cdot i \right). \end{aligned} \tag{2.4}$$

The RM has a finite sample breakdown point of $\lfloor n/2 \rfloor / n$.

In a simulation study, in which LMS, LTS, DR, and RM are compared, Gather et al. (2006) find that the bias behaviour of RM regression is worse than that of LMS and LTS if more than 30% of the data points are affected by contamination. Furthermore, RM and DR tend to blur location shifts and are more efficient than LMS and LTS under normality, as compared to OLS. RM regression performs slightly better than DR regarding efficiency and outlier robustness. It is also Lipschitz continuous for equidistant design points. This ensures locally stable estimates when there are small changes in the observations because of, for example, noise (Fried et al., 2007). A fast algorithm allows for updating the RM estimates in linear time when using a moving time window (Bernholt and Fried, 2003).

Modifications of repeated-median filtering

Because of its promising properties, RM regression has received further attention, in particular for the filtering of medical time series like the one introduced in Subsection 2.2.1. We will now present some modifications proposed in the literature and point out the advantages over ordinary RM regression.

Trimmed repeated median (TRM) regression extends the idea of the MTM to the regression context (Bernholt et al., 2006). Starting with a robust estimate of μ_t, β_t, and σ_t, all observations, for which the deviation from the initial regression line exceeds a

specific threshold $d_t = d \cdot \hat{\sigma}_t^{(\mathrm{MAD})}$, $d \geq 0$, are removed from the sample. Here, $\hat{\sigma}_t^{(\mathrm{MAD})}$ is computed from the regression residuals. Then, OLS regression is applied to the remaining observations to get an estimator which is efficient under the normality assumption. The slope and level estimators are given by

$$\begin{aligned}
\hat{\beta}_t^{(\mathrm{TRM})} &= \frac{\sum_{i=-h}^{h} \left(i - \bar{i}_t\right) \cdot \left(Y_{t+i} - \overline{Y}_t\right) \cdot I_{J_t}(i)}{\sum_{i=-h}^{h} \left(i - \bar{i}_t\right)^2 \cdot I_{J_t}(i)}, \\
\hat{\mu}_t^{(\mathrm{TRM})} &= \overline{Y}_t - \hat{\beta}_t^{(\mathrm{TRM})} \cdot \bar{i}_t, \\
\overline{Y}_t &= \frac{1}{|J_t|} \sum_{i=-h}^{h} Y_{t+i} \cdot I_{J_t}(i), \quad \bar{i}_t = \frac{1}{|J_t|} \sum_{i=-h}^{h} i \cdot I_{J_t}(i), \\
J_t &= \left\{ i = -h, \ldots, h \colon \; |Y_{t+i} - \left(\hat{\mu}_t^{(\mathrm{RM})} + \hat{\beta}_t^{(\mathrm{RM})} \cdot i\right)| \leq d_t \right\}.
\end{aligned}$$

Similarly, the DWMTM can be adapted when computing the RM level and slope from a shorter window of width $2 \cdot m + 1 < 2 \cdot h + 1$ (Bernholt et al., 2006). This *double-window trimmed repeated median* (DWTRM) preserves sudden shifts better because the bias effects from the inclusion of shifted and non-shifted observations in the initial fit are less spread out. These trimmed filters are, unlike the RM, not Lipschitz continuous, so that slight changes in the data can have a large effect on the estimation.

Robust extensions of the linear median hybrid filters are *repeated median hybrid* filters (Fried et al., 2006). They can be designed to preserve specific signal characteristics. Like the RM, they are Lipschitz continuous. Their computation time, but also their robustness, is smaller than for the RM. An example is the *predictive repeated median hybrid* (PRMH) filter, which corresponds to the median of: (i) the central observation in the sample, (ii) the one-sided RM forecast, computed from $Y_{t-h}, \ldots, Y_{t-1}$, and (iii) the one-sided RM backcast, computed from $Y_{t+1}, \ldots, Y_{t+h}$. Sudden slope changes in an otherwise linear trend can be exactly preserved in the absence of noise, but the estimator has a rather low efficiency under normally distributed noise. The estimators are given by

$$\begin{aligned}
\hat{\beta}_t^{(\mathrm{RM},F)} &= \underset{i=-h,\ldots,-1}{\mathrm{median}} \left(\underset{\substack{j=-h,\ldots,-1 \\ j \neq i}}{\mathrm{median}} \left(\frac{Y_{t+i} - Y_{t+j}}{i - j} \right) \right), \\
\hat{\mu}_t^{(\mathrm{RM},F)} &= \underset{i=-h,\ldots,-1}{\mathrm{median}} \left(Y_{t+i} - \hat{\beta}_t^{(\mathrm{RM},F)} \cdot i \right), \\
\hat{\beta}_t^{(\mathrm{RM},B)} &= \underset{i=1,\ldots,h}{\mathrm{median}} \left(\underset{\substack{j=1,\ldots,h \\ j \neq i}}{\mathrm{median}} \left(\frac{Y_{t+i} - Y_{t+j}}{i - j} \right) \right), \\
\hat{\mu}_t^{(\mathrm{RM},B)} &= \underset{i=1,\ldots,h}{\mathrm{median}} \left(Y_{t+i} - \hat{\beta}_t^{(\mathrm{RM},B)} \cdot i \right), \\
\hat{\mu}_t^{(\mathrm{PRMH})} &= \mathrm{median} \left(\hat{\mu}_t^{(\mathrm{RM},F)}, Y_t, \hat{\mu}_t^{(\mathrm{RM},B)} \right),
\end{aligned}$$

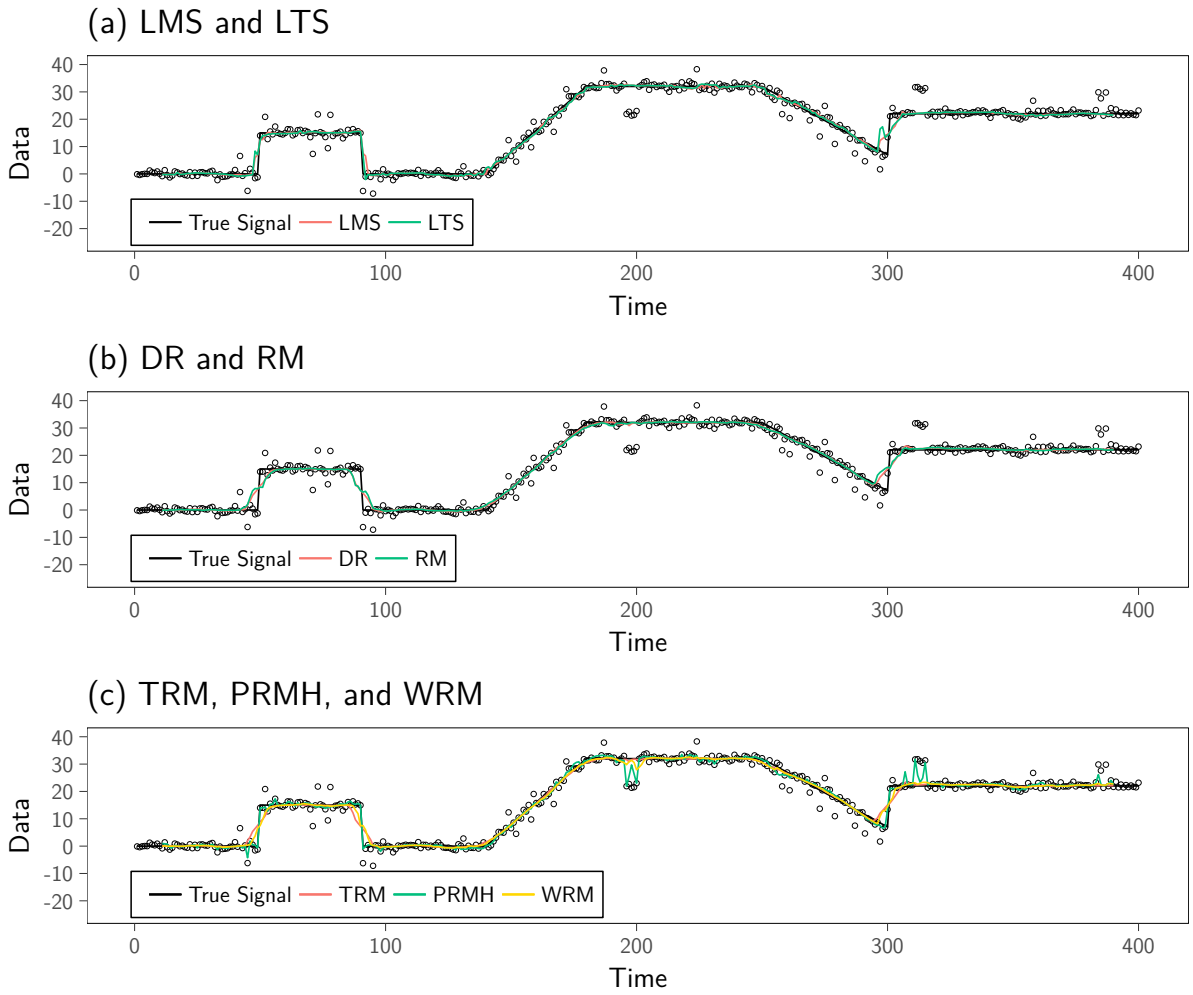

Figure 2.5: Level estimates of selected regression-based filters with window width $h = 21$ on a simulated time series with standard normal noise and 7.5% additive outliers at random time points as well as two outlier patches.

where the superscripts F and B denote the forecast and the backcast.

Weighted repeated median (WRM) regression (Fried et al., 2007) improves the noise reduction by weighting observations when calculating the RM. Giving higher weights to observations close to the estimation time t reduces the influence of remote shifts so that it is possible to use a larger window for the regression. It is also Lipschitz continuous. Let $w_i, \tilde{w}_i$, $i = -h, \ldots, h$, be two weighting schemes with $w_i, \tilde{w}_i \in \mathbb{N}$. By $w_i \diamond Y_i$ we denote that Y_i is replicated w_i times. The WRM estimators are

$$\hat{\beta}_t^{(\mathrm{WRM})} = \underset{i=-h,\ldots,h}{\mathrm{median}} \left(\tilde{w}_i \diamond \left(\underset{\substack{j=-h,\ldots,h \\ j \neq i}}{\mathrm{median}} \left(\tilde{w}_j \diamond \frac{Y_{t+i} - Y_{t+j}}{i - j} \right) \right) \right),$$

$$\hat{\mu}_t^{(\mathrm{WRM})} = \underset{i=-h,\ldots,h}{\mathrm{median}} \left(w_i \diamond \left(Y_{t+i} - \hat{\beta}_t^{(\mathrm{WRM})} \cdot i \right) \right).$$

A possible weighting scheme are the triangular weights $w_i = \tilde{w}_i = (h+1) - |i|$, $i = -h, \ldots, h$.

Fried et al. (2007) conclude from a study, that the WRM is a good compromise between

stable estimates, preservation of location shifts, robustness against outliers, and efficiency under normally distributed noise.

Figure 2.5 shows the level estimates of the robust regression filters with $h = 21$ for the simulated time series from before. The PRMH filter tracks the level shifts well but is easily affected by outliers, similar to the FMH and PFMH filters. Furthermore, its estimates are rather unstable, even in periods without outliers. This observation can also be made for the LMS and the LTS, although to a lesser extent. These filters also track the first level shifts prior to $t = 100$ quite well. In contrast, the DR and RM filters smooth these location shifts. In addition, all but the PRMH filter seem to have problems with the location shift at $t = 301$. Except for the PRMH, the filters are reasonably robust against the outliers.

Adaptive selection of the window width

As mentioned at the beginning of this section, the choice of the window width is based on a compromise between robustness, efficiency, and a small delay for tracking level and trend changes. The previously described filters use a fixed window width. Some effort has been made to construct RM-based filters which adjust the window width automatically over time. They adapt to local signal and noise characteristics.

One way for a data-adaptive choice of the window width is using a goodness-of-fit test based on the signs of the most recent residuals in the window (Gather and Fried, 2004). In case of a good regression fit, the signs of the residuals should be balanced. If the numbers of negative and positive residuals in the foremost part of the window differ substantially, the estimators should be calculated again with a reduced window width. This should be repeated until the signs of the residuals are appropriately balanced.

An online version of this approach, called *adaptive online repeated median* (aoRM), is introduced by Schettlinger et al. (2010). To decide whether the current window width should be decreased or increased, it is suggested to use a sign-based goodness-of-fit test on the most recent residuals to a rather large significance level. The assumptions on the noise distribution are quite weak but, because of using a sign test, the procedure is not very efficient under normality.

Based on the idea of the aoRM, Borowski and Fried (2014) propose the *slope-comparing adaptive repeated median* (SCARM) for data-adaptive window-width selection. Similar to the aoRM, it uses a test to decide if the assumption of a locally linear signal is justifiable within the time window for the currently used width. The window $(Y_{t-n+1}, \ldots, Y_t)$ of width $n = h + k$ is split into a left part $(Y_{t-n+1}, \ldots, Y_{t-k})$ of width h and a right part $(Y_{t-k+1}, \ldots, Y_t)$ of width k. For each subwindow, the RM slope estimators $\hat{\beta}_t^{(\text{RM, left})}$ and $\hat{\beta}_t^{(\text{RM, right})}$ are calculated separately. The test statistic is the standardised difference of

these estimators and given by

$$T_t^{(\text{SCARM})} = \frac{\hat{\beta}_t^{(\text{RM, left})} - \hat{\beta}_t^{(\text{RM, right})}}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}_t^{(\text{RM, left})} - \hat{\beta}_t^{(\text{RM, right})}\right)}}, \tag{2.5}$$

where the denominator is the estimated standard deviation of the difference. This scale estimator is immune to underlying trends and can be modified to work well with different types of noise, for example, from a first-order autoregression (AR(1)). The null distribution can be approximated by a t-distribution, where the degrees of freedom depend on h and k. Starting with a minimal window width, the window width is increased by one for each new observation if the null hypothesis cannot be rejected. Simulations by Borowski and Fried (2014) indicate that the SCARM test is more efficient, in terms of detection rate and speed, in detecting signal changes than the sign test of the aoRM. The SCARM reduces the window width to the minimal value as soon as the test yields a significant result. The estimation is then performed with the minimal window width. This improves the preservation of relevant signal characteristics further compared to the aoRM, which tests repeatedly and reduces the window width slowly.

2.4 Online detection of shifts and volatility changes

In addition to signal extraction, detecting relevant patterns like level shifts, the beginning of trends, trend changes, or volatility changes is often of interest. These tasks are closely related to *Statistical Process Control* (SPC).

In this thesis, the detection of sudden level shifts is of foremost interest. In Subsection 2.4.1, we concentrate on this aspect. We briefly comment on the identification of volatility changes in Subsection 2.4.2 because the procedures discussed in Chapters 3 and 4 can be adapted to this task as well.

2.4.1 Detection of location shifts

In this subsection, we present selected approaches for the real-time detection of level shifts under a restricted false-alarm frequency. Thus, in contrast to many existing sequential methods for change-point detection, for example by Zeileis (2005), Eichinger and Kirch (2018), or Aue and Horváth (2013), we are not interested in keeping a global significance level.

Basic principles of a control chart

Control charts are popular tools for the detection of unusual deviations of one or more process characteristics from a fixed target value during online monitoring. We will embed the procedures of the Chapters 3 and 4 into the control-chart context, which is why we

will describe briefly the general ideas of this concept and give some examples to point out some drawbacks of ordinary control-chart principles. Our main reference for the sequel is the book "Statistical Quality Control: A Modern Introduction" by Montgomery (2009, p. 180ff.).

When the goal is to detect level changes, a target value represents the level of the observations when the process operates *in control*, that is, in the absence of relevant changes. Then, it is only affected by unavoidable random noise. Using the conventional assumptions of the control-chart context, this also means that the signal is globally constant, so that $\mu_t = \mu_0$ for all $t \in \mathbb{N}$, where $\mu_0 \in \mathbb{R}$ is the target value. Other types of disturbances may affect the process, leading to observations that cannot be explained by random noise, for example a level shift. In this case, the process is said to be *out of control*. In the following, we call the time point at which a structural break occurs *change point*.

Based on sequentially taken samples from the process, a control chart aims to identify an out-of-control situation as soon as possible after its occurrence. For each new incoming sample, the value of a control statistic is computed and compared to control limits, set in a fixed distance from the target value. If the control statistic realises within these limits, the process is assumed to be in control. Otherwise, it is said to be out of control and an alarm is given. Typically, some kind of action is taken to eliminate the cause for the disturbance and bring the process back into control.

The control limits are often chosen so that the process fulfils certain requirements on the *run length*, which corresponds to the duration between two subsequent alarms (Basseville and Nikiforov, 1993, p. 151ff.). The general idea is that the duration until an alarm is given should be long for an in-control process, while it should be small for an out-of-control process. We will go into more detail in Subsection 3.3.1, where we discuss two typical run-length-based criteria: the average run length and the median run length.

A control chart typically operates in two phases (Montgomery, 2009, p. 198f.). In *phase I*, historical data are analysed to make sure that the process is in control before the monitoring starts. Moreover, they are used to estimate the target value and other process parameters to establish the control limits if the parameters are unknown. The actual monitoring is called *phase II*. During this phase, new observations are analysed to detect deviations from the in-control behaviour as learned in phase I.

Traditional control-chart principles

Well-known control schemes for the mean are the *Shewhart*, the *cumulative sum* (CUSUM), and the *exponentially weighted moving average* (EWMA) control chart. The underlying principles are often used as the basis for more specialised control charts, which is why the general ideas are described shortly in the following, assuming a constant process variance $\sigma_t^2 = \sigma^2$ for all $t \in \mathbb{N}$.

The Shewhart control chart uses the sample mean as its control statistic (Shewhart,

1931, p. 249ff.). Typically, the samples are non-overlapping. This can lead to a large detection delay if the sample sizes are large and the structural break happens between two consecutive samples. A simple modification would be to use a moving average as described in Subsection 2.3.1. This would allow for a decision with each new incoming observation (Montgomery, 2009, p. 428). The Shewhart control chart is useful for detecting large and long-term location shifts. On the other hand, it is weak in finding small to moderate-sized shifts because it only uses information from the current sample (Montgomery, 2009, p. 402).

The CUSUM control chart uses all in-control information since the last detected change point for its decision, which leads to an increased detection speed for small location shifts (Page, 1954). Basically, the control statistic is the cumulative sum of the differences between each new observation and the target value.

Like the CUSUM chart, the EWMA control chart also has a memory (Roberts, 1959). Its control statistic at time t is a weighted mean between the mean of the current sample and the previous value of the control statistic. Thus, it is possible to specify how important the process history and the new observations are for the decision of the control chart by choosing an appropriate value for the corresponding tuning parameter.

The CUSUM and EWMA charts have similar run-length properties. Unlike the Shewhart control chart, they are good at detecting small to moderate-sized location shifts fast. However, they can be slower in identifying large shifts (Montgomery, 2009, p. 424).

To set up the control limits, all three control charts need a fixed target value μ_0 and the process variance σ^2. If they are unknown before the monitoring, they have to be estimated in phase I. Jensen et al. (2006) conclude from a literature survey on parameter estimation in the control-chart context that a large amount of in-control data is necessary to ensure that a control chart keeps the desired restrictions on the run length and detects structural breaks reliably. Chakraborti (2000), for example, recommends using 2 500 to 5 000 observations to estimate the in-control parameters of the Shewhart control chart. However, in many applications, for example the intensive care context, such a large amount of data, or any in-control data at all, are not available (Imhoff et al., 2002). By using a self-starting control chart, it is possible to overcome the need for many in-control observations (Keefe et al., 2015). The idea is to start with initial estimates for the in-control parameters, obtained from a small historical data set. These are updated with each new incoming observation during phase II until an alarm is given. The risk is that an undetected out-of-control situation may lower the performance of the control chart as it can worsen the estimation of the process parameters.

The dependence on a fixed target value is a general problem of ordinary control charts in the applications we have in mind. In addition to the problem of its availability, a fixed value does often not reflect potential changes in the process dynamics (Imhoff et al., 2002). Thus, they are likely to confuse even a slowly varying in-control signal with an out-of-control signal. Examples for such data are the time-series of heart-rate measurements and the

crack-width measurements in Section 2.2.

The choice of the control limits of the aforementioned charts depends on distributional assumptions that have to be fulfilled by the data, for example that they stem from a normal distribution. This is necessary to ensure that the desired run-length properties are valid. In real-world applications, the data-generating distribution is often unknown, which makes it likely that distributional assumptions are violated. This can weaken the performance of the charts dramatically. For example, the Shewhart control chart leads to a larger number of false alarms than desired if the underlying distribution is heavy-tailed or asymmetric instead of a normal distribution (Amin et al., 1995). Similar results are reported for the CUSUM and the EWMA chart (Qiu and Li, 2011; Human et al., 2011).

Distribution-free control charts

Unlike parametric charts as those described before, distribution-free, also called non-parametric, control charts aim at bypassing this difficulty by choosing the control limits so that the chart has the same in-control run-length properties under every continuous distribution (Chakraborti et al., 2011). Distribution-free control charts can have better detection capabilities than parametric charts if the distributional assumptions of the latter are not valid. Moreover, they do not depend on the process variance, avoiding the necessity to estimate it. Chakraborti et al. (2001), Chakraborti and Graham (2008), and Chakraborti et al. (2011) give extensive summaries on distribution-free control charts. Most of them are based on the principles underlying the Shewhart, CUSUM, or EWMA chart, but the control statistic is replaced with a statistic whose in-control distribution is independent of the data-generating distribution. Like for the parametric charts, the distribution-free charts in general still require a fixed target value or a reference set of observations. An example of such a chart is the Mann-Whitney chart by Chakraborti and van de Wiel (2008). A two-sample Mann-Whitney test is performed between the new sample from the monitoring phase and a reference sample collected during phase I.

Also based on the Mann-Whitney statistic, Hawkins and Deng (2010) propose a distribution-free control chart that does not need in-control reference data. At time t, each time point in the set $\{1, \ldots, t\}$ is considered as a possible change point. The complete process $Y_1, \ldots, Y_t$ is divided into two samples. The first contains all observations before the selected change-point candidate, the second sample consists of all behind it. Then, the two-sample Mann-Whitney statistic is computed between both samples. This is repeated for all time points in $\{1, \ldots, t\}$. To identify a possible change point, the maximum over all realised values of the test statistic is compared to control limits derived via simulation. Ross et al. (2011) provide a computationally fast implementation of this procedure in the `R` package `cpm` (Ross, 2015). However, the chart assumes a constant in-control signal.

Hackl and Ledolter (1992) present a distribution-free EWMA chart which is based

on sequential ranks. The sequential rank of a new observation is its rank among all observations in a time window of a fixed length that covers the most recent observations. The control chart uses an EWMA statistic in which the original observations are replaced with the sequential ranks. These follow a discrete uniform distribution for an in-control process, leading to the natural target value of $\mu_0 = 0$. From their simulation studies, Hackl and Ledolter (1992) conclude that the chart leads to good results during trend periods, but is rather weak in finding location shifts.

Robust control charts

Distribution-free control charts possess an inherent robustness against outliers. This is an important property in our considered applications. However, it is difficult to tune the charts so that they resist a desired number of outliers. This also applies to several robust control charts (Fried, 2007). Moreover, they are based on similar assumptions as the Shewhart, CUSUM, and EWMA charts. In general, the control statistics are replaced with robust versions. Hence, such charts share the general drawbacks of ordinary control schemes in our considered setting (Lucas and Crosier, 1982; Rocke, 1989; Nazir et al., 2013).

The methods studied in Chapter 3 and 4 are motivated by the outlined drawbacks of the control schemes described above. We aim for control charts that are able to distinguish outlier patches from location shifts while being, at least approximately, distribution free without the need to collect large amounts of historical in-control data.

2.4.2 Detection of volatility changes

Similar to the detection of location shifts, one is often interested in finding change points in the variability. They can indicate a changing uncertainty in the data. Again, many control charts exist for this setting, where a majority follows similar principles as those for the shift detection; see for example Montgomery (2009, p. 228) and Ross et al. (2011).

Fried (2012) adapts robust, efficient location tests to this problem by applying them to the logarithm of squared centred observations in a moving time window. The resulting procedures perform better than traditional tests, like the F-test, the Mood test, or the Ansari-Bradley test, in case of outliers and under asymmetric distributions. However, they have a slightly smaller efficiency under symmetric, normally-tailed distributions. We will briefly investigate this approach in the control-chart context on the crack-width time series in Chapter 4.

An alternative approach would be to use a robust, efficient scale estimator to construct two-sample tests for variability changes. A comparison of different approaches regarding their advantages and drawbacks is a topic for future research and beyond the scope of this thesis.

2.5 Discussion

We do not provide an exhaustive description of online-monitoring approaches in this chapter. It is restricted to the methods and terminology we refer to in the remainder of this thesis. A more detailed overview containing, among others, procedures for real-time segmentation of time series or multivariate pattern detection can be found in the article by Fried et al. (2017) mentioned at the beginning of this chapter and the references cited therein.

In the sequel of this thesis, we will stay in the online-monitoring context. Our goal is to find robust and distribution-free procedures that are able to find abrupt changes in the signal of a time series with a time-varying trend.

The discussed shortcomings of ordinary control schemes, like the need for a fixed target value or historical in-control data, motivates the control charts studied in Chapter 3, which are two-sample location tests applied in a moving time window to detect sudden location shifts.

In Chapter 4, we combine these charts with robust signal-extraction procedures to identify location or trend changes in the signal of a time series with a non-linear in-control signal.

3 Robust control charts for the mean based on repeated two-sample location tests

This chapter is based on the publication "Control charts for the mean based on robust two-sample tests" by Abbas and Fried (2017), which appeared in the *Journal of Statistical Computation and Simulation*.

For this thesis, the article has been extended and changed in the following ways:

- The performance of the control charts is evaluated for additional sample sizes and noise distributions to investigate a possible generalisation of the results.
- We add a brief simulation study to analyse the influence of outliers in an out-of-control setting.
- In addition to the out-of-control run lengths of the control charts, we also investigate the detection rates for structural breaks.
- Tables instead of figures are used to show the influence of non-normality on the in-control run lengths. This change has been made to better quantify how non-normality affects the performance of the procedures.
- The in-control analysis of the control charts under normality has been extended. We discuss the formula to approximate the significance level for achieving a desired in-control average run length in more detail.
- We identify the run length by the number of tests until an alarm and not by the number of observations as in the original article. Using the definition of this chapter, the out-of-control run length corresponds to the detection delay. Both definitions only differ by an additive constant. This does not change the overall conclusions.

Minor changes involve rephrasing and extended explanations. Moreover, Introduction and Discussion have been adjusted to their role in this thesis and to avoid repetitions.

3.1 Introduction

In this chapter, we continue the considerations made in Subsection 2.4.1. We propose control charts designed for detecting location shifts in time series with a slowly time-varying in-control signal. By making local assumptions, we avoid the need to collect large historical in-control data sets and the specification of a fixed target value.

Examples for potential applications are the time series of heart-rate measurements and the PAMONO time series, both introduced in Section 2.2.

The general idea of the control charts we investigate in the sequel is to apply a two-sample location test in a moving time window to the time series. The first part of the window forms a reference set, while the second part is a test set. By comparing both windows to each other, it is possible to decide whether a location shift occurs between the current observations and their direct predecessors.

This principle allows for an adaptation to the local signal course and makes it possible to automatically distinguish between small, irrelevant, system-inherent changes and sudden large location shifts. The sample sizes for both sets control the robustness to prevent confusion between level shifts and outlier patches. Hence, the charts can be set up to resist a specific number of consecutive outliers.

The general scheme of the charts is similar to the one of some distribution-free control charts as outlined in Subsection 2.4.1, for example the Mann-Whitney chart by Chakraborti and van de Wiel (2008). As opposed to that, we use a moving reference sample, improving the versatility.

Moving two-sample tests to detect change points in a time series with a locally constant signal have been studied in several publications, for example by Fried and Gather (2007) and Fried (2007). Abbas et al. (2016) investigate them as part of an analysis pipeline for the PAMONO data. Morell (2012) uses two-sample tests for the smoothing of time series which are affected by step changes. The two latter references apply the principle retrospectively so that the complete time series has already been observed before searching for change points. In these references, the general interest lies in keeping a global significance level.

To our knowledge, none of the past works has dealt with these procedures in the control-chart context. Thus, we extend the existing work by studying the run-length properties of the approach. The considered criteria are the influence of the data-generating distribution on the in-control run length, the robustness against outliers, and the ability to rapidly detect location shifts.

In our studies, we find that rank-based tests lead to control charts with a distribution-free in-control run length. Tests based on a randomisation principle to determine the null distribution can lead to approximately distribution-free control charts. Moreover, if the test statistic is robust against outliers, the control chart inherits this property. Control charts based on efficient tests are also efficient for the detection of structural breaks in terms of detection speed and quality.

This chapter is structured as follows: In Section 3.2, we introduce the general model for the local two-sample location problem. Afterwards, in Section 3.3, we describe the two-sample location tests. They are compared in several in- and out-of-control settings via simulations in Section 3.4. The studies are followed by investigations of the performance of the control charts on the time series of heart-rate measurements and the PAMONO time series in Section 3.5. We discuss our main findings in Section 3.6.

3.2 Model

Let $(Y_t\colon\ t \in \mathbb{N})$ be a time series that can be described by the additive components model

$$Y_t = \mu_t + \varepsilon_t + \eta_t,\ t \in \mathbb{N},$$

introduced in Section 2.3. In this chapter, we assume that the variance of the noise is constant over time so that $\operatorname{Var}(\varepsilon_t) = \sigma^2$ for all $t \in \mathbb{N}$.

Following Fried and Gather (2007), we use a local two-sample location test to detect sudden shifts in $(\mu_t\colon\ t \in \mathbb{N})$, which we assume to be locally approximately constant as in Subsection 2.3.1, except for rarely occurring abrupt location shifts. For the following motivation of this idea, we assume $\eta_t = 0$ for all $t \in \mathbb{N}$.

The time window $\boldsymbol{Y}_t^{(n)} = (Y_{t-n+1}, \ldots, Y_t)$, $t \geq n$, of the most recent $n = h + k$ observations is split into the two subwindows

$$\boldsymbol{Y}_{t-} = \left(Y_{t,1}^-, \ldots, Y_{t,h}^-\right) \quad \text{and} \quad \boldsymbol{Y}_{t+} = \left(Y_{t,1}^+, \ldots, Y_{t,k}^+\right)$$

of widths h and k, where

$$Y_{t,i}^- = Y_{t-n+i},\ i = 1, \ldots, h, \quad \text{and} \quad Y_{t,j}^+ = Y_{t-k+j},\ j = 1, \ldots, k.$$

We call $\boldsymbol{Y}_{t-}$ the *reference window* and $\boldsymbol{Y}_{t+}$ the *test window*. Assuming constant signals in each window leads to

$$\operatorname{E}\left(Y_{t,i}^-\right) = \mu_{t-n+i} = \mu_{t-},\ i = 1, \ldots, h, \quad \text{and} \quad \operatorname{E}\left(Y_{t,j}^+\right) = \mu_{t-k+j} = \mu_{t+},\ j = 1, \ldots, k,$$

where $\mu_{t-}, \mu_{t+} \in \mathbb{R}$ are constants with $\mu_{t+} = \mu_{t-} + \Delta_t$ and $\Delta_t \in \mathbb{R}$ is the shift height between the time points $t-k$ and $t-k+1$. The variances in the subwindows are given by

$$\operatorname{Var}\left(Y_{t,i}^-\right) = \sigma^2,\ i = 1, \ldots, h, \quad \text{and} \quad \operatorname{Var}\left(Y_{t,j}^+\right) = \sigma^2,\ j = 1, \ldots, k.$$

We assume the random variables in each subwindow to be independent and identically distributed (i.i.d.). Moreover, if all higher-order moments of the distributions underlying both subwindows are equal, the distributions differ at most in location so that

$$Y_{t,1}^-, \ldots, Y_{t,h}^- \overset{\text{i.i.d.}}{\sim} F \quad \text{and} \quad Y_{t,1}^+, \ldots, Y_{t,k}^+ \overset{\text{i.i.d.}}{\sim} G,$$

where $F,\ G\colon \mathbb{R} \to [0,1]$ are the cumulative distribution functions of the underlying continuous distributions with $G(x) = F(x - \Delta_t)$ for all $x \in \mathbb{R}$.

Consequently, the problem of detecting a sudden shift in $\boldsymbol{Y}_t^{(n)}$ can be understood as a

testing problem for the hypotheses

$$H_{0,t}\colon \Delta_t = 0 \quad \text{vs.} \quad H_{1,t}\colon \Delta_t \neq 0. \tag{3.1}$$

Hence, under $H_{0,t}$ the signal is constant in the full time window, whereas under $H_{1,t}$ a location shift occurs between the time points $t-k$ and $t-k+1$. If $H_{0,t}$ is rejected, we call $t-k+1$ the change point.

In this thesis, we concentrate on the two-sided problem as shown in (3.1). Modification of the control charts described in this chapter to the one-sided cases can be achieved by replacing the two-sided tests with one-sided tests.

Applying the tests as control charts in a moving time window implies a non-avoidable detection delay which increases in k. This is because a certain number of observations in the test window has to be shifted before $H_{0,t}$ can be rejected. Hence, the procedure can be set up to distinguish between patches of subsequent outliers and persistent location shifts (Fried, 2007). Similarly, a large value of h protects against outliers in the reference window. If the assumption of a locally constant signal is valid, long subwindows improve the estimation of the process level and can increase the efficiency of the procedure. However, in case of a violation, the test is likely to confuse small trends with location shifts, even though the process is in control. Thus, when choosing the window widths, there is a trade-off between the tolerable detection delay, the robustness, the power of the local test, and the justifiability of the local assumptions.

3.3 Methods

In this section, we introduce different two-sample tests and performance criteria. In Subsection 3.3.1, we define the run length and related measures used to set up and evaluate the control charts. Selected two-sample location tests, which are the basis for the control charts, are presented in Subsection 3.3.2.

3.3.1 Measuring the performance of a control chart

In online-monitoring applications, it is important to detect change points as soon as possible while avoiding too many false alarms (Basseville and Nikiforov, 1993, p. 6). In this subsection, we introduce our main performance criterion. If not noted otherwise, we follow Basseville and Nikiforov (1993, p. 151ff.).

To facilitate the descriptions in the sequel, we make two assumptions:

- For an in-control process, the time between two subsequent false alarms is the same as the time until the first false alarm.
- For an out-of-control process, we adopt the commonly used assumption that the structural break occurs directly at the beginning of the monitoring (Frisén, 2003).

Therefore, we can restrict the definitions in this subsection to the case of a single structural break of size Δ.

Average run length

The most popular criterion for setting up and comparing control charts is the *average run length* (ARL). We define the *run length* as a random variable RL, which represents the number of analysed samples until the first alarm is given. The ARL is a function $\mathrm{ARL}\colon \mathbb{R} \to \mathbb{R}^+$ that maps the true shift height Δ to the expected value of RL. Thus, $\mathrm{ARL}(\Delta) = \mathrm{E}_\Delta(RL)$, where E_Δ denotes the dependence of the expectation on Δ. For $\Delta = 0$, the ARL is called *in-control ARL* and abbreviated by ARL_0. It corresponds to the expected duration until the first false alarm. If $\Delta \neq 0$, the ARL measures the expected detection delay, commonly abbreviated by ARL_1. The dependence of the ARL on the shift height is suppressed in the following for ease of notation. For an in-control process, the ARL should be large, but small for an out-of-control process.

The control limits of a control chart are typically chosen such that it achieves a desired ARL_0. The out-of-control performance is measured by the ARL_1. To compare different control charts, the charts are tuned to have the same ARL_0. They are then compared with respect to their ARL_1. A chart that minimises the ARL_1 for all $\Delta \in \mathbb{R}$ given a fixed ARL_0 is called *(ARL-)optimal*.

This is similar to the comparison of statistical hypothesis tests. The ARL_0 is an analogue to the significance level $\alpha \in (0, 1)$ and is connected to the type I error, which corresponds to a false rejection of the null hypothesis. Similarly, the ARL_1 can be understood as the power of the chart. The relationship between the ARL_0 and the ARL_1 can be compared to the one between α and the power of a test. A large ARL_0 leads to wide control limits and reduces the number of false alarms at the cost of a lower detection quality. Tuning a control chart to a small ARL_0-value results in a narrow control region and hence improves the ability to detect change points. However, false alarms are more likely (Montgomery, 2009, p. 189, Aroian and Levene, 1950).

Analogously to statistical tests, we call a control chart *conservative* if its true ARL_0 is larger than the desired value. This means that the duration until a false alarm is given exceeds the nominal ARL_0 and leads to a smaller number of false alarms than intended. If the ARL_0 falls below the specified value, we call the chart *anti-conservative*. Then, it is likely to get more false alarms than the chart is designed for. Analogously to hypothesis tests, conservative control charts can be less powerful than charts that attain the desired value or are anti-conservative.

The ARL is often criticised as a performance measure because the run-length distribution of an in-control process is skewed to the right. Hence, the ARL_0 might not appropriately represent it (Montgomery, 2009, p. 192). Furthermore, the skewness of the run-length distribution can depend on the shift height. For example, with increasing Δ, the distribution

may become more symmetric. By this, the interpretation of the ARL_0 is different from the one of the ARL_1 (Gan, 1993).

Median run length

A commonly proposed alternative to the ARL is the *median run length* (MRL), which is given by the median of the random variable RL, so that $\mathrm{MRL} = \mathrm{median}_\Delta(RL)$. A specific value of the MRL means that, with a probability of 50%, the run length will be smaller, regardless of whether the process is in- or out-of-control. Such an interpretation is not possible for the ARL (Gan, 1993; Chakraborti, 2007). We denote the *in-control MRL* by MRL_0 and the *out-of-control MRL* by MRL_1. Setting up and comparing control charts with respect to the MRL is analogous to the ARL.

Optimality criterion based on ARL_0 and MRL_1

In this thesis, we combine ARL_0 and MRL_1 to a new criterion. For an in-control process, we design the control charts to keep a specific ARL_0. To compare the out-of-control performance, we use the MRL_1. This is because, for an in-control process, the ARL_0 gives more information on the frequency of false alarms, since it is more easily influenced by large values than the MRL_0. The latter can be misleading for in-control processes. Thus, we aim at control charts which have a *minimal MRL_1 for a fixed ARL_0* for all $\Delta \in \mathbb{R}$.

To motivate the criterion, we consider two control charts. Both achieve the run length $RL = 100$ with probability 60%. The first chart has a probability of 40% for $RL = 1$, the second for $RL = 99$. Therefore, they have the same $\mathrm{MRL}_0 = 100$. The first chart has $\mathrm{ARL}_0 = 60.4$, the second one $\mathrm{ARL}_0 = 99.6$. Thus, the ARL_0 reflects in-control differences between the charts better. In our intended applications, it is important that change points are detected quickly. Then, the probabilistic interpretation of the MRL is more informative. This is why we will compare the charts with respect to their MRL_1-values to assess the out-of-control performance.

However, using different criteria for setting up and comparing the charts can be problematic in some cases. The same ARL_0 for two charts does not mean that they have the same MRL_0. Hence, a small MRL_1-value can sometimes be explained by a small MRL_0. This is important when interpreting the results. In general, such differences will only have a large impact for small values of Δ. We study this in more detail in the simulations in Section 3.4.

When using a two-sample test as a control chart, the run length depends on the local significance level. We expect the ARL and the MRL to increase with decreasing value of α. We use the critical values of the local test as the control limits. A small value for α increases the distance between these control limits, so that the null hypothesis is less likely to be rejected.

There exists a variety of other performance metrics such as alarm probabilities (Woodall

and Montgomery, 2014). Here, we use the run length because the detection speed is our foremost interest.

3.3.2 Selected two-sample tests for the location problem

We consider two types of test statistics for the construction of the control charts: test statistics that standardise an estimator for the location difference by an appropriate scale estimator and rank statistics.

A well-known test following the first principle is the *two-sample t-test*, which is one of the most popular tests for the location problem described by the hypotheses in (3.1). We discuss its shortcomings and present some rank-based and robust alternatives in the sequel.

The two-sample t-test

The test statistic of the two-sample t-test is given by

$$T_t^{(t)} = \sqrt{\frac{h \cdot k}{h+k}} \cdot \frac{\hat{\Delta}_t^{(t)}}{\hat{S}_t^{(0)}}, \tag{3.2}$$

where

$$\hat{\Delta}_t^{(t)} = \overline{Y}_{t+} - \overline{Y}_{t-}$$

is the difference of the sample means

$$\overline{Y}_{t-} = \frac{1}{h}\sum_{i=1}^{h} Y_{t,i}^{-} \quad \text{and} \quad \overline{Y}_{t+} = \frac{1}{k}\sum_{j=1}^{k} Y_{t,j}^{+},$$

and

$$\hat{S}_t^{(0)} = \sqrt{\frac{1}{h+k-2} \cdot \left(\sum_{i=1}^{h} \left(Y_{t,i}^{-} - \overline{Y}_{t-} \right)^2 + \sum_{j=1}^{k} \left(Y_{t,j}^{-} - \overline{Y}_{t+} \right)^2 \right)}$$

is the pooled empirical standard deviation (Lehmann and Romano, 2005, p. 160f.). If the normality assumption is valid for both subwindows, $T_t^{(t)}$ follows a t-distribution with $h+k-2$ degrees of freedom under $H_{0,t}$. We call the control chart based on this definition of the t-test *ordinary t-chart* in the following.

The popularity of the t-test stems from the central limit theorem. It protects against non-normal distributions under $H_{0,t}$, given large subwindow sizes h and k, and an existing second moment of the data-generating distribution. However, non-normality can lead to a reduced power (Wilcox, 2003, p. 242). The test is also known to be vulnerable to outliers. Even a single outlier can lead to a violation of the significance level and a substantial loss of power (Fried and Gather, 2007).

Rank tests

Distribution-free alternatives to the t-test can be constructed with rank statistics. We denote the ranks of the observations in the full window $\boldsymbol{Y}_t^{(n)}$ by $R_{t,1}^-, \ldots, R_{t,h}^-, R_{t,1}^+, \ldots, R_{t,k}^+$. As the underlying distribution is assumed to be continuous, the probability of assigning the same rank to two observations is zero. Nevertheless, the observations are typically rounded in applications, so that some of them can be equal. We will assign the ranks randomly in such cases.

A well-known rank-based alternative to the t-test is the two-sample *Wilcoxon rank-sum test* (Hollander et al., 2013, p. 116). The test statistic is the sum of the ranks in the test window, given by

$$T_t^{(\mathrm{W})} = \sum_{j=1}^{k} R_{t,j}^+. \tag{3.3}$$

The null distribution can be derived by a permutation principle, so that it is independent of the data-generating distribution. The resulting control chart will be called *Wilcoxon chart* in the remainder of this thesis.

A quantity to compare the power of hypothesis tests is the *Pitman asymptotic relative efficiency* (PARE) (Hodges and Lehmann, 1956). It compares the sample sizes which are necessary so that both tests have the same power. A brief definition can be found in Appendix D.3. The PARE of the Wilcoxon test compared to the t-test is $3/\pi \approx 0.955$ under normality and never drops below 0.864. Thus, the loss in power when using the Wilcoxon test instead of the t-test under normality is quite small (Hodges and Lehmann, 1956).

Another rank-based test for the location problem is the *two-sample Median test*. The test statistic is the number of observations in the test window that are larger than the median of the full window, so that

$$T_t^{(\mathrm{M})} = \sum_{j=1}^{k} I_{\left(\frac{n+1}{2},\infty\right)}\left(R_{j,t}^+\right). \tag{3.4}$$

The test statistic follows a hypergeometric distribution under $H_{0,t}$ (Daniel, 1978, p. 76ff.). We refer to the resulting control chart as *Median chart*.

The PARE of the Median test compared to the t-test is $2/\pi \approx 0.637$ under normality (Mood, 1954). It can be more powerful under non-normal distributions. However, the small efficiency under the normal distribution indicates that the test may not have a convincing performance for distributions which have a shape that is similar to the normal distribution.

The null distributions of the Median test and the Wilcoxon test are discrete, making the tests conservative. To achieve exact significance levels, we use randomisation (Hájek, 1969, p. 24).

For small samples, Fried and Gather (2007) show that the Wilcoxon and the Median

test can be nearly as vulnerable against outliers as the t-test. They suggest a robustification in which the critical values and the observations in the test window are adjusted. This adjustment depends on the underlying data-generating distribution and the desired significance level. We will not consider this approach in the following, because we aim for methods that can reliably deal with unknown data-generating distributions. Moreover, for the adjustment, we would have to perform additional simulations for each considered value of the significance level in our simulations. This is quite time-consuming as opposed to the expected benefit.

Robust alternatives to the t-test

Fried and Dehling (2011) construct outlier-resistant alternatives to the t-test by replacing the sample means and the pooled empirical standard deviation with robust counterparts which we will describe in the following.

A natural way to achieve robustness is to replace the mean with the *sample median*; see Subsection 2.3.1. The location difference can be estimated by

$$\hat{\Delta}_t^{(\mathrm{MD})} = \widetilde{Y}_{t+} - \widetilde{Y}_{t-},$$

where

$$\widetilde{Y}_{t-} = \underset{i=1,\dots,h}{\mathrm{median}} \left(Y_{t,i}^-\right) \quad \text{and} \quad \widetilde{Y}_{t+} = \underset{j=1,\dots,k}{\mathrm{median}} \left(Y_{t,j}^+\right)$$

are the sample medians of $\boldsymbol{Y}_{t-}$ and $\boldsymbol{Y}_{t+}$. We consider two test statistics based on $\hat{\Delta}_t^{(\mathrm{MD})}$. For the first, the standard deviation within the samples is measured by

$$\hat{S}_t^{(1)} = 2 \cdot \mathrm{median} \left(|Y_{t,1}^- - \widetilde{Y}_{t-}|, \dots, |Y_{t,h}^- - \widetilde{Y}_{t-}|, |Y_{t,1}^+ - \widetilde{Y}_{t+}|, \dots, |Y_{t,k}^+ - \widetilde{Y}_{t+}|\right).$$

The second is the sum of the MADs of both subwindows; see Equation (2.2). We denote the MADs for the reference window by $\hat{\sigma}_{t-}^{(\mathrm{MAD})}$ and for the test window by $\hat{\sigma}_{t+}^{(\mathrm{MAD})}$. The scale estimator is

$$\hat{S}_t^{(2)} = \hat{\sigma}_{t-}^{(\mathrm{MAD})} + \hat{\sigma}_{t+}^{(\mathrm{MAD})}.$$

The test statistics are

$$T_t^{(\mathrm{MD1})} = \frac{\hat{\Delta}_t^{(\mathrm{MD})}}{\hat{S}_t^{(1)}} \quad \text{and} \quad T_t^{(\mathrm{MD2})} = \frac{\hat{\Delta}_t^{(\mathrm{MD})}}{\hat{S}_t^{(2)}}. \tag{3.5}$$

We call the corresponding control charts *MD1-chart* and *MD2-chart* in the following.

By being based on the sample median, the ARE between $\hat{\Delta}_t^{(\mathrm{MD})}$ and $\hat{\Delta}_t^{(t)}$ is also $2/\pi$ under normality.

A compromise between the efficiency of the sample mean and the robustness of the sample median is the *one-sample Hodges-Lehmann estimator* (HL1-estimator) (Hodges

and Lehmann, 1963). The HL1-estimators for the subwindows are defined as

$$\widehat{Y}_{t-} = \underset{1\le i<j\le h}{\text{median}} \left(\frac{Y_{t,i}^- + Y_{t,j}^-}{2} \right) \quad \text{and} \quad \widehat{Y}_{t+} = \underset{1\le i<j\le k}{\text{median}} \left(\frac{Y_{t,i}^+ + Y_{t,j}^+}{2} \right).$$

The magnitude of the location shift can be estimated by

$$\hat{\Delta}_t^{(\text{HL1})} = \widehat{Y}_{t+} - \widehat{Y}_{t-}.$$

The HL1-estimator has a breakdown point of about 0.293 (Serfling, 2011). Hodges and Lehmann (1963) show that its ARE compared to the sample mean is $3/\pi$ under normality. It never drops below 0.864 for symmetric distributions. This corresponds to the PARE between the t-test and the Wilcoxon test. Fried and Dehling (2011) suggest the scale estimators

$$\hat{S}_t^{(3)} = \underset{\substack{1\le i<j\le h \\ 1\le i'<j'\le k}}{\text{median}} \left(|Y_{t,i}^- - Y_{t,j}^-|,\ |Y_{t,i'}^+ - Y_{t,j'}^+| \right) \quad \text{and} \quad \hat{S}_t^{(4)} = \underset{1\le i<j\le n}{\text{median}} \left(|Z_{t,i} - Z_{t,j}| \right),$$

where $(Z_{t,1}, \ldots, Z_{t,n}) = \left(Y_{t,1}^- - \widetilde{Y}_{t-}, \ldots, Y_{t,h}^- - \widetilde{Y}_{t-}, Y_{t,1}^+ - \widetilde{Y}_{t+}, \ldots, Y_{t,k}^+ - \widetilde{Y}_{t+} \right)$.
The estimator $\hat{S}_t^{(3)}$ calculates the median of absolute pairwise differences within the subwindows, whereas $\hat{S}_t^{(4)}$ is the median of absolute pairwise differences within the joint sample after centring the random variables by the respective subwindow median. The test statistics are

$$T_t^{(\text{HL11})} = \frac{\hat{\Delta}_t^{(\text{HL1})}}{\hat{S}_t^{(3)}} \quad \text{and} \quad T_t^{(\text{HL12})} = \frac{\hat{\Delta}_t^{(\text{HL1})}}{\hat{S}_t^{(4)}} \tag{3.6}$$

and we call the control charts *HL11-chart* and *HL12-chart*.

Alloway and Raghavachari (1991) and Pappanastos and Adams (1996) describe control charts based on the HL1-estimator. However, the charts depend on phase-I knowledge and a distributional assumption.

The *two-sample Hodges-Lehmann estimator* (HL2-estimator) estimates Δ_t directly from the full window (Hodges and Lehmann, 1963). It is defined as the median of the pairwise differences between the samples and given by

$$\hat{\Delta}_t^{(\text{HL2})} = \underset{\substack{i=1,\ldots,k \\ j=1,\ldots,h}}{\text{median}} \left(Y_{t,i}^+ - Y_{t,j}^- \right).$$

The HL2-estimator can be combined with the scale estimators $\hat{S}_t^{(3)}$ and $\hat{S}_t^{(4)}$, so that the test statistics are

$$T_t^{(\text{HL21})} = \frac{\hat{\Delta}_t^{(\text{HL2})}}{\hat{S}_t^{(3)}} \quad \text{and} \quad T_t^{(\text{HL22})} = \frac{\hat{\Delta}_t^{(\text{HL2})}}{\hat{S}_t^{(4)}}, \tag{3.7}$$

leading to the *HL21-chart* and the *HL22-chart*. The HL2-estimator has the same ARE compared to $\hat{\Delta}_t^{(t)}$ as $\hat{\Delta}_t^{(\text{HL1})}$. Høyland (1965) compares $\hat{\Delta}_t^{(\text{HL1})}$ to $\hat{\Delta}_t^{(\text{HL2})}$. He shows that

their ARE equals 1 for symmetric distributions. However, for asymmetric distributions $\hat{\Delta}_t^{(\mathrm{HL2})}$ can be infinitely more efficient than $\hat{\Delta}_t^{(\mathrm{HL1})}$.

Fried and Dehling (2011) compare tests based on the aforementioned statistics in simulation studies. They find that the tests based on the HL-estimators have similar power as the Wilcoxon test under different distributions while being less vulnerable to outliers. Tests based on the HL2-estimator have benefits under very skewed distributions. The MD-tests provide even better robustness in the studies but suffer from a loss of power under normality and moderately heavy-tailed or skewed distributions. From their simulations, the authors conclude that efficient estimators for the location difference lead to efficient tests.

Control limits of the robust control charts

The finite-sample null distributions of the robust test statistics are unknown. Using the asymptotic distributions of the location-difference estimators, which is a normal distribution in each case, allows the construction of asymptotic tests which are approximately distribution free. From their simulation studies, Fried and Dehling (2011) conclude that the sample sizes for the subwindows should be at least $h = k = 30$ to keep a significance level of $\alpha = 0.05$ under several distributions. We will not investigate this approach further. The local significance level determines the run length of the resulting control chart. In general, much smaller values than $\alpha = 0.05$ are needed to obtain reasonably large values for the ARL_0. Thus, we would need much larger windows to get a reliable test and a control chart that has the desired ARL_0, imposing the risk of violating the assumption of a locally constant signal.

Fried and Dehling (2011) discuss using a permutation principle. This would lead to distribution-free local tests. The general idea is to compute the value of the test statistic for all possible $\binom{n}{k}$ splits of the window $\boldsymbol{Y}_t^{(n)}$ into two subwindows to obtain a permutation distribution. Under $H_{0,t}$ the splits are equiprobable, so that the critical values of the local tests can be obtained as quantiles from the permutation distribution. Permutation tests can be extremely powerful. For example, using the permutation distribution instead of the t-distribution for the t-statistic in Equation (3.2) leads to a test that is as powerful under normality as the ordinary t-test (Lehmann, 2009).

The necessity to compute the permutation distribution for each new time point is a serious deficiency, especially when the data are gathered with a high frequency. For example, even for the quite small subwindow widths $h = k = 10$, there exist 184 756 splits for which the value of the test statistic has to be computed.

A commonly used simplification is the randomisation principle, not to be confused with randomisation to achieve an exact test. Instead of using all possible splits, a random subset of size $b << \binom{n}{k}$, $b \in \mathbb{N}$, is drawn. The observed sample is added to this set. The splits can be randomly drawn with or without replacement. Some authors, for example

Ernst (2004), do not distinguish between both ways. Others, such as Phipson and Smyth (2010), point out that drawing with replacement can lead to slightly conservative tests because a single split can occur multiple times. This seems to be particularly relevant in multiple-testing scenarios where it is important to maintain a global significance level. In our case, this is negligible because we use the ARL_0. Thus, we concentrate on drawing with replacement, which is computationally faster because we do not have to check whether a split has already been drawn.

Although a randomisation test for a single time window improves the speed substantially, using the test sequentially as a control chart can still lead to unacceptable computation times. That is why we consider two alternative strategies which avoid the calculation of the null distribution with each new incoming observation.

A first strategy is based on simulating the null distribution under a distributional assumption. We draw $N \in \mathbb{N}$ random samples of size $n = h + k$ from the assumed distribution. The value of the test statistic is computed for each sample. Critical values are chosen as appropriate quantiles from the simulated distribution. This approach has the benefit that, once computed, the same critical values can be used as control limits for every time series. However, this comes at the cost that the local tests are not distribution free. Nevertheless, this approach provides a simple possibility to obtain a distribution of the test statistics. In the following, we use the standard normal distribution and refer to the resulting control charts as *simulative charts*.

The sample mean, the sample median, and the one-sample Hodges-Lehmann estimator are location and scale equivariant. Thus, the estimators $\hat{\Delta}_t^{(t)}$, $\hat{\Delta}_t^{(\mathrm{MD})}$, and $\hat{\Delta}_t^{(\mathrm{HL1})}$ are invariant to location changes and scale equivariant. This is also true for $\hat{\Delta}_t^{(\mathrm{HL2})}$. Moreover, all scale estimators are scale equivariant and location invariant. Hence, all test statistics are invariant to linear transformations. The definitions of these terms are given in Appendix D.4. Thus, using the standard normal distribution imposes no restriction on the simulation of the critical values compared to an arbitrary normal distribution.

Another approach is a simplification of the randomisation principle. The randomisation distribution is computed only for the first sample $Y_1, \ldots, Y_n$. The resulting critical values are used for each of the following test decisions on the time series. Implicitly, the underlying assumption is that all observations in the time series come from the same time-invariant distributional class. Moreover, the critical values calculated for the first sample must be representative for all following samples. These assumptions might seem counter-intuitive given our criticism on ordinary control charts that need historical in-control data. However, the sample size used for computing the critical values corresponds to the width of the time window and is therefore rather short. Moreover, we can use the distribution to perform the test on the first sample. Additionally, the reference sample, with which we compare the test sample, is not fixed. We need the critical values only to decide whether the location difference is large enough to indicate a structural break between the test and the reference window. The control charts using this principle are called *simplified-randomised control*

charts from now on.

To illustrate the benefits of the simplified randomisation and the simulative approach over the permutation principle and the ordinary randomisation, we conduct a small simulation study using the t-statistic. The charts are applied to ten time series of length 1 000 from a standard normal distribution with $h = k = 10$, leading to 981 tests on each time series. For the randomisation principles, we use $b = 10\,000$. To compute the necessary splits for the permutation principle, we use the function `combinations` from the `R` package `gtools` (Warnes et al., 2015). The simulative distribution is computed from $N = 100\,000$ samples. The permutation principle needs about nine hours on average to perform all tests. Ordinary randomisation reduces this time to nearly thirty minutes, whereas the simplified randomisation principle and the simulative approach need less than five seconds.

Even though the computing time with the ordinary randomisation principle does not seem to be too large, it has to be considered that we analyse tens of thousands of time series in our in- and out-of-control analyses in Section 3.4. Although we use a high-performance computing cluster, the number of long-running jobs would be enormous due to many parameter combinations of h, k, α, different data-generating distributions, and test statistics. Moreover, the robust test statistics are based on the sample median and thus require more computing time than the t-statistic. Therefore, we only use the simplified randomisation principle and the simulative approach in the following so that the simulations can be finished within a reasonable time frame.

A randomisation distribution is discrete. Similar to the rank tests, randomisation can be used to achieve exact significance levels. According to Fried and Dehling (2011), this is not necessary, since the probability of two different splits leading to the same realised values of the test statistic is quite small.

3.4 Simulations

In this section, we present and discuss the results of several simulation studies in which we compare control charts based on the two-sample tests introduced in Subsection 3.3.2 in a variety of in- and out-of-control scenarios. Table 3.1 shows the test statistics used in the studies and by which principle the control limits are computed.

For the simplified randomisation principle, we use $b = 10\,000$ randomly drawn splits. The control limits of the simulative charts are computed from $N = 100\,000$ samples of size n from a standard normal distribution. As subwindow widths for our studies, we choose $(h, k) = (10, 10), (20, 20), (20, 10)$. In the main part of this section, we show the results for $h = k = 10$ and present those for $h = k = 20$ and $h = 20,\ k = 10$ in the Appendix. We keep the description of our observations as general as possible and comment on differences between the individual subwindow widths explicitly.

The simulation data are generated according to the additive components model given in Equation (2.1). We use the following distributions for the noise $(\varepsilon_t \colon\ t \in \mathbb{N})$:

Table 3.1: Test statistics and principles to compute the control limits used in the simulation studies.

Test statistic	Equation	Control limits		
		Ordinary	Simulation under $\mathcal{N}(0,1)$	Simplified randomisation
t-test	(3.2)	✓	✓	✓
Wilcoxon test	(3.3)	✓		
Median test	(3.4)	✓		
MD-tests	(3.5)		✓	✓
HL1-tests	(3.6)		✓	✓
HL2-tests	(3.7)		✓	✓

- Standard normal distribution (abbreviation: $\mathcal{N}(0,1)$),
- t-distributions with 5 and 2 degrees of freedom (abbreviations: t_5 and t_2),
- χ^2-distributions with 3 and 1 degrees of freedom (abbreviations: χ^2_3 and χ^2_1).

The χ^2-distributions do not have expectation zero. Due to the location invariance of the test statistics, this does not influence the local test decisions. The t- and the χ^2-distributions represent departures from the normality assumption in terms of heavy-tailed or skewed distributions. We use rather extreme cases. By this, we implicitly study the effect of outliers in an in-control process. The variance of the t_2-distribution does not exist, so that the requirements of the central limit theorem are not fulfilled.

If not mentioned otherwise, we generate 10 000 time series for each simulation setting. A detailed description of how a single time series for studying a specific question is constructed is given in the corresponding subsection. The ARL- and the MRL-values are estimated from the run lengths on the simulated time series by the sample mean and the sample median.

In Subsection 3.4.1, we investigate the in-control performance of the control charts under the normal distribution and describe the functional relationship between the ARL_0 and the significance level α. Afterwards, in Subsection 3.4.2, we study the impact of non-normality on the ARL_0 and comment briefly on the MRL_0. The control charts are compared with respect to their detection quality in Subsection 3.4.3. We conclude this section with a short investigation of how the out-of-control performance is influenced by outliers.

3.4.1 In-control comparison under the $\mathcal{N}(0,1)$-distribution

To achieve a desired ARL_0, denoted by ARL_0^* in the following, the value of the corresponding significance level α^* for the local tests needs to be specified. Except for distribution-free charts, the relationship between the ARL_0 and α depends on the data-generating distribution. In practice, it is generally unknown. Later in this work, we use the $\mathcal{N}(0,1)$-distribution to select the value of α^* because, as will be motivated, a distributional

assumption is inevitable for most of our control charts. For a distribution-free chart, this is no restriction, since we obtain the same ARL_0 for a fixed value of α under every distribution. In Subsection 3.4.2, we study the loss we have to expect when the true distribution is not normal.

The leading question in this subsection is how the functional relationship between the ARL_0 and the significance level α can be modelled if the data stem from a normal distribution.

We generate time series of length 20 000 under the assumption that $\mu_t = 0$ for all $t = 1, \ldots, 20\,000$. Because of the location invariance of the test statistics, this is no restriction compared to an arbitrary constant value for μ_t. Hence, a single simulated time series follows the model

$$Y_t = \varepsilon_t, \ t = 1, \ldots, 20\,000,$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $t = 1, \ldots, 20\,000$. The charts are applied with local significance levels $\alpha = 0.0025, 0.005, \ldots, 0.015, 0.02, \ldots, 0.05$.

Although the probability for a finite run length is 1, it is possible that no alarm is given within the considered time range. We replace such missing run lengths with $20\,001 - n + 1$, which is a lower bound for the true run length. We do analogous replacements in the other considered settings in the remainder of this chapter. In general, this is needed in less than 1% of the simulation runs, so that the effect of underestimating the true ARL_0 is negligible.

We first study the ARL_0 for $\alpha = 0.005, 0.02, 0.05$. Table 3.2 shows the ARL_0-values and the standard errors for the subwindow widths $h = k = 10$. The corresponding results for $h = k = 20$ and $h = 20,\ k = 10$ can be found in the Tables C.1 and C.2 in Appendix C. Our main observations are the following:

- For the considered equally-sized subwindows, the robust simplified-randomised charts lead to a higher uncertainty than the simulative versions. Especially the simplified-randomised MD-charts stand out. The estimates of the simplified-randomised HL-charts become more precise when increasing the value of α or the window widths. For the simplified-randomised MD-charts, the standard errors become smaller when the window widths are unequal. They are still larger than for the other charts but do not exceed the values as much as for the equal subwindow widths.

- Taking the standard errors into account, the ARL_0-values for the t-charts, the Wilcoxon chart and the simulative HL-charts are quite similar for the considered values of α, including those not displayed. The ARL_0-values of the Median chart are generally the smallest among the considered charts.

- Under subwindows of equal size, the simplified-randomised HL-charts have somewhat larger ARL_0-values than the ordinary t-chart but approach its values when the subwindow widths or α increase. Other than for unequal window widths, there is

Table 3.2: ARL_0 for selected values of α under normality for the subwindow widths $h = k = 10$. The values in brackets are the standard errors. All values are rounded to one decimal place.

Control chart		Significance level α		
		0.005	0.02	0.05
t-chart	ordinary	338.2 (3.4)	100.8 (1.0)	44.9 (0.5)
	simpl. random.	343.6 (3.6)	101.2 (1.0)	45.0 (0.5)
	simulative	341.2 (3.5)	102.7 (1.1)	44.9 (0.5)
HL11-chart	simpl. random.	405.1 (8.7)	109.3 (2.4)	45.5 (0.6)
	simulative	352.3 (3.6)	102.2 (1.0)	45.6 (0.5)
HL12-chart	simpl. random.	398.0 (7.8)	106.2 (1.6)	45.3 (0.5)
	simulative	342.7 (3.5)	104.1 (1.1)	45.5 (0.5)
HL21-chart	simpl. random.	426.4 (10.0)	110.5 (1.5)	46.3 (0.5)
	simulative	362.8 (3.7)	103.3 (1.1)	45.3 (0.5)
HL22-chart	simpl. random.	408.8 (8.8)	109.2 (1.5)	45.8 (0.5)
	simulative	365.6 (3.8)	104.3 (1.1)	45.2 (0.5)
MD1-chart	simpl. random.	504.4 (16.9)	156.0 (6.6)	63.0 (3.5)
	simulative	380.8 (3.9)	112.3 (1.1)	47.0 (0.5)
MD2-chart	simpl. random.	497.6 (17.2)	159.2 (6.7)	64.1 (3.1)
	simulative	385.6 (3.9)	113.0 (1.2)	47.7 (0.5)
Rank chart	Wilcoxon	334.6 (3.4)	100.3 (1.0)	44.1 (0.5)
	Median	269.4 (2.7)	97.7 (1.0)	33.6 (0.3)

no big difference in the results between the selected scale estimators. For unequal window widths, the scale estimator $\hat{S}_t^{(4)}$ leads to ARL_0-values that are nearer to those of the ordinary t-chart than $\hat{S}_t^{(3)}$, but are slightly smaller. The differences between the scale estimators become smaller when α increases.

- The simplified-randomised MD-charts lead for most considered values of α to the largest ARL_0-values for equally-sized subwindows. In contrast, their ARL_0-values are, in general, among the smallest for $h = 20,\ k = 10$.

- Except for the robust simplified-randomised charts, the ARL_0-values increase with the width of the full window. For the former, this seems to be only the case when both subwindow widths are equal.

A possible reason for the observed similarities between the t-chart and the Wilcoxon chart could be that they are based on similarly efficient tests. The simulative t-chart has the same distributional assumption as the ordinary t-chart, which explains why the results are nearly equal in all considered cases. This could also be a reason why the randomised t-chart performs similarly to the ordinary t-chart. Due to the high efficiency of the sample mean under the normal distribution, the control limits computed from the first window

Table 3.3: ARL_0 for selected values of α under normality for the subwindow widths $h = k = 10$ when using the ordinary randomisation principle. The values in brackets are the standard errors. All values are rounded to one decimal place.

	Significance level α		
Control chart	0.005	0.02	0.05
HL12-chart	329.4 (3.3)	100.1 (1.0)	44.4 (0.5)
HL22-chart	328.7 (3.3)	99.7 (1.0)	43.7 (0.5)
MD2-chart	312.7 (3.1)	95.9 (1.0)	42.0 (0.4)

are very representative for those that would be obtained from the randomisation principle when performed for later tests on the time series.

The HL-estimators for shift are also known to be very efficient under normality (Hodges and Lehmann, 1963), which could be an explanation of why the simulative HL-charts perform similarly to the t- and the Wilcoxon chart. The simulative MD-charts, by being based on a less efficient estimator under normality, need larger sample sizes to lead to similar results.

Different efficiencies could also be the reason for our observations on the simplified-randomised robust charts. For the subwindow widths $h = k = 10$, we consider 10 000 random splits, which is about 5% of the possible splits. Together with the quite small subwindow widths, the control limits from the first sample might not be very representative, especially for the MD-charts. However, increasing the subwindow widths to $h = k = 20$ leads, at least for the HL-charts, to ARL_0-values and standard errors that are comparable to those of the t-charts. Although the relative amount of splits is smaller than for $h = k = 10$, we can expect more reliable control limits due to the larger sample sizes. For $h = 20,\ k = 10$, one reason for the simplified-randomised HL12- and HL22-chart to have ARL_0-values that are closer to the one of the t-charts could be that the scale estimator $\hat{S}_t^{(4)}$ uses a larger set of observations than $\hat{S}_t^{(3)}$, improving its efficiency.

For comparison, we also computed the ARL_0 and the standard errors for the HL12-, the HL22-, and the MD2-chart, when using the ordinary randomisation principle for $\alpha = 0.005, 0.02, 0.035, 0.05$ and $h = k = 10$. The results for $\alpha = 0.005, 0.02, 0.05$ are shown in Table 3.3. The ARL_0-values of each chart deviate by less than 10% from those of the t-charts and the Wilcoxon chart. Furthermore, the standard errors are much smaller than for the simplified randomisation.

Simplified randomisation increases the uncertainty because it uses a comparatively small set of splits to compute the control limits. Particularly for small values of α, the limits are affected by quite large variabilities. This makes a good approximation for the critical values of the local tests difficult, especially when the test statistics lack efficiency.

Based on our findings, it does not seem to be possible to describe the functional relationship between the ARL_0 and α for all charts equally well by a single function. It

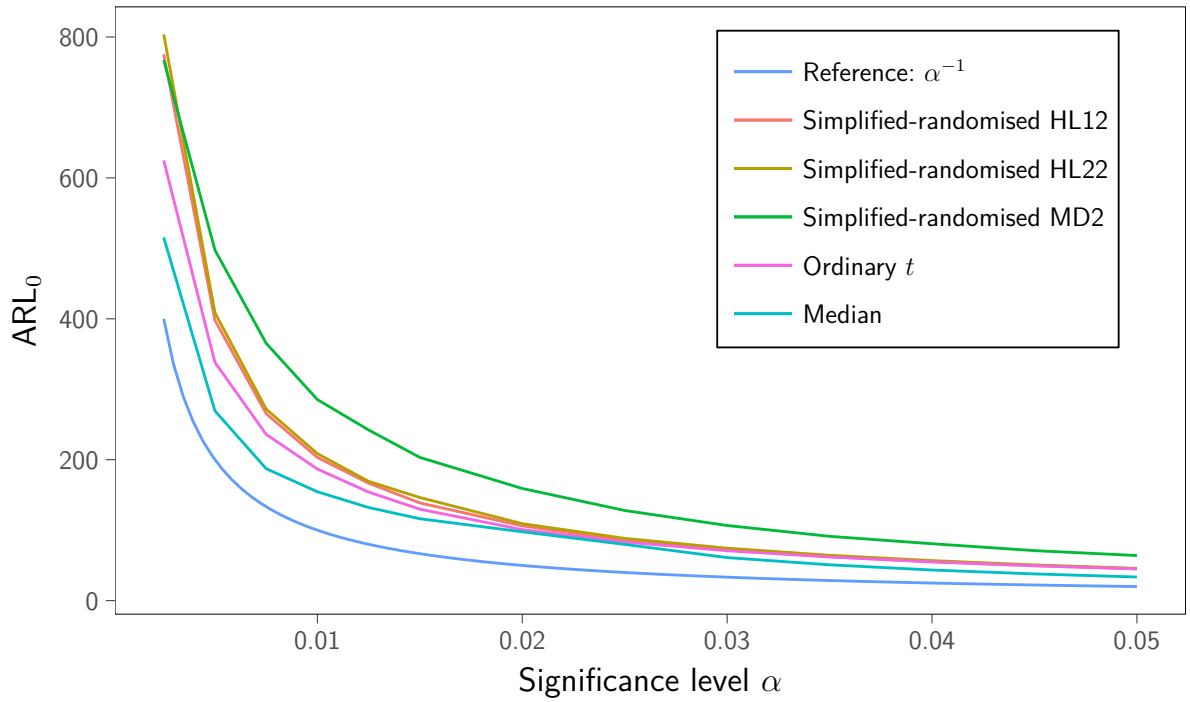

Figure 3.1: ARL_0-curves of selected control charts under normality with subwindow widths $h = k = 10$. For reference, the curve of the function α^{-1} is shown.

has to be specified for each control chart separately. If the samples were non-overlapping, the run length would follow a geometric distribution with success probability α for an in-control process, so that $\mathrm{ARL}_0 = \alpha^{-1}$ (Page, 1954).

Figure 3.1 illustrates the estimated ARL_0 for selected control charts with subwindow widths $h = k = 10$ together with the function α^{-1} as a reference. The dependence induced by using overlapping samples causes the ARL_0 to be larger than for non-overlapping samples. However, the shapes of the curves are similar to the reference curve. This leads to the impression that the relationship between the ARL_0 and α for our control charts can be generalised by

$$\mathrm{ARL}_0 = \gamma_0 \cdot \alpha^{\gamma_1}, \ \gamma_0 > 0 \text{ and } \gamma_1 < 0. \tag{3.8}$$

For the sake of a less complex notation, we suppress the dependency of Equation (3.8) on the subwindow widths and the control chart. In case of non-overlapping samples, the parameters would be $\gamma_0 = 1$ and $\gamma_1 = -1$.

Equation (3.8) can be linearised by applying the logarithm to both sides, so that

$$\log\left(\mathrm{ARL}_0\right) = \log\left(\gamma_0\right) + \gamma_1 \cdot \log\left(\alpha\right). \tag{3.9}$$

Then, we can determine α^* for a desired ARL_0^* by

$$\alpha^* = \exp\left(\frac{\log\left(\mathrm{ARL}_0^*\right) - \log\left(\gamma_0\right)}{\gamma_1}\right). \tag{3.10}$$

Table 3.4: Estimated regression coefficients for the linearised relationship between the ARL_0 and α in Equation (3.9) under normality for the subwindow widths $h = k = 10$, rounded to two decimal places. The values in brackets are the standard errors and R^2 denotes the coefficient of determination, both rounded to four decimal places.

Control chart		$\widehat{\log(\gamma_0)}$	$\hat{\gamma}_1$	R^2
t-chart	ordinary	1.19 (0.0079)	-0.88 (0.0019)	0.9999
	simpl. random.	1.16 (0.0084)	-0.89 (0.0020)	0.9999
	simulative	1.17 (0.0147)	-0.88 (0.0035)	0.9998
HL11-chart	simpl. random.	0.97 (0.0150)	-0.95 (0.0036)	0.9998
	simulative	1.18 (0.0107)	-0.88 (0.0026)	0.9999
HL12-chart	simpl. random.	0.96 (0.0130)	-0.95 (0.0031)	0.9999
	simulative	1.19 (0.0116)	-0.88 (0.0028)	0.9999
HL21-chart	simpl. random.	0.95 (0.0142)	-0.96 (0.0034)	0.9999
	simulative	1.15 (0.0152)	-0.89 (0.0036)	0.9998
HL22-chart	simpl. random.	0.98 (0.0149)	-0.95 (0.0036)	0.9998
	simulative	1.13 (0.0188)	-0.90 (0.0045)	0.9997
MD1-chart	simpl. random.	1.61 (0.0969)	-0.87 (0.0232)	0.9922
	simulative	1.15 (0.0238)	-0.91 (0.0057)	0.9996
MD2-chart	simpl. random.	1.68 (0.0879)	-0.85 (0.0210)	0.9934
	simulative	1.19 (0.0254)	-0.90 (0.0061)	0.9995
Rank chart	Wilcoxon	1.18 (0.0099)	-0.87 (0.0024)	0.9999
	Median	0.99 (0.1125)	-0.88 (0.0269)	0.9899

We estimate the unknown parameters $\log(\gamma_0)$ and γ_1 by ordinary least squares regression with Equation (3.9) from our simulation results.

Table 3.4 shows the parameter estimates, the standard errors and the coefficient of determination for the subwindow widths $h = k = 10$. The corresponding values for the subwindow widths $h = k = 20$ and $h = 20,\ k = 10$ can be found in Table C.3 in Appendix C. For all considered cases, the coefficient of determination is larger than 0.98, indicating a very good approximation of the relationship. Over all considered subwindow widths, the Median chart leads to the largest standard errors. Thus, it might be more difficult to estimate the significance level reliably.

Like the ARL_0-values discussed before indicate, the estimates show that some control charts have a similar functional relationship between the ARL_0 and α.

For an impression of how good the approximation actually is, we estimate α^* for $\mathrm{ARL}_0^* = 250,370$ from our simulation results. The value $\mathrm{ARL}_0^* = 370$ is often used in SPC (Montgomery, 2009, p. 346). The control charts are applied with the calculated values of α^* to 10 000 new time series of length 20 000 from a $\mathcal{N}(0,1)$-distribution. Figure 3.2 shows the asymptotic 95%-confidence intervals obtained from the simulation results for

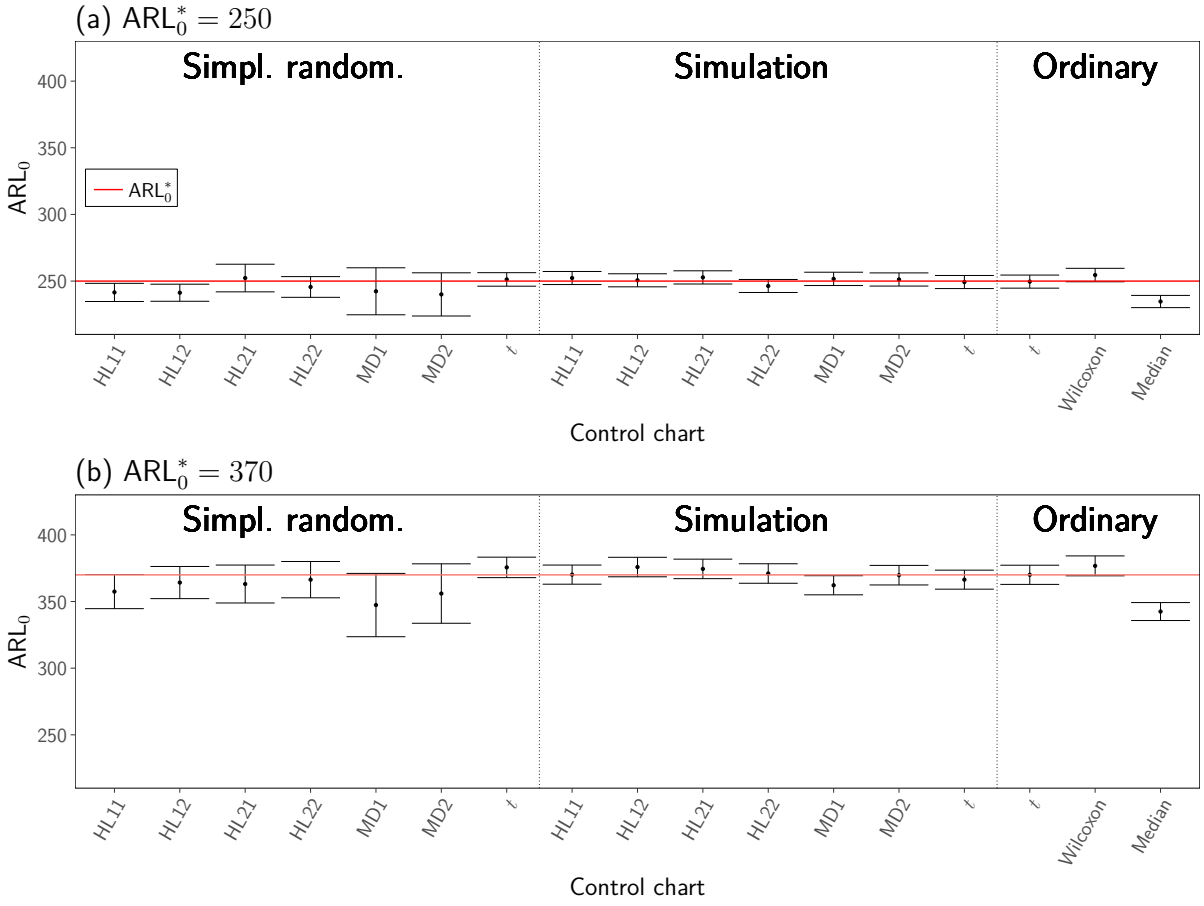

Figure 3.2: Asymptotic 95%-confidence intervals for the ARL_0 after computing α^* from Equation (3.10) for the subwindow widths $h = k = 10$ under the normality assumption. The vertical dotted lines separate between the different groups of control charts.

the subwindow widths $h = k = 10$. The plots for the subwindow widths $h = k = 20$ and $h = 20,\ k = 10$ can be found in Figure B.1 in Appendix B. The general observations are similar among the different subwindow widths. For nearly all control charts, the confidence intervals cover ARL_0^*. Large differences can only be observed for the Median chart. For example, with $h = 20,\ k = 10$ and $ARL_0^* = 370$, the estimated ARL_0 for the Median chart is about 10% larger than ARL_0^*. This could originate in the large standard deviations of the regression estimates. The simplified-randomised robust charts lead to wider confidence intervals than the non-randomised charts, particularly for $h = k = 10$. This reflects the larger standard errors of the estimated ARL_0-values. The width of each confidence interval increases in ARL_0^* so that we have to expect a larger uncertainty when we aim at a large ARL_0-value.

Concluding, the formula in Equation (3.10) seems well-suited to determine α^*. On the downside, we have to simulate the ARL_0 for several values of α because it is hardly possible to derive the relationship analytically. Especially for randomised charts, these simulations can be time-consuming. Given the similarities between the functional relationships for different charts, an approximation to a satisfying degree can be possible by using the

Table 3.5: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the t-charts with subwindow widths $h = k = 10$. All values are rounded to two decimal places.

	t-charts											
	ordinary				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.38	2.70	1.48	3.24	1.39	2.73	1.49	3.27	1.08	1.12	1.09	1.22
0.0050	1.29	2.28	1.43	2.66	1.29	2.29	1.43	2.68	1.06	1.12	1.09	1.20
0.0075	1.26	2.07	1.38	2.35	1.26	2.07	1.38	2.35	1.07	1.12	1.08	1.18
0.0100	1.21	1.90	1.31	2.10	1.22	1.91	1.31	2.12	1.05	1.10	1.08	1.16
0.0125	1.18	1.77	1.27	1.97	1.19	1.79	1.27	1.98	1.06	1.10	1.06	1.15
0.0150	1.18	1.72	1.27	1.90	1.19	1.74	1.27	1.91	1.05	1.10	1.04	1.15
0.0200	1.16	1.64	1.22	1.74	1.16	1.65	1.22	1.75	1.06	1.11	1.05	1.13

corresponding local significance level α^* of a suitable non-randomised chart for which the simulations can be performed much faster. For example, one could use α^* of the ordinary t-chart for the HL1- or the HL2-charts. In general, such approximations are better if ARL_0^* is small. Moreover, equally-sized windows seem to be beneficial for the approximation. Furthermore, we expect good results when the local tests are similarly efficient.

3.4.2 ARL_0 under non-normal distributions

We now investigate what happens to the ARL_0 when α is specified under the $\mathcal{N}(0,1)$-distribution, but the true distribution is non-normal.

We repeat the simulations as described at the beginning of Subsection 3.4.1 with noise from the t_5-, t_2-, χ^2_3-, and χ^2_1-distribution. To quantify the impact of non-normality, we divide the ARL_0-value of each non-normal distribution for each value of α by the corresponding ARL_0-value under the $\mathcal{N}(0,1)$-distribution. This allows us to investigate whether the ARL_0-value is larger or smaller than under the $\mathcal{N}(0,1)$-distribution, which is in accordance with the terms conservative and anti-conservative introduced in Subsection 3.3.1.

We only use selected distributions, so that we cannot generalise the results to all continuous distributions. Nevertheless, our results should provide a good intuition on how the ARL_0 of a control chart behaves under different distributional shapes.

The simulations were also performed for the t_{10}- and the χ^2_2-distribution. We found that each chart performs similarly under the t_{10}-distribution as under normality. Under the χ^2_2-distribution, the ARL_0-values lie between those for the χ^2_3- and the χ^2_1-distribution, which is why we consider only the less and the more extreme cases in the following.

We concentrate on $\alpha \leq 0.02$ as large ARL_0-values are often of interest in practical applications so that α is required to be small.

Table 3.5 displays the results for the t-charts with the subwindow widths $h = k = 10$. The ARL_0-ratios for $h = k = 20$ and $h = 20,\ k = 10$ can be found in Tables C.4 and C.5

Table 3.6: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the rank-based charts with subwindow widths $h = k = 10$. All values are rounded to two decimal places.

	Rank charts							
	Wilcoxon				Median			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.04	1.04	1.05	1.05	1.02	0.99	1.00	1.00
0.0050	1.02	1.01	1.03	1.02	0.99	1.00	1.01	1.00
0.0075	1.04	1.01	1.04	1.02	1.03	1.02	1.01	1.01
0.0100	1.03	1.01	1.03	1.02	1.00	0.99	1.01	0.98
0.0125	1.01	1.00	1.02	1.00	1.00	0.99	1.00	0.98
0.0150	1.03	1.00	1.03	1.01	0.99	1.00	1.00	1.00
0.0200	1.02	1.01	1.00	1.00	1.00	1.00	1.00	0.99

in Appendix C. Our main observations are as follows:

- In general, all three charts are conservative under the considered distributions so that we have to expect an ARL_0-value that is larger than the nominal value.
- The underlying normality assumption causes the charts based on the ordinary and the simulative t-test to perform worse in terms of conservatism than the simplified-randomised chart.
- Simplified randomisation protects comparatively well against moderate heavy-tailedness. For larger subwindow widths, the ARL_0 of the simplified-randomised t-chart even seems to be nearly distribution-free among the considered distributions.
- Longer subwindows also reduce the conservatism of the ordinary and the simulative t-charts, but they still perform worse than under simplified randomisation.

In Table 3.6, we show the ARL_0-ratios for the Wilcoxon and the Median chart for $h = k = 10$. The results indicate a nearly distribution-free ARL_0, which we also observe for the subwindow widths $h = k = 20$ and $h = 20,\ k = 10$. The following Proposition 1 states that the in-control run-length distribution of a control chart, which is based on a two-sample rank test, is indeed distribution free. Concerning the in-control performance, the rank charts represent the ideal case of a distribution-free control chart.

Proposition 1 (In-control run length of control charts based on rank tests)
Let $(Y_t\colon\ t \in \mathbb{N})$ *be a sequence of independent and identically distributed random variables following a continuous distribution with cumulative distribution function* F*. Let furthermore* $(T_t\colon\ t \geq n)$*,* $n \in \mathbb{N}$*, be a sequence of test statistics, where* T_t *is a rank statistic based on* $\boldsymbol{Y}_t^{(n)} = (Y_{t-n+1}, \ldots, Y_t)$*. Then, the in-control run length of a control chart based on this sequence of test statistics is distribution free.*

The proof is deferred to Appendix A. The main idea is that the ranks in the current sample are a transformation of the ranks in the time series up to the current time point.

In Table 3.7, we report the ARL_0-ratios for the control charts based on the robust test statistics with $h = k = 10$. For $h = k = 20$ and $h = 20,\ k = 10$, we refer to the Tables C.6 and C.7 in Appendix C. Our main observations can be summarised as follows:

- The ARL_0-values of the simulative charts depend heavily on the underlying distribution. Here, it is difficult to generalise over all considered subwindow widths in terms of conservatism and anti-conservatism. For example, the simulative HL21-chart is strongly anti-conservative under the χ^2_1-distribution for $h = k = 10$, but conservative for $h = k = 20$ for $\alpha \geq 0.01$.

- Simplified randomisation can improve the results substantially. Especially the HL22-chart benefits from this principle by having a nearly distribution-free ARL_0 among the considered distributions for all the three combinations of subwindow widths. We make a similar observation for the MD2-chart.

- However, simplified randomisation does not always ensure this property. Moreover, the behaviour of the charts can change from being conservative to becoming anti-conservative, or the other way around, with increasing value of α. This is, for example, the case for the HL1-charts under the χ^2-distributions.

In general, the results indicate that properties of the underlying tests transfer to the control charts. This explains why the t-charts are mainly conservative. Moreover, as stated in Subsection 3.4.1, the efficiency of the underlying test statistics seems to be important for a good in-control performance of the charts. This could explain the worse performance of the HL1-charts under skewed distributions compared to the HL2-charts. The HL1-estimator has a smaller efficiency. By being more efficient, the initial control limits computed for the HL2-charts may be less affected by random variations than those of the HL1-charts, leading to more representative control limits.

For comparison, we also computed the ratios for selected ordinary-randomised charts analogously to Subsection 3.4.1. The results are shown in Table 3.8. The ARL_0-values never deviate much more than 5% from the values obtained under normality. We cannot expect complete distribution independence of the in-control run length, since the randomised tests are not distribution-free themselves. However, the results indicate that approximate distribution independence is possible. Compared to these results, the performance loss of the simplified-randomised HL22- and MD2-charts seems to be acceptable, considering the enormous reduction of computing time.

Comment on the MRL_0

For the interpretation of the out-of-control results in the next subsection, we discuss briefly the behaviour of the MRL_0 under non-normality.

For the non-randomised charts, our conclusions are the same as those for the ARL_0. Moreover, the simplified-randomised MD- and HL2-charts lead, in general, to the same

Table 3.7: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the robust control charts with subwindow widths $h = k = 10$. All values are rounded to two decimal places.

	HL11-charts								HL12-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.26	1.13	0.34	0.08	0.97	0.84	0.85	0.55	1.32	1.40	0.54	0.14	0.98	0.87	0.86	0.34
0.0050	1.20	1.13	0.39	0.11	0.99	1.01	0.95	0.68	1.24	1.33	0.59	0.18	0.99	0.98	0.90	0.46
0.0075	1.19	1.11	0.43	0.14	1.00	1.09	0.99	0.79	1.20	1.27	0.61	0.21	1.03	1.05	0.93	0.57
0.0100	1.15	1.09	0.45	0.15	1.01	1.18	1.04	0.92	1.17	1.23	0.61	0.23	1.02	1.12	0.95	0.64
0.0125	1.14	1.07	0.47	0.17	1.03	1.23	1.08	0.97	1.17	1.20	0.62	0.24	1.01	1.18	0.94	0.71
0.0150	1.13	1.07	0.48	0.18	1.02	1.26	1.09	1.03	1.15	1.18	0.63	0.26	1.02	1.22	0.99	0.79
0.0200	1.11	1.07	0.50	0.20	1.04	1.36	1.13	1.11	1.12	1.16	0.64	0.27	1.02	1.29	1.02	0.90

	HL21-charts								HL22-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.39	1.99	0.61	0.34	1.03	0.94	1.08	1.04	1.46	2.40	0.96	1.16	1.00	0.94	1.03	1.11
0.0050	1.31	1.79	0.69	0.44	1.05	0.99	1.18	1.23	1.33	2.04	1.01	1.31	1.05	0.97	1.11	1.19
0.0075	1.26	1.69	0.75	0.51	1.05	1.00	1.25	1.43	1.28	1.87	1.04	1.44	1.05	1.00	1.14	1.28
0.0100	1.24	1.61	0.79	0.57	1.04	1.01	1.23	1.38	1.25	1.74	1.05	1.48	1.03	0.98	1.10	1.20
0.0125	1.23	1.56	0.81	0.62	1.02	1.00	1.11	1.11	1.24	1.68	1.07	1.53	1.04	0.98	1.04	1.10
0.0150	1.22	1.51	0.83	0.67	1.03	1.01	1.13	1.10	1.23	1.63	1.08	1.59	1.02	0.94	1.03	1.08
0.0200	1.19	1.45	0.86	0.74	1.04	1.02	1.07	1.13	1.20	1.55	1.10	1.63	1.04	0.98	1.02	1.08

	MD1-charts								MD2-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.33	1.99	0.51	0.17	0.97	0.95	0.92	0.91	1.43	2.46	0.81	0.62	0.98	0.98	0.93	0.92
0.0050	1.28	1.85	0.56	0.21	0.96	0.94	0.93	0.90	1.34	2.22	0.83	0.65	0.95	0.94	0.93	0.92
0.0075	1.26	1.75	0.58	0.24	0.96	0.96	0.95	0.94	1.33	2.08	0.84	0.68	0.96	0.94	0.94	0.96
0.0100	1.22	1.67	0.60	0.26	0.96	0.92	0.92	0.91	1.32	2.00	0.86	0.69	0.97	0.97	0.94	0.98
0.0125	1.22	1.63	0.62	0.28	0.98	0.91	0.90	0.87	1.30	1.92	0.85	0.69	0.94	0.93	0.90	0.96
0.0150	1.20	1.61	0.63	0.30	0.98	0.91	0.91	0.88	1.28	1.84	0.85	0.69	0.96	0.90	0.93	0.98
0.0200	1.18	1.55	0.66	0.32	1.02	0.89	0.93	0.91	1.25	1.79	0.85	0.69	0.96	0.86	0.92	0.98

impressions as based on the ARL_0. Notable differences occur for the simplified-randomised HL1-charts and the simplified-randomised t-chart, both of which tend to be strongly anti-conservative. For the simplified-randomised HL1-charts, this is the case under the t_2- and the χ^2-distributions. In case of the simplified-randomised t-chart, the chart is anti-conservative under the t_2-distribution. For both charts, the results can be improved by enlarging the subwindows to $h = k = 20$.

In total, control charts that show a desirable in-control performance with respect to the ARL_0 also show a good one for the MRL_0. In addition to the rank-based charts, these are the MD2- and the HL22-chart.

Table 3.8: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality when using the ordinary randomisation principle with subwindow widths $h = k = 10$. All values are rounded to two decimal places.

	Ordinary-randomised charts					
	HL12		HL22		MD2	
α	t_2	χ^2_1	t_2	χ^2_1	t_2	χ^2_1
0.0050	1.02	0.96	1.01	1.00	1.00	0.99
0.0200	0.99	0.94	1.00	1.01	0.99	1.00
0.0350	0.98	0.93	0.99	1.00	1.00	1.00
0.0500	0.98	0.92	1.00	1.01	1.00	1.00

3.4.3 Out-of-control analysis

We now evaluate the out-of-control performance of the control charts with respect to the detection speed, as measured by the MRL_1, and the detection rate, both for a fixed ARL_0.

The time series are generated from the model

$$Y_t = \begin{cases} \varepsilon_t, & t = 1, \dots, n \\ \varepsilon_t + \Delta \cdot q_{\text{diff}}, & t = n+1, \dots, 20\,001, \end{cases}$$

where q_{diff} is the difference between the 84.13%- and the 50%-quantile of the distribution of ε_t. For the $\mathcal{N}(0,1)$-distribution, its value is $q_{\text{diff}} = 1$. To improve the comparability between different distributions, we choose the jump heights as multiples $\Delta = 0.5, 1, 1.5, 2$ of q_{diff}. This difference is used as a substitute for the commonly used standard deviation, which does not exist for the t_2-distribution. Here, we concentrate on positive location shifts. Due to symmetry, the results for negative ones will be similar.

Figure 3.3 illustrates the construction of the time series. For given subwindow widths h and k with $h + k = n$, a sustained location shift starts at time point $t = n + 1$. The first n observations of the time series are used to compute the control limits for the simplified-randomised control charts from a clean sample that is not affected by the location shift. The first time window in the monitoring covers $(Y_2, \dots, Y_{n+1})$ so that only the rightmost observation in the time window is shifted.

The location shift can only be detected during the first $n - 1$ tests. Starting with the n-th test, the full time window is beyond the location shift. Thus, the control charts perform like in the in-control situation as the expected values for both subwindows are equal again. Every alarm can then be attributed to the type I error of the local test. Because of this, we do not use ordinary control charts for comparison in the following. They would have an unfair advantage, since they always compare the value of the control statistic to a fixed reference value. Although we study a rather simple scenario in which their requirements are fulfilled, we aim for applications where this is not the case. Thus, we concentrate on a comparison of the different test statistics described so far.

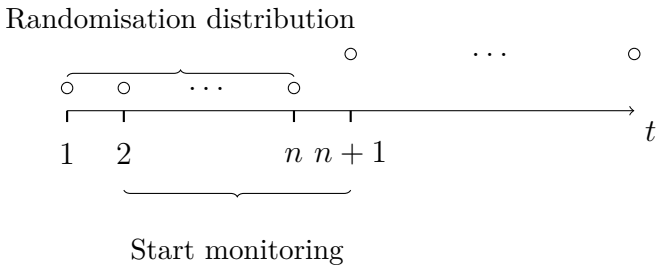

Figure 3.3: Schematic representation of the set-up for the out-of-control simulations. A persistent location shift starts at time $t = n + 1$. The first n observations are used to calculate the control limits for the simplified-randomised charts. The monitoring begins with the window $(Y_2, \ldots, Y_{n+1})$.

We select the local significance levels for the different charts to achieve $\mathrm{ARL}_0^* = 250, 370$ under normality using Equation (3.10) with our in-control-simulation results. However, even though the charts are tuned to have the same ARL_0^*, this does not imply that the corresponding MRL_0 is also equal. For example, for the subwindow widths $h = k = 10$ and $\mathrm{ARL}_0^* = 250$, the MRL_0-values of the simplified-randomised MD-charts are about 80, which is drastically smaller than the MRL_0-values of roughly 170 for the non-randomised charts and 140 for the simplified-randomised HL-charts. This leads to the following problem: As will be shown in the sequel of this subsection, small shifts are likely to be missed by a chart. With respect to the MRL_1, the simplified-randomised MD-charts are likely to cause alarms faster in such cases.

To quantify the detection speed of the control charts over a wide range of distributions, we define an efficiency measure which takes values in $[0, 1]$. The idea is to compare the MRL_1 of each control chart to the smallest MRL_1 in a set of distributions. Let F be the cumulative distribution function of a noise distribution and $\mathcal{C}$ the set of all considered control charts in our study. By $\mathrm{MRL}_1(C, \Delta, F)$ we denote the MRL_1 of a control chart $C \in \mathcal{C}$ for the jump-height factor Δ under the cumulative distribution function F. We define the *relative efficiency* (RE) of C for Δ under F by

$$\mathrm{RE}(C, \Delta, F) = \frac{\min\limits_{\tilde{C} \in \mathcal{C}} \mathrm{MRL}_1\left(\tilde{C}, \Delta, F\right)}{\mathrm{MRL}_1(C, \Delta, F)}.$$

A value close to 1 means that C is nearly as fast in detecting the structural break as the fastest chart in the set $\mathcal{C}$. Thus, RE reflects how much slower a control chart is compared to the fastest chart in $\mathcal{C}$, as measured by the MRL_1.

Our interest lies in finding control charts with a good overall performance. In a similar spirit as Morgenthaler and Tukey (1991, p. 69), we define a *minimal relative efficiency* (MRE), which reflects the worst-case behaviour of a control chart over a set $\mathcal{F}$ of cumulative

Table 3.9: Minimal MRL_1-values for different noise distributions, separated by shift-height factor Δ and nominal ARL_0^* under normality, over all control charts for the subwindow widths $h = k = 10$.

		Shift-height factor Δ			
ARL_0^*	Distribution	0.5	1	1.5	2
250	$\mathcal{N}(0,1)$	66	25	9	8
	t_5	63	16	9	8
	t_2	61	12	9	7
	χ^2_3	44	9	7	6
	χ^2_1	10	6	6	5
370	$\mathcal{N}(0,1)$	88	48	10	8
	t_5	87	33	9	8
	t_2	88	17	9	8
	χ^2_3	66	10	8	6
	χ^2_1	10	7	6	6

distribution functions and is given by

$$\mathrm{MRE}\left(C, \Delta, \mathcal{F}\right) = \min_{F \in \mathcal{F}} \mathrm{RE}\left(C, F, \Delta\right).$$

Here, the worst case is restricted to the distributions used in our simulations.

An aspect that has to be kept in mind when interpreting the MRE is that slight differences between small values are more emphasised than between large values. For example, if the minimal MRL_1 was 6 and a control chart had the MRL_1 8 for a specific value of Δ, the chart would have a relative efficiency of only 0.75. On the other hand, if the minimal MRL_1 was 60 and the chart had the MRL_1 58, the MRE would be 0.97.

We start by reporting the minimal MRL_1-values, separated by distribution, jump-height factor, and ARL_0^*. The results for $h = k = 10$ are shown in Table 3.9. The corresponding results for $h = k = 20$ and $h = 20,\ k = 10$ can be found in Tables C.8 and C.9 in Appendix C. These values are helpful for interpreting the MRE-values.

The minimal MRL_1-values for the skewed distributions are substantially smaller than for the symmetric ones, particularly for the χ^2_1-distribution. This can be attributed to the anti-conservatism of some procedures under skewed distributions, for example the HL1-charts. In general, the MRL_1-values decrease with increasing shift height. However, even for quite large jumps, a small delay is unavoidable. Even for values of Δ larger than $\Delta = 2$, there is not much room for improving the speed.

Therefore, anti-conservative charts may have a large impact on the MRE if we consider a group of distributions that includes asymmetric distributions. Thus, alarms cannot solely be attributed to a better detection quality of a chart. Instead, it is likely that they originate from the anti-conservatism. This will mainly affect small shift heights, where it is likely that the shifts are missed by a control chart. Moreover, apart from a possible

anti-conservatism, the charts are rather slow in detecting small changes. This is not a big problem for our intended applications, because we are foremost interested in finding structural breaks of large magnitudes that cannot be explained by the natural course of the signal.

The minimal MRL_1-values also emphasise the resistance of the charts to short-term location shifts of a certain length, making them insensitive to outlier patches. Our results for $h = k = 20$ and $h = 20,\ k = 10$ indicate that the minimal MRL_1 tends to become larger when increasing the size of the test window for large values of Δ.

In Figure 3.4, we show the MRE-values for the subwindow widths $h = k = 10$ and $\text{ARL}_0^* = 250$. The results for $\text{ARL}_0^* = 370$, $h = k = 20$, and $h = 20,\ k = 10$ can be found in the Figures B.2 to B.6 in Appendix B.

We separate the distributions into three classes: all considered distributions ($\mathcal{N}(0,1)$-, t_5-, t_2-, χ_3^2-, and χ_1^2-distribution), the symmetric distributions ($\mathcal{N}(0,1)$-, t_5-, and t_2-distribution), and the skewed distributions (χ_3^2- and χ_1^2-distribution).

With respect to the comments made on small jump heights, we concentrate on large location shifts with $\Delta \geq 1.5$. Moreover, we focus on the group containing all distributions, which is shown in panel (a) of the corresponding figures. The most important observations for $h = k = 10$ are:

- For large values of Δ, the different control charts show only marginal differences with respect to the detection speed; see Figure 3.4(a).
- The ordinary and simulative t-chart are negatively affected by the t_2-distribution, causing the MRE-values to be worse than for the other charts when considering only symmetric distributions; see Figure 3.4(c). Wilcoxon and Median chart show a weaker performance under the skewed distributions; see Figure 3.4(e).
- The simplified-randomised charts perform nearly equally well among all considered groups of distributions. Only the HL11-chart falls somewhat short compared to the others because of the asymmetric distributions.
- The robust simulative charts, particularly the HL-charts, also show an appealing behaviour over the different distributions.
- A larger ARL_0^* can decrease the performance because of the smaller value of the significance level that is selected.

Increasing the subwindow widths does not necessarily improve the MRE-results; see Figures B.3(a) and B.5(a) for $h = k = 20$ and $h = 20,\ k = 10$ with $\text{ARL}_0^* = 250$. All but the simulative HL11-chart seem to lose much of their efficiency. This is because of the strong anti-conservatism of this chart under the χ_1^2-distribution. Removing this distribution from the set leads to a similar impression as for $h = k = 10$.

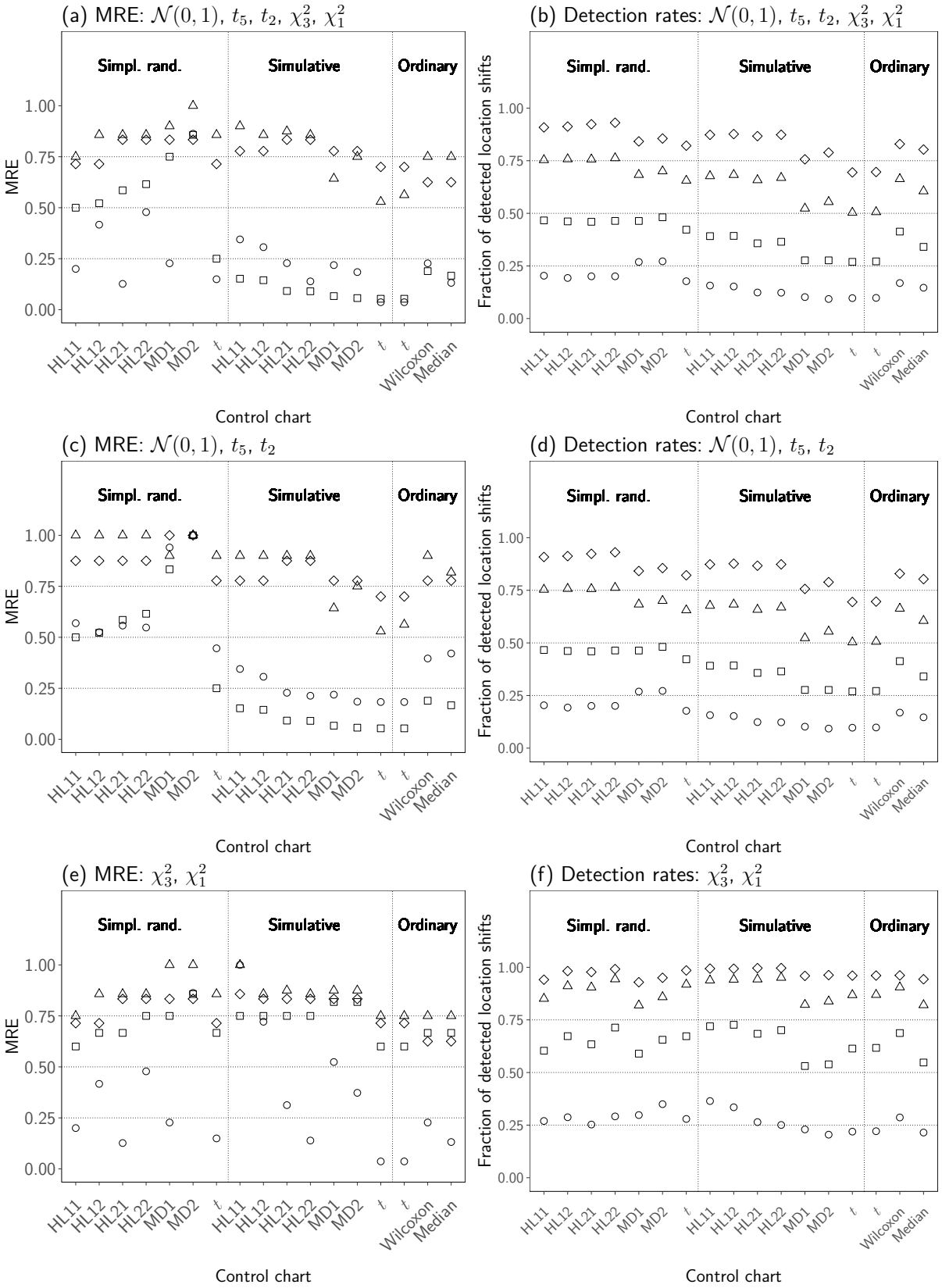

Figure 3.4: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality and subwindow widths $h = k = 10$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

As the MRE compares the MRL_1 of a specific control chart to the minimal MRL_1, the MRE-values are dominated by anti-conservative charts. For example, the simulative HL1-charts are very anti-conservative under skewed distributions, which reduces the MRE-values for other control charts. However, anti-conservatism is an undesirable property, particularly for in-control processes.

To put the MRE-values into perspective, we additionally consider worst-case detection rates. We define the *detection rate* by the fraction of time series where an alarm is given within the first $n-1$ tests. As this value is in $[0,1]$, we calculate the minimal detection rate over a set of distributions directly instead of comparing it to the best value. The results for $h = k = 10$ and $\mathrm{ARL}_0^* = 250$ are shown in the panels (b), (d), and (f) of Figure 3.4. The results for the other parameter choices are shown in the aforementioned figures in Appendix B. Our main observations for $h = k = 10$ are:

- The simplified-randomised HL-charts have the overall largest worst-case detection rates; see Figure 3.4(b). For $\Delta = 2$, they are close to 1, indicating that the change point is detected with a high probability. The simulative charts are not much worse.
- The simplified-randomised MD-charts are slightly worse than the HL-charts. However, the simplified randomisation still leads to somewhat better results than the simulative approach. The main reason for the weaker performance of the MD-charts is the $\mathcal{N}(0,1)$-distribution; see Figure 3.4(d).
- Among the ordinary charts, the Wilcoxon chart is marginally better than the Median chart. Both are somewhat superior to the ordinary t-chart when considering all distributions.
- The weak spots of the non-randomised charts, the simulative, and the simplified-randomised t-chart are the heavy-tailed distributions.
- Similar to the MRE-values, the detection rates decrease for a larger value of ARL_0^*.

The detection rates generally improve when the window widths increase, leading to the same ordering of the procedures as for $h = k = 10$.

How a distribution affects a chart's out-of-control performance depends on the local test. The simplified-randomised HL-charts provide good performance over a wide range of distributions, followed by the rank-based charts. Anti-conservatism can distort the results somewhat. Charts with anti-conservative behaviour are more likely to cause false alarms in the in-control setting. This is mainly the case for the simulative HL-charts. Nevertheless, they can be reasonable replacements for the simplified-randomised charts in case of symmetric distributions when simplified or ordinary randomisation is not feasible.

Comment on the ARL_1

The MRE can also be computed for the ARL_1. We discuss the results briefly, to gain a fuller picture of the performance of the charts. The Figures B.7 to B.9 in Appendix B depict the MRE-values of the charts with respect to the ARL_1 for $ARL_0^* = 250$ under normality and all three subwindow-width combinations. We focus again on the subwindow widths $h = k = 10$ at first.

Compared to the MRL_1, the MRE-values computed from the ARL_1 are drastically reduced for most charts. For the simplified-randomised charts, the ordinary t-, and the rank-based charts, they do not exceed 0.25 in most cases when using all noise distributions for the comparison. The simulative HL1-charts, on the other hand, have very large MRE-values of about 0.75 for all considered shift-height factors. This resembles the results for the MRL_1. Possible explanations are the following:

- Again, anti-conservatism of the simulative charts under the χ^2-distributions gives them an advantage over their competitors. This can be seen when looking at the results for the symmetric and the skewed distributions separately. Considering only the latter, the MRE-values are quite similar to those when taking all distributions into account. Computing the MRE only from the symmetric distributions improves MRE-values, especially for the simplified-randomised HL- and the rank-based charts.

- Another reason for the smaller MRE-values of the simplified-randomised charts are occasional very large run lengths in the simulation. The in-control run-length distribution is generally skewed to the right. This is particularly relevant when a structural break is missed because the setting then corresponds to the in-control case of equal levels in both subwindows. Under the normal distribution, for example, simplified randomisation causes longer run lengths than using an exact distribution; see Table 3.2. This is reflected in the ARL_1-values because the average of the simulated run lengths is easily affected by large values. For example if $\Delta = 2$, the ordinary t-chart misses the structural break in about 4% of the simulation runs, the simplified-randomised HL22-chart in nearly 6%, which is not much larger. However, for the HL22-chart, about 10% of the simulated run lengths in which the structural break is missed are larger than 1 000. For the ordinary t-chart, this is the case in only 2% of the simulation runs.

Enlarging the subwindow widths to $h = k = 20$ increases the MRE-values. For example, the HL2-charts perform now similarly well as with respect to the MRL_1. However, for small subwindow widths, the results obtained from the ARL_1 diminish the good impression the simplified-randomised HL-charts left with respect to the MRL_1. Given the explanations, the ARL_1-performance of these charts might be improved by modifying the randomisation principle.

In conclusion, both ARL and MRL have weak spots as representatives for the out-of-control performance of the charts. The ARL puts much weight on large run lengths, whereas the MRL favours charts with a smaller MRL_0 for a fixed ARL_0^*. Hence, to gain adequate insight into the overall performance of the control charts, multiple criteria should be considered.

3.4.4 Influence of outliers in the out-of-control setting

We will now briefly investigate if, and how strong, additive outliers can worsen the out-of-control performance of the control charts. Here, we use the subwindow widths $h = k = 10$ and $\mathrm{ARL}_0^* = 250$ under normality for an illustration. The evaluation is based on the MRL_1 and the detection rates in the following example.

We investigate a setting with a single positive location shift of a fixed size. The time series are generated from the model

$$Y_t = \begin{cases} \varepsilon_t, & t = 1, \ldots, 20 \\ \varepsilon_t + 3, & t = 21, \ldots, 20\,001, \end{cases}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. We consider two scenarios: one negative outlier at time $t = 26$ and two subsequent negative outliers at the time points $t = 26$ and $t = 27$. We use $\eta_t = -5, -10, -15, -20$, so that for two outliers their size is equal.

The selected jump height leads to detection rates of more than 0.95 and MRL_1-values smaller than 9 for each control chart. In the uncontaminated setting, we observe the highest alarm rates at the time points, at which we place the outliers.

From Figure 3.5, which illustrates the relative efficiencies and the detection rates for both scenarios, we make the following observations:

- All t-charts are heavily influenced even by a single outlier. Starting with $\eta = -10$, the detection rate never exceeds 0.25. Moreover, the MRE-values are less than 0.05.
- The rank-based charts can deal with a single outlier quite well. However, as soon as we add a second outlier, their performance also worsens. The Median chart shows a somewhat better performance than the Wilcoxon chart. This is because of its increased information reduction.
- All robust charts resist the outliers in both settings. For two outliers, the detection rates are marginally reduced because 20% of the observations in the reference window are contaminated.

This short study hints that charts based on robust test statistics also provide robustness against outliers in an out-of-control setting. An improvement is possible by increasing the subwindow widths. These observations coincide with those made by Fried and Dehling (2011) in their studies of the underlying two-sample tests.

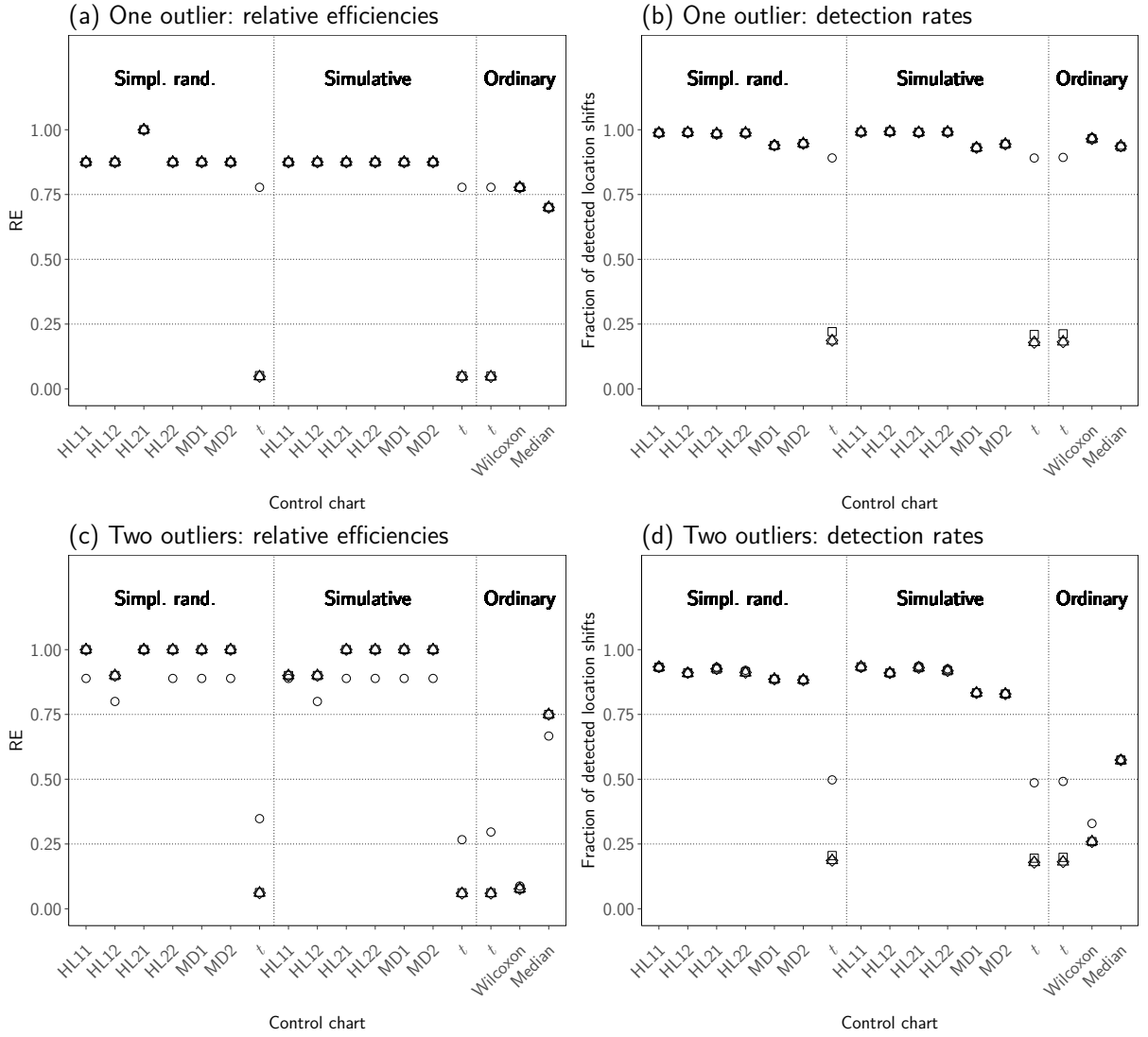

Figure 3.5: Relative efficiencies (RE) with respect to MRL_1 and detection rates under normality in outlier scenarios for an out-of-control process. The control charts are applied with subwindow widths $h = k = 10$ and tuned to $\mathrm{ARL}_0^* = 250$. The vertical dotted lines separate between the different groups of control charts. Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

3.5 Applications

In this section, we show how the control charts perform in selected real-world applications. We use the PAMONO time series (Subsection 2.2.2) and the heart-rate measurements (Subsection 2.2.1).

The PAMONO time series has a piecewise nearly constant signal, so that it provides a good testing ground for our control-chart principle as the assumption of a locally constant signal is nearly exactly valid. Opposed to that, the heart-rate measurements are affected by a time-varying trend and several disturbing structures of which not all may be clinically relevant, making the data more challenging for the charts.

We apply the charts based on the ordinary t-test, the Wilcoxon test, and the robust

chart based on the HL22-statistic. For the latter, we use the simplified randomisation and, for comparison, the ordinary randomisation. Our choice fell on these control charts for the following reasons: The ordinary t-test is the standard procedure for the local two-sample problem considered here and the Wilcoxon test is its most popular non-parametric competitor. The HL22-chart turned out to be promising regarding distribution independence of the ARL_0 and robustness in our simulation studies. Considering the discussed restrictions of our performance criteria, the HL22-chart led to a very good overall performance.

We tune all charts to $\mathrm{ARL}_0^* = 250$ under normality and apply them with the subwindow widths $h = k = 10$.

For the simplified and the ordinary randomisation of the HL22-chart, we choose $b = 10\,000$ random splits. A problem arises when trying to specify α^* for the ordinary-randomised HL22-chart. The required simulations are too time-consuming to be performed in a reasonable time frame. Hence, we approximate α^* by using the corresponding value obtained for the ordinary t-chart. The motivation behind this is that we obtained similar ARL_0-values for selected values of α in our exemplary in-control simulation for the ordinary-randomised HL22-chart and the ordinary t-chart; see Table 3.3.

In Subsection 3.5.1, we will briefly describe a strategy on how the control charts can be set up for use in practical applications. The results for the PAMONO time series are presented in Subsection 3.5.2. In Subsection 3.5.3, we discuss the time series of heart-rate measurements.

3.5.1 Setting up the control charts for applications

Before using the control charts on real data sets, several choices have to be made. The main questions are:

- How many consecutive observations have to be shifted so that a level shift can be assumed?
- What time delay is acceptable before a change point needs to be detected?
- For what time period can the signal be considered to be approximately constant?
- How long should the duration between two subsequent false alarms be?
- Is anything known about the distributional shape of the data? For example, is it symmetric or asymmetric?

The answers to these questions determine which control chart, subwindow widths, and ARL_0^* should be used. For example, if it is known that no outliers occur in the data, the Wilcoxon chart is a good choice because of its distribution-free in-control run length and good detection quality under several types of distributions. If some robustness is required, the HL-based charts could be more reasonable.

Specifying the correct α^* is the most time-consuming part because we do not have any analytical results on the relationship between the ARL_0 and α. Hence, it has to be determined by simulation. For the window widths considered in this thesis, Equation (3.10) can be used with the parameter estimates given in the corresponding tables. If other window widths are desired, a simulation similar to the one for our in-control investigations could be performed.

Combining the previous thoughts, we suggest determining α^* using Scheme 1.

Scheme 1 (Simulation of α^*)
Given a control chart with subwindow widths h and k, the significance level α^* to achieve a desired ARL_0 can be computed by simulation using the following steps:

1) *Specification of a grid of reasonable values for the significance level α, covering the range in which α^* is suspected.*

2) *Generation of 10 000 time series of length 20 000.*

   a) *If the control chart has an approximately or completely distribution-free ARL_0, it suffices to perform the simulations under the $\mathcal{N}(0,1)$-distribution, since the results between the different distributions will vary only marginally.*

   b) *For a control chart with a distribution-dependent ARL_0, it is necessary to specify a data-generating distribution.*

3) *Application of the control chart to each time series with each value of α on the grid.*

4) *Estimation of the ARL_0 for each value of α by the sample mean of the individual run lengths.*

5) *Estimation of the functional relationship between ARL_0 and α according to Equation (3.9).*

6) *Estimation of α^* by Equation (3.10).*

Specifying a data-generating distribution in 2) is necessary because we cannot determine the needed parameters in Equation (3.10) analytically. For a completely or approximately distribution-free control chart, this does not impose a restriction. The functional relationship between the ARL_0 and α needs to be calculated only once and can be used for arbitrary and even unknown data-generating distributions.

The range in which α^* lies can be determined by carrying out a pre-run, where the control chart is applied to a few time series with some selected values of α.

If the nominal ARL_0 should be large, it is advisable to use a narrow grid of possible values for α to improve the approximation. The smaller α is, the more influential are slight changes in a decimal with respect to the ARL_0; see Figure 3.1.

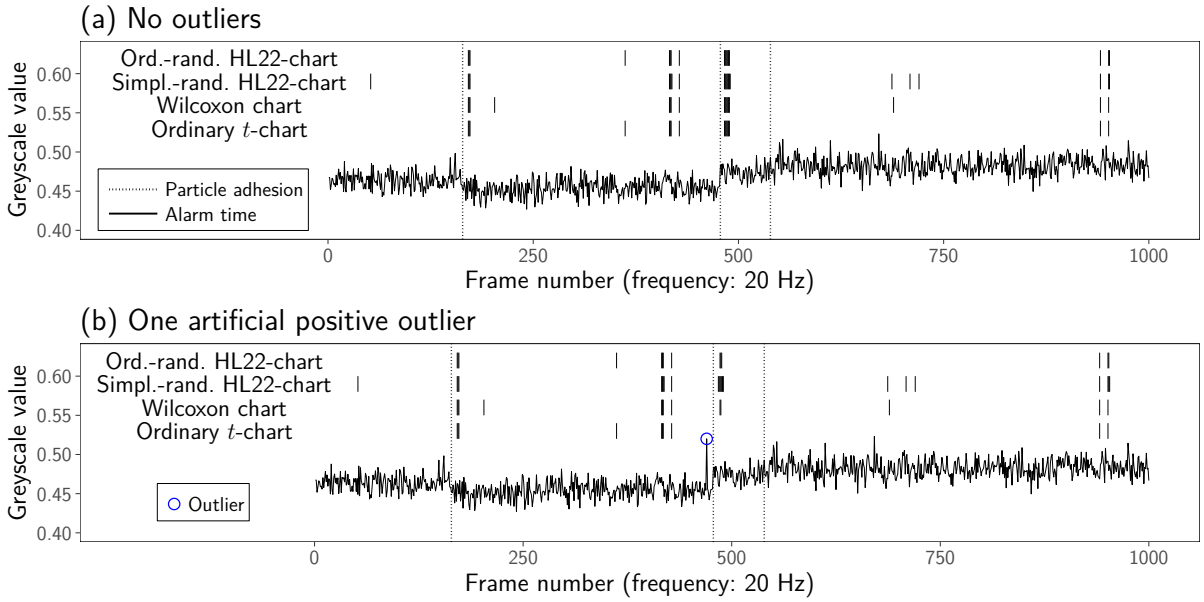

Figure 3.6: Application of selected control charts to the original PAMONO time series (panel (a)) and the same time series with an artificial additive outlier (panel (b)). The charts are applied with the subwindow widths $h = k = 10$ and tuned to $\mathrm{ARL}_0^* = 250$ under normality.

3.5.2 PAMONO data

As we use synthetically generated sensor data, the particle-induced change points are known, so that we can distinguish relevant from irrelevant location shifts. In their analysis, Abbas et al. (2016) conclude that the noise in these time series can be adequately modelled by a normal distribution.

We evaluate the detection performance of the charts on the original time series and on the time series after replacing one observation with an artificial outlier. Figure 3.6 shows the alarm times, symbolised by black, vertical lines.

In general, the alarms occur in groups of several subsequent alarms. This is because of the moving-window nature. To explain this, we consider an ideal shift at time t^*, where all observations at time points $t \geq t^*$ are larger than those at time points $t < t^*$. In the sample $\boldsymbol{Y}_{t^*}^{(n)}$, only the rightmost observation is shifted. For each new time point $t > t^*$ in the window, the number of shifted observations increases. Hence, at some time point, the fraction of observations affected by the shift in the test window suffices to cause a rejection of the null hypothesis, even though not all observations in the test window are affected by the structural break. Consequently, moving the window makes it likely that the local null hypothesis for next time windows will be rejected as well. This is the case until enough observations in the reference window are shifted and there is no indication of a location difference between both samples.

From Figure 3.6(a) we see that two of the three particle-induced location shifts are detected by each control chart. The third relevant location shift is missed because it is rather small. The fraction of alarm times which are not associated with a location shift is

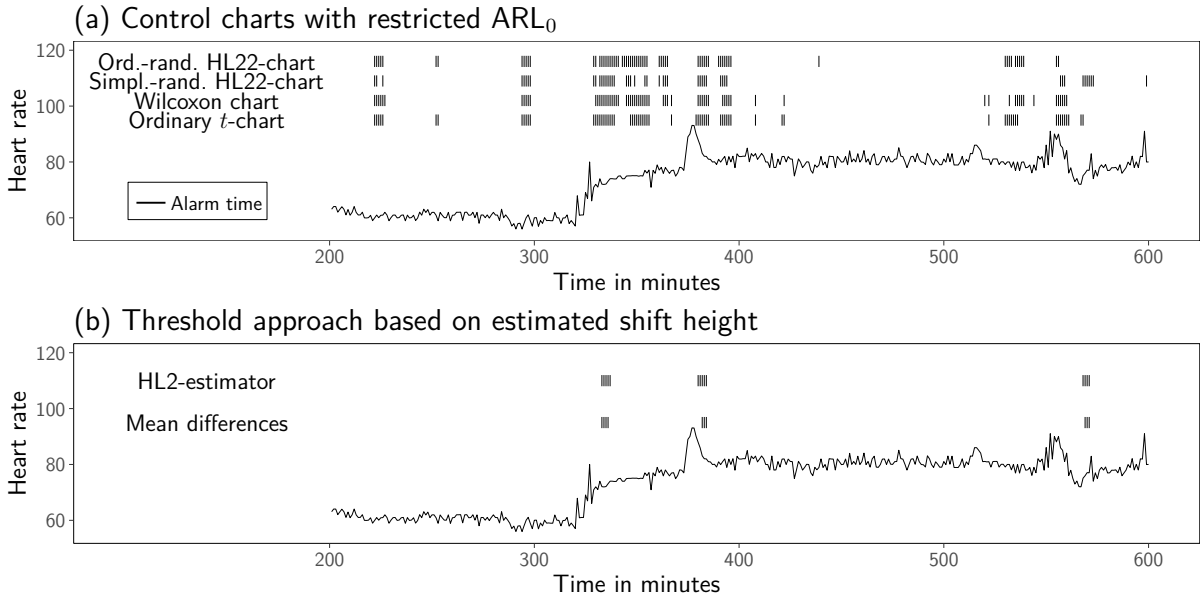

Figure 3.7: Application of selected control procedures to the time series of heart-rate measurements. The procedures are applied with the subwindow widths $h = k = 10$. In panel (a), the two-sample tests are applied as control charts with $\mathrm{ARL}_0^* = 250$ under normality. In panel (b), an alarm is given when the absolute value of the estimated shift height $\hat{\Delta}_t$ exceeds a specified threshold value, here if $|\hat{\Delta}_t| > 10$.

about 0.01 for all charts and therefore quite small. Both HL22-charts lead to very similar results. Hence, in this rather clean scenario, the simplified-randomised HL22-chart seems to be a quite good substitute for the ordinary-randomised HL22-chart.

Figure 3.6(b) depicts the alarm times when we replace the measurement at time $t = 470$, which is 0.44, with the larger value 0.52 to create a positive outlier prior to the second particle-induced location shift. In accordance with our results obtained from studying the outlier scenarios in Subsection 3.4.4, the t-chart misses the location shift because of the outlier. The other charts remain nearly unaffected.

In an evaluation of a complete PAMONO data set, the change-point detection in the pixel time series is only one of several analysis steps to decide whether a structure in an image sequence is caused by a particle. Thus, rare false alarms and missed location shifts are no major drawbacks. As the spots affect several adjacent pixel coordinates, it is likely that the change points are detected in another time series, so that this problem can be caught in the further processing; see Abbas et al. (2016).

3.5.3 Time series of heart-rate measurements

In the intensive-care time series, the clinically relevant change points are unknown. Intuitively, we would classify the large location shift starting at $t = 321$ as relevant. Nevertheless, alarms due to the small deviations from a constant signal are likely.

The alarm times of the control charts are shown in Figure 3.7(a). The large shift at $t = 321$ is found by all procedures. However, we can also note a comparatively large

number of alarms caused by other structures. The simplified-randomised HL22-chart seems to be slightly more resistant to small changes. Unlike the other charts, it ignores the peak at $t = 515$. Again, the results for simplified and ordinary randomisation are quite similar. In general, changes can be detected quite reliably. Nevertheless, the number of alarms due to irrelevant structures seems to be too large.

One possibility for improvement would be to compare the estimated shift height for each time point to a specific threshold to determine the relevance of the structural break. A simple initial idea is the following: Given the test window $\boldsymbol{Y}_{t+}$ and the reference window $\boldsymbol{Y}_{t-}$, we compute the estimated location difference $\hat{\Delta}_t$. In the two-sided case, the absolute value is compared to a fixed value $c > 0$, which represents the magnitude of location shifts that should be detected. An alarm would then be given if $|\hat{\Delta}_t| > c$.

As an example of this approach, we use $c = 10$ so that only sudden differences in the heart rate by more than 10 units should cause an alarm. The results for the differences of the sample means and the two-sample Hodges-Lehmann estimator, which correspond to the t-chart and the HL22-chart used before, are shown in Figure 3.7(b). The Wilcoxon statistic cannot be considered here, since the rank sum does not give information on the height of the location shift.

Compared to the results in 3.7(a) there are far fewer alarms. They now concentrate on the location shift at $t = 321$, the large peak at $t = 375$, and the structure at $t = 550$.

3.6 Discussion

Our simulations in this chapter indicate that the discussed control-chart approach of using two-sample tests in a moving time window is a promising step to monitor a process robustly and nearly distribution free in real time. It allows overcoming some deficiencies of ordinary control schemes for applications where the time series are affected by a time-varying signal. Moreover, by making only local assumptions, the resulting charts are quite flexible and allow for adaptation to the course of the signal.

We come to the following conclusions:

- The rank-based control charts have a distribution-free in-control run length.
- When using a simplified randomisation principle, it is possible to obtain charts with an approximately distribution-free in-control ARL. Especially the MD- and the HL2-charts show promising results over several subwindow widths. The HL1-charts have problems with skewed distributions.
- Control charts based on robust statistics are insensitive to outliers. Moreover, using test statistics of tests which are powerful over a wide range of distributions yields control charts with good out-of-control properties regarding the detection speed as measured by the MRL_1 and the detection rate.

- For small subwindow widths, the simplified-randomised charts show a rather weak performance with respect to the ARL_1. Enlarging the windows leads to better results and similar conclusions as for the MRL_1.
- Using the normality assumption to simulate the null distribution of the robust charts makes them far more prone to different data-generating distributions than the simplified randomisation.
- Two application examples indicate that our approach is well suited for time series with a piecewise constant signal. However, the control charts react sensitively to violations of the assumption of a locally constant signal.
- Based on our simulation studies we would recommend the Wilcoxon test and the HL2-estimator as a basis for a control chart. The Wilcoxon chart is distribution free on an in-control process and shows good detection properties in uncontaminated data. The HL2-estimator leads to a control chart which has an approximately distribution-free ARL_0, is powerful in detecting change points, and robust against outliers.

To improve the computational feasibility of the robust control charts, we use a special version of the randomisation principle to compute the control limits. Instead of recomputing the null distribution at each new time point, we calculate it only once for each time series by using only the observations in the first time window. This approach requires that the distributional structure of the observations does not change over time. For our real-world examples, the results obtained with this simplified approach do not differ much from the ordinary randomisation. However, the simplified approach increases the dependence of the control charts on the data-generating distribution for an in-control process. In addition, it leads occasionally to very large run lengths, which negatively affects the ARL_1-performance. Thus, there is still room for further adjustments. We will consider one modification in Subsection 4.3.4, where we update the randomisation distribution after a specific amount of time.

Speeding up the computations of the robust charts by simulating the null distribution under a distributional assumption turns out to be problematic. The in-control performance of the charts deteriorates if the distributional assumption is violated. However, the estimators are robust against outliers, so that the charts provide a reasonable alternative to the randomised charts if the data justify the distributional assumption.

In Chapter 4, we investigate residual-based versions of the control charts. The goal is to detect sudden location and trend changes in time series with a time-varying signal, for which the assumption of a locally constant signal is not justifiable.

4 Robust control charts for the mean of locally linear time series

This chapter is based on the manuscript "Control charts for the mean of a locally linear time series" by Abbas and Fried (2019), submitted for publication.

Most of this chapter corresponds to the submitted version. Some major changes are the following:

- We illustrate why the charts from Chapter 3 are not reliably applicable under the model assumptions considered here.
- We comment on a control chart based on the SCARM statistic from Equation (2.5).
- The update-randomisation principle described in this chapter is combined with the control charts from Chapter 3.
- We add a short simulation study in which we investigate the capability to detect sudden trend changes with the presented control charts.
- A comment on the ARL_1-performance is given.
- An additional proposition states some properties of one-step-ahead forecast errors of a local linear regression.
- Some rephrasing and shortening was necessary to avoid redundancy of aspects already mentioned and studied in Chapter 3.

4.1 Introduction

The control charts studied in Chapter 3 are able keep the desired ARL_0 even in case of a slowly varying signal, as long as it is locally nearly constant. As an example, we consider the time series shown in Figure 4.1(a), generated from the model

$$Y_t = \sin\left(0.8 \cdot \left(-2\pi + \frac{t-1}{100}\right)\right) + \varepsilon_t,\ t = 1, \dots 1257, \tag{4.1}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. We apply the ordinary t-chart with subwindow widths $h = k = 10$, tuned to $\text{ARL}_0^* = 250$ under normality. Given the selected value $\alpha^* \approx 0.007$, the alarm proportion of about 0.007 is quite small. Reducing the variance of the noise to $1/100$ emphasises the trend in the time series. This is shown in Figure 4.1(b). Here, the alarm

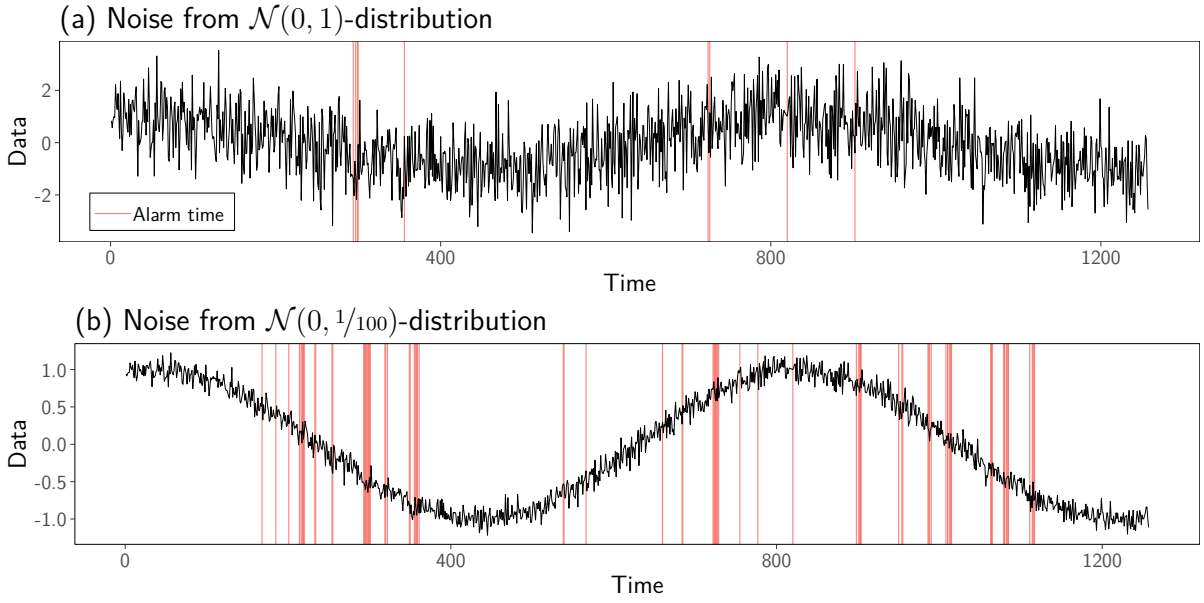

Figure 4.1: Application of the ordinary t-chart with subwindow widths $h = k = 10$ and $\text{ARL}_0^* = 250$ under normality to a time series with a sine-wave signal and additive normally distributed noise.

proportion increases to nearly 7%. The alarms occur mainly in areas where the slope of the signal is very steep, indicating that the chart confuses natural trends with location shifts. Thus, alarms due to structural breaks cannot be distinguished from alarms caused by the trend.

This impression is backed up by a small simulation study. We generate 1 000 time series from the model in Equation (4.1). For $\mathcal{N}(0,1)$-distributed noise, the estimated ARL_0 is 248 (standard error 7.7). If the noise variance is set to $1/100$, the estimated ARL_0 drops to 94 (standard error 1.5). Because the test statistic is scale invariant, the difference in the ARL-values can only be caused by the trend.

In this chapter, we extend the control charts from Chapter 3 so that they work reliably on time series with a strongly non-linear in-control signal. Examples for such time series are the time series of crack widths and the time series of heart-rate measurements introduced in Section 2.2.

Under the assumption of a locally linear signal, we aim at control charts that have the same properties as stated in the previous chapter, such as being robust against outliers, stable over a wide range of distributions, and fast in detecting change points. We broaden the definition of a change point so that we are not only interested in finding location shifts, but also sudden trend changes. Again, we aim at procedures that are able to adapt to the local signal behaviour, as described in the previous chapter.

In the SPC literature, there are many proposals to perform process-control tasks on time series with a time-varying signal. These are generally based on the assumption that the process behaviour can be described by a time-series model. In general, there exist two ways to modify ordinary control charts and make them applicable in such situations: *residual*

control charts and *modified control charts* (Knoth and Schmid, 2004). A modified control chart is a control chart in which the control limits are adjusted to the underlying time-series model. A residual control chart fits a time-series model to the data and performs the monitoring on one-step-ahead forecast errors. In this chapter, we focus on residual control charts.

It is important to specify the correct model to ensure adequate performance of the chart (Zhou and Goh, 2016). However, sufficiently large data sets to estimate the model parameters may not exist, as motivated in Chapter 3. That is why we propose residual control charts in which the one-step-ahead forecast errors are computed based on a local robust regression fit. We use the repeated median from Equation (2.4) to compute the one-step-ahead forecast errors from a moving time window. We apply two-sample location tests to them as described in Chapter 3. Abrupt trend or location shifts lead to level shifts in the time series of forecast errors, which can then be detected by the tests.

In this chapter, we concentrate on a local simple linear regression model as introduced in Subsection 2.3.2. We conduct several simulation studies in which we compare the control charts in different in- and out-of-control settings. Similar to the results obtained in Chapter 3, we find that the two-sample Hodges-Lehmann estimator leads to very promising results regarding our performance criteria. The rank-based charts lose the distribution independence of their ARL_0. Additionally, we also look briefly at a local AR(1)-model. Here, we find that the distribution independence of the run length for in-control processes is not retained, particularly, when the autocorrelation is high. However, under distributional assumptions, the charts are still able to detect change points quickly for small to moderate autocorrelations.

In the literature on control charts, residual charts for global time-series models have received much attention; see Alwan and Roberts (1988), Montgomery and Mastrangelo (1991), Garthoff et al. (2014), and Garthoff et al. (2015). Similar to ordinary control charts, they do not allow for natural changes in the model parameters. Local model assumptions are made in the field of profile monitoring, where a functional relationship between a target variable and explanatory variables is assumed. However, historical data sets are needed to estimate a standard in-control profile and the relationship is assumed to be the same in each sample (Saghaei and Noorossana, 2011). Croux et al. (2011) use forecast errors of robustified Holt-Winters smoothing. The drawback here is that it needs historical data for the parameter tuning and the control limits are derived under the normality assumption. Moreover, it is not clear how to make the procedure robust against a desired number of consecutive outliers.

This chapter is organised as follows: In Section 4.2, we describe the basic model assumptions and connect the problem to the one discussed in Chapter 3. The construction of the residual control charts is described in Section 4.3. The results of our simulation studies are discussed in Secion 4.4. In Section 4.5, we apply the charts to some exemplary time series and discuss our main findings in Section 4.6.

4.2 Model

Again, our basic model assumption is that a real-valued time series $(Y_t\colon\ t \in \mathbb{N})$ can be described by the additive components model

$$Y_t = \mu_t + \varepsilon_t + \eta_t,\ t \in \mathbb{N},$$

with $\mathrm{Var}\,(\varepsilon_t) = \sigma^2$ for all $t \in \mathbb{N}$.

In contrast to Chapter 3, we now assume that the in-control process can be locally approximated by a linear regression model in a time window $\boldsymbol{Y}_t^{(\ell)} = (Y_{t-\ell+1}, \ldots, Y_t)$ of length $\ell \in \mathbb{N}$. We call $\boldsymbol{Y}_t^{(\ell)}$ *regression window at time* t in the following. The simplest example of such a model is

$$\mu_{t+i} \approx \mu_t + \beta_t \cdot i,\ i = -\ell+1, \ldots, 0,\ t \geq \ell, \tag{4.2}$$

where μ_t is the signal value at time t and β_t the constant slope in the window. This is similar to the definition in Equation (2.3), with the difference that we shifted the design points to take the online setting considered here into account. The one-step-ahead forecast for Y_{t+1} based on $\boldsymbol{Y}_t^{(\ell)}$ is given by $\hat{Y}_{t+1} = \mu_t + \beta_t$. We define the corresponding one-step-ahead forecast error by

$$e_{t+1} = Y_{t+1} - \hat{Y}_{t+1} = \varepsilon_{t+1} + \eta_{t+1} + (\mu_{t+1} - (\mu_t + \beta_t))\,,\ t \geq \ell. \tag{4.3}$$

Therefore, a structural break with $\mu_{t+1} \neq \mu_t + \beta_t$ in $(Y_t\colon\ t \in \mathbb{N})$ leads to a structural break in the sequence $(e_t\colon\ t \geq \ell + 1)$ of one-step-ahead forecast errors. Moreover, in the absence of change points, this sequence does not contain a trend.

The basic idea of our approach described in Section 4.3 is to apply the control charts introduced in Chapter 3 to the one-step-ahead forecast errors in order to detect structural breaks in the trend-free time series.

Let now $(\mu_t^*\colon\ \mu_t^* = \mathrm{E}\,(e_t)\,, t \geq \ell + 1)$ be the mean function of the sequence of one-step-ahead forecast errors. Similar to Chapter 3, our goal is to detect location shifts in the mean function between the time points $t-k$ and $t-k+1$ using the time window $\boldsymbol{e}_t^{(n)} = (e_{t-n+1}, \ldots, e_t)$, $t \geq \ell + n$, of the n most recent one-step-ahead forecast errors. In the time window, the mean function can be described by

$$\mu_{t+i}^* = \begin{cases} \mu_{t-}^*, & i = -n+1, \ldots, -k \\ \mu_{t+}^*, & i = -k+1, \ldots, 0, \end{cases} \tag{4.4}$$

with $\mu_{t+}^* = \mu_{t-}^* + \Delta_t$, where $\Delta_t \in \mathbb{R}$ is the size of a potential location shift between the time points $t-k$ and $t-k+1$. Thus, we are in the same situation as described in Section 3.2.

4.3 Methods

Similar to the approach from Chapter 3, we split the time window $\boldsymbol{e}_t^{(n)} = (e_{t-n+1}, \ldots, e_t)$ into two subwindows

$$\boldsymbol{e}_{t-} = \left(e_{t,1}^-, \ldots, e_{t,h}^-\right) \quad \text{and} \quad \boldsymbol{e}_{t+} = \left(e_{t,1}^+, \ldots, e_{t,k}^+\right),$$

where

$$e_{t,i}^- = e_{t-n+i},\ i = 1, \ldots, h, \quad \text{and} \quad e_{t,j}^+ = e_{t-k+j},\ j = 1, \ldots, k,$$

of widths $h, k \in \mathbb{N}$ with $n = h + k$. The reference window $\boldsymbol{e}_{t-}$ and the test window $\boldsymbol{e}_{t+}$ are then compared by a two-sample location test for a change point at time $t - k + 1$.

4.3.1 Removing trends by local regression

Equation (4.3) requires the true regression parameters μ_t and β_t in the time window $\boldsymbol{Y}_t^{(\ell)}$ to be known. This ensures that the one-step-ahead forecast errors are i.i.d. for an in-control process. In practice, the parameters are unknown, so that we estimate them from the time window. Using the assumption of a locally linear signal in Equation (4.2), we estimate the parameters by linear regression. The control charts are applied as described by the following Scheme 2.

Scheme 2 (Residual control charts based on local regression)
Let $(Y_t\colon\ t \in \mathbb{N})$ be a time series, $\ell \in \mathbb{N}$ the width of the regression window, and h and k the subwindow widths for the two-sample test with $h + k = n$. The procedure starts at time point $t = \ell + 1$:

1. *Fit a linear function to the time window $\boldsymbol{Y}_{t-1}^{(\ell)}$, providing estimators $\hat{\mu}_{t-1}$ and $\hat{\beta}_{t-1}$ of level and slope in the time window.*

2. *Calculate the one-step-ahead forecast error $\hat{e}_t = Y_t - \left(\hat{\mu}_{t-1} + \hat{\beta}_{t-1}\right)$.*

3. *If $t < \ell + n$, there are too few observations to perform a two-sample test with the specified subwindow widths. Set $t = t + 1$ and return to 1.*

4. *If $t \geq \ell + n$, perform a two-sample test on the sample $(\hat{e}_{t-n+1}, \ldots, \hat{e}_t)$ to identify a change point at the time point $t - k + 1$, using $(\hat{e}_{t-n+1}, \ldots, \hat{e}_{t-k})$ as reference and $(\hat{e}_{t-k+1}, \ldots, \hat{e}_t)$ as test window. Then, set $t = t + 1$ and return to 1.*

As the regression estimators for subsequent one-step-ahead forecast errors are computed from overlapping regression windows, the forecast errors are ℓ-dependent. With the following Proposition 2, we show this for the simple case of a linear trend with constant parameters. We defer the proof to Appendix A.

Proposition 2 (Properties of the one-step-ahead forecast errors)
Let $(Y_t\colon\ t \in \mathbb{Z})$ *be a time series with* $Y_t = \mu_t + \varepsilon_t$, *where* $\mu_t = \mu + \beta \cdot t$, $t \in \mathbb{Z}$. *The process* $(\varepsilon_t\colon\ t \in \mathbb{Z})$ *of independent random variables is assumed to be strictly stationary with expectation* $E(\varepsilon_t) = 0$ *and variance* $Var(\varepsilon_t) = \sigma^2$ *for all* $t \in \mathbb{Z}$. *Let moreover* $\hat{\mu}_{t-1}$ *and* $\hat{\beta}_{t-1}$ *be unbiased and regression equivariant regression estimators for* μ_{t-1} *and* β, *obtained from a time window* $(Y_{t-\ell}, \ldots, Y_{t-1})$ *with a fixed design* $i = -\ell + 1, \ldots, 0$. *Then, the following results hold for the expectation, the variance, and the autocovariance function of the sequence of one-step-ahead forecast errors* $(\hat{e}_t\colon\ t \in \mathbb{Z})$*:*

(a) $E(\hat{e}_t) = 0$ *for all* $t \in \mathbb{Z}$.

(b) $Var(\hat{e}_t) = \sigma^2 + Var\left(\hat{\mu}_{t-1}\right) + Var\left(\hat{\beta}_{t-1}\right) + 2 \cdot Cov\left(\hat{\mu}_{t-1}, \hat{\beta}_{t-1}\right)$ *for all* $t \in \mathbb{Z}$.

(c)

$$Cov(\hat{e}_t, \hat{e}_{t+s}) = \begin{cases} Var(\hat{e}_t)\,, & s = 0 \\ Cov\left(\hat{\mu}_{t-1}, \hat{\mu}_{t+s-1}\right) + Cov\left(\hat{\mu}_{t-1}, \hat{\beta}_{t+s-1}\right) + & \\ Cov\left(\hat{\beta}_{t-1}, \hat{\mu}_{t+s-1}\right) + Cov\left(\hat{\beta}_{t-1}, \hat{\beta}_{t+s-1}\right) - & \\ Cov\left(\varepsilon_t, \hat{\mu}_{t+s-1}\right) - Cov\left(\varepsilon_t, \hat{\beta}_{t+s-1}\right), & 1 \le s \le \ell \\ 0, & s > \ell \end{cases}$$

for all $t \in \mathbb{Z}$.

(d) $(\hat{e}_t\colon\ t \in \mathbb{Z})$ *is weakly stationary.*

Even though the trend is removed from the time series, the independence assumption underlying the two-sample tests considered here is violated. Therefore, the results obtained in Chapter 3 are not directly transferable; see Section 4.4.

4.3.2 Comparison of repeated median and ordinary least squares

In principle, every linear regression estimator is legitimate for estimating the regression parameters. Based on the results summarised in Chapter 2, we use RM regression in the following; see Equation (2.4).

To motivate the preference of a robust regression estimator over a non-robust one, like OLS regression, we consider two examples.

First, we study how a level or a slope change affects the one-step-ahead forecast errors considering only the signal without additive noise. We compute the signal from the model $\mu_t = 0.01 \cdot t$, $t = 1, \ldots, 80$. Figure 4.2(a) shows the time series with a level shift at time point $t = 51$. The one-step-ahead forecast errors are computed with $\ell = 20$. Figure 4.2(b) displays them for RM regression, Figure 4.2(c) those for OLS regression. Before the change point, the forecast errors are zero for both procedures. The main difference is that the RM forecast errors show a location shift that begins at $t = 51$ and persists for ten time points. OLS regression leads to a peak at $t = 51$. The forecast errors decrease linearly directly

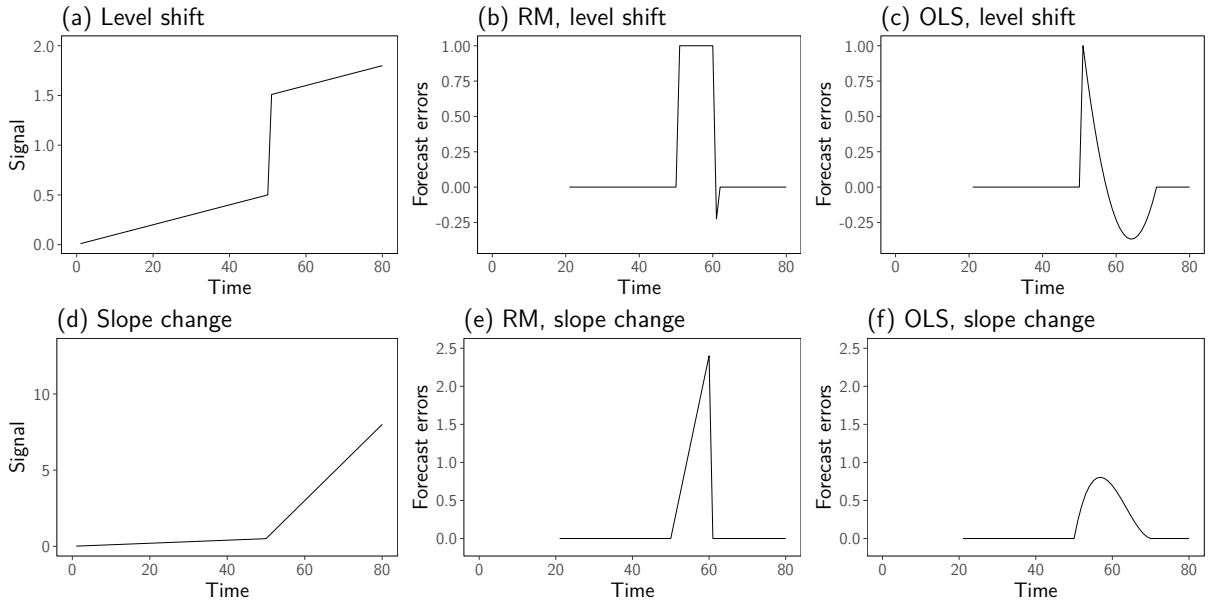

Figure 4.2: One-step-ahead forecast errors obtained from RM and OLS regression when the signal is affected by a level shift (panel (a)) or a slope change (panel (d)). The panels (b), (c), (e), and (f) show the one-step-ahead forecast errors for both situations. The width of the regression window is $\ell = 20$.

after the change point. For both regression estimators, they become negative before going back to zero. For OLS regression, a parabola can be observed for the values below zero, whereas the RM forecast errors are below zero only for one time point.

The results harmonise with the intuition from the breakdown properties of the estimators. As RM regression has a breakdown point of about 50%, many observations in the window have to be shifted before the estimator is affected substantially by the structural break. This causes a sequence of large forecast errors resulting in a location shift. OLS regression is influenced much earlier by the outliers, so that the forecast errors drop quickly and the level shift is not retained as well as for RM regression. Moreover, because of the parabolic shape the forecast errors describe when they are negative, the estimated location difference may be close to zero, so that the control chart might ignore the change point. RM regression reduces this risk. The negative values can be explained by an overestimation of the true signal because the time window contains observations from before and after the location shifts, leading to a large positive slope.

Figures 4.2(e) and 4.2(f) show the one-step-ahead forecast errors for a slope change. The original signal is shown in Figure 4.2(d). The RM forecast errors increase linearly before dropping abruptly to zero. Those for OLS regression describe nearly a bell shape.

Again, this can be attributed to the breakdown properties, which cause the increasing trend to influence the OLS forecast errors far earlier than the RM forecast errors. Here, OLS regression seems to have some benefits as the shape of the forecast errors resembles a location shift more than for the RM forecast errors. However, using suitable subwindow widths h and k for the tests makes it possible to detect the trend change because the

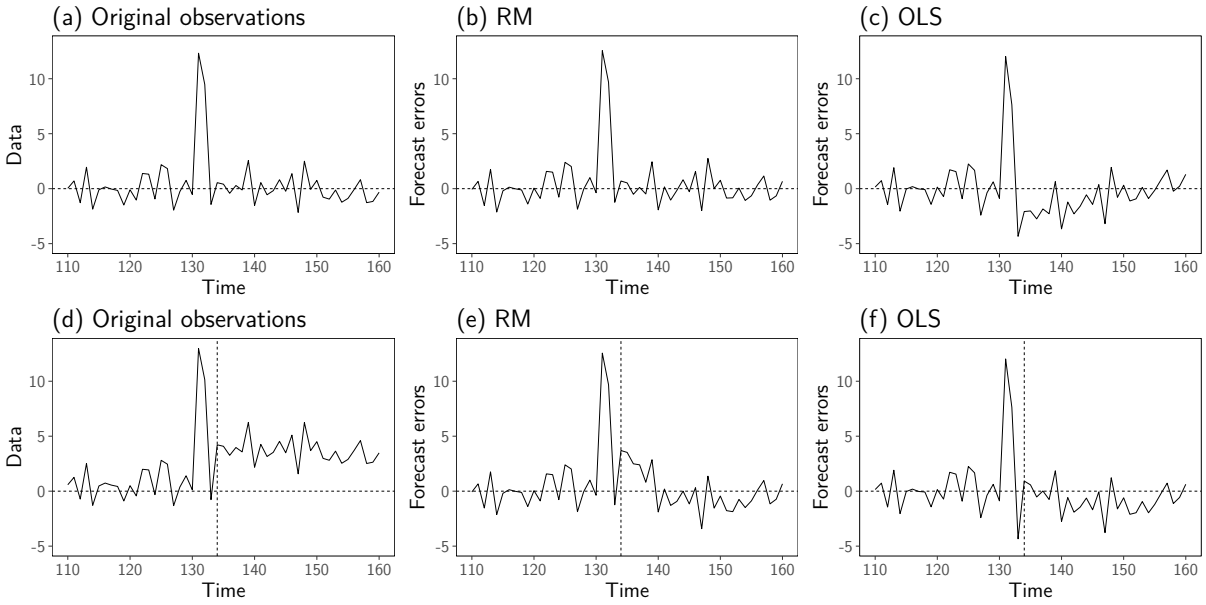

Figure 4.3: One-step-ahead forecast errors obtained from RM and OLS regression. Panels (b) and (c) show the forecast errors for the in-control time series in (a), which is corrupted by two subsequent outliers. Panels (e) and (f) show the forecast errors in the vicinity of a location shift directly behind the outliers; see panel (d). The width of the regression window is $\ell = 30$. The horizontal line is for orientation and shows the zero position, the vertical line marks the structural break.

linear increase in the forecast errors leads to a natural location difference between both subwindows.

In a second example, we illustrate how outliers can influence the forecast errors. Here, we generate an uncontaminated time series from the model $Y_t = \varepsilon_t$, $t = 1, \ldots, 160$, where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Again, we consider two scenarios.

In the first, we add outliers of size 10 at the time points $t = 131$ and $t = 132$. In the second case, we add a location shift of size 3 at $t = 134$, starting nearly instantaneously behind the outliers to mimic an out-of-control scenario. We compute the forecast errors shown in Figure 4.3 with $\ell = 30$.

The one-step-ahead OLS forecast errors are influenced by the outliers. The level shortly after the outliers is smaller than before. In contrast, the errors obtained by RM regression are not visibly influenced and resemble the original time series; see Figure 4.3(a) to (c).

Similarly, outliers can also prevent the detection of structural breaks when the forecast errors are computed by OLS regression; see Figure 4.3(d) to (f). The RM forecast errors retain the location shift better than the OLS forecast errors. The level of the latter does not seem to be much larger than prior to the shift.

Concluding, from the two procedures only RM regression retains the location shift in the time series of forecast errors. Moreover, the illustrations indicate that it is able to provide reliable results even under the influence of outliers. As these results stem from the robustness of the RM, other robust estimators may provide similar results.

The detection of trend changes seems to be more difficult because they do not necessarily lead to a location shift in the error sequence. Therefore, our assumption in Equation (4.4) is only justified for location shifts.

The decreasing forecast errors result because, at some point, even a robust estimator will be affected by a change point, so that the absolute values of the forecast errors become smaller. This behaviour implies the risk that the control charts may signal two structural breaks after a change point, of which only the first one corresponds to the true change point. A simple way to prevent this would be to stop the monitoring after a structural break has been signalled. Similar to the approach of Borowski and Fried (2014), it would start again after ℓ observations behind the alarm time are available, assuming that none of the observations in the new regression window is affected by the change point. However, this also means that no change points can be detected during this time period. In this thesis, we will not take any specific action to deal with the second location shift. Our interest lies in studying the general applicability of the control charts and additional rules would complicate the comparison of the procedures.

The declining trend also indicates that the subwindow widths h and k used for the two-sample test should be smaller than the width of the regression window ℓ. This is to prevent that the location shift in the sequence of forecast errors is missed. It can be confused with an outlier sequence if the test is performed with subwindow widths that are too large. The number of consecutive shifted forecast errors increases in ℓ because the regression method is robust against more outliers.

4.3.3 Improving the computation time of the repeated median

The calculation of the RM estimates is time-consuming for a large regression window. Computing them straightforwardly needs at least $O(\ell^2)$ computation time. To reduce it, Fried and Gather (2002) suggest a two-step approach which we outline briefly in the following.

The regression window $\boldsymbol{Y}_t^{(\ell)}$ is split into $\ell_2 \in \mathbb{N}$ blocks of $\ell_1 \in \mathbb{N}$ consecutive observations, where $\ell = \ell_1 \cdot \ell_2$ and ℓ_1 is odd. For each block, the sample median is computed, leading to the medians $\left(\widetilde{Y}_{t,1}, \ldots, \widetilde{Y}_{t,\ell_2}\right)$, where $\widetilde{Y}_{t,i}$ is the sample median of the i-th block, $i = 1, \ldots, \ell_2$. The new design points are the centre indices of the blocks.

If $\ell_1 << \ell$, the two-step RM estimates will be very similar to the results for ordinary RM estimation. The robustness of these new estimators will be slightly reduced. The effect on our control charts should be negligible as reasonable robustness suffices for retaining structural breaks in the forecast errors.

4.3.4 Modification of the simplified randomisation

The simplified randomisation described in Subsection 3.3.2 depends on a good approximation of the control limits by the critical values obtained from the first time window. This

assumption may be questionable when the number b of randomly drawn splits is small compared to the number of possible splits.

We now modify the simplified randomisation principle to an *update-randomisation principle*. The general idea is to start with an initial randomisation distribution and update it regularly after a specific duration. This improves the approximation of the critical values over time if all observations come from the same distributional class. Moreover, this approach adapts to possible trends in the time series.

The following Scheme 3 gives a sketch of this principle.

Scheme 3 (Update-randomisation principle)
Let $(e_t\colon\ t \geq \ell + 1)$ *be a time series of one-step-ahead forecast errors,* $n \in \mathbb{N}$ *the width of the moving time window for the tests,* $\ell \in \mathbb{N}$ *the width of the regression window,* $\alpha \in (0, 1)$ *the significance level, and* $d > n$, $d \in \mathbb{N}$, *a limit for the number of subsequent tests before an update is performed.*

1. *Compute a randomisation distribution for the first time window* $\boldsymbol{e}_{\ell+n}^{(n)}$ *of one-step-ahead forecast errors. The* $\alpha/2$*- and* $(1 - \alpha/2)$*-quantiles are the initial control limits.*

2. *Set counter* $j \in \mathbb{N}$ *to* $j = 1$.

3. *For all time points* $t \geq \ell + n$*:*

   i) *If* $j \leq d$, *perform a test on the time window* $\boldsymbol{e}_t^{(n)}$ *and set* $j = j + 1$.

   ii) *Else if* $j = d + 1$, *compute a randomisation distribution for* $\boldsymbol{e}_{t-n}^{(n)}$ *and add the values of the test statistic to the previously computed randomisation distribution. The new control limits are the corresponding* $\alpha/2$*- and* $(1 - \alpha/2)$*-quantiles. Perform the test on* $\boldsymbol{e}_t^{(n)}$ *and set* $j = 1$.

By using the randomisation distribution of $\boldsymbol{e}_{t-n}^{(n)}$ for updating the critical values, we avoid that a potential change point in $\boldsymbol{e}_t^{(n)}$ affects the new control limits.

This principle might be further improved by additional strategies. For example, to reduce the risk of possibly missed structural breaks affecting the randomisation distribution or prevent the randomisation distribution from becoming too large, we could drop old distributions from the set of all distributions after some time. Moreover, depending on the data, it could be advisable to compute a completely new set of distributions after the detection of a change point.

In this thesis, we do not use any additional rules for the update-randomisation principle, because in our simulations, we assume that all observations come from the same distributional class, the change points are placed at time points that do not affect the computation of the critical values, and we are only interested in the duration until the first alarm.

Table 4.1 displays the ARL_0-ratios for the robust control charts obtained from a simulation performed analogously to the one in Subsection 3.4.2 for the subwindow widths $h = k = 10$ under the assumption of a locally constant signal. We applied the charts to

Table 4.1: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality using the update-randomisation principle with subwindow widths $h = k = 10$. All values are rounded to two decimal places.

	HL11-chart				HL12-chart			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	0.99	0.85	0.79	0.47	1.00	0.88	0.88	0.35
0.0050	1.01	0.93	0.87	0.59	1.02	0.93	0.89	0.45
0.0075	1.02	0.97	0.90	0.66	1.02	0.96	0.90	0.52
0.0100	1.03	0.99	0.93	0.73	1.01	0.99	0.92	0.57
0.0125	1.03	1.01	0.95	0.77	1.01	0.99	0.92	0.62
0.0150	1.02	1.03	0.98	0.82	1.01	1.00	0.93	0.66
0.0200	1.03	1.06	1.01	0.88	1.03	1.02	0.94	0.73
	HL21-chart				HL22-chart			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.00	0.98	0.95	0.90	1.01	0.97	0.99	0.94
0.0050	1.03	0.99	1.01	1.01	1.02	0.98	1.01	1.00
0.0075	1.03	1.02	1.03	1.05	1.03	1.00	1.02	1.04
0.0100	1.03	1.01	1.05	1.06	1.03	0.98	1.02	1.05
0.0125	1.04	1.02	1.06	1.03	1.03	0.98	1.01	1.02
0.0150	1.03	1.02	1.05	1.03	1.03	0.98	1.01	1.01
0.0200	1.02	1.01	1.02	1.03	1.03	0.98	1.00	1.02
	MD1-chart				MD2-chart			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	0.99	1.01	0.95	0.89	1.01	1.03	0.94	0.93
0.0050	1.03	1.01	0.99	0.95	1.02	1.01	0.96	0.94
0.0075	1.01	1.00	0.98	0.96	1.02	1.00	0.96	0.94
0.0100	1.02	1.01	0.99	0.99	1.03	0.99	0.98	0.96
0.0125	1.02	0.99	0.99	0.98	1.03	1.00	0.99	0.98
0.0150	1.00	0.97	0.98	0.97	1.02	0.99	0.98	0.98
0.0200	1.00	0.98	0.98	0.99	0.99	0.99	0.98	0.99

the original observations instead of one-step-ahead forecast errors. The duration between two updates is set to $d = 200$.

For the HL2- and the MD-charts, the results are improved somewhat in terms of distribution independence of the ARL_0. Still, we cannot expect complete distribution independence of the ARL_0 as can be seen for the HL1-charts.

These results indicate that the update-randomisation principle is likely to reduce conservatism. This can be explained by smaller run lengths as compared to the simplified randomisation. A further improvement to reduce the anti-conservatism can be achieved by reducing the value of d at the cost of a larger computation time.

4.3.5 Using one-sample tests on the forecast errors

According to Proposition 2, for an unbiased estimation of the regression parameters, the one-step-ahead forecast errors of an in-control process vary regularly around zero. It is reasonable to ask whether an ordinary control chart with a fixed target value could be

Table 4.2: Average number of false alarms after applying a residual control chart based on a one-sample test (sample size $k = 10$) and a two-sample test (sample sizes $h = k = 10$) to time series with a sine-wave signal with additive $\mathcal{N}(0, 1/400)$-distributed noise for different widths ℓ of the regression window. The control charts are tuned to $\mathrm{ARL}_0^* = 370$ under normality, the forecast errors are computed by RM regression. The values are rounded to one decimal place.

ℓ	One-sample Wilcoxon signed rank	Two-sample Wilcoxon rank sum
50	13.5	6.2
75	51.3	6.3
100	192.4	7.4
125	447.4	8.0

applied to them safely. However, this property only holds when the signal in the time window follows an exact linear relationship. This can be doubtful when the signal is non-linear and ℓ is too large to justify an approximately linear signal in the regression window. A two-sample test can deal with slight non-linearities in the sequence of forecast errors, which may be the results of model misspecification.

We illustrate this by an example: In 1 000 replications, we generate time series with a sign-wave signal like in Equation (4.1) with additive $\mathcal{N}(0, 1/400)$-distributed noise.

We compare the Wilcoxon chart described in Subsection 3.3.2 with subwindow widths $h = k = 10$ to a control chart that uses the one-sample Wilcoxon signed-rank test in a moving time window of width $k = 10$. Control charts based on signed-rank statistics are, for example, investigated by Bakir (2006) and Chakraborti and Eryilmaz (2007). For our small example, we use the textbook definition of the signed-rank test; see Hollander et al. (2013, p. 40). The target value is set to zero. We compute the one-step-ahead forecast for $\ell = 50, 75, 100, 125$. The significance levels are chosen so that $\mathrm{ARL}_0^* = 370$ under normality.

Table 4.2 shows the average number of false alarms on the complete time series for both control charts and all window widths. The violation of the local linearity assumption becomes more severe with increasing value of ℓ. This is reflected by an increasing average number of false alarms for both charts. For the two-sample Wilcoxon test, it increases comparatively slowly from about 6 to 8, whereas for the signed-rank test, the values rise from 13 ($\ell = 50$) to 447 ($\ell = 125$).

Therefore, for a one-sample procedure it seems much more important that the local model is correctly specified. Otherwise, structures that are not eliminated by the regression are likely to be confused with an out-of-control situation. Using a two-sample test with a moving reference window compensates for this, which is why we only consider this approach in the following.

4.4 Simulations

Similar to Section 3.4, we study the run-length performance of the control charts in different in- and out-of-control scenarios. In most parts, the general simulation set-up is the same as in Section 3.4:

- Noise distributions: $\mathcal{N}(0,1)$, t_5, t_2, χ^2_3, χ^2_1.
- Subwindow widths for the two-sample tests: $(h,k) = (10,10), (20,20), (20,10)$.
- Number of randomly drawn splits for the update-randomisation principle: $b = 10\,000$.

We essentially compare the same control charts as in Section 3.4. However, we replace the simplified randomisation with the update randomisation. Furthermore, we do not include the simulative charts in our comparison. The reason is that their in-control performance in the simulations in Section 3.4 was not very convincing under a strong violation of the normality assumption. Moreover, because of the new parameter ℓ for the width of the regression window, this reduces the complexity in the presentation of the results.

The RM is not necessarily unbiased under skewed distributions (Siegel, 1982). Due to the regression equivariance of the RM and the location invariance of the test statistics, this does not lead to any problems in our analysis. A definition of regression equivariance can be found in Appendix D.4.

The new parameters in the simulations are chosen as follows:

- Width of the regression window: $\ell = 50, 75, 100, 125$.
- Block size for RM regression in the two-step approach: $\ell_1 = 5$ if $\ell \geq 100$ and $\ell_1 = 1$ otherwise.
- Update of distribution after every $d = 200$ consecutive tests for the update-randomisation principle.

In the following subsections, we generate 10 000 time series from each of the discussed simulation models.

4.4.1 In-control comparison under the $\mathcal{N}(0,1)$-distribution

Following the reasoning of Subsection 3.4.1, we start the analysis by studying the functional relationship between the ARL_0 and α under the $\mathcal{N}(0,1)$-distribution.

The data are generated from the model

$$Y_t = \varepsilon_t,\ t = 1, \ldots, 20\,000 + \ell,\ \ell = 50, 75, 100, 125,$$

where the first ℓ observations are used to fit the initial linear model and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Assuming that $\mu_t = \beta_t = 0$ for all $t = 1, \ldots, 20\,000 + \ell$ is no restriction compared to an

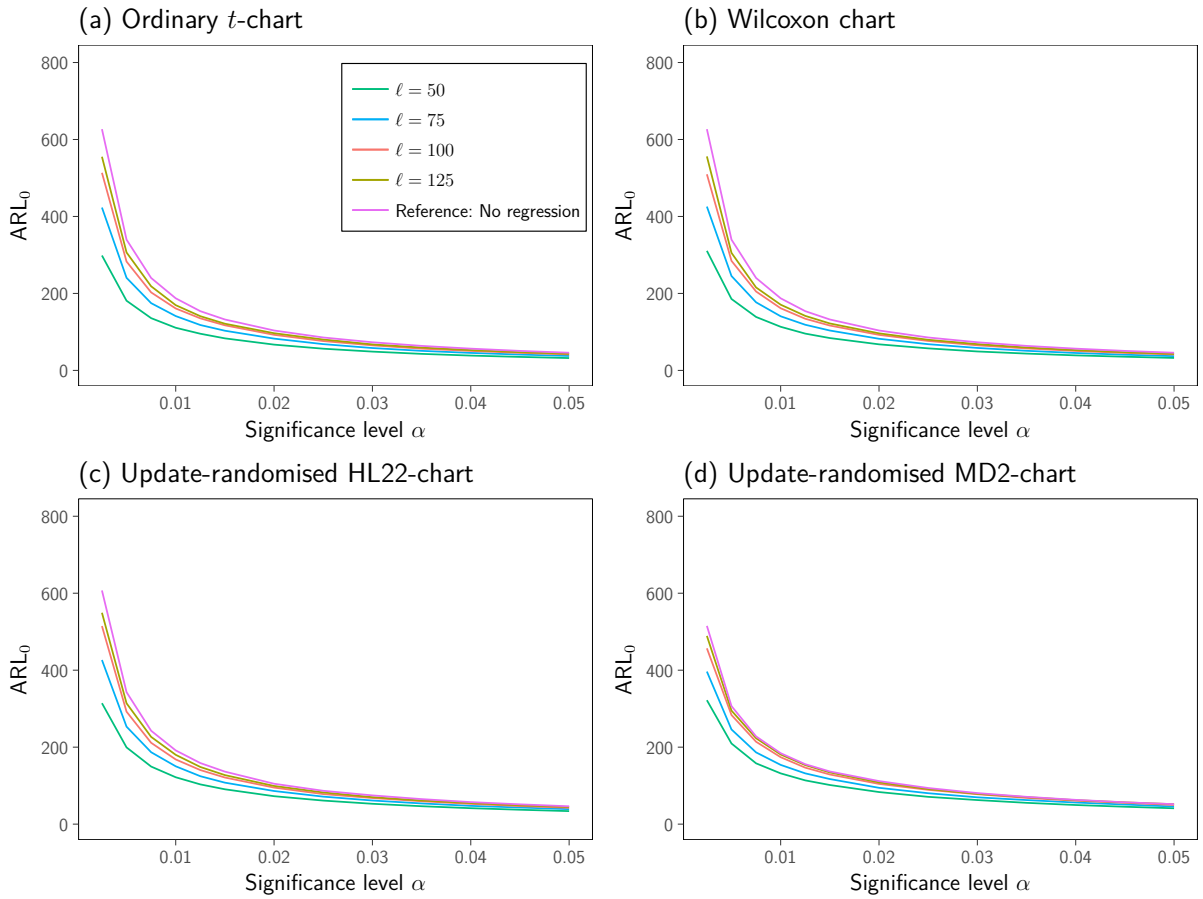

Figure 4.4: ARL_0-curves of selected control charts under normality with subwindow widths $h = k = 10$ and different widths ℓ of the regression window. For reference, the corresponding ARL_0-values obtained without a prior regression are shown.

arbitrary linear trend, because of the equivariance and invariance properties. However, using the two-step approach from Subsection 4.3.3 will lead to slightly different results compared to the ordinary RM estimation. For $\ell_1 << \ell$, the corresponding results differ only marginally. We illustrate this by a brief example at the end of this subsection.

Similar to the simulations in Section 3.4, we replace missing run lengths with a lower bound. Here, we use $20\,001 - \ell - n + 1$ and act analogously for the following simulations. Again, the fraction of missing run lengths is less than 1%, so that the effect of this replacement is negligible.

In Figure 4.4, we show the estimated ARL_0-curves for the ordinary t-chart, the Wilcoxon chart, the HL22-, and the MD2-chart for the different values of ℓ and the subwindow widths $h = k = 10$. Moreover, we added the ARL_0-curves obtained without a prior regression. The ARL_0-values of the residual charts are always smaller than when using the control charts from Chapter 3. For a fixed value of α, the ARL_0-value increases monotonously in the width ℓ of the regression window and approaches the value obtained without a prior regression. The course of the values resembles the functional relationship in Equation (3.8) in Subsection 3.4.1.

Thus, using the control limits obtained for locally constant signals in Subsection 3.4.1

Table 4.3: ARL_0 for selected values of α under normality for the subwindow widths $h = k = 10$ and different widths ℓ of the regression window. The values in brackets are the standard errors. All values are rounded to one decimal place.

ℓ	Control chart		Significance level α 0.005	0.02	0.05
50	t-charts	ordinary	181.0 (1.8)	67.0 (0.7)	32.1 (0.3)
		update random.	184.7 (1.8)	67.5 (0.7)	32.3 (0.4)
	HL11-chart		197.2 (2.2)	71.0 (0.8)	33.7 (0.4)
	HL12-chart		197.7 (2.1)	71.4 (0.8)	33.5 (0.4)
	HL21-chart		200.5 (2.2)	74.2 (0.8)	34.2 (0.4)
	HL22-chart		199.4 (2.2)	72.6 (0.8)	33.8 (0.4)
	MD1-chart		199.4 (3.1)	82.1 (1.2)	41.3 (0.6)
	MD2-chart		209.5 (3.4)	83.4 (1.2)	41.2 (0.6)
	Rank charts	Wilcoxon	185.6 (1.8)	67.7 (0.7)	32.4 (0.3)
		Median	182.7 (1.9)	72.6 (0.7)	28.1 (0.3)
125	t-charts	ordinary	306.4 (3.1)	96.6 (1.0)	43.1 (0.5)
		update random.	310.9 (3.2)	97.0 (1.0)	43.0 (0.4)
	HL11-chart		310.8 (3.5)	97.8 (1.1)	44.6 (0.5)
	HL12-chart		315.1 (3.4)	98.3 (1.0)	43.7 (0.5)
	HL21-chart		314.0 (3.6)	100.4 (1.1)	44.8 (0.5)
	HL22-chart		314.0 (3.5)	99.2 (1.1)	44.5 (0.5)
	MD1-chart		275.8 (4.4)	105.6 (1.6)	49.9 (0.7)
	MD2-chart		295.0 (5.1)	108.3 (1.6)	51.8 (0.8)
	Rank charts	Wilcoxon	305.4 (3.0)	96.3 (1.0)	42.8 (0.4)
		Median	256.9 (2.6)	95.8 (1.0)	33.5 (0.3)

seems only to be reliable if $\ell >> 125$. The correlation in the one-step-ahead forecast errors causes the local tests to be anti-conservative so that alarms can be expected to occur more frequently than desired. However, using a large value of ℓ requires the time span in which the signal can be seen as approximately linear to be rather long.

As the ARL_0-values for $\ell = 75$ and $\ell = 100$ are between those for $\ell = 50$ and $\ell = 125$, we restrict the considerations in the sequel to $\ell = 50$ and $\ell = 125$ for a more concise presentation of the results.

Table 4.3 allows a more detailed insight into the ARL_0-values for selected values of α, the subwindow widths $h = k = 10$ and $\ell = 50, 125$. The results for the subwindow widths $h = k = 20$ and $h = 20,\ k = 10$ can be found in the Tables C.10 and C.11 in Appendix C. Other than for the ARL_0-values shown in Table 3.2, the standard errors for all control charts are similar. Again, the t-charts, the HL-charts, and the Wilcoxon chart lead to similar ARL_0-values for the considered values of α. Moreover, the scale estimator used for the robust control charts seems to be more important for the actual ARL_0-value when the

Table 4.4: Estimated regression coefficients for the linearised relationship between the ARL_0 and α in Equation (3.9) under normality for the subwindow widths $h = k = 10$ and different widths ℓ of the regression window, rounded to two decimal places. The values in brackets are the standard errors and R^2 denotes the coefficient of determination, both rounded to four decimal places.

ℓ	Control chart		$\widehat{\log(\gamma_0)}$	$\hat{\gamma}_1$	R^2
50	t-chart	ordinary	1.27 (0.0284)	-0.74 (0.0068)	0.9991
		update. random.	1.25 (0.0288)	-0.75 (0.0069)	0.9991
	HL11-chart		1.29 (0.0315)	-0.75 (0.0075)	0.9989
	HL12-chart		1.27 (0.0334)	-0.76 (0.0080)	0.9988
	HL21-chart		1.35 (0.0422)	-0.75 (0.0101)	0.9980
	HL22-chart		1.33 (0.0435)	-0.75 (0.0104)	0.9979
	MD1-chart		1.81 (0.0433)	-0.66 (0.0104)	0.9973
	MD2-chart		1.72 (0.0360)	-0.68 (0.0086)	0.9983
	Rank chart	Wilcoxon	1.26 (0.0226)	-0.75 (0.0054)	0.9994
		Median	1.05 (0.0959)	-0.79 (0.0229)	0.9909
125	t-chart	ordinary	1.24 (0.0134)	-0.85 (0.0032)	0.9998
		update. random.	1.24 (0.0142)	-0.85 (0.0034)	0.9998
	HL11-chart		1.29 (0.0115)	-0.84 (0.0028)	0.9999
	HL12-chart		1.26 (0.0160)	-0.85 (0.0038)	0.9998
	HL21-chart		1.31 (0.0250)	-0.84 (0.0060)	0.9994
	HL22-chart		1.30 (0.0215)	-0.84 (0.0051)	0.9996
	MD1-chart		1.80 (0.0412)	-0.72 (0.0099)	0.9980
	MD2-chart		1.74 (0.0263)	-0.75 (0.0063)	0.9992
	Rank chart	Wilcoxon	1.24 (0.0147)	-0.85 (0.0035)	0.9998
		Median	1.03 (0.1101)	-0.86 (0.0263)	0.9899

subwindow widths h and k are unequal.

Thus, we make similar observations as for the charts introduced in Chapter 3. The update-randomisation principle leads to a substantial decrease of the standard errors, particularly for small values of α. This hints at a better possibility to estimate a significance level for a desired ARL_0. However, for the smallest considered values of α, all charts obtain only comparatively small ARL_0-values. Similar to the impressions we got from Subsection 3.4.1, it seems possible to approximate the ARL_0 for one of the update-randomised HL-charts under the $\mathcal{N}(0,1)$-distribution by the t-chart, which facilitates the computation.

Table 4.4 shows the regression coefficients of Equation (3.9), estimated from the simulation results for the subwindow widths $h = k = 10$. The corresponding values for $h = k = 20$ and $h = 20,\ k = 10$ are shown in Appendix C in the Tables C.12 and C.13. The coefficient of determination is always larger than 0.98, so that we can expect a very good approximation of the true functional relationship by the formula.

Having specified the functional relationship gives us the possibility to compare the ARL_0-values of the ordinary RM regression to the two-step approach we use for large values of ℓ. We conduct a small simulation study for the ordinary t-chart with $\ell = 125$ and the subwindow widths $h = k = 10$. We aim at $\text{ARL}_0^* = 370$ under normality and use the two-step approach with $\ell_1 = 5$. Estimating the ARL_0 from 1 000 time series from the model shown at the beginning of the section with $\mathcal{N}(0, 1)$-distributed noise leads to an estimated ARL_0 of 360 (standard error 11.7) for the ordinary RM regression and 357 (standard error 11.6) for the two-step approach. Hence, both values do not deviate by much from each other and from the nominal ARL_0.

4.4.2 ARL_0 under non-normal distributions

We proceed analogously to Subsection 3.4.2 and quantify the impact of non-normality on the control charts by computing ARL_0-ratios which compare the ARL_0 under normality to the ARL_0 under non-normality.

We show the results for $h = k = 10$ in Table 4.5. The corresponding results for $h = k = 20$ and $h = 20,\ k = 10$ can be found in Tables C.14 and C.15 in Appendix C. As in Subsection 3.4.2, we focus on $\alpha \leq 0.02$.

We omit the results for the ordinary t-chart in the tables because it is always strongly conservative, particularly for small values of α.

We start by describing the results for $h = k = 10$ in more detail and point out the general differences for the other subwindow widths. Our general observations are the following:

- Under the symmetric distributions, all but the update-randomised t-chart have similar ARL_0-values as under normality. The update-randomised t-chart is conservative by more than 10% under the t_2-distribution but leads to comparable results as under normality for the t_5- and the χ^2_3-distribution.

- Substantial deviations to the ARL_0-values under normality can be observed for the skewed distributions, especially for the χ^2_1-distribution. For example, for $\ell = 50$, the Wilcoxon chart is anti-conservative by about 15% and the HL12-chart by nearly 40% for small values of α. Opposed to this, the HL21-, the update-randomised t- and the MD1-chart are strongly conservative.

- For most charts, the ARL_0-ratios approach 1 for the different values of α when ℓ increases. Exceptions are the HL1-charts, which are still anti-conservative. For the HL11-chart, the performance seems even to deteriorate as its ARL_0-ratios are quite close to 1 for $\ell = 50$ but are much smaller when $\ell = 125$.

- Concerning the general performance of the charts, the Median chart and the HL22-chart show a quite stable behaviour with ARL_0-values close to 1 for all considered values of ℓ and each value of α.

- Increasing the subwindow widths to $h = k = 20$ leads generally to worse results if ℓ is small for all but the t_5-distribution. Compared to $h = k = 10$, the Median chart is strongly conservative under the χ^2_1-distribution with relative deviations of more than 30% for $\alpha \leq 0.0075$ when $\ell = 50$. In contrast, the HL22-chart still performs reasonably well.

- Using the larger regression window $\ell = 125$ for $h = k = 20$ leads to better results under the t_2- and the χ^2_3-distribution. However, the χ^2_1-distribution is still problematic for most charts. Exceptions are the Median chart, which shows a similar performance as for $h = k = 10$ and, to a limited degree, the HL22-chart. It is anti-conservative by roughly 10%.

- Under unequal subwindow widths $h = 20,\ k = 10$, all charts again have similar ARL_0-values as under normality for the t_5- and the χ^2_3-distribution, like for $h = k = 10$. Increasing ℓ improves the outcome under the t_2-distribution. We can again note difficulties under the χ^2_1-distribution. Similar to $h = k = 10$, the Median chart performs very well over all considered distributions and the HL22-chart also provides good results.

From our simulation results it seems that, like in Chapter 3, large discrepancies to the ARL_0-values under normality occur mostly for small values of α. This is particularly important when a large ARL_0-value is desired. If the charts are not too anti-conservative, it is still ensured to some degree that the true ARL_0-value for a specific value of α is large. Compared to the results from Chapter 3, the in-control run lengths of the rank-based charts do not have a distribution-free ARL_0. This can be explained by the correlation in the one-step-ahead forecast errors. In general, our results indicate that the more similar a distribution is to the normal distribution, the more likely it is that the ARL_0-values will also be similar, even for comparatively small values of ℓ. However, in order to ensure this, the subwindow widths need to be small compared to the widths of the regression window. In addition, it seems that the charts approach their behaviour observed in Chapter 3 with increasing value of ℓ. This would explain why, for example, the HL1-charts become more anti-conservative when the regression window is large; see Table 4.1. Similar to Chapter 3, the HL22-chart leads to a good overall performance among the robust charts

Comment on the MRL_0

Similar to Subsection 3.4.2, we briefly discuss the MRL_0. Again, when considering the out-of-control performance, the MRL_0 is mainly relevant for structural breaks of a small magnitude.

The ordinary t-chart, the Wilcoxon, and the Median chart show a similar performance as observed for the ARL_0-ratios. The update-randomised charts show the largest deviations between normality and non-normality again for the χ^2_1-distribution. The MRL_0-ratios of

Table 4.5: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the residual charts with subwindow widths $h = k = 10$ and different widths ℓ of the regression window. All values are rounded to two decimal places.

Control chart	α	ℓ =50				ℓ =125			
		t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
update random. t-chart	0.0025	1.02	1.04	1.00	1.11	0.99	0.92	0.95	0.88
	0.0050	1.03	1.09	1.04	1.17	1.00	0.97	0.99	0.96
	0.0075	1.04	1.11	1.05	1.20	1.01	1.04	1.02	1.01
	0.0100	1.04	1.12	1.04	1.22	1.00	1.05	1.02	1.04
	0.0125	1.04	1.12	1.03	1.21	1.01	1.07	1.02	1.05
	0.0150	1.03	1.14	1.03	1.22	1.01	1.07	1.01	1.06
	0.0200	1.04	1.15	1.03	1.24	0.99	1.06	1.01	1.06
Median chart	0.0025	0.97	0.99	1.00	0.97	1.01	1.03	1.05	0.97
	0.0050	0.99	0.98	1.02	1.00	1.00	1.00	1.02	0.96
	0.0075	1.00	0.97	1.02	1.00	1.00	0.98	1.00	0.95
	0.0100	0.99	0.97	1.01	1.00	0.98	0.98	0.99	0.94
	0.0125	1.01	0.97	1.03	1.01	1.00	0.98	1.00	0.96
	0.0150	0.99	0.97	1.02	1.02	0.99	0.98	1.00	0.95
	0.0200	0.99	0.97	1.02	1.01	0.99	0.97	0.99	0.95
Wilcoxon chart	0.0025	1.05	1.10	0.94	0.84	1.01	1.01	0.99	0.89
	0.0050	1.04	1.08	0.95	0.84	1.02	1.04	0.99	0.90
	0.0075	1.03	1.07	0.94	0.84	1.02	1.04	0.98	0.91
	0.0100	1.01	1.05	0.94	0.83	1.00	1.02	0.98	0.90
	0.0125	1.01	1.05	0.94	0.83	0.99	1.01	0.98	0.90
	0.0150	1.01	1.04	0.94	0.84	0.98	1.01	0.97	0.90
	0.0200	1.00	1.03	0.94	0.84	0.97	1.00	0.96	0.90
HL11-chart	0.0025	0.97	0.96	1.05	0.92	0.96	0.86	0.87	0.54
	0.0050	0.96	0.96	1.02	0.97	0.95	0.91	0.90	0.63
	0.0075	0.97	0.97	1.02	0.99	0.94	0.93	0.91	0.70
	0.0100	0.98	0.98	1.02	1.04	0.97	0.96	0.93	0.76
	0.0125	0.98	0.98	1.03	1.06	0.97	0.96	0.95	0.80
	0.0150	0.98	0.98	1.03	1.09	0.98	0.97	0.95	0.84
	0.0200	0.98	1.01	1.06	1.12	0.97	0.98	0.97	0.91
HL12-chart	0.0025	0.97	0.94	1.00	0.62	0.98	0.88	0.91	0.42
	0.0050	0.96	0.95	0.96	0.68	0.97	0.93	0.91	0.50
	0.0075	0.98	0.96	0.95	0.74	0.97	0.95	0.89	0.57
	0.0100	0.99	0.97	0.96	0.79	0.97	0.95	0.90	0.62
	0.0125	0.98	0.97	0.96	0.82	0.97	0.95	0.91	0.67
	0.0150	0.98	0.97	0.96	0.86	0.98	0.96	0.91	0.71
	0.0200	0.98	0.98	0.94	0.92	0.98	0.97	0.91	0.78
HL21-chart	0.0025	0.97	1.00	1.18	1.58	0.98	1.01	1.00	1.00
	0.0050	0.97	1.00	1.14	1.47	0.98	1.02	1.02	1.04
	0.0075	0.98	0.98	1.11	1.37	0.97	0.99	1.01	1.04
	0.0100	0.99	0.99	1.11	1.36	0.98	0.99	1.02	1.05
	0.0125	0.99	0.98	1.09	1.27	0.98	0.99	1.01	1.04
	0.0150	1.00	0.97	1.07	1.26	0.99	0.99	1.02	1.03
	0.0200	0.98	0.95	1.06	1.20	0.99	0.99	1.02	1.05
HL22-chart	0.0025	0.98	0.98	1.10	1.17	0.99	1.00	0.99	1.00
	0.0050	0.97	0.98	1.04	1.10	0.99	1.00	1.00	1.00
	0.0075	0.99	0.98	1.03	1.08	0.99	0.99	1.01	1.00
	0.0100	0.98	0.97	1.02	1.05	0.98	0.99	1.01	0.99
	0.0125	0.99	0.97	1.00	1.01	0.97	0.98	1.00	0.97
	0.0150	0.97	0.96	1.00	1.00	0.98	0.99	0.99	0.98
	0.0200	0.98	0.96	1.00	0.99	0.99	0.99	0.99	0.99
MD1-chart	0.0025	1.00	0.96	1.18	1.67	1.02	1.01	1.05	1.29
	0.0050	0.99	0.94	1.11	1.41	1.01	1.00	1.04	1.16
	0.0075	1.00	0.95	1.09	1.34	1.01	0.98	1.02	1.12
	0.0100	1.00	0.94	1.07	1.27	1.02	0.98	1.01	1.09
	0.0125	1.00	0.95	1.06	1.22	1.01	0.97	0.99	1.05
	0.0150	0.99	0.95	1.07	1.22	1.00	0.96	1.01	1.02
	0.0200	0.97	0.94	1.05	1.18	0.99	0.95	1.00	1.03
MD2-chart	0.0025	0.97	0.92	1.03	1.12	1.04	1.04	1.02	1.28
	0.0050	0.96	0.92	1.00	1.06	1.01	1.00	1.00	1.16
	0.0075	0.97	0.95	1.02	1.06	0.98	0.98	0.97	1.10
	0.0100	0.97	0.94	1.00	1.03	0.98	0.99	0.97	1.07
	0.0125	0.98	0.94	1.00	1.02	0.97	0.99	0.99	1.06
	0.0150	0.97	0.93	0.99	1.00	0.97	0.98	0.99	1.04
	0.0200	0.96	0.92	0.97	0.99	0.97	0.97	0.99	1.02

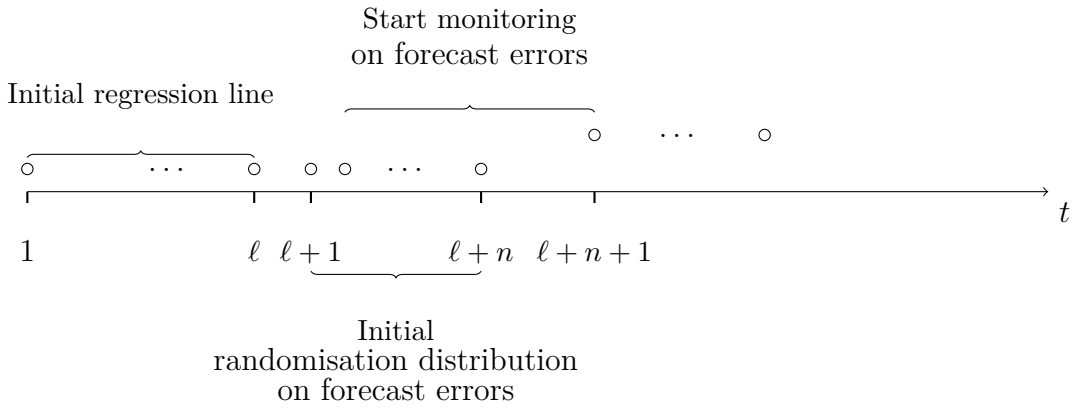

Figure 4.5: Schematic representation of the set-up for the out-of-control simulations. A persistent location shift starts at time $t = \ell + n + 1$. The first ℓ observations are used to calculate the first regression line. The initial randomisation distribution is computed from the forecast errors at the time points $\ell + 1, \ldots, \ell + n$. The monitoring begins with the window $(Y_{\ell+2}, \ldots, Y_{\ell+n+1})$.

the HL22-chart are similar to the corresponding ARL_0-ratios, underlining its stability as opposed to the other randomised charts, which show an MRL_0-behaviour other than for the ARL_0. For example, for $h = k = 10$ the MRL_0-ratios of the MD1-chart are close to 1 for all considered distributions, whereas the chart is strongly conservative with respect to the ARL_0 under the χ^2_1-distribution.

4.4.3 Out-of-control analysis

We now turn to the out-of-control analysis and proceed analogously to Subsection 3.4.3. First, we study the performance under the shift alternative among the different noise distributions. Afterwards, we briefly investigate the capabilities of the residual charts to detect sudden trend changes. Concluding this subsection, we show how outliers can influence the detection performance.

Based on our experiences from the simulations for a locally constant signal in Chapter 3, we restrict our analysis to the subwindow widths $h = k = 10$. For larger subwindows, we expect higher detection rates due to the increased power of the local tests, as long as the regression window is sufficiently long.

Basic settings for the simulations

Figure 4.5 illustrates the basic set-up for the out-of-control simulations. In general, we use the first ℓ observations of the generated time series to compute the first regression line. The observations at the time points $\ell + 1, \ldots, \ell + n$ are used to calculate the first randomisation distribution for the update-randomisation principle. The structural break occurs at time point $\ell + n + 1$ and we perform the first test on the time points $\ell + 2, \ldots, \ell + n + 1$ so that only the rightmost observation in the sample is affected by the structural break.

As in Subsection 3.4.3, we select α^* to achieve $\mathrm{ARL}_0^* = 250, 370$ under normality. Like before, this does not necessarily mean that the charts have the same MRL_0. Again, the MD-charts have the smallest MRL_0-values under normality, about 110. The MRL_0-values of the other charts vary between 160 and 170.

Shift alternative

To investigate the detection quality for location shifts, we generate the time series from the model

$$Y_t = \begin{cases} \varepsilon_t, & t = 1, \ldots, \ell + n \\ \varepsilon_t + \Delta \cdot q_{\text{diff}}, & t = \ell + n + 1, \ldots, 20\,001 + \ell, \end{cases}$$

where q_{diff} again denotes the difference between the 84.13%- and the 50%-quantile of the noise distribution. We use multiples $\Delta = 0.5, 1, 1.5, 2$ of q_{diff} as shift height.

To summarise the detection speed, we use the MRE and the worst-case detection rates; see Subsection 3.4.3.

Table 4.6 shows the minimal MRL_1-values for each distribution and ARL_0^*-value separately for each shift height and $\ell = 50, 125$. Figure 4.6 displays the MRE-values and the worst-case detection rates for $\mathrm{ARL}_0^* = 250$ and $\ell = 125$. The corresponding results for $\mathrm{ARL}_0^* = 370$ and for $\ell = 50$ can be found in Figures B.10 to B.12 in Appendix B. We split the distributions into the same groups as in Subsection 3.4.3: all distributions, the symmetric distributions, and the asymmetric distributions. Again, we concentrate on the large shift heights with $\Delta \geq 1.5$.

We can summarise our main observations as follows:

- From Table 4.6 we can see that the charts are quite insensitive to small location shifts. This was also noted in Chapter 3; see Table 3.9. It seems that the residual charts are less influenced by small changes than those for the locally constant signal.

- Regarding the minimal MRL_1, the width of the regression window seems to affect mainly the detection of small shifts. For the medium-sized shift height $\Delta = 1.5$, the detection is somewhat faster for $\ell = 125$ than for $\ell = 50$ under the symmetric distributions. Under the skewed distributions, there are no such big differences, which is because of the anti-conservatism of the charts.

- The ordering of the residual charts with respect to the worst-case detection rates, shown in Figure 4.6(b), is similar to the ordering observed in Chapter 3; see Figure 3.4(b). The HL-charts show the overall highest worst-case detection rates and are followed by the update-randomised t-chart, the Wilcoxon and the Median chart. All these charts outperform the ordinary t-chart, which shows its weakest performance under the t_2-distribution. While the HL-charts provide a nearly equally well performance over all considered noise distributions, the MD- and the Median

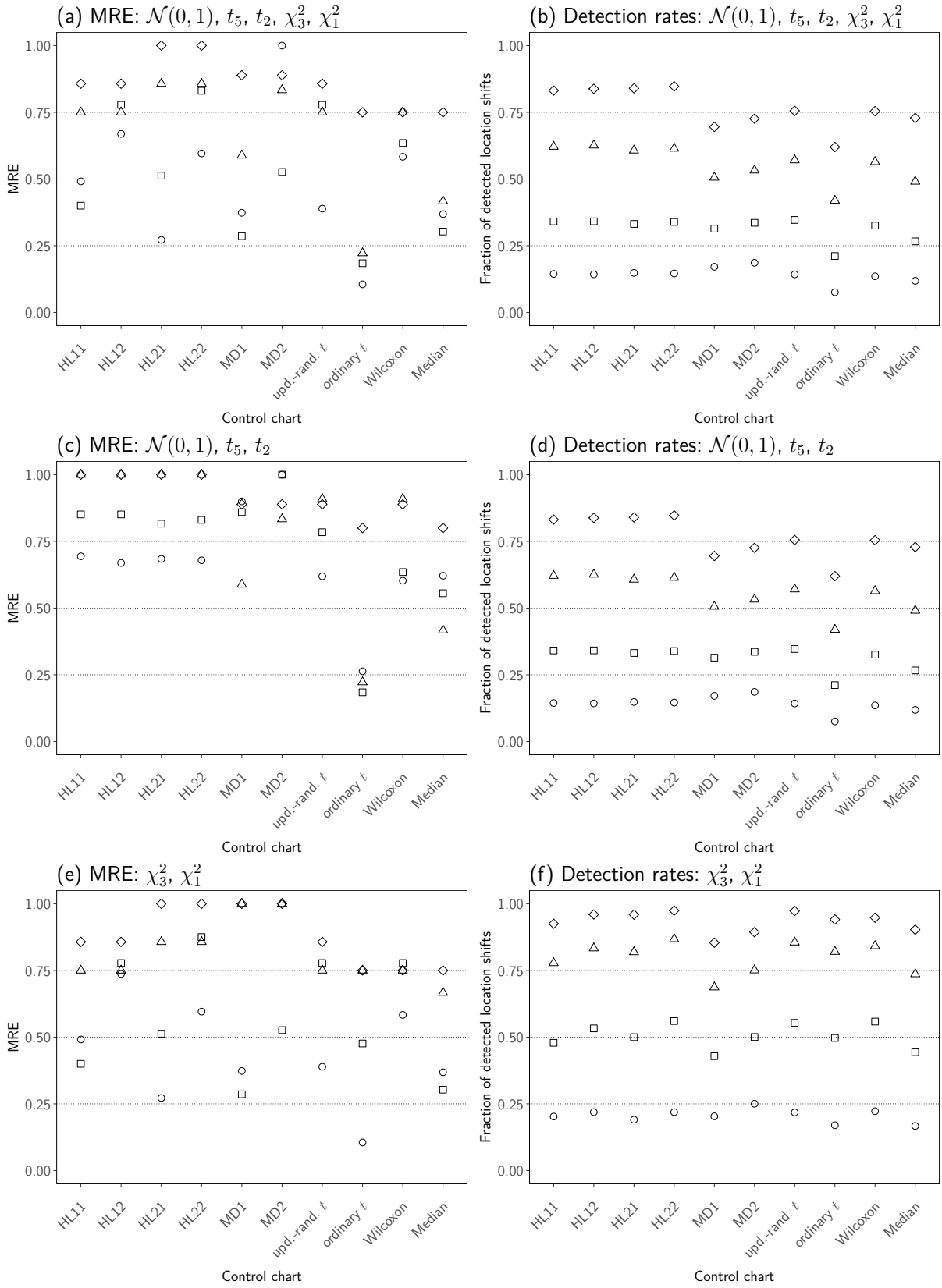

Figure 4.6: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates for the shift alternative over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality, subwindow widths $h = k = 10$, and $\ell = 125$. Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

Table 4.6: Minimal MRL_1-values for different noise distributions, separated by shift-height factor Δ, width ℓ of the regression window, and nominal ARL_0^* under normality, over all control charts for the subwindow widths $h = k = 10$.

			Shift-height factor Δ			
ℓ	ARL_0^*	Distribution	0.5	1	1.5	2
50	250	$\mathcal{N}(0,1)$	92	62	23	10
		t_5	92	57	24	10
		t_2	87	46	21	10
		χ^2_3	80	28	10	8
		χ^2_1	31	8	7	6
	370	$\mathcal{N}(0,1)$	121	88	32	10
		t_5	121	83	30	11
		t_2	117	68	24	10
		χ^2_3	105	37	10	8
		χ^2_1	44	9	7	6
125	250	$\mathcal{N}(0,1)$	93	54	10	8
		t_5	91	49	10	8
		t_2	90	40	10	8
		χ^2_3	78	10	8	7
		χ^2_1	28	7	6	6
	370	$\mathcal{N}(0,1)$	125	78	10	9
		t_5	127	70	11	9
		t_2	125	60	10	9
		χ^2_3	108	25	8	7
		χ^2_1	41	8	7	6

charts deteriorate under the $\mathcal{N}(0,1)$-distributions. The Wilcoxon chart also loses some power under the heavy-tailed distributions.

- In general, a large regression window improves the worst-case detection rates compared to a small one. For example, for $\ell = 50$ and $\Delta = 2$, the worst-case values of the HL-charts are about 0.6 under $\mathrm{ARL}_0^* = 250$ compared to nearly 0.8 for $\ell = 125$.

- Regarding the MRE, all charts perform nearly equally well for $\Delta = 2$, like observed in Chapter 3; see Figures 4.6(a) and 3.4(a). The update-randomised charts have a slight advantage and perform somewhat better than the ordinary charts under the heavy-tailed distributions. However, similar to Chapter 3, the MD-charts benefit from the reduced MRL_0 under normality.

- The only setting in which the HL-charts perform rather weak is when ℓ is small and ARL_0^* is large. Then, the MRE-values are smaller than 0.6, which is because of the symmetric distributions.

- The individual MRE-values are slightly larger than in Chapter 3 because we omitted

Table 4.7: Minimal MRL_1-values for the SCARM chart for different noise distributions, separated by shift-height factor Δ and nominal ARL_0^* under normality, for the subwindow widths $h = k = 10$.

		Shift-height factor Δ			
ARL_0^*	Distribution	0.5	1	1.5	2
250	$\mathcal{N}(0,1)$	167	157	141	114
	t_5	181	170	154	131
	t_2	198	190	174	153
	χ^2_3	171	152	119	71
	χ^2_1	174	139	82	15
370	$\mathcal{N}(0,1)$	255	246	223	192
	t_5	268	259	239	213
	t_2	288	279	262	237
	χ^2_3	254	234	197	141
	χ^2_1	240	203	142	56

the simulative charts. These were anti-conservative under the skewed distributions, giving them very small MRL_1-values under such distributions.

The results show that charts that perform reasonably well in case of a locally constant signal without a prior regression also lead to good residual charts. However, a large regression window should be preferred to improve the detection quality. A possible reason is a decreasing variability of the forecast errors for an increasing value of ℓ, so that location shifts are less likely to be covered by noise. In addition, for small values of ℓ, a comparatively small number of shifted observations suffices to affect the RM estimators. Thus, the location shift is less well represented in the sequence of forecast errors.

An intuitive alternative to the herein discussed residual charts is a control chart that makes use of the SCARM test as defined in Equation (2.5). The test can also be applied in a moving time window to detect location shifts. However, in some additional simulation studies, we found that the resulting control chart is not competitive in terms of detection quality. The RM slope estimator reacts only slowly to structural breaks, which leads to a comparatively large detection delay even for large jump heights; see Table 4.7.

Comment on the ARL_1 under the shift alternative

Results for the MRE based on the ARL_1 are shown in Figures B.13 and B.14 in Appendix B for the subwindow widths $h = k = 10$, $\ell = 50, 125$, and $ARL_0^* = 250$ under normality.

Like for the locally constant signal, the MRE-values for most charts are smaller than those for the MRL_1, albeit to a lesser extent. One reason could be the omission of the simulative charts in the comparison. Moreover, as noted in Subsection 4.3.4, update randomisation apparently leads to smaller in-control run lengths than simplified randomisation, which can be advantageous for the simulation runs in which the structural break is missed. Another

possible explanation is the second structural break in the sequence of forecast errors, which can also lead to an alarm and, as a consequence, to smaller run lengths; see Figure 4.2. The results for both values of ℓ are very much alike.

The HL22-chart, which leads to quite convincing results with respect to the MRL_1 and the detection rate, shows a similar performance under the ARL_1, particularly for symmetric distributions. The Wilcoxon chart achieves the highest MRE-values. The weak spot for most charts seems to be asymmetry.

Trend alternative

In addition to detecting sudden level shifts, we can also use the residual charts to find abrupt trend changes that cannot be explained by the inherent variation of the signal. In analogously performed simulations like before, we replace the location shift with a slope change and generate the observations from the model

$$Y_t = \begin{cases} \varepsilon_t, & t = 1, \ldots, \ell + n \\ b \cdot (t - \ell - n), & t = \ell + n + 1, \ldots, 20\,001 + \ell \end{cases}$$

in 10 000 replications with $b = 0.05, 0.1, 0.2, 0.5, 0.8$. We only consider the case $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ in the following. As motivated by Figure 4.2, sudden trend changes in the original time series also lead to structural breaks in the sequence of forecast errors. Therefore, we expect a similar ordering as observed for the shift alternative of the control charts when considering all distributions.

Figure 4.7 shows the relative efficiencies and the detection rates. Large changes can be found quite reliably by our control charts. However, due to the structure of the forecast errors, a strong slope change is needed for a reliable detection; see Figure 4.2.

Outliers under the shift alternative

We now evaluate briefly how the control charts behave when outliers are present in the neighbourhood of a location shift in otherwise clean data. The observations are generated from

$$Y_t = \begin{cases} \varepsilon_t, & t = 1, \ldots, 20 + \ell \\ \varepsilon_t + 3, & t = 21 + \ell, \ldots, 20\,001 + \ell, \end{cases}$$

where we assume that $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, which corresponds to the model used in Subsection 3.4.4. Again, we consider the settings of one and two negative outliers of decreasing size $\eta_t = -5, -10, -15, -20$ after the location shift at the time points $t = \ell + 26$ and $\ell + 27$.

Figure 4.8 shows the relative efficiencies and the detection rates for $\ell = 125$ and $\mathrm{ARL}_0^* = 250$. We gain similar conclusions as for the case of a locally constant signal: The robust control charts are able to resist the outliers rather well in both scenarios. The

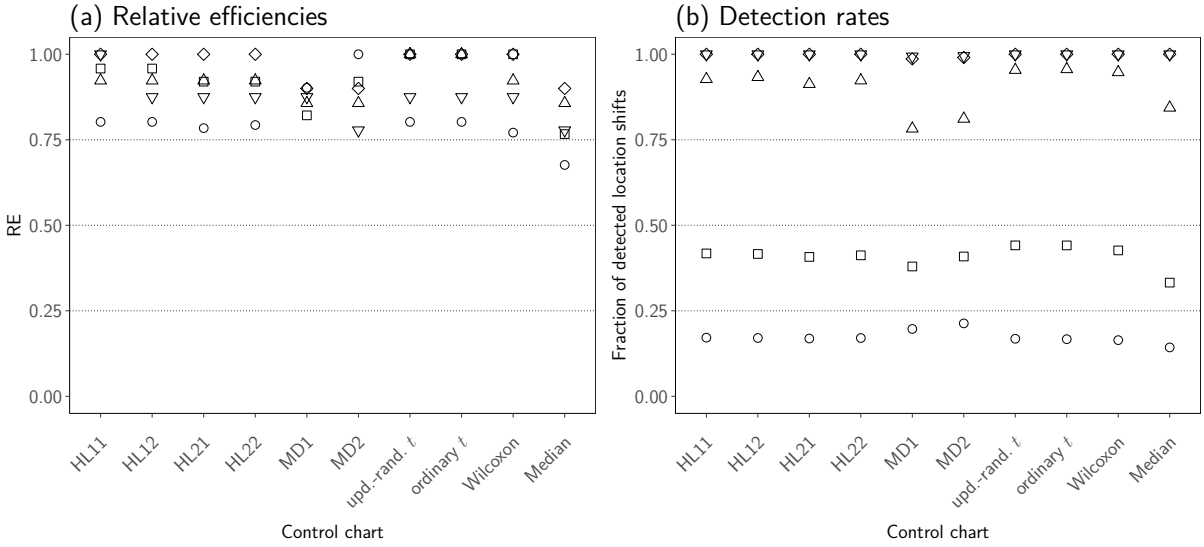

Figure 4.7: Relative efficiencies (RE) with respect to MRL_1 and detection rates for the trend alternative under normality. The control charts are applied with window widths $h = k = 10$ and $\ell = 125$, and tuned to $\mathrm{ARL}_0^* = 250$.
Slope: 0.05 (○), 0.1 (□), 0.2 (△), 0.5 (◇), 0.8 (▽).

Wilcoxon and the Median chart can deal with one outlier reasonably well, whereas the t-charts deteriorate in both settings.

4.4.4 Local AR(1)-models

In this subsection, we briefly investigate how our residual-chart principle performs when the time series can locally be described by an AR(1)-model instead of the simple linear model in Equation (4.2). We assume that

$$\begin{aligned} Y_{t+i} &= X_{t+i} + \eta_{t+i}, \text{ with} \\ X_{t+i} &= c_t + \phi_t \cdot X_{t+i-1} + \varepsilon_{t+i}, \; i = -\ell + 1, \ldots, 0, \end{aligned} \tag{4.5}$$

where $c_t \in \mathbb{R}$ is the local level, $\phi_t \in (-1, 1)$ is the value of the AR-parameter in the window, and ε_{t+i} and η_{t+i} denote noise and outliers as before. Our goal is to detect sudden changes in c_t under the assumption that the AR-parameter is also slowly time-varying, but does not exhibit abrupt changes.

Using procedures for i.i.d. observations can lead to a substantially larger false-alarm frequency than desired (Montgomery and Mastrangelo, 1991; Knoth and Schmid, 2004).

In the sequel, we provide some comments on how our proposed approach works on such data, considering just some special cases.

To illustrate that the residual charts for the simple linear model in Equation (4.2) are not directly applicable without some loss, we exemplarily apply the ordinary t-chart with subwindow widths $h = k = 10$, $\ell = 50$, and $\mathrm{ARL}_0^* = 250$ under normality to 1 000 time

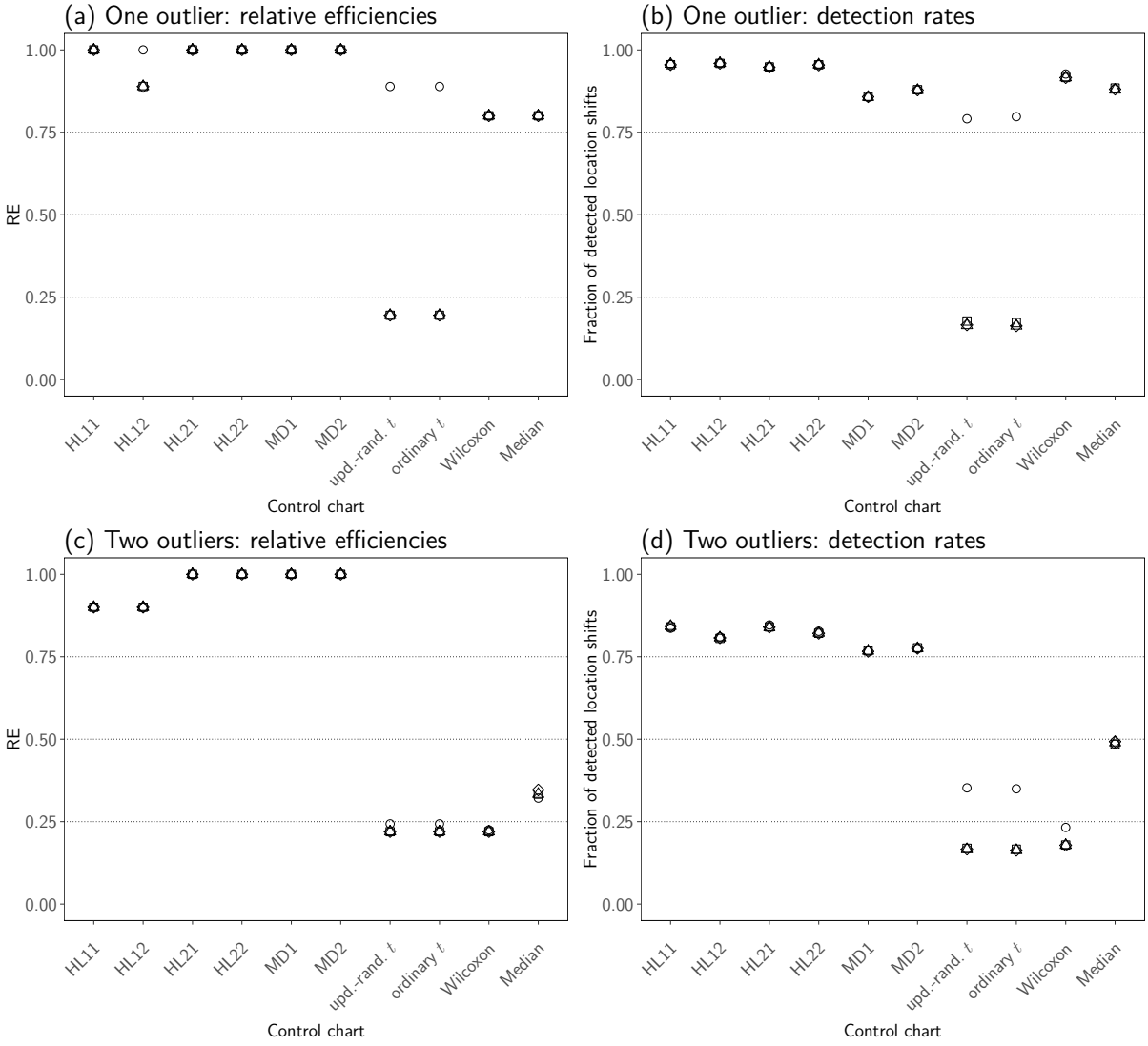

Figure 4.8: Relative efficiencies (RE) with respect to MRL_1 and detection rates under normality in outlier scenarios under the shift alternative. The control charts are applied with window widths $h = k = 10$ and $\ell = 125$, and tuned to $\mathrm{ARL}_0^* = 250$.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

series generated from the model

$$Y_t = \phi \cdot Y_{t-1} + \varepsilon_t,\ t = 1, \ldots, 20\,000,$$

where $\phi = 0.005, 0.2, 0.8$ and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

From the ARL_0-values shown in Table 4.8, it can be seen that the chart is unable to keep the desired ARL_0 and becomes anti-conservative even for a small autocorrelation. The degree of anti-conservatism increases in the value of ϕ.

To account for the new model, we replace the robust regression with a robust estimation of ϕ. The value of the AR-parameter corresponds to the lag 1 autocorrelation. Dürre et al. (2015) discuss various robust estimators for the autocorrelation function (ACF). In

Table 4.8: ARL_0 of the residual chart based on the ordinary t-test assuming a simple linear model when applied to data from an AR(1)-model with different values of the AR-parameter ϕ and additive $\mathcal{N}(0,1)$-distributed noise. The window widths are $h = k = 10$ and $\ell = 50$. The chart is tuned to $\text{ARL}_0^* = 250$. The values in brackets are the standard errors. All values are rounded to two decimal places.

$\phi = 0.005$	$\phi = 0.2$	$\phi = 0.8$
194.18 (5.93)	96.90 (3.12)	8.91 (0.34)

the following, we use the implementation from the `R` package `robts` (Dürre et al., 2017), which is freely available on R-Forge. We use the default settings and estimate the ACF by a Gnanadesikan-Kettenring estimator.

4.4.5 In-control performance for AR(1)-models

To study the in-control relationship between the ARL_0 and α, we perform simulations analogous to those for the simple linear model. We first study the behaviour under normality before turning to the stability under non-normal distributions.

In-control performance under the $\mathcal{N}(0,1)$-distribution

In 10 000 replications, we generate observations from the model

$$Y_t = \phi \cdot Y_{t-1} + \varepsilon_t,\ t = 1, \ldots, 20\,000 + \ell,$$

with $\phi = 0.2, 0.5, 0.8, \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, and $\ell = 50, 75, 100, 125$.

Table 4.9 shows the ARL_0-values and the standard errors for selected values of $\alpha, \ell = 125$, and $\phi = 0.2, 0.8$. In general, the ARL_0 decreases in ϕ for a fixed value of α. Although operating on one-step-ahead forecast errors, the value of the AR(1)-parameter has a large impact on the run length, unlike the regression parameters for the simple linear model. This can be explained by the missing equivariance structure of the AR-models and the bias towards zero of the correlation estimators, which increases in ϕ. We obtain similar results for the other values of ℓ. Hence, setting up the charts to obtain a desired ARL_0 requires the specification of ϕ.

For a fixed value of ϕ, it is again possible to describe the functional relationship between α and the ARL_0 by Equation (3.8) to a satisfying degree with a coefficient of determination of at least 0.98 for the linearised relationship in Equation (3.9).

Stability of the ARL_0 under non-normal distributions

To investigate the influence of the noise distribution on the ARL_0, we generate time series from the above model using i.i.d. noise from the t_5-, t_2-, χ^2_3-, and χ^2_1-distribution.

We exemplarily select α^* to achieve $\text{ARL}_0^* = 250$ under normality by Equation (3.10).

Table 4.9: ARL_0 for selected values of α under different AR(1)-models with additive $\mathcal{N}(0,1)$-distributed noise. The window widths are $h = k = 10$ and $\ell = 125$. The values in brackets are the standard errors. All values are rounded to one decimal place.

Control chart		Significance level α					
		0.005		0.02		0.05	
		$\phi = 0.2$	$\phi = 0.8$	$\phi = 0.2$	$\phi = 0.8$	$\phi = 0.2$	$\phi = 0.8$
t-charts	ordinary	280.7	120.7	85.9	48.2	39.7	25.0
		(2.7)	(1.2)	(0.9)	(0.5)	(0.4)	(0.3)
	update random.	283.9	124.2	87.1	48.9	39.9	25.3
		(2.8)	(1.2)	(0.9)	(0.5)	(0.4)	(0.3)
HL11-chart		285.8	134.1	91.6	52.6	40.4	26.4
		(3.2)	(1.6)	(1.0)	(0.6)	(0.5)	(0.3)
HL12-chart		288.8	137.6	89.9	52.9	39.5	26.3
		(3.1)	(1.5)	(0.9)	(0.6)	(0.4)	(0.3)
HL21-chart		289.1	141.1	93.4	55.1	40.9	27.3
		(3.3)	(1.7)	(1.0)	(0.6)	(0.4)	(0.3)
HL22-chart		289.9	142.7	90.9	54.4	40.3	27.1
		(3.2)	(1.6)	(1.0)	(0.6)	(0.4)	(0.3)
MD1-chart		263.8	167.1	98.1	68.5	46.5	34.2
		(4.3)	(2.8)	(1.5)	(1.0)	(0.7)	(0.5)
MD2-chart		279.1	179.3	100.4	70.4	47.9	35.1
		(4.8)	(3.2)	(1.5)	(1.1)	(0.7)	(0.5)
Rank charts	Wilcoxon	279.7	127.0	86.5	50.7	39.3	26.1
		(2.7)	(1.2)	(0.9)	(0.5)	(0.4)	(0.3)
	Median	242.8	138.1	87.7	56.8	31.7	24.6
		(2.4)	(1.3)	(0.9)	(0.6)	(0.3)	(0.3)

The estimated ARL_0-values and corresponding standard errors for $\ell = 125$ as well as all considered values of ϕ can be found in Table 4.10. All charts keep ARL_0^* under the $\mathcal{N}(0,1)$-distribution reasonably well, but none of the charts has a distribution-free ARL_0. However, as long as ϕ is not too large and the underlying distribution is symmetric and not very heavy-tailed, the deviations are moderate. Our observations are similar for other values of ℓ.

4.4.6 Out-of-control performance for AR(1)-models

Our out-of-control study is performed under the normality assumption. The data are generated from the model

$$Y_t = \begin{cases} X_t, & t = 1, \ldots, \ell + n \\ X_t + \Delta, & t = \ell + n + 1, \ldots, 20\,001 + \ell, \end{cases}$$

Table 4.10: ARL_0 under different distributions for different AR(1)-models. The control charts are tuned to $ARL_0^* = 250$ under normality and applied with window widths $h = k = 10$ and $\ell = 125$. The values in brackets are the standard errors. All values are rounded to one decimal place.

			Distribution				
ϕ	Control chart		$\mathcal{N}(0,1)$	t_5	t_2	χ^2_3	χ^2_1
0.2	t-charts	ordinary	248.8 (2.4)	282.1 (2.8)	406.4 (4.0)	254.3 (2.5)	271.7 (2.7)
		update	247.5 (2.4)	229.4 (2.3)	195.0 (2.3)	196.7 (2.0)	139.1 (1.6)
		random.					
	HL11-chart		248.4 (2.8)	223.4 (2.5)	160.2 (2.5)	170.4 (2.1)	107.9 (1.7)
	HL12-chart		249.7 (2.7)	225.7 (2.5)	169.7 (2.7)	179.4 (2.1)	106.7 (2.0)
	HL21-chart		248.4 (2.8)	230.6 (2.6)	185.8 (2.2)	198.1 (2.2)	150.3 (1.9)
	HL22-chart		248.1 (2.8)	230.6 (2.6)	189.5 (2.2)	208.7 (2.3)	159.1 (1.9)
	MD1-chart		248.1 (3.9)	242.2 (3.9)	212.8 (3.3)	212.8 (3.5)	151.1 (2.2)
	MD2-chart		248.8 (4.1)	244.7 (4.1)	232.9 (3.9)	230.9 (3.9)	184.3 (3.2)
	Rank charts	Wilcoxon	249.2 (2.4)	221.9 (2.1)	166.4 (1.7)	184.2 (1.8)	114.0 (1.2)
		Median	236.2 (2.3)	215.8 (2.1)	184.7 (1.9)	195.6 (2.0)	137.7 (1.4)
0.5	t-charts	ordinary	250.8 (2.5)	251.2 (2.5)	263.0 (2.6)	191.8 (1.9)	119.7 (1.2)
		update	251.3 (2.5)	211.3 (2.2)	149.4 (1.7)	163.4 (1.7)	83.8 (0.9)
		random.					
	HL11-chart		249.3 (2.8)	196.1 (2.4)	107.5 (3.0)	133.9 (1.6)	74.2 (1.1)
	HL12-chart		249.0 (2.7)	208.1 (2.4)	127.5 (3.7)	155.3 (1.8)	84.6 (1.8)
	HL21-chart		245.1 (2.8)	208.4 (2.6)	119.5 (1.4)	147.1 (1.7)	81.6 (1.0)
	HL22-chart		246.5 (2.8)	217.6 (2.6)	146.4 (1.8)	170.0 (1.9)	95.2 (1.1)
	MD1-chart		246.6 (4.0)	244.1 (4.1)	169.0 (2.7)	163.1 (2.5)	100.7 (1.5)
	MD2-chart		246.8 (4.3)	250.7 (4.5)	245.9 (4.8)	194.4 (3.1)	133.3 (2.2)
	Rank charts	Wilcoxon	251.3 (2.5)	203.2 (2.0)	123.3 (1.2)	161.2 (1.6)	87.8 (0.9)
		Median	243.4 (2.4)	213.5 (2.1)	152.5 (1.5)	178.1 (1.7)	103.6 (1.0)
0.8	t-charts	ordinary	243.4 (2.4)	136.4 (1.4)	50.5 (0.5)	89.4 (0.9)	34.9 (0.4)
		update	244.8 (2.5)	124.7 (1.3)	43.1 (0.5)	85.3 (0.9)	32.4 (0.4)
		random.					
	HL11-chart		238.7 (2.9)	109.0 (1.3)	43.2 (2.5)	77.7 (0.9)	34.6 (0.5)
	HL12-chart		243.6 (2.9)	121.6 (1.4)	49.5 (2.6)	87.2 (1.0)	34.9 (0.4)
	HL21-chart		237.1 (3.0)	110.5 (1.3)	38.6 (0.5)	75.7 (0.9)	33.2 (0.5)
	HL22-chart		241.0 (3.0)	122.4 (1.4)	42.9 (0.5)	84.1 (1.0)	34.7 (0.4)
	MD1-chart		239.7 (4.3)	129.4 (2.0)	46.0 (0.6)	85.0 (1.3)	42.8 (0.7)
	MD2-chart		242.3 (4.5)	142.8 (2.3)	53.1 (0.8)	92.3 (1.4)	45.7 (0.8)
	Rank charts	Wilcoxon	244.9 (2.4)	131.2 (1.3)	48.2 (0.5)	91.9 (0.9)	36.2 (0.4)
		Median	246.8 (2.4)	148.7 (1.5)	58.2 (0.6)	106.3 (1.1)	42.5 (0.4)

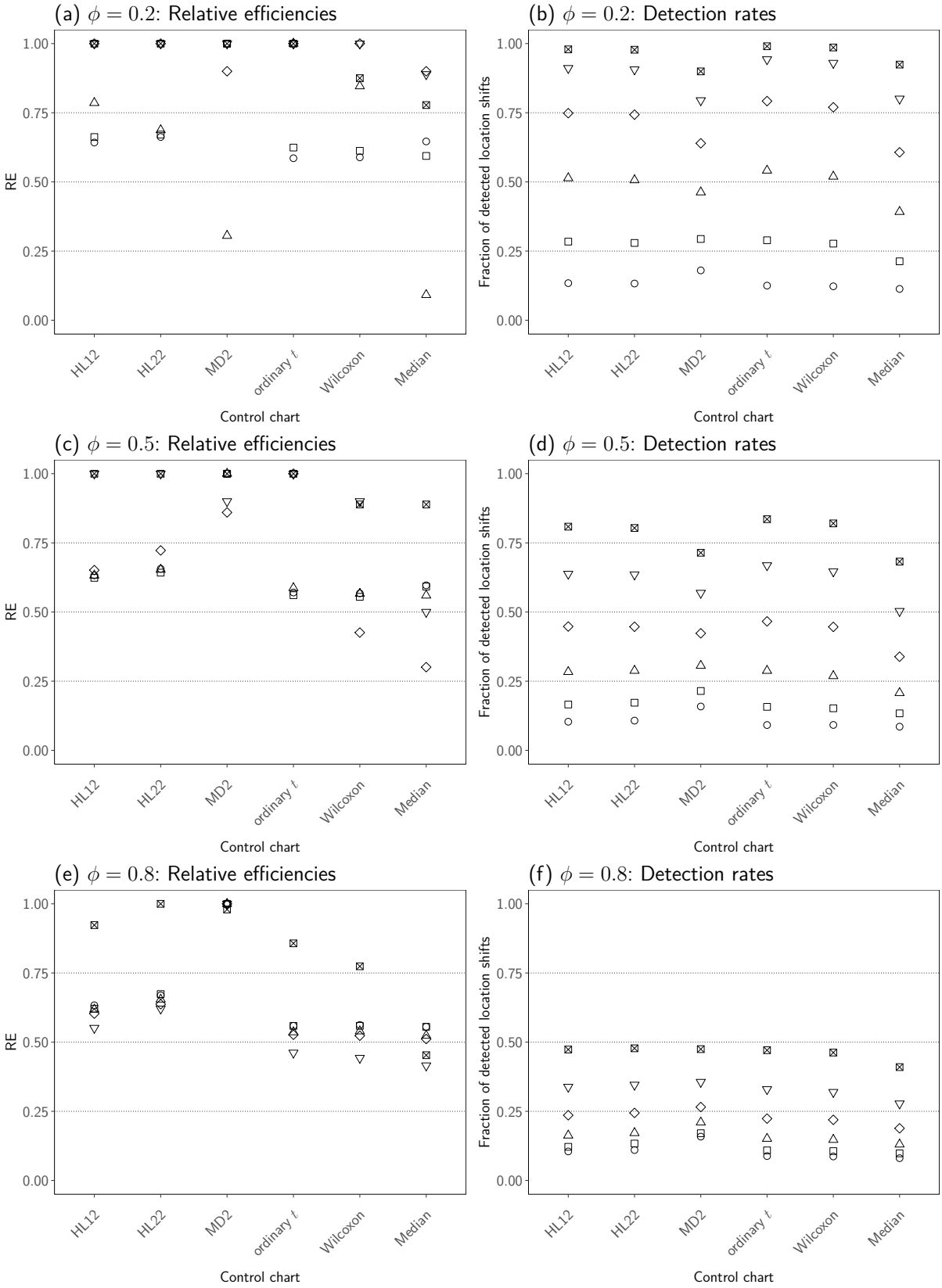

Figure 4.9: Relative efficiencies (RE) with respect to MRL_1 and detection rates under the $\mathcal{N}(0,1)$-distribution assuming a local AR(1)-model under the shift alternative with $\mathrm{ARL}_0^* = 250$ under normality. The window widths are $h = k = 10$ and $\ell = 125$.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇), 2.5 (▽), 3 (⊠).

where $\Delta = 0.5, 1, \ldots, 3$, $\ell = 50, 125$, and

$$X_t = \phi \cdot X_{t-1} + \varepsilon_t,\ t = 1, \ldots, 20\,001 + \ell,$$

with $\phi = 0.2, 0.5, 0.8$ and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Again, we tune the charts to $\text{ARL}_0^* = 250$ under normality. As the results are only for illustration, we consider only the charts based on the HL12-, HL22-, MD2-, and the ordinary t-test, the Wilcoxon and the Median test.

Figure 4.9 shows the relative efficiencies for the three values of ϕ and $\ell = 125$. In general, the ordering of the procedures is similar to the one observed for the simple linear model under normality. The main difference is that location shifts are more difficult to detect, the larger the value of ϕ is. Detecting shifts in highly positively correlated data is difficult, since the shift height needs to be large enough to stand out from other correlation-induced structures.

From this initial study, we conclude that the charts lead to acceptable results for large location shifts and small to moderate autocorrelations. However, there is still room for improvement. For example, Knoth and Schmid (2004) compare residual-based to modified CUSUM and EWMA charts. They conclude that the modified versions are better suited for positive autocorrelations.

4.5 Applications

In this section, we apply selected residual control charts to time series from the real-world applications introduced in Section 2.2 and a simulated time series. Like in Section 3.5, we intend to show that the principles are widely applicable, even though certain modifications for concrete applications may be necessary.

Again, we reduce the discussion to the ordinary t-chart, the Wilcoxon chart and the update-randomised HL22-chart. Similar to Section 3.5, we also applied the ordinary-randomised HL22-chart. The results turned out to be very similar to those of the update-randomised chart, so that we do not show the results here.

In Subsection 4.5.1, we investigate the performance on the PAMONO time series. We present the outcome on a simulated time series in Subsection 4.5.2. In Subsection 4.5.3, we discuss the application on the time series of heart-rate measurements. Lastly, in Subsection 4.5.4, we show how the control charts can be modified so that they can be used to detect sudden variability changes.

To set up the control charts, we follow Scheme 1 from Subsection 3.5.1.

4.5.1 PAMONO data

We start by applying the residual charts to the PAMONO time series. As in Chapter 3, we use the subwindow widths $h = k = 10$ and choose $\text{ARL}_0^* = 250$ under normality. Moreover,

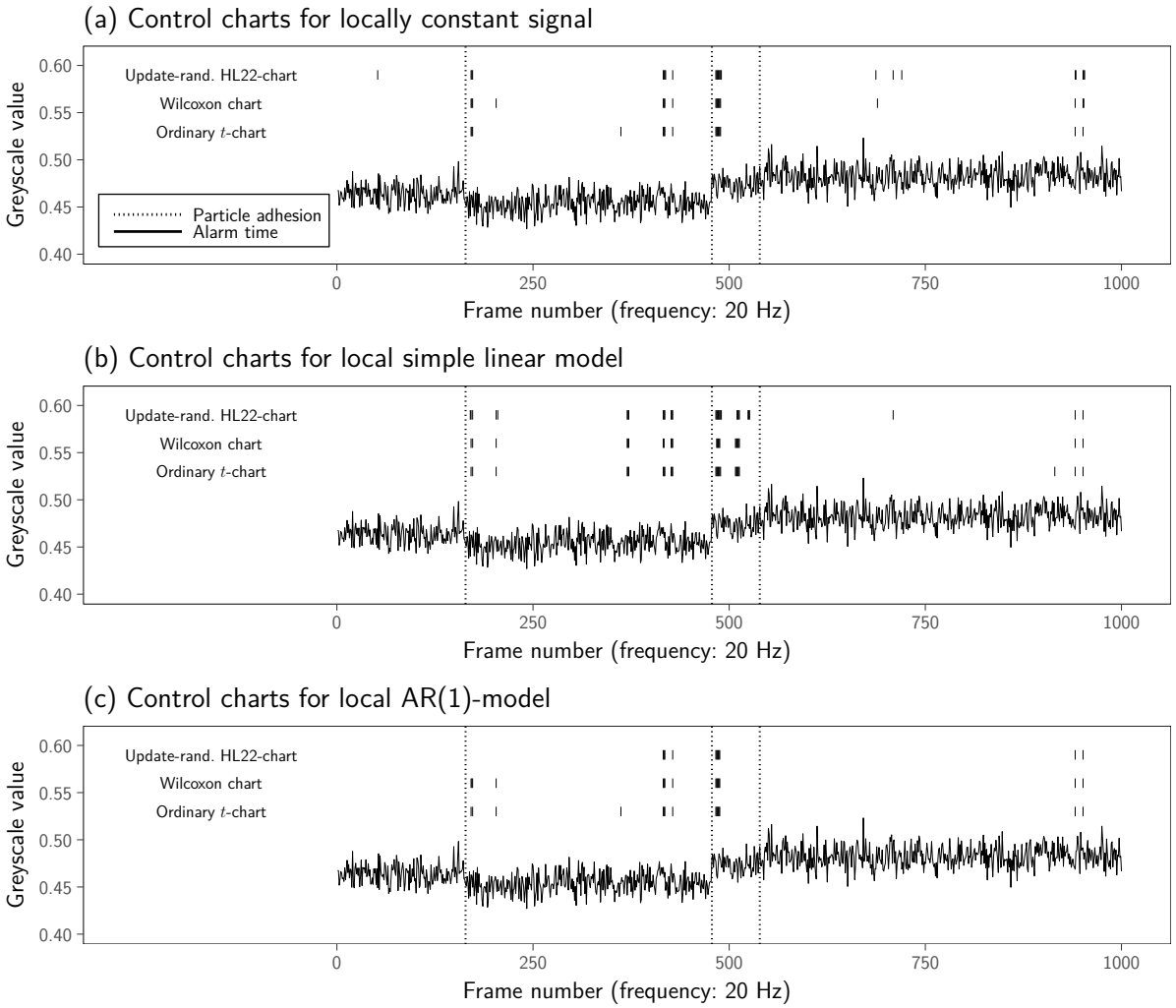

Figure 4.10: Application of selected residual control charts to the PAMONO time series. The charts are applied with the subwindow widths $h = k = 10$ and tuned to $\mathrm{ARL}_0^* = 250$ under normality. Panel (a) shows the results of charts assuming a locally constant signal for comparison; see Chapter 3. The residual charts in panels (b) and (c) are applied with a regression window of length $\ell = 100$.

we select the length $\ell = 100$ for the regression window. For the local AR(1)-model, we use $\phi = 0.2$, based on the estimated ACF from the first 100 observations.

In Figure 4.10, we show the alarm times for the residual charts (panels (b) and (c)) together with the results obtained in Subsection 3.5.2 (panel (a)). The residual charts based on the simple linear model lead to similar results as the corresponding charts without a prior regression. The only obvious difference is that the charts for the locally linear signal lead to a second alarm block behind the second location shift.

More differences can be found when assuming a local AR(1)-model. Then, the HL22-chart does not detect the first location shift, whereas the results for the t- and the Wilcoxon chart are similar to those for the simple linear model. One reason could be that we misspecified the AR(1)-parameter.

This example gives a first hint that the residual charts based on the simple linear model

can be a suitable alternative to the control charts from Chapter 3 by being applicable under weaker assumptions than they were designed for. However, more knowledge on the process might be required when using the residual charts based on the local AR(1)-model.

4.5.2 Time series with sine-wave signal

To study the benefits of the residual charts over the charts from Chapter 3, we consider a time series with a strong non-linear trend in form of a sine-wave, generated from

$$\mu_t = \sin\left(0.8 \cdot \left(-2\pi + \frac{t-1}{100}\right)\right) + \Delta_t + \frac{\varepsilon_t}{20} + \eta_t,\ t = 1, \ldots, 1257,$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} t_2$, with location shifts and outlier-generating process

$$\begin{aligned}\Delta_t &= -0.56 \cdot I_{\{492,\ldots,1257\}}(t) + 0.3 \cdot I_{\{1003,\ldots,1257\}}(t) \text{ and}\\ \eta_t &= I_{\{494\}}(t) + 0.5 \cdot I_{\{495\}}(t) + 1.5 \cdot I_{\{497\}}(t),\ t = 1, \ldots, 1257.\end{aligned}$$

This time series is similar to the one shown in Figure 4.1. We use heavy-tailed noise and outliers to challenge particularly the non-robust charts more.

All charts are applied with subwindow widths $h = k = 10$ and $\text{ARL}_0^* = 250$ under normality. For the residual charts, we choose $\ell = 100$. In addition, we select α^* under the assumption $\phi = 0.8$ for the residuals charts based on the local AR(1)-model. We obtained similar results as those shown here for other assumed values of ϕ. The alarm times are shown in Figure 4.2.

The charts for the locally constant signal (Figure 4.2(a)) seem to be unsuitable on time series with strong trends, even if the time window is rather short. This supports the impressions obtained from Figure 4.1.

Using residual charts reduces the number of false alarms drastically. There are two notable differences between the charts for the simple linear model (Figure 4.2(b)) and those for the AR(1)-model (Figure 4.2(c)): The latter do not detect the second location shift and all alarms occur only behind the first shift. Thus, missing the shift may be caused by conservatism due to using independent noise or the heavy-tailed distribution.

On the RM forecast errors, a quick detection of both location shifts is only achieved by the HL22-chart. The Wilcoxon chart detects the second shift. Both, the t-chart and the Wilcoxon chart show a large delay before alarms are given for the first location shift. This indicates that both procedures only cause alarms because of the slowly decreasing level in the sequence of forecast errors, similar to the one shown in Figure 4.2.

In total, this example illustrates the benefits of the residual charts, particularly in combination with a robust and efficient test. Together with the impressions from the PAMONO time series, they seem to be more versatile and should be preferred when the signal is known to be non-linear. However, it is important to verify the model assumptions. Further studies are necessary to work out the properties of the AR(1)-based charts under

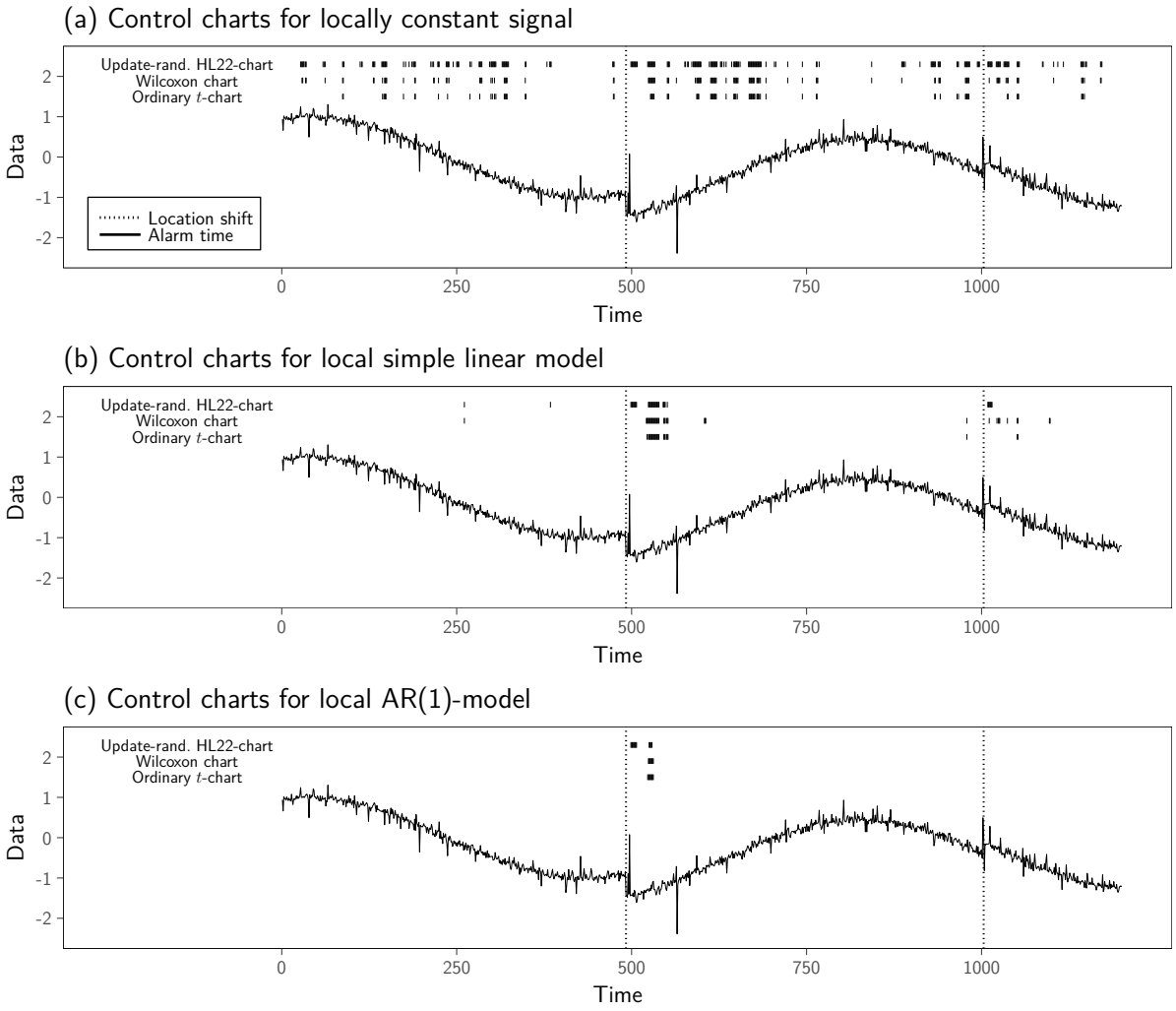

Figure 4.11: Application of selected control charts to a time series with a sine-wave signal. The charts are applied with the subwindow widths $h = k = 10$ and tuned to $\mathrm{ARL}_0^* = 250$ under normality. For the residual charts in panels (b) and (c), the regression window has the length $\ell = 100$.

independent noise.

In the following, we only use the residual charts.

4.5.3 Time series of heart-rate measurements

We now consider the intensive-care time series, which provides a bigger challenge due to its unclear structure. The time series is rather short with 400 observations, so that we set $\ell = 50$. The desired ARL_0 under normality is again $\mathrm{ARL}_0^* = 250$ and as subwindow widths we choose $h = k = 10$. For the AR(1)-model, we assume $\phi = 0.5$, based on an estimation of the autocorrelation function from the first 50 observations, where no notable change point is noticeable.

Figure 4.12 shows the results. Both model assumptions lead to similar alarm times. The AR(1)-assumption seems to set the focus on the larger notable structures. Compared to the application of the charts for a locally constant signal shown in Figure 3.7, the residual

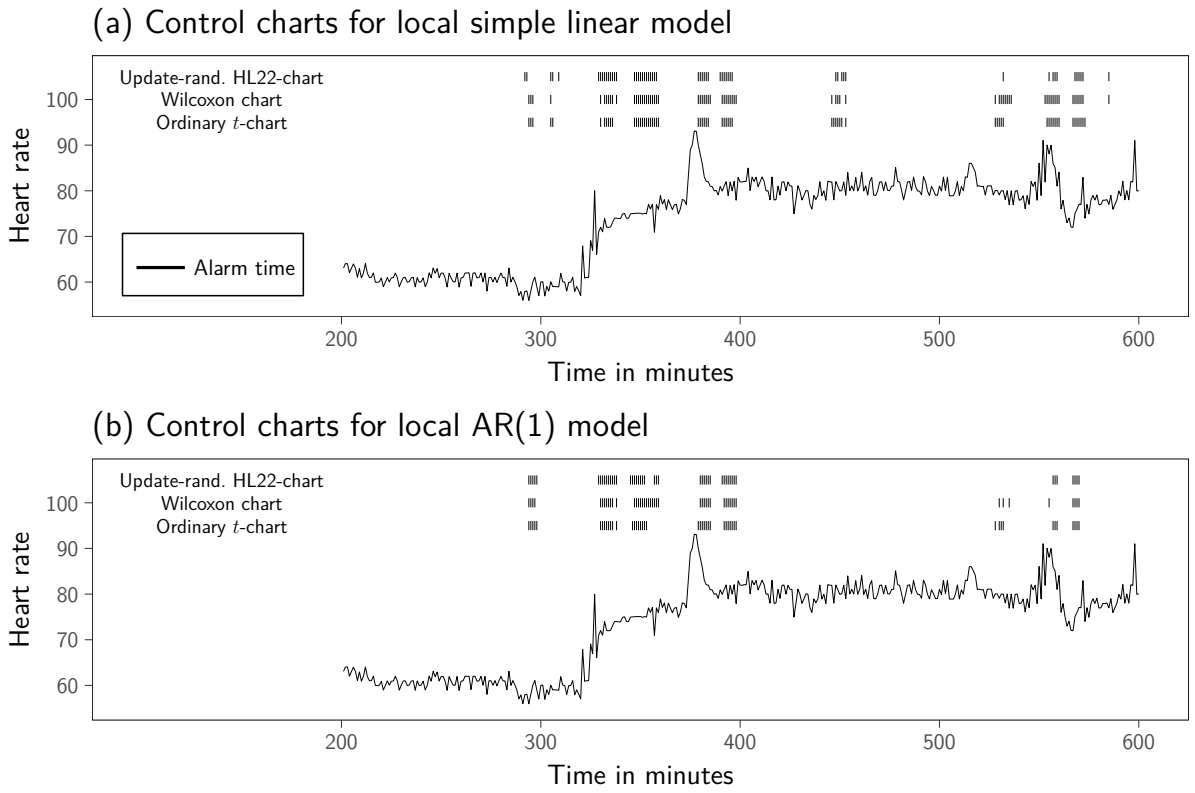

Figure 4.12: Application of selected residual control charts with $\mathrm{ARL}_0^* = 250$ under normality to the time series of heart-rate measurements. The window widths are $h = k = 10$ and $\ell = 50$.

charts for the local simple linear model do not provide a big advantage. Similar to the suggestion in Subsection 3.5.3, an improvement is possible by using a threshold approach. This, however, requires that the height of potential location shifts is known. Another way would be to develop the AR(1)-approach further as the initial result seems to be rather promising. However, this single time series provides only a limited data base.

4.5.4 Time series of crack-width measurements

Using a suitable data-transformation, the residual charts can be modified to detect sudden variability changes in a time series with a time-varying signal. To illustrate this, we use the time series of crack-width measurements.

Data transformation

We follow the approach of Fried (2012) as outlined in Subsection 2.4.2 and apply the control charts to log-transformed squared observations. We outline the general idea behind it briefly.

Similar to the time-varying signal $(\mu_t \colon\ t \in \mathbb{N})$ in Equation (2.1), the variance $(\sigma_t^2 \colon\ t \in \mathbb{N})$ might fluctuate slowly over time with only a few sudden changes. The model in Equation

(2.1) can be reformulated to

$$Y_t = \mu_t + \sigma_t \cdot \varepsilon_t,\ t \in \mathbb{N}, \tag{4.6}$$

when we ignore the outlier-generating process. Here, $(\mu_t\colon\ t \in \mathbb{N})$ is again locally approximately linear, but does not exhibit sudden trend or level changes. The sequence $(\sigma_t\colon\ t \in \mathbb{N})$ denotes the slowly varying true, but unknown, standard deviation, which we assume to be locally constant. The noise sequence $(\varepsilon_t\colon\ t \in \mathbb{N})$ is assumed to be white noise with variance $\mathrm{Var}\,(\varepsilon_t) = 1$. Thus, from Equation (4.6), we obtain $\mathrm{Var}\,(Y_t) = \sigma_t^2$.

The main difference to the previous considerations in this work is that we now are interested in detecting structural breaks in $(\sigma_t\colon\ t \in \mathbb{N})$ instead of $(\mu_t\colon\ t \in \mathbb{N})$. Given a time window at time t of width n, the standard deviation can be described by

$$\sigma_{t+i} = \begin{cases} \sigma_{t-}, & i = -n+1, \ldots, -k \\ \sigma_{t+}, & i = -k+1, \ldots, 0, \end{cases}$$

where $\sigma_{t+} = \nu_t \cdot \sigma_{t-}$, $\nu_t \in \mathbb{R} \setminus \{0\}$. The testing problem is now $H_{0,t}\colon \nu_t = 1$ vs. $H_{0,t}\colon \nu_t \neq 1$.

A general assumption of Fried (2012) is $\mu_t = 0$ for all $t \in \mathbb{N}$. Then, the observations can be transformed to

$$Z_t = \log\left(Y_t^2\right) = \log\left(\sigma_t^2\right) + \log\left(\varepsilon_t^2\right).$$

A scale change of magnitude ν_t then corresponds to a location shift in $(Z_t\colon\ t \in \mathbb{N})$ of size $\Delta_t = \log\left(\sigma_{t+}^2\right) - \log\left(\sigma_{t-}^2\right)$.

Time series exhibiting a trend do not fulfil the assumption of a centred process. We modify the approach by applying the log-transformation to the squares of the one-step-ahead forecast errors of the RM regression. Then, the two-sample tests are applied to $\log\left(\hat{e}_t^2\right)$ in Scheme 2 in Subsection 4.3.1.

Application to the time series

For our example, we only consider the forecast errors computed from the simple linear model. Due to the isolated peaks, using a robust control chart seems reasonable. We use a control chart based on the HL22-test. However, using the update-randomisation principle as proposed in Subsection 4.3.4 is not possible. Due to the length of the time series, the set of randomly drawn splits becomes very large and causes memory problems on the HPC cluster. As this example serves only as a proof of concept, we leave suitable modifications of this approach to future research and rely on the simplified randomisation principle as described in Subsection 3.3.2.

From some prior experiments, we found that small windows lead to alarms that are scattered all over the time range and make it difficult to draw conclusions about the true change points. Therefore, we choose rather large subwindows of widths $h = k = 200$. Such long windows make it time-consuming to determine the relationship between the ARL_0 and α in order to estimate α^*. Hence, we adopt the two-step approach used for RM

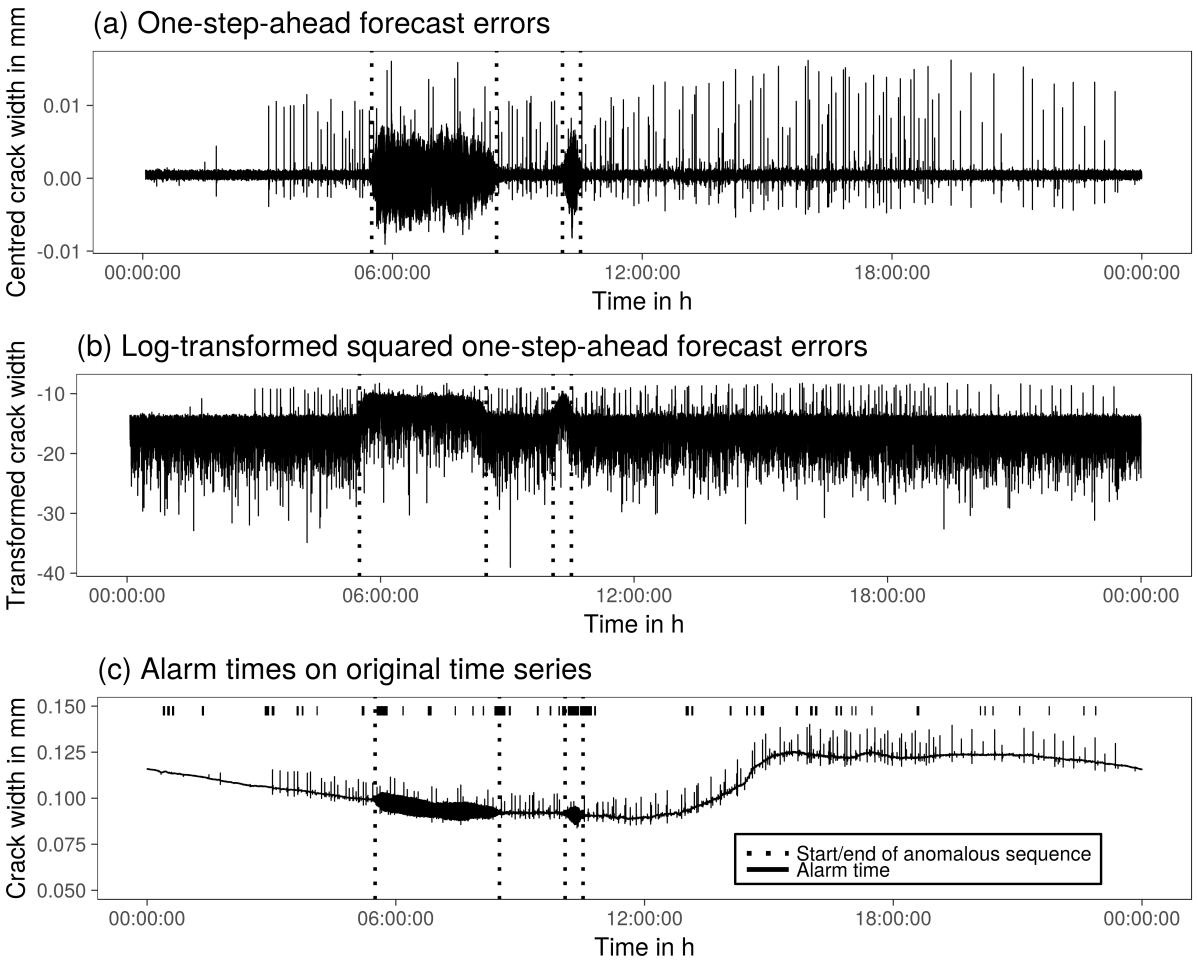

Figure 4.13: Detection of variability changes using the time series of crack-width measurements. The HL22-chart, tuned to $\mathrm{ARL}_0^* = 370$ under normality, is applied to log-transformed squared one-step-ahead forecast errors from RM regression. Panel (c) shows the results for the window widths $h = k = 200$ and $\ell = 125$.

regression; see Subsection 4.3.3. We split the subwindows into 20 blocks, each containing 10 observations and compute the sample median from each block. The tests are then performed on two samples of size 20, formed by the sample medians.

We tune the chart to $\mathrm{ARL}_0^* = 370$ under normality for $\ell = 125$. The choice of a larger ARL_0^* is motivated by the length of the time series, which requires more tests to be performed than in the previously considered examples.

The results are shown in Figure 4.13. Roughly 80% of the alarms are given between 5:30 and 11:30 o'clock, the time range in which the two anomalous sequences occur. Especially for the first sequence, which is quite long compared to the second one, the alarms are mainly given at the beginning and the end. Thus, as noted for the other application examples, the residual charts seem helpful as a first analysis step that provides a basis for post-processing. Concrete examples for post-processing in the bridge-monitoring application are discussed in Abbas et al. (2018).

4.6 Summary

In this chapter, we extend the control charts introduced in Chapter 3 to time series where the signal can be locally described by a regression model. We construct residual control charts where the two-sample tests are applied to one-step-ahead forecast errors of a robust estimation of the regression parameters.

The major findings can be summarised as follows:

- Due to the correlation of the forecast errors, the rank-based tests do not ensure distribution-free control charts.
- Using a robust regression estimator helps to retain structural breaks in the time series more reliably than a non-robust estimator.
- Most of the charts are anti-conservative under skewed distributions. However, an approximate distribution independence for in-control processes can still be achieved.
- The general principle is flexible and can also be useful for time series with a locally constant signal.
- Again, the HL22-chart leads to good overall results with respect to MRL_1, detection rate, and robustness.
- The introduced update-randomisation principle can improve the in-control run-length behaviour and the general ARL_1-performance of the charts as compared to the simplified randomisation.
- A large width of the regression window seems to be favourable in terms of the detection quality.
- In case of a local AR(1)-model, none of the charts is even approximately distribution free. Under a distributional assumption and known AR-parameter, we can still expect good detection results if the autocorrelation is not too large.

5 Summary and Outlook

In this chapter, we summarise the main results of this thesis (Section 5.1) and give some ideas of possible starting points for further research (Section 5.2).

5.1 Summary

We develop new control charts by applying two-sample location tests in a moving time window. The goal is to detect abrupt changes of the level or trend in the underlying, possibly slowly time-varying, signal of the time series under a restricted duration between two false alarms as measured by the run length.

Compared to ordinary control charts, the herein studied methods do not depend on a fixed target value. It is not necessary to estimate the process parameters from historical in-control data or make global parametric model assumptions. Only the most recent observations are relevant for the decision of whether a change point occurs. This allows for an adaptation to the signal course, avoiding the confusion of natural, process-inherent fluctuations with relevant changes that may point at a failure of the monitored system. Moreover, the charts can be easily set up to resist a specific number of consecutive outliers for an in-control process. This makes it possible to distinguish persistent structural breaks from short-term contamination.

We start our analysis under the assumption of a locally constant signal. Then, ordinary two-sample location tests can be used within our control-chart framework. We consider the well-known t-test and Wilcoxon rank-sum test. Moreover, we add selected robust competitors of the t-test to our comparison, which are based on a robust estimator for the location difference. These are a test based on the difference of the sample medians (MD), one that uses the difference of the one-sample Hodges-Lehmann estimators (HL1), and a test based on the two-sample Hodges-Lehmann estimator (HL2). From theoretical considerations, we find that control charts based on rank tests have a distribution-free in-control run-length distribution. Moreover, extensive simulations show that control charts using MD- or HL2-based statistics in combination with a randomisation principle make it possible to obtain charts that have an approximately distribution-free in-control average run length (ARL_0). Hence, for fixed values of the control limits, we can expect a similar ARL_0 under a broad spectrum of distributions. We also find that properties of the underlying tests are inherited by the control charts. Thus, a test which is powerful over a wide range of distributions leads to overall good detection properties. A big advantage of the robust charts over rank-based ones is their reduced sensitivity towards outliers. While

the latter lead to a convincing detection quality, they can deteriorate even for a small amount of contamination. Control charts based on robust two-sample tests can deal with a reasonable amount of outliers.

The in-control performance of a sequentially applied two-sample location test becomes worse when using it on a time series with strong trends, as these are confused with relevant location shifts. Assuming a locally linear signal, we combine the tests with robust regression techniques and obtain residual control charts that operate on one-step-ahead forecast errors of a local regression model. The performance of the charts depends on the assumptions made on the process. Using a simple linear model, it is possible to obtain similar properties as under the assumption of a locally constant model. However, the sample used for the regression has to be large to reduce the influence of the correlation in the forecast errors. We are confronted with more problems, when we assume an autoregressive model of order 1 (AR(1)-model). Here, we compute the one-step-ahead forecast errors from a robust estimator of the AR-parameter. The resulting charts are not even approximately distribution free. Nevertheless, if the underlying distribution is known and the autocorrelation is not too large, we can still expect a rather good detection quality.

From our simulations regarding distribution independence for an in-control process, robustness against outliers, and fast detection of a structural break, we conclude that the HL2-estimator in combination with a randomisation principle to compute the critical values leads to very promising results. Additionally, for computing the one-step-ahead forecast errors, it seems advisable to use a robust estimator not only because it is less sensitive to outliers, but it also helps to retain location shifts better in the sequence of forecast errors than a non-robust estimator.

5.2 Outlook

Our experiments show that the general principle of using two-sample tests as a new type of control scheme has some desirable detection properties. However, from applications to real-world data it becomes clear that some modifications may be necessary to have procedures that can deal with the specific peculiarities of the data. In the sequel, we give some ideas of how the control charts can be adjusted to some goals that may be of interest in practice.

Estimation of the change point

A structural break in the monitored time series typically leads to a sequence of consecutive alarms. We can combine our approach with techniques to estimate change points in order to determine the time point of the structural break. Examples for such procedures are the methods of Wu and Chu (1993) and Qiu and Yandell (1998).

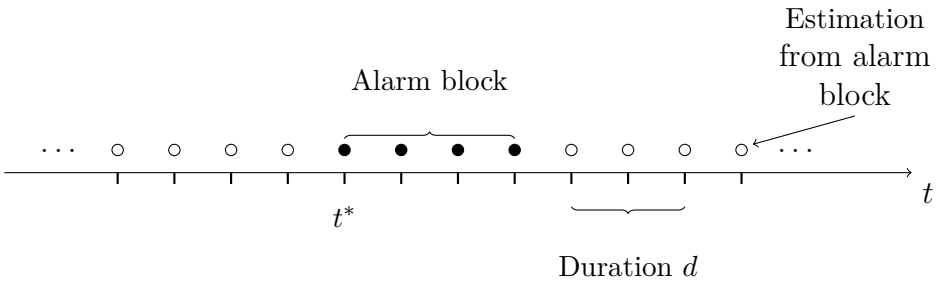

Figure 5.1: Visual representation for possible change-point estimation. The time of a change point is estimated from an alarm block which begins with t^*, after at least d time points without an alarm have passed.

The basic idea is to estimate the change point by the time index in an alarm block, where the absolute value of the test statistic is maximal; see for example Morell (2012, p. 33). This technique can also be used to retrospectively thin out the number of alarms in order to facilitate further processing of the data as in Abbas et al. (2016) or Abbas et al. (2018).

For the online context, one could use the following idea: Let t^* be the time point of an alarm which is possibly followed by further alarms. One could now wait for a specific duration after the last alarm in this sequence and estimate the true alarm time retrospectively from this block. A pictorial representation can be found in Figure 5.1.

Additional rules

In an application of the control charts to a time series from an online monitoring in an intensive care unit, the procedures do not protect against all irregular structures.

In that case, further rules might be necessary to determine whether an alarm can be traced back to an undesired effect or is a false alarm. One such approach is replacing the local significance test with a relevance test which compares the estimated height of the location shift to a specific threshold. This value includes prior knowledge on the magnitude of relevant structural breaks.

Another way to avoid the triggering of false alarms could be to include an additional parameter for the number of consecutive alarms. Then, an alarm would only be given if a minimal number of realisations of the test statistic is beyond the control limits.

Adaptation of the window widths

For the residual charts, we found that the performance can be improved by using longer time windows for the estimation of the regression parameters. However, at the beginning of the monitoring, it might not be clear how long the window should be in order to justify the local model assumptions. Moreover, the duration in which they can be safely assumed may change throughout the monitoring.

In such cases, it might be advisable to start with a short regression window and increase

it whilst there is no indication that the local assumptions may be violated. For this, the regression could be combined with adaptive techniques similar to the one proposed by Borowski and Fried (2014).

Improving the performance under the local AR(1)-model

Especially for a local AR(1)-model, we found that the procedures are very sensitive to the noise distribution. Perhaps another estimator for the autocorrelation function could be useful. Here, we used a Gnanadesikan-Kettenring estimator. However, as we observed for the simple linear model, the procedure with which the forecast errors are computed can strongly influence the performance of the chart.

Another way might be to choose control limits for which the charts have an ARL_0 that is larger than the desired value over a wide range of distributions. Even though the charts would be conservative, this concept would avoid giving more false alarms than desired. This is related to control charts with a guaranteed ARL-performance; see for example Weiß et al. (2018).

Other characteristics

As indicated by an application example in which we apply the charts to transformed forecast errors in order to detect variability instead of location changes, other characteristics of a process can be monitored by this principle as well. The two-sample location tests can be replaced with other suitable tests.

A Proofs

Proposition 1 (In-control run length of control charts based on rank tests)
Let $(Y_t\colon\ t \in \mathbb{N})$ *be a sequence of independent and identically distributed random variables following a continuous distribution with cumulative distribution function* F*. Let furthermore* $(T_t\colon\ t \geq n)$*,* $n \in \mathbb{N}$*, be a sequence of test statistics, where* T_t *is a rank statistic based on* $\boldsymbol{Y}_t^{(n)} = (Y_{t-n+1}, \ldots, Y_t)$*. Then, the in-control run length of a control chart based on this sequence of test statistics is distribution free.*

Proof
Let RL be the random variable which describes the run length of the control chart. In the following, we define the run length as the time point at which an alarm is given. This is different from the main part of this thesis, where the run length is defined as the number of tests until an alarm is triggered, given by $RL - n + 1$.

Let P_F be the probability measure under a continuous distribution with cumulative distribution function F. We have to show that $P_F(RL = t)$ does not depend on F for all $t \geq n$.

We consider the sequence $(I_t\colon\ t \geq n)$ of indicator variables with

$$I_t = \begin{cases} 1, & \text{alarm at sample } (Y_{t-n+1}, \ldots, Y_t) \\ 0, & \text{no alarm at sample } (Y_{t-n+1}, \ldots, Y_t). \end{cases}$$

Then

$$P_F\left(RL = t\right) = P_F\left(I_n = 0, I_{n+1} = 0, \ldots, I_{t-1} = 0, I_t = 1\right).$$

Each I_t depends on T_t, which is a function of the ranks

$$\tilde{\boldsymbol{R}}_t = \left(\tilde{R}(Y_{t-n+1}), \tilde{R}(Y_{t-n+2}), \ldots, \tilde{R}(Y_t)\right)$$

in $\boldsymbol{Y}_t^{(n)}$. Then, we can write $I_t = h\left(\tilde{\boldsymbol{R}}_t\right)$, where $h\colon \mathbb{N}^n \to \{0, 1\}$ is a function. Now, let

$$\boldsymbol{R}_t = (R(Y_1), R(Y_2), \ldots, R(Y_t))$$

be the vector of the ranks of all observations up to time point t. Obviously, $\tilde{\boldsymbol{R}}_t$ is a function $g_t\colon \mathbb{N}^t \to \mathbb{N}^n$ of $\boldsymbol{R}_t$ with $\tilde{\boldsymbol{R}}_t = g\left(\boldsymbol{R}_t\right)$, so that $I_t = h\left(g_t\left(\boldsymbol{R}_t\right)\right)$ and

$$\left(I_n, \ldots, I_t\right) = \left(h\left(g_n\left(\boldsymbol{R}_n\right)\right), \ldots, h\left(g_t\left(\boldsymbol{R}_t\right)\right)\right).$$

This implies that the distribution of $(I_n, \ldots, I_t)$ does not depend on F, since this is true for

$\boldsymbol{R}_t$, and hence the probability of $\{RL = t\}$ does not depend on F either for all $t \geq n$. $\square$

Proposition 2 (Properties of the one-step-ahead forecast errors)
Let $(Y_t\colon\ t \in \mathbb{Z})$ be a time series with $Y_t = \mu_t + \varepsilon_t$, where $\mu_t = \mu + \beta \cdot t$, $t \in \mathbb{Z}$. The process $(\varepsilon_t\colon\ t \in \mathbb{Z})$ of independent random variables is assumed to be strictly stationary with expectation $E(\varepsilon_t) = 0$ and variance $Var(\varepsilon_t) = \sigma^2$ for all $t \in \mathbb{Z}$. Let moreover $\hat{\mu}_{t-1}$ and $\hat{\beta}_{t-1}$ be unbiased and regression equivariant regression estimators for μ_{t-1} and β, obtained from a time window $(Y_{t-\ell}, \ldots, Y_{t-1})$ with a fixed design $i = -\ell + 1, \ldots, 0$. Then, the following results hold for the expectation, the variance, and the autocovariance function of the sequence of one-step-ahead forecast errors $(\hat{e}_t\colon\ t \in \mathbb{Z})$:

(a) $E(\hat{e}_t) = 0$ *for all* $t \in \mathbb{Z}$.

(b) $Var(\hat{e}_t) = \sigma^2 + Var\left(\hat{\mu}_{t-1}\right) + Var\left(\hat{\beta}_{t-1}\right) + 2 \cdot Cov\left(\hat{\mu}_{t-1}, \hat{\beta}_{t-1}\right)$ *for all* $t \in \mathbb{Z}$.

(c)

$$Cov(\hat{e}_t, \hat{e}_{t+s}) = \begin{cases} Var(\hat{e}_t)\,, & s = 0 \\ Cov\left(\hat{\mu}_{t-1}, \hat{\mu}_{t+s-1}\right) + Cov\left(\hat{\mu}_{t-1}, \hat{\beta}_{t+s-1}\right) + & \\ Cov\left(\hat{\beta}_{t-1}, \hat{\mu}_{t+s-1}\right) + Cov\left(\hat{\beta}_{t-1}, \hat{\beta}_{t+s-1}\right) - & \\ Cov\left(\varepsilon_t, \hat{\mu}_{t+s-1}\right) - Cov\left(\varepsilon_t, \hat{\beta}_{t+s-1}\right), & 1 \leq s \leq \ell \\ 0, & s > \ell \end{cases}$$

for all $t \in \mathbb{Z}$.

(d) $(\hat{e}_t\colon\ t \in \mathbb{Z})$ *is weakly stationary.*

Proof
For the one-step-ahead forecast $\hat{Y}_t = \hat{\mu}_{t-1} + \hat{\beta}_{t-1}$, we can write the one-step-ahead forecast error at time point $t \in \mathbb{Z}$ as

$$\hat{e}_t = Y_t - \hat{Y}_t = \mu + \beta \cdot t + \varepsilon_t - \left(\hat{\mu}_{t-1} + \hat{\beta}_{t-1}\right).$$

(a) As $\hat{\mu}_{t-1}$ is an unbiased estimator for μ_{t-1}, it is

$$\mathrm{E}\left(\hat{\mu}_{t-1}\right) = \mu_{t-1} = \mu + \beta \cdot (t-1)\,.$$

Then

$$\begin{aligned}
\mathrm{E}\left(\hat{e}_t\right) &= \mathrm{E}\left(\mu + \beta \cdot t + \varepsilon_t - \hat{\mu}_{t-1} - \hat{\beta}_{t-1}\right) \\
&= \mu + \beta \cdot t + \mathrm{E}\left(\varepsilon_t\right) - \mathrm{E}\left(\hat{\mu}_{t-1}\right) - \mathrm{E}\left(\hat{\beta}_{t-1}\right) \\
&= \mu + \beta \cdot t - \mu_{t-1} - \beta \\
&= \mu + \beta \cdot t - \mu - \beta \cdot (t-1) - \beta \\
&= \mu + \beta \cdot t - \mu - \beta \cdot t + \beta - \beta \\
&= 0.
\end{aligned}$$

(b)

$$\begin{aligned}
\mathrm{Var}\left(\hat{e}_t\right) &= \mathrm{Var}\left(\mu + \beta \cdot t + \varepsilon_t - \hat{\mu}_{t-1} - \hat{\beta}_{t-1}\right) \\
&= \mathrm{Var}\left(\varepsilon_t\right) + \mathrm{Var}\left(\hat{\mu}_{t-1}\right) + \mathrm{Var}\left(\hat{\beta}_{t-1}\right) + 2 \cdot \mathrm{Cov}\left(\hat{\mu}_{t-1}, \hat{\beta}_{t-1}\right) \\
&= \sigma^2 + \mathrm{Var}\left(\hat{\mu}_{t-1}\right) + \mathrm{Var}\left(\hat{\beta}_{t-1}\right) + 2 \cdot \mathrm{Cov}\left(\hat{\mu}_{t-1}, \hat{\beta}_{t-1}\right),
\end{aligned}$$

because ε_t and $\left(\hat{\mu}_{t-1}, \hat{\beta}_{t-1}\right)$ are independent.

(c) For the lag $s \in \mathbb{Z}$, the covariance between $\hat{e}_t$ and $\hat{e}_{t+s}$ is given by

$$\begin{aligned}
\mathrm{Cov}\left(\hat{e}_t, \hat{e}_{t+s}\right) &= \mathrm{Cov}\Big(\mu + \beta \cdot t + \varepsilon_t - \hat{\mu}_{t-1} - \hat{\beta}_{t-1}, \\
&\qquad \mu + \beta \cdot (t+s) + \varepsilon_{t+s} - \hat{\mu}_{t+s-1} - \hat{\beta}_{t+s-1}\Big) \\
&= \mathrm{Cov}\left(\varepsilon_t - \hat{\mu}_{t-1} - \hat{\beta}_{t-1}, \varepsilon_{t+s} - \hat{\mu}_{t+s-1} - \hat{\beta}_{t+s-1}\right) \\
&= \mathrm{Cov}\left(\hat{\mu}_{t-1}, \hat{\mu}_{t+s-1}\right) + \mathrm{Cov}\left(\hat{\beta}_{t-1}, \hat{\beta}_{t+s-1}\right) + \\
&\quad \mathrm{Cov}\left(\hat{\mu}_{t-1}, \hat{\beta}_{t+s-1}\right) + \mathrm{Cov}\left(\hat{\beta}_{t-1}, \hat{\mu}_{t+s-1}\right) + \\
&\quad \mathrm{Cov}\left(\varepsilon_t, \varepsilon_{t+s}\right) - \mathrm{Cov}\left(\varepsilon_t, \hat{\mu}_{t+s-1}\right) - \mathrm{Cov}\left(\varepsilon_t, \hat{\beta}_{t+s-1}\right) - \\
&\quad \mathrm{Cov}\left(\hat{\mu}_{t-1}, \varepsilon_{t+s}\right) - \mathrm{Cov}\left(\hat{\beta}_{t-1}, \varepsilon_{t+s}\right).
\end{aligned}$$

- For $s = 0$, we get the result in (b), since

$$\mathrm{Cov}\left(\varepsilon_t, \hat{\mu}_{t-1}\right) = \mathrm{Cov}\left(\varepsilon_t, \hat{\beta}_{t-1}\right) = 0$$

  because ε_t is independent of $\hat{\mu}_{t-1}$ and $\hat{\beta}_{t-1}$.

- For $|s| > \ell$, we get

$$\mathrm{Cov}\left(\hat{e}_t, \hat{e}_{t+s}\right) = 0.$$

  The time window $(Y_{t-\ell}, \ldots, Y_t)$ is used to calculate $\hat{e}_t$. Analogously, we use $(Y_{t+s-\ell}, \ldots, Y_{t+s})$ for $\hat{e}_{t+s}$. If $|s| > \ell$, both windows are disjoint, so that the two one-step-ahead forecast errors are uncorrelated.

- For $1 < s \leq \ell$, we get

$$\begin{aligned}\operatorname{Cov}(\hat{e}_t, \hat{e}_{t+s}) = {} & \operatorname{Cov}\left(\hat{\mu}_{t-1}, \hat{\mu}_{t+s-1}\right) + \operatorname{Cov}\left(\hat{\beta}_{t-1}, \hat{\beta}_{t+s-1}\right) + \\ & \operatorname{Cov}\left(\hat{\mu}_{t-1}, \hat{\beta}_{t+s-1}\right) + \operatorname{Cov}\left(\hat{\beta}_{t-1}, \hat{\mu}_{t+s-1}\right) - \\ & \operatorname{Cov}\left(\varepsilon_t, \hat{\mu}_{t+s-1}\right) - \operatorname{Cov}\left(\varepsilon_t, \hat{\beta}_{t+s-1}\right),\end{aligned}$$

because

$$\operatorname{Cov}(\varepsilon_t, \varepsilon_{t+s}) = \operatorname{Cov}\left(\hat{\mu}_{t-1}, \varepsilon_{t+s}\right) = \operatorname{Cov}\left(\hat{\beta}_{t-1}, \varepsilon_{t+s}\right) = 0,$$

as ε_t and ε_{t+s} are independent and ε_{t+s} is not used to calculate $\hat{\mu}_{t-1}$ and $\hat{\beta}_{t-1}$.

Let $\boldsymbol{\varepsilon}_{t-1}^{(\ell)}$ be the vector of random errors in the time window $\boldsymbol{Y}_{t-1}^{(\ell)}$. We can write the regression estimators as

$$\hat{\mu}_{t-1} = \mu_{t-1} + g_\mu\left(\boldsymbol{\varepsilon}_{t-1}^{(\ell)}\right) \quad \text{and} \quad \hat{\beta}_{t-1} = \beta_{t-1} + g_\beta\left(\boldsymbol{\varepsilon}_{t-1}^{(\ell)}\right)$$

with functions $g_\mu,\ g_\beta \colon \mathbb{R}^\ell \to \mathbb{R}$ because they are assumed to be regression equivariant and the Y_t are linear transformations of the ε_t with

$$Y_t = \varepsilon_t + (1, t) \cdot \begin{pmatrix} \mu \\ \beta \end{pmatrix} \text{ for all } t \in \mathbb{Z}.$$

Due to the strict stationarity of $(\varepsilon_t \colon\ t \in \mathbb{Z})$, the sequence $\left(\left(\hat{\mu}_{t-1}, \hat{\beta}_{t-1}, \varepsilon_{t-1}\right) \colon\ t \in \mathbb{Z}\right)$ is weakly stationary, implying that the sequence $(\hat{e}_t \colon\ t \in \mathbb{Z})$ is covariance stationary.

Therefore, we can summarise the autocovariance function by

$$\operatorname{Cov}(\hat{e}_t, \hat{e}_{t+s}) = \begin{cases} \sigma^2 + \operatorname{Var}\left(\hat{\mu}_{t-1}\right) + \operatorname{Var}\left(\hat{\beta}_{t-1}\right) + 2 \cdot \operatorname{Cov}\left(\hat{\mu}_{t-1}, \hat{\beta}_{t-1}\right), & s = 0 \\ \operatorname{Cov}\left(\hat{\mu}_{t-1}, \hat{\mu}_{t+s-1}\right) + \operatorname{Cov}\left(\hat{\mu}_{t-1}, \hat{\beta}_{t+s-1}\right) + & \\ \operatorname{Cov}\left(\hat{\beta}_{t-1}, \hat{\mu}_{t+s-1}\right) + \operatorname{Cov}\left(\hat{\beta}_{t-1}, \hat{\beta}_{t+s-1}\right) - & \\ \operatorname{Cov}\left(\varepsilon_t, \hat{\mu}_{t+s-1}\right) - \operatorname{Cov}\left(\varepsilon_t, \hat{\beta}_{t+s-1}\right), & 1 \leq s \leq \ell \\ 0, & s > \ell. \end{cases}$$

None of the covariances depends on the actual time point t, since the strictly stationary time series is always processed locally in the same way.

(d) The results (a), (b), and (c) do not depend on the time point t. Hence, it follows that $(\hat{e}_t \colon\ t \in \mathbb{Z})$ is weakly stationary.

□

B Figures

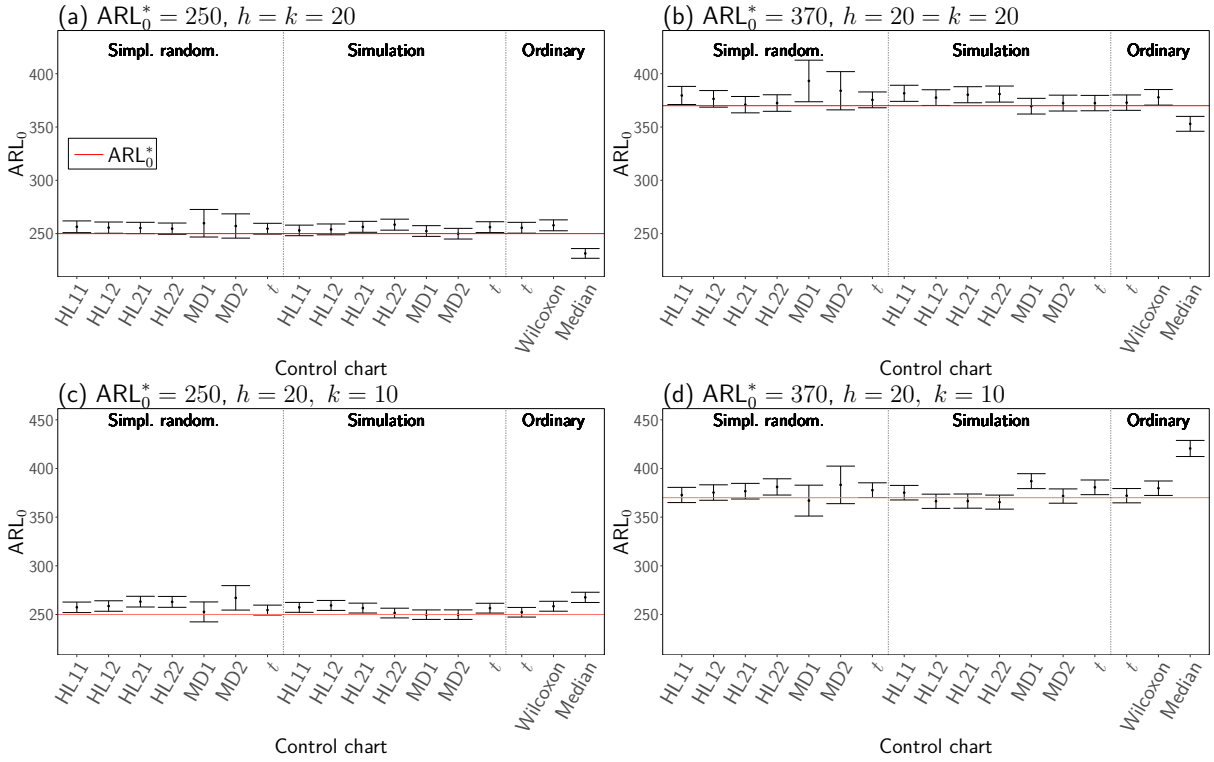

Figure B.1: Asymptotic 95%-confidence intervals for the ARL_0 after computing α^* from Equation (3.10) for different subwindow widths under the normality assumption. The vertical dotted lines separate between the different groups of control charts.

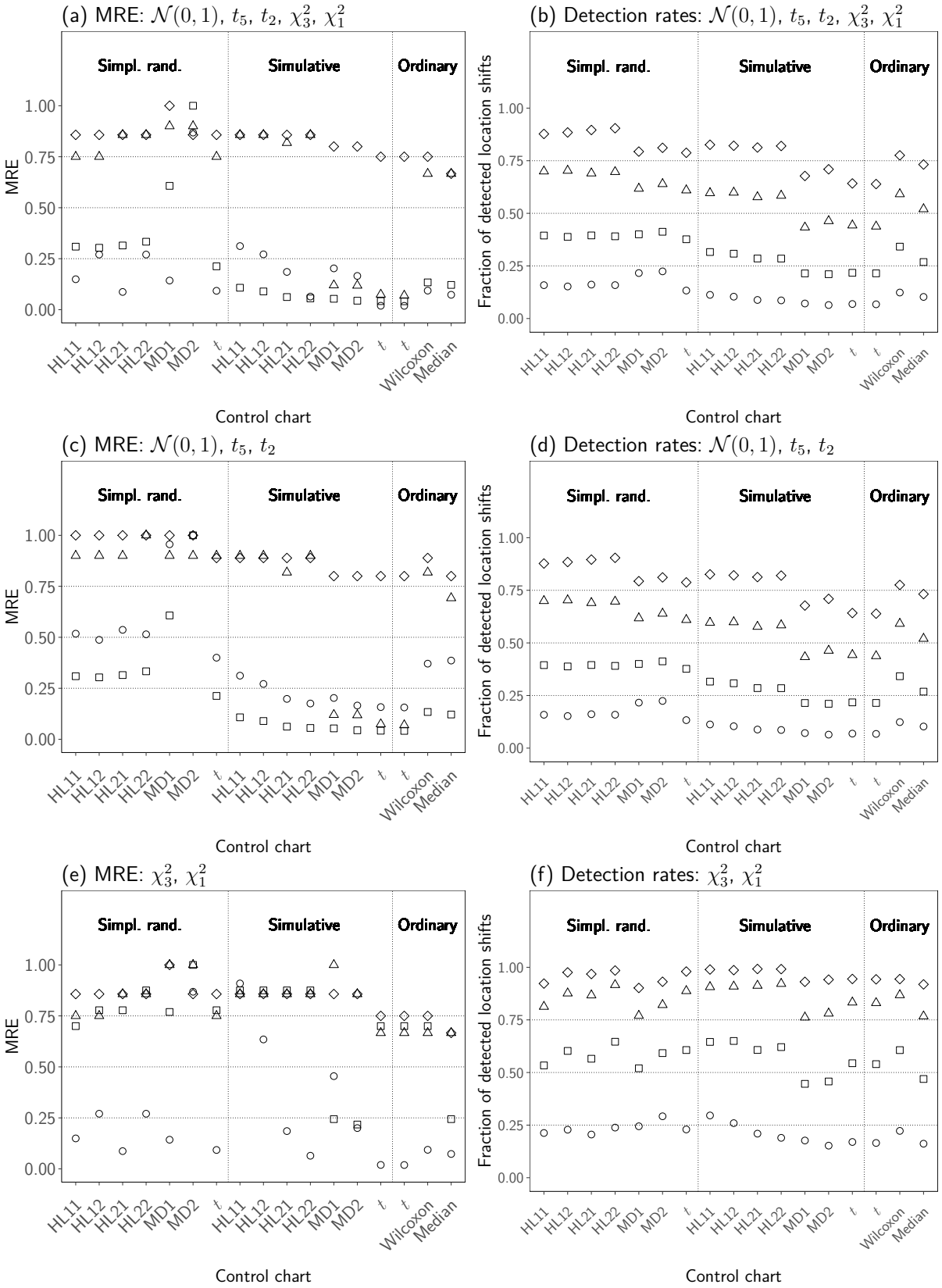

Figure B.2: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates over different groups of distributions for $\mathrm{ARL}_0^* = 370$ under normality and subwindow widths $h = k = 10$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

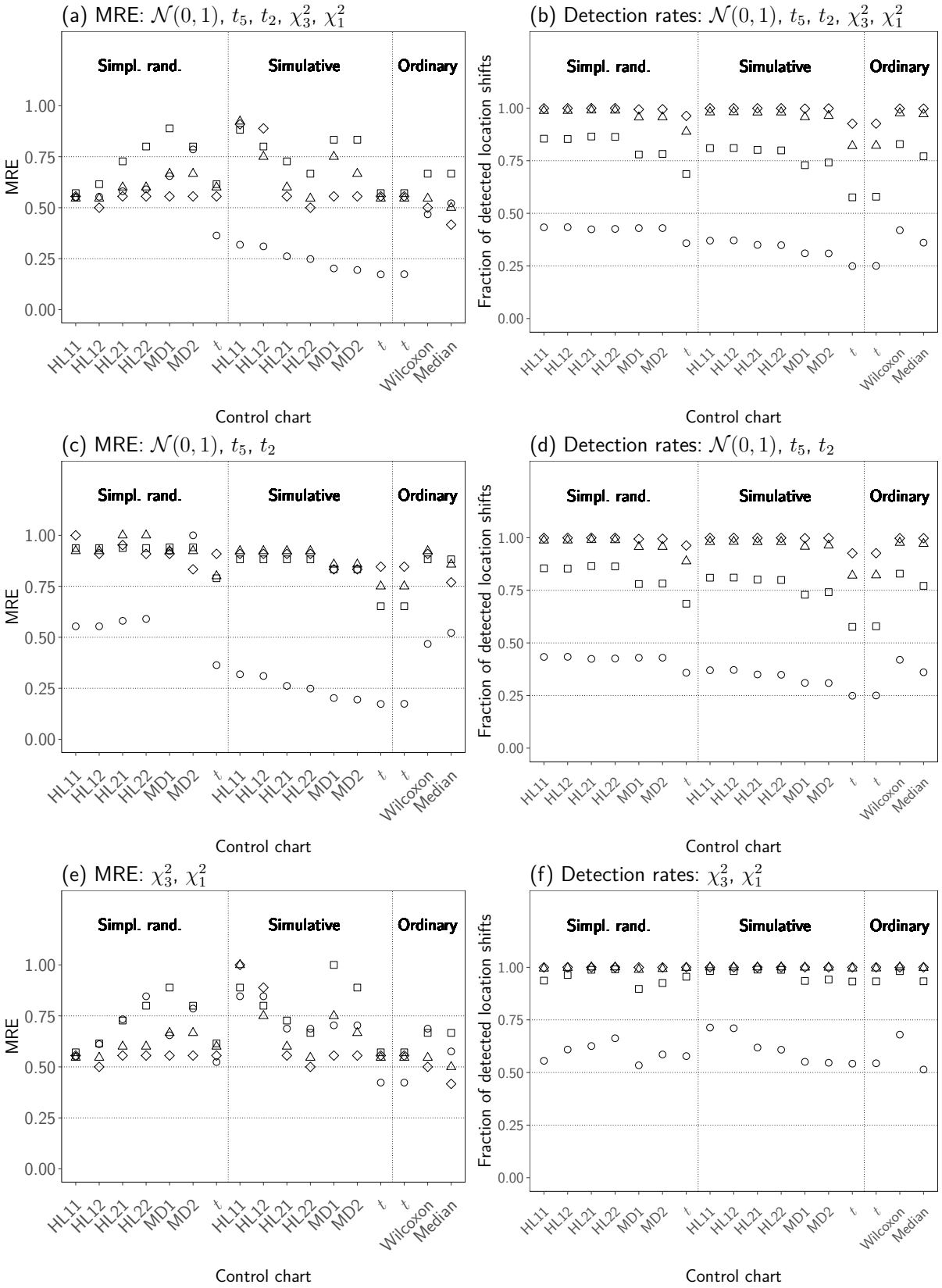

Figure B.3: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality and subwindow widths $h = k = 20$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

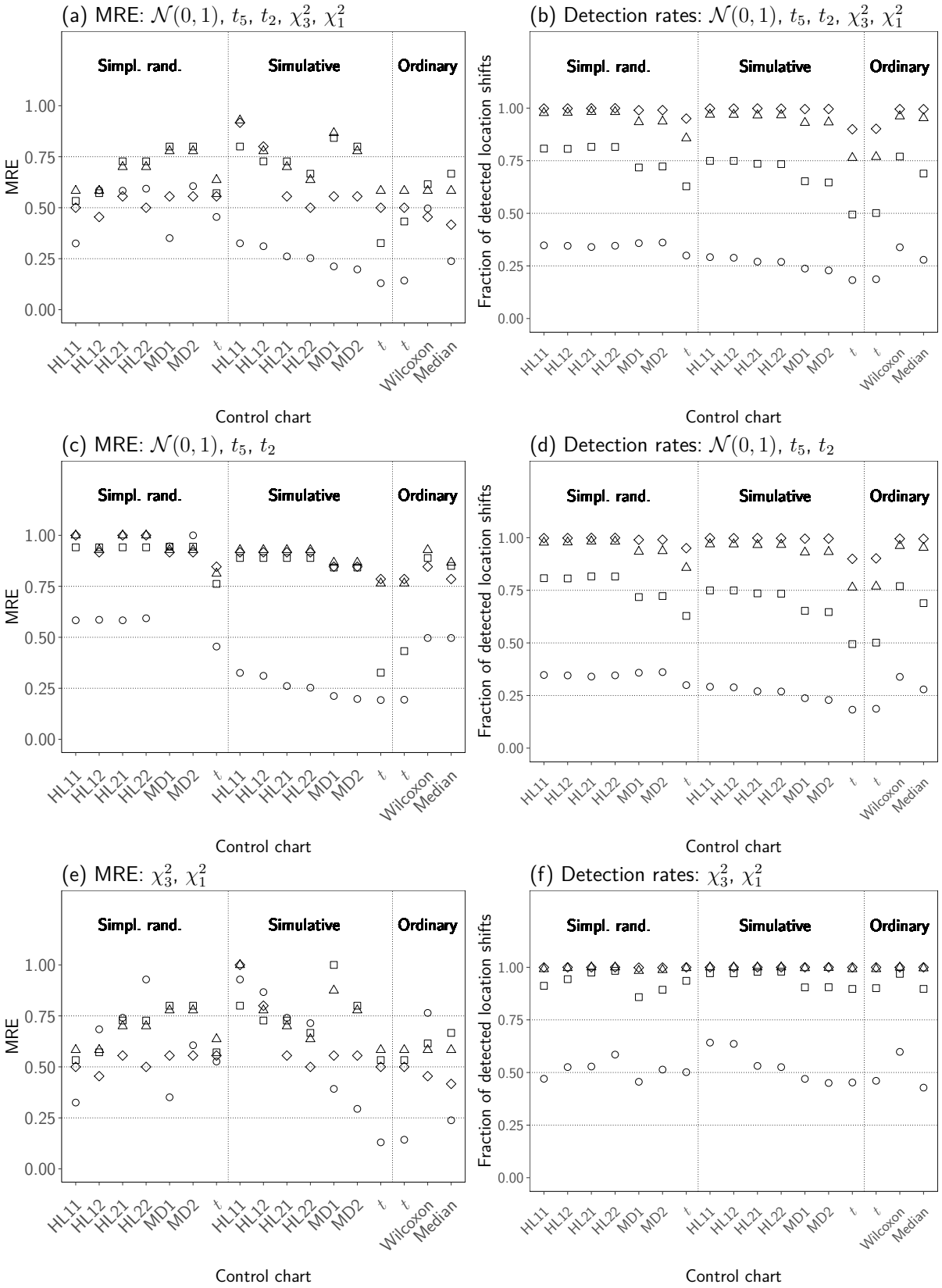

Figure B.4: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates over different groups of distributions for $\mathrm{ARL}_0^* = 370$ under normality and subwindow widths $h = k = 20$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

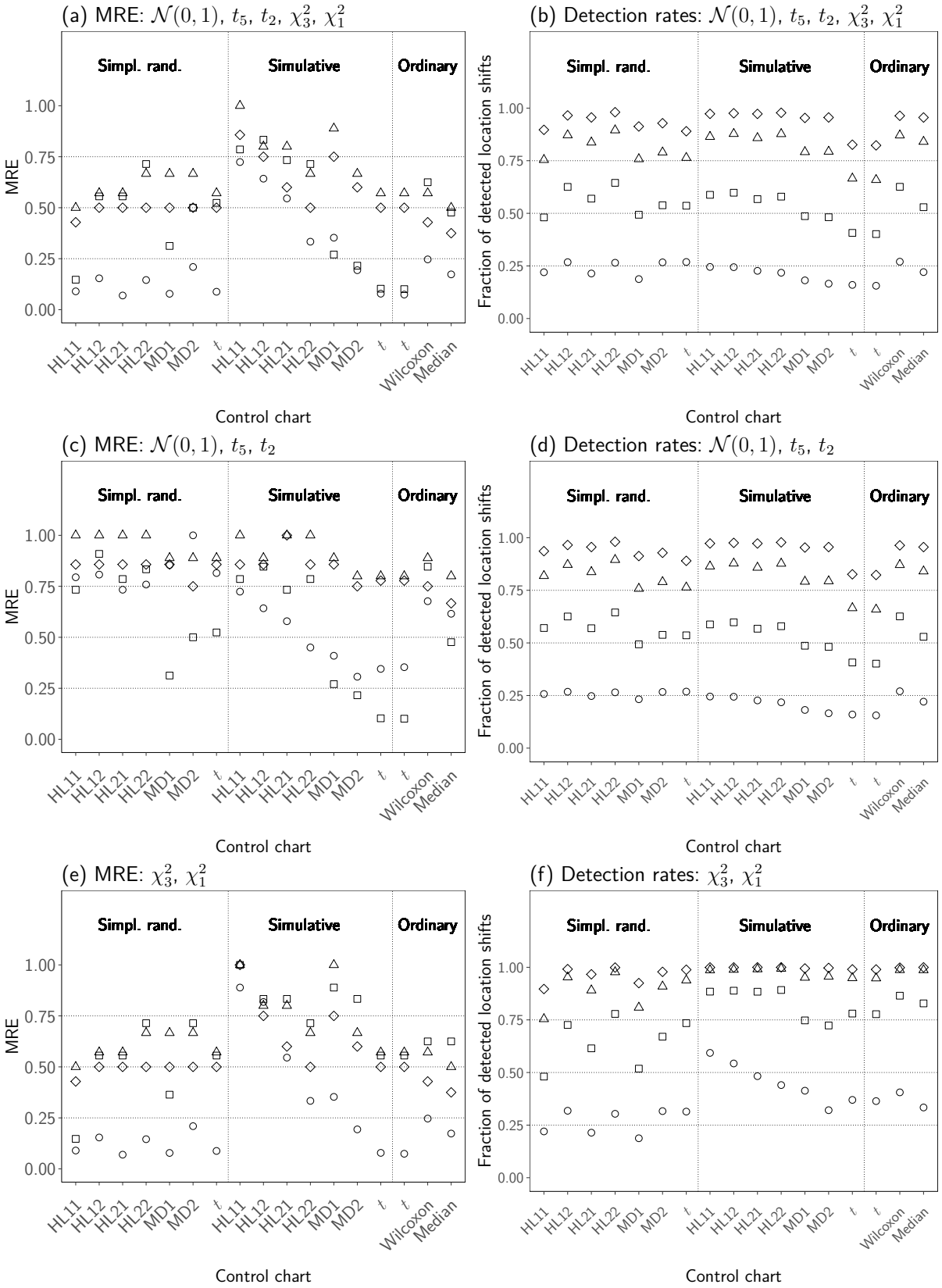

Figure B.5: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality and subwindow widths $h = 20,\ k = 10$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

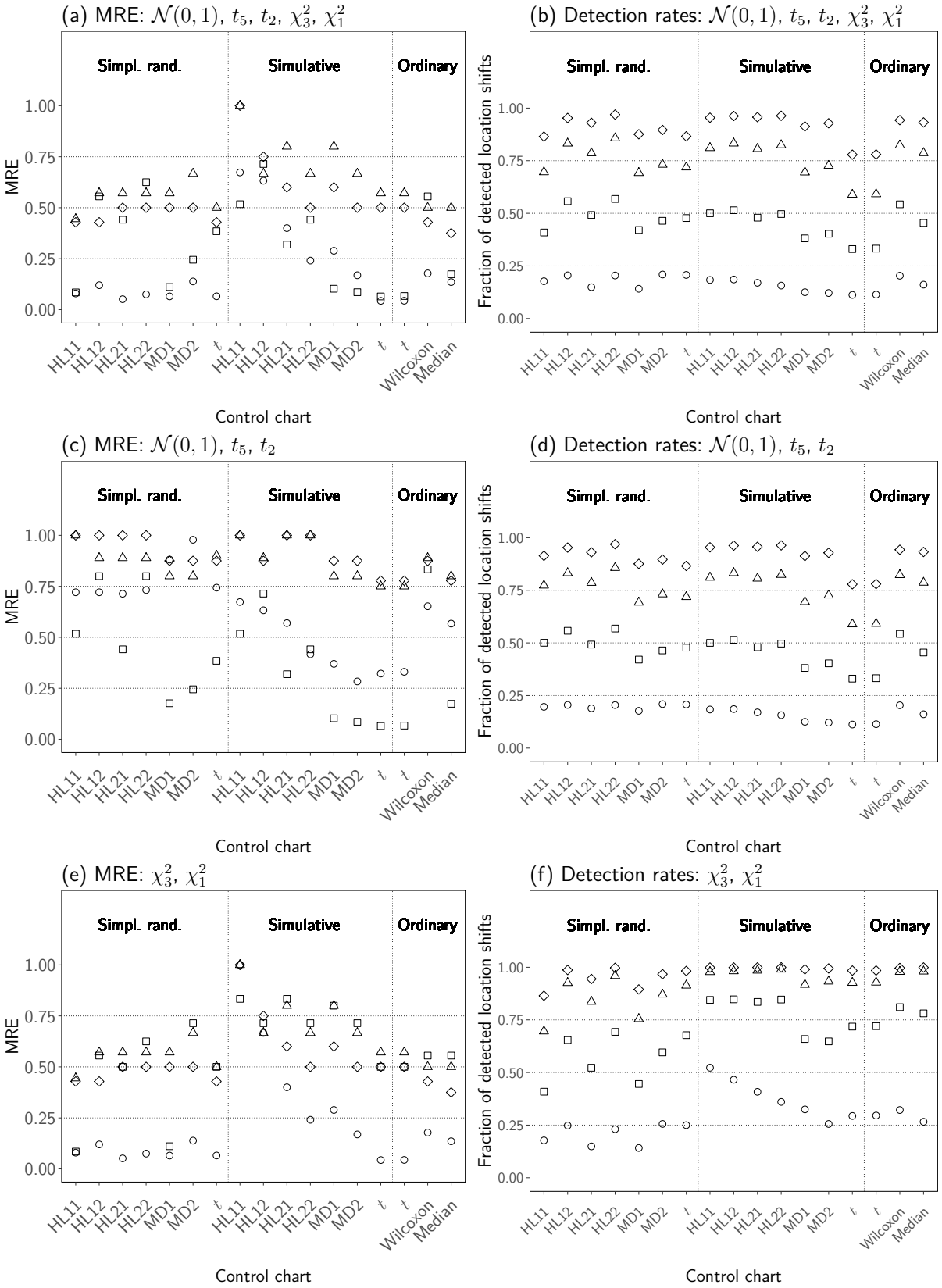

Figure B.6: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates over different groups of distributions for $\mathrm{ARL}_0^* = 370$ under normality and subwindow widths $h = 20,\ k = 10$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

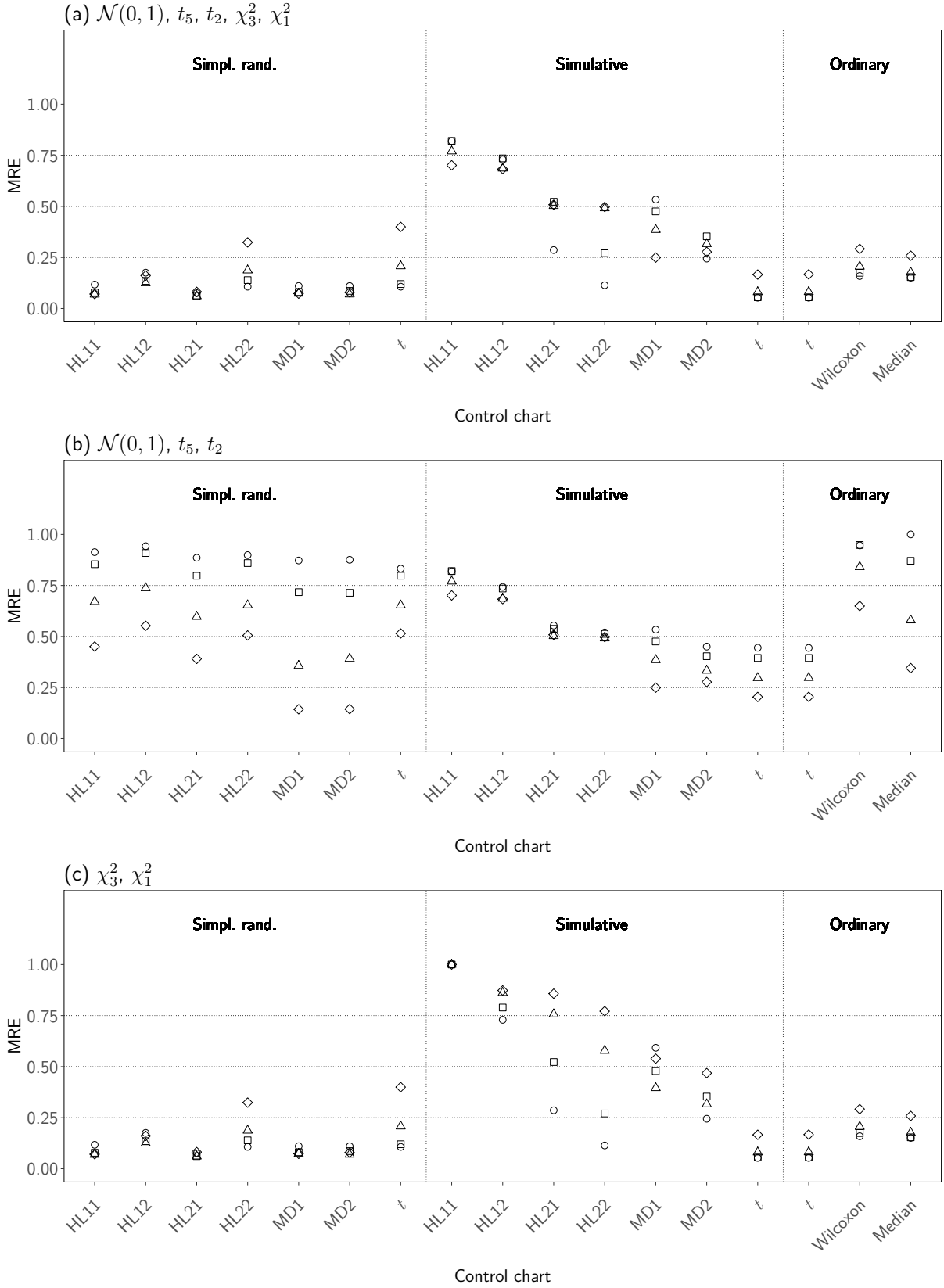

Figure B.7: Minimal relative efficiencies (MRE) with respect to ARL_1 over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality and subwindow widths $h = k = 10$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

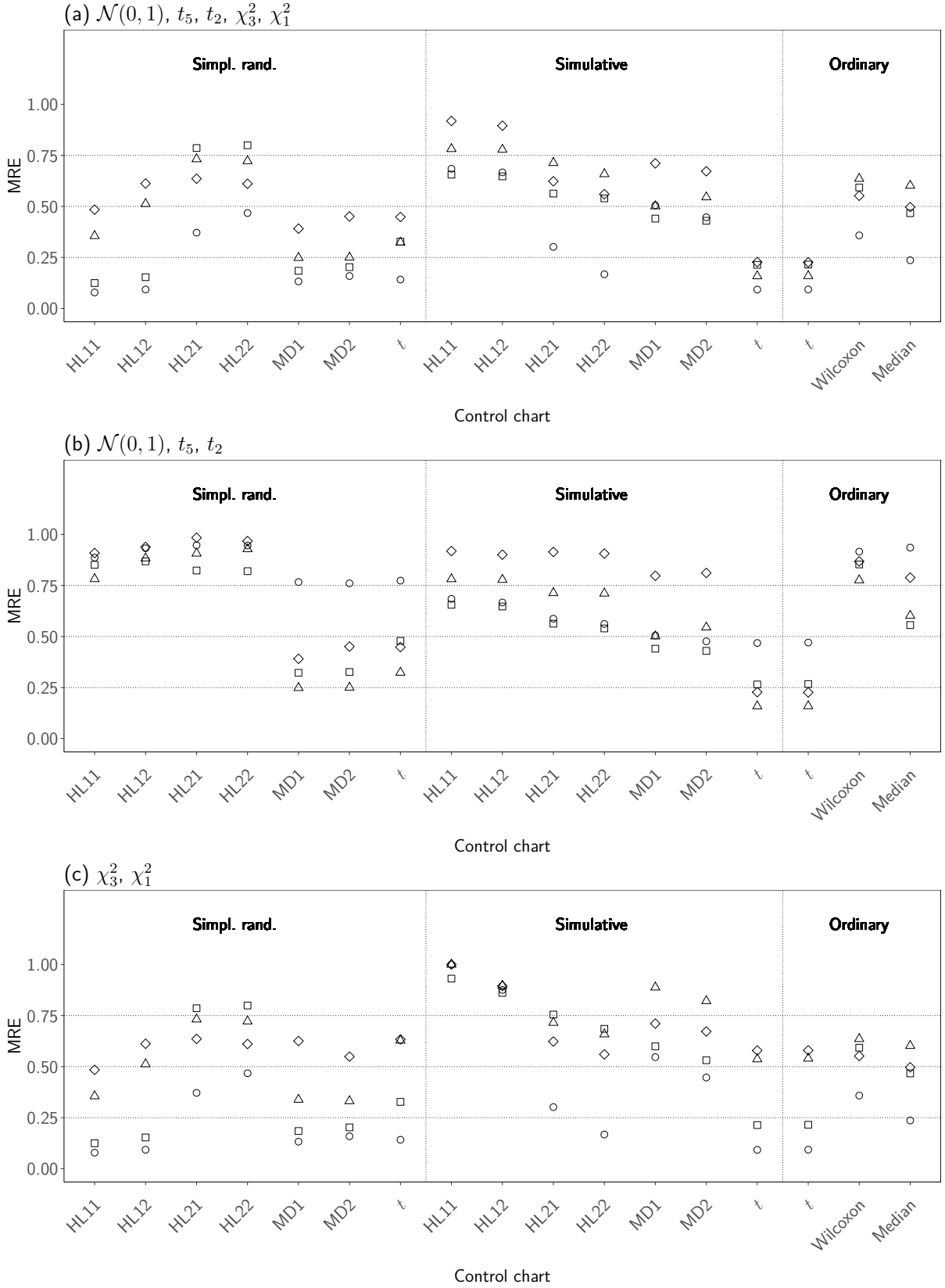

Figure B.8: Minimal relative efficiencies (MRE) with respect to ARL_1 over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality and subwindow widths $h = k = 20$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

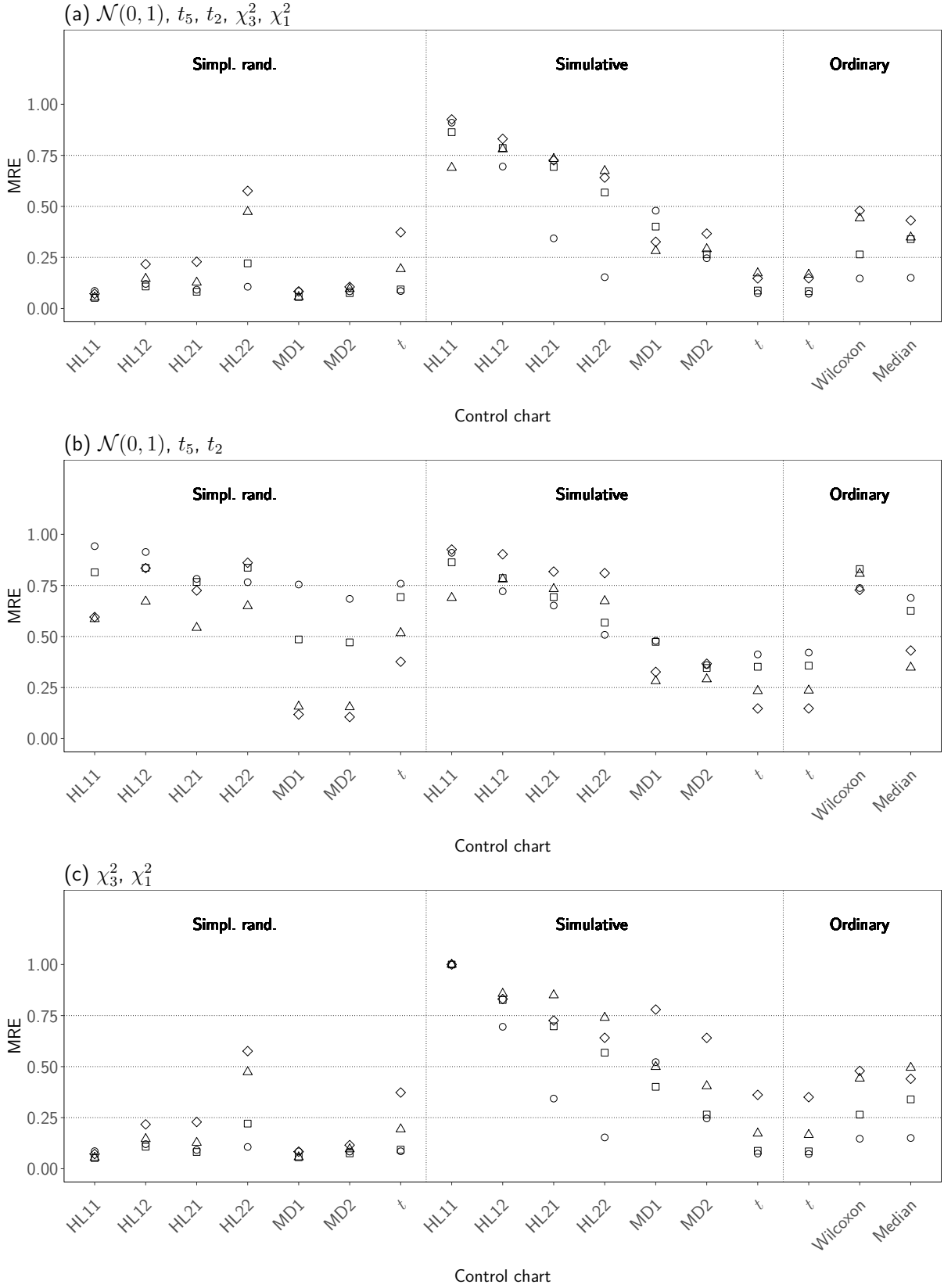

Figure B.9: Minimal relative efficiencies (MRE) with respect to ARL_1 over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality and subwindow widths $h = 20$, $k = 10$. The vertical dotted lines separate between the different groups of control charts.
Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

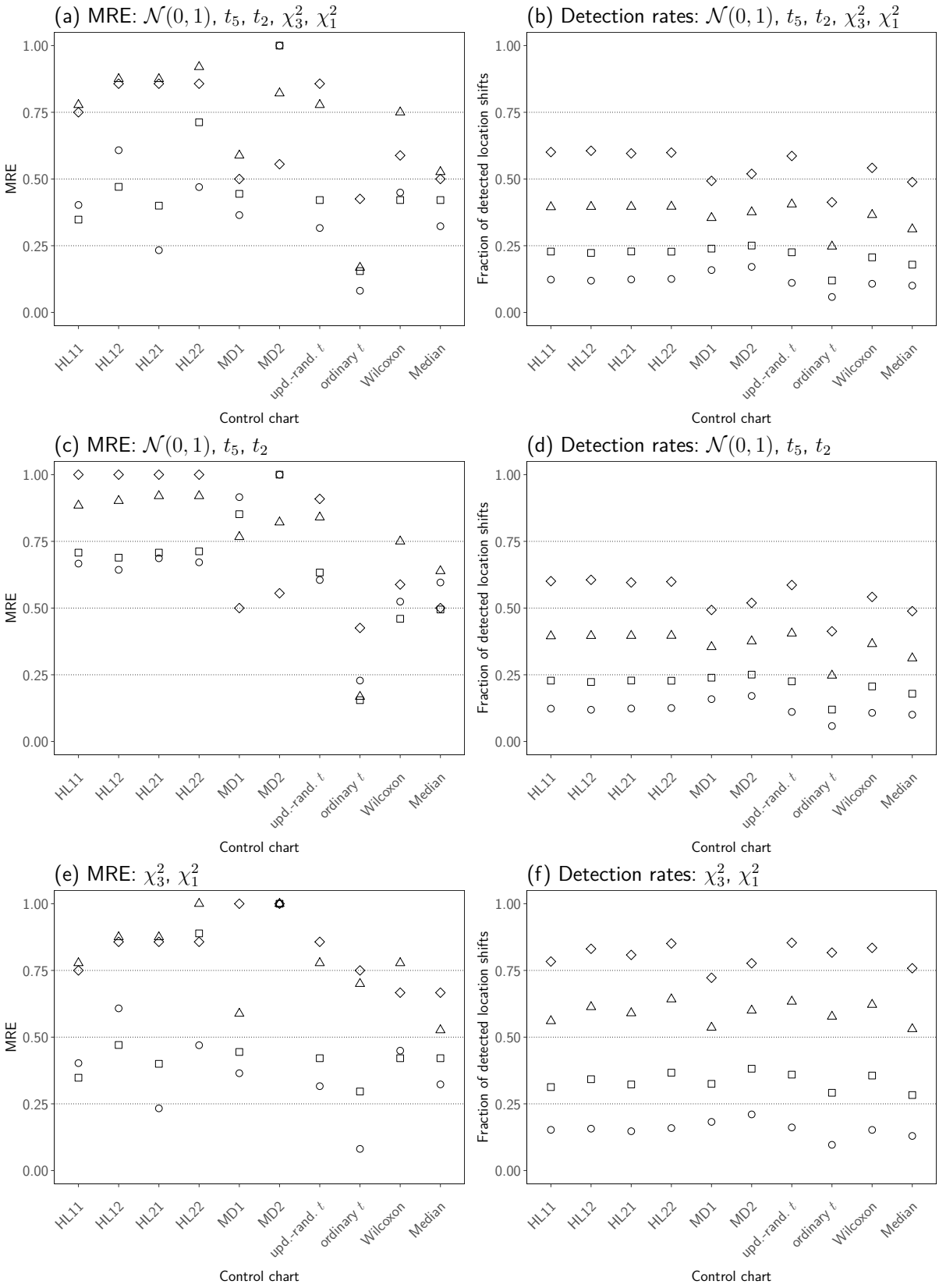

Figure B.10: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates for the shift alternative over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality, subwindow widths $h = k = 10$, and $\ell = 50$. Shift-height factors: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

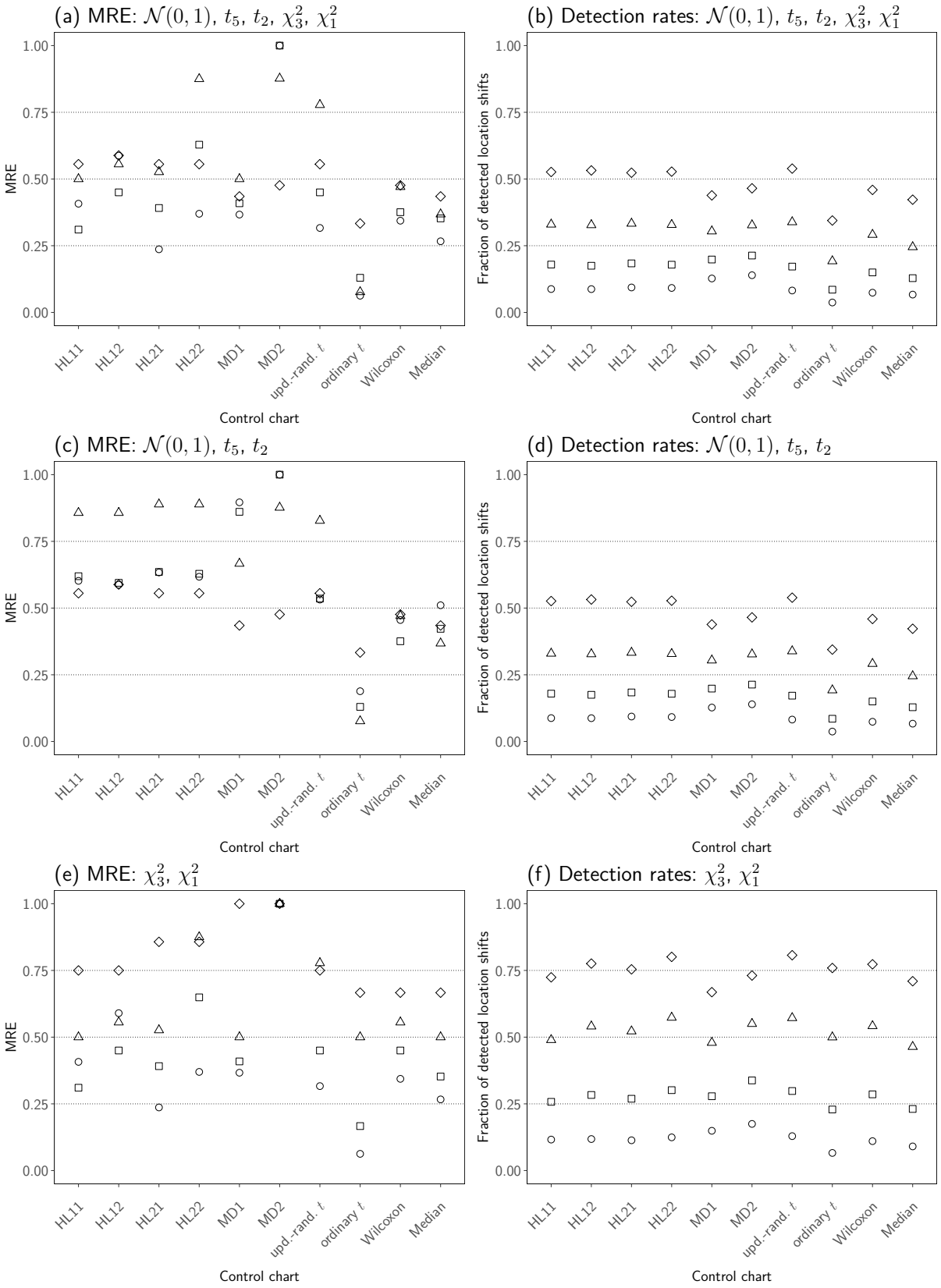

Figure B.11: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates for the shift alternative over different groups of distributions for $\mathrm{ARL}_0^* = 370$ under normality, subwindow widths $h = k = 10$, and $\ell = 50$. Shift-height factors: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

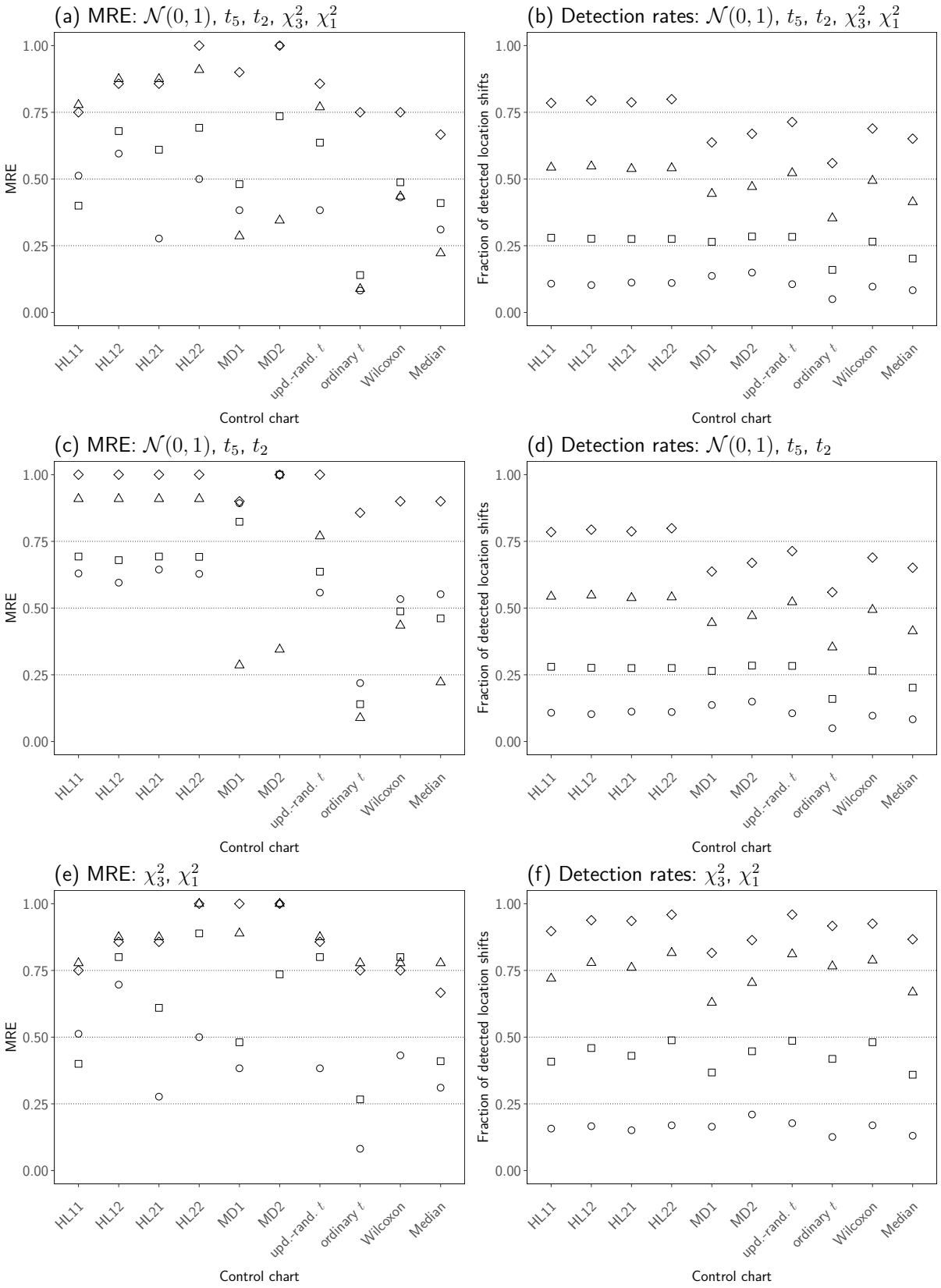

Figure B.12: Minimal relative efficiencies (MRE) with respect to MRL_1 and worst-case detection rates for the shift alternative over different groups of distributions for $\mathrm{ARL}_0^* = 370$ under normality, subwindow widths $h = k = 10$, and $\ell = 125$. Shift-height factor: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

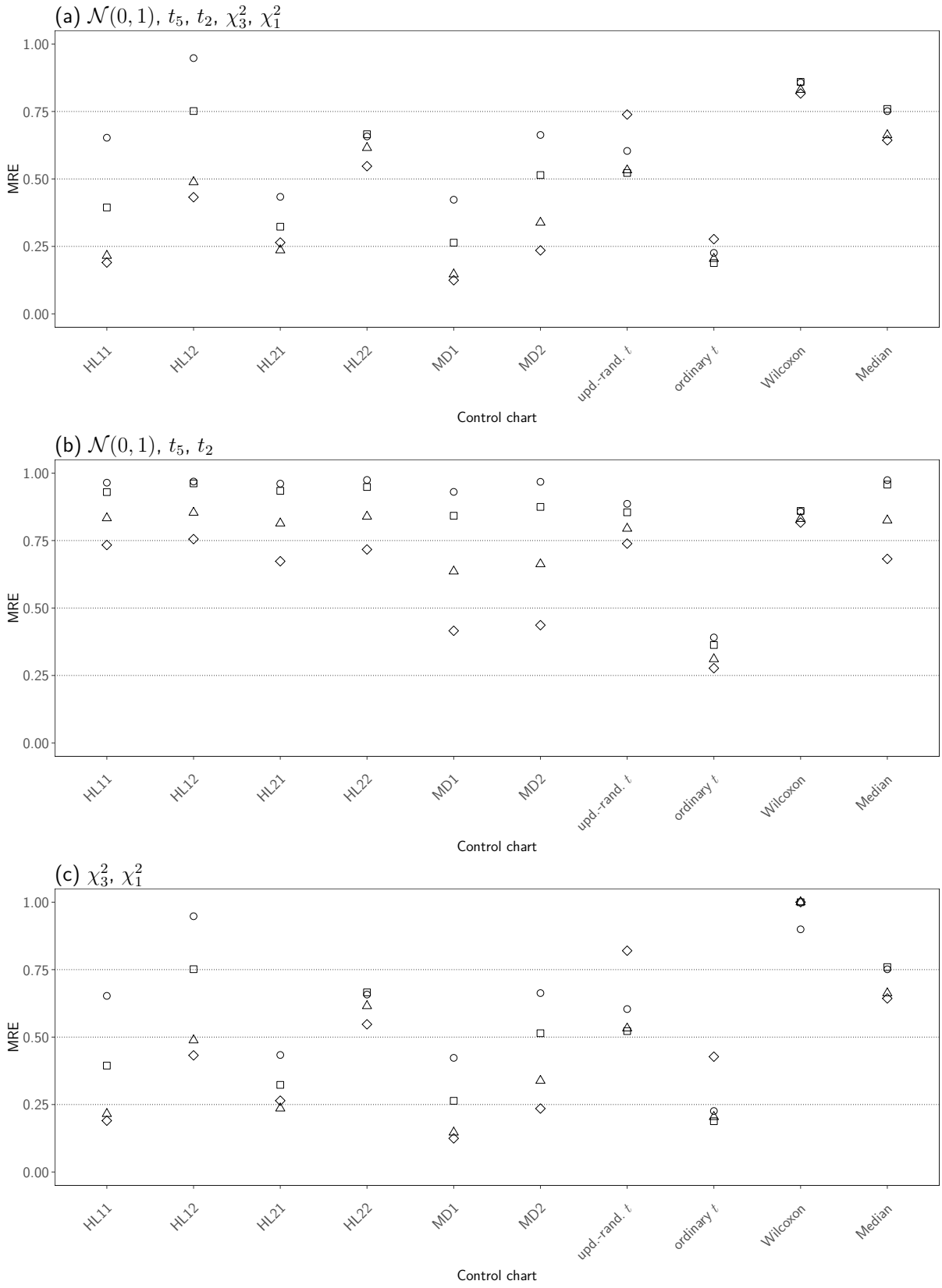

Figure B.13: Minimal relative efficiencies (MRE) with respect to ARL_1 for the shift alternative over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality, subwindow widths $h = k = 10$, and $\ell = 50$.
Shift-height factors: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

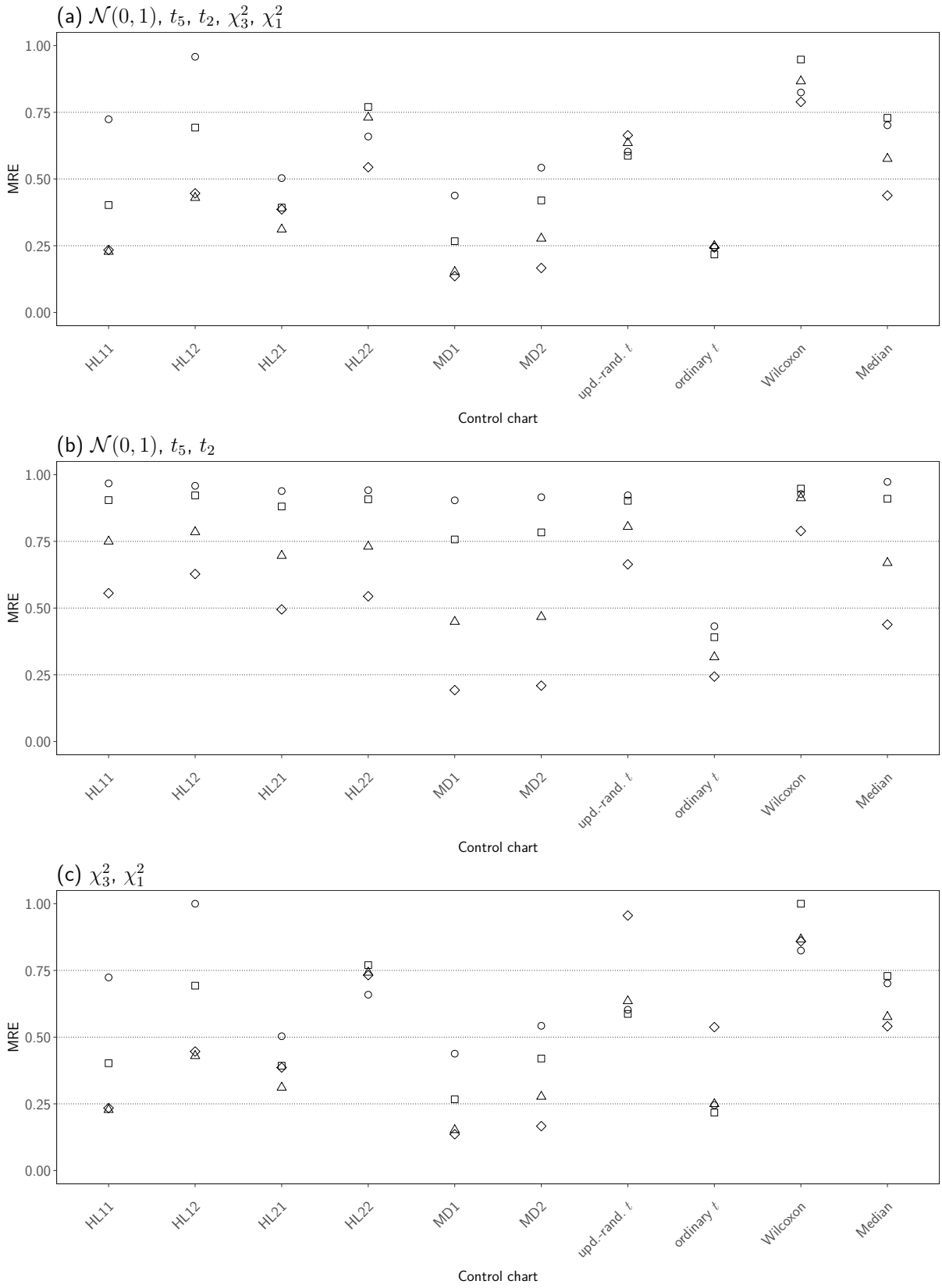

Figure B.14: Minimal relative efficiencies (MRE) with respect to ARL_1 for the shift alternative over different groups of distributions for $\mathrm{ARL}_0^* = 250$ under normality, subwindow widths $h = k = 10$, and $\ell = 125$.
Shift-height factors: 0.5 (○), 1 (□), 1.5 (△), 2 (◇).

C Tables

Table C.1: ARL_0 for selected values of α under normality for the subwindow widths $h = k = 20$. The values in brackets are the standard errors. All values are rounded to one decimal place.

Control chart		Significance level α		
		0.005	0.02	0.05
t-chart	ordinary	562.3 (5.6)	169.1 (1.7)	73.8 (0.8)
	simpl. random.	566.9 (5.8)	168.0 (1.7)	74.1 (0.8)
	simulative	569.2 (5.7)	169.1 (1.8)	70.8 (0.8)
HL11-chart	simpl. random.	573.2 (6.6)	168.9 (1.8)	72.8 (0.8)
	simulative	544.0 (5.5)	165.9 (1.7)	71.5 (0.8)
HL12-chart	simpl. random.	560.9 (6.1)	166.9 (1.7)	73.1 (0.8)
	simulative	542.2 (5.5)	164.2 (1.7)	71.0 (0.8)
HL21-chart	simpl. random.	567.3 (6.3)	165.6 (1.7)	73.0 (0.8)
	simulative	526.0 (5.3)	165.4 (1.7)	71.5 (0.8)
HL22-chart	simpl. random.	563.8 (6.3)	166.9 (1.8)	72.9 (0.8)
	simulative	525.7 (5.3)	165.4 (1.7)	71.1 (0.8)
MD1-chart	simpl. random.	647.4 (15.0)	191.5 (3.7)	80.8 (1.3)
	simulative	549.8 (5.6)	169.3 (1.7)	74.7 (0.8)
MD2-chart	simpl. random.	653.7 (15.1)	198.3 (4.4)	81.2 (1.4)
	simulative	543.0 (5.6)	170.1 (1.7)	74.5 (0.8)
Rank chart	Wilcoxon	548.8 (5.5)	166.1 (1.7)	72.5 (0.8)
	Median	471.0 (4.9)	149.8 (1.6)	55.6 (0.6)

Table C.2: ARL_0 for selected values of α under normality for the subwindow widths $h = 20,\ k = 10$. The values in brackets are the standard errors. All values are rounded to one decimal place.

Control chart		Significance level α		
		0.005	0.02	0.05
t-chart	ordinary	383.8 (3.9)	113.0 (1.2)	49.2 (0.5)
	simpl. random.	375.1 (3.9)	112.4 (1.2)	49.3 (0.5)
	simulative	386.5 (3.9)	110.6 (1.1)	48.1 (0.5)
HL11-chart	simpl. random.	324.4 (3.6)	107.4 (1.2)	47.5 (0.5)
	simulative	377.8 (3.8)	112.0 (1.2)	48.5 (0.5)
HL12-chart	simpl. random.	356.8 (3.9)	110.0 (1.2)	47.7 (0.5)
	simulative	365.6 (3.7)	110.7 (1.2)	48.3 (0.5)
HL21-chart	simpl. random.	322.5 (3.5)	107.4 (1.1)	48.4 (0.5)
	simulative	374.3 (3.8)	112.2 (1.1)	48.5 (0.5)
HL22-chart	simpl. random.	352.0 (3.9)	111.0 (1.2)	48.7 (0.5)
	simulative	367.3 (3.7)	112.3 (1.2)	49.0 (0.5)
MD1-chart	simpl. random.	234.5 (4.8)	93.2 (1.8)	46.9 (0.6)
	simulative	418.0 (4.3)	115.0 (1.2)	50.0 (0.5)
MD2-chart	simpl. random.	300.2 (6.3)	109.3 (2.1)	51.8 (0.8)
	simulative	407.1 (4.2)	117.4 (1.2)	50.3 (0.5)
Rank chart	Wilcoxon	374.6 (3.8)	112.0 (1.2)	48.6 (0.5)
	Median	374.9 (3.8)	89.3 (0.9)	49.5 (0.5)

Table C.3: Estimated regression coefficients for the linearised relationship between the ARL_0 and α in Equation (3.9) under normality for the subwindow widths $h = k = 20$ and $h = 20,\ k = 10$, rounded to two decimal places. The values in brackets are the standard errors and R^2 denotes the coefficient of determination, both rounded to four decimal places.

Window widths	Control chart		$\widehat{\log(\gamma_0)}$	$\hat{\gamma}_1$	R^2
$h = k = 20$	t-chart	ordinary	1.69 (0.0142)	-0.88 (0.0034)	0.9998
		simpl. random.	1.68 (0.0108)	-0.88 (0.0026)	0.9999
		simulative	1.65 (0.0304)	-0.88 (0.0073)	0.9993
	HL11-chart	simpl. random.	1.63 (0.0116)	-0.89 (0.0028)	0.9999
		simulative	1.70 (0.0215)	-0.87 (0.0051)	0.9996
	HL12-chart	simpl. random.	1.66 (0.0098)	-0.88 (0.0023)	0.9999
		simulative	1.69 (0.0214)	-0.87 (0.0051)	0.9996
	HL21-chart	simpl. random.	1.63 (0.0071)	-0.89 (0.0017)	1.0000
		simulative	1.74 (0.0290)	-0.85 (0.0069)	0.9993
	HL22-chart	simpl. random.	1.65 (0.0094)	-0.89 (0.0022)	0.9999
		simulative	1.72 (0.0315)	-0.86 (0.0075)	0.9992
	MD1-chart	simpl. random.	1.79 (0.0440)	-0.88 (0.0105)	0.9984
		simulative	1.71 (0.0159)	-0.87 (0.0038)	0.9998
	MD2-chart	simpl. random.	1.79 (0.0457)	-0.89 (0.0109)	0.9983
		simulative	1.74 (0.0172)	-0.86 (0.0041)	0.9997
	Rank chart	Wilcoxon	1.69 (0.0153)	-0.87 (0.0037)	0.9998
		Median	1.42 (0.1134)	-0.89 (0.0271)	0.9899
$h = 20,\ k = 10$	t-chart	ordinary	1.25 (0.0096)	-0.89 (0.0023)	0.9999
		simpl. random.	1.28 (0.0103)	-0.88 (0.0025)	0.9999
		simulative	1.23 (0.0287)	-0.89 (0.0069)	0.9993
	HL11-chart	simpl. random.	1.44 (0.0323)	-0.82 (0.0077)	0.9990
		simulative	1.21 (0.0200)	-0.89 (0.0048)	0.9997
	HL12-chart	simpl. random.	1.31 (0.0219)	-0.86 (0.0052)	0.9996
		simulative	1.24 (0.0222)	-0.89 (0.0053)	0.9996
	HL21-chart	simpl. random.	1.49 (0.0329)	-0.81 (0.0079)	0.9990
		simulative	1.21 (0.0236)	-0.89 (0.0056)	0.9996
	HL22-chart	simpl. random.	1.36 (0.0246)	-0.85 (0.0059)	0.9995
		simulative	1.23 (0.0266)	-0.89 (0.0064)	0.9994
	MD1-chart	simpl. random.	1.81 (0.0291)	-0.69 (0.0070)	0.9989
		simulative	1.17 (0.0166)	-0.91 (0.0040)	0.9998
	MD2-chart	simpl. random.	1.75 (0.0419)	-0.75 (0.0100)	0.9980
		simulative	1.21 (0.0112)	-0.90 (0.0027)	0.9999
	Rank chart	Wilcoxon	1.24 (0.0104)	-0.88 (0.0025)	0.9999
		Median	1.22 (0.1146)	-0.86 (0.0274)	0.9889

Table C.4: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the t-charts with subwindow widths $h = k = 20$. All values are rounded to two decimal places.

	t-charts											
	ordinary				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.22	2.38	1.35	2.36	1.21	2.35	1.35	2.33	1.00	1.07	1.04	1.09
0.0050	1.15	1.99	1.24	1.91	1.15	1.98	1.24	1.91	0.99	1.05	1.02	1.06
0.0075	1.12	1.83	1.20	1.70	1.12	1.83	1.20	1.70	1.00	1.07	1.02	1.06
0.0100	1.12	1.76	1.18	1.61	1.12	1.75	1.17	1.59	1.01	1.07	1.02	1.06
0.0125	1.11	1.67	1.15	1.50	1.10	1.66	1.14	1.50	1.02	1.08	1.02	1.07
0.0150	1.10	1.61	1.14	1.45	1.10	1.61	1.14	1.45	1.02	1.06	1.02	1.06
0.0200	1.09	1.53	1.12	1.37	1.09	1.52	1.12	1.36	1.03	1.08	1.03	1.06

Table C.5: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the t-charts with subwindow widths $h = 20,\ k = 10$. All values are rounded to two decimal places.

	t-charts											
	ordinary				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.22	2.02	1.18	1.46	1.22	1.99	1.20	1.47	0.99	0.90	1.02	1.02
0.0050	1.18	1.80	1.20	1.47	1.18	1.80	1.19	1.46	1.01	0.93	1.03	1.06
0.0075	1.15	1.67	1.18	1.45	1.15	1.66	1.15	1.39	1.01	0.94	1.03	1.07
0.0100	1.15	1.60	1.17	1.44	1.15	1.60	1.16	1.41	1.02	0.97	1.03	1.09
0.0125	1.14	1.53	1.16	1.41	1.14	1.53	1.16	1.39	1.03	0.97	1.02	1.08
0.0150	1.13	1.48	1.14	1.37	1.13	1.47	1.12	1.35	1.02	0.96	1.01	1.07
0.0200	1.12	1.41	1.13	1.34	1.12	1.40	1.11	1.32	1.03	0.97	1.02	1.06

Table C.6: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the robust control charts with subwindow widths $h = k = 20$. All values are rounded to two decimal places.

	HL11-charts								HL12-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.20	1.49	0.33	0.09	0.99	1.01	1.02	0.86	1.21	1.64	0.49	0.13	1.01	1.01	0.96	0.65
0.0050	1.16	1.38	0.37	0.11	1.00	1.03	1.05	1.04	1.18	1.49	0.50	0.15	1.01	1.03	0.98	0.88
0.0075	1.16	1.35	0.41	0.13	1.02	1.06	1.08	1.16	1.15	1.41	0.52	0.17	1.01	1.05	1.00	1.07
0.0100	1.14	1.31	0.43	0.15	1.01	1.08	1.09	1.23	1.14	1.38	0.54	0.19	0.99	1.04	1.01	1.19
0.0125	1.12	1.28	0.45	0.16	1.01	1.08	1.11	1.29	1.13	1.35	0.55	0.20	1.01	1.05	1.03	1.28
0.0150	1.12	1.28	0.47	0.17	1.02	1.08	1.12	1.31	1.13	1.33	0.56	0.20	1.03	1.06	1.07	1.37
0.0200	1.11	1.26	0.48	0.19	1.02	1.08	1.13	1.37	1.11	1.30	0.57	0.22	1.03	1.05	1.07	1.42
	HL21-charts								HL22-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.27	1.84	0.83	0.76	1.00	0.97	1.03	1.10	1.26	1.95	1.23	2.69	1.00	0.98	1.02	1.00
0.0050	1.23	1.67	0.92	0.92	1.01	0.99	1.02	1.03	1.25	1.77	1.26	2.65	1.02	1.00	1.01	1.00
0.0075	1.21	1.59	0.96	1.02	1.01	0.98	1.00	1.01	1.22	1.66	1.26	2.57	1.03	1.01	1.02	1.01
0.0100	1.20	1.55	0.98	1.11	1.02	1.02	1.01	0.99	1.20	1.59	1.26	2.50	1.02	1.02	1.02	0.99
0.0125	1.17	1.52	1.01	1.16	1.01	1.01	1.01	0.97	1.18	1.57	1.26	2.45	1.02	1.02	1.01	1.01
0.0150	1.16	1.48	1.03	1.21	1.02	1.02	1.01	0.99	1.16	1.52	1.25	2.38	1.02	1.03	1.01	1.01
0.0200	1.14	1.44	1.05	1.28	1.02	1.03	1.01	0.99	1.14	1.47	1.25	2.32	1.02	1.04	1.01	1.00
	MD1-charts								MD2-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.35	2.13	0.51	0.18	0.99	1.01	1.05	1.01	1.38	2.33	0.78	0.46	0.99	1.00	1.02	0.97
0.0050	1.29	1.90	0.57	0.23	1.01	1.00	1.07	1.01	1.31	2.06	0.80	0.46	0.99	1.02	1.04	1.04
0.0075	1.27	1.80	0.60	0.26	1.01	1.04	1.10	1.02	1.27	1.94	0.80	0.46	0.99	1.01	1.06	1.07
0.0100	1.25	1.74	0.62	0.29	0.99	1.02	1.09	0.98	1.26	1.85	0.81	0.46	0.98	1.00	1.06	1.06
0.0125	1.23	1.68	0.64	0.31	0.99	1.00	1.07	0.99	1.25	1.79	0.81	0.46	0.99	0.99	1.06	1.07
0.0150	1.21	1.65	0.65	0.32	0.99	1.01	1.08	0.98	1.23	1.75	0.81	0.46	0.99	1.00	1.04	1.04
0.0200	1.20	1.58	0.66	0.33	0.99	1.02	1.08	0.98	1.22	1.66	0.81	0.45	0.97	1.01	1.02	1.02

Table C.7: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the robust control charts with subwindow widths $h = 20,\ k = 10$. All values are rounded to two decimal places.

	HL11-charts								HL12-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	0.99	0.61	0.24	0.07	0.90	0.60	0.90	0.62	1.16	0.95	0.41	0.11	0.98	0.69	0.91	0.46
0.0050	1.00	0.72	0.31	0.10	0.91	0.68	0.94	0.76	1.12	0.98	0.48	0.15	0.98	0.76	0.94	0.58
0.0075	1.00	0.75	0.36	0.13	0.93	0.73	0.98	0.84	1.10	1.00	0.50	0.18	0.97	0.81	0.97	0.68
0.0100	1.01	0.78	0.38	0.14	0.95	0.77	1.01	0.89	1.09	1.00	0.52	0.19	0.97	0.83	0.97	0.75
0.0125	1.02	0.81	0.41	0.16	0.95	0.79	1.02	0.94	1.11	1.02	0.53	0.21	0.99	0.84	0.98	0.81
0.0150	1.01	0.82	0.43	0.17	0.96	0.80	1.02	0.96	1.09	1.01	0.55	0.22	1.00	0.87	0.96	0.86
0.0200	1.01	0.85	0.47	0.20	0.95	0.82	1.03	1.02	1.09	1.01	0.56	0.24	1.01	0.88	0.98	0.94

	HL21-charts								HL22-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.11	1.07	0.48	0.23	0.99	0.91	0.98	0.86	1.27	1.68	0.80	0.70	1.00	0.96	1.00	0.98
0.0050	1.08	1.09	0.58	0.34	0.99	0.92	0.98	0.87	1.20	1.56	0.91	0.92	1.02	0.98	1.03	0.99
0.0075	1.10	1.11	0.65	0.41	1.01	0.93	0.98	0.85	1.19	1.49	0.95	1.05	1.03	0.98	1.03	0.99
0.0100	1.11	1.14	0.70	0.48	1.02	0.94	1.00	0.87	1.19	1.45	0.99	1.14	1.03	0.97	1.02	0.99
0.0125	1.11	1.13	0.75	0.54	1.01	0.93	1.00	0.88	1.19	1.42	1.02	1.21	1.04	0.98	1.02	0.99
0.0150	1.12	1.13	0.78	0.59	1.02	0.95	1.00	0.90	1.18	1.40	1.03	1.27	1.03	0.97	1.01	0.99
0.0200	1.12	1.14	0.83	0.66	1.01	0.94	0.98	0.92	1.17	1.36	1.07	1.36	1.04	0.98	1.01	0.99

	MD1-charts								MD2-charts							
	simulative				simpl. randomised				simulative				simpl. randomised			
α	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
0.0025	1.26	1.63	0.45	0.10	1.00	0.96	1.00	0.79	1.43	2.43	0.79	0.52	1.07	1.10	1.02	1.00
0.0050	1.24	1.55	0.52	0.13	1.01	0.98	1.04	0.82	1.35	2.16	0.82	0.57	1.05	1.07	1.02	1.04
0.0075	1.23	1.52	0.56	0.17	0.99	0.97	1.02	0.83	1.33	1.99	0.85	0.59	1.06	1.04	1.05	1.05
0.0100	1.22	1.46	0.59	0.19	0.99	0.96	1.02	0.86	1.31	1.89	0.85	0.60	1.05	1.04	1.06	1.08
0.0125	1.19	1.41	0.61	0.21	1.00	0.95	1.01	0.87	1.29	1.82	0.85	0.59	1.03	1.03	1.05	1.07
0.0150	1.17	1.38	0.62	0.23	1.00	0.95	1.02	0.89	1.27	1.77	0.85	0.59	1.03	1.00	1.05	1.07
0.0200	1.15	1.36	0.65	0.26	1.00	0.96	1.01	0.90	1.25	1.68	0.85	0.58	1.02	0.99	1.05	1.10

Table C.8: Minimal MRL_1-values for different noise distributions, separated by shift-height factor Δ and nominal ARL_0^* under normality, over all control charts for the subwindow widths $h = k = 20$.

		Shift-height factor Δ			
ARL_0^*	Distribution	0.5	1	1.5	2
250	$\mathcal{N}(0,1)$	62	16	12	10
	t_5	52	16	12	10
	t_2	36	15	12	11
	χ^2_3	19	12	9	8
	χ^2_1	11	8	6	5
370	$\mathcal{N}(0,1)$	100	17	13	11
	t_5	89	17	13	11
	t_2	70	16	13	11
	χ^2_3	20	13	10	8
	χ^2_1	13	8	7	5

Table C.9: Minimal MRL_1-values for different noise distributions, separated by shift-height factor Δ and nominal ARL_0^* under normality, over all control charts for the subwindow widths $h = 20,\ k = 10$.

		Shift-height factor Δ			
ARL_0^*	Distribution	0.5	1	1.5	2
250	$\mathcal{N}(0,1)$	98.5	10	8	6
	t_5	97	11	8	6
	t_2	88	11	8	7
	χ^2_3	18	8	6	5
	χ^2_1	8	5	4	3
370	$\mathcal{N}(0,1)$	142	12	8	7
	t_5	141	15	9	7
	t_2	135	15	9	7
	χ^2_3	26	8	6	5
	χ^2_1	9	5	4	3

Table C.10: ARL_0 for selected values of α under normality for the subwindow widths $h = k = 20$ and different widths ℓ of the regression window. The values in brackets are the standard errors. All values are rounded to one decimal place.

ℓ	Control chart		Significance level α 0.005	0.02	0.05
50	t-charts	ordinary	167.8 (1.8)	73.2 (0.8)	38.2 (0.5)
		update random.	169.6 (1.8)	73.4 (0.8)	38.2 (0.5)
	HL11-chart		192.7 (2.1)	80.1 (0.9)	41.3 (0.5)
	HL12-chart		191.7 (2.0)	80.2 (0.9)	41.2 (0.5)
	HL21-chart		196.1 (2.1)	81.3 (0.9)	41.7 (0.5)
	HL22-chart		194.5 (2.1)	80.8 (0.9)	41.9 (0.5)
	MD1-chart		243.9 (3.1)	104.0 (1.3)	53.4 (0.7)
	MD2-chart		244.9 (3.1)	104.5 (1.3)	52.6 (0.7)
	Rank charts	Wilcoxon	180.3 (1.9)	77.1 (0.9)	40.6 (0.5)
		Median	239.2 (2.5)	96.0 (1.0)	40.9 (0.5)
125	t-charts	ordinary	340.0 (3.5)	118.4 (1.2)	57.4 (0.6)
		update random.	338.3 (3.4)	118.8 (1.2)	57.5 (0.6)
	HL11-chart		346.0 (3.6)	122.5 (1.3)	58.1 (0.6)
	HL12-chart		344.9 (3.6)	121.7 (1.3)	58.1 (0.6)
	HL21-chart		349.4 (3.7)	121.0 (1.3)	58.1 (0.6)
	HL22-chart		347.0 (3.6)	121.9 (1.3)	58.2 (0.6)
	MD1-chart		354.6 (4.6)	132.1 (1.6)	64.7 (0.8)
	MD2-chart		360.7 (4.8)	134.7 (1.7)	65.2 (0.9)
	Rank charts	Wilcoxon	344.0 (3.5)	120.2 (1.3)	57.7 (0.6)
		Median	342.4 (3.5)	120.6 (1.3)	48.5 (0.5)

Table C.11: ARL_0 for selected values of α under normality for the subwindow widths $h = 20,\ k = 10$ and different widths ℓ of the regression window. The values in brackets are the standard errors. All values are rounded to one decimal place.

			Significance level α		
ℓ	Control chart		0.005	0.02	0.05
50	t-charts	ordinary	150.0 (1.5)	60.1 (0.6)	30.9 (0.3)
		update random.	148.1 (1.5)	60.5 (0.7)	30.7 (0.3)
	HL11-chart		145.1 (1.6)	60.2 (0.6)	31.0 (0.3)
	HL12-chart		156.4 (1.7)	62.7 (0.7)	31.5 (0.4)
	HL21-chart		145.3 (1.6)	61.2 (0.7)	31.6 (0.4)
	HL22-chart		154.8 (1.7)	63.2 (0.7)	32.1 (0.4)
	MD1-chart		132.5 (1.7)	60.1 (0.7)	33.5 (0.4)
	MD2-chart		164.4 (2.3)	69.8 (0.9)	36.6 (0.5)
	Rank charts	Wilcoxon	161.5 (1.6)	62.7 (0.7)	31.6 (0.4)
		Median	198.3 (2.0)	62.1 (0.6)	36.8 (0.4)
125	t-charts	ordinary	292.0 (2.9)	96.3 (1.0)	44.1 (0.5)
		update random.	288.4 (2.9)	96.0 (1.0)	43.7 (0.5)
	HL11-chart		270.2 (3.0)	90.4 (0.9)	43.0 (0.4)
	HL12-chart		294.8 (3.2)	94.1 (1.0)	43.5 (0.5)
	HL21-chart		272.1 (3.1)	91.6 (0.9)	43.1 (0.4)
	HL22-chart		292.0 (3.2)	94.5 (1.0)	43.7 (0.5)
	MD1-chart		201.2 (2.7)	80.0 (0.9)	42.7 (0.5)
	MD2-chart		255.6 (3.7)	93.3 (1.2)	46.5 (0.6)
	Rank charts	Wilcoxon	298.4 (3.0)	96.8 (1.0)	44.2 (0.5)
		Median	308.4 (3.1)	80.5 (0.8)	46.2 (0.5)

Table C.12: Estimated regression coefficients for the linearised relationship between the ARL_0 and α in Equation (3.9) under normality for the subwindow widths $h = k = 20$ and different widths ℓ of the regression window, rounded to two decimal places. The values in brackets are the standard errors and R^2 denotes the coefficient of determination, both rounded to four decimal places.

ℓ	Control chart		$\widehat{\log(\gamma_0)}$	$\hat{\gamma}_1$	R^2
50	t-chart	ordinary	1.80 (0.0365)	-0.63 (0.0087)	0.9979
		update. random.	1.80 (0.0373)	-0.63 (0.0089)	0.9978
	HL11-chart		1.81 (0.0336)	-0.65 (0.0080)	0.9983
	HL12-chart		1.80 (0.0338)	-0.65 (0.0081)	0.9983
	HL21-chart		1.81 (0.0304)	-0.66 (0.0073)	0.9986
	HL22-chart		1.82 (0.0281)	-0.65 (0.0067)	0.9988
	MD1-chart		2.09 (0.0401)	-0.64 (0.0096)	0.9976
	MD2-chart		2.05 (0.0429)	-0.66 (0.0103)	0.9973
	Rank chart	Wilcoxon	1.84 (0.0294)	-0.64 (0.0070)	0.9987
		Median	1.64 (0.1080)	-0.72 (0.0258)	0.9861
125	t-chart	ordinary	1.79 (0.0192)	-0.76 (0.0046)	0.9996
		update. random.	1.78 (0.0191)	-0.76 (0.0046)	0.9996
	HL11-chart		1.78 (0.0237)	-0.77 (0.0057)	0.9994
	HL12-chart		1.78 (0.0225)	-0.77 (0.0054)	0.9995
	HL21-chart		1.75 (0.0204)	-0.78 (0.0049)	0.9996
	HL22-chart		1.76 (0.0183)	-0.77 (0.0044)	0.9996
	MD1-chart		2.02 (0.0203)	-0.73 (0.0049)	0.9995
	MD2-chart		2.02 (0.0209)	-0.73 (0.0050)	0.9995
	Rank chart	Wilcoxon	1.78 (0.0168)	-0.77 (0.0040)	0.9997
		Median	1.54 (0.1046)	-0.81 (0.0250)	0.9896

Table C.13: Estimated regression coefficients for the linearised relationship between the ARL_0 and α in Equation (3.9) under normality for the subwindow widths $h = 20,\ k = 10$ and different widths ℓ of the regression window, rounded to two decimal places. The values in brackets are the standard errors and R^2 denotes the coefficient of determination, both rounded to four decimal places.

ℓ	Control chart		$\widehat{\log(\gamma_0)}$	$\hat{\gamma}_1$	R^2
50	t-chart	ordinary	1.42 (0.0236)	-0.68 (0.0056)	0.9992
		update. random.	1.43 (0.0297)	-0.68 (0.0071)	0.9988
	HL11-chart		1.51 (0.0328)	-0.66 (0.0079)	0.9984
	HL12-chart		1.44 (0.0267)	-0.68 (0.0064)	0.9990
	HL21-chart		1.54 (0.0310)	-0.65 (0.0074)	0.9986
	HL22-chart		1.48 (0.0308)	-0.67 (0.0074)	0.9987
	MD1-chart		1.80 (0.0241)	-0.58 (0.0058)	0.9989
	MD2-chart		1.72 (0.0312)	-0.64 (0.0075)	0.9985
	Rank chart	Wilcoxon	1.41 (0.0266)	-0.69 (0.0064)	0.9991
		Median	1.38 (0.0740)	-0.72 (0.0177)	0.9933
125	t-chart	ordinary	1.35 (0.0145)	-0.82 (0.0035)	0.9998
		update. random.	1.37 (0.0186)	-0.81 (0.0044)	0.9997
	HL11-chart		1.37 (0.0105)	-0.80 (0.0025)	0.9999
	HL12-chart		1.30 (0.0114)	-0.83 (0.0027)	0.9999
	HL21-chart		1.38 (0.0129)	-0.80 (0.0031)	0.9998
	HL22-chart		1.33 (0.0140)	-0.82 (0.0033)	0.9998
	MD1-chart		1.76 (0.0086)	-0.67 (0.0021)	0.9999
	MD2-chart		1.62 (0.0184)	-0.74 (0.0044)	0.9996
	Rank chart	Wilcoxon	1.33 (0.0131)	-0.83 (0.0031)	0.9998
		Median	1.31 (0.0994)	-0.81 (0.0238)	0.9906

Table C.14: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the residual charts with subwindow widths $h = k = 20$ and different widths ℓ of the regression window. All values are rounded to two decimal places.

Control chart	α	ℓ =50				ℓ =125			
		t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
update random. t-chart	0.0025	1.17	1.52	1.19	2.11	1.09	1.17	1.01	1.17
	0.0050	1.17	1.50	1.19	2.00	1.07	1.17	1.03	1.18
	0.0075	1.18	1.53	1.20	1.98	1.08	1.18	1.04	1.19
	0.0100	1.17	1.52	1.18	1.92	1.08	1.19	1.03	1.20
	0.0125	1.17	1.51	1.18	1.87	1.09	1.20	1.04	1.20
	0.0150	1.16	1.50	1.17	1.85	1.07	1.19	1.03	1.20
	0.0200	1.14	1.50	1.16	1.77	1.07	1.21	1.04	1.20
Median chart	0.0025	0.94	0.93	1.09	1.45	0.98	0.97	0.99	0.94
	0.0050	0.96	0.93	1.08	1.36	0.98	0.98	1.00	0.96
	0.0075	0.96	0.95	1.07	1.32	1.00	0.99	1.00	0.97
	0.0100	0.97	0.95	1.07	1.29	1.00	0.99	0.99	0.95
	0.0125	0.96	0.95	1.07	1.30	1.00	1.02	0.99	0.97
	0.0150	0.96	0.95	1.07	1.27	1.00	0.99	1.00	0.97
	0.0200	0.96	0.95	1.06	1.26	1.00	1.00	0.99	0.97
Wilcoxon chart	0.0025	1.09	1.29	0.91	0.93	1.05	1.12	0.88	0.69
	0.0050	1.10	1.25	0.90	0.89	1.04	1.09	0.88	0.71
	0.0075	1.08	1.24	0.90	0.87	1.03	1.10	0.91	0.75
	0.0100	1.08	1.23	0.90	0.85	1.02	1.08	0.90	0.75
	0.0125	1.09	1.21	0.90	0.84	1.03	1.07	0.89	0.76
	0.0150	1.08	1.21	0.91	0.83	1.04	1.07	0.90	0.76
	0.0200	1.07	1.19	0.89	0.81	1.04	1.06	0.91	0.77
HL11-chart	0.0025	1.01	1.12	1.38	2.88	1.00	1.07	1.08	1.16
	0.0050	1.04	1.14	1.27	2.30	1.02	1.07	1.06	1.12
	0.0075	1.03	1.13	1.22	2.11	1.00	1.06	1.02	1.11
	0.0100	1.03	1.12	1.20	1.97	1.00	1.06	1.02	1.11
	0.0125	1.02	1.12	1.18	1.89	1.00	1.07	1.02	1.12
	0.0150	1.02	1.13	1.16	1.83	1.00	1.05	1.01	1.11
	0.0200	1.03	1.13	1.15	1.79	1.01	1.05	1.03	1.13
HL12-chart	0.0025	1.02	1.11	1.09	1.51	1.01	1.07	0.97	0.80
	0.0050	1.03	1.13	1.06	1.43	1.02	1.06	0.96	0.85
	0.0075	1.04	1.12	1.05	1.41	1.01	1.06	0.97	0.90
	0.0100	1.04	1.13	1.05	1.40	1.00	1.06	0.96	0.92
	0.0125	1.03	1.12	1.04	1.38	1.01	1.05	0.95	0.95
	0.0150	1.03	1.11	1.03	1.38	1.00	1.04	0.95	0.97
	0.0200	1.03	1.12	1.03	1.38	1.02	1.05	0.96	1.01
HL21-chart	0.0025	1.01	1.08	1.18	2.10	1.01	1.05	1.07	1.25
	0.0050	1.01	1.09	1.12	1.69	1.01	1.05	1.04	1.15
	0.0075	1.02	1.09	1.08	1.52	1.01	1.05	1.02	1.10
	0.0100	1.03	1.10	1.07	1.43	0.99	1.03	1.00	1.06
	0.0125	1.03	1.10	1.05	1.37	0.99	1.03	0.99	1.04
	0.0150	1.03	1.10	1.05	1.33	1.01	1.04	1.00	1.03
	0.0200	1.00	1.09	1.00	1.25	1.02	1.04	0.99	1.02
HL22-chart	0.0025	1.01	1.07	0.97	1.07	1.00	1.03	0.97	0.86
	0.0050	1.02	1.07	0.96	1.06	1.01	1.05	0.96	0.89
	0.0075	1.02	1.08	0.96	1.05	1.00	1.05	0.96	0.91
	0.0100	1.04	1.10	0.96	1.05	1.00	1.04	0.95	0.92
	0.0125	1.04	1.11	0.97	1.06	1.00	1.04	0.94	0.91
	0.0150	1.03	1.11	0.96	1.06	1.02	1.04	0.96	0.92
	0.0200	1.02	1.10	0.95	1.05	1.02	1.04	0.96	0.92
MD1-chart	0.0025	0.93	0.84	1.35	3.49	0.96	0.94	1.09	1.76
	0.0050	0.93	0.85	1.24	2.42	0.96	0.94	1.06	1.46
	0.0075	0.93	0.85	1.19	1.99	0.96	0.94	1.04	1.29
	0.0100	0.94	0.85	1.15	1.77	0.97	0.96	1.03	1.21
	0.0125	0.94	0.85	1.12	1.59	0.97	0.97	1.02	1.16
	0.0150	0.93	0.86	1.09	1.49	0.97	0.96	1.01	1.12
	0.0200	0.92	0.86	1.07	1.34	0.99	0.96	1.01	1.07
MD2-chart	0.0025	0.92	0.81	1.10	1.42	0.99	0.93	1.01	1.22
	0.0050	0.94	0.84	1.05	1.21	0.98	0.95	0.99	1.09
	0.0075	0.93	0.84	1.00	1.10	0.98	0.95	0.98	1.01
	0.0100	0.93	0.86	0.99	1.04	0.96	0.95	0.96	0.97
	0.0125	0.93	0.86	0.98	1.00	0.97	0.96	0.97	0.97
	0.0150	0.93	0.86	0.98	0.98	0.96	0.95	0.97	0.96
	0.0200	0.93	0.86	0.98	0.95	0.96	0.95	0.96	0.92

Table C.15: Ratios of the ARL_0 under a non-normal distribution compared to the ARL_0 under normality for the residual charts with subwindow widths $h = 20$, $k = 10$ and different widths ℓ of the regression window. All values are rounded to two decimal places.

Control chart	α	ℓ =50				ℓ =125			
		t_5	t_2	χ^2_3	χ^2_1	t_5	t_2	χ^2_3	χ^2_1
update random. t-chart	0.0025	1.10	1.22	1.02	1.28	1.03	1.01	0.93	0.89
	0.0050	1.12	1.29	1.06	1.39	1.04	1.03	0.98	0.97
	0.0075	1.08	1.26	1.07	1.39	1.03	1.03	0.99	1.01
	0.0100	1.08	1.27	1.07	1.41	1.02	1.02	0.98	1.03
	0.0125	1.08	1.26	1.07	1.40	1.02	1.04	1.00	1.05
	0.0150	1.07	1.25	1.08	1.39	1.02	1.03	0.99	1.05
	0.0200	1.07	1.25	1.07	1.40	1.01	1.03	0.99	1.06
Median chart	0.0025	0.97	0.97	1.02	1.05	0.99	1.01	0.99	0.93
	0.0050	0.99	0.98	1.04	1.09	1.00	1.01	1.00	0.93
	0.0075	0.99	0.99	1.03	1.08	0.99	1.00	0.99	0.96
	0.0100	1.00	0.99	1.05	1.10	0.97	0.99	0.98	0.93
	0.0125	1.00	0.98	1.03	1.10	0.99	0.99	0.99	0.96
	0.0150	1.00	0.97	1.06	1.09	1.01	1.00	1.00	0.97
	0.0200	1.01	0.96	1.04	1.07	1.00	1.00	0.99	0.95
Wilcoxon chart	0.0025	1.07	1.18	0.95	0.95	1.03	1.05	0.96	0.81
	0.0050	1.05	1.14	0.94	0.91	1.03	1.05	0.96	0.84
	0.0075	1.04	1.13	0.94	0.90	1.03	1.06	0.96	0.85
	0.0100	1.03	1.13	0.95	0.89	1.03	1.05	0.96	0.85
	0.0125	1.04	1.13	0.95	0.89	1.02	1.04	0.96	0.84
	0.0150	1.04	1.13	0.96	0.88	1.01	1.03	0.96	0.85
	0.0200	1.04	1.12	0.95	0.86	1.01	1.03	0.95	0.85
HL11-chart	0.0025	1.05	1.18	0.98	0.98	1.02	0.92	0.86	0.63
	0.0050	1.03	1.11	0.99	1.01	1.01	0.90	0.91	0.72
	0.0075	1.01	1.07	1.01	1.05	0.99	0.91	0.91	0.78
	0.0100	1.01	1.05	1.01	1.06	0.98	0.89	0.91	0.81
	0.0125	1.02	1.05	1.02	1.08	0.97	0.90	0.93	0.85
	0.0150	1.02	1.04	1.02	1.10	0.97	0.91	0.94	0.88
	0.0200	1.00	1.03	1.03	1.12	1.00	0.94	0.96	0.93
HL12-chart	0.0025	1.02	1.10	0.99	0.85	1.03	0.99	0.93	0.54
	0.0050	1.03	1.08	0.99	0.90	1.01	0.94	0.91	0.62
	0.0075	1.01	1.06	0.99	0.91	1.00	0.93	0.92	0.66
	0.0100	1.03	1.06	0.99	0.94	1.00	0.94	0.91	0.70
	0.0125	1.02	1.06	0.99	0.96	0.98	0.93	0.91	0.74
	0.0150	1.00	1.04	0.98	0.97	1.00	0.95	0.93	0.78
	0.0200	1.00	1.02	0.99	0.99	1.01	0.97	0.94	0.83
HL21-chart	0.0025	1.07	1.28	0.98	1.10	1.09	1.24	0.90	0.81
	0.0050	1.05	1.20	0.99	1.10	1.05	1.14	0.93	0.84
	0.0075	1.03	1.17	1.00	1.09	1.04	1.10	0.94	0.87
	0.0100	1.03	1.14	1.00	1.09	1.02	1.06	0.94	0.88
	0.0125	1.02	1.13	0.99	1.09	1.03	1.04	0.94	0.89
	0.0150	1.01	1.12	0.99	1.11	1.01	1.03	0.95	0.90
	0.0200	1.00	1.08	0.98	1.10	1.02	1.03	0.96	0.93
HL22-chart	0.0025	1.05	1.17	0.99	1.10	1.06	1.15	0.92	0.88
	0.0050	1.03	1.13	1.01	1.13	1.04	1.08	0.95	0.93
	0.0075	1.02	1.11	1.00	1.12	1.04	1.08	0.97	0.94
	0.0100	1.03	1.11	1.01	1.13	1.03	1.05	0.96	0.94
	0.0125	1.03	1.11	1.00	1.12	1.03	1.03	0.97	0.95
	0.0150	1.01	1.09	0.99	1.12	1.02	1.03	0.97	0.94
	0.0200	0.99	1.06	0.98	1.11	1.02	1.04	0.99	0.97
MD1-chart	0.0025	0.97	1.01	1.03	1.11	1.09	1.16	0.99	1.02
	0.0050	0.97	0.99	1.00	1.02	1.04	1.06	0.98	0.94
	0.0075	0.97	0.99	1.00	1.00	1.03	1.04	0.98	0.92
	0.0100	0.97	0.99	0.99	0.99	1.02	1.00	0.98	0.92
	0.0125	0.96	0.97	0.99	0.98	1.01	0.99	0.98	0.92
	0.0150	0.97	0.97	0.99	0.97	1.01	1.00	0.99	0.92
	0.0200	0.97	0.96	0.99	0.97	1.02	1.00	1.00	0.93
MD2-chart	0.0025	0.98	0.96	0.98	0.91	1.05	1.10	0.97	1.19
	0.0050	0.97	0.96	0.94	0.88	1.06	1.06	0.99	1.09
	0.0075	0.96	0.94	0.94	0.88	1.05	1.06	1.00	1.06
	0.0100	0.98	0.95	0.96	0.89	1.02	1.04	1.01	1.05
	0.0125	0.97	0.95	0.96	0.90	1.00	1.02	1.00	1.02
	0.0150	0.97	0.94	0.96	0.90	1.01	1.02	1.01	1.01
	0.0200	0.96	0.94	0.95	0.87	1.00	1.00	1.01	0.99

D Supplementary Definitions

D.1 Asymptotic relative efficiency

The *asymptotic relative efficiency* (ARE) compares the asymptotic variances of two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ for the same parameter $\theta \in \Theta$, where $\Theta \subset \mathbb{R}$ denotes the parameter space. We follow Theorem 2.1 of Lehmann (1997, p. 345).

Let $\left(\hat{\theta}_{1,n} \colon\ n \in \mathbb{N}\right)$ and $\left(\hat{\theta}_{2,n} \colon\ n \in \mathbb{N}\right)$ be two sequences of estimators for θ, where n denotes the size of the sample used for the estimation. Both estimators are assumed to be asymptotically normal with expectation θ and variance ${}^{\sigma_i^2(\theta)}/_{n}$, $i = 1, 2$, so that

$$\sqrt{n} \cdot \left(\hat{\theta}_{i,n} - \theta\right) \overset{\text{asy}}{\sim} \mathcal{N}(0, \sigma_i^2(\theta)),\ i = 1, 2.$$

Then, the ARE of $\hat{\theta}_2$ with respect to $\hat{\theta}_1$ is given by

$$\operatorname{ARE}\left(\hat{\theta}_2, \hat{\theta}_1\right) = \frac{\sigma_1^2(\theta)}{\sigma_2^2(\theta)}.$$

If $\operatorname{ARE}\left(\hat{\theta}_2, \hat{\theta}_1\right) < 1$, then $\hat{\theta}_2$ is said to be less efficient than $\hat{\theta}_1$. For $\operatorname{ARE}\left(\hat{\theta}_2, \hat{\theta}_1\right) = 1$, both estimators are equally efficient.

D.2 Breakdown point

The breakdown point is a concept to quantify the robustness of an estimator $\hat{\theta}$ for a parameter $\theta \in \Theta$ against outliers, where $\Theta \subset \mathbb{R}$ denotes the parameter space. We concentrate on the *replacement finite-sample breakdown point* as introduced by Donoho and Huber (1983).

Let $\boldsymbol{y}^{(n)} \in \mathbb{R}^n$ be an uncontaminated, observed finite sample of size $n \in \mathbb{N}$. Following the notation of Maronna et al. (2006, p. 61), we define a set $\mathcal{Z}_m$, $m \in \mathbb{N}, m < n$, which contains all possible observed samples $\boldsymbol{z}^{(n)}$ of size n, in which $n - m$ elements coincide with those in $\boldsymbol{y}^{(n)}$. Denoting the estimates based on $\boldsymbol{y}^{(n)}$ and $\boldsymbol{z}^{(n)}$ by $\hat{\theta}\left(\boldsymbol{y}^{(n)}\right)$ and $\hat{\theta}\left(\boldsymbol{z}^{(n)}\right)$, the replacement finite-sample breakdown point is defined as

$$\inf_m \left\{ \frac{m}{n} \colon\ \sup_{\boldsymbol{z} \in \mathcal{Z}_m} |\hat{\theta}(\boldsymbol{y}) - \hat{\theta}(\boldsymbol{z})| = \infty \right\}.$$

It represents the smallest fraction ${}^{m}/_{n}$ of elements in $\boldsymbol{y}^{(n)}$ that have to be replaced in order to make $\hat{\theta}\left(\boldsymbol{z}^{(n)}\right)$ realise values arbitrarily far away from $\hat{\theta}\left(\boldsymbol{y}^{(n)}\right)$.

D.3 Pitman asymptotic relative efficiency

We consider the general testing problem $H_0: \ \theta \in \Theta_0$ vs. $H_1: \ \theta \in \Theta_1$ for a parameter $\theta \in \Theta$, where Θ_0 and Θ_1 are the parameter spaces under the null and alternative hypothesis with $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$, and $\Theta \subset \mathbb{R}$ is the full parameter space.

Given two hypothesis tests T_1 and T_2 to the significance level α, the *Pitman asymptotic relative efficiency* (PARE) compares their power to each other in large samples. We follow the definition given in Hodges and Lehmann (1956).

Let $\zeta_n^{(1)}(\theta)$ and $\zeta_n^{(2)}(\theta)$ be the power functions of T_1 and T_2 for the same $n \in \mathbb{N}$ observations. Let $\zeta \in (\alpha, 1)$ be a nominal power and $(\theta_n: \ n \in \mathbb{N})$ a sequence of parameter values in Θ_1, so that

$$\zeta_n^{(1)}(\theta_n) \overset{n\to\infty}{\longrightarrow} \zeta.$$

Moreover, let $(n^* = g(n): \ n \in \mathbb{N})$, where $g: \ \mathbb{N} \to \mathbb{N}$, be a sequence of sample sizes so that

$$\zeta_{n^*}^{(2)}(\theta_n) \overset{n\to\infty}{\longrightarrow} \zeta.$$

The PARE of T_2 with respect to T_1 is defined as

$$\operatorname{PARE}(T_2, T_1) = \lim_{n\to\infty} \frac{n}{n^*}$$

if the limit exists. It is independent of the sequences $(\theta_n: \ n \in \mathbb{N})$ and $(n^* = g(n): \ n \in \mathbb{N})$, and of α and ζ.

The test T_2 is called less efficient than T_1 if $\operatorname{PARE}(T_2, T_1) < 1$. Both tests are equally efficient if $\operatorname{PARE}(T_2, T_1) = 1$.

D.4 Selected equivariance and invariance properties

Equivariance and invariance of an estimator describe how it behaves under transformations of the data. The following definitions for univariate estimators are taken from Rousseeuw and Leroy (1987, p. 158f.), and for regression estimators from Rousseeuw and Leroy (1987, p. 116f.)

Univariate estimators

Let $\boldsymbol{Y}^{(n)} \in \mathbb{R}^n$ be a random sample of size $n \in \mathbb{N}$ and $\hat{\theta}\left(\boldsymbol{Y}^{(n)}\right)$ an estimator for a parameter $\theta \in \Theta$, where $\Theta \subset \mathbb{R}$ denotes the parameter space.

A location estimator is called *shift equivariant* if

$$\hat{\theta}\left(\boldsymbol{Y}^{(n)} + \boldsymbol{a}\right) = \hat{\theta}\left(\boldsymbol{Y}^{(n)}\right) + \boldsymbol{a} \quad \text{for all } \boldsymbol{a} = (a, \ldots, a) \in \mathbb{R}^n.$$

A location estimator is called *scale equivariant* if

$$\hat{\theta}\left(b \cdot \boldsymbol{Y}^{(n)}\right) = b \cdot \hat{\theta}\left(\boldsymbol{Y}^{(n)}\right) \quad \text{for all } b \in \mathbb{R} \setminus \{0\}.$$

A scale estimator is called *scale equivariant* if

$$\hat{\theta}\left(b \cdot \boldsymbol{Y}^{(n)}\right) = |b| \cdot \hat{\theta}\left(\boldsymbol{Y}^{(n)}\right) \quad \text{for all } b \in \mathbb{R} \setminus \{0\}.$$

A scale estimator is called *location invariant* if

$$\hat{\theta}\left(\boldsymbol{Y}^{(n)} + \boldsymbol{a}\right) = \hat{\theta}\left(\boldsymbol{Y}^{(n)}\right) \quad \text{for all } \boldsymbol{a} = (a, \ldots, a) \in \mathbb{R}^n.$$

Regression estimators

The concept of location equivariance can be adapted to the regression context. Let $\boldsymbol{Y}^{(n)} = (Y_1, \ldots, Y_n) \in \mathbb{R}^n$ be a random sample and $\boldsymbol{x}_i = (1, x_i)$, $x_i \in \mathbb{R}$, a vector of fixed regressors for the i-th sample element, $i = 1, \ldots, n$.

A regression estimator

$$\hat{\boldsymbol{\theta}}\left(\{(\boldsymbol{x}_i, Y_i) : \ i = 1, \ldots, n\}\right) = \left(\hat{\theta}_1, \hat{\theta}_2\right)'$$

is called *regression equivariant* if

$$\hat{\theta}\left(\{(\boldsymbol{x}_i, Y_i + \boldsymbol{x}_i \cdot \boldsymbol{d}) : \ i = 1, \ldots, n\}\right) = \left(\hat{\theta}_1, \hat{\theta}_2\right)' + \boldsymbol{d} \quad \text{for all } \boldsymbol{d} = (d_1, d_2)' \in \mathbb{R}^2.$$

List of Figures

2.1 Time series of heart-rate measurements 7
2.2 Time series of greyscale values 9
2.3 Time series of crack-width measurements 10
2.4 Level estimates of location-based filters 15
2.5 Level estimates of regression-based filters 19

3.1 ARL_0-curves under normality, locally constant 44
3.2 Asymptotic 95%-confidence intervals for ARL_0, $h = k = 10$, locally constant 46
3.3 Set-up for out-of-control simulations, locally constant 52
3.4 MRL_1-MRE and worst-case detection rates, $ARL_0^* = 250$, $h = k = 10$, locally constant 55
3.5 MRL_1-RE and detection rates in outlier setting, $ARL_0^* = 250$, locally constant 59
3.6 Control charts for locally constant signal on PAMONO time series 62
3.7 Control charts for locally constant signal on time series of heart-rate measurements 63

4.1 Control chart for locally constant signal on time series with sine-wave signal 68
4.2 One-step-ahead forecast errors from RM and OLS regression without noise 73
4.3 One-step-ahead forecast errors from RM and OLS regression in outlier setting 74
4.4 ARL_0-curves under normality, $h = k = 10$, locally linear 80
4.5 Set-up for out-of-control simulations, locally linear 86
4.6 MRL_1-MRE and worst-case detection rates for shift alternative, $ARL_0^* = 250$, $\ell = 125$, locally linear 88
4.7 MRL_1-RE and detection rates for trend alternative, $ARL_0^* = 250$, $\ell = 125$, locally linear 92
4.8 MRL_1-RE and detection rates in outlier setting, $ARL_0^* = 250$, $\ell = 125$, locally linear 93
4.9 MRL_1-RE and detection rates under local AR(1)-model, $ARL_0^* = 250$, $\ell = 125$ 97
4.10 Residual control charts on PAMONO time series 99
4.11 Residual control charts on time series with sine-wave signal 101
4.12 Residual control charts on time series of heart-rate measurements 102
4.13 Residual control charts on time series of crack-width measurements 104

5.1 Online estimation of a change point 109

B.1 Asymptotic 95%-confidence intervals for ARL_0, $h = k = 20$, $h = 20,\ k = 10$, locally constant 115
B.2 MRL_1-MRE and worst-case detection rates, $\mathrm{ARL}_0^* = 370$, $h = k = 10$, locally constant 116
B.3 MRL_1-MRE and worst-case detection rates, $\mathrm{ARL}_0^* = 250$, $h = k = 20$, locally constant 117
B.4 MRL_1-MRE and worst-case detection rates, $\mathrm{ARL}_0^* = 370$, $h = k = 20$, locally constant 118
B.5 MRL_1-MRE and worst-case detection rates, $\mathrm{ARL}_0^* = 250$, $h = 20,\ k = 10$, locally constant 119
B.6 MRL_1-MRE and worst-case detection rates, $\mathrm{ARL}_0^* = 370$, $h = 20,\ k = 10$, locally constant 120
B.7 ARL_1-MRE, $\mathrm{ARL}_0^* = 250$, $h = k = 10$, locally constant 121
B.8 ARL_1-MRE, $\mathrm{ARL}_0^* = 250$, $h = k = 20$, locally constant 122
B.9 ARL_1-MRE, $\mathrm{ARL}_0^* = 250$, $h = 20,\ k = 10$, locally constant 123
B.10 MRL_1-MRE and worst-case detection rates for shift alternative, $\mathrm{ARL}_0^* = 250$, $\ell = 50$, locally linear 124
B.11 MRL_1-MRE and worst-case detection rates for shift alternative, $\mathrm{ARL}_0^* = 370$, $\ell = 50$, locally linear 125
B.12 MRL_1-MRE and worst-case detection rates for shift alternative, $\mathrm{ARL}_0^* = 370$, $\ell = 125$, locally linear 126
B.13 ARL_1-MRE for shift alternative, $\mathrm{ARL}_0^* = 250$, $\ell = 50$, locally linear 127
B.14 ARL_1-MRE for shift alternative, $\mathrm{ARL}_0^* = 250$, $\ell = 125$, locally linear 128

List of Tables

3.1 Test statistics used in simulation studies 40
3.2 ARL_0 for selected α, $h = k = 10$, locally constant 42
3.3 ARL_0 for selected α, $h = k = 10$, ordinary randomisation, locally constant 43
3.4 Coefficients for relationship between ARL_0 and α, $h = k = 10$, locally constant 45
3.5 ARL_0-ratios for t-charts, $h = k = 10$, locally constant 47
3.6 ARL_0-ratios for rank-based charts, $h = k = 10$, locally constant 48
3.7 ARL_0-ratios for robust charts, $h = k = 10$, locally constant 50
3.8 ARL_0-ratios using ordinary randomisation, $h = k = 10$, locally constant . . 51
3.9 Minimal MRL_1, $h = k = 10$, locally constant 53

4.1 ARL_0-ratios for robust charts using update-randomisation principle, locally constant 77
4.2 Comparison of residual charts based on one- and two-sample tests, locally linear 78
4.3 ARL_0 for selected α, $h = k = 10$, locally linear 81
4.4 Coefficients for relationship between ARL_0 and α, $h = k = 10$, locally linear 82
4.5 ARL_0-ratios for residual charts, $h = k = 10$, locally linear 85
4.6 Minimal MRL_1, locally linear 89
4.7 Minimal MRL_1 of SCARM chart, locally linear 90
4.8 ARL_0 of residual t-chart under local AR(1)-model 94
4.9 ARL_0 for selected α, local AR(1)-model 95
4.10 ARL_0 under different distributions, local AR(1)-model 96

C.1 ARL_0 for selected α, $h = k = 20$, locally constant 129
C.2 ARL_0 for selected α, $h = 20,\ k = 10$, locally constant 130
C.3 Coefficients for relationship between ARL_0 and α, $h = k = 20$, $h = 20,\ k = 10$, locally constant 131
C.4 ARL_0-ratios for t-charts, $h = k = 20$, locally constant 132
C.5 ARL_0-ratios for t-charts, $h = 20,\ k = 10$, locally constant 132
C.6 ARL_0-ratios for robust charts, $h = k = 20$, locally constant 133
C.7 ARL_0-ratios for robust charts, $h = 20,\ k = 10$, locally constant 134
C.8 Minimal MRL_1, $h = k = 20$, locally constant 135
C.9 Minimal MRL_1, $h = 20,\ k = 10$, locally constant 135
C.10 ARL_0 for selected α, $h = k = 20$, locally linear 136

C.11 ARL_0 for selected α, $h = 20,\ k = 10$, locally linear 137
C.12 Coefficients for relationship between ARL_0 and α, $h = k = 20$, locally linear 138
C.13 Coefficients for relationship between ARL_0 and α, $h = 20,\ k = 10$, locally linear 139
C.14 ARL_0-ratios for residual charts, $h = k = 20$, locally linear 140
C.15 ARL_0-ratios for residual charts, $h = 20,\ k = 10$, locally linear 141

References

Abbas, S. (2013). Detektion von Nanoobjekten in Graustufenbildern und Bildsequenzen mittels robuster Zeitreihenmethoden zur Strukturbrucherkennung. Master's Thesis. Fakultät Statistik, Technische Universität Dortmund.

Abbas, S. and Fried, R. (2017). Control charts for the mean based on robust two-sample tests. *Journal of Statistical Computation and Simulation* 87(1), pp. 138–155.

Abbas, S. and Fried, R. (2019). Control charts for the mean of a locally linear time series. *Manuscript submitted for publication.*

Abbas, S., Fried, R., and Gather, U. (2016). Detection of local intensity changes in grayscale images with robust methods for time-series analysis. In: *Solving Large Scale Learning Tasks: Challenges and Algorithms: Essays Dedicated to Katharina Morik on the Occasion of Her 60th Birthday.* Ed. by S. Michaelis, N. Piatkowski, and M. Stolpe. Cham: Springer, pp. 251–271.

Abbas, S., Fried, R., Heinrich, J., Horn, M., Jakubzik, M., Kohlenbach, J., Maurer, R., Michels, A., and Müller, C. H. (2018). *Detection of anomalous sequences in crack data of a bridge monitoring.* SFB Discussion Paper 29. Dortmund.

Alloway, J. A. and Raghavachari, M. (1991). Control chart based on the Hodges-Lehmann estimator. *Journal of Quality Technology* 23(4), pp. 336–347.

Alwan, L. C. and Roberts, H. V. (1988). Time-series modeling for statistical process control. *Journal of Business & Economic Statistics* 6(1), pp. 87–95.

Amin, R. W., Reynolds, M., and Bakir, S. T. (1995). Nonparametric quality control charts based on the sign statistic. *Communications in Statistics - Theory and Methods* 24(6), pp. 1597–1623.

Aroian, L. A. and Levene, H. (1950). The effectiveness of quality control charts. *Journal of the American Statistical Association* 45(252), pp. 520–529.

Astola, J., Heinonen, P., and Neuvo, Y. (1989). Linear median hybrid filters. *IEEE Transactions on Circuits and Systems* 36(11), pp. 1430–1438.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis* 34(1), pp. 1–16.

Auguie, B. (2016). *gridExtra: Miscellaneous functions for "Grid" graphics*. R package version 2.2.1.

Bakir, S. T. (2006). Distribution-free quality control charts based on signed-rank-like statistics. *Communications in Statistics - Theory and Methods* 35(4), pp. 743–757.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, New Jersey: Prentice-Hall.

Bengtsson, H. (2018). *future.apply: Apply function to elements in parallel using futures*. R package version 0.2.0.

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

Bernholt, T., Fried, R., Gather, U., and Wegener, I. (2006). Modified repeated median filters. *Statistics and Computing* 16(2), pp. 177–192.

Borowski, M. and Fried, R. (2014). Online signal extraction by robust regression in moving windows with data-adaptive width selection. *Statistics and Computing* 24(4), pp. 597–613.

Borowski, M., Görges, M., Fried, R., Such, O., Wrede, C. E., and Imhoff, M. (2011). VDE Position Paper: Medical device alarms. *Biomedizinische Technik* 56, pp. 73–83.

Chakraborti, S., Human, S. W., and Graham, M. A. (2011). Nonparametric (distribution-free) quality control charts. In: *Handbook of Methods and Applications of Statistics: Engineering, Quality Control, and Physical Sciences*. Ed. by N. Balakrishnan. Hoboken, New Jersey: Wiley, pp. 298–329.

Chakraborti, S., Van der Laan, P., and Bakir, S. T. (2001). Nonparametric control charts: An overview and some results. *Journal of Quality Technology* 33(3), pp. 304–315.

Chakraborti, S. (2000). Run length, average run length and false alarm rate of Shewhart X-bar chart: Exact derivations by conditioning. *Communications in Statistics - Simulation and Computation* 29(1), pp. 61–81.

Chakraborti, S. (2007). Run length distribution and percentiles: The Shewhart X-bar chart with unknown parameters. *Quality Engineering* 19(2), pp. 119–127.

Chakraborti, S. and Eryilmaz, S. (2007). A nonparametric Shewhart-type signed-rank control chart based on runs. *Communications in Statistics - Theory and Methods* 36(2), pp. 335–356.

Chakraborti, S. and Graham, M. A. (2008). Control charts, Nonparametric. In: *Encyclopedia of Statistics in Quality and Reliability*. Ed. by F. Ruggeri, R. S. Kennett, and F. Faltin. Chichester: Wiley.

Chakraborti, S. and van de Wiel, M. A. (2008). A nonparametric control chart based on the Mann-Whitney statistic. In: *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*. Ed. by N. Balakrishnan, E. A. Peña, and M. J. Silvapulle. Beachwood: Institute of Mathematical Statistics, pp. 156–172.

Croux, C., Gelper, S., and Mahieu, K. (2011). Robust control charts for time series data. *Expert Systems with Applications* 38(11), pp. 13810–13815.

Dahl, D. B. (2016). *xtable: Export tables to LaTeX or HTML*. R package version 1.8-2.

Daniel, W. W. (1978). *Applied Nonparametric Statistics*. Boston: Houghton Mifflin.

Davies, P. L. and Gather, U. (2005). The breakdown point - Examples and counterexamples. *RevStat - Statistical Journal* 5(1), pp. 1–17.

Davies, P. L., Fried, R., and Gather, U. (2004). Robust signal extraction for on-line monitoring data. *Journal of Statistical Planning and Inference* 122(1-2), pp. 65–78.

Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In: *A Festschrift for Erich Lehmann*. Ed. by P. J. Bickel, K. Doksum, and J. L. Hodges. Belmont: Wadsworth, pp. 157–184.

Dowle, M. and Srinivasan, A. (2017). *data.table: Extension of 'data.frame'*. R package version 1.10.4.

Dürre, A., Fried, R., and Liboschik, T. (2015). Robust estimation of (partial) autocorrelation. *WIREs Computational Statistics* 7(3), pp. 205–222.

Dürre, A., Fried, R., Liboschik, T., and Rathjens, J. (2017). *robts: Robust time series analysis*. R package version 0.3.0/r221.

Eichinger, B. and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* 24(1), pp. 526–564.

Ernst, M. D. (2004). Permutation methods: A basis for exact inference. *Statistical Science* 19(4), pp. 676–685.

Fried, R. (2007). On the robust detection of edges in time series filtering. *Computational Statistics & Data Analysis* 52(2), pp. 1063–1074.

Fried, R. (2012). On the online estimation of piecewise constant volatilities. *Computational Statistics & Data Analysis* 56(11), pp. 3080–3090.

Fried, R., Abbas, S., Borowski, M., and Imhoff, M. (2017). Online analysis of medical time series. *Annual Review of Statistics and Its Application* 4(1), pp. 169–188.

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

Fried, R. and Dehling, H. (2011). Robust nonparametric tests for the two sample location problem. *Statistical Methods & Applications* 20(4), pp. 409–422.

Fried, R., Einbeck, J., and Gather, U. (2007). Weighted repeated median smoothing and filtering. *Journal of the American Statistical Association* 102(480), pp. 1300–1308.

Fried, R. and Gather, U. (2002). Fast and robust filtering of time series with trends. In: *Compstat*. Ed. by W. Härdle and B. Rönz. Heidelberg: Physica, pp. 367–372.

Fried, R. and Gather, U. (2007). On rank tests for shift detection in time series. *Computational Statistics & Data Analysis* 52, pp. 221–233.

Fried, R., Schettlinger, K., and Borowski, M. (2014). *robfilter: Robust time series filters*. R package version 4.1.

Frisén, M. (2003). Statistical surveillance. Optimality and methods. *International Statistical Review* 71(2), pp. 403–434.

Gan, F. F. (1993). An optimal design of EWMA control charts based on median run length. *Journal of Statistical Computation and Simulation* 45(3-4), pp. 169–184.

Garthoff, R., Okhrin, I., and Schmid, W. (2014). Statistical surveillance of the mean vector and the covariance matrix of nonlinear time series. *AStA Advances in Statistical Analysis* 98(3), pp. 225–255.

Garthoff, R., Okhrin, I., and Schmid, W. (2015). Control charts for multivariate nonlinear time series. *RevStat - Statistical Journal* 13(2), pp. 131–144.

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

Gather, U., Schettlinger, K., and Fried, R. (2006). Online signal extraction by robust linear regression. *Computational Statistics* 21(1), pp. 33–51.

Hackl, P. and Ledolter, J. (1992). A new nonparametric quality control technique. *Communications in Statistics - Simulation and Computation* 21(2), pp. 423–443.

Hájek, J. (1969). *Nonparametric Statistics*. San Francisco: Holden-Day.

Hampel, F. R. (1975). Beyond location parameters: Robust concepts and methods. *Bulletin of the International Statistical Institute* 46(1), pp. 375–382.

Hawkins, D. M. and Deng, Q. (2010). A nonparametric change-point control chart. *Technometrics* 42(2), pp. 165–173.

Heinonen, P. and Neuvo, Y. (1987). FIR-median hybrid filters. *IEEE Transaction on Acoustics, Speech, and Signal Processing* 35(6), pp. 832–838.

Heinonen, P. and Neuvo, Y. (1988). FIR-median hybrid filters with predictive FIR substructures. *IEEE Transaction on Acoustics, Speech, and Signal Processing* 36(6), pp. 892–899.

Hodges, J. L. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *The Annals of Mathematical Statistics* 27(2), pp. 324–335.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics* 34(2), pp. 598–611.

Hollander, M., Wolfe, D. A., and Chicken, E. (2013). *Nonparametric Statistical Methods*. 3rd ed. Hoboken, New Jersey: Wiley.

Høyland, A. (1965). Robustness of the Hodges-Lehmann estimates for shift. *The Annals of Mathematical Statistics* 36(1), pp. 174–197.

Human, S. W., Kritzinger, P., and Chakraborti, S. (2011). Robustness of the EWMA control chart for individual observations. *Journal of Applied Statistics* 38(10), pp. 2071–2087.

Imhoff, M., Bauer, M., Gather, U., and Fried, R. (2002). Pattern detection in intensive care monitoring time series with autoregressive models: Influence of the model order. *Biometrical Journal* 44(6), pp. 746–761.

Imhoff, M., Kuhls, S., Gather, U., and Fried, R. (2009). Smart alarms from medical devices in the OR and ICU. *Best Practice & Research: Clinical Anaesthesiology* 23(1), pp. 39–50.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., and Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology* 38(4), pp. 349–364.

Keefe, M. J., Woodall, W. H., and Jones-Farmer, L. A. (2015). The conditional in-control performance of self-starting control charts. *Quality Engineering* 27(4), pp. 488–499.

Knoth, S. and Schmid, W. (2004). Control charts for time series: A review. In: *Frontiers in Statistical Quality Control*. Vol. 7. Heidelberg: Physica, pp. 210–236.

Lang, M., Bischl, B., and Surmann, D. (2017). batchtools: Tools for R to work on batch systems. *The Journal of Open Source Software* 2(10).

Lee, Y. and Kassam, S. (1985). Generalized median filtering and related nonlinear filtering techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33(3), pp. 672–683.

Lehmann, E. L. (1997). *Theory of Point Estimation*. New York: Springer.

Lehmann, E. L. (2009). Parametric versus nonparametric: Two alternative methodologies. *Journal of Nonparametric Statistics* 21(4), pp. 397–405.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypothesis*. 3rd ed. New York: Springer.

Lucas, J. M. and Crosier, R. B. (1982). Robust CUSUM: A robustness study for CUSUM quality control schemes. *Communications in Statistics - Theory and Methods* 11(23), pp. 2669–2687.

Mäkivirta, A., Koski, E., Kari, A., and Sukuvaara, T. (1991). The median filter as a preprocessor for a patient monitor limit alarm system in intensive care. *Computer Methods and Programs in Biomedicine* 34(2 - 3), pp. 139–144.

Maronna, R. A., Martin, D. R., and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Chichester: Wiley.

Montgomery, D. C. (2009). *Statistical Quality Control: A Modern Introduction*. 6th ed. Hoboken, New Jersey: Wiley.

Montgomery, D. C. and Mastrangelo, C. M. (1991). Some statistical process control methods for autocorrelated data. *Journal of Quality Technology* 23(3), pp. 179–204.

Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *The Annals of Mathematical Statistics* 25(3), pp. 514–522.

Morell, O. (2012). On nonparametric methods for robust jump-preserving smoothing and trend detection. Dissertation. Fakultät Statistik, Technische Universität Dortmund.

Morgenthaler, S. and Tukey, J. W. (1991). *Configural Polysampling: A Route to Statistical Robustness*. New York: Wiley.

Nazir, H. Z., Riaz, M., Does, R. J. M. M., and Abbas, N. (2013). Robust CUSUM control charting. *Quality Engineering* 25(3), pp. 211–224.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* 41(1/2), pp. 100–115.

Pappanastos, E. A. and Adams, B. M. (1996). Alternative designs of the Hodges-Lehmann control chart. *Journal of Quality Technology* 28(2), pp. 213–223.

Phipson, B. and Smyth, G. K. (2010). Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology* 9(1), Article 39.

Qiu, P. and Li, Z. (2011). Distribution-free monitoring of univariate processes. *Statistics and Probability Letters* 81(12), pp. 1833–1840.

Qiu, P. and Yandell, B. (1998). A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics* 40(2), pp. 141–152.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics* 3(1), pp. 239–250.

Rocke, D. M. (1989). Robust control charts. *Technometrics* 31(2), pp. 173–184.

Ross, G. J. (2015). Parametric and nonparametric sequential change detection in R: The cpm package. *Journal of Statistical Software* 66(3), pp. 1–20.

Ross, G. J., Tasoulis, D. K., and Adams, N. M. (2011). Nonparametric monitoring of data streams for changes in location and scale. *Technometrics* 53(4), pp. 379–389.

Rousseeuw, P. J. and Hubert, M. (1999). Regression depth. *Journal of the American Statistical Association* 94(446), pp. 388–402.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association* 79(388), pp. 871–880.

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

Saghaei, A. and Noorossana, R. (2011). Introduction to profile monitoring. In: *Statistical Analysis of Profile Monitoring*. Ed. by R. Noorossana, A. Saghaei, and A. Amiri. Hoboken, New Jersey: Wiley, pp. 1–18.

Schettlinger, K., Fried, R., and Gather, U. (2006). Robust filters for intensive care monitoring: Beyond the running median. *Biomedizinische Technik* 51(2), pp. 49–56.

Schettlinger, K., Fried, R., and Gather, U. (2010). Real time signal processing by adaptive repeated median filters. *International Journal of Adaptive Control and Signal Processing* 24(5), pp. 346–362.

Serfling, R. (2011). Asymptotic relative efficiency in estimation. In: *International Encyclopedia of Statistical Science*. Ed. by M. Lovric. Berlin: Springer, pp. 68–72.

Sharpsteen, C. and Bracken, C. (2018). *tikzDevice: R graphics output in LaTeX format*. R package version 0.11.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. New York: D. Van Nostrand Company, Inc.

Siedhoff, D., Libuschewski, P., Weichert, F., Marwedel, P., and Müller, H. (2014a). Modellierung und Optimierung eines Biosensors zur Detektion viraler Strukturen. In: *Bildverarbeitung für die Medizin*. Berlin: Springer, pp. 108–113.

Siedhoff, D., Weichert, F., Libuschewski, P., and Timm, C. (2011). Detection and classification of nano-objects in biosensor data. *International Conference on Systems Biology - Microscopic Image Analysis with Applications in Biology (MIAAB 2011)*, pp. 1–6.

Siedhoff, D., Zybin, A., Shpacovitch, V., and Libuschewski, P. (2014b). PAMONO Sensor Data 200nm_10Apr13.SFB876.doi:10.15467/e9ofqnvl6o.

Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika* 69(1), pp. 242–244.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley.

Warnes, G. R., Bolker, B., and Lumley, T. (2015). *gtools: Various R programming tools*. R package version 3.5.0.

Weichert, F., Gaspar, M., Timm, C., Zybin, A., Gurevich, E. L., Engel, M., Müller, H., and Marwedel, P. (2010). Signal analysis and classification for plasmon assisted microscopy of nanoobjects. *Sensors and Actuators B: Chemical* 151(1), pp. 281–290.

Weiß, C. H., Steuer, D., Jentsch, C., and Testik, M. C. (2018). Guaranteed conditional ARL performance in the presence of autocorrelation. *Computational Statistics & Data Analysis* 128, pp. 367–379.

Wichman, R., Astola, J. T., Heinonen, P. J., and Neuvo, Y. A. (1990). FIR-Median hybrid filters with excellent transient response in noisy conditions. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38(12), pp. 2108–2117.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software* 40(1), pp. 1–29.

Wickham, H. (2017). *scales: Scale functions for visualization*. R package version 0.5.0.

Wickham, H., Francois, R., Henry, L., and Müller, K. (2017). *dplyr: A grammar of data manipulation*. R package version 0.7.2.

Wickham, H. and Henry, L. (2017). *tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. R package version 0.7.1.

Wilcox, R. R. (2003). *Applying Contemporary Statistical Techniques*. Amsterdam: Academic Press.

Woodall, W. H. and Montgomery, D. C. (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology* 46(1), pp. 78–94.

Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics* 21(3), pp. 1545–1566.

Zeileis, A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Economic Reviews* 24(4), pp. 445–466.

Zhou, M. and Goh, T. N. (2016). Effects of model accuracy on residual control charts. *Quality and Reliability Engineering International* 32(5).

Zybin, A., Kuritsyn, Y. A., Gurevich, E. L., Temchura, V. V., Überla, K., and Niemax, K. (2010). Real-time detection of single immobilized nanoparticles by surface plasmon resonance imaging. *Plasmonics* 5(1), pp. 31–35.