technische universität
dortmund

Blockwise estimation of parameters under abrupt changes in the mean

DISSERTATION

in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften

presented to the

Faculty of Statistics
TU Dortmund University

by

Ieva Axt geb. Zelo

Submitted:	Dortmund, January 2020
Advisor and primary referee:	Prof. Dr. Roland Fried
Secondary referee:	Prof. Dr. Christine Müller
Day of the oral examination:	12th February 2020

Contents

1 Introduction

Structural change in time series is an important issue in statistics. For example changes in exchange rates (Indonesian Rupiah and US Dollar, see e.g. [33]) and stock market prices (see e.g. Chapter 2.2.4 in [11]), sudden climate or temperature changes (see e.g. [95]), unexpected hydrological phenomena or construction of dams resulting in changes in monthly discharge of a river (see e.g. [34] or [37]) make parameter estimation difficult, since many estimators, such as the sample variance, rely on the assumption that the underlying process does not change its properties over time. We consider a special case of structural changes – abrupt shifts in the mean. Our goal is to estimate other parameters properly in this scenario. To achieve satisfying results under shifts in the mean we propose to segregate the data into non-overlapping or overlapping blocks and to estimate the parameter of interest in each block separately. Subsequently the blocks estimates are combined to obtain the final estimate of the parameter. This thesis consists of two parts, which are described in the following.

Estimation of the Hurst parameter under shifts in the mean

Chapter 2 of this thesis is based on the article "Estimation methods for the LRD parameter under a change in the mean" [72], where we focus on estimation of the Hurst parameter H, which characterizes the intensity of long range dependence (LRD) in time series. This type of dependence is described by a slowly decaying autocovariance function, while short range dependent processes, such as the ARMA process, exhibit exponentially decaying autocovariances. Ordinary estimators of the LRD parameter, such as the local Whittle ([50]) or the Geweke and Porter-Hudak (GPH, proposed by [28]) estimators cannot distinguish between LRD and jumps in the mean and tend to overestimate the parameter H in the presence of jumps. As a consequence, change-point tests, such as the Wilcoxon change-point test proposed by [19], often do not reject the hypothesis of no changes in the mean. This is due to the fact that the corresponding test statistic involves the true parameter H, which has to be estimated properly under level shifts. To cope with this problem we investigate approaches based on segregating the sample into blocks and estimating the LRD parameter on each block separately. The resulting estimates are then combined to get a final estimate of H. We examine segregation into two blocks considering all possible positions or using the test statistic of the Wilcoxon change-point test to estimate the shift location. An average value of estimates obtained from the two blocks yields the final estimate of H. As this procedure is designed for the case of at most one level shift we propose to segregate the data into many overlapping or non-overlapping blocks obtaining many estimates, which are then combined by averaging. This procedure yields satisfying results in our simulation studies and is MSE-consistent as is shown in this thesis. This feature implies weak consistency, which is useful when employing the estimator in the test statistic of the Wilcoxon change-point test to preserve convergence in distribution.

Scale estimation under shifts in the mean

The results of Chapter 3 of this thesis originate from the articles “On variance estimation under shifts in the mean” [4] and “Robust scale estimation under shifts in the mean” [3]. The focus of this part is on estimation of the variance or standard deviation under level shifts in the possible presence of outliers. We use the same idea of segregating the data into many blocks, since this procedure improves the performance of ordinary estimation techniques considerably in the context of LRD.

If only a few level shifts are present we propose usage of the average value of sample variances, obtained from many non-overlapping blocks. Under some conditions on the number of change-points and the number of blocks we show strong consistency of this estimator under independence. For weakly dependent stationary processes weak consistency is shown. In the absence of level shifts the blocks estimator is asymptotically normal with the same parameters as in the case of the ordinary sample variance. In the presence of many changes in the mean this procedure is highly biased in finite samples. Therefore, we investigate an adaptively trimmed mean of blockwise estimates. This procedure is designed for independent data and yields very good results in our simulations, even for autocorrelated data.

In the presence of outliers these are no longer suitable estimators, since they are not robust. Thus, we propose a modified version of the median absolute deviation (MAD), where the sample median is calculated in blocks rather than on the whole sample. By doing so, we make sure that absolute differences from only few blocks are biased, while the rest are not. The proposed estimator yields very good results in our simulation study. We show strong consistency of the modified MAD under some conditions on the number of change-points and the number of blocks. When no level shifts are present the estimator is shown to be asymptotically normal, with the same asymptotic variance as in the case of the ordinary MAD.

In Chapter 4 the results of this thesis are summarized and an outlook on future work is given. Some proofs, supplementary figures and tables can be found in the Appendix.

2 Estimation methods for the LRD parameter under a change in the mean

This chapter is based on the article "Estimation methods for the LRD parameter under a change in the mean" [72], published in *Statistical Papers*. This article is joint work with Aeneas Rooch (Ruhr-University Bochum) and Roland Fried (TU Dortmund University). Estimation of the Hurst parameter, which describes the strength of dependence in time series with long memory, under shifts in the mean is in the focus of this article.

The blockwise estimation techniques discussed in Section 2.3 are considered by Aeneas Rooch in his dissertation, see [71]. The focus is on the improvement of the performance of the change-point tests, such as the Wilcoxon change-point test, see [19]. The corresponding test statistic involves the true Hurst parameter, which has to be estimated properly under shifts in the mean. Aeneas Rooch segregates the data into several blocks and estimates the Hurst parameter using the Whittle estimator, a parametric estimation technique, which requires knowledge about the true data generating process. Also the Box-periodogram estimator is used, which is not based on the knowledge of the true model, but is highly biased in many situations.

In this thesis we focus on estimating the Hurst parameter properly. Different estimators, such as the Geweke and Porter-Hudak (GPH) estimator (see [28]), which do not require knowledge about the true model, are employed to estimate the Hurst parameter. Extensive simulations are performed to compare the performance of the estimators with each other as well as with other estimators from literature. Scenarios where data exhibit short range dependence are included. Furthermore, the estimation techniques are refined using an ARMA correction procedure to improve the results when dealing with an ARFIMA (fractionally integrated ARMA) model, which involves components of short and long range dependence. Asymptotic MSE-consistency, and therefore convergence in probability, of the non-overlapping and overlapping blocks approaches when using the local Whittle or the GPH estimator in every block is proven. Moreover, the assumption of at most one jump in the mean is relaxed by allowing for multiple changes in the mean. It is also shown that the test statistic of the Wilcoxon change-point test has the same asymptotic distribution when using the blockwise estimator for the Hurst parameter instead of the true one. Finally, the estimation techniques are applied to real data sets.

For better reading minor changes have been incorporated in the article. A major change is the consideration of a modified version of the GPH estimator in each block to prove the consistency of the blocks estimators, see Section 2.4. Since the Hurst parameter $H \in (0,1)$ has to be estimated and the regression based GPH estimator $\widehat{H}$ may take values outside the interval $(0,1)$, we consider the modified estimator $\widehat{H}\mathbb{I}_{(0,1)}(\widehat{H}) + \mathbb{I}_{[1,\infty)}(\widehat{H})$. Moreover, the consistency of the blocks estimation technique is also shown to hold when using the local Whittle estimator, see Section 2.4.

2.1 Introduction

A stochastic process $(\xi_t)_{t\geq 1}$ is said to exhibit long range dependence (LRD) or long memory if the dependence between ξ_t and ξ_s vanishes slowly as the distance $|t-s|$ between the variables increases. The autocovariance of short range dependent processes, such as the ARMA (autoregressive moving average) process, decreases exponentially as $|t-s| \to 0$, see e.g. Chapter 3 in [36]. As opposed to this, the autocovariance of a process with long memory has a much slower decay (see (2.1)).

Long-range dependence is an issue in several fields of time series analysis, ranging from climate sciences [86] and network data traffic [23], [52], [47] to economics and finance. In the latter context, volatilities may exhibit long memory [10], and there are discussions whether long-range dependence can be found in stock market prices [100], [12], [7], [15], [53]; for a survey on long-range dependence in economics, see [5].

We consider a discrete-time second order stationary stochastic process $(\epsilon_t)_{t\geq 1}$ defined by

$$\epsilon_t = G(\xi_t), \quad t \geq 1,$$

where $(\xi_t)_{t\geq 1}$ is a stationary Gaussian process with $E(\xi_t) = 0$ and $E(\xi_t^2) = 1$ $\forall t$ and autocovariance function γ with

$$Cov(\xi_t, \xi_s) = \gamma(|t-s|) = \gamma(k) = k^{-D}L(k), \quad k \geq 1, \tag{2.1}$$

where $k = |t-s|$, $0 < D < 1$ and $L(k)$ is a slowly varying function with $L(bx)/L(x) \to 1$ for $x \to \infty$ $\forall b > 0$. $G : \mathbb{R} \to \mathbb{R}$ is a measurable function, which satisfies $E(G(\xi_t)) = 0$. The process $(\xi_t)_{t\geq 1}$ exhibits LRD due to the slow decay and therefore the non-summability of the autocovariances. Equivalently, stochastic processes with long memory can be characterized using the spectral density f:

$$\lim_{\lambda \to 0} \frac{f(\lambda)}{c_f |\lambda|^{-\beta}} = 1,$$

with $\beta \in (0,1)$ and a constant $c_f > 0$, see Chapter 2.1 in [8].

The exponent $D \in (0,1)$ in (2.1) describes the intensity of the LRD. It is more common to use the fractional differencing parameter $d = (1-D)/2 \in (0, 1/2)$ in the context of fractionally integrated ARMA (ARFIMA) processes or the Hurst parameter $H = 1 - D/2 = d + 1/2 \in (1/2, 1)$ in the context of LRD and stationary time series. The higher H is, the slower is the decay of the autocovariance and the stronger is the dependence of the random variables.

One example of a long memory process is the fractional Gaussian noise (fGn), a process of first order differences of the fractional Brownian motion. Furthermore, a stochastic process described by an ARFIMA(p, d, q) model with AR parameter p, MA parameter q and a differencing parameter d exhibits long range dependence. See Appendix A for a brief introduction of those processes and [8] for more information.

Remark 1. While $H \in (1/2, 1)$ corresponds to long range dependent time series, the value $H \leq 1/2$ implies short range dependence. $H = 1$ results in time series with correlation between two random variables Y_t and Y_s equal to one for all s, t, i.e., $Cov(X_t, X_s)/\sqrt{Var(X_t)}\sqrt{Var(X_s)} = 1, \forall s, t$. The correlation between the random variables of the process diverges to infinity in the case $H > 1$, which is a contradiction to the fact that the correlation between two random variables must be in the interval $[-1, 1]$. The case $H \geq 1$ may occur for stochastic processes with infinite variances, which is not of interest in this thesis. In the case of $H < 0$ the stochastic process is not measurable. See Chapter 2.3 in [8] for more information. Hence, only $H \in (0, 1)$ is considered in this thesis, specifically $H \in (1/2, 1)$, since we are interested in long range dependent processes.

The rest of this chapter is organized as follows: In Section 2.2 we present the model of the data generating process. In Section 2.3 we introduce our adaption techniques together with the methods of [56] and [44] for estimating H under a jump in the mean. In Section 2.4 we prove the consistency of the blocks estimators, and in Section 2.5 we demonstrate the usefulness of the proposed estimators when using the Wilcoxon change-point test. In Section 2.6 we analyse the performance of our methods in a simulation study. In Section 2.7 we apply the estimation methods to real data. In Section 2.8 we give a summary and an outlook.

2.2 Model

In this thesis we consider K change-points of possibly different heights $h_1, ..., h_K$ at unknown locations $t_1, ..., t_K$, i.e., a process $(Y_t)_{t\geq 1}$ of the type

$$Y_t = \sum_{k=1}^{K} h_k I_{t\geq t_k} + \epsilon_t = \sum_{k=1}^{K} h_k I_{t\geq t_k} + G(\xi_t), \tag{2.2}$$

for $t = 1, \ldots, N$, where N denotes the sample size. There are other change-point problems: changes in the marginal distribution of the observations [30], in the coefficients of linear regression models with long range dependent errors [49] or in the dependence structure [35], [65]. All these procedures require knowledge of a LRD parameter, e.g. the Hurst parameter. More information on change-point analysis see e.g. [14] or [45].

2.3 Estimation methods

The literature offers various approaches to estimate the Hurst parameter. In [42] a rescaled range (R/S) statistic for estimation of the Hurst parameter was introduced. A parametric Whittle approach involving the approximate log-Likelihood was introduced by [99]. Based on this estimation technique [50] proposed the semiparametric local Whittle estimator, while consistency and normality of this estimator was shown by [69]. Consistency and the distribution of the local Whittle estimator in the nonstationary case were considered by [92] and [93] involving data tapers. Dependence of the convergence rate and the limit distribution on H was introduced in [81]. Cases in which the local Whittle estimator is not consistent are discussed in [63]. Another semiparametric estimator of the Hurst parameter is the Geweke and Porter-Hudak (GPH) estimator based on the log-periodogram regression, proposed by [28]. In contrast to the local Whittle estimator the GPH estimator has a closed form. While

[28] used all frequencies and the corresponding periodogram values in the formula of the estimator, [70] introduced a trimmed version of this estimation procedure by truncating the lowest and the highest frequencies to improve the performance of the estimator. For further overview on estimators of the LRD parameter see [88] or [89], where e.g. the Box-Periodogram method, Peng's method and estimation involving absolute values of aggregated series are described.

Unfortunately, ordinary estimators usually overestimate the Hurst Parameter under changes in the mean. In [82] it was discussed how level shifts lead to a misspecification of the dependence structure and therefore to a possible overestimation of the Hurst parameter.

Thus, the Hurst parameter H needs to be estimated properly taking the level shifts into account. This is e.g. crucial for change-point tests, such as the Wilcoxon change-point test proposed by [19], since the corresponding test statistic involves the true parameter H, which has to be estimated.

To estimate the Hurst parameter under shifts in the mean, [40] investigated usage of the local Whittle estimator, which requires knowledge of the number of change-points. [56] used a trimmed version of the Geweke and Porter-Hudak estimator to remove the effect of the change in the mean by trimming the lower frequencies in the spectral domain. The estimation method of [44] is also based on trimming the lowest frequencies, applied to the local Whittle estimator, originally proposed by [50]. We will restrict ourselves on the latter two approaches in the simulations (see Section 2.6), since the method proposed by [40] requires knowledge about the number of level shifts.

We investigate estimation methods based on estimating the parameter in blocks and then combining the estimates. This general idea can be combined with any estimation technique, such as the GPH estimator. We examine the overlapping and the non-overlapping blocks approach along with two further variants of this idea for comparison and verify that they improve the respective standard estimators, which do not take jumps in the mean into account. We compare our proposals with the adapted approaches of [56] and [44].

2.3.1 Generic estimation techniques

In what follows we will involve the two following generic techniques in the estimation of the Hurst parameter.

GPH estimator

The regression based GPH estimator, which was first introduced in [28], uses the spectral representation of the long memory processes. The regression is based on the representation

$$\log\left(f(\lambda)\right) \approx c - (2H-1)\log(\lambda)$$

of the spectral density $f(\lambda)$ as $\lambda \to 0$, see [56]. The corresponding periodogram $I(\lambda_j)$ at frequency $\lambda_j = 2\pi j/N$, $j = 1, ..., N$, is an estimator of the spectral density $f(\lambda_j)$ and is defined as follows:

$$I(\lambda_j) = w(\lambda_j)w(-\lambda_j), \quad \text{where}$$
$$w(\lambda_j) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} Y_t \exp(i\lambda_j t),$$

see [44]. An estimator $\widehat{d}$ of the slope d in the regression

$$\log\left(I(\lambda_j)\right) = c + dz_j + e_j$$

yields the GPH estimator $\hat{H}_{\text{GPH}} = \widehat{d} + 0.5$, i.e.,

$$\hat{H}_{\text{GPH}} = -0.5 \sum_{j=1}^{b} (z_j - \bar{z}) \log(I(\lambda_j)) / \sum_{j=1}^{b} (z_j - \bar{z})^2 + 0.5,$$

where $z_j = \log|1 - \exp(-i\lambda_j)|$, $\bar{z} = b^{-1} \sum_{j=1}^{b} z_j$ and $b = \lfloor N^u \rfloor$, $u > 0$, is the bandwidth parameter, which satisfies $b \log(b)/N = o(1)$ for the consistency of the parameter estimator (see [43]). We will use $u = 2/3$ in the following.

Local Whittle estimator

The local Whittle (LW) estimator $\hat{H}_{\text{LW}}$ proposed in [50] minimizes the approximate frequency domain likelihood function

$$R(H) = \log\left(\frac{1}{b}\sum_{j=1}^{b} \lambda_j^{2H-1} I(\lambda_j)\right) - (2H-1)\frac{1}{b}\sum_{j=1}^{b} \log(\lambda_j)$$

with respect to $H \in (0,1)$, where b is a bandwidth parameter, where the relationship

$$f(\lambda) = G\lambda^{2H-1}$$

as the frequency λ decreases to zero with $G \in (0,\infty)$ and $H \in (0,1)$ is assumed, see [69]. We set $b = 0.8N^{0.79}$, as is done in [44].

2.3.2 Pre-estimating the jump

The following approach is based on the assumption of at most one level shift of height h_1 at location t_1, i.e., $K \leq 1$, see model (2.2). An intuitive approach is removal of a jump before application of any of the previous methods. First, the time point t_1 has to be estimated. Then we remove the level shift and estimate the Hurst parameter H on the corrected time series. Another possibility is to estimate H before and after the level shift and average over the two estimates to get an overall estimate. We use the Wilcoxon change-point test proposed by [19] (see Section 2.5), which checks the null hypothesis of no change in the mean using the following test statistic

$$W_n = \max_{1 \leq k \leq n} |W_{k,n}|,$$

with $W_{k,n}$ as defined in Section 2.5. See [19] for further information. For large values of W_n the null hypothesis is rejected. The estimated time point t_1 of the level shift is then defined as follows:

$$\hat{t}_1 = \underset{1 \le k \le N-1}{\arg\max} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{N} \left(I_{\{X_i \le X_j\}} - \frac{1}{2} \right) \right|.$$

The estimator $\hat{t}_1/N$ is a weakly consistent estimator for the true fraction $\tau \in (0, 1)$ of observations after which the jump occurs as can be proven using the arguments of [39], see [9]. In the absence of level shifts the pre-estimation of $\hat{t}_1$ does not influence the following estimation approach considerably.

We segregate the sequence $Y_1, \ldots, Y_N$ into two blocks

$$Y_1, \ldots, Y_{\hat{t}_1} \quad \text{and} \quad Y_{\hat{t}_1+1}, \ldots, Y_N.$$

The height h_1 is then estimated as

$$\hat{h}_1 = \frac{1}{N - \hat{t}_1}(Y_{\hat{t}_1+1} + \ldots + Y_N) - \frac{1}{\hat{t}_1}(Y_1 + \ldots + Y_{\hat{t}_1}),$$

yielding the corrected time series

$$Y_1, \ldots, Y_{\hat{t}_1}, Y_{\hat{t}_1+1} - \hat{h}_1, \ldots, Y_n - \hat{h}_1. \tag{2.3}$$

Subsequently, we estimate the Hurst parameter H on the corrected time series (2.3). The corresponding estimator is denoted by $\hat{H}_{\text{pre},1}$.

Another possibility is estimating H before and after the estimated jump location $\hat{t}_1$, resulting in the two estimates $\hat{H}^{(1)}_{\hat{t}_1}$ and $\hat{H}^{(2)}_{\hat{t}_1}$. The average over those two values is an overall estimate of H:

$$\hat{H}_{\text{pre},2} = \frac{\hat{H}^{(1)}_{\hat{t}_1} + \hat{H}^{(2)}_{\hat{t}_1}}{2}.$$

Remark 2. Since estimation under the assumption of LRD requires many observations, we can avoid dealing with too small blocks by setting

$$\tilde{t} = \underset{t \in \mathcal{K}}{\arg\min} \left| t - \hat{t}_1 \right|,$$

where

$$\mathcal{K} = \{k_{\text{low}}, k_{\text{low}} + 1, \ldots, k_{\text{up}}\}, \tag{2.4}$$

is a set of candidate points and

$$k_{\text{low}} = \max\{\lfloor N/10 \rfloor, 10\} + 1,$$
$$k_{\text{up}} = N - k_{\text{low}}.$$

2.3.3 Estimation from two blocks

Instead of applying a change-point test statistic for splitting the sample into two subsets, we can consider any splitting of the data into two blocks of reasonable size,

$$Y_1, \dots, Y_k \quad \text{and} \quad Y_{k+1}, \dots, Y_N$$

and estimate H on each block. This is repeated for all candidate points k. Early or late time points resulting in small blocks are not considered. We consider $\mathcal{K}$ from (2.4) as a set of candidate points yielding $|\mathcal{K}| = k_{\text{up}} - k_{\text{low}} + 1$ tuples $(\hat{H}_k^{(1)}, \hat{H}_k^{(2)})$, obtained before and after the candidate time point, respectively.

We consider two possible methods to estimate H based on the estimates $\hat{H}_k^{(1)}, \hat{H}_k^{(2)}, k \in \mathcal{K}$: the average value of all obtained estimates

$$\hat{H}_{\text{me}} = \frac{1}{2|K|} \sum_{k \in K} \left(\hat{H}_k^{(1)} + \hat{H}_k^{(2)} \right)$$

and the average value of the left and the right estimate

$$\hat{H}_{\text{diff}} = \frac{\hat{H}_{k^*}^{(1)} + \hat{H}_{k^*}^{(2)}}{2}$$

obtained by segregating the observations at a candidate point k^*, which yields the least difference between the two estimates:

$$k^* = \arg\min_{k \in K} \left| \hat{H}_k^{(2)} - \hat{H}_k^{(1)} \right| .$$

2.3.4 Estimation from many blocks

The techniques presented in the previous subsections rely on the existence of only one jump and finding an adequate splitting into two blocks. Our main proposal extends this idea by splitting the data into many blocks

$$Y_{j-w}, Y_{j-w+1}, \dots, Y_j, \dots, Y_{j+w}$$

of size $n = \bar{w} = 2w + 1$ around the observation Y_j and estimating H on this window, leading to $N - 2w$ estimates $\check{H}_{n,j}$ for H, where $j = w + 1, ..., N - w$. Subsequently, we average over all values obtaining the estimate

$$\hat{H}_{o,n} = \frac{1}{N - 2w} \sum_{j=w+1}^{N-w} \check{H}_{n,j}. \tag{2.5}$$

We choose the flank length $w = w(N)$ (resulting in a block size $n = \bar{w} = 2w + 1$) depending on the sample size N,

$$w = w(N) = \max \left\{ \lfloor \sqrt{N} \rfloor, 10 \right\} , \tag{2.6}$$

since the block size should be reasonably large when estimating under the LRD assumption. Moreover, we also investigate the choice $n = w$, i.e., shorter windows.

Furthermore, we investigate segregation of the observations $Y_1, ..., Y_N$ into $m = \lfloor N/n \rfloor$ non-overlapping blocks of size $n = \bar{w} = 2w + 1$

$$Y_1, \ldots, Y_n, \qquad Y_{n+1}, \ldots, Y_{2n}, \qquad Y_{2n+1}, \ldots, Y_{3n}, \quad \ldots \quad Y_{(\lfloor N/n \rfloor - 1)n+1}, \ldots, Y_{\lfloor N/n \rfloor n},$$

yielding $m = \lfloor N/n \rfloor$ estimates $\check{H}_{n,k}$, $k = 1, \ldots, m$. As an estimate for H we consider the average of the estimates

$$\hat{H}_{no,n} = \frac{1}{m} \sum_{k=1}^{m} \check{H}_{n,k}. \tag{2.7}$$

We also consider the smaller block size $n = w$ with w as in (2.6) for comparison.

2.3.5 Trimmed estimators

A trimmed version of the regression based GPH estimator was investigated by [56] where the lowest frequencies are trimmed, as these are affected by changes in the mean. Let $I(\lambda_j)$ be the value of the periodogram at frequency $\lambda_j = 2\pi j/N$. The trimmed GPH estimator is

$$\hat{H}^{\text{tr}}_{\text{GPH}} = -0.5 \sum_{j=l}^{b} (z_j - \bar{z}) \log(I(\lambda_j)) / \sum_{j=l}^{b} (z_j - \bar{z})^2 + 0.5,$$

where $z_j = \log|1 - \exp(-i\lambda_j)|$ and $\bar{z} = (b - l + 1)^{-1} \sum_{j=l}^{b} z_l$. The trimming parameter is denoted by $l = \lfloor N^{1/2+\epsilon} \rfloor$, $\epsilon > 0$, while $b = \lfloor N^u \rfloor$, $u > 0$, is the bandwidth parameter. Moreover, the adapted GPH estimation technique was proposed, where the trimming parameter is computed recursively. Let $l_i = \lfloor N^{(2-2\hat{H}_{i-1})/(3-2\hat{H}_{i-1})+\epsilon} \rfloor$, where $\hat{H}_1$ is computed using $l_1 = \lfloor N^{1/2+\epsilon} \rfloor$. The value $\hat{H}_i$ denotes the estimate computed using the trimming parameter l_i. Then the Hurst parameter is estimated by $\hat{H}_i$ if $|\hat{H}_i - \hat{H}_{i-1}| < 0.01$ or $i > 9$. The authors suggest values $\epsilon = 0.05$ and $u = 0.8$. We will use the adapted trimmed estimator for comparison in this thesis.

Similarly, [44] proposed the trimmed version of the local Whittle estimator, where the computation of the estimate is also based only on the higher frequencies – this shall take into account a change in the mean. The trimmed local Whittle estimator $\hat{H}^{\text{tr}}_{\text{LW}}$ minimizes the function

$$R(H) = \log\left(\frac{1}{b - l + 1} \sum_{j=l}^{b} \lambda_j^{2H-1} I(\lambda_j)\right) - (2H - 1) \frac{1}{b - l + 1} \sum_{j=l}^{b} \log(\lambda_j)$$

with respect to $H \in (0, 1)$, where l is the trimming parameter and b is the bandwidth parameter. Suggested values are $l = \lfloor 1 + 0.2N^{0.62} \rfloor$ and $b = \lfloor 0.8N^{0.79} \rfloor$.

2.4 Inheriting asymptotic MSE-consistency

In Section 2.6 we will see that our main proposal to split the data into many blocks and estimate H separately improves estimation of the Hurst parameter considerably. In this section we investigate conditions for the asymptotic MSE-consistency of this approach. The estimators of H constructed using overlapping or non-overlapping blocks inherit asymptotic MSE-consistency in the absence or presence of one or even several jumps, as is demonstrated in the following.

Theorem 1. *Let $(\xi_t)_{t\geq 1}$ be a stationary Gaussian LRD process, with mean 0, variance 1 and autocovariances satisfying* (2.1) *and consider observations $Y_1, \ldots, Y_N$ of the type $Y_t = \sum_{k=1}^{K} h_k I_{t\geq t_k} + G(\xi_t)$ for a measurable function $G : \mathbb{R} \to \mathbb{R}$ satisfying $E(G(\xi_t)) = 0$, see model (2.2). Consider the overlapping blocks estimator* (2.5) *with a monotone increasing window length $n = 2w + 1 = o(N)$ with $n \to \infty$, where w is the flank length. As generic estimation method $\check{H}_{n,j}$ on the j-th window choose any estimator $\check{H}_{n,j}$ for H, which is asymptotically MSE-consistent in the absence of level shifts. Moreover, we require that the variance of the estimator $\check{H}_{n,\tilde{k}}$ in a jump-contaminated block $\tilde{k}$ grows sufficiently slowly, i.e.,*

$$Var(\check{H}_{n,\tilde{k}}) = o\left(\frac{N^2}{K^2 n^2}\right). \tag{2.8}$$

Let $\tilde{K}$ denote the index set of the windows, which include at least one jump, with $|\tilde{K}| \leq K$. For the number of change-points we assume that

$$K = o\left(\frac{N}{n \log^2(N)}\right) = o\left(\frac{m}{\log^2(N)}\right) \tag{2.9}$$

holds, i.e., the number of level shifts grows slower than the number of blocks m.

Then the overlapping blocks estimator (2.5) *is asymptotically MSE-consistent.*

Proof of Theorem 1 for the local Whittle estimator. In the following we will use the local Whittle estimator $\hat{H}_{\mathrm{LW}}$, see Section 2.3.1. This estimator is the solution of a minimization problem over a bounded interval $(0, 1)$. Therefore, the variance of this estimator is naturally bounded by one. Moreover, the variance and the bias of $\hat{H}_{\mathrm{LW}}$ tend to zero in the absence of level shifts, see [68, p. 223].

We assume that we are dealing with $K \cdot n$ jump-contaminated overlapping windows, which is the maximum number and therefore represents the worst case scenario. Due to the stationarity of the $(Y_t)_{t\geq 1}$ in the absence of level shifts (i.e., $Y_t = G(\xi_t)$, see model (2.2)) we have

$$E\left(\hat{H}_{o,n} - H\right) = \frac{1}{N-n+1} \sum_{j=w+1}^{N-w} E(\check{H}_{n,j} - H) = \frac{N-n+1}{N-n+1} E(\check{H}_{N,1} - H) \to 0.$$

In the presence of level shifts we obtain

$$\begin{aligned} E\left(\hat{H}_{o,n} - H\right) &= \frac{1}{N-n+1} \sum_{j=w+1, j\notin \tilde{K}}^{N-w} E\left(\check{H}_{n,j} - H\right) + \frac{1}{N-n+1} \sum_{\tilde{k}\in \tilde{K}} E\left(\check{H}_{n,\tilde{k}} - H\right) \\ &= \frac{N-n+1-Kn}{N-n+1} E(\check{H}_{n,1} - H) + \frac{Kn}{N-n+1} O\left(1\right) \to 0, \end{aligned} \tag{2.10}$$

since (2.9) holds and $E(\check{H}_{n,\tilde{k}})$, $E(H)$ are both in the interval $(0,1)$. This proves the asymptotic unbiasedness. Moreover, we obtain for the variance

$$\begin{aligned} Var(\hat{H}_{o,n}) &= \frac{1}{(N-n+1)^2}\left(\sum_{k=w+1}^{N-w} Var(\check{H}_{n,k}) + \sum_{k=w+1, j\neq k}^{N-w} Cov(\check{H}_{n,k}, \check{H}_{n,j})\right) \\ &\leq \frac{1}{(N-n+1)^2}\left(\sum_{k=w+1}^{N-w} Var(\check{H}_{n,k}) + \sum_{k=w+1, j\neq k}^{N-w} \sqrt{Var(\check{H}_{n,k})}\sqrt{Var(\check{H}_{n,j})}\right) \\ &= \frac{Kn}{(N-n+1)^2} Var(\check{H}_{n,\tilde{k}}) && (2.11) \\ &+ \frac{N-n+1-Kn}{(N-n+1)^2} Var(\check{H}_{n,1}) && (2.12) \\ &+ \frac{2Kn(N-n+1-Kn)}{(N-n+1)^2}\sqrt{Var(\check{H}_{n,\tilde{k}})}\sqrt{Var(\check{H}_{n,1})} && (2.13) \\ &+ \frac{Kn(Kn-1)}{(N-n+1)^2} Var(\check{H}_{n,\tilde{k}}) && (2.14) \\ &+ \frac{(N-n+1-Kn)(N-n-Kn)}{(N-n+1)^2} Var(\check{H}_{n,1}) && (2.15) \\ &\to 0. \end{aligned}$$

The terms in (2.12) and (2.15) tend to zero as $N \to \infty$, since we are dealing with a MSE-consistent estimator of the Hurst parameter. The terms (2.11), (2.13) and (2.14) represent the decomposition of the covariance part and tend to zero, since the variance of the local Whittle estimator is bounded by one satisfying condition (2.8).

The MSE-consistency follows from the convergence of the bias and the variance to zero. □

Remark 3. 1. The above Theorem 1 can also be proven using the non-overlapping estimator (2.7).

2. We can also use the GPH estimator in the above proof with a slight modification:

   As noted in Remark 1 the Hurst parameter H lies in the interval $(0,1)$. However some estimators, such as the GPH estimator proposed by [28], may take values which are outside this range. Therefore, for simplicity, this estimator can be modified by bounding its possible outcomes. I.e., we consider an estimator of the type $\widehat{H}\mathbb{I}_{(0,1)}(\widehat{H}) + \mathbb{I}_{[1,\infty)}(\widehat{H})$, where $\hat{H}$ is the GPH estimator. The variance of such estimator is bounded and lies in the interval $[0,1]$ satisfying condition (2.8). The variance and the bias of the ordinary GPH estimator $\hat{H}$ tends to zero as $N \to \infty$ implying convergence in distribution.

Moreover, we have that $\mathbb{I}_{(0,1)}(\hat{H}) \to 1$ and $\mathbb{I}_{[1,\infty)}(\widehat{H}) \to 0$ in probability. Applying Slutsky's Theorem we get that $\widehat{H}\mathbb{I}_{(0,1)}(\widehat{H}) + \mathbb{I}_{[1,\infty)}(\widehat{H}) \to H$ in distribution. Uniform integrability of the random variable $\widehat{H}\mathbb{I}_{(0,1)}(\widehat{H}) + \mathbb{I}_{[1,\infty)}(\widehat{H})$ follows from the bounded variance, see Lemma 1.4A in [79].

Convergence in distribution and uniform integrability of $\widehat{H}\mathbb{I}_{(0,1)}(\widehat{H}) + \mathbb{I}_{[1,\infty)}(\widehat{H})$ imply convergence of the first moment, i.e., $E\left(\widehat{H}\mathbb{I}_{(0,1)}(\widehat{H}) + \mathbb{I}_{[1,\infty)}(\widehat{H})\right) \to H$, see Theorem 1.4A in [79].

Therefore, the bias $E\left(\widehat{H}\mathbb{I}_{(0,1)}(\widehat{H}) + \mathbb{I}_{[1,\infty)}(\widehat{H})\right) - H$ tends to zero and Theorem 1 is proven, since (2.10) in the above proof holds.

2.5 Application to the Wilcoxon change-point test

If one wants to test whether a sample $Y_1, \ldots, Y_N$ includes a change-point, i.e., a point at which the structure of the data changes, knowledge about H is essential since it is involved in the scaling of test statistics. So is the function L, which appears in (2.1), but we do not concentrate on this issue in this thesis.

Change-point problems for LRD time series have been in the focus of research for many years. They may concern a possible change in the mean. Here, one wants to test the null hypothesis

$$H_0 : \mu_1 = \ldots = \mu_n$$

of a constant mean against the alternative

$$A : \mu_1 = \ldots = \mu_k \neq \mu_{k+1} = \ldots = \mu_N \quad \text{for some } k \in \{1, \ldots, N-1\}$$

that there is a change-point at which the level changes, yielding the test problem (H_0, A). CUSUM-type tests for the test problem (H_0, A) have been studied e.g. by [39], [48] or [96]. In [97] a test procedure based on Wilcoxon rank statistics is analysed, while [19] proposed a test, which is based on the Wilcoxon two-sample test statistic:

$$W_{k,N} := \frac{1}{N\, d_N} \sum_{t=1}^{k} \sum_{s=k+1}^{N} \left(1_{\{Y_t \leq Y_s\}} - \frac{1}{2}\right)$$

where

$$d_N^2 \sim c_r\, N^{2-r(2-2H)} L^r(N) \tag{2.16}$$

with the Hermite rank $r \in \mathbb{N}$ of the class of functions $\{1_{G(\xi_t) \leq x} - F(x), x \in \mathbb{R}\}$ and $c_r = \lfloor 2r!/(1-(2-2H)r)(2-(2-2H)r) \rfloor$ being the right scaling for $W_{k,n}$ to have a non-degenerate limit distribution under H_0 if the Y_t have a continuous distribution function F. The proposed change-point test for the test problem (H_0, A) rejects the null hypothesis H_0 for large values of the following test statistic

$$W_N := \max_{1 \leq k \leq N} |W_{k,N}|\,.$$

Under H_0 [19] proved that W_N with $r = 1$ converges in distribution to

$$\frac{1}{2\sqrt{\pi}} \sup_{0\leq\nu\leq1} |Z_1(\nu) - \nu Z_1(1)|,$$

where $(Z_1(\nu))_{\nu\geq0}$ denotes the standard fractional Brownian motion with Hurst parameter H if the observations are strictly monotone functions of a stationary Gaussian process $(\xi_t)_{t\geq1}$ with mean zero, variance 1 and autocovariance function as in (2.1). Upper quantiles of this asymptotic distribution can be used as critical values for the test. For details and critical values, see [19].

Since the scaling function d_N^2 in (2.16) depends on the Hurst parameter H, which is not known in practice, one has to use a suitable estimator of H, such as the overlapping blocks estimator $\hat{H}_{o,n}$, to apply the test. The modified test statistic is then defined as follows:

$$\widetilde{W}_N = \sqrt{\frac{N^{2-r(2-2H)}}{N^{2-r(2-2\hat{H}_{o,n})}}} \max_{1\leq k\leq N} |W_{k,N}| = N^{r(H-\hat{H}_{o,n})} \max_{1\leq k\leq N} |W_{k,N}| .$$

Taking the logarithm of $\widetilde{W}_N$ yields

$$\log(\widetilde{W}_N) = r(H - \hat{H}_{o,n}) \cdot \log(N) + \log(\max_{1\leq k\leq n} |W_{k,N}|). \tag{2.17}$$

The term $(H-\hat{H}_{o,n})\cdot\log(N)$ converges in probability to zero as long as the generic estimator used in every block has a convergence rate faster than $\log(N)$ and the assumption (2.9) on the number of level shifts is satisfied. This is e.g. satisfied by the local Whittle estimator, see [68, p. 223]. To be more precise, we have (with (2.10) and (2.11) - (2.15)) that

$$E\Big((H - \hat{H}_{o,n}) \cdot \log(N)\Big) = O(1)E(\check{H}_{n,1} - H)\log(N) + \frac{Kn}{N-n+1}\log(N)O\left(1\right) \to 0$$

and

$$Var((H - \hat{H}_{o,n}) \cdot \log(N)) = \left(O(Var(\check{H}_{n,1})) + O\left(\frac{K}{m}\right)\right) \cdot \log^2(N) \to 0$$

yielding

$$E\left(\left|(H - \hat{H}_{o,n}) \cdot \log(N)\right|^2\right) \to 0.$$

Convergence in second mean implies convergence in probability.

Combining the Continuous Mapping Theorem with Slutsky's theorem it is obvious from (2.17) that $\widetilde{W}_N$ has the same limit distribution as W_N.

2.6 Simulations

In this section we analyse the performance of the adaption techniques described in Section 2.3. Moreover we compare these estimators with the techniques proposed by [56] and [44], see Section 2.3.5.

2.6.1 Simulation scenarios

As generic (not adapted) estimation methods for the Hurst parameter H we choose the local Whittle estimator and the GPH estimator (see Section 2.3.1) and compare the results with the trimmed counterparts. We simulate fGn time series $Y_1, \ldots, Y_N$ with $H = 0.5$, 0.7, 0.9 of length $N = 500$, 1000, 1500 by generating $\xi_1, \ldots, \xi_N$ from fGn with the corresponding Hurst parameter H, using the `fArma` package in `R`, and choosing G as the identity $G(t) = t$ or the quantile transform

$$G(t) = (3/4)^{-(1/2)} \left((\Phi(t))^{-1/3} - 3/2 \right),$$

where Φ is the distribution function of the standard normal distribution. The resulting random variables follow the Pareto(3,1)-distribution and are therefore heavy-tailed.

We add a jump of height $h = 0.5, 1, 2$ after $0.1 \cdot N$ or after $0.5 \cdot N$ observations (i.e., after a proportion λ with $\lambda \in \{0.1, 0.5\}$) to the generated sequence of observations. We also look at time series in the jump-free scenario. The adapted estimators presented in Section 2.3 are applied to the generated data. For comparison, we also apply the two estimation techniques $\hat{H}^{\text{tr}}_{\text{GPH}}$ and $\hat{H}^{\text{tr}}_{\text{LW}}$ proposed by [56] and [44], described in Section 2.3.5, to the generated time series.

The described procedure is repeated 1000 times for each scenario, leading to 1000 estimates $\tilde{H}_1 \ldots, \tilde{H}_{1000}$ for each combination of the generic estimation method, adaption technique $\tilde{H}$ from Section 2.3 and scenario. We use the root mean-squared error (RMSE) in order to compare the quality of the estimates. The RMSE is widely used as evaluation criterion also for the memory parameter (see e.g. the simulation studies of [56] and [44]) as it takes bias and variance of the estimators into account, while being on the same scale as the parameter to be estimated.

For the case of multiple breaks we add three jumps after proportions $\lambda = 0.1, 0.5, 0.75$. Moreover, we simulate from a fractionally integrated ARMA model, i.e., ARFIMA(p, d, q) model

$$(1 - \varphi B)(1 - B)^d x_t = (1 - \theta B)\varepsilon_t$$

with $p = 0$ (therefore, $\phi = 0$), $q = 1$ (i.e., ARFIMA(0, d, 1)) and $\theta = -0.6$, using the `arfima` package in `R` and following the simulation study of [56], where $d = H - 0.5$ is the differencing parameter. Finally, we consider the short range dependent series from the ARMA model. Since fractional Gaussian noise (fGn) and ARFIMA(0, d, 0) are equivalent models (see e.g. [28]) the latter model is included implicitly in this simulation study.

2.6.2 Simulation results

We will discuss in detail the results for $H = 0.7$ in the following and comment on the results for $H = 0.5$ and $H = 0.9$ thereafter.

Simulation results for $H = 0.7$

Tables 2.1 (fGn) and 2.2 (Pareto) report the RMSE when applying our adaption techniques with the local Whittle estimator. We summarize the results in the following:

- For fGn time series
  - if there is no jump, all estimation procedures yield similar results.
  - the overlapping blocks technique $\hat{H}_{o,n}$ with block length $n = \bar{w} = 2w+1 = 2\sqrt{N}+1$ performs better than the other estimators when dealing with a small number of observations and low jumps.
  - the overlapping blocks estimator $\hat{H}_{o,w}$ with window length $n = w = \sqrt{N}$ yields the best results in all other cases, followed by the non-overlapping blocks estimator $\hat{H}_{no,w}$. The performance of the blocks estimators depends on the sample size rather than the jump height.
- For the Pareto(3,1)-transformed fGn
  - with the estimation on two blocks $\widehat{H}_{\mathrm{me}}$ we obtain the best results in the case of low jumps in the mean.
  - the overlapping blocks estimator $\hat{H}_{o,n}$ dominates if the jump in the mean is high and in the jump-free scenario.
  - most of the estimators yield better results when dealing with high jumps rather than low jumps. Figure D.1 in Appendix D suggests that the Hurst parameter is being underestimated under the heavy-tailed distribution, which is compensated by the overestimation, originating from a high jump in the mean.

Simulation results for $H = 0.5$ and $H = 0.9$

For the case of $H = 0.5$ and $H = 0.9$ we report the results in Figures 2.1 and 2.2. Here we restrict the graphical illustration of the boxplots to the case of fGn and $N = 1000$. The results are similar to the case $H = 0.7$. The overlapping blocks estimator $\hat{H}_{o,n}$ and the non-overlapping blocks approach $\hat{H}_{no,n}$ exhibit small variability and yield rather small bias in the case of short memory, i.e., $H = 0.5$. The estimators $\hat{H}_{\mathrm{pre},1}$ and $\hat{H}_{\mathrm{pre},2}$ perform well for jumps in the middle. The trimmed version of the local Whittle estimator $\hat{H}_{\mathrm{LW,tr}}$ overestimates the parameter if the jump is high. Both $\hat{H}_{\mathrm{LW,tr}}$ and $\hat{H}_{\mathrm{GPH,tr}}$ exhibit rather high variability in comparison to the blocks estimators. For $H = 0.9$, i.e., very strong dependence, again the blocks and the pre-estimation approaches yield good results and outperform the trimmed versions of the local Whittle and the GPH estimators.

In the case of the GPH estimator as a generic estimation technique we present the boxplots of the estimates for $H = 0.7$ in Figure 2.3. We can observe that the pre-estimation techniques yield the least bias in most cases, while having rather high variability. When dealing with early and high jumps the overlapping blocks estimator $\hat{H}_{o,n}$ exhibits the least bias and variability. In the case of $H = 0.5$ and $H = 0.9$ the corresponding figures show very similar results and are omitted in this thesis.

h	$\hat{H}_{\text{LW}}$	$\hat{H}_{\text{me}}$	$\hat{H}_{\text{diff}}$	$\hat{H}_{o,w}$	$\hat{H}_{o,n}$	$\hat{H}_{no,w}$	$\hat{H}_{no,n}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}^{\text{tr}}_{\text{LW}}$	$\hat{H}^{\text{tr}}_{\text{GPH}}$
					$N = 500$						
0	5.52	5.13	8.10	4.88	**4.79**	5.60	5.39	6.20	6.14	8.46	10.93
$\lambda = 0.1$											
0.5	5.92	5.54	9.12	4.82	**4.77**	5.53	5.42	5.81	6.16	8.48	10.94
1	8.44	7.31	11.43	**4.69**	4.85	5.39	5.47	5.79	7.15	8.65	11.58
2	17.20	12.23	19.22	**4.47**	5.18	5.14	5.66	8.72	10.68	9.92	14.10
$\lambda = 0.5$											
0.5	7.05	5.44	8.33	484	**4.80**	5.56	5.41	5.97	5.52	8.53	10.92
1	12.69	7.08	10.11	**4.72**	4.87	5.44	5.58	5.68	5.31	8.79	11.31
2	23.94	11.81	13.67	**4.53**	5.13	5.23	5.93	5.64	5.38	10.33	12.99
					$N = 1000$						
0	4.19	3.94	6.48	**3.21**	3.56	3.63	3.85	4.46	4.42	6.25	7.16
$\lambda = 0.1$											
0.5	4.73	4.46	6.97	**3.20**	3.60	3.61	3.92	4.26	4.76	6.33	7.29
1	7.59	6.47	11.07	**3.18**	3.74	3.59	4.11	4.55	5.78	6.59	7.69
2	16.09	11.49	19.38	**3.17**	4.09	3.59	4.50	7.74	9.67	8.06	9.47
$\lambda = 0.5$											
0.5	6.01	4.39	6.59	**3.21**	3.61	3.63	3.88	4.37	4.06	6.35	7.29
1	11.59	6.27	8.32	**3.19**	3.74	3.61	4.01	4.28	4.03	6.61	7.61
2	21.86	10.87	11.79	**3.18**	4.08	3.60	4.38	4.24	4.07	8.06	9.22
					$N = 1500$						
0	3.41	3.26	5.59	**2.59**	3.29	3.00	3.50	3.57	3.58	5.38	5.69
$\lambda = 0.1$											
0.5	4.09	3.84	6.01	**2.58**	3.35	2.99	3.52	3.43	3.92	5.45	5.80
1	7.16	5.97	11.14	**2.59**	3.53	3.00	3.58	3.95	5.32	5.62	6.16
2	15.42	10.92	20.26	**2.64**	3.89	3.03	3.81	6.91	8.92	6.60	7.57
$\lambda = 0.5$											
0.5	5.48	3.84	5.91	**2.59**	3.37	3.00	3.63	3.45	3.25	5.44	5.78
1	11.01	5.81	7.49	**2.61**	3.54	3.02	3.86	3.40	3.28	5.62	6.10
2	20.74	10.27	10.61	**2.65**	3.89	3.08	4.23	3.40	3.32	6.60	7.45

Table 2.1: Estimated RMSE$\cdot 10^2$ of different estimators for the Hurst parameter $H = 0.7$ in time series without ($h = 0$) and with change-point (jump of height h after a proportion of λ of the data), each based on 1000 simulation runs with $N = 500$, $N = 1000$ and $N = 1500$ values from fGn, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

h	$\hat{H}_{\text{LW}}$	$\hat{H}_{\text{me}}$	$\hat{H}_{\text{diff}}$	$\hat{H}_{o,w}$	$\hat{H}_{o,n}$	$\hat{H}_{no,w}$	$\hat{H}_{no,n}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}^{\text{tr}}_{\text{LW}}$	$\hat{H}^{\text{tr}}_{\text{GPH}}$
					$N = 500$						
0	8.18	7.60	10.63	8.63	**7.50**	9.05	7.92	10.45	9.24	10.86	14.76
$\lambda = 0.1$											
0.5	6.61	**6.00**	12.44	8.28	6.92	8.68	7.65	9.04	8.66	10.54	14.74
1	8.31	**6.09**	15.48	7.90	6.17	8.22	7.06	7.98	8.94	9.92	14.58
2	18.88	10.00	20.73	7.55	**5.44**	7.81	6.17	10.74	10.81	9.92	15.51
$\lambda = 0.5$											
0.5	6.50	**5.60**	8.75	8.31	6.94	8.76	7.15	9.34	7.94	10.57	14.71
1	13.39	**5.27**	8.11	7.95	6.22	8.34	6.29	9.20	7.83	10.00	14.46
2	25.01	9.25	11.34	7.61	**5.51**	7.92	5.64	9.14	7.81	10.27	14.91
					$N = 1000$						
0	7.21	6.72	9.13	7.12	**6.42**	7.32	6.73	8.73	7.83	9.08	11.39
$\lambda = 0.1$											
0.5	5.31	**4.98**	10.24	6.88	6.00	7.10	6.15	7.55	7.03	8.81	11.25
1	6.70	**4.83**	15.81	6.61	5.43	6.81	5.47	6.69	7.04	8.16	10.94
2	16.74	9.20	21.59	6.34	**4.83**	6.49	4.94	8.89	9.59	7.47	10.78
$\lambda = 0.5$											
0.5	5.08	**4.73**	6.78	6.88	6.01	7.10	6.30	8.11	7.07	8.80	11.24
1	11.36	**4.17**	5.79	6.59	5.45	6.80	5.71	8.01	7.03	8.14	11.06
2	22.71	8.33	9.32	6.33	**4.88**	6.47	5.10	7.95	6.99	7.41	10.81
					$N = 1500$						
0	6.81	6.43	8.01	6.55	**5.64**	6.68	5.79	8.14	7.35	8.47	10.12
$\lambda = 0.1$											
0.5	4.67	**4.48**	9.64	6.36	5.31	6.58	5.66	6.92	6.50	8.29	10.11
1	5.82	**3.98**	15.93	6.13	4.85	6.39	5.38	6.03	6.39	7.77	9.88
2	15.48	8.39	22.24	5.91	**4.36**	6.11	4.84	7.77	8.96	6.49	9.46
$\lambda = 0.5$											
0.5	**4.30**	4.34	5.29	6.36	5.32	6.43	5.36	7.48	6.65	8.26	10.09
1	10.29	**3.42**	5.23	6.14	4.86	6.14	4.83	7.42	6.61	7.71	9.90
2	21.17	7.46	8.04	5.92	**4.40**	5.92	4.41	7.39	6.57	6.45	9.42

Table 2.2: Estimated RMSE$\cdot 10^2$ of different estimators for the Hurst parameter $H = 0.7$ in time series without ($h = 0$) and with change-point (jump of height h after a proportion of λ of the data), each based on 1000 simulation runs with $N = 500$, $N = 1000$ and $N = 1500$ values from Pareto(3,1)-transformed fGn, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

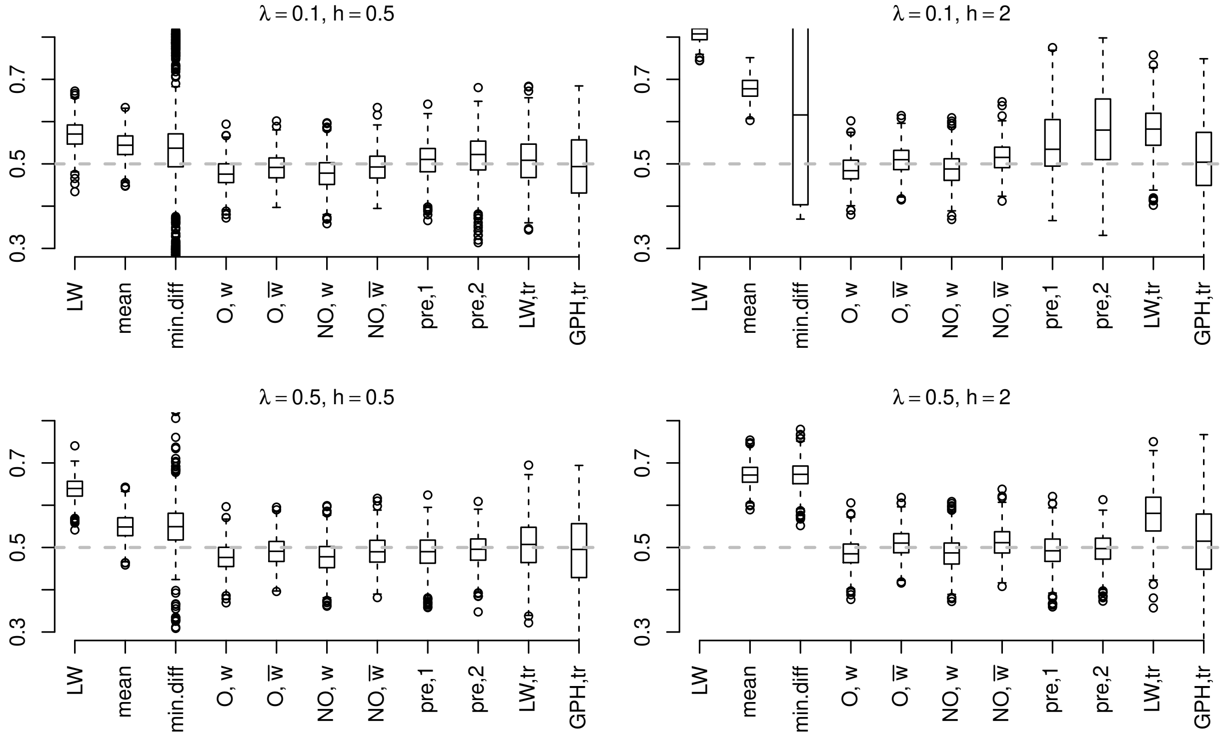

Figure 2.1: Boxplots of the estimators based on the local Whittle estimator for the Hurst parameter $H = 0.5$ in time series with one jump of height h after a proportion of λ of the data, based on each 1000 simulation runs with $N = 1000$ realizations of fGn.

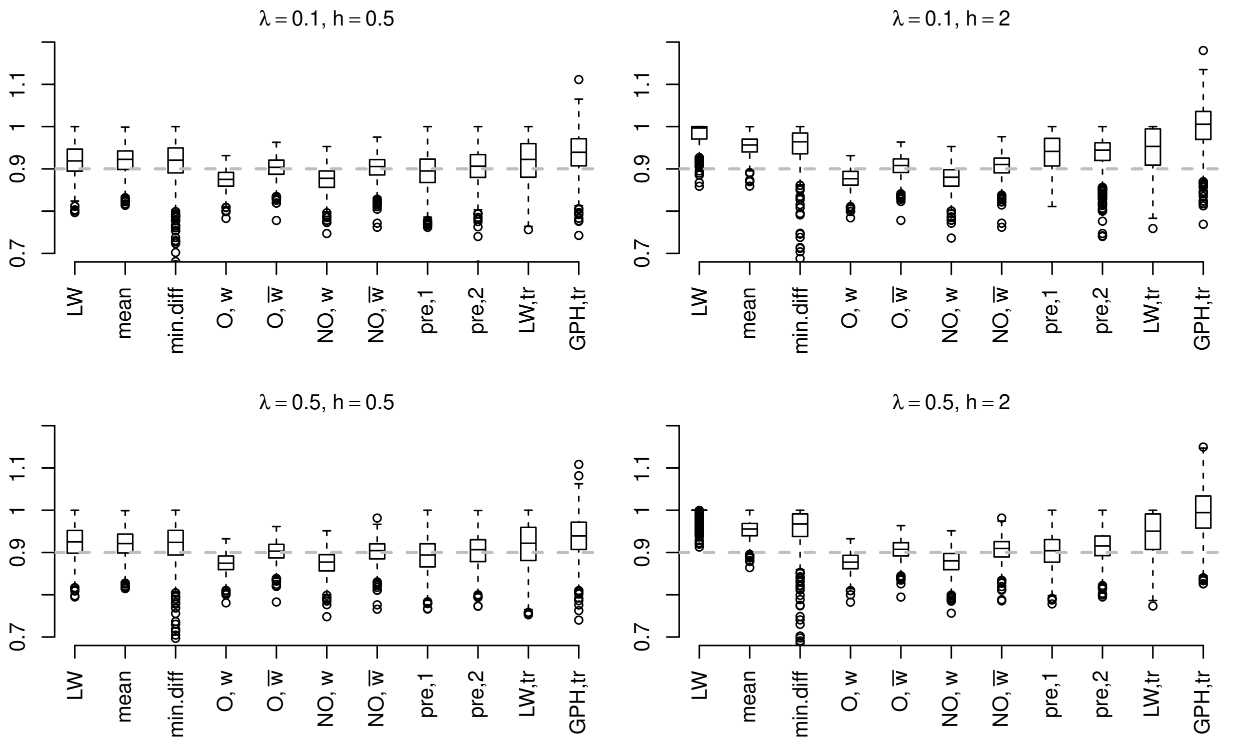

Figure 2.2: Boxplots of the estimators based on the local Whittle estimator for the Hurst parameter $H = 0.9$ in time series with one jump of height h after a proportion of λ of the data, based on each 1000 simulation runs with $N = 1000$ realizations of fGn.

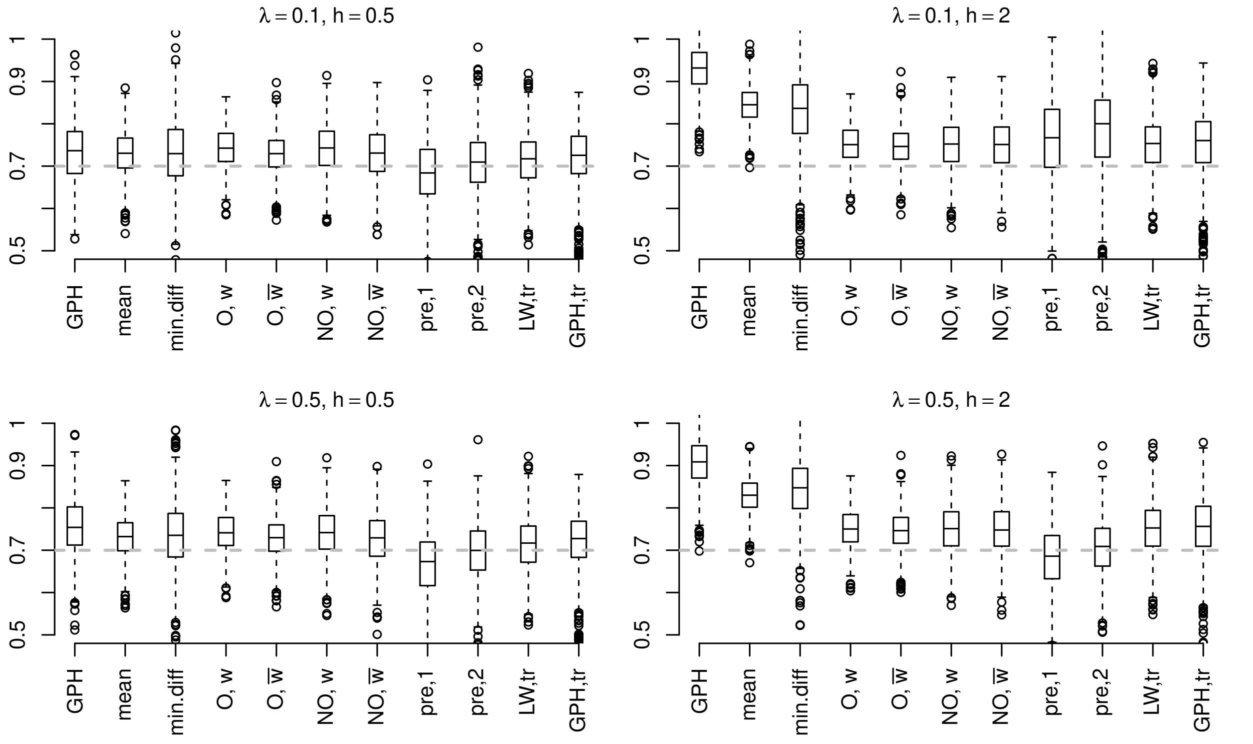

Figure 2.3: Boxplots of the estimators based on the GPH estimator for the Hurst parameter $H = 0.7$ in time series with one jump of height h after a proportion of λ of the data, based on each 1000 simulation runs with $N = 1000$ realizations of fGn.

2.6.3 Multiple breaks

It is also interesting to investigate how the proposed methods perform in the case of multiple breaks. For this purpose we generate time series $\xi_1, \dots, \xi_N$ from fGn with the Hurst parameter $H = 0.7$ choosing G as the identity $G(t) = t$. We add three jumps of equal height $h = 0.5, 1, 2$ after $0.1 \cdot N$, $0.5 \cdot N$ and $0.75 \cdot N$ observations to these time series. To the resulting sequences of observations we apply the adapted estimators. For comparison we also apply the two adapted estimation techniques $\hat{H}^{\text{tr}}_{\text{LW}}$ and $\hat{H}^{\text{tr}}_{\text{GPH}}$ proposed by [44] and [56] to the generated time series. We repeat these simulations 1000 times for each scenario, leading to 1000 estimates $\tilde{H}_1 \dots, \tilde{H}_{1000}$ for each estimation method and each scenario.

Table 2.3 reports the estimated RMSE for fGn time series. In the case of three change-points, the overlapping and the non-overlapping blocks estimators with window length $n = w = \sqrt{N}$ perform best. The overlapping blocks estimator $\hat{H}_{o,w}$ outperforms the other estimation methods. This can be explained by the fact that only a few blocks are contaminated with the jumps in the mean, so the biased estimates of the corresponding blocks do not influence the resulting estimator dramatically. The other methods suffer from a high number of change points and are highly biased then. The methods proposed by [44] and [56] are outperformed by our blocks techniques in every case. Taking into account the fast computation and the good performance of the overlapping blocks estimator it can be recommended in our survey.

h	$\hat{H}_{\text{LW}}$	$\hat{H}_{\text{me}}$	$\hat{H}_{\text{diff}}$	$\hat{H}_{o,w}$	$\hat{H}_{o,n}$	$\hat{H}_{no,w}$	$\hat{H}_{no,n}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}^{\text{tr}}_{\text{LW}}$	$\hat{H}^{\text{tr}}_{\text{GPH}}$
						$N = 500$					
0.5	12.77	7.57	10.22	**4.75**	4.84	5.47	5.48	5.73	5.95	8.94	11.62
1	24.63	14.84	16.86	**4.44**	5.47	5.20	6.17	12.01	12.08	11.47	15.90
2	29.99	24.63	21.84	**4.20**	7.55	5.01	8.38	25.47	24.44	21.76	32.58
						$N = 1000$					
0.5	11.77	6.82	8.54	**3.19**	3.74	3.60	4.10	4.85	5.03	6.81	7.89
1	22.72	13.86	14.99	**3.19**	4.38	3.61	4.90	11.49	11.17	9.22	10.94
2	29.99	24.16	23.38	**3.43**	6.00	3.93	6.78	23.75	22.59	18.93	23.91

Table 2.3: Estimated RMSE$\cdot 10^2$ of different estimators for the Hurst parameter $H = 0.7$ in time series with three change-points (jumps of height h after a proportion of $\lambda = 0.25$, 0.5 and 0.75 of the data), each based on 1000 simulation runs with $N = 500$ and $N = 1000$ values from fGn, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

2.6.4 ARFIMA model

As already mentioned in [56] and [66], the estimators of the differencing parameter $d = H - 1/2$ tend to be positively biased in the ARFIMA(p, d, q) model, i.e., when short range dependence components are present in the model. We simulate ARFIMA(0, d, 1) time series with $\theta = -0.6$ for $d = 0.2$ ($H = 0.7$) and $d = 0$ ($H = 0.5$) of length $N = 1000$. Again, jumps of various heights are added at the beginning or in the middle of the data in order to compare the performance of the estimators in the absence and presence of level shifts. In Table 2.4 we can observe that the blocks procedures fail to estimate the parameter correctly when the block size is small. The best results are obtained by $\hat{H}_{\text{pre},1}$. In the case of short memory, i.e., ARFIMA(0, 0, 1), again, the pre-estimating procedure $\hat{H}_{\text{pre},1}$ performs best, since it involves all data points in the estimation procedure.

The overlapping and the non-overlapping blocks approaches from Subsection 2.3.4 perform similarly and do not yield desirable results for the ARFIMA model when a small block size is chosen. In Tables E.1 and E.2 in Appendix E we show the results corresponding to further block sizes of the non-overlapping blocks estimator (the results for the overlapping blocks approach are similar throughout the simulation study, though the 1000 simulation runs require more computation time). We observe that its performance can be improved by choosing a larger block size n. Moreover, in the fGn-scenario, selecting a higher value for n does not worsen the performance of $\hat{H}_{\text{O},n}$ considerably compared to the case of the block length $\sqrt{N}$, see Table E.2. Still, the results are not satisfying when dealing with both, long and short memory.

h	$\hat{H}_{\text{LW}}$	$\hat{H}_{\text{me}}$	$\hat{H}_{\text{diff}}$	$\hat{H}_{o,w}$	$\hat{H}_{o,n}$	$\hat{H}_{no,w}$	$\hat{H}_{no,n}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}^{\text{tr}}_{\text{LW}}$	$\hat{H}^{\text{tr}}_{\text{GPH}}$
					ARFIMA(0, 0.2, 1)						
0	6.38	8.56	9.15	21.50	17.68	21.71	17.82	**5.37**	7.43	10.29	12.74
$\lambda = 0.1$											
0.5	6.98	8.97	9.30	21.52	17.61	21.65	17.79	**5.52**	7.99	10.16	12.79
1	8.24	10.25	10.32	21.51	17.73	21.50	18.03	**5.95**	9.42	10.64	13.47
2	13.19	13.51	13.93	21.62	18.07	21.79	18.34	**8.09**	12.10	11.33	15.70
$\lambda = 0.5$											
0.5	7.67	8.70	9.60	21.57	17.70	21.61	17.64	**5.19**	7.51	10.25	12.97
1	10.56	10.16	11.70	21.62	17.79	21.50	17.87	**5.58**	7.30	10.53	13.66
2	17.53	13.26	15.54	21.66	18.20	21.76	18.21	**5.70**	7.62	11.40	15.30
					ARFIMA(0, 0, 1)						
0	5.60	7.43	8.74	26.29	17.17	26.61	17.33	**4.86**	7.28	9.14	14.69
$\lambda = 0.1$											
0.5	8.18	9.25	9.93	26.50	17.45	26.44	17.45	**5.35**	8.78	8.99	14.69
1	13.61	12.68	14.49	26.46	17.77	26.62	18.02	**7.03**	11.24	9.61	15.11
2	24.31	18.72	22.64	26.73	18.67	26.91	19.03	**10.18**	15.40	11.65	16.08
$\lambda = 0.5$											
0.5	12.01	9.65	11.81	26.33	17.38	26.46	17.38	**5.00**	6.49	9.60	15.01
1	20.38	12.98	15.40	26.54	17.68	26.67	17.85	**5.38**	6.58	9.68	15.05
2	32.34	18.56	19.16	26.64	18.64	26.84	18.83	**5.28**	6.75	11.74	16.35

Table 2.4: Estimated RMSE$\cdot 10^2$ of different estimators for the differencing parameters $d = 0.2$ ($H = 0.7$) and $d = 0$ ($H = 0.5$) in time series without ($h = 0$) and with change-point (jump of height h after a proportion of λ of the data), each based on 1000 simulation runs with $N = 1000$ values from the ARFIMA(0, d, 1) model with $\theta = -0.6$, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

Therefore, we followed the ARMA correction procedure proposed by [66], where the differencing parameter d in the ARFIMA(p, d, q) model is estimated recursively in several steps:

1. Obtain an initial estimate $\hat{d}$ of d in the ARFIMA(p, d, q) model.

2. Remove the LRD structure: calculate $\hat{Y}_t = (1 - B)^{\hat{d}} Y_t$.

3. Identify the order and estimate the parameters in the ARMA(p, q) model $\Phi(B)\hat{Y}_t = \Theta(B)\varepsilon_t$.

4. Remove the ARMA structure: calculate $\hat{Z}_t = \frac{\widehat{\Phi}(B)}{\widehat{\Theta}(B)} Y_t$.

5. Obtain new estimate $\hat{d}$ of d in the ARFIMA(0, d, 0) model $(1 - B)^d \hat{Z}_t = \varepsilon_t$.

6. Repeat steps 2 to 5, until the difference between two consecutive estimates of d is less than 0.01 or the number of steps exceeds 10.

The RMSE values obtained using the above procedure are given in Tables 2.5 (for the overlapping approach) and 2.6 (for the non-overlapping approach). We observe that the results of the non-overlapping blocks approach (Table 2.6) improve considerably when following the ARMA correction procedure (see Table 2.4 for comparison). For $d = 0$ ($H = 0.5$), i.e., short memory, higher window values are preferred, while for LRD with $d = 0.2$ ($H = 0.7$) the block length $\sqrt{N}$ yields the lowest RMSE. Moreover, the results are nearly the same when dealing with fGn, i.e., when no short memory components are involved. In the case of the overlapping blocks approach we show the results for $N = 1000$, $\lambda = 0.5$ and $h = 2$ in Table 2.5. We can observe that even lower RMSE values can be achieved when choosing overlapping instead of non-overlapping blocks, hence it can be recommended in our survey. To be on the safe side, we suggest choosing a block length n equal to a multiple of $\sqrt{N}$, preferably $n > 100$ in order to obtain desirable results.

h	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$
	$H = 0.5$ ($d = 0$)					$H = 0.7$ ($d = 0.2$)				
2	11.81	**6.51**	7.06	8.54	10.46	**3.69**	4.79	5.43	6.61	7.22

Table 2.5: Estimated RMSE$\cdot 10^2$ of the overlapping blocks estimator based on the local Whittle estimator with different block sizes for the differencing parameters $d = 0$ ($H = 0.5$) and $d = 0.2$ ($H = 0.7$) in time series with change-point (jump of height h after a proportion of $\lambda = 0.5$ of the data), each based on 1000 simulation runs with $N = 1000$ values from the ARFIMA(0, d, 1) model with $\theta = -0.6$. The ARMA-structure is removed in several steps.

h	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$
	$H=0.5$ ($d=0$)					$H=0.7$ ($d=0.2$)				
					ARFIMA(0, d, 1)					
					$N=1000$					
0	12.06	5.95	**5.52**	6.33	5.57	5.54	**5.52**	5.69	6.64	6.33
$\lambda=0.1$										
0.5	11.83	6.39	**5.69**	6.64	6.40	**5.28**	5.34	5.80	6.62	6.71
1	11.96	7.08	**6.29**	7.76	8.71	**5.27**	5.66	5.99	6.92	7.08
2	11.90	8.06	**6.91**	10.29	12.71	**4.88**	5.97	5.86	7.67	8.69
$\lambda=0.5$										
0.5	11.67	6.07	**5.90**	6.46	6.13	**5.13**	5.31	6.03	6.77	6.73
1	11.91	6.68	**6.22**	7.01	7.83	**5.12**	5.53	5.87	6.77	6.89
2	11.90	**7.17**	7.59	7.58	10.97	**5.26**	5.63	6.56	7.33	8.54
					$N=1500$					
0	7.45	5.16	5.12	5.09	**4.76**	**4.12**	4.85	5.22	5.38	5.18
$\lambda=0.1$										
0.5	7.51	5.36	5.58	5.50	**5.19**	**4.13**	5.14	5.45	5.79	5.29
1	7.83	5.85	6.43	**5.71**	7.06	**4.05**	5.05	5.62	5.88	5.66
2	7.81	6.36	7.69	**6.24**	10.33	**4.06**	5.33	6.16	5.81	7.26
$\lambda=0.5$										
0.5	7.71	5.70	5.63	5.50	**5.11**	**4.05**	5.02	5.46	5.74	5.33
1	7.89	**6.14**	6.36	6.23	6.21	**4.24**	5.06	5.87	5.92	5.46
2	8.12	**6.81**	7.50	7.76	9.03	**4.12**	5.59	6.28	6.59	6.79
					fGn					
					$N=1000$					
$\lambda=0.5$										
2	**4.17**	4.54	5.17	5.61	10.28	**3.98**	4.56	5.07	5.63	8.18
					$N=1500$					
$\lambda=0.5$										
2	**3.39**	3.96	4.71	5.54	8.80	**3.35**	4.15	4.55	4.93	6.08

Table 2.6: Estimated RMSE$\cdot 10^2$ of the non-overlapping blocks estimator based on the local Whittle estimator with different block sizes for the differencing parameters $d=0$ ($H=0.5$) and $d=0.2$ ($H=0.7$) in time series without ($h=0$) and with change-point (jump of height h after a proportion of λ of the data), each based on 1000 simulation runs with $N=1000$ and $N=1500$ values from the ARFIMA(0, d, 1) model with $\theta=-0.6$ and fGn. The ARMA-structure is removed in several steps.

2.7 Application

We illustrate our improved estimation techniques with the help of two real data sets. We use the famous *Nile River minima*, yearly minimal water level of the Nile River for the years from 773 to 1281; the data are available e.g. in the R-package `wavelets`[1]. Nile River data is a widely-used example for LRD time series, see e.g. Example 5.9 in [62]. Log-periodogram values of the Nile data were considered in [8, p. 21], where evidence of long range dependence was found. A broad analysis of the monthly river flows was carried out by [58], where the authors applied several techniques to estimate the LRD parameter taking the seasonality of the river data into account.

We also apply the estimation techniques to the seasonally adjusted *global temperature for the northern hemisphere* for the years 1854–1989; the data are available in the R-package `longmemo`[2]. The data consist of differences between average monthly temperature and the corresponding monthly average over the years 1950–1979. This dataset was discussed in [8, p. 29], where a linear trend rather than a possible structural change in the data was considered. This was supported by [20], and [46] assumed that the temperature growth is rather due to impact of human activity or climate change. Evidence of long range dependence was found by [8, p. 173] and [29]. A possible change point in the data was detected by [95].

2.7.1 Nile River minima

The data are displayed in Figure 2.4. In order to estimate the LRD parameter H from the data, we have applied the local Whittle estimator to it as well as the adaption methods investigated in this thesis. The results are given in Table 2.7. The blocks estimators with ARMA correction from Subsection 2.6.4 are denoted by $\hat{H}_{o,w}^{\text{ARMA}}$ and $\hat{H}_{no,w}^{\text{ARMA}}$. For analysing the data [62, p. 386] use wavelets and the maximum-likelihood method yielding 0.95 as an estimate for the Hurst parameter, which is similar to the value $\hat{H}_{\text{LW}} = 0.97$ obtained by the ordinary local Whittle estimator. As opposed to this, the overlapping blocks approach with ARMA correction and the window length $\sqrt{N}$ yields the smaller value $\hat{H}_{o,w}^{\text{ARMA}} = 0.78$. The two estimation techniques $\hat{H}_{\text{GPH}}^{\text{tr}}$ and $\hat{H}_{\text{LW}}^{\text{tr}}$ described in Section 2.3.5 yield the higher values 1.00 and 1.01. The discrepancy between the estimates supports the possibility of a level shift.

$\hat{H}_{\text{LW}}$	$\hat{H}_{\text{me}}$	$\hat{H}_{\text{diff}}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}_{o,w}$	$\hat{H}_{no,w}$	$\hat{H}_{o,w}^{\text{ARMA}}$	$\hat{H}_{no,w}^{\text{ARMA}}$	$\hat{H}_{\text{LW}}^{\text{tr}}$	$\hat{H}_{\text{GPH}}^{\text{tr}}$
0.97	0.96	0.96	0.93	0.95	0.84	0.87	0.78	0.83	1.00	1.01

Table 2.7: Estimated LRD parameter H of the Nile River Minima data by different adaption methods, based on the local Whittle estimator.

[1] `http://artax.karlin.mff.cuni.cz/r-help/library/wavelets/html/nile.html`
[2] `http://cran.r-project.org/web/packages/longmemo/longmemo.pdf`

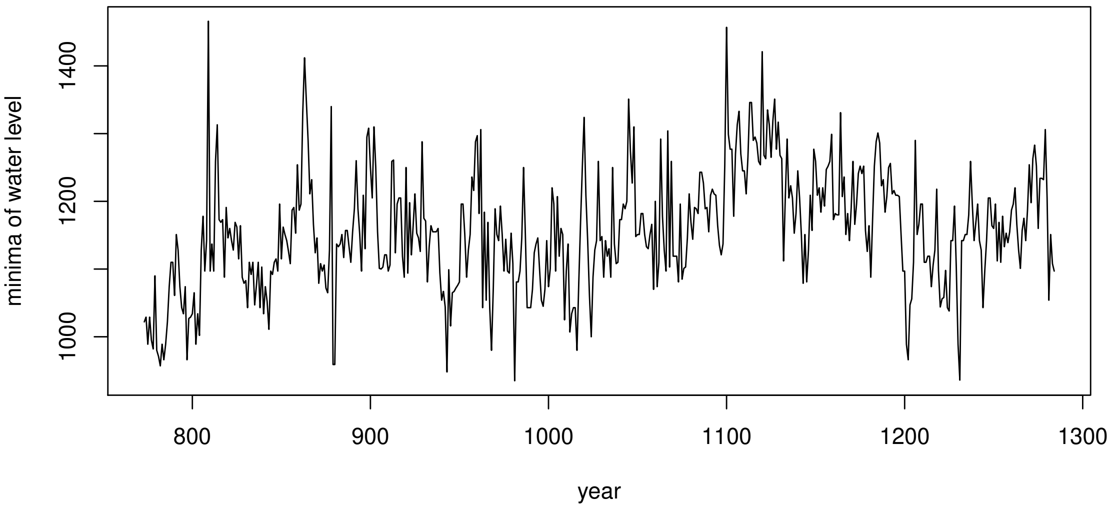

Figure 2.4: Annual minima of the water level of the Nile River near Cairo.

2.7.2 Global temperature for the northern hemisphere

Figure 2.5 shows the seasonally adjusted global temperature for the northern hemisphere. A presence of a level shift after the year 1923 was suggested by [95]. Moreover, we can observe a change in the scale after the year 1880. Since we are interested in changes in the mean only, we reduce the data to the years 1880-1989 yielding 1320 monthly observations.

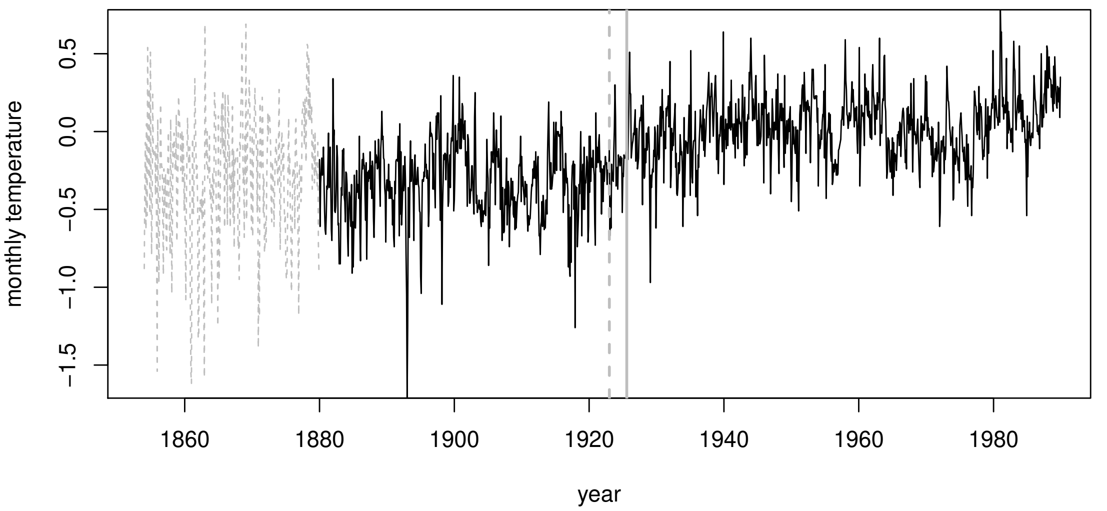

Figure 2.5: Monthly global temperature for the northern hemisphere for the years 1854-1989. The vertical lines denote a change-point, suggested by [95] (dashed) and detected using the Wilcoxon change-point test (bold). The grey part of the time series was excluded from the estimation procedure.

We have applied the local Whittle estimator together with the adaption methods to this data. The results are given in Table 2.8. The ordinary local Whittle estimator yields $\hat{H}_{\text{LW}} = 0.94$, while the other estimates, obtained by the adapted methods, are smaller. The overlapping blocks approach with ARMA correction and window length $\sqrt{N}$ yields $\hat{H}_{O,w}^{\text{ARMA}} = 0.84$, which is similar to the value obtained by $\hat{H}_{\text{pre},1} = 0.85$. It can be assumed that the pre-estimation technique was able to determine the location of a level shift adequately and estimated the Hurst parameter precisely in the two blocks with considerably large number of observations. Hence, a hypothesis of a change in the mean is supported by the results.

$\hat{H}_{\text{LW}}$	$\hat{H}_{\text{me}}$	$\hat{H}_{\text{diff}}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}_{o,w}$	$\hat{H}_{no,w}$	$\hat{H}_{o,w}^{\text{ARMA}}$	$\hat{H}_{no,w}^{\text{ARMA}}$	$\hat{H}_{\text{LW}}^{\text{tr}}$	$\hat{H}_{\text{GPH}}^{\text{tr}}$
0.94	0.91	0.90	0.85	0.87	0.85	0.86	0.84	0.86	0.91	0.93

Table 2.8: Estimated LRD parameter H of the global temperature for the northern hemisphere by different adaption methods, based on the local Whittle estimator.

2.8 Conclusion and outlook

When dealing with time series with long range dependence (LRD) it is crucial to estimate the LRD parameter, such as the Hurst parameter H, properly, e.g. for inference. Different estimation techniques, such as the Geweke and Porter-Hudak (GPH) estimator proposed by [28], can be found in the literature. However, most of these estimators may overestimate the LRD parameter under shifts in the mean. Under short range dependence this fact could lead to the false assumption of an LRD structure, see e.g. [85].

Many change-point tests, e.g. the Wilcoxon change-point test by [19], require knowledge of the LRD-parameter H for standardisation. Rejection of the hypothesis of a change in the mean could fail due to overestimation of H.

Therefore, we have investigated techniques based on the idea to segregate the data into blocks to adapt estimation procedures to time series with level shifts. We conducted a simulation study to asses the performance of these proposals for different tuning parameters and in different situations, using the local Whittle and the GPH estimators as generic estimation techniques. We conclude that our blockwise estimation techniques yield better results than ordinary estimators, such as the GPH estimator proposed by [28], in the change-point scenario.

We compared the performance of the adapted estimators with the trimmed versions of the local Whittle and the GPH estimators proposed by [44] and [56], respectively. We could observe that our proposals outperform the trimmed counterparts of the two generic estimation techniques used in this thesis in most simulation scenarios.

We recommend estimation of H by averaging estimates obtained from overlapping blocks with length approximately equal to a multiple of $\sqrt{N}$, where N is the sample size. This adaption inherits asymptotic MSE-consistency from its underlying estimation method when using the local Whittle or the GPH estimator, while offering fast computation. A minimum of 100 observations per block should be available for the estimation when dealing with ARFIMA models. In this case the estimation should be carried out using the ARMA correction procedure from Section 2.6.4.

Our approach opens up some ideas for future research:

- One could explore possible improvements by e.g. choosing other monotone increasing subsequences of the sample size N as window sizes for the overlapping or the non-overlapping blocks approaches or by considering a weighted average of estimates obtained from two blocks, taking the different block lengths into account.
- In the overlapping or the non-overlapping blocks approach, one could use a trimmed mean in order to trim the estimations on blocks, which are influenced by the jump.
- The derivation of the asymptotic distribution of the blockwise approaches might be of interest for conducting inference. However, it involves new theory on triangular arrays for long range dependent data. This is beyond the scope of this thesis.

3 Scale estimation under shifts in the mean

3.1 Introduction

In many applications the variance and the standard deviation of the observations need to be estimated properly, e.g. for standardization. In the presence of outliers and jumps in the mean suitable estimation procedures are required, since ordinary estimators, such as the sample variance or the sample standard deviation, are biased in this situation. We propose estimation techniques based on segregating the data $Y_1, \ldots, Y_N$ into m non-overlapping blocks of size $n = \lfloor N/m \rfloor$, since such techniques improve parameter estimation under shifts in the mean in the context of long range dependence, see Chapter 2.

If the data are contaminated by level shifts we use the sample variance in each block $j = 1, ..., m$. When only a few level shifts are expected to occur, we propose to combine the estimates by averaging them to obtain an estimate of the variance. The blocks-estimator performs well in change-point scenarios, as can be seen in a simulation study. Theoretical results, such as strong consistency and asymptotic normality, are shown under certain assumptions on the number of blocks and the number of jumps. Moreover, suggestions on the choice of the block size are given. In the case of many changes in the mean we consider an adaptively trimmed mean of the blockwise sample variances, where the trimming fraction is chosen adaptively, depending on the observed data.

If the data are contaminated by outliers a robust measure of scale, such as the median absolute deviation (MAD), is considered instead of the sample variance or the sample standard deviation. However, this estimator is strongly biased under shifts in the mean, since the data are centred by the median of all observations, which itself is a biased estimator of location in that case. In this thesis we propose a modified MAD as a scale estimator. The data are divided into m non-overlapping blocks. In each block the sample median is calculated. Subsequently, the data are centred by the blockwise medians before calculating the median of the absolute differences. In this way we expect that only a small fraction of the absolute differences is influenced by the jumps in the mean reducing the bias of the estimator. We show strong consistency and asymptotic normality of this estimator. Moreover, suggestions on the choice of the block size are given. We compare the performance of the proposed estimation procedure with that of other robust estimation techniques.

This chapter is organized as follows: In Section 3.2 we introduce the model of the data generating process. In Section 3.3 the blocks-estimator of the variance under shifts in the mean is presented. Section 3.4 deals with scale estimation in the change-point and outlier scenario.

3.2 Model

We consider a sequence of random variables $Y_1, \dots, Y_N$ generated by the model

$$Y_t = X_t + \sum_{k=1}^{K} h_k I_{t \geq t_k} + \gamma_t U_t, \tag{3.1}$$

where X_t are i.i.d. with continuous distribution function F, $E(X_t) = \mu$ and $Var\,(X_t) = \sigma^2$, $t = 1, ..., N$. Without loss of generality we will assume $\mu = 0$ in the following. In this thesis we consider the case of K level shifts of potentially different heights h_k at different change locations $t_1, \dots, t_K$. Moreover, we allow for the presence of outliers by including the term $\gamma_t U_t$, $t = 1, \dots, N$, e.g. with $\gamma_t \sim \mathrm{N}(\gamma, \rho^2)$, the average absolute outlier magnitude $\gamma \geq 0$, $\rho \geq 0$, and $U_t \in \{-1, 0, 1\}$ with respective probabilities $p/2$, $1 - p$ and $p/2$ with $0 \leq p \leq 1$.

In the following Sections 3.3 and 3.4 we will discuss approaches based on segregating the data $Y_1, ..., Y_N$ into m blocks of size n.

Remark 4.
1. The number of blocks m and the block size n may both depend on the sample size N, i.e., $m = m(N)$ and $n = n(N)$.
2. The heights $h_1, ..., h_K$ of the jumps are assumed to be positive, which is worse than both, positive and negative jumps, resulting in a higher bias for most scale estimators.
3. If N is not divisible by m, we choose $n = \lfloor N/m \rfloor$ and disregard the remaining observations.

3.3 Variance estimation under shifts in the mean

This section is based on the article "On variance estimation under shifts in the mean" [4], under review in *AStA – Advances in Statistical Analysis*, published as SFB Discussion Paper.

In Section 3.3.1 we analyse estimators of σ^2 from the sequence of observations $Y_1, ..., Y_N$ by averaging estimates obtained from splitting the data into several blocks. Without the need of explicit distributional assumptions the mean of the blockwise estimates turns out to be consistent if the size and the number of blocks increases, and the number of jumps increases slower than the number of blocks. If many jumps in the mean are expected to occur, an adaptively trimmed mean of the blockwise estimates can be used, see Section 3.3.2. Section 3.3.4 treats estimation of σ in a similar way. In Section 3.3.3 a simulation study is conducted to assess the performance of the proposed approaches. In Section 3.3.5 the estimation procedures are applied to real data sets, while Section 3.3.6 summarizes the results.

3.3.1 Estimation of the variance by averaging

The blocks-estimator $\widehat{\sigma^2}_{\text{Mean}}$ of the variance investigated here is defined as

$$\widehat{\sigma^2}_{\text{Mean}} = \frac{1}{m}\sum_{j=1}^{m} S_j^2, \tag{3.2}$$

where $S_j^2 = \frac{1}{n-1}\sum_{t=1}^{n}(Y_{j,t} - \overline{Y}_j)^2$, $\overline{Y}_j = \frac{1}{n}\sum_{t=1}^{n} Y_{j,t}$ and $Y_{j,1}, \ldots, Y_{j,n}$ are the observations in the j-th block. We are interested in finding the block size n, which yields a low mean squared error (MSE) under certain assumptions.

We will use some algebraic rules for derivation of the expectation and the variance of quadratic forms in order to calculate the MSE of $\widehat{\sigma^2}_{\text{Mean}}$, see [78]. Let B be the number of blocks with jumps in the mean and $K \geq B$ the total number of jumps. The expected value and the variance of $\widehat{\sigma^2}_{\text{Mean}}$ are given as follows:

$$E\left(\widehat{\sigma^2}_{\text{Mean}}\right) = \sigma^2 + \frac{1}{m(n-1)}\sum_{j=1}^{B}\mu_j^T A \mu_j,$$

$$Var\left(\widehat{\sigma^2}_{\text{Mean}}\right) = \frac{1}{m}\left(\frac{\nu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}\right) + \frac{4\sigma^2}{m^2(n-1)^2}\sum_{j=1}^{B}\mu_j^T A \mu_j,$$

where $\nu_4 = E\left(X_1^4\right)$, $A = \mathbb{I}_n - \frac{1}{n} 1_n 1_n^T$, $\mathbb{I}_n$ is the unit matrix, $1_n = (1, \ldots, 1)^T$ and μ_j contains the expected values of the random variables in the perturbed block $j = 1, \ldots, B$, i.e., $\mu_j = (\mu_{j,1}, \ldots, \mu_{j,n})^T = (E(Y_{j,1}), \ldots, E(Y_{j,n}))^T$. The term $\mu_j^T A \mu_j/(n-1)$ is the empirical variance of the expected values $E(Y_{j,1}), \ldots, E(Y_{j,n})$ in block j. In a jump-free block we have $\mu_j^T A \mu_j = 0$, since all expected values and therefore the elements of μ_j are equal.

The blocks-estimator (3.2) estimates the variance consistently if the number of blocks grows sufficiently fast as is shown in Theorem 2.

Theorem 2. *Let $Y_1, \ldots, Y_N$ with $Y_t = X_t + \sum_{k=1}^{K} h_k I_{t \geq t_k}$ from Model (3.1) be segregated into m blocks of size n, where $t_1, ..., t_K$ are the time points of the jumps of size $h_1, ..., h_K$,* *respectively. Let B out of m blocks be contaminated by $\widetilde{K}_1, \ldots, \widetilde{K}_B$ jumps, respectively, with $\sum_{j=1}^{B} \widetilde{K}_j = K = K(N)$. Moreover, let $E(|X_1|^4) < \infty$, $K\left(\sum_{k=1}^{K} h_k\right)^2 = o(m)$ and $m \to \infty$, whereas the block size n can be fixed or increasing as $N \to \infty$. Then $\widehat{\sigma^2}_{Mean} = \frac{1}{m}\sum_{j=1}^{m} S_j^2 \to \sigma^2$ almost surely.*

Proof. Without loss of generality assume that the first B out of m blocks are contaminated by $\widetilde{K}_1, \ldots, \widetilde{K}_B$ jumps, respectively. Let the term $S_{j,0}^2$ denote the empirical variance of the uncontaminated data in block j, while $S_{j,h}^2$ is the empirical variance when $\widetilde{K}_j$ level shifts are present. Moreover, $Y_{j,1}, \ldots, Y_{j,n}$ are the observations in the j-th block, $\mu_{j,t} = E\left(Y_{j,t}\right)$ and $\overline{\mu}_j = \frac{1}{n}\sum_{t=1}^{n} E\left(Y_{j,t}\right)$. Then we have

$$\begin{aligned}
\widehat{\sigma^2}_{\text{Mean}} &= \frac{1}{m}\sum_{j=1}^{m} S_j^2 = \frac{1}{m}\sum_{j=B+1}^{m} S_{j,0}^2 + \frac{1}{m}\sum_{j=1}^{B} S_{j,h}^2 \\
&= \frac{1}{m}\sum_{j=B+1}^{m} S_{j,0}^2 + \frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(X_{j,t} + \mu_{j,t} - \overline{X}_j - \overline{\mu}_j\right)^2 \\
&= \frac{1}{m}\sum_{j=1}^{m} S_{j,0}^2 + \frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}(X_{j,t} - \overline{X}_j)(\mu_{j,t} - \overline{\mu}_j) \\
&+ \frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t} - \overline{\mu}_j\right)^2.
\end{aligned} \tag{3.3}$$

For the second term in the last equation (3.3) we have almost surely

$$\begin{aligned}
&\left|\frac{2}{m(n-1)}\sum_{j=1}^{B}\sum_{t=1}^{n}(X_{j,t} - \overline{X}_j)(\mu_{j,t} - \overline{\mu}_j)\right| \\
&\leq \frac{2}{m(n-1)}\sum_{j=1}^{B}\sum_{t=1}^{n}\left|(X_{j,t} - \overline{X}_j)\right|\left|(\mu_{j,t} - \overline{\mu}_j)\right| \\
&\leq \frac{2}{m(n-1)}\sum_{j=1}^{B}\sum_{t=1}^{n}\left|(X_{j,t} - \overline{X}_j)\right|\left|\sum_{k=1}^{K} h_k\right| \\
&= B\left|\sum_{k=1}^{K} h_k\right|\frac{2}{m}\frac{n}{n-1}\frac{1}{B}\sum_{j=1}^{B}\frac{1}{n}\sum_{t=1}^{n}\left|(X_{j,t} - \overline{X}_j)\right| \longrightarrow 0.
\end{aligned} \tag{3.4}$$

The term $\frac{1}{B}\sum_{j=1}^{B}\frac{1}{n}\sum_{t=1}^{n}\left|(X_{j,t} - \overline{X}_j)\right|$ in (3.4) is a random variable with finite moments if n and B are fixed. This random variable converges to $E\left(\frac{1}{n}\sum_{t=1}^{n}\left|(X_{j,t} - \overline{X}_j)\right|\right)$ almost surely if $B \to \infty$. In the case of $B \to \infty$ and $n \to \infty$ this term converges to $E\left(|X_1|\right)$ almost surely due to Theorem 2 of [41] and the condition $E(|X_1|^4) < \infty$, since $S_j^2 - E\left(S_j^2\right)$ are uniformly bounded with $P(|S_j^2 - E\left(S_j^2\right)| > t) \to 0\,\forall t$ due to Chebyshev's inequality and $Var(S_j^2) \to 0$. Moreover, we used the fact that $B\left|\sum_{k=1}^{K} h_k\right| \leq K\left|\sum_{k=1}^{K} h_k\right| = o(m)$.

The following is valid for the third term in (3.3):

$$\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t}-\overline{\mu}_j\right)^2 \leq \frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\sum_{k=1}^{K}h_k\right)^2$$
$$= \frac{B}{m}\frac{n}{n-1}\left(\sum_{k=1}^{K}h_k\right)^2 \longrightarrow 0.$$

The first term of the last equation in (3.3) converges almost surely to σ^2 due to the results on triangular arrays in Theorem 2 of [41], assuming that the condition $E(|X_1|^4) < \infty$ holds, since $S_j^2 - E\left(S_j^2\right)$ are uniformly bounded with $P(|S_j^2 - E\left(S_j^2\right)| > t) \to 0\,\forall t$ due to Chebyshev's inequality and $Var(S_j^2) \to 0$. Application of Slutsky's Theorem proves the result. □

Remark 5. 1. If the jump heights are bounded by a constant $h \geq h_k$, $k = 1,\ldots,K$, the strongest restriction arises if all heights equal this upper bound resulting in the constraint $K\left(\sum_{k=1}^{K}h_k\right)^2 = K^3h^2 = o(m)$. Consistency is thus guaranteed if the number of blocks grows faster than K^3.

2. By the Central Limit Theorem the estimator $\widehat{\sigma^2}_{\text{Mean}}$ is asymptotically normal if no level shifts are present and the block size n is fixed. Its asymptotic efficiency relative to the ordinary sample variance is

$$\frac{Var(S^2)}{Var\left(\widehat{\sigma^2}_{\text{Mean}}\right)} = \frac{\frac{\nu_4}{N}-\frac{\sigma^4(N-3)}{N(N-1)}}{\frac{1}{m}\left(\frac{\nu_4}{n}-\frac{\sigma^4(n-3)}{n(n-1)}\right)} = \frac{\nu_4-\frac{\sigma^4(N-3)}{(N-1)}}{\nu_4-\frac{\sigma^4(n-3)}{(n-1)}} \xrightarrow{N\to\infty} \frac{\nu_4-\sigma^4}{\nu_4-\frac{\sigma^4(n-3)}{(n-1)}}$$

in case of i.i.d data with finite fourth moments, where

$$Var\left(\widehat{\sigma^2}_{\text{Mean}}\right) = Var\left(\frac{1}{m}\sum_{j=1}^{m}S_j^2\right) = \frac{1}{m}Var\left(S_1^2\right) = \frac{1}{m}\left(\frac{\nu_4}{n}-\frac{\sigma^4(n-3)}{n(n-1)}\right)$$

(see [2] for the variance $Var(S_1^2)$ of the sample variance). E.g., under normality the efficiency of the blocks estimater with fixed n is $(n-1)n^{-1} < 1$.

3. The asymptotic efficiency of the blocks-estimator relative to the sample variance is 1 if $n \to \infty$.

The next Theorem shows that $\widehat{\sigma^2}_{\text{Mean}}$ is asymptotically not only normal but even fully efficient in case of a growing block size.

Theorem 3. *Assume that the i.i.d. random variables $Y_1,\ldots,Y_N$ are segregated into m blocks of size n, with $m, n \to \infty$ such that $m = o(n)$, $n = o(N)$. Moreover, assume that $\nu_4 = E\left(X_1^4\right) < \infty$. Then we have*

$$\sqrt{N}\left(\widehat{\sigma^2}_{Mean} - \sigma^2\right) \xrightarrow{d} N\left(0, \nu_4 - \sigma^4\right).$$

Proof. Rewriting the estimator $\widehat{\sigma^2}_{\text{Mean}}$ we get

$$\frac{\widehat{\sigma^2}_{\text{Mean}} - \sigma^2}{\sqrt{Var\left(\widehat{\sigma^2}_{\text{Mean}}\right)}} = \frac{\frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{t=1}^{n} \left(X_{j,t} - \overline{X}_j\right)^2 - \sigma^2}{\sqrt{\frac{1}{m}\left(\frac{\nu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}\right)}}$$

$$= \sqrt{N} \frac{\frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{t=1}^{n} X_{j,t}^2 - \frac{n}{m(n-1)} \sum_{j=1}^{m} \overline{X}_j^2 - \sigma^2}{\sqrt{\nu_4 - \frac{\sigma^4(n-3)}{n-1}}}. \tag{3.5}$$

For the second term of the numerator in (3.5) we have that

$$E\left(\left|\sqrt{N}\frac{n}{m(n-1)}\sum_{j=1}^{m}\overline{X}_j^2\right|\right) = \frac{\sqrt{N}}{m}\frac{n}{n-1}\sum_{j=1}^{m} E\left(\overline{X}_j^2\right) = \sqrt{N}\frac{n}{n-1}\frac{\sigma^2}{n}$$

$$= \sqrt{mn}\frac{n}{n-1}\frac{\sigma^2}{n} = \sqrt{\frac{m}{n}}\frac{n}{n-1}\sigma^2 \to 0, \tag{3.6}$$

since $m = o(n)$. Convergence of the term (3.6) in mean implies convergence in probability to zero. Application of the Central Limit Theorem to the remaining terms of (3.5) yields the desired result. □

Remark 6. In the proof of Theorem 3 we have assumed that $m = o(n)$, i.e., the block size grows faster than the number of blocks. This condition can be dropped using the Lyapunov condition under the assumption of finite eighth moments, as will be shown in the following. We set

$$T_{i,j} = \frac{S_{i,j}^2 - \sigma^2}{\sqrt{\frac{m_i}{n_i}\left(\nu_4 - \frac{\sigma^4(n_i-3)}{n_i-1}\right)}},$$

with $E(T_{i,j}) = 0$ and $\sum_{j=1}^{m_i} E(T_{i,j}^2) = 1\ \forall i$, where i denotes the i-th row of the triangular array. The Lyapunov condition (see corollary 1.9.3 in [79]) is the following:

$$\exists \delta > 0 : \lim_{i\to\infty} \sum_{j=1}^{m_i} E\left(|T_{i,j}|^{2+\delta}\right) = 0$$

With $\delta = 2$ and existing moments $\nu_1, ..., \nu_8$ of X_1 we get

$$\sum_{j=1}^{m_i} E\left(|T_{i,j}|^4\right) = \sum_{j=1}^{m_i} \frac{1}{\frac{m_i^2}{n_i^2}\left(\nu_4 - \frac{\sigma^4(n_i-3)}{n_i-1}\right)^2} \cdot E\left(\left(S_{i,j}^2 - \sigma^2\right)^4\right)$$

$$= m_i \frac{1}{\frac{m_i^2}{n_i^2}\left(\nu_4 - \frac{\sigma^4(n_i-3)}{n_i-1}\right)^2} \cdot O\left(\frac{1}{n_i^2}\right) \cdot g(\nu_1, ..., \nu_8) = o\left(\frac{1}{m_i}\right) \to 0,$$

where g is a function of the existing moments $\nu_1, ..., \nu_8$ with $g(\nu_1, ..., \nu_8) = O(1)$. See [2] for the fourth central moment of the sample variance. Therefore, the condition $m = o(n)$ can be dropped.

Choice of the block size

When choosing blocks of length $n = 2$, the estimator $\hat{\sigma}^2_{\text{Mean}}$ results in a difference-based estimator, which considers $\lfloor N/2 \rfloor$ consecutive non-overlapping differences:

$$\widehat{\sigma^2}_{\text{Mean},n=2} = \frac{1}{2\lfloor N/2 \rfloor} \sum_{j=1}^{\lfloor N/2 \rfloor} \left(Y_{2j} - Y_{2j-1}\right)^2 .$$

Difference-based estimators have been considered in many papers, see [94], [67], [26], [32], [21], [59], [91], among many others. [17] discussed estimation approaches based on differences in nonparametric regression context, [98] considered an estimation technique, which involves differences of second order, while [90] proposed a difference-based estimator for m-dependent data. An ordinary difference-based estimator of first order, which considers all $N-1$ consecutive differences, is (see e.g. [94]):

$$\widehat{\sigma^2}_{\text{Diff}} = \frac{1}{2(N-1)} \sum_{j=1}^{N-1} \left(Y_{j+1} - Y_j\right)^2 = \frac{1}{2(N-1)} Y^T A Y, \tag{3.7}$$

where $A = \tilde{A}^T \tilde{A}$ and $\tilde{A}$ such that $\tilde{A}Y = \left(Y_2 - Y_1, \ldots, Y_N - Y_{N-1}\right)^T$. Theorems 1.5 and 1.6 in [78] are used to calculate the expectation and the variance of the estimator $\widehat{\sigma^2}_{\text{Diff}}$:

$$\begin{aligned} E\left(\widehat{\sigma^2}_{\text{Diff}}\right) &= \sigma^2 + \frac{1}{2(N-1)} \mu^T A \mu, \\ Var\left(\widehat{\sigma^2}_{\text{Diff}}\right) &= \frac{1}{4(N-1)^2} \left(\nu_4(4N-6) + 2\sigma^4 + 4\sigma^2 \mu^T A^2 \mu + 4\nu_3 \mu^T A a\right), \end{aligned}$$

where $\mu = E(Y)$, $\nu_i = E\left(X_1^i\right)$ and a is a vector of the diagonal elements of A.

When no changes in the mean are present both, the difference-based and the averaging estimators, are unbiased. For the variance of the estimators we have

$$\begin{aligned} Var\left(\widehat{\sigma^2}_{\text{Diff}}\right) &= \nu_4 \frac{4N-6}{4(N-1)^2} + \frac{\sigma^4}{2(N-1)^2}, \\ Var\left(\widehat{\sigma^2}_{\text{Mean}}\right) &= \frac{\nu_4}{N} - \frac{\sigma^4(n-3)}{N(n-1)}. \end{aligned}$$

E.g. for $N = 100$ and a block size $n = 10$ we get $Var\left(\widehat{\sigma^2}_{\text{Mean}}\right) = 0.0222$, while $Var\left(\widehat{\sigma^2}_{\text{Diff}}\right) = 0.0302$. Therefore, when no changes in the mean are present the blocks estimator can have smaller variance than the difference-based estimator. In Section 3.3.3 we compare the performance of the proposed estimation procedures with that of the difference-based estimator (3.7) in different change-point scenarios.

In the following we investigate the proper choice of the block size for the estimator $\widehat{\sigma^2}_{\text{Mean}}$. All calculations in this thesis have been performed with the statistical software `R`, version 3.5.2, [64].

For known jump positions the MSE of the blocks-variance estimator $\widehat{\sigma^2}_{\text{Mean}}$ can be determined analytically. The position of the jump is relevant for the performance of this approach. Therefore, it is reasonable to consider different positions of the K jumps to get an overall assessment of the performance of the blocks-estimator. For every $K \in \{1,3,5\}$, we generate K jumps of equal heights $h = \delta \cdot \sigma$, with $\delta \in \{0, 0.1, 0.2, \ldots, 4.9, 5\}$, at positions sampled randomly from a uniform distribution on the values $\max_n (N - \lfloor N/n \rfloor n) + 1, \ldots, N - \max_n (N - \lfloor N/n \rfloor n)$ (if $N \neq mn$ then $N - mn$ observations are left out at the beginning and at the end of the sequence of observations) without replacement, and calculate the MSE for every reasonable block size $n \in \{2, 3, 4, \ldots, \lfloor N/2 \rfloor\}$. This is repeated 1000 times, leading to 1000 MSE values for every h and n based on different jump positions. The average of these MSE-values is taken for each h and n. Data are generated from the standard normal or the t_5-distribution.

Panel (a) of Fig. 3.1 shows the block size n_{opt}, which yields the least theoretical MSE value of the estimator $\widehat{\sigma^2}_{\text{Mean}}$ depending on the jump height $h = \delta \cdot \sigma$ with $K \in \{1,3,5\}$ jumps for $N = 1000$ observations and normal distribution. We observe that n_{opt} decreases for $\widehat{\sigma^2}_{\text{Mean}}$ as the jump height grows. Blocks of size 2 (resulting in a non-overlapping difference-based estimator) are preferred when $h \approx 4\sigma$ and $K = 5$, while larger blocks lead to better results in case of smaller or less jumps.

Panel (b) of Fig. 3.1 depicts the MSE of $\widehat{\sigma^2}_{\text{Mean}}$ for the respective MSE-optimal block size n_{opt}. In all three scenarios the MSE increases if the jump height gets larger.

Different values for the optimal block-size n_{opt} are obtained in different scenarios, i.e., for different K and $h = \delta \cdot \sigma$. As the true number and height of the jumps in the mean are usually not known in practice, we wish to choose a block size, which yields good results in many scenarios. We do not consider very high jumps any further, since they can be detected easily and are thus not very interesting. The square root of the sample size N has proven to be a good choice for the block size in many applications, see e.g. [72]. Moreover, in the upper panel of Figure 3.1 we observe that smaller block sizes n are preferred when the number of change-points is high. If the estimation of the variance is in the focus of the application, we suggest to choose the block size depending on K:

$$n = \max \left\{ \left\lfloor \frac{\sqrt{N}}{K+1} \right\rfloor, 2 \right\}, \tag{3.8}$$

which gets smaller if the number of jumps increases, resulting in many jump-free and only few contaminated blocks. Otherwise, if testing is of interest in view of Theorem 3 we suggest a block size, which grows slightly faster than $\sqrt{N}$, e.g. $n = \max\left\{\left\lfloor \frac{N^{6/10}}{K+1} \right\rfloor, 2\right\}$, which yields similar results as (3.8).

Remark 7. For large N we get $m = N/n = \sqrt{N}(K+1)$. In this case the number of jumps needs to satisfy $K = o\left(m^{1/3}\right) = o\left(N^{1/6}K^{1/3}\right)$, i.e., $K = o\left(N^{1/4}\right)$ can be tolerated, see Remark 5. An even larger rate of shifts can be tolerated by choosing $m = N/c$ for some constant c, i.e., a fixed block length n. However, this reduces the efficiency of the resulting estimator in the presence of a small rate of shifts. E.g., for the standard normal distribution the asymptotic relative efficiency of the blocks-estimator $\widehat{\sigma^2}_{\text{Mean}}$ is $(n-1)n^{-1} < 1$ and that of the difference-based estimator (as defined in (3.7)) is $2/3$ (see e.g. [78]).

Panel (c) of Fig. 3.1 shows the MSE of the estimator $\widehat{\sigma^2}_{\text{Mean}}$ with the block size n chosen according to (3.8). For $K \in \{3, 5\}$ there is only a moderate loss of performance when choosing n according to (3.8) instead of n_{opt}, which depends on the number of jumps K and the height $h = \delta \cdot \sigma$. In case of $K = 1$ change in the mean the performance of the blocks-estimator worsens slightly.

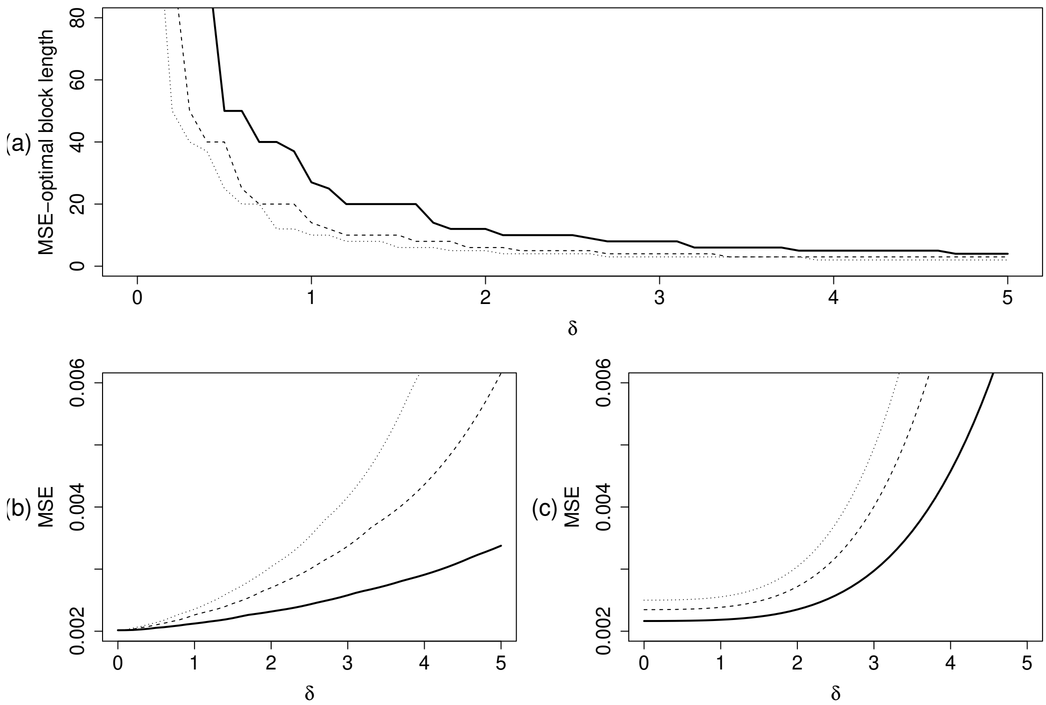

Figure 3.1: (a) MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$, (b) MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ and (c) MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for $K = 1$ (—), $K = 3$ (- - -) and $K = 5$ (• • •) with $N = 1000$, $Y_t = X_t + \sum_{k=1}^{K} hI_{t \geq t_k}$, where $X_t \sim \text{N}(0, 1)$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$.

Table 3.1 shows the average MSE of the ordinary sample variance for normally distributed data and different values of K and h, with $N = 1000$. Again, 1000 simulation runs are performed where jumps are added to the generated data at randomly chosen positions. We observe that the MSE becomes very large when the number and height of the level shifts increases. Obviously, the blocks-estimator $\widehat{\sigma^2}_{\text{Mean}}$ performs much better than the sample variance.

		h					
		0	1	2	3	4	5
	1	0.20	3.62	54.04	272.02	858.45	2094.59
K	3	-	115.40	1838.69	9303.10	293.7.04	71764.25
	5	-	678.29	10837.86	54854.82	173355.69	423217.58

Table 3.1: MSE$\cdot 10^2$ of the sample variance for normally distributed data, $N = 1000$ and different values of K and h.

The results for data from the t_5-distribution are similar to those obtained for the normal distribution, see Fig. D.2 in Appendix D. Again, the blocks-estimator with the block size (3.8) (Panel (c)) performs well and does not lose too much performance compared to Panel (b) of Fig. D.2, where the optimal block size is considered. Similar results are obtained for $N = 2500$, see Fig. D.4 and D.5 in Appendix D.

As the number of change-points K is not known exactly in real applications, there are several possibilities to set K in formula (3.8):

1. Use a block size $n = \max\left\{\left\lfloor\frac{\sqrt{N}}{K+1}\right\rfloor, 2\right\}$ with a large guess on the value of K.
2. Pre-estimate K with an appropriate procedure.
3. Use prior knowledge about plausible values of K.

We will discuss the first two approaches in the following two subsections.

Using a large guess on the number of jumps

If many change-points are present, a small block size should be chosen, while larger blocks are preferred in the case of only a few level shifts. If a practitioner does not have knowledge about the number of jumps in the mean we recommend choosing a rather high value K in the formula (3.8), which results in small blocks. Doing so we are usually on the safe side, since choosing too few blocks can result in a very high MSE, while the performance of the estimator does not worsen so much when choosing many blocks.

As an example we generate 1000 time series of length $N = 1000$ with $K = 3$ jumps at random positions. Figure 3.2 shows the MSE of $\widehat{\sigma^2}_{\text{Mean}}$ depending on the jump height $h = \delta\sigma$, $\delta \in \{0, 0.1, ..., 5\}$, when choosing $n = \frac{\sqrt{N}}{K+1}$ with values $K \in \{0, 1, ..., 6\}$. We observe that choosing a too small number of blocks (i.e., a too large block size) results in large MSE values if the jumps are rather high. On the other hand, the results do not worsen as much when choosing unnecessarily many and thus small blocks. Figure D.3 in Appendix D shows similar results for $K = 5$ jumps.

Choosing $n = 2$ results in a non-overlapping difference-based estimator, which performs also well but loses efficiency compared to the blocks-estimator with growing block size, which can be fully efficient, see Remarks 5 and 7.

We will not consider the blocks-estimator with the block size n depending on the choice of some large value of K instead of the true one in the following investigation, since it is a subjective choice of a practitioner.

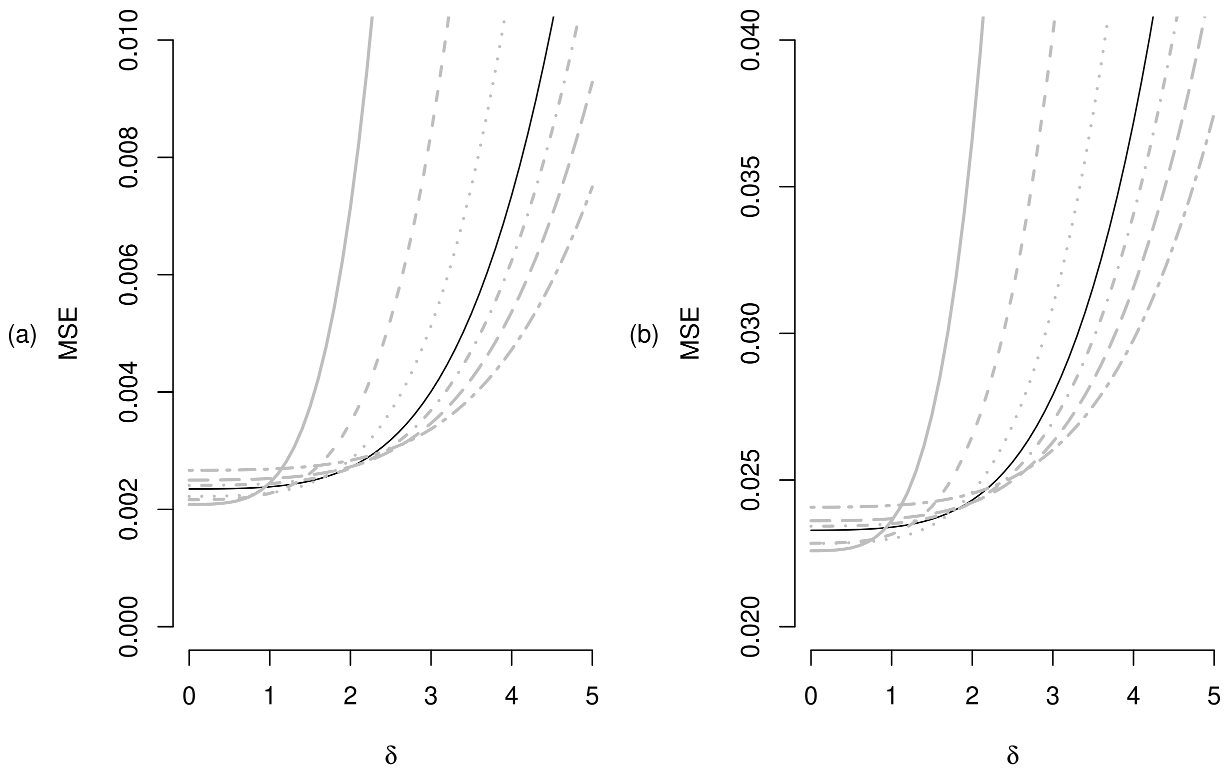

Figure 3.2: MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for true $K = 3$ (—), $K = 0$ (—) $K = 1$ (- - -) $K = 2$ (· · · ·) $K = 4$ (- · · -) $K = 5$ (– – –) and $K = 6$ (– - –) with $N = 1000$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$, $Y_t = X_t + \sum_{k=1}^{3} hI_{t \geq t_k}$, where (a) $X_t \sim \mathrm{N}(0, 1)$ and (b) $X_t \sim t_3$.

Pre-estimation of the number of jumps

We will investigate the MOSUM procedure for the detection of multiple change-points proposed by [22] to pre-estimate the number of change-points K. According to the simulations in the aforementioned paper this procedure yields very good results in comparison to many other procedures for change-point estimation. We will describe the procedure briefly in the following.

At time point t a statistic $T_{t,N}$ is calculated as

$$T_{t,N}(G) = \frac{1}{\sqrt{2G}} \left(\sum_{i=t+1}^{t+G} Y_i - \sum_{t-G+1}^{t} Y_i \right),$$

where $G = G(N)$ is a bandwidth parameter and $G \leq t \leq N - G$. In what follows we will set the bandwidth parameter to $G = \sqrt{N}$. The estimated number of change-points $\widehat{K}$ is the number of pairs (v_i, w_i), which fulfil

$$\begin{aligned}
&\frac{|T_{t,N}(G)|}{\widehat{\tau}_{t,N}} \geq D_N(G, \delta_N) \quad \text{for} \quad t = v_i, \dots, w_i \\
&\frac{|T_{t,N}(G)|}{\widehat{\tau}_{t,N}} < D_N(G, \delta_N) \quad \text{for} \quad t = v_i - 1, w_i + 1 \\
&w_i - v_i \geq \eta G \quad \text{with } 0 < \eta < 1/2 \text{ arbitrary but fixed,}
\end{aligned}$$

where $D_N(G, \delta_N)$ is a critical value depending on the bandwidth parameter G and a sequence $\delta_N \to 0$ and $\widehat{\tau^2}_{t,N}$ is a local estimator of the variance at location t. In a window of length $2G$ the location t is treated as a possible change-point location. A mean correction is performed before and after the time point t for computing $\widehat{\tau^2}_{t,N}$.

The corresponding estimated change-point locations $\widehat{t}_1, \dots, \widehat{t}_{\widehat{K}}$ are

$$\widehat{t}_i = \arg \max_{v_i \leq t \leq w_i} \frac{|T_{t,N}(G)|}{\widehat{\tau}_{t,N}}.$$

More information on the procedure can be found in [22].

We will use the MOSUM procedure to estimate the number of change-points K in the formula (3.8) for the block size n. The corresponding blocks-estimator is denoted as $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$. In Section 3.3.3 we will see that the performance of the two blocks-estimators $\widehat{\sigma^2}_{\text{Mean}}$ (as defined in (3.2)) and $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ is similar in many cases.

Moreover, we introduce an additional estimation procedure, which is fully based on the MOSUM method, for comparison. We divide the data into $\widehat{K} + 1$ blocks at the estimated locations $\widehat{t}_1, \dots, \widehat{t}_{\widehat{K}}$ of the level shifts. In every block $j = 1, \dots, \widehat{K} + 1$ the empirical variance S_j^2 is calculated. A weighted average $\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$ of those values can be computed to estimate the variance, i.e.,

$$\widehat{\sigma^2}^{\text{mosum}}_{\text{W}} = \sum_{j=1}^{\widehat{K}} \frac{n_j}{N} S_j^2, \tag{3.9}$$

where n_j is the size of block $j = 1, \dots, \widehat{K}$.

Extension to short-range dependent data

Since many real datasets exhibit autocorrelation, we investigate the averaging estimation procedure under dependence. We consider a strictly stationary linear process $(X_t)_{t \geq 1}$ with $X_t = \sum_{i=0}^{\infty} \psi_i \zeta_{t-i}$, where ζ_t are i.i.d. with mean zero and finite variance and $(\psi_i)_{i \geq 0}$ is a sequence of constants with $\sum_{i=0}^{\infty} \psi_i < \infty$. Moreover, the autocovariance function is defined as $\gamma(\delta) = E((X_t - E(X_t))(X_{t+\delta} - E(X_{t+\delta})))$, $\delta \in \mathbb{N}$, and is assumed to be absolutely summable, i.e., $\sum_{\delta=0}^{\infty} |\gamma(\delta)| < \infty$.

Theorem 4. *Consider a strictly stationary linear process $(X_t)_{t \geq 1}$ with an absolutely summable autocovariance function and $Y_t = X_t + \sum_{k=1}^{K} h_k I_{t \geq t_k}$, i.e., data with K level shifts. Moreover, let $K \left(\sum_{k=1}^{K} h_k\right)^2 = o(m \log(N)^{-1})$. Then $\widehat{\sigma^2}_{Mean} = \frac{1}{m} \sum_{j=1}^{m} S_j^2 \to \sigma^2$ in probability.*

Proof. See Appendix B. □

Remark 8. If the jump heights are bounded by a constant $h \geq h_k$, $k = 1, \ldots, K$, the strongest restriction arises if all heights equal this upper bound resulting in the constraint $K^3 h^2 = o\left(m \log(N)^{-1}\right)$.

3.3.2 Trimmed estimator of the variance

So far we have considered cases where the number of changes in the mean K is rather small with respect to the number of blocks m and thus to the number of observations N. However, there might be situations in which level shifts occur frequently. Asymptotically, the blocks-estimator $\widehat{\sigma^2}_{\text{Mean}}$ (see (3.2)) is still a good choice for the estimation of the variance as long as the number of level shifts grows slowly, see e.g. Theorem 2. However, if many jumps are present in a finite sample the blocks-estimator is no longer a good choice and will become strongly biased.

We propose an asymmetric trimmed mean of the blockwise estimates instead of their ordinary average, i.e., large estimates are removed and the average value of the remaining ones is calculated. We do not consider a symmetric trimmed mean, since the sample variance is positively biased in the presence of level shifts, so that estimates from blocks containing a level shift are expected to show up as upper outliers. Moreover, we suggest using rather many small blocks to account for potentially many level shifts. In the next Subsections 3.3.2 and 3.3.2 the choice of the trimming fraction is discussed.

Estimation with a fixed trimming fraction

The trimmed blocks-estimator is given as

$$\widehat{\sigma^2}_{\text{Tr},\alpha} = C_{N,\text{Tr},\alpha} \frac{1}{m - \lfloor \alpha m \rfloor} \sum_{j=1}^{m - \lfloor \alpha m \rfloor} S^2_{(j)}, \tag{3.10}$$

where $S^2_{(1)} \leq \ldots \leq S^2_{(m)}$ are the ordered blockwise estimates, m is the number of blocks and $C_{N,\text{Tr},\alpha}$ is a sample and distribution dependent correction factor to ensure unbiasedness in the absence of level shifts. In practice this constant can be simulated under the assumption of observing data from a known location-scale family. E.g., for standard normal distribution, $\alpha = 0.2$, $N = 1000$ and $n = 20$ ($m = 50$) we generate 1000 samples of length $N = 1000$ and calculate the average of the uncorrected trimmed variance estimates. The reciprocal of this average value yields $C_{1000,\text{Tr},0.2} = 1.198$.

As an example, we generate 1000 time series of length $N \in \{1000, 2500\}$ from normal and t_5-distribution. We add $K = N \cdot p$ jumps to the generated data at randomly chosen positions, as was done in Subsection 3.3.1, with $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$ and height $h \in \{0, 2, 3, 5\}$. We choose $n = 20$ to ensure that the number of jump-contaminated blocks is sufficiently smaller than the total number of blocks.

Table 3.2 shows the simulated MSE of the trimmed estimator (3.10) for $\alpha \in \{0.1, 0.3, 0.5\}$. Clearly, the performance of the trimmed estimator depends on the number of jumps in the mean and the trimming parameter α. Larger values of α are required when dealing with many jumps but lead to an increased MSE if there are only a few jumps. Therefore, it is reasonable to choose α adaptively, as will be described in the next Subsection 3.3.2.

		N(0, 1)			t_5		
K	h	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$
0	0	0.24	0.28	0.35	2.39	4.56	7.05
2	2	0.28	0.30	0.38	1.89	3.96	6.39
	8	0.26	0.28	0.35	1.50	3.51	6.01
4	2	0.35	0.34	0.39	1.49	3.12	5.50
	8	0.53	0.43	0.46	1.61	2.76	5.13
6	2	0.54	0.46	0.48	1.22	2.47	4.81
	8	2.80	0.54	0.51	8.74	1.99	4.32
10	2	1.14	0.76	0.67	2.03	1.58	3.56
	8	69.38	1.04	0.77	199.96	1.43	2.79

Table 3.2: Simulated MSE$\cdot 10^2$ of $\widehat{\sigma^2}_{\mathrm{Tr},\alpha}$ for normally and t_5-distributed data with $N = 1000$ and different α, h and K.

Adaptive choice of the trimming fraction

Instead of using a fixed trimming fraction we can choose α adaptively, yielding the adaptive trimmed estimator $\widehat{\sigma^2}_{\mathrm{Tr,ad}}$ with

$$\widehat{\sigma^2}_{\mathrm{Tr,ad}} = C_{N,\mathrm{Tr,ad}} \frac{1}{m - \lfloor \alpha_{\mathrm{adapt}} m \rfloor} \sum_{j=1}^{m - \lfloor \alpha_{\mathrm{adapt}} m \rfloor} S_{(j)}^2, \tag{3.11}$$

where $S_{(1)}^2 \leq \ldots \leq S_{(m)}^2$ are the ordered blockwise sample variances and α_{adapt} is the adaptively chosen percentage of the blocks-estimates, which will be removed.

Estimation under normality

We use the approach for outlier detection discussed in [18] to determine α_{adapt}, assuming that the underlying distribution is normal. In this case the distribution of the sample variance is well known, i.e., in block j we have that

$$T_j := \frac{(n-1)S_j^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Since the true variance σ^2 is not known we propose to replace σ^2 by an appropriate initial estimate, such as the median of the blocks-estimates, i.e.,

$$\widehat{\sigma^2}_{\mathrm{Med}} = C_{N,\mathrm{Med}} \cdot \mathrm{med}\{S_1^2, \ldots, S_m^2\}, \tag{3.12}$$

where $C_{N,\text{Med}}$ is a finite sample correction factor. Subsequently, we remove those values

$$\widehat{T}_j = \frac{(n-1)S_j^2}{\widehat{\sigma^2}_{\text{Med}}}, \tag{3.13}$$

which exceed $q_{\chi^2_{n-1},1-\beta_m}$, the $(1-\beta_m)$-quantile of the χ^2_{n-1}-distribution, with

$$\beta_m = 1-(1-\beta)^{1/m} \quad \text{and} \quad \beta \in (0,1). \tag{3.14}$$

We will refer to the adaptively trimmed estimator based on the approach of [18] under normality as $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$.

The choice of β_m results in a probability of $1-\beta$ that no observation (block in our case) is trimmed if $T_1,\dots,T_m$ are i.i.d. χ^2_{n-1}-distributed, i.e.,

$$P(T_1 \le q_{\chi^2_{n-1},1-\beta_m},\dots,T_m \le q_{\chi^2_{n-1},1-\beta_m}) = \left(P(T_1 \le q_{\chi^2_{n-1},1-\beta_m})\right)^m$$
$$= (1-\beta_m)^m = \left((1-\beta)^{1/m}\right)^m = 1-\beta.$$

Furthermore, we expect that roughly $m\cdot\beta_m$ blocks are trimmed on average in the absence of level shifts. The following simulations suggest that the adaptive trimming fraction is slightly larger than β_m, which can be explained by the fact that we need to use an estimate such as $\widehat{\sigma^2}_{\text{Med}}$ instead of the unknown σ^2. We generated 10000 sequences of observations of size $N\in\{1000,2500\}$ for $\beta\in\{0.05,\ 0.1\}$. Table 3.3 shows the average number of trimmed blocks in the absence of level shifts.

		N	
		1000	2500
β	0.05	0.0647	0.0583
	0.1	0.1268	0.1208

Table 3.3: Average number of trimmed blocks in the absence of level shifts for normally distributed data and different N and β for the estimator $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$.

We suggest choosing a small block size, e.g. $n=20$, to cope with a possibly large number of change-points. In this way it is ensured that the number of uncontaminated blocks is much larger than the number of perturbed blocks. Moreover, we choose $\beta=0.05$.

The correction factor in (3.11) needs to be simulated taking into account that the percentage of the omitted block-estimates is no longer fixed. Therefore, for given N and β we generate 1000 sequences of observations. In each simulation run we calculate the block estimates S_j^2, $j=1,\dots,m$, and the initial estimate of the variance $\widehat{\sigma^2}_{\text{Med}}$. Subsequently, we remove the values $\widehat{T}_j=(n-1)S_j^2/\widehat{\sigma^2}_{\text{Med}}$, which exceed the quantile $q_{\chi^2_{n-1},1-\beta_m}$. Then the average value of the remaining block-estimates is computed. The procedure yields 1000 estimates. The correction factor is the reciprocal of the average of these values. For $N=1000$ and $\beta=0.05$ the simulated correction factor is $C^{0.05}_{1000,\text{Tr,ad}}=1.0020$, while for $N=2500$ we have

$C^{0.05}_{2500,\mathrm{Tr,ad}} = 1.0009$, so both are nearly 1 and could be neglected with little loss.

Estimation under unknown distribution

When no distributional assumptions are made and the block size n is large, one can use that $\sqrt{n}\left(S_j^2 - \sigma^2\right)/\sqrt{\nu_4 - \sigma^4}$ is approximately standard normal if the fourth moment exists.

The fourth central moment $\nu_4 = E\left((X_1 - E(X_1))^4\right)$ needs to be estimated properly in the presence of level shifts. We can estimate this quantity in blocks and then compute the median of the blocks-estimates $\widehat{\mu}_{4,\mathrm{Med}}$, as was done in (3.12). Then, values

$$\widehat{T}_j = \sqrt{n}\left(S_j^2 - \widehat{\sigma^2}_{\mathrm{Med}}\right)/\sqrt{\widehat{\mu}_{4,\mathrm{Med}} - \widehat{\sigma^2}^2_{\mathrm{Med}}}, \tag{3.15}$$

which exceed the $(1-\beta_m)$-quantile of the standard normal distribution are removed. The corresponding adaptively trimmed estimator is denoted as $\widehat{\sigma^2}^{\mathrm{other}}_{\mathrm{Tr,ad}}$, where, again, $\beta = 0.05$ (see (3.14)) will be used in what follows.

Table 3.4 shows the average number of trimmed blocks in the absence of level shifts for normally distributed data, analogously to Table 3.3. We observe that the average number of trimmed blocks-estimates is much larger than the values for the trimming procedure, which is based on the normality assumption.

			N	
			1000	2500
$n = 20$	β	0.05	1.2327	2.1748
		0.1	1.6007	2.8159

			N	
			1000	2500
$n = 40$	β	0.05	0.4018	0.6036
		0.1	0.5676	0.8829

Table 3.4: Average number of trimmed blocks in the absence of level shifts for normally distributed data and different values of N and β for the estimator $\widehat{\sigma^2}^{\mathrm{other}}_{\mathrm{Tr,ad}}$.

Remark 9. We do not use distribution dependent correction factors for the estimators $\widehat{\mu}_{4,\mathrm{Med}}$ and $\widehat{\sigma^2}_{\mathrm{Med}}$, since the underlying distribution is not known. The asymptotic distribution of the blockwise empirical variances is normal and therefore, for large block sizes, the distribution is expected to be roughly symmetric. A correction factor can then be neglected with little loss, since the sample median of symmetrically distributed random variables estimates their expected value.

3.3.3 Simulations

In this section we compare the variance estimators $\widehat{\sigma^2}_{\mathrm{Mean}}$, $\widehat{\sigma^2}_{\mathrm{Diff}}$, $\widehat{\sigma^2}_{\mathrm{Tr},0.5}$ with the block size $n = 20$, $\widehat{\sigma^2}^{\mathrm{normal}}_{\mathrm{Tr,ad}}$ with the block size $n = 20$ and $\beta = 0.05$, $\widehat{\sigma^2}^{\mathrm{other}}_{\mathrm{Tr,ad}}$ with the block size $n = 40$ and $\beta = 0.05$, $\widehat{\sigma^2}^{\mathrm{mosum}}_{\mathrm{Mean}}$ and $\widehat{\sigma^2}^{\mathrm{mosum}}_{\mathrm{W}}$ in different scenarios. We generate 1000 sequences of observations of length $N \in \{200, 1000, 2500\}$ from the standard normal and the t_5 distribution. We add K jumps of heights $h \in \{0, 2, 3, 5, 8\}$ to the data at randomly chosen positions as was done in Subsection 3.3.1. K is chosen dependent on the number of observations, i.e., $K = p \cdot N$ with $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$.

Table 3.5 shows the simulated MSE for the normal distribution. The estimators $\widehat{\sigma^2}_{\text{Mean}}$ and $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ yield similar results. We conclude that the estimation of the number of jumps K (required in the rule (3.8)) does not have a large effect on the estimator. The estimators $\widehat{\sigma^2}_{\text{Mean}}$ and $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ yield the best results if the jump heights are not very large, i.e., $h \leq 2$. However, the MSE of taking the ordinary average is much larger than that of the other estimators if the jump heights are large. Large jumps result in large blockwise estimates, which have a strong impact on the ordinary average.

The trimmed estimator $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$ yields the best results among all methods considered here for normally distributed data if the jumps are rather high. When many small level shifts are present $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ outperforms $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$, although the latter makes use of the exact normality assumption. The estimator $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ tends to remove more block-estimates than $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$ in the absence of level shifts, see Tables 3.3 and 3.4. Therefore, we also expect that more blocks are trimmed away by $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ if level shifts are present, reducing the risk of including perturbed blocks in the trimmed estimator $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$.

The trimmed estimator $\widehat{\sigma^2}_{\text{Tr,0.5}}$ with a fixed trimming fraction also yields good results. However, this estimation procedure requires the knowledge of the underlying distribution to compute the finite sample correction factor, see Subsection 3.3.2. The difference-based estimator $\widehat{\sigma^2}_{\text{Diff}}$ performs well as long as the jumps are moderately high.

Table 3.6 shows the results for the t_5 distribution. The estimation procedures $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$ and $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ yield the best results in this scenario.

In Table 3.7 the simulated MSE is presented when the data is generated from the autoregressive (AR) model with $\phi = 0.5$, i.e., the data is positively correlated. The performance of the difference-based estimator worsens considerably then. This is due to the fact that this estimation procedure makes explicit use of the assumption of uncorrelatedness. While $\widehat{\sigma^2}_{\text{Diff}}$ underestimates the true variance drastically (resulting in a high MSE value) when no changes in the mean are present, the performance seems to improve slightly when dealing with many high jumps. This can be explained by the fact that the positive bias, which arises from the jumps, compensates for the negative bias, which arises from the (incorrect) assumption of uncorrelatedness. The blocks-estimator $\widehat{\sigma^2}_{\text{Mean}}$ exhibits a similar behaviour, since the block size is small when the number of jumps is high, while correlated data require large block sizes to ensure satisfying results. For dependent data the best results are obtained when using the adaptively trimmed estimators.

K	h	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr},0.5}$	$\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$
					$N = 200$			
0	0	1.10	1.51	1.64	1.10	1.38	1.19	0.98
1	2	1.34	1.54	1.77	1.54	1.79	1.29	2.92
	3	1.81	1.60	1.87	1.43	1.50	1.73	1.01
	5	5.23	2.03	1.94	1.32	1.46	5.00	0.95
	8	26.31	4.41	2.01	1.30	1.48	24.27	0.99
2	2	1.56	1.57	2.28	2.15	3.11	1.59	2.13
	3	2.25	1.76	2.45	2.13	2.37	2.40	1.11
	5	7.15	3.21	2.52	1.55	1.94	7.27	2.40
	8	37.09	12.17	2.38	1.34	1.77	40.25	5.95
					$N = 1000$			
0	0	0.21	0.30	0.35	0.23	0.27	0.22	0.18
2	2	0.25	0.30	0.38	0.29	0.28	0.27	0.21
	3	0.36	0.31	0.35	0.25	0.25	0.40	0.21
	5	1.23	0.37	0.36	0.23	0.26	1.28	0.30
	8	6.47	0.72	0.35	0.20	0.24	6.69	1.04
4	2	0.29	0.31	0.39	0.39	0.34	0.29	0.23
	3	0.46	0.33	0.41	0.38	0.26	0.47	0.30
	5	1.79	0.56	0.42	0.25	0.27	1.96	0.79
	8	10.40	1.95	0.46	0.24	0.27	10.57	4.48
6	2	0.32	0.32	0.48	0.59	0.49	0.33	0.33
	3	0.51	0.38	0.49	0.49	0.35	0.55	0.63
	5	2.03	0.87	0.47	0.26	0.27	2.29	2.31
	8	11.68	4.01	0.51	0.25	0.28	13.25	15.74
10	2	0.46	0.34	0.67	1.16	1.16	0.40	0.55
	3	0.66	0.50	0.71	0.88	0.58	0.71	1.79
	5	2.28	1.87	0.72	0.32	0.38	3.07	11.23
	8	12.36	10.57	0.77	0.27	0.32	18.10	69.41
					$N = 2500$			
0	0	0.08	0.12	0.13	0.08	0.10	0.09	0.08
5	2	0.11	0.12	0.15	0.14	0.11	0.12	0.09
	3	0.17	0.13	0.14	0.11	0.10	0.18	0.14
	5	0.70	0.18	0.15	0.10	0.10	0.76	0.36
	8	4.08	0.53	0.15	0.09	0.10	4.40	2.14
10	2	0.13	0.13	0.17	0.27	0.18	0.13	0.16
	3	0.21	0.15	0.19	0.20	0.12	0.24	0.44
	5	0.87	0.37	0.21	0.11	0.11	1.02	2.39
	8	4.99	1.76	0.18	0.09	0.10	6.43	12.63
15	2	0.15	0.13	0.23	0.47	0.33	0.15	0.38
	3	0.26	0.19	0.27	0.37	0.15	0.27	1.27
	5	1.19	0.68	0.28	0.13	0.11	1.27	8.19
	8	7.03	3.81	0.24	0.08	0.11	7.54	53.93
25	2	0.21	0.16	0.47	1.22	0.96	0.19	2.00
	3	0.39	0.32	0.51	0.94	0.32	0.38	7.90
	5	1.86	1.68	0.54	0.17	0.17	1.82	52.63
	8	11.14	10.37	0.48	0.11	0.13	11.16	356.78

Table 3.5: Simulated MSE$\cdot 10^2$ of $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr},0.5}$, $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ and $\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$ for normally distributed data and different sample sizes N, jump heights $h \cdot \sigma$ and number of jumps $K = p \cdot N$ with $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$.

K	h	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr},0.5}$	$\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$
				$N = 200$				
0	0	11.56	12.53	7.80	8.16	9.38	10.60	8.93
1	2	12.24	12.62	8.57	7.78	10.85	12.69	17.70
	3	13.48	12.80	9.22	8.75	11.63	12.88	11.62
	5	23.04	13.98	9.85	8.89	10.49	21.03	9.35
	8	79.79	20.62	9.33	8.51	9.86	78.54	10.94
2	2	12.67	12.70	10.82	8.28	17.25	10.50	17.78
	3	14.42	13.23	11.72	11.60	27.34	15.19	11.15
	5	28.38	17.27	13.08	10.16	31.33	29.36	16.02
	8	111.37	42.16	12.35	8.84	16.83	129.94	21.69
				$N = 1000$				
0	0	2.26	2.50	1.40	2.24	2.80	1.87	2.01
2	2	2.37	2.51	1.61	1.82	2.38	2.43	2.32
	3	2.66	2.53	1.64	1.99	2.65	2.52	1.93
	5	5.02	2.69	1.70	2.33	2.88	4.72	2.69
	8	20.46	3.68	1.64	2.25	2.60	21.87	3.90
4	2	2.48	2.52	1.80	1.59	2.19	2.06	2.37
	3	2.94	2.60	1.94	1.91	2.51	2.98	2.96
	5	6.69	3.21	1.86	2.09	2.62	6.59	3.88
	8	29.63	7.10	1.82	2.17	2.54	33.05	9.51
6	2	2.56	2.54	2.30	1.64	2.46	2.25	2.42
	3	3.09	2.71	2.22	1.86	2.47	2.96	3.37
	5	7.34	4.08	2.48	2.05	2.39	7.76	6.59
	8	33.98	12.80	2.34	2.06	2.60	39.50	39.14
10	2	2.95	2.61	3.07	2.26	4.61	2.49	3.15
	3	3.52	3.07	3.40	2.98	4.90	3.82	6.11
	5	7.89	6.86	3.71	2.13	3.17	11.64	26.66
	8	35.75	31.04	3.93	2.20	2.89	50.17	175.67
				$N = 2500$				
0	0	0.89	1.00	0.58	1.32	1.81	0.83	0.94
5	2	0.97	1.01	0.65	0.79	1.28	0.90	0.79
	3	1.15	1.02	0.70	0.99	1.37	1.19	0.87
	5	2.67	1.18	0.69	1.20	1.45	2.67	2.14
	8	12.06	2.14	0.73	1.25	1.55	12.94	7.97
10	2	1.02	1.02	0.92	0.55	0.93	0.87	1.12
	3	1.24	1.09	1.03	0.74	1.20	1.48	2.03
	5	3.03	1.70	0.99	1.02	1.23	3.53	6.21
	8	14.40	5.56	0.95	1.23	1.38	17.91	36.37
15	2	1.09	1.04	1.24	0.65	1.02	1.17	1.64
	3	1.40	1.20	1.38	0.86	1.05	1.39	4.19
	5	4.00	2.57	1.40	1.00	1.10	4.24	24.28
	8	20.18	11.25	1.34	1.15	1.24	21.27	159.02
25	2	1.25	1.11	2.45	1.63	3.28	1.15	7.77
	3	1.75	1.56	2.85	2.23	2.30	1.67	21.70
	5	5.77	5.35	2.94	0.89	1.19	5.95	145.89
	8	32.04	29.47	2.78	1.11	1.18	31.27	951.03

Table 3.6: Simulated MSE$\cdot 10^2$ of $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr},0.5}$, $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ and $\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$ for t_5-distributed data and different sample sizes N, jump heights $h \cdot \sigma$ and number of jumps $K = p \cdot N$ with $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$.

K	h	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr},0.5}$	$\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$
				$N = 200$				
0	0	5.21	44.79	7.18	5.92	4.95	27.83	5.02
1	2	9.75	43.01	6.08	4.78	5.15	32.52	6.05
	3	7.84	41.08	5.94	5.30	5.31	30.28	5.17
	5	4.69	34.70	6.00	5.31	5.06	24.75	5.35
	8	19.66	21.60	5.74	5.14	4.73	18.98	5.51
2	2	20.08	41.81	5.50	4.82	6.08	35.01	5.84
	3	15.88	37.36	5.14	5.29	6.45	31.92	5.31
	5	7.75	26.00	4.98	5.30	5.25	21.20	6.05
	8	18.98	6.77	5.00	5.37	4.86	16.33	14.93
				$N = 1000$				
0	0	1.31	44.58	5.32	2.72	1.90	18.58	1.02
2	2	5.12	43.84	4.59	2.16	1.53	24.74	1.11
	3	4.24	43.00	4.53	2.40	1.77	22.73	1.05
	5	2.10	40.32	4.54	2.57	1.78	20.18	1.12
	8	2.60	33.96	4.54	2.60	1.71	14.36	1.76
4	2	11.48	43.01	4.05	1.87	1.50	29.51	1.12
	3	9.41	41.39	3.87	1.98	1.51	26.50	0.99
	5	4.66	36.13	3.92	2.39	1.61	21.68	1.85
	8	1.98	24.60	3.72	2.42	1.68	11.80	8.29
6	2	20.15	42.53	3.61	1.51	1.23	33.68	1.02
	3	17.05	39.82	3.46	1.80	1.32	30.27	1.07
	5	9.76	32.04	3.49	2.45	1.37	22.29	3.69
	8	1.90	17.06	3.56	2.64	1.66	10.47	19.47
10	2	40.98	40.84	2.81	1.10	1.11	38.31	0.89
	3	37.03	37.04	2.49	1.36	1.38	34.43	2.06
	5	25.41	25.13	2.59	2.32	1.38	23.44	14.21
	8	7.94	6.07	2.44	2.52	1.35	7.12	122.52
				$N = 2500$				
0	0	0.51	44.46	4.83	2.16	1.41	13.95	0.38
5	2	7.50	43.79	4.30	1.65	1.05	22.84	0.38
	3	6.45	42.93	4.32	1.79	1.18	21.52	0.40
	5	3.68	40.15	4.34	2.09	1.24	18.08	0.75
	8	0.64	33.80	4.32	2.20	1.22	12.27	3.50
10	2	20.78	43.04	3.78	1.23	0.78	31.08	0.36
	3	18.76	41.45	3.72	1.44	0.92	29.04	0.60
	5	13.30	36.13	3.74	1.93	1.08	23.45	3.77
	8	3.93	24.69	3.59	2.09	1.08	12.97	22.39
15	2	28.58	42.44	3.38	0.94	0.59	37.94	0.45
	3	25.87	39.86	3.16	1.06	0.80	35.29	1.75
	5	17.91	32.15	2.98	1.76	0.97	27.69	14.74
	8	5.13	16.92	3.10	2.08	0.96	13.55	78.80
25	2	41.03	41.11	2.46	0.47	0.47	40.96	1.97
	3	36.83	36.80	2.09	0.59	0.59	36.77	11.09
	5	25.48	25.06	1.99	1.49	0.73	25.45	89.71
	8	6.84	5.78	2.07	2.00	0.88	6.84	556.63

Table 3.7: Simulated MSE$\cdot 10^2$ of $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr},0.5}$, $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ and $\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$ for data generated from the AR(1) model with $\phi = 0.5$ and different sample sizes N, jump heights $h \cdot \sigma$ and number of jumps $K = p \cdot N$ with $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$.

Since the variance σ^2 is underestimated when the data are dependent, the values $\widehat{T}_j$ in (3.13) and (3.15) get larger resulting in a higher trimming parameter α_{adapt}. Therefore, more blocks are trimmed away ensuring that the perturbed ones are not involved in the calculation of the overall estimate.

Figure 3.3 shows the simulated MSE dependent on the AR-parameter $\phi \in \{0.1, \dots, 0.8\}$ in four different scenarios: no jumps, few small jumps, many small jumps and many high jumps. We observe that the performance of all estimators $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr},0.5}$, $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ and $\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$ worsens when the strength of the correlation (expressed by the parameter ϕ) grows. When no changes in the mean are present (see Panel (a)) the estimators underestimate the variance more as ϕ gets larger.

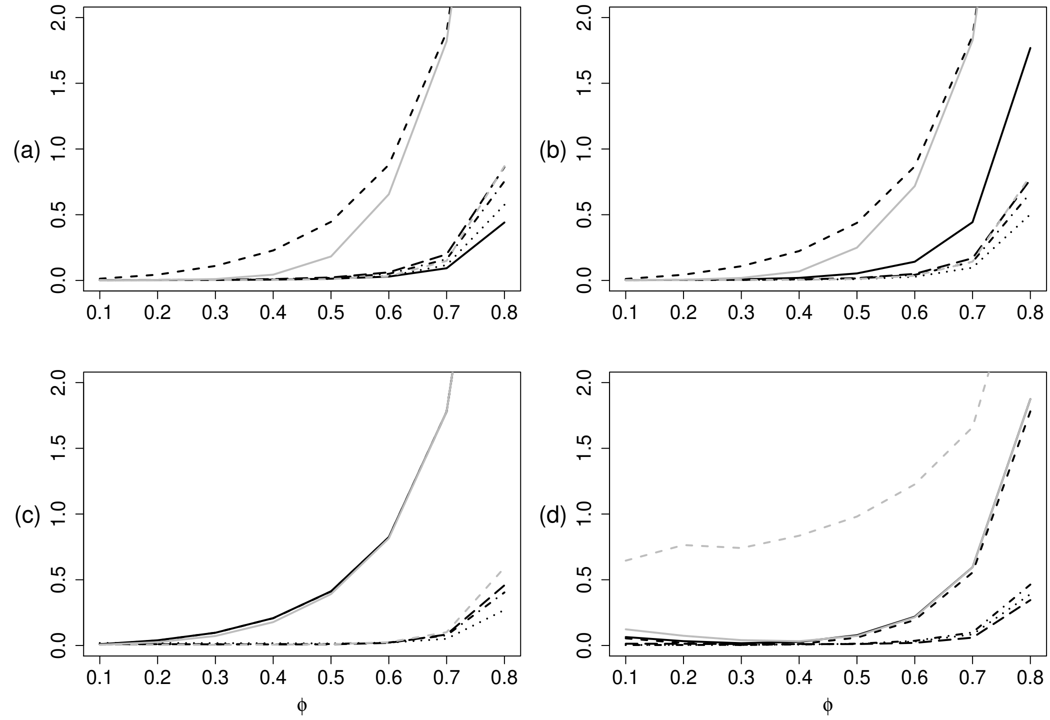

Figure 3.3: Simulated MSE of $\widehat{\sigma^2}_{\text{Mean}}$ (—), $\widehat{\sigma^2}_{\text{Diff}}$ (- - -), $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$ (· · ·), $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$ (- · -), $\widehat{\sigma^2}_{\text{Tr},0.5}$ (— — —), $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ with bandwidth $G = \sqrt{N}$ (—) and $\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$ with bandwidth $G = \sqrt{N}$(- - -) for (a) $K = 0, h = 0$, (b) $K = 2, h = 2$, (c) $K = 10, h = 2$ and (d) $K = 10, h = 8$ with $N = 1000$, $Y_t = X_t + \sum_{k=1}^{K} hI_{t \geq t_k}$, where X_t originates from the AR(1)-process with parameter $\phi \in \{0.1, 0.2, \dots, 0.8\}$.

On the other hand, large level shifts result in a positive bias of the estimators. Therefore, when many high jumps are present (Panel (d)) the results are better than in the case of many small jumps (Panel (c)). This is more obvious for the estimators $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$ and $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$. The trimmed estimation procedures do not suffer much from the increasing strength of correlation in our example. The averaging approach $\widehat{\sigma^2}_{\text{Mean}}$ yields good results if the dependence is rather weak and if only few small jumps are present. The performance of the weighted average $\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$ worsens drastically when many high jumps are present. During this estimation procedure the data is segregated into few blocks at estimated change-point locations. The approach is highly biased in the case of many and high level shifts if the number of change-points is underestimated. This is sometimes the case if two level shifts are not sufficiently far away from each other, see [22].

Based on the simulation results we recommend using the averaging estimation procedure $\widehat{\sigma^2}_{\text{Mean}}$ if the data are not expected to be correlated or heavy-tailed and the jump heights are rather small. Otherwise, if either no information on the distribution, the dependence structure of the data and the jump heights is given or the jumps are expected to be rather large, the adaptively trimmed procedure $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$ can be recommended.

3.3.4 Blockwise estimation of the standard deviation

In many applications we do not wish to estimate the variance σ^2 but rather the standard deviation σ, e.g. for standardization.

Estimation by the blockwise average

We will consider the two blocks-estimators

$$\widehat{\sigma}^{\text{corr}}_{\text{Mean},1} = C_{N,1}\widehat{\sigma}_{\text{Mean},1} = C_{N,1}\frac{1}{m}\sum_{j=1}^{m} S_j \quad \text{and} \tag{3.16}$$

$$\widehat{\sigma}^{\text{corr}}_{\text{Mean},2} = C_{N,2}\widehat{\sigma}_{\text{Mean},2} = C_{N,2}\sqrt{\frac{1}{m}\sum_{j=1}^{m} S_j^2} = C_{N,2}\sqrt{\widehat{\sigma^2}_{\text{Mean}}}, \tag{3.17}$$

where $C_{N,1}$ and $C_{N,2}$ are sample dependent correction factors to ensure unbiasedness when no changes in the mean are present. The block size n can be chosen according to the rule (3.8). If the number of change-points is not known in practice, it can be estimated as is done in Subsection 3.3.1.

For normally distributed data the correction factors $C_{N,1}$ and $C_{N,2}$ can be determined analytically. To derive the correction factor $C_{N,1}$ for the estimator $\widehat{\sigma}^{\text{corr}}_{\text{Mean},1}$ we will first consider the exact distribution of the empirical variance when dealing with jumps in the mean in order to derive the distribution of $\widehat{\sigma}_{\text{Mean},1}$ in (3.16). Given independent identically normally distributed data $X_{j,1}, \ldots, X_{j,n}$ it is well known that $\frac{(n-1)S_j^2}{\sigma^2} \sim \chi^2_{n-1}$ in a j-th jump-free block with n observations. The situation is different in the presence of jumps. Without loss of generality the following lemma is expressed in terms of the first block consisting of the observation times $t = 1, \ldots, n$ and containing $\widetilde{K}_1 \leq K$ jumps.

Lemma 5. *Assume that* $X_1, \ldots, X_n \sim \mathcal{N}(0, \sigma^2)$ *and* $Y_t = X_t + \sum_{k=1}^{\widetilde{K}_1} h_k I_{t \geq t_k}$ *for* $t = 1, \ldots, n$. *Then we have for* $S_1^2 = \frac{1}{n-1}\sum_{t=1}^{n}(Y_t - \overline{Y}_1)^2$ *that*

$$\frac{n-1}{\sigma^2} S_1^2 \sim \chi^2_{n-1,\lambda_1} \quad (\textit{the non-central chi-squared distribution}),$$

where $\lambda_1 = \frac{1}{\sigma^2}\sum_{t=1}^{n}\left(\mu_{1,t} - \overline{\mu}_1\right)^2$, $\mu_{1,t} = \sum_{k=1}^{\widetilde{K}_1} h_k I_{t \geq t_k}$ *and* $\overline{\mu}_1 = \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{\widetilde{K}_1} h_k I_{t \geq t_k}$.

Proof. $\overline{Y}_1 = \overline{X}_1 + \overline{\mu}_1$ and $Y_t - \overline{Y}_1 = X_t - \overline{X}_1 - \overline{\mu}_1 + \mu_{1,t}$, $t = 1, \ldots, n$, are independent, since $\overline{X}_1$ and $X_t - \overline{X}_1$ are independent and the remaining terms are deterministic constants. Hence, S_1^2 and $\overline{Y}_1$ are independent. Furthermore,

$$\begin{aligned}\sum_{t=1}^{n}\left(\frac{Y_t - \overline{\mu}_1}{\sigma}\right)^2 &= \sum_{t=1}^{n}\left(\frac{Y_t - \overline{Y}_1 + \overline{Y}_1 - \overline{\mu}_1}{\sigma}\right)^2 \\ &= \sum_{t=1}^{n}\left(\frac{Y_t - \overline{Y}_1}{\sigma}\right)^2 + \sum_{t=1}^{n}\left(\frac{\overline{Y}_1 - \overline{\mu}_1}{\sigma}\right)^2 \\ &+ 2\left(\frac{\overline{Y}_1 - \overline{\mu}_1}{\sigma}\right)\sum_{t=1}^{n}\left(\frac{Y_t - \overline{Y}_1}{\sigma}\right) \\ &= \frac{n-1}{\sigma^2}S_1^2 + \frac{n}{\sigma^2}\left(\overline{Y}_1 - \overline{\mu}_1\right)^2 + 0 = \frac{n-1}{\sigma^2}S_1^2 + \frac{n}{\sigma^2}\overline{X}_1^2\end{aligned}$$

with $\sum_{t=1}^{n}\left(\frac{Y_t - \overline{\mu}_1}{\sigma}\right)^2 \sim \chi^2_{n,\lambda_1}$, since $Y_t\,\forall t$ are independent and $\frac{n}{\sigma^2}\overline{X}_1^2 \sim \chi^2_1$. The moment generating function at $z \in \mathbb{R}$ of both sides and the independence of S_1^2 and $\overline{Y}_1$ yield:

$$\begin{aligned}(1 - 2\cdot z)^{-n/2}\exp\left(\frac{\lambda_1 z}{1-2z}\right) &= M_{\chi^2_{n,\lambda_1}}(z) = M_{\frac{n-1}{\sigma^2}S_1^2}(z)\cdot M_{\chi^2_1}(z) \\ &= M_{\frac{n-1}{\sigma^2}S_1^2}(z)\cdot(1-2\cdot z)^{-1/2}\end{aligned}$$

$$\begin{aligned}&\Leftrightarrow M_{\frac{n-1}{\sigma^2}S_1^2}(z) = (1-2\cdot z)^{-(n-1)/2}\exp\left(\frac{\lambda_1 z}{1-2z}\right) = M_{\chi^2_{n-1,\lambda_1}}(z) \\ &\Rightarrow \frac{n-1}{\sigma^2}S_1^2 \sim \chi^2_{n-1,\lambda_1}.\end{aligned}$$

□

In the following we assume that $B \leq K$ blocks are contaminated by $\widetilde{K}_1, \ldots, \widetilde{K}_B$ jumps, respectively, with $\sum_{k=1}^{B}\widetilde{K}_k = K$. Without loss of generality assume that the jumps are contained in the first B blocks, while the last $m - B > 0$ blocks do not contain any jumps. The square root of a χ^2_{n-1,λ_j}-distributed random variable $(n-1)S_j^2/\sigma^2$ is χ-distributed with $n-1$ degrees of freedom and non-centrality parameter $\sqrt{\lambda_j}$, see e.g. [57]. We hence have $\sqrt{n-1}S_j/\sigma \sim \chi_{n-1,\sqrt{\lambda_j}}$, $j = 1, \ldots, m$, where $\lambda_j = 0$ for the last blocks $j = B+1, \ldots, m$, i.e., $\sqrt{n-1}S_j/\sigma \sim \chi_{n-1}$. The expected value of S_j is given as

$$E(S_j) = \sigma\frac{\sqrt{2}}{\sqrt{n-1}}\frac{\Gamma(0.5n)}{\Gamma(0.5(n-1))}F_{1,1}(-0.5, 0.5(n-1), -0.5\lambda_j) =: \sigma C_{n,\lambda_j}, \tag{3.18}$$

where $F_{1,1}(a,b,z)$ represents the generalized hypergeometric function, see [61] for more details. When no changes in the mean are present we have that $\lambda_j = 0\,\forall\, j$ and therefore $F_{1,1}(-0.5, 0.5(n-1), -0.5\lambda_j) = 1$. The exact finite sample correction factor is given as

$$C_{N,1} = \frac{\sqrt{n-1}}{\sqrt{2}}\frac{\Gamma(0.5(n-1))}{\Gamma(0.5n)},$$

which is the reciprocal of the term C_{n,λ_j} in (3.18) when no level shifts are present, since $F_{1,1}(-0.5, 0.5(n-1), -0.5\lambda_j) = 1$, $j = 1, ..., m$, in this case.

For the second estimator (3.17) we have the following statements on its expectation and a suitable finite sample correction factor:

$$
\begin{aligned}
\widehat{\sigma}^{\text{corr}}_{\text{Mean},2} &= C_{N,2}\frac{\sigma}{\sqrt{m(n-1)}}\left(\frac{n-1}{\sigma^2}\sum_{j=B+1}^{m} S_j^2 + \frac{n-1}{\sigma^2}\sum_{j=1}^{B} S_j^2\right)^{1/2}, \\
E\left(\widehat{\sigma}^{\text{corr}}_{\text{Mean},2}\right) &= C_{N,2}\frac{\sigma\sqrt{2}}{\sqrt{m(n-1)}}\frac{\Gamma(0.5(m(n-1)+1))}{\Gamma(0.5m(n-1))} \\
&\quad \times F_{1,1}\left(-0.5, 0.5m(n-1), -0.5\sum_{j=1}^{B}\lambda_j\right) \\
&:= C_{N,2}\sigma D_{n,\lambda_1,\ldots,\lambda_B}, \\
C_{N,2} &= \frac{\sqrt{m(n-1)}}{\sqrt{2}}\frac{\Gamma(0.5m(n-1))}{\Gamma(0.5(m(n-1)+1))}.
\end{aligned}
$$

We used the fact that $\widehat{\sigma}_{\text{Mean},2}$ follows a scaled $\chi_{u,v}$ distribution with $u = m(n-1)$ degrees of freedom and the non-centrality parameter $v = \sqrt{\sum_{j=1}^{B}\lambda_j}$, since $\frac{n-1}{\sigma^2}\sum_{j=B+1}^{m} S_j^2 \sim \chi^2_{(m-B)(n-1)}$ and $\frac{n-1}{\sigma^2}\sum_{j=1}^{B} S_j^2 \sim \chi^2_{B(n-1),\sum_{j=1}^{B}\lambda_j}$. Using this information we can determine the expectation of the estimator straightforwardly, see [57] and [61]. The correction factor is the reciprocal of $D_{n,\lambda_1,\ldots,\lambda_B}$, where $F_{1,1}\left(-0.5, 0.5m(n-1), -0.5\sum_{j=1}^{B}\lambda_j\right) = 1$ in the absence of level shifts.

The following consistency statements are valid for the two introduced uncorrected estimators $\widehat{\sigma}_{\text{Mean},1} = \frac{1}{m}\sum_{j=1}^{m} S_j$ (as defined in (3.16)) and $\widehat{\sigma}_{\text{Mean},2} = \sqrt{\frac{1}{m}\sum_{j=1}^{m} S_j^2}$ (as defined in (3.17)) of σ:

Corollary 1. *Under the conditions of Theorem 2 the estimators $\widehat{\sigma}_{Mean,1}$ and $\widehat{\sigma}_{Mean,2}$ converge almost surely to σ, as $N \to \infty$.*

Proof. The strong consistency of $\widehat{\sigma}_{\text{Mean},2}$ follows immediately from the Continuous Mapping Theorem.
For $\widehat{\sigma}_{\text{Mean},1}$, we have due to Theorem 2 of [41] that

$$
\frac{1}{m}\sum_{j=1}^{m}\left(S_j - E\left(S_j\right)\right) \to 0 \quad \text{almost surely},
$$

since $S_j - E\left(S_j\right)$ are uniformly bounded with $P(|S_j - E\left(S_j\right)| > t) \to 0\,\forall t$ due to Chebyshev's inequality and $Var(S_j) \to 0$.

Let $S_{j,h}$ be the sample standard deviation in the perturbed block while $S_{j,0}$ is the estimate in the uncontaminated block. We have that

$$
\frac{1}{m}\sum_{j=1}^{m}\left(S_j - E\left(S_j\right)\right) = \widehat{\sigma}_{\text{Mean},1} - \frac{1}{m}\left(\sum_{j=B+1}^{m} E\left(S_{j,0}\right) + \sum_{j=1}^{B} E\left(S_{j,h}\right)\right),
$$

i.e., it suffices to show $\frac{1}{m}\left(\sum_{j=B+1}^{m} E\left(S_{j,0}\right) + \sum_{j=1}^{B} E\left(S_{j,h}\right)\right) \to \sigma$. For the first of these two terms we have

$$\frac{1}{m}\sum_{j=B+1}^{m} E\left(S_{j,0}\right) = \frac{m-B}{m} E\left(S_{1,0}\right) \longrightarrow \sigma \quad \text{as} \quad N \longrightarrow \infty,$$

since the consistency and the decreasing variance of $S_{1,0}$ implies convergence of the expectation, see Lemma 1.4A in [79]. Using Jensen's inequality we get for the second term

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{B} E\left(S_{j,h}\right) = \frac{1}{m}\sum_{j=1}^{B} E\left(\sqrt{S_{j,h}^2}\right) \\
&\leq \frac{1}{m}\sum_{j=1}^{B} \sqrt{E\left(S_{j,0}^2 + \sum_{t=1}^{n} \frac{2(X_{j,t}-\overline{X}_j)(\mu_{j,t}-\overline{\mu}_j) + \left(\mu_{j,t}-\overline{\mu}_j\right)^2}{n-1}\right)} \\
&= \frac{1}{m}\sum_{j=1}^{B} \sqrt{E\left(S_{j,0}^2\right) + \sum_{t=1}^{n} \frac{\left(\mu_{j,t}-\overline{\mu}_j\right)^2}{n-1}} = \frac{1}{m}\sum_{j=1}^{B} \sqrt{\sigma^2 + \sum_{t=1}^{n} \frac{\left(\mu_{j,t}-\overline{\mu}_j\right)^2}{n-1}} \\
&\leq \frac{B}{m}\sqrt{\sigma^2 + \frac{n}{n-1}\left(\sum_{k=1}^{K} h_k\right)^2} \longrightarrow 0,
\end{aligned}$$

where $\mu_{j,t}$ and $\overline{\mu}_j$ are defined in the proof of Theorem 2. □

Remark 10. The correction factors $C_{N,1}$ and $C_{N,2}$ from (3.16) and (3.17) satisfy

$$C_{N,1} \to 1 \quad \text{and} \quad C_{N,2} \to 1 \quad \text{as} \quad N \to \infty,$$

where $C_{N,1} = \sigma/E(\widehat{\sigma}_{\text{Mean},1})$ and $C_{N,2} = \sigma/E(\widehat{\sigma}_{\text{Mean},2})$ in the absence of level shifts. This can be shown with Lemma 1.4A in [79], since $\widehat{\sigma}_{\text{Mean},1}$ and $\widehat{\sigma}_{\text{Mean},2}$ are consistent estimators and their variances tend to zero, which implies convergence of the means and thus the above statement. Therefore, for large N and n we can neglect the correction factors and use the estimators $\widehat{\sigma}_{\text{Mean},1}$ and $\widehat{\sigma}_{\text{Mean},2}$ instead of $\widehat{\sigma}^{\text{corr}}_{\text{Mean},1}$ and $\widehat{\sigma}^{\text{corr}}_{\text{Mean},1}$ with block sizes $n \to \infty$.

Trimmed estimation

When dealing with a large number of level shifts, as is discussed in Section 3.3.2, the square root of the variance estimator $\widehat{\sigma^2}_{\text{Tr,ad}}$ from (3.11) can be used to estimate the standard deviation σ. For large N and n a correction factor to ensure unbiasedness when no changes in the mean are present can be neglected. Table 3.8 shows the simulated finite sample correction factors for normally and t_5-distributed data as well as for the stationary AR(1)-process with normal errors and parameter $\phi \in \{0.3, 0.6\}$. We observe that the correction factors are nearly one except for strongly correlated data, i.e., AR-process with parameter $\phi = 0.6$.

		$N(0,1)$	t_5	AR(0.3)	AR(0.6)
$\sqrt{\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}}$	$N = 1000$, $n = 50$	1.0025	1.0660	1.0213	1.0939
	$N = 5000$, $n = 100$	0.9999	1.0413	1.0108	1.0450
$\sqrt{\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}}$	$N = 1000$, $n = 50$	1.0082	1.0666	1.0310	1.1179
	$N = 5000$, $n = 100$	1.0014	1.0334	1.0135	1.0550

Table 3.8: Simulated finite sample correction factors for the adaptively trimmed estimation procedures for normally and t_5-distributed data as well as for the stationary AR(1)-process with normal errors and parameter $\phi \in \{0.3, 0.6\}$, denoted by AR(0.3) and AR(0.6).

3.3.5 Application

In this section we apply the blocks-approach to two datasets in order to estimate the variance.

Nile river flow data

The first dataset contains the widely discussed Nile river flow records in Aswan from 1871 to 1984, see e.g. [34], [37], [87], among many others. We consider the $N = 114$ annual maxima of the average monthly discharge in m^3/s, since these values are often assumed to be independent in hydrology. The maxima are determined from January to December. The flooding season is from July to September, see [34]. Fig. 3.4 shows the annual maxima of the average monthly discharge of the Nile river for the years 1871 – 1984.

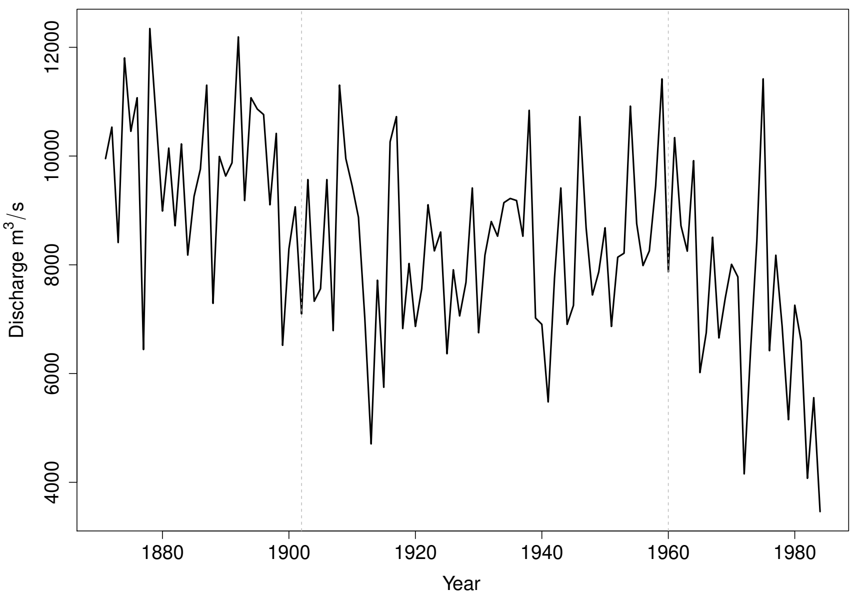

Figure 3.4: Maximal monthly discharge of the Nile river at Aswan in the period 1871 – 1984.

The construction of the two Aswan dams in 1902 and from 1960 to 1969 obviously caused changes in the river flow, see [34] and [37]. We used Levene's test (see Section 12.4.2 in [24]) to check the three segments of the data (divided by the years 1902 and 1960) for equality of variances. The null hypothesis of equal variances was not rejected with a p-value of $p = 0.40$.

A Q-Q plot of the data indicates that the deviation from a normal distribution is not large, see Fig. D.6 in Appendix D. With $\beta = 0.05$ and $n = 10$ ($m = 11$ blocks) no blocks are trimmed away during the trimming procedures $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$ and $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$. The ordinary sample variance of the entire data yields the value 3243866, see Table 3.9. For the blocks-estimator of the variance from (3.2) we choose the block size according to (3.8) with $K = 2$ getting $n = \lfloor\sqrt{114}/3\rfloor = 3$. All blockwise estimators examined in this section yield much smaller variance estimates for the whole observation period, ranging from 2075819 to 2684368.

We conclude that the procedures $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr,0.5}}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$, $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ and $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$ perform better than the ordinary sample variance, since the estimated values on the whole dataset are similar to those for the period 1903 – 1960 in between the changes.

S^2	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr,0.5}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$	$\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$	$\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$
				1871 – 1984			
3244	2123	2219	2564	2122	2122	2684	2076
				1903 – 1960			
2129	2396	1815	2041	2278	2278	2003	2129

Table 3.9: Rounded estimates ($\cdot 10^{-3}$) of the variance for the annual maxima of the average monthly discharge of the Nile river in Aswan.

PAMONO data

In the second example we use data obtained from the PAMONO (Plasmon Assisted Microscopy of Nano-Size Objects) biosensor, see [84]. This technique is used for detection of small particles, e.g. viruses, in a sample fluid. For more details, see [83]. PAMONO data sets are sequences of grayscale images. A particle adhesion causes a sustained local intensity change. This results in an obvious level shift in the time series of grayscale values for each corresponding pixel coordinate. To the best of our knowledge, a change of the variance after a jump in the mean is not expected to occur. A Q-Q plot of the data indicates that the assumption of a normal distribution is reasonable, see Fig. D.7 in Appendix D. Moreover, the PAMONO data were analysed by [1], where the authors analyse these data assuming a normal distribution.

In Panel (a) of Fig. 3.5 we see a time series corresponding to one pixel, which exhibits a virus adhesion, therefore revealing several level shifts in the mean of the time series. $N = 1000$ observations are available. Panel (b) of Fig. 3.5 shows a boxplot of 101070 values of the ordinary sample variance for time series, which correspond to pixels without virus adhesion. Since changes in the mean are not expected there, we use these data to get some insight into the typical value range of the variance. The sample variance of the contaminated data (upper panel) is $1.59 \cdot 10^{-4}$, which is not within the typical range of values, since it exceeds the upper whisker of the boxplot.

The other estimation procedures discussed in this section yield values within the interval $[1.1 \cdot 10^{-4}, 1.2 \cdot 10^{-4}]$, which are well within the interquartile range. We conclude that these approaches yield reasonable estimates for these data.

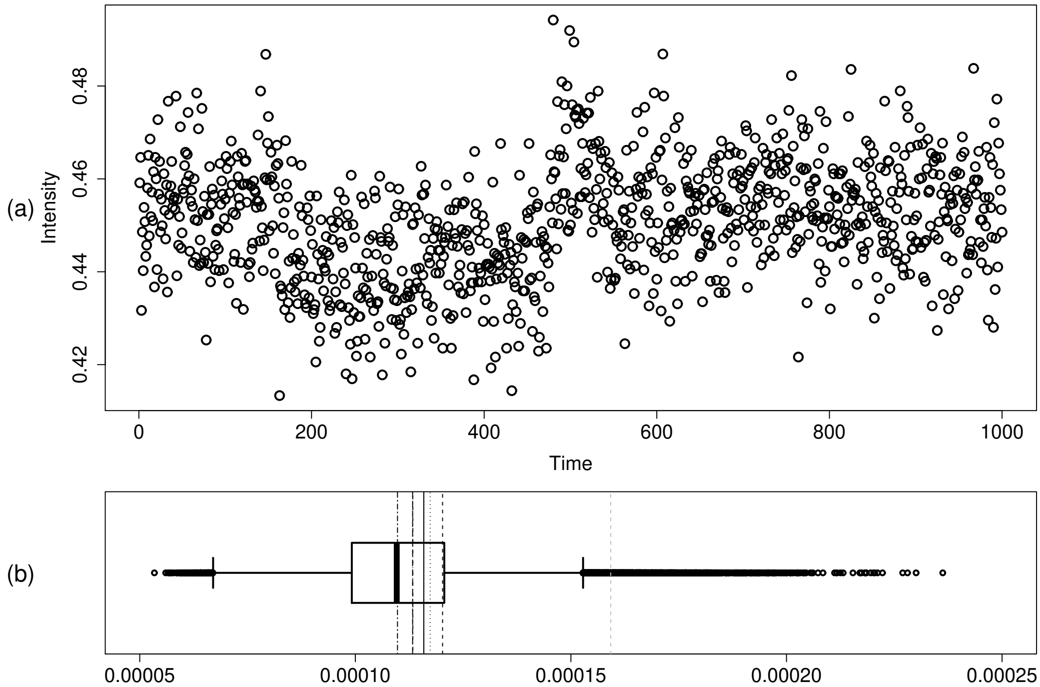

Figure 3.5: (a) Intensity over time for one pixel and (b) a boxplot of variances for the virus-free pixels together with the ordinary sample variance of the above data (- - -) and values of $\widehat{\sigma^2}_{\text{Mean}}$ and $\widehat{\sigma^2}_{\text{Diff}}$ (—), $\widehat{\sigma^2}_{\text{Tr},0.5}$ (- - -), $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$ (· · · ·), $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ (- · · -), $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ (− − −) and $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$ (− - −).

PAMONO data with trend

Again, we consider a PAMONO dataset, see Subsection 3.3.5. Panel (a) of Fig. 3.6 shows a time series corresponding to a pixel, which seems to exhibit a virus adhesion as well as a linear trend. Panel (b) of Fig. 3.6 shows the differenced data, i.e., $Y_t - Y_{t-1}$, $t = 2, ..., 388$. The differences of first order appear to be independent and scattered around a fixed mean. Few large differences can be observed, which presumably originate from the jumps in the mean at the corresponding time points. The existence of the trend could be explained by the fact that the surface, on which the fluid for virus adhesion is placed, was heated up over time. $N = 388$ observations are available. We apply the estimation procedures $\widehat{\sigma^2}_{\text{Mean}}$ (using $K \in \{1, ..., 5\}$ in the formula (3.8)), $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr},0.5}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{normal}}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$, $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ and $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$ to the data and get estimated values for the variance, which range from $\widehat{\sigma^2}_{\text{Mean}} = 0.95 \cdot 10^{-6}$ (with $K = 5$) to $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}} = 1.66 \cdot 10^{-6}$. The empirical variance of the observations has the value $26.48 \cdot 10^{-6}$, which is much larger than the other estimates. According to our experience the PAMONO data can be assumed to be uncorrelated after differencing. The sample variance of differenced data is $1.93 \cdot 10^{-6}$, which is an estimator of $2\sigma^2$, yielding the value $0.97 \cdot 10^{-6}$ as an estimate for σ^2, which is near the estimated value of $\widehat{\sigma^2}_{\text{Mean}}$.

We conclude that the proposed procedures yield reasonable results even in this situation, where a linear trend is present.

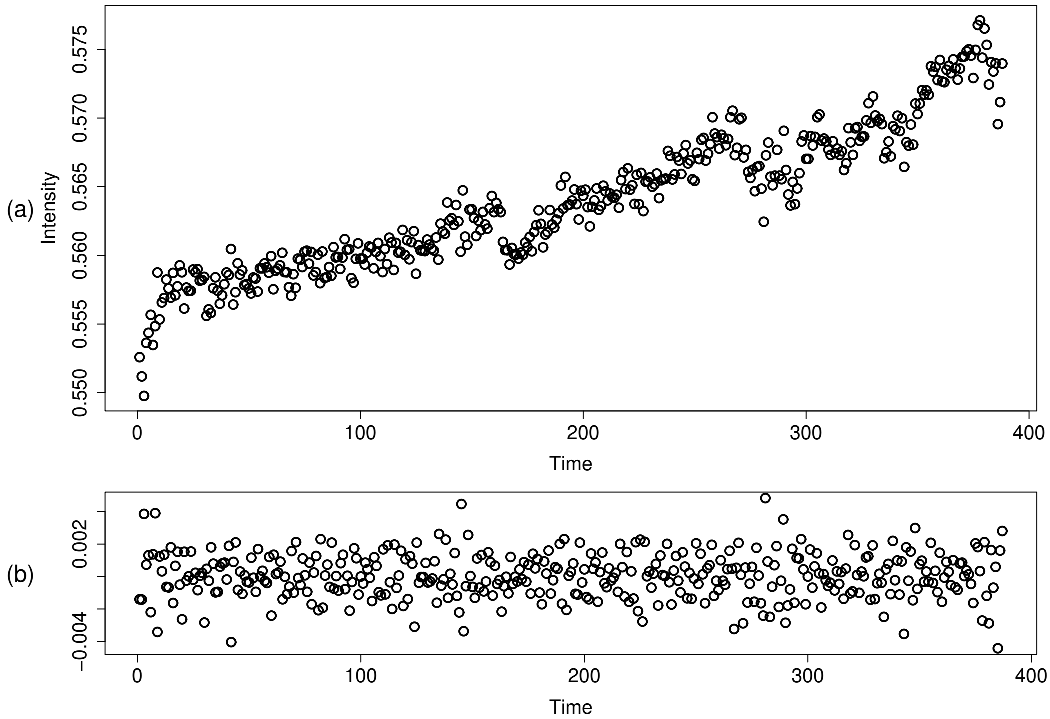

Figure 3.6: (a) Intensity over time for one pixel and (b) corresponding differenced series.

3.3.6 Conclusion

In the presence of level shifts ordinary variance estimators like the empirical variance perform poorly. In this section we considered several estimation procedures in order to account for possible changes in the mean.

Estimation of σ^2 based on pairwise differences is popular in nonparametric regression and works well in the presence of level shifts and an unknown error distribution if the data are independent and the fraction of shifts is asymptotically negligible. However, we have identified scenarios where estimation based on longer blocks is to be preferred.

If only a few small level shifts are expected in a long sequence of observations our recommendation is to use the mean of the blocks-variances $\widehat{\sigma^2}_{\text{Mean}}$. This estimation procedure does not require knowledge of the underlying distribution, performs well in the aforementioned situation and is asymptotically even as efficient as the ordinary sample variance if there are no level shifts.

If many or large level shifts are expected to occur we recommend using the adaptive trimmed estimators $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$ and $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$. These procedures are constructed for independent data and use either the exact χ^2-distribution or the asymptotic normal distribution of the blockwise estimates, where the second and the fourth moments need to be estimated. We have found these trimming approaches to work reasonably well even under moderate autocorrelations, although many blocks are trimmed away then, presumably due to the underestimation of the unknown variance in the formula (3.15). Therefore, when no changes in the mean are present the trimmed estimators suffer efficiency loss. On the other hand, we expect that many perturbed blocks are trimmed away in the presence of level shifts reducing the bias of the estimator. The trimming approach could be extended to dependent data in future work.

In many applications we rather wish to estimate the standard deviation σ, e.g. for standardization. If only few jumps of moderate heights are expected to occur, either the average value of the blockwise standard deviations or the square root of the blocks-variance estimator $\widehat{\sigma^2}_{\text{Mean}}$ can be used. Otherwise, the square root of the trimmed estimator $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$ can be recommended. For a large sample size N the finite sample correction factors can be neglected with little loss.

An interesting extension will be to consider situations where not only the level but also the variability of the data can change. Suitable approaches for such scenarios might be constructed by combining the ideas discussed here with those presented by [101], where tests for changes in variability have been investigated using blockwise approaches, assuming a constant mean. This will be an issue for future work.

3.4 Robust scale estimation under shifts in the mean

This section is based on the preprint "Robust scale estimation under shifts in the mean" [3]. We are grateful for the help of Alexander Dürre, one of the co-authors of the article, who gave lots of advice on mathematical tools to derive asymptotic theory for the proposed estimation procedure.

In Section 3.4.1 we present non-overlapping blocks approaches for estimating the scale measure σ in the change-point and outlier scenario. In Section 3.4.2 we discuss theoretical properties of the modified MAD. In Section 3.4.3 we discuss the choice of the block size for the blockwise estimators. Further methods for comparison are presented in Section 3.4.4. In Section 3.4.5 we analyse and compare the performance of the methods in a simulation study, and in Section 3.4.6 we apply them to a real data set. In Section 3.4.7, we give a brief summary and an outlook.

3.4.1 Robust measures of scale

There is a broad variety of literature on robust measures of scale. [76] introduced an estimator which is based on the shortest half of the observations. [74] proposed two alternatives to the well known median absolute deviation (MAD). The estimators S_n and Q_n use quantiles of absolute differences of all pairs of observations. Some further robust scale estimators are examined in [51]. Moreover, one can conduct an estimation of the variability in terms of M-, S- or τ-estimation procedures, see e.g. [13], [73] and [102]. Still, the MAD is presumably the most popular robust scale estimator.

[4] consider blockwise estimation procedures based on the sample variance to estimate σ^2 under shifts in the mean. [90] investigate a difference based approach under m-dependence when the underlying signal is discontinuous. A possible outlier contamination is not considered in both papers. Difference-based estimators have been widely discussed in the literature, see e.g. [94], [67], [26], [32], [21], [59], [91], [90]. Robust transformations of the vertical heights, formed by three consecutive observations, in order to deal with both, level shifts and outliers are considered in [75] and [27]. [6] use the median of the absolute differences of first or second order to estimate σ^2 under changes in the mean.

Our proposal is the estimator $\widehat{\sigma}_{\mathrm{M,mod}}$ of the standard deviation σ, a modification of the MAD, which will be investigated in this thesis. If jumps in the location are present, the median of the total sample is not a suitable estimator. The modified estimator $\widehat{\sigma}_{\mathrm{M,mod}}$ is based on the idea of estimating the location in blocks rather than using the whole sample. For this purpose the data is segregated into m blocks of size n, and in each block the median value is calculated. Subsequently, the corrected modified MAD is calculated as follows:

$$\widehat{\sigma}_{\mathrm{M,mod}} = C_N^{\mathrm{mod}} \widehat{\zeta}_{N,n} \quad \text{with} \tag{3.19}$$

$$\widehat{\zeta}_{N,n} = \mathrm{med}\{|Y_{1,1} - \widehat{\nu}_{N,1}|, \ldots, |Y_{1,n} - \widehat{\nu}_{N,1}|, |Y_{2,1} - \widehat{\nu}_{N,2}|, \ldots, |Y_{m,n} - \widehat{\nu}_{N,m}|\} \quad \text{and} \tag{3.20}$$

$$\widehat{\nu}_{N,j} = \mathrm{med}\{Y_{j,1}, \ldots, Y_{j,n}\},\ j = 1, \ldots, m, \tag{3.21}$$

where $Y_{j,1}, \ldots, Y_{j,n}$ are the n observations in the j-th block and C_N^{mod} is a distribution dependent finite sample correction factor, which can be simulated. Choosing $m = 1$ results in the ordinary (corrected) MAD.

There are other possibilities to utilize ordinary robust estimators, such as the MAD, to estimate the scale parameter in the change-point scenario and we will include such possibilities for comparison. The data can be segregated into m blocks, as was suggested above. A robust measure of scale, e.g. the MAD, can then be calculated in every block. Subsequently, the resulting m estimates $\mathrm{M}_1, \ldots, \mathrm{M}_m$ can be combined to get an overall estimate of scale. One possibility is averaging the block estimates, as was done e.g. in [16] in the context of repeated measurements, in [72] for estimation of the Hurst parameter or in [4] for non-robust estimation of the variance. The resulting estimator (average MAD) is:

$$\widehat{\sigma}_{\mathrm{M,me}} = C_N^{\mathrm{me}} \frac{1}{m} \sum_{j=1}^{m} \mathrm{M}_j. \tag{3.22}$$

Further possibilities to combine the block estimates are the median MAD and the trimmed MAD,

$$\widehat{\sigma}_{\mathrm{M,med}} = C_N^{\mathrm{M}} \mathrm{med}\left\{\mathrm{M}_1, \ldots, \mathrm{M}_m\right\} \quad \text{and} \tag{3.23}$$

$$\widehat{\sigma}^{\alpha}_{\mathrm{M,tr}} = C_N^{\mathrm{T}} \frac{1}{m - \lfloor \alpha m \rfloor} \sum_{j=1}^{m-\lfloor \alpha m \rfloor} \mathrm{M}_{(j)}, \tag{3.24}$$

where C_N^{M} and C_N^{T} are the finite sample correction factors. We do not consider a symmetric trimmed mean of the block estimates since the MAD is positively biased in the presence of level shifts. Therefore, we are only interested in trimming the block estimates, which are rather large.

The asymptotic properties of the proposed estimator $\widehat{\sigma}_{\mathrm{M,mod}}$ can be investigated straightforwardly in contrast to the estimators in (3.22) – (3.24). Moreover, the simulation results indicate that the performance of the three robust MAD-based estimators (3.19), (3.23) and (3.24) is similar, while the estimator (3.22) exhibits rather higher MSE in the presence of level shifts and outliers. Therefore, we recommend the estimator $\widehat{\sigma}_{\mathrm{M,mod}}$.

3.4.2 Theoretical properties of the modified MAD

We will now discuss some asymptotic properties of $\widehat{\zeta}_{N,n}$, the modified version of the ordinary MAD, see (3.20). Many theoretical results, which are valid for the ordinary MAD, can be adapted to the estimator $\widehat{\zeta}_{N,n}$ due to their similar construction. The theoretical results in this section make use of the contributions of [55], [79] and [80]. The proofs of the following results can be found in Appendix B.

Consistency

Let $X_1, \ldots, X_N$ be i.i.d. with continuous cumulative distribution function (CDF) F. We assume in the following that F is strictly monotone increasing around the median ν with

$\nu := F^{-1}(0.5)$. Let G be the CDF of $|X_1 - \nu|$ with

$$G(x) := P(|X_1 - \nu| \le x) = F(\nu + x) - F(\nu - x) \quad \forall x \in \mathbb{R}, \tag{3.25}$$

which is assumed to be strictly monotone increasing around the median

$$\zeta := G^{-1}(0.5). \tag{3.26}$$

We consider $\widehat{\zeta}_{N,n}$ as defined in (3.20), where m is the number of blocks and $n = \lfloor N/m \rfloor$ is the block size. The following Lemma 6 yields an exponential probability inequality for $\widehat{\zeta}_{N,n}$, which is later on used to show the almost sure convergence to the theoretical MAD in the change-point scenario (i.e., model (3.1) with $\gamma_t = 0\,\forall\, t$). The probability that the difference between the modified MAD $\widehat{\zeta}_{N,n}$ and the true value ζ is larger than ϵ is bounded by three terms, which all converge to zero.

Lemma 6. *Let $Y_t = X_t + \sum_{k=1}^{K} h_k I_{t \ge t_k}$, where the X_t are i.i.d. with $E(X_t) = 0$, $E(X_t^2) = \sigma^2$ and $h_k \ge 0$ for $k = 1, ..., K$ (see model (3.1) with $\gamma_t = 0$). We assume that the number of blocks m satisfies $m \to \infty$, $m = o(N)$ and $K = o(m)$, where $n = N/m$ is the number of observations in a block with $n \to \infty$. Then, for every $\epsilon > 0$:*

$$P\left(|\widehat{\zeta}_{N,n} - \zeta| > \epsilon\right) \le \exp\{-2N\Delta_{2,\epsilon,N}^2\} + \exp\{-2N\Delta_{3,\epsilon,N}^2\} + 4m\exp\left\{-2\frac{N}{m}\delta_{\epsilon,n}^2\right\} \tag{3.27}$$

with

$$\Delta_{2,\epsilon,N} = \left(F(\nu + \zeta + \epsilon/2) - F(\nu - \zeta - \epsilon/2) - \frac{B}{m} - \left\lfloor \frac{N+1}{2} \right\rfloor \frac{1}{N}\right)^+, \tag{3.28}$$

$$\Delta_{3,\epsilon,N} = \left\lfloor \frac{N+1}{2} \right\rfloor \frac{1}{N} - F(\nu + \zeta - \epsilon/2) + F(\nu - \zeta + \epsilon/2) - \frac{B}{m}, \tag{3.29}$$

$$\delta_{\epsilon,n} = \min\{a_0, b_0\}, \tag{3.30}$$

$$a_0 = a_0(\epsilon) = \left(F(\nu + \epsilon/2) - \left(\left\lfloor \frac{n+1}{2} \right\rfloor - 1\right)\frac{1}{n}\right)^+,$$

$$b_0 = b_0(\epsilon) = \left(\left\lfloor \frac{n+1}{2} \right\rfloor \frac{1}{n} - F(\nu - \epsilon/2)\right)^+,$$

where $x^+ = x$ if $x > 0$ and $x^+ = 0$ otherwise.

Proof. See Appendix B. □

Proposition 7. *Under the conditions of Lemma 6 $\widehat{\zeta}_{N,n}$ converges almost surely to ζ as $N \to \infty$.*

Proof. See Appendix B. □

Bahadur representation

The Bahadur representation for $\widehat{\zeta}_{N,n}$ (as defined in (3.20)) in the outlier- and jump-free scenario is given as follows (see proof of Theorem 8):

$$\widehat{\zeta}_{N,n} - \zeta = \frac{1/2 - \left(\widehat{F}_N(\nu+\zeta) - \widehat{F}_N(\nu-\zeta)\right)}{G'(\zeta)} - \frac{F'(\nu-\zeta) - F'(\nu+\zeta)}{G'(\zeta)} \frac{1/2 - \widehat{F}_N(\nu)}{F'(\nu)} + \Delta_N, \tag{3.31}$$

where $\widehat{F}_N$ is the empirical distribution function of the i.i.d. random variables $X_1, ..., X_N$, F is the CDF of X_1, G is the CDF of $|X_1 - \nu|$ as defined in (3.25), and ν, ζ are the population median and the population MAD, respectively. The first two terms of (3.31) are the same as in the case of the ordinary MAD and are dominating for $N \to \infty$, see [55].

In the following Theorem 8 we show the weak convergence of the error term Δ_N to zero with a convergence rate of at least $\sqrt{N}$.

Theorem 8. *(Weak Bahadur representation)*
Let $X_1, \dots, X_N$ be i.i.d. random variables and F, the CDF of X_1, be twice continuously differentiable with $F'(\nu) > 0$, $G'(\zeta) = F'(\nu+\zeta) + F'(\nu-\zeta) > 0$ and $F''(x) \leq M \, \forall x \in \mathbb{R}$, $M > 0$. When choosing $m = o\left(N^{1/3}\right)$ with $m \to \infty$ we then have for Δ_N from (3.31)

$$\Delta_N = o_P(N^{-1/2}). \tag{3.32}$$

Proof. See Appendix B. □

To prove Theorem 8 we will use the following Lemma, where the inequality for the absolute difference $\left|\left(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}\right) - (\nu+\zeta)\right|$, $j = 1, ..., m$, is presented. In contrast to [55], we show this inequality using the blockwise medians $\widehat{\nu}_{N,j}$ instead of the sample median of the entire data. Moreover, a slight modification of the constant D_1 in the proof of [55] is necessary in order to ensure the validity of the inequalities stated in the following Lemma. We consider $D_1 = \max\{8/F'(\nu), 8/G'(\zeta)\}$, while the minimum of the two values is used in [55].

Lemma 9. *Let $X_1, \dots, X_N$ be i.i.d. random variables and F, the CDF of X_1, twice continuously differentiable at x in the neighbourhood of $\nu \pm \zeta$, with $F'(\nu) > 0$ and $G'(\zeta) = F'(\nu+\zeta) + F'(\nu-\zeta) > 0$. Moreover, let $\widehat{\nu}_{N,j}$ be the median of the observations $X_{j,1}, ..., X_{j,n}$ in block $j \in \{1, \dots, m\}$, where m is the total number of blocks, which satisfies $m = o(N^{2/3})$ with $m \to \infty$, and n is the block size. Then, with*

$$D_1 = \max\{8/F'(\nu), 8/G'(\zeta)\}, \tag{3.33}$$

the following result is valid almost surely:

$$\left|\left(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}\right) - (\nu+\zeta)\right| \leq 2D_1 \frac{\log(n)^{1/2}}{n^{1/2}}, \tag{3.34}$$

$$\left|\left(\widehat{\nu}_{N,j} - \widehat{\zeta}_{N,n}\right) - (\nu-\zeta)\right| \leq 2D_1 \frac{\log(n)^{1/2}}{n^{1/2}}, \tag{3.35}$$

$j = 1, \dots, m$, for sufficiently large N.

Proof. See Appendix B. □

Corollary 2. *Under the conditions of Theorem 8 the limiting distribution of the estimator $\widehat{\zeta}_N$ is the same as that of the ordinary MAD in the outlier- and jump-free scenario when choosing $m = o(N^{1/3})$:*

$$\sqrt{N}\left(\widehat{\zeta}_{N,n} - \zeta\right) \xrightarrow{d} N(0, \vartheta^2), \tag{3.36}$$

where $\vartheta^2 = \frac{1}{4(G'(\zeta))^2}\left(1 + \frac{\gamma}{F'(\nu)^2}\right)$, with $\gamma = \beta^2 - 4(1-\alpha)F'(\nu)$, $\alpha = F(\nu - \zeta) + F(\nu + \zeta)$, $\beta = F'(\nu - \zeta) - F'(\nu + \zeta)$ and $\nu = F^{-1}(0.5)$, as defined in [55].

Proof. See Appendix B. □

Remark 11.

1. Under shifts in the mean the random variable $\sqrt{N}\left(\widehat{\zeta}_{N,n} - \zeta\right)$ diverges in distribution when choosing $n = \frac{N^{1-\delta}}{K+1}$, i.e., $m = N^{\delta}(K+1)$, with $\delta = 1/3.1$ (we choose m as large as possible without violating the necessary assumption $m = o(N^{1/3})$, since this should yield an estimator which is least influenced under shifts in mean.). As an example Figure 3.7 shows the histograms of 1000 simulated values of $\sqrt{N}\left(\widehat{\zeta}_{N,n} - \zeta\right)$ when dealing with three jumps of height $h = 5$ after $\tau_k N$ observations, with $\tau_k = k/4$ and $k \in \{1, 2, 3\}$. We observe that the mean of $\sqrt{N}\left(\widehat{\zeta}_{N,n} - \zeta\right)$ diverges as $N \to \infty$.

2. In Section 3.4.5 we will see that $\widehat{\sigma}_{\mathrm{M,mod}}$ with a block size n, which is a function of $\sqrt{N}$ (see (3.37)), yields very good results. However, simulations suggest that $\sqrt{N}\left(\widehat{\zeta}_{N,n} - \zeta\right)$ diverges in distribution under shifts in the mean, see Figure 3.8 for an example with $K = 3$ and $h = 5\sigma$.

3. The estimator $\widehat{\sigma}_{\mathrm{M,mod}}$ is affine equivariant, i.e., applying $\widehat{\sigma}_{\mathrm{M,mod}}$ to transformed observations $ay_1 + b, ..., ay_N + b$, $a, b \in \mathbb{R}$, results in multiplying the original estimate by the factor $|a|$.

4. Simulations suggest that the correction factor C_N^{mod} approaches the population correction factor $C_\infty^{\mathrm{mod}} = \sqrt{Var\,(X)}/\zeta$ with ζ as defined in (3.26), which equals $C_{\mathrm{MAD},\infty}$, the population correction factor of the MAD, if the block size n grows to infinity as $N \to \infty$. Table 3.10 shows the results for the block sizes $n = N^{1/2}/(K+1)$ and $n = N^{1-1/3.1}/(K+1)$ with $K = 0$.

	N(0,1)					t_5				
$q \setminus N$	500	1000	2500	5000	10000	500	1000	2500	5000	10000
$1/2$	1.5646	1.5362	1.5154	1.5060	1.4988	1.8573	1.8288	1.8084	1.8002	1.7922
$1-\frac{1}{3.1}$	1.5045	1.4988	1.4904	1.4886	1.4859	1.8075	1.7876	1.7854	1.7786	1.7790

Table 3.10: Correction factor C_N^{mod} for $n = N^q$ with $q \in \{1/2, 1-1/3.1\}$, $K = 0$ and $N \in \{500, 1000, 2500, 5000, 10000\}$. Data are generated from the normal ($C_{\mathrm{MAD},\infty} = 1.4826$) and the t_5-distribution ($C_{\mathrm{MAD},\infty} = 1.7765$).

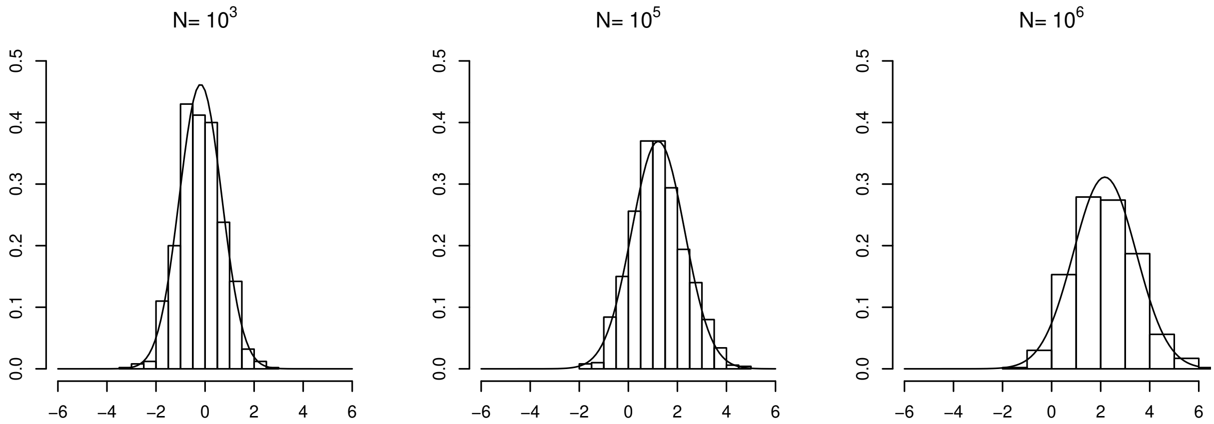

Figure 3.7: 1000 simulated values of $\widehat{\zeta}_{N,n}$ based on $N \in \{10^3, 10^5, 10^6\}$ observations from $Y_t = X_t + \sum_{k=1}^{3} 5 I_{t \geq t_k}$, where $X_t \sim \mathrm{N}(0,1)$ and $n = \frac{N^{1-1/3.1}}{K+1}$.

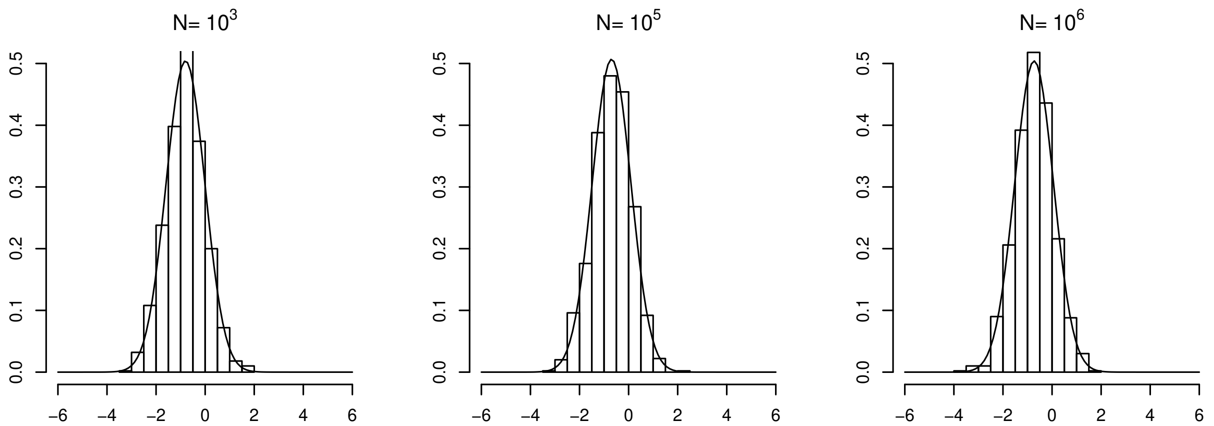

Figure 3.8: 1000 simulated values of $\widehat{\zeta}_{N,n}$ based on $N \in \{10^3, 10^5, 10^6\}$ observations from $Y_t = X_t$, where $X_t \sim \mathrm{N}(0,1)$, $n = \sqrt{N}$.

Robustness against outliers

We will consider the finite sample explosion breakdown point $\text{fsebp}_N(\widehat{\Theta}, N)$ of an estimator $\widehat{\Theta}$ given a sample $y_1, ..., y_N$. It is the minimum fraction of observations which can lead to arbitrarily large values of the estimator irrespective of the other data (see e.g. Section 3.2.5 in [54]). The fsebp of the sample variance or the standard deviation is equal to $1/N$, since arbitrarily large values can be achieved with a single outlier.

The asymptotic explosion breakdown point $\text{aebp}(\widehat{\Theta})$ is then defined as

$$\text{aebp}(\widehat{\Theta}) = \lim_{N\to\infty} \text{fsebp}_N(\widehat{\Theta}, N).$$

The ordinary MAD has the finite sample explosion breakdown point $\text{fsebp}(\text{MAD}, N) = \lfloor \frac{N+1}{2} \rfloor / N$, which results in the largest possible asymptotic explosion breakdown point $\text{aebp}(\text{MAD}) = 1/2$ for equivariant estimators (see [74] and Section 3.8.2 in [54]).

For the robust estimation procedures under level shifts we get the following results on the fsebp:

- Modified MAD
  For the modified MAD $\widehat{\zeta}_{N,n}$ (see (3.20)) we get $\text{fsebp}(\widehat{\zeta}_{N,n}, N) = \lfloor \frac{n+1}{2} \rfloor \lfloor \frac{m+1}{2} \rfloor / (mn)$, which yields $\text{aebp}(\widehat{\zeta}_{N,n}) = 1/4$ for any block size n. This is due to the fact that $\lfloor \frac{n+1}{2} \rfloor$ of the observations need to be contaminated in a block to achieve arbitrarily large values of the blockwise sample medians $\widehat{\nu}_{N,j}$ from (3.21), and $\lfloor \frac{m+1}{2} \rfloor$ blocks need to contain contaminated medians to achieve arbitrary values of the modified MAD.

- Average MAD
  To achieve arbitrarily large values of the average MAD as defined in (3.22) only one blockwise MAD needs to be perturbed. This is the case when $\lfloor \frac{n+1}{2} \rfloor$ outliers are present in one block. Therefore the aebp is $\lim_{N\to\infty} \lfloor \frac{n+1}{2} \rfloor / (mn) = 0$ if $m \to \infty$.

- Median MAD
  The median of the blockwise MAD's as defined in (3.23) has an aebp of 1/4, since $\lfloor \frac{m+1}{2} \rfloor$ blocks, each with $\lfloor \frac{n+1}{2} \rfloor$ outliers, are required to obtain an arbitrarily large estimate.

- Trimmed MAD
  Similar considerations can be made for the trimmed mean of the blockwise MAD's: $\lfloor \alpha m \rfloor + 1$ blocks, each with $\lfloor \frac{n+1}{2} \rfloor$ contaminated observations, cause the trimmed estimator to explode resulting in an aebp of $\alpha/2$.

  By choosing a different quantile than the median one could achieve a larger fsebp than 0.25 at the expense of a smaller efficiency and implosion break down point.

The aebp of the robust triangle-based estimators discussed in [77], which are considered in the simulation study for comparison (see Section 3.4.5), is 1/4, like for the modified and the median MAD.

Robustness against level shifts

Besides robustness against outliers we are also interested in the robustness against level shifts. For simplicity we will consider the case of even sample size N, block size n and number of blocks m.

The ordinary MAD can take arbitrarily large values then if a single level shift occurs after exactly 50% of the observations. Therefore, the smallest number of level shifts, which results in a severely biased estimate, is one. As opposed to this, the modified MAD can get arbitrarily large if half of the blocks are perturbed by a change in the mean, i.e., $\frac{m}{2}$ shifts.

To compare both quantities we will consider the smallest number of change-points needed to distort an estimate relative to the sample size and refer to this measure as the finite sample breakdown point under level shifts (fsbpul).

- Modified MAD
  We have $\text{fsbpul}(\widehat{\zeta}_{N,n}, N) = \frac{m}{2}/(mn) = 1/(2n)$ and therefore $\text{fsbpul}(\text{MAD}, N) = 1/(mn) = o\left(\text{fsbpul}(\widehat{\zeta}_{N,n}, N)\right)$ if $m \to \infty$ as $N \to \infty$.

- Average MAD
  For the average MAD we get an fsbpul of $1/(mn)$ since one perturbed block can lead to arbitrarily large estimates.

- Median MAD
  The fsbpul of the median MAD is the same as that for the modified MAD, i.e., $1/(2n)$.

- Trimmed MAD
  The trimmed MAD has a fsbpul of $(\lfloor \alpha m \rfloor + 1)/(mn)$, which is smaller than that of the modified and the median MAD if $\alpha < 0.5$.

For odd N, n or m the consideration of fsbpul gets somewhat more complicated. In case of odd n, for instance, the discussion above still applies if the zero deviation arising from the median itself is excluded from the outer median in the definition of the MAD.

3.4.3 Choice of the block size

In this section we investigate suitable choices for the block size of the robust estimators introduced in Section 3.4.1. The simulation scenarios are summarized in Table 3.11:

Distribution	$\text{N}(0,1)$, t_3
Sample size	$N \in \{1000, 2500\}$
Number of change-points	$K \in \{1, 2, 3, 4\}$
Jump height	$h \in \{1\sigma, 3\sigma, 5\sigma\}$
Average absolute outlier height	$g \in \{6\sigma, 10\sigma\}$
Number of simulation runs	1000

Table 3.11: Simulation scenarios for the choice of the block size.

Our goal is to determine an appropriate block size n, which performs well in all considered scenarios. The performance of the blocks-estimator $\widehat{\sigma}_{\mathrm{M,mod}}$ as well as that of $\widehat{\sigma}_{\mathrm{M,me}}$, $\widehat{\sigma}_{\mathrm{M,med}}$ and $\widehat{\sigma}^{\alpha}_{\mathrm{M,tr}}$ depends on the position of the jumps. We consider different positions of the K jumps to evaluate the performance of the blocks-estimators. For every $K \in \{1,2,3,4\}$, we generate K jumps of equal heights $h \in \{0, 1\sigma, 3\sigma, 5\sigma\}$ at positions sampled randomly from a uniform distribution on the values $\max_n (N - \lfloor N/n \rfloor n) + 1, \ldots, N - \max_n (N - \lfloor N/n \rfloor n)$ (if $N \neq mn$ then $N - mn$ observations are left out at the beginning and at the end of the sequence of observations) without replacement, and calculate the estimate for every block size $n \in \{2, 3, 4, \ldots, \lfloor N/2 \rfloor\}$. We repeat this procedure 1000 times and get 1000 estimates $\widehat{\sigma}_1, ..., \widehat{\sigma}_{1000}$ for each estimator $\widehat{\sigma}$ and every h and n. Subsequently, the MSE of $\widehat{\sigma}$ for given values of h and n is estimated using these values, i.e.,

$$\widehat{MSE}\left(\widehat{\sigma}\right) = \frac{1}{1000}\sum_{i=1}^{1000}\left(\widehat{\sigma}_i - \frac{1}{1000}\sum_{i=1}^{1000}\widehat{\sigma}_i\right)^2 + \left(\frac{1}{1000}\sum_{i=1}^{1000}\widehat{\sigma}_i - \sigma\right)^2.$$

Data are generated from different distributions, see Table 3.11.

Choice of the block size for $\widehat{\sigma}_{\mathrm{M,mod}}$

Figure 3.9 shows the simulated RMSE for the modified MAD in case of $N = 1000$ and different outlier and change-point scenarios. For both, the normal and the heavy tailed t_3 distribution, we observe a similar behaviour of the curves, except that in the latter case the RMSE is uniformly higher. Similar results are obtained for $N = 2500$, see Figure D.8 in Appendix D.

Larger blocks lead to smaller RMSE in case of a few small jumps, but the differences are not large in such scenarios. Shorter blocks are preferred as the number of change-points increases.

The square root of the sample size N has proven to be a good choice for the block length in many applications, see e.g. [4] (and Chapter 3.3) and [72]. We propose to choose the block size according to the formula

$$n_1 = n_1(N, K) = \max\left\{\left\lfloor \frac{\sqrt{N}}{K+1} \right\rfloor, 2\right\}, \tag{3.37}$$

if one is mainly interested in a good performance of the estimator. The corresponding estimator is denoted as $\widehat{\sigma}^{2}_{\mathrm{M,mod}}$. In the simulation study in Section 3.4.5 we will also investigate the performance of the modified MAD when choosing

$$n_2 = n_2(N, K) = \max\left\{\left\lfloor \frac{N^{1-1/3.1}}{K+1} \right\rfloor, 2\right\}, \tag{3.38}$$

in view of the results in Corollary 2, which requires $m = o(N^{1/3})$ for the convergence to the centered normal distribution. The corresponding estimator is denoted by $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$.

The number K of jumps in the mean is usually not known exactly. There are several possibilities to deal with this problem:

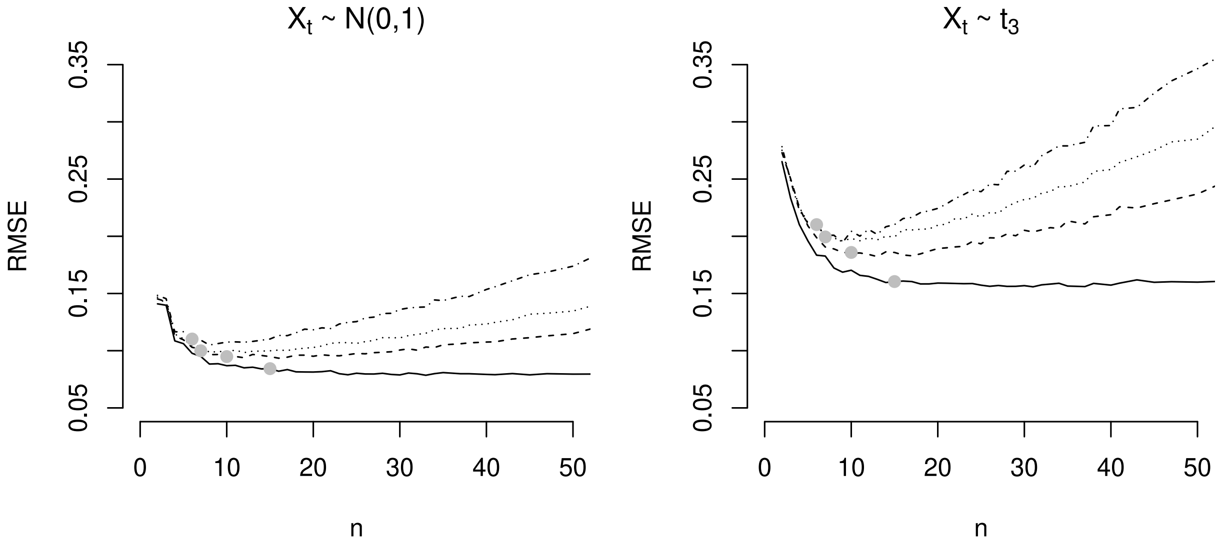

Figure 3.9: RMSE of $\widehat{\sigma}_{\mathrm{M,mod}}$ for $K = 1$, $h = 1\sigma$, $g = 6$ (——) $K = 2$, $h = 3\sigma$, $g = 6$ (- - -), $K = 3$, $h = 3\sigma$, $g = 10$ (···) and $K = 4$, $h = 5\sigma$, $g = 10$ (- · -). $N = 1000$ observations from $Y_t = X_t + \sum_{k=1}^{K} h\sigma I_{t \geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \mathrm{N}(g\sigma, 0.5)$, $X_t \sim \mathrm{N}(0,1)$ (left) and $X_t \sim \mathrm{t}_3$ (right). The grey dots denote the RMSE corresponding to the block size $n = \sqrt{N}/(K+1)$.

1. Use prior knowledge about the possible number of changes in the mean.

2. Pre-estimate the number of change-points with an appropriate procedure (e.g. robust regression trees, see [25]).

3. Choose the block size according to one of the rules

$$n_3 = n_1(N, \widetilde{K}) = \max\left\{\left\lfloor \frac{\sqrt{N}}{\widetilde{K}+1} \right\rfloor, 2\right\} \tag{3.39}$$

$$n_4 = n_2(N, \widetilde{K}) = \max\left\{\left\lfloor \frac{N^{1-1/3.1}}{\widetilde{K}+1} \right\rfloor, 2\right\} \tag{3.40}$$

with fixed $\widetilde{K}$, which do not require knowledge of the number of change-points. In Section 3.4.5 we will see that the modified MAD yields satisfying results when choosing n according to (3.39) and (3.40).

Remark 12. Choosing the smallest possible block size, i.e., $n = 2$, results in the median of absolute differences:

$$\widehat{\zeta}_{N,2} = \mathrm{med}\left\{\frac{|Y_2 - Y_1|}{2}, \frac{|Y_4 - Y_3|}{2}, \dots\right\}, \tag{3.41}$$

which is a median of i.i.d. random variables when no level shifts or outliers are present. A correction factor $c_{N,2}$ is required to ensure almost sure convergence of $\widehat{\zeta}_{N,2}$ to the population

MAD ζ. Estimators based on differences of observations have gained attention in the context of scale estimation, see e.g. [74], [90] or [27], where differences of first or second order are used. We will now show that the estimator $\widehat{\zeta}_{N,2}$ is less efficient than $\widehat{\zeta}_{N,n}$ under normality, with n chosen according to (3.38).

When $Y_1, \ldots, Y_N \sim \mathrm{N}(\mu, \sigma^2)$ i.i.d. we have that $|Y_{2j} - Y_{2j-1}|2^{-1/2} \sim \mathrm{HN}(\sigma)$, where $\mathrm{HN}(\sigma)$ denotes the half-normal distribution with parameter σ. The median of this distribution is $\sigma\sqrt{2}\mathrm{erf}^{-1}(1/2) = \Phi^{-1}(3/4) = \zeta$, where erf^{-1} is the inverse error function and Φ is the CDF of the standard normal distribution. Therefore, the appropriate correction factor for $\widehat{\zeta}_{N,2}$, as defined in (3.41), is $c_{N,2} = \sqrt{2}$ in order to achieve consistency. The asymptotic distribution of the sample median of $N/2$ i.i.d. random variables is well known, see e.g. Theorem 2.3.3A in [79]:

$$\begin{aligned} &\sqrt{N}\left(\sqrt{2}\widehat{\zeta}_{N,2} - \zeta\right) \xrightarrow{d} \mathrm{N}\left(0, \eta^2\right) \quad \text{with} \\ &\eta^2 = \frac{1}{2\left(f_{\frac{|Y_1 - Y_2|}{\sqrt{2}}}(\zeta)\right)^2}. \end{aligned}$$

E.g. for the standard normal distribution the random variables $|Y_{2j} - Y_{2j-1}|2^{-1/2}$ are $\mathrm{HN}(1)$-distributed and we get the asymptotic variance $\eta^2 = 1.24$. On the other hand, if the block size n is growing and the number of blocks m satisfies $m = o(N^{1/3})$ the asymptotic variance is $\vartheta^2 = 0.62$, which is much smaller (only half of the value) than in the case of blocks with fixed size $n = 2$.

Choice of the block size for further robust estimators

In this part of the thesis we will discuss an appropriate choice of the block size for further robust estimation procedures.

Choice of the block size for $\widehat{\sigma}_{\mathrm{M,me}}$

Since the ordinary MAD is a robust estimator we conclude that the height of the change-points and the magnitude as well as the amount of outliers might be not as crucial for the choice of the block size as the number of change-points is. Intuitively, the more changes in the mean occur the larger should the number of blocks be.

Figures 3.10 and D.9 confirm that the optimal block size n_{opt}, which yields the smallest RMSE, decreases as the number of jumps grows. We observe that the behaviour of the RMSE for the median MAD is similar to that of the modified MAD. Therefore we suggest choosing the block size according to the rule in (3.37).

Remark 13. The computation of the exact MSE requires the exact CDF of the MAD under a change in the mean. It can be derived analytically (see Appendix C for odd n and one change in the mean), using similar considerations as in the i.i.d. case, see e.g. [60]. However, the implementation and computation of the MSE is very time consuming and CPU-intensive, since numerical integration of products of distribution functions is involved in the computation. We conduct a simulation study instead to get an idea on the proper choice of the block size n.

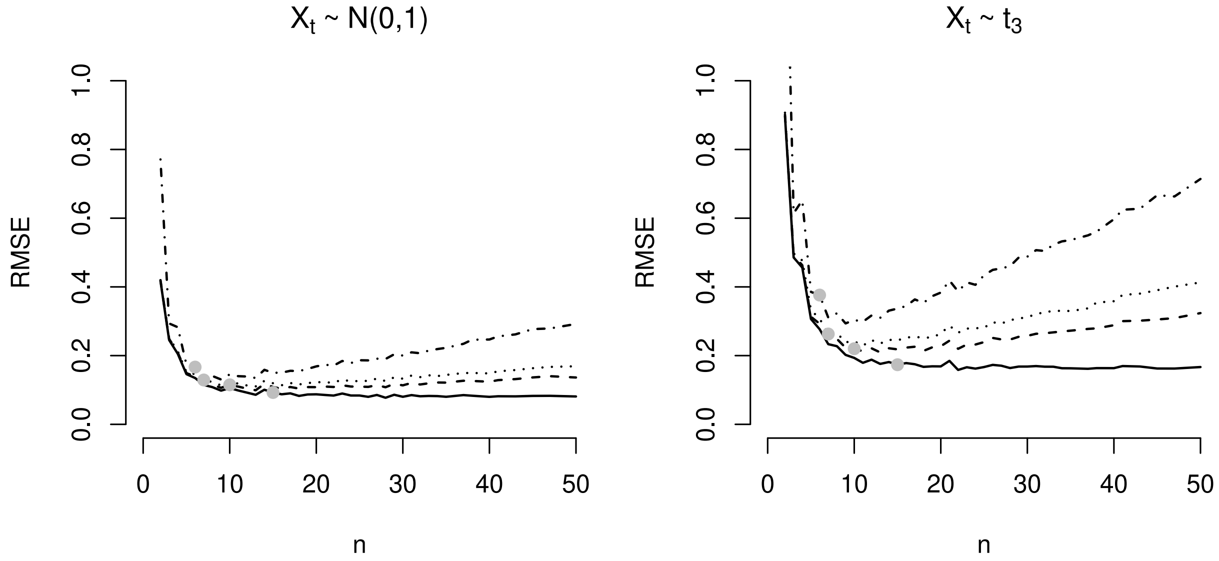

Figure 3.10: RMSE of $\widehat{\sigma}_{\mathrm{M,me}}$ for $K = 1$, $h = 1$, $\gamma = 6\sigma$ (——) $K = 2$, $h = 3$, $\gamma = 6\sigma$ (- - -), $K = 3$, $h = 3$, $\gamma = 6\sigma$ (• • •) and $K = 4$, $h = 5\sigma$, $\gamma = 10$ (- • -). $N = 1000$ observations from $Y_t = X_t + \sum_{k=1}^{K} h\sigma I_{t \geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \mathrm{N}(\gamma\sigma, 0.5)$, $X_t \sim \mathrm{N}(0,1)$ (left) and $X_t \sim \mathrm{t}_3$ (right). The grey dots denote the RMSE corresponding to the block size $n = \sqrt{N}/(K+1)$.

Choice of the trimming parameter α for $\widehat{\sigma}^{\alpha}_{\mathrm{M,tr}}$

For the estimator $\widehat{\sigma}^{\alpha}_{\mathrm{M,tr}}$ from (3.24) we will investigate the reasonable choice of the trimming parameter α first. Figure 3.11 shows the RMSE of $\widehat{\sigma}^{\alpha}_{\mathrm{M,tr}}$ for $\alpha \in \{0.1, 0.3, 0.5\}$ when $X_t \sim \mathrm{N}(0,1)$ or $X_t \sim \mathrm{t}_3$ (see model (3.1)). Clearly, $\alpha = 0.5$ yields the best results since more jump contaminated blocks are trimmed away. Hence, in what follows we will use $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$ with $\alpha = 0.5$.

Choice of the block size for $\widehat{\sigma}_{\mathrm{M,med}}$ and $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$

We will now investigate the proper choice of the block size n for the estimators $\widehat{\sigma}_{\mathrm{M,med}}$ and $\widehat{\sigma}^{\alpha}_{\mathrm{M,tr}}$. Figures 3.12 and D.10 in Appendix D show the RMSE of $\widehat{\sigma}_{\mathrm{M,med}}$ and $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$ in different scenarios where the X_t are either normally or t_3-distributed. The RMSE-curves for the estimator $\widehat{\sigma}_{\mathrm{M,med}}$ resemble those of $\widehat{\sigma}_{\mathrm{M,me}}$ and $\widehat{\sigma}_{\mathrm{M,mod}}$.

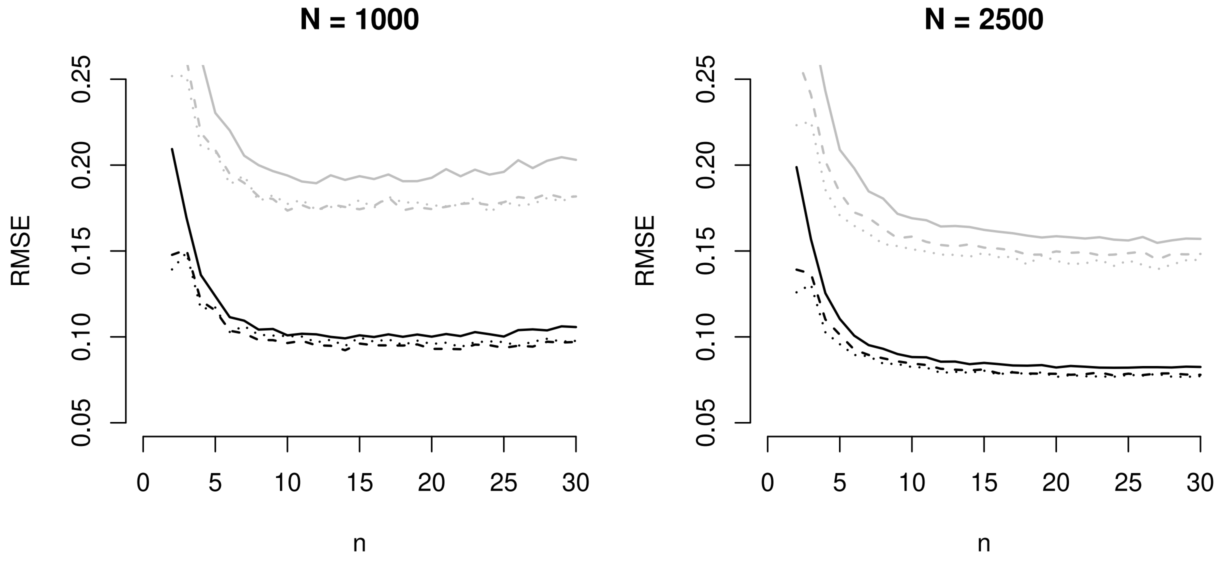

Figure 3.11: RMSE of $\widehat{\sigma}^{\alpha}_{\mathrm{M,tr}}$ for $\alpha = 0.1$ (——) $\alpha = 0.3$ (- - -), $\alpha = 0.5$ (· · ·). $N = 1000$ (left) and $N = 2500$ (right) observations from $Y_t = X_t + \sum_{k=1}^{3} 3\sigma I_{t \geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \mathrm{N}(6, 0.5)$, $X_t \sim \mathrm{N}(0, 1)$ (black) and $X_t \sim \mathrm{t}_3$ (grey).

Therefore, the block size can be chosen according to (3.37). For the estimator $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$ we suggest choosing the block size according to

$$n = n(N, K) = \max\left\{\left\lfloor\sqrt{\frac{N}{K+1}}\right\rfloor, 2\right\}, \tag{3.42}$$

since the impact of K seems to be slightly smaller in this case, see e.g. Figure D.10 in Appendix D. Moreover, we observe that $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$ performs slightly better than $\widehat{\sigma}_{\mathrm{M,med}}$. In Section 3.4.5 the performance of all estimators with the chosen block sizes will be investigated and compared with each other as well as with the robust scale estimators proposed by [27].

3.4.4 Further methods for comparison

The methods introduced in Section 3.4.1 will be compared with each other as well as with the approaches proposed by [4], [75] and [27] in a simulation study.

Now we review the blocks-estimators and the adaptively trimmed estimators of the standard deviation proposed in [4] (see Sections 3.3.4 and 3.3.2 for the details) as well as the three robust model-free approaches proposed in [75] and [27].

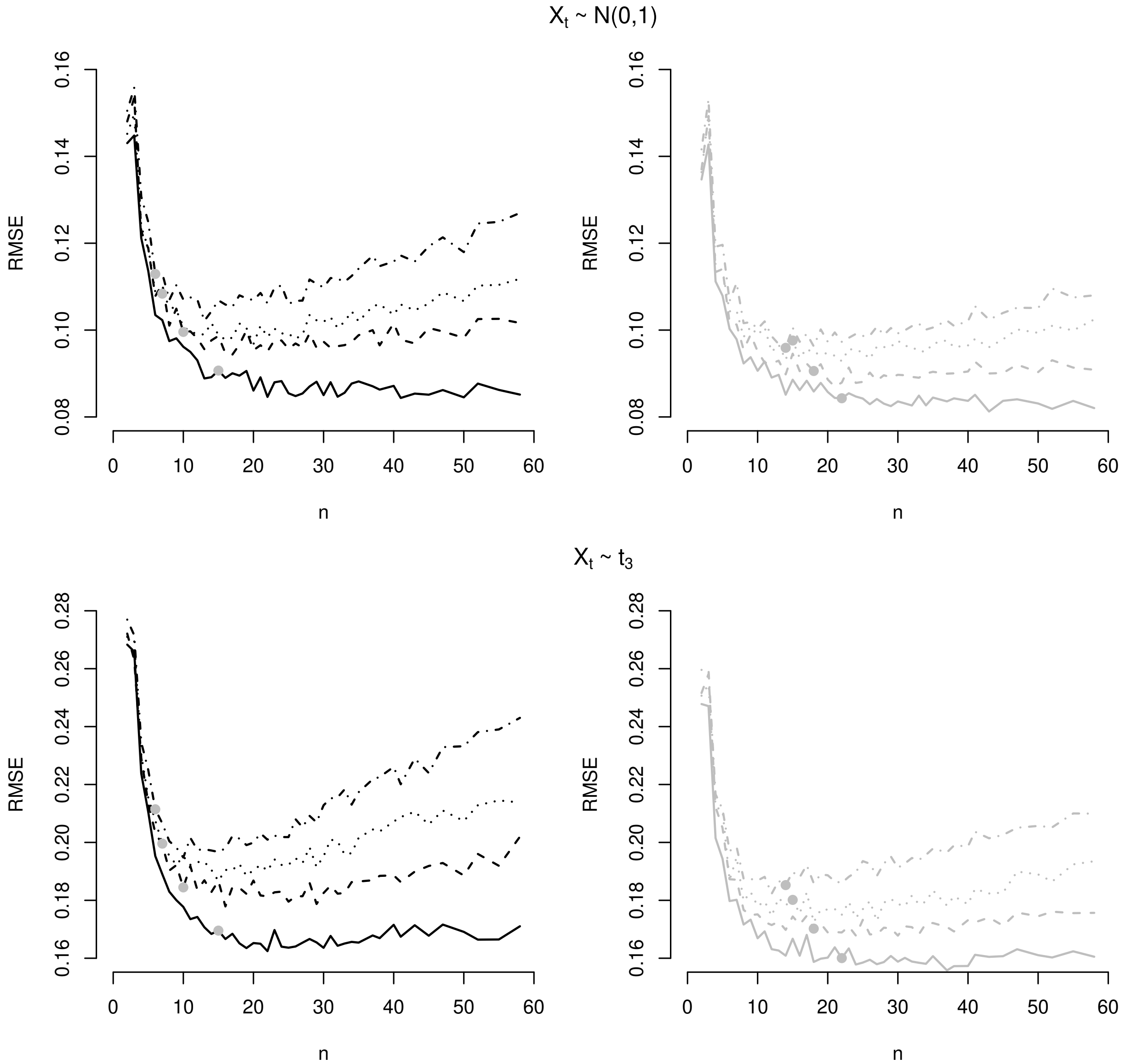

Figure 3.12: RMSE of $\widehat{\sigma}_{\mathrm{M,med}}$ (black) and $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$ (grey) for $K = 1$, $h = 1\sigma$, $g = 6$ (——) $K = 2$, $h = 3\sigma$, $g = 6$ (- - -), $K = 3$, $h = 3\sigma$, $g = 10$ (· · ·) and $K = 4$, $h = 5\sigma$, $g = 10$ (- · · -). $N = 1000$ observations from $Y_t = X_t + \sum_{k=1}^{K} h\sigma I_{t\geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \mathrm{N}(g\sigma, 0.5)$, $X_t \sim \mathrm{N}(0,1)$ (upper panel) and $X_t \sim \mathrm{t}_3$ (lower panel).

Blocks-estimators

The blocks-estimators proposed in [4] are defined as follows:

$$\widehat{\sigma}_{\mathrm{me},1} = C_{N,1}\frac{1}{m}\sum_{j=1}^{m} S_j \quad \text{and} \quad \widehat{\sigma}_{\mathrm{me},2} = C_{N,2}\sqrt{\frac{1}{m}\sum_{j=1}^{m} S_j^2}, \tag{3.43}$$

where S_j is the standard deviation in block j, $C_{N,1}$ and $C_{N,2}$ are the finite sample correction factors, which can be neglected for large N. We will choose n according to the rule (3.37).

Adaptively trimmed estimators

The adaptively trimmed estimator also introduced in [4] is the following:

$$\widehat{\sigma}_{\text{Tr,ad}} = C_{N,\text{Tr,ad}} \sqrt{\frac{1}{m - \lfloor \alpha_{\text{adapt}} m \rfloor} \sum_{j=1}^{m - \lfloor \alpha_{\text{adapt}} m \rfloor} S^2_{(j)}}, \tag{3.44}$$

where $S^2_{(1)} \leq \ldots \leq S^2_{(m)}$ are the ordered sample variances in the corresponding blocks and α_{adapt} the adaptively chosen percentage of the trimmed blocks-estimates. The term $C_{N,\text{Tr,ad}}$ is a finite sample correction factor, which can be neglected in large samples. The trimming parameter α_{adapt} is determined using the outlier detection procedure discussed in [18], since blocks containing a level shift will often lead to outlying sample variances. For more details see [4] and Section 3.3.2. We will refer to the adaptively trimmed estimator based on the normality assumption as $\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$. The trimmed estimator is denoted as $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$ if no distributional assumptions are made.

Robust triangle-based approaches

For the three robust model-free approaches first the $N-2$ adjacent triangle heights have to be calculated:

$$h_t^{\text{adj}} = \left| y_{t+1} - \frac{y_t + y_{t+2}}{2} \right|, \; t = 1, ..., N-2.$$

Then the estimators are given as follows:

$$Q^{\beta}_{adj} = c^q_n \{h_1^{\text{adj}}, \ldots, h_{N-2}^{\text{adj}}\}_{(\lfloor \beta(N-2) \rfloor)}, \tag{3.45}$$

$$M^{\beta}_{adj} = c^m_n \frac{1}{\lfloor \beta(N-2) \rfloor} \sum_{t=1}^{\lfloor \beta(N-2) \rfloor} h^{\text{adj}}_{(t)}, \tag{3.46}$$

$$MS^{\beta}_{adj} = c^s_n \sqrt{\frac{1}{\lfloor \beta(N-2) \rfloor} \sum_{t=1}^{\lfloor \beta(N-2) \rfloor} \left(h^{\text{adj}}_{(t)}\right)^2}. \tag{3.47}$$

We set $\beta = 0.5$ as was recommended by the authors (therefore $Q^{0.5}_{adj}, M^{0.5}_{adj}$ and $MS^{0.5}_{adj}$ are considered). The factors c^q_n, c^m_n and c^s_n are finite sample correction constants, which are required in order to achieve unbiasedness.

We will only consider Q^{β}_{adj} and MS^{β}_{adj} in the following since the performance of the estimator MS^{β}_{adj} is very similar to that of M^{β}_{adj}.

Further approaches

The triangle heights are based on differences of second order. We also consider estimation of scale based on the median of the first order differences. When using non-overlapping differences, the estimator is the following:

$$\widehat{\sigma}_{\mathrm{M,mod,2}} = \frac{\sqrt{2}}{\Phi^{-1}(0.75)}\widehat{\zeta}_{N,2} = \frac{\sqrt{2}}{\Phi^{-1}(0.75)}\mathrm{med}\left\{\frac{|Y_2 - Y_1|}{2}, \frac{|Y_4 - Y_3|}{2}, \ldots\right\}. \tag{3.48}$$

It is a special case of the estimator $\widehat{\sigma}_{\mathrm{M,mod}}$ with $n = 2$, which was already presented in (3.41). A finite sample correction factor will not be used since the samples considered in the further simulations are large and $\widehat{\sigma}_{\mathrm{M,mod,2}}$ is a consistent estimator of σ. The estimator based on overlapping differences is defined as follows:

$$\widehat{\sigma}_{\mathrm{Diff,med}} = C_N^{\mathrm{Diff}}\mathrm{med}\left\{|Y_2 - Y_1|, |Y_3 - Y_2|, \ldots\right\}, \tag{3.49}$$

where C_N^{Diff} is the finite sample correction factor, which can be simulated.

Finally, we consider a slight modification of the estimator $\widehat{\sigma}_{\mathrm{M,mod}}$ from (3.19) by using moving blocks of size n instead of non-overlapping (separate) ones. The corresponding overlapping MAD is defined as follows:

$$\begin{aligned} \widehat{\sigma}_{\mathrm{M,mod,\ over}} &= C_N^{\mathrm{mod,o}}\widehat{\zeta}_{N,n} \quad \text{with} \\ \widehat{\zeta}_{N,n,\ \mathrm{over}} &= \mathrm{med}\left\{|Y_t - \widehat{\nu}_{N,t}| : t = \left\lfloor\frac{n}{2}\right\rfloor + 1, \ldots, N - \left\lfloor\frac{n}{2}\right\rfloor\right\} \quad \text{and} \\ \widehat{\nu}_{N,t} &= \mathrm{med}\left\{Y_{t-\lfloor\frac{n}{2}\rfloor}, \ldots, Y_{t+\lfloor\frac{n}{2}\rfloor}\right\}, \; t = \left\lfloor\frac{n}{2}\right\rfloor + 1, \ldots, N - \left\lfloor\frac{n}{2}\right\rfloor, \end{aligned} \tag{3.50}$$

where $C_N^{\mathrm{mod,o}}$ is a distribution dependent finite sample correction factor, which can be simulated.

3.4.5 Simulations

We will now compare the estimators introduced in Section 3.4.1 with each other as well as with the robust triangle-based estimators (3.45) and (3.46) proposed by [75] and [27], the non-robust estimators (3.43), (3.44) and the robust difference-based estimators (3.48) and (3.49). In the simulation study we consider independent as well as positively correlated data.

Independent data

The simulation scenarios considered in case of independent data are summarized in Table 3.12. Throughout this thesis we will simulate the data from $\mathrm{N}(0,1)$, t_3, t_5, $\mathrm{Gum}(0,1)$ and $\mathrm{Lap}(1,3)$ distributions. The $\mathrm{Gum}(0,1)$ represents the Gumbel distribution with parameters $\mu = 0$ and $\sigma = 1$. The $\mathrm{Lap}(1,3)$ is the Laplace distribution with location parameter $\mu = 1$ and scale parameter $b = 3$.

Distribution	N(0, 1), t_3, t_5, Gum(0, 1), Lap(1, 3)
Sample size	$N \in \{500, 1000, 2500\}$
Number of change-points	$K \in \{1, 2, 3, 4\}$
Jump height	$h \in \{0, 1\sigma, 3\sigma, 5\sigma\}$
Outlier probability	$p = 0.05$
Average absolute outlier height	$g \in \{0, 6\sigma, 10\sigma\}$
Number of simulation runs	1000

Table 3.12: Simulation scenarios considered for independent data.

We generate 1000 datasets consisting of N observations from model (3.1) for each scenario and present the bias and the RMSE in Tables 3.13, 3.14 as well as in E.3, E.4 and E.5 in Appendix E. Again, the positions of the jumps are chosen randomly, as was done in Section 3.4.3.

Non-robust and difference-based methods

For the non-robust blocks-estimators (3.43), the adaptively trimmed estimator (3.44) and the robust difference-based estimators (3.48) and (3.49) we conduct some simulations to illustrate that these estimators do not perform well in comparison to the proposed robust estimation procedures (see Section 3.4.1) when level shifts and outliers are present. We only display the results for the modified MAD $\widehat{\sigma}_{\mathrm{M,mod}}$ from Section 3.4.1 for comparison, since we recommend using this approach in what follows. Table 3.13 shows the corresponding simulation results. We observe that these estimators yield very good results if zero or three jumps in the mean are present. If the data are additionally contaminated by outliers, the robust difference-based estimators (3.48) and (3.49) perform much better than the non-robust ones (3.43) and (3.44), but are outperformed by the modified MAD (and other robust estimators from Section 3.4.1, see Table 3.14 for comparison).

Moreover, we consider the overlapping MAD $\widehat{\sigma}_{\mathrm{M,mod,\ over}}$ from (3.50) for comparison. We observe that the performance of this estimator is very similar to that of the non-overlapping $\widehat{\sigma}_{\mathrm{M,mod}}$. Therefore, further investigation of $\widehat{\sigma}_{\mathrm{M,mod,\ over}}$ is omitted in our simulations.

Scenarios									
N		500			1000			2500	
K	0	3	3	0	3	3	0	3	3
ρ	0	0	0.5	0	0	0.5	0	0	0.5
g	0	0	6	0	0	6	0	0	6
RMSE									
$\widehat{\sigma}_{\text{me},1}$	3.52	5.41	42.02	2.46	3.68	46.99	1.56	1.98	53.46
$\widehat{\sigma}_{\text{me},2}$	3.31	4.65	69.62	2.22	3.09	69.33	1.42	1.86	68.48
$\widehat{\sigma^2}^{\text{normal}}_{\text{Tr,ad}}$	6.48	7.87	87.07	4.69	5.29	69.01	3.00	3.00	63.58
$\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$	7.22	7.53	83.79	5.14	5.32	72.42	3.27	3.17	66.00
$\widehat{\sigma}_{\text{M,mod},2}$	7.48	7.29	16.46	5.23	5.22	14.60	3.33	3.21	13.53
$\widehat{\sigma}_{\text{Diff,med}}$	5.86	5.72	15.34	4.19	4.11	14.01	2.65	2.55	13.26
$\widehat{\sigma}_{\text{M,mod,over}}$	5.60	8.73	13.40	3.84	5.09	10.00	2.37	2.81	8.05
$\widehat{\sigma}_{\text{M,mod}}$	5.37	7.25	12.45	3.88	4.54	10.58	2.32	2.67	8.17

Table 3.13: RMSE$\cdot 10^2$ of different estimators in different scenarios. Data are generated from the standard normal distribution.

Robust methods

Table 3.14 shows the results for $N = 1000$. We observe that the robust estimators, introduced in this thesis, perform well and lead to rather similar results. In the outlier-free change-point scenario ($K = 3$, $h = 3\sigma$, $g = 0$, $\rho = 0$) the estimators $\widehat{\sigma}_{\text{M,me}}$ and $\widehat{\sigma}_{\text{M,mod}}$ perform slightly better than $\widehat{\sigma}_{\text{M,med}}$ and $\widehat{\sigma}^{0.5}_{\text{M,tr}}$. The triangle-based robust estimators $Q^{0.5}_{adj}$ and $MS^{0.5}_{adj}$, proposed by [75] and [27], yield satisfying results as well.

In the presence of outliers and level shifts ($K = 3$, $h = 3\sigma$, $g = 6$, $\rho = 0.5$) robust combinations of block estimates, i.e., $\widehat{\sigma}_{\text{M,med}}$, $\widehat{\sigma}^{0.5}_{\text{M,tr}}$ or the modified MAD $\widehat{\sigma}_{\text{M,mod}}$ are preferable, since they outperform the approaches proposed by [75] and [27] as well as the estimator $\widehat{\sigma}_{\text{M,me}}$.

When the data is neither contaminated by structural changes nor by outliers ($K = 0$, $g = 0$, $\rho = 0$) $\widehat{\sigma}_{\text{M,mod}}$ yields slightly better results than $\widehat{\sigma}_{\text{M,med}}$ and $\widehat{\sigma}^{0.5}_{\text{M,tr}}$. We can also observe the similarity between the modified and the ordinary MAD, as expected.

The RMSE of the estimator $\widehat{\sigma}^{3.1}_{\text{M,mod}}$ is slightly higher than that of $\widehat{\sigma}^{2}_{\text{M,mod}}$ if changes in the mean are present. This can be explained by the fact that less blocks are used when n is chosen according to the rule in (3.38). E.g., when $N = 1000$ and $K \in \{0, 1, \dots, 5\}$ the resulting number of blocks m ranges from 18 to 55, while we have $m \in [63, 189]$ when choosing n according to (3.37).

Tables E.3 and E.4 in Appendix E show similar results for $N = 500$ and $N = 2500$. Results for the case of $N = 500, 1000, 2500$ with $K = 5$ and $h = 5\sigma$ can be found in Table E.5 in Appendix E, where the same statements on the performance of the estimators can be made. In all Tables 3.14, E.3, E.4 and E.5 we observe that the bias is the dominating part of the RMSE when changes in the mean are present which are not very small.

Distribution	Sd	MAD	$\widehat{\sigma}_{\mathrm{M,me}}$	$\widehat{\sigma}_{\mathrm{M,med}}$	$\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$	$\widehat{\sigma}^{2}_{\mathrm{M,mod}}$	$\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$	$Q^{0.5}_{adj}$	$MS^{0.5}_{adj}$
				$K = 0,\ g = 0,\ \rho = 0$					
Bias									
$\mathrm{N}(0,1)$	-0.00	0.09	0.44	0.35	0.30	0.13	0.06	-0.06	-0.09
t_3	-1.57	-0.14	0.43	0.40	0.02	0.01	-0.02	-0.14	-0.20
t_5	-0.00	0.21	-0.08	0.63	0.66	0.22	0.29	0.15	0.14
$\mathrm{Gum}(0,1)$	0.14	-0.10	0.23	0.04	0.06	0.09	-0.07	0.47	0.61
$\mathrm{Lap}(1,3)$	1.19	0.75	-0.20	-0.20	0.53	1.25	1.63	0.70	0.67
RMSE									
$\mathrm{N}(0,1)$	2.22	3.77	3.84	4.68	4.94	3.88	3.80	4.51	4.89
t_3	21.34	6.96	7.20	8.58	8.67	7.10	7.05	8.69	9.13
t_5	5.51	4.94	5.28	6.46	6.60	5.33	4.97	6.00	6.37
$\mathrm{Gum}(0,1)$	4.11	4.98	5.18	6.48	6.34	5.28	5.01	6.11	6.61
$\mathrm{Lap}(1,3)$	16.05	19.02	19.27	23.41	23.52	19.77	19.30	20.87	22.14
				$K = 3,\ h = 3\sigma,\ g = 0,\ \rho = 0$					
Bias									
$\mathrm{N}(0,1)$	214.63	209.55	1.28	0.69	1.22	0.85	2.92	0.47	0.40
t_3	377.86	606.14	2.78	1.24	1.93	1.31	6.44	0.87	0.82
t_5	271.78	326.95	1.94	1.24	1.20	1.13	4.49	0.42	0.38
$\mathrm{Gum}(0,1)$	278.41	316.36	2.28	0.82	1.25	1.08	3.91	0.65	0.56
$\mathrm{Lap}(1,3)$	903.11	$> 10^3$	4.24	4.08	4.02	5.06	17.29	2.28	2.28
RMSE									
$\mathrm{N}(0,1)$	221.65	237.89	4.34	5.38	5.15	4.54	4.99	4.56	4.94
t_3	390.60	698.62	8.57	9.94	9.86	8.25	10.54	8.73	9.09
t_5	281.35	377.98	5.88	7.21	6.97	6.15	7.26	5.93	6.36
$\mathrm{Gum}(0,1)$	287.69	361.87	6.18	7.46	6.82	6.25	6.76	6.05	6.41
$\mathrm{Lap}(1,3)$	934.63	$> 10^3$	21.44	27.37	25.57	23.31	28.46	21.82	22.87
				$K = 3,\ h = 3\sigma,\ g = 6,\ \rho = 0.5$					
Bias									
$\mathrm{N}(0,1)$	243.84	223.81	11.95	9.31	8.06	9.25	10.00	19.19	17.86
t_3	429.86	640.60	24.47	16.74	14.68	17.09	20.00	37.88	33.78
t_5	312.15	358.90	16.66	12.22	10.46	12.33	14.04	26.46	24.24
$\mathrm{Gum}(0,1)$	313.20	332.12	16.21	11.25	9.76	11.50	12.83	25.53	23.40
$\mathrm{Lap}(1,3)$	$> 10^3$	$> 10^3$	55.32	40.80	37.45	43.16	52.97	91.55	82.38
RMSE									
$\mathrm{N}(0,1)$	249.36	251.94	12.92	11.10	9.91	10.58	10.96	20.15	18.96
t_3	439.78	729.40	26.58	20.03	17.97	19.51	21.96	39.80	35.84
t_5	319.43	404.61	18.02	14.64	12.96	14.04	15.43	27.80	25.67
$\mathrm{Gum}(0,1)$	320.15	372.46	17.60	13.80	12.42	13.34	14.33	26.86	24.88
$\mathrm{Lap}(1,3)$	$> 10^3$	$> 10^3$	61.07	50.15	46.61	49.81	58.10	96.19	87.38

Table 3.14: Bias$\cdot 10^2$ and RMSE$\cdot 10^2$ of different estimators for $N = 1000$ under independence.

Therefore, our recommendation of the modified version of the MAD has been confirmed in the simulation study. If one is only interested in estimating the scale parameter properly, the estimator $\widehat{\sigma}^2_{\mathrm{M,mod}}$ with the block size $n = \max\left\{\frac{\sqrt{N}}{K+1}, 2\right\}$ is a good choice. When testing hypotheses on σ is in the focus of the application, $n = \max\left\{\frac{N^{1-1/3.1}}{K+1}, 2\right\}$ can be chosen.

Block size independent of K

The above choice of the block size n depends on the number K of change-points, see rules (3.37) and (3.38). However, K is usually not known. Therefore, we investigate the performance of the modified MAD choosing a block size, which does not depend on the unknown value of K. Figure 3.13 shows the RMSE of the estimators $\widehat{\sigma}^2_{\mathrm{M,mod}}$ and $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$ using block sizes (3.37) and (3.38), respectively, together with the estimators $\widetilde{\widetilde{\sigma}}^2_{\mathrm{M,mod}}$ and $\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$ using block sizes $n = \max\{\lfloor\sqrt{N}/(\widetilde{K}+1)\rfloor, 2\}$ and $n = \max\{N^{1-1/3.1}/(\widetilde{K}+1), 2\}$ with $\widetilde{K} \in \{0, 1, 2, 4, 6\}$ (i.e., setting $K = 0, 1, 2, 4, 6$ in the formulae (3.37) and (3.38)), which do not depend on the true number of jumps $K \in \{1, 3, 5, 7, 9\}$.

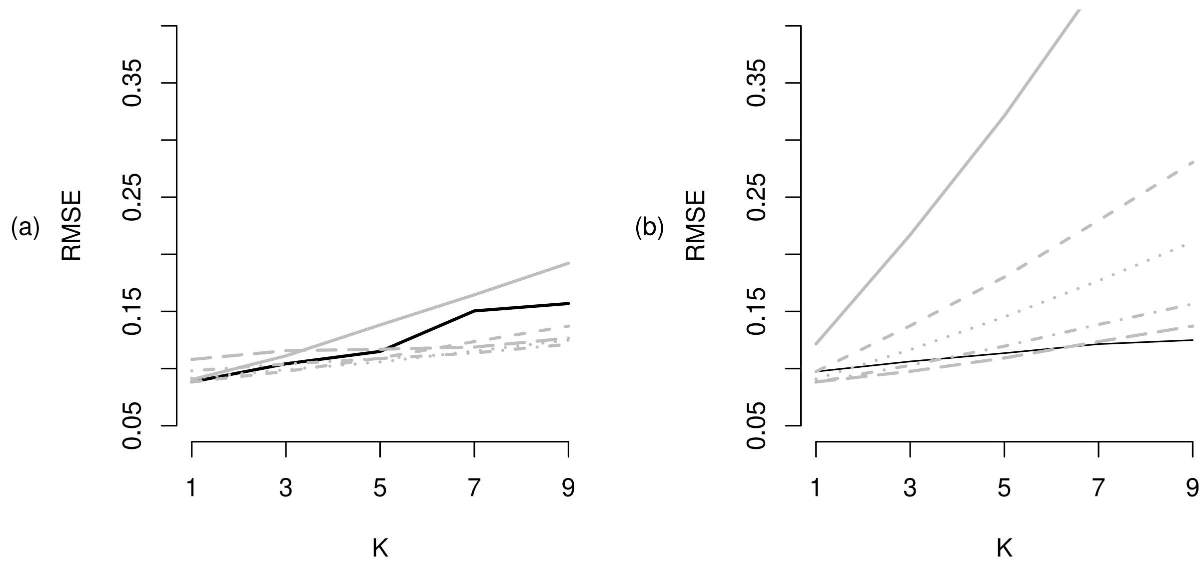

Figure 3.13: RMSE of (a) $\widehat{\sigma}^2_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor\sqrt{N}/(K+1)\rfloor, 2\}$ and $\widetilde{\widetilde{\sigma}}^2_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor\sqrt{N}/(\widetilde{K}+1)\rfloor, 2\}$, (b) $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(K+1)\rfloor, 2\}$ and $\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(\widetilde{K}+1)\rfloor, 2\}$ based on true K (——), and on $\widetilde{K} = 0$ (——), $\widetilde{K} = 1$ (- - -), $\widetilde{K} = 2$ (· · ·), $\widetilde{K} = 4$ (- · -), $\widetilde{K} = 6$ (– – –) for $N = 1000$ observations from $Y_t = X_t + \sum_{k=1}^{K} 3\sigma I_{t\geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \mathrm{N}(6\sigma, 0.5)$, $X_t \sim \mathrm{N}(0, 1)$.

We consider data from the normal distribution with 5% outliers, where $N = 1000$, $h = 3$ and $K = 3$. We observe that the performance of the estimators $\widehat{\sigma}^2_{\mathrm{M,mod}}$ and $\widetilde{\widetilde{\sigma}}^2_{\mathrm{M,mod}}$ (independent of K) is similar for any fixed $\widetilde{K}$, while $\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$ (with fixed $\widetilde{K}$) yields much higher RMSE values than the estimator $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$ if the fixed value $\widetilde{K}$ is low (e.g. $\widetilde{K} = 0$). We conclude that in this situation choosing a large value of $\widetilde{K}$, such as $\widetilde{K} = 6$, is beneficial. Similar conclusions can be made for the case $N = 2500$ with $h = 3$ and $N = 1000$ with $h = 5$ for normally distributed data and the case $N = 1000$ with $h = 3$ for t_3-distributed data, see Figure D.11 and the upper panel of Figure D.12 in Appendix D.

The results for the small jump height $h = 1$ and the normal distribution are given in the lower panel of Figure D.12 in Appendix D. In this case estimation based on any block size with fixed $\widetilde{K}$ yields similar results. We conclude that the block size $n = \sqrt{N}/(\widetilde{K}+1)$ with a fixed value $\widetilde{K}$, e.g. $\widetilde{K} = 2$, is an appropriate choice if the value of K is not known. If the focus is on testing hypotheses on σ (in light of Corollary 2) a block size $n = \max\{N^{1-1/3.1}/(\widetilde{K}+1), 2\}$ with a rather large value $\widetilde{K} \geq 6$ should be chosen.

Correlated data

We also consider positively correlated data generated by the first order autoregressive (AR) model with normal errors and parameter $\phi \in \{0, 0.1, \ldots, 0.9\}$, which describes the strength of the correlation. Independent normally distributed data are considered when $\phi = 0$. Table 3.15 shows the simulation results. The bias and the RMSE of the ordinary sample standard deviation and the MAD are not displayed in the table, since these values are much larger than the rest. We observe that high positive autocorrelation causes a negative bias, i.e., the parameter σ is underestimated. On the other hand the estimators are positively biased when level shifts and outliers are present. This explains why the absolute value of the bias decreases for moderate autocorrelation and increases as ϕ grows.

If the correlation is moderately high the performance of the proposed robust estimators does not worsen much. Choosing the block size $n = \max\{\lfloor\sqrt{N}/(2+1)\rfloor, 2\}$ instead of $n = \max\{\lfloor\sqrt{N}/(K+1)\rfloor, 2\}$, i.e., considering the estimator $\widetilde{\widetilde{\sigma}}^{2}_{\mathrm{M,mod}}$, improves the performance slightly for higher values of ϕ, since a larger block size is beneficial when dealing with strongly correlated data. Even better results are obtained by $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$ with the block size $n = \max\{\lfloor N^{1-1/3.1}/(K+1)\rfloor, 2\}$ for large ϕ. Choosing a block size $n = \max\{\lfloor N^{1-1/3.1}/(6+1)\rfloor, 2\}$ (i.e., considering $\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$) does not improve the performance of the estimator considerably. The trimmed estimator $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$ yields slightly better results than the modified MAD for small values of ϕ. For large values of ϕ the modified MAD is a better choice.

Similar results are obtained for $N = 2500, h = 5, K = 5$ and $N = 1000, h = 3, K = 3$, see Tables E.6 and E.7 in Appendix E.

ϕ	$\widehat{\sigma}_{\mathrm{M,me}}$	$\widehat{\sigma}_{\mathrm{M,med}}$	$\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$	$\widehat{\sigma}^{2}_{\mathrm{M,mod}}$	$\stackrel{\approx}{\sigma}^{2}_{\mathrm{M,mod}}$	$\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$	$\stackrel{\approx}{\sigma}^{3.1}_{\mathrm{M,mod}}$	$Q^{0.5}_{adj}$	$MS^{0.5}_{adj}$
				$K = 5,\ h = 5\sigma,\ g = 0,\ \rho = 0$					
Bias									
0	1.98	1.66	1.69	1.47	2.53	3.84	3.25	1.17	1.17
0.1	-0.17	-0.99	0.68	-1.11	1.22	3.03	2.60	-5.51	-5.64
0.3	-6.08	-7.48	-2.79	-7.56	-1.83	1.04	0.32	-20.58	-20.62
0.5	-16.48	-18.93	-9.56	-19.16	-8.70	-3.12	-4.46	-39.96	-39.99
0.7	-40.22	-43.83	-28.04	-44.19	-26.86	-15.84	-18.35	-72.12	-72.13
0.9	-128.33	-133.81	-109.35	-134.24	-107.18	-84.26	-90.35	-168.08	-168.19
RMSE									
0	4.63	5.91	5.27	5.10	5.11	5.73	5.43	4.59	4.94
0.1	4.30	5.71	4.95	5.08	4.42	5.32	5.00	6.90	7.19
0.3	7.39	9.34	5.84	9.00	4.96	4.56	4.40	20.87	20.96
0.5	17.07	19.74	11.04	19.77	9.95	5.78	6.64	40.07	40.12
0.7	40.50	44.23	28.75	44.50	27.45	17.04	19.36	72.18	72.19
0.9	128.43	133.95	109.61	134.34	107.40	84.71	90.74	168.10	168.21
				$K = 5,\ h = 5\sigma,\ g = 6,\ \rho = 0.5$					
Bias									
0	16.07	10.86	8.76	10.43	10.44	11.16	10.87	20.19	18.81
0.1	13.33	8.11	7.63	7.74	9.10	10.43	9.99	12.53	11.14
0.3	6.86	1.02	4.32	0.59	5.46	8.30	7.56	-4.09	-5.43
0.5	-3.00	-10.28	-2.88	-10.94	-0.95	4.73	3.07	-24.96	-26.19
0.7	-25.98	-34.94	-20.10	-35.85	-18.30	-6.70	-9.67	-58.12	-59.28
0.9	-111.72	-124.37	-100.62	-125.41	-97.92	-72.83	-79.57	-155.55	-156.63
RMSE									
0	17.06	12.61	10.40	11.83	11.46	12.07	11.86	21.07	19.82
0.1	14.46	10.39	9.47	9.57	10.29	11.45	11.08	13.85	12.70
0.3	8.82	6.48	7.14	5.46	7.39	9.69	9.09	6.58	7.51
0.5	6.69	12.18	6.58	12.29	5.51	7.41	6.37	25.34	26.55
0.7	26.82	35.61	21.30	36.31	19.27	9.77	11.71	58.26	59.43
0.9	112.03	124.54	100.94	125.54	98.18	73.46	80.06	155.59	156.67

Table 3.15: Bias$\cdot 10^2$ and RMSE$\cdot 10^2$ of different estimators for $N = 1000$, $K = 5$, $h = 5\sigma$ and different AR-parameters $\phi \in \{0, 0.1, \ldots, 0.9\}$.

3.4.6 Application

In this section we apply the robust estimation techniques to a dataset in order to estimate the standard deviation.

Data with level shifts

We analyse data obtained from the PAMONO (Plasmon Assisted Microscopy of Nano-Size Objects) biosensor. For more details regarding the data, see Section 3.3.5 as well as [83] and [84].

In Panel (a) of Fig. 3.14 a time series of length $N = 1000$ corresponding to one pixel with a virus adhesion is shown. Three level shifts in the mean of the time series can be observed. Panel (b) of Fig. 3.14 presents a boxplot of 101070 standard deviations for time series, which correspond to pixels without virus adhesion, i.e., without level shifts. We use these data to get an idea about the range of typical values of the standard deviation. The values of the sample standard deviation and the ordinary MAD of the contaminated data (upper panel) are not within this typical range and exceed the upper whisker of the boxplot. As opposed to this, the non-robust and robust blocks-estimators $\widehat{\sigma}_{\mathrm{me},2}$, $\widehat{\sigma}_{\mathrm{Tr,ad}}$, $\widehat{\sigma}_{\mathrm{M,mod},2}$, $\widehat{\sigma}_{\mathrm{Diff,med}}$, $\widehat{\sigma}_{\mathrm{M,me}}$, $\widehat{\sigma}_{\mathrm{M,med}}$, $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$, $\widehat{\sigma}^{2}_{\mathrm{M,mod}}$ and $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$, $\tilde{\tilde{\sigma}}^{2}_{\mathrm{M,mod}}$, $\tilde{\tilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$ yield values, which are well within the interquartile range. More precisely, the modified MAD yields values from $1.05 \cdot 10^{-2}$ to $1.09 \cdot 10^{-2}$. We conclude that the blocks-approach yields reasonable estimates for these data.

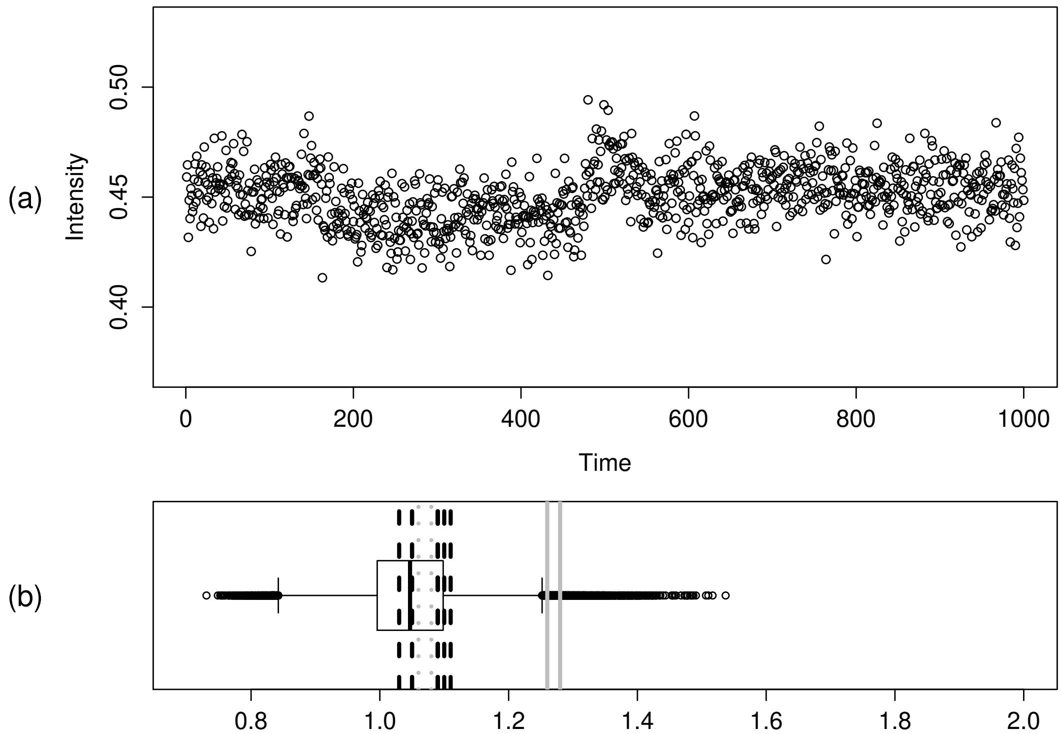

Figure 3.14: (a) Intensity over time for one pixel; (b) A boxplot of standard deviations (multiplied by 10^2) for the virus-free pixels together with the ordinary sample standard deviation and MAD (——), the non-robust estimators $\widehat{\sigma}_{\mathrm{me},2}$ and $\widehat{\sigma}_{\mathrm{Tr,ad}}$ ($\cdots$) and the robust estimators $\widehat{\sigma}_{\mathrm{M,mod},2}$, $\widehat{\sigma}_{\mathrm{Diff,med}}$, $\widehat{\sigma}_{\mathrm{M,me}}$, $\widehat{\sigma}_{\mathrm{M,med}}$, $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$, $\widehat{\sigma}^{2}_{\mathrm{M,mod}}$, $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$, $\tilde{\tilde{\sigma}}^{2}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor \sqrt{N}/(2+1) \rfloor, 2\}$ and $\tilde{\tilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(6+1) \rfloor, 2\}$ (- - -) applied to the above data in (a).

Data with level shifts and outliers

Panel (a) of Fig. 3.15 shows the PAMONO data where additive outliers with an occurrence probability of 5% have been added to the observations. The height of the outliers is chosen to be normally distributed with mean $6 \cdot \overline{\mathrm{Std}}$ and a standard deviation 0.0005, where $\overline{\mathrm{Std}}$ denotes the average value of the standard deviations for PAMONO time series without level shifts. The sample standard deviation, the MAD as well as the non-robust estimators $\widehat{\sigma}_{\mathrm{me},2}$ and $\widehat{\sigma}_{\mathrm{Tr,ad}}$ apparently overestimate the standard deviation of the data strongly as the corresponding estimates are much higher than the upper whisker of the boxplot. The robust estimates are smaller than the upper whisker of the boxplot and look more plausible. The modified MAD yields the smallest values ranging from $1.17 \cdot 10^{-2}$ to $1.20 \cdot 10^{-2}$.

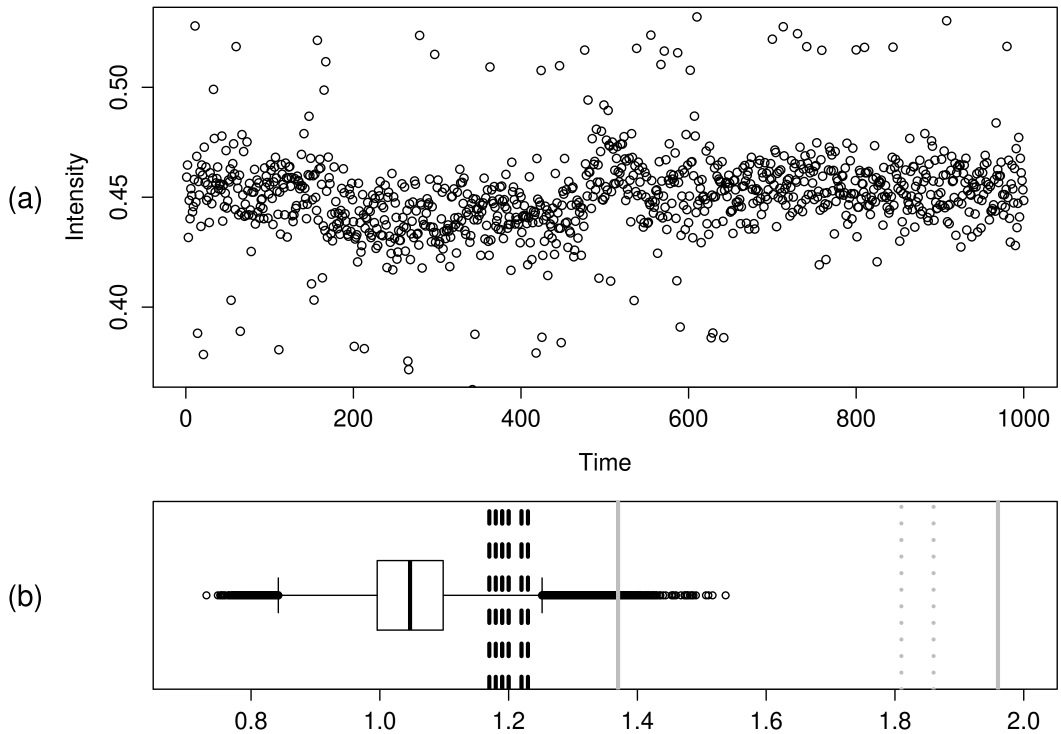

Figure 3.15: (a) Intensity over time for one pixel with additive outliers; (b) A boxplot of standard deviations (multiplied by 10^2) for the virus-free pixels together with the ordinary sample standard deviation and MAD (——), the non-robust estimators $\widehat{\sigma}_{\mathrm{me},2}$ and $\widehat{\sigma}_{\mathrm{Tr,ad}}$ ($\cdots$) and the robust estimators $\widehat{\sigma}_{\mathrm{M,mod},2}$, $\widehat{\sigma}_{\mathrm{Diff,med}}$, $\widehat{\sigma}_{\mathrm{M,me}}$, $\widehat{\sigma}_{\mathrm{M,med}}$, $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$, $\widehat{\sigma}^{2}_{\mathrm{M,mod}}$, $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$, $\widetilde{\widetilde{\sigma}}^{2}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor\sqrt{N}/(2+1)\rfloor, 2\}$ and $\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(6+1)\rfloor, 2\}$ (- - -) applied to the above data in (a).

Data with level shifts and trend

We consider another PAMONO dataset, which exhibits a linear trend. The increase in the grayscale intensity over time presumably originates from the fact that the surface on which the fluid for virus adhesion is placed was heated over time. The corresponding data were discussed in Section 3.3.5. In Panel (a) of Fig. 3.6 (see Section 3.3.5) we see $N = 388$ observations corresponding to a pixel with a virus adhesion and a linear trend. Panel (b) of Fig. 3.6 shows the differenced data $Y_t - Y_{t-1}$, $t = 2, \ldots, 388$. We do not find relevant correlations among the differences. Few large differences indicate the existence of level shifts at the corresponding time points.

We apply the estimators $\widehat{\sigma}_{\mathrm{me},2}$ and $\widehat{\sigma}_{\mathrm{Tr,ad}}$ and the robust procedures $\widehat{\sigma}_{\mathrm{M,mod},2}$, $\widehat{\sigma}_{\mathrm{Diff,med}}$, $\widehat{\sigma}_{\mathrm{M,me}}$, $\widehat{\sigma}_{\mathrm{M,med}}$, $\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$, $\widehat{\sigma}^{2}_{\mathrm{M,mod}}$, $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$ (choosing $K=3$), $\widetilde{\widetilde{\sigma}}^{2}_{\mathrm{M,mod}}$ with block size $n=\max\{\lfloor\sqrt{N}/(2+1)\rfloor,2\}$ and $\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$ with block size $n=\max\{\lfloor N^{1-1/3.1}/(6+1)\rfloor,2\}$ to the PAMONO-data. The non-robust estimators yield values in the range from $1.03\cdot10^{-3}$ to $1.14\cdot10^{-3}$, the robust estimates are between $0.94\cdot10^{-3}$ and $1.17\cdot10^{-3}$. The values of the modified MAD range from $1.03\cdot10^{-3}$ to $1.17\cdot10^{-3}$, while the ordinary MAD and the standard deviation yield values larger than $5\cdot10^{-3}$. The sample standard deviation of differenced observations is $1.38\cdot10^{-3}$, which estimates $\sqrt{2\sigma^2}$. This implies an estimate of $1.38\cdot10^{-3}/\sqrt{2}=0.98\cdot10^{-3}$ for σ, which is close to the values obtained by the blockwise estimators, as e.g. the modified MAD with block size $\lfloor\sqrt{N}/(2+1)\rfloor$ yields the value $1.03\cdot10^{-3}$. This example indicates that the blockwise estimators cope well not only with jumps in the mean but also with trends.

3.4.7 Conclusion

In the presence of level shifts and outliers ordinary scale estimators like the sample standard deviation or the MAD are biased. In this thesis we have investigated several approaches in order to estimate the parameter σ in such situations.

Our basic idea is to segregate the data into many non-overlapping blocks and to estimate the scale in every block. Subsequently, the block-estimates are combined to get a robust estimate of scale. We propose a modified version of the MAD, where the median is estimated in blocks rather than on the whole sample. Alternatively, we can estimate the scale parameter in blocks and take the average, the median or the trimmed mean of the values.

To get an overall assessment of the performance of the proposed estimators we have conducted a simulation study. Further robust estimation techniques based on taking the median or a trimmed average of the blockwise MAD values have been involved for comparison. These robust estimation techniques have performed similarly well and provided good results in many scenarios.

Our recommendation is to use the modified version of the MAD, since we have asymptotic theory on it available. If the focus is on consistent estimation we recommend choosing the block size as a multiple of the square root of N, i.e., $n=\max\left\{\left\lfloor\sqrt{N}/(K+1)\right\rfloor,2\right\}$. Using $n=\max\left\{\left\lfloor\sqrt{N}/(2+1)\right\rfloor,2\right\}$ yields similar results. If the focus is on testing then choosing larger block sizes is recommended, to satisfy the condition $m=o(N^{1/3})$. Theoretical results for this approach were obtained under independence assumption. In future work further asymptotic theory could be developed for dependent data, e.g. in the case of stationary ARMA processes.

An advantage of the median and the trimmed MAD is their smaller computation time, which is $O(\max\{m,n\})$. In case of $n=O(\sqrt{N})=m$ the computation time is $O(\sqrt{N})$ and thus smaller than the $O(N)$ needed for the ordinary or the modified MAD (see [74]).

4 Summary and Outlook

In this chapter the results of this thesis are summarised and an outlook on future work is given.

4.1 Summary

In this thesis we have been dealing with estimation of parameters under shifts in the mean. The results of this work are based on the articles “Estimation methods for the LRD parameter under a change in the mean” [72], “On variance estimation under shifts in the mean” [4] and “Robust scale estimation under shifts in the mean” [3].

Estimation of the Hurst parameter under shifts in the mean

In the context of long range dependent (LRD) stochastic processes the main task is estimation of the Hurst parameter H, which describes the strength of dependence. When data are contaminated by level shifts ordinary estimators of H, such as the Geweke and Porter-Hudak (GPH) estimator, may fail to distinguish between LRD and structural changes, such as jumps in the mean. As a consequence, the estimator may suffer from positive bias and overestimate the intensity of the LRD. This fact is e.g. a major issue when testing for changes in the mean. The Wilcoxon change-point test, proposed by [19], involves the exact value of H, which is used for the standardization of the test statistic. Overestimation of the LRD parameter may lead to false test decisions, a structural change in a sequence of observations could be overlooked. Therefore, the need to estimate H properly arises in the change-point context.

To overcome this problem, we have proposed to segregate the sample of size N into blocks and then to estimate H on each block separately. Estimates, obtained from different blocks, were then combined and a final estimate of the Hurst parameter was obtained. We have investigated several possibilities of segregating the data and assessed their performance in a simulation study.

One possibility was segregation into two blocks. The position at which the data are separated into two parts could either be estimated using the Wilcoxon change-point test or chosen at any point, yielding estimates, which were combined by averaging. When dealing with only one jump these procedures performed well. Under several shifts in the mean the results worsened drastically, since these procedures are designed for the case of one level shift.

Better results could be achieved by dividing the sequence of observations into many overlapping or non-overlapping blocks of size n, chosen as a multiple of $\sqrt{N}$, and estimating H by averaging estimates from these blocks. In the presence of several jumps this procedure performed much better. When dealing with processes with long memory and short range dependence, such as the fractionally integrated ARMA process (ARFIMA), the blockwise estimators did not yield satisfying results. Therefore, we followed an ARMA correction procedure and estimated the Hurst parameter in several recursive steps, using the overlapping or the non-overlapping blocks approach. This strategy improved the results considerably.

The proposed estimators were applied to two real data sets, the Nile river minima and the global temerature at the northern hemisphere, revealing a possible existence of level shifts. High values of H were obtained by ordinary estimators, while our techniques delivered lower estimates, indicating that the data might be contaminated by level shifts.

Scale estimation under shifts in the mean

In the context of LRD we have seen that segregation into many blocks has improved the ordinary estimators of H considerably under abrupt changes in the mean. We followed this same idea of segregation to estimate the variance of independent or weakly dependent processes under level shifts. When dealing with a few level shifts in finite samples we proposed usage of the ordinary average of sample variances, obtained from many non-overlapping blocks. We proved strong consistency and asymptotic normality under independence, where full asymptotic efficiency compared to the ordinary sample variance was shown. For weakly correlated processes we proved weak consistency of the blocks estimator. This estimator performed well when the number of level shifts was moderately low. In the presence of many level shifts even better results were obtained by an adaptive trimmed mean of the sample variances from non-overlapping blocks. The fraction of trimmed blockwise estimates was chosen adaptively, where extraordinary high sample variances were removed before calculating the average value. Even though this procedure was developed under the assumption of independence, it performed well also under weak dependence, e.g. when dealing with AR processes.

If the data are additionally contaminated by outliers the proposed estimators fail to estimate the variance properly, since they are not robust. Therefore, we investigated a modified version of the well known median absolute deviation (MAD) to account for both sources of contamination – level shifts and outliers. The formula of the MAD involves the sample median, which is not a good estimator of location in the presence of level shifts, especially if those occur approximately after 50% of the observations. Our proposal was to calculate the sample median in non-overlapping blocks and to consider absolute differences involving blockwise medians instead of only one median calculated on the whole sample. In this way only some blocks are affected by level shifts and the resulting modified MAD is robust against outliers and level shifts simultaneously. We proved strong consistency and asymptotic normality under some conditions on the number of change-points and the number of blocks for independent random variables. The Bahadur representation of the proposed estimator was shown to be the same as in the case of the ordinary MAD, resulting in the same asymptotic variance. In a simulation study the modified MAD provided very good results. The proposed estimator performed well as compared to other robust methods, which were discussed for comparison, in many simulation scenarios.

Application of the robust and the non-robust blockwise estimators to a real jump-contaminated PAMONO data set indicated that the segregation technique improved the results considerably, since smaller estimates of scale were obtained than in the case of ordinary scale estimators, such as the sample standard deviation or the ordinary MAD. We also added outliers to the PAMONO data and observed that the robust blockwise techniques performed much better than the non-robust ones then. Finally, data with a linear trend were analysed. The results indicated that our segregation techniques provided reasonable results even in such scenarios.

4.2 Outlook

In this thesis we came to the conclusion that segregation into many blocks improved ordinary estimation techniques considerably when dealing with abrupt level shifts. Our work opens up some ideas for future research.

In the context of long range dependent processes we considered an average value of blockwise estimates of the Hurst parameter. It would be interesting to develop estimation techniques based on trimmed averaging. The choice of the block size could also be investigated more extensively in additional simulation studies. Finally, further theoretical results on the overlapping and the non-overlapping blocks approaches, such as derivation of an asymptotic distribution for inference, might be interesting.

For estimation of the variance under shifts in the mean we proposed usage of blockwise techniques involving the sample variance in every block. Theoretical results, such as consistency and asymptotic normality, could be derived straightforwardly for the average of the blocks estimates. For the adaptively trimmed average of sample variances this seems to be more difficult and is an interesting task for future work. Moreover, the trimmed estimator was developed assuming independent random variables. Although this approach worked reasonably well for autoregressive processes, it is advisable to adapt it to dependent data to improve its performance.

In the presence of outliers we developed a modified version of the MAD to cope with level shifts and outliers simultaneously. Theoretical results were developed under the assumption of independence. In future work asymptotic theory could be developed for dependent data.

In the simulation study the performance of the approaches based on segregating data into many blocks was compared with each other. Observations from an autoregressive model were generated to asses the performance of the estimators under dependence. Other models, such as the ARMA model, could be considered in further investigations. Moreover, we applied the estimators to a sequence of observations with a linear trend to get an impression of how they work in the presence of a trend. An extensive analysis of the performance under other trend scenarios could be carried out. Finally, investigation of the performance of the proposed estimators assuming other kinds of structural changes, such as changes in scale or gradual changes in the mean, seem interesting.

A fBm, fGn and ARFIMA

In this Section we will briefly introduce the definition of fractional Brownian motion (fBm) and fractional Gaussian noise (fGn). For more information see [8].

Definition 1 (Brownian motion)**.** A Brownian motion $B(t)$ is a stochastic process with the following properties:

1. $B(t)$ is a Gaussian process (i.e., $B(t_1), ..., B(t_k)$ has multivariate normal distribution for any set $t_1, ..., t_k$, $k < \infty$),
2. $B(0) = 0$ almost surely,
3. The corresponding process of increments $B(t) - B(s)$ is independent,
4. $E(B(t)) = E(B(s))$,
5. $Var(B(t) - B(s)) = \sigma^2 |t - s|$.

If $\sigma^2 = 1$, then $B(t)$ is called a standardized Brownian motion.

Definition 2 (Fractional Brownian motion)**.** Let w_H be a weight function with

$$w_H(t,u) = \begin{cases} 0, & t \leq u, \\ (t-u)^{H-1/2}, & 0 \leq u \leq t, \\ (t-u)^{H-1/2} - (-u)^{H-1/2}, & u < 0, \end{cases}$$

where $0 < H < 1$. Moreover, consider the standardized Brownian motion $B(t)$. Let $s > 0$ be a constant. Then, the fractional Brownian motion $B_H(t)$ with self-similarity parameter H is defined as follows:

$$B_H(t) = s \int w_H(t,u) dB(u). \tag{A.1}$$

$B_H(t)$ is a stochastic integral, which is convergent in the L^2 sense with respect to the Lebesgue measure. The process $B_H(t)$ is a Brownian motion in case of $H = 1/2$.

Definition 3 (Fractional Gaussian noise)**.** The stationary incremental process of the fractional Brownian motion $(\xi_t)_{t \geq 1} = B_H(t) - B_H(t-1)$, $t \in \mathbb{Z}$ is called fractional Gaussian noise.

Definition 4 (ARFIMA process)**.** A fractionally integrated ARMA process (ARFIMA(p,d,q)) X_t with MA parameter p, AR parameter q and fractional differencing parameter $d = H - 1/2 \in$

$(-1/2, 1/2)$, where H is the Hurst parameter, is an extension of the integrated process ARIMA(p, d, q). It is defined by the equation

$$\Phi(B)(1-B)^d X_t = \Theta(B)\epsilon_t,$$

where B is the backshift operator with $B\epsilon_t = \epsilon_{t-1}$, $\Phi(B)$ and $\Theta(B)$ are the AR- and MA-filters of order p and q, respectively, and ϵ_t are i.i.d. with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) < \infty$.

Remark 14. The fGn and the ARFIMA processes exhibit long range dependence, since both have a slowly decaying autocovariance function. See [8] for more details.

B Proofs

Proof of Theorem 4.
Without loss of generality assume that the first $B \leq K$ out of m blocks are perturbed by K jumps in the mean. Furthermore, let $\mu = E(X_{j,t}) \, \forall j,t$, $\mu_{j,t} = E\left(Y_{j,t}\right)$ and $\overline{\mu}_j = \frac{1}{n}\sum_{t=1}^{n} E\left(Y_{j,t}\right)$. Let the term $S_{j,0}^2$ denote the empirical variance of the data $X_{j,1},\ldots,X_{j,n}$ in the uncontaminated block $j = 1,\ldots,m$, while $S_{j,h}^2$ is the empirical variance in the perturbed block $j = 1,\ldots,B$. Then

$$
\begin{aligned}
&P\left(\left|\widehat{\sigma^2}_{\text{Mean}} - \sigma^2\right| > \epsilon\right) = P\left(\left|\frac{1}{m}\sum_{j=B+1}^{m} S_{j,0}^2 + \frac{1}{m}\sum_{j=1}^{B} S_{j,h}^2 - \sigma^2\right| > \epsilon\right)\\
&= P\left(\left|\frac{1}{m}\sum_{j=B+1}^{m} S_{j,0}^2 + \frac{1}{m}\sum_{j=1}^{B} S_{j,0}^2\right.\right.\\
&+ \frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}(X_{j,t} - \overline{X}_j)(\mu_{j,t} - \overline{\mu}_j)\\
&\left.\left.+\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t} - \overline{\mu}_j\right)^2 - \sigma^2\right| > \epsilon\right)\\
&= P\left(\left|\frac{1}{m}\sum_{j=1}^{m} S_{j,0}^2 - \sigma^2 + \frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n} X_{j,t}(\mu_{j,t} - \overline{\mu}_j)\right.\right.\\
&\left.\left.+\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t} - \overline{\mu}_j\right)^2\right| > \epsilon\right)\\
&\leq P\left(\left|\frac{1}{m}\sum_{j=1}^{m} S_{j,0}^2 - \sigma^2\right| > \frac{\epsilon}{3}\right)\\
&+ P\left(\left|\frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n} X_{j,t}(\mu_{j,t} - \overline{\mu}_j)\right| > \frac{\epsilon}{3}\right)\\
&+ P\left(\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t} - \overline{\mu}_j\right)^2 > \frac{\epsilon}{3}\right)\\
&=: A_{1,N} + A_{2,N} + A_{3,N}
\end{aligned}
\tag{B.1}
$$

For the first term $A_{1,N}$ of (B.1) we have:

$$
\begin{aligned}
A_{1,N} =& P\left(\left|\frac{1}{m}\sum_{j=1}^{m} S_{j,0}^2 - \sigma^2\right| > \frac{\epsilon}{3}\right)\\
&= P\left(\left|\frac{1}{m}\sum_{j=1}^{m}\frac{1}{n-1}\sum_{t=1}^{n}(X_{j,t} - \overline{X}_j)^2 - \sigma^2\right| > \frac{\epsilon}{3}\right)
\end{aligned}
\tag{B.2}
$$

$$
\begin{aligned}
&= P\left(\left|\frac{1}{m}\sum_{j=1}^{m}\frac{1}{n-1}\sum_{t=1}^{n}(X_{j,t}-\mu)^2+\frac{1}{m}\sum_{j=1}^{m}\frac{n}{n-1}(\mu-\overline{X}_j)^2\right.\right.\\
&\left.\left.+\frac{2}{m(n-1)}\sum_{j=1}^{m}\sum_{t=1}^{n}(X_{j,t}-\mu)(\mu-\overline{X}_j)-\sigma^2\right|>\frac{\epsilon}{3}\right)\\
&\le P\left(\left|\frac{N}{m(n-1)}\frac{1}{N}\sum_{j=1}^{m}\sum_{t=1}^{n}(X_{j,t}-\mu)^2-\sigma^2\right|>\frac{\epsilon}{6}\right)\\
&+P\left(\left|-\frac{n}{m(n-1)}\sum_{j=1}^{m}(\overline{X}_j-\mu)^2\right|>\frac{\epsilon}{6}\right)
\end{aligned}
\tag{B.3}
$$

The first term of (B.3) converges to zero, since the term $\frac{1}{N}\sum_{j=1}^{m}\sum_{t=1}^{n}(X_{j,t}-\mu)^2$ is a consistent estimator of σ^2 and $\frac{N}{m(n-1)}=\frac{N}{N-m}\to 1$. For the second term of (B.3) we have

$$
\begin{aligned}
&E\left(\left|-\frac{n}{m(n-1)}\sum_{j=1}^{m}(\overline{X}_j-\mu)^2\right|\right)=\frac{n}{m(n-1)}\sum_{j=1}^{m}E\left((\overline{X}_j-\mu)^2\right)\\
&=\frac{n}{m(n-1)}\sum_{j=1}^{m}Var(\overline{X}_j)\le\frac{n}{(n-1)}\frac{\gamma(0)+2\sum_{i=1}^{\infty}\gamma(i)}{n}\\
&\longrightarrow 0,
\end{aligned}
$$

due to absolute summability of the autocovariance function γ, which implies convergence in probability. Therefore, the second term of (B.3) converges to zero. Altogether, the sum (B.3) converges to zero.

The second and the third terms $A_{2,N}$ and $A_{3,N}$ of (B.1) converge to zero due to the following considerations:

$$
\begin{aligned}
&E\left(\frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j)\right)\\
&=\frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}E(X_{j,t})(\mu_{j,t}-\overline{\mu}_j)\\
&=\frac{1}{m}\sum_{j=1}^{B}\frac{2\mu}{n-1}\sum_{t=1}^{n}(\mu_{j,t}-\overline{\mu}_j)=0\\
&Var\left(\frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j)\right)\\
&=\frac{4}{m^2(n-1)^2}Var\left(\sum_{j=1}^{B}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j)\right)\\
&=\frac{4}{m^2(n-1)^2}Cov\left(\sum_{j=1}^{B}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j),\sum_{l=1}^{B}\sum_{s=1}^{n}X_{l,s}(\mu_{l,s}-\overline{\mu}_l)\right)\\
&\le\frac{4}{m^2(n-1)^2}\sum_{j=1}^{B}\sum_{t=1}^{n}\sum_{l=1}^{B}\sum_{s=1}^{n}|\mu_{j,t}-\overline{\mu}_j||\mu_{l,s}-\overline{\mu}_l||Cov(X_{j,t},X_{l,s})|
\end{aligned}
$$

$$
\begin{aligned}
&\le \frac{4}{m^2(n-1)^2}\left(\sum_{k=1}^{K} h_k\right)^2 \sum_{j=1}^{m}\sum_{t=1}^{n}\sum_{l=1}^{m}\sum_{s=1}^{n} |Cov(X_{j,t}, X_{l,s})| \\
&= \frac{4}{m^2(n-1)^2}\left(\sum_{k=1}^{K} h_k\right)^2 \left(NVar(X_{1,1}) + 2\sum_{u=1}^{N-1} |\gamma(u)|(N-u)\right) \\
&\le \frac{4N}{m^2(n-1)^2}\left(\sum_{k=1}^{K} h_k\right)^2 \left(Var(X_{1,1}) + 2\sum_{u=1}^{\infty} |\gamma(u)|\right) \\
&\longrightarrow 0,
\end{aligned}
$$

due to absolute summability of the autocovariance function γ and the condition $K\left(\sum_{k=1}^{K} h_k\right)^2 = o(m\log(N)^{-1})$. Therefore, the term $A_{2,N}$ converges to zero in r-th mean with $r = 2$, which implies convergence in probability. The argument in the probability $A_{3,N}$ is deterministic. Set $\epsilon_N = \log(N)^{-1}$. For large N we have

$$
\begin{aligned}
A_{3,N} &\le P\left(\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t} - \overline{\mu}_j\right)^2 > \frac{\epsilon_N}{3}\right) \\
&= P\left(\log(N)\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t} - \overline{\mu}_j\right)^2 > \frac{1}{3}\right) \\
&\le P\left(\log(N)\frac{B}{m}\frac{n}{n-1}\left(\sum_{k=1}^{K} h_k\right)^2 > \frac{1}{3}\right) \\
&\le P\left(\log(N)\frac{K}{m}\frac{n}{n-1}\left(\sum_{k=1}^{K} h_k\right)^2 > \frac{1}{3}\right) = 0,
\end{aligned}
$$

since $K\left(\sum_{k=1}^{K} h_k\right)^2 = o(m\log(N)^{-1})$ holds. Altogether we get the result $\widehat{\sigma^2}_{\text{Mean}} \xrightarrow{p} \sigma^2$. □

Proof of Lemma 6.
We will use Hoeffding's inequality to show the result. Without loss of generality we assume that the first $B \le K$ out of m blocks are contaminated by $\widetilde{K}_1, \ldots, \widetilde{K}_B$ jumps, respectively, with $\sum_{j=1}^{B} \widetilde{K}_j = K$, each with the corresponding empirical CDF $\widehat{G}_{N,j,h}(y)$ of the absolute differences, $j \in \{1, \ldots, B\}$, e.g., for the first contaminated block we have

$$
\widehat{G}_{N,1,h}(y) = \frac{1}{n}\sum_{t=1}^{n} I_{\{W_t \le y\}}, \quad \text{with} \quad W_t = \left| X_t + \sum_{k=1}^{\widetilde{K}_1} h_k I_{t \ge t_k} - \widehat{\nu}_{N,1,h} \right|, \; t = 1, \ldots, n,
$$

where $\widehat{\nu}_{N,1,h}$ is the sample median in this block. The empirical CDF $\widehat{G}_{N,j,0}(y)$ in a jump-free block j is defined similarly. The empirical CDF $\widehat{G}_N$ of the absolute differences in the change-point scenario can be written as follows:

$$\widehat{G}_N(y) = \frac{1}{N}\sum_{t=1}^{N} I_{\{W_t \le y\}} = \frac{1}{m}\sum_{j=1}^{m}\widehat{G}_{N,j}(y) = \frac{1}{m}\left(\sum_{j=1}^{B}\widehat{G}_{N,j,h}(y) + \sum_{j=B+1}^{m}\widehat{G}_{N,j,0}(y)\right),\ y \in \mathbb{R}.$$

Let $\epsilon > 0$, then:

$$P(|\widehat{\zeta}_{N,n} - \zeta| > \epsilon) = P(\widehat{\zeta}_{N,n} < \zeta - \epsilon) + P(\widehat{\zeta}_{N,n} > \zeta + \epsilon). \tag{B.4}$$

Consider one of the two terms in (B.4). With

$\widehat{F}_{N,j,0}$ the empirical CDF of $X_{j,1}, \ldots, X_{j,n}$ in j-th block, $j = 1, \ldots, m$,
$\widehat{\nu}_{N,j,0}$ the sample median of $X_{j,1}, \ldots, X_{j,n}$ in j-th block, $j = 1, \ldots, m$,

$$\alpha_N = \left\lfloor \frac{N+1}{2} \right\rfloor \frac{1}{N},$$

$$A_N = \{\alpha_N + \frac{B}{m} > \frac{1}{m}\sum_{j=1}^{m}\widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0} + \zeta + \epsilon) - \widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0} - \zeta - \epsilon)\},$$

$$B_N = \{\alpha_N + \frac{B}{m} > \frac{1}{m}\sum_{j=1}^{m}\widehat{F}_{N,j,0}(\nu + \zeta + \epsilon/2) - \widehat{F}_{N,j,0}(\nu - \zeta - \epsilon/2)\},$$

$C_N = \{$at least one of $|\widehat{\nu}_{N,j,0} - \nu|$ is larger than $\epsilon/2\}$ and
$C_N^C = \{|\widehat{\nu}_{N,1,0} - \nu| \le \epsilon/2\} \cap \ldots \cap \{|\widehat{\nu}_{N,m,0} - \nu| \le \epsilon/2\}$

we have

$$\begin{aligned}
P\left(\widehat{\zeta}_{N,n} > \zeta + \epsilon\right) &= P\left(\widehat{G}_N^{-1}(\alpha_N) > \zeta + \epsilon\right) = P\left(\alpha_N > \widehat{G}_N(\zeta + \epsilon)\right) \\
&= P\left(\alpha_N > \frac{1}{m}\sum_{j=1}^{B}\widehat{G}_{N,j,h}(\zeta + \epsilon) + \frac{1}{m}\sum_{j=B+1}^{m}\widehat{G}_{N,j,0}(\zeta + \epsilon)\right) \\
&= P\left(\alpha_N > \frac{1}{m}\sum_{j=1}^{B}\underbrace{\left(\widehat{G}_{N,j,h}(\zeta + \epsilon) - \widehat{G}_{N,j,0}(\zeta + \epsilon)\right)}_{\in[-1,1]} + \frac{1}{m}\sum_{j=1}^{m}\widehat{G}_{N,j,0}(\zeta + \epsilon)\right) \\
&\le P\left(\alpha_N > -\frac{B}{m} + \frac{1}{m}\sum_{j=1}^{m}\widehat{G}_{N,j,0}(\zeta + \epsilon)\right) \\
&= P\left(\alpha_N + \frac{B}{m} > \frac{1}{m}\sum_{j=1}^{m}(\widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0} + \zeta + \epsilon) - \widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0} - \zeta - \epsilon -))\right) \\
&\le P(A_N) = P(A_N \cap C_N^C) + P(A_N \cap C_N) \le P(B_N) + P(C_N).
\end{aligned} \tag{B.5}$$

The inequality in (B.5) holds due to the following considerations:

$\widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0} + \zeta + \epsilon) - \widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0} - \zeta - \epsilon) \ge \widehat{F}_{N,j,0}(\nu + \zeta + \epsilon/2) - \widehat{F}_{N,j,0}(\nu - \zeta - \epsilon/2)$, when $|\widehat{\nu}_{N,j,0} - \nu| \le \epsilon/2 \Leftrightarrow \nu - \epsilon/2 \le \widehat{\nu}_{N,j,0} \le \nu + \epsilon/2$, yielding

$\widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0}+\zeta+\epsilon) \geq \widehat{F}_{N,j,0}(\nu-\epsilon/2+\zeta+\epsilon)$ and
$\widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0}-\zeta-\epsilon) \leq \widehat{F}_{N,j,0}(\nu+\epsilon/2-\zeta-\epsilon)$.

Therefore:

$$
\begin{aligned}
&P\left(A_N \cap C_N^C\right) \\
&= P\left(\left\{\alpha_N + \frac{B}{m} > \frac{1}{m}\sum_{j=1}^{m} \widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0}+\zeta+\epsilon) - \widehat{F}_{N,j,0}(\widehat{\nu}_{N,j,0}-\zeta-\epsilon)\right\} \middle| C_N^C\right) \cdot P\left(C_N^C\right) \\
&\leq P\left(\left\{\alpha_N + \frac{B}{m} > \frac{1}{m}\sum_{j=1}^{m} \widehat{F}_{N,j,0}(\nu+\zeta+\epsilon/2) - \widehat{F}_{N,j,0}(\nu-\zeta-\epsilon/2)\right\} \middle| C_N^C\right) \cdot P\left(C_N^C\right) \\
&= P\left(B_N | C_N^C\right) \cdot P\left(C_N^C\right) \leq P(B_N)
\end{aligned}
$$

With $Z_t = I_{\{\nu-\zeta-\epsilon/2 < X_t \leq \nu+\zeta+\epsilon/2\}} \in [0,1]$ we have the following result for B_N:

$$
\begin{aligned}
P\left(B_N\right) &= P\left(\alpha_N + \frac{B}{m} > \frac{1}{m}\sum_{j=1}^{m} \widehat{F}_{N,j,0}(\nu+\zeta+\epsilon/2) - \widehat{F}_{N,j,0}(\nu-\zeta-\epsilon/2)\right) \\
&= P\left(\alpha_N + \frac{B}{m} > \widehat{F}_{N,0}(\nu+\zeta+\epsilon/2) - \widehat{F}_{N,0}(\nu-\zeta-\epsilon/2))\right) \\
&= P\left(N\left(\alpha_N + \frac{B}{m}\right) > \sum_{t=1}^{N} I_{\{\nu-\zeta-\epsilon/2 < X_t \leq \nu+\zeta+\epsilon/2\}}\right) = P\left(N\left(\alpha_N + \frac{B}{m}\right) > \sum_{t=1}^{N} Z_t\right) \\
&\overset{Z_t \text{ i.i.d.}}{=} P\left(N\left(\alpha_N + \frac{B}{m} - E(Z_1)\right) > \sum_{t=1}^{N}(Z_t - E(Z_t))\right) \\
&\overset{[38]}{\leq} \exp\{-2N\Delta_{2,\epsilon,N}^2\} \text{ with } \Delta_{2,\epsilon,N} = (E(Z_1) - \frac{B}{m} - \alpha_N)^+, \qquad \text{(B.6)}
\end{aligned}
$$

where $\widehat{F}_{N,0}(x) = \frac{1}{m}\sum_{j=1}^{m} \widehat{F}_{N,j,0}(x)$. We have that $E(Z_1) = F(\nu+\zeta+\epsilon/2) - F(\nu-\zeta-\epsilon/2) > 1/2$ since $G(\zeta) = 1/2$, and $\alpha_N \to 1/2$. Thus, $\exists\, N_0 \in \mathbb{N}$ such that $\alpha_N + \frac{B}{m} - E(Z_1) < 0 \, \forall N \geq N_0$ (and thus $\Delta_{2,\epsilon,N} = (E(Z_1) - \frac{B}{m} - \alpha_N)^+ = E(Z_1) - \frac{B}{m} - \alpha_N \, \forall N \geq N_0$). Therefore, Hoeffding's Inequality can be applied in (B.6). Furthermore,

$$
\begin{aligned}
P\left(C_N\right) &= P\left(\{\text{at least one of } |\widehat{\nu}_{N,j,0} - \nu| \text{ is larger than } \epsilon/2\}\right) \\
&= P\left(\bigcup_{j=1}^{m} \{|\widehat{\nu}_{N,j,0} - \nu| > \epsilon/2\}\right) \leq \sum_{j=1}^{m} P\left(|\widehat{\nu}_{N,j,0} - \nu| > \epsilon/2\right) \\
&\overset{[79]}{\leq} \sum_{j=1}^{m} 2\exp\left\{-2\frac{N}{m}\delta_{\epsilon,n}^2\right\} = 2m\exp\left\{-2\frac{N}{m}\delta_{\epsilon,n}^2\right\}.
\end{aligned}
$$

Therefore, we obtain

$$
P\left(\widehat{\zeta}_{N,n} > \zeta + \epsilon\right) \leq \exp\{-2N\Delta_{2,\epsilon,N}^2\} + 2m\exp\left\{-2\frac{N}{m}\delta_{\epsilon,n}^2\right\}.
$$

Similarly,

$$P\left(\widehat{\zeta}_{N,n} < \zeta - \epsilon\right) \leq \exp\{-2N\Delta^2_{3,\epsilon,N}\} + 2m\exp\left\{-2\frac{N}{m}\delta^2_{\epsilon,n}\right\},$$

which proves the above Lemma. □

Proof of Proposition 7.
We will show that the infinite sum of the probabilities $P(|\widehat{\zeta}_{N,n} - \zeta| > \epsilon)$ from (3.27) is finite and use the Borell-Cantelli Lemma to prove the almost sure convergence. We observe that the third term of the right hand side of (3.27) converges to zero as long as $2(N/m)(\delta_{\epsilon,n})^2 > \log(m)$, where $\delta_{\epsilon,n} = \delta_{\epsilon,N/m}$. By definition of $\delta_{\epsilon,n}$ from (3.30) and due to uniqueness of ν (see [79]) we have

$$\delta_{\epsilon,n} \overset{N\to\infty}{\to} \delta_\epsilon = \min\{F(\nu+\epsilon/2) - 1/2,\, 1/2 - F(\nu-\epsilon/2)\} > 0$$

Therefore, $\exists N_0$ such that

$$\delta_{\epsilon,n} = \delta_{\epsilon,N/m} > \delta_\epsilon/2 > 0 \,\forall\, N \geq N_0 \tag{B.7}$$

and therefore,

$$2m\exp\left\{-2\frac{N}{m}\delta^2_{\epsilon,n}\right\} \leq 2m\exp\left\{-2\frac{N}{m}\left(\frac{\delta_\epsilon}{2}\right)^2\right\} = 2\exp\left\{\log(m) - 2\frac{N}{m}\left(\frac{\delta_\epsilon}{2}\right)^2\right\}.$$

Moreover (keeping in mind that $m = m(N)$),

$$\sum_{N=N_0}^{\infty} 2m\exp\left\{-2\frac{N}{m}\delta^2_{\epsilon,n}\right\} \leq \sum_{N=N_0}^{\infty} 2\exp\left\{\log(m) - 2\frac{N}{m}\left(\frac{\delta_\epsilon}{2}\right)^2\right\} < \infty,$$

for sufficiently large N_0 (see (B.7)), is ensured if the following condition is satisfied:

$$\frac{2\exp\left\{\log(m) - 2\frac{N}{m}\left(\frac{\delta_\epsilon}{2}\right)^2\right\}}{1/N^p} = 2\exp\left\{\log(N^p m) - 2\frac{N}{m}\left(\frac{\delta_\epsilon}{2}\right)^2\right\} \to 0,$$

for any $p > 1$. This is for example true when choosing m as a multiple of any power of N, i.e., $m = cN^q$, $q \in (0,1)$, $c > 0$, since

$$2\exp\left\{\log(N^p cN^q) - 2\frac{N}{cN^q}\left(\frac{\delta_\epsilon}{2}\right)^2\right\} = 2\exp\left\{\log(c) + (p+q)\log(N) - 2\frac{N^{(1-q)}}{c}\left(\frac{\delta_\epsilon}{2}\right)^2\right\} \to 0.$$

Let

$$\alpha_N = \left\lfloor \frac{N+1}{2} \right\rfloor \frac{1}{N}.$$

For the first term of the right hand side of (3.27) we have the following considerations:

$$\begin{aligned}
&F(\nu+\zeta+\epsilon/2) - F(\nu-\zeta-\epsilon/2) > 1/2,\ \alpha_N \to 1/2 \text{ and } \frac{B}{m} \to 0 \\
&\Rightarrow \exists\, N_0 \in \mathbb{N}, \ddot{\epsilon} > 0 \text{ such that } F(\nu+\zeta+\epsilon/2) - F(\nu-\zeta-\epsilon/2) - \alpha_N + \frac{B}{m} \geq \ddot{\epsilon} \\
&\Rightarrow \sum_{N=N_0}^{\infty} \exp\{-2N\Delta^2_{2,\epsilon,N}\} \leq \sum_{N=N_0}^{\infty} \exp\{-2N\ddot{\epsilon}^2\} \to 0.
\end{aligned}$$

Similar argumentations are valid for the second term. Together with the Borel-Cantelli Lemma the almost sure convergence of $\widehat{\zeta}_N$ to ζ in the change-point scenario is ensured. □

Proof of Lemma 9.

$$\begin{aligned}
&P\left(\left|\left(\widehat{\nu}_{N,i} + \widehat{\zeta}_{N,n}\right) - (\nu+\zeta)\right| > \epsilon\right) = P\left(\left|(\widehat{\nu}_{N,i} - \nu) + \left(\widehat{\zeta}_{N,n} - \zeta\right)\right| > \epsilon\right) \\
&\leq P\left(\left|(\widehat{\nu}_{N,i} - \nu)\right| > \epsilon/2\right) + P\left(\left|\left(\widehat{\zeta}_{N,n} - \zeta\right)\right| > \epsilon/2\right) \\
&\leq (2+4m)\exp\left\{-2\frac{N}{m}\delta^2_{\epsilon/2,n}\right\} + \exp\{-2N\Delta^2_{2,\epsilon/2,N}\} + \exp\{-2N\Delta^2_{3,\epsilon/2,N}\},
\end{aligned} \tag{B.8}$$

with $\delta_{\epsilon/2,n}$, $\Delta_{2,\epsilon/2,N}$, $\Delta_{3,\epsilon/2,N}$ as defined in (3.28) - (3.30). Here we used the fact that $P\left(\left|(\widehat{\nu}_{N,i} - \nu)\right|\right) \leq 2\exp\left\{-2\frac{N}{m}\delta^2_{\epsilon/2,n}\right\}$, see [79]. With

$$\epsilon_n := 2D_1 \log(n)^{1/2}/n^{1/2} \tag{B.9}$$

and Taylor's theorem we get

$$\begin{aligned}
a_0(\epsilon_n/2) &= F(\nu+\epsilon_n/4) - \left(\left\lfloor \frac{n+1}{2} \right\rfloor - 1\right)\frac{1}{n} \\
&= F(\nu+\epsilon_n/4) - \frac{1}{2} + O\left(\frac{1}{n}\right) = \frac{F'(\nu)}{4}\epsilon_n + o(\epsilon_n) + O\left(\frac{1}{n}\right) \\
&= \frac{F'(\nu)}{4}\max\left\{\frac{8}{F'(\nu)}, \frac{8}{G'(\zeta)}\right\}\frac{2\log(n)^{1/2}}{n^{1/2}} + o(\epsilon_n) + O\left(\frac{1}{n}\right) \\
&> \frac{2\log(n)^{1/2}}{n^{1/2}} \text{ for suffciently large } n.
\end{aligned}$$

Similarly, we get

$$b_0(\epsilon_n/2) > \frac{2\log(n)^{1/2}}{n^{1/2}} \text{ for suffciently large } n.$$

$$\begin{aligned}
\Delta_{2,\epsilon_n/2,N} &= \left(F(\nu+\zeta+\epsilon_n/4) - F(\nu-\zeta-\epsilon_n/4) - \left\lfloor \frac{N+1}{2} \right\rfloor /N\right)^+ \\
&= \left(G(\zeta+\epsilon_n/4) - G(\zeta) + O\left(\frac{1}{N}\right)\right)^+ \\
&= \left(G'(\zeta)\epsilon_n/4 + o\left(\epsilon_n\right) + O\left(\frac{1}{N}\right)\right)^+ \\
&> \frac{2\log(n)^{1/2}}{n^{1/2}} \text{ for suffciently large } n. \\
\Delta_{3,\epsilon_n/2,N} &= \left\lfloor \frac{N+1}{2} \right\rfloor /N - F(\nu+\zeta-\epsilon/2) - F(\nu-\zeta+\epsilon/2) \\
&> \frac{2\log(n)^{1/2}}{n^{1/2}} \text{ for suffciently large } n.
\end{aligned}$$

Similarly, we get for sufficiently large n, that

$$(\epsilon_n/2) > \frac{2\log(n)^{1/2}}{n^{1/2}}, \; \Delta_{2,\epsilon_n/2,N} > \frac{2\log(n)^{1/2}}{n^{1/2}} \quad \text{and} \quad \Delta_{3,\epsilon_n/2,N} > \frac{2\log(n)^{1/2}}{n^{1/2}}.$$

We conclude that

$$\min\{a_0(\epsilon_n/2), b_0(\epsilon_n/2), \Delta_{2,\epsilon_n/2,N}, \Delta_{3,\epsilon_n/2,N}\} > \frac{2\log(n)^{1/2}}{n^{1/2}} \text{ for suffciently large } n. \tag{B.10}$$

Therefore,

$$\begin{aligned}
&P\left(\left|\left(\widehat{\nu}_{N,i} + \widehat{\zeta}_{N,n}\right) - (\nu+\zeta)\right| > \epsilon_n\right) \\
&\leq (2+4m)\exp\left\{-2\frac{N}{m}\delta^2_{\epsilon_n/2,n}\right\} + \exp\{-2N\Delta^2_{2,\epsilon_n/2,N}\} + \exp\{-2N\Delta^2_{3,\epsilon_n/2,N}\} \\
&\leq (2+4m)\exp\{-2\cdot 4\log(n)\} + 2\exp\{-2m\cdot 4\log(n)\} \\
&= (2+4m)n^{-8} + 2n^{-8m}.
\end{aligned} \tag{B.11}$$

Summability of (B.11), i.e.,

$$\sum_{N=1}^{\infty}\left(2\left(\frac{m}{N}\right)^8 + \frac{4m^9}{N^8} + 2\left(\frac{m}{N}\right)^{8m}\right) < \infty, \tag{B.12}$$

is ensured when all terms approach zero faster than $\frac{1}{N^p}$, for any $p > 1$. We consider only the slowest term:

$$\frac{\frac{4m^9}{N^8}}{1/N^p} = \frac{4m^9}{N^{8-p}} \to 0,$$

if $m = o(N^{2/3})$ (for $p = 2$). Furthermore, for any $\epsilon > 0 \; \exists N_0$ such that $\epsilon_n < \epsilon \, \forall \, n = n(N) > n(N_0)$ and we have

$$P\left(\left|\left(\widehat{\nu}_{N,i} + \widehat{\zeta}_{N,n}\right) - (\nu+\zeta)\right| > \epsilon\right) \leq P\left(\left|\left(\widehat{\nu}_{N,i} + \widehat{\zeta}_{N,n}\right) - (\nu+\zeta)\right| > \epsilon_n\right)$$

with ϵ_n from (B.9). The Borel-Cantelli Lemma yields the desired result due to the fact that the probability on the right hand side satisfies (B.12). The same result holds for (3.35).

□

Proof of Theorem 8.
Let F be the CDF of X_1 and G the CDF of $|X_1 - \nu|$. Furthermore, let $\widehat{F}_{N,j}$ be the empirical CDF in block j, whereas $\widehat{G}_{N,j}$ with $\widehat{G}_{N,j}(y) = \widehat{F}_{N,j}(\widehat{\nu}_{N,j} + y) - \widehat{F}_{N,j}(\widehat{\nu}_{N,j} - y-)$ is the empirical CDF of $|X_{j,1} - \widehat{\nu}_{N,j}|, \ldots, |X_{j,n} - \widehat{\nu}_{N,j}|$, where $\widehat{\nu}_{N,j} = \text{med}\{X_{j,1}, \ldots, X_{j,n}\}$, $j = 1, \ldots, m$, and $\widehat{F}_{N,j}(x-) = \lim_{z\uparrow x} \widehat{F}_{N,j}(z)$. Moreover, let $\widehat{F}_N = \frac{1}{m}\sum_{j=1}^m \widehat{F}_{N,j}$ be the empirical CDF of the whole sample $X_1, \ldots, X_N$ and $\widehat{\nu}_N = \frac{1}{m}\sum_{j=1}^m \widehat{\nu}_{N,j}$.

Similarly to [55], we decompose Δ_N into many components $\Delta_N^{(1)}, \ldots, \Delta_N^{(6)}$, where the last five terms are sums of blockwise error terms, i.e., $\Delta_N^{(l)} = \frac{1}{m}\sum_{j=1}^m \Delta_{N,j}^{(l)}$, $l = 2, \ldots, 6$, getting

$$\Delta_N = \frac{\Delta_N^{(1)} + O\left(\frac{1}{N}\right) + \Delta_N^{(3)} + \Delta_N^{(4)} - \Delta_N^{(2)} - \Delta_N^{(5)}}{G'(\zeta)} + \frac{F'(\nu - \zeta) - F'(\nu + \zeta)}{G'(\zeta)}\Delta_N^{(6)}.$$

The components $\Delta_N^{(1)}, \ldots, \Delta_N^{(6)}$ are defined by the following equations:

$$\widehat{G}_N(\widehat{\zeta}_{N,n}) = \frac{1}{m}\sum_{j=1}^m \widehat{F}_{N,j}(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\widehat{\nu}_{N,j} - \widehat{\zeta}_{N,n}-) = \frac{1}{2} + \Delta_N^{(1)}, \tag{B.13}$$

$$\frac{1}{m}\sum_{j=1}^m (F(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}) - F(\nu + \zeta)) = \frac{1}{m}\sum_{j=1}^m (F'(\nu + \zeta)(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n} - \nu - \zeta) + \Delta_{N,j}^{(2)}) \tag{B.14}$$

$$= F'(\nu + \zeta)(\widehat{\nu}_N + \widehat{\zeta}_{N,n} - \nu - \zeta) + \Delta_N^{(2)},$$

$$\frac{1}{m}\sum_{j=1}^m (F(\widehat{\nu}_{N,j} - \widehat{\zeta}_{N,n}) - F(\nu - \zeta)) = \frac{1}{m}\sum_{j=1}^m (F'(\nu - \zeta)(\widehat{\nu}_{N,j} - \widehat{\zeta}_{N,n} - \nu + \zeta) + \Delta_{N,j}^{(3)}) \tag{B.15}$$

$$= F'(\nu - \zeta)(\widehat{\nu}_N - \widehat{\zeta}_{N,n} - \nu + \zeta) + \Delta_N^{(3)},$$

$$\frac{1}{m}\sum_{j=1}^m (F(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}) - F(\nu + \zeta)) = \frac{1}{m}\sum_{j=1}^m (\widehat{F}_{N,j}(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\nu + \zeta)) + \Delta_N^{(4)}, \tag{B.16}$$

$$\frac{1}{m}\sum_{j=1}^m (F(\widehat{\nu}_{N,j} - \widehat{\zeta}_{N,n}) - F(\nu - \zeta)) = \frac{1}{m}\sum_{j=1}^m (\widehat{F}_{N,j}(\widehat{\nu}_{N,j} - \widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\nu - \zeta)) + \Delta_N^{(5)}, \tag{B.17}$$

$$\widehat{\nu}_N = \frac{1}{m}\sum_{j=1}^m \widehat{\nu}_{N,j} = \frac{1}{m}\sum_{j=1}^m \left(\nu + \frac{1/2 - \widehat{F}_{N,j}(\nu)}{F'(\nu)} + \Delta_{N,j}^{(6)}\right) = \nu + \frac{1/2 - \widehat{F}_N(\nu)}{F'(\nu)} + \Delta_N^{(6)}. \tag{B.18}$$

Combining (B.13) - (B.17), as was done by [55], we get

$$\begin{aligned}
&F'(\nu + \zeta)(\widehat{\nu}_N + \widehat{\zeta}_{N,n} - \nu - \zeta) - F'(\nu - \zeta)(\widehat{\nu}_N - \widehat{\zeta}_{N,n} - \nu + \zeta) \\
&= \frac{1}{m}\sum_{j=1}^m \widehat{F}_{N,j}(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\widehat{\nu}_{N,j} - \widehat{\zeta}_{N,n}) \\
&- (\widehat{F}_N(\nu + \zeta) - \widehat{F}_N(\nu - \zeta)) + \Delta_N^{(3)} + \Delta_N^{(4)} - \Delta_N^{(2)} - \Delta_N^{(5)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{m}\sum_{j=1}^{m} \widehat{F}_{N,j}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\widehat{\nu}_{N,j}-\widehat{\zeta}_{N,n}-) + O\left(\frac{1}{N}\right)\\
&- (\widehat{F}_N(\nu+\zeta) - \widehat{F}_N(\nu-\zeta)) + \Delta_N^{(3)} + \Delta_N^{(4)} - \Delta_N^{(2)} - \Delta_N^{(5)}\\
&= \frac{1}{2} + \Delta_N^{(1)} + O\left(\frac{1}{N}\right) - (\widehat{F}_N(\nu+\zeta) - \widehat{F}_N(\nu-\zeta)) + \Delta_N^{(3)} + \Delta_N^{(4)} - \Delta_N^{(2)} - \Delta_N^{(5)}\\
\Leftrightarrow &- (\widehat{\nu}_N - \nu)(F'(\nu-\zeta) - F'(\nu+\zeta)) + (\widehat{\zeta}_N - \zeta)G'(\zeta)\\
&= \frac{1}{2} - (\widehat{F}_N(\nu+\zeta) - \widehat{F}_N(\nu-\zeta)) + \Delta_N^{\star},
\end{aligned}
$$

with $\Delta_N^{\star} = \Delta_N^{(1)} + O\left(\frac{1}{N}\right) + \Delta_N^{(3)} + \Delta_N^{(4)} - \Delta_N^{(2)} - \Delta_N^{(5)}$, which yields

$$
\begin{aligned}
(\widehat{\zeta}_N - \zeta) &= \frac{1/2 - (\widehat{F}_N(\nu+\zeta) - \widehat{F}_N(\nu-\zeta))}{G'(\zeta)} + \frac{F'(\nu-\zeta) - F'(\nu+\zeta)}{G'(\zeta)}(\widehat{\nu}_N - \nu) + \frac{\Delta_N^{\star}}{G'(\zeta)}\\
&= \frac{1/2 - (\widehat{F}_N(\nu+\zeta) - \widehat{F}_N(\nu-\zeta))}{G'(\zeta)} + \frac{\Delta_N^{\star}}{G'(\zeta)}\\
&+ \frac{F'(\nu-\zeta) - F'(\nu+\zeta)}{G'(\zeta)}\left(\frac{1/2 - \widehat{F}_N(\nu)}{F'(\nu)} + \Delta_N^{(6)}\right)\\
&= \frac{1/2 - (\widehat{F}_N(\nu+\zeta) - \widehat{F}_N(\nu-\zeta))}{G'(\zeta)} + \frac{F'(\nu-\zeta) - F'(\nu+\zeta)}{G'(\zeta)}\left(\frac{1/2 - \widehat{F}_N(\nu)}{F'(\nu)}\right) + \Delta_N,
\end{aligned}
$$

with

$$
\Delta_N = \frac{\Delta_N^{(1)} + O\left(\frac{1}{N}\right) + \Delta_N^{(3)} + \Delta_N^{(4)} - \Delta_N^{(2)} - \Delta_N^{(5)}}{G'(\zeta)} + \frac{F'(\nu-\zeta) - F'(\nu+\zeta)}{G'(\zeta)}\Delta_N^{(6)}.
$$

We now have to prove

$$
\Delta_N^{(l)} = o_P(N^{-1/2}) \quad \forall l = 1, \dots, 6 \text{ (see (B.13)-(B.18))},
$$

which will be done in the following steps 1. - 6.

1. Following [55], we have for the error term $\Delta_N^{(1)}$ in (B.13)

$$
\widehat{G}_N(\widehat{\zeta}_{N,n}) = \left\lfloor \frac{N+1}{2} \right\rfloor \Big/ N = \frac{1}{2} + O\left(\frac{1}{N}\right).
$$

   Ties between W_j and W_k, $j \neq k$ are excluded, since the underlying CDF F is assumed to be continuous. Therefore, $\sqrt{N}\Delta_N^{(1)} = \sqrt{N}O\left(\frac{1}{N}\right)$ converges to zero in probability.

2. With Taylors theorem we get for the error term $\Delta_N^{(2)}$ in (B.14)

$$
\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}(F(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n}) - F(\nu+\zeta)) = \frac{1}{m}\sum_{j=1}^{m} F'(\nu+\zeta)(\widehat{\nu}_{N,j} + \widehat{\zeta}_{N,n} - \nu - \zeta)\\
&+ \frac{1}{m}\sum_{j=1}^{m}\frac{F''(c_j)}{2}(\widehat{\nu}_{N,i} + \widehat{\zeta}_{N,n} - \nu - \zeta)^2.
\end{aligned}
$$

Let $\epsilon_N := \dfrac{D_1^2}{\log(N)}$, $D_1 = \max\{8/F'(\nu), 8/G'(\zeta)\}$ (see (3.33)) and $\epsilon > 0$.

Then, for large N we have

$$
\begin{aligned}
&P\left(\left|\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}\frac{F''(c_j)}{2}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}-\nu-\zeta)^2\right| > \epsilon\right)\\
&\leq P\left(\left|\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}\frac{F''(c_j)}{2}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}-\nu-\zeta)^2\right| > \epsilon_N\right)\\
&\leq P\left(\left|\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}\frac{M}{2}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}-\nu-\zeta)^2\right| > \epsilon_N\right)\\
&= P\left(\left|\sum_{j=1}^{m}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}-\nu-\zeta)^2\right| > \frac{2mD_1^2}{\log(N)\sqrt{N}M}\right)\\
&\leq \sum_{j=1}^{m} P\left(\left|(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}-\nu-\zeta)^2\right| > \frac{2D_1^2}{\log(N)\sqrt{N}M}\right)\\
&\overset{(\star)}{=} mP\left(\left|\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}-\nu-\zeta\right| > \underbrace{\sqrt{\frac{2D_1^2}{\log(N)\sqrt{N}M}}}_{:=\tilde{\epsilon}_N}\right)\\
&\overset{(B.8)}{\leq} m\left((2+4m)\exp\left\{-2\frac{N}{m}\delta^2_{\tilde{\epsilon}_N/2,n}\right\} + \exp\{-2N\Delta^2_{2,\tilde{\epsilon}_N/2,N}\} + \exp\left\{-2N\Delta^2_{3,\tilde{\epsilon}_N/2,N}\right\}\right)\\
&\overset{(B.10)}{\leq} m\left((2+4m)\exp\left\{-2\frac{N}{m}\frac{2}{M\sqrt{N}\log(N)}\right\} + 2\exp\left\{-2N\frac{2}{M\sqrt{N}\log(N)}\right\}\right)\\
&= m\left((2+4m)\exp\left\{-2\frac{\sqrt{N}}{m}\frac{2}{M\log(N)}\right\} + 2\exp\left\{-2\sqrt{N}\frac{2}{M\log(N)}\right\}\right)\\
&\longrightarrow 0 \iff m = o(N^{1/3}).
\end{aligned}
$$

In $(\star)$ we used the fact that $\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}$ are identically distributed. Hence, the error term $\Delta_N^{(2)}$ converges to zero in probability.

3. Analogously to 2., the error term $\Delta_N^{(3)}$ from (B.15) converges to zero in probability.

4. We will utilize the proof of Lemma 2.5.4E in [79] where stochastic sequences $K_{N,j}$, $j = 1,\dots,m$, and deterministic sequences $\beta_{N,j} = \beta_{N,1}$, $j = 1,\dots,m$, are defined with the following properties:

$$
\begin{aligned}
&\beta_{N,1} = \mathcal{O}\left(n^{-3/4}\right),\\
&P\left(K_{N,j} \geq \gamma_N\right) = \mathcal{O}\left(n^{-3/2}\right),\\
&\gamma_N = c_1 n^{-3/4}\left(\log(n)^{(1/2)(q+1)}\right),
\end{aligned}
$$

where $q \geq 1/2$ and c_1 is a constant, which depends on $F'(\nu)$. Moreover, we will use Lemma 9 where $|\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}-\theta| =: |x_j| \leq \epsilon_n$ a.s. with $\theta := (\nu+\zeta)$. Then with $\epsilon > 0$,

$\epsilon_N := \log(N)^{-1}$ and large N we get for the error term $\Delta_N^{(4)}$ in (B.16)

$$
\begin{aligned}
& P\left(\left|\sqrt{N}\Delta_N^{(4)}\right| > \epsilon\right) \le P\left(\left|\sqrt{N}\Delta_N^{(4)}\right| > \epsilon_N\right) \\
&= P\left(\left|\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}(\widehat{F}_{N,j}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\nu+\zeta)) - (F(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}) - F(\nu+\zeta))\right| > \epsilon_N\right) \\
&\le P\left(\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}\left|(\widehat{F}_{N,j}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\nu+\zeta)) - (F(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}) - F(\nu+\zeta))\right| > \epsilon_N\right) \\
&\le \sum_{j=1}^{m} P\left(\sqrt{N}\left|(\widehat{F}_{N,j}(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}) - \widehat{F}_{N,j}(\nu+\zeta)) - (F(\widehat{\nu}_{N,j}+\widehat{\zeta}_{N,n}) - F(\nu+\zeta))\right| > \epsilon_N\right) \\
&\overset{(\triangle)}{=} mP\left(\sqrt{N}\left|(\widehat{F}_{N,1}(\widehat{\nu}_{N,1}+\widehat{\zeta}_{N,n}) - \widehat{F}_{N,1}(\nu+\zeta)) - (F(\widehat{\nu}_{N,1}+\widehat{\zeta}_{N,n}) - F(\nu+\zeta))\right| > \epsilon_N\right) \\
&\le mP\left(\sqrt{N}\underbrace{\sup_{|x_1|\le\epsilon_n}\left|(\widehat{F}_{N,1}(\theta+x_1) - \widehat{F}_{N,1}(\theta)) - (F(\theta+x_1) - F(\theta))\right|}_{\le K_{N,1}+\beta_{N,1},\ \text{see [79]}} > \epsilon_N\right) \\
&\le mP\left(K_{N,1}+\beta_{N,1} > \frac{\epsilon_N}{\sqrt{N}}\right) \\
&= mP\left(K_{N,1} > \frac{\epsilon_N}{\sqrt{N}} - \mathcal{O}\left(n^{-3/4}\right)\right) \\
&\le mP\left(K_{N,1} > \gamma_N\right) \qquad \left(\Leftrightarrow m = o\left(N^{1/3}\right)\right) \\
&\le m\mathcal{O}\left(n^{-3/2}\right) \longrightarrow 0.
\end{aligned}
$$

Equality $(\triangle)$ holds since the underlying random variables are exchangeable. Hence, the error term $\Delta_N^{(4)}$ converges to zero in probability.

5. Analogously to 4, the error term $\Delta_N^{(5)}$ from (B.17)converges to zero in probability.

6. The following result is due to the well known Bahadur representation of the median and further results on the behaviour of the corresponding remainder term in Theorems 2.5.5B and 2.5.5C in [79]. We will prove the convergence in mean square. Let $\epsilon > 0$, $m = o\left(N^{1/3}\right)$ and $n = N/m$. Then we get for the error term $\Delta_N^{(6)}$ in (B.18):

$$
\begin{aligned}
E\left(\left|\sqrt{N}\Delta_N^{(6)} - 0\right|^2\right) &= E\left(\left|\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}\Delta_{N,i}^{(6)}\right|^2\right) \\
&= Var\left(\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}\Delta_{N,j}^{(6)}\right) + E\left(\sqrt{N}\frac{1}{m}\sum_{j=1}^{m}\Delta_{N,j}^{(6)}\right)^2 \\
&= \frac{N}{m}Var\left(\Delta_{N,1}^{(6)}\right) + E\left(\sqrt{N}\Delta_{N,1}^{(6)}\right)^2 \\
&\le \frac{N}{m}E\left(\left(\Delta_{N,1}^{(6)}\right)^2\right) + E\left(\left(\sqrt{N}\Delta_{N,1}^{(6)}\right)^2\right)
\end{aligned}
$$

$$= N\frac{m+1}{m}\frac{\sqrt{\frac{1}{2\pi}}+o\left(n^{-3/4+\epsilon}\right)}{n^{3/2}F'(\nu)^2} \longrightarrow 0.$$

The convergence in mean square implies convergence of $\Delta_N^{(6)}$ to zero in probability.

Combining the results 1. - 6. for the error terms $\Delta_N^{(1)}, \dots, \Delta_N^{(6)}$ the Theorem 8 is proven. □

Proof of Corollary 2.
We have that

$$\sqrt{N}\left(\widehat{\zeta}_{N,n}-\zeta\right) = \sqrt{N}\Bigg(\frac{1/2-\left(\widehat{F}_N(\nu+\zeta)-\widehat{F}_N(\nu-\zeta)\right)}{G'(\zeta)} - \frac{F'(\nu-\zeta)-F'(\nu+\zeta)}{G'(\zeta)}\frac{1/2-\widehat{F}_N(\nu)}{F'(\nu)}\Bigg) + \sqrt{N}\Delta_N.$$

The term $\sqrt{N}\Delta_N$ converges to zero in probability due to Theorem 8. The remaining term is the same as in the case of the ordinary sample MAD and has asymptotic normal distribution with mean zero and variance ϑ^2, see [55]. The result is proven with Slutsky's Theorem. □

C Exact CDF of the MAD for odd sample size in the change-point scenario

Let $Y_t = X_t + hI_{t>t_1}$, $t = 1, \ldots, n = 2k-1$, $k \geq 1$, where X_t i.i.d., $h \geq 0$, $t_1 \leq k-1$. Hence, $Y_1, \ldots, Y_{t_1}$ are i.i.d. and so are $Y_{t_1+1}, \ldots, Y_n$. Let F_0 with $F_0(x) = P\left(X_1 \leq x\right)$ and F_h with $F_h(x) = P\left(X_{n-\lfloor n\tau\rfloor+1} + h \leq x\right)$ denote the corresponding CDFs, respectively. Then, using similar argumentations as in [60] the CDF of the sample MAD of $Y_1, \ldots, Y_n$ is given as follows:

$$
\begin{aligned}
F_{\mathrm{MAD}}(x) = \int_{-\infty}^{\infty} \Bigg[\Bigg\{ & \sum_{a=\max(0,t_1-k)}^{t_1} \sum_{\substack{b=\max(0,\\ t_1-a-1)}}^{t_1-a} \sum_{l=k}^{2k-1} \sum_{\substack{d=\max(0,\\ l-w+b-k)}}^{\min(b,l-w+1)} \sum_{w=l-k}^{k-1} \sum_{\substack{b=[k-1-s\\ -(k-1-w)]^+}}^{\min(k-1-s,w)} \\
& \binom{t_1}{a}\binom{2k-1-t_1}{k-1-a}\binom{t_1-a}{b}\binom{k-t_1+a}{k-1-b}\binom{a}{c}\binom{k-1-a}{w-c}\binom{b}{d}\binom{k-1-b}{l-1-w-d} \\
& \left[F_0(m+x) - F_0(m)\right]^c \left[F_h(m+x) - F_h(m)\right]^{w-c} \\
& \cdot \left[1 - F_0(m+x)\right]^{a-c} \left[1 - F_h(m+x)\right]^{k-1-a-w+c} \\
& \cdot \left[F_0(m) - F_0(m-x)\right]^d \left[F_h(m) - F_h(m-x)\right]^{l-1-w-d} \\
& \cdot F_0(m-x)^{b-d} F_h(m-x)^{k-b-l+w+d} \Bigg\} \Bigg/ \\
& \Bigg\{ \sum_{i=0}^{t_1} \sum_{\substack{j=\max(0,\\ t_1-i-1)}}^{\min(t_1-i,k-1)} \binom{t_1}{i}\binom{2k-1-t_1}{k-1-i}\binom{t_1-i}{j}\binom{k-t_1+i}{k-1-j} \\
& F_0(m)^i F_h(m)^{k-1-i} (1-F_0(m))^j (1-F_h(m))^{k-1-j} \Bigg\} \cdot f_{\mathrm{MED}}(m) \Bigg] \mathrm{d}m,
\end{aligned}
$$

with $x \geq 0$. The CDF F_{MED} of the sample median of $Y_1, \ldots, Y_n$ can be derived easily as follows:

$$
\begin{aligned}
F_{\mathrm{MED}}(x) &= P\left(\mathrm{med}\{Y_1, \ldots, Y_n\} \leq x\right) = P\left(\text{at least } k \text{ of the } Y_1, \ldots, Y_n \text{ are } \leq x\right) \\
&= \sum_{l=k}^{n} P\left(\text{exactly } l \text{ of the } Y_1, \ldots, Y_n \text{ are } \leq x\right)
\end{aligned}
$$

$$= \sum_{l=k}^{n} \sum_{\substack{i=\max(0,\\ l-2k+1+t_1)}}^{\min(t_1,l)} \binom{t_1}{i} \binom{2k-1-t_1}{l-i} F_0(x)^i F_h(x)^{l-i}$$
$$\cdot \left(1 - F_0(x)\right)^{t_1 - i} \left(1 - F_h(x)\right)^{2k-1-t_1-l+i}.$$

The density function f_{MED} is the derivative of F_{MED}.
Alternatively, one can use the paper of [31], where the exact CDF of order statistics was derived, in order to determine the CDF F_{MAD}.

D Figures

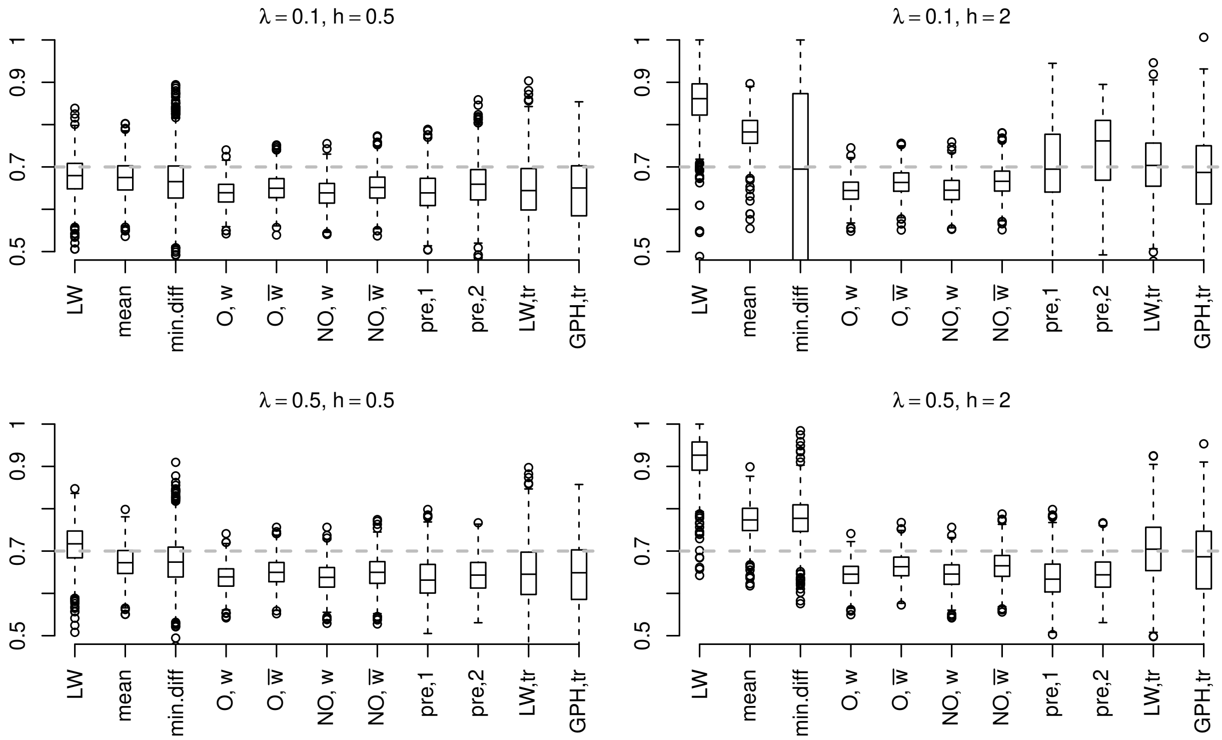

Figure D.1: Boxplots of the estimators based on the local Whittle estimator for the Hurst parameter $H = 0.7$ in time series with one jump of height h after a proportion of λ of the data, each based on 1000 simulation runs with $N = 1000$ realizations of Pareto(3,1)-transformed fGn.

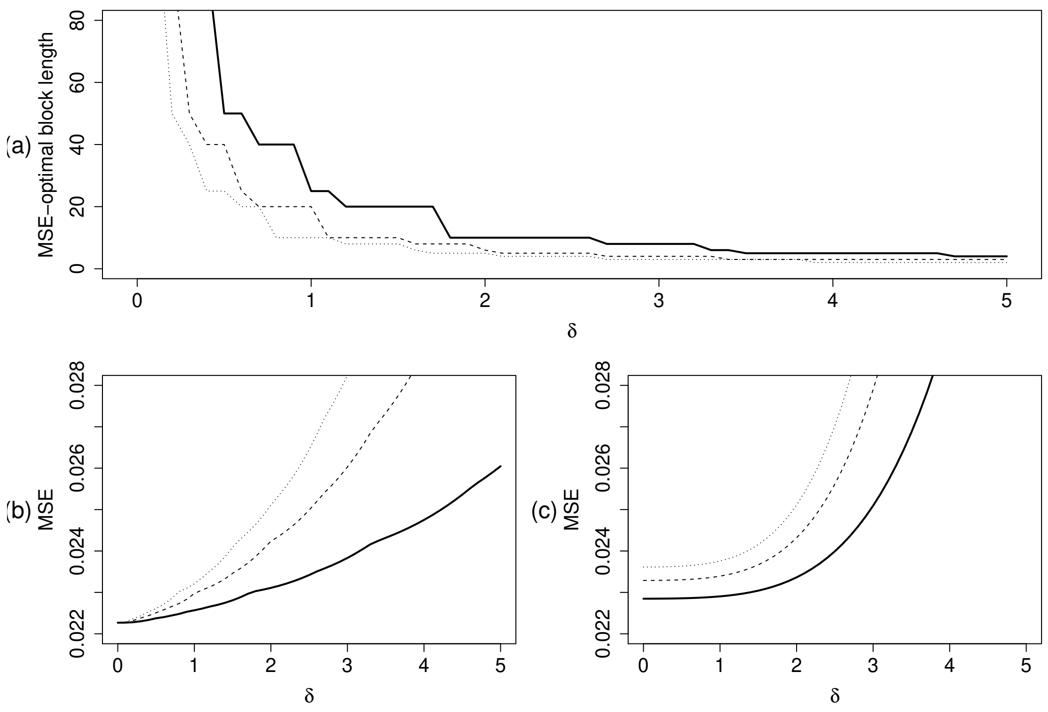

Figure D.2: (a) MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$, (b) MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ and (c) MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for $K = 1$ (—), $K = 3$ (- - -) and $K = 5$ (• • •) with $N = 1000$, $Y_t = X_t + \sum_{k=1}^{K} hI_{t \geq t_k}$, where $X_t \sim t_5$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$.

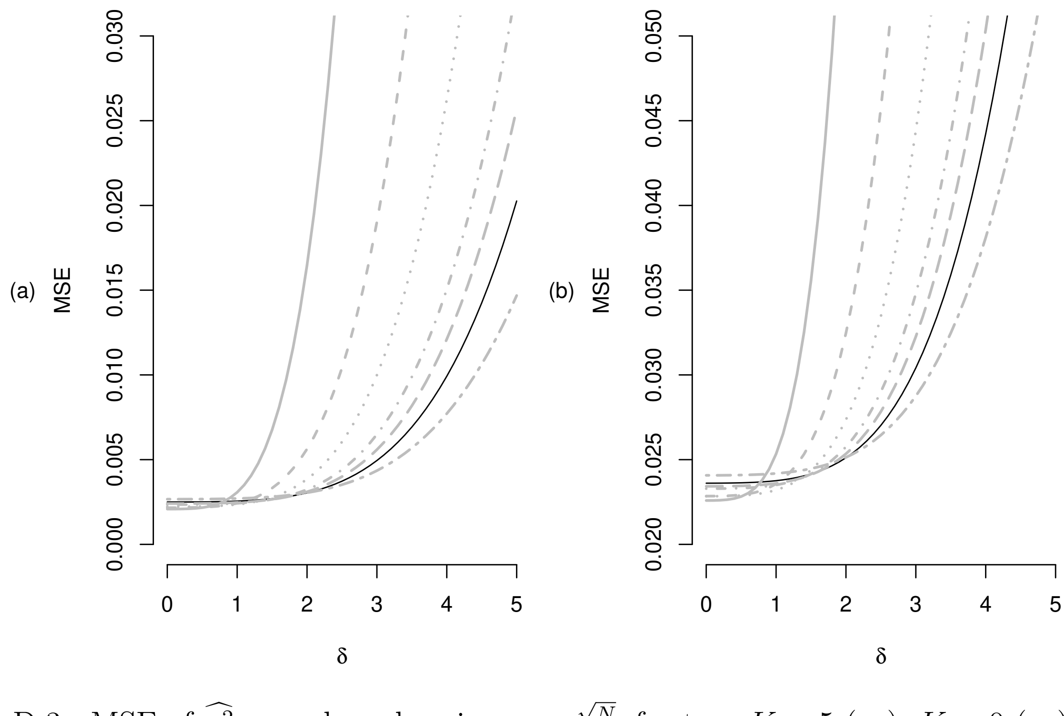

Figure D.3: MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for true $K = 5$ (—), $K = 0$ (—) $K = 1$ (- - -) $K = 2$ (· · · ·) $K = 3$ (- · · -) $K = 4$ (– – –) and $K = 6$ (– - –) with $N = 1000$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$, $Y_t = X_t + \sum_{k=1}^{5} hI_{t \geq t_k}$, where (a) $X_t \sim \mathrm{N}(0, 1)$ and (b) $X_t \sim t_3$.

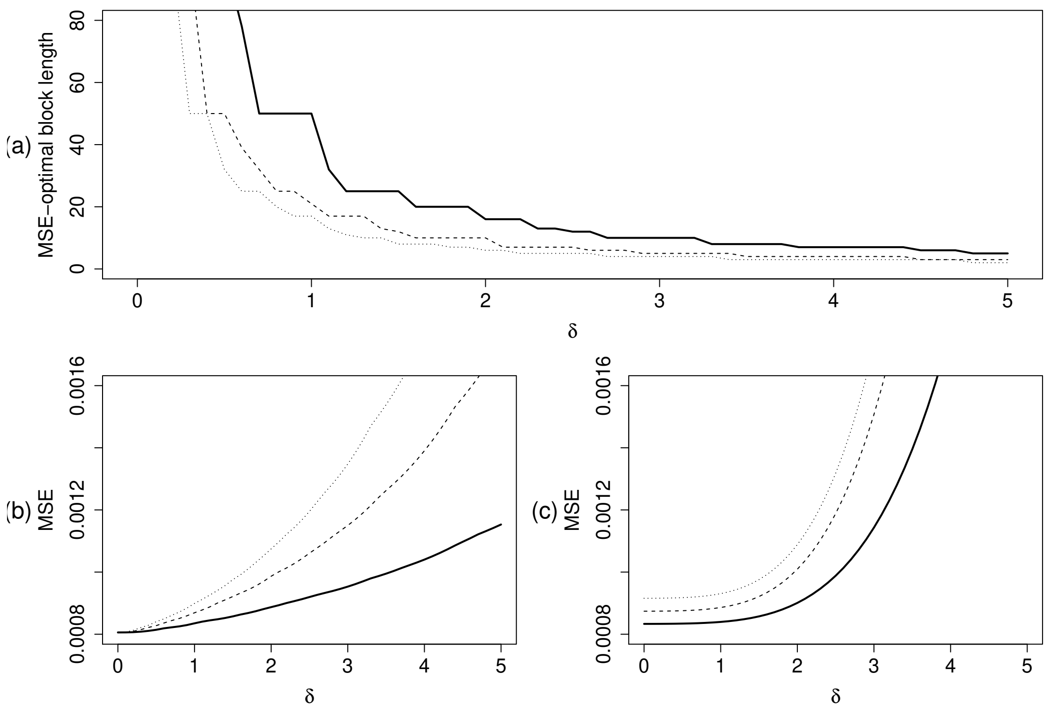

Figure D.4: (a) MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$, (b) MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ and (c) MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for $K = 1$ (—), $K = 3$ (- - -) and $K = 5$ (• • •) with $N = 2500$, $Y_t = X_t + \sum_{k=1}^{K} hI_{t \geq t_k}$, where $X_t \sim \text{N}(0, 1)$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$.

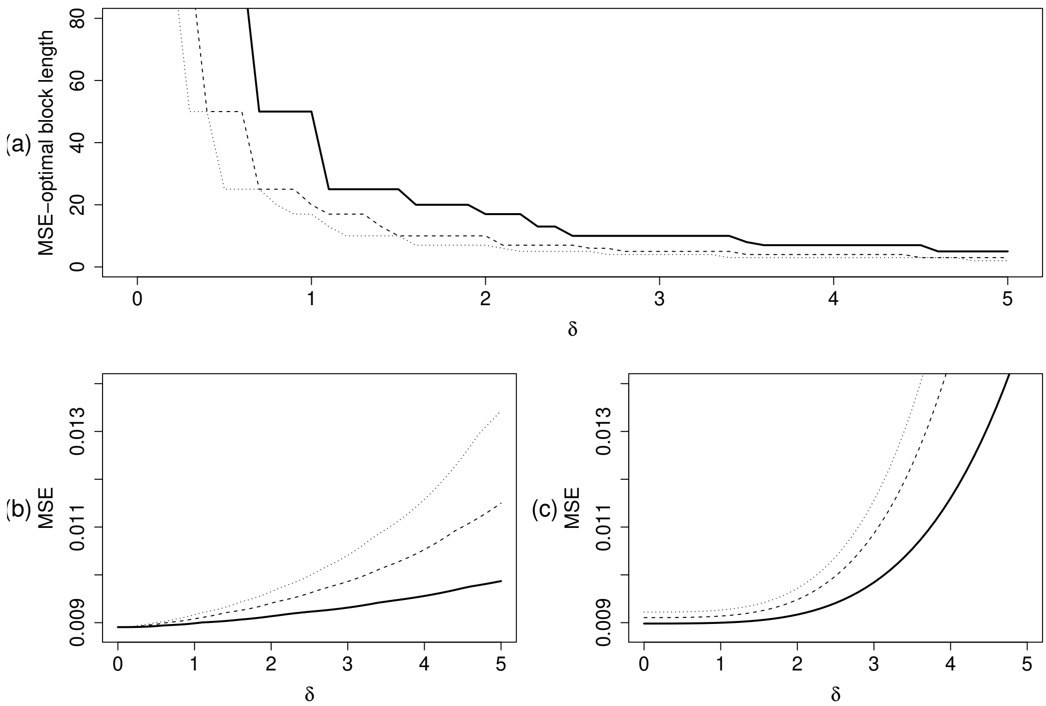

Figure D.5: (a) MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$, (b) MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ and (c) MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for $K = 1$ (—), $K = 3$ (- - -) and $K = 5$ (• • •) with $N = 2500$, $Y_t = X_t + \sum_{k=1}^{K} h I_{t \geq t_k}$, where $X_t \sim t_5$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$.

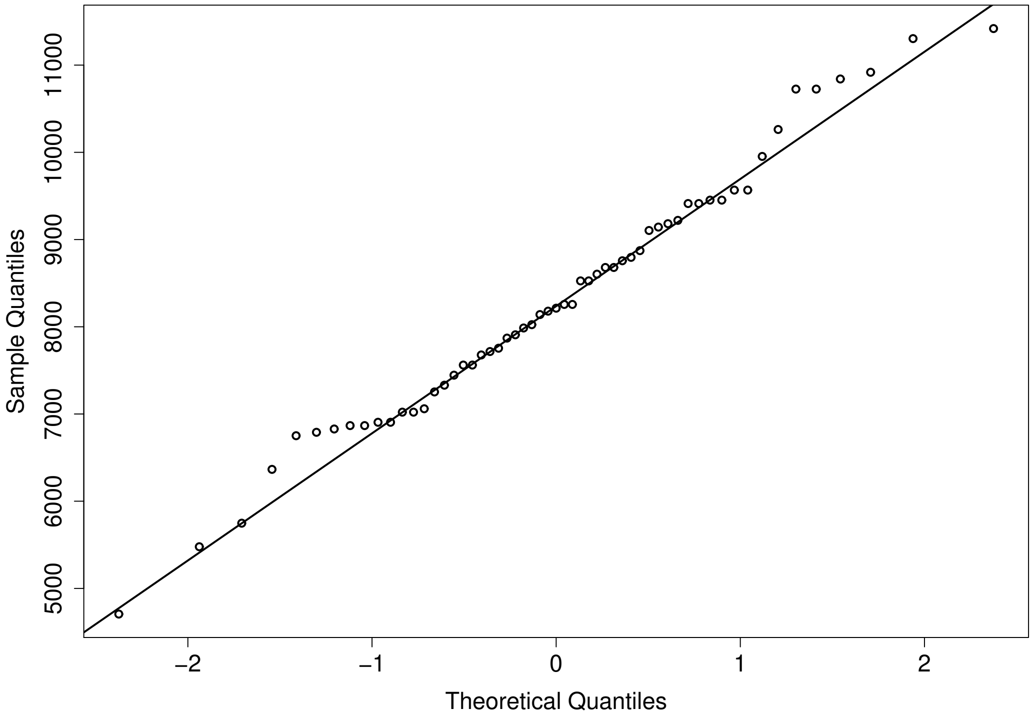

Figure D.6: Q-Q plot of the maximal monthly discharge of the Nile river in the period 1871 – 1984.

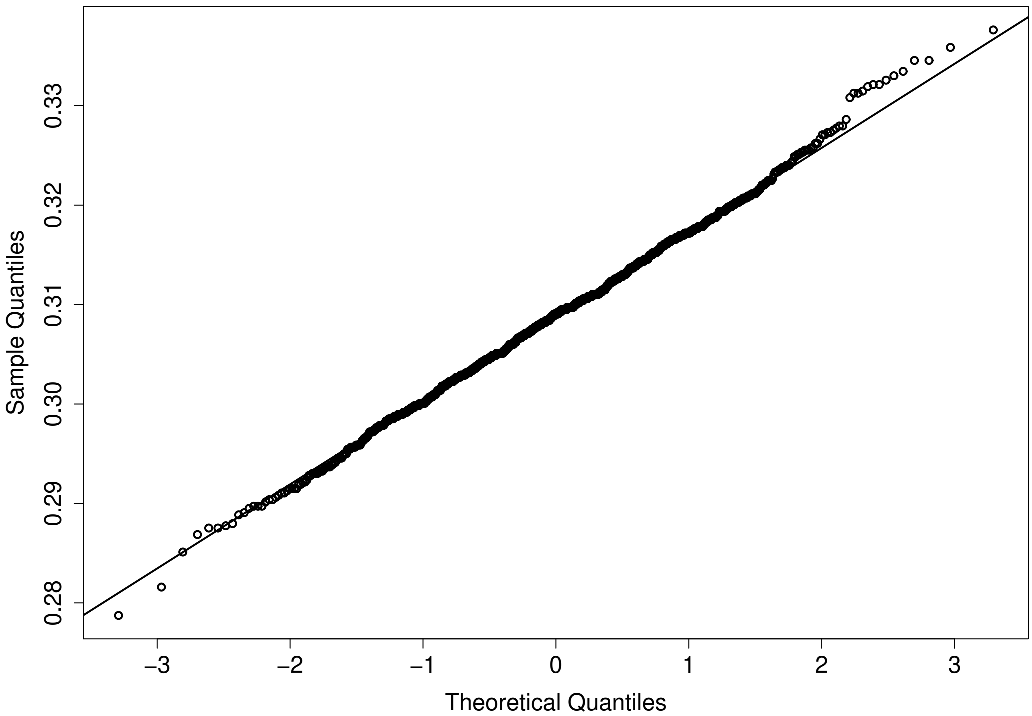

Figure D.7: Q-Q plot of a pixel with a virus adhesion from the PAMONO data.

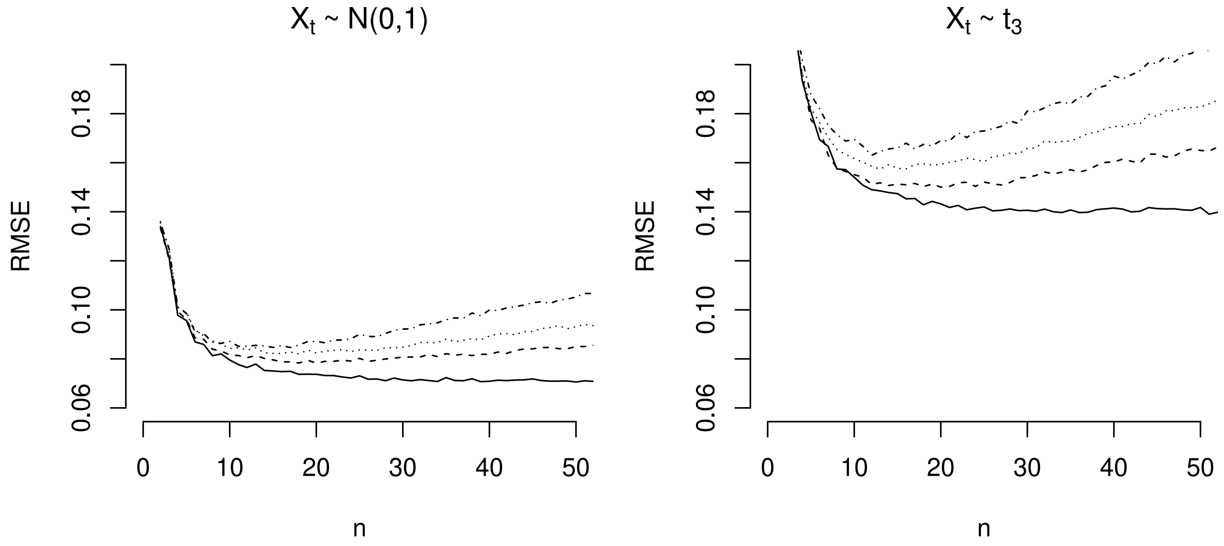

Figure D.8: RMSE of $\widehat{\sigma}_{\text{M,mod}}$ for $K = 1$, $h = 1\sigma$, $g = 6$ (——) $K = 2$, $h = 3\sigma$, $g = 6$ (- - -), $K = 3$, $h = 3\sigma$, $g = 10$ (· · ·) and $K = 4$, $h = 5\sigma$, $g = 10$ (- · -). $N = 2500$ observations from $Y_t = X_t + \sum_{k=1}^{K} h\sigma I_{t \geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \text{N}(g\sigma, 0.5)$, $X_t \sim \text{N}(0,1)$ (left) and $X_t \sim \text{t}_3$ (right).

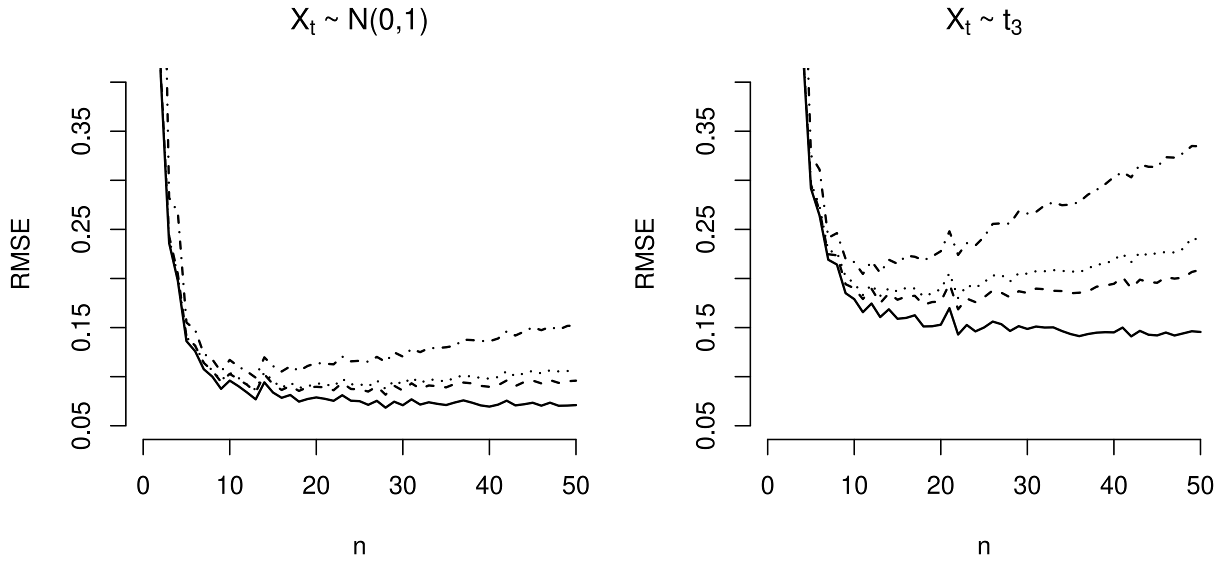

Figure D.9: RMSE of $\widehat{\sigma}_{\text{M,me}}$ for $K = 1$, $h = 1$, $\gamma = 6\sigma$ (——) $K = 2$, $h = 3$, $\gamma = 6\sigma$ (- - -), $K = 3$, $h = 3$, $\gamma = 6\sigma$ (· · ·) and $K = 4$, $h = 5\sigma$, $\gamma = 10$ (- · -). $N = 2500$ observations from $Y_t = X_t + \sum_{k=1}^{K} h\sigma I_{t \geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \text{N}(\gamma\sigma, 0.5)$, $X_t \sim \text{N}(0,1)$ (left) and $X_t \sim \text{t}_3$ (right).

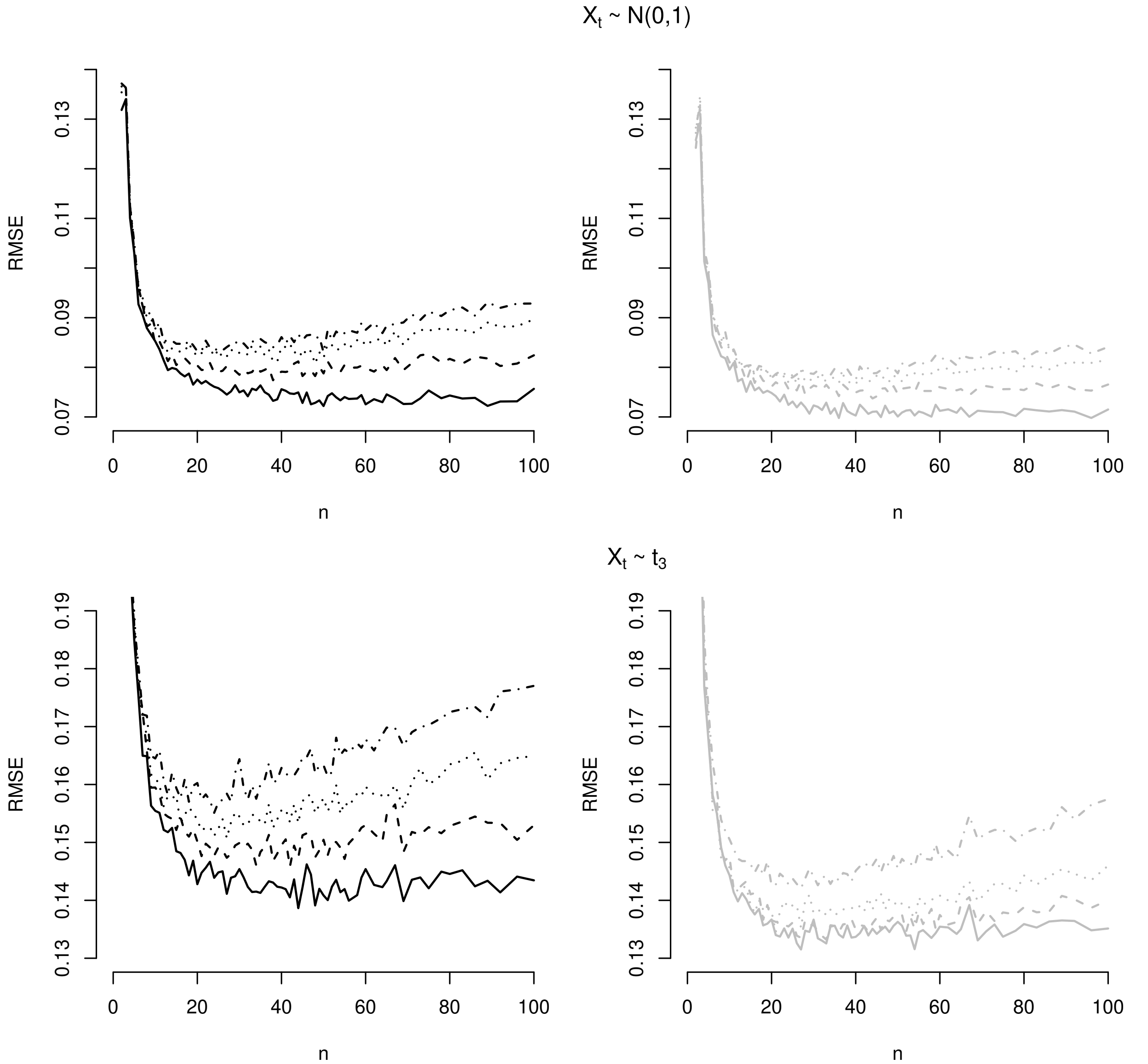

Figure D.10: RMSE of $\widehat{\sigma}_{\text{M,med}}$ (black) and $\widehat{\sigma}^{0.5}_{\text{M,tr}}$ (grey) for $K = 1$, $h = 1\sigma$, $g = 6$ (——) $K = 2$, $h = 3\sigma$, $g = 6$ (- - -), $K = 3$, $h = 3\sigma$, $g = 10$ (· · ·) and $K = 4$, $h = 5\sigma$, $g = 10$ (- · · -). $N = 2500$ observations from $Y_t = X_t + \sum_{k=1}^{K} h\sigma I_{t \geq t_k} + \gamma_t U_t$ with 5% outliers, where $\gamma_t \sim \text{N}(g\sigma, 0.5)$, $X_t \sim \text{N}(0, 1)$ (upper panel) and $X_t \sim \text{t}_3$ (lower panel).

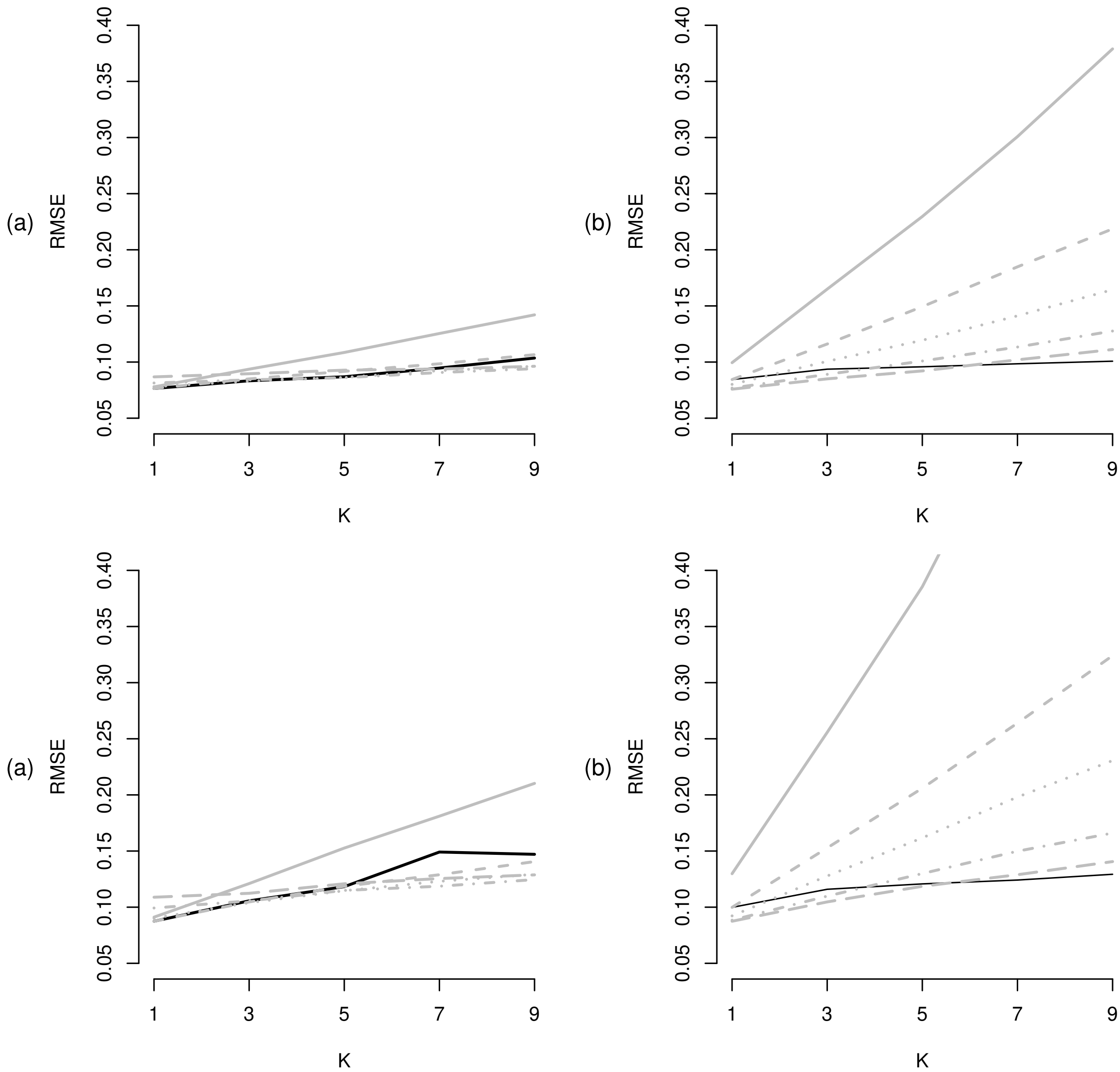

Figure D.11: RMSE of (a) $\widehat{\sigma}^2_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor\sqrt{N}/(K+1)\rfloor, 2\}$ and $\widetilde{\widehat{\sigma}}^2_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor\sqrt{N}/(\widetilde{K}+1)\rfloor, 2\}$, (b) $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(K+1)\rfloor, 2\}$ and $\widetilde{\widehat{\sigma}}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(\widetilde{K}+1)\rfloor, 2\}$ based on true K (——), and on $\widetilde{K} = 0$ (——), $\widetilde{K} = 1$ (- - -), $\widetilde{K} = 2$ (· · ·), $\widetilde{K} = 4$ (- · -), $\widetilde{K} = 6$ (— — —) for observations from $Y_t = X_t + \sum_{k=1}^{K} K\sigma I_{t\geq t_k} + \gamma_t U_t$, where (upper panel) $N = 1000$, $K = 3$, $X_t \sim t_3$ and (lower panel) $N = 1000$, $K = 5$, $X_t \sim \mathrm{N}(0,1)$, with 5% outliers, $\gamma_t \sim \mathrm{N}(6\sigma, 0.5)$.

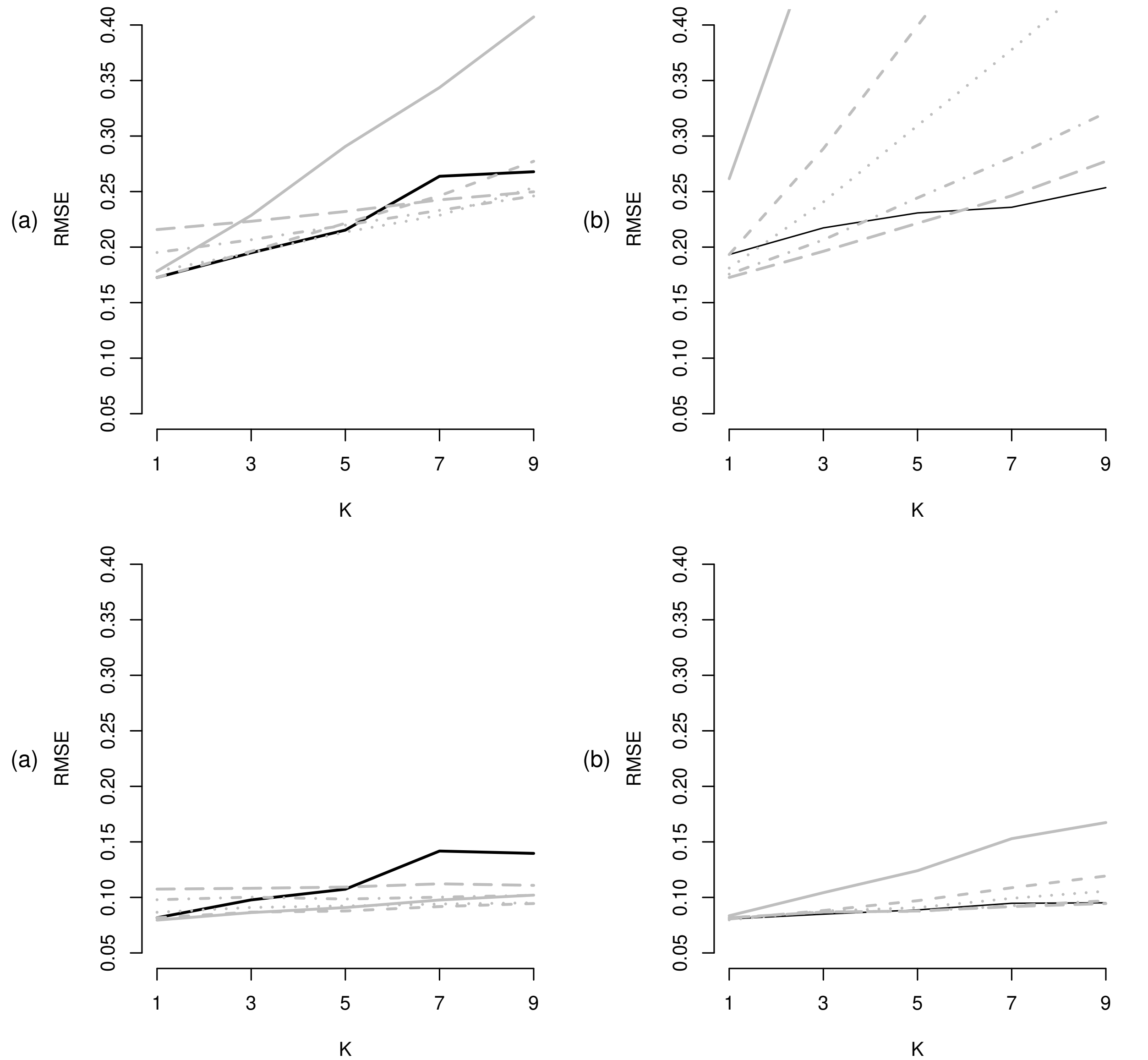

Figure D.12: RMSE of (a) $\widehat{\sigma}^2_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor\sqrt{N}/(K+1)\rfloor, 2\}$ and $\widetilde{\widehat{\sigma}}^2_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor\sqrt{N}/(\widetilde{K}+1)\rfloor, 2\}$, (b) $\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(K+1)\rfloor, 2\}$ and $\widetilde{\widehat{\sigma}}^{3.1}_{\mathrm{M,mod}}$ with block size $n = \max\{\lfloor N^{1-1/3.1}/(\widetilde{K}+1)\rfloor, 2\}$ based on true K (——), and on $\widetilde{K} = 0$ (——), $\widetilde{K} = 1$ (- - -), $\widetilde{K} = 2$ (· · ·), $\widetilde{K} = 4$ (- · · -), $\widetilde{K} = 6$ (— — —) for observations from $Y_t = X_t + \sum_{k=1}^{K} K\sigma I_{t \geq t_k} + \gamma_t U_t$, where (upper panel) $N = 2500$, $K = 3$ and (lower panel) $N = 1000$, $K = 5$, with 5% outliers, $\gamma_t \sim \mathrm{N}(6\sigma, 0.5)$, $X_t \sim \mathrm{N}(0, 1)$.

E Tables

h	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$
			$H = 0.5$ ($d = 0$)					$H = 0.7$ ($d = 0.2$)		
					$N = 1000$					
0	26.61	17.33	13.99	12.00	**9.35**	21.71	17.82	15.48	13.90	**10.59**
$\lambda = 0.1$										
0.5	26.44	17.45	14.04	12.67	**10.40**	21.65	17.79	15.44	13.87	**11.16**
1	26.62	18.02	14.15	13.46	**12.50**	21.50	18.03	15.72	14.21	**12.03**
2	26.91	19.03	**14.63**	15.35	16.18	21.79	18.34	15.78	15.15	**13.97**
$\lambda = 0.5$										
0.5	26.46	17.38	14.47	12.21	**10.20**	21.61	17.64	15.58	13.78	**11.00**
1	26.67	17.85	15.08	12.51	**11.98**	21.50	17.87	15.87	13.87	**11.80**
2	26.84	18.83	16.36	**13.80**	15.39	21.76	18.21	16.52	14.44	**13.76**
					$N = 1500$					
0	23.22	15.06	12.30	10.69	**8.07**	20.51	16.26	14.08	12.57	**9.62**
$\lambda = 0.1$										
0.5	23.37	15.15	12.77	10.95	**9.26**	20.55	16.22	14.25	12.64	**9.86**
1	23.30	15.24	13.24	**10.87**	10.95	20.53	16.31	14.35	12.64	**10.54**
2	23.56	15.88	14.60	**11.36**	13.41	20.54	16.52	15.14	12.98	**12.15**
$\lambda = 0.5$										
0.5	23.43	15.45	12.80	11.21	**9.13**	20.57	16.30	14.30	12.69	**9.69**
1	23.67	15.89	13.51	11.73	**10.33**	20.54	16.46	14.25	12.99	**10.43**
2	23.80	16.64	14.54	13.16	**12.93**	20.57	16.84	15.22	13.65	**11.98**

Table E.1: Estimated RMSE of the non-overlapping blocks estimator with different block sizes for the differencing parameters $d = 0$ ($H = 0.5$) and $d = 0.2$ ($H = 0.7$) in time series without ($h = 0$) and with change-point (jump of height h after a proportion of λ of the data, each based on 1000 simulation runs with $n = 1000$ and $n = 1500$ values from the ARFIMA(0, d, 1) model with $\theta = -0.6$), based on the local Whittle estimator.

h	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$	$\sqrt{N}$	$2\sqrt{N}+1$	$3\sqrt{N}+1$	$4\sqrt{N}+1$	$7\sqrt{N}+1$
	$H = 0.5$ ($d = 0$)					$H = 0.7$ ($d = 0.2$)				
	$N = 1000$									
$\lambda = 0.5$										
2	4.03	**3.86**	4.42	4.59	9.68	**3.60**	4.38	4.96	4.95	7.63
	$N = 1500$									
$\lambda = 0.5$										
2	**3.24**	3.32	4.02	4.71	7.55	**3.08**	4.23	4.58	5.08	6.32

Table E.2: Estimated RMSE of the non-overlapping blocks estimator with different block sizes for the differencing parameters $d = 0$ ($H = 0.5$) and $d = 0.2$ ($H = 0.7$) in time series without ($h = 0$) and with change-point (jump of height h after a proportion of λ of the data, each based on 1000 simulation runs with $n = 1000$ and $n = 1500$ values from fGn), based on the local Whittle estimator.

Distribution	$\widehat{\sigma}_{\text{M,me}}$	$\widehat{\sigma}_{\text{M,med}}$	$\widehat{\sigma}^{0.5}_{\text{M,tr}}$	$\widehat{\sigma}^{2}_{\text{M,mod}}$	$\widehat{\sigma}^{3.1}_{\text{M,mod}}$	$Q^{0.5}_{adj}$	$MS^{0.5}_{adj}$
			$K=0,\ g=0,\ \rho=0$				
Bias							
$N(0,1)$	-0.11	-0.08	0.01	0.14	-0.13	0.23	0.30
t_3	-1.11	-0.32	-0.04	-0.13	-0.31	-0.72	-0.60
t_5	-0.66	-0.24	-0.09	0.01	-0.18	-0.17	-0.08
$Gum(0,1)$	-0.57	-0.75	-1.28	0.03	-0.13	-0.35	-0.36
$Laplace(1,3)$	-3.36	-1.43	-2.28	0.69	0.89	0.50	0.68
RMSE							
$N(0,1)$	5.09	6.46	6.63	5.37	5.37	6.28	6.71
t_3	10.07	12.15	12.54	10.52	9.90	11.83	12.33
t_5	7.16	8.69	9.07	7.44	7.05	8.71	9.02
$Gum(0,1)$	7.07	8.91	8.92	7.50	7.25	8.31	8.75
$Laplace(1,3)$	27.64	34.23	33.15	29.33	27.35	29.23	30.59
			$K=3,\ h=3\sigma,\ g=0,\ \rho=0$				
Bias							
$N(0,1)$	1.90	1.80	1.78	1.42	3.85	0.81	0.91
t_3	5.52	1.95	2.72	2.36	8.59	2.75	2.47
t_5	2.47	1.49	2.60	1.53	5.75	1.36	1.32
$Gum(0,1)$	1.81	0.26	1.33	1.49	5.15	1.03	1.00
$Laplace(1,3)$	4.71	3.45	7.32	7.23	22.42	4.69	4.49
RMSE							
$N(0,1)$	6.21	8.52	7.61	7.25	7.05	6.68	7.08
t_3	12.95	15.00	14.02	12.80	14.38	12.31	12.69
t_5	8.66	10.74	10.50	9.40	10.05	8.70	9.20
$Gum(0,1)$	8.67	10.90	10.35	9.64	9.76	8.89	9.33
$Laplace(1,3)$	29.66	39.24	37.05	34.99	38.70	30.42	32.42
			$K=3,\ h=3\sigma,\ g=6,\ \rho=0.5$				
Bias							
$N(0,1)$	14.97	11.08	8.61	9.85	10.56	19.28	17.79
t_3	33.55	20.39	15.78	19.80	22.58	39.13	35.35
t_5	21.95	14.27	12.43	13.99	16.04	27.46	25.20
$Gum(0,1)$	20.37	12.66	10.97	13.63	14.59	26.97	24.43
$Laplace(1,3)$	71.11	48.86	40.05	51.27	60.27	93.65	85.13
RMSE							
$N(0,1)$	16.82	14.38	11.80	12.45	12.51	21.00	19.72
t_3	37.61	26.71	21.75	24.94	26.05	42.59	39.16
t_5	24.40	18.79	16.67	17.50	18.23	30.05	27.85
$Gum(0,1)$	23.09	17.44	15.64	17.46	17.20	29.35	27.12
$Laplace(1,3)$	80.98	65.27	55.89	63.34	68.85	102.36	94.31

Table E.3: Bias$\cdot 10^2$ and RMSE$\cdot 10^2$ of different estimators for $N = 500$ under independence.

Distribution	$\widehat{\sigma}_{\text{M,me}}$	$\widehat{\sigma}_{\text{M,med}}$	$\widehat{\sigma}^{0.5}_{\text{M,tr}}$	$\widehat{\sigma}^{2}_{\text{M,mod}}$	$\widehat{\sigma}^{3.1}_{\text{M,mod}}$	$Q^{0.5}_{adj}$	$MS^{0.5}_{adj}$
			$K=0,\ g=0,\ \rho=0$				
Bias							
N(0, 1)	-0.14	-0.19	-0.00	0.01	-0.04	-0.10	-0.10
t_3	0.28	-0.13	0.06	0.03	0.15	0.15	-0.02
t_5	0.31	0.01	0.01	0.25	0.11	0.05	0.02
Gum(0, 1)	0.30	0.19	0.08	-0.08	-0.17	-0.09	-0.18
Laplace(1, 3)	-2.27	-0.43	0.77	-0.20	-0.17	-0.03	-0.45
RMSE							
N(0, 1)	2.24	2.80	2.81	2.32	2.26	2.79	2.96
t_3	4.36	5.55	5.31	4.49	4.44	5.44	5.73
t_5	3.27	4.18	4.07	3.38	3.19	3.82	4.04
Gum(0, 1)	3.29	3.96	3.93	3.30	3.17	3.64	3.96
Laplace(1, 3)	12.57	15.16	15.14	12.80	12.77	13.52	14.31
			$K=3,\ h=3\sigma,\ g=0,\ \rho=0$				
Bias							
N(0, 1)	0.87	0.57	0.51	0.60	2.31	0.19	0.03
t_3	2.76	0.79	1.01	1.37	5.14	0.21	0.08
t_5	1.31	0.44	0.88	0.95	3.56	0.39	0.28
Gum(0, 1)	0.96	0.59	0.54	0.66	2.95	0.14	0.04
Laplace(1, 3)	4.68	3.20	2.53	2.99	13.45	0.43	-0.06
RMSE							
N(0, 1)	2.58	3.08	3.09	2.70	3.48	2.89	3.08
t_3	5.51	5.79	5.65	5.07	7.29	5.42	5.67
t_5	3.62	4.18	4.38	3.77	5.08	3.87	4.05
Gum(0, 1)	3.32	4.04	4.19	3.59	4.59	3.79	3.99
Laplace(1, 3)	12.88	15.38	15.53	13.43	19.12	13.36	14.22
			$K=3,\ h=3\sigma,\ g=6,\ \rho=0.5$				
Bias							
N(0, 1)	8.94	7.75	7.09	7.78	8.92	18.90	17.55
t_3	18.84	14.49	12.52	14.94	17.51	37.08	32.99
t_5	12.30	10.21	9.34	10.65	12.35	25.81	23.45
Gum(0, 1)	11.34	9.78	8.55	9.92	11.32	25.15	23.03
Laplace(1, 3)	46.80	39.63	34.92	39.86	48.32	89.22	80.07
RMSE							
N(0, 1)	9.36	8.45	7.89	8.33	9.37	19.29	17.97
t_3	19.60	15.85	13.85	15.90	18.35	37.79	33.76
t_5	12.89	11.20	10.41	11.38	13.04	26.37	24.06
Gum(0, 1)	11.89	10.74	9.59	10.62	11.95	25.67	23.62
Laplace(1, 3)	48.77	42.99	38.49	42.37	50.56	91.12	82.19

Table E.4: Bias$\cdot 10^2$ and RMSE$\cdot 10^2$ of different estimators for $N = 2500$ under independence.

Distribution	$\widehat{\sigma}_{\text{M,me}}$	$\widehat{\sigma}_{\text{M,med}}$	$\widehat{\sigma}^{0.5}_{\text{M,tr}}$	$\widehat{\sigma}^{2}_{\text{M,mod}}$	$\widehat{\sigma}^{3.1}_{\text{M,mod}}$	$Q^{0.5}_{adj}$	$MS^{0.5}_{adj}$
			$K=0,\ g=0,\ \rho=0$				
Bias							
N(0, 1)	26.83	14.21	10.38	13.28	13.14	20.91	19.57
t_3	52.64	24.58	17.66	22.50	24.25	41.72	37.24
t_5	35.94	18.10	13.36	16.72	16.99	28.79	26.36
Gum(0, 1)	36.25	18.12	13.26	16.50	16.91	28.72	26.22
Laplace(1, 3)	117.54	60.79	46.34	54.85	66.73	99.74	90.21
RMSE							
N(0, 1)	28.67	18.31	13.57	17.64	14.94	22.69	21.51
t_3	57.35	32.45	23.99	30.71	27.80	45.51	41.31
t_5	38.80	23.55	17.78	22.35	19.49	31.22	28.98
Gum(0, 1)	39.21	23.61	17.25	22.31	19.27	31.35	28.95
Laplace(1, 3)	127.61	80.05	61.70	75.47	76.39	108.76	99.80
			$N=1000,\ K=5,\ h=5\sigma,\ g=6,\ \rho=0.5$				
Bias							
N(0, 1)	16.07	10.86	8.76	10.43	11.16	20.19	18.81
t_3	33.58	19.27	15.51	18.79	21.34	39.33	34.94
t_5	21.92	13.91	11.56	13.40	15.03	27.22	24.77
Gum(0, 1)	21.09	13.45	10.53	13.65	14.26	26.83	24.54
Laplace(1, 3)	71.99	48.10	41.49	49.54	58.94	94.45	85.20
RMSE							
N(0, 1)	17.06	12.61	10.40	11.83	12.07	21.07	19.82
t_3	35.57	22.75	18.51	21.38	23.17	41.01	36.75
t_5	23.32	16.44	13.73	15.36	16.27	28.49	26.14
Gum(0, 1)	22.53	16.19	12.92	15.63	15.75	28.05	25.92
Laplace(1, 3)	77.31	57.13	48.64	56.26	63.81	99.01	90.13
			$N=2500,\ K=5,\ h=5\sigma,\ g=6,\ \rho=0.5$				
Bias							
N(0, 1)	11.52	8.74	7.29	8.47	9.58	19.23	17.89
t_3	24.31	15.83	13.10	15.92	18.15	37.22	33.10
t_5	15.65	11.54	9.67	11.64	13.34	26.45	24.11
Gum(0, 1)	15.05	10.71	8.96	10.95	11.96	25.43	23.35
Laplace(1, 3)	58.80	42.82	37.33	43.50	51.11	91.42	81.52
RMSE							
N(0, 1)	11.85	9.33	8.00	8.97	9.98	19.61	18.32
t_3	25.05	17.02	14.44	16.90	19.01	38.01	33.91
t_5	16.23	12.47	10.71	12.40	13.94	27.00	24.73
Gum(0, 1)	15.56	11.64	9.92	11.70	12.54	26.00	23.98
Laplace(1, 3)	60.58	46.11	40.89	46.15	53.43	93.17	83.49

Table E.5: Bias$\cdot 10^2$ and RMSE$\cdot 10^2$ of different estimators for $N = 500, 1000, 2500$, $K = 5$, $h = 5\sigma$, $g = 6$ and $\rho = 0.5$ under independence.

ϕ	$\widehat{\sigma}_{\mathrm{M,me}}$	$\widehat{\sigma}_{\mathrm{M,med}}$	$\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$	$\widehat{\sigma}^{2}_{\mathrm{M,mod}}$	$\widetilde{\widetilde{\sigma}}^{2}_{\mathrm{M,mod}}$	$\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$	$\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$	$Q^{0.5}_{adj}$	$MS^{0.5}_{adj}$
				$K=5,\ h=5\sigma,\ g=0,\ \rho=0$					
Bias									
0	1.81	0.69	0.81	0.75	4.37	2.88	5.59	0.45	0.30
0.1	0.33	-0.86	0.04	-0.79	4.05	2.39	5.51	-6.39	-6.46
0.3	-3.11	-4.85	-2.10	-4.84	3.44	1.46	5.02	-21.21	-21.27
0.5	-9.90	-13.11	-6.81	-12.72	2.52	-0.48	4.54	-40.61	-40.73
0.7	-28.07	-33.76	-20.81	-32.99	-1.97	-7.61	1.32	-72.63	-72.74
0.9	-108.55	-118.83	-92.64	-117.28	-36.67	-55.83	-25.56	-168.45	-168.56
RMSE									
0	3.07	3.15	3.23	2.98	5.25	3.94	6.39	2.97	3.08
0.1	2.51	3.20	3.03	2.82	4.98	3.57	6.38	6.92	7.05
0.3	4.05	5.84	3.82	5.58	4.64	3.16	5.97	21.34	21.42
0.5	10.30	13.50	7.67	13.06	4.30	3.33	6.02	40.66	40.79
0.7	28.27	33.96	21.23	33.16	5.11	8.72	5.32	72.65	72.76
0.9	108.63	118.91	92.83	117.35	37.94	56.34	27.66	168.46	168.57
				$K=5,\ h=5\sigma,\ g=6,\ \rho=0.5$					
Bias									
0	11.52	8.74	7.29	8.47	11.25	9.58	12.54	19.23	17.89
0.1	9.80	7.02	6.77	6.82	10.96	9.21	12.30	11.66	10.22
0.3	6.52	3.26	4.85	2.90	10.78	8.69	12.37	-4.44	-5.87
0.5	-0.14	-5.13	0.23	-5.07	10.11	6.97	12.17	-25.37	-26.70
0.7	-17.91	-25.46	-13.05	-24.90	7.22	1.18	10.95	-58.93	-60.14
0.9	-97.22	-109.75	-83.59	-108.87	-23.72	-44.21	-11.83	-156.20	-157.29
RMSE									
0	11.85	9.33	8.00	8.97	11.68	9.98	12.99	19.61	18.32
0.1	10.23	7.86	7.54	7.51	11.43	9.68	12.80	12.23	10.91
0.3	7.18	4.81	6.02	4.22	11.30	9.24	12.91	5.47	6.69
0.5	3.35	6.46	3.99	6.11	10.90	7.81	12.88	25.53	26.84
0.7	18.30	25.81	13.90	25.19	8.96	5.05	12.38	58.98	60.19
0.9	97.33	109.85	83.83	108.95	25.80	44.97	16.52	156.22	157.31

Table E.6: Bias$\cdot 10^2$ and RMSE$\cdot 10^2$ of different estimators for $N = 2500$, $K = 5$, $h = 5\sigma$ and different AR-parameters $\phi \in \{0, 0.1, \dots, 0.9\}$.

ϕ	$\widehat{\sigma}_{\mathrm{M,me}}$	$\widehat{\sigma}_{\mathrm{M,med}}$	$\widehat{\sigma}^{0.5}_{\mathrm{M,tr}}$	$\widehat{\sigma}^{2}_{\mathrm{M,mod}}$	$\widetilde{\widetilde{\sigma}}^{2}_{\mathrm{M,mod}}$	$\widehat{\sigma}^{3.1}_{\mathrm{M,mod}}$	$\widetilde{\widetilde{\sigma}}^{3.1}_{\mathrm{M,mod}}$	$Q^{0.5}_{adj}$	$MS^{0.5}_{adj}$
				$K=3,\ h=3\sigma,\ g=0,\ \rho=0$					
Bias									
0	1.56	1.18	1.35	1.04	3.62	3.26	4.40	0.51	0.47
0.1	-0.14	-0.67	0.46	-0.66	3.04	2.43	3.76	-6.43	-6.53
0.3	-4.18	-5.48	-2.49	-5.40	2.13	1.14	2.90	-21.18	-21.16
0.5	-12.02	-14.55	-8.51	-14.77	-0.37	-1.95	0.56	-40.52	-40.59
0.7	-31.82	-36.71	-24.00	-36.50	-6.82	-9.59	-4.84	-72.45	-72.56
0.9	-115.63	-123.14	-101.90	-123.29	-56.57	-65.21	-50.97	-168.32	-168.36
RMSE									
0	4.44	5.51	5.31	4.71	5.58	5.30	6.23	4.58	5.08
0.1	3.98	5.41	5.30	4.48	5.18	4.88	5.70	7.63	7.97
0.3	5.88	7.82	5.91	7.10	4.89	4.58	5.33	21.49	21.52
0.5	12.86	15.65	10.30	15.58	5.17	5.49	5.26	40.64	40.73
0.7	32.23	37.28	24.93	36.89	9.84	11.74	8.43	72.50	72.62
0.9	115.76	123.31	102.26	123.41	57.87	66.17	52.52	168.34	168.38
				$K=3,\ h=3\sigma,\ g=6,\ \rho=0.5$					
Bias									
0	11.70	9.25	7.75	9.01	10.12	9.56	10.58	19.37	18.10
0.1	10.27	7.31	7.04	7.27	9.85	9.05	10.47	11.76	10.36
0.3	6.08	2.47	4.41	2.37	9.34	8.35	10.04	-4.58	-5.82
0.5	-1.73	-6.35	-1.20	-6.72	7.55	6.29	8.65	-25.28	-26.61
0.7	-21.02	-28.16	-16.94	-28.52	1.50	-1.47	3.46	-58.48	-59.68
0.9	-103.50	-114.50	-93.43	-114.92	-45.65	-53.99	-39.29	-155.61	-156.72
RMSE									
0	12.70	11.14	9.71	10.42	11.11	10.62	11.65	20.27	19.13
0.1	11.47	9.55	9.08	8.89	10.90	10.19	11.52	12.99	11.73
0.3	7.88	6.48	7.11	5.48	10.52	9.59	11.20	6.84	7.73
0.5	5.67	8.93	6.38	8.54	9.37	8.43	10.44	25.66	26.98
0.7	21.86	28.96	18.42	29.07	7.45	7.49	8.49	58.61	59.81
0.9	103.78	114.73	93.87	115.09	47.47	55.39	41.65	155.65	156.76

Table E.7: Bias$\cdot 10^2$ and RMSE$\cdot 10^2$ of different estimators for $N = 1000$, $K = 3$, $h = 3\sigma$ and different AR-parameters $\phi \in \{0, 0.1, \ldots, 0.9\}$.

Bibliography

[1] Abbas, S., Fried, R., and Gather, U. Detection of local intensity changes in grayscale images with robust methods for time-series analysis. In *Solving Large Scale Learning Tasks. Challenges and Algorithms*. Springer, 2016, pp. 251–271.

[2] Angelova, J. A. On moments of sample mean and variance. *International Journal of Pure and Applied Mathematics 79*, 1 (2012), 67–85.

[3] Axt, I., Duerre, A., and Fried, R. Robust scale estimation under shifts in the mean. *Preprint* (2020).

[4] Axt, I., and Fried, R. On scale estimation under a change in the mean. *SFB 823 Discussion Paper 06/19, Dortmund* (2019).

[5] Baillie, R. T. Long memory processes and fractional integration in econometrics. *Journal of Econometrics 73*, 1 (1996), 5–59.

[6] Baranowski, R., Chen, Y., and Fryzlewicz, P. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*, 3 (2019), 649–672.

[7] Barkoulas, J. T., Baum, C. F., and Travlos, N. Long memory in the greek stock market. *Applied Financial Economics 10*, 2 (2000), 177–184.

[8] Beran, J. *Statistics for Long-Memory Processes*. Chapman & Hall, New York, 1994.

[9] Betken, A., et al. Change point estimation based on wilcoxon tests in the presence of long-range dependence. *Electronic Journal of Statistics 11*, 2 (2017), 3633–3672.

[10] Breidt, F. J., Crato, N., and De Lima, P. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics 83*, 1-2 (1998), 325–348.

[11] Chen, J., and Gupta, A. K. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Springer Science & Business Media, 2011.

[12] Cheung, Y.-W., and Lai, K. S. A search for long memory in international stock market returns. *Journal of International Money and Finance 14*, 4 (1995), 597–615.

[13] Croux, C., and Dehon, C. Robust estimation of location and scale. *Wiley StatsRef: Statistics Reference Online* (2014).

[14] Csorgo, M., and Horváth, L. *Limit Theorems in Change-Point Analysis*. John Wiley & Sons Chichester, 1997.

[15] Cutland, N. J., Kopp, P. E., and Willinger, W. Stock price returns and the joseph effect: a fractional version of the black-scholes model. In *Seminar on Stochastic Analysis, Random Fields and Applications* (1995), Springer, pp. 327–351.

[16] Dai, W., Ma, Y., Tong, T., and Zhu, L. Difference-based variance estimation in nonparametric regression with repeated measurement data. *Journal of Statistical Planning and Inference 163* (2015), 1–20.

[17] Dai, W., and Tong, T. Variance estimation in nonparametric regression with jump discontinuities. *Journal of Applied Statistics 41*, 3 (2014), 530–545.

[18] Davies, L., and Gather, U. The identification of multiple outliers. *Journal of the American Statistical Association 88*, 423 (1993), 782–792.

[19] Dehling, H., Rooch, A., and Taqqu, M. S. Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics 40*, 1 (2013), 153–173.

[20] Deo, R. S., and Hurvich, C. M. Linear trend with fractionally integrated errors. *Journal of Time Series Analysis 19*, 4 (1998), 379–397.

[21] Dette, H., Munk, A., and Wagner, T. Estimating the variance in nonparametric regression – what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*, 4 (1998), 751–764.

[22] Eichinger, B., Kirch, C., et al. A mosum procedure for the estimation of multiple random change points. *Bernoulli 24*, 1 (2018), 526–564.

[23] Erramilli, A., Narayan, O., and Willinger, W. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking 4*, 2 (1996), 209–223.

[24] Fox, J. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications, 2015.

[25] Galimberti, G., Pillati, M., and Soffritti, G. Robust regression trees based on m-estimators. *Statistica 67*, 2 (2007), 173–190.

[26] Gasser, T., Sroka, L., and Jennen-Steinmetz, C. Residual variance and residual pattern in nonlinear regression. *Biometrika 73*, 3 (1986), 625–633.

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

[28] Geweke, J., and Porter-Hudak, S. The estimation and application of long memory time series models. *Journal of Time Series Analysis 4*, 4 (1983), 221–238.

[29] Gil-Alana, L. A. Statistical modeling of the temperatures in the northern hemisphere using fractional integration techniques. *Journal of Climate 18*, 24 (2005), 5357–5369.

[30] Giraitis, L., Leipus, R., and Surgailis, D. The change-point problem for dependent observations. *Journal of Statistical Planning and Inference 53*, 3 (1996), 297–310.

[31] Glueck, D. H., Karimpour-Fard, A., Mandel, J., Hunter, L., and Muller, K. E. Fast computation by block permanents of cumulative distribution functions of order statistics from several populations. *Communications in Statistics – Theory and Methods 37*, 18 (2008), 2815–2824.

[32] Hall, P., Kay, J., and Titterinton, D. Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika 77*, 3 (1990), 521–528.

[33] Hardi, A. S., Kawai, K.-i., Lee, S., and Maekawa, K. Change point analysis of exchange rates using bootstrapping methods: An application to the indonesian rupiah 2000–2008. *Asia-Pacific Financial Markets 22*, 4 (2015), 429–444.

[34] Hassan, F. A. Historical nile floods and their implications for climatic change. *Science 212*, 4499 (1981), 1142–1145.

[35] Hassler, U., and Scheithauer, J. Detecting changes from short to long memory. *Statistical Papers 52*, 4 (2011), 847–870.

[36] Hens, T., and Schenk-Hoppé, K. R. *Handbook of Financial Markets: Dynamics and Evolution.* Elsevier, 2009.

[37] Hipel, K. W., and McLeod, A. I. *Time Series Modelling of Water Resources and Environmental Systems*, vol. 45. Elsevier, 1994.

[38] Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association 58*, 301 (1963), 13–30.

[39] Horváth, L., and Kokoszka, P. The effect of long-range dependence on change-point estimators. *Journal of Statistical Planning and Inference 64*, 1 (1997), 57–81.

[40] Hsu, C.-C. Long memory or structural changes: An empirical examination on inflation rates. *Economics Letters 88*, 2 (2005), 289–294.

[41] Hu, T.-C., Moricz, F., and Taylor, R. Strong laws of large numbers for arrays of rowwise independent random variables. *Acta Mathematica Hungarica 54*, 1-2 (1989), 153–162.

[42] Hurst, H. E. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng. 116* (1951), 770–799.

[43] Hurvich, C. M., Deo, R., and Brodsky, J. The mean squared error of geweke and porter-hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis 19*, 1 (1998), 19–46.

[44] Iacone, F. Local Whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis 31*, 1 (2010), 37–49.

[45] Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. Inference for single and multiple change-points in time series. *Journal of Time Series Analysis 34*, 4 (2013), 423–446.

[46] Jarušková, D. Some problems with application of change-point detection methods to environmental data. *Environmetrics: The Official Journal of the International Environmetrics Society 8*, 5 (1997), 469–483.

[47] Karagiannis, T., Faloutsos, M., and Riedi, R. H. Long-range dependence: now you see it, now you don't! In *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE* (2002), vol. 3, IEEE, pp. 2165–2169.

[48] Kokoszka, P., and Leipus, R. Change-point in the mean of dependent observations. *Statistics & Probability Letters 40*, 4 (1998), 385–393.

[49] Krämer, W., and Sibbertsen, P. Testing for structural change in the presence of long memory. *International Journal of Business and Economics, 1*, 3 (2002), 235–242.

[50] Kunsch, H. R. Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society, 1987* (1987), vol. 1, VNU Science Press, pp. 67–74.

[51] Lax, D. A. Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association 80*, 391 (1985), 736–741.

[52] Li, Q., and Mills, D. L. On the long-range dependence of packet round-trip delays in internet. In *ICC'98. 1998 IEEE International Conference on Communications. Conference Record. Affiliated with SUPERCOMM'98 (Cat. No. 98CH36220)* (1998), vol. 2, IEEE, pp. 1185–1191.

[53] Lo, A. W. Long-term memory in stock market prices. vol. 59, The Econometric Society, pp. 1279–1313.

[54] Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. *Robust Statistics: Theory and Methods (with R).* John Wiley & Sons, 2019.

[55] Mazumder, S., and Serfling, R. Bahadur representations for the median absolute deviation and its modifications. *Statistics & Probability Letters 79*, 16 (2009), 1774–1783.

[56] McCloskey, A., and Perron, P. Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory 29*, 6 (2013), 1196–1237.

[57] Miller, K. S. *Multidimensional Gaussian Distributions.* John Wiley & Sons Inc, 1964.

[58] Montanari, A., Rosso, R., and Taqqu, M. S. A seasonal fractional arima model applied to the nile river monthly flows at aswan. *Water Resources Research 36*, 5 (2000), 1249–1259.

[59] Munk, A., Bissantz, N., Wagner, T., and Freitag, G. On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*, 1 (2005), 19–41.

[60] Nagatsuka, H., Kawakami, H., Kamakura, T., and Yamamoto, H. The exact finite-sample distribution of the median absolute deviation about the median of continuous random variables. *Statistics & Probability Letters 83*, 4 (2013), 999–1005.

[61] Olver, F. W., Lozier, D. W., Boisvert, R. F., and Clark, C. W. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

[62] Percival, D. B., and Walden, A. T. *Wavelet Methods for Time Series Analysis*, vol. 4. Cambridge university press, 2000.

[63] Phillips, P. C., Shimotsu, K., et al. Local whittle estimation in nonstationary and unit root cases. *The Annals of Statistics 32*, 2 (2004), 656–692.

[64] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[65] Rachinger, H. Multiple breaks in long memory time series1. *Job Market Paper, Universidad Carlos III de Madrid* (2011).

[66] Reisen, V., Abraham, B., and Lopes, S. Estimation of parameters in arfima processes: A simulation study. *Communications in Statistics – Simulation and Computation 30*, 4 (2001), 787–803.

[67] Rice, J. Bandwidth choice for nonparametric regression. *The Annals of Statistics 12*, 4 (1984), 1215–1230.

[68] Robinson, P., and Rosenblatt, M. *Athens Conference on Applied Probability and Time Series Analysis: Volume II: Time Series Analysis In Memory of EJ Hannan*, vol. 115. Springer Science & Business Media, 2012.

[69] Robinson, P. M., et al. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics 23*, 5 (1995), 1630–1661.

[70] Robinson, P. M., et al. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics 23*, 3 (1995), 1048–1072.

[71] Rooch, A. Change-point tests for long-range dependent data. *Dissertation, Ruhr Universität Bochum* (2013).

[72] Rooch, A., Zelo, I., and Fried, R. Estimation methods for the LRD parameter under a change in the mean. *Statistical Papers 60*, 1 (2019), 313–347.

[73] Rousseeuw, P., and Yohai, V. Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis*. Springer, 1984, pp. 256–272.

[74] Rousseeuw, P. J., and Croux, C. Alternatives to the median absolute deviation. *Journal of the American Statistical Association 88*, 424 (1993), 1273–1283.

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

[76] Rousseeuw, P. J., and Leroy, A. M. A robust scale estimator based on the shortest half. *Statistica Neerlandica 42*, 2 (1988), 103–116.

[77] Schettlinger, K., Gelper, S., Gather, U., and Croux, C. Regression-based, regression-free and model-free approaches for robust online scale estimation. *Journal of Statistical Computation and Simulation 80*, 9 (2010), 1023–1040.

[78] Seber, G. A., and Lee, A. J. *Linear Regression Analysis*, vol. 329. John Wiley & Sons, 2012.

[79] Serfling, R. *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.

[80] Serfling, R., and Mazumder, S. Exponential probability inequality and convergence results for the median absolute deviation and its modifications. *Statistics & Probability Letters 79*, 16 (2009), 1767–1773.

[81] Shimotsu, K., and Phillips, P. C. Exact local whittle estimation of fractional integration. *The Annals of Statistics 33*, 4 (2005), 1890–1933.

[82] Sibbertsen, P. Long memory versus structural breaks: An overview. *Statistical Papers 45*, 4 (2004), 465–515.

[83] Siedhoff, D., Weichert, F., Libuschewski, P., and Timm, C. Detection and classification of nano-objects in biosensor data. In *Proceedings of the International Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB)* (2011), vol. 9.

[84] Siedhoff, D., Zybin, A., Shpacovitch, V., and Libuschewski, P. Pamono sensor data 200nm_10apr13.sfb876.doi:10.15467/e9ofqnvl6o. `https://sfb876.tu-dortmund.de/auto?self=%24e45xtiwhs0`, 2014.

[85] Smith, A. Level shifts and the illusion of long memory in economic time series. *Journal of Business & Economic Statistics 23*, 3 (2005), 321–335.

[86] Smith, R. L., Barnett, V., and Turkman, K. Long-range dependence and global warming. In *Applications of Statistics to Modeling the Earth's Climate System*. Citeseer, 1993, pp. 89–92.

[87] Syvitski, J. P., and Saito, Y. Morphodynamics of deltas under the influence of humans. *Global and Planetary Change 57*, 3-4 (2007), 261–282.

[88] Taqqu, M. S., and Teverovsky, V. On estimating the intensity of long-range dependence in finite and infinite variance time series. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (1998), 177 – 218.

[89] Taqqu, M. S., Teverovsky, V., and Willinger, W. Estimators for long-range dependence: an empirical study. *Fractals 3*, 04 (1995), 785–798.

[90] Tecuapetla-Gómez, I., and Munk, A. Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. *Scandinavian Journal of Statistics 44*, 2 (2017), 346–368.

[91] Tong, T., Ma, Y., and Wang, Y. Optimal variance estimation without estimating the mean function. *Bernoulli 19*, 5A (2013), 1839–1854.

[92] Velasco, C. Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis 20*, 1 (1999), 87–127.

[93] Velasco, C., and Robinson, P. M. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association 95*, 452 (2000), 1229–1243.

[94] Von Neumann, J., Kent, R., Bellinson, H., and Hart, B. t. The mean square successive difference. *The Annals of Mathematical Statistics 12*, 2 (1941), 153–162.

[95] Wang, L. Change-point estimation in long memory nonparametric models with applications. *Communications in Statistics – Simulation and Computation 37*, 1 (2007), 48–61.

[96] Wang, L. Change-in-mean problem for long memory time series models with applications. *Journal of Statistical Computation and Simulation 78*, 7 (2008), 653–668.

[97] Wang, L. Change-point detection with rank statistics in long-memory time-series models. *Australian & New Zealand Journal of Statistics 50*, 3 (2008), 241–256.

[98] Wang, W., Lin, L., and Yu, L. Optimal variance estimation based on lagged second-order difference in nonparametric regression. *Computational Statistics 32*, 3 (2017), 1047–1063.

[99] Whittle, P. Estimation and information in stationary time series. *Arkiv för matematik 2*, 5 (1953), 423–434.

[100] Willinger, W., Taqqu, M. S., and Teverovsky, V. Stock market prices and long-range dependence. *Finance and Stochastics 3*, 1 (1999), 1–13.

[101] Wornowizki, M., Fried, R., and Meintanis, S. G. Fourier methods for analyzing piecewise constant volatilities. *AStA Advances in Statistical Analysis 101*, 3 (2017), 289–308.

[102] Yohai, V. J., and Zamar, R. H. High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association 83*, 402 (1988), 406–413.