SFB 823

On scale estimation under shifts in the mean

Ieva Axt, Roland Fried

Discussion Paper

Nr. 6/2019

On scale estimation under shifts in the mean

Ieva Axt[1] and Roland Fried

In many situations it is crucial to estimate the variance properly. Ordinary variance estimators perform poorly in the presence of shifts in the mean. We investigate an approach based on non-overlapping blocks, which yields good results in this change-point scenario. We show the strong consistency and the asymptotic normality of such blocks-estimators of the variance under rather general conditions. For estimation of the standard deviation a blocks-estimator based on average standard deviations turns out to be preferable over the square root of the average variances. We provide recommendations on the appropriate choice of the block size and compare this blocks-approach with difference-based estimators. If level shifts occur rather frequently even better results can be obtained by adaptive trimming of the blocks under the assumption of normality.

Keywords: change-point; variance; estimation; blocks

1 Introduction

We consider a sequence of observations $(Y_t)_{t\geq 1}$ generated by

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

where $(X_t)_{t\in\mathbb{N}}$ are i.i.d. random variables with $E[X_t] = \mu$ and $Var[X_t^2] = \sigma^2$. W.l.o.g we will assume $\mu = 0$ in the following. This means that the observed data $y_1, ..., y_N$ are affected by K level shifts of possibly different heights at different time points $t_1, ..., t_K$. Our goal is estimation of the variance σ^2 or the standard deviation σ.

In Section 2 we analyse estimators of σ^2 from the sequence of observations $(Y_t)_{t\geq 1}$ by combining estimates obtained from splitting the data into several blocks. Without the need of explicit distributional assumptions the mean of the block-wise estimates turns out to be consistent if the size and the number of blocks increases and the fraction of jumps is asymptotically negligible. Otherwise, an adaptively trimmed mean of the block-wise estimates can be used if distributional assumptions can be made. Section 3 treats estimation of σ in a similar way. In Section 4 the estimation procedures are applied to real data sets, while Section 5 summarizes the results of this paper.

2 Estimation of the variance

When dealing with independent identically distributed data the sample variance is the common choice for estimation of σ^2. However, if we are aware of a possible presence of level shifts at unknown locations, it is reasonable to divide the sample $Y_1, ..., Y_N$ into m non-overlapping blocks of size $n = \lfloor N/m \rfloor$ and to calculate the average of the m sample variances derived from the different blocks. A similar approach has been used in Dai et al. (2015) in the context of repeated measurements data and in Rooch et al. (2016) for estimation of the Hurst parameter.

[1]Corresponding author. TU Dortmund University, Faculty of Statistics, 44221 Dortmund, Germany

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

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

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 $Y_{j,1},...,Y_{j,n}$ are the observations in the j-th block. We are interested in finding the MSE-optimal block size n to achieve desirable results under certain assumptions.

In what follows, we will concentrate on the situation where all jump heights are positive. This is a worse scenario than having both, positive and negative jumps, since the data is more spread in the former case resulting in a larger positive bias of most scale estimates.

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 Seber and Lee (2012). 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:

$$\begin{aligned} E\left(\widehat{\sigma^2}_{\text{Mean}}\right) &= \frac{1}{m}\left(tr\left(A\Sigma\right) + \sum_{j=1}^{B}\Theta_j^T A \Theta_j\right), \\ Var\left(\widehat{\sigma^2}_{\text{Mean}}\right) &= \frac{1}{m^2}\left(m\left(\mu_4 - 3\mu_2^2\right)a^T a + 2m\mu_2^2 tr(A^2) + \sum_{j=1}^{B}\left(4\mu_2\Theta_j^T A^2\Theta_j + 4\mu_3\Theta_j^T A a\right)\right), \end{aligned}$$

where $\mu_i = E\left(X_1^i\right)$, $\Sigma = \mu_2\mathbb{I}_n$, $A = \mathbb{I}_n - \frac{1}{n}1_n 1_n^T$, Θ_j contains the expected values of the random variables in the perturbed block $j = 1,...,B$ and a is a vector of the diagonal elements of A. The blocks-estimator (2) estimates the variance consistently if the number of blocks grows sufficiently fast.

Theorem 1. *Let $Y_1,...,Y_N$ from Model (1) be segregated into m blocks of size n. Let $B = B_N$ out of m blocks be contaminated by $\widetilde{K}_1,...,\widetilde{K}_B$ jumps, respectively, with $\sum_{j=1}^{B}\widetilde{K}_j = K$. Moreover, let $B\left(\sum_{k=1}^{K} h_k\right)^2 = o(m)$ and $m \to \infty$. Then $\widehat{\sigma^2}_{Mean} = \frac{1}{m}\sum_{j=1}^{m} S_j^2 \to \sigma^2$ almost surely.*

Proof. W.l.o.g. assume that the last B out of m blocks are contaminated by $\widetilde{K}_1,...,\widetilde{K}_B$ 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, $Y_{j,1},...,Y_{j,n}$ are the observations in the j-th block. Furthermore, let $\Delta_{j,t} = E\left(Y_{j,t}\right)$ and $\overline{\Delta}_j = \frac{1}{n}\sum_{t=1}^{n} E\left(Y_{j,t}\right)$, e.g. in block $j = m-B+1$ we have $\Delta_{m-B+1,t} = \sum_{k=1}^{\widetilde{K}_1} h_k I_{(m-B)\cdot n+t \geq t_k}$, $t = 1,...,n$, and $\overline{\Delta}_{m-B+1} = \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{\widetilde{K}_1} h_k I_{(m-B)\cdot n+t\geq t_k}$. Then we have

$$\begin{aligned} \widehat{\sigma^2}_{\text{Mean}} &= \frac{1}{m}\sum_{j=1}^{m} S_j^2 = \frac{1}{m}\sum_{j=1}^{m-B} S_{j,0}^2 + \frac{1}{m}\sum_{j=m-B+1}^{m} S_{j,h}^2 \\ &= \frac{1}{m}\sum_{j=1}^{m-B} S_{j,0}^2 + \frac{1}{m}\sum_{j=m-B+1}^{m}\frac{1}{n-1}\sum_{t=1}^{n}\left(X_{j,t} - \bar{X}_j + \Delta_{j,t} - \overline{\Delta}_j\right)^2 \\ &= \frac{1}{m}\sum_{j=1}^{m} S_{j,0}^2 + \frac{1}{m}\sum_{j=m-B+1}^{m}\frac{2}{n-1}\sum_{t=1}^{n}(X_{j,t} - \bar{X}_j)(\Delta_{j,t} - \overline{\Delta}_j) \\ &+ \frac{1}{m}\sum_{j=m-B+1}^{m}\frac{1}{n-1}\sum_{t=1}^{n}\left(\Delta_{j,t} - \overline{\Delta}_j\right)^2. \end{aligned} \tag{3}$$

For the second term in the last equation (3) we have

$$
\begin{aligned}
&\left|\frac{2}{m(n-1)}\sum_{j=m-B+1}^{m}\sum_{t=1}^{n}(X_{j,t}-\bar{X}_j)(\Delta_{j,t}-\overline{\Delta}_j)\right| \\
&\le \frac{2}{m(n-1)}\sum_{j=m-B+1}^{m}\sum_{t=1}^{n}\left|(X_{j,t}-\bar{X}_j)\right|\left|(\Delta_{j,t}-\overline{\Delta}_j)\right| \\
&\le \frac{2}{m(n-1)}\sum_{j=m-B+1}^{m}\sum_{t=1}^{n}\left|(X_{j,t}-\bar{X}_j)\right|\left|\sum_{k=1}^{K}h_k\right| \\
&= \underbrace{B\left|\sum_{k=1}^{K}h_k\right|}_{=o(m)}\frac{2}{m}\frac{n}{n-1}\qquad\underbrace{\frac{1}{B}\sum_{j=m-B+1}^{m}\frac{1}{n}\sum_{t=1}^{n}\left|(X_{j,t}-\bar{X}_j)\right|}_{\to E(|X_1|)\text{ if }n\to\infty\text{, see Hu et al. (1989), otherwise }<\infty}\qquad\longrightarrow 0.
\end{aligned}
$$

The following is valid for the third term in (3):

$$
\begin{aligned}
\frac{1}{m}\sum_{j=m-B+1}^{m}\frac{1}{n-1}\sum_{t=1}^{n}\left(\Delta_{j,t}-\overline{\Delta}_j\right)^2 &\le \frac{1}{m}\sum_{j=m-B+1}^{m}\frac{1}{n-1}\sum_{t=1}^{n}\left(\sum_{k=1}^{K}h_k\right)^2 \\
&= \frac{B}{m}\frac{n}{n-1}\left(\sum_{k=1}^{K}h_k\right)^2\longrightarrow 0.
\end{aligned}
$$

The first term of the last equation (3) converges almost surely to σ^2 due to the results on triangular arrays in Hu et al. (1989). Application of Slutsky's Theorem proves the result. □

Remark 2. 1. *If the jump heights are bounded by a constant $h \ge h_k$, $k = 1, ..., K$, the worst restriction arises if all heights equal this upper bound resulting in the constraint $B\left(\sum_{k=1}^{K}h_k\right)^2 = BK^2h^2 = o\,(m)$, which is worst for $B = K$ (at most one jump is located in one block). Consistency is thus guaranteed if the number of blocks grows faster than K^3.*

2. *The proof uses $B\left|\sum_{k=1}^{K}h_k\right| = \mathcal{O}\left(B\left|\sum_{k=1}^{K}h_k\right|^2\right) = o(m)$ since we assume that all jump heights are positive.*

3. *The result is valid for constant block size n as well as for $n \to \infty$.*

4. *Application of the ordinary Central Limit Theorem shows that $\widehat{\sigma^2}_{Mean}$ is asymptotically normal if the block size n is fixed, and its efficiency relative to the ordinary sample variance is $(\mu_4 - \sigma^4)\left[\mu_4 - \sigma^4(n-3)/(n-1)\right]^{-1}$ in case of i.i.d data with finite fourth moments, which is $(n-1)n^{-1}$ under normality. The next Theorem shows that $\widehat{\sigma^2}_{Mean}$ is not only asymptotically normal but even fully efficient in case of a growing block size.*

Theorem 3. *Assume that $Y_1 = X_1, ..., Y_N = X_N$ are segregated into m blocks of size n, with $m, n \to \infty$ such that $m = o\,(n)$, $n = o\,(N)$. Moreover, assume that $\mu_4 = E\left(X_t^4\right) < \infty$. Then we have*

$$
\sqrt{N}\left(\widehat{\sigma^2}_{Mean} - \sigma^2\right) \xrightarrow{d} N\left(0, \mu_4 - \sigma^4\right).
$$

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

$$\frac{\widehat{\sigma^2}_{\text{Mean}} - \sigma^2}{\sqrt{Var\left(\widehat{\sigma^2}_{\text{Mean}}\right)}} = \frac{\frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{t=1}^{n} \left(X_{j,t} - \overline{X}_j\right)^2 - \sigma^2}{\sqrt{\frac{1}{m}\left(\frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}\right)}}$$

$$= \sqrt{N} \frac{\frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{t=1}^{n} X_{j,t}^2 - \frac{n}{m(n-1)} \sum_{j=1}^{m} \overline{X}_j^2 - \sigma^2}{\sqrt{\mu_4 - \frac{\sigma^4(n-3)}{n-1}}}, \tag{4}$$

see Angelova (2012) for $Var\left(\widehat{\sigma^2}_{\text{Mean}}\right)$. For the second term of the numerator in (4) we have that

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

$$= \frac{\sqrt{mn}}{n}\frac{n}{n-1}\sigma^2 = \sqrt{\frac{m}{n}}\frac{n}{n-1}\sigma^2 \to 0, \tag{5}$$

since $m = o(n)$. Convergence of the term (5) in mean implies convergence in probability to zero. Application of the standard Central Limit Theorem to the remaining terms of (4) yields the desired result. □

Remark 4. *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 sixth moments.*

2.1 Choice of the block size

When considering blocks of length $n = 2$, the estimator $\widehat{\sigma}^2_{\text{Mean}}$ 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 Von Neumann et al. (1941), Rice et al. (1984), Gasser et al. (1986), Hall et al. (1990), Dette et al. (1998), Munk et al. (2005), Tong et al. (2013), Tecuapetla-Gómez and Munk (2017), among many others. An ordinary difference-based estimator of first order, which considers all $N-1$ consecutive differences, is (see e.g. Von Neumann et al. (1941)):

$$\widehat{\sigma^2}_{\text{Diff}} = \frac{1}{2(N-1)} \sum_{j=1}^{N-1} \left(Y_{j+1} - Y_j\right)^2. \tag{6}$$

W.l.o.g. we assume that the K jumps in the mean occur in the last K observations, since the position of the jumps has no influence on the performance of the estimator in (6). The expectation and the variance of quadratic forms can be calculated using Seber and Lee (2012), leading to the following formulae for the expectation and the variance of the estimator $\widehat{\sigma^2}_{\text{Diff}}$:

$$
\begin{aligned}
E\left(\widehat{\sigma^2}_{\text{Diff}}\right) &= \frac{1}{2(N-1)}\left(tr\left(A\Sigma\right)+\Theta^T A\Theta\right),\\
Var\left(\widehat{\sigma^2}_{\text{Diff}}\right) &= \frac{1}{4(N-1)^2}\left(\left(\mu_4-3\mu_2^2\right)a^T a+2\mu_2^2 tr(A^2)\right.\\
&\quad\left.+4\mu_2\Theta^T A^2\Theta+4\mu_3\Theta^T Aa\right),
\end{aligned}
$$

where $\Theta = E(Y) = (0,...,0,h_1,h_1+h_2,...,h_1+...+h_K)^T$, $\mu_i = E\left(X_1^i\right)$, $\Sigma = \mu_2\mathbb{I}_N$, $A = \tilde{A}^T\tilde{A}$, $\tilde{A}Y = (Y_2-Y_1,...,Y_N-Y_{N-1})^T$ and a is a vector of the diagonal elements of A.

In the following we investigate in which scenarios the estimator $\widehat{\sigma^2}_{\text{Mean}}$ is preferable over $\widehat{\sigma^2}_{\text{Diff}}$ defined in (6). All calculations in this paper have been performed with the statistical software `R`, version 3.5.2, R Core Team (2018).

For known jump positions the MSE of both variance estimators can be determined analytically. The position of the jump is relevant for the performance of the estimator $\widehat{\sigma^2}_{\text{Mean}}$. 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 $\delta \in \{0,0.1,0.2,...,4.9,5\}$, at positions sampled randomly from a uniform distribution on the values $\max_n\left(N-\lfloor N/n\rfloor n\right)+1,...,N-\max_n\left(N-\lfloor N/n\rfloor n\right)$ without replacement, and calculate the MSE for every reasonable block size $n \in \{2,3,4,...,\lfloor N/2\rfloor\}$. 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.

The first panel of Figure 1 shows the MSE-optimal block size n_{opt} 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. 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 $K = 5$ ($\cdot\cdot\cdot$), while larger blocks lead to better results in case of smaller or less jumps.

The second panel of Figure 1 depicts the MSE of $\widehat{\sigma^2}_{\text{Mean}}$ for the respective MSE-optimal block size n_{opt}. The MSE of $\widehat{\sigma^2}_{\text{Diff}}$ is shown in grey for comparison. When dealing with only one jump (—), the blocks-estimator outperforms the difference based estimator as long as the jump height is below $h = 4.5\sigma$. The corresponding block size satisfies $n_{\text{opt}} \geq 5$, depending on the height of the jump. For $K = 3$ (- - -), the estimator $\widehat{\sigma^2}_{\text{Mean}}$ yields better results when the jump height is less than $h \approx 3\cdot\sigma$, and for $K = 5$ when the jump height is at most $h \approx 2\sigma$.

Different values for the optimal block-size n_{opt} are obtained in different scenarios. 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. If the estimation of the variance is in the focus of the application, we suggest the following block size, depending on K:

$$
n = \max\left\{\left\lfloor\frac{\sqrt{N}}{K+1}\right\rfloor, 2\right\}. \tag{7}
$$

Therefore, for large N, we get $m = N/n = \sqrt{N}(K+1)$. In this case the number of jumps K needs to satisfy $K = o\left(m^{1/3}\right) = o\left(N^{1/6}K^{1/3}\right)$, i.e. $K = o\left(N^{1/4}\right)$, see Remark 2. A larger rate

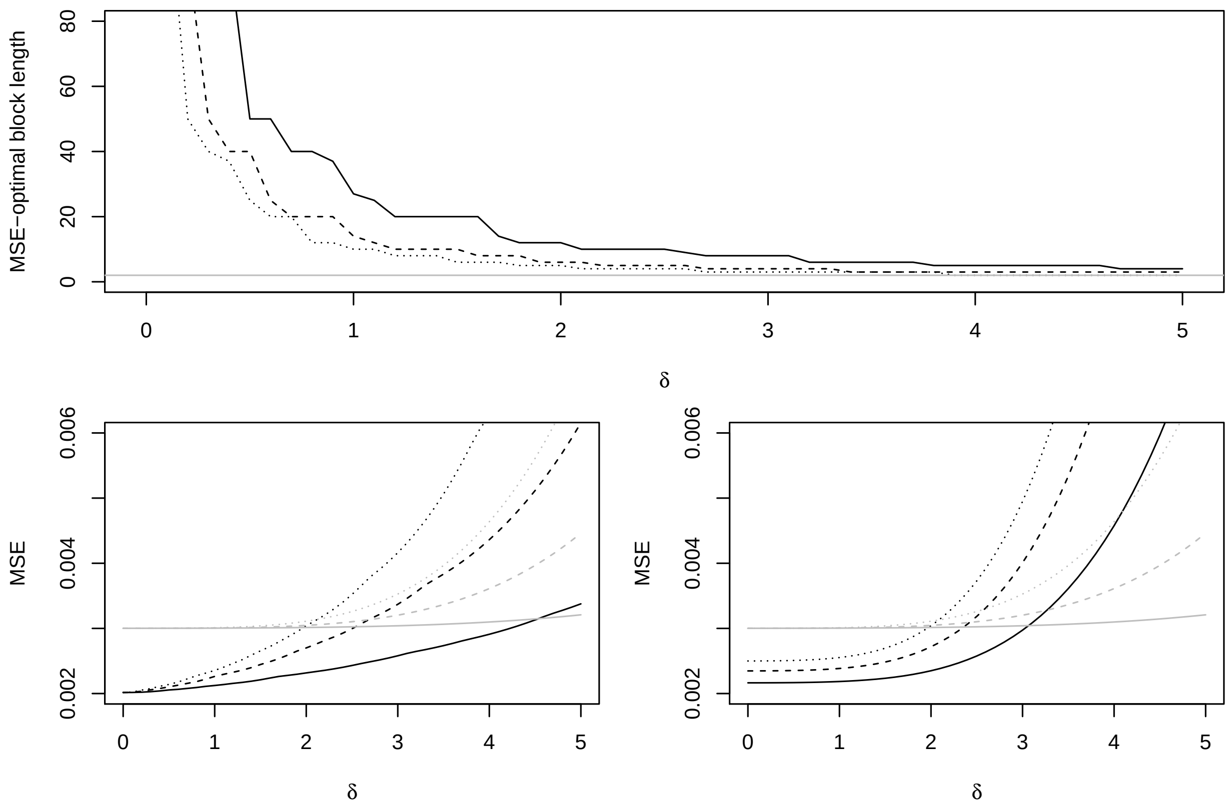

Figure 1: MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ (top), exact MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ together with $\widehat{\sigma^2}_{\text{Diff}}$ (bottom left) and exact MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ (bottom right) 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 $X_t \sim \mathrm{N}(0, 1)$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, ..., 5\}$.

$K = o\left(N^{1/3}\right)$ can be dealt with by choosing $m = N/c$ for some constant c, i.e. a fixed block length n. Otherwise, if testing is of interest in view of Theorem 3, we suggest choosing 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 (7). As K is not known exactly in real applications, there are several possibilities to set K in this formula:

1. Use prior knowledge about the possible value of K.
2. Determine a reasonable upper bound for K.
3. Pre-estimate K with any appropriate procedure (e.g. robust regression trees, see Galimberti et al. (2007)).

The third panel of Figure 1 shows the MSE of the estimator $\widehat{\sigma^2}_{\text{Mean}}$ with the block size n chosen according to (7). For $K \in \{3, 5\}$ there is only a moderate loss of performance when choosing n according to (7) instead of n_{opt} which depends on the number of jumps K and the height $h = \delta \cdot \sigma$. When dealing with $K = 1$ changes in the mean the performance of the blocks-estimator worsens slightly, but still yields better results than the difference-based method when the jump height is at most $h \approx 3\sigma$.

Table 1 shows the exact MSE of the ordinary sample variance for normally distributed data and different values of K and h, with $N = 1000$. We observe that the MSE gets very large when the number and height of the level shifts increases. Obviously, the blocks- and the difference-based estimators perform much better than the sample variance.

Table 1: Exact MSE of the sample variance for normally distributed data and different K and h, $N = 1000$.

		h					
		0	1	2	3	4	5
	1	0.0020	0.0362	0.5404	2.7202	8.5845	20.9459
K	3	-	1.1540	18.3869	93.0310	293.9704	717.6425
	5	-	6.7829	108.3786	548.5482	1733.5569	4232.1758

The results for data from the t_5-distribution are similar to those obtained for the normal distribution, see Figure 2. Again, the blocks-estimator with the block size which is a function of the square root of N performs well and does not lose too much performance compared to the second panel of Figure 2, where the optimal block size is considered.

Similar results are obtained for $N = 2500$, see Figures 6 and 7 in the Appendix.

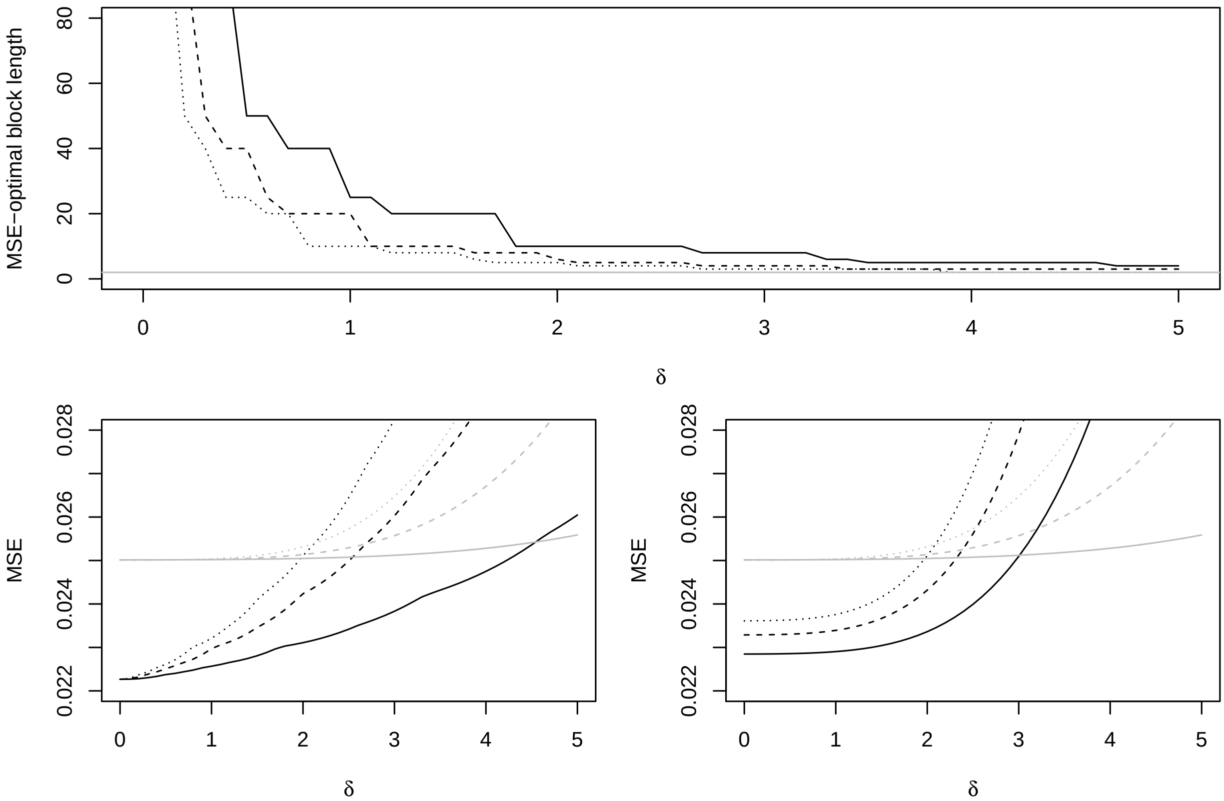

Figure 2: MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ (top), exact MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ together with $\widehat{\sigma^2}_{\text{Diff}}$ (bottom left) and exact MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ (bottom right) 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 $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, ..., 5\}$.

2.2 Estimation of the variance under frequently occurring level shifts

So far we have discussed the case where the number of changes in the mean K is asymptotically negligible with respect to the number of the blocks m and thus the number of observations N. However, there might be situations in which level shifts occur frequently, say, with frequency $p = 1/P$, e.g. $p = 1/1000$ corresponds to one jump every $P = 1000$ points on average. The choice of a growing block size according to the rule in (7) is no longer reasonable in this case.

2.2.1 Estimation with a fixed trimming fraction

Situations with frequent level shifts can be dealt with using an asymmetric trimmed mean of the block-wise estimates instead of their ordinary average. Thereby we assume that the practitioner knows this frequency, that it was pre-estimated in a prior study or a lower bound for P was determined. The corresponding trimmed blocks-estimator is given as

$$\widehat{\sigma^2}_{\mathrm{Tr},\alpha} = C_{N,\mathrm{Tr},\alpha} \frac{1}{m - \lfloor \alpha m \rfloor} \sum_{j=1}^{m-\lfloor \alpha m \rfloor} S^2_{(j)}, \tag{8}$$

where m is the number of blocks and $C_{N,\mathrm{Tr},\alpha}$ is a sample and distribution dependent correction factor to ensure unbiasedness in the absence of level shifts. In practice this constant can be simulated. E.g., for the 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 $C_{1000,\mathrm{Tr},0.2} = 1.1980$

As an example, we generate 1000 time series of length $N \in \{1000, 2500\}$. We add $K = N \cdot p$ jumps, with $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$ of height $h \in \{0, 2, 3, 5\}$ to the generated data at randomly chosen positions, as was done in Subsection 2.1. We choose $n = 20$ in order to ensure that the number of jump-contaminated blocks is sufficiently smaller than the total number of blocks.

Table 2 shows the simulated MSE of the trimmed estimator (8) for $\alpha \in \{0.1, 0.2, 0.5\}$, as well as the MSE of $\widehat{\sigma^2}_{\mathrm{Mean}}$, $\widehat{\sigma^2}_{\mathrm{Diff}}$ and $\widehat{\sigma^2}^{\beta}_{\mathrm{Tr,ad}}$ (see Subsection 2.2.2). We can observe that the trimmed estimator $\widehat{\sigma^2}_{\mathrm{Tr}}$ performs better than $\widehat{\sigma^2}_{\mathrm{Mean}}$ when dealing with many high jumps in the mean. When zero or few level shifts are present the averaging approach yields the lowest MSE value. 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 2.2.2.

2.2.2 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

$$\widehat{\sigma^2}_{\mathrm{Tr,ad}} = C_{N,\mathrm{Tr,ad}} \frac{1}{m - \lfloor \alpha_{\mathrm{adapt}} m \rfloor} \sum_{j=1}^{m-\lfloor \alpha_{\mathrm{adapt}} m \rfloor} S^2_{(j)} \tag{9}$$

and α_{adapt} the adaptively chosen percentage of the blocks-estimates which will be removed. We use the approach for outlier detection discussed in Davies and Gather (1993) 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 $(n-1)S_j^2/\sigma^2 \sim \chi^2_{n-1}$. Since the true variance

Table 2: Simulated MSE of $\widehat{\sigma^2}_{\text{Tr},\alpha}$, $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$ and $\widehat{\sigma^2}^{\beta}_{\text{Tr,ad}}$ for normally distributed data and different N, h and $K = p \cdot N$ with $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$.

K	h	$\widehat{\sigma^2}_{\text{Tr},0.1}$	$\widehat{\sigma^2}_{\text{Tr},0.2}$	$\widehat{\sigma^2}_{\text{Tr},0.5}$	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}^{0.05}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{0.1}_{\text{Tr,ad}}$
					$N = 1000$			
0	0	0.0021	0.0023	0.0031	0.0021	0.0030	0.0020	0.0021
	2	0.0025	0.0026	0.0034	0.0025	0.0030	0.0026	0.0027
2	3	0.0028	0.0028	0.0037	0.0037	0.0031	0.0026	0.0026
	5	0.0032	0.0031	0.0038	0.0122	0.0037	0.0025	0.0024
	2	0.0038	0.0036	0.0040	0.0029	0.0031	0.0042	0.0037
4	3	0.0047	0.0040	0.0043	0.0045	0.0033	0.0034	0.0034
	5	0.0050	0.0040	0.0042	0.0179	0.0056	0.0025	0.0027
	2	0.0053	0.0046	0.0047	0.0032	0.0032	0.0058	0.0053
6	3	0.0082	0.0055	0.0049	0.0051	0.0038	0.0047	0.0043
	5	0.0135	0.0062	0.0050	0.0202	0.0087	0.0029	0.0029
	2	0.0120	0.0088	0.0068	0.0046	0.0034	0.0125	0.0109
10	3	0.0288	0.0130	0.0077	0.0066	0.0050	0.0098	0.0075
	5	0.1330	0.0147	0.0075	0.0222	0.0187	0.0033	0.0032
					$N = 2500$			
0	0	0.0010	0.0010	0.0014	0.0008	0.0012	0.0009	0.0009
	2	0.0012	0.0012	0.0014	0.0011	0.0012	0.0013	0.0013
5	3	0.0014	0.0013	0.0015	0.0018	0.0013	0.0013	0.0011
	5	0.0015	0.0014	0.0016	0.0069	0.0018	0.0010	0.0009
	2	0.0023	0.0020	0.0020	0.0013	0.0013	0.0029	0.0025
10	3	0.0030	0.0022	0.0019	0.0021	0.0015	0.0021	0.0017
	5	0.0034	0.0022	0.0019	0.0086	0.0037	0.0011	0.0011
	2	0.0039	0.0029	0.0024	0.0015	0.0013	0.0049	0.0044
15	3	0.0064	0.0037	0.0025	0.0027	0.0019	0.0035	0.0030
	5	0.0107	0.0042	0.0028	0.0119	0.0068	0.0013	0.0012
	2	0.0100	0.0069	0.0042	0.0021	0.0016	0.0117	0.0102
25	3	0.0268	0.0105	0.0050	0.0039	0.0032	0.0092	0.0069
	5	0.1239	0.0130	0.0054	0.0186	0.0168	0.0019	0.0015

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.,

$$\widehat{\sigma^2}_{\text{Med}} = C_{N,\text{Med}} \cdot \text{med}\{S_1^2, ..., S_m^2\}, \tag{10}$$

where $C_{N,\text{Med}}$ is a correction factor in order to ensure unbiasedness in the absence of level shifts. Subsequently, we remove those values $(n-1)S_j^2/\widehat{\sigma^2}_{\text{Med}}$ which exceed $q_{\chi^2_{n-1},1-\beta_m}$, the $(1-\beta_m)$-quantile of the χ^2_{n-1}-distribution, with $\beta_m = 1-(1-\beta)^{1/m}$ and $\beta \in (0,1)$. We will refer to the adaptively trimmed estimator based on the approach of Davies and Gather (1993) as $\widehat{\sigma^2}^{\beta}_{\text{Tr,ad}}$.

Assuming that the expected percentage of level shifts in the data is at most 1%, i.e. the occurrence frequency is $p = 1/P = 1/100$, we suggest choosing the block size $n = 20$. In this way it is ensured that the number of uncontaminated blocks is much larger than the number of perturbed blocks. Moreover, we choose $\beta \in \{0.05, 0.1\}$.

Table 3: Average number of trimmed blocks in the absence of level shifts for normally distributed data and different N and β.

		N	
		1000	2500
β	0.05	0.0647	0.0583
	0.1	0.1268	0.1208

In the absence of level shifts slightly more than $\beta_m \cdot m$ blocks will be removed on average, as the following simulation suggests. We generated 10000 sequences of observations of size $N \in \{1000, 2500\}$ for $\beta \in \{0.05, 0.1\}$. Table 3 shows the average number of trimmed blocks in the absence of level shifts.

In order to assess the performance of the adaptive trimmed estimator $\widehat{\sigma^2}^{\beta}_{\text{Tr,ad}}$ we conduct a simulation study. We will consider $N \in \{1000, 2500\}$, $h \in \{0, 2, 3, 5\}$, $K \in \{0, 2, 5, 10\}$ and the block size $n = 20$. The correction factor in (9) has 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, $j = 1, ..., m$, and the initial estimate of the variance $\widehat{\sigma^2}_{\text{Med}}$. Subsequently, we remove the values $(n-1)S_j^2/\widehat{\sigma^2}_{\text{Med}}$ which exceed the quantile $q_{\chi^2_{19},1-\beta_m}$. 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 $C^{0.05}_{2500,\text{Tr,ad}} = 1.0009$, so both are are nearly 1 and could be neglected with little loss. For $\beta = 0.1$ similar values are obtained.

The simulated MSE for the adaptive trimmed variance estimator $\widehat{\sigma^2}^{\beta}_{\text{Tr,ad}}$ is given in Table 2. We see that the estimator $\widehat{\sigma^2}^{\beta}_{\text{Tr,ad}}$ performs better than $\widehat{\sigma^2}_{\text{Tr},\alpha}$, $\widehat{\sigma^2}_{\text{Mean}}$ and $\widehat{\sigma^2}_{\text{Diff}}$ when the number or the height of the jumps is large. Large level shifts are detected easily and removed by the adaptive approach. When there are only a few level shifts, which are not very large, the estimator $\widehat{\sigma^2}_{\text{Mean}}$ yields slightly better results.

Remark 5. *When other distributional assumptions are made, one can use the fact that the term* $\sqrt{n}\left(S_j^2 - \sigma^2\right)/\sqrt{E\left(X_1^4\right) - \sigma^4}$ *is approximately standard normal, when* n *is large enough, i.e. at least 50 observations per block are recommended. In this case the amount of level shifts should be reasonably smaller to ensure that the number of uncontaminated blocks is much larger than the number of perturbed blocks, i.e.* p *should be not larger than* 0.5% *of* N.

The fourth non-central moment $\mu_4 = E\left(X_1^4\right)$ *has to be estimated properly in the presence of level shifts then. We can estimate this quantity in blocks and then compute the median of the blocks-estimates* $\widehat{\mu}_{4,Med}$*, as was done in (10). Then, blocks-values* $\sqrt{n}\left(S_j^2 - \widehat{\sigma^2}_{Med}\right)/\sqrt{\widehat{\mu}_{4,Med} - \widehat{\sigma^2}^{2}_{Med}}$ *which exceed the* $(1-\beta_m)$*-quantile of the standard normal distribution are removed.*

3 Block-wise 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.

We will consider the two blocks-estimators

$$\widehat{\sigma}_{\text{Mean},1} = C_{N,1}\frac{1}{m}\sum_{j=1}^{m} S_j \quad \text{and} \tag{11}$$

$$\widehat{\sigma}_{\text{Mean},2} = C_{N,2}\sqrt{\frac{1}{m}\sum_{j=1}^{m} S_j^2} = C_{N,2}\sqrt{\widehat{\sigma^2}_{\text{Mean}}}, \tag{12}$$

where $C_{N,1}$ and $C_{N,2}$ are sample dependent correction factors to ensure unbiasedness when no changes in the mean are present. For the first estimator the expected value and the variance are needed for S_j rather than for S_j^2 in every block $j = 1, ..., m$.

For this purpose we will first consider the exact distribution of the empirical variance when dealing with jumps in the mean. Given independent identically normally distributed data $X_{j,1}, ..., X_{j,n}$ 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 6. *Assume that $X_1, ..., X_n \sim \mathcal{N}(0, \sigma^2)$ and $Y_t = X_t + \sum_{k=1}^{\widetilde{K}_1} h_k I_{t \geq t_k}$ for $t = 1, ..., n$. 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 $\lambda_1 = \frac{1}{\sigma^2}\sum_{t=1}^{n}\left(\Delta_{1,t} - \overline{\Delta}_1\right)^2$, $\Delta_{1,t} = \sum_{k=1}^{\widetilde{K}_1} h_k I_{t\geq t_k}$ and $\overline{\Delta}_1 = \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{\widetilde{K}_1} h_k I_{t \geq t_k}$.

Proof. $\overline{Y}_1 = \overline{X}_1 + \overline{\Delta}_1$ and $Y_t - \overline{Y}_1 = X_t - \overline{X}_1 - \overline{\Delta}_1 + \Delta_{1,t}$, $t = 1, ..., n$, 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,

$$\begin{aligned}
\underbrace{\sum_{t=1}^{n}\left(\frac{Y_t - \overline{\Delta}_1}{\sigma}\right)^2}_{\sim\chi^2_{n,\lambda_1}\ (Y_t\ \forall t \text{ independent})} &= \sum_{t=1}^{n}\left(\frac{Y_t - \overline{Y}_1 + \overline{Y}_1 - \overline{\Delta}_1}{\sigma}\right)^2 \\
&= \sum_{t=1}^{n}\left(\frac{Y_t - \overline{Y}_1}{\sigma}\right)^2 + \sum_{t=1}^{n}\left(\frac{\overline{Y}_1 - \overline{\Delta}_1}{\sigma}\right)^2 \\
&\quad + 2\left(\frac{\overline{Y}_1 - \overline{\Delta}_1}{\sigma}\right)\sum_{t=1}^{n}\left(\frac{Y_t - \overline{Y}_1}{\sigma}\right) \\
&= \frac{n-1}{\sigma^2} S_1^2 + \frac{n}{\sigma^2}\left(\overline{Y}_1 - \overline{\Delta}_1\right)^2 = \frac{n-1}{\sigma^2} S_1^2 + \underbrace{\frac{n}{\sigma^2}\overline{X}_1^2}_{\sim\chi^2_1}
\end{aligned}$$

The moment generating function at $z \in \mathbb{R}$ of both sides and the independence of S_1^2 and $\overline{Y}_1$ yield:

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

$$= M_{\frac{n-1}{\sigma^2}S_1^2}(z) \cdot (1 - 2 \cdot z)^{-1/2}$$

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

$$\Rightarrow \frac{n-1}{\sigma^2} S_1^2 \sim \chi^2_{n-1,\lambda_1}.$$

$\square$

In the following we assume that $B \leq K$ blocks are contaminated by $\widetilde{K}_1, ..., \widetilde{K}_B$ jumps, respectively, with $\sum_{k=1}^{B} \widetilde{K}_k = K$. W.l.o.g. assume that the jumps are contained in the last B blocks, while the first $m - B > 0$ 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. Lax (1985) and Miller (1964). 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 first blocks $j = 1, \ldots, m - B$, i.e., $\sqrt{n-1}S_j/\sigma \sim \chi_{n-1}$. The moments of S_j are given by

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

$$Var(S_j) = \sigma^2 + \frac{\sigma^2}{n-1}\left(\lambda_j - \left(E(S_j)\frac{\sqrt{n-1}}{\sigma}\right)^2\right),$$

where $F_{1,1}(a, b, z)$ represents the generalized hypergeometric function, see Olver et al. (2010) for more details. The exact finite sample correction factor to ensure unbiasedness when no change in the mean is present, i.e., $K = 0$, is

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

Therefore, the MSE of the estimator $\widehat{\sigma}_{\text{Mean},1}$ (see (11)) is given as

$$\text{MSE}(\widehat{\sigma}_{\text{Mean},1}, \sigma) = \left(\frac{C_{N,1}}{m}\left((m-B)E(S_1) + \sum_{j=m-B+1}^{m} E\left(S_j\right)\right) - \sigma\right)^2 + C_{N,1}^2\left(\frac{m-B}{m^2} Var\left(S_1\right) + \frac{1}{m^2}\sum_{j=m-B+1}^{B} Var\left(S_j\right)\right),$$

assuming that S_1 arises from a block without jumps, while S_j, $j = m-B+1, ..., m$, are the estimates in contaminated blocks.

For the second estimator (12) we have the following statements on its expectation and its variance as well as a suitable finite sample correction factor in the absence of jumps.

$$\widehat{\sigma}_{\text{Mean},2} = C_{N,2} \frac{\sigma}{\sqrt{m(n-1)}} \underbrace{\sqrt{\underbrace{\frac{n-1}{\sigma^2}\sum_{j=1}^{m-B} S_j^2}_{\sim\chi^2_{(m-B)(n-1)}} + \underbrace{\frac{n-1}{\sigma^2}\sum_{j=m-B+1}^{m} S_j^2}_{\sim\chi^2_{B(n-1),\sum_{j=m-B+1}^{m}\lambda_j}}}}_{\sim\chi_{m(n-1),\sqrt{\sum_{j=m-B+1}^{m}\lambda_j}}}$$

$$E\left(\widehat{\sigma}_{\text{Mean},2}\right) = C_{N,2}\frac{\sigma\sqrt{2}}{\sqrt{m(n-1)}}\frac{\Gamma(0.5(m(n-1)+1))}{\Gamma(0.5m(n-1))}F_{1,1}\left(-0.5, 0.5m(n-1), \sum_{j=m-B+1}^{m}\frac{-\lambda_j}{2}\right),$$

$$Var\left(\widehat{\sigma}_{\text{Mean},2}\right) = C_{N,2}^2\left(\sigma^2 + \frac{\sigma^2}{m(n-1)}\left(\sum_{j=m-B+1}^{m}\lambda_j - \left(\frac{E(\widehat{\sigma}_{\text{Mean},2})}{C_{N,2}}\frac{\sqrt{m(n-1)}}{\sigma}\right)^2\right)\right),$$

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

The following consistency statements are valid for the two introduced estimators $\widehat{\sigma}_{\text{Mean},1} = C_{N,1}\frac{1}{m}\sum_{j=1}^{m}S_j$ (as defined in (11)) and $\widehat{\sigma}_{\text{Mean},2} = C_{N,2}\sqrt{\frac{1}{m}\sum_{j=1}^{m}S_j^2}$ (as defined in (12)) of σ:

Corollary 7. *Under the conditions of Theorem 1 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 Hu et al. (1989) that

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

where $C_{N,1} \to 1$, since the sample standard deviation is a consistent estimator for σ in every block when no changes in the mean are present, see Remark 8. 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

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

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

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

and for the second

$$\frac{1}{m}\sum_{j=m-B+1}^{m} E\left(S_{j,h}C_{N,1}\right) = \frac{C_{N,1}}{m}\sum_{j=m-B+1}^{m} E\left(\sqrt{S_j^2}\right)$$

$$\leq \frac{C_{N,1}}{m}\sum_{j=m-B+1}^{m}\sqrt{E\left(S_{j,0}^2 + \sum_{t=1}^{n}\frac{2(X_{j,t}-\bar{X}_j)(\Delta_{j,t}-\overline{\Delta}_j) + \left(\Delta_{j,t}-\overline{\Delta}_j\right)^2}{n-1}\right)}$$

$$= \frac{C_{N,1}}{m}\sum_{j=m-B+1}^{m}\sqrt{E\left(S_{j,0}^2 + \sum_{t=1}^{n}\frac{\left(\Delta_{j,t}-\overline{\Delta}_j\right)^2}{n-1}\right)}$$

$$= \frac{C_{N,1}}{m}\sum_{j=m-B+1}^{m}\underbrace{\sqrt{\sigma^2 + \sum_{t=1}^{n}\frac{\left(\Delta_{j,t}-\overline{\Delta}_j\right)^2}{n-1}}}_{\in\left[\sigma,\sqrt{\sigma^2+\frac{n}{n-1}\left(\sum_{k=1}^{K}h_k\right)^2}\right]}$$

$$\leq C_{N,1}\frac{B}{m}\sqrt{\sigma^2 + \frac{n}{n-1}\left(\sum_{k=1}^{K}h_k\right)^2} \longrightarrow 0,$$

where $\Delta_{j,t}$ and $\overline{\Delta}_j$ are defined in the proof of Theorem 1. □

Remark 8. *The correction factors $C_{N,1}$ and $C_{N,2}$ satisfy*

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

This can be shown with Lemma 1.4A in Serfling (2009). Therefore, for large N and n we can neglect the correction factors when using the estimators $\widehat{\sigma}_{Mean,1}$, $\widehat{\sigma}_{Mean,2}$ with block sizes n which are a function of N with $n \to \infty$.

We will now investigate the performance of both estimators under normality, where the exact MSE can be considered. Again, we will generate jumps at random positions as was done in Section 2.

In the first panel of Figure 3 we see that in many cases the estimator $\widehat{\sigma}_{\text{Mean},1}$ requires larger block sizes than $\widehat{\sigma}_{\text{Mean},2}$. The second panel of Figure 3 shows the MSE arising from the optimal block size. For this choice, the estimator $\widehat{\sigma}_{\text{Mean},1}$ apparently works better since its MSE is smaller than that of the square root of the blocks-variance estimator. This is plausible since jumps cause large positive biases in the corresponding blocks which affect the mean even more after taking squares. Otherwise, when choosing the block size according to the rule (7), this conclusion remains valid when the jump height is rather large, while the estimator $\widehat{\sigma}_{\text{Mean},2}$ is preferable in case of small jumps, see the third panel of Figure 3. All in all, the results worsen only slightly when choosing the block size according to the rule (7) instead of the optimal one.

Remark 9. *When dealing with frequently occurring level shifts, as is discussed in Section 2.2, the square root of the variance estimator $\widehat{\sigma^2}^{\beta}_{Tr,ad}$ from (9) can be used in order to estimate the standard deviation σ. For large N a correction factor to ensure unbiasedness under no level shifts can be neglected.*

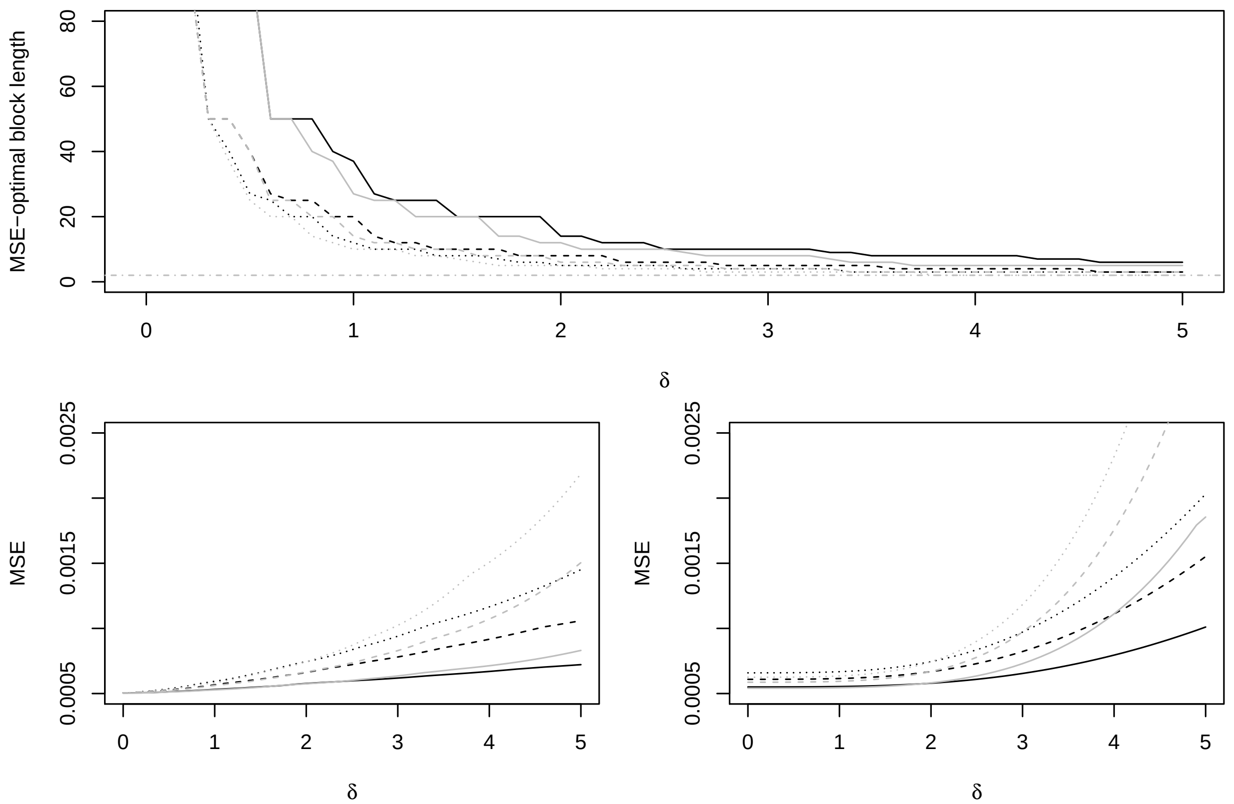

Figure 3: MSE-optimal block length n_{opt} of $\widehat{\sigma}_{\text{Mean},1}$ and $\widehat{\sigma}_{\text{Mean},2}$ (top), exact MSE regarding n_{opt} of $\widehat{\sigma}_{\text{Mean},1}$ together with $\widehat{\sigma}_{\text{Mean},2}$ (bottom left) and exact MSE of $\widehat{\sigma}_{\text{Mean},1}$ and $\widehat{\sigma}_{\text{Mean},2}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ (bottom right) for $K = 1$ (—), $K = 3$ (- - -) and $K = 5$ ($\cdot\cdot\cdot$) with $N = 1000$, $Y_t = X_t + \sum_{k=1}^{K} h I_{t \geq t_k}$, where $X_t \sim \mathrm{N}(0, 1)$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, ..., 5\}$.

4 Application

In this section we apply the blocks-approach to two datasets in order to estimate the variance (or the standard deviation).

4.1 Nile river flow data

The first dataset contains the widely discussed Nile river flow records in Aswan from 1871 to 1984, see e.g. Hassan (1981), Hipel and McLeod (1994), Syvitski and Saito (2007), 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 Hassan (1981). The construction of the two Aswan dams in 1902 and from 1960 to 1969 obviously caused changes in the river flow, see Hassan (1981) and Hipel and McLeod (1994). In order to verify that the annual maximal values are independent we plot the ACF and the PACF of the values between the two possible change-points, i.e. for the period 1903 – 1960, see the lower panel of Figure 4, which indicates that the assumption of independence is justified. Moreover, we used Levene's test 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 $p = 0.40$.

The ordinary sample variance of the entire data yields the value 3243866, the corresponding sample standard deviation is 1801.07, see Table 4. For the blocks-estimator of the variance from (2) we choose the block size according to (7) with $K = 2$ getting $n = \lfloor\sqrt{114}/3\rfloor = 3$. The corresponding estimate of the variance is $\widehat{\sigma^2}_{\mathrm{Mean}} = 2123327.60$. In order to avoid determining a correction factor for the estimator of the standard deviation we use the estimator (12) since the number of observations is not large enough to use the estimator (11) without the correction factor. We get $\widehat{\sigma}_{\mathrm{Mean},2} = 1457.16$. This can be compared to the ordinary sample variance of the values between the years 1903 and 1960 which is 2129229.70, and the sample standard deviation of 1459.19. We conclude that the blocks-procedures from (2) and (12) perform better than their ordinary counterparts since the estimated values on the whole dataset are similar to those for the period 1903 – 1960 in between the changes.

Table 4: Estimates of the variance and the standard deviation for the annual maxima of the average monthly discharge of the Nile river in Aswan.

Period	K	S^2	S	$\widehat{\sigma^2}_{\mathrm{Mean}}$	$\widehat{\sigma}_{\mathrm{Mean},2}$
1871 – 1984	2	3243866	1801.07	2123327.60	1457.16
1903 – 1960	0	2129229.70	1459.19	2003210.50	1415.35

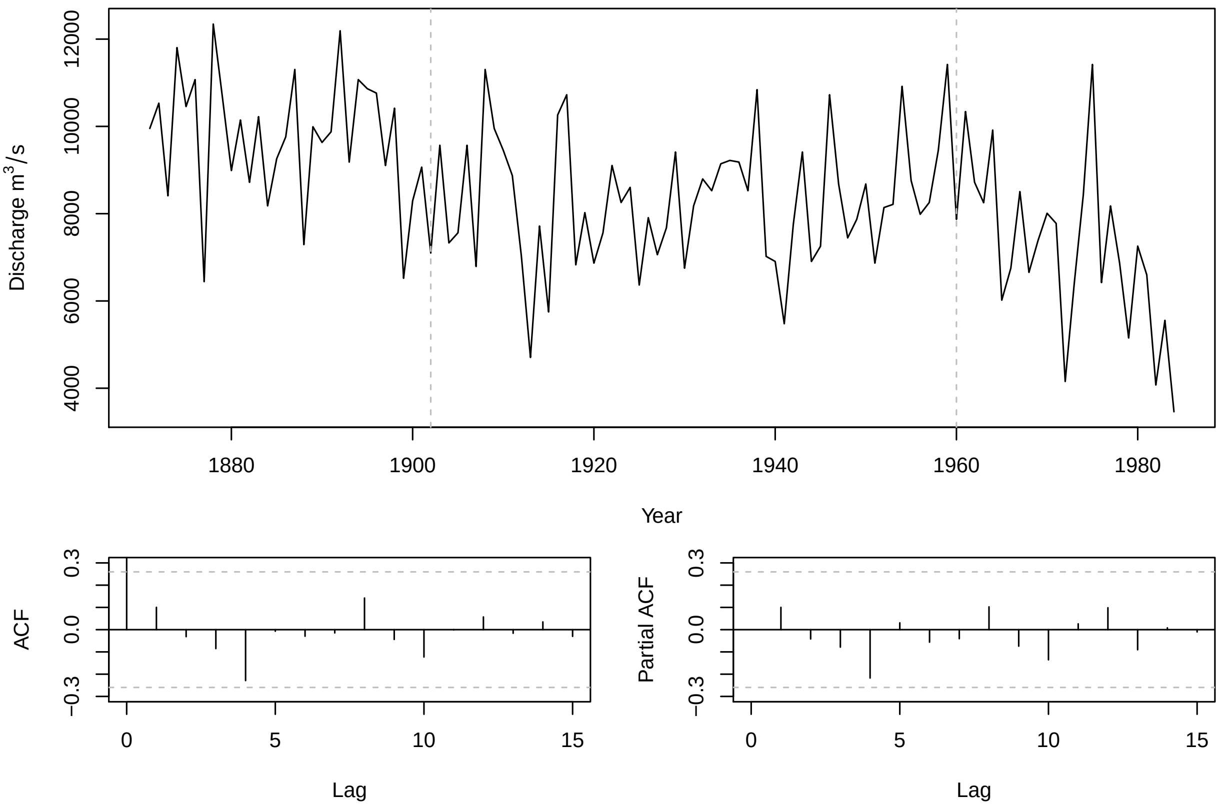

Figure 4: Maximal monthly discharge of the Nile river at Aswan in the period 1871 – 1984 (upper panel); the corresponding ACF (left lower panel) and PACF (right lower panel) for the years 1903 – 1960.

Furthermore, we apply the adaptive trimmed estimator $\widehat{\sigma^2}^{\beta}_{\mathrm{Tr,ad}}$ to the data for comparison, although frequent level shifts are not assumed to appear in the river flow data. A Q-Q plot of the data indicates that the deviation from a normal distribution is not very large, see Figure 8 in the Appendix. With $\beta = 0.05$ and $n = 10$ ($m = 11$ blocks) no blocks are trimmed away during the procedure. We get an estimate $\widehat{\sigma^2}^{0.05}_{\mathrm{Tr,ad}} = 2260282$ which is much smaller than the ordinary sample variance of the data and not far from $\widehat{\sigma^2}_{\mathrm{Mean}}$ with $n = 3$.

4.2 PAMONO data

In the next example we use data from the PAMONO (Plasmon Assisted Microscopy of Nano-Size Objects) biosensor, provided by the Leibniz- Institut für Analytische Wissenschaften – ISAS – e.V. in Dortmund, see Zybin et al. (2010). The PAMONO biosensor is used for detection of particles, primarily viruses, see e.g. Siedhoff et al. (2011), Timm et al. (2011) and Zybin et al. (2010). A time series of grayscale images is obtained during the process of recording. An obvious level shift in the data is caused by the adhesion of a virus on the sensor surface. A change of the variance after a jump in the mean is not expected to occur. More information on this can be found in Timm et al. (2011) and Abbas et al. (2016).

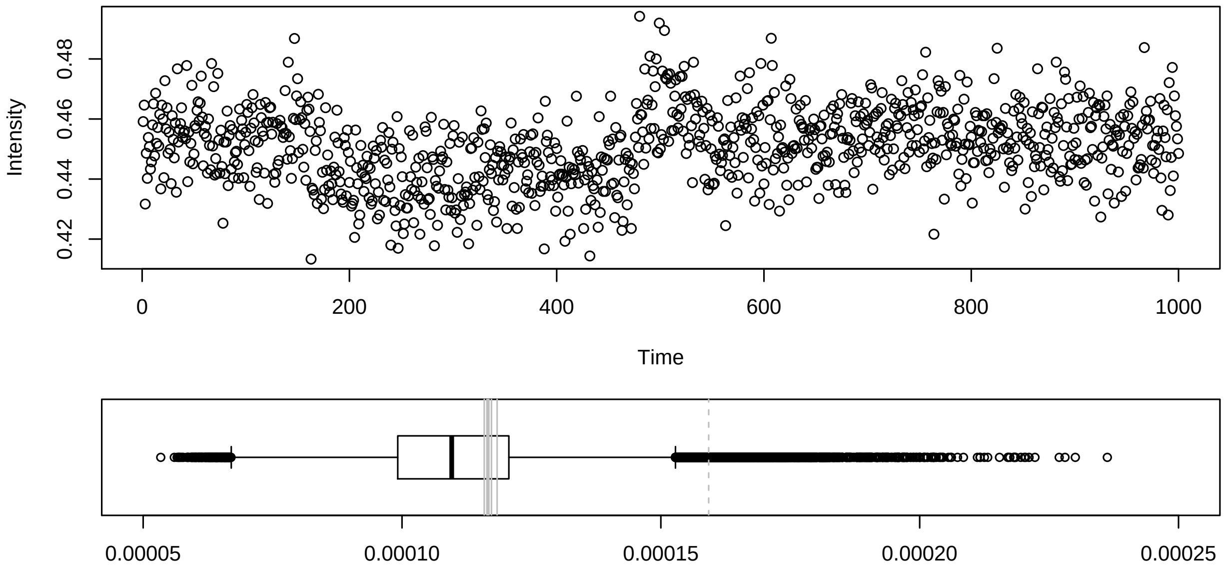

Figure 5: Intensity over time for one pixel (upper panel) and a boxplot of variances for the virus-free pixels (lower panel) together with the ordinary sample variance of the above data (- - -) and blocks-variance estimates for $K \in 0, 1, ..., 5$ (—).

In the upper panel of Figure 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. The lower panel 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 in order to have an 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 blocks-estimator $\widehat{\sigma^2}_{\mathrm{Mean}}$ with n chosen according to the rule (7), with $K \in 0, 1, ..., 5$, yields values within the interval $[1.15 \cdot 10^{-4}, 1.18 \cdot 10^{-4}]$ (—) which are well within the interquartile range. We conclude that the blocks-approach yields reasonable estimates for these data.

Again, we apply the adaptive trimmed estimator $\widehat{\sigma^2}^{\beta}_{\mathrm{Tr,ad}}$ for comparison, although level shifts occur only rarely in this application. A Q-Q plot of the data indicates that the assumption of a normal distribution is reasonable, see Figure 9 in the Appendix. With $n = 20$ and $\beta = 0.05$ one block is trimmed away during the procedure and we get $\widehat{\sigma^2}^{0.05}_{\mathrm{Tr,ad}} = 1.11 \cdot 10^{-4}$. This value is within the interquartile range and also a reasonable estimate of the variance.

5 Conclusion

In the presence of level shifts ordinary variance estimators perform poorly. In this paper we considered several estimation procedures in order to account for possible changes in the mean. If only few level shifts are expected in a long sequence of observations our recommendation is to use the blocks-variance estimator $\widehat{\sigma^2}_{\mathrm{Mean}}$. This estimation procedure does not require knowledge of the underlying distribution and performs well in the aforementioned situation.

If level shifts are expected to occur frequently and the practitioner is willing to make a distributional assumption we recommend using the adaptive trimmed procedure $\widehat{\sigma^2}^{\beta}_{\mathrm{Tr,ad}}$. Under normality the χ^2-distribution of the sample variance is used in order to choose the trimming parameter properly. An appropriate estimate of the fourth moment of the data is required when other distributions are assumed.

If no distributional assumptions can be made in the case of frequently occurring level shifts the difference-based estimator, as defined in (6) can be used.

Acknowledgments

This work has been supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG), which is gratefully acknowledged. The authors would like to thank Sermad Abbas for providing the R code to extract the PAMONO time series.

References

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

Angelova JA (2012) On moments of sample mean and variance. Int J Pure Appl Math 79(1):67–85

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

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

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

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

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

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

Hassan FA (1981) Historical nile floods and their implications for climatic change. Science 212(4499):1142–1145

Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems, vol 45. Elsevier

Hu TC, Moricz F, Taylor R (1989) Strong laws of large numbers for arrays of rowwise independent random variables. Acta Mathematica Hungarica 54(1-2):153–162

Lax DA (1985) Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. Journal of the American Statistical Association 80(391):736–741

Miller KS (1964) Multidimensional gaussian distributions. John Wiley & Sons Inc

Munk A, Bissantz N, Wagner T, Freitag G (2005) 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):19–41

Olver FW, Lozier DW, Boisvert RF, Clark CW (2010) NIST handbook of mathematical functions hardback and CD-ROM. Cambridge University Press

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`

Rice J, et al. (1984) Bandwidth choice for nonparametric regression. The Annals of Statistics 12(4):1215–1230

Rooch A, Zelo I, Fried R (2016) Estimation methods for the lrd parameter under a change in the mean. Statistical Papers pp 1–35

Seber GA, Lee AJ (2012) Linear regression analysis, vol 329. John Wiley & Sons

Serfling RJ (2009) Approximation theorems of mathematical statistics, vol 162. John Wiley & Sons

Siedhoff D, Weichert F, Libuschewski P, Timm C (2011) Detection and classification of nano-objects in biosensor data. In: Proceedings of the International Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB), p 9

Syvitski JP, Saito Y (2007) Morphodynamics of deltas under the influence of humans. Global and Planetary Change 57(3-4):261–282

Tecuapetla-Gómez I, Munk A (2017) Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. Scandinavian Journal of Statistics 44(2):346–368

Timm C, Libuschewski P, Siedhoff D, Weichert F, Müller H, Marwedel P (2011) Improving nanoobject detection in optical biosensor data. In: Proceedings of the 5th International Symposium on Bio-and Medical Information and Cybernetics (BMIC 2011), vol 2, pp 236–240

Tong T, Ma Y, Wang Y, et al. (2013) Optimal variance estimation without estimating the mean function. Bernoulli 19(5A):1839–1854

Von Neumann J, Kent R, Bellinson H, Hart Bt (1941) The mean square successive difference. The Annals of Mathematical Statistics 12(2):153–162

Zybin A, Kuritsyn YA, Gurevich EL, Temchura VV, Überla K, Niemax K (2010) Real-time detection of single immobilized nanoparticles by surface plasmon resonance imaging. Plasmonics 5(1):31–35

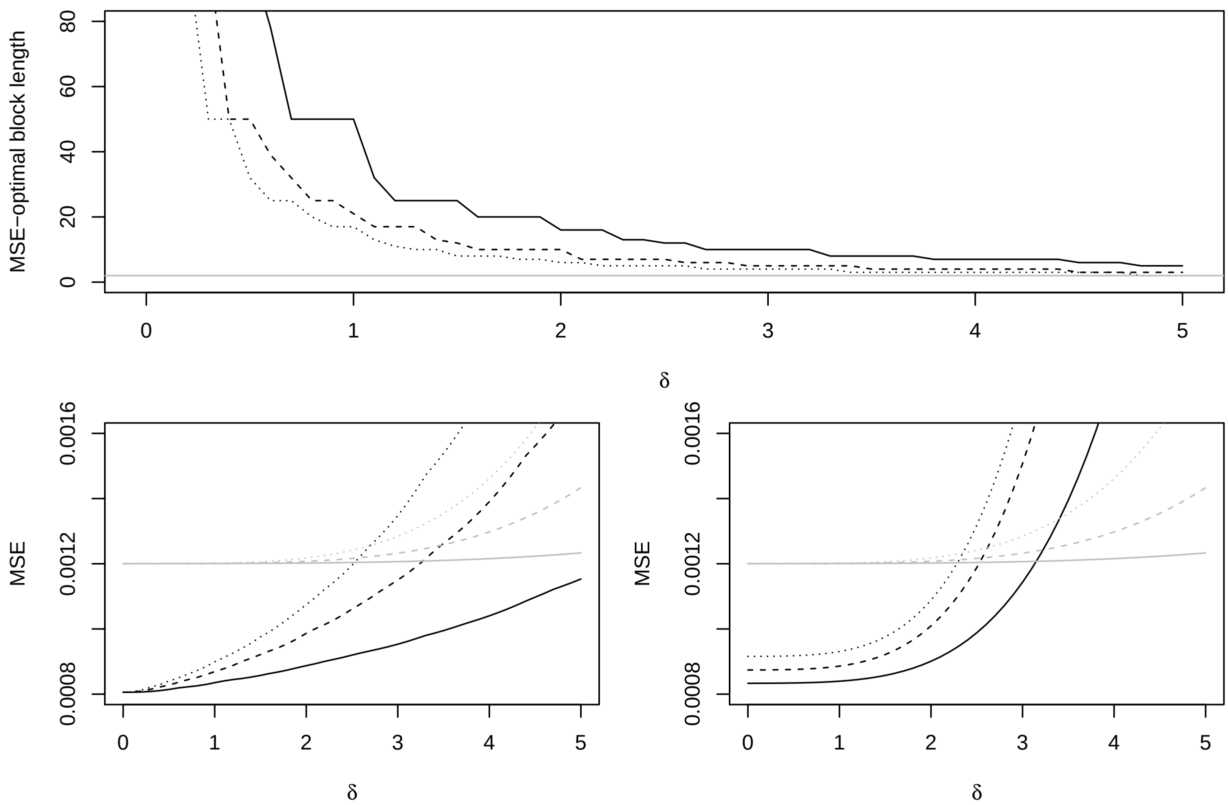

Figure 6: MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ (top), exact MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ together with $\widehat{\sigma^2}_{\text{Diff}}$ (bottom left) and exact MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ (bottom right) 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 $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, ..., 5\}$.

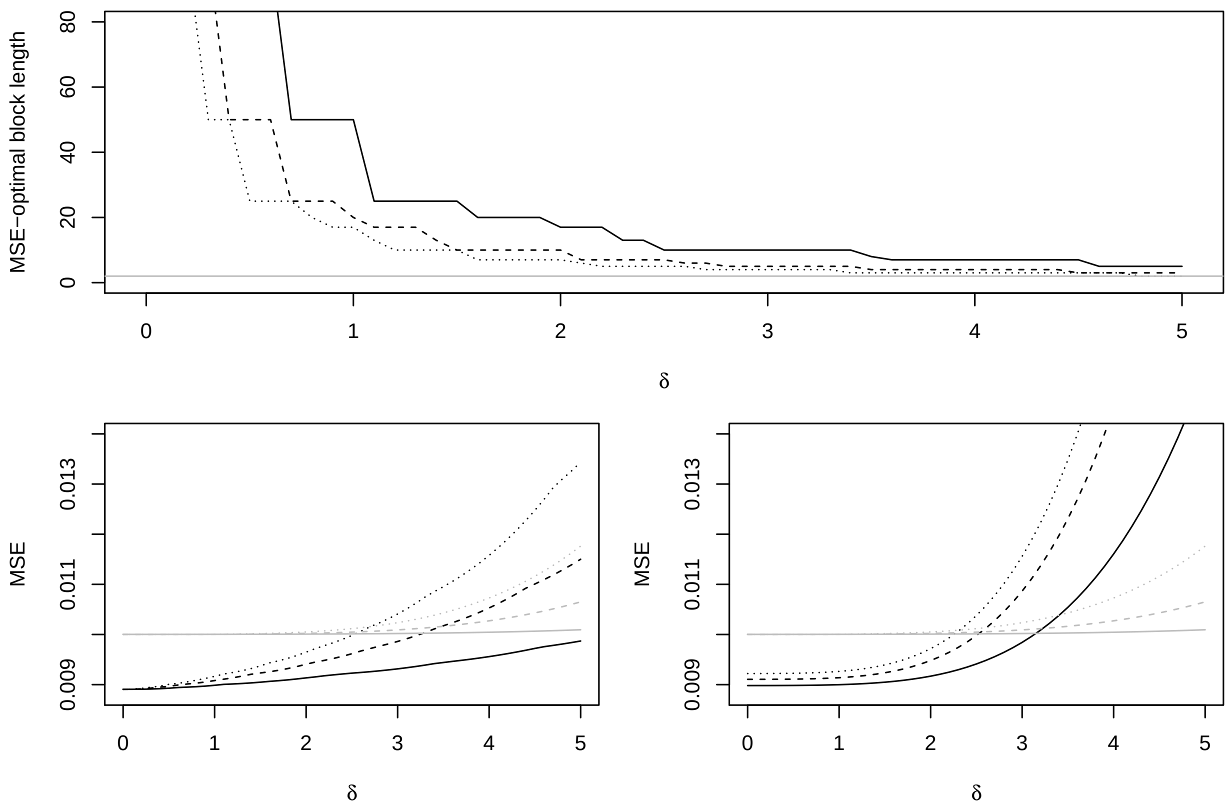

Figure 7: MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ (top), exact MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ together with $\widehat{\sigma^2}_{\text{Diff}}$ (bottom left) and exact MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ (bottom right) 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 $X_t \sim t_5$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, ..., 5\}$.

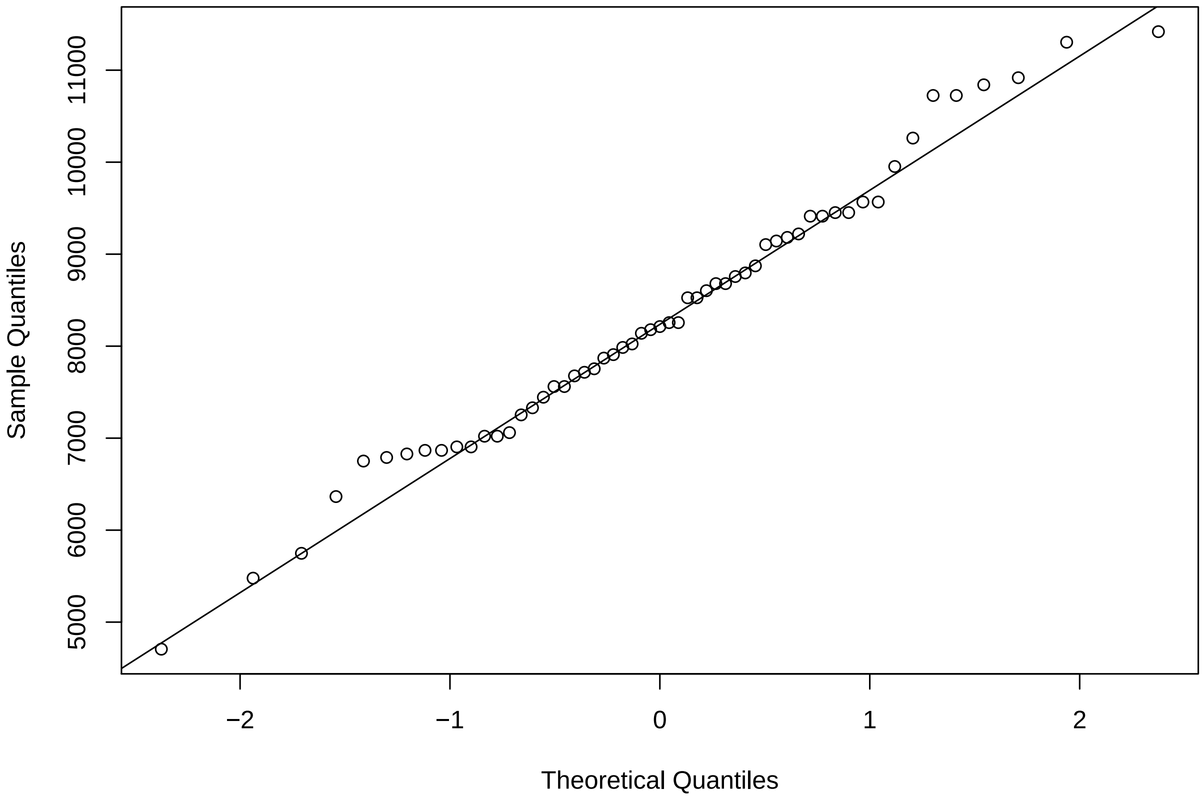

Figure 8: Q-Q plot of the maximal monthly discharge of the Nile river in the period 1871 – 1984.

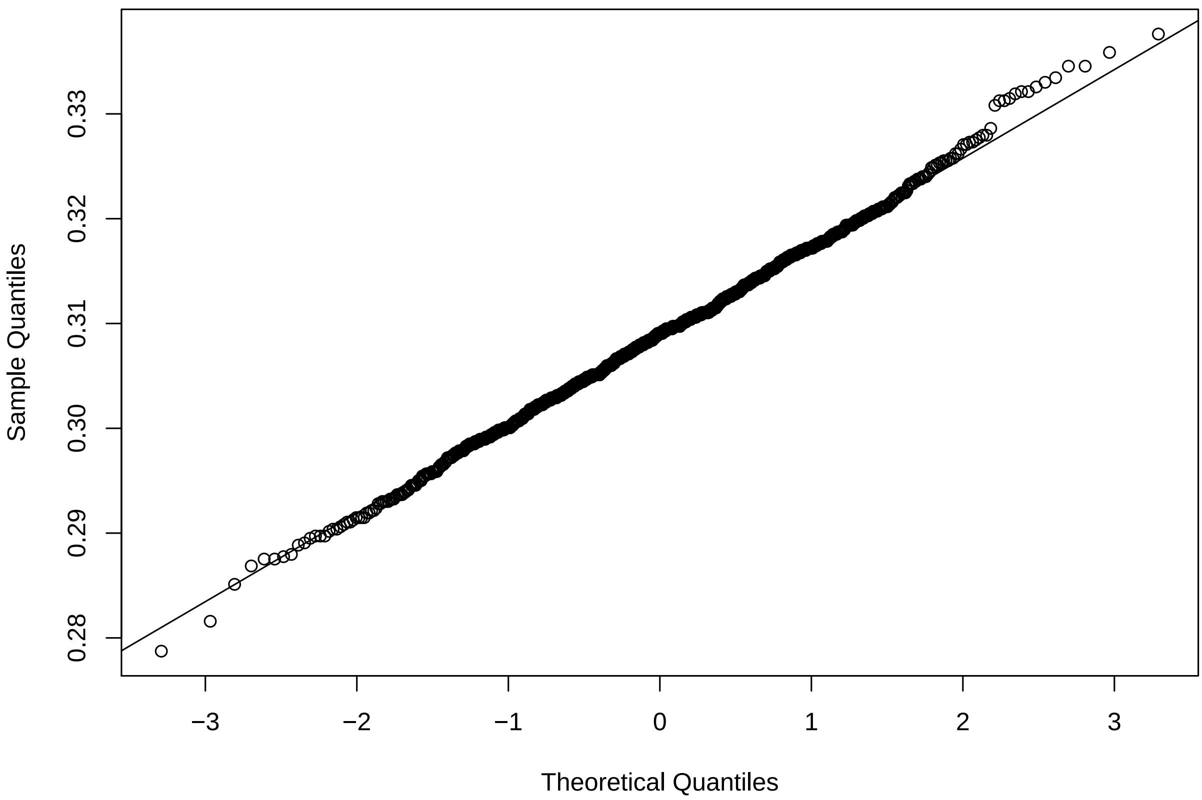

Figure 9: Q-Q plot of a pixel with a virus adhesion from the PAMONO data.