Metrika (2012) 75:673–708
DOI 10.1007/s00184-011-0347-7

Bootstrapping sequential change-point tests for linear regression

Marie Hušková · Claudia Kirch

Received: 10 September 2009 / Published online: 17 February 2011

Abstract Bootstrap methods for sequential change-point detection procedures in linear regression models are proposed. The corresponding monitoring procedures are designed to control the overall significance level. The bootstrap critical values are updated constantly by including new observations obtained from the monitoring. The theoretical properties of these sequential bootstrap procedures are investigated, showing their asymptotic validity. Bootstrap and asymptotic methods are compared in a simulation study, showing that the studentized bootstrap tests hold the overall level better especially for small historic sample sizes while having a comparable power and run length.

Keywords Bootstrap · Sequential test · Change-point analysis · Linear regression

1 Introduction

In recent years an increasing number of data sets are collected automatically and without significant costs in such a way that the observations arrive steadily. Examples

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.

M. Hušková
Department of Statistics, Charles University of Prague, Sokolovská 83, 18675 Praha 8, Czech Republic
e-mail: huskova@karlin.mff.cuni.cz

C. Kirch (✉)
Institute for Stochastics, Karlsruhe Institute of Technology (KIT), Kaiserstr. 89,
76133 Karlsruhe, Germany
e-mail: claudia.kirch@kit.edu

include financial data sets e.g. in risk management (Andreou and Ghysels 2006) or CAPM models (Aue et al. 2009) as well as medical data sets e.g. monitoring intensive care patients (Fried and Imhoff 2004). 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 may not be appropriate anymore or the condition of the patient in intensive medical care may have changed.

The consideration of such data sets leads to sequential statistical analysis, which is sometimes also called on-line monitoring. Model parameters are estimated from a historic data set without change before monitoring starts. Similarly to classical statistical analysis the sequential tests we are interested in control the overall significance level of the monitoring procedure. Asymptotic critical values are obtained by letting the size of the historic data set go to infinity. In previous papers (Chu et al. 1996; Horváth et al. 2004) monitoring is assumed to continue for infinity (open-end procedure) if no alarm is raised. Since it is more realistic in many situations that monitoring is stopped after a finite time horizon even if no change is detected (closed-end procedure), we generalize their results to include this case. Using the critical values from the open-end procedure for a closed-end procedure distorts the size of the test resulting in a loss of power.

Frequently, asymptotic tests perform unsatisfactory for small sample sizes due to a slow convergence to the limit distribution, which may not even be known explicitly. This lead to the development of permutation and bootstrap tests, which usually work better for small samples. For a thorough introduction we refer to Good (2005). In change-point analysis permutation methods were first suggested by Antoch and Hušková (2001) and later pursued by others (for a recent survey confer Hušková 2004). Berkes et al. (2004) 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 the classical situation, where the complete data set has been observed before conducting the test.

In this paper, we develop variations of bootstrap methods for linear regression models that are suitable in a sequential setting. The bootstrap estimate for the critical value is updated with every new observation leading to an improvement as more observations are being used. From a practical point of view the update step should include all available information but still be computationally fast. From a theoretical point of view the influence of time-varying critical values on the asymptotic size and power of the test needs to be checked.

The literature on bootstrapping methods for sequential tests is very scarce. Steland (2006) used a bootstrap in a sequential unit-root test and Kirch (2008) explored several possibilities of sequential bootstrapping for the detection of mean changes in i.i.d. data. As in classical statistics, those sequential bootstrap tests behave better for small historic sample sizes than the corresponding asymptotic tests.

In this paper we use the same approach to sequential bootstrapping but some additional problems arise due to the more complicated data structure. We focus on the linear regression model

$$y(i) = \mathbf{x}(i)^T \boldsymbol{\beta}_{i,m} + e(i), \quad i \geq 1, \tag{1}$$

where $\mathbf{x}(i)$ is a $p \times 1$ random vector and $\boldsymbol{\beta}_{i,m}$ 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.

Model estimation is based on a historic sequence of observations, where one assumes that no change in the regression coefficient has occurred, i.e.

$$\boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, \quad 1 \leq i \leq m. \tag{2}$$

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

$$H_0 : \boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, \quad m < i < m + T_m + 1 \tag{3}$$

against the alternative of a change in the regression coefficient

$$\begin{aligned} H_1 &: \text{there is a } k_m^\circ \geq 1 \quad \text{such that } \boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, \quad m < i \leq m + k_m^\circ \\ &\text{and} \quad \boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_m^0 \neq \boldsymbol{\beta}_0, \quad m + k_m^\circ < i < m + T_m + 1. \end{aligned} \tag{4}$$

T_m is the observation horizon which can be finite or infinite but has to converge to infinity with m. The values of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_m^0$ and k_m° are not known.

We allow the regression coefficient after the change-point $\boldsymbol{\beta}_{i,m}, i \geq m + k_m^\circ$, to depend on m in order to cover both fixed as well as local alternatives, for which $\mathbf{d}_m := \boldsymbol{\beta}_m^0 - \boldsymbol{\beta}_0 \to \mathbf{0}$ as $m \to \infty$ at a certain rate.

The test is based on the following statistic

$$\begin{aligned} \Gamma(m,k,\gamma) &= \sum_{m<i\leq m+k} \left(y(i) - \mathbf{x}(i)^T \widehat{\boldsymbol{\beta}}_m \right) / g(m,k,\gamma), \\ \text{where } g(m,k,\gamma) &= m^{1/2} \left(1 + \frac{k}{m}\right) \left(\frac{k}{m+k}\right)^\gamma \end{aligned} \tag{5}$$

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

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

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

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

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

$$\widehat{\sigma}_m^2 = \frac{1}{m-p} \sum_{i=1}^{m} \left(y(i) - \mathbf{x}(i)^T \widehat{\boldsymbol{\beta}}_m \right)^2. \tag{6}$$

The null hypothesis is rejected at the following stopping time

$$\tau(m) = \begin{cases} \inf\{1 \le k < T_m + 1 : \frac{1}{\widehat{\sigma}_m} |\Gamma(m,k,\gamma)| \ge c\}, \\ \infty, \quad \text{if } \frac{1}{\widehat{\sigma}_m} |\Gamma(m,k,\gamma)| < c \text{ for all } 1 \le k < T_m + 1, \end{cases} \tag{7}$$

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

$$\lim_{m\to\infty} P(\tau(m) < \infty) = \alpha \tag{8}$$

for some given level $0 < \alpha < 1$. This shows that the monitoring procedure keeps the overall significance level α asymptotically as in classical test theory. Under the alternative H_1 we require that

$$\lim_{m\to\infty} P(\tau(m) < \infty) = 1, \tag{9}$$

in other words, the test has asymptotic power one.

The paper is organized as follows: In Sect. 2 some known results on the limit behavior of the test statistic under the null as well as alternative hypotheses are summarized and extended. In Sect. 3 the so called regression bootstrap is introduced in a sequential setting and its asymptotic equivalence to the asymptotic procedure shown. This proves the validity of the sequential bootstrap in this setting. In Sect. 4 a second type of bootstrap used in linear regression namely the pair bootstrap is introduced in this sequential setting and corresponding results are obtained. In Sect. 5 some simulations illustrate the usefulness of the bootstrap methods. Finally the proofs are given in Sects. 6 and 7 for the regression and pair bootstrap, respectively.

2 Assumptions and limit behavior of the test statistic

In this section we formulate the required assumptions and consider the limit behavior of the monitoring procedure. For the open-end procedure those results under the null hypothesis as well as under fixed alternatives were obtained by Horváth et al. (2004). Here, we also allow for local changes and explicitly consider the asymptotics for the closed-end procedure which is very important for the consideration of bootstrap results, where statistics can only be calculated as closed-end approximations to possible open-end procedures.

We consider model (1) satisfying the following assumptions:

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) $\{e(i) : 1 \leq i < \infty\}$ 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 $\{\boldsymbol{x}(i) = (1, x_2(i), \ldots, x_p(i))^T : 1 \leq i < \infty\}$ there exists a positive definite matrix $\mathbf{C}$ and a constant $0 < \rho \leq 1/2$ 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 $\{e(i) : 1 \leq i < \infty\}$ and $\{\boldsymbol{x}(i) : 1 \leq i < \infty\}$ are independent,

(iv) $0 \leq \gamma < \rho$.

Horváth et al. (2004) proved the following asymptotics for the open-end procedure under the null distribution (i.e. (2) and (3)) if Assumption $\mathcal{A}$.1 holds:

$$\lim_{m\to\infty} P\left(\sup_{1\leq k<\infty} \frac{|\Gamma(m,k,\gamma)|}{\widehat{\sigma}_m} \leq y \right) = P\left(\sup_{0\leq t\leq 1} \frac{|W(t)|}{t^{\gamma}} \leq y \right) \tag{10}$$

for all $y \in \mathbb{R}$, where $\{W(t) : 0 \leq t < \infty\}$ denotes a Wiener process. The explicit form of the limit distribution is known explicitly only for $\gamma = 0$ and has to be simulated otherwise. Using quantiles of the limit distribution as critical values c in (7) ensures that (8) holds, i.e. the corresponding asymptotic open-end-test controls the overall false-rejection rate.

We now consider the closed-end procedure with

Assumption $\mathcal{A}$.2 $T_m < \infty$ with $\lim_{m\to\infty} T_m/m = N$ for some $0 < N < \infty$ or $\lim_{m\to\infty} T_m/m = \infty$.

Under this assumption a slight modification of the proof of the above results shows that it holds for all $y \in \mathbb{R}$ as $m \to \infty$

$$P\left(\sup_{1\leq k<T_m+1} \frac{|\Gamma(m,k,\gamma)|}{\widehat{\sigma}_m} \leq y \right)$$
$$= P\left(\sup_{1\leq k<T_m+1} \frac{|W_1(k/m) - k/m W_2(1)|}{(1+k/m)(k/(k+m))^{\gamma}} \leq y \right) + o_P(1), \tag{11}$$

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

Going through the proof of Theorem 2.1 in Horváth et al. (2004), the following result can be seen: The distribution of the right hand side of (11) converges to the same limit as given in (10) if $T_m/m \to \infty$ and it converges to the distribution of

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

if $T_m/m \to N$. Again the distribution on the right hand side of (11) is not explicitly known and needs to be simulated. Simulations concerning the location model in Kirch (2008) show that this distribution is very close to the one in (10) if $N \ge 10$. However, for a smaller observation horizon it is not recommendable to use critical values from the distribution in (10) (for a detailed discussion we refer to Kirch 2008).

Using the quantiles of the distribution on the right hand side of (11) as critical value c in (7) guarantees that (8) holds. The corresponding monitoring procedure will be called **asymptotic closed-end procedure**.

In the next sections we explore possibilities of obtaining critical values using bootstrap methods.

For the asymptotic open-end procedure for fixed alternatives, where $\boldsymbol{\beta}_m^0 = \boldsymbol{\beta}^0$ and $\boldsymbol{c}_1^T(\boldsymbol{\beta}^0 - \boldsymbol{\beta}_0) \ne 0$ ($\boldsymbol{c}_1$ is the first column of $\boldsymbol{C}$) Horváth et al. (2004) have proven that

$$\frac{1}{\widehat{\sigma}_m} \sup_{1\le k<\infty} |\Gamma(m,k,\gamma)| \xrightarrow{P} \infty. \tag{12}$$

This shows that the corresponding monitoring procedure has asymptotic power one, i.e. fulfills (9). Their proofs can be extended to the class of alternatives fulfilling

$$k_m^\circ = O(m), \qquad \lim_{m\to\infty} \sqrt{m}|\boldsymbol{c}_1^T\boldsymbol{d}_m| = \infty, \tag{13}$$

where $\mathbf{d}_m = \boldsymbol{\beta}_m^0 - \boldsymbol{\beta}_0$, which includes certain local alternatives.

Furthermore, their result can be extended to the closed-end procedure if additionally $\limsup_{m\to\infty} k_m^\circ/m < N$, where $N = \lim_{m\to\infty} T_m/m$.

For open-end procedures corresponding results have also been proven in a more general setup including e.g. heteroscedastic errors, cf. Aue et al. (2006). Bootstrap methods for this more general situation can in principle also be developed, however the proofs would become much more complex and even less transparent.

Furthermore, a different class of test statistics based on L_1 estimators and related partial sums of residuals instead of the corresponding L_2 procedures above has been considered by Hušková and Koubková (2005), Hušková and Koubková (2006) and Koubková (2008).

The test based on the test statistic (5) is only consistent under (13) which is quite restrictive, as essentially the change needs to imply a mean change of $y(\cdot)$. Hušková and Koubková (2005) introduced test procedures based on quadratic forms of weighted partial sums of residuals, which yield consistent tests for all fixed alternatives removing the restriction $\boldsymbol{c}_1^T(\boldsymbol{\beta}^0 - \boldsymbol{\beta}_0) \ne 0$.

Extensions of the bootstrapping techniques developed in this paper to these test statistics and more general setups 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. We begin with the discussion of the regression bootstrap in a sequential setup and will consider the pair bootstrap in the next section.

The general idea of the regression bootstrap is to resample the estimated residuals but keep the regressors in their original order. As a result, in the bootstrap world one deals with a regression with a fixed design rather than a stochastic one. In a sequential setup the additional problem arises that future regressors are not known yet.

To understand the bootstrap statistic, observe that under the null hypothesis as well as for $\ell \le k_m^\circ$ under the alternative

$$y(\ell) - \mathbf{x}(\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} y(j)\mathbf{x}(j) = e(\ell) - \mathbf{x}(\ell)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} e(j)\mathbf{x}(j). \quad (14)$$

Thus

$$\Gamma(m, \ell, \gamma) = \left(\sum_{i=m+1}^{m+\ell} e(i) - \sum_{i=m+1}^{m+\ell} \mathbf{x}(i)^T \boldsymbol{C}_m^{-1} \sum_{j=1}^{m} \mathbf{x}(j)e(j) \right) / g(m, \ell, \gamma). \quad (15)$$

This is the version we will use in the bootstrap as its distribution is equal to the null distribution which is the one that needs to be approximated by the bootstrap.

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.

Suppose that $m+k$ observations have been taken (including the m from the historic data set) and the null hypothesis has not been rejected yet. Based on those $m+k$ observations $(y(i), \mathbf{x}(i))$, $i = 1, \dots, m+k$, a bootstrap statistic can now be constructed in the following way:
First, replace $e(i)$ in the formula on the right hand side of (15) by the bootstrap estimates $e^*_{m,k}(i)$ below and keep $\mathbf{x}(i)$ for $1 \le i \le k+m$.

The test statistic $\sup_{1 \le \ell < T_m+1} |\Gamma(m, \ell, \gamma)|$ (with $\Gamma(m, \ell, \gamma)$ as on the right-hand side of (15)) depends additionally on the future regressors $\mathbf{x}(\ell)$, $\ell \ge k$, 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 1 b) as well as

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

To sum up, for the calculation of bootstrap critical values we propose to use $\sup_{1\le\ell<T_m+1}|\widetilde{\Gamma}(m,\ell,\gamma)|$ where

$$\widetilde{\Gamma}(m,\ell,\gamma)(e(1),\ldots,e(m+\ell))$$
$$=\left(\sum_{i=m+1}^{m+\ell} e(i)-\mathbf{c}_1(m,k,\ell)^T \boldsymbol{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)e(j)\right)\Big/ g(m,\ell,\gamma),$$
$$\text{with } \mathbf{c}_1(m,k,\ell)=\begin{cases}\sum_{i=m+1}^{m+\ell}\mathbf{x}(i), & \ell\le k,\\ \sum_{i=m+k-\ell+1}^{m+k}\mathbf{x}(i), & k<\ell<k+m,\\ \frac{\ell}{m+k}\sum_{i=1}^{m+k}\mathbf{x}(i), & \ell\ge m+k.\end{cases} \tag{16}$$

for $1\le\ell<T_m+1$. Let the bootstrap errors be defined by

$$e^*_{m,k}(i)=\widehat{e}_{m,k}(U_{m,k}(i)), \quad \text{where} \quad \widehat{e}_{m,k}(j)=y(j)-\mathbf{x}(j)^T\widehat{\boldsymbol{\beta}}_{m+k}, \tag{17}$$

$i=1,\ldots,m+T_m$, $j=1,\ldots,m+k$, where $\{U_{m,k}(i):1\le i\le m+T_m\}$ are i.i.d. random variables with $P(U_{m,k}(1)=j)=1/(m+k)$, $j=1,\ldots,m+k$, independent of $\{y(i):1\le i\le m+T_m\}$ and $\{\mathbf{x}(i):1\le i\le m+T_m\}$. By $P^*_{m,k}$, $\mathrm{E}^*_{m,k}$, $\mathrm{var}^*_{m,k}$ etc. we denote the conditional probability, expectation, variance etc. given $\{(y(i),\mathbf{x}(i)^T):1\le i\le m+k\}$, i.e. with respect to $\{U_{m,k}(i):1\le i\le m+T_m\}$.

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

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

where by (14)

$$\left(\widehat{\sigma}^{(R*)}_{m,k}\right)^2=\frac{1}{m-p}\sum_{i=1}^{m}\left(e^*_{m,k}(i)-\mathbf{x}(i)^T\boldsymbol{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)e^*_{m,k}(j)\right)^2 \tag{18}$$

is the bootstrap version of (6).

However, it is often 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 (2006), that has also proven to work well in Kirch (2008). 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 Lth observation. Secondly, and maybe even more importantly, we use a convex combination of the latest M bootstrap distributions. Thus, in applications an empirical distribution function not only based on the newest bootstrap samples but also on older samples is used. As a result only a fraction of the bootstrap samples need to be generated each time the critical values are updated: e.g. for $\alpha_i = 1/M$ below, only $t_1 = t/M$ new samples are needed in each update to calculate the empirical quantiles based on t samples. Therefore, the procedure is significantly accelerated even if critical values are updated after each new observation ($L = 1$). Let for $j \geq 1$, $\sum_{i=0}^{M-1} \alpha_i = 1$ and $\alpha_i \geq 0$

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

Then, the bootstrap critical values $c_{m,k}^{(R)}$ at time $k + m$ are obtained as follows

$$\widetilde{F}_{m,k}^{(R)}(c_{m,k}^{(R)}) \geq 1 - \alpha, \tag{19}$$

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

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

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

Theorem 1 *Let (1), Assumption $\mathcal{A}.1$ and $\mathcal{A}.2$ hold true. Then it holds as $m \to \infty$*

a) under the null hypothesis,

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

b) If (13) in addition to $\mathbf{d}_m = O(1)$ holds, then

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

Remark 1 To obtain bootstrap critical values for the open-end procedure obviously only a finite observation horizon can be used. In this case the assertions of Theorem 1 remain true for the open-end procedure, if critical values are obtained from a bootstrap based on a statistic with horizon $\widetilde{T}_m < \infty$ fulfilling $\widetilde{T}_m/m \to \infty$.

Remark 2 Clearly, the test procedure based on the bootstrap approximation of critical values has the desired properties (8) and (9).

Moreover, under H_0 and local alternatives the bootstrap provides an asymptotically correct approximation for the critical value in the sense given in equation (45).

Under alternatives the proof of the theorem only shows that the bootstrap critical values are uniformly bounded, cf. (46). However, it is to be expected that using more technical proofs it can be shown that the bootstrap critical values under alternatives are asymptotically correct in a P-stochastic sense corresponding to (45). In fact, the term that is responsible for the weaker result is $\frac{1}{m+k}\mathbf{C}_{k_m^\circ,k}$ (cf. e.g. (34)), which is without further knowledge only uniformly bounded. As soon as we can prove that it converges to 0 uniformly for $1 \le k \le \widetilde{\tau}(m)$, where $\widetilde{\tau}(m)$ is the stopping time of the procedure, the stronger result follows. Aue et al. (2006) show for the stopping time $\tau(m)$ of the asymptotic open-end procedure that

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

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 $\sup_{1\le k\le\tau(m)} \|\frac{1}{m+k} C_{k_m^\circ,k}\|_\infty = o_P(1)$ as desired.

4 Pair bootstrap

In this section we consider a second popular bootstrap in linear regression models in a sequential setup. This bootstrap 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 pair bootstrap directly preserves the dependence between $\mathbf{x}(i)$ and $y(i)$. Therefore, 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.

The specifics about the update step in the sequential setup are the same as for the regression bootstrap.

Precisely the pairs $\{(y(i), \mathbf{x}(i)) : 1 \le i \le m + k\}$ are bootstrapped, i.e. one considers

$$y^*_{m,k}(i) = y(U_{m,k}(i)), \qquad \mathbf{x}^*_{m,k}(i) = \mathbf{x}(U_{m,k}(i)), \tag{20}$$

where $\{U_{m,k}(i) : 1 \le i \le m + T_m\}$ are i.i.d. random variables with $P(U_{m,k}(1) = j) = 1/(m + k)$, $j = 1, \ldots, m + k$, independent of $\{y(i) : 1 \le i \le m + T_m\}$ and $\{\mathbf{x}(i) : 1 \le i \le m + T_m\}$ as above. The bootstrap statistic is given by

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

Analogously to Sect. 3 we define

$$F^{(P*)}_{m,k}(x) = P_{m,k}\left(\frac{1}{\widehat{\sigma}^{(P*)}_{m,k}}\sup_{1\le\ell<T_m+1}\left|\Gamma(m,\ell,\gamma)^*_{m,k}\right| \le x\right), \tag{21}$$

where

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

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

is the bootstrap version of (6) and

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

Finally, the critical value at time $k+m$ is calculated as

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

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

To obtain validity of the pair bootstrap somewhat stronger assumptions than before are needed. Precisely we either assume:

Assumption $\mathcal{A}$.3 Let the observation horizon T_m, on which the bootstrap is based, fulfill

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

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

This assumption is no restriction for the closed-end procedure. For the open-end procedure the calculation of bootstrap critical values needs to be based on a statistic with horizon $\widetilde{T}_m$ fulfilling Assumption $\mathcal{A}.3$ in addition to $\widetilde{T}_m/m \to \infty$. Similarly as in Remark 1, the monitoring procedure based on the bootstrap critical values fulfill (8) and (9) (cf. also Remark 3).

Alternatively, we can put some stronger assumptions on the regressors.

Assumption $\mathcal{A}$.4 The regressors fulfill for some $r > 1$ as $k \to \infty$

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

Usually the rate in Assumption $\mathcal{A}.1$ (ii) is obtained by some higher moment assumptions which imply Assumption $\mathcal{A}.4$ as well.

The following theorem shows that (8) and (9) hold for the monitoring scheme based on the pair bootstrap.

Theorem 2 *Let (1) and Assumptions $\mathcal{A}.1$ as well as $\mathcal{A}.2$ hold true in addition to either $\mathcal{A}.3$ or $\mathcal{A}.4$. Then, it holds as $m \to \infty$*

a) *under the null hypothesis,*

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

b) *If (13) in addition to $\mathbf{d}_m = O(1)$ holds, then*

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

Remark 3 The assertion in Remarks 1 and 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 (2008).

5 Some simulations

In the previous sections the asymptotic validity of the bootstrap tests has been established.

The following simulation study compares the two bootstrap procedures with the asymptotic closed-end (CE) procedure to get an impression how well the procedures work for small historic sample sizes.

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

We visualize these features by the following plots:

Size-power curves
The line corresponding to the null hypothesis ($\mathbf{d} = (0,0)^T$) shows the empirical size of the corresponding monitoring procedures for a nominal size as given by the x-axis. The lines corresponding to specific alternatives ($\mathbf{d} = (1,0)^T, (0,1)^T, (1,1)^T$) show the empirical power of the monitoring procedure for the nominal size as given by the x-axis. As a result the plots give a visualization of $\mathcal{C}.1$ and $\mathcal{C}.2$ above. 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.

Size-Power Curves are easily obtained by plotting the empirical distribution function of the p-values obtained from 1000 random realizations of model (1). The p-value in a sequential setting with possibly varying critical values can be calculated as follows: Let the critical value at point $m + k$ be given by $c_k = G_k^{-1}(1 - \alpha)$, where

$G_k = \widetilde{F}^{(R)}_{m,k}$ ($G_k = \widetilde{F}^{(P)}_{m,k}$) for the regression (pair) bootstrap and $G_k = G$ is equal to the distribution function on the right hand side of (11) for the asymptotic closed-end procedure. Define the p-value at point $m+k$ by $p_k = 1 - G_k(|\Gamma(m,k,\gamma)|/\widehat{\sigma}_m -)$, so that the null hypothesis is rejected at time point $m+k$ if $p_k < \alpha$. Then the p-value of this sequential procedure is defined by $p := \inf_{1\le k<T_m+1} p_k$ and the null hypothesis rejected if $p < \alpha$.

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 $\mathbf{d}_m = (1,1)^T$. This visualizes $\mathcal{C}$.3.

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 the power of two tests can only be reasonably compared if the true size (not the nominal one) is equal. The estimation of the density of the run length 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 (2008). The following model is considered

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

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 $U[-1,1]$ distributed innovations and coefficients $-0.5, 0.5$ respectively (Scenarios 2 resp. 3).
- $\mathbf{d}_m^T = (d_0, d_1) = (0,1), (1,0), (1,1)$, i.e. changes in the slope resp. intercept only as well as in both
- $m = 10, 20, 50$
- $T_m = Nm$ with $N = 1, 2, 5, 10$
- $k_m^\circ = \lfloor \vartheta m \rfloor$ 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 are also considered (each with $x_2(\cdot)$ i.i.d. $U[0,2]$):

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

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 1 000 bootstrap samples while the plots are based on 1 000 repetitions of the procedure. In Figs. 1, 2, 3 some selected size-power curves and density plots of the run length can be found.

It can be concluded from the simulations that both bootstrap methods perform better than the asymptotic closed-end procedure for small historic sample sizes:

- The bootstrap methods hold the level much better.
- The run-time of the bootstrap methods is a bit longer, however a smaller percentage of all rejections takes place before the change. This can especially be seen in Fig. 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. Fig. 3).
- All three methods become approximately equivalent for $m \geq 50$ (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 (Fig. 3, Scenarios 2 and 3), where one would expect the regression bootstrap to be better.
- The bootstrap methods become very close for $m \geq 20$.

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. (2004).

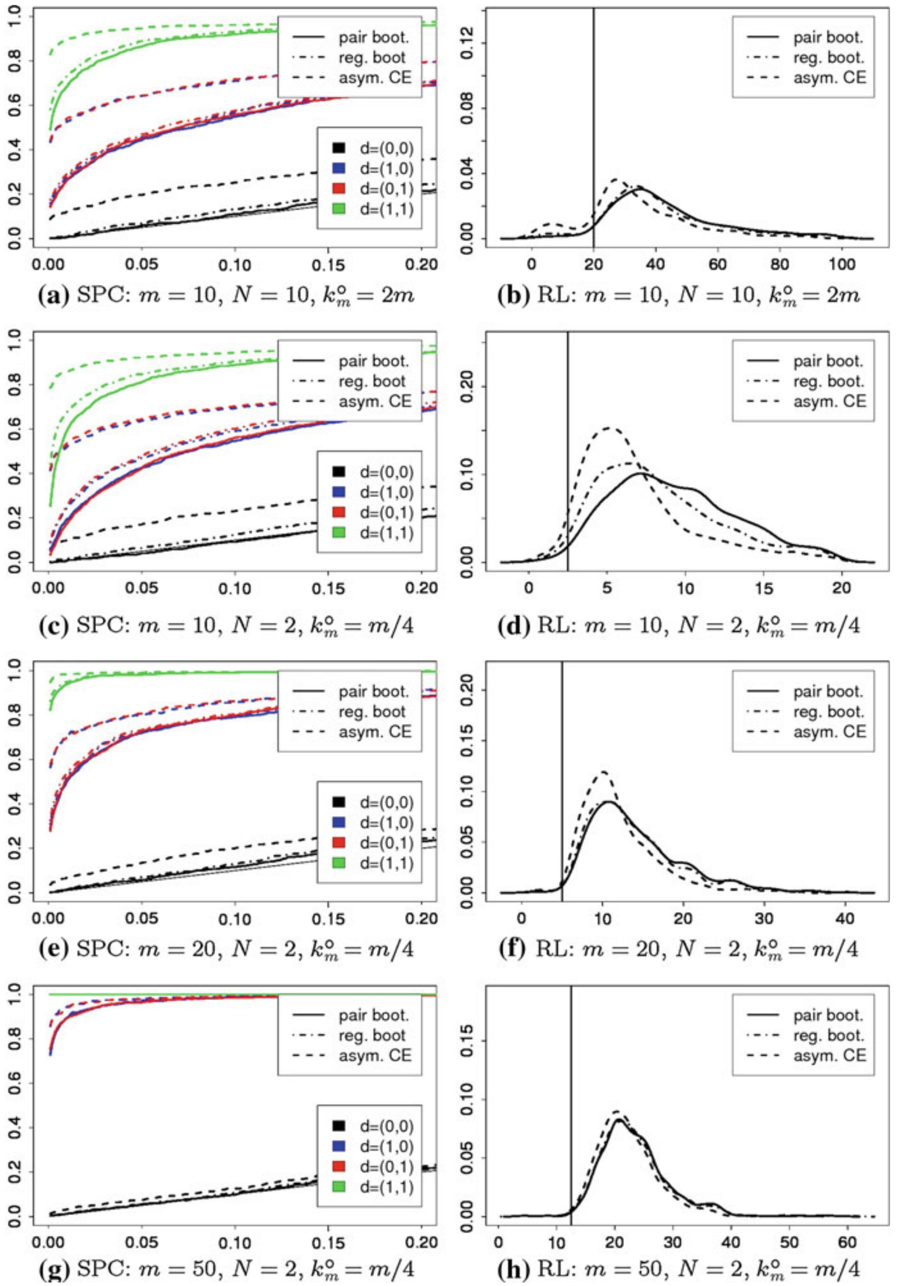
Lemma 1 *Under the Assumption $\mathcal{A}.1$ (ii) it holds as $m \to \infty$*

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

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

Our aim is to prove that the bootstrap critical values are uniformly asymptotically correct under the null hypothesis and bounded under alternatives (see Eqs. (45) resp. (46) below).

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



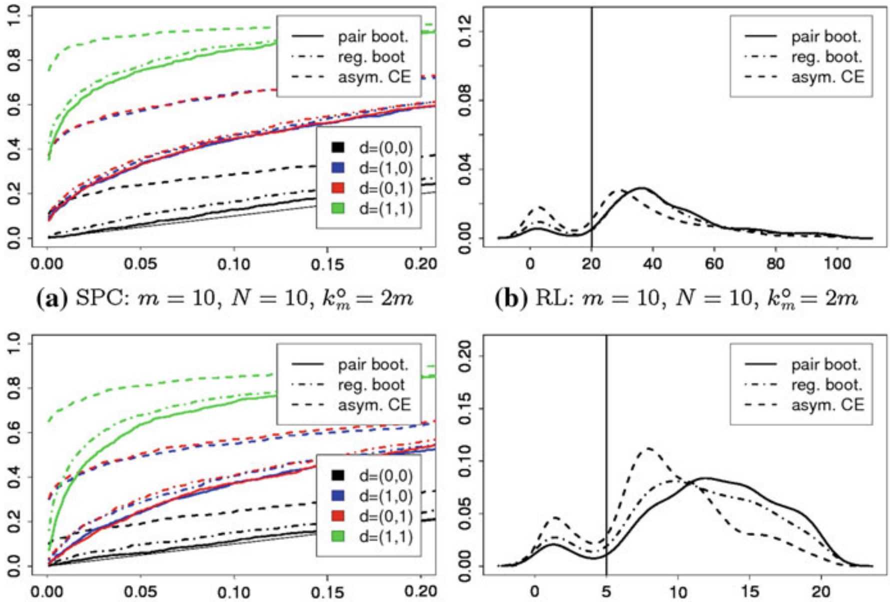
(a) SPC: $m = 10$, $N = 10$, $k_m^\circ = 2m$ **(b)** RL: $m = 10$, $N = 10$, $k_m^\circ = 2m$

(c) SPC: $m = 10$, $N = 2$, $k_m^\circ = m/4$ **(d)** RL: $m = 10$, $N = 2$, $k_m^\circ = m/4$

(e) SPC: $m = 20$, $N = 2$, $k_m^\circ = m/4$ **(f)** RL: $m = 20$, $N = 2$, $k_m^\circ = m/4$

(g) SPC: $m = 50$, $N = 2$, $k_m^\circ = m/4$ **(h)** RL: $m = 50$, $N = 2$, $k_m^\circ = m/4$

Fig. 1 Size-power curves and plots of estimated density: Scenario 1, centered exponential errors, $\gamma = 0$, in RL: $\mathbf{d} = (1, 1)$, $\alpha = 0.05$



(a) SPC: $m = 10$, $N = 10$, $k_m^\circ = 2m$ (b) RL: $m = 10$, $N = 10$, $k_m^\circ = 2m$

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

Fig. 2 Size-power curves and plots of estimated density: Scenario 1, centered exponential errors, $\gamma = 0.49$, in RL: $\mathbf{d} = (1, 1)$, $\alpha = 0.05$

Lemma 2 *Let c, $c_k(m)$ be such that $P(Y > c) = \alpha$, respectively, $P^*_{m,k}(Y_k(m) > c_k(m)) \le \alpha$ 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 $m \to \infty$)

$$\sup_{1\le k<\infty} \left|P^*_{m,k}(Y_k(m) \le x) - P\,(Y \le x)\right| \to 0 \qquad P-a.s. \tag{22}$$

Then, as $m \to \infty$,

$$\sup_{1\le k<\infty} |c_k(m) - c| \to 0 \qquad P-a.s. \tag{23}$$

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

$$\sup_{1\le k<\infty} \left|P^*_{m,k}(Y_k(m) \ge A)\right| \le \epsilon + o(1) \qquad P-a.s. \tag{24}$$

then

$$\sup_{1\le k<\infty} |c_k(m)| = O(1) \quad P-a.s. \tag{25}$$

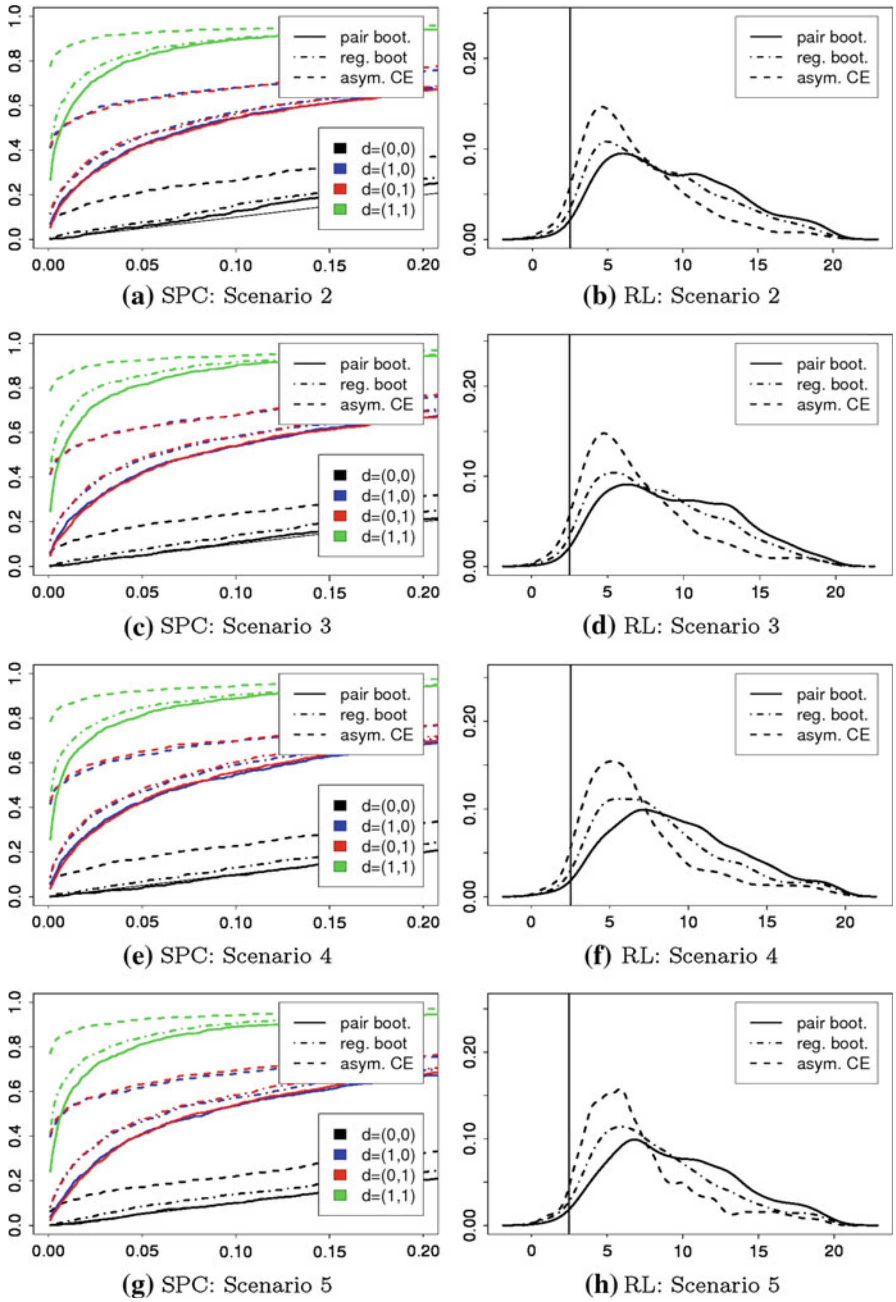


Fig. 3 Size-power curves and plots of estimated density: centered exponential errors, $\gamma = 0$, $m = 10$, $N = 2$, $k_m^{\circ} = m/4$, in RL: $\mathbf{d} = (1, 1)$, $\alpha = 0.05$

Proof For a) we refer to the proof of Kirch (2008), Lemma A.1. Concerning b), consider the set of all $\omega \in M$ with $P(M) = 1$ and such that (24) holds. We prove (25) for all $\omega \in M$ by contradiction. If (25) 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 $P^*_{m,f(\beta(m))}\left(Y_{f(\beta(m))}(m) > c_{f(\beta(m))}\right) \le \alpha$ we get by the minimality of $c_{f(\beta(m))}(m)$ by (24) that

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

which is a contradiction.

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

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

and the bootstrap variance (18). Note that

$$\widehat{e}_{m,k}(i) = e(i) - \boldsymbol{x}(i)^T \boldsymbol{C}_{m+k}^{-1} \sum_{j=1}^{m+k} \boldsymbol{x}(j)e(j)$$
$$+1_{\{i>m+k^\circ_m\}}\boldsymbol{x}(i)^T \boldsymbol{d}_m - 1_{\{k>k^\circ_m\}}\boldsymbol{x}(i)^T \boldsymbol{C}_{m+k}^{-1}\boldsymbol{C}_{k^\circ_m,k}\boldsymbol{d}_m, \tag{26}$$

and for $k > k^\circ_m$

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

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:

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

where

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

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

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

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

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

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

In order to prove (22) under H_0 resp. (24) 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 3 *Let (1) and Assumption A.1 hold true and either H_0 or $\mathbf{d}_m = O(1)$.*

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

$$(i) \quad \sup_{1\le k<\infty} P_{m,k}^* \left(\max_{1\le \ell < T_m+1} \frac{|I_3(m,k,\ell)|}{g(m,\ell,\gamma)} \ge \epsilon \right) \to 0 \quad P-a.s.,$$

$$(ii) \quad \sup_{1\le k<\infty} P_{m,k}^* \left(\max_{1\le \ell < T_m+1} \frac{|I_4(m,k,\ell)|}{g(m,\ell,\gamma)} \ge \epsilon \right) \to 0 \quad P-a.s.$$

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

$$(i) \quad \sup_{1\le k<\infty} P_{m,k}^* \left(\max_{1\le \ell < T_m+1} \frac{|I_5(m,k,\ell)|}{g(m,\ell,\gamma)} \ge \epsilon \right) \to 0 \quad P-a.s.,$$

$$(ii) \quad \sup_{1\le k<\infty} P_{m,k}^* \left(\max_{1\le \ell < T_m+1} \frac{|I_6(m,k,\ell)|}{g(m,\ell,\gamma)} \ge \epsilon \right) \to 0 \quad P-a.s.$$

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

$$(i) \quad \sup_{1\le k<\infty} P_{m,k}^* \left(\max_{1\le \ell < T_m+1} \frac{|I_5(m,k,\ell)|}{g(m,\ell,\gamma)} \ge A \right) \le \epsilon + o(1) \quad P-a.s.,$$

$$(ii) \quad \sup_{1\le k<\infty} P_{m,k}^* \left(\max_{1\le \ell < T_m+1} \frac{|I_6(m,k,\ell)|}{g(m,\ell,\gamma)} \ge A \right) \le \epsilon + o(1) \quad P-a.s.$$

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 $\mathrm{var}(Z) \le \mathrm{E}(Z^2)$)

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

Standard decoupling arguments yield

$$\sup_{k\ge 1}\frac{1}{m+k}\left|\sum_{j=1}^{m+k}\mathbf{x}(j)e(j)\right| = o(1) \qquad P-a.s., \tag{27}$$

since conditioned on $\{\mathbf{x}(\cdot)\}$ the sequence fulfills the Kolmogorov condition for a strong LLN. This together with Lemma 1 shows that, as $m \to \infty$,

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

Denote

$$Z_{m,k}(U_{m,k}(j)) = \boldsymbol{x}(U_{m,k}(j))^T \boldsymbol{C}_{m+k}^{-1} \sum_{v=1}^{m+k} \boldsymbol{x}(v)e(v) - \frac{1}{m+k}\sum_{j=1}^{m+k} e(j). \tag{28}$$

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

$$\begin{aligned} \mathrm{E}^*_{m,k}\, Z_{m,k}(U_{m,k}(1)) &= 0, \\ \sup_k \mathrm{var}^*_{m,k}\, Z_{m,k}(U_{m,k}(1)) &= o(1) \qquad P-a.s. \end{aligned} \tag{29}$$

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

$$g(m,\ell,\gamma) \ge \begin{cases} D_1 m^{1/2-\gamma}\,\ell^\gamma, & \ell \le m, \\ D_1 m^{-1/2}\,\ell, & \ell > m, \end{cases} \tag{30}$$

yielding for some $D_2 > 0$

$$\sum_{l=1}^{T_m} \frac{1}{g^2(m,\ell,\gamma)} \le D_1^{-2} m^{-1+2\gamma} \sum_{\ell=1}^{m} \frac{1}{\ell^{2\gamma}} + D_1^{-2} m \sum_{\ell=m+1}^{T_m} \frac{1}{\ell^2} \le D_2. \tag{31}$$

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

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

uniformly in k which finishes the proof of a)(i) by (29).

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 ℓ while the other one is (conditionally) random and does not depend on ℓ. 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 jth 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 1 a)

$$\begin{aligned}
&\mathrm{E}^*_{m,k}\left(\sqrt{m}\mathbf{u}_j^T\mathbf{C}_m^{-1}\sum_{i=1}^m \mathbf{x}(i)Z_{m,k}(U_{m,k}(i))\right)^2\\
&= \mathrm{var}^*_{m,k}\left(\sqrt{m}\sum_{i=1}^m \mathbf{u}_j^T\mathbf{C}_m^{-1}\mathbf{x}(i)Z_{m,k}(U_{m,k}(i))\right)\\
&= \left(\mathrm{var}^*_{m,k}\, Z_{m,k}(U_{m,k}(1))\right)\, \mathbf{u}_j^T\, m\mathbf{C}_m^{-1}\, \sum_{i=1}^m \mathbf{x}(i)\mathbf{x}(i)^T\, \mathbf{C}_m^{-1}\,\mathbf{u}_j\\
&= \left(\mathrm{var}^*_{m,k}\, Z_{m,k}(U_{m,k}(1))\right)\, \mathbf{u}_j^T\, (\mathbf{C}^{-1} + o(1))\,\mathbf{u}_j = o(1) \qquad P-a.s. \qquad (32)
\end{aligned}$$

uniformly in k. By Lemma 1 b) in addition to (30) we additionally get for some $D_3 > 0$

$$\sup_{k\geq 1}\ \sup_{1\leq \ell < T_m+1}\left\|\frac{\mathbf{c}_1(m,k,\ell)^T}{\sqrt{m}\, g(m,\ell,\gamma)}\right\|_\infty \leq D_3 + o(1) \qquad P-a.s. \qquad (33)$$

Together this yields that

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

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

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

Direct calculations give

$$\begin{aligned}
&\mathrm{E}^*_{m,k}\,\widetilde{Z}_{m,k}(U_{m,k}(i)) = 0,\\
&\mathrm{E}^*_{m,k}\left(\widetilde{Z}_{m,k}(U_{m,k}(i))\right)^2 = \frac{1}{m+k}\boldsymbol{d}_m^T(\boldsymbol{C}_{k_m^\circ,k} - \boldsymbol{C}_{k_m^\circ,k}\boldsymbol{C}_{m+k}^{-1}\boldsymbol{C}_{k_m^\circ,k})\boldsymbol{d}_m \qquad (34)\\
&\leq \frac{1}{m+k}\boldsymbol{d}_m^T\boldsymbol{C}_{k_m^\circ,k}\boldsymbol{d}_m \leq \frac{1}{m+k}\boldsymbol{d}_m^T\boldsymbol{C}_{m+k}\boldsymbol{d}_m \leq \boldsymbol{d}_m^T\boldsymbol{C}\boldsymbol{d}_m\ (1+o(1)) \qquad P-a.s.
\end{aligned}$$

uniformly in k by Lemma 1 a).

Assertions b) (i) and c) (i) follow now by the Hájek-Rényi inequality and (31).

Furthermore $(P-a.s.)$

$$\mathrm{E}^*_{m,k}\left(\sqrt{m}\,\mathbf{u}_j^T\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)\widetilde{Z}_{m,k}(U_{m,k}(j))\right)^2\leq \boldsymbol{d}_m^T\mathbf{C}\boldsymbol{d}_m\,\mathbf{u}_j^T\mathbf{C}^{-1}\mathbf{u}_j\ (1+o(1)) \tag{35}$$

uniformly in k. This, in addition to (33), 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 4 *Let (1) and Assumption $\mathcal{A}.1$ hold true. Then for all $\epsilon>0$ it holds*

$$\sup_{1\leq k<\infty}P^*_{m,k}\left(\max_{1\leq \ell<T_m+1}\frac{|I_2(m,k,\ell)-\frac{-\ell}{m}\sum_{j=1}^{m}e(U_{m,k}(j))|}{g(m,\ell,\gamma)}\geq\epsilon\right)\to 0\qquad P-a.s.$$

Proof The proof is a slight modification of Lemma 5.2 in Horváth et al. (2004). Denote

$$\bar{e}_{m,k}=\frac{1}{m+k}\sum_{i=1}^{m+k}e(i),\qquad \widehat{\sigma}^2_{m,k}=\frac{1}{m+k}\sum_{i=1}^{m+k}\left(e(i)-\bar{e}_{m,k}\right)^2. \tag{36}$$

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

$$\sup_k|m^{1/2-(\rho-\gamma)}\bar{e}_{m,k}|\to 0\quad P-a.s.,\qquad \sup_k|\widehat{\sigma}^2_{m,k}-\sigma^2|\to 0\quad P-a.s., \tag{37}$$

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

$$\begin{aligned}
&\mathrm{E}^*_{m,k}\left(\frac{1}{m^{1/2+\rho-\gamma}}\mathbf{u}_j^T\sum_{i=1}^{m}\mathbf{x}(i)e(U_{m,k}(i))\right)^2\\
&=\frac{1}{m^{1+2(\rho-\gamma)}}\sum_{i=1}^{m}(\mathbf{u}_j^T\mathbf{x}(i))^2\,\mathrm{var}^*\left(e(U_{m,k}(1))\right)\\
&\quad+\left(\frac{\mathbf{u}_j^T}{m^{1/2+\rho-\gamma}}\sum_{i=1}^{m}\mathbf{x}(i)\,\mathrm{E}^*(e(U_{m,k}(1))\right)^2\\
&=\frac{1}{m^{1+2(\rho-\gamma)}}\mathbf{u}_j^T\mathbf{C}_m\mathbf{u}_j\,\widehat{\sigma}^2_{m,k}+(\mathbf{u}_j^T\mathbf{c}_1+o(1))^2\left(m^{1/2-(\rho-\gamma)}\bar{e}_{m,k}\right)^2\\
&=o(1)\qquad P-a.s.
\end{aligned} \tag{38}$$

By Lemmas 1 and (30) we get

$$\sup_k \sup_{1\le \ell < T_m+1} \left\| \frac{\mathbf{c}_1(m,k,\ell)^T\, mC_m^{-1} - \ell \mathbf{c}_1^T \mathbf{C}^{-1}}{m^{1/2-\rho+\gamma}\, g(m,\ell,\gamma)} \right\|_\infty$$

$$= O(1) \sup_{1\le \ell < T_m+1} \left| \frac{(m+\ell)^{1-\rho} + \ell m^{-\rho}}{m^{1/2-\rho+\gamma}\, g(m,\ell,\gamma)} \right| = O(1) \qquad P-a.s.$$

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 5 *Let (1) and Assumption $\mathcal{A}.1$ hold true and either H_0 or $\boldsymbol{d}_m = O(1)$, T_m is as in Assumption $\mathcal{A}.2$. Let $\widehat{\sigma}^2_{m,k}$ as in (36).*

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

$$\sup_{1\le k<T_m+1} \left| P^*_{m,k}\left(\frac{1}{\widehat{\sigma}_{m,k}} \sup_{1\le \ell < T_m+1} \frac{\widetilde{\Gamma}(m,\ell,\gamma)(e^*_{m,k}(1),\ldots,e^*_{m,k}(m+\ell))}{g(m,\ell,\gamma)} \le x \right) \right.$$
$$\left. - P\left(\sup_{1\le k<T_m+1} \frac{|W_1(k/m) - k/m W_2(1)|}{(1+k/m)(k/(k+m))^\gamma} \le x \right) \right| \to 0 \qquad P-a.s.$$

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

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

Proof The proof of Theorem 2.3 in Kirch (2008) shows that (note that this corresponds to the null hypothesis there)

$$\sup_{1\le k<T_m+1} \left| P^*_{m,k}\left(\frac{1}{\widehat{\sigma}_{m,k}} \sup_{1\le \ell < T_m+1} \frac{\left| \sum_{i=m+1}^{m+\ell} \left(e(U_{m,k}(i)) - \frac{1}{m}\sum_{j=1}^m e(U_{m,k}(j)) \right) \right|}{g(m,\ell,\gamma)} \le x \right) \right.$$
$$\left. - P\left(\sup_{1\le k<T_m+1} \frac{|W_1(k/m) - k/m W_2(1)|}{(1+k/m)(k/(k+m))^\gamma} \le x \right) \right| \to 0 \qquad P-a.s.$$

Putting this together with Lemmas 3 and 4, as well as (37) 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 *Let (1) and Assumption $\mathcal{A}.1$ hold true. For any $\frac{1}{2} < \xi < 1$ and any $\epsilon > 0$ $(P - a.s.)$*

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

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

If for the pair bootstrap additionally Assumption $\mathcal{A}.4$ holds, then we even get the assertions for $\xi = \frac{1}{2}$.

Proof By the von Bahr–Esseen inequality (cf. Theorem 3 in (1965)) with $1/\xi$ we get for some constant $D > 0$ and for any $c > 0$

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

since conditioned on $\{\mathbf{x}(\cdot)\}$ the sequence fulfills condition (1) in Theorem 5.2.1 in Chow and Teicher (1997) similarly to (27). This proves a) but b) is analogous.

The same arguments also holds for $\xi = \frac{1}{2}$ if the stronger Assumption $\mathcal{A}.4$ holds.

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

Lemma 7 *Let (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 (36).*

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

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

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

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

Proof By (26) it holds

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

where

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

where $\bar{e}_{m,k}$ is as in (36). 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 $k \leq k_m^\circ$.

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 (2008) (this corresponds to the null hypothesis there) it holds

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

which in addition to (37) yields

$$\sup_{k \geq 1} P_{m,k}^* \left(\left| \frac{\frac{1}{m-p} \sum_{i=1}^{m} J_1^2(m,k,i)}{\widehat{\sigma}_{m,k}^2} - 1 \right| \geq \epsilon \right) \to 0 \qquad P - a.s. \tag{39}$$

Denote by $\mathbf{u}_j$ again the jth unit vector. By Lemma 1 and (38) it holds by the Cauchy–Schwarz inequality

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

which yields by an application of the Markov inequality

$$\sup_{k\geq 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^{m} J_2^2(m,k,i)\geq \epsilon\right)\to 0 \qquad P-a.s. \tag{40}$$

Noting that $J_3(m,k,i) = -Z_{m,k}(U_{m,k}(i))$ as in the proof of Lemma 3, hence by an application of the Markov inequality in addition to (29)

$$\sup_{k\geq 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^{m} J_3^2(m,k,i)\geq \epsilon\right)\to 0 \qquad P-a.s. \tag{41}$$

Similarly by (32)

$$\begin{aligned}
&\sup_{k\geq 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^{m} J_4^2(m,k,i)\geq \epsilon\right)\\
&\quad\leq \sup_{k\geq 1}\frac{1}{\epsilon}\mathrm{E}^*_{m,k}\left(\frac{1}{m-p}\sum_{j=1}^{m}\mathbf{x}(j)^T Z_{m,k}(U_{m,k}(j))\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)Z_{m,k}(U_{m,k}(j))\right)\\
&\quad\leq \frac{1}{\epsilon}\left\|\frac{1}{m}\mathbf{C}_m\right\|_\infty \frac{p^2}{m-p}\sup_{k\geq 1}\sup_{j=1,\ldots,p}\mathrm{E}^*_{m,k}\left(\sqrt{m}\mathbf{u}_j^T\mathbf{C}_m^{-1}\sum_{v=1}^{m}\mathbf{x}(v)Z_{m,k}(U_{m,k}(v))\right)^2\\
&\quad\to 0 \qquad P-a.s.
\end{aligned} \tag{42}$$

Putting (39) to (41) together with (37) yields assertion a) under H_0.

Now we prove the assertions under alternatives. First, note that it is sufficient to consider $k > k_m^\circ$, since otherwise J_5 and J_6 are equal to 0. First we prove that J_6 is negligible: By (35) it holds

$$\begin{aligned}
&\sup_{k\geq 1} P^*_{m,k}\left(\frac{1}{m}\sum_{i=1}^{m} J_6^2(m,k,i)\geq \epsilon\right)\\
&\quad\leq \frac{1}{\epsilon}\frac{p^2}{m-p}\left\|\frac{1}{m}\mathbf{C}_m\right\|_\infty \sup_{j=1,\ldots,p}\mathrm{E}^*_{m,k}\left(\sqrt{m}\mathbf{u}_j^T\mathbf{C}_m^{-1}\sum_{j=1}^{m}\mathbf{x}(j)\widetilde{Z}_{m,k}(U_{m,k}(j))\right)^2\\
&\quad= o(1) \qquad P-a.s.
\end{aligned} \tag{43}$$

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

$$\sup_{k\geq 1} P^*_{m,k}\left(\frac{1}{m}\sum_{i=1}^{m} J_5^2(m,k,i)\geq c\right)\leq \frac{\mathbf{d}_m^T\mathbf{C}\mathbf{d}_m}{c^2}+o(1) \qquad P-a.s. \tag{44}$$

by (34), since $J_5(m,k,i) = \widetilde{Z}_{m,k}(U_{m,k}(i))$ as in the proof of Lemma 3. This yields a) under local alternatives.

For fixed alternatives note that

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

The square terms are negligible except for J_1^2 and J_5^2 by (40)–(43). By the Cauchy–Schwarz inequality, (39) and (44) the same holds true for the mixed terms except $J_1\, J_5$ but the latter one is also negligible due to Lemmas 1 and 6 since

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

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 $\frac{1}{m}\sum_{i=1}^{m}(J_1^2(m,k,i)+J_5^2(m,k,i)) \geq \frac{1}{m}\sum_{i=1}^{m}J_1^2(m,k,i)$ assertion b) follows by (39).

Putting the above lemmas together we easily obtain Theorem 1.

Proof of Theorem 1 Putting together Lemmas 5 and 7 we obtain under H_0 as well as local alternatives

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

By Lemma 2 this yields

$$\sup_{k\geq 1}|c_{m,k}^{(R)}-c|\to 0 \qquad P-a.s., \tag{45}$$

where c is the asymptotic critical value obtained from the distribution of $\sup_{0\leq t\leq 1-\frac{1}{N+1}}\frac{|W(t)|}{t^\gamma}$. Together with (11) this implies a).

Under H_1, by Lemmas 5 and 7 for every $\epsilon>0$ there exists a constant $A>0$ such that $(P-a.s.)$

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

By Lemma 2 this yields

$$\sup_{k\geq 1} |c_{m,k}^{(R)}| = O(1) \qquad P-a.s. \tag{46}$$

Together with (12) this implies b).

7 Proofs of Section 4

Denote

$$\begin{aligned}
A_1(m,k,\ell) &= \sum_{i=m+1}^{m+\ell} e^*_{m,k}(i)\\
\boldsymbol{A}_2(m,k,\ell) &= \sum_{i=m+1}^{m+\ell} \boldsymbol{x}^*_{m,k}(i)\\
\boldsymbol{A}_3(m,k) &= \sum_{j=1}^{m} \boldsymbol{x}^*_{m,k}(j)\boldsymbol{x}^{*T}_{m,k}(j)\\
\boldsymbol{A}_4(m,k) &= \sum_{s=1}^{m} \boldsymbol{x}^*_{m,k}(s)e^*_{m,k}(s)\\
\mathbf{A}_5(m,k,\ell) &= \sum_{i=m+1}^{m+\ell} \boldsymbol{x}^*_{m,k}(i)1_{\{U_{m,k}(i)>m+k_m^\circ\}}\\
\boldsymbol{A}_6(m,k) &= \sum_{s=1}^{m} \boldsymbol{x}^*_{m,k}(s)\boldsymbol{x}^{*T}_{m,k}(s)\mathbf{1}_{\{U_{m,k}(s)>m+k_m^\circ\}}
\end{aligned}$$

where $e^*_{m,k}(i) = e(U_{m,k}(i))$, which is different from the bootstrapped residuals in the regression bootstrap. Similarly to (14) 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

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

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

Lemma 8 *Let (1) and Assumption $\mathcal{A}.1$ hold true.*

a) *Under either Assumption $\mathcal{A}.3$ or $\mathcal{A}.4$ we get for any $\epsilon > 0$*

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

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

b) *Under either Assumption $\mathcal{A}.3$ or $\mathcal{A}.4$ we get for any $\epsilon > 0$*

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

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

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

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

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

which yields

$$\|\mathbf{x}(j)\mathbf{x}(j)^T\|_\infty = O(j^{1-\rho}) \qquad P-a.s. \tag{47}$$

Further note that

$$\mathrm{E}^*_{m,k}(\mathbf{x}^*_{m,k}(i)\mathbf{x}^*_{m,k}(i)^T) = \frac{1}{m+k}\sum_{j=1}^{m+k}\mathbf{x}(i)\mathbf{x}(i)^T = \mathbf{C} + O(m^{-\rho}) \qquad P-a.s. \tag{48}$$

uniformly in k by Assumption $\mathcal{A}.1$. This, (47) and an application of the Chebyshev inequality yields now (the square of the matrix is meant componentwise)

$$\begin{aligned}
&\sup_{1\le k<T_m+1} P^*_{m,k}\left(\frac{1}{m^{1-\eta}}\left\|\sum_{i=1}^{m}(\mathbf{x}^*_{m,k}(i)\mathbf{x}^*_{m,k}(i)^T-\mathbf{C})\right\|_\infty \ge \epsilon\right)\\
&\le O(m^{2\eta-2\rho}) + O(1)\,\frac{1}{m^{1-2\eta}}\sup_{1\le k<T_m+1}\frac{1}{m+k}\left\|\sum_{i=1}^{m+k}(\mathbf{x}(i)\mathbf{x}(i)^T)^2\right\|_\infty\\
&\le O(m^{2\eta-2\rho}) + O\left(\frac{(m+T_m)^{1-\rho}}{m^{1-2\eta}}\right)\sup_{k\ge1}\sup_{j=1,\dots,p}\frac{1}{m+k}\sum_{i=1}^{m+k}x_j^2(i)\\
&= O(m^{2\eta-2\rho}+m^{2\eta-\epsilon}) = o(1) \qquad P-a.s.
\end{aligned}$$

under Assumption $\mathcal{A}.3$ for some $\rho > 0$, which yields a). A similar argument but using $\mathcal{A}.4$ and the von Bahr–Esseen inequaltiy (cf. Theorem 3 in (1965)) also yields assertion a).

Analogously we obtain b).

The next lemma is the analogue to Lemmas 3 and 4 for the regression bootstrap.

Lemma 9 *Let (1) and Assumption $\mathcal{A}.1$ hold true and either H_0 or $\boldsymbol{d}_m = O(1)$.*

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

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

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

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

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

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

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

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

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

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

Direct calculations give

$$E^*_{m,k}\boldsymbol{A}_2^T(m,k,\ell)(E^*_{m,k}\boldsymbol{A}_3(m,k))^{-1} = \frac{\ell}{m+k}\sum_{i=1}^{m+k}\boldsymbol{x}(i)\left(\frac{m}{m+k}\sum_{i=1}^{m+k}\boldsymbol{x}(i)\boldsymbol{x}(i)^T\right)^{-1}$$
$$= \frac{\ell}{m}(1,0,\ldots,0)^T \tag{49}$$

Therefore

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

By Lemmas 6 and 8 as well as (48) we get for any $\epsilon > 0$

$$\sup_{k\geq 1} P^*_{m,k}\left(m^{\eta-1}\|\mathbf{A}_3(m,k) - \mathrm{E}^*_{m,k}\,\mathbf{A}_3(m,k)\|_\infty \geq \epsilon\right) \to 0 \qquad P-a.s. \tag{50}$$
$$\sup_{k\geq 1} P^*_{m,k}\left(m^{1-\xi}\|\mathbf{A}_3(m,k)^{-1}\mathbf{A}_4(m,k)\|_\infty \geq \epsilon\right) \to 0 \qquad P-a.s. \tag{51}$$

for some $\eta > 0$ and for ξ as in Lemma 8.

By (30) we get

$$\sup_{\ell\geq 1}\frac{\ell}{m\, g(m,\ell,\gamma)} = O(m^{-1/2}), \tag{52}$$

which together with (49), (50) and (51) yields

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

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

$$\mathrm{var}^*_{m,k}(\mathbf{x}^*_{m,k}(i)) \leq \frac{1}{m+k}\sum_{i=1}^{m+k}\mathbf{x}(i)\mathbf{x}(i)^T = \mathbf{C} + o(1) \qquad P-a.s. \tag{53}$$

uniformly in k. An application of the Hájek-Rényi inequality, (31) and (51) yields

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

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 $k \geq 1$

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

An application of the Hájek-Rényi inequality, (31) and Lemma 8 yields for any $c > 0$

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

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

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

Since

$$\mathrm{E}^*_{m,k} \mathbf{x}^*_{m,k}(1) \mathbf{x}^*_{m,k}(1)^T 1_{\{U_{m,k}(1) > m+k^\circ_m\}} = \frac{1}{m+k} \sum_{i=m+k^\circ_m}^{m+k} \mathbf{x}(i)\mathbf{x}(i)^T$$
$$\leq D + o(1) \qquad P - a.s. \tag{54}$$

uniformly in k, we obtain from Lemma 8 that for each $\epsilon > 0$ there exists $A > 0$ such that

$$\sup_{k \geq 1} P^*_{m,k} \left(\|\mathbf{A}_3(m,k)^{-1} \mathbf{A}_6(m,k)\|_\infty \geq A \right) \leq \epsilon + o(1) \qquad P - a.s. \tag{55}$$

This in addition to an application of the Hájek-Rényi inequality and (31) yields for any $c > 0$

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

By (49) we get

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

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

$$\sup_{k\ge 1} P^*_{m,k}\left(\frac{1}{\sqrt{m}}\left|\sum_{j=1}^{m}(\mathbf{x}^*_{m,k}(j)^T 1_{\{U_{m,k}(j)>m+k_m^\circ\}}-\mathrm{E}^*\,\mathbf{x}^*_{m,k}(j)^T 1_{\{U_{m,k}(j)>m+k_m^\circ\}})\mathbf{d}_m\right|\ge c\right)$$
$$\le\frac{\|\mathbf{d}_m\|_\infty^2}{c^2}\sup_{k\ge 1}\frac{1}{m+k}\left\|\sum_{i=m+k_m^\circ}^{m+k}\mathbf{x}(i)\mathbf{x}(i)^T\right\|_\infty\le D\,\frac{\|\mathbf{d}_m\|_\infty^2}{c^2}+o(1)\qquad P-a.s.\tag{56}$$

Together with (52) this yields

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

Finally by (49)

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

By (52), (55) and an analogous argument to (56) using (53) we finally obtain

$$\sup_{1\le k<T_m+1} P^*_{m,k}\left(\max_{1\le \ell<T_m+1}\frac{|B_{4,3}(m,k,\ell)|}{g(m,\ell,\gamma)}\ge c\right)\le \|C\|_\infty\,\frac{\|\mathbf{d}_m\|_\infty^2}{c^2}+o(1)\quad P-a.s.,$$

which completes the proof. □

Now, we prove the equivalent of Lemma 7.

Lemma 10 *Let (1), Assumption $\mathcal{A}.1$, and either Assumption $\mathcal{A}.3$ or $\mathcal{A}.4$ hold true. Let $\widehat{\sigma}^2_{m,k}$ be as in (36).*

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

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

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

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

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

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

By (39) it holds $(D_1(m,k,i) = J_1(m,k,i))$

$$\sup_{k\geq 1} P^*_{m,k}\left(\left|\frac{\frac{1}{m-p}\sum_{i=1}^m D_1^2(m,k,i)}{\widehat{\sigma}^2_{m,k}} - 1\right| \geq \epsilon\right) \to 0 \qquad P-a.s. \tag{57}$$

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 and 8

$$\sup_{k\geq 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^m D_2^2(m,k,i) \geq \epsilon\right) \to 0 \qquad P-a.s. \tag{58}$$

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

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

which implies due to Lemma 8 and (54) for every $c > 0$ for some constant $D > 0$

$$\sup_{k\geq 1} P^*_{m,k}\left(\frac{1}{m-p}\sum_{i=1}^{m} D_3^2(m,k,i) \geq c\right) \leq D\,\frac{\|\mathbf{d}_m\|_\infty^2}{c^2} + o(1) \qquad P-a.s. \quad (59)$$

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

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

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

which is also negligible due to Lemmas 6 and 8. We can now finish the proof for fixed alternatives analogously to the proof of Lemma 7.

Proof of Theorem 2 Due to Lemma 9 we obtain the analogous assertion for the pair bootstrap to what is given for the regression bootstrap in Lemma 5. We can then conclude as in the proof of Theorem 1 using Lemma 10.

References

Andreou E, Ghysels E (2006) Monitoring disruptions in financial markets. J Economet 135:77–124

Antoch J, Hušková M (2001) Permutation tests for change point analysis. Stat Probab Lett 53:37–46

Aue A, Horváth L, Hušková M, Kokoszka P (2006) Change-point monitoring in linear models. Economet J 9:373–403

Aue A, Hörmann S, Horváth L, Hušková M (2009) Sequential testing for the stability of portfolio betas. In preparation

Berkes I, Horváth L, Hušková M, Steinebach J (2004) Applications of permutations to the simulations of critical values. J Nonparametr Stat 16:197–216

Chow YS, Teicher H (1997) Probability Theory—Independence, Interchangeability, Martingales. 3. Springer, New York

Chu C-SJ, Stinchcombe M, White H (1996) Monitoring structural change. Econometrica 64:1045–1065

Fried R, Imhoff M (2004) On the online detection of monotonic trends in time series. Biom J 46:90–102

Good P (2005) Permutation, Parametric, and Bootstrap Tests of Hypothesis. 3. Springer, New York

Horváth L, Hušková M, Kokoszka P, Steinebach J (2004) Monitoring changes in linear models. J Stat Plann Inference 126:225–251

Hušková M (2004) Permutation principle and bootstrap in change point analysis. Fields Inst Commun 44:273–291

Hušková M, Koubková A (2005) Monitoring jump changes in linear models. J Stat Res 39:59–78

Hušková M, Koubková A (2006) Sequential procedures for detection of changes in autoregressive sequences. In: Hušková M, Lachout P (eds) Proceedings of the Prague stochstics, pp 437–447

Kirch C (2008) Bootstrapping sequential change-point tests. Seq Anal 27:330–349

Koubková A (2008) Change detection in the slope parameter of a linear regression model. Tatra Mt Math Publ 39:245–253

Steland A (2006) A bootstrap view on Dickey-Fuller control charts for AR(1) series. Aust J Stat 35:339–346

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