# Bootstrapping sequential change-point tests for linear regression *

Marie Hušková†, Claudia Kirch‡

May 12, 2010

**Abstract**

In this paper we propose some bootstrapping methods to obtain critical values for sequential change-point tests for linear regression models. 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 $\alpha$– resp. $\beta$–errors and its run length.

**Keywords:** Bootstrap, sequential test, change-point analysis, linear regression

**AMS Subject Classification 2000:** 62L10, 62G09, 62J05

## 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 [9]. For a thorough introduction into permutation and bootstrap tests we refer to Good [8].

In change-point analysis this approach was first suggested by Antoch and Hušková [2] and later pursued by others (for a recent survey confer Hušková [11]). Berkes et. al. [5]

*The work was supported by DFG-Grant KI 1443/2-1, the work of the first author was supported by MSM 0021620839 and GACR 201/09/J006 and the position of the second author was financed by the Stifterverband für die Deutsche Wissenschaft by funds of the Claussen-Simon-trust.

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

‡Karlsruhe Institute of Technology (KIT), Institute for Stochastics, Kaiserstr. 89, D – 76133 Karlsruhe, Germany; claudia.kirch@kit.edu

showed that the bootstrap provides better approximations for the critical values than asymptotics in a number of change-point situations. All of those papers, however, deal with a posteriori tests, i.e. tests, where we have observed the complete data set already.

In recent years an increasing number of data sets are collected automatically or without significant costs in such a way that the observations arrive steadily. Examples include financial data sets e.g. in risk management (Andreou and Ghysels [1]) or CAPM models (Aue et al. [3]) as well as medical data sets e.g. monitoring intensive care patients (Fried and Imhoff [7]). More applications can be found in different areas of applied statistics. With each new observation the question arises whether the model is still capable of explaining the data. If this is not the case an alarm needs to be raised, for example the financial models might not be appropriate anymore or the condition of the patient in intensive medical care might have changed.

The consideration of such data sets leads to sequential statistical analysis, which is sometimes also called on-line monitoring. 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.

In the classical statistical setting, where we have observed the complete data set already, the bootstrap has turned out to be very useful in many applications. Therefore, we are interested in developing variations of the bootstrap that are appropriate in a sequential setting. However, it is not obvious how best to do that. 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 [16] used a bootstrap in sequential testing of the unit-root problem. Kirch [14] considered several possibilities to do sequential bootstrapping in the simple case of a mean change for i.i.d. errors. Not surprisingly it turned out that the bootstrap versions of the test behave better for small sample sizes than the asymptotic tests. While the validity of the bootstrap that is only based on the historic data sequence is easily proven, this method only behaves well for larger historic data sequences. For smaller sample sizes the power is smaller than for the bootstrap that takes every observation into account. The latter bootstrap version recalculates the critical values after each new observation using all observation up to that time. The drawback is that it is computationally very expensive. This is why in practical situations it is better to update the critical values only from time to time and also to use new bootstrap samples as well as old ones for this procedure. The simulation study in Kirch [14] showed that this almost yields the same results as if one updates the critical values after each observation, but it is much faster. Moreover it turned out that only the studentized test statistics did hold the level right for small samples.

In this paper we would like to follow that ideas but some additional problems arise due

to the much more complicated data structure. We focus on the linear regression model

$$y(i) = \mathbf{x}(i)^T \boldsymbol{\beta}_i + e(i), \quad i \geqslant 1, \tag{1.1}$$

where $\mathbf{x}(i)$ is a $p \times 1$ random vector and $\boldsymbol{\beta}_i$ is a $p \times 1$ vector. Furthermore we assume that the error sequence is i.i.d. and independent of the regressors. However, the proposed version of the bootstrap can be extended to other sequential setups including regression models for dependent data or nonlinear models in a similar fashion as the bootstrap can be extended to such settings in a classical off-line model.

We assume that we have a historic sequence of observations, where no change in the regression coefficient occurred, i.e.

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad 1 \leqslant i \leqslant m. \tag{1.2}$$

Now we are interested in testing the null hypothesis of no change in the monitoring period

$$H_0 : \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad m < i < m + N(m) + 1 \tag{1.3}$$

against the alternative of a change in the regression coefficient

$$\begin{aligned} H_1 :&\ \text{there is a } k^\circ \geqslant 1 \text{ such that } \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad m < i \leqslant m + k^\circ \\ &\text{and } \boldsymbol{\beta}_i = \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_0, \quad m + k^\circ < i < m + N(m) + 1. \end{aligned} \tag{1.4}$$

$N(m)$ is the observation horizon which can be finite or infinite but has to converge to infinity with $m$. The values of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ and $k^\circ$ are not known and can depend on $m$. This includes local changes for which $\boldsymbol{d}_m := \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0 \to 0$. The test is then based on the following statistic

$$\begin{aligned} &\Gamma(m, k, \gamma) = \sum_{m < i \leqslant m+k} \left( y(i) - \mathbf{x}(i)^T \widehat{\boldsymbol{\beta}}_m \right) / g(m, k, \gamma), \\ &\text{where } g(m, k, \gamma) = m^{1/2} \left( 1 + \frac{k}{m} \right) \left( \frac{k}{m+k} \right)^\gamma \end{aligned} \tag{1.5}$$

for $0 \leqslant \gamma < 1/2$ and

$$\widehat{\boldsymbol{\beta}}_m = \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \mathbf{x}(j) y(j), \qquad \text{where } \boldsymbol{C}_m = \sum_{i=1}^{m} \mathbf{x}(i) \mathbf{x}(i)^T,$$

is the least squares estimator of the regression coefficient based on the historic data set $y(1), \ldots, y(m)$. The statistic is then given by

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

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

$$\widehat{\sigma}_m^2 = \frac{1}{m-p} \sum_{i=1}^{m} \left( y(i) - \mathbf{x}(i)^T \widehat{\boldsymbol{\beta}}_m \right)^2 . \tag{1.6}$$

We reject the null hypothesis at the following stopping time

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

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

$$\lim_{m\to\infty} P(\tau(m) < \infty) = \alpha \tag{1.7}$$

for some given level $0 < \alpha < 1$. We require that under the alternative $H_1$

$$\lim_{m\to\infty} P(\tau(m) < \infty) = 1. \tag{1.8}$$

The paper is organized as follows: In Section 2 we summarize some known results on the limit behavior of the test statistic under the null as well as alternative hypotheses. In Section 3 we introduce the so called regression bootstrap in a sequential setting and show that the corresponding bootstrap test is asymptotically equivalent to the procedure based on asymptotic critical value. In Section 4 the corresponding result is given for another type of bootstrap namely the pair bootstrap.
In Section 5 some simulations illustrate the usefulness of the bootstrap methods. Finally the proofs are given in Sections 6 and 7 for the regression and pair bootstrap, respectively.

## 2 Assumptions and limit behavior of the test statistic

Here, we shortly summarize the assumptions and the results proved in Horváth et al. [10] on the behavior of the test statistics related to the above introduced test procedure.

**Assumption $\mathcal{A}$. 1.** *We assume that the sequence of vectors of regressors $\{\boldsymbol{x}(i)\}$ and the sequence of random errors $\{e(i)\}$ satisfy*

*(i)* *$\{e(i) : 1 \leqslant i < \infty\}$ are independent identically distributed (i.i.d.) random variables with*

$$\mathrm{E}\, e(i) = 0, \quad 0 < \operatorname{var} e(i) = \sigma^2, \quad \mathrm{E}\,|e(i)|^\nu < \infty \quad \text{for some } \nu > 2,$$

*(ii)* *for the sequence of vectors $\{\boldsymbol{x}(i) = (1, x_2(i), \ldots, x_p(i))^T : 1 \leqslant i < \infty\}$ there exists a positive definite matrix $\mathbf{C}$ and a constant $0 < \rho \leqslant 1/2$ such that*

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}(i)\mathbf{x}(i)^T - \mathbf{C} \right\|_\infty = O(n^{-\rho}) \quad a.s.,$$

*where $\|\cdot\|_\infty$ denotes the maximum norm of matrices,*

*(iii)* *the sequences $\{e(i) : 1 \leqslant i < \infty\}$ and $\{\boldsymbol{x}(i) : 1 \leqslant i < \infty\}$ are independent.*

Next we formulate the main results proved in Horváth et al. [10] and also some of their useful simpler modifications.

The following theorem gives the null asymptotics, which have been proven for the open-end procedure in Horváth et al. [10]. The results for the closed-end procedure can be obtained analogously but are not stated in that paper. However, for the bootstrap techniques discussed in this paper, it is the closed-end procedures that play he crucial role.

**Theorem 2.1.** *Let* (1.1) *and Assumption* $\mathcal{A}.1$ *hold true. Let*

$$0 \leqslant \gamma < \min(\rho, 1/2). \tag{2.1}$$

*Then, under the null hypothesis (i.e.* (1.2) *and* (1.3)*), for all* $y \in \mathbb{R}$

$$\lim_{m\to\infty} P\left(\sup_{1\leqslant k<\infty} \frac{|\Gamma(m,k,\gamma)|}{\widehat{\sigma}_m} \leqslant y\right) = P\left(\sup_{0\leqslant t\leqslant 1} \frac{|W(t)|}{t^\gamma} \leqslant y\right), \tag{2.2}$$

*holds true (open-end procedure). Here* $\{W(t) : 0 \leqslant t < \infty\}$ *denotes a Wiener process.*

*For* $N(m) < \infty$ *(closed-end procedure) with* $\lim_{m\to\infty} N(m)/m = N$ *for some* $0 < N < \infty$ *or* $\lim_{m\to\infty} N(m)/m = \infty$ *it holds as* $m \to \infty$

$$P\left(\sup_{1\leqslant k<N(m)+1} \frac{|\Gamma(m,k,\gamma)|}{\widehat{\sigma}_m} \leqslant y\right) = P\left(\sup_{1\leqslant k<N(m)+1} \frac{|W_1(k/m) - k/mW_2(1)|}{(1+k/m)(k/(k+m))^\gamma} \leqslant y\right) + o_P(1), \tag{2.3}$$

*where* $\{W_1(\cdot)\}$, $\{W_2(\cdot)\}$ *are independent Wiener processes.*

**Proof.** Assertion (2.2) with $N(m) = \infty$ is proven in Horváth et al. [10]. Going through the proof of Theorem 2.1 in Horváth et al. [10] we realize that assertion (2.3) holds true. ■

**Remark 2.1.** a) The limit distribution in (2.2) is explicitly only known for $\gamma = 0$.

b) The distribution on the right hand side of (2.3) converges to the same limit as given in (2.2) for $N(m)/m \to \infty$ and it converges to

$$\sup_{0\leqslant t\leqslant N/(N+1)} \frac{|W(t)|}{t^\gamma}$$

for $N(m)/m \to N$ (cf. the proof of Theorem 2.1 in Horváth et al. [10]).

c) In the simulation study we compare our bootstrap procedures with the closed-end procedure where the critical values are obtained from simulated quantiles of the distribution on the right hand side of (2.3). This will be called **asymptotic closed-end procedure**. Simulations concerning the location model in Kirch [14] show that this distribution is very close to the one in (2.2) if $N \geqslant 10$. For a smaller observation horizon it is not recommendable to use critical values from the distribution in (2.2) (for a detailed discussion we refer to Kirch [14]).

**Remark 2.2.** Horváth et al. [10] pointed out that a value of $\gamma$ 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 $\gamma$, but a $\gamma$ close to $1/2$ has a higher probability of a false alarm well before the change occurred.

The following theorem has been proven in Horváth et al. [10] for the open-end procedures and fixed alternatives. The same proof techniques can be applied to obtain the following result also including closed-end procedures as well as local changes.

**Theorem 2.2.** *Let* (1.1)*, Assumption* $\mathcal{A}.1$*, and* (2.1) *hold true. Let*

$$k^\circ = O(m), \qquad \lim_{m\to\infty} \sqrt{m}|\boldsymbol{c}_1^T \boldsymbol{d}_m| = \infty, \tag{2.4}$$

*hold, where* $\boldsymbol{c}_1$ *is the first column of* $\boldsymbol{C}$ *and* $\boldsymbol{d}_m = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$*. Then, under* $H_1$*, as* $m \to \infty$*,*

$$\frac{1}{\widehat{\sigma}_m} \sup_{1\leqslant k<N(m)} |\Gamma(m,k,\gamma)| \xrightarrow{P} \infty, \tag{2.5}$$

*for* $\lim_{m\to\infty} N(m)/m = \infty$ *or* $N(m) = \infty$*.*
*Assertion* (2.5) *remains true if* $\lim_{m\to\infty} N(m)/m = N$ *for some* $0 < N < \infty$ *and if* $\limsup_{m\to\infty} k^\circ/m < N$*.*

**Proof.** The assertion follows analogously to the proof of Theorem 2.2 in Horváth et al. [10]. ■

The assertions in Theorems 2.1 and 2.2 were proved in a more general setup including e.g. heteroscedastic errors, cf. Aue et al. [4]. However, bootstrapping methods need to be adapted in order to work well in such situations. Furthermore, one can consider a different class of test statistics, e.g. Hušková and Koubková [12, 13] and Koubková [15] developed and studied the limit behavior of test procedures based on $L_1$ estimators and related partial sums of residuals instead of the corresponding $L_2$ procedures above.

The test based on the test statistic (1.5) is only consistent under (2.4) which is quite restrictive, since essentially it means that our change somehow implies a mean change of $y(\cdot)$. Hušková and Koubková [12] introduced test procedures based on quadratic forms of weighted partial sums of residuals, which yield consistent tests for all fixed alternatives of the above type. Extensions of the bootstrapping techniques developed in this paper to these test statistics are in principle possible but quite technical and will be considered elsewhere.

# 3 Regression Bootstrap

In linear regression there are essentially two main approaches to bootstrapping, namely the regression or fixed design bootstrap and the pair bootstrap. We will use the index $R$ for the regression bootstrap and the index $P$ for the pair bootstrap. In this section we discuss the first one.

The general idea of the regression bootstrap is that we resample the estimated residuals but keep the regressors in their original order, so in the bootstrap world we deal with a regression with a fixed design rather than a stochastic one. In sequential bootstrapping this yields the problem that we have only observed the regressors up to the current time point but we also need the future ones.

First, observe that under the null hypothesis respectively for $\ell \leqslant k^\circ$ under the alternative

$$y(\ell) - \mathbf{x}(\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} y(j)\mathbf{x}(j) = e(\ell) - x(\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} e(j)\mathbf{x}(j), \tag{3.1}$$

thus under the null hypothesis and for $\ell \leqslant k^\circ$ under alternatives it holds

$$\Gamma(m,\ell,\gamma) = \left( \sum_{i=m+1}^{m+\ell} e(i) - \sum_{i=m+1}^{m+\ell} \mathbf{x}(i)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \mathbf{x}(j) e(j) \right) / g(m,\ell,\gamma). \tag{3.2}$$

This is the version we will use in the bootstrap as it splits the errors and regressors in addition to being closer to the null hypothesis even under alternatives.

The general idea in sequential bootstrapping is to repeat the bootstrap procedure at several times during the monitoring in order to incorporate the increased knowledge obtained from the additional observations in the bootstrap. Suppose now that we are at point $m+k$ in the monitoring, i.e. we know the $m$ observations from the historic data set in addition to $k$ observations that we have already obtained during the monitoring (without rejecting yet). Based on those $m+k$ observations $(y(i), \mathbf{x}(i))$, $i = 1, \ldots, m+k$, we construct a bootstrap statistic in the following way:
First, we replace $e(i)$ in the formula on the right hand side of (3.2) by the bootstrap estimates $e^*_{m,k}(i)$ below and keep $\mathbf{x}(i)$ for $1 \leqslant i \leqslant k+m$. The problem is that the test statistic $\sup_{1 \leqslant \ell < N(m)+1} |\Gamma(m,\ell,\gamma)|$ (with $\Gamma(m,\ell,\gamma)$ as in the right-hand side of (3.2)) depends additionally on the future regressors $\mathbf{x}(\ell), \ell \geqslant k$, which have not been observed yet (at least in the more interesting case of a random design). More precisely it depends on the term $\sum_{i=m+1}^{m+\ell} \mathbf{x}(i)^T$ which contains unknown regressors if $\ell > k$. In order to use as much information as possible and still be close to the original statistic (also in the situation where $k$ is very small), we propose to replace this term by $\mathbf{c}_1(m,k,\ell)$ below. Different choices are possible as long as they fulfill Lemma 6.1 b) as well as

$$\mathbf{c}_1(m,k,\ell)^T (1,0,\ldots,0) = 1.$$

To sum up, for the calculation of bootstrap critical values we use the 'test statistic'[1] $\sup_{1 \leqslant \ell < N(m)+1} |\widetilde{\Gamma}(m,\ell,\gamma)|$ where

$$\widetilde{\Gamma}(m,\ell,\gamma)(e(1),\ldots,e(m+\ell)) = \left( \sum_{i=m+1}^{m+\ell} e(i) - \mathbf{c}_1(m,k,\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \mathbf{x}(j) e(j) \right) / g(m,\ell,\gamma),$$

$$\text{where } \mathbf{c}_1(m,k,\ell) = \begin{cases} \sum_{i=m+1}^{m+\ell} \mathbf{x}(i), & \ell \leqslant k, \\ \sum_{i=m+k-\ell+1}^{m+k} \mathbf{x}(i), & k < \ell < k+m, \\ \frac{\ell}{m+k} \sum_{i=1}^{m+k} \mathbf{x}(i), & \ell \geqslant m+k. \end{cases} \tag{3.3}$$

for $1 \leqslant \ell < N(m)+1$. Let

$$e^*_{m,k}(i) = \widehat{e}_{m,k}(U_{m,k}(i)), \qquad \text{where} \quad \widehat{e}_{m,k}(j) = y(j) - \mathbf{x}(j)^T \widehat{\boldsymbol{\beta}}_{m+k}, \tag{3.4}$$

$i = 1, \ldots, m+N(m)$, $j = 1, \ldots, m+k$, where $\{U_{m,k}(i) : 1 \leqslant i \leqslant m+N(m)\}$ are i.i.d. random variables with $P(U_{m,k}(1) = j) = 1/(m+k)$, $j = 1, \ldots, m+k$, independent of $\{y(i) : 1 \leqslant i \leqslant m+N(m)\}$ and $\{\mathbf{x}(i) : 1 \leqslant i \leqslant m+N(m)\}$. By $P^*_{m,k}$, $\mathrm{E}^*_{m,k}$,

[1] This is not a test statistic as the errors $e(i)$ are not observable, but it is quite useful for the regression bootstrap where we artificially create bootstrap errors.

$\text{var}^*_{m,k}$ etc. we denote the conditional probability, expectation, variance etc. given $\{(y(i), \mathbf{x}(i)^T) : 1 \leqslant i \leqslant m+k\}$, i.e. with respect to $\{U_{m,k}(i) : 1 \leqslant i \leqslant m + N(m)\}$.

Now, we are ready to discuss the sequential bootstrap more precisely. A first idea is to calculate critical values at time $m+k$ based on the distribution $P^*_{m,k}$, i.e. based on the quantiles of

$$F^{(R*)}_{m,k}(x) = P^*_{m,k}\left[\frac{1}{\widehat{\sigma}^{(R*)}_{m,k}} \sup_{1\leqslant \ell < N(m)+1} \left|\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1),\ldots,e^*_{m,k}(m+\ell))\right| \leqslant x\right],$$

where

$$\left(\widehat{\sigma}^{(R*)}_{m,k}\right)^2 = \frac{1}{m-p}\sum_{i=1}^{m}\left(e^*_{m,k}(i) - \mathbf{x}(i)^T \boldsymbol{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)e^*_{m,k}(j)\right)^2 \tag{3.5}$$

is the bootstrap version of (1.6) by (3.1).
However, it is computationally too expensive to generate new bootstrap samples after each new incoming observation in the above way and calculate the critical values based on these.

Therefore, we follow an approach by Steland [16], that has also proven to work well in Kirch [14]. 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.

We apply two modifications to reduce computation time significantly. First, we calculate new critical values only after each $L$th 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 bootstrap samples but also on older samples. This is why we need to generate only a fraction of the bootstrap samples each time we update: E.g. for $\alpha_i = 1/M$ 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 ($L = 1$).
Let for $j \geqslant 1$, $\sum_{i=0}^{M-1}\alpha_i = 1$ and $\alpha_i \geqslant 0$

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

Then, we calculate the critical values $c^{(R)}_{m,k}$ at time $k+m$ as follows

$$\widetilde{F}^{(R)}_{m,k}(c^{(R)}_{m,k}) \geqslant 1-\alpha, \tag{3.6}$$

$c^{(R)}_{m,k}$ minimal.

For the simulations in this paper we use the convex combinations with equal weights $\alpha_i = \frac{1}{M}$, $L = m/5$ and $M = 5$. Thus, after monitoring for $m$ observations we have completely replaced the bootstrap samples.

Now we are ready to state the main theorem of this section:

**Theorem 3.1.** *Let* (1.1)*, Assumption* $\mathcal{A}$*.1, and* (2.1) *hold true, let* $N(m) < \infty$ *(closed-end procedure) with either* $\lim_{m\to\infty} \frac{N(m)}{m} = N$*,* $0 < N < \infty$*, or* $\lim_{m\to\infty} \frac{N(m)}{m} = \infty$*. Then we have as* $m \to \infty$

*a) under the null hypothesis,*

$$P\left(\frac{1}{\widehat{\sigma}_m} \sup_{1\leqslant k<N(m)+1} \frac{|\Gamma(m,k,\gamma)|}{c^{(R)}_{m,k}} > 1\right) \to \alpha.$$

*b) Under the assumptions of Theorem 2.2 and if additionally* $\boldsymbol{d}_m = O(1)$*, then*

$$P\left(\frac{1}{\widehat{\sigma}_m} \sup_{1\leqslant k<N(m)+1} \frac{|\Gamma(m,k,\gamma)|}{c^{(R)}_{m,k}} > 1\right) \to 1.$$

**Remark 3.1.** The assertions of Theorem 3.1 remain true, if we use the open-end procedure, i.e. $N(m) = \infty$, and critical values based on a bootstrap with horizon $\widetilde{N}(m) < \infty$ 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.

**Remark 3.2.** Clearly, the test procedure based on the bootstrap approximation of critical values have the desired properties (1.7) and (1.8).
Moreover, under $H_0$ and local alternatives the bootstrap provides an asymptotically correct approximation for the critical value in the sense given in equation (6.24).
Under alternatives we can only obtain that the bootstrap critical values are uniformly bounded, cf. (6.25). However, it is to be expected that more technical proofs also yield that the bootstrap critical values under alternatives are asymptotically correct in a $P$-stochastic sense corresponding to (6.24). The reason is as follows: Checking the proofs it can be seen that the term that is responsible for the weaker result is $\frac{1}{m+k}\mathbf{C}_{k^\circ,k}$ (cf. e.g. (6.13)), which is without further knowledge only uniformly bounded. As soon as we can prove that it converges to 0 uniformly for $1 \leqslant k \leqslant \widetilde{\tau}(m)$, where $\widetilde{\tau}(m)$ is the stopping time of the procedure, the stronger result follows. But Aue et al. [4] show for the stopping time $\tau(m)$ of the procedure with the asymptotic critical values from Theorem 2.1 that

$$\tau(m) - k^\circ = o_P(m + k^\circ).$$

It is to be expected that this result remains true for a sequence of critical values as in the bootstrap as long as this sequence is uniformly bounded, which in turn implies $\sup_{1\leqslant k\leqslant\tau(m)} \|\frac{1}{m+k} C_{k^\circ,k}\|_\infty = o_P(1)$ as desired.

# 4 Pair Bootstrap

In this section we introduce another bootstrapping scheme in linear regression. This is especially suitable but not restricted to situations where $(e(1), \mathbf{x}(1)), (e(2), \mathbf{x}(2)), \ldots$ are i.i.d. vectors. For dependent situations a block version also seems suitable.

The advantage of the pair bootstrap is that the dependence structure between $\mathbf{x}(i)$ and $y(i)$ is directly preserved in the bootstrap, thus we expect it to be more robust in situations where the regression is not purely linear. In fact, in the simulations of

the misspecified Scenarios 4 and 5 the pair bootstrap does behave quite well, but the regression bootstrap does not behave too badly either.

The specifics about sequential bootstrapping are the same as for the regression bootstrap.

Precisely we bootstrap the pairs $\{(y(i), \mathbf{x}(i)) : 1 \leqslant i \leqslant m+k\}$, i.e. let

$$y^*_{m,k}(i) = y(U_{m,k}(i)), \qquad \mathbf{x}^*_{m,k}(i) = \mathbf{x}(U_{m,k}(i)), \tag{4.1}$$

where $\{U_{m,k}(i) : 1 \leqslant i \leqslant m + N(m)\}$ are i.i.d. random variables with $P(U_{m,k}(1) = j) = 1/(m+k)$, $j = 1, \ldots, m+k$, independent of $\{y(i) : 1 \leqslant i \leqslant m + N(m)\}$ and $\{\mathbf{x}(i) : 1 \leqslant i \leqslant m + N(m)\}$ as above. Now we bootstrap the statistic and obtain as bootstrap statistic

$$\begin{aligned} &g(m,\ell,\gamma)\,\Gamma(m,\ell,\gamma)^{(P*)}_{m,k} \\ &= \sum_{m<i\leqslant m+\ell} \left( y^*_{m,k}(i) - \mathbf{x}^*_{m,k}(i)^T \left( \sum_{j=1}^{m} \mathbf{x}^*_{m,k}(j)\mathbf{x}^*_{m,k}(j)^T \right)^{-1} \sum_{j=1}^{m} \mathbf{x}^*_{m,k}(j)^T y^*_{m,k}(j) \right) \end{aligned}$$

Analogously to Section 3 we define

$$F^{(P*)}_{m,k}(x) = P_{m,k}\left( \frac{1}{\widehat{\sigma}^{(P*)}_{m,k}} \sup_{1\leqslant \ell < N(m)+1} \left|\Gamma(m,\ell,\gamma)^*_{m,k}\right| \leqslant x \right), \tag{4.2}$$

where

$$\left(\widehat{\sigma}^{(P*)}_{m,k}\right)^2 = \frac{1}{m-p} \sum_{i=1}^{m} \left( y^*_{m,k}(i) - \mathbf{x}^*_{m,k}(i)^T \left(\mathbf{C}^*_{m,k}\right)^{-1} \sum_{l=1}^{m} \mathbf{x}^*_{m,k}(l)^T y^*_{m,k}(l) \right)^2,$$

$$\text{where } \mathbf{C}^*_{m,k} = \sum_{j=1}^{m} \mathbf{x}^*_{m,k}(j)\mathbf{x}^*_{m,k}(j)^T,$$

is the bootstrap version of (1.6) and

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

Then we calculate the critical value at time $k+m$ as

$$\widetilde{F}^{(P)}_{m,k}\left(c^{(P)}_{m,k}\right) \geqslant 1-\alpha,$$

$c^{(P)}_{m,k}$ minimal.

In order for the pair bootstrap to be valid we need somewhat stronger assumptions than before. Precisely one of the following two assumptions is needed.

**Assumption $\mathcal{A}$. 2.** *Let the observation horizon $N(m)$, on which the bootstrap is based, fulfill*

$$\frac{N(m)^{1-\rho}}{m} = O(m^{-\kappa})$$

*for some $\kappa > 0$ and $\rho$ is as in Assumption $\mathcal{A}$.1 (ii).*

This assumption is only a restriction on the bootstrap observation horizon, meaning we can still use an original statistic, which is based on an observation horizon which converges to infinity with a faster rate or more importantly is equal to infinity. In this case the result of the following theorem remains true as long as the observation horizon that is used in the bootstrap converges to infinity with a rate faster than $m$ and fulfills the above assumption (cf. also Remark 4.1).

Alternatively we can put some stronger assumptions on the regressors.

**Assumption $\mathcal{A}$. 3.** *The regressors fulfill for some $r > 1$ as $k \to \infty$*

$$k^{-1} \sum_{i=1}^{k} \|\boldsymbol{x}(i)\boldsymbol{x}(i)^T\|_\infty^r = O(1) \qquad P - a.s.$$

This assumption is also not very strong, since usually the rate in Assumption $\mathcal{A}$.1 (ii) is obtained by some higher moment assumptions which also imply the above.

Now we are ready to state the main theorem of this section, namely that the pair bootstrap is valid.

**Theorem 4.1.** *Let* (1.1)*, Assumption $\mathcal{A}$.1, and* (2.1) *hold true. Furthermore assume either $\mathcal{A}$.2 or $\mathcal{A}$.3. Then, we have as $m \to \infty$*

*a) under the null hypothesis,*

$$P\left(\frac{1}{\widehat{\sigma}_m} \sup_{1 \leqslant k < N(m)+1} \frac{|\Gamma(m,k,\gamma)|}{c_{m,k}^{(P)}} > 1\right) \to \alpha.$$

*b) Under the assumptions of Theorem 2.2 and if additionally $\boldsymbol{d}_m = O(1)$, then*

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

**Remark 4.1.** The assertion in Remarks 3.1 and 3.2 remain true for the pair bootstrap.

We would like to point out that for $p = 1$ both procedures coincide with the bootstrap in the location model considered in Kirch [14].

# 5 Some simulations

In the previous sections 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 where we compare the two bootstrap procedures with the asymptotic closed-end (CE) procedure (from Remark 2.1 c)).

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

$\mathcal{C}.1$ The actual level ($\alpha$-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 $\beta$-error should be small.

$\mathcal{C}.3$ The stopping time $\tau(m)$ should be shortly after the change-point. This is often also called run-length of the test.

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 a test under the null hypothesis as well as under various alternatives, where the $p$-values represent the smallest level for which this realization is rejected. What we get is a plot that shows the empirical size and power (i.e. the empirical $\alpha$–errors resp. $1-(\beta$–errors)) on the $y$-axis for the chosen level on the $x$-axis (thus visualizing $\mathcal{C}.1$ 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 close to 1.
The question is how to calculate the $p$-value in a sequential setting with possibly varying critical values. At point $m+k$ the null hypothesis is rejected if $|\Gamma(m,k,\gamma)|/\widehat{\sigma}_m$ is greater than the critical value $c_k$ which is obtained as a quantile with respect to some distribution $Z_k$. In case of the regression bootstrap, for example, $Z_k$ is determined by the distribution function $\widetilde{F}^{(R)}_{m,k}$. As usual one can calculate the $p$-value $p_k$ of $|\Gamma(m,k,\gamma)|/\widehat{\sigma}_m$ with respect to $Z_k$. This is the smallest level for which the null hypothesis is rejected at point $m+k$. The monitoring stops as soon as this happens for the first time, so that the smallest level for which the sequential test rejects the null hypothesis is the **smallest** one of the $p_k$. Hence the $p$-value of a sequential test is given by $\min_{1\leqslant k<N(m)+1} p_k$.

**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 Silverman's rule of thumb (Silverman (1986, 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 occurred. In the plots given here, we use the specific alternative $\boldsymbol{d}_m = (1,1)^T$. This visualizes $\mathcal{C}.3$.

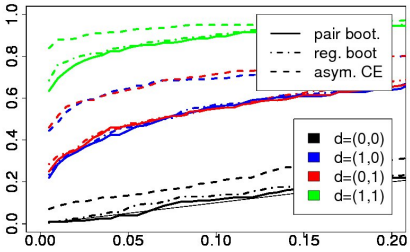
Note that only a combination of the three criteria can result in a reliable judgment of the quality of the test, and the emphasize on the criteria may also depend on the application. For example the actual power is higher if the actual level is higher, so that two tests are comparable in terms of their power only if the true size (not the nominal one) is equal. Concerning the estimated density of the run length it is important to note that it is based only on those simulations where the null hypothesis was indeed rejected. The percentage of rejected samples can be found in the SPC-plot right next to it (green line at nominal 5% level) and needs to be taken into account.

For the simulation study we use a model for $p = 2$, the results for $p = 1$ can be found in Kirch [14]. The following model is considered

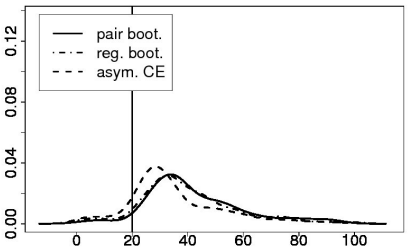
$$Y(i) = x_2(i) + d_0 1_{\{i>k^\circ\}} + d_1 1_{\{i>k^\circ\}} x_2(i) + \epsilon(i)$$

with parameters

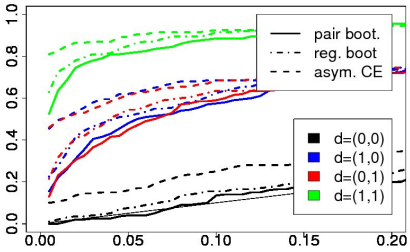
- $x_2(i)$ i.i.d. $U[0,2]$ (Scenario 1), $x_2(i) = 1+\widetilde{x}(i)$, where $\widetilde{x}(\cdot)$ is an AR(1) process with



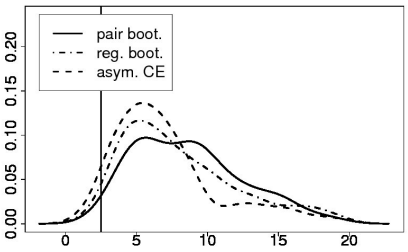

(a) SPC: $m = 10$, $N = 10$, $k^\circ = 2m$



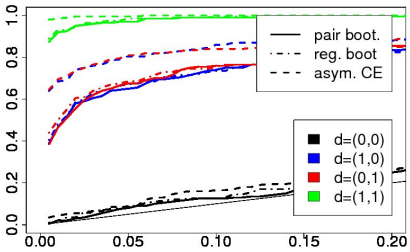

(b) RL: $m = 10$, $N = 10$, $k^\circ = 2m$



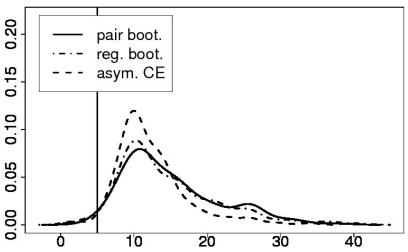

(c) SPC: $m = 10$, $N = 2$, $k^\circ = m/4$



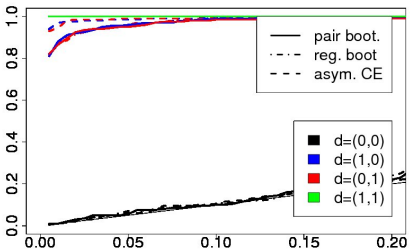

(d) RL: $m = 10$, $N = 2$, $k^\circ = m/4$



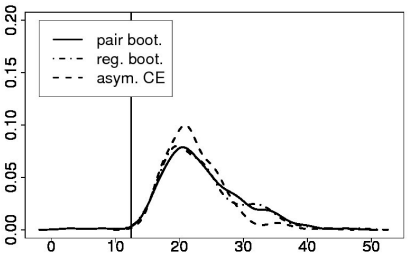

(e) SPC: $m = 20$, $N = 2$, $k^\circ = m/4$




(f) RL: $m = 20$, $N = 2$, $k^\circ = m/4$

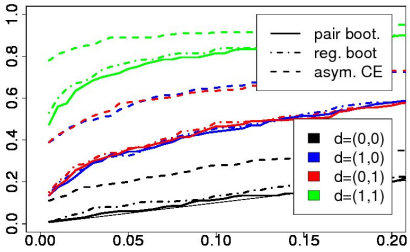



(g) SPC: $m = 50$, $N = 2$, $k^\circ = m/4$

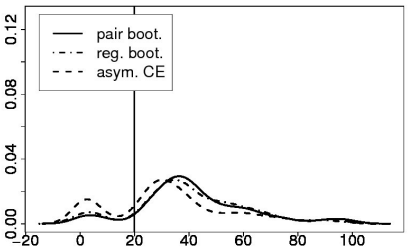



(h) RL: $m = 50$, $N = 2$, $k^\circ = m/4$

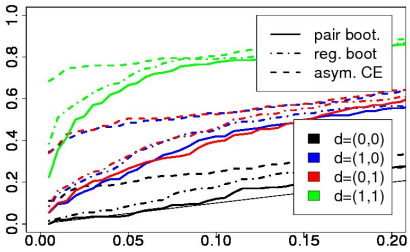
Figure 5.1: Size-power curves and plots of estimated density: Scenario 1, centered exponential errors, $\gamma = 0$, in RL: $d = (1, 1)$, $\alpha = 0.5$.



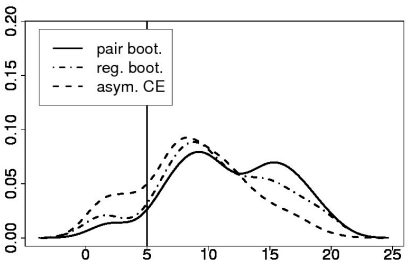

(a) SPC: $m = 10$, $N = 10$, $k^\circ = 2m$




(b) RL: $m = 10$, $N = 10$, $k^\circ = 2m$




(c) SPC: $m = 10$, $N = 2$, $k^\circ = m/2$




(d) RL: $m = 10$, $N = 2$, $k^\circ = m/2$

Figure 5.2: Size-power curves and plots of estimated density: Scenario 1, centered exponential errors, $\gamma = 0.49$, in RL: $d = (1, 1)$, $\alpha = 0.5$.

$U[-1, 1]$ distributed innovations and coefficients $-0.5, 0.5$ respectively (Scenarios 2 resp. 3).

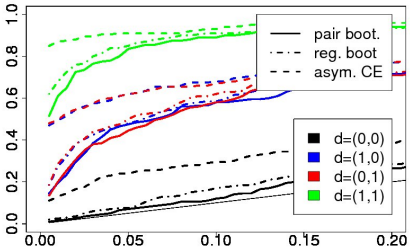
- $\boldsymbol{d}_m^T = (d_0, d_1) = (0, 1), (1, 0), (1, 1)$, i.e. changes in the slope resp. intercept only as well as in both
- $m = 10, 20, 50$
- $N(m) = Nm$ with $N = 1, 2, 5, 10$
- $k^\circ = \lfloor \vartheta m \rfloor$ with $\vartheta = 0.25, 0.5, 2, 5$
- standard normally distributed errors and centered exponentially distributed errors
- $\gamma = 0, 0.49$

The following misspecified models were also considered (each with $x_2(\cdot)$ i.i.d. $U[0, 2]$):

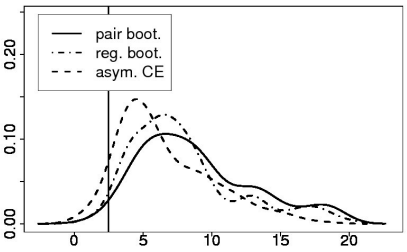
$$Y(i) = x_2(i) + 0.1\, x_2(i)^2 + d_0 1_{\{i>k^\circ\}} + d_1 1_{\{i>k^\circ\}} x_2(i) + \epsilon(i) \quad \text{(Scenario 4)}$$
$$Y(i) = x_2(i)^2 + d_0 1_{\{i>k^\circ\}} + d_1 1_{\{i>k^\circ\}} x_2(i) + \epsilon(i) \quad \text{(Scenario 5)}$$

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 authors (pdf-File, 51 p., 9 MB).

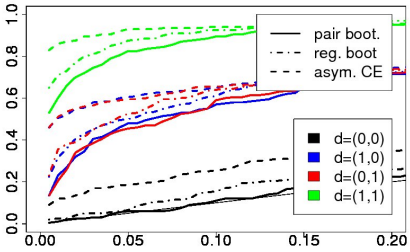
For the pair and regression bootstrap we use $L = m/5$, $M = 5$, furthermore the bootstraps are based on 500 bootstrap samples while the plots are based on 200 repetitions of the procedure. These relatively small numbers of repetitions are necessary due to rather long computation times even when using the sequential bootstrap based on the convex combination. In Figures 5.1-5.3 some selected size-power curves and density plots of the run length can be found.



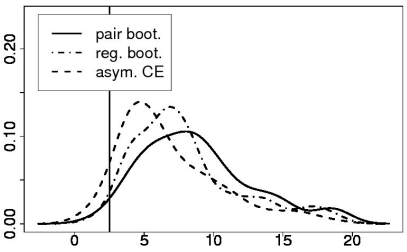

(a) SPC: Scenario 2



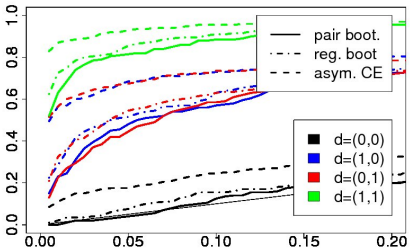

(b) RL: Scenario 2



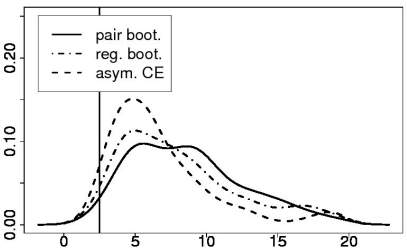

(c) SPC: Scenario 3



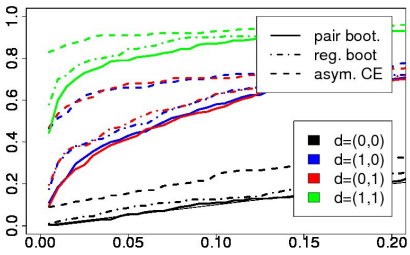

(d) RL: Scenario 3



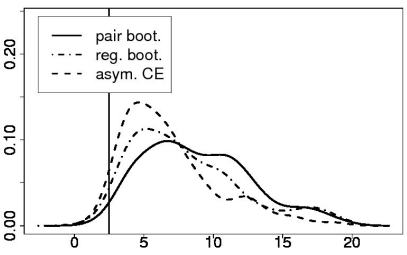

(e) SPC: Scenario 4




(f) RL: Scenario 4




(g) SPC: Scenario 5




(h) RL: Scenario 5

Figure 5.3: Size-power curves and plots of estimated density: centered exponential errors, $\gamma = 0$, $m = 10$, $N = 2$, $k^\circ = m/4$, in RL: $d = (1, 1)$, $\alpha = 0.5$.

It can be concluded from the simulations that both bootstrap methods perform better than the asymptotic closed-end procedure for two reasons:

- The bootstrap methods hold the level much better.
- The run-time of the bootstrap methods is a bit longer, however a much smaller percentage of all rejections takes place before the change. This can especially be seen in Figure 5.2 and may be caused partly by the smaller level of the bootstrap methods.
- All three methods still work well under the misspecified scenarios 3 and 4 (cf. Figure 5.3).
- All three methods become approximately equivalent for $m \geqslant 50$ (for $p = 2$).
- The bootstrap methods already work well for a very small historic data length of $m = 10$ (for $p = 2$).

Concerning a comparison of the two bootstrap methods the following can be noticed:

- The pair bootstrap holds the level better consequently has a somewhat smaller power and higher run-length. Interestingly, this remains true even if the regressors are correlated (Figure 5.3, Scenarios 2 and 3), where one would expect the regression bootstrap to be better.
- The bootstrap methods become very close for $m \geqslant 20$.

# 6 Proofs of Section 3

The following lemma summarizes some results on $\mathbf{C}$ and $\mathbf{c}_1(m, k, \ell)$. It follows immediately from Lemma 5.1 in Horváth et al. [10].

**Lemma 6.1.** *Under the Assumption $\mathcal{A}.1$ (ii) it holds as $m \to \infty$*

$$a) \qquad \left\| m\mathbf{C}_m^{-1} - \mathbf{C}^{-1} \right\|_\infty = O(m^{-\rho}) \quad P - a.s.$$

$$b) \qquad \sup_{\ell \geqslant 1} \sup_{k \geqslant 1} \frac{\|\mathbf{c}_1(m,k,\ell) - \ell \mathbf{c}_1\|_\infty}{(m+\ell)^{1-\rho} + \ell m^{-\rho}} = O(1) \qquad P - a.s.$$

Our aim is to prove that the bootstrap critical values are uniformly asymptotically correct under the null hypothesis and bounded under alternatives (see equations (6.24) resp. (6.25) below).

In view of the following lemma it is clear that for this it is sufficient to prove the correct asymptotic behavior of $\sup_k F_{m,k}^{(R*)}(x)$.

**Lemma 6.2.** *Let $c$, $c_k(m)$ be such that $P(Y > c) = \alpha$ respectively $P^*_{m,k}(Y_k(m) > c_k(m)) \leqslant \alpha$ 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 $c$.*

*a) Moreover let for all $x$ in $K$ (as $m \to \infty$)*

$$\sup_{1 \leqslant k < \infty} \left| P^*_{m,k}(Y_k(m) \leqslant x) - P\left(Y \leqslant x\right) \right| \to 0 \qquad P - a.s. \tag{6.1}$$

*Then, as* $m \to \infty$,

$$\sup_{1 \leqslant k < \infty} |c_k(m) - c| \to 0 \qquad P - a.s. \tag{6.2}$$

*b) If instead we only have that for each* $\epsilon > 0$ *there exists a constant* $A = A(\epsilon) > 0$, *s.t.*

$$\sup_{1 \leqslant k < \infty} \left| P^*_{m,k}(Y_k(m) \geqslant A) \right| \leqslant \epsilon + o(1) \qquad P - a.s. \tag{6.3}$$

*then*

$$\sup_{1 \leqslant k < \infty} |c_k(m)| = O(1) \quad P - a.s. \tag{6.4}$$

**Proof.** For a) we refer to the proof of Kirch [14], Lemma A.1. Concerning b), consider the set of all $\omega \in M$ with $P(M) = 1$ and such that (6.3) holds. We prove (6.4) for all $\omega \in M$ by contradiction. If (6.4) does not hold, we find a subsequence $\beta(\cdot)$ and a function $f$, such that $c_{f(\beta(m))}(\beta(m)) \to \infty$. On the other hand since $P^*_{m,f(\beta(m))}\left(Y_{f(\beta(m))}(m) > c_{f(\beta(m))}\right) \leqslant \alpha$ we get by the minimality of $c_{f(\beta(m))}(m)$ by (6.3) that

$$c_{f(\beta(m))}(m) \leqslant A(\alpha/2),$$

which is a contradiction. ■

Note that $F^{(R*)}_{m,k}$ is determined mainly by the distribution of $(\ell = 1, \ldots, N(m))$

$$\begin{aligned} &g(m,\ell,\gamma)\,\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1), \ldots, e^*_{m,k}(m+\ell)) \\ &= \sum_{i=m+1}^{m+\ell} \widehat{e}_{m,k}(U_{m,k}(i)) - \boldsymbol{c}_1(m,k,\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j) \widehat{e}_{m,k}(U_{m,k}(i)) \end{aligned}$$

and the bootstrap variance (3.5). Note that

$$\begin{aligned} \widehat{e}_{m,k}(i) =& e(i) - \boldsymbol{x}(i)^T \boldsymbol{C}_{m+k}^{-1} \sum_{j=1}^{m+k} \boldsymbol{x}(j) e(j) \\ &+ 1_{\{i > m+k^\circ\}} \boldsymbol{x}(i)^T \boldsymbol{d}_m - 1_{\{k > k^\circ\}} \boldsymbol{x}(i)^T \boldsymbol{C}_{m+k}^{-1} \boldsymbol{C}_{k^\circ,k} \boldsymbol{d}_m, \end{aligned} \tag{6.5}$$

and for $k > k^\circ$

$$\boldsymbol{C}_{k^\circ,k} = \sum_{i=m+k^\circ+1}^{m+k} \mathbf{x}(i)\mathbf{x}(i)^T = \mathbf{C}_k - \mathbf{C}_{k^\circ}.$$

From this we can decompose $g(m,\ell,\gamma)\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1), \ldots, e^*_{m,k}(m+\ell))$ as follows:

$$\begin{aligned} &g(m,\ell,\gamma)\,\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1), \ldots, e^*_{m,k}(m+\ell)) \\ &= I_1(m,k,\ell) + I_2(m,k,\ell) + I_3(m,k,\ell) + I_4(m,k,\ell) + I_5(m,k,\ell) + I_6(m,k,\ell), \end{aligned}$$

where

$$I_1(m,k,\ell) = \sum_{i=m+1}^{m+\ell} e(U_{m,k}(i)),$$

$$I_2(m,k,\ell) = -\boldsymbol{c}_1(m,k,\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j) e(U_{m,k}(j)),$$

$$I_3(m,k,\ell) = \sum_{i=m+1}^{m+\ell} \boldsymbol{x}(U_{m,k}(i))^T \boldsymbol{C}_{m+k}^{-1} \sum_{j=1}^{m+k} \boldsymbol{x}(j) e(j) - \frac{\ell}{m+k} \sum_{i=1}^{m+k} e(i),$$

$$I_4(m,k,\ell) = -\boldsymbol{c}_1(m,k,\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j) \boldsymbol{x}(U_{m,k}(j))^T \boldsymbol{C}_{m+k}^{-1} \sum_{v=1}^{m+k} \boldsymbol{x}(v) e(v) + \frac{\ell}{m+k} \sum_{i=1}^{m+k} e(i),$$

$$I_5(m,k,\ell) = \sum_{i=m+1}^{m+\ell} \Big( 1_{\{U_{m,k}(j) > m+k^\circ\}} \boldsymbol{x}(U_{m,k}(j))^T \boldsymbol{d}_m - 1_{\{k>k^\circ\}} \boldsymbol{x}(U_{m,k}(j))^T \boldsymbol{C}_{m+k}^{-1} \boldsymbol{C}_{k^\circ,k} \boldsymbol{d}_m \Big)$$

$$I_6(m,k,\ell) = -\boldsymbol{c}_1(m,k,\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j) \Big( 1_{\{U_{m,k}(j) > m+k^\circ\}} \boldsymbol{x}(U_{m,k}(j))^T \boldsymbol{d}_m - 1_{\{k>k^\circ\}} \boldsymbol{x}(U_{m,k}(j))^T \boldsymbol{C}_{m+k}^{-1} \boldsymbol{C}_{k^\circ,k} \boldsymbol{d}_m \Big).$$

In order to prove (6.1) under $H_0$ resp. (6.3) under $H_1$, we show that $g(m,\ell,\gamma)\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1),\ldots,e^*_{m,k}(m+\ell))$ is asymptotically determined by $I_1(m,k,\ell)$ respectively $I_2(m,k,\ell)$. Precisely, the following lemma shows that $I_j(m,k,\ell)$, $j=3,4$, converge uniformly to 0 and the terms $I_j(m,k,\ell)$, $j=5,6$, which are nonzero only under alternatives, are uniformly bounded.

**Lemma 6.3.** *Let* (1.1), *Assumption* $\mathcal{A}.1$, *and* (2.1) *hold true and either* $H_0$ *or* $\boldsymbol{d}_m = O(1)$.

*a) Then for all* $\epsilon > 0$ *it holds:*

$$(i) \qquad \sup_{1 \leqslant k < \infty} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|I_3(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon \right) \to 0 \quad P-a.s.,$$

$$(ii) \qquad \sup_{1 \leqslant k < \infty} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|I_4(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon \right) \to 0 \quad P-a.s.$$

*b) Under* $H_0$ *it holds that* $I_j(m,k,\ell) = 0$, $j = 5,6$, *under local alternatives, i.e. if* $\boldsymbol{d}_m = o(1)$, *it holds for all* $\epsilon > 0$ *that*

$$(i) \qquad \sup_{1 \leqslant k < \infty} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|I_5(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon \right) \to 0 \quad P-a.s.,$$

$$(ii) \qquad \sup_{1 \leqslant k < \infty} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|I_6(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon \right) \to 0 \quad P-a.s.$$

*c) For alternatives, for which only* $\boldsymbol{d}_m = O(1)$, *we get only the following weaker assertion: For every* $\epsilon > 0$ *there exists* $A > 0$ *such that*

$$(i) \qquad \sup_{1 \leqslant k < \infty} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|I_5(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant A \right) \leqslant \epsilon + o(1) \quad P-a.s.,$$

$$(ii) \qquad \sup_{1 \leqslant k < \infty} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|I_6(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant A \right) \leqslant \epsilon + o(1) \quad P-a.s.$$

**Proof.** All terms are sums of i.i.d. random vectors. Therefore, it suffices in all cases to calculate the variance matrices and to apply the Hájek-Rényi or Markov inequality. By direct calculations

$$E^*_{m,k}\boldsymbol{x}(U_{m,k}(i)) = \frac{1}{m+k}\sum_{j=1}^{m+k}\boldsymbol{x}(j),$$

and since

$$\sum_{j=1}^{m+k}\boldsymbol{x}(j)^T\boldsymbol{C}^{-1}_{m+k}\mathbf{x}(v) = (1,0,\ldots,0)\mathbf{x}(v) = 1$$

we obtain

$$\mathrm{E}^*_{m,k}\left(\boldsymbol{x}(U_{m,k}(i))^T\boldsymbol{C}^{-1}_{m+k}\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right) = \frac{1}{m+k}\sum_{j=1}^{m+k}e(j),$$

showing that $I_3(m,k,\ell)$ is centered. Moreover

$$\mathrm{E}^*_{m,k}\left(\boldsymbol{x}(U_{m,k}(i))\boldsymbol{x}(U_{m,k}(i))^T\right) = \frac{1}{m+k}\boldsymbol{C}_{m+k},$$

hence (by $\mathrm{var}(Z) \leqslant \mathrm{E}(Z^2)$)

$$\begin{aligned}
&\mathrm{var}^*_{m,k}\left(\boldsymbol{x}(U_{m,k}(i))^T\boldsymbol{C}^{-1}_{m+k}\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right)\\
&\leqslant \frac{1}{m+k}\left(\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right)^T\boldsymbol{C}^{-1}_{m+k}\boldsymbol{C}_{m+k}\boldsymbol{C}^{-1}_{m+k}\left(\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right)\\
&= \frac{1}{m+k}\left(\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right)^T\boldsymbol{C}^{-1}_{m+k}\left(\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right).
\end{aligned}$$

Standard decoupling arguments yield

$$\sup_{k\geqslant 1}\frac{1}{m+k}\left|\sum_{j=1}^{m+k}\mathbf{x}(j)e(j)\right| = o(1) \qquad P-a.s., \tag{6.6}$$

since conditioned on $\{\mathbf{x}(\cdot)\}$ the sequence fulfills the Kolmogorov condition for a strong LLN. This together with Lemma 6.1 shows that, as $m\to\infty$,

$$\sup_{1\leqslant k<\infty}\frac{1}{m+k}\left(\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right)^T\boldsymbol{C}^{-1}_{m+k}\left(\sum_{j=1}^{m+k}\boldsymbol{x}(j)e(j)\right) = o(1) \quad P-a.s.$$

Denote

$$Z_{m,k}(U_{m,k}(j)) = \boldsymbol{x}(U_{m,k}(j))^T\boldsymbol{C}^{-1}_{m+k}\sum_{v=1}^{m+k}\boldsymbol{x}(v)e(v) - \frac{1}{m+k}\sum_{j=1}^{m+k}e(j). \tag{6.7}$$

Conditionally $\{Z_{m,k}(U_{m,k}(j)\}$ are i.i.d. random variable with

$$\begin{aligned}&\mathrm{E}^*_{m,k}\, Z_{m,k}(U_{m,k}(1)) = 0,\\ &\sup_k \mathrm{var}^*_{m,k}\, Z_{m,k}(U_{m,k}(1)) = o(1) \qquad P-a.s.\end{aligned} \tag{6.8}$$

We start by proving assertion a)(i): For some $D_1 > 0$

$$g(m,\ell,\gamma) \geqslant \begin{cases} D_1 m^{1/2-\gamma}\, \ell^{\gamma}, & \ell \leqslant m,\\ D_1 m^{-1/2}\, \ell, & \ell > m,\end{cases} \tag{6.9}$$

yielding for some $D_2 > 0$

$$\sum_{l=1}^{N(m)} \frac{1}{g^2(m,\ell,\gamma)} \leqslant D_1^{-2} m^{-1+2\gamma} \sum_{\ell=1}^{m} \frac{1}{\ell^{2\gamma}} + D_1^{-2} m \sum_{\ell=m+1}^{N(m)} \frac{1}{\ell^2} \leqslant D_2. \tag{6.10}$$

Then, by the Hájek-Rényi inequality for any $\epsilon > 0$

$$P^*_{m,k}\left(\max_{1\leqslant \ell < N(m)+1} \frac{|I_3(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon\right) \leqslant \epsilon^{-2} D_2\ \mathrm{var}^*_{m,k}\, Z_{m,k}(U_{m,k}(1)) \to 0 \qquad P-a.s.$$

uniformly in $k$ which finishes the proof of a)(i) by (6.8).

Now we prove a) (ii). Notice that $I_4(m,k,\ell)$ can be expressed as the product of two terms one of them is (conditionally) nonrandom and depends on $\ell$ while the other one is (conditionally) random and does not depend on $\ell$. We will make use of this fact, which is why the proof differs from the one of a)(i). By $\mathbf{x}(i)^T \mathbf{C}_m^{-1} \sum_{j=1}^m \mathbf{x}(j) = 1$ it holds

$$I_4(m,k,\ell) = -\mathbf{c}_1(m,k,\ell)^T \mathbf{C}_m^{-1} \sum_{j=1}^{m} \mathbf{x}(j) Z_{m,k}(U_{m,k}(j)).$$

Denote by $\mathbf{u}_j$ the $j$th unit vector i.e. the $p$-dimensional vector $\mathbf{u}_j = (u_{1,j},\ldots,u_{p,j})^T$ with $\mathbf{u}_{i,j} = 1_{\{i=j\}}$, then by Lemma 6.1 a)

$$\begin{aligned}&\mathrm{E}^*_{m,k}\left(\sqrt{m}\mathbf{u}_j^T \mathbf{C}_m^{-1} \sum_{i=1}^{m} \mathbf{x}(i) Z_{m,k}(U_{m,k}(i))\right)^2\\ &= \mathrm{var}^*_{m,k}\left(\sqrt{m}\sum_{i=1}^{m} \mathbf{u}_j^T \mathbf{C}_m^{-1}\mathbf{x}(i) Z_{m,k}(U_{m,k}(i))\right)\\ &= \left(\mathrm{var}^*_{m,k}\, Z_{m,k}(U_{m,k}(1))\right)\ \mathbf{u}_j^T\, m\mathbf{C}_m^{-1} \sum_{i=1}^{m} \mathbf{x}(i)\mathbf{x}(i)^T\, \mathbf{C}_m^{-1}\, \mathbf{u}_j\\ &= \left(\mathrm{var}^*_{m,k}\, Z_{m,k}(U_{m,k}(1))\right)\ \mathbf{u}_j^T\, (\mathbf{C}^{-1} + o(1))\, \mathbf{u}_j = o(1) \qquad P-a.s.\end{aligned} \tag{6.11}$$

uniformly in $k$. By Lemma 6.1 b) in addition to (6.9) we additionally get for some $D_3 > 0$

$$\sup_{k\geqslant 1} \sup_{1\leqslant \ell < N(m)+1} \left\| \frac{\mathbf{c}_1(m,k,\ell)^T}{\sqrt{m}\, g(m,\ell,\gamma)} \right\|_\infty \leqslant D_3 + o(1) \qquad P-a.s. \tag{6.12}$$

Together this yields that

$$\mathrm{E}^*_{m,k}\left(\max_{1\leqslant\ell<N(m)+1}\frac{|I_4(m,k,\ell)|}{g(m,\ell,\gamma)}\right)^2=o(1)\qquad P-a.s.,$$

which gives the assertion by the Markov inequality. Now, we prove b) and c). First note that it suffices to consider $k>k^\circ$, since for $k\leqslant k^\circ$ it holds $I_j(m,k,\ell)=0$, $j=5,6$. Denote

$$\widetilde{Z}_{m,k}(U_{m,k}(i))=1_{\{U_{m,k}(i)>m+k^\circ\}}\boldsymbol{x}^T(U_{m,k}(i))\boldsymbol{d}_m-1_{\{k>k^\circ\}}\boldsymbol{x}^T(U_{m,k}(i))\boldsymbol{C}^{-1}_{m+k}\boldsymbol{C}_{k^\circ,k}\boldsymbol{d}_m.$$

Direct calculations give

$$\begin{aligned}&\mathrm{E}^*_{m,k}\,\widetilde{Z}_{m,k}(U_{m,k}(i))=0,\\&\mathrm{E}^*_{m,k}\left(\widetilde{Z}_{m,k}(U_{m,k}(i))\right)^2=\frac{1}{m+k}\boldsymbol{d}_m^T(\boldsymbol{C}_{k^\circ,k}-\boldsymbol{C}_{k^\circ,k}\boldsymbol{C}^{-1}_{m+k}\boldsymbol{C}_{k^\circ,k})\boldsymbol{d}_m\\&\leqslant\frac{1}{m+k}\boldsymbol{d}_m^T\boldsymbol{C}_{k^\circ,k}\boldsymbol{d}_m\leqslant\frac{1}{m+k}\boldsymbol{d}_m^T\boldsymbol{C}_{m+k}\boldsymbol{d}_m\leqslant\boldsymbol{d}_m^T\boldsymbol{C}\boldsymbol{d}_m\,(1+o(1))\qquad P-a.s.\end{aligned}\tag{6.13}$$

uniformly in $k$ by Lemma 6.1 a).

Assertions b) (i) and c) (i) follow now by the Hájek-Rényi inequality and (6.10).

Furthermore ($P-a.s.$)

$$\mathrm{E}^*_{m,k}\left(\sqrt{m}\,\mathbf{u}_j^T\mathbf{C}_m^{-1}\sum_{j=1}^m\mathbf{x}(j)\widetilde{Z}_{m,k}(U_{m,k}(j))\right)^2\leqslant\boldsymbol{d}_m^T\mathbf{C}\boldsymbol{d}_m\,\mathbf{u}_j^T\mathbf{C}^{-1}\mathbf{u}_j\,(1+o(1))\tag{6.14}$$

uniformly in $k$. This, in addition to (6.12), yields b) (ii) and c) (ii) by the Markov inequality. ■

The next lemma allows us to replace $I_2(m,k,\ell)$ by a simpler expression.

**Lemma 6.4.** *Let* (1.1), *Assumption* $\mathcal{A}.1$, *and* (2.1) *hold true. Then for all* $\epsilon>0$ *it holds*

$$\sup_{1\leqslant k<\infty}P^*_{m,k}\left(\max_{1\leqslant\ell<N(m)+1}\frac{|I_2(m,k,\ell)-\frac{-\ell}{m}\sum_{j=1}^m e(U_{m,k}(j))|}{g(m,\ell,\gamma)}\geqslant\epsilon\right)\to0\qquad P-a.s.$$

**Proof.** The proof is a slight modification of Lemma 5.2 in Horváth et al. [10]. Denote

$$\bar{e}_{m,k}=\frac{1}{m+k}\sum_{i=1}^{m+k}e(i),\qquad\widehat{\sigma}^2_{m,k}=\frac{1}{m+k}\sum_{i=1}^{m+k}\left(e(i)-\bar{e}_{m,k}\right)^2.\tag{6.15}$$

By Assumption $\mathcal{A}.1$ and the law of iterated logarithm we get uniformly in $k$

$$\sup_k|m^{1/2-(\rho-\gamma)}\bar{e}_{m,k}|\to0\qquad P-a.s.,\qquad\sup_k|\widehat{\sigma}^2_{m,k}-\sigma^2|\to0\qquad P-a.s.,\tag{6.16}$$

since by assumption $\rho-\gamma>0$ for $\rho$ from Assumption $\mathcal{A}.1$. Let $\mathbf{u}_j$ denote again the $j$th unit vector. Then, we get uniformly in $k$

$$\begin{aligned}
&\mathrm{E}^*_{m,k}\left(\frac{1}{m^{1/2+\rho-\gamma}}\mathbf{u}_j^T\sum_{i=1}^m \mathbf{x}(i)e(U_{m,k}(i))\right)^2\\
&=\frac{1}{m^{1+2(\rho-\gamma)}}\sum_{i=1}^m(\mathbf{u}_j^T\mathbf{x}(i))^2\,\mathrm{var}^*\left(e(U_{m,k}(1))\right)+\left(\frac{\mathbf{u}_j^T}{m^{1/2+\rho-\gamma}}\sum_{i=1}^m\mathbf{x}(i)\,\mathrm{E}^*(e(U_{m,k}(1))\right)^2\\
&=\frac{1}{m^{1+2(\rho-\gamma)}}\mathbf{u}_j^T\mathbf{C}_m\mathbf{u}_j\,\widehat{\sigma}^2_{m,k}+(\mathbf{u}_j^T\mathbf{c}_1+o(1))^2\left(m^{1/2-(\rho-\gamma)}\bar{e}_{m,k}\right)^2\\
&=o(1)\qquad P-a.s.
\end{aligned}\tag{6.17}$$

By Lemma 6.1 and (6.9) we get

$$\begin{aligned}
&\sup_k\sup_{1\leqslant\ell<N(m)+1}\left\|\frac{\mathbf{c}_1(m,k,\ell)^T\,mC_m^{-1}-\ell\mathbf{c}_1^T\mathbf{C}^{-1}}{m^{1/2-\rho+\gamma}\,g(m,\ell,\gamma)}\right\|_\infty\\
&=O(1)\sup_{1\leqslant\ell<N(m)+1}\left|\frac{(m+\ell)^{1-\rho}+\ell m^{-\rho}}{m^{1/2-\rho+\gamma}\,g(m,\ell,\gamma)}\right|=O(1)\qquad P-a.s.
\end{aligned}$$

Since $\mathbf{c}_1^T\mathbf{C}^{-1}\sum_{i=1}^m\mathbf{x}(j)e(U_{m,k}(i))=\sum_{i=1}^m e(U_{m,k}(i))$ we obtain the assertion by an application of the Markov inequality. ■

We are now ready to state the asymptotics of $\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1),\ldots,e^*_{m,k}(m+\ell))$.

**Lemma 6.5.** *Let* (1.1), *Assumption* $\mathcal{A}.1$, *and* (2.1) *hold true and either* $H_0$ *or* $\boldsymbol{d}_m=O(1)$, $N(m)$ *is as in Theorem 2.1. Let* $\widehat{\sigma}^2_{m,k}$ *as in* (6.15).

*a) Under* $H_0$ *and for local alternatives* $\boldsymbol{d}_m=o(1)$ *it holds*

$$\begin{aligned}
\sup_{1\leqslant k<N(m)+1}\Bigg|P^*_{m,k}\left(\frac{1}{\widehat{\sigma}_{m,k}}\sup_{1\leqslant\ell<N(m)+1}\frac{\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1),\ldots,e^*_{m,k}(m+\ell))}{g(m,\ell,\gamma)}\leqslant x\right)\\
-P\left(\sup_{1\leqslant k<N(m)+1}\frac{|W_1(k/m)-k/mW_2(1)|}{(1+k/m)(k/(k+m))^\gamma}\leqslant x\right)\Bigg|\to 0\qquad P-a.s.
\end{aligned}$$

*b) Under* $H_1$ *for every* $\epsilon>0$ *there exists a constant* $A>0$ *such that* $(P-a.s.)$

$$\sup_{1\leqslant k<N(m)+1}\left|P^*_{m,k}\left(\frac{1}{\widehat{\sigma}_{m,k}}\sup_{1\leqslant\ell<N(m)+1}\frac{\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1),\ldots,e^*_{m,k}(m+\ell))}{g(m,\ell,\gamma)}\geqslant A\right)\right|\leqslant\epsilon+o(1).$$

**Proof.** The proof of Theorem 2.3 in Kirch [14] shows that (note that this corresponds to the null hypothesis there)

$$\begin{aligned}
\sup_{1\leqslant k<N(m)+1}\Bigg|P^*_{m,k}\left(\frac{1}{\widehat{\sigma}_{m,k}}\sup_{1\leqslant\ell<N(m)+1}\frac{\left|\sum_{i=m+1}^{m+\ell}\left(e(U_{m,k}(i))-\frac{1}{m}\sum_{j=1}^m e(U_{m,k}(j))\right)\right|}{g(m,\ell,\gamma)}\leqslant x\right)\\
-P\left(\sup_{1\leqslant k<N(m)+1}\frac{|W_1(k/m)-k/mW_2(1)|}{(1+k/m)(k/(k+m))^\gamma}\leqslant x\right)\Bigg|\to 0\qquad P-a.s.
\end{aligned}$$

Putting this together with Lemmas 6.3 and 6.4, as well as (6.16) yields the assertion. ■

Before we finally deal with the bootstrapped variance, we need a small auxiliary lemma which will also be crucial for the proof of the pair bootstrap.

**Lemma 6.6.** *Let* (1.1) *and Assumption* $\mathcal{A}.1$ *hold true. For any* $\frac{1}{2} < \xi < 1$ *and any* $\epsilon > 0$ *(*$P-a.s.$*)*

$$a) \qquad \sup_{k \geqslant 1} P^*_{m,k}\left( m^{-\xi} \left\| \sum_{s=1}^{m} \mathbf{x}(U_{m,k}(s))^T e(U_{m,k}(s)) \right\|_\infty > \epsilon \right) \to 0$$

$$b) \qquad \sup_{k \geqslant 1} P^*_{m,k}\left( m^{-\xi} \left\| \sum_{s=1}^{m} \mathbf{x}(U_{m,k}(s))^T e(U_{m,k}(s)) 1_{\{U_{m,k}(s) > m+k^\circ\}} \right\|_\infty > \epsilon \right) \to 0$$

*If for the pair bootstrap additionally Assumption* $\mathcal{A}.3$ *holds, then we even get the assertions for* $\xi = \frac{1}{2}$.

**Proof.** By the von Bahr-Esseen inequality (cf. Theorem 3 in [17]) with $1/\xi$ we get for some constant $D > 0$ and for any $c > 0$

$$\begin{aligned} &\sup_{k \geqslant 1} P^*_{m,k}\left( m^{-\xi} \left\| \sum_{s=1}^{m} \boldsymbol{x}(U_{m,k}(s))^T e(U_{m,k}(s)) \right\|_\infty \geqslant c \right) \\ &\leqslant \frac{D}{c^{1/\xi}} \sup_{k \geqslant 1} \frac{1}{m+k} \sum_{j=1}^{m+k} \|\mathbf{x}(i)e(i)\|_\infty^{1/\xi} = O(1) \qquad P-a.s., \end{aligned}$$

since conditioned on $\{\mathbf{x}(\cdot)\}$ the sequence fulfills condition (1) in Theorem 5.2.1 in Chow and Teicher [6] similarly to (6.6). This proves a) but b) is analogous.

The same arguments also holds for $\xi = \frac{1}{2}$ if the stronger assumption $\mathcal{A}.3$ holds. ■

Finally we deal with the bootstrapped variance in the following lemma:

**Lemma 6.7.** *Let* (1.1) *and Assumption* $\mathcal{A}.1$ *hold true and either* $H_0$ *or* $\boldsymbol{d}_m = O(1)$. *Let* $\widehat{\sigma}^2_{m,k}$ *be as in* (6.15).

*a) Under* $H_0$ *or local alternatives (*$\boldsymbol{d}_m = o(1)$*) it holds for all* $\epsilon > 0$

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

*b) Under* $H_1$ *for every* $\epsilon > 0$ *there exists* $A > 0$ *such that*

$$\sup_{k} P^*_{m,k}\left( \left| \frac{\widehat{\sigma}_{m,k}}{\widehat{\sigma}^{(R*)}_{m,k}} \right| \geqslant A \right) \leqslant \epsilon + o(1) \qquad P-a.s.$$

**Proof.** By (6.5) it holds

$$\begin{aligned} &e^*_{m,k}(i) - \mathbf{x}(i)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \mathbf{x}(j) e^*_{m,k}(j) \\ &= J_1(m,k,i) + J_2(m,k,i) + J_3(m,k,i) + J_4(m,k,i) + J_5(m,k,i) + J_6(m,k,i), \end{aligned}$$

where

$$
\begin{aligned}
J_1(m,k,i) &= e(U_{m,k}(i)),\\
J_2(m,k,i) &= -\mathbf{x}(i)^T\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)e(U_{m,k}(j)),\\
J_3(m,k,i) &= -\mathbf{x}(U_{m,k}(i))^T\mathbf{C}_{m+k}^{-1}\sum_{j=1}^{m+k}\mathbf{x}(j)e(j)+\bar{e}_{m,k},\\
J_4(m,k,i) &= \mathbf{x}(i)^T\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)\left(\mathbf{x}(U_{m,k}(j))^T\mathbf{C}_{m+k}^{-1}\sum_{v=1}^{m+k}\mathbf{x}(v)e(v)-\bar{e}_{m,k}\right),\\
J_5(m,k,i) &= 1_{\{U_{m,k}(i)>m+k^\circ\}}\mathbf{x}(U_{m,k}(i))^T\boldsymbol{d}_m - 1_{\{k>k^\circ\}}\mathbf{x}(U_{m,k}(i))^T\mathbf{C}_{m+k}^{-1}\mathbf{C}_{k^\circ,k}\boldsymbol{d}_m,\\
J_6(m,k,i) &= -\mathbf{x}(i)^T\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)\mathbf{x}(U_{m,k}(j))^T\Big(1_{\{U_{m,k}(j)>m+k^\circ\}}-1_{\{k>k^\circ\}}\mathbf{C}_{m+k}^{-1}\mathbf{C}_{k^\circ,k}\Big)\boldsymbol{d}_m,
\end{aligned}
$$

where $\bar{e}_{m,k}$ is as in (6.15). Note that $J_5(m,k,i)$ and $J_6(m,k,i)$ are equal to 0 under the null hypothesis and under alternatives for $k \leqslant k^\circ$.

The following relations hold true for any fixed $\epsilon > 0$ as $m \to \infty$.

By Lemma A.3 and the proof of Theorem 2.3 in Kirch [14] (this corresponds to the null hypothesis there) it holds

$$
\sup_{k\geqslant 1} P_{m,k}^*\left(\left|\frac{\frac{1}{m}\sum_{i=1}^m(e(U_{m,k}(i))-\bar{e}_{m,k})^2}{\widehat{\sigma}_{m,k}^2}-1\right|\geqslant\epsilon\right)\to 0 \qquad P-a.s.,
$$

which in addition to (6.16) yields

$$
\sup_{k\geqslant 1} P_{m,k}^*\left(\left|\frac{\frac{1}{m-p}\sum_{i=1}^m J_1^2(m,k,i)}{\widehat{\sigma}_{m,k}^2}-1\right|\geqslant\epsilon\right)\to 0 \qquad P-a.s. \tag{6.18}
$$

Denote by $\mathbf{u}_j$ again the $j$th unit vector. By Lemma 6.1 and (6.17) it holds by the Cauchy-Schwarz inequality

$$
\begin{aligned}
&\mathrm{E}_{m,k}^*\left(\frac{1}{m}\sum_{i=1}^m J_2^2(m,k,i)\right)\\
&=\mathrm{E}_{m,k}^*\left(\frac{1}{m}\sum_{i=1}^m\left(\mathbf{x}(i)^T\mathbf{C}_m^{-1}\sum_{j=1}^m\mathbf{x}(j)e(U_{m,k}(j))\right)^2\right)\\
&\leqslant p^2\|mC_m^{-1}\|_\infty\max_{j=1,\ldots,p}\mathrm{E}_{m,k}^*\left(\frac{\mathbf{u}_j^T}{m}\sum_{i=1}^m\mathbf{x}(i)e(U_{m,k}(i))\right)^2\to 0 \qquad P-a.s.,
\end{aligned}
$$

which yields by an application of the Markov inequality

$$
\sup_{k\geqslant 1} P_{m,k}^*\left(\frac{1}{m-p}\sum_{i=1}^m J_2^2(m,k,i)\geqslant\epsilon\right)\to 0 \qquad P-a.s. \tag{6.19}
$$

Noting that $J_3(m,k,i) = -Z_{m,k}(U_{m,k}(i))$ as in the proof of Lemma 6.3, hence by an application of the Markov inequality in addition to (6.8)

$$\sup_{k\geqslant 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^{m} J_3^2(m,k,i) \geqslant \epsilon\right) \to 0 \qquad P-a.s. \tag{6.20}$$

Similarly by (6.11)

$$\begin{aligned}
&\sup_{k\geqslant 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^{m} J_4^2(m,k,i) \geqslant \epsilon\right)\\
&\leqslant \sup_{k\geqslant 1}\frac{1}{\epsilon}\,\mathrm{E}^*_{m,k}\left(\frac{1}{m-p}\sum_{j=1}^{m}\mathbf{x}(j)^T Z_{m,k}(U_{m,k}(j))\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)Z_{m,k}(U_{m,k}(j))\right)\\
&\leqslant \frac{1}{\epsilon}\left\|\frac{1}{m}\mathbf{C}_m\right\|_\infty \frac{p^2}{m-p}\sup_{k\geqslant 1}\sup_{j=1,\dots,p}\mathrm{E}^*_{m,k}\left(\sqrt{m}\mathbf{u}_j^T\mathbf{C}_m^{-1}\sum_{v=1}^{m}\mathbf{x}(v)Z_{m,k}(U_{m,k}(v))\right)^2\\
&\to 0 \qquad P-a.s.
\end{aligned} \tag{6.21}$$

Putting (6.18) to (6.20) together with (6.16) yields assertion a) under $H_0$.

Now we prove the assertions under alternatives. First, note that it is sufficient to consider $k > k^\circ$, since otherwise $J_5$ and $J_6$ are equal to 0. First we prove that $J_6$ is negligible: By (6.14) it holds

$$\begin{aligned}
&\sup_{k\geqslant 1} P^*_{m,k}\left(\frac{1}{m}\sum_{i=1}^{m} J_6^2(m,k,i) \geqslant \epsilon\right)\\
&\leqslant \frac{1}{\epsilon}\frac{p^2}{m-p}\left\|\frac{1}{m}\mathbf{C}_m\right\|_\infty \sup_{j=1,\dots,p}\mathrm{E}^*_{m,k}\left(\sqrt{m}\mathbf{u}_j^T\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)\widetilde{Z}_{m,k}(U_{m,k}(j))\right)^2\\
&= o(1) \qquad P-a.s.
\end{aligned} \tag{6.22}$$

$J_5$ is only negligible for local alternatives but still bounded for fixed alternatives. Precisely for any $c > 0$,

$$\sup_{k\geqslant 1} P^*_{m,k}\left(\frac{1}{m}\sum_{i=1}^{m} J_5^2(m,k,i) \geqslant c\right) \leqslant \frac{\boldsymbol{d}_m^T\mathbf{C}\boldsymbol{d}_m}{c^2} + o(1) \qquad P-a.s. \tag{6.23}$$

by (6.13), since $J_5(m,k,i) = \widetilde{Z}_{m,k}(U_{m,k}(i))$ as in the proof of Lemma 6.3. This yields a) under local alternatives.

For fixed alternatives note that

$$\left(\widehat{\sigma}^{(R*)}_{m,k}\right)^2 = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{6} J_j^2(m,k,i) + \sum_{u\neq v} J_u(m,k,i)J_v(m,k,i)\right).$$

The square terms are negligible except for $J_1^2$ and $J_5^2$ by (6.19) – (6.22). By the Cauchy-Schwarz inequality, (6.18) and (6.23) the same holds true for the mixed terms except

$J_1\, J_5$ but the latter one is also negligible due to Lemmas 6.1 and 6.6 since

$$\frac{1}{m}\sum_{s=1}^{m}(J_1(m,k,s)J_5(m,k,s)) = \frac{1}{m}\sum_{s=1}^{m}\mathbf{x}(U_{m,k}(s))^T e(U_{m,k}(s))1_{\{U_{m,k}(s)>m+k^\circ\}}\boldsymbol{d}_m - \frac{1}{m}\sum_{s=1}^{m}\mathbf{x}(U_{m,k}(s))^T e(U_{m,k}(s))\mathbf{C}_{m+k}^{-1}\mathbf{C}_{k^\circ,k}\boldsymbol{d}_m.$$

This shows that the only influential terms are $\frac{1}{m}\sum_{i=1}^{m}(J_1^2(m,k,i)+J_5^2(m,k,i))$. But since $\widehat{\sigma}_{m,k}^{(R*)}$ in Lemma 6.6. b) is in the denominator and $\frac{1}{m}\sum_{i=1}^{m}(J_1^2(m,k,i)+J_5^2(m,k,i)) \geqslant \frac{1}{m}\sum_{i=1}^{m}J_1^2(m,k,i)$ assertion b) follows by (6.18). ■

Putting the above lemmas together we easily obtain Theorem 3.1.

**Proof Proof of Theorem 3.1.** Putting together Lemmas 6.5 and 6.7 we obtain under $H_0$ as well as local alternatives

$$\sup_{1\leqslant k<N(m)+1}\left|P_{m,k}^*\left(\frac{1}{\widehat{\sigma}_{m,k}^{(R*)}}\sup_{1\leqslant \ell<N(m)+1}\frac{\widetilde{\Gamma}(m,\ell,\gamma)(e_{m,k}^*(1),\ldots,e_{m,k}^*(m+\ell))}{g(m,\ell,\gamma)}\leqslant x\right) - P\left(\sup_{1\leqslant k<N(m)+1}\frac{|W_1(k/m)-k/mW_2(1)|}{(1+k/m)(k/(k+m))^\gamma}\leqslant x\right)\right| \to 0 \qquad P-a.s.$$

By Lemma 6.2 this yields

$$\sup_{k\geqslant 1}|c_{m,k}^{(R)}-c| \to 0 \qquad P-a.s., \tag{6.24}$$

where $c$ is the asymptotic critical value obtained from the distribution of $\sup_{0\leqslant t\leqslant 1-\frac{1}{N+1}}\frac{|W(t)|}{t^\gamma}$. Together with Theorem 2.1 this implies a).

Under $H_1$, by Lemmas 6.5 and 6.7 for every $\epsilon>0$ there exists a constant $A>0$ such that $(P-a.s.)$

$$\sup_{1\leqslant k<N(m)+1}\left|P_{m,k}^*\left(\frac{1}{\widehat{\sigma}_{m,k}^{(R*)}}\sup_{1\leqslant \ell<N(m)+1}\frac{\widetilde{\Gamma}(m,\ell,\gamma)(e_{m,k}^*(1),\ldots,e_{m,k}^*(m+\ell))}{g(m,\ell,\gamma)}\geqslant A\right)\right| \leqslant \epsilon+o(1)$$

By Lemma 6.2 this yields

$$\sup_{k\geqslant 1}|c_{m,k}^{(R)}| = O(1) \qquad P-a.s. \tag{6.25}$$

Together with Theorem 2.2 this implies b). ■

# 7 Proofs of Section 4

Denote

$$A_1(m,k,\ell) = \sum_{i=m+1}^{m+\ell} e_{m,k}^*(i)$$

$$\boldsymbol{A}_2(m,k,\ell) = \sum_{i=m+1}^{m+\ell} \boldsymbol{x}^*_{m,k}(i)$$
$$\boldsymbol{A}_3(m,k) = \sum_{j=1}^{m} \boldsymbol{x}^*_{m,k}(j)\boldsymbol{x}^{*T}_{m,k}(j)$$
$$\boldsymbol{A}_4(m,k) = \sum_{s=1}^{m} \boldsymbol{x}^*_{m,k}(s)e^*_{m,k}(s)$$
$$\mathbf{A}_5(m,k,\ell) = \sum_{i=m+1}^{m+\ell} \boldsymbol{x}^*_{m,k}(i)1_{\{U_{m,k}(i)>m+k^\circ\}}$$
$$\boldsymbol{A}_6(m,k) = \sum_{s=1}^{m} \boldsymbol{x}^*_{m,k}(s)\boldsymbol{x}^{*T}_{m,k}(s)\mathbf{1}_{\{U_{m,k}(s)>m+k^\circ\}}$$

where $e^*_{m,k}(i) = e(U_{m,k}(i))$, which is different from the bootstrapped residuals in the regression bootstrap. Similarly to (3.1) it holds

$$g(m,k,\gamma)\,\Gamma(m,\ell,\gamma)^*_{m,k} = B_1(m,k,\ell) + B_2(m,k,\ell) + B_3(m,k,\ell) + B_4(m,k,\ell),$$

where

$$B_1(m,k,\ell) = A_1(m,k,\ell),$$
$$B_2(m,k,\ell) = -\boldsymbol{A}_2^T(m,k,\ell)\boldsymbol{A}_3^{-1}(m,k)\boldsymbol{A}_4(m,k)$$
$$B_3(m,k,\ell) = \left(\mathbf{A}_5(m,k,\ell) - \ell\,\mathrm{E}^*_{m,k}\,\mathbf{x}^*_{m,k}(1)1_{\{U_{m,k}(1)>m+k^\circ\}}\right)\boldsymbol{d}_m,$$
$$B_4(m,k,\ell) = -\left(\boldsymbol{A}_2^T(m,k,\ell)\boldsymbol{A}_3^{-1}(m,k)\boldsymbol{A}_6(m,k) - \ell\,\mathrm{E}^*_{m,k}\,\mathbf{x}^*_{m,k}(1)1_{\{U_{m,k}(1)>m+k^\circ\}}\right)\boldsymbol{d}_m.$$

The next lemma gives some properties of the terms $\boldsymbol{A}_j$.

**Lemma 7.1.** *Let* (1.1) *and Assumption* $\mathcal{A}$*.1 hold true.*

*a) Under either Assumption* $\mathcal{A}$*.2 or* $\mathcal{A}$*.3 we get for any* $\epsilon > 0$

$$\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\frac{1}{m^{1-\eta}}\left\|\boldsymbol{A}_3(m,k) - m\mathbf{C}\right\|_\infty \geqslant \epsilon\right) \to 0$$

*$P-a.s.$ for some $\eta > 0$.*

*b) Under either Assumption* $\mathcal{A}$*.2 or* $\mathcal{A}$*.3 we get for any* $\epsilon > 0$

$$\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\left\|\frac{1}{m}\boldsymbol{A}_6(m,k) - \mathrm{E}^*\,\mathbf{x}^*_{m,k}(1)\mathbf{x}^*_{m,k}(1)^T 1_{\{U_{m,k}(i)>m+k^\circ\}}\right\|_\infty \geqslant \epsilon\right) \to 0 \qquad P\text{–}a.s.$$

**Proof.** First note that from Assumptions $\mathcal{A}$.1 (ii) we get for every $\omega \in M$ with $P(M) = 1$ the existence of a constant $D(\omega)$, such that

$$\left\|\sum_{i=1}^{j}(\mathbf{x}(i)(\omega)\mathbf{x}(i)^T(\omega) - \mathbf{C})\right\|_\infty \leqslant D(\omega)j^{-\rho}$$

for each $j$. Subtracting the term for $j$ and $j-1$ we get

$$\|\mathbf{x}(j)(\omega)\mathbf{x}(j)^T(\omega)\|_\infty \leqslant \|\mathbf{C}\|_\infty + 2D(\omega)j^{1-\rho},$$

which yields

$$\|\mathbf{x}(j)\mathbf{x}(j)^T\|_\infty = O(j^{1-\rho}) \qquad P-a.s. \tag{7.1}$$

Further note that

$$\mathrm{E}^*_{m,k}(\mathbf{x}^*_{m,k}(i)\mathbf{x}^*_{m,k}(i)^T) = \frac{1}{m+k}\sum_{j=1}^{m+k}\mathbf{x}(i)\mathbf{x}(i)^T = \mathbf{C} + O(m^{-\rho}) \qquad P-a.s. \tag{7.2}$$

uniformly in $k$ by Assumption $\mathcal{A}.1$. This, (7.1) and an application of the Chebyshev inequality yields now (the square of the matrix is meant componentwise)

$$\begin{aligned}
&\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\frac{1}{m^{1-\eta}}\left\|\sum_{i=1}^{m}(\mathbf{x}^*_{m,k}(i)\mathbf{x}^*_{m,k}(i)^T - \mathbf{C})\right\|_\infty \geqslant \epsilon\right)\\
&\leqslant O(m^{2\eta-2\rho}) + O(1)\,\frac{1}{m^{1-2\eta}}\sup_{1\leqslant k<N(m)+1}\frac{1}{m+k}\left\|\sum_{i=1}^{m+k}(\mathbf{x}(i)\mathbf{x}(i)^T)^2\right\|_\infty\\
&\leqslant O(m^{2\eta-2\rho}) + O\left(\frac{(m+N(m))^{1-\rho}}{m^{1-2\eta}}\right)\sup_{k\geqslant 1}\sup_{j=1,\ldots,p}\frac{1}{m+k}\sum_{i=1}^{m+k}x_j^2(i)\\
&= O(m^{2\eta-2\rho} + m^{2\eta-\epsilon}) = o(1) \qquad P-a.s.
\end{aligned}$$

under Assumption $\mathcal{A}.2$ for some $\rho > 0$, which yields a). A similar argument but using $\mathcal{A}.3$ and the von Bahr-Esseen inequaltiy (cf. Theorem 3 in [17]) also yields assertion a).

Analogously we obtain b). ■

The next lemma is the analogue to Lemmas 6.3 and 6.4 for the regression bootstrap.

**Lemma 7.2.** *Let* (1.1), *Assumption* $\mathcal{A}.1$, *and* (2.1) *hold true and either* $H_0$ *or* $\boldsymbol{d}_m = O(1)$.

*a) Then for all* $\epsilon > 0$ *it holds:*

$$\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant \ell<N(m)+1}\frac{|B_2(m,k,\ell) - \frac{-\ell}{m}\sum_{j=1}^{m}e^*_{m,k}(j)|}{g(m,\ell,\gamma)} \geqslant \epsilon\right) \to 0,$$

*b) Under* $H_0$ *it holds that* $B_j(m,k,\ell) = 0$, $j = 3,4$, *under local alternatives, i.e. if* $\boldsymbol{d}_m = o(1)$, *it holds for all* $\epsilon > 0$ *that*

$$(i) \qquad \sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant \ell<N(m)+1}\frac{|B_3(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon\right) = o(1) \qquad P-a.s.,$$

$$(ii) \qquad \sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant \ell<N(m)+1}\frac{|B_4(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon\right) = o(1) \qquad P-a.s.$$

*c) For fixed alternatives, for which* $\boldsymbol{d}_m = O(1)$, *we get only the following weaker assertion: For every* $\epsilon > 0$ *there exists* $A > 0$ *such that*

$$(i) \qquad \sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant \ell<N(m)+1}\frac{|B_3(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant A\right) \leqslant \epsilon + o(1) \qquad P-a.s.,$$

$$(ii) \qquad \sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant \ell<N(m)+1}\frac{|B_4(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant A\right) \leqslant \epsilon + o(1) \qquad P-a.s.$$

**Proof.**

$$-B_2(m,k,\ell) = B_{2,1}(m,k,\ell) + B_{2,2}(m,k,\ell) + B_{2,3}(m,k,\ell),$$

where

$$\begin{aligned}
B_{2,1}(m,k,\ell) &= (\boldsymbol{A}_2(m,k,\ell) - E^*_{m,k}\boldsymbol{A}_2(m,k,\ell)^T)\boldsymbol{A}_3(m,k)^{-1}\boldsymbol{A}_4(m,k)\\
B_{2,2}(m,k,\ell) &= E^*_{m,k}\boldsymbol{A}_2^T(m,k,\ell)(E^*_{m,k}\boldsymbol{A}_3(m,k))^{-1}\boldsymbol{A}_4(m,k)\\
B_{2,3}(m,k,\ell)&\\
&= E^*_{m,k}\boldsymbol{A}_2^T(m,k,\ell)(E^*_{m,k}\boldsymbol{A}_3(m,k))^{-1}(E^*_{m,k}\boldsymbol{A}_3(m,k) - \boldsymbol{A}_3(m,k))\boldsymbol{A}_3(m,k)^{-1}\boldsymbol{A}_4(m,k).
\end{aligned}$$

Direct calculations give

$$\begin{aligned}
E^*_{m,k}\boldsymbol{A}_2^T(m,k,\ell)(E^*_{m,k}\boldsymbol{A}_3(m,k))^{-1} &= \frac{\ell}{m+k}\sum_{i=1}^{m+k}\boldsymbol{x}(i)\Big(\frac{m}{m+k}\sum_{i=1}^{m+k}\boldsymbol{x}(i)\boldsymbol{x}(i)^T\Big)^{-1}\\
&= \frac{\ell}{m}(1,0,\ldots,0)^T \qquad (7.3)
\end{aligned}$$

Therefore

$$B_{2,2}(m,k,\ell) = \frac{\ell}{m}\sum_{i=1}^{m} e^*_{m,k}(i).$$

By Lemmas 6.6 and 7.1 as well as (7.2) we get for any $\epsilon > 0$

$$\sup_{k\geqslant 1} P^*_{m,k}\left(m^{\eta-1}\|\mathbf{A}_3(m,k) - \mathrm{E}^*_{m,k}\,\mathbf{A}_3(m,k)\|_\infty \geqslant \epsilon\right) \to 0 \qquad P-a.s. \qquad (7.4)$$

$$\sup_{k\geqslant 1} P^*_{m,k}\left(m^{1-\xi}\|\mathbf{A}_3(m,k)^{-1}\mathbf{A}_4(m,k)\|_\infty \geqslant \epsilon\right) \to 0 \qquad P-a.s. \qquad (7.5)$$

for some $\eta > 0$ and for $\xi$ as in Lemma 7.1.

By (6.9) we get

$$\sup_{\ell\geqslant 1}\frac{\ell}{m\,g(m,\ell,\gamma)} = O(m^{-1/2}), \qquad (7.6)$$

which together with (7.3), (7.4) and (7.5) yields

$$\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant\ell<N(m)+1}\frac{|B_{2,3}(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon\right) \to 0 \qquad P-a.s.$$

Finally note that by Assumption $\mathcal{A}.1$ (ii)

$$\mathrm{var}^*_{m,k}(\mathbf{x}^*_{m,k}(i)) \leqslant \frac{1}{m+k}\sum_{i=1}^{m+k}\mathbf{x}(i)\mathbf{x}(i)^T = \mathbf{C} + o(1) \qquad P-a.s. \qquad (7.7)$$

uniformly in $k$. An application of the Hájek-Rényi inequality, (6.10) and (7.5) yields

$$\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant\ell<N(m)+1}\frac{|B_{2,1}(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant \epsilon\right) \to 0 \qquad P-a.s.$$

This completes the proof of a).

In the following let $D > 0$ be some (non-random) constant which can differ in every occurrence. Concerning b) and c), first note that by Assumption $\mathcal{A}.1$ we get uniformly in $k \geqslant 1$

$$\left\| \mathrm{var}^*_{m,k}(\mathbf{x}^*_{m,k}(i) 1_{\{U_{m,k}(i) > m+k^\circ\}}) \right\|_\infty = \frac{1}{m+k} \left\| \sum_{i=m+k^\circ}^{m+k} \mathbf{x}(i)\mathbf{x}(i)^T \right\|_\infty \leqslant D + o(1) \qquad P-a.s.$$

An application of the Hájek-Rényi inequality, (6.10) and Lemma 7.1 yields for any $c > 0$

$$\sup_{1 \leqslant k < N(m)+1} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|B_3(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant c \right) \leqslant D \frac{\|\boldsymbol{d}_m\|^2_\infty}{c^2} + o(1) \qquad P - a.s.,$$

which proves b) (i) and c) (i).

Concerning $B_4(m,k,\ell)$ we need an analogous decomposition as for $B_2(m,k,\ell)$ above.

$$- B_4(m,k,\ell) = B_{4,1}(m,k,\ell) + B_{4,2}(m,k,\ell) + B_{4,3}(m,k,\ell),$$

where

$$\begin{aligned}
&B_{4,1}(m,k,\ell) = \left( \boldsymbol{A}_2^T(m,k,\ell) - E^*_{m,k}\boldsymbol{A}_2^T(m,k,\ell) \right) \boldsymbol{A}_3(m,k)^{-1} \boldsymbol{A}_6(m,k) \boldsymbol{d}_m \\
&B_{4,2}(m,k,\ell) \\
&= \left( E^*_{m,k}\boldsymbol{A}_2^T(m,k,\ell)(E^*_{m,k}\boldsymbol{A}_3(m,k))^{-1}\boldsymbol{A}_6(m,k) - \ell\, \mathrm{E}^*_{m,k}\, \mathbf{x}^*_{m,k}(1) 1_{\{U_{m,k}(1) > m+k^\circ\}} \right) \boldsymbol{d}_m \\
&B_{4,3}(m,k,\ell) \\
&= \left( E^*_{m,k}\boldsymbol{A}_2^T(m,k,\ell)(E^*_{m,k}\boldsymbol{A}_3(m,k))^{-1}(E^*_{m,k}\boldsymbol{A}_3(m,k) - \boldsymbol{A}_3(m,k))\boldsymbol{A}_3(m,k)^{-1}\boldsymbol{A}_6(m,k) \right) \boldsymbol{d}_m.
\end{aligned}$$

Since

$$\mathrm{E}^*_{m,k}\, \mathbf{x}^*_{m,k}(1)\mathbf{x}^*_{m,k}(1)^T 1_{\{U_{m,k}(1) > m+k^\circ\}} = \frac{1}{m+k} \sum_{i=m+k^\circ}^{m+k} \mathbf{x}(i)\mathbf{x}(i)^T \leqslant D + o(1) \qquad P-a.s. \tag{7.8}$$

uniformly in $k$, we obtain from Lemma 7.1 that for each $\epsilon > 0$ there exists $A > 0$ such that

$$\sup_{k \geqslant 1} P^*_{m,k} \left( \|\mathbf{A}_3(m,k)^{-1}\mathbf{A}_6(m,k)\|_\infty \geqslant A \right) \leqslant \epsilon + o(1) \qquad P - a.s. \tag{7.9}$$

This in addition to an application of the Hájek-Rényi inequality and (6.10) yields for any $c > 0$

$$\sup_{1 \leqslant k < N(m)+1} P^*_{m,k} \left( \max_{1 \leqslant \ell < N(m)+1} \frac{|B_{4,1}(m,k,\ell)|}{g(m,\ell,\gamma)} \geqslant c \right) \leqslant D\, \frac{\|\boldsymbol{d}_m\|^2_\infty}{c^2} + o(1) \qquad P - a.s.$$

By (7.3) we get

$$B_{4,2}(m,k,\ell) = \frac{\ell}{m} \sum_{j=1}^{m} (\mathbf{x}^*_{m,k}(j)^T 1_{\{U_{m,k}(j) > m+k^\circ\}} - \mathrm{E}^*\, \mathbf{x}^*_{m,k}(j) 1_{\{U_{m,k}(j) > m+k^\circ\}}) \boldsymbol{d}_m.$$

An application of the Chebyshev inequality yields for any $c > 0$

$$\sup_{k\geqslant 1} P^*_{m,k}\left(\frac{1}{\sqrt{m}}\left|\sum_{j=1}^{m}(\mathbf{x}^*_{m,k}(j)^T 1_{\{U_{m,k}(j)>m+k^\circ\}} - \mathrm{E}^*\,\mathbf{x}^*_{m,k}(j)^T 1_{\{U_{m,k}(j)>m+k^\circ\}})\boldsymbol{d}_m\right| \geqslant c\right)$$
$$\leqslant \frac{\|\boldsymbol{d}_m\|^2_\infty}{c^2}\sup_{k\geqslant 1}\frac{1}{m+k}\left\|\sum_{i=m+k^\circ}^{m+k}\mathbf{x}(i)\mathbf{x}(i)^T\right\|_\infty \leqslant D\,\frac{\|\boldsymbol{d}_m\|^2_\infty}{c^2} + o(1) \qquad P-a.s. \tag{7.10}$$

Together with (7.6) this yields

$$\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant \ell<N(m)+1}\frac{|B_{4,2}(m,k,\ell)|}{g(m,\ell,\gamma)}\geqslant c\right) \leqslant D\,\frac{\|\boldsymbol{d}_m\|^2_\infty}{c^2} + o(1) \qquad P-a.s.$$

Finally by (7.3)

$$B_{4,3}(m,k,\ell) = -\frac{\ell}{m}\sum_{j=1}^{m}(\mathbf{x}^*_{m,k}(j)^T - \mathrm{E}^*_{m,k}\,\mathbf{x}^*_{m,k}(j)^T)\mathbf{A}_3(m,k)^{-1}\mathbf{A}_6(m,k)\boldsymbol{d}_m$$

By (7.6), (7.9) and an analogous argument to (7.10) using (7.7) we finally obtain

$$\sup_{1\leqslant k<N(m)+1} P^*_{m,k}\left(\max_{1\leqslant \ell<N(m)+1}\frac{|B_{4,3}(m,k,\ell)|}{g(m,\ell,\gamma)}\geqslant c\right) \leqslant \|C\|_\infty\,\frac{\|\boldsymbol{d}_m\|^2_\infty}{c^2} + o(1) \qquad P-a.s.,$$

which completes the proof. ■

Now, we prove the equivalent of Lemma 6.7.

**Lemma 7.3.** *Let* (1.1)*, Assumption A.1, and either Assumption A.2 or A.3 hold true. Let* $\widehat{\sigma}^2_{m,k}$ *be as in* (6.15)*.*

*a) Under* $H_0$ *or local alternatives (*$\boldsymbol{d}_m = o(1)$*) it holds for all* $\epsilon > 0$

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

*b) Under* $H_1$ *with* $\boldsymbol{d}_m = O(1)$ *for every* $\epsilon > 0$ *there exists* $A > 0$ *such that*

$$\sup_k P^*_{m,k}\left(\left|\frac{\widehat{\sigma}_{m,k}}{\widehat{\sigma}^{(P*)}_{m,k}}\right| \geqslant A\right) \leqslant \epsilon + o(1) \qquad P-a.s.$$

**Proof.** Note that

$$y^*_{m,k}(i) - \mathbf{x}^*_{m,k}(i)^T\left(\sum_{j=1}^{m}\mathbf{x}^*_{m,k}(j)\mathbf{x}^*_{m,k}(j)^T\right)^{-1}\sum_{l=1}^{m}\mathbf{x}^*_{m,k}(l)^T y^*_{m,k}(l)$$
$$= D_1(m,k,i) + D_2(m,k,i) + D_3(m,k,i),$$

where

$$D_1(m,k,i) = e^*_{m,k}(i),$$
$$D_2(m,k,i) = -\mathbf{x}^{*T}_{m,k}(i)\mathbf{A}_3(m,k)^{-1}\mathbf{A}_4(m,k),$$
$$D_3(m,k,i) = \mathbf{x}^{*T}_{m,k}(i)1_{\{U_{m,k}(i)>m+k^\circ\}}\boldsymbol{d}_m - \mathbf{x}^{*T}_{m,k}(i)\mathbf{A}_3(m,k)^{-1}\mathbf{A}_6(m,k)\boldsymbol{d}_m.$$

By (6.18) it holds ($D_1(m,k,i) = J_1(m,k,i)$)

$$\sup_{k\geqslant 1} P^*_{m,k}\left(\left|\frac{\frac{1}{m-p}\sum_{i=1}^m D_1^2(m,k,i)}{\widehat{\sigma}^2_{m,k}} - 1\right| \geqslant \epsilon\right) \to 0 \qquad P-a.s. \tag{7.11}$$

Furthermore since

$$\sum_{i=1}^m D_2^2(m,k,i) = \mathbf{A}_4(m,k)\mathbf{A}_3^{-1}(m,k)\mathbf{A}_4(m,k),$$

for every $\epsilon > 0$ by Lemmas 6.6 and 7.1

$$\sup_{k\geqslant 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^m D_2^2(m,k,i) \geqslant \epsilon\right) \to 0 \qquad P-a.s. \tag{7.12}$$

This shows that asymptotically this summand is negligible as is the mixed term of $D_1$ and $D_2$ due to the Cauchy-Schwarz inequality. Since $D_3 = 0$ under $H_0$ this proves a) under $H_0$.

Concerning alternatives it holds

$$\sum_{j=1}^m D_3^2(m,k,i) = \boldsymbol{d}_m^T\mathbf{A}_6(m,k)\boldsymbol{d}_m - \boldsymbol{d}_m^T\mathbf{A}_6(m,k)(\mathbf{A}_3(m,k))^{-1}\mathbf{A}_6(m,k)\boldsymbol{d}_m,$$

which implies due to Lemma 7.1 and (7.8) for every $c > 0$ for some constant $D > 0$

$$\sup_{k\geqslant 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^m D_3^2(m,k,i) \geqslant c\right) \leqslant D\,\frac{\|\boldsymbol{d}_m\|^2_\infty}{c^2} + o(1) \qquad P-a.s. \tag{7.13}$$

proving a) for local alternatives, since the mixed terms are again negligible due to the Cauchy-Schwarz inequality.

Finally for $\widetilde{\mathbf{A}}_4(m,k) = \sum_{s=1}^m \mathbf{x}^*_{m,k}(s)e^*_{m,k}(s)1_{\{U_{m,k}(s)>m+k^\circ\}}$

$$\sum_{i=1}^m D_1(m,k,i)D_3(m,k,i) = \widetilde{\mathbf{A}}_4(m,k)\boldsymbol{d}_m - \mathbf{A}_4(m,k)(\mathbf{A}_3(m,k))^{-1}\mathbf{A}_6(m,k)\boldsymbol{d}_m,$$

which is also negligible due to Lemmas 6.6 and 7.1. We can now finish the proof for fixed alternatives analogously to the proof of Lemma 6.7. ■

**Proof Proof of Theorem 4.1.** Due to Lemma 7.2 we obtain the analogous assertion for the pair bootstrap to what is given for the regression bootstrap in Lemma 6.5. We can then conclude as in the proof of Theorem 3.1 using Lemma 7.3. ■

# References


[1] Andreou, E., and Ghysels, E. Monitoring disruptions in financial markets. *J. Econometrics*, 135:77–124, 2006.

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

[3] Aue, A., Hörmann, S., Horváth, L., and Hušková, M. Sequential testing for the stability of portfolio betas. 2009. In preparation.

[4] Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. Change-point monitoring in linear models. *Econometrics Journal*, 9:373–403, 2006.

[5] Berkes, I., Horváth, L., Hušková, M., and Steinebach, J. Applications of permutations to the simulations of critical values. *J. Nonparametr. Stat.*, 16:197–216, 2004.

[6] Chow, Y.S., and Teicher, H. *Probability Theory – Independence, Interchangeability, Martingales*. Springer, New York, third edition, 1997.

[7] Fried, R., and Imhoff, M. On the online detection of monotonic trends in time series. *Biom. J.*, 46:90–102, 2004.

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

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

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

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

[12] Hušková, M. and Koubková, A. Monitoring jump changes in linear models. *J. Statist. Res.*, 39:59–78, 2005.

[13] Hušková, M. and Koubková, A. Sequential procedures for detection of changes in autoregressive sequences. In Hušková, M. and Lachout, P., editors, *Proceedings of the Prague Stochstics 2006*, pages 437–447, 2006.

[14] Kirch, C. Bootstrapping sequential change-point tests. *Sequential Anal.*, 27:330–349, 2008.

[15] Koubková, A. Change detection in the slope parameter of a linear regression model. *Tatra Mt. Math. Publ.*, 39:245–253, 2008.

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

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