Bootstrapping sequential change-point tests*

Claudia Kirch†

February 2008

Abstract

In this paper we propose some bootstrapping methods to obtain critical values for sequential change-point tests. We consider a change in the mean with i.i.d. errors. Theoretical results show the asymptotic validity of the proposed bootstrap procedures. A simulation study compares the bootstrap and the asymptotic tests and shows that the studentized bootstrap test behaves generally better than asymptotic tests if measured by α– resp. β–errors and its run length.

Keywords: Bootstrap, sequential test, change-point analysis, mean change

AMS Subject Classification 2000: 62G09, 62L10

1 Introduction

For many testing procedures in change-point analysis the calculation of critical values is based on the limit behavior of the test statistic under the null hypothesis. However, the convergence to the limit distribution of the test statistic is frequently rather slow, in other cases the explicit form is unknown. For time series models it can also happen that the limit distribution does not take the small sample dependency structure sufficiently into account. Therefore permutation and bootstrap tests have been developed. Some guidelines for bootstrap hypothesis testing are given by Hall and Wilson [5]. For a thorough introduction into permutation and bootstrap tests we refer to Good [3].

In change-point analysis this approach was first suggested by Antoch and Hušková [1] and later pursued by others (for a recent survey confer Hušková [7]). Kirch [9, 10] proposed some permutation procedures to obtain critical values for change-point tests if the errors are no longer independent.

*This work was supported by the DFG Graduiertenkolleg 'Mathematik und Praxis'.

†University of Kaiserslautern, Department of Mathematics, Erwin-Schrödinger-Straße, D–67 653 Kaiserslautern, Germany; ckirch@mathematik.uni-kl.de

All of those papers, however, deal with a posteriori tests, i.e. tests, where we have observed the complete data set already. In many situations sequential or on-line monitoring tests are much more realistic. In those models data arrive steadily and with each new observation the question arises whether the model is still capable of explaining the data. Critical values in this setting are also frequently based on asymptotics. Additionally to the problems of a-posteriori tests the asymptotics usually assume that the monitoring goes on for an infinite time horizon. In many situations it is much more realistic to monitor data only for a finite time horizon (maybe as long or twice as long as the historical data set, on which the preliminary assumptions are based). If the calculation of the critical values is based on an infinite observation period but in fact it is finite, one necessarily looses some power.

It is not obvious how best to do bootstrapping in a sequential setting. New data arrive steadily, so we could use these new observations in the bootstrap and hopefully improve the estimate of the critical values. From a practical point of view this is computationally expensive, so one might think of alternatives, which are less expensive and still good enough. From a theoretical point of view this means that we have new critical values with each incoming observation, so the question is whether this procedure remains consistent. The literature on bootstrapping methods for sequential tests is very scarce. Steland [12] used a bootstrap in sequential testing of the unit-root problem.

In the a posteriori setting permutation tests are often preferable to bootstrap tests if the underlying error structure is exchangeable, because those tests are exact whereas bootstrap tests are only asymptotically consistent. In the sequential setting this approach is not possible because the test statistic is based on more observations than what we have already observed.

In this paper we concentrate on examining and comparing different bootstrap procedures for sequential change-point tests. To this end we use the same model as Antoch and Hušková [1], when first considering permutation tests for change-point analysis but we use it in a sequential setting. In forthcoming work this will be extended to more difficult models.

We consider the following mean change problem. Let

$$X(i) = \mu(i) + \epsilon(i), \tag{1.1}$$

where $\{\epsilon(i) : i \geqslant 1\}$ is a sequence of i.i.d. random variables with

$$\mathrm{E}\,\epsilon(1) = 0, \qquad 0 < \sigma^2 = \mathrm{var}\,\epsilon(1) < \infty, \qquad \mathrm{E}\,|\epsilon(1)|^\nu < \infty \text{ for some } \nu > 2. \tag{1.2}$$

We assume that we have observed a historic data set of length m, where no change has occurred, i.e.

$$\mu(i) = \mu_0, \quad 1 \leqslant i \leqslant m. \tag{1.3}$$

Now, we are interested in monitoring the future incoming observations sequentially for a change in the mean, i.e. we want to test the null hypothesis

$$H_0 : \mu(i) = \mu_0, \quad i > m,$$

against the alternative

$$\begin{aligned} H_1 &: \text{there exists } k^* \geqslant 0 \text{ such that } \mu(i) = \mu_0,\ m < i \leqslant m + k^*, \\ &\quad \text{but } \mu(i) = \mu_0 + d,\ i > m + k^*,\ d \neq 0. \end{aligned}$$

The values of μ_0 and d are not specified, $d = d(m)$ and $k^* = k^*(m)$ may depend on m.

Note that this is a special case of the linear model discussed in Horváth et al. [6]. We also use their monitoring statistic but with a possibly finite monitoring horizon $N(m)$. It is based on

$$\begin{aligned}\Gamma(m,k,\gamma) &= \Gamma(m,k,\gamma)(X(1),X(2),\ldots,X(m+k)) \\ &= \left|\sum_{m<i\leqslant m+k}\left(X(i)-\frac{1}{m}\sum_{j=1}^{m}X(j)\right)\right| / g(m,k,\gamma), \\ \text{where } g(m,k,\gamma) &= m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}\end{aligned} \tag{1.4}$$

for $0 \leqslant \gamma < 1/2$. The statistic is then given by

$$\frac{1}{\widehat{\sigma}_m}\sup_{1\leqslant k\leqslant N(m)}\Gamma(m,k,\gamma),$$

where $N(m)/m \to \infty$, $N(m)/m \to N > 0$, or $N(m) = \infty$, as $m \to \infty$, and $\widehat{\sigma}_m^2 - \sigma^2 = o_P(1)$ is a consistent estimator of σ^2 and only depends on the historical data set. In this paper we use the following variance estimator on the historic data set

$$\widehat{\sigma}_m^2 = \widehat{\sigma}_m^2(X(1),\ldots,X(m)) = \frac{1}{m-1}\sum_{i=1}^{m}\left(X(i)-\frac{1}{m}\sum_{j=1}^{m}X(j)\right)^2. \tag{1.5}$$

We reject the null hypothesis at

$$\tau(m) = \begin{cases}\inf\{k \geqslant 1 : \frac{1}{\widehat{\sigma}_m}\Gamma(m,k,\gamma) \geqslant c\}, \\ \infty, \quad \text{if } \frac{1}{\widehat{\sigma}_m}\Gamma(m,k,\gamma) < c,\ k = 1,2,\ldots,N(m),\end{cases}$$

where c is chosen in such a way that we control the false alarm rate, i.e. that under the null hypothesis

$$\lim_{m\to\infty} P(\tau(m) < \infty) = \alpha$$

for some given level $0 < \alpha < 1$. Under the alternative the limit should converge to 1.

The paper is organized as follows. In the next chapter we give the asymptotics based on which the critical value has been chosen by Horváth et al. [6] and propose some alternative methods based on bootstrapping. Theoretical results proving the asymptotic consistency of the proposed procedures under the null hypothesis as well as alternatives are also given. In Section 3 a simulation study compares the different bootstrap methods with the asymptotic method. Finally the proofs can be found in Section 4.

2 Bootstrapping procedures and consistency of the corresponding tests

Before we discuss some possible bootstrap procedures we first give the asymptotic distribution for an infinite horizon (Theorem 2.1 a) below) and an improved version for a

finite time horizon (Theorem 2.1 b) below). This second approach can also be viewed as a pseudo-parametric bootstrap. That means we assume normality and estimate the variance from the historical data set. This last estimate is a problem for small historical sample sizes. When we talk about parametric bootstrap in the following, this is the procedure we mean.

Horváth et al. [6] propose to calculate the critical values according to the asymptotics given in a) of the following theorem with $N(m) = \infty$.

Theorem 2.1. *Let* (1.1) – (1.3) *hold,* $0 \leqslant \gamma < 1/2$, *and* $\widehat{\sigma}_m - \sigma = o_P(1)$. *Then we have, as* $m \to \infty$, *under the null hypothesis,*

a) if $N(m)/m \to \infty$ *or* $N(m) = \infty$,

$$\frac{1}{\widehat{\sigma}_m} \sup_{1 \leqslant k \leqslant N(m)} \Gamma(m,k,\gamma)(X(1),\ldots,X(m+k)) \xrightarrow{\mathcal{D}} \sup_{0<t<1} \frac{|W(t)|}{t^\gamma},$$

where $\{W(t) : 0 \leqslant t \leqslant 1\}$ *is a standard Wiener process,*

b) if $N(m)/m \to \infty$ *or* $N(m)/m \to N$ *for a constant* $N > 0$,

$$\begin{aligned}&\frac{1}{\widehat{\sigma}_m} \sup_{1 \leqslant k \leqslant N(m)} \Gamma(m,k,\gamma)(X(1),\ldots,X(m+k)) \\ &= \max_{1 \leqslant k \leqslant N(m)} \left| W_1\left(\frac{k}{m}\right) - \frac{k}{m} W_2(1) \right| \Big/ \left(\left(1 + \frac{k}{m}\right) \left(\frac{k}{k+m}\right)^\gamma \right) + o_P(1),\end{aligned}$$

where $\{W_1(t) : t \geqslant 0\}$ *is a standard Wiener process,* $W_2(1)$ *is standard normally distributed and independent of* $\{W_1(\cdot)\}$.

Remark 2.1. a) The limit distribution in a) is explicitly only known for $\gamma = 0$.

b) The distribution of the right hand side of b) is the actual distribution of $\Gamma(m,k,\gamma)$ under the null hypothesis for normal errors with variance 1. Note that it converges to the same limit as in a) for $N(m)/m \to \infty$ and it converges to

$$\sup_{0<t\leqslant N} \frac{|W_1(t) - tW_2(1)|}{(1+t)(t/(1+t))^\gamma}$$

for $N(m)/m \to N$ (cf. the proof of Theorem 2.1 in Horváth et al. [6]).

Remark 2.2. Horváth et al. [6] pointed out that a value of γ close to $1/2$ has the shortest detection delay time for early changes, however the probability of a false alarm (before the change occurred) is higher. If the change occurs well after the monitoring started, the detection delay time is similar for all values of γ, but a γ close to $1/2$ has a significant probability of a false alarm well before the change occurred.

The next theorem shows the asymptotic consistency of the tests under alternatives, where we have chosen the critical values according to Theorem 2.1 a) or b) above.

Theorem 2.2. *Let* (1.1) – (1.3) *hold,* $0 \leqslant \gamma < 1/2$, *and* $\sqrt{m}\, d(m) \to \infty$. *Note that the last assumption includes fixed as well as local changes. Furthermore let either*

a) $k^*/m = O(1)$, *if* $N(m)/m \to \infty$ *or* $N(m) = \infty$, ***or***

b) $k^*/m \to \tilde{N}$ *with constants* $0 \leqslant \tilde{N} < N$, *if* $N(m)/m \to N$.

Then we have, as $m \to \infty$, $sup_{1 \leqslant k \leqslant N(m)} \Gamma(m,k,\gamma)(X(1), X(2), \ldots) \xrightarrow{P} \infty$.

Next we discuss some non-parametric bootstrap procedures, which we will compare in the simulation study in Section 3. It turns out that bootstrap version B.3c is best in most situations.

For the bootstrap we need that $N(m)$ is finite. However, the bootstrap procedures are also asymptotically correct for an infinite time horizon if $N(m)/m \to \infty$. First we discuss three possible bootstraps that take different observations into account. Different possibilities to choose the variance estimator $\widehat{\sigma}^2_{m,k}(V)$ in the bootstrap descriptions below will be discussed later.

Bootstrap B.1
The first bootstrap uses only the historical data set. Consider the bootstrap statistic

$$\Gamma^*(m,l,\gamma)^{(1)}_{1,m} = \Gamma(m,l,\gamma)(X(U_{1,m}(1)), \ldots, X(U_{1,m}(l+m)),$$

where $\{U_{1,m}(i) : 1 \leqslant i \leqslant m + N(m)\}$ are i.i.d. random variables with $P(U_{1,m}(1) = j) = 1/m$, $j = 1, \ldots, m$, independent of $\{X(i) : 1 \leqslant i \leqslant N(m)\}$.

We choose the critical value $c_k^{(1V)}(m) = c^{(1V)}(m)$ for a given level α such that

$$P^*_{1,m}\left(\frac{1}{\widehat{\sigma}_{1,m}(V)} \sup_{1 \leqslant l \leqslant N(m)} \Gamma^*(m,l,\gamma)^{(1)}_{1,m} > c^{(1V)}(m)\right) \leqslant \alpha, \tag{2.1}$$

$c^{(1V)}(m)$ minimal, $V = a, b, c$ denotes one of the variance versions below. $P^*_{1,m}$ denotes the conditional probability given $X(1), \ldots, X(m)$. In applications one does not calculate the above conditional probability but instead draws t, say $t = 1\,000$, random bootstrap samples and uses the empirical distribution.

This bootstrap has essentially two advantages. First we only have to calculate the critical values once at the beginning of the observation period. Theoretical results are also easy to obtain because we do not have to deal with new critical values for each incoming observation. Another advantage is that we already know that the observations follow the null hypothesis. The disadvantage is that we have only few observations to base on a bootstrap for a much longer time series.

Bootstrap B.2
The second bootstrap uses all data available up to this point. This means that the critical value now depends on the time point. At time $m + k$ consider the bootstrap statistic

$$\Gamma^*(m,l,\gamma)^{(2)}_{k,m} = \Gamma(m,l,\gamma)(X(U_{k,m}(1)), \ldots, X(U_{k,m}(l+m))),$$

where $\{U_{k,m}(i) : 1 \leqslant i \leqslant m + N(m)\}$ are i.i.d. random variables with $P(U_{k,m}(1) = j) = 1/(m+k-1)$, $j = 1, \ldots, m+k-1$, independent of $\{X(i) : 1 \leqslant i \leqslant m + N(m)\}$.

We reject the null hypothesis at a given level α at time $m+k$ if $\Gamma(m,k,\gamma)/\widehat{\sigma}_m > c_k^{(2V)}(m)$, where

$$P_{k,m}^*\left(\frac{1}{\widehat{\sigma}_{k,m}(V)}\sup_{1\leqslant l\leqslant N(m)}\Gamma^*(m,l,\gamma)_{k,m}^{(2)} > c_k^{(2V)}(m)\right)\leqslant\alpha, \tag{2.2}$$

$c_k^{(2V)}(m)$ minimal, $V=a,b,c$ denotes one of the variance versions below. $P_{k,m}^*$ denotes the conditional probability given $X(1),\ldots,X(m+k-1)$. In applications one uses the empirical distribution function to calculate the critical value.

This bootstrap is computationally very expensive, since we calculate a new critical value after each incoming observation. Moreover a small portion of the observations may already follow the alternative. It has the advantage, that we use as much data as possible for the bootstrap.

Bootstrap B.3

The third bootstrap is essentially a combination of the above two. The idea is that only the older bootstrap samples do not represent the current data well enough whereas the newer ones are still reasonably good. This bootstrap procedure was suggested by Steland [12].

In Bootstrap B.2 we calculate critical values at time $m+k$ based on the distribution $P_{k,m}^*$. For Bootstrap B.3 we use two modifications to reduce the computation time significantly.

First of all we only calculate new critical values after each Lth observation.

Secondly and maybe even more importantly we use a convex combination of the latest M bootstrap distributions. Thus, in applications we use an empirical distribution function not only based on the newest samples (which are generated from $X(1),\ldots,X(m+k-1)$) as in B.2 but also on older samples. This is why we need to generate only a fraction of the bootstrap samples needed for Bootstrap B.2 in each step. For example for $\beta_i = 1/M$ below, i.e. equal weights in the convex combination, we only need $t_1 = t/M$ new samples each time we update the critical values to get an empirical distribution function based on t samples. Therefore the procedure is significantly accelerated even if we calculate new critical values after each new observation ($L=1$).

For a theoretical description denote by

$$\begin{aligned}&F_{k,m}^{*(V)}(x)\\&= P_{k,m}^*\left(\frac{1}{\widehat{\sigma}_{k,m}(V)}\sup_{1\leqslant l\leqslant N(m)}\Gamma^*(m,l,\gamma)_{k,m}^{(2)}\leqslant x\right)\end{aligned}$$

the conditional distribution functions from B.2. Let for $j\geqslant 1$, $\sum_{i=0}^{M-1}\beta_i = 1$ and $\beta_i\geqslant 0$

$$\widetilde{F}_{k,m}^{(V)} = \sum_{i=0}^{M-1}\beta_i F^{*(V)}_{\max((j-i)\,L,m),m}, \qquad \text{for } k = j\,L,\ldots,(j+1)L-1.$$

Then we calculate the critical value $c_k^{(3V)}(m)$ at time $k+m$ as follows:

$$\widetilde{F}_{k,m}^{(V)}\left(c_k^{(3V)}(m)\right)\geqslant 1-\alpha, \tag{2.3}$$

$c_k^{(3V)}(m)$ minimal, $V = a, b, c$ denotes one of the variance versions below.

In simulations we use the convex combinations with equal weights $\beta_i = \frac{1}{M}$, $L = m/5$ and $M = 5$. Thus after monitoring for m observations we have completely replaced the bootstrap samples. With this parameters this bootstrap is computationally only little more expensive than Bootstrap B.1 based on the historical data sequence, since we only replace all t bootstrap samples after m observations. The simulations show that it yields almost the same results as Bootstrap B.2 but is indeed a lot faster.

Furthermore we give several possibilities on how to deal with the variance in the bootstrap.

Variance a
In this version we normalize each bootstrap which is based on the first $m + k - 1$ observations with the conditional variance $\widehat{\sigma}^2_{k,m}(a) = \frac{1}{m+k-1} \sum_{i=1}^{m+k-1} (X(i) - \bar{X}_{m+k-1})^2$. This is the bootstrap that Lemma 4.2 yields initially.

Variance b
In this version we use no variance estimator neither for the bootstrap nor the original statistic. This is equivalent to using the same for both. That means we choose $\widehat{\sigma}_{k,m}(b) = \widehat{\sigma}_m$.

Variance c
This version uses pivoting with respect to the historical data set, i.e. we also bootstrap the variance estimator $\widehat{\sigma}_m$, hence

$$\widehat{\sigma}_{k,m}(c) = \widehat{\sigma}_m(X(U_{k,m}(1)), \ldots, X(U_{k,m}(m))).$$

Hall and Wilson [5] suggest that in most situations for a posteriori tests pivoted bootstrap procedures, i.e. studentized test statistics, hold the level best. In fact the simulations show that this is also true in this situation.

The next theorem shows the asymptotic consistency of the test based on the above bootstrap critical values under the null hypothesis as well as under alternatives.

Theorem 2.3. *Let* (1.1) – (1.3) *hold,* $0 \leqslant \gamma < 1/2$, $\widehat{\sigma}_m - \sigma = o_P(1)$. *Then we have as* $m \to \infty$

a) under the null hypothesis,

$$P\left(\frac{1}{\widehat{\sigma}_m} \sup_{1 \leqslant k \leqslant N(m)} \frac{\Gamma(m, k, \gamma)}{c_k^{(jV)}(m)} > 1\right) \to \alpha,$$

where $j = 1, 2, 3$, $V = a, b, c$.

b) If additionally the assumptions of Theorem 2.2 are fulfilled and $d = d(m) = O(1)$, *then*

$$P\left(\frac{1}{\widehat{\sigma}_m} \sup_{1 \leqslant k \leqslant N(m)} \frac{\Gamma(m, k, \gamma)}{c_k^{(jV)}(m)} > 1\right) \to 1.$$

where $j = 1, 2, 3$, $V = a, b, c$.

Remark 2.3. The assertions of Theorem 2.3 remain true, if we use an infinite observation horizon and critical values based on a bootstrap with horizon $\widetilde{N}(m)$ fulfilling $\widetilde{N}(m)/m \to \infty$. This is important because a computer can obviously not calculate critical values based on an infinite monitoring horizon.

3 Simulation study

In the previous chapter we have established the asymptotic validity of the bootstrap tests. The question remains how well the procedures work for small sample sizes.

We will establish the answer to this question in the following simulation study. First we compare the different bootstrap procedures, then we compare the best bootstrap procedure with the asymptotic test as well as parametric bootstrap (of Theorem 2.1).

The goodness of sequential tests can essentially be determined by three criteria:

$\mathcal{C}.1$ The actual level (α-error) of the test should be close to the nominal level.

$\mathcal{C}.2$ The power of the test should be large, preferably close to 1, i.e. the β-error should be small.

$\mathcal{C}.3$ The run length of the test, i.e. the time after monitoring starts until the null hypothesis (if at all) is rejected, should be shortly after the change-point.

We visualize these qualities by the following plots:

Size-Power Curves

Size-power curves are plots of the empirical distribution function of the p-values of the test under the null hypothesis as well as under various alternatives. In the sequential setting with varying critical values it is the minimum of the p-values for each step, i.e. the minimum of $p_k, 1 \leqslant k \leqslant N(m) < \infty$, where p_k is the p-value of $\Gamma(m,k,\gamma)/\widehat{\sigma}_m$ with respect to the distribution from which we obtain the critical value $c_k(m)$.
What we get is a plot that shows the empirical size and power (i.e. the empirical α–errors resp. $1-(\beta$–errors$)$) on the y-axis for the chosen level on the x-axis. So, the graph for the null hypothesis should be close to the diagonal (which is given by the dotted line) and for the alternatives it should be as steep as possible.

Plot of the Average Run Length

We store the run lengths for nominal test levels $0.01, 0.02, \ldots, 0.1$ based on t runs and plot the mean of those runs where the test did reject, linearly interpolated. Note that runs in which the test did not reject the null hypothesis are not used in the calculation. This can lead to effects where the average run length is longer for a higher level due to the fact that more runs are rejected. This effect is especially visible if the monitoring period ends shortly after the change. The horizontal line in the plot indicates where the change occurred.

It is sometimes argued that the median of the run length is a better criteria than the average run length. But for the purpose of comparing the different procedures here, the results are qualitatively very similar. Thus we concentrate on the average run length which is more frequently used.

Plot of the estimated density of the run length

For the density estimation we use the standard R procedure which uses a Gaussian kernel, where the bandwidth is chosen according to Silvermans rule of thumb ([11], p. 48 eq. (3.31)). The estimation is based on only those simulations where the null hypothesis was rejected at the 5% level. The vertical line in the plot indicates where the change occured.

For the simulation study we used the following parameters:

- $m = 10, 20, 50$
- $N(m) = Nm$ with $N = 1, 2, 5, 10$
- $k^* = \vartheta m$ with $\vartheta = 0.25, 0.5, 2, 5$
- $d = 0.5, 1, 2$
- standard normally distributed errors and exponentially distributed errors
- $\gamma = 0, 0.25, 0.49$

Due to limitations of space and similarity of results we will only present a small selection of plots here, the complete simulation results can be obtained from the author (pdf-File, 215 p., 25 MB).

3.1 Variance estimators for the bootstrap

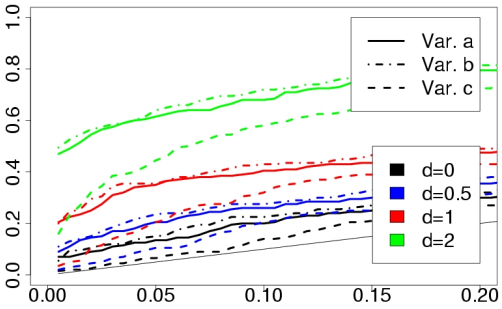
In this section we compare the different variance estimators a-c for Bootstraps B.1 and B.2. Bootstrap B.3 is an intermediary of the two as will be seen in the next section. Because Bootstrap B.2 is computationally quite expensive we will calculate the bootstrap quantiles based on 500 random bootstrap samples and calculate the empirical size and power respectively average run length based on 200 runs. All plots are based on the same underlying error sequences.

Some results for exponential errors, $\gamma = 0$, and $k^* = 1/2m$ are given in Figure 3.1. The other results are very similar in view of the behavior concerning the variance estimator and therefore omitted.

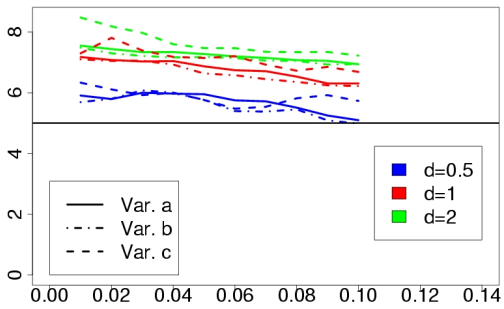
Most importantly the plots show clearly that variance estimator $\widehat{\sigma}_{k,m}(c)$ holds the level best. This is in accordance with a posteriori tests where the studentized version usually behaves best in this regard (cf. e.g. Hall and Wilson [5]). The empirical power of the different versions is comparable if one takes the actual level into account. The same holds true for the average run length. The average run length is in some cases shorter for variance versions a and b, but this can be explained by the higher false alarm rate.

What also might seem odd at first is the fact that the average run length for small changes ($d = 0.5$) is smaller than for more obvious changes ($d = 2$) for a short monitoring period. This is explained by the fact that we reject fewer runs for less obvious changes. For longer monitoring periods (after the change) this phenomenon disappears because we reject most runs.

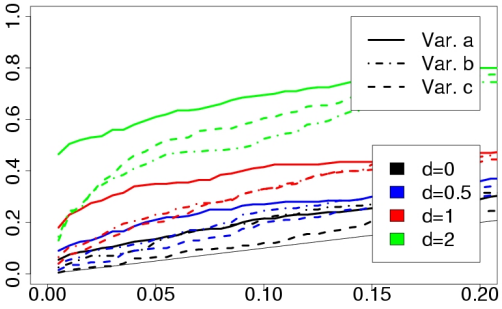
The results for Bootstraps B.1 and B.2 are similar. For $m = 50$ all procedures behave approximately the same.



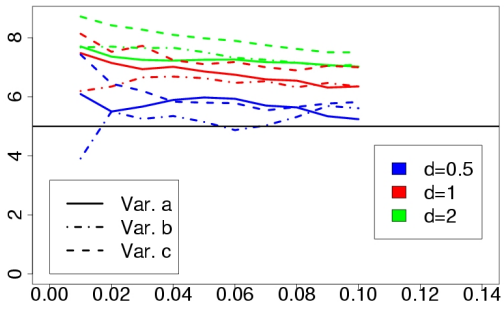
(1) SPC: B.1, $m = 10$, $N = 1$



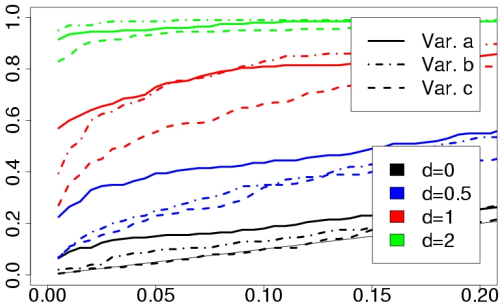
(2) ARL: B.1, $m = 10$, $N = 1$



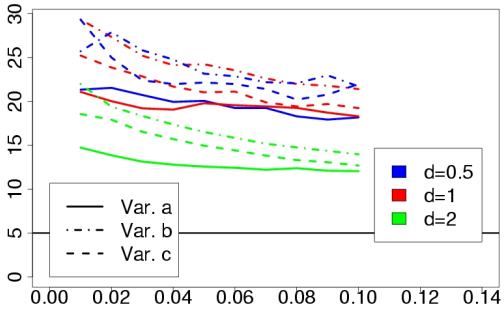
(3) SPC: B.2, $m = 10$, $N = 1$



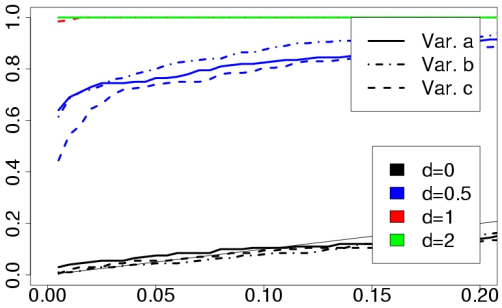
(4) ARL: B.2, $m = 10$, $N = 1$



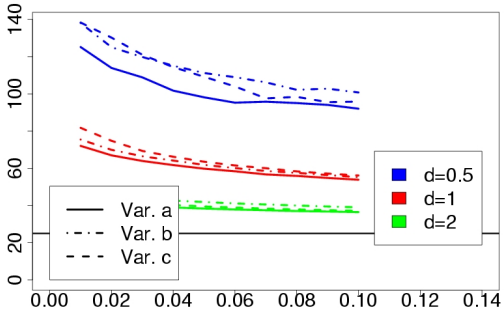
(5) SPC: B.2, $m = 10$, $N = 5$



(6) ARL: B.2, $m = 10$, $N = 5$



(7) SPC: B.2, $m = 50$, $N = 5$



(8) ARL: B.2, $m = 50$, $N = 5$

Figure 3.1: Size-power curves and ARL: exponential errors, $k^* = 1/2m$, $\gamma = 0$

3.2 Different bootstrap procedures

In this section we compare Bootstraps B.1c, B.2c and B.3c. Because Bootstrap B.2 is computationally quite expensive we will calculate the bootstrap quantiles based on 500 random bootstrap samples and calculate the empirical size and power respectively average run length based on 200 runs. All plots are based on the same underlying error sequences.

Concerning Bootstrap B.3 we use $L = m/5$ and $M = 5$.

Some selected plots can be found in Figure 3.2. The plots for other parameters are similar with respect to the behavior concerning different bootstrap procedures and are therefore omitted.

First of all the plots show that the level is approximately equal for all three procedures and close to the nominal one. However, the power is much better for Bootstrap B.2 and B.3 than for B.1, this is especially obvious for exponential errors. In fact the size-power curves for Bootstrap B.3 are surprisingly close to the ones for B.2, even though we only completely replace the samples after we have observed as many new observations as the historic length m was. Thus the procedure is computationally sufficiently cheap while it still has the advantages of Bootstrap B.2.

Concerning the average run length no clear tendency can be found.

As already with the variances the procedures become essentially equivalent for $m = 50$.

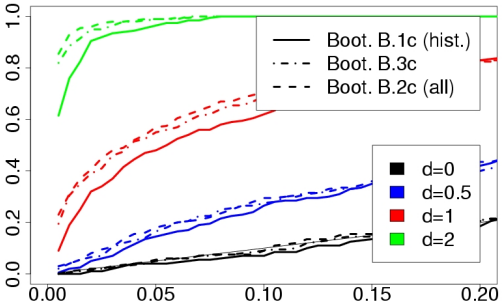
3.3 Comparison of bootstrap and asymptotic methods

In this section we finally compare Bootstrap B.3c ($L = m/5,\ M = 5$), which has turned out to have the best overall behavior, with the parametric bootstrap and the asymptotic test from Theorem 2.1. Because Bootstrap B.3 is not that computationally expensive we will now use $t = 1\,000$ random bootstrap samples to calculate the critical values and base the plots on $1\,000$ runs. Again we use the same underlying error sequence for different parameters.

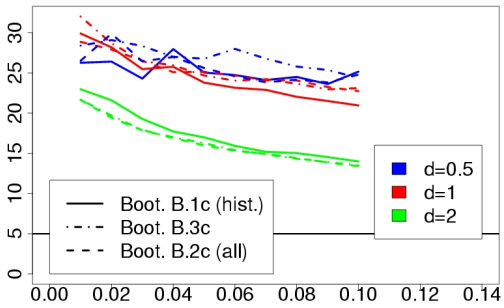
In Figures 3.3(i)-3.3(ii) some selected size-power curves and plots of the average run length can be found.

We would like to point out that for normal errors the parametric bootstrap gives the actual distribution for the statistic if the variance estimator is correct. Surprisingly it still does not hold the level for normal errors but in fact has a higher level than the nominal one for small historical data sets. For exponential errors it is even worse because the variance estimator for small samples is skewed with a median much smaller than the mean (which is the variance we are interested in).

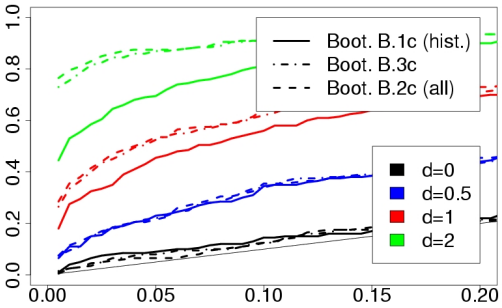
Not surprisingly the level of the asymptotic test is much too small for small observation sequences. For exponential errors the level sometimes is fine, but this is only due to the fact that the variance estimator is too small. Heuristically the remainder of the monitoring period can be neglected if the asymptotic test and the parametric bootstrap become similar, this happens approximately around $N = 10$.



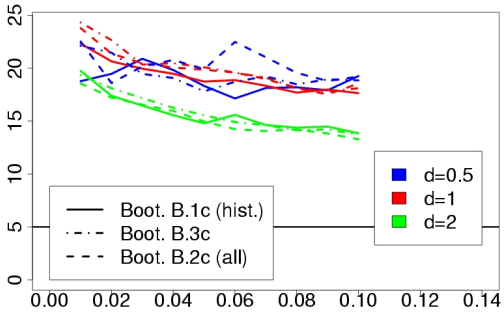
(1) SPC: normal errors, $m = 10$, $N = 5$



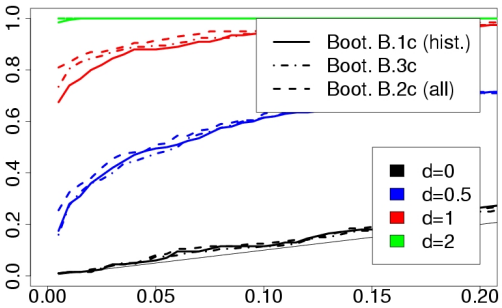
(2) ARL: normal errors, $m = 10$, $N = 5$



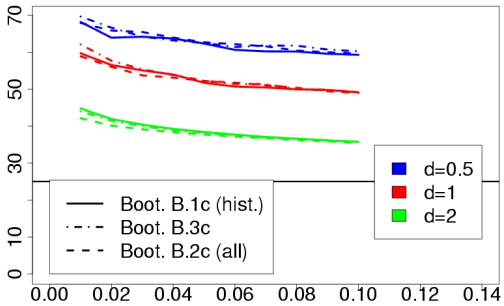
(3) SPC: exp. errors, $m = 10$, $N = 5$



(4) ARL: exp. errors, $m = 10$, $N = 5$



(5) SPC: exp. errors, $m = 50$, $N = 2$



(6) ARL: exp. errors, $m = 50$, $N = 2$

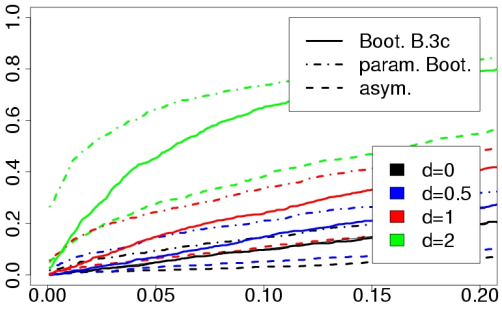
Figure 3.2: Size-power curves and ARL: $k^* = 1/2m$, $\gamma = 0$

The power of all tests becomes better the longer the monitoring period is (after the change).

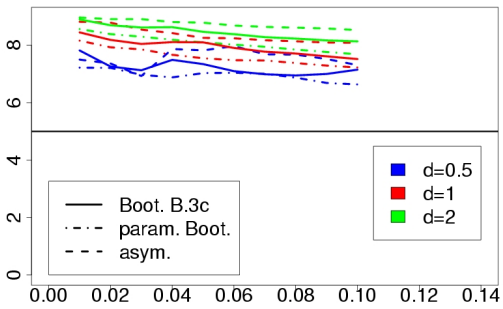
The bootstrap test is the only test that holds the level well in all cases. The power and average run length are comparable (maybe slightly better for the asymptotic methods) for all three tests if one takes the actual level of the test into account.

Again the parametric bootstrap and Bootstrap B.3c are approximately equivalent for $m = 50$, the asymptotic bootstrap then is approximately equivalent for $N \geqslant 10$.

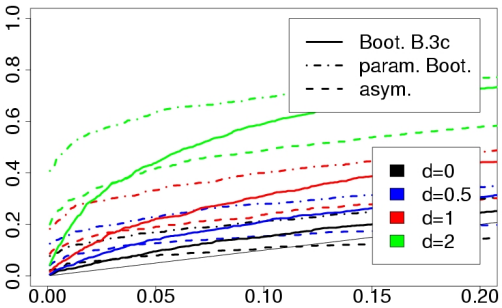
Some plots of the densities of the run length can be found in Figure 3.4.
The densities are similar if one takes the different levels into account. We would like to point out that the false alarm rate before a change occurred, which is not negligible for



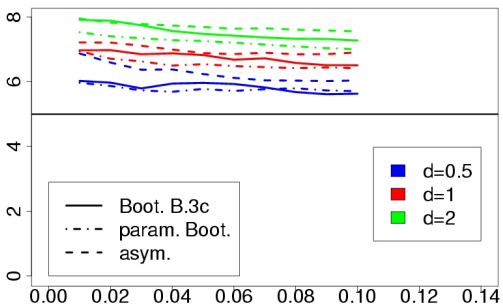
(1) SPC: normal errors, $N = 1$



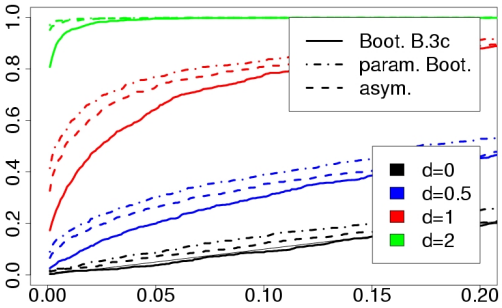
(2) ARL: normal errors, $N = 1$



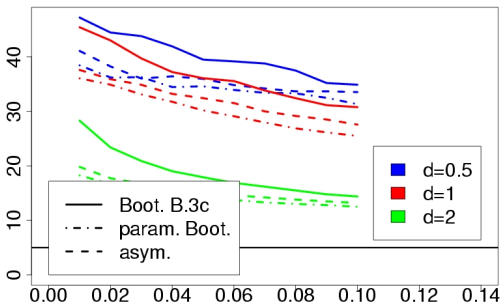
(3) SPC: exp. errors, $N = 1$



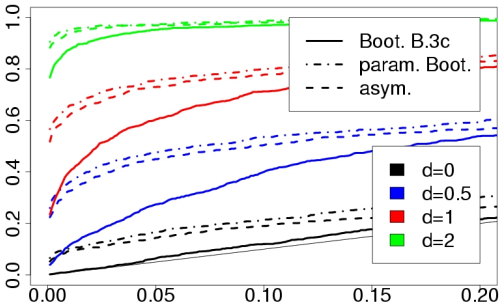
(4) ARL: exp. errors, $N = 1$



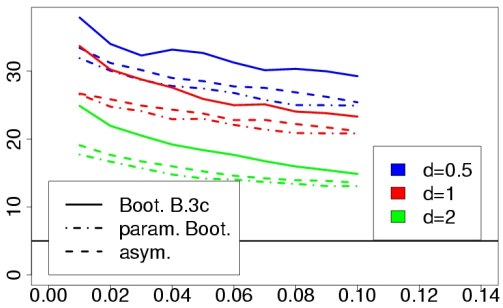
(5) SPC: normal errors, $N = 10$



(6) ARL: normal errors, $N = 10$

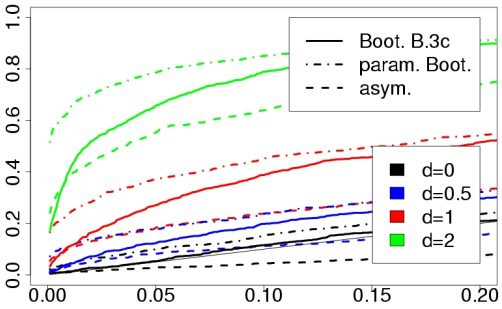


(7) SPC: exp. errors, $N = 10$

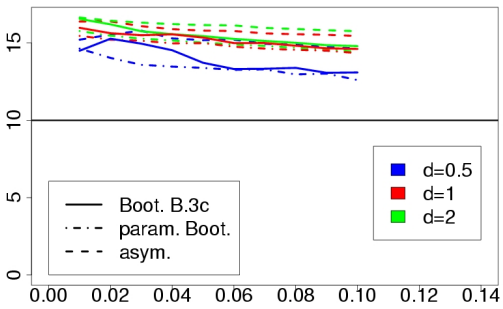


(8) ARL: exp. errors, $N = 10$

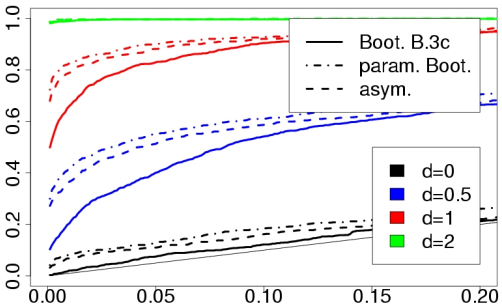
Figure 3.3(i): Size-power curves and ARL: $m = 10$, $k^* = 1/2m$, $\gamma = 0$



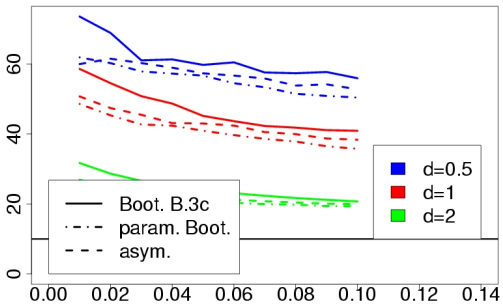
(9) SPC: $m = 20$, $N = 1$



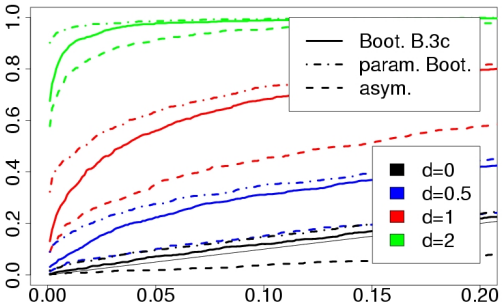
(10) ARL: $m = 20$, $N = 1$



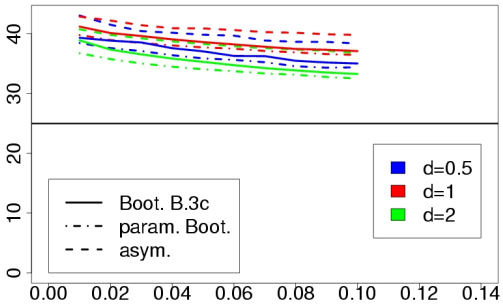
(11) SPC: $m = 20$, $N = 10$



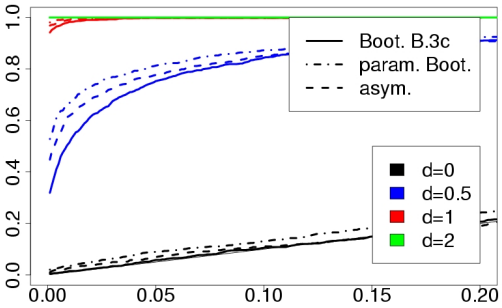
(12) ARL: $m = 20$, $N = 10$



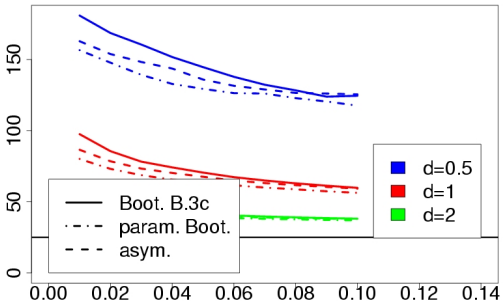
(13) SPC: $m = 50$, $N = 1$



(14) ARL: $m = 50$, $N = 1$

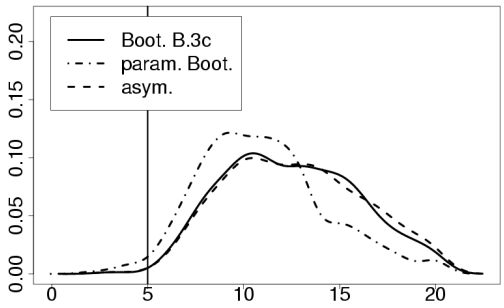


(15) SPC: $m = 50$, $N = 10$

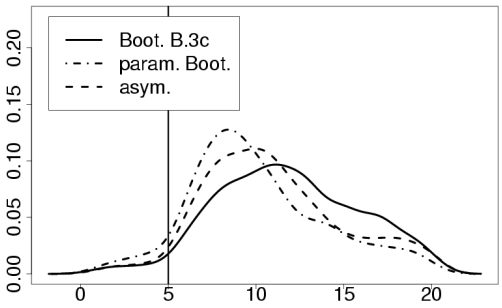


(16) ARL: $m = 50$, $N = 10$

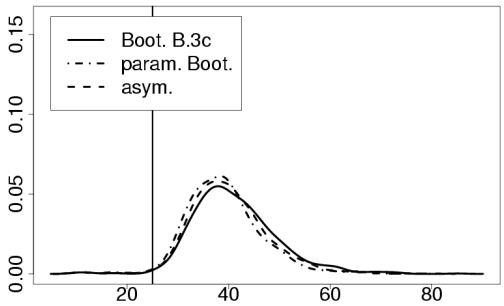
Figure 3.3(ii): Size-power curves and ARL: $k^* = 1/2m$, $\gamma = 0$ and exponential errors



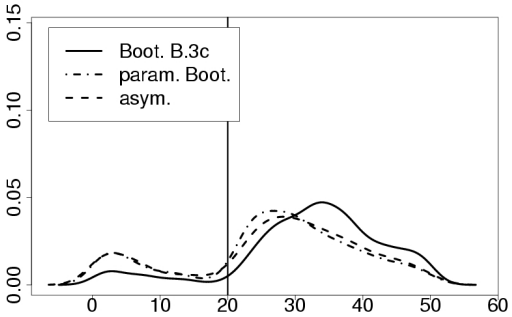
(1) normal errors, $m = 10$, $N = 2$, $k^* = 1/2m$, $\gamma = 0$



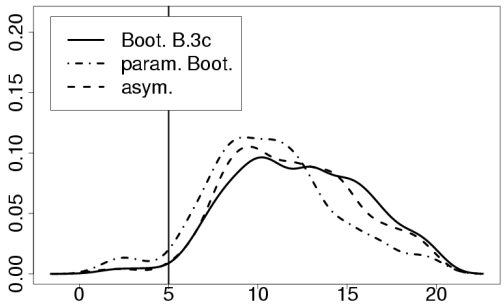
(2) exp. errors, $m = 10$, $N = 2$, $k^* = 1/2m$, $\gamma = 0$



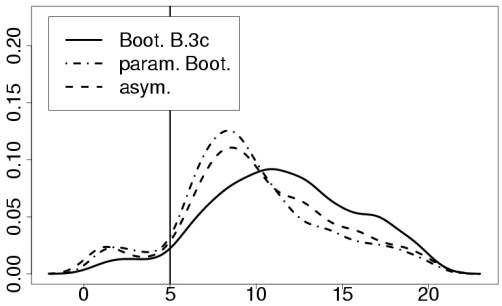
(3) exp. errors, $m = 50$, $N = 10$, $k^* = 1/2m$, $\gamma = 0$



(4) exp. errors, $m = 10$, $N = 5$, $k^* = 2m$, $\gamma = 0.25$



(5) normal errors, $m = 10$, $N = 2$, $k^* = 1/2m$, $\gamma = 0.25$



(6) exp. errors, $m = 10$, $N = 2$, $k^* = 1/2m$, $\gamma = 0.25$

Figure 3.4: Density estimates of run length, $d = 2$

$\gamma = 0.25$ or even $\gamma = 0.49$, is smaller for the bootstrap. The difference is most visible for exponential errors.

3.4 Conclusions

We have seen that the studentized versions of the bootstraps (variance version c) hold the level best. Moreover the bootstraps taking more observations into account have a better power. Thus taking the computation time into account, the bootstrap with the best overall performance is Bootstrap B.3c.

For a historic data sequence larger than $m = 50$ all procedures (except the asymptotic test) become more or less equivalent, thus to save computation time one should either use the historic Bootstrap B.1c or even the parametric bootstrap. The asymptotic test should only be applied if one indeed plans to monitor for quite some time (approximately $N(m) \geqslant 10m$).

In all other cases we recommend to use Bootstrap B.3c with the above parameters ($L = m/5,\ M = 5$) because it has the best overall performance and is computationally reasonably cheap.

4 Proofs

Proof of Theorem 2.1. It follows immediately from the proof of Theorem 2.1 in Horváth et al. [6]. ■

The following proof is similar to the proof of Theorem 2.2 in Horváth et al. [6].

Proof of Theorem 2.2. First note that for $\tilde{k} > k^*$ it holds by Theorem 2.1 and Remark 2.1

$$\begin{aligned}\Gamma(m,\tilde{k},\gamma) &= \left(\sum_{i=m+1}^{m+\tilde{k}} \epsilon(i) - \frac{\tilde{k}}{m}\sum_{i=1}^{m}\epsilon(i)\right)/g(m,\tilde{k},\gamma) + d(m)(\tilde{k}-k^*)/g(m,\tilde{k},\gamma)\\ &= O_P(1) + d(m)(\tilde{k}-k^*)/g(m,\tilde{k},\gamma).\end{aligned}$$

Under a) let $\tilde{k} = k^* + m = O(m)$ ($\leqslant N(m)$ for m large enough), and under b) $\tilde{k} = N(m) = O(m)$. Then, $g(m,\tilde{k},\gamma)/\sqrt{m} = O(1)$, and thus

$$d(m)(\tilde{k}-k^*)/g(m,k,\gamma) = \sqrt{m}\,d(m)\left(\frac{\tilde{k}}{m} - \frac{k^*}{m}\right)\frac{\sqrt{m}}{g(m,\tilde{k},\gamma)} \to \infty,$$

which yields the assertion. ■

To prove Theorem 2.3 we first need three auxiliary lemmas. When dealing with bootstrap procedures of a posteriori tests (or for Bootstrap B.1) it usually suffices to obtain conditionally the same asymptotical distribution of the bootstrap statistic as for the original statistic in a P-stochastic sense. This yields the stochastic convergence of the bootstrap critical values to the asymptotic ones. Here, however, we have a triangular scheme of critical values and thus need the convergence in a uniform way. This is established by the next lemma under the condition that we also have the distributional convergence uniformly.

Lemma 4.1. *Let c, $c_k(m)$ be such that $P(Y > c) = \alpha$ respectively $P^*(Y_k(m) > c_k(m)) \leqslant \alpha$ for some $0 < \alpha < 1$ ($c_k(m)$ minimal), where $Y_k(m)$ is some statistic and Y is a random variable with strictly monotone and continuous distribution function in a compact neighborhood K of c. Moreover let for all x in K (as $m \to \infty$)*

$$\sup_{1\leqslant k<\infty} \left|P^*_{k,m}(Y_k(m) \leqslant x) - P\left(Y \leqslant x\right)\right| \xrightarrow{P} 0. \tag{4.1}$$

Then,

$$\sup_{1\leqslant k<\infty} |c_k(m) - c| \xrightarrow{P} 0 \qquad \text{as } m \to \infty. \tag{4.2}$$

Proof. Using standard arguments one can conclude that

$$\sup_{x\in K}\sup_{1\leqslant k<\infty}\left|P^*_{k,m}(Y_k(m)\leqslant x)-P\left(Y\leqslant x\right)\right|\xrightarrow{P}0. \tag{4.3}$$

We show that assertion (4.2) holds true almost surely if we have (4.3) almost surely, the P-stochastic result then follows via the subsequence principle. Thus there exists a set N^C with $P(N^C)=1$ such that (4.3) holds for any $\omega\in N^C$. We will now prove that (4.2) holds for any $\omega\in N^C$ by contradiction. Therefore we assume the existence of a subsequence $\alpha(\cdot)$, such that $\sup_k|c_k(\alpha(m))-c|>\epsilon$ for some $\epsilon>0$ with $c\pm\epsilon\in K$. Then, there is a function f with $|c_{f(\alpha(m))}(\alpha(m))-c|>\epsilon$, hence another subsequence $\beta(m)$ with

$$c_{f(\beta(m))}(\beta(m))\leqslant c-\epsilon\qquad\text{or}\qquad c_{f(\beta(m))}(\beta(m))>c+\epsilon.$$

In the first case we have

$$\begin{aligned}
&0<1-\alpha-P(Y\leqslant c-\epsilon)\leqslant P^*_{f(\beta(m)),\beta(m)}\left(Y_{f(\beta(m))}(\beta(m))\leqslant c-\epsilon\right)-P(Y\leqslant c-\epsilon)\\
&\leqslant\sup_k\left|P^*_{k,\beta(m)}\left(Y_k(\beta(m))\leqslant c-\epsilon\right)-P(Y\leqslant c-\epsilon)\right|\\
&\leqslant\sup_{x\in K}\sup_k\left|P^*_{k,\beta(m)}\left(Y_k(\beta(m))\leqslant x\right)-P(Y\leqslant x)\right|\to 0.
\end{aligned}$$

This is a contradiction. The other case is analogous since $P^*_{k,m}(Y_k(m)>c+\epsilon)>\alpha$ for $c_k(m)>c+\epsilon$ due to the minimality of $c_k(m)$. ■

To obtain (4.1) we need the next lemma.

Lemma 4.2. *Let $a_{k,m}(i), 1\leqslant i\leqslant m+k-1$, be scores satisfying $\frac{1}{m+k-1}\sum_{i=1}^{m+k-1}a_{k,m}(i)=0$ and*

$$\frac{1}{m+k-1}\sum_{i=1}^{m+k-1}(a_{k,m}(i))^2=1 \tag{4.4}$$

$$\frac{1}{m+k-1}\sum_{i=1}^{m+k-1}|a_{k,m}(i)|^\nu\leqslant D<\infty, \tag{4.5}$$

for $2<\nu\leqslant 4$ and a constant $D>0$. Then, on a rich enough probability space, there exists a sequence of stochastic processes

$$\{\xi_{k,m}(l):l\geqslant 1\}\stackrel{\mathcal{D}}{=}\left\{\sum_{i=1}^{l}a_{k,m}(U_{k,m}(i)):l\geqslant 1\right\},$$

where $U_{k,m}(\cdot)$ is i.i.d. uniformly distributed on $\{1,\ldots,m+k-1\}$, such that for a fixed Wiener process $\{W(t):t\geqslant 0\}$ and all $0<\mu<\nu$ we find for each $\epsilon>0$ a $c=c(\epsilon)$ so that for $m\geqslant N(\epsilon)$ it holds

$$\sup_{k\geqslant 1}P\left(\sup_{1\leqslant l<\infty}\frac{1}{l^{1/\mu}}\left|\xi_{k,m}(l)-W(l)\right|\geqslant c\right)\leqslant\epsilon.$$

Proof. The proof is very close to the proof of Theorem 1 of Einmahl and Mason [2] but much simpler. By the Skorohod embedding (cf. e.g. Hall and Heyde [4], Theorem

A.1) we find a Wiener process and stopping times $\{\tau_{k,m}(i) : i \geqslant 1\}$ i.i.d. with mean 1 and $\mathrm{E}\,|\tau_{k,m}(1)|^{\nu/2} \leqslant CD$, where C is a constant only depending on ν, such that for $\xi_{k,m}(l) := W\left(\sum_{i=1}^{l} \tau_{k,m}(i)\right)$ it holds

$$\{\xi_{k,m}(l) : l \geqslant 1\} \stackrel{\mathcal{D}}{=} \left\{\sum_{i=1}^{l} a_{k,m}(U_{k,m}(i)) : l \geqslant 1\right\}.$$

The Wiener process can be chosen independently of k and m, because we can use the same Wiener process to construct the stopping times. Lemma D.1 in Kirch [8] (a generalization of the Hájek-Rényi inequality) remains true for an infinite maximum and an infinite sum (continuity arguments), the square-integrability of the martingale sequence is also not needed because the von Bahr-Esseen inequality remains true for martingales (cf. von Bahr and Esseen [13]). Thus we get for $2 < \mu' < \nu$

$$P\left(\sup_{1\leqslant l<\infty} \frac{1}{l^{2/\mu'}} \left|\sum_{i=1}^{l} \tau_{k,m}(i) - l\right| \geqslant \rho\right) \leqslant \frac{2^{\nu/2-1}CD}{\rho^{\nu/2}} \sum_{l\geqslant 1} \frac{1}{l^{\nu/\mu'}}. \tag{4.6}$$

This yields for some ρ depending only on ϵ

$$\sup_k P(A_{k,m}(\rho)) \leqslant \epsilon, \text{ for } A_{k,m}(\rho) = \left\{\sup_{1\leqslant l<\infty} \frac{1}{l^{2/\mu'}} \left|\sum_{i=1}^{l} \tau_{k,m}(i) - l\right| > \rho\right\},$$

on the other hand we get for $0 < \mu < \mu'$

$$\begin{aligned}
&P\left(\left\{\sup_{1\leqslant l<\infty} \frac{1}{l^{1/\mu}} \left|W\left(\sum_{i=1}^{l} \tau_{k,m}(i)\right) - W(l)\right| \geqslant c\right\} \cap A_{k,m}^c(\rho)\right) \\
&\leqslant 2\sum_{l\geqslant 1} P\left(\sup_{0\leqslant t\leqslant \rho l^{2/\mu'}} W(t) \geqslant c l^{1/\mu}\right) \\
&\leqslant 4\sum_{l\geqslant 1} P\left(W\left(\rho l^{2/\mu'}\right) \geqslant c l^{1/\mu}\right) = 4\sum_{l\geqslant 1} P\left(W(1) \geqslant c l^{1/\mu - 1/\mu'} \rho^{-1/2}\right) \\
&\leqslant 4\sum_{l\geqslant 1} \exp\left(-\frac{1}{2} c^2 l^{\frac{2(\mu'-\mu)}{\mu\mu'}} \rho^{-1}\right) \leqslant \epsilon \text{ for } c \text{ large enough,}
\end{aligned}$$

uniformly in k, which yields the assertion. ■

The next lemma is needed to get the validity of the bootstrap with variance version c.

Lemma 4.3. *Let $a_{k,m}(i), 1 \leqslant i \leqslant m+k-1$, be scores satisfying the conditions of Lemma 4.2. Then we have for all $\epsilon > 0$*

$$\sup_{k\geqslant 1} P\left(\frac{1}{m}\left|\sum_{i=1}^{m} a_{k,m}^2(U_{k,m}(i)) - 1\right| \geqslant \epsilon\right) \to 0,$$

where $U_{k,m}(\cdot)$ is i.i.d. uniformly distributed on $\{1, \ldots, m+k-1\}$.

Proof. The Chebyshev and von Bahr-Esseen inequality (confer e.g. von Bahr and

Esseen [13]) yield ($2 < \nu \leqslant 4$)

$$
\begin{aligned}
&P\left(\frac{1}{m}\left|\sum_{i=1}^{m} a_{k,m}^2(U_{k,m}(i)) - 1\right| \geqslant \epsilon\right) \\
&\leqslant \frac{1}{\epsilon^{\nu/2}} \frac{1}{m^{\nu/2}} \mathrm{E}\left|\sum_{i=1}^{m} (a_{k,m}^2(U_{k,m}(i)) - 1)\right|^{\nu/2} \\
&\leqslant \frac{2^{\nu/2-1}}{\epsilon^{\nu/2}} \frac{1}{m^{\nu/2}} \sum_{i=1}^{m} \mathrm{E}\left|a_{k,m}^2(U_{k,m}(i)) - 1\right|^{\nu/2} = O\left(m^{1-\nu/2}\right) = o(1)
\end{aligned}
$$

uniformly in k. ■

Proof of Theorem 2.3. Let $c > 0$ be the asymptotic critical value to level α (confer Theorem 2.1 and Remark 2.1b)), i.e.

$$
\begin{cases}
P\left(\sup_{0<t<1} \frac{|W(t)|}{t^\gamma} \geqslant c\right) = \alpha, & \text{if } N(m) = \infty \text{ or } N(m)/m \to \infty, \\
P\left(\sup_{0<t\leqslant N} \frac{|W_1(t) - tW_2(1)|}{(1+t)(t/(1+t))^\gamma} \geqslant c\right) = \alpha, & \text{if } N(m)/m \to N,
\end{cases}
$$

where the notation is as in Theorem 2.1. Our goal is to prove that

$$
\sup_k \left|c_k^{(jV)}(m) - c\right| = o_P(1) \tag{4.7}
$$

under the null hypothesis and at least

$$
\sup_k \left|c_k^{(jV)}(m)\right| = O_P(1) \tag{4.8}
$$

under alternatives, where $j = 1, 2, 3,\ V = a, b, c$.

To this end we first apply Lemma 4.2 with

$$
a_{k,m}(i) = \frac{X(i) - \bar{X}_{m+k-1}}{\sqrt{\frac{1}{m+k-1}\sum_{j=1}^{m+k-1}(X(j) - \bar{X}_{m+k-1})^2}},
$$

where $\bar{X}_{m+k-1} = \frac{1}{m+k-1}\sum_{i=1}^{m+k-1} X(i)$. Indeed (4.5) is fulfilled almost surely, since the law of large numbers yields

$$
\begin{aligned}
&\frac{1}{m+k-1} \sum_{i=1}^{m+k-1} (X(i) - \bar{X}_{m+k-1})^2 \\
&= \frac{1}{m+k-1} \sum_{i=1}^{m+k-1} (\epsilon(i) - \bar{\epsilon}_{m+k-1})^2 \\
&\quad - 1_{\{k^*<k-1\}} 2d \frac{1}{m+k-1} \sum_{i=1}^{m+k^*} (\epsilon(i) - \bar{\epsilon}_{m+k-1}) \\
&\quad + 1_{\{k^*<k-1\}} d^2 \frac{(m+k^*)(k-1-k^*)}{(m+k-1)^2} \\
&\geqslant \sigma^2 + o(1) \qquad P-a.s. \qquad \text{uniformly in } k, k^* \geqslant 1
\end{aligned} \tag{4.9}
$$

and similarly

$$\begin{aligned}
&\frac{1}{m+k-1}\sum_{i=1}^{m+k-1}\left|X(i)-\bar{X}_{m+k-1}\right|^{\nu} \\
&= O\left(\frac{1}{m+k-1}\sum_{i=1}^{m+k-1}|\epsilon(i)-\bar{\epsilon}_{m+k-1}|^{\nu}+|d|^{\nu}\right) \\
&= O(1) \qquad P-a.s. \qquad \text{uniformly in } k \geqslant 1.
\end{aligned} \tag{4.10}$$

From Lemma 4.2 we can conclude (4.1) $P-a.s.$ for Bootstrap B.2a where

$$Y_k(m)=\frac{1}{\widehat{\sigma}_{k,m}(a)}\sup_{1\leqslant l\leqslant N(m)}\Gamma(m,l,\gamma)(X(U_{k,m}(1)),\ldots,X(U_{k,m}(l+m)))$$

and Y is the limit distribution of Theorem 2.1 a) resp. as in Remark 2.1.

The proof is analogous to the proof of Theorem 2.1 in Horváth et al. [6]. Note that by replacing the results of Komlós et al. there by Lemma 4.2 we get the convergence rates of Lemma 5.3 in [6] uniformly in k. This gives finally also the distributional convergence uniformly in k.

By Lemma 4.2 we get (4.7) under the null hypothesis as well as under alternatives for $V=a$ and $j=1,2,3$. Note that (4.1) implies an equivalent asertion for $\widetilde{P}_{k,m}(\cdot)$.

In order to obtain (4.7) respectively (4.8) for variance versions b, the law of large numbers yields analogously to (4.9) and (4.10)

$$\sup_k\left|\frac{\widehat{\sigma}^2_{k,m}(a)}{\widehat{\sigma}^2_{k,m}(b)}-1\right|=\begin{cases}o(1) & P-a.s. \quad \text{under } H_0,\\ O(1) & P-a.s. \quad \text{under } H_1.\end{cases}$$

Since $c_k^{(jb)}(m)=\widehat{\sigma}_{k,m}(a)/\widehat{\sigma}_{k,m}(b)c_k^{(ja)}(m)$ we get (4.7) under the null hypothesis and (4.8) under alternatives.

The Chebyshev inequality shows that

$$\sup_{k\geqslant 1}P^*\left(\frac{1}{m}\left|\sum_{i=1}^{m}\frac{X(U_{k,m}(i))-\bar{X}_{m+k-1}}{\left(\frac{1}{m+k-1}\sum_{j=1}^{m+k-1}(X(j)-\bar{X}_{m+k-1})^2\right)^{1/2}}\right|\geqslant\epsilon\right)\to 0 \quad P-a.s.$$

Together with Lemma 4.3 this yields

$$\sup_{k\geqslant 1}P^*\left(\left|\frac{\widehat{\sigma}^2_{k,m}(c)}{\widehat{\sigma}^2_{k,m}(a)}-1\right|\geqslant\epsilon\right)\to 0 \quad P-a.s.$$

Assertion (4.1) remains true for $Y_k(m)/Z_k(m)$ if $\sup_k P^*(|Z_k(m)-1|\geqslant\epsilon)\xrightarrow{P}0$. Thus Lemma 4.2 yields (4.7) under the null hypothesis as well as alternatives.

Theorem 2.1 and (4.7) yield assertion a), Theorem 2.2 and (4.8) assertion b). ■

References

[1] Antoch, J., and Hušková, M. Permutation tests for change point analysis. *Statist. Probab. Lett.*, 53:37–46, 2001.

[2] Einmahl, U., and Mason, D.M. Approximations to permutation and exchangeable processes. *J. Theoret. Probab.*, 5:101–126, 1992.

[3] Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypothesis*. Springer, New York, third edition, 2005.

[4] Hall, P., and Heyde, C.C. *Martingale Limit Theory and its Application*. Academic Press, New York, 1980.

[5] Hall, P. and Wilson, S.R. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47:757–762, 1991.

[6] Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. Monitoring changes in linear models. *J. Statist. Plann. Inference*, 126:225–251, 2004.

[7] Hušková, M. Permutation principle and bootstrap in change point analysis. *Fields Inst. Commun.*, 44:273–291, 2004.

[8] Kirch, C. *Resampling Methods for the Change Analysis of Dependent Data*. PhD thesis, University of Cologne, Cologne, 2006. http://kups.ub.uni-koeln.de/volltexte/2006/1795/.

[9] Kirch, C. Block permutation principles for the change analysis of dependent data. *J. Statist. Plann. Inference*, 137:2453–2474, 2007.

[10] Kirch, C. Resampling in the frequency domain of time series to determine critical values for change-point tests. *Statist. Decisions*, 2007. Accepted.

[11] Silverman, B. W. *Density Estimation*. Chapman and Hall, London, 1986.

[12] Steland, A. A bootstrap view on Dickey-Fuller control charts for AR(1) series. *Austrian Journal of Statistics*, 35:339–346, 2006.

[13] von Bahr, B., and Esseen, C.-G. Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Stat.*, 36:299–303, 1965.