Bootstrapping Sequential Change-Point Tests

Claudia Kirch[1] and Marie Hušková[2]

[1] University of Kaiserslautern
Mathematical Institute
Erwin-Schrödinger-Straße
67663 Kaiserslautern, Germany
ckirch@mathematik.uni-kl.de

[2] Charles University of Prague
Department of Statistics
Sokolovská 83
CZ – 186 75 Praha 8, Czech Republic
huskova@karlin.mff.cuni.cz

Abstract. Frequently, in change-point analysis the calculation of critical values is based on distributional asymptotics of the test statistic under the null hypothesis. Since convergence is often rather slow, bootstrapping methods to obtain critical values promise better small sample behavior. We will discuss several approaches to bootstrapping in a sequential context, where new data arrives steadily which can improve the bootstrap approximation. Taking these new samples into account, however, results in critical values that are not constant through time. This leads to some additional theoretic problems and the practical question of how to improve computation time for the bootstrap critical values.
Keywords. Bootstrap, sequential test, change-point analysis, linear regression.

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.

In change-point analysis this approach was first suggested by Antoch and Hušková (2001) and later pursued by others (for a recent survey confer Hušková, 2004).

All of those papers, however, deal with a posteriori tests. In many situations sequential or on-line monitoring tests are much more realistic, where 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 historic 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 (2006) used a bootstrap in sequential testing of the unit-root problem.

The aim is to examine and compare different bootstrap procedures for sequential change-point tests. To this end we use the same model as Antoch and Hušková (2001) when first considering permutation tests for change-point analysis but we use it in a sequential setting and discuss some possible extensions to linear regression models later on. We consider the following mean change problem. Let

$$X(i) = \mu(i) + \varepsilon(i), \tag{1}$$

where $\{\varepsilon(i) : i \geq 1\}$ is a sequence of i.i.d. random variables with $\mathrm{E}\varepsilon(1) = 0$, $0 < \mathrm{E}|\varepsilon(1)|^\nu < \infty$ for some $\nu > 2$. We assume that we have observed a historic data set of length m, where no change has occurred, i.e.

$$\mu(i) = \mu_0, \quad 1 \leq i \leq m. \tag{2}$$

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

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

The simplest CUSUM-type monitoring statistic is given by

$$\Gamma(m,k)(X(1), X(2), \ldots, X(m+k)) = \frac{\left|\sum_{m<i\leq m+k} \left(X(i) - \frac{1}{m}\sum_{j=1}^m X(j)\right)\right|}{m^{1/2}\left(1 + \frac{k}{m}\right)}. \tag{3}$$

We reject the null hypothesis at

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

for

$$\widehat{\sigma}_m^2 = \widehat{\sigma}_m^2(X(1), \ldots, X(m)) = \frac{1}{m-1}\sum_{i=1}^m (X(i) - \bar{X}_m)^2$$

and 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 a large class of alternatives the limit converges to 1.

2 Bootstrapping methods in a sequential setting

The goal is to find a suitable bootstrap method to approximate the critical value c. In all cases we will use a studentized bootstrap because it has shown to work best.

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.

Historic Bootstrap B.1

The simplest idea is to use a bootstrap only based on the historic data set $X(1), \ldots, X(m)$. Precisely we use the bootstrap sample $X^{(1)*}(i) = X(U_m(i))$, $i = 1, \ldots, m + N(m)$, where $\{U_m(\cdot)\}$ are i.i.d. (and independent of $X(\cdot)$) with $P(U_m(1) = k) = 1/m$, $k = 1, \ldots, m$. Then, we calculate the critical value $c_m^{(1)}$ based on the conditional distribution of the studentized statistic

$$\max_{1\leq k\leq N(m)} \frac{\Gamma(m,k)(X^{(1)*}(1), \ldots, X^{(1)*}(m+k))}{\widehat{\sigma}_m(X^{(1)*}(1), \ldots, X^{(1)*}(m))}.$$

This bootstrap yields an asymptotic consistent test in the sense that

$$\lim_{m\to\infty} P_{H_0}\left(\frac{1}{\widehat{\sigma}_m}\sup_{1\leq k\leq N(m)} \Gamma(m,k) \geq c_m^{(1)}\right) = \alpha,$$

$$\lim_{m\to\infty} P_{H_1}\left(\frac{1}{\widehat{\sigma}_m}\sup_{1\leq k\leq N(m)} \Gamma(m,k) \geq c_m^{(1)}\right) = 1$$

for a large class of alternatives.

The simulations below show that it holds the level much better than the test based on asymptotic critical values at least for small m but that we loose some power.

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 sample $X^{(2)*}(i) = X(U_{m,k}(i))$, $i = 1, \ldots, m+N(m)$, where $\{U_{k,m}(\cdot)\}$ are i.i.d. (and independent of $X(\cdot)$) with $P(U_{k,m}(1) = j) = 1/(m+k-1)$, $j = 1, \ldots, m+k-1$.

We calculate the critical value $c_m^{(2)}(k)$ at time $m+k$ based on the conditional distribution

$$F_{k,m}^{(2)*}(x) = P_{m,k}^{*}\left(\frac{1}{\widehat{\sigma}_m(X^{(2)*}(1), \ldots, X^{(2)*}(m))} \sup_{1 \le l \le N(m)} \Gamma(m,l)(X^{(2)*}(1), \ldots, X^{(2)*}(m+l) \le x \right),$$

where $P_{m,k}^{*}(\cdot) = P(\cdot | X(1), \ldots, X(m+k-1))$ is the conditional probability given $X(1), \ldots, X(m+k-1)$.

This bootstrap also yields an asymptotic consistent test and holds the level quite well for small sample sizes. Furthermore, the power is much better than for the historic bootstrap above as some simulations show.

The disadvantage is that it is computationally very expensive, since we calculate a new critical value after each incoming observation.

Bootstrap B.3

The third bootstrap tries to use new information but at the same time reduce the computation time. The idea is that 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 (2006).

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 the parameters below 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 we calculate the critical values $c_m^{(3)}(k)$ at time $m+k$ based on the conditional distribution function

$$F_{k,m}^{(3)*}(x) = \frac{1}{M} \sum_{i=0}^{M-1} F^{(2)*}_{\max((j-i)\,L,0),m}, \qquad \text{for } k = j\,L, \ldots, (j+1)L-1.$$

This modification yields also asymptotically correct critical values in the sense that

$$\lim_{m \to \infty} P_{H_0}\left(\frac{1}{\widehat{\sigma}_m} \sup_{1 \le k \le N(m)} \frac{\Gamma(m,k)}{c_m^{(3)}(k)} \ge 1 \right) = \alpha,$$
$$\lim_{m \to \infty} P_{H_1}\left(\frac{1}{\widehat{\sigma}_m} \sup_{1 \le k \le N(m)} \frac{\Gamma(m,k)}{c_m^{(3)}(k)} \ge 1 \right) = 1$$

for a large class of alternatives.

In the simulations below we use $L = m/5$ 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 the computations are indeed a lot faster.

3 Selected Simulations

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 test should stop 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 $p_k, 1 \leq k \leq N(m) < \infty$, where p_k is the p-value of $\Gamma(m,k)/\widehat{\sigma}_m$ with respect to the distribution from which we obtain the critical value $c_m(k)$.
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, thus visualizes $\mathcal{C}.1$ and $\mathcal{C}.2$. 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 estimated density of the run length

We give a plot of the (estimated) density of the run length for $d = 2$, i.e. the time (after monitoring starts) at which the null hypothesis is rejected at the 5% level. Runs for which the null hypothesis is never rejected are not taken into account. The vertical line in the plot indicates where the change occurred. This is a visualization of $\mathcal{C}.3$.

We only give some examples to show that the claims in the previous section are indeed true.

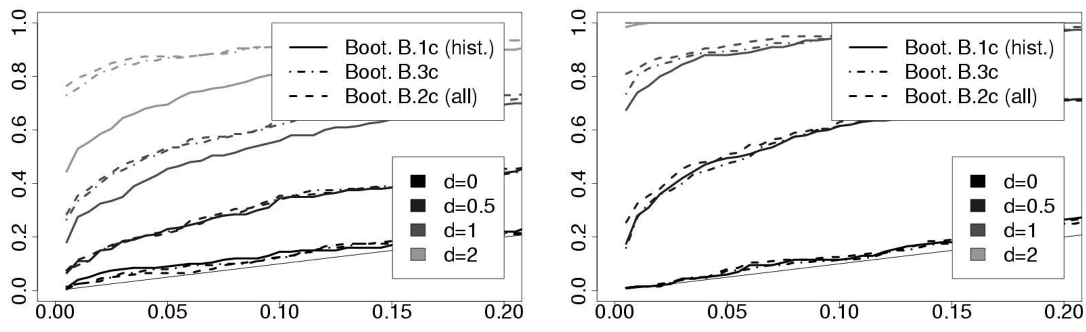

Fig. 1. Different bootstrap procedures for exponential errors and $m = 10, N(m) = 5m, k^* = 5$ (left) respectively $m = 50, N(m) = 2m, k^* = 25$ (right).

Bootstrap $B.2$ works best, but $B.3$ is almost as good yet the computation time is much better. Therefore we only consider Bootstrap $B.3$ in the following.

For the parametric bootstrap we used the actual distribution of the statistic under the assumption that the observations are normal with a variance that is equal to the estimated variance (this is the asymptotic distribution for fixed observation horizon $N(m)$). The asymptotic critical value are based on the asymptotics of the test statistic for an infinite observation horizon.

One can see that for $m = 50$ all methods become approximately equivalent and that the parametric bootstrap and the asymptotic test become approximately equivalent for $N(m) = 10m$ (cf. Figures 1 –3).

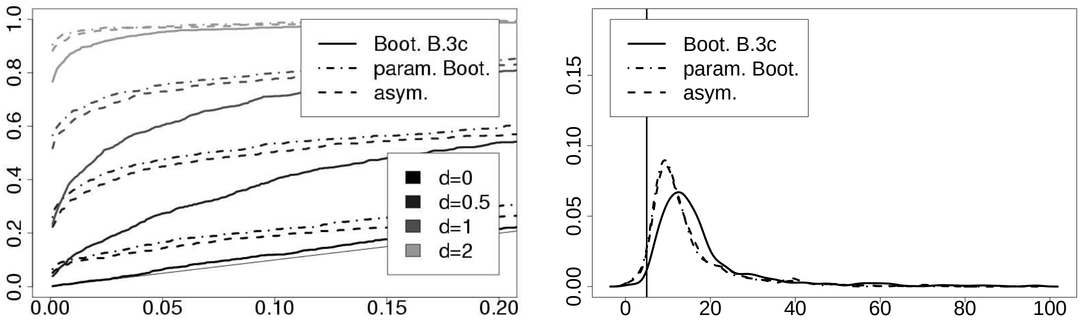

Fig. 2. Bootstrap and asymptotic tests for exponential errors and $m = 10, N(m) = 10m, k^* = 5$: SPC (left), plot of estimated density (right).

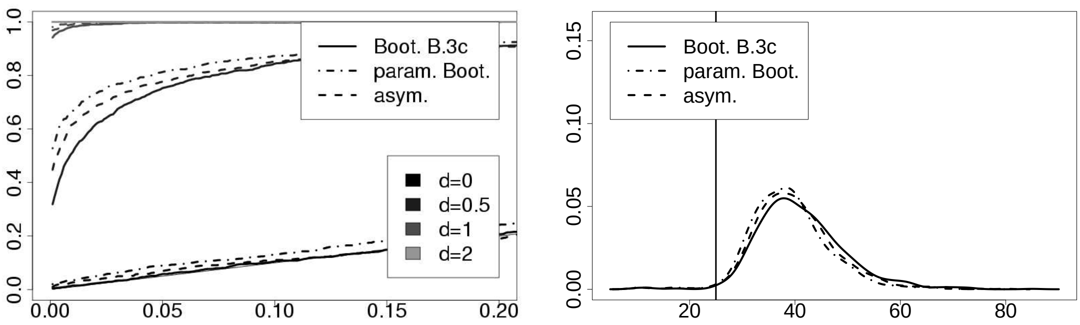

Fig. 3. Bootstrap and asymptotic tests for exponential errors and $m = 50, N(m) = 10m, k^* = 25$: SPC (left), plot of estimated density (right).

If we choose the statistic slightly different, namely we divide by $(k/(m+k))^\gamma$, $0 \leq \gamma < \frac{1}{2}$, then the asymptotic statistic has for larger values of γ a high tendency to reject the null hypothesis at the very beginning of the observation period even if nothing has changed yet. Figure 4 shows that the bootstrap also behaves much better with respect to this phenomenon.

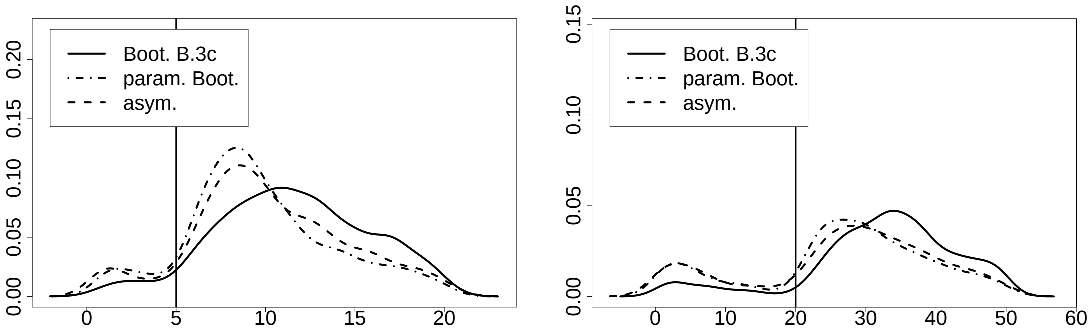

Fig. 4. Density plots of the run length for bootstrap and asymptotic tests for exponential errors and $m = 10, \gamma = 0.25$ as well as $N(m) = 2m, k^* = 5$ (left) respectively $N(m) = 5, k = 20$ (right).

4 Testing for a change in a linear regression model

Instead of finding a mean change we would like to find a change in the regression coefficient $\boldsymbol{\beta} = \boldsymbol{\beta}_i$ for a linear regression $y(i) = \boldsymbol{x}(i)^T \boldsymbol{\beta}_i + \epsilon(i)$. The statistic of interest is given by

$$\Gamma(m,k,\gamma) = \frac{\sum_{i=m+1}^{m+k} y(i) - \sum_{i=m+1}^{m+k} \boldsymbol{x}(i)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j)y(j)}{m^{1/2}\left(1+\frac{k}{m}\right)},$$

where

$$\boldsymbol{C}_m = \sum_{i=1}^{m} \boldsymbol{x}(i)\boldsymbol{x}(i)^T.$$

Under the null hypothesis and for $k \leq k^*$ the statistic is equal to

$$\Gamma(m,k,\gamma) = \frac{\sum_{i=m+1}^{m+k} e(i) - \sum_{i=m+1}^{m+k} \boldsymbol{x}(i)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j)e(j)}{m^{1/2}\left(1+\frac{k}{m}\right)},$$

which is the version that we will bootstrap.

A typical bootstrap in this context is the so called regression bootstrap, where the regressors $\boldsymbol{x}(i)$ are kept constant and only the estimated errors

$$\widehat{\epsilon}_{m,k}(i) = y(i) - \boldsymbol{x}(j)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j)y(j)$$

are bootstrapped, i.e. $\epsilon^*_{m,k}(i) = \widehat{\epsilon}_{m,k}(U_{m,k}(i))$.

We can then use the same techniques as in the previous section. The additional problem arising is the term $\sum_{i=m+1}^{m+l} \boldsymbol{x}(i)^T$ which depends on regressors that we have not yet observed for $l \geq k$. Therefore in the bootstrap statistic we replace it at time point $m+k$ by

$$\mathbf{c}_1(m,k,l) = \begin{cases} \sum_{i=m+1}^{m+l} \boldsymbol{x}(i), & l \leq k, \\ \sum_{i=m+k-l+1}^{m+k} \boldsymbol{x}(i), & k < l < k+m, \\ \frac{l}{m+k} \sum_{i=1}^{m+k} \boldsymbol{x}(i), & l \geq m+k. \end{cases}$$

We can then again show that the corresponding sequential bootstrap test has asymptotic level α and power 1 for the same class of alternatives as the original asymptotic test.

References

Antoch, J. and Hušková, M. (2001). Permutation tests for change point analysis. *Statist. Probab. Lett.*, **53**, 37–46.
Hušková, M. (2004). Permutation principle and bootstrap in change point analysis. *Fields Inst. Commun.*, **44**, 273–291.
Hušková, M. and Kirch, C. (2009+). Bootstrapping sequential change-point tests for linear regression. In preparation.
Kirch, C. (2008). Bootstrapping sequential change-point tests. *Sequential Anal.*, **27**, 330-349.
Steland, A. (2006). A bootstrap view on Dickey-Fuller control charts for AR(1) series. *Austrian Journal of Statistics*, **35**, 339–346.