On variance estimation under shifts in the mean

Ieva Axt[1] · Roland Fried[1]

Abstract

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 change-point scenarios. We show the strong consistency and the asymptotic normality of such blocks-estimators of the variance under independence. Weak consistency is shown for short-range dependent strictly stationary data. We provide recommendations on the appropriate choice of the block size and compare this blocks-approach with difference-based estimators. If level shifts occur frequently and are rather large, the best results can be obtained by adaptive trimming of the blocks.

Keywords Blockwise estimation · Change-point · Trimmed mean

1 Introduction

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

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

Most of the time we assume that $X_1, \ldots, X_N$ are i.i.d. random variables with $E(X_t) = \mu$ and $\mathrm{Var}(X_t) = \sigma^2$, but this will be relaxed occasionally to allow for a short-range dependent strictly stationary sequence. The observed data $y_1, \ldots, y_N$ are affected by an unknown number K of level shifts of possibly different heights $h_1, \ldots, h_K$ at different time points $t_1, \ldots, t_K$. Our goal is the estimation of the variance σ^2. Without loss of generality, we will set $\mu = 0$ in the following.

✉ Ieva Axt
axt@statistik.tu-dortmund.de

Roland Fried
fried@statistik.tu-dortmund.de

[1] TU Dortmund University, Dortmund, Germany

In Sect. 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 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 Sect. 3. In Sect. 4 a simulation study is conducted to assess the performance of the proposed approaches. In Sect. 5 the estimation procedures are applied to real data sets, while Sect. 6 summarizes the results of this paper.

2 Estimation of the variance by averaging

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 $Y_1, \ldots, Y_N$ 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. (2019) for estimation of the Hurst parameter.

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 $Y_{j,1}, \ldots, Y_{j,n}$ are the observations in the jth block. We are interested in finding the block size n which yields a low mean squared error (MSE) 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 are more spread in the former case resulting in a larger positive bias of most scale estimators.

2.1 Asymptotic properties

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:

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

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

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

The blocks-estimator (2) estimates the variance consistently if the number of blocks grows sufficiently fast as is shown in Theorem 1.

Theorem 1 *Let $Y_1, \ldots, Y_N$ with $Y_t = X_t + \sum_{k=1}^{K} h_k I_{t \geq t_k}$ from Model (1) be segregated into m blocks of size n, where $t_1, \ldots, t_K$ are the time points of the jumps of size $h_1, \ldots, h_K$, 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 $\widehat{\sigma^2}_{\text{Mean}} = \frac{1}{m} \sum_{j=1}^{m} S_j^2 \to \sigma^2$ 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 jth block, $\mu_{j,t} = E(Y_{j,t})$ and $\overline{\mu}_j = \frac{1}{n} \sum_{t=1}^{n} E(Y_{j,t})$. Then we have

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

For the second term in the last Eq. (3), we have almost surely

$$\left| \frac{2}{m(n-1)} \sum_{j=1}^{B} \sum_{t=1}^{n} (X_{j,t} - \overline{X}_j)(\mu_{j,t} - \overline{\mu}_j) \right|$$

$$\leq \frac{2}{m(n-1)} \sum_{j=1}^{B} \sum_{t=1}^{n} \left| (X_{j,t} - \overline{X}_j) \right| \left| (\mu_{j,t} - \overline{\mu}_j) \right|$$

$$\leq \frac{2}{m(n-1)} \sum_{j=1}^{B} \sum_{t=1}^{n} \left| (X_{j,t} - \overline{X}_j) \right| \left| \sum_{k=1}^{K} h_k \right| \qquad (4)$$

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

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 (4) is a random variable with finite moments if n and B are fixed. This random variable converges to the term $E\left(\frac{1}{n} \sum_{t=1}^{n} \left| (X_{j,t} - \overline{X}_j) \right| \right)$ almost surely if $B \to \infty$. In the case of $B \to \infty$ and $n \to \infty$ this term converges to $E(|X_1|)$ almost surely due to Theorem 2 of Hu et al. (1989) 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 $\mathrm{Var}(S_j^2) \to 0$. 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):

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

$$= \frac{B}{m} \frac{n}{n-1} \left(\sum_{k=1}^{K} h_k \right)^2 \longrightarrow 0.$$

The first term of the last equation in (3) converges almost surely to σ^2 due to the results on triangular arrays in Theorem 2 of Hu et al. (1989), 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 $\mathrm{Var}(S_j^2) \to 0$. Application of Slutsky's Theorem proves the result.   □

Remark 2

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 $K\left(\sum_{k=1}^{K} h_k \right)^2 = K^3 h^2 = o(m)$. 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

$$\frac{\text{Var}(S^2)}{\text{Var}\left(\widehat{\sigma^2}_{\text{Mean}}\right)} = \frac{\frac{v_4}{N} - \frac{\sigma^4(N-3)}{N(N-1)}}{\frac{1}{m}\left(\frac{v_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}\right)} = \frac{v_4 - \frac{\sigma^4(N-3)}{(N-1)}}{v_4 - \frac{\sigma^4(n-3)}{(n-1)}} \xrightarrow{N\to\infty} \frac{v_4 - \sigma^4}{v_4 - \frac{\sigma^4(n-3)}{(n-1)}}$$

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

$$\text{Var}\left(\widehat{\sigma^2}_{\text{Mean}}\right) = \text{Var}\left(\frac{1}{m}\sum_{j=1}^{m} S_j^2\right) = \frac{1}{m}\text{Var}\left(S_1^2\right) = \frac{1}{m}\left(\frac{v_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}\right)$$

(see Angelova (2012) for the variance $\text{Var}(S_1^2)$ 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 $v_4 = E\left(X_1^4\right) < \infty$. Then we have*

$$\sqrt{N}\left(\widehat{\sigma^2}_{\text{Mean}} - \sigma^2\right) \xrightarrow{d} N\left(0, v_4 - \sigma^4\right).$$

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

$$\frac{\widehat{\sigma^2}_{\text{Mean}} - \sigma^2}{\sqrt{\text{Var}\left(\widehat{\sigma^2}_{\text{Mean}}\right)}} = \frac{\frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{t=1}^{n}\left(X_{j,t} - \overline{X}_j\right)^2 - \sigma^2}{\sqrt{\frac{1}{m}\left(\frac{v_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}\right)}}$$

$$= \sqrt{N}\frac{\frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{t=1}^{n}X_{j,t}^2 - \frac{n}{m(n-1)}\sum_{j=1}^{m}\overline{X}_j^2 - \sigma^2}{\sqrt{v_4 - \frac{\sigma^4(n-3)}{n-1}}}.$$

(5)

For the second term of the numerator in (5), we have that

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

$$= \sqrt{mn}\frac{n}{n-1}\frac{\sigma^2}{n} = \sqrt{\frac{m}{n}\frac{n}{n-1}}\sigma^2 \to 0, \tag{6}$$

since $m = o(n)$. Convergence of the term (6) in mean implies convergence in probability to zero. Application of the Central Limit Theorem to the remaining terms of (5) 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 eighth moments, as will be shown in the following. We set

$$T_{i,j} = \frac{S_{i,j}^2 - \sigma^2}{\sqrt{\frac{m_i}{n_i}\left(v_4 - \frac{\sigma^4(n_i-3)}{n_i-1}\right)}},$$

with $E(T_{i,j}) = 0$ and $\sum_{j=1}^{m_i} E(T_{i,j}^2) = 1 \ \forall i$, where i denotes the ith row of the triangular array. The Lyapunov condition [see corollary 1.9.3 in Serfling (1980)] 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 $v_1, \ldots, v_8$ of X_1 we get

$$\sum_{j=1}^{m_i}E\left(|T_{i,j}|^4\right) = \sum_{j=1}^{m_i}\frac{1}{\frac{m_i^2}{n_i^2}\left(v_4 - \frac{\sigma^4(n_i-3)}{n_i-1}\right)^2}\cdot E\left(\left(S_{i,j}^2 - \sigma^2\right)^4\right)$$

$$= \frac{m_i}{\frac{m_i^2}{n_i^2}\left(v_4 - \frac{\sigma^4(n_i-3)}{n_i-1}\right)^2}\cdot O\left(\frac{1}{n_i^2}\right)\cdot g(v_1, \ldots, v_8)$$

$$= o\left(\frac{1}{m_i}\right) \to 0,$$

where g is a function of the existing moments $v_1, \ldots, v_8$ with $g(v_1, \ldots, v_8) = O(1)$. See Angelova (2012) for the fourth central moment of the sample variance. Therefore, the condition $m = o(n)$ can be dropped.

2.2 Choice of the block size

When choosing 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 (1984), Gasser et al. (1986), Hall et al. (1990), Dette et al. (1998), Munk et al. (2005), Tong et al. (2013), among many others. Dai and Tong (2014) discussed estimation approaches based on differences in nonparametric regression context, Wang et al. (2017) considered an estimation technique which involves differences of second order, while Tecuapetla-Gómez and Munk (2017) 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., Von Neumann et al. (1941)]:

$$\widehat{\sigma}^2_{\text{Diff}} = \frac{1}{2(N-1)} \sum_{j=1}^{N-1} \left(Y_{j+1} - Y_j\right)^2 = \frac{1}{2(N-1)} Y^{\text{T}} A Y, \tag{7}$$

where $A = \tilde{A}^{\text{T}}\tilde{A}$ and $\tilde{A}$ such that $\tilde{A}Y = \left(Y_2 - Y_1, \ldots, Y_N - Y_{N-1}\right)^{\text{T}}$. Theorems 1.5 and 1.6 in Seber and Lee (2012) are used to calculate the expectation and the variance of the estimator $\widehat{\sigma}^2_{\text{Diff}}$:

$$E\left(\widehat{\sigma}^2_{\text{Diff}}\right) = \sigma^2 + \frac{1}{2(N-1)} \mu^{\text{T}} A \mu,$$

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

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

$$\text{Var}\left(\widehat{\sigma}^2_{\text{Diff}}\right) = \nu_4 \frac{4N-6}{4(N-1)^2} + \frac{\sigma^4}{2(N-1)^2},$$

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

For example, for $N = 100$ and a block size $n = 10$, we get $\text{Var}\left(\widehat{\sigma}^2_{\text{Mean}}\right) = 0.0222$, while $\text{Var}\left(\widehat{\sigma}^2_{\text{Diff}}\right) = 0.0302$. Therefore, when no changes in the mean are present the block-estimator can have smaller variance than the difference-based estimator. In Sect. 4 we compare the performance of the proposed estimation procedures with that of the difference-based estimator (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 paper have been performed with the statistical software +R+, version 3.5.2, R Core Team (2018).

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 $\max_n (N - \lfloor N/n \rfloor n) + 1, \ldots, N - \max_n (N - \lfloor N/n \rfloor n)$ 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. 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.
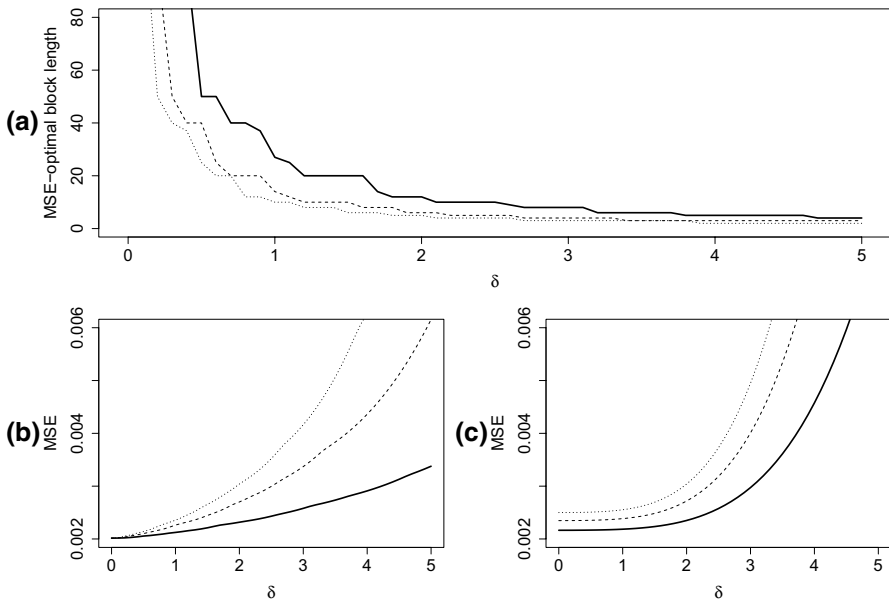
Fig. 1 **a** MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$, **b** MSE regarding n_{opt} of $\widehat{\sigma^2}_{\text{Mean}}$ and **c** MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for $K = 1$ (solid line), $K = 3$ (dashed line) and $K = 5$ (dotted line) with $N = 1000$, $Y_t = X_t + \sum_{k=1}^{K} h I_{t \geq t_k}$, where $X_t \sim N(0, 1)$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$.

Panel (b) of Fig. 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. Rooch et al. (2019). Moreover, in the upper panel of Fig. 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:

$$n = \max \left\{ \left\lfloor \frac{\sqrt{N}}{K+1} \right\rfloor, 2 \right\}, \tag{8}$$

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 (8).

Remark 5 For large N we get $m = N/n = \sqrt{N}(K+1)$. In this case the number of jumps needs to satisfy $K = o(m^{1/3}) = o(N^{1/6}K^{1/3})$, i.e., $K = o(N^{1/4})$ can be tolerated, see Remark 2. 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, see Fig. 1. For example, 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 (7)] is 2/3 [see e.g. Seber and Lee (2012)].

Panel (c) of Fig. 1 shows the MSE of the estimator $\widehat{\sigma^2}_{\text{Mean}}$ with the block size n chosen according to (8). For $K \in \{3, 5\}$ there is only a moderate loss of performance when choosing n according to (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.

Table 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.

The results for data from the t_5-distribution are similar to those obtained for the normal distribution, see Fig. 7 in "Appendix." Again, the blocks-estimator with the block size (8) [Panel (c)] performs well and does not lose too much performance compared to Panel (b) of Fig. 7, where the optimal block size is

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

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

considered. Similar results are obtained for $N = 2500$, see Figs. 9 and 10 in "Appendix".

As the number of change-points K is not known exactly in real applications, there are several possibilities to set K in formula (8):

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

2.3 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 (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 2 shows the MSE of $\widehat{\sigma^2}_{\text{Mean}}$ depending on the jump height $h = \delta\sigma$, $\delta \in \{0, 0.1, \dots, 5\}$, when choosing $n = \frac{\sqrt{N}}{K+1}$ with values $K \in \{0, 1, \dots, 6\}$. 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 8 in "Appendix" 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 block-estimator with growing block size, which can be fully efficient, see Remarks 2 and 5.

We will not consider the block-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.
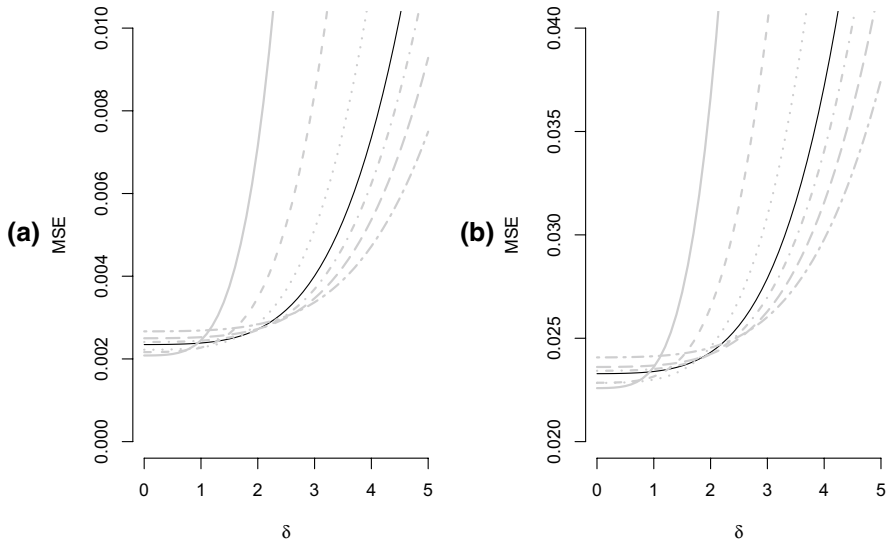
Fig. 2 MSE of $\widehat{\sigma^2}_{\text{Mean}}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for true $K = 3$ (—), $K = 0$ (——) $K = 1$ (- - -) $K = 2$ (· · ·) $K = 4$ (‒ · ‒) $K = 5$ (– – –) and $K = 6$ (– · –) with $N = 1000$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$, $Y_t = X_t + \sum_{k=1}^{3} h I_{t \geq t_k}$, where **a** $X_t \sim N(0, 1)$ and **b** $X_t \sim t_3$

2.4 Pre-estimation of the number of jumps

We will investigate the MOSUM procedure for the detection of multiple change-points proposed by Eichinger and Kirch (2018) 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 with 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

$$\frac{|T_{t,N}(G)|}{\widehat{\tau}_{t,N}} \geq D_N(G, \delta_N) \quad \text{for} \quad t = v_i, \ldots, w_i$$

$$\frac{|T_{t,N}(G)|}{\widehat{\tau}_{t,N}} < D_N(G, \delta_N) \quad \text{for} \quad t = v_i - 1, w_i + 1$$

$$w_i - v_i \geq \eta G \quad \text{with} 0 < \eta < 1/2 \text{arbitrary but fixed,}$$

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 $\widehat{t}_1, \ldots, \widehat{t}_{\widehat{K}}$ 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 Eichinger and Kirch (2018).

We will use the MOSUM procedure to estimate the number of change-points K in the formula (8) for the block size n. The corresponding blocks-estimator is denoted as $\widehat{\sigma}^2{}_{\text{Mean}}^{\text{mosum}}$. In Sect. 4 we will see that the performance of the two blocks-estimators $\widehat{\sigma}^2{}_{\text{Mean}}$ [as defined in (2)] and $\widehat{\sigma}^2{}_{\text{Mean}}^{\text{mosum}}$ is similar in many cases.

Moreover, we introduce another 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 $\widehat{t}_1, \ldots, \widehat{t}_{\widehat{K}}$ of the level shifts. In every block $j = 1, \ldots, \widehat{K} + 1$, the empirical variance S_j^2 is calculated. A weighted average $\widehat{\sigma}^2{}_{\text{W}}^{\text{mosum}}$ of those values can be computed to estimate the variance, i.e.,

$$\widehat{\sigma}^2{}_{\text{W}}^{\text{mosum}} = \sum_{j=1}^{\widehat{K}} \frac{n_j}{N} S_j^2, \tag{9}$$

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

2.5 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 $\sum_{i=0}^{\infty} \psi_i < \infty$. 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 absolutely summable in this case, i.e., $\sum_{\delta=0}^{\infty} |\gamma(\delta)| < \infty$.

Theorem 6 *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 $\widehat{\sigma}^2{}_{\text{Mean}} = \frac{1}{m} \sum_{j=1}^{m} S_j^2 \to \sigma^2$ in probability.*

Proof 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 $\mu = E(X_{j,t}) \, \forall j, t$, $\mu_{j,t} = E(Y_{j,t})$ and $\overline{\mu}_j = \frac{1}{n} \sum_{t=1}^{n} E(Y_{j,t})$. Let the term $S_{j,0}^2$ denote the empirical variance of the uncontaminated data $X_{j,1}, \ldots, X_{j,n}$ in the block $j = 1, \ldots, m$, while $S_{j,h}^2$ is the empirical variance in the perturbed block $j = 1, \ldots, B$.

Then

$$
P\left(\left| \widehat{\sigma^2}_{\mathrm{Mean}} - \sigma^2 \right| > \epsilon \right) = P\left(\left| \frac{1}{m} \sum_{j=B+1}^{m} S_{j,0}^2 + \frac{1}{m} \sum_{j=1}^{B} S_{j,h}^2 - \sigma^2 \right| > \epsilon \right)
$$

$$
= P\left(\left| \frac{1}{m} \sum_{j=B+1}^{m} S_{j,0}^2 + \frac{1}{m} \sum_{j=1}^{B} S_{j,0}^2 \right. \right.
$$

$$
+ \frac{1}{m} \sum_{j=1}^{B} \frac{2}{n-1} \sum_{t=1}^{n} (X_{j,t} - \overline{X}_j)(\mu_{j,t} - \overline{\mu}_j)
$$

$$
\left. \left. + \frac{1}{m} \sum_{j=1}^{B} \frac{1}{n-1} \sum_{t=1}^{n} \left(\mu_{j,t} - \overline{\mu}_j \right)^2 - \sigma^2 \right| > \epsilon \right)
$$

$$
= P\left(\left| \frac{1}{m} \sum_{j=1}^{m} S_{j,0}^2 - \sigma^2 + \frac{1}{m} \sum_{j=1}^{B} \frac{2}{n-1} \sum_{t=1}^{n} X_{j,t}(\mu_{j,t} - \overline{\mu}_j) \right. \right. \tag{10}
$$

$$
\left. \left. + \frac{1}{m} \sum_{j=1}^{B} \frac{1}{n-1} \sum_{t=1}^{n} \left(\mu_{j,t} - \overline{\mu}_j \right)^2 \right| > \epsilon \right)
$$

$$
\leq P\left(\left| \frac{1}{m} \sum_{j=1}^{m} S_{j,0}^2 - \sigma^2 \right| > \frac{\epsilon}{3} \right)
$$

$$
+ P\left(\left| \frac{1}{m} \sum_{j=1}^{B} \frac{2}{n-1} \sum_{t=1}^{n} X_{j,t}(\mu_{j,t} - \overline{\mu}_j) \right| > \frac{\epsilon}{3} \right)
$$

$$
+ P\left(\frac{1}{m} \sum_{j=1}^{B} \frac{1}{n-1} \sum_{t=1}^{n} \left(\mu_{j,t} - \overline{\mu}_j \right)^2 > \frac{\epsilon}{3} \right)
$$

$$
=: A_{1,N} + A_{2,N} + A_{3,N}
$$

For the first term $A_{1,N}$ of (10), we have:

$$A_{1,N} = P\left(\left| \frac{1}{m} \sum_{j=1}^{m} S_{j,0}^2 - \sigma^2 \right| > \frac{\epsilon}{3} \right)$$

$$= P\left(\left| \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n-1} \sum_{t=1}^{n} (X_{j,t} - \overline{X}_j)^2 - \sigma^2 \right| > \frac{\epsilon}{3} \right)$$

$$= P\left(\left| \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n-1} \sum_{t=1}^{n} (X_{j,t} - \mu)^2 + \frac{1}{m} \sum_{j=1}^{m} \frac{n}{n-1} (\mu - \overline{X}_j)^2 \right. \right.$$

$$\left. \left. + \frac{2}{m(n-1)} \sum_{j=1}^{m} \sum_{t=1}^{n} (X_{j,t} - \mu)(\mu - \overline{X}_j) - \sigma^2 \right| > \frac{\epsilon}{3} \right) \qquad (11)$$

$$\leq P\left(\left| \frac{N}{m(n-1)} \frac{1}{N} \sum_{j=1}^{m} \sum_{t=1}^{n} (X_{j,t} - \mu)^2 - \sigma^2 \right| > \frac{\epsilon}{6} \right)$$

$$+ P\left(\left| -\frac{n}{m(n-1)} \sum_{j=1}^{m} (\overline{X}_j - \mu)^2 \right| > \frac{\epsilon}{6} \right)$$

The first term of (11) 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 (11) we have

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

$$= \frac{n}{m(n-1)} \sum_{j=1}^{m} \mathrm{Var}(\overline{X}_j) \leq \frac{n}{(n-1)} \frac{\gamma(0) + 2 \sum_{i=1}^{\infty} \gamma(i)}{n}$$

$$\longrightarrow 0,$$

due to absolute summability of the autocovariance function γ, which implies convergence in probability. Therefore, the second term of (11) converges to zero. Altogether, the sum (11) converges to zero.

The second and the third terms $A_{2,N}$ and $A_{3,N}$ of (10) converge to zero due to the following considerations:

$$E\left(\frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j)\right)$$

$$=\frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}E(X_{j,t})(\mu_{j,t}-\overline{\mu}_j)$$

$$=\frac{1}{m}\sum_{j=1}^{B}\frac{2\mu}{n-1}\sum_{t=1}^{n}(\mu_{j,t}-\overline{\mu}_j)=0$$

$$\text{Var}\left(\frac{1}{m}\sum_{j=1}^{B}\frac{2}{n-1}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j)\right)$$

$$=\frac{4}{m^2(n-1)^2}\text{Var}\left(\sum_{j=1}^{B}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j)\right)$$

$$=\frac{4}{m^2(n-1)^2}Cov\left(\sum_{j=1}^{B}\sum_{t=1}^{n}X_{j,t}(\mu_{j,t}-\overline{\mu}_j),\sum_{l=1}^{B}\sum_{s=1}^{n}X_{l,s}(\mu_{l,s}-\overline{\mu}_l)\right)$$

$$\leq\frac{4}{m^2(n-1)^2}\sum_{j=1}^{B}\sum_{t=1}^{n}\sum_{l=1}^{B}\sum_{s=1}^{n}|\mu_{j,t}-\overline{\mu}_j||\mu_{l,s}-\overline{\mu}_l||Cov(X_{j,t},X_{l,s})|$$

$$\leq\frac{4}{m^2(n-1)^2}\left(\sum_{k=1}^{K}h_k\right)^2\sum_{j=1}^{m}\sum_{t=1}^{n}\sum_{l=1}^{m}\sum_{s=1}^{n}|Cov(X_{j,t},X_{l,s})|$$

$$=\frac{4}{m^2(n-1)^2}\left(\sum_{k=1}^{K}h_k\right)^2\left(NVar(X_{1,1})+2\sum_{u=1}^{N-1}|\gamma(u)|(N-u)\right)$$

$$\leq\frac{4N}{m^2(n-1)^2}\left(\sum_{k=1}^{K}h_k\right)^2\left(Var(X_{1,1})+2\sum_{u=1}^{\infty}|\gamma(u)|\right)$$

$$\longrightarrow 0,$$

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 rth 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

$$A_{3,N} \leq P\left(\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t}-\overline{\mu}_j\right)^2 > \frac{\epsilon_N}{3}\right)$$

$$= P\left(\log(N)\frac{1}{m}\sum_{j=1}^{B}\frac{1}{n-1}\sum_{t=1}^{n}\left(\mu_{j,t}-\overline{\mu}_j\right)^2 > \frac{1}{3}\right)$$

$$\leq P\left(\log(N)\frac{B}{m}\frac{n}{n-1}\left(\sum_{k=1}^{K}h_k\right)^2 > \frac{1}{3}\right)$$

$$\leq P\left(\log(N)\frac{K}{m}\frac{n}{n-1}\left(\sum_{k=1}^{K}h_k\right)^2 > \frac{1}{3}\right) = 0,$$

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

Remark 7 If the jump heights are bounded by a constant $h \geq h_k$, $k = 1, \dots, K$, 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 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 the 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 (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 1. 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 Sects. 3.1 and 3.2, the choice of the trimming fraction is discussed.

3.1 Estimation with a fixed trimming fraction

The trimmed blocks-estimator is given as

$$\widehat{\sigma^2}_{\text{Tr},\alpha} = C_{N,\text{Tr},\alpha} \frac{1}{m - \lfloor \alpha m \rfloor} \sum_{j=1}^{m-\lfloor \alpha m \rfloor} S_{(j)}^2, \tag{12}$$

where $S_{(1)}^2 \leq \cdots \leq S_{(m)}^2$ 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. For example, 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 Sect. 2.2, 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 2 shows the simulated MSE of the trimmed estimator (12) 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 Sect. 3.2.

3.2 Adaptive choice of the trimming fraction

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

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

K	h	$N(0,1)$			t_5		
		$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$
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

$$\widehat{\sigma^2}_{\text{Tr,ad}} = C_{N,\text{Tr,ad}} \frac{1}{m - \lfloor \alpha_{\text{adapt}} m \rfloor} \sum_{j=1}^{m - \lfloor \alpha_{\text{adapt}} m \rfloor} S_{(j)}^2, \tag{13}$$

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

3.2.1 Estimation under normality

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

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

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.,

$$\widehat{\sigma^2}_{\text{Med}} = C_{N,\text{Med}} \cdot \text{med}\{S_1^2, \ldots, S_m^2\}, \tag{14}$$

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

$$\widehat{T}_j = \frac{(n-1)S_j^2}{\widehat{\sigma^2}_{\text{Med}}}, \tag{15}$$

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

$$\beta_m = 1 - (1 - \beta)^{1/m} \quad \text{and} \quad \beta \in (0, 1). \tag{16}$$

We will refer to the adaptively trimmed estimator based on the approach of Davies and Gather (1993) under normality as $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$.

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

$$P(T_1 \leq q_{\chi_{n-1}^2, 1-\beta_m}, \ldots, T_m \leq q_{\chi_{n-1}^2, 1-\beta_m}) = \left(P(T_1 \leq q_{\chi_{n-1}^2, 1-\beta_m})\right)^m$$
$$= \left(1 - \beta_m\right)^m = \left((1 - \beta)^{1/m}\right)^m = 1 - \beta.$$

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 10,000 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.

Table 3 Average number of trimmed blocks in the absence of level shifts for normally distributed data and different N and β for the estimator $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$

β	N	
	1000	2500
0.05	0.0647	0.0583
0.1	0.1268	0.1208

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 (13) 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 $S_j^2, j = 1, \ldots, m$, 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 $C_{1000,\text{Tr,ad}}^{0.05} = 1.0020$, while for $N = 2500$ we have $C_{2500,\text{Tr,ad}}^{0.05} = 1.0009$, so both are nearly 1 and could be neglected with little loss.

3.2.2 Estimation under unknown distribution

When no distributional assumptions are made and the block size n is large, one can use that $\sqrt{n}\left(S_j^2 - \sigma^2\right)/\sqrt{v_4 - \sigma^4}$ is approximately standard normal if the fourth moment exists.

The fourth central moment $v_4 = E\left(\left(X_1 - E(X_1)\right)^4\right)$ needs to be estimated properly in the presence of level shifts. We can estimate it in blocks and then compute the median of the blocks-estimates $\widehat{\mu}_{4,\text{Med}}$, as was done in (14). Then, values

$$\widehat{T}_j = \sqrt{n}\left(S_j^2 - \widehat{\sigma^2}_{\text{Med}}\right)/\sqrt{\widehat{\mu}_{4,\text{Med}} - \widehat{\sigma}^2_{\text{Med}}}, \tag{17}$$

which exceed the $(1 - \beta_m)$-quantile of the standard normal distribution are removed. The corresponding adaptively trimmed estimator is denoted as $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$, where, again, $\beta = 0.05$ [see (16)] will be used in what follows.

Table 4 shows the average number of trimmed blocks in the absence of level shifts for normally distributed data, analogously to Table 3. We observe that the average number of trimmed block-estimates is much larger than the values for the trimming procedure which is based on the normality assumption.

Remark 8 We do not use distribution dependent correction factors for the estimators $\widehat{\mu}_{4,\text{Med}}$ and $\widehat{\sigma^2}_{\text{Med}}$, since the underlying distribution is not known. The asymptotic

Table 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 $\widehat{\sigma}^{2}{}^{\text{other}}_{\text{Tr,ad}}$

β	N	
	1000	2500
$n = 20$		
0.05	1.2327	2.1748
0.1	1.6007	2.8159
$n = 40$		
0.05	0.4018	0.6036
0.1	0.5676	0.8829

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 Choice of the trimming fraction based on the MOSUM procedure

In Sect. 2.4 we have used the MOSUM procedure proposed by Eichinger and Kirch (2018) to estimate the unknown number of jumps K. The corresponding estimate $\widehat{K}$ can also be used to determine the trimming fraction α of the trimmed estimator in (12), i.e., we can set $\alpha = \widehat{K}/m$. We will denote the corresponding estimator as $\widehat{\sigma}^{2}{}^{\text{mos}}_{\text{Tr,ad}}$. We compare the average number of trimmed blocks for the three estimators $\widehat{\sigma}^{2}{}^{\text{mos}}_{\text{Tr,ad}}$, $\widehat{\sigma}^{2}{}^{\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma}^{2}{}^{\text{other}}_{\text{Tr,ad}}$ in a simulation study. We generate 1000 sequences of observations of length $N = 1000$ for different values of K and h. To every sequence we add K jumps of height h at randomly chosen positions, as was done in Sect. 2.2. Table 5 shows the average number of trimmed blocks for data sampled from the standard normal distribution and the corresponding MSE in different scenarios.

Table 5 Average number of trimmed blocks ($\widehat{K}$) and resulting MSE for normally distributed data, $N = 1000$ and different values of K and h

K	h	$\overline{\widehat{K}}$			MSE			
		$\widehat{\sigma}^{2}{}^{\text{mos}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2}{}^{\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2}{}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2}{}^{\text{mos}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2}{}^{\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2}{}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2}_{\text{Mean}}$
0	0	0.09	0.11	0.41	0.23	0.23	0.26	0.22
2	2	2.06	0.45	1.38	0.25	0.26	0.29	0.25
2	5	2.06	1.82	2.13	0.24	0.23	0.26	1.20
2	8	2.07	1.96	2.20	0.23	0.23	0.25	6.71
6	2	5.52	1.10	3.05	0.30	0.53	0.47	0.33
6	5	5.63	5.16	5.26	0.29	0.26	0.29	1.91
6	8	5.68	5.51	5.43	0.43	0.23	0.27	11.20
10	2	8.40	1.66	4.20	0.34	1.11	1.11	0.46
10	5	8.72	8.16	7.89	0.37	0.33	0.39	2.18
10	8	8.78	8.72	8.22	0.84	0.27	0.32	11.65

Moreover, the MSE of the averaging estimator $\widehat{\sigma}^2_{\text{Mean}}$ is displayed in the table for comparison.

We observe that $\widehat{\sigma}^{2\,\text{mos}}_{\text{Tr,ad}}$ trims more blocks on average and yields a smaller MSE than $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$ if the jumps are small. The MSE of the averaging estimator $\widehat{\sigma}^2_{\text{Mean}}$ is similar to that of $\widehat{\sigma}^{2\,\text{mos}}_{\text{Tr,ad}}$ in that case. As opposed to this, the three trimmed estimators trim similarly many blocks on average if the jumps are large, with $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$ leading to smaller MSE values than $\widehat{\sigma}^{2\,\text{mos}}_{\text{Tr,ad}}$. This can be explained by the fact that the former methods base the decision whether to trim a block or not directly on the criterion whether the empirical variance in a block is unusual or not. The averaging estimator $\widehat{\sigma}^2_{\text{Mean}}$ is outperformed by the three trimmed estimators when dealing with large jumps.

For the heavy-tailed t_5-distribution, the results are found in Table 6. The MOSUM-based procedure yields slightly better results for large jump heights, i.e., $h > 2$, since the adaptively trimmed estimators overestimate the number of contaminated blocks in many cases. For small jumps, the adaptively trimmed procedures yield smaller MSE values. The three trimmed estimators perform better than $\widehat{\sigma}^2_{\text{Mean}}$ in every case.

The results for the data generated by the AR(1) model with $\phi = 0.5$ are displayed in Table 7. The MOSUM-based procedure trims considerably more blocks than the adaptively trimmed estimators $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$ yielding higher MSE values. Moreover, the averaging estimator $\widehat{\sigma}^2_{\text{Mean}}$ is outperformed by the adaptively trimmed estimators $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$ in every case.

We conclude that the estimators $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$ perform better when dealing with large jumps and normally distributed data. For small jumps less contaminated blocks are trimmed by the adaptively trimmed estimators resulting in a higher MSE values in many scenarios. In this case the MOSUM-based estimator $\widehat{\sigma}^{2\,\text{mos}}_{\text{Tr,ad}}$ yields the best results which are similar to those of $\widehat{\sigma}^2_{\text{Mean}}$. For heavy-tailed data, the MOSUM-based procedure is preferable under large jumps but is inferior to the other

Table 6 Average number of trimmed blocks ($\widehat{\bar{K}}$) and resulting MSE for t_5-distributed data, $N = 1000$ and different values of K and h

K	h	$\widehat{\bar{K}}$			MSE			
		$\widehat{\sigma}^{2\,\text{mos}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2\,\text{mos}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma}^2_{\text{Mean}}$
0	0	0.07	2.04	1.56	2.17	2.98	2.81	2.11
2	2	2.06	2.26	1.85	2.16	2.54	2.33	2.29
2	5	2.05	3.55	2.99	1.63	2.72	2.46	4.01
2	8	2.04	3.73	3.06	1.74	2.90	2.61	8.80
6	2	5.45	2.38	2.26	3.81	1.91	1.91	1.85
6	5	5.57	6.12	5.43	1.98	2.58	2.51	3.93
6	8	5.65	6.86	5.87	2.19	2.78	2.56	15.48
10	2	8.32	2.60	2.79	4.97	1.61	2.94	2.87
10	5	8.64	8.48	7.41	2.04	2.43	3.74	4.57
10	8	8.74	9.63	8.18	2.49	2.53	3.06	13.99

Table 7 Average number of trimmed blocks ($\widehat{K}$) and resulting MSE for data generated from the AR(1) model with $\phi = 0.5$ and $N = 1000$ for different values of K and h

K	h	$\overline{\widehat{K}}$			MSE			
		$\widehat{\sigma^2}^{\text{mos}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{mos}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}_{\text{Mean}}$
0	0	5.08	0.66	1.27	6.01	2.69	1.84	1.17
2	2	6.79	1.04	1.84	6.77	2.51	1.64	5.46
2	5	6.47	2.29	2.75	5.94	2.69	1.80	2.65
2	8	6.37	2.48	2.83	5.72	2.72	1.72	1.32
6	2	9.41	1.54	2.82	7.26	1.74	1.16	20.85
6	5	9.00	5.33	5.55	5.43	2.50	1.64	12.41
6	8	8.96	5.80	5.71	5.26	2.58	1.54	3.29
10	2	11.53	1.99	3.65	7.37	1.22	1.04	42.02
10	5	11.26	8.02	7.87	4.72	2.06	1.37	29.70
10	8	11.30	8.92	8.27	4.52	2.40	1.35	13.75

trimmed estimates if the jumps are small. When dealing with positively correlated processes the adaptively trimmed estimators $\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$ perform best. Therefore, the choice of the most appropriate estimator depends on the knowledge about the underlying distribution and the height of the jumps.

In the simulation study in Sect. 4, we concentrate on the adaptively trimmed estimators $\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$ since they yield very good results in many scenarios, and on the averaging estimator $\widehat{\sigma^2}_{\text{Mean}}$ since it performs well when dealing with a few small jumps and independent data.

4 Simulations

In this section we compare the estimators $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr,0.5}}$ with the block size $n = 20$, $\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$ with the block size $n = 20$ and $\beta = 0.05$, $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$ with the block size $n = 40$ and $\beta = 0.05$, $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ and $\widehat{\sigma^2}_{\text{W}}$ 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 Sect. 2.2. 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 8 shows the simulated MSE for the normal distribution. The estimators $\widehat{\sigma^2}_{\text{Mean}}$ and $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ yield similar results. We conclude that the estimation of the number of jumps K [required in the rule (8)] does not have a large effect on the estimator. The estimators $\widehat{\sigma^2}_{\text{Mean}}$ and $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ 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 $\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$ yields the best results among all methods considered here for normally distributed data if the jumps are rather high. When many

Table 8 Simulated MSE·10^2 of $\widehat{\sigma}^2_{\text{Mean}}$, $\widehat{\sigma}^2_{\text{Diff}}$, $\widehat{\sigma}^2_{\text{Tr},0.5}$, $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$, $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma}^{2\,\text{mosum}}_{\text{Mean}}$ 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	$\widehat{\sigma}^2_{\text{Mean}}$	$\widehat{\sigma}^2_{\text{Diff}}$	$\widehat{\sigma}^2_{\text{Tr},0.5}$	$\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma}^{2\,\text{mosum}}_{\text{Mean}}$	$\widehat{\sigma}^{2\,\text{mosum}}_{\text{W}}$
		$N = 200$						
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

Table 8 (continued)

K	h	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr,0.5}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$	$\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$	$\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$
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

small level shifts are present $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ outperforms $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$, 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 $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$ in the absence of level shifts, see Tables 3 and 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 Sect. 3.1. The difference-based estimator $\widehat{\sigma^2}_{\text{Diff}}$ performs well as long as the jumps are moderately high.

Table 9 shows the results for the t_5 distribution. The estimation procedures $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$ and $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ yield the best results in this scenario.

In Table 10 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. Since the variance σ^2 is underestimated when the data are dependent, the values $\widehat{T}_j$ in (15) and (17) 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.

Table 9 Simulated MSE·10^2 of $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr,0.5}}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$, $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ and $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$ 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	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr,0.5}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$	$\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$	$\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$
		$N = 200$						
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

Table 9 (continued)

K	h	$\widehat{\sigma}^2_{\text{Mean}}$	$\widehat{\sigma}^2_{\text{Diff}}$	$\widehat{\sigma}^2_{\text{Tr,0.5}}$	$\widehat{\sigma}^2{}^{\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma}^2{}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma}^2{}^{\text{mosum}}_{\text{Mean}}$	$\widehat{\sigma}^2{}^{\text{mosum}}_{\text{W}}$
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

Figure 3 shows the simulated MSE dependent on the AR-parameter $\phi \in \{0.1, \ldots, 0.8\}$ in four different scenarios: no jumps, few small jumps, many small jumps, and many high jumps. We observe that the performance of all estimators $\widehat{\sigma}^2_{\text{Mean}}$, $\widehat{\sigma}^2_{\text{Diff}}$, $\widehat{\sigma}^2_{\text{Tr,0.5}}$, $\widehat{\sigma}^2{}^{\text{norm}}_{\text{Tr,ad}}$, $\widehat{\sigma}^2{}^{\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma}^2{}^{\text{mosum}}_{\text{Mean}}$ 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. 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 are 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 Eichinger and Kirch (2018).

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.
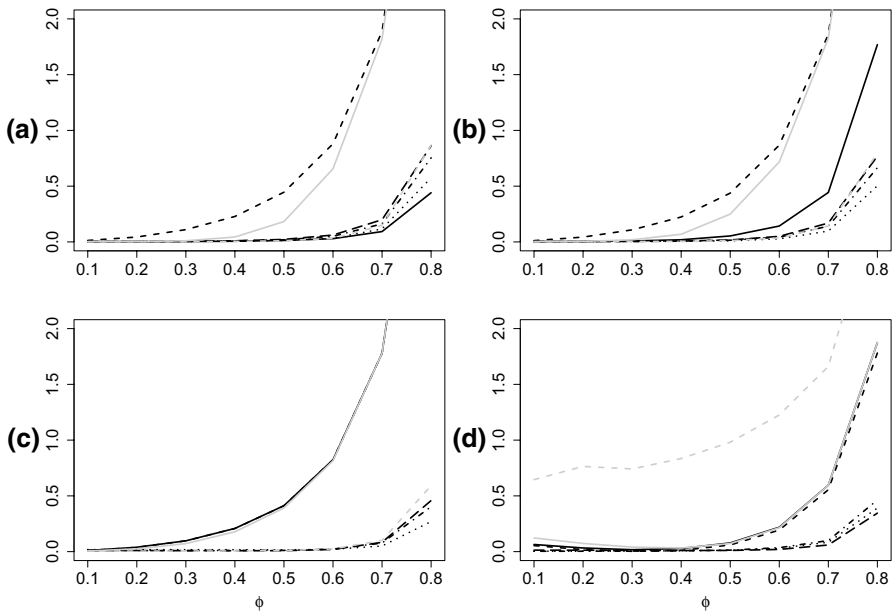
5 Application

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

Table 10 Simulated MSE$\cdot 10^2$ of $\widehat{\sigma^2}_{\text{Mean}}$, $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr,0.5}}$, $\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$ 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 $p \in \{0, 2/1000, 4/1000, 6/1000, 10/1000\}$

K	h	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr,0.5}}$	$\widehat{\sigma^2}^{\text{norm}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{other}}_{\text{Tr,ad}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{Mean}}$	$\widehat{\sigma^2}^{\text{mosum}}_{\text{W}}$
		$N = 200$						
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

Table 10 (continued)

K	h	$\widehat{\sigma^2}_{\text{Mean}}$	$\widehat{\sigma^2}_{\text{Diff}}$	$\widehat{\sigma^2}_{\text{Tr,0.5}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$	$\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$	$\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$	$\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$
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

Fig. 3 Simulated MSE of $\widehat{\sigma^2}_{\text{Mean}}$ (——), $\widehat{\sigma^2}_{\text{Diff}}$ (- - -), $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$ ($\cdots$) , $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ (— · —), $\widehat{\sigma^2}_{\text{Tr,0.5}}$ (— — —), $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ with bandwidth $G = \sqrt{N}$ (——) and $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$ 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 $N = 1000$, $Y_t = X_t + \sum_{k=1}^{K} h I_{t \geq t_k}$, where X_t originates from the AR(1)-process with parameter $\phi \in \{0.1, 0.2, \ldots, 0.8\}$

5.1 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
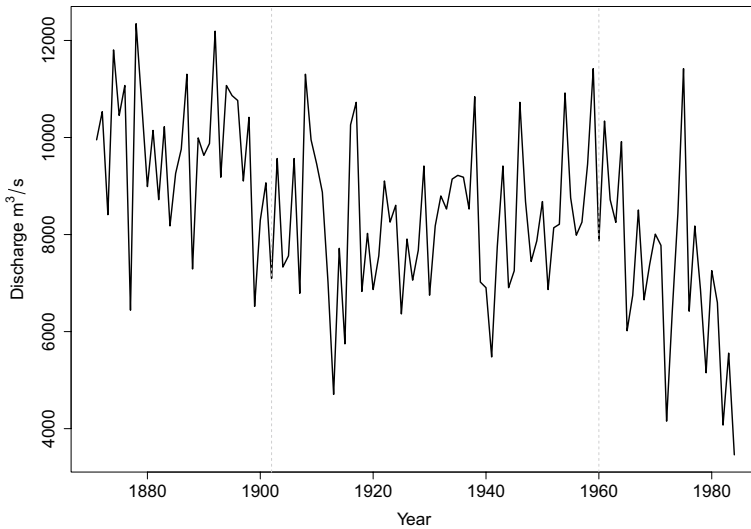
Fig. 4 Maximal monthly discharge of the Nile river at Aswan in the period 1871–1984

Table 11 Rounded estimates ($\cdot 10^{-3}$) of the variance for the annual maxima of the average monthly discharge of the Nile river in Aswan

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

to be independent in hydrology. The maxima are determined from January to December. The flooding season is from July to September, see Hassan (1981). Figure 4 shows the annual maxima of the average monthly discharge of the Nile river for the years 1871–1984.

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). We used Levene's test [see Section 12.4.2 in Fox (2015)] 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. 11 in "Appendix". With $\beta = 0.05$ and $n = 10$ ($m = 11$ blocks) no blocks are trimmed away during the trimming procedures $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$ and $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$. The ordinary sample variance of the entire data yields the value $3{,}243{,}866$, see Table 11. For the blocks-estimator of the variance from (2) we choose the block size according to (8) with $K = 2$ getting $n = \lfloor\sqrt{114/3}\rfloor = 3$. All blockwise estimators examined in this paper yield much smaller variance estimates for the whole observation period, ranging from $2{,}075{,}819$ to $2{,}684{,}368$.

We conclude that the procedures $\widehat{\sigma}^2_{\text{Mean}}$, $\widehat{\sigma}^2_{\text{Diff}}$, $\widehat{\sigma}^2_{\text{Tr},0.5}$, $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$, $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$, $\widehat{\sigma}^{2\,\text{mosum}}_{\text{Mean}}$ and $\widehat{\sigma}^{2\,\text{mosum}}_{\text{W}}$ 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.

5.2 PAMONO data

In the second example, we use data obtained from the PAMONO (Plasmon Assisted Microscopy of Nano-Size Objects) biosensor, see Siedhoff et al. (2014). This technique is used for detection of small particles, e.g. viruses, in a sample fluid. For more details, see Siedhoff et al. (2011). 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. 12 in "Appendix".

In Panel (a) of Fig. 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. 5 shows a boxplot
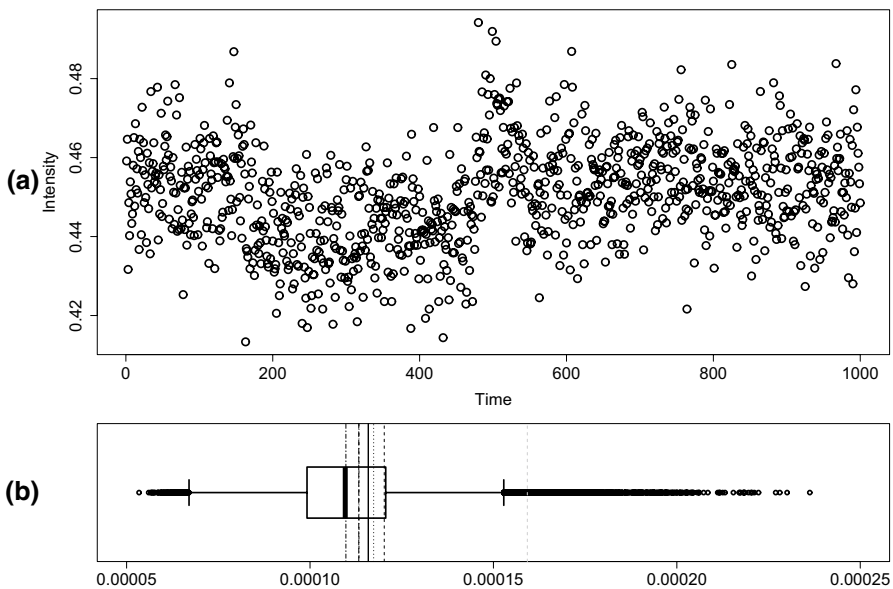
Fig. 5 a Intensity over time for one pixel and **b** a boxplot of variances for the virus-free pixels together with the ordinary sample variance of the above data (– – –) and values of $\widehat{\sigma}^2_{\text{Mean}}$ and $\widehat{\sigma}^2_{\text{Diff}}$ (—), $\widehat{\sigma}^2_{\text{Tr},0.5}$ (- - -), $\widehat{\sigma}^{2\,\text{norm}}_{\text{Tr,ad}}$ (· · ·), $\widehat{\sigma}^{2\,\text{other}}_{\text{Tr,ad}}$ (- · -), $\widehat{\sigma}^{2\,\text{mosum}}_{\text{Mean}}$ (– – –) and $\widehat{\sigma}^{2\,\text{mosum}}_{\text{W}}$ (– –)

of 101,070 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 1.59×10^{-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 paper yield values within the interval $[1.1 \times 10^{-4}, \ 1.2 \times 10^{-4}]$ which are well within the interquartile range. We conclude that these approaches yield reasonable estimates for these data.

5.3 PAMONO data with trend

Again, we consider a PAMONO dataset, see Sect. 5.2. Panel (a) of Fig. 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. 6 shows the differenced data, i.e., $Y_t - Y_{t-1}$, $t = 2, \dots, 388$.

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 $K \in \{1, \dots, 5\}$ in the formula (8)], $\widehat{\sigma^2}_{\text{Diff}}$, $\widehat{\sigma^2}_{\text{Tr},0.5}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$, $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$, $\widehat{\sigma^2}_{\text{Mean}}^{\text{mosum}}$ and $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}}$ to the data and get estimated
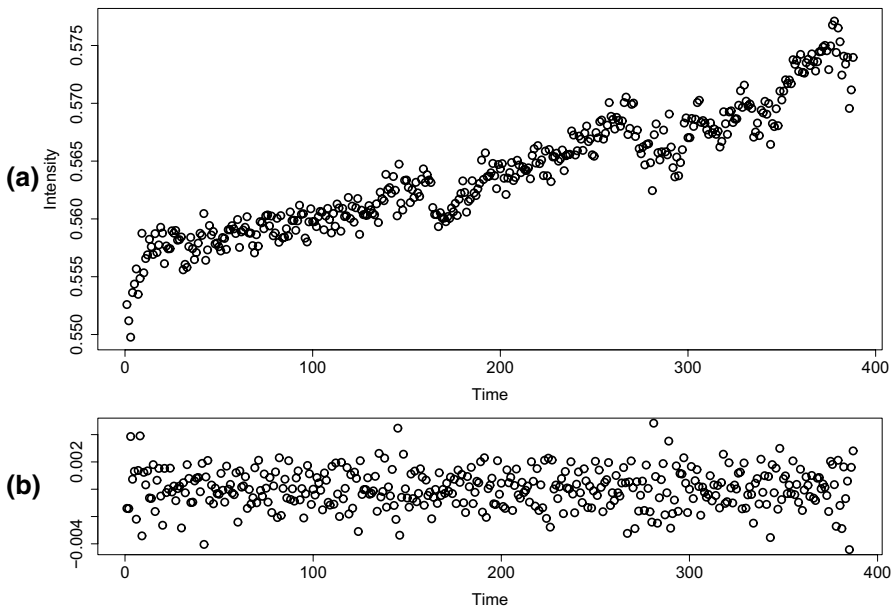
Fig. 6 **a** Intensity over time for one pixel and **b** corresponding differenced series

values for the variance, which range from $\widehat{\sigma^2}_{\text{Mean}} = 0.95 \times 10^{-6}$ (with $K = 5$) to $\widehat{\sigma^2}_{\text{W}}^{\text{mosum}} = 1.66 \times 10^{-6}$. The empirical variance of the observations has the value 26.48×10^{-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 1.93×10^{-6}, which is an estimator of $2\sigma^2$, yielding the value 0.97×10^{-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.

6 Conclusion

In the presence of level shifts, ordinary variance estimators like the empirical variance perform poorly. In this paper, 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 $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{norm}}$ and $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$. 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 (17). 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 $\widehat{\sigma^2}_{\text{Tr,ad}}^{\text{other}}$ can be recommended. For a large sample size N the finite sample correction factors can be neglected with little loss, see "Appendix".

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 Wornowizki et al. (2017), where tests for changes in variability have been investigated using blockwise approaches, assuming a constant mean. This will be an issue for future work.

Appendix

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

$$\widehat{\sigma}_{\text{Mean},1}^{\text{corr}} = C_{N,1}\widehat{\sigma}_{\text{Mean},1} = C_{N,1}\frac{1}{m}\sum_{j=1}^{m} S_j \quad \text{and} \tag{18}$$

$$\widehat{\sigma}_{\text{Mean},2}^{\text{corr}} = C_{N,2}\widehat{\sigma}_{\text{Mean},2} = C_{N,2}\sqrt{\frac{1}{m}\sum_{j=1}^{m} S_j^2} = C_{N,2}\sqrt{\widehat{\sigma^2}_{\text{Mean}}}, \tag{19}$$

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 (8). If the number of change-points is not known in practice, it can be estimated as is done in Sect. 2.4.

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 $\widehat{\sigma}_{\text{Mean},1}^{\text{corr}}$, 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 (18). Given

independent identically normally distributed data $X_{j,1}, \dots, X_{j,n}$ it is well known that $\frac{(n-1)S_j^2}{\sigma^2} \sim \chi_{n-1}^2$ in a jth 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 $t = 1, \dots, n$ and containing $\widetilde{K}_1 \leq K$ jumps.

Lemma 9 *Assume that $X_1, \dots, X_n \sim \mathcal{N}(0, \sigma^2)$ and $Y_t = X_t + \sum_{k=1}^{\widetilde{K}_1} h_k I_{t \geq t_k}$ for $t = 1, \dots, n$. Then we have for $S_1^2 = \frac{1}{n-1} \sum_{t=1}^{n} (Y_t - \overline{Y}_1)^2$ that*

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

where $\lambda_1 = \frac{1}{\sigma^2} \sum_{t=1}^{n} (\mu_{1,t} - \overline{\mu}_1)^2$, $\mu_{1,t} = \sum_{k=1}^{\widetilde{K}_1} h_k I_{t \geq t_k}$ 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 $Y_t - \overline{Y}_1 = X_t - \overline{X}_1 - \overline{\mu}_1 + \mu_{1,t}$, $t = 1, \dots, n$, 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,

$$\sum_{t=1}^{n} \left(\frac{Y_t - \overline{\mu}_1}{\sigma} \right)^2 = \sum_{t=1}^{n} \left(\frac{Y_t - \overline{Y}_1 + \overline{Y}_1 - \overline{\mu}_1}{\sigma} \right)^2$$

$$= \sum_{t=1}^{n} \left(\frac{Y_t - \overline{Y}_1}{\sigma} \right)^2 + \sum_{t=1}^{n} \left(\frac{\overline{Y}_1 - \overline{\mu}_1}{\sigma} \right)^2$$

$$+ 2 \left(\frac{\overline{Y}_1 - \overline{\mu}_1}{\sigma} \right) \sum_{t=1}^{n} \left(\frac{Y_t - \overline{Y}_1}{\sigma} \right)$$

$$= \frac{n-1}{\sigma^2} S_1^2 + \frac{n}{\sigma^2} \left(\overline{Y}_1 - \overline{\mu}_1 \right)^2 + 0 = \frac{n-1}{\sigma^2} S_1^2 + \frac{n}{\sigma^2} \overline{X}_1^2$$

with $\sum_{t=1}^{n} \left(\frac{Y_t - \overline{\mu}_1}{\sigma} \right)^2 \sim \chi_{n,\lambda_1}^2$, since $Y_t \, \forall t$ are independent and $\frac{n}{\sigma^2} \overline{X}_1^2 \sim \chi_1^2$. The moment-generating function at $z \in \mathbb{R}$ of both sides and the independence of S_1^2 and $\overline{Y}_1$ yield:

$$(1 - 2 \cdot z)^{-n/2} \exp \left(\frac{\lambda_1 z}{1 - 2z} \right) = M_{\chi_{n,\lambda_1}^2}(z) = M_{\frac{n-1}{\sigma^2} S_1^2}(z) \cdot M_{\chi_1^2}(z)$$

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

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

$$\exp \left(\frac{\lambda_1 z}{1 - 2z} \right) = M_{\chi_{n-1,\lambda_1}^2}(z)$$

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

In the following, we assume that $B \leq K$ blocks are contaminated by $\widetilde{K}_1, \dots, \widetilde{K}_B$ 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. 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 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

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

(20)

where $F_{1,1}(a, b, z)$ represents the generalized hypergeometric function, see Olver et al. (2010) 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 (20) when no level shifts are present, since $F_{1,1}(-0.5, 0.5(n-1), -0.5\lambda_j) = 1$, $j = 1, \ldots, m$, in this case.

For the second estimator (19), we have the following statements on its expectation and a suitable finite sample correction factor:

$$\hat{\sigma}_{\text{Mean},2}^{\text{corr}} = C_{N,2} \frac{\sigma}{\sqrt{m(n-1)}} \left(\frac{n-1}{\sigma^2} \sum_{j=B+1}^{m} S_j^2 + \frac{n-1}{\sigma^2} \sum_{j=1}^{B} S_j^2 \right)^{1/2},$$

$$E\left(\hat{\sigma}_{\text{Mean},2}^{\text{corr}} \right) = C_{N,2} \frac{\sigma\sqrt{2}}{\sqrt{m(n-1)}} \frac{\Gamma(0.5(m(n-1)+1))}{\Gamma(0.5m(n-1))}$$

$$\times F_{1,1}\left(-0.5, 0.5m(n-1), -0.5 \sum_{j=1}^{B} \lambda_j \right)$$

$$:= C_{N,2} \sigma D_{n,\lambda_1,\ldots,\lambda_B},$$

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

We used the fact that $\hat{\sigma}_{\text{Mean},2}$ 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 Miller (1964) and Olver et al. (2010). The correction factor is the reciprocal of $D_{n,\lambda_1,\ldots,\lambda_B}$, where we have $F_{1,1}\left(-0.5, 0.5m(n-1), -0.5 \sum_{j=1}^{B} \lambda_j \right) = 1$ 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 (18)] and $\widehat{\sigma}_{\text{Mean},2} = \sqrt{\frac{1}{m} \sum_{j=1}^{m} S_j^2}$ [as defined in (19)] of σ:

Corollary 10 *Under the conditions of Theorem 1 the estimators* $\widehat{\sigma}_{\text{Mean},1}$ *and* $\widehat{\sigma}_{\text{Mean},2}$ *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 Hu et al. (1989) 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 $\text{Var}(S_j) \to 0$.

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 Serfling (1980).

Using Jensen's inequality we get for the second term

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

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

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

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

Table 12 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)

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

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

Remark 11 The correction factors $C_{N,1}$ and $C_{N,2}$ from (18) and (19) satisfy

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

where $C_{N,1} = \sigma/E(\hat{\sigma}_{\text{Mean},1})$ and $C_{N,2} = \sigma/E(\hat{\sigma}_{\text{Mean},2})$ in the absence of level shifts.

This can be shown with Lemma 1.4A in Serfling (1980), since $\hat{\sigma}_{\text{Mean},1}$ and $\hat{\sigma}_{\text{Mean},2}$ 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 $\hat{\sigma}_{\text{Mean},1}$ and $\hat{\sigma}_{\text{Mean},2}$ instead of $\hat{\sigma}_{\text{Mean},1}^{\text{corr}}$ and $\hat{\sigma}_{\text{Mean},1}^{\text{corr}}$ with block sizes $n \to \infty$.

Trimmed estimation

When dealing with a large number of level shifts, as is discussed in Sect. 3, the square root of the variance estimator $\widehat{\sigma^2}_{\text{Tr,ad}}$ from (13) 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 12 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 $\phi = 0.6$.
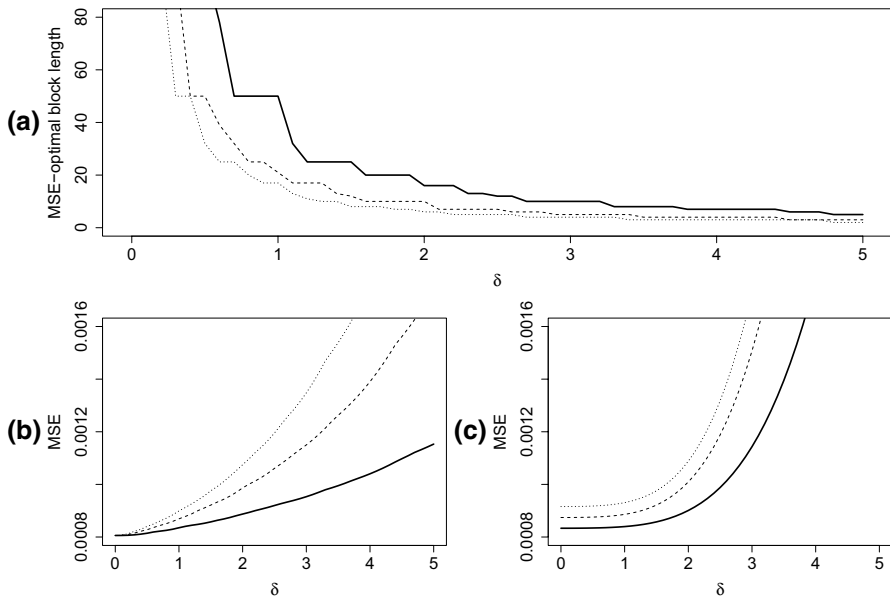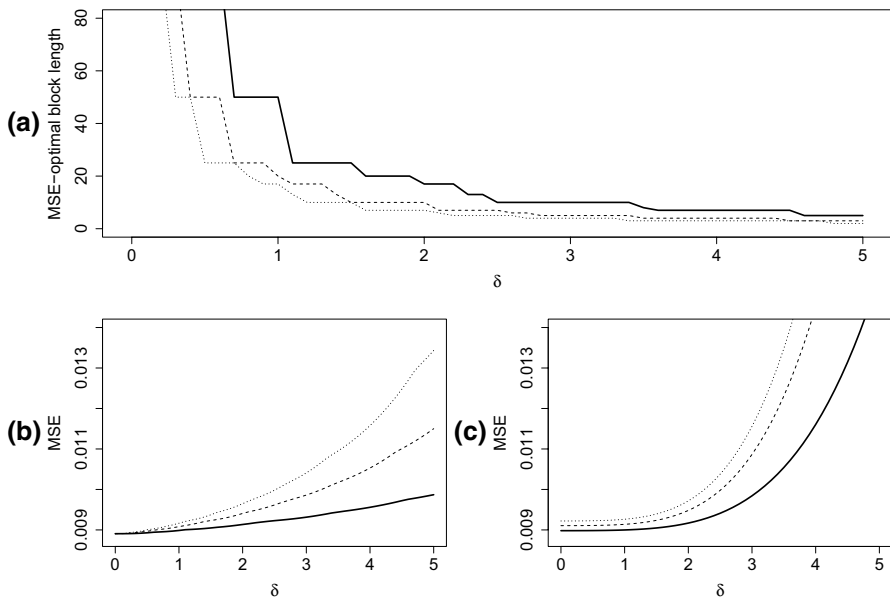
See Figs. 7, 8, 9, 10, 11 and 12.
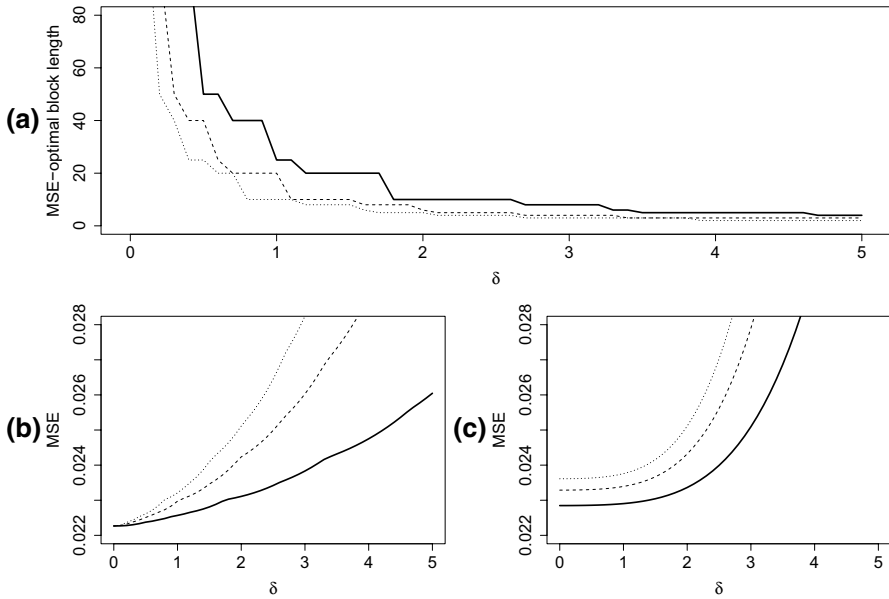
Fig. 7 a MSE-optimal block length n_{opt} of $\widehat{\sigma^2}_{Mean}$, **b** MSE regarding n_{opt} of $\widehat{\sigma^2}_{Mean}$ and **c** MSE of $\widehat{\sigma^2}_{Mean}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for $K = 1$ (—), $K = 3$ (- - -) and $K = 5$ ($\cdots$) with $N = 1000$, $Y_t = X_t + \sum_{k=1}^{K} hI_{t \geq t_k}$, where $X_t \sim t_5$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$
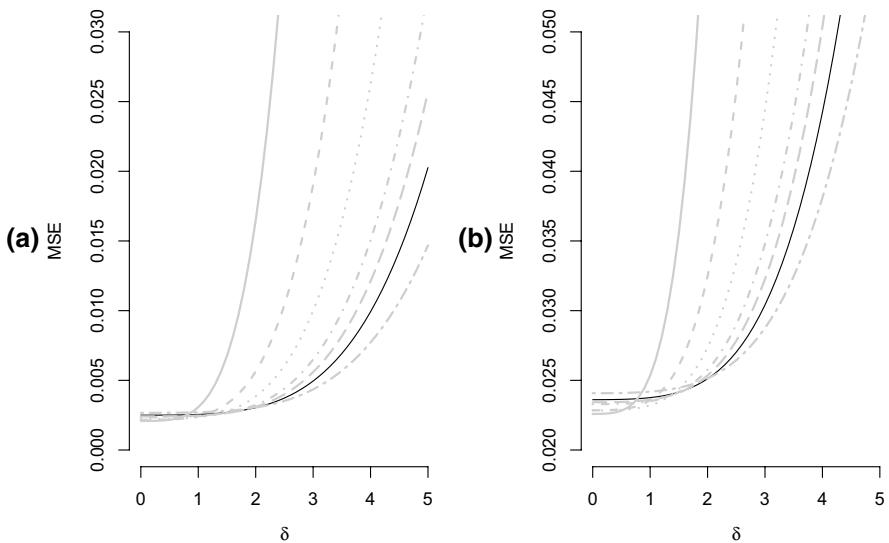

Fig. 8 MSE of $\widehat{\sigma^2}_{Mean}$ when choosing $n = \frac{\sqrt{N}}{K+1}$ for true $K = 5$ (—), $K = 0$ (—) $K = 1$ (- - -) $K = 2$ ($\cdots$) $K = 3$ (- · -) $K = 4$ (– – –) and $K = 6$ (– - –) with $N = 1000$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \ldots, 5\}$, $Y_t = X_t + \sum_{k=1}^{5} hI_{t \geq t_k}$, where **a** $X_t \sim N(0, 1)$ and **b** $X_t \sim t_3$

Fig. 9 **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 $X_t \sim N(0, 1)$ and $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \dots, 5\}$

Fig. 10 **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 $h = \delta \cdot \sigma$, $\delta \in \{0, 0.1, \dots, 5\}$
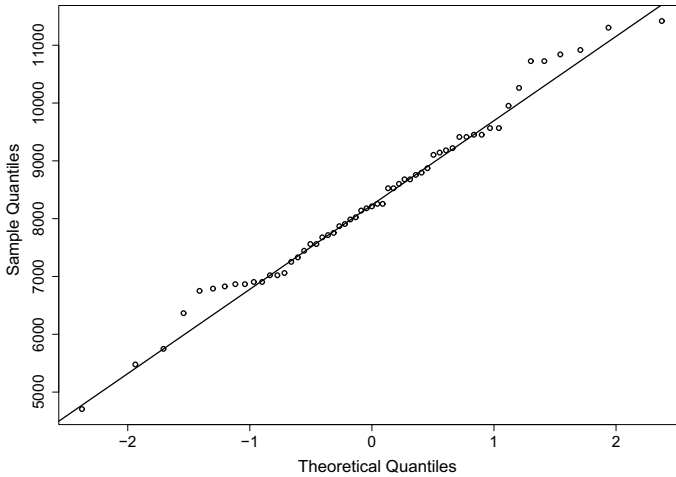
Fig. 11 Q–Q plot of the maximal monthly discharge of the Nile river in the period 1871–1984
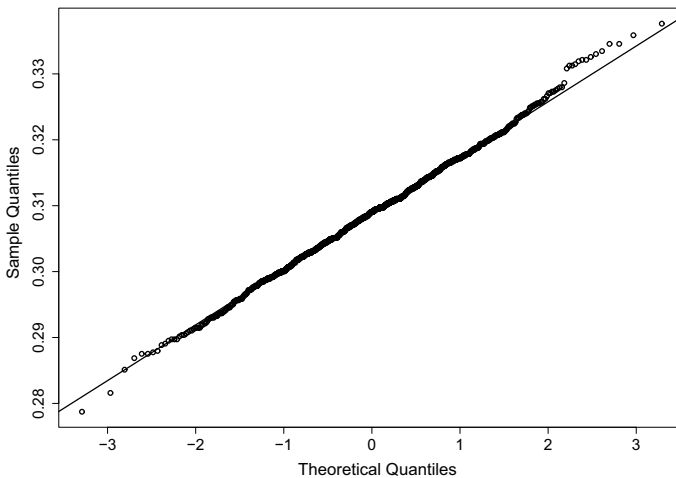
Fig. 12 Q–Q plot of a pixel with a virus adhesion from the PAMONO data

References

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

Dai, W., Tong, T.: Variance estimation in nonparametric regression with jump discontinuities. J. Appl. Stat. **41**(3), 530–545 (2014)

Dai, W., Ma, Y., Tong, T., Zhu, L.: Difference-based variance estimation in nonparametric regression with repeated measurement data. J. Stat. Plan. Inference **163**, 1–20 (2015)

Davies, L., Gather, U.: The identification of multiple outliers. J. Am. Stat. Assoc. **88**(423), 782–792 (1993)

Dette, H., Munk, A., Wagner, T.: Estimating the variance in nonparametric regression—what is a reasonable choice? J. R. Stat. Soc. Ser. B (Stat. Methodol.) **60**(4), 751–764 (1998)

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

Fox, J.: Applied Regression Analysis and Generalized Linear Models. Sage Publications, Thousand Oaks (2015)

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

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

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

Hipel, K.W., McLeod, A.I.: Time Series Modelling of Water Resources and Environmental Systems, vol. 45. Elsevier, Amsterdam (1994)

Hu, T.C., Moricz, F., Taylor, R.: Strong laws of large numbers for arrays of rowwise independent random variables. Acta Math. Hung. **54**(1–2), 153–162 (1989)

Miller, K.S.: Multidimensional Gaussian Distributions. Wiley, New York (1964)

Munk, A., Bissantz, N., Wagner, T., Freitag, G.: On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **67**(1), 19–41 (2005)

Olver, F.W., Lozier, D.W., Boisvert, R.F., Clark, C.W.: NIST Handbook of Mathematical Functions Hardback and CD-ROM. Cambridge University Press, Cambridge (2010)

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

Rice, J., et al.: Bandwidth choice for nonparametric regression. Ann. Stat. **12**(4), 1215–1230 (1984)

Rooch, A., Zelo, I., Fried, R.: Estimation methods for the lrd parameter under a change in the mean. Stat. Pap. **60**(1), 313–347 (2019)

Seber, G.A., Lee, A.J.: Linear Regression Analysis, vol. 329. Wiley, New York (2012)

Serfling, R.: Approximation Theorems of Mathematical Statistics. Wiley, New York (1980)

Siedhoff, D., Weichert, F., Libuschewski, P., 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), p. 9 (2011)

Siedhoff, D., Zybin, A., Shpacovitch, V., Libuschewski, P.: Pamono sensor data 200nm_10apr13.sfb876 (2014). https://doi.org/10.15467/e9ofqnvl6o. https://sfb876.tu-dortmund.de/auto?self=%24e45xtiwhs0

Syvitski, J.P., Saito, Y.: Morphodynamics of deltas under the influence of humans. Glob. Planet. Change **57**(3–4), 261–282 (2007)

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

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

Von Neumann, J., Kent, R., Bellinson, H., Bt, Hart: The mean square successive difference. Ann. Math. Stat. **12**(2), 153–162 (1941)

Wang, W., Lin, L., Yu, L.: Optimal variance estimation based on lagged second-order difference in nonparametric regression. Comput. Stat. **32**(3), 1047–1063 (2017)

Wornowizki, M., Fried, R., Meintanis, S.G.: Fourier methods for analyzing piecewise constant volatilities. AStA Adv. Stat. Anal. **101**(3), 289–308 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.