Change-Point Procedures for Multivariate Dependent Data

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

an der Fakultät für Mathematik
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Dipl.-Math. oec. Silke Martina Weber
aus Karlsruhe

Tag der mündlichen Prüfung: 8. Februar 2017

Referentin: Prof. Dr. Claudia Kirch
Korreferent: Prof. Dr. Norbert Henze

Danksagung

Als erstes möchte ich meiner Doktormutter Prof. Dr. Claudia Kirch für die kontinuierliche Unterstützung meiner Arbeit danken, durch permantentes Lesen von Beweisen und ausführliche Korrekturen, sowie durch regelmäßige Gespräche über mathematische Probleme als sie noch am KIT war. Nachdem sie an die Universität Magdeburg gewechselt hat, bestand die kontinuierliche Unterstützung weiter durch Telefonate und regen E-Mail Kontakt. Aber ich möchte mich auch dafür bedanken, dass sie mich schon während des Studiums gefördert hat bis hin zur Promotion.

Zudem möchte ich meinem Korreferenten Herrn Henze danken für seine vielen sprachlichen Verbesserungen und Tipps bezüglich wissenschaflichen Schreibens in Englisch.

Dem ganzen Institut für Stochastik möchte ich danken für die angenehme Atmosphäre in den letzten Jahren. Insbesondere möchte ich Anton Popp für die seelische und moralische Unterstützung danken und meinem Zimmerkollegen Fabian Gieringer für die gute Versorgung.

Auch bedanken möchte ich mich bei meinen ehemaligen Zimmerkollegen im Allianzgebäude Ines Ziebarth, Julia Hörmann und Christina Stöhr für eine angenehme Atmosphäre.

Weiter danke ich meinen Eltern für ihre stetige Unterstützung und ihre unkomplizierte Art. Außerdem möchte ich meinem Fabian danken.

Karlsruhe, Januar 2017

Meine Stelle wurde vom Ministerium für Wissenschaft, Forschung und Kunst vom Land Baden-Württemberg durch Az. 7635.52/Kirch/1 finanziert.

Mein Aufenthalt an der Universität Lancaster (GB) wurde vom Karlsruhe House of Young Scientists (KHYS) finanziell unterstützt.

Contents

List of Figures

Introduction

In recent years, the analysis of data ordered in time becomes more and more popular due to the increasing amount of automatically collected data over time in many different fields, for example in economics, banks, medicine or engineering. Data collected over time are called time series, and an important question is whether the structure of such data is changing over time or whether the model still describe the data well enough. A change in the data structure means a change in the unknown model parameter of the underlying data set. The most popular changes are changes in the mean or in the variance. The points in time in which changes occur, regardless of the special kind of change, are called the change-points. The main interests are to decide by means of suitable change-point tests if there are changes in the collected data set as well as to locate the changes as precisely as possible. The mathematical field of searching changes in a data structure is called change-point analysis.

There are two different kinds of change point procedures, the sequential method and the a-posteriori method. In the first case we collect data sequentially and test after each new observation whether a change has occurred or not. The procedure stops if the test detects a change. In case of the a-posteriori method a complete data set has been observed without collecting new observations. Then we search for a change over time in the given data set. Using this method we obtain estimators for the change-points. In case of the sequential method we want to detect the change very fast after it has occurred. The procedure mostly stops with delay after the change has happened, but sometimes it stops earlier as the change occurs. The latter case is a false alarm, and this is discussed in detail in the simulation study in the first part of this work. The time period of monitoring the data till the test stops is called the run length.

If the sequential procedure is used, there are the options to monitor new incoming observations during a given time period, i.e. the time horizon is finite, or the time horizon is not restricted in time, which means that the monitoring period is infinite. The first method is called the closed-end procedure and the second one the open-end procedure.

The sequential procedure is useful in situations in which it is important to react as soon as possible after a change has occurred. Examples are the monitoring of medical observations. Here, it is important that the doctors intervene very fast after a change happened. Furthermore, stock prices are collected permanently and the stockbroker has to adapt the purchase and sale strategy immediately after the change in the stock prices.

If a statistical test is conducted, we want to decide whether the null hypothesis of no changes should be rejected or whether the data still follow the proposed model. Therefore we first have to calculate the statistic, and we need critical values that decide about rejection or acceptance of the null hypothesis. Usually, change-point analysis employs asymptotic procedures. Here, critical values are determined by the asymptotic distribution of the test statistic under the null hypothesis. Further there is another way to obtain critical values, the bootstrap method. Here, a new data set with the same length as the observed data set is created by sampling with replacement from the observations. This proceeding is replicated very often to mimic the distribution of the test statistic. The obtained time series are called the bootstrap series. Based on the resulting bootstrap series the critical values are calculated. This procedure usually works better with small sample sizes, and it is necessary in higher dimensions where the estimator of the covariance matrix, involved in the test statistic, is not stable anymore.

This dissertation is divided in three parts. The first part studies the sequential procedure, and the other two parts deal with the a-posteriori method whereat one of the last to parts focuses on bootstrap methods.

In the first part we introduce a general setting of sequential testing of changes in possibly multivariate time series. Contrary to the existing literature, we do not suppose a specific data model, but we only propose some weak assumptions which are fulfilled by the data models already discussed in the literature and actually extend the theory to many new data models which have not yet been considered. Kirch and Kamgaing (2015) is the only publication that has already introduced such a common setting by using the most popular CUSUM (cumulative sum) statistic. We will extend it to the modified MOSUM (moving sum), Page-CUSUM and standard MOSUM statistic which are useful in situations where the CUSUM does not work very well, for example for late changes with respect to the time horizon. The different behaviour of the statistics depending on the time point of change and some other parameters is illustrated by simulations.

Afterwards we switch to the a-posteriori procedure and concentrate on different types of bootstrap methods in a multivariate mean change model, namely the well-known block-bootstrap, the wild-bootstrap and the AR-sieve-bootstrap. These procedures differ in the way of sampling with replacement from the observed data set to obtain the bootstrap series. The block-bootstrap has already been discussed in the literature but only in the univariate case. We develop the theory that the bootstrap series provide useful critical values for the multivariate block-bootstrap. In addition, there follows an extensive simulation study with the mentioned bootstrap methods to compare them concerning the empirical size and power.

In the third part we again suppose the a-posteriori procedure for multivariate series but with an epidemic mean change, which possibly occurs in the components, i.e. at some point in time the mean of a component changes to a higher or lower level, and later returns back again. There is some literature about epidemic changes even for multivariate time series. However, the procedures only allow the changes to occur in each component at the same time point. In our approach the epidemic mean changes need not occur at the same time point in each component, but there is a functional relationship between the changes in the different components. To locate these changes we use the standard multivariate statistic and a projection statistic, for which the multivariate data have to be projected first. An application of this change-point procedure is locating sources of gas emissions. Somewhere under the ground is a gas source emitting gas which is carried jointly with the wind and thus a gas plume arises. Then data are collected from inside and outside of the gas plume along a flight path of a plane. The gas concentration inside is higher than outside of the plume. So the observed data can be transformed into a multivariate time series, in which the change-points are the points where the gas concentration jumps to a higher level and decreases back when the plane flies out of the plume. The performance of our procedure is discussed in a simulation study, and we apply our method to a real data example.

In the Appendix the assumptions and propositions under the null hypothesis used in the first part about the sequential procedure are merged as an overview to ease the reading of the corresponding sections. Moreover, we state some results of probability theory.

Part I.

Multivariate Sequential Procedures

1. Introduction to the Sequential Setting

The procedure of observing data sequentially and stopping as soon as a change has been detected, is clearly useful when we want to detect changes very quickly after they have occurred. An example in economics is the fluctuation of prices at the stock market, which should be detected quickly to conform the strategies for purchase and sale. It can also be helpful in medicine where data of patients are collected automatically and the doctors have to act very fast after a rapid change has occurred. But even in engineering it is a useful method for instance to detect increases in exhaust gas values.

We use the approach of Chu et al. (1996), which uses the assumption of a historical data set without a change to estimate the unknown parameter in the supposed model. Afterwards we start monitoring and search for a change of this parameter. To develop the asymptotics under the null hypothesis of no change as well as under the alternative of an existing change in the unknown parameter, the length of the historical data set grows to infinity. We treat both the closed-end and the open-end procedure. The first method has a finite time horizon in which the change is searched and the second one has no time limit to monitor the new incoming data.

The asymptotic distribution under the null hypothesis is used to control the asymptotic type-I-error, where the test stops and falsely detects a change. Under the alternative that there is a change in the time series, the procedure should have asymptotic power one, i.e. the change is asymptotically always detected.

The first statistic that was analysed in the sequential setup was the CUSUM statistic that uses all observations after the historical data set for the summation. It has first been proposed in a linear model in Horváth et al. (2004) with i.i.d. errors, then in Aue et al. (2006b) with uncorrelated errors. In Schmitz and Steinebach (2010) the errors have other dependency structures like being strongly mixing or autoregressive. The CUSUM has already been considered even in further data structures as in non-linear time series in Ciuperca (2013) and Kirch and Kamgaing (2011) as well as in Kirch and Kamgaing (2015) in the same general setup, as we will do in the next chapters. This means that the existing literature about several models and dependency structures about the sequential CUSUM statistic is included in the setup, and these authors extended it to many new examples.

However, the main disadvantage of the CUSUM is a potentially very long time between the occurrence of the change and its detection by the procedure, the so-called run-length, in particular for late changes, since the CUSUM sums up all new incoming data. Consequently, a lot of data after the change are necessary to override the already summed up observations before the change, until the procedure will stop.

Hence, there arose the idea of statistics that do not use all data after the historical data set. These are the modified MOSUM, Page-CUSUM and the MOSUM statistics which we will use for testing in this work. They only differ in the number of observations that are included in the statistic.

We will develop the asymptotics under the null hypothesis and the alternative hypothesis for these statistics in a very general setup, thereby unifying and extending the existing literature. The modified MOSUM and the Page-CUSUM procedures have only been considered in the linear model in Chen and Tian (2010) and Fremdt (2014), respectively, the MOSUM even only in the location model in Horváth et al. (2012) and Aue et al. (2008). Our results include new examples as diverse as the non-linear, the binary and the Poisson autoregressive model. Only the CUSUM statistic has already been considered in such a general setup in Kirch and Kamgaing (2015) as mentioned above. In a simulation study, we finally compare all types of statistics including the CUSUM statistic in terms of empirical size, power and run length.

In Chapter 2 we will specify the sequential testing in the general setup and state some regularity conditions which will be used in Chapter 3 and 4 to develop the asymptotics under the null and the alternative hypothesis. In Chapter 5 we will apply the procedure to the models already discussed in the existing literature, and we will give some new examples. The final Chapter 6 presents a detailed simulation study that compares the empirical behaviour of the different types of statistics.

2. Sequential Testing Problem based on Estimating Functions

This chapter explains the general monitoring scheme and introduces the types of sequential statistics which we analyse concerning their asymptotic behaviour under the null and alternative hypothesis in the next chapters, as well as in a simulation study in Chapter 6.

Related to this chapter, the well-known CUSUM statistic has been proposed in this general setup in Kirch and Kamgaing (2015). It will also be introduced in this chapter for the sake of a comparison of the statistics in the simulation study.

First of all we require a historical data set or a so-called training period in which the time series has no change in the unknown parameter. This assumption is often called the "noncontamination" assumption. Based on this initial observations, we estimate the unknown parameter θ_0 using an estimating function. Then we start monitoring and check after each new incoming observation using a monitoring function whether the model with the estimated parameter describes the data still well enough. The observations $\boldsymbol{X}_t, t = 1, 2, \ldots$ can be multivariate and dependent. The choice of the null hypothesis H_0 of no change against the alternative hypothesis H_1 of an existing change depends on the type of change for which is searching for, e.g. a mean change or a variance change or a change in the linear regression parameter with random regressors etc.

The unknown parameter θ_0 has to be estimated based on the historical data set such that the estimator fulfills the equation

$$\sum_{t=1}^{m} G(\boldsymbol{X}_t, \widehat{\theta}_m) \stackrel{!}{=} 0, \tag{2.1}$$

where $\boldsymbol{X}_t, t = 1, ..., m$, is the training period and G is the estimating function and has values in $\mathbb{R}^d$.

In the classical non-sequential procedure the length of the data set grows to infinity. But this does not work in the sequential case because we do not know the whole data set for an analysis. Therefore the length m of the historical data set converges to infinity, which means that the estimated parameter $\widehat{\theta}_m$ becomes more and more precise, i.e. $\widehat{\theta}_m \stackrel{P}{\longrightarrow} \theta_0$, where θ_0 is the true parameter. In the correctly specified case, i.e. if the correct model is used, the true parameter is the unique solution of $E(G(\boldsymbol{X}_t, \theta_0)) = 0, \ t = 1, ..., m$. However we do not need the correct model. In the misspecified case θ_0 also fulfills the mentioned equation and can be interpreted as the best approximating parameter.

For example in the mean change model if we use the least-square-estimator, the function G is given by $G(X_t, \theta) = X_t - \theta$. In the linear regression model, if we want to search a change in the unknown parameter θ, the least squares estimator is received by using the function $G(X_t, \theta) = x_t(X_t - \theta^T x_t)$, where x_t are the regressors.

To control the new incoming data we use detectors based on a monitoring function H which also has values in $\mathbb{R}^d$, however not necessarily with the same dimension as the estimating function

G but with dimension $\widetilde{d} \leq d$. The following sums lead to the statistics:

$$\begin{aligned}
\text{CUSUM:} \quad & \boldsymbol{S}_1(m,k) = \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t, \widehat{\theta}_m); \\
\text{modified MOSUM:} \quad & \boldsymbol{S}_2(m,k,h) = \sum_{t=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_t, \widehat{\theta}_m), \ h \in (0,1); \\
\text{MOSUM:} \quad & \boldsymbol{S}_3(m,k,h) = \sum_{t=m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \widehat{\theta}_m), \\
& \text{where } h \text{ is the so-called window size with } 0 \leq h \leq m, h \in \mathbb{N} \\
& \text{and } h = h(m) \overset{m\to\infty}{\longrightarrow} \infty; \\
\text{Page-CUSUM:} \quad & \boldsymbol{S}_4(m,k,i) = \sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t, \widehat{\theta}_m),
\end{aligned}$$

where $\boldsymbol{S}_j, j = 1,2,3,4$, can be multivariate and $\widehat{\theta}_m$ is the estimator satisfying equation (2.1).

By using the monitoring function, we test the null hypothesis of no change against the alternative hypothesis of a change at a time point k^* later than m.

Already discussed in Kirch and Kamgaing (2015), the underlying idea is, that $E(H(X_t, \widehat{\theta}_m)) \approx E(H(\boldsymbol{X}_t, \theta_0)) = 0$ if there is no change in θ_0 and $E(H(\boldsymbol{X}_t, \widehat{\theta}_m)) \approx E(H(\boldsymbol{X}_t, \theta_0)) \neq 0$ for $t > k^*$, where k^* is the change-point. So a small value of $\boldsymbol{S}_j, \ j = 1,2,3,4$ gives evidence to the null hypothesis and a value away from zero to the alternative.

Throughout this chapter $\|\cdot\|_{\boldsymbol{A}}$ denotes the vector norm defined by

$$\|\boldsymbol{Y}\|^2_{\boldsymbol{A}} = \boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y}, \quad \text{for any vector } \boldsymbol{Y} \text{ and for any positive definite matrix } \boldsymbol{A}.$$

Since $\boldsymbol{S}_j, j = 1,2,3,4$, may be multivariate, our monitoring schemes are in quadratic form, which read

$$\begin{aligned}
&\|\boldsymbol{S}_j(m,k)\|^2_{\boldsymbol{A}} && \text{for} \quad j = 1, \\
&\|\boldsymbol{S}_j(m,k,h)\|^2_{\boldsymbol{A}} && \text{for} \quad j = 2,3, \\
\max_{1\leq i\leq k} &\|\boldsymbol{S}_j(m,k,i)\|^2_{\boldsymbol{A}} && \text{for} \quad j = 4.
\end{aligned}$$

The positive definite matrix $\boldsymbol{A}$ can be replaced by a consistent estimator.

Note that the difference between the statistics is the upper and lower bound of the sums. The CUSUM starts directly after the historical data set to add the monitoring function depending on the observations and the estimated parameter $\widehat{\theta}_m$.

The modified MOSUM and Page-CUSUM start later, except for the case that the time point k is close to m or the parameter h of the modified MOSUM is near by zero. The explanation is, the sum $\boldsymbol{S}_2$ includes a bandwidth parameter h which influences the lower bound of the sum and consequently the amount of added observations. If h is near to one, there are only few observations in the sum and otherwise, if h is near to zero, we add almost all up to actually the full set of observations after the training period, but it always depends on k. In the last mentioned case the value of the modified MOSUM would be equal to the one received by the CUSUM. Since the lower bound of the sum depends on k and h is a fixed value, the difference between the lower and upper bound of the sum getting greater with k is growing. So the amount of added observations is increasing.

By usage of the Page-CUSUM the lower bound of the sum is attained by maximizing over all possible points in time from $m+1$ to k. So according to the modified MOSUM, the lower bound of the sum in case of the Page-CUSUM depends on k as well.

The MOSUM uses a monitoring window of fixed length h. Thus the bounds of the sum are independent of k, and if we start monitoring and consequently k is small, the monitoring window extends into the training period.

The null hypothesis of no change will be rejected at the first point in time, denoted by k, for which holds

- for $j = 1$: $\quad w^2(m,k)\|\boldsymbol{S}_1(m,k)\|^2_{\boldsymbol{A}} \geq c$,
- for $j = 2$: $\quad w^2(m,k)\|\boldsymbol{S}_2(m,k,h)\|^2_{\boldsymbol{A}} \geq c$,
- for $j = 3$: $\quad w^2_M(h,k)\|\boldsymbol{S}_3(m,k,h)\|^2_{\boldsymbol{A}} \geq c$,
- for $j = 4$: $\quad w^2(m,k)\max\limits_{1\leq i\leq k}\|\boldsymbol{S}_4(m,k,i)\|^2_{\boldsymbol{A}} \geq c$.

Here, c is the critical value which can be obtained by the limit distribution under the null hypothesis. The functions $w(m,k)$ respectively $w_M(m,h)$ are suitable so-called boundary functions or weight functions. The left-hand sides of the conditions above are the test statistics. Otherwise, if these conditions are not yet fulfilled, we continue monitoring and check the condition after the next new incoming observation again. The procedure stops at the first point in time at which the null hypothesis is rejected. So the stopping time of the sequential procedure is defined as

- for $j = 1$:

$$\tau_m = \begin{cases} \min\{k : 1 \leq k < N(m),\ w^2(m,k)\|\boldsymbol{S}_1(m,k,h)\|^2_{\boldsymbol{A}} \geq c\} \\ \infty, \quad w^2(m,k)\boldsymbol{S}_1(m,k)^T\boldsymbol{A}\boldsymbol{S}_1(m,k) < c \quad \forall\ k \in [1, N(m)) \end{cases},$$

- for $j = 2$:

$$\tau_m = \begin{cases} \min\{k : 1 \leq k < N(m),\ w^2(m,k)\|\boldsymbol{S}_2(m,k,h)\|^2_{\boldsymbol{A}} \geq c\} \\ \infty, \quad w^2(m,k)\|\boldsymbol{S}_2(m,k,h)\|^2_{\boldsymbol{A}} < c \quad \forall\ k \in [1, N(m)) \end{cases},$$

- for $j = 3$:

$$\tau_m = \begin{cases} \min\{k : 1 \leq k < N(m),\ w^2_M(h,k)\|\boldsymbol{S}_3(m,k,h)\|^2_{\boldsymbol{A}} \geq c\} \\ \infty, \quad w^2_M(h,k)\|\boldsymbol{S}_2(m,k,h)\|^2_{\boldsymbol{A}} < c \quad \forall\ k \in [1, N(m)) \end{cases},$$

- for $j = 4$:

$$\tau_m = \begin{cases} \min\{k : 1 \leq k < N(m),\ w^2(m,k)\max\limits_{1\leq i\leq k}\|\boldsymbol{S}_4(m,k,i)\|^2_{\boldsymbol{A}} \geq c\} \\ \infty, \quad w^2(m,k)\max\limits_{1\leq i\leq k}\|\boldsymbol{S}_4(m,k,i)\|^2_{\boldsymbol{A}} < c \quad \forall\ k \in [1, N(m)). \end{cases}$$

Since we distinguish between the open-end and closed-end procedure, the observation horizon is $N(m) = Nm + 1, N > 0$ in the closed-end and $N(m) = \infty$ in the open-end case.

More precisely, the critical value $c = c(\alpha)$ has to be chosen to have

$$\lim_{m\to\infty} P_{H_0}(\tau_m < \infty) = \alpha$$

under the null hypothesis and

$$\lim_{m\to\infty} P_{H_1}(\tau_m < \infty) = 1$$

under the alternative hypothesis. The first equation controls the asymptotic type-I-error and the second one ensures asymptotic power one or the procedure is consistent which means that the change is asymptotically detected with probability one.

3. Asymptotics under the Null Hypothesis

In the previous chapter we introduced an asymptotic level-α-test, so we need critical values such that the probability of a false rejection of the null hypothesis converges to α. Thus the critical values are the appropriate quantiles of the limit distribution depending on the level.

Hence in this chapter we develop the limit distributions under the null hypothesis for the modified MOSUM, the Page-CUSUM and the MOSUM statistic.

The required assumptions are the same as needed for the CUSUM statistic analysed in Kirch and Kamgaing (2015). Depending on the statistic some weaker assumptions are sufficient but to unify them for all statistics under the null hypothesis, we propose the same assumptions as in Kirch and Kamgaing (2015).

First, we need some regularity conditions for the boundary function specified in Assumption 3.1. With Assumption 3.1a) we could to start monitoring after a_m observations, where a_m has to be relatively small compared to the length of the historical data set m. So we will see in this chapter that it is allowed to start monitoring later. This can be useful in some situations, a detailed explanation follows in the simulation study in Chapter 6.

Furthermore, by Assumption 3.1 we control the behaviour of the boundary function at zero and infinity to ensure that the limit distributions are well-defined. For the closed-end procedure it is adequate to control the behaviour at zero as in part a) of this assumption and for the open-end case we additionally require the behaviour at infinity of the boundary function, as satisfied in part b).

Assumption 3.1

a) The weight function has the form

$$w(m,k) = m^{-\frac{1}{2}}\widetilde{w}(m,k)$$

with

$$\widetilde{w}(m,k) = \begin{cases} \rho\left(\frac{k}{m}\right), & k \geq a_m \\ 0, & k < a_m \end{cases}$$

and $\frac{a_m}{m} \to 0$ as $m \to \infty$. In addition, we assume that ρ is continuous and that

$$\lim_{t\to 0} t^{\gamma}\rho(t) < \infty \quad \text{for some} \quad 0 \leq \gamma < \frac{1}{2}.$$

b) For the open-end procedure we additionally assume

$$\lim_{t\to\infty} t\rho(t) < \infty.$$

Assumption 3.2 indicates that the monitoring function H, inserted with the estimated parameter based on the training period, can be replaced asymptotically by the monitoring function, inserted with the true parameter, adjusted by its fluctuation in the historical data set. The matrix $\boldsymbol{B}(\cdot)$ guarantees that the monitoring function H and the term $\boldsymbol{B}(\cdot)G$, including the estimating function of possibly higher dimension than H, have the same dimension.

Assumption 3.2

The following approximation holds under H_0, where $N(m)$ is the observation horizon and can be infinite:

$$\sup_{1\leq k<N(m)} \frac{1}{\sqrt{m}} \min\left(\frac{1}{m^{-\gamma}k^{\gamma}}, \frac{m}{k}\right) \left\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0)\right)\right\| = o_P(1),$$

for some θ_0 and γ as in Assumption 3.1a).

The following assumption includes the required behaviour of the corresponding partial sum process. More precisely, it has to fulfill a multivariate functional central limit theorem as stated in Assumption 3.3a). Since the replacement mentioned above is allowed due to Assumption 3.2, we need the Hájék-Rényi-type inequalities as denoted in b) and c), to control the behaviour of the cumulative sum of the monitoring function, started after the training period, at zero and infinity. Similar to the behaviour of the weight function, we merely need Assumption 3.3b) for the closed-end procedure because we merely have to control the behaviour at zero. If there is a infinite time horizon, the behaviour at infinity is also relevant, thus we additionally need Assumption 3.3c).

Assumption 3.3

a) The partial sum process

$$\left\{\frac{1}{\sqrt{m}}\sum_{t=1}^{\lfloor ms\rfloor}\left(H(\boldsymbol{X}_t,\theta_0), \boldsymbol{B}(\theta_0)G(\boldsymbol{X}_t,\theta_0)\right) : 1\leq s\leq T\right\}$$

fulfills a functional limit theorem for any $T>0$:

$$\left\{\frac{1}{\sqrt{m}}\sum_{t=1}^{\lfloor ms\rfloor}\left(H(\boldsymbol{X}_t,\theta_0), \boldsymbol{B}(\theta_0)G(\boldsymbol{X}_t,\theta_0)\right) : 1\leq s\leq T\right\} \xrightarrow{D} \left\{(\boldsymbol{W}_1(s), \boldsymbol{W}_2(s)) : 1\leq s\leq T\right\},$$

where $(\boldsymbol{W}_1(s), \boldsymbol{W}_2(s))$ is a multivariate Wiener process with covariance matrix $\boldsymbol{\Sigma}$ =

$$\begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{C} \\ \boldsymbol{C}^T & \boldsymbol{\Sigma}_2 \end{pmatrix}.$$

b) The following Hájék-Rényi-type inequality holds for all $0 < \alpha < \frac{1}{2}$:

$$\max_{1 \le k \le m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1) \quad (m \to \infty).$$

c) For the open-end procedure the following Hájék-Rényi-type inequality holds for any sequence $k_m > 0$

$$\max_{k \ge k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1) \quad (m \to \infty).$$

3.1. Null Asymptotics of the Modified MOSUM

We start with the modified MOSUM statistic, introduced in Chen and Tian (2010), to develop the limit distribution under the null hypothesis.

In Proposition 3.1 we show that Assumption 3.2 implies a corresponding statement for the modified MOSUM. Further, by Propositions 3.2 and 3.4, the Hájék-Rényi-type inequalities, as stated in Assumption 3.3b) and c) for the CUSUM, also holds for the modified MOSUM. These propositions are relevant for the development of the limit distribution as an auxiliary tool to show the asymptotic neglibility of the observations before a time point τm with $\tau \to 0$ and after a time point Tm, $T \to \infty$, both converge uniformly in m. The equivalent statements for the limiting processes are given in Propositions 3.3 and 3.5 to ease the proof so that only the limiting process of a finite time horizon is relevant.

By the definition of the weight function in Assumption 3.1 it is allowed to start monitoring after a_m observations with $\frac{a_m}{m} \to 0,\ m \to \infty$ and by Theorem 3.1 the limit distribution under the null hypothesis still remains the same. Heuristically, it is useful to start monitoring after a_m observations in view of the effect that the procedure gives an alarm and stops before the changes occurs. This means that in the beginning of the monitoring time and especially if the value of h is near to 1, the statistic consists of a few observations only. So, depending on certain parameter constellations, the boundary function jumps randomly over the value of the statistic and leads to a false alarm under the null hypothesis. The effect of too early detection will be discussed in detail by the simulation study in Chapter 6

Proposition 3.1

Under the null hypothesis let Assumptions 3.1 and 3.2 hold. Then we obtain:

$$\sup_{1 \le k < N(m)} w(m,k) \left\| \sum_{i=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) \right.$$

$$- \left(\sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k - \lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| = o_P(1).$$

Proof. By Assumption 3.1 and 3.2 we have

$$\begin{aligned}
&\sup_{1 \le k < N(m)} w(m,k) \Bigg\| \sum_{i=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) \\
&\qquad - \left(\sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k - \lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&= \sup_{1 \le k < N(m)} w(m,k) \Bigg\| \left(\sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \sum_{i=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_i, \hat{\theta}_m) \right) \\
&\quad - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) - \left(\sum_{j=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_j, \theta_0) - \frac{\lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \right) \Bigg\| \\
&\le \sup_{1 \le k < N(m)} w(m,k) \Bigg\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&\quad + \sup_{1 \le k < N(m)} w(m,k) \Bigg\| \sum_{i=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_j, \theta_0) - \frac{\lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&\le \sup_{1 \le k < N(m)} \frac{1}{\sqrt{m}} \rho\left(\frac{k}{m}\right) \Bigg\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&\quad + \sup_{1 \le k < N(m)} \frac{1}{\sqrt{m}} \rho\left(\frac{k}{m}\right) \Bigg\| \sum_{i=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_j, \theta_0) - \frac{\lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&= \sup_{1 \le k < N(m)} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k} \right) \max\left(\left(\frac{k}{m}\right)^{\gamma}, \frac{k}{m} \right) \rho\left(\frac{k}{m}\right) \\
&\qquad \Bigg\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&\quad + \sup_{1 \le k < N(m)} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k} \right) \max\left(\left(\frac{k}{m}\right)^{\gamma}, \frac{k}{m} \right) \rho\left(\frac{k}{m}\right) \\
&\qquad \Bigg\| \sum_{i=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_j, \theta_0) - \frac{\lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\|
\end{aligned}$$

$$\leq \sup_{1\leq k<N(m)} \max\left(\left(\frac{k}{m}\right)^{\gamma}, \frac{k}{m}\right)\rho\left(\frac{k}{m}\right)$$
$$\left(\sup_{1\leq k<N(m)} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k}\right)\right.$$
$$\left\|\sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|$$
$$+ \sup_{1\leq k<N(m)} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k}\right)$$
$$\left.\left\|\sum_{i=m+1}^{m+\lfloor kh\rfloor} H(\boldsymbol{X}_i,\hat{\theta}_m) - \left(\sum_{j=m+1}^{m+\lfloor kh\rfloor} H(\boldsymbol{X}_j,\theta_0) - \frac{\lfloor kh\rfloor}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\right)$$

$$\leq \sup_{1\leq k<N(m)} \max\left(\left(\frac{k}{m}\right)^{\gamma}, \frac{k}{m}\right)\rho\left(\frac{k}{m}\right)$$
$$\left(\sup_{1\leq k<N(m)} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k}\right)\right.$$
$$\left\|\sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|$$
$$+ \sup_{1\leq k<N(m)} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k}\right)$$
$$\left.\max_{1\leq i\leq k}\left\|\sum_{i=m+1}^{m+i} H(\boldsymbol{X}_i,\hat{\theta}_m) - \left(\sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j,\theta_0) - \frac{i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\right)$$

$$= \sup_{1\leq k<N(m)} \max\left(\left(\frac{k}{m}\right)^{\gamma}, \frac{k}{m}\right)\rho\left(\frac{k}{m}\right)$$
$$\left(\sup_{1\leq k<N(m)} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k}\right)\right.$$
$$\left\|\sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|$$
$$+ \sup_{1\leq k<N(m)} \max_{0\leq i\leq k} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k}\right)$$
$$\left.\left\|\sum_{i=m+1}^{m+i} H(\boldsymbol{X}_i,\hat{\theta}_m) - \left(\sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j,\theta_0) - \frac{i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\right)$$

$$\leq 2 \sup_{1\leq k<N(m)} \max_{0\leq i\leq k} \frac{1}{\sqrt{m}} \min\left(\left(\frac{k}{m}\right)^{-\gamma}, \frac{m}{k}\right)$$

$$\left\| \sum_{i=m+1}^{m+i} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) - \frac{i}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \right\|$$

$$= 2 \sup_{1 \leq k < N(m)} \max \left(\left(\frac{k}{m} \right)^{\gamma}, \frac{k}{m} \right) \rho \left(\frac{k}{m} \right)$$

$$\sup_{1 \leq k < N(m)} \frac{1}{\sqrt{m}} \min \left(\left(\frac{k}{m} \right)^{-\gamma}, \frac{m}{k} \right)$$

$$\left\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \right\|$$

$$= o_P(1).$$

□

Proposition 3.2

If Assumption 3.3b) holds, then the following Hájék-Rényi-type inequality is valid for all $0 < \alpha < \frac{1}{2}$:

$$\max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1).$$

Proof. From Assumption 3.3b) we conclude

$$\max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\| \leq \max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} \lfloor kh \rfloor^{\alpha}} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$\leq \max_{1 \leq \lfloor kh \rfloor \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} \lfloor kh \rfloor^{\alpha}} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\| = \max_{1 \leq l \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} l^{\alpha}} \left\| \sum_{t=m+1}^{m+l} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$= O_P(1).$$

Furthermore, by Assumption 3.3b) and the following inequality the assertion follows:

$$\max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$\leq \max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| + \max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$= O_P(1)$$

□

Proposition 3.3

If $\{\boldsymbol{W}_1(t) : t \geq 0\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$ then

$$\sup_{0<t\leq\tau} \frac{1}{t^\alpha} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)\| = o_P(1), \quad \tau \to 0,$$

where $0 < \alpha < \frac{1}{2}$.

Proof. Since

$$\frac{\sqrt{2t \log\log \frac{1}{t}}}{t^\alpha} = o(1), \quad t \to 0 \quad \text{and} \qquad \frac{\sqrt{2th \log\log \frac{1}{th}}}{(th)^\alpha} = o(1), \quad t \to 0.$$

The law of iterated logarithm yields

$$\begin{aligned}
&\sup_{0<t\leq\tau} \frac{1}{t^\alpha} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1) - \boldsymbol{W}_1(1+th) + \boldsymbol{W}_1(1)\| \\
&\leq \sup_{0<t\leq\tau} \frac{1}{t^\alpha} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1)\| + \sup_{0<t\leq\tau} \frac{1}{t^\alpha} \|\boldsymbol{W}_1(1+th) - \boldsymbol{W}_1(1)\| \\
&\overset{D}{=} \sup_{0<t\leq\tau} \frac{1}{t^\alpha} \|\boldsymbol{W}_1(t)\| + \sup_{0<t\leq\tau} \frac{1}{t^\alpha} \|\boldsymbol{W}_1(th)\| \\
&\leq \sup_{0<t\leq\tau} \frac{\sqrt{2t \log\log \frac{1}{t}}}{t^\alpha} \frac{\|\boldsymbol{W}_1(t)\|}{\sqrt{2t \log\log \frac{1}{t}}} + \sup_{0<t\leq\tau} \frac{\sqrt{2th \log\log \frac{1}{th}}}{t^\alpha} \frac{\|\boldsymbol{W}_1(th)\|}{\sqrt{2th \log\log \frac{1}{th}}} \\
&\leq \sup_{0<t\leq\tau} \frac{\sqrt{2t \log\log \frac{1}{t}}}{t^\alpha} \sup_{0<t\leq\tau} \frac{\|\boldsymbol{W}_1(t)\|}{\sqrt{2t \log\log \frac{1}{t}}} + \sup_{0<t\leq\tau} \frac{\sqrt{2th \log\log \frac{1}{th}}}{(th)^\alpha} \sup_{0<t\leq\tau} \frac{\|\boldsymbol{W}_1(th)\|}{\sqrt{2th \log\log \frac{1}{th}}} \\
&= o_P(1), \quad \tau \to 0.
\end{aligned}$$

$\square$

Proposition 3.4

Let Assumption 3.3c) hold, then a Hájék-Rényi-type inequality for the open-end procedure is fulfilled for any sequence $k_m > 0$:

$$\max_{k\geq k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1).$$

Proof. By Assumption 3.3c) we obtain

$$\max_{k\geq k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+1}^{m+\lfloor kh\rfloor} H(\boldsymbol{X}_t, \theta_0) \right\| \leq \max_{k\geq k_m} \frac{h\sqrt{k_m}}{\lfloor kh\rfloor} \left\| \sum_{t=m+1}^{m+\lfloor kh\rfloor} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$\le \max_{kh \ge \lfloor k_m h \rfloor} \frac{h\sqrt{k_m}}{\lfloor kh \rfloor} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\| = \max_{kh \ge \lfloor k_m h \rfloor} \sqrt{h \frac{k_m h}{\lfloor k_m h \rfloor}} \frac{\sqrt{\lfloor k_m h \rfloor}}{\lfloor kh \rfloor} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$= \max_{\lfloor kh \rfloor \ge \lfloor k_m h \rfloor} \sqrt{h \frac{k_m h}{\lfloor k_m h \rfloor}} \frac{\sqrt{\lfloor k_m h \rfloor}}{\lfloor kh \rfloor} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$= \max_{\ell \ge \lfloor k_m h \rfloor} \sqrt{h \frac{k_m h}{\lfloor k_m h \rfloor}} \frac{\sqrt{\lfloor k_m h \rfloor}}{\ell} \left\| \sum_{t=m+1}^{m+\ell} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$= O_P(1).$$

Then we conclude

$$\max_{k \ge k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$\le \max_{k \ge k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| + \max_{k \ge k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+1}^{m+\lfloor kh \rfloor} H(\boldsymbol{X}_t, \theta_0) \right\|$$

$$= O_P(1).$$

□

Proposition 3.5

If $\{\boldsymbol{W}_1(t) : t \ge 0\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$ then

$$\max_{t \ge T} \frac{1}{t} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)\| = o_P(1), \quad T \to \infty.$$

Proof. Notice that

$$\max_{t \ge T} \frac{\sqrt{2t \log \log t}}{t} = o(1), \quad T \to \infty$$

and

$$\max_{t \ge T} \frac{\sqrt{2th \log \log (th)}}{t} \le \max_{t \ge T} \frac{\sqrt{2th \log \log th}}{th} = o(1), \quad T \to \infty.$$

By the law of the iterated logarithm we have

$$\max_{t \ge T} \frac{1}{t} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)\|$$

$$= \max_{t \ge T} \frac{1}{t} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1) - \boldsymbol{W}_1(1+th) + \boldsymbol{W}_1(1)\|$$

$$\le \max_{t \ge T} \frac{1}{t} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1)\| + \max_{t \ge T} \frac{1}{t} \|\boldsymbol{W}_1(1+th) - \boldsymbol{W}_1(1)\|$$

$$\overset{D}{=} \max_{t \ge T} \frac{1}{t} \|\boldsymbol{W}_1(t)\| + \max_{t \ge T} \frac{1}{t} \|\boldsymbol{W}_1(th)\|$$

$$\le \max_{t\ge T} \frac{\sqrt{2t\log\log t}}{t} \max_{t\ge T} \frac{\|\boldsymbol{W}_1(t)\|}{\sqrt{2t\log\log t}} + \max_{t\ge T} \frac{\sqrt{2th\log\log th}}{t} \max_{t\ge T} \frac{\|\boldsymbol{W}_1(th)\|}{\sqrt{2th\log\log(th)}}$$
$$= o_P(1), \quad T\to\infty.$$

□

Now we are able to develop the asymptotics for the modified MOSUM under the null hypothesis for the closed-end as well as the open-end procedure.

Theorem 3.1

Let Assumption 3.2 and the null hypothesis hold.

a) Closed-end procedure:
Suppose Assumptions 3.1a) and 3.3a) hold, and that the function ρ figuring in Assumption 3.1a) is bounded. Then for any symmetric positive semi-definite matrix $\boldsymbol{A}$, we have

$$\sup_{1\le k\le Nm} w^2(m,k)\,\|\boldsymbol{S}_2(m,k,h)\|^2_{\boldsymbol{A}} \xrightarrow{D} \sup_{0<t\le N} \rho^2(t)\,\|\boldsymbol{W}_1(t)-\boldsymbol{W}_1(th)-t(1-h)\boldsymbol{W}_2(1)\|^2_{\boldsymbol{A}},$$

where $\{\boldsymbol{W}_1(t): t\ge 0\}$ and $\{\boldsymbol{W}_2(t): t\ge 0\}$ are independent Wiener processes with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as in Assumption 3.3a).

The assertion is also true for a more general weight function as stated in Assumption 3.1a) if additionally Assumption 3.3b) holds.

b) Open-end procedure:
If Assumptions 3.1a)-b) and 3.3a)-c) hold, then

$$\sup_{1\le k<\infty} w^2(m,k)\,\|\boldsymbol{S}_2(m,k,h)\|^2_{\boldsymbol{A}} \xrightarrow{D} \sup_{t>0} \rho^2(t)\,\|\boldsymbol{W}_1(t)-\boldsymbol{W}_1(th)-t(1-h)\boldsymbol{W}_2(1)\|^2_{\boldsymbol{A}},$$

where $\{\boldsymbol{W}_1(t): t\ge 0\}$ and $\{\boldsymbol{W}_2(t): t\ge 0\}$ retain their meanings from a).

The assertions remain true if we replace the matrix $\boldsymbol{A}$ by a consistent estimator.

Proof. Proposition 3.1 gives

$$\sup_{1\le k<N(m)} w^2(m,k)\,\|\boldsymbol{S}_2(m,k,h)\|^2_{\boldsymbol{A}}$$
$$= \sup_{1\le k<N(m)} w^2(m,k) \left\| \sum_{j=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k-\lfloor kh\rfloor}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right\|^2_{\boldsymbol{A}} + o_P(1).$$

a) (i) If ρ is bounded:
By the functional limit theorem in Assumption 3.3a) we have for any $N>0$

$$\sup_{1\le k\le Nm} w^2(m,k) \left\| \sum_{j=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k-\lfloor kh\rfloor}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right\|^2_{\boldsymbol{A}}$$

$$= \sup_{1\le k\le Nm} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \rho^2(t) \Bigg\| \frac{1}{\sqrt{m}} \sum_{j=m+\lfloor\lfloor mt\rfloor h\rfloor+1}^{m+\lfloor mt\rfloor} H(\boldsymbol{X}_j,\theta_0) - \frac{\lfloor mt\rfloor - \lfloor\lfloor mt\rfloor h\rfloor}{m} \frac{1}{\sqrt{m}} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \Bigg\|_{\boldsymbol{A}}^2$$

$$\xrightarrow{D} \sup_{0<t\le N} \rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) - (t-th)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 . \tag{3.1}$$

Furthermore,

$$\sup_{0<t\le N} \rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1) - \boldsymbol{W}_1(1+th) + \boldsymbol{W}_1(1) - (t-th)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2$$

$$\stackrel{D}{=} \sup_{0<t\le N} \rho^2(t) \left\| \widetilde{\boldsymbol{W}}_1(t) - \widetilde{\boldsymbol{W}}_1(th) - (t-th)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 , \tag{3.2}$$

where $\{\widetilde{\boldsymbol{W}}_1(t) = \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1) : t \ge 0\}$ has the covariance matrix $\boldsymbol{\Sigma}_1$ and is independent of $\{\boldsymbol{W}_2(1)\}$. So far we have

$$\sup_{1\le k\le Nm} w^2(m,k) \left\| \boldsymbol{S}_2(m,k,h) \right\|_{\boldsymbol{A}}^2$$

$$= \sup_{1\le k\le Nm} w^2(m,k) \left\| \sum_{j=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k-\lfloor kh\rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 + o_P(1),$$

and

$$\sup_{1\le k\le Nm} w^2(m,k) \left\| \sum_{j=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k-\lfloor kh\rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2$$

$$\xrightarrow{D} \sup_{0<t\le N} \rho^2(t) \left\| \widetilde{\boldsymbol{W}}_1(t) - \widetilde{\boldsymbol{W}}_1(th) - (t-th)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 ,$$

and the assertion follows.

(ii) For a more general weight function ρ:
On the limited time intervall $[\tau m, Nm]$ of the observations the function ρ is also bounded so we can show analogously to (3.1) for any $\tau > 0$

$$\sup_{\tau m\le k\le Nm} w^2(m,k) \left\| \boldsymbol{S}_2(m,k,h) \right\|_{\boldsymbol{A}}^2$$

$$\xrightarrow{D} \sup_{\tau\le t\le N} \rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) - (t-th)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 . \tag{3.3}$$

Now we look at the asymptotical behaviour of the first τm observations.
By Assumption 3.1a) and Proposition 3.2 we obtain for a constant $C > 0$ and $\gamma < \alpha < \frac{1}{2}$

$$\sup_{1\le k<\tau m} w^2(m,k) \left\| \sum_{j=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k-\lfloor kh\rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2$$

$$
\begin{aligned}
&\leq \sup_{1\leq k<\tau m} \frac{1}{m}\widetilde{w}^2(m,k) \left\| \sum_{j=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\qquad + \sup_{1\leq k<\tau m} \frac{(k-\lfloor kh\rfloor)^2}{m^3}\widetilde{w}^2(m,k) \left\| \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\leq \sup_{\frac{1}{m}\leq t<\tau} \frac{1}{m}\rho(t) \left\| \sum_{j=m+\lfloor mth\rfloor+1}^{m+\lfloor mt\rfloor} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\qquad + \sup_{\frac{1}{m}\leq t<\tau} \frac{(t-\lfloor th\rfloor)^2}{m}\rho(t) \left\| \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{B}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\leq \sup_{\frac{1}{m}\leq t<\tau} t^{2\alpha}\rho(t)\frac{1}{mt^{2\alpha}} \left\| \sum_{j=m+\lfloor mth\rfloor+1}^{m+\lfloor mt\rfloor} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\qquad + \sup_{\frac{1}{m}\leq t<\tau} t^{2\alpha}\rho(t)\frac{t^2}{mt^{2\alpha}} \left\| \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\leq \sup_{\frac{1}{m}\leq t<\tau} t^{2\alpha}\rho(t) \left(\sup_{\frac{1}{m}\leq t<\tau} \frac{1}{mt^{2\alpha}} \left\| \sum_{j=m+\lfloor mth\rfloor+1}^{m+\lfloor mt\rfloor} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \right. \\
&\qquad \left. + \sup_{\frac{1}{m}\leq t<\tau} \frac{t^2}{mt^{2\alpha}} \left\| \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \right) \\
&\leq \sup_{0<t<\tau} t^{2\alpha}\rho^2(t) \left(\sup_{\frac{1}{m}\leq t<\tau} \frac{1}{mt^{2\alpha}} \left\| \sum_{j=m+\lfloor mth\rfloor+1}^{m+\lfloor mt\rfloor} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \right. \\
&\qquad \left. +\tau^{2-2\alpha} \left\| \frac{1}{\sqrt{m}}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \right) \\
&= o_P(1), \quad \tau\to 0 \quad \text{uniformly in } m. \qquad (3.4)
\end{aligned}
$$

By Proposition 3.3 and Assumption 3.1a) we obtain an analogous assertion for the limiting Wiener process:

$$
\begin{aligned}
&\sup_{0<t<\tau} \rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) - (t-th)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 \\
&\leq \sup_{0<t<\tau} \rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) \right\|_{\boldsymbol{A}}^2 + \sup_{0<t<\tau} \rho^2(t)(t-th)^2 \left\| \boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 \\
&= \sup_{0<t<\tau} t^{2\alpha}\rho^2(t)\frac{1}{t^{2\alpha}} \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) \right\|_{\boldsymbol{A}}^2 + \sup_{0<t<\tau} t^{2\alpha}\rho^2(t)\frac{(t-th)^2}{t^{2\alpha}} \left\| \boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 \\
&\leq \sup_{0<t<\tau} t^{2\alpha}\rho^2(t) \left(\sup_{0<t<\tau} \frac{1}{t^{2\alpha}} \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) \right\|_{\boldsymbol{A}}^2 + \sup_{0<t<\tau} \frac{(t-th)^2}{t^{2\alpha}} \left\| \boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 \right)
\end{aligned}
$$

$$\le \sup_{0<t<\tau} t^{2\alpha}\rho^2(t) \left(\sup_{0<t<\tau} \frac{1}{t^{2\alpha}} \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) \right\|_{\boldsymbol{A}}^2 + \tau^{2-2\alpha} \left\| \boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 \right)$$
$$= o_P(1), \quad \tau \to 0 \quad \text{uniformly in } m. \tag{3.5}$$

Carefully putting together (3.3), (3.4), (3.5) and (3.2), the assertion follows. The way of putting together the equations will be explained in the proof of part b).

b) We get by Assumptions 3.1b) and Proposition 3.4

$$\sup_{k\ge Tm} w^2(m,k) \left\| \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2$$
$$\le \sup_{k\ge Tm} \left(\frac{k}{m}\right)^2 \rho^2\left(\frac{k}{m}\right) \sup_{k\ge Tm} \frac{1}{m\left(\frac{k}{m}\right)^2} \left\| \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2$$
$$\le \sup_{t\ge T} t^2\rho^2(t) \frac{1}{T} \sup_{t\ge T} \frac{T}{mt^2} \left\| \sum_{j=m+\lfloor mth \rfloor+1}^{m+\lfloor mt \rfloor} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2$$
$$= o_P(1), \quad T \to \infty \quad \text{uniformly in } m. \tag{3.6}$$

An analogous assertion holds for the limiting Wiener process by Assumption 3.1b) and Proposition 3.5, since

$$\sup_{t\ge T} \rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) \right\|_{\boldsymbol{A}}^2$$
$$\le \sup_{t\ge T} t^2\rho^2(t) \sup_{t\ge T} \frac{1}{t^2} \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) \right\|_{\boldsymbol{A}}^2$$
$$= o_P(1), \quad T \to \infty. \tag{3.7}$$

Now we use again Assumptions 3.3a) and 3.1b) and we obtain for $m \to \infty$ and T fixed

$$\sup_{k\ge \tau m} \left\| w(m,\min(k,mT)) \sum_{j=m+\lfloor \min((k,mT)h) \rfloor+1}^{m+\min(k,mT)} H(\boldsymbol{X}_j,\theta_0) - w(m,k)\frac{k-\lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2$$

$$\xrightarrow{D} \sup_{t\ge\tau} \left\| \rho(\min(t,T))(\boldsymbol{W}_1(1+\min(t,T)) - \boldsymbol{W}_1(1+\min(t,T)h)) - t(1-h)\rho(t)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2 . \tag{3.8}$$

Invoking (3.6), the first term in (3.8), reads

$$\sup_{k\ge \tau m} \Bigg\| w(m,\min(k,mT)) \sum_{j=m+\lfloor \min(k,Tm)h \rfloor+1}^{m+\min(k,Tm)} H(\boldsymbol{X}_j,\theta_0) - w(m,k)\frac{k-\lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)$$

$$- \left(w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - w(m,k)\frac{k-\lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \Bigg\|_{\boldsymbol{A}}^2$$

$$
\begin{aligned}
&= \sup_{k \geq \tau m} \left\| w(m, \min(k, mT)) \sum_{j=m+\lfloor \min(k,Tm)h \rfloor+1}^{m+\min(k,Tm)} H(\boldsymbol{X}_j, \theta_0) - w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\leq \sup_{\tau m \leq k < Tm} \left\| w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\quad + \sup_{k \geq Tm} \left\| w(m, mT) \sum_{j=m+\lfloor Tmh \rfloor+1}^{m+Tm} H(\boldsymbol{X}_j, \theta_0) - w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 \\
&= \sup_{k \geq Tm} \left\| w(m, mT) \sum_{j=m+\lfloor mTh \rfloor+1}^{m+Tm} H(\boldsymbol{X}_j, \theta_0) - w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\leq \left\| w(m, mT) \sum_{j=m+\lfloor mTh \rfloor+1}^{m+Tm} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 + \sup_{k \geq Tm} \left\| w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\leq 2 \sup_{k \geq Tm} \left\| w(m,k) \sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 \\
&= o_P(1), \quad T \to \infty \quad \text{uniformly in } m.
\end{aligned}
$$

Analogously we proceed with the limiting process in equation (3.8) and obtain by equation (3.7):

$$
\begin{aligned}
&\sup_{t \geq \tau} \| \rho(\min(t,T))(\boldsymbol{W}_1(1+\min(t,T)) - \boldsymbol{W}_1(1+\min(t,T)h)) - t(1-h)\rho(t)\boldsymbol{W}_2(1) \\
&\quad - (\rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) - t(1-h)\rho(t)\boldsymbol{W}_2(1)) \|_{\boldsymbol{A}}^2 \\
&= \sup_{t \geq \tau} \| \rho(\min(t,T))(\boldsymbol{W}_1(1+\min(t,T)) - \boldsymbol{W}_1(1+\min(t,T)h)) \\
&\quad - \rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) \|_{\boldsymbol{A}}^2 \\
&\leq \sup_{\tau \leq t < T} \| \rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) - \rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) \|_{\boldsymbol{A}}^2 \\
&\quad + \sup_{t \geq T} \| \rho(T)(\boldsymbol{W}_1(1+T) - \boldsymbol{W}_1(1+Th)) - \rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) \|_{\boldsymbol{A}}^2 \\
&= \sup_{t \geq T} \| \rho(T)(\boldsymbol{W}_1(1+T) - \boldsymbol{W}_1(1+Th)) - \rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) \|_{\boldsymbol{A}}^2 \\
&\leq \| \rho(T)(\boldsymbol{W}_1(1+T) - \boldsymbol{W}_1(1+Th)) \|_{\boldsymbol{A}}^2 + \sup_{t \geq T} \| \rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) \|_{\boldsymbol{A}}^2 \\
&\leq 2 \sup_{t \geq T} \| \rho(t)(\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)) \|_{\boldsymbol{A}}^2 \\
&= o_P(1), \quad T \to \infty.
\end{aligned}
$$

We define

$$
\widetilde{X}(m,T,k) := \left\| w(m, \min(k, Tm)) \sum_{j=m+\lfloor \min(k,Tm)h \rfloor+1}^{m+\min(k,Tm)} H(\boldsymbol{X}_j, \theta_0) \right.
$$

$$-w(m,k)\frac{k-\lfloor kh\rfloor}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2,$$

$$X(m,k) := w^2(m,k)\left\|\sum_{j=m+\lfloor kh\rfloor+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)-\frac{k-\lfloor kh\rfloor}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2,$$

$$\widetilde{Y}(t,T) := \|\rho(\min(t,T))(\boldsymbol{W}_1(1+\min(t,T))-\boldsymbol{W}_1(1+\min(t,T)h))-t(1-h)\rho(t)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2$$

and

$$Y(t) := \rho^2(t)\,\|(\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+th))-t(1-h)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2.$$

Our results so far using these definitions are

$$\sup_{k\geq\tau m}\left|\widetilde{X}(m,T,k)-X(m,k)\right| = o_P(1),\quad T\to\infty\quad\text{uniformly in } m, \tag{3.9}$$

and

$$\sup_{t\geq\tau}\left|\widetilde{Y}(t,T)-Y(t)\right| = o_P(1),\quad T\to\infty \tag{3.10}$$

and with (3.8) for $m\to\infty$ and T fixed

$$\sup_{k\geq\tau m}\widetilde{X}(m,T.k)\xrightarrow{D}\sup_{t\geq\tau}\widetilde{Y}(t). \tag{3.11}$$

In the proof of part a)ii) we also have shown

$$\sup_{1\leq k\leq\tau m}X(m,k) = o_P(1)\quad\text{as } \tau\to 0\quad\text{uniformly in } m \tag{3.12}$$

and

$$\sup_{0<t\leq\tau}Y(t) = o_P(1)\quad\text{as } \tau\to 0. \tag{3.13}$$

The following lines show the way how we obtain the assertion by the results so far. They come from Christina Stöhr (University of Magdeburg, Institute of Mathematical Stochastics).

Let $\epsilon>0$ be arbitrary but fixed.

By (3.9) and (3.10) we obtain for fixed $\tau,\delta>0$ that there are integers $T_1=T_1(\epsilon.\delta,\tau)$ and $T_2=T_2(\epsilon.\delta,\tau)$ such that

$$P\left(\sup_{k>\tau m}\left|\widetilde{X}(m,T,k)-X(m,k)\right|>\delta\right)<\epsilon\quad\forall\, T\geq T_1,\ m\in\mathbb{N}, \tag{3.14}$$

$$P\left(\sup_{t>\tau}\left|\widetilde{Y}(t,T)-Y(t)\right|>\delta\right)<\epsilon\quad\forall\quad T\geq T_2,\ m\in\mathbb{N}. \tag{3.15}$$

By (3.11) and the continuity of $\sup_{t>\tau} \widetilde{Y}(t,T)$ for all $T \in \mathbb{N},\ \tau > 0$, we find for a fixed integer T and $\tau \geq 0$ an integer $M_0 = M_0(T, \epsilon, \tau)$ with

$$\sup_{z\in\mathbb{R}} \left| P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \leq z\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z\right)\right| < \epsilon \quad \forall\ m \geq M_0.$$

The inequalities (3.14) and (3.15) hold true for all m if we choose $T \geq \max(T_1, T_2)$. Given the continuity of $\sup_{t>\tau} \widetilde{Y}(t,T)$ for all $T \in \mathbb{N},\ \tau > 0$, there exists a $\delta_0 > 0$ such that

$$\sup_{z\in\mathbb{R}} \left| P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z+\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z-\delta\right)\right| < \epsilon \quad \forall\ \delta \leq \delta_0.$$

For a fixed $\delta \leq \delta_0$ and T as above, there is an integer M_0 such that

$$\begin{aligned}
&\left| P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \leq z+\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z-\delta\right)\right| \\
&\leq \left| P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \leq z+\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z+\delta\right)\right| \\
&\quad + \left| P\left(\sup_{k>\tau m} Y(t,T) \leq z+\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z-\delta\right)\right| \\
&\leq 2\epsilon \quad \forall m\ \geq M_0
\end{aligned}$$

and

$$\begin{aligned}
&\left| P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \leq z-\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z+\delta\right)\right| \\
&\leq \left| P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \leq z-\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z-\delta\right)\right| \\
&+ \left| P\left(\sup_{k>\tau m} \widetilde{Y}(t,T) \leq z+\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \leq z-\delta\right)\right| \\
&\leq 2\epsilon \quad \forall m\ \geq M_0.
\end{aligned}$$

Then we get for each fixed $m \geq M_0$

$$\begin{aligned}
&\left| P\left(\sup_{k>\tau m} X(m,k) \leq z\right) - P\left(\sup_{t>\tau} Y(t) \leq z\right)\right| \qquad (3.16)\\
&= \Bigg| P\left(\sup_{k>\tau m} X(m,k) \leq z, \left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| \leq \delta\right) \\
&\quad + P\left(\sup_{k>\tau m} X(m,k) \leq z, \left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) \\
&\quad - P\left(\sup_{t>\tau} Y(t) \leq z, \left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| \leq \delta\right) \\
&\quad - P\left(\sup_{t>\tau} Y(t) \leq z, \left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right)\Bigg| \\
&= \max\Bigg(P\left(\sup_{k>\tau m} X(m,k) \leq z, \left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| \leq \delta\right) \\
&\quad + P\left(\sup_{k>\tau m} X(m,k) \leq z, \left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right)
\end{aligned}$$

$$
\begin{aligned}
&-P\left(\sup_{t>\tau} Y(t) \le z, \left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| \le \delta\right) \\
&-P\left(\sup_{t>\tau} Y(t) \le z, \left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right), \\
&-P\left(\sup_{k>\tau m} X(m,k) \le z, \left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| \le \delta\right) \\
&-P\left(\sup_{k>\tau m} X(m,k) \le z, \left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) \\
&+P\left(\sup_{t>\tau} Y(t) \le z, \left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| \le \delta\right) \\
&+P\left(\sup_{t>\tau} Y(t) \le z, \left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right)\Bigg) \\
\le{}& \max\Bigg(P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \le z+\delta\right) + P\left(\left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) \\
&-\left(P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \le z-\delta\right) - P\left(\left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right)\right), \\
&-\left(P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \le z-\delta\right) - P\left(\left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right)\right) \\
&+P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \le z+\delta\right) + P\left(\left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right)\Bigg) \\
\le{}& \max\Bigg(P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \le z+\delta\right) - P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \le z-\delta\right) \\
&+P\left(\left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) + P\left(\left|\sup_{t>\tau} \widetilde{Y}(t,T) + \sup_{t>\tau} Y(t)\right| > \delta\right), \\
&-P\left(\sup_{k>\tau m} \widetilde{X}(m,T,k) \le z-\delta\right) + P\left(\sup_{t>\tau} \widetilde{Y}(t,T) \le z+\delta\right) \\
&+P\left(\left|\sup_{k>\tau m} \widetilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) + P\left(\left|\sup_{t>\tau} \widetilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right)\Bigg) \\
\le{}& 4\epsilon.
\end{aligned}
$$

By (3.12) and (3.13) there exist $\tau_1(z,\epsilon), \tau_2(z,\epsilon) \in \mathbb{Q}$ such that

$$P\left(\sup_{1\le k\le \tau m} X(m,k) > z\right) < \epsilon \quad \forall\, \tau \le \tau_1, m \in \mathbb{N},$$

and

$$P\left(\sup_{0<t\le\tau} Y(t) > z\right) < \epsilon \quad \forall\, \tau \le \tau_2, m \in \mathbb{N}.$$

Both inequalities hold for each m and each $\tau \le \min(\tau_1, \tau_2)$. For this τ inequality (3.16) holds for an $M_0 \in \mathbb{N}$. Finally we obtain

$$\left|P\left(\sup_{k\ge 1} X(m,k) \le z\right) - P\left(\sup_{t>0} Y(t) \le z\right)\right|$$

$$= \left| P\left(\max\left(\sup_{1\le k\le \tau m} X(m,k), \sup_{k>\tau m} X(m,k)\right) \le z\right) - P\left(\max\left(\sup_{0<t\le\tau} Y(t), \sup_{t>\tau} Y(t)\right) \le z\right)\right|$$

$$= \left| P\left(\sup_{1\le k\le \tau m} X(m,k) \le z, \sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{0<t\le\tau} Y(t) \le z, \sup_{t>\tau} Y(t) \le z\right)\right|$$

$$= \Bigg| P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{1\le k\le \tau m} X(m,k) > z, \sup_{k>\tau m} X(m,k) \le z\right)$$
$$-P\left(\sup_{t>\tau} Y(t) \le z\right) + P\left(\sup_{0<t\le\tau} Y(t) > z, \sup_{t>\tau} Y(t) \le z\right)\Bigg|$$
$$\le \left| P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{t>\tau} Y(t) \le z\right)\right|$$
$$+ P\left(\sup_{1\le k\le \tau m} X(m,k) > z, \sup_{k>\tau m} X(m,k) \le z\right) + P\left(\sup_{0<t\le\tau} Y(t) > z, \sup_{t>\tau} Y(t) \le z\right)$$
$$\le \left| P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{t>\tau} Y(t) \le z\right)\right|$$
$$+ P\left(\sup_{1\le k\le \tau m} X(m,k) > z\right) + P\left(\sup_{0<t\le\tau} Y(t) > z\right)$$
$$\le 6\epsilon \quad \forall\, m \ge M_0.$$

Consequently

$$\sup_{k\ge 1} w^2(m,k) \left\| \sum_{j=m+\lfloor kh\rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k-\lfloor kh\rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2$$
$$\xrightarrow{D} \sup_{t>0} \rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) - t(1-h)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2$$

and

$$\rho^2(t) \left\| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th) - t(1-h)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2$$
$$\stackrel{D}{=} \rho^2(t) \left\| \widetilde{\boldsymbol{W}}_1(t) - \widetilde{\boldsymbol{W}}_1(th) - t(1-h)\boldsymbol{W}_2(1) \right\|_{\boldsymbol{A}}^2,$$

where $\{\widetilde{\boldsymbol{W}}(\cdot)\}$ is defined as in the proof of part a)(i).

A consistent estimator $\widehat{\boldsymbol{A}}$ of matrix $\boldsymbol{A}$ satisfies

$$\left| \widehat{\boldsymbol{A}} - \boldsymbol{A} \right| = o_P(1).$$

Then it follows

$$\left| \sup_{1\le k\le Nm} w^2(m,k) \left\| S_2(m,k,h) \right\|_{\widehat{\boldsymbol{A}}}^2 - \sup_{1\le k\le Nm} w^2(m,k) \left\| S_2(m,k,h) \right\|_{\boldsymbol{A}}^2 \right|$$
$$= \left| \sup_{1\le k\le Nm} w^2(m,k) \left\| \widehat{\boldsymbol{A}}^{-\frac{1}{2}} S_2(m,k,h) \right\|^2 - \sup_{1\le k\le Nm} w^2(m,k) \left\| \boldsymbol{A}^{-\frac{1}{2}} S_2(m,k,h) \right\|^2 \right|$$

$$
\begin{aligned}
&\leq \sup_{1\leq k\leq Nm} w^2(m,k)\left\|\widehat{\boldsymbol{A}}^{-\frac{1}{2}}S_2(m,k,h)-\boldsymbol{A}^{-\frac{1}{2}}S_2(m,k,h)\right\|^2\\
&= \sup_{1\leq k\leq Nm} w^2(m,k)\left\|\left(\widehat{\boldsymbol{A}}^{-\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}-\mathrm{Id}\right)\boldsymbol{A}^{-\frac{1}{2}}S_2(m,k,h)\right\|^2\\
&\leq \left\|\widehat{\boldsymbol{A}}^{-\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}-\mathrm{Id}\right\|^2 \sup_{1\leq k\leq Nm} w^2(m,k)\left\|\boldsymbol{A}^{-\frac{1}{2}}S_2(m,k,h)\right\|^2\\
&= \left\|\widehat{\boldsymbol{A}}^{-\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}-\mathrm{Id}\right\|^2 \sup_{1\leq k\leq Nm} w^2(m,k)\left\|S_2(m,k,h)\right\|_{\boldsymbol{A}}^2 = o_P(1).
\end{aligned}
$$

□

The choice

$$
\rho\left(\frac{k}{m}\right)=\left(1+\frac{k}{m}\right)^{-1}\left(\frac{k}{m+k}\right)^{-\gamma} \tag{3.17}
$$

of the boundary function is very popular because we can simplify the limit distribution for the open-end procedure as will be shown in the next theorem.

Theorem 3.2

If $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$ is the covariance matrix of the independent Wiener processes $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$, then for any $0\leq\gamma<\frac{1}{2}$

$$
\sup_{t>0}\left\|\frac{(\boldsymbol{W}_1(t)-\boldsymbol{W}_1(th)-t(1-h)\boldsymbol{W}_2(1))}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2 \stackrel{D}{=} \sup_{0<s<1}\left\|\frac{\boldsymbol{W}(s)}{s^\gamma}-(1-(1-h)s)\frac{\boldsymbol{W}\left(\frac{hs}{1-(1-h)s}\right)}{s^\gamma}\right\|_{\boldsymbol{A}}^2,
$$

where $\{\boldsymbol{W}(t):0<t<1\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$.

Proof. By Hušková and Koubková (2005), proof of Theorem 1 and

$$
\frac{1+th}{1+t}=\frac{1+th}{t}\frac{t}{1+t}=\left(\frac{1}{t}+h\right)\frac{t}{1+t}=\frac{1}{1+t}+h\frac{t}{1+t}=1-\frac{t}{1+t}+h\frac{t}{1+t}=1-(1-h)\frac{t}{1+t},
$$

we obtain

$$
\sup_{t>0}\left\|\frac{(\boldsymbol{W}_1(t)-\boldsymbol{W}_1(th)-t(1-h)\boldsymbol{W}_2(1))}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2 = \sup_{t>0}\left\|\frac{\boldsymbol{W}_1(t)-t\boldsymbol{W}_2(1)-(\boldsymbol{W}_1(th)-th\boldsymbol{W}_2(1))}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2
$$

$$
\stackrel{D}{=}\sup_{t>0}\left\|\frac{(1+t)\boldsymbol{W}\left(\frac{t}{1+t}\right)-(1+th)\boldsymbol{W}\left(\frac{th}{1+th}\right)}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2 \stackrel{D}{=} \sup_{0<s<1}\left\|\frac{\boldsymbol{W}(s)}{s^\gamma}-(1-(1-h)s)\frac{\boldsymbol{W}\left(\frac{hs}{1-(1-h)s}\right)}{s^\gamma}\right\|_{\boldsymbol{A}}^2,
$$

where $\{\boldsymbol{W}(t):0<t<1\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$. The last equality in distribution holds since the mapping $[0,\infty)\ni t\mapsto\frac{t}{1+t}\in[0,1)$ is bijective. □

In the situation of Theorem 3.2 and if we additionally choose the inverse of the covariance matrix $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ of the Wiener processes $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$ as matrix $\boldsymbol{A}$, we receive a pivotal limit, in the sense that the limit distribution does not depend on unknown parameters.

Corollary 3.1

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ is the covariance matrix of the independent Wiener processes $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$ and $\boldsymbol{A} = \boldsymbol{\Sigma}_1^{-1}$, then for any $0 \leq \gamma < \frac{1}{2}$

$$\sup_{t>0} \left\| \frac{(\boldsymbol{W}_1(t) - \boldsymbol{W}_1(th) - t(1-h)\boldsymbol{W}_2(1))}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}} \right\|^2_{\boldsymbol{\Sigma_1}^{-1}} \stackrel{D}{=} \sup_{0<s<1} \left\| \frac{\boldsymbol{W}(s)}{s^{\gamma}} - (1-(1-h)s)\frac{\boldsymbol{W}\left(\frac{hs}{1-(1-h)s}\right)}{s^{\gamma}} \right\|^2,$$

where $\{\boldsymbol{W}(t) : 0 < t < 1\}$ is a standard Wiener process and $\| \cdot \|$ is the l_2- norm.

3.2. Null Asymptotics of the Page-CUSUM

The Page-CUSUM statistic is traced back to an idea of Page (1954), who gives its name.

The way to develop the asymptotic distribution under the null hypothesis of the Page-CUSUM is analogous to the one of the modified MOSUM.

First we give an analogon to Assumption 3.2 for the Page-CUSUM, and in Proposition 3.7 we prove a Hájék-Rényi-type inequality for the Page-CUSUM following from the one in Assumption 3.3b). Propositions 3.8 and 3.9 guarantee that the limiting process is well-defined. We do not give a corresponding proposition to Proposition 3.3 for the Page-CUSUM because such a proposition follows immediately by the assumptions, so we include it directly in the proof of the main Theorem 3.3 of this section.

As well as by using the modified MOSUM statistic, it is possible to cut the first a_m observations with $\frac{a_m}{m} \to 0,\ m \to \infty$ and still get the same asymptotics under the null hypothesis, by using the Page-CUSUM statistic.

However to wait for a_m observations is not necessarily essential. The reason is the Page-CUSUM maximizes over all possible time points till k, and chooses a very early time point for the lower bound of the sum of the statistic if k is small. Hence false alarms in the beginning of the monitoring time will occur less often. This effect is illustrated in Figures 6.3a)-b) and will be explained in detail in the simulation study.

Proposition 3.6

Under the null hypothesis let Assumptions 3.1 and 3.2 hold. Then

$$\sup_{1\leq k<N(m)} w(m,k) \left| \max_{0\leq i\leq k} \left\| \sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t, \hat{\theta}_m) \right\| - \max_{0\leq i\leq k} \left\| \sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) - \frac{k-i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right\| \right| = o_P(1).$$

Proof. By Assumptions 3.1 and 3.2 we have

$$
\begin{aligned}
&\sup_{1\le k<N(m)} w(m,k)\left|\max_{0\le i\le k}\left\|\sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)\right\|\right.\\
&\qquad\qquad\left.-\max_{0\le i\le k}\left\|\sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t,\theta_0)-\frac{k-i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right\|\right|\\
&=\sup_{1\le k<N(m)} w(m,k)\left|\max_{0\le i\le k}\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)-\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\hat{\theta}_m)\right\|\right.\\
&\qquad\qquad-\max_{0\le i\le k}\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\theta_0)-\frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right.\\
&\qquad\qquad\qquad\left.\left.-\left(\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\theta_0)-\frac{i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\right|\\
&\le\sup_{1\le k<N(m)} w(m,k)\max_{0\le i\le k}\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)-\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\hat{\theta}_m)\right.\\
&\quad\left.-\left(\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\theta_0)-\frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)-\left(\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\theta_0)-\frac{i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right)\right\|\\
&\le\sup_{1\le k<N(m)} w(m,k)\max_{0\le i\le k}\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)-\left(\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\theta_0)-\frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\\
&\quad+\sup_{1\le k<N(m)} w(m,k)\max_{0\le i\le k}\left\|\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\hat{\theta}_m)-\left(\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\theta_0)-\frac{i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\\
&\le\sup_{1\le k<N(m)}\max\left(\left(\frac{k}{m}\right)^{\gamma},\frac{k}{m}\right)\rho\left(\frac{k}{m}\right)\\
&\quad\left(\sup_{1\le k<N(m)}\frac{1}{\sqrt{m}}\min\left(\left(\frac{k}{m}\right)^{-\gamma},\frac{m}{k}\right)\right.\\
&\qquad\qquad\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)-\left(\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\theta_0)-\frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\\
&\quad+\sup_{1\le k<N(m)}\frac{1}{\sqrt{m}}\min\left(\left(\frac{k}{m}\right)^{-\gamma},\frac{m}{k}\right)
\end{aligned}
$$

$$\max_{1\le i\le k}\left\|\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\hat{\theta}_m)-\left(\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\theta_0)-\frac{i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\Bigg)$$

$$= \sup_{1\le k<N(m)} \max\left(\left(\frac{k}{m}\right)^{\gamma},\frac{k}{m}\right)\rho\left(\frac{k}{m}\right)$$
$$\Bigg(\sup_{1\le k<N(m)}\frac{1}{\sqrt{m}}\min\left(\left(\frac{k}{m}\right)^{-\gamma},\frac{m}{k}\right)$$
$$\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)-\left(\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\theta_0)-\frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|$$

$$+\sup_{1\le k<N(m)}\max_{1\le i\le k}\frac{1}{\sqrt{m}}\min\left(\left(\frac{k}{m}\right)^{-\gamma},\frac{m}{k}\right)$$
$$\left\|\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\hat{\theta}_m)-\left(\sum_{t=m+1}^{m+i} H(\boldsymbol{X}_t,\theta_0)-\frac{i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|\Bigg)$$

$$= 2\sup_{1\le k<N(m)} \max\left(\left(\frac{k}{m}\right)^{\gamma},\frac{k}{m}\right)\rho\left(\frac{k}{m}\right)$$
$$\sup_{1\le k<N(m)}\frac{1}{\sqrt{m}}\min\left(\left(\frac{k}{m}\right)^{-\gamma},\frac{m}{k}\right)$$
$$\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)-\left(\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\theta_0)-\frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|$$

$$= o_P(1).$$

□

Proposition 3.7

If Assumptions 3.3b) and c) hold, we have the Hájék-Rényi-type inequality

$$\sup_{k\ge Tm}\frac{\sqrt{Tm}}{k}\max_{1\le i\le k}\left\|\sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j,\theta_0)\right\| = O_P(1).$$

Proof. With Assumptions 3.3b) and c) we conclude

$$\sup_{k\ge Tm}\frac{\sqrt{Tm}}{k}\max_{1\le i\le k}\left\|\sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j,\theta_0)\right\|$$

$$
\begin{aligned}
&\leq \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \left(\max_{1 \leq i \leq m} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| + \max_{m < i < Tm} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \right. \\
&\qquad \left. + \max_{k \geq i \geq Tm} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \right) \\
&\leq \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \max_{1 \leq i \leq m} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| + \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \max_{m < i < Tm} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\quad + \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \max_{k \geq i \geq Tm} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\leq \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \max_{1 \leq i \leq m} m^{\frac{1}{2}-\alpha} i^{\alpha} \frac{1}{m^{\frac{1}{2}-\alpha} i^{\alpha}} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\quad + \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \max_{1 \leq i \leq Tm} (Tm)^{\frac{1}{2}-\alpha} i^{\alpha} \frac{1}{(Tm)^{\frac{1}{2}-\alpha} i^{\alpha}} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\quad + \max_{i \geq Tm} \frac{\sqrt{Tm}}{i} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\leq \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \max_{1 \leq i \leq m} m^{\frac{1}{2}-\alpha} i^{\alpha} \max_{1 \leq l \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} l^{\alpha}} \left\| \sum_{j=m+1}^{m+l} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\quad + \sup_{k \geq Tm} \frac{\sqrt{Tm}}{k} \max_{1 \leq i \leq Tm} (Tm)^{\frac{1}{2}-\alpha} i^{\alpha} \max_{1 \leq l \leq Tm} \frac{1}{(Tm)^{\frac{1}{2}-\alpha} l^{\alpha}} \left\| \sum_{j=m+1}^{m+l} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\quad + \max_{i \geq Tm} \frac{\sqrt{Tm}}{i} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| \\
&\leq \frac{\sqrt{Tm}}{Tm} m^{\frac{1}{2}-\alpha} m^{\alpha} O_P(1) + \frac{\sqrt{Tm}}{Tm} (Tm)^{\frac{1}{2}-\alpha} (Tm)^{\alpha} O_P(1) + O_P(1) \\
&= \frac{1}{\sqrt{T}} O_P(1) + O_P(1) + O_P(1) = O_P(1).
\end{aligned}
$$

□

Proposition 3.8

For a Wiener process $\{\boldsymbol{W}_1(\cdot)\}$ with covariance matrix $\boldsymbol{\Sigma}_1$ and $0 < \alpha < \frac{1}{2}$, we have

$$
\sup_{0 < t \leq \tau} \frac{1}{t^{\alpha}} \max_{0 \leq s \leq t} \| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+s) \| = o_P(1), \quad \tau \to 0.
$$

Proof. The law of iterated logarithm (confer Theorem 1.3.1 in Csörgő and Révész (1981)) gives

$$\begin{aligned}
&\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\| \\
&=\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1)-\boldsymbol{W}_1(1+s)+\boldsymbol{W}_1(1)\| \\
&\le\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1)\|+\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+s)-\boldsymbol{W}_1(1)\| \\
&\stackrel{D}{=}\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\|\boldsymbol{W}_1(t)\|+\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\max_{0\le s\le t}\|\boldsymbol{W}_1(s)\| \\
&\le\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\|\boldsymbol{W}_1(t)\|+\sup_{0<t\le\tau}\max_{0\le s\le t}\frac{1}{s^{\alpha}}\|\boldsymbol{W}_1(s)\| \\
&=2\sup_{0<t\le\tau}\frac{1}{t^{\alpha}}\|\boldsymbol{W}_1(t)\| \\
&=2\sup_{0<t\le\tau}\frac{\sqrt{2t\log\log\frac{1}{t}}}{t^{\alpha}}\sup_{0<t\le\tau}\frac{\|\boldsymbol{W}_1(t)\|}{\sqrt{2t\log\log\frac{1}{t}}} \\
&=o_P(1),\quad \tau\to 0,
\end{aligned}$$

where $0<\alpha<\frac{1}{2}$. □

Proposition 3.9

For a Wiener process $\{\boldsymbol{W}_1(\cdot)\}$ with covariance matrix $\boldsymbol{\Sigma}_1$, we have

$$\sup_{t\ge T}\frac{1}{t}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\|=o_P(1),\quad T\to\infty.$$

Proof. Applying the law of iterated logarithm (confer Theorem 1.3.1 in Csörgő and Révész (1981)), we obtain

$$\sup_{t\ge T}\frac{1}{t}\|\boldsymbol{W}_1(t)\|=o_P(1),\quad T\to\infty.$$

Invoking Theorem 1.3.1* in Csörgő and Révész (1981), we have

$$\sup_{t\ge T}\frac{1}{t}\max_{0\le s\le t}\|\boldsymbol{W}_1(s)\|\le\sup_{t\ge T}\frac{\sqrt{2t\log\log t}}{t}\sup_{t\ge T}\max_{0\le s\le t}\frac{\|\boldsymbol{W}_1(s)\|}{\sqrt{2t\log\log t}}=o_P(1),\quad T\to\infty.$$

Analogously to the proof of Proposition 3.5 in case of the modified MOSUM statistic we obtain

$$\begin{aligned}
&\sup_{t\ge T}\frac{1}{t}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\| \\
&=\sup_{t\ge T}\frac{1}{t}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1)-\boldsymbol{W}_1(1+s)+\boldsymbol{W}_1(1)\| \\
&\stackrel{D}{=}\sup_{t\ge T}\frac{1}{t}\max_{0\le s\le t}\|\boldsymbol{W}_1(t)-\boldsymbol{W}_1(s)\|
\end{aligned}$$

$$\leq \sup_{t \geq T} \frac{1}{t} \|\boldsymbol{W}_1(t)\| + \sup_{t \geq T} \frac{1}{t} \max_{0 \leq s \leq t} \|\boldsymbol{W}_1(s)\|$$
$$= o_P(1), \quad T \to \infty.$$

□

Theorem 3.3

Let Assumption 3.2 and the null hypothesis hold.

a) Closed-end procedure:
If Assumptions 3.1a) and 3.3a) hold and the function ρ figuring in Assumption 3.1a) is bounded, then for any symmetric positive semi-definite matrix $\boldsymbol{A}$, we get

$$\sup_{1 \leq k \leq Nm} w^2(m,k) \max_{0 \leq i \leq k} \|\boldsymbol{S}_4(m,k)\|^2_{\boldsymbol{A}}$$
$$\xrightarrow{D} \sup_{0 < t \leq N} \rho^2(t) \max_{0 \leq s \leq t} \|\boldsymbol{W}_1(t) - \boldsymbol{W}_1(s) - (t-s)\boldsymbol{W}_2(1)\|^2_{\boldsymbol{A}},$$

where $\{\boldsymbol{W}_1(t) : t \geq 0\}$ and $\{\boldsymbol{W}_2(t) : t \geq 0\}$ are independent Wiener processes with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as in Assumption 3.3a).

The assertion is also true for a more general weight function as denoted in Assumption 3.1a) if additionally Assumption 3.3b) holds.

b) Open-end procedure:
If Assumptions 3.1a)-b) and 3.3a)-c) hold, then

$$\sup_{1 \leq k < \infty} w^2(m,k) \max_{0 \leq i \leq k} \|\boldsymbol{S}_4(m,k)\|^2_{\boldsymbol{A}}$$
$$\xrightarrow{D} \sup_{t > 0} \rho^2(t) \max_{0 \leq s \leq t} \|\boldsymbol{W}_1(t) - \boldsymbol{W}_1(s) - (t-s)\boldsymbol{W}_2(1)\|^2_{\boldsymbol{A}},$$

where $\{\boldsymbol{W}_1(t) : t \geq 0\}$ and $\{\boldsymbol{W}_2(t) : t \geq 0\}$ are given in a).

The assertions still holds true if we replace the matrix $\boldsymbol{A}$ by a consistent estimator.

Proof. Proposition 3.6 yields

$$\sup_{1 \leq k < N(m)} w^2(m,k) \max_{0 \leq i \leq k} \|\boldsymbol{S}_4(m,k)\|^2_{\boldsymbol{A}}$$
$$= \sup_{1 \leq k < N(m)} w^2(m,k) \max_{0 \leq i \leq k} \left\| \sum_{j=m+i+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k-i}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right\|^2_{\boldsymbol{A}} + o_P(1).$$

a) i) If ρ is bounded:
By the functional limit theorem in Assumption 3.3a), we have for any $N > 0$

$$\sup_{1 \leq k \leq Nm} w^2(m,k) \max_{0 \leq i \leq k} \left\| \sum_{j=m+i+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k-i}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right\|^2_{\boldsymbol{A}}$$

$$
\begin{aligned}
&= \sup_{\frac{1}{m}\le t\le N} \rho^2(t) \max_{0\le s\le t} \Bigg\| \frac{1}{\sqrt{m}} \sum_{j=m+\lfloor ms\rfloor+1}^{m+\lfloor mt\rfloor} H(\boldsymbol{X}_j,\theta_0) \\
&\qquad\qquad - \frac{\lfloor mt\rfloor - \lfloor ms\rfloor}{m} \frac{1}{\sqrt{m}} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \Bigg\|_{\boldsymbol{A}}^2 \\
&\xrightarrow{D} \sup_{0<t\le N} \rho^2(t) \max_{0\le s\le t} \| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+s) - (t-s)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2 \qquad (3.18)
\end{aligned}
$$

and furthermore

$$
\begin{aligned}
&\sup_{0<t\le N} \rho^2(t) \max_{0\le s\le t} \| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+s) - (t-s)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2 \\
&= \sup_{0<t\le N} \rho^2(t) \max_{0\le s\le t} \| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1) - \boldsymbol{W}_1(1+s) + \boldsymbol{W}_1(1) - (t-s)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2 \\
&\stackrel{D}{=} \sup_{0<t\le N} \rho^2(t) \max_{0\le s\le t} \left\| \widetilde{\boldsymbol{W}}_1(t) - \widetilde{\boldsymbol{W}}_1(s) - (t-s)\boldsymbol{W}_2(1)\right\|_{\boldsymbol{A}}^2 ,
\end{aligned}
$$

where $\{\widetilde{\boldsymbol{W}}_1 : t \ge 0\}$ and $\{\boldsymbol{W}_2 : t \ge 0\}$ are independent and have the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. The assertion follows.

ii) For a more general weight function ρ:
On the limited time interval $[\tau m, Nm]$ the function ρ is also bounded so we can show analogously to (3.18) for any $\tau > 0$

$$
\begin{aligned}
&\sup_{\tau m\le k\le Nm} w^2(m,k)\, \|\boldsymbol{S}_4(m,k)\|_{\boldsymbol{A}}^2 \\
&\xrightarrow{D} \sup_{\tau\le t\le N} \rho^2(t)\, \| \boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+s) - (t-s)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2 . \qquad (3.19)
\end{aligned}
$$

Now we look at the asymptotical behaviour of the first τm observations.
By Assumptions 3.1a) and 3.3b) we obtain

$$
\begin{aligned}
&\sup_{1\le k<\tau m} w^2(m,k) \max_{0\le i\le k} \left\| \sum_{j=m+i+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k-i}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\le \sup_{1\le k<\tau m} \widetilde{w}^2(m,k) \max_{0\le i\le k} \frac{1}{m} \left\| \sum_{j=m+i+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\qquad + \sup_{1\le k<\tau m} \widetilde{w}^2(m,k) \max_{0\le i\le k} \frac{(k-i)^2}{m^2} \left\| \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \\
&\le \sup_{1\le k<\tau m} \left(\frac{k}{m}\right)^{2\gamma} \rho^2\left(\frac{k}{m}\right) \left(\sup_{1\le k<\tau m} \frac{1}{\left(\frac{k}{m}\right)^{2\gamma}} \max_{0\le i\le k} \frac{1}{m} \left\| \sum_{j=m+i+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \right. \\
&\qquad \left. + \left\| \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right\|_{\boldsymbol{A}}^2 \sup_{1\le k<\tau m} \frac{1}{\left(\frac{k}{m}\right)^{2\gamma}} \max_{0\le i\le k} \frac{(k-i)^2}{m^3} \right)
\end{aligned}
$$

$$
\begin{aligned}
&\le \sup_{1\le k<\tau m}\left(\frac{k}{m}\right)^{2\gamma}\rho^2\left(\frac{k}{m}\right)\left(\sup_{1\le k<\tau m}\frac{1}{m\left(\frac{k}{m}\right)^{2\gamma}}\left\|\sum_{j=m+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\right.\\
&\quad+\sup_{1\le k<\tau m}\frac{1}{m\left(\frac{k}{m}\right)^{2\gamma}}\max_{0\le i\le k}\left\|\sum_{j=m+1}^{m+i}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\\
&\quad\left.+\left\|\frac{1}{\sqrt{m}}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\sup_{1\le k<\tau m}\left(\frac{k}{m}\right)^{2-2\gamma}\right)\\
&\le \sup_{1\le k<\tau m}\left(\frac{k}{m}\right)^{2\gamma}\rho^2\left(\frac{k}{m}\right)\left(\sup_{1\le k<\tau m}\frac{1}{m\left(\frac{k}{m}\right)^{2\gamma}}\left\|\sum_{j=m+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\right.\\
&\quad+\sup_{1\le k<\tau m}\max_{0\le i\le k}\frac{1}{m\left(\frac{i}{m}\right)^{2\gamma}}\left\|\sum_{j=m+1}^{m+i}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\\
&\quad\left.+\left\|\frac{1}{\sqrt{m}}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\sup_{1\le k<\tau m}\left(\frac{k}{m}\right)^{2-2\gamma}\right)\\
&\le \sup_{1\le k<\tau m}\left(\frac{k}{m}\right)^{2\gamma}\rho^2\left(\frac{k}{m}\right)\left(2\sup_{1\le k<\tau m}\frac{1}{m\left(\frac{k}{m}\right)^{2\gamma}}\left\|\sum_{j=m+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\right.\\
&\quad\left.+\left\|\frac{1}{\sqrt{m}}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\sup_{1\le k<\tau m}\left(\frac{k}{m}\right)^{2-2\gamma}\right)\\
&\le \sup_{\frac{1}{m}\le t<\tau}t^{2\gamma}\rho^2(t)\left(O_P(1)+\tau^{2-2\gamma}\left\|\frac{1}{\sqrt{m}}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\right)\\
&= o_P(1),\quad \tau\to 0 \text{ uniformly in } m.
\end{aligned}
\tag{3.20}
$$

For the limiting Wiener process we also obtain that the first τm observations are asymptotically negligible. Therefore we need Assumption 3.1a) and Proposition 3.8 to conclude

$$
\begin{aligned}
&\sup_{0<t<\tau}\rho^2(t)\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)-(t-s)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2\\
&=\sup_{0<t<\tau}t^{2\gamma}\rho^2(t)\frac{1}{t^{2\gamma}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)-(t-s)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2\\
&\le \sup_{0<t<\tau}t^{2\gamma}\rho^2(t)\left(\sup_{0<t<\tau}\frac{1}{t^{2\gamma}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\|_{\boldsymbol{A}}^2\right.\\
&\qquad\qquad\left.+\sup_{0<t<\tau}\frac{1}{t^{2\gamma}}\max_{0\le s\le t}(t-s)^2\|\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2\right)\\
&\le \sup_{0<t<\tau}t^{2\gamma}\rho^2(t)\left(\sup_{0<t<\tau}\frac{1}{t^{2\gamma}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\|_{\boldsymbol{A}}^2+\sup_{0<t<\tau}t^{2-2\gamma}\|\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2\right)
\end{aligned}
$$

$$\le \sup_{0<t<\tau} t^{2\gamma}\rho^2(t)\left(\sup_{0<t<\tau}\frac{1}{t^{2\gamma}}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\|_{\boldsymbol{A}}^2+\tau^{2-2\gamma}\|\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2\right)$$
$$= o_P(1), \quad \tau\to 0. \tag{3.21}$$

With (3.19), (3.20) und (3.21) we get the assertion in the same way as in the proof of Theorem 3.1 in the section on the modified MOSUM.

b) By Assumptions 3.1b), 3.3c) and Proposition 3.7, it follows that

$$\begin{aligned}
&\sup_{k\ge Tm} w^2(m,k)\max_{0\le i\le k}\left\|\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\\
&\le \sup_{k\ge Tm}\left(\frac{k}{m}\right)^2\rho^2\left(\frac{k}{m}\right)\sup_{k\ge Tm}\frac{1}{m\left(\frac{k}{m}\right)^2}\max_{0\le i\le k}\left\|\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\\
&= \sup_{k\ge Tm}\left(\frac{k}{m}\right)^2\rho^2\left(\frac{k}{m}\right)\sup_{k\ge Tm}\frac{m}{k^2}\max_{0\le i\le k}\left\|\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\\
&= \sup_{k\ge Tm}\left(\frac{k}{m}\right)^2\rho^2\left(\frac{k}{m}\right)\frac{1}{T}\left(\sup_{k\ge Tm}\frac{Tm}{k^2}\left\|\sum_{j=m+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\right.\\
&\qquad\left.+\sup_{k\ge Tm}\frac{Tm}{k^2}\max_{1\le i\le k}\left\|\sum_{j=m+1}^{m+i}H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\right)\\
&= O(1)o(1)(O_P(1)+O_P(1))\\
&= o_P(1), \quad T\to\infty \quad \text{uniformly in } m.
\end{aligned}\tag{3.22}$$

Analogously, for the limiting process Proposition 3.9 yields

$$\begin{aligned}
&\sup_{t\ge T}\rho^2(t)\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\|_{\boldsymbol{A}}^2\\
&\le \sup_{t\ge T}t^2\rho^2(t)\sup_{t\ge T}\frac{1}{t^2}\max_{0\le s\le t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\|_{\boldsymbol{A}}^2\\
&= o_P(1) \quad T\to\infty.
\end{aligned}\tag{3.23}$$

Now we want to combine the previous results to complete the proof. But first we need the following two statements:

$$\begin{aligned}
&\sup_{1\le k\le N(m)} w(m,k)^2\max_{0\le i\le k}\left\|\sum_{t=m+i+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)-\frac{k-i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2\\
&= \sup_{1\le k\le N(m)}\max_{0\le i\le k}\left\|w(m,k)\sum_{t=m+i+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)-w(m,k)\frac{k-i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2
\end{aligned}$$

and

$$\sup_{k\ge\tau m}\max_{0\le i\le k}\left\|w(m,\min(k,mT))\sum_{j=m+\min(i,mT)+1}^{m+\min(k,mT)}H(\boldsymbol{X}_j,\theta_0)-w(m,k)\frac{k-i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2$$

$$\xrightarrow{D} \sup_{t\geq\tau}\max_{0\leq s\leq t}\|\rho(\min(t,T))(\boldsymbol{W}_1(1+\min(t,T))-\boldsymbol{W}_1(1+\min(s,T)))-(t-s)\rho(t)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2$$

Then we conclude with (3.22)

$$\sup_{k\geq\tau m}\max_{0\leq i\leq k}\Bigg\|w(m,\min(k,mT))\sum_{j=m+\min(i,mT)+1}^{m+\min(k,mT)}H(\boldsymbol{X}_j,\theta_0)-w(m,k)\frac{k-i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)$$

$$-\left(w(m,k)\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)-w(m,k)\frac{k-i}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right)\Bigg\|_{\boldsymbol{A}}^2$$

$$=\sup_{k\geq\tau m}\max_{0\leq i\leq k}\Bigg\|w(m,\min(k,mT))\sum_{j=m+\min(i,mT)+1}^{m+\min(k,mT)}H(\boldsymbol{X}_j,\theta_0)-w(m,k)\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2$$

$$\leq\sup_{\tau m\leq k<Tm}\max_{0\leq i\leq k}\Bigg\|w(m,\min(k,mT))\sum_{j=m+\min(i,mT)+1}^{m+\min(k,mT)}H(\boldsymbol{X}_j,\theta_0)-w(m,k)\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2$$

$$+\sup_{k\geq Tm}\max_{0\leq i\leq k}\Bigg\|w(m,\min(k,mT))\sum_{j=m+\min(i,mT)+1}^{m+\min(k,mT)}H(\boldsymbol{X}_j,\theta_0)-w(m,k)\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2$$

$$=\sup_{k\geq Tm}\max_{0\leq i\leq k}\Bigg\|w(m,mT)\sum_{j=m+\min(i,mT)+1}^{m+mT}H(\boldsymbol{X}_j,\theta_0)-w(m,k)\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2$$

$$\leq\sup_{k\geq Tm}\max_{0\leq i\leq k}\Bigg\|w(m,mT)\sum_{j=m+\min(i,mT)+1}^{m+mT}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2+\sup_{k\geq Tm}\max_{0\leq i\leq k}\Bigg\|w(m,k)\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2$$

$$\leq w^2(m,mT)\max_{0\leq i\leq mT}\Bigg\|\sum_{j=m+i+1}^{m+mT}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2+\sup_{k\geq Tm}w^2(m,k)\max_{0\leq i\leq k}\Bigg\|\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2$$

$$\leq 2\sup_{k\geq Tm}w^2(m,k)\max_{0\leq i\leq k}\Bigg\|\sum_{j=m+i+1}^{m+k}H(\boldsymbol{X}_j,\theta_0)\Bigg\|_{\boldsymbol{A}}^2$$

$$=o_P(1),\quad T\to\infty\quad\text{uniformly in } m \text{ by(3.22).}$$

In an analogous way, by using (3.23), we obtain for the limiting Wiener process

$$\begin{aligned}&\sup_{t\geq T}\max_{0\leq s\leq t}\|\rho(\min(t,T)(\boldsymbol{W}_1(1+\min(t,T))-\boldsymbol{W}_1(1+\min(s,T)))-(t-s)\rho(t)\boldsymbol{W}_2(1)\\&\quad-\rho(t)(\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s))-(t-s)\rho(t)\boldsymbol{W}_2(1)\|_{\boldsymbol{A}}^2\\&=o_P(1),\quad T\to\infty.\end{aligned}$$

Finally we combine the approximations in the same way as in part b) of the proof of Theorem 3.1 to get the assertion.

□

As we see in the next Theorem the boundary function stated in (3.17) is also useful to simplify the limit process of the Page-CUSUM statistic. Fremdt (2014) also used this weight function to simplify the limit process.

Theorem 3.4

If $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$ is the covariance matrix of independent Wiener processes $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$, then for $0\leq\gamma<\frac{1}{2}$

$$\sup_{t>0}\max_{0\leq s\leq t}\left\|\frac{\boldsymbol{W}_1(t)-\boldsymbol{W}_1(s)-(t-s)\boldsymbol{W}_2(1)}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2\overset{D}{=}\sup_{0<t<1}\max_{0\leq s\leq t}\frac{1}{t^{2\gamma}}\left\|\boldsymbol{W}(t)-\frac{t-1}{s-1}\boldsymbol{W}(s)\right\|_{\boldsymbol{A}}^2,$$

where $\{\boldsymbol{W}(\cdot)\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$.

Proof. Since the independent Wiener processes $\boldsymbol{W}_1(\cdot)$ and $\boldsymbol{W}_2(\cdot)$ have covariance matrix $\boldsymbol{\Sigma}_1$ and

$$\frac{1+s}{1+t}=\frac{1+s}{s}\frac{s}{t}\frac{t}{1+t}=\frac{1+s}{s}\frac{\frac{t+1}{t}-1}{\frac{s+1}{s}-1}\frac{t}{1+t}=\frac{1-\frac{t}{1+t}}{1-\frac{s}{1+s}}=\frac{\frac{t}{1+t}-1}{\frac{s}{1+s}-1},$$

we have

$$\begin{aligned}&\sup_{t>0}\max_{0\leq s\leq t}\left\|\frac{\boldsymbol{W}_1(t)-\boldsymbol{W}_1(s)-(t-s)\boldsymbol{W}_2(1)}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2=\sup_{t>0}\max_{0\leq s\leq t}\left\|\frac{\boldsymbol{W}_1(t)-t\boldsymbol{W}_2(1)-(\boldsymbol{W}_1(s)-s\boldsymbol{W}_2(1))}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2\\&\overset{D}{=}\sup_{t>0}\max_{0\leq s\leq t}\left\|\frac{(1+t)\boldsymbol{W}\left(\frac{t}{1+t}\right)-(1+s)\boldsymbol{W}\left(\frac{s}{1+s}\right)}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2=\sup_{t>0}\max_{0\leq s\leq t}\left\|\frac{\boldsymbol{W}\left(\frac{t}{1+t}\right)}{\left(\frac{t}{1+t}\right)^{\gamma}}-\frac{(1+s)}{(1+t)}\frac{\boldsymbol{W}\left(\frac{s}{1+s}\right)}{\left(\frac{t}{1+t}\right)^{\gamma}}\right\|_{\boldsymbol{A}}^2\\&\overset{D}{=}\sup_{0<t<1}\max_{0\leq s\leq t}\frac{1}{t^{2\gamma}}\left\|\boldsymbol{W}(t)-\frac{t-1}{s-1}\boldsymbol{W}(s)\right\|_{\boldsymbol{A}}^2,\end{aligned}$$

where $\{\boldsymbol{W}(\cdot)\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$. And the last equality in distribution holds since the mapping $[0,\infty)\ni t\mapsto\frac{t}{1+t}\in[0,1)$ is bijective and increasing. □

Like with the modified MOSUM, a further simplification arises by using the inverse of the covariance matrix $\boldsymbol{\Sigma}_1$ of a Wiener process $\{\boldsymbol{W}_1(\cdot)\}$ as matrix $\boldsymbol{A}$, still under the condition that the Wiener processes $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$ have the same covariance matrix. The further simplification results again in a pivotal limit, i.e. there are no unknowns in the limit behaviour.

Corollary 3.2

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ is the covariance matrix of $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$ and $\boldsymbol{A} = {\boldsymbol{\Sigma}_1}^{-1}$, then, for $0 \leq \gamma < \frac{1}{2}$, we have

$$\sup_{t>0} \max_{0\leq s\leq t} \left\| \frac{\boldsymbol{W}_1(t) - \boldsymbol{W}_1(s) - (t-s)\boldsymbol{W}_2(1)}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}} \right\|^2_{\boldsymbol{\Sigma}_1} \overset{D}{=} \sup_{0<t<1} \max_{0\leq s\leq t} \frac{1}{t^{2\gamma}} \left\| \boldsymbol{W}(t) - \frac{t-1}{s-1}\boldsymbol{W}(s) \right\|^2,$$

where $\{\boldsymbol{W}(\cdot)\}$ is a standard Wiener process and $\|\cdot\|$ is the l_2-norm.

3.3. Null Asymptotics of the MOSUM

The MOSUM uses a constant bandwidth h to monitor the new incoming data. More precisely, the statistic always sums up a fixed number of observations independent of the actual time point k, in contrast to the modified MOSUM and Page-CUSUM, where the number of added observations depends on k. Thus the boundary function jumping randomly over the value of the statistic mentioned in the context of using the modified MOSUM statistic caused by a few data points included in the statistic. However, if we use the MOSUM statistic, in the beginning of the monitoring time where k is small, the window width extends into the training period. So the effect of too early detection still only exists in less extent.

We distinguish between two cases. In the first case, the bandwidth has the same order as the length of the historical data set, so $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta, \beta \in (0, 1]$, where h depends on m and converges to infinity while m is growing to infinity. The second case is that $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$, i.e. the bandwidth h is very small relatively to the length m of the training period.

Since the MOSUM uses a fixed window width and accordingly it is not necessary to start monitoring later than $m + 1$, we need a bounded weight function ρ as denoted in Assumption 3.4a). The behaviour of $\rho(t)$ if t goes to infinity is the same as in case of the modified MOSUM and Page-CUSUM statistic, but only if $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta$. If $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$ we need a somewhat weaker condition as stated in Assumption 3.4c).

We will see in Assumption 3.4a) that we can replace m to h in comparison to the weight function of the previous statistics, so $w_M(h, k) = \frac{1}{\sqrt{h}}\rho_M\left(\frac{k}{h}\right)$. If we suppose a similar boundary function as in case of the modified MOSUM and Page-CUSUM with $w_M(m, k) = \frac{1}{\sqrt{m}}\rho_M\left(\frac{k}{m}\right)$, we would obtain a reasonable limiting process in case of $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta$ specified in Theorem 2.1 in Horváth et al. (2012). However, if $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$, the limit distribution would be identically to zero as shown for the mean change model in Theorem 2.2 in Horváth et al. (2012), which is useless to calculate critical values. Hence we assume the weight function as in Assumption 3.4a).

In this section, the case $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta$ is studied first, followed by the case $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$.

If $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta$, Proposition 3.10 is an analogon to Propositions 3.1 and 3.6 for the previous statistics. Since we propose a bounded weight function, we do not need a Hájék-Rényi-type inequality controlling the behaviour at zero of the MOSUM, but we have to control its behaviour at infinity by a Hájék-Rényi- type inequality as stated in Proposition 3.11 as well as the behaviour at infinity of its limiting process, as we will show in Proposition 3.12.

Assumption 3.4

Using the MOSUM statistic to test for a structural change in a time series,

a) the weight function has the form

$$w_M(h,k) = h^{-\frac{1}{2}} \rho_M\left(\frac{k}{h}\right),$$

where ρ_M is bounded and continuous and $h = h(m) \to \infty$, as $m \to \infty$.

b) In the open-end procedure if $\frac{h}{m} \to \beta$, as $m \to \infty$, for some $\beta \in (0,1]$, suppose that

$$\lim_{t\to\infty} t\rho_M(t) < \infty.$$

c) If $\frac{h}{m} \to 0$, as $m \to \infty$, we need the following weaker condition:

$$\limsup_{t\to\infty} t^{\frac{1}{\nu}} \rho_M(t) < \infty, \quad \text{for some} \quad \nu > 2.$$

Assumptions 3.4b) and c) are equivalent to Assumption 3.1b) on the boundary function of the modified MOSUM and the Page-CUSUM statistics.

Proposition 3.10

Let the null hypothesis hold as well as $h(m) \overset{m\to\infty}{\longrightarrow} \infty$ and $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta,\ \beta \in (0,1]$. The boundary function ρ satisfies Assumption 3.2 with $\gamma = 0$ and Assumption 3.4a)-b). Then we have

$$\begin{aligned}\sup_{1\le k<N(m)} w_M(h,k) \Bigg\| &\sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \hat{\theta}_m) \\ &- \left(\sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| = o_P(1).\end{aligned}$$

Proof. Analogously to the proof of Proposition 3.1 as well as Proposition 3.6, by Assumption 3.1 with $\gamma = 0$ and Assumption 3.4a)-b) it follows that

$$\begin{aligned}\sup_{1\le k<N(m)} w_M(h,k) \Bigg\| &\sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) \\ &- \left(\sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\|\end{aligned}$$

$$
\begin{aligned}
&= \sup_{1\le k<N(m)} w_M(h,k) \Bigg\| \left(\sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat\theta_m) - \sum_{i=m+1}^{m+k-h} H(\boldsymbol{X}_i,\hat\theta_m) \right) \\
&\quad - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) - \left(\sum_{j=m+1}^{m+k-h} H(\boldsymbol{X}_j,\theta_0) - \frac{k-h}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \right) \Bigg\| \\
&\le \sup_{1\le k<N(m)} \frac{1}{\sqrt{h}} \rho_M\left(\frac{k}{h}\right) \left\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat\theta_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \right\| \\
&\quad + \sup_{1\le k<N(m)} \frac{1}{\sqrt{h}} \rho_M\left(\frac{k}{h}\right) \left\| \sum_{i=m+1}^{m+k-h} H(\boldsymbol{X}_i,\hat\theta_m) - \left(\sum_{j=m+1}^{m+k-h} H(\boldsymbol{X}_j,\theta_0) - \frac{k-h}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \right\| \\
&\le \sup_{1\le k<N(m)} \max\left(\sqrt{m}, \frac{k}{\sqrt{m}}\right) \frac{1}{\sqrt{h}} \rho_M\left(\frac{k}{h}\right) \\
&\quad \Bigg(\sup_{1\le k<N(m)} \min\left(\frac{1}{\sqrt{m}}, \frac{\sqrt{m}}{k}\right) \\
&\qquad \left\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat\theta_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \right\| \\
&\quad + \sup_{1\le k<N(m)} \max_{0\le i\le k} \min\left(\frac{1}{\sqrt{m}}, \frac{\sqrt{m}}{k}\right) \\
&\qquad \left\| \sum_{i=m+1}^{m+i} H(\boldsymbol{X}_i,\hat\theta_m) - \left(\sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j,\theta_0) - \frac{i}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \right\| \Bigg) \\
&= 2 \sup_{1\le k<N(m)} \max\left(\sqrt{m}, \frac{k}{\sqrt{m}}\right) \frac{1}{\sqrt{h}} \rho_M\left(\frac{k}{h}\right) \\
&\quad \sup_{1\le k<N(m)} \min\left(\frac{1}{\sqrt{m}}, \frac{\sqrt{m}}{k}\right) \\
&\qquad \left\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat\theta_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \right\| \\
&= 2 \sup_{1\le k<N(m)} \max\left(\sqrt{\frac{m}{h}}, \frac{k}{h}\sqrt{\frac{h}{m}}\right) \rho_M\left(\frac{k}{h}\right) \\
&\quad \sup_{1\le k<N(m)} \min\left(\frac{1}{\sqrt{m}}, \frac{\sqrt{m}}{k}\right) \\
&\qquad \left\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i,\hat\theta_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0) \right) \right\| \\
&= o_P(1).
\end{aligned}
$$

□

Proposition 3.11

Under Assumption 3.3c) hold, we have a Hájék-Rényi-type inequality in the open-end procedure, namely

$$\max_{k\geq k_m+h}\frac{\sqrt{k_m}}{k}\left\|\sum_{t=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)\right\|=O_P(1)\quad\text{for any sequence } k_m\geq 0.$$

Proof. By Assumption 3.3c),

$$\begin{aligned}
&\max_{k\geq k_m+h}\frac{\sqrt{k_m}}{k}\left\|\sum_{t=m+1}^{m+k-h}H(\boldsymbol{X}_t,\theta_0)\right\|\leq\max_{k\geq k_m+h}\frac{\sqrt{k_m}}{k-h}\left\|\sum_{t=m+1}^{m+k-h}H(\boldsymbol{X}_t,\theta_0)\right\|\\
&=\max_{k-h\geq k_m}\frac{\sqrt{k_m}}{k-h}\left\|\sum_{t=m+1}^{m+k-h}H(\boldsymbol{X}_t,\theta_0)\right\|=\max_{l\geq k_m}\frac{\sqrt{k_m}}{l}\left\|\sum_{t=m+1}^{m+l}H(\boldsymbol{X}_t,\theta_0)\right\|\\
&=O_P(1),
\end{aligned}$$

and the assertion follows from

$$\begin{aligned}
&\max_{k\geq k_m+h}\frac{\sqrt{k_m}}{k}\left\|\sum_{t=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)\right\|\\
&\leq\max_{k\geq k_m+h}\frac{\sqrt{k_m}}{k}\left\|\sum_{t=m+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)\right\|+\max_{k\geq k_m+h}\frac{\sqrt{k_m}}{k}\left\|\sum_{t=m+1}^{m+k-h}H(\boldsymbol{X}_t,\theta_0)\right\|\\
&=O_P(1).
\end{aligned}$$

□

Proposition 3.12

Let $\{\boldsymbol{W}(\cdot)\}$ be a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$. Then

$$\sup_{t\geq T}\frac{1}{t}\left\|\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right)-\boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right\|=o_P(1),\quad T\to\infty.$$

Proof. Analogously to the proof of Proposition 3.5 and 3.9 we get

$$\begin{aligned}
&\sup_{t\geq T}\frac{1}{t}\left\|\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right)-\boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right\|\\
&=\sup_{t\geq T}\frac{1}{t}\left\|\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right)-\boldsymbol{W}_1\left(\frac{1}{\beta}\right)-\boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)+\boldsymbol{W}_1\left(\frac{1}{\beta}\right)\right\|\\
&\leq\sup_{t\geq T}\frac{1}{t}\left\|\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right)-\boldsymbol{W}_1\left(\frac{1}{\beta}\right)\right\|+\sup_{t\geq T}\frac{1}{t}\left\|\boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)+\boldsymbol{W}_1\left(\frac{1}{\beta}\right)\right\|
\end{aligned}$$

$$
\begin{aligned}
&\stackrel{D}{=} \sup_{t\geq T} \frac{1}{t} \left\| \boldsymbol{W}_1(t) \right\| + \sup_{t\geq T} \frac{1}{t} \left\| \boldsymbol{W}_1(t-1) \right\| \\
&= o_P(1), \quad T \to \infty.
\end{aligned}
$$

To obtain the last equality we use the law of iterated logarithm analogously to Proposition 3.5. □

Theorem 3.5

Assume $h \to \infty$ as $m \to \infty$ and $\frac{h}{m} \to \beta$ for some $\beta \in (0, 1]$.
Furthermore, let Assumption 3.2 with $\gamma = 0$, Assumptions 3.3a) and 3.4a) hold under the null hypothesis. Then, for any symmetric positive semi-definite matrix $\boldsymbol{A}$, we have:

a) in the closed-end procedure

$$
\begin{aligned}
&\sup_{1\leq k\leq Nm} w_M^2(h,k) \left\| \boldsymbol{S}_3(m,k,h) \right\|_{\boldsymbol{A}}^2 \\
&\stackrel{D}{\longrightarrow} \sup_{0<t\leq \frac{N}{\beta}} \rho_M^2(t) \left\| \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right) \right\|_{\boldsymbol{A}}^2,
\end{aligned}
$$

where $\{\boldsymbol{W}_1(t) : t \geq 0\}$ and $\{\boldsymbol{W}_2(t) : t \geq 0\}$ are Wiener processes defined in Assumption 3.3a),

b) in the open-end procedure, if additionally Assumption 3.3c) holds,

$$
\begin{aligned}
&\sup_{1\leq k<\infty} w_M^2(h,k) \left\| \boldsymbol{S}_3(m,k,h) \right\|_{\boldsymbol{A}}^2 \\
&\stackrel{D}{\longrightarrow} \sup_{0<t<\infty} \rho_M^2(t) \left\| \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right) \right\|_{\boldsymbol{A}}^2
\end{aligned}
$$

where $\{\boldsymbol{W}_1(t) : t \geq 0\}$ and $\{\boldsymbol{W}_2(t) : t \geq 0\}$ are given in a).

The assertions remain true if we replace the matrix $\boldsymbol{A}$ by a consistent estimator.

Proof. First note with Proposition 3.10

$$
\begin{aligned}
&\sup_{1\leq k<N(m)} w_M^2(h,k) \left\| \boldsymbol{S}_3(m,k,h) \right\|_{\boldsymbol{A}}^2 \\
&= \sup_{1\leq k<N(m)} w_M^2(h,k) \left\| \sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 + o_P(1).
\end{aligned}
$$

a) If $\frac{h}{m} \stackrel{m\to\infty}{\longrightarrow} \beta, \quad \beta \in (0, 1]$ and Assumption 3.3a) holds, then

$$
\sup_{1\leq k\leq Nm} w_M^2(h,k) \left\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2
$$

$$= \sup_{\frac{1}{h}\leq t\leq \frac{Nm}{h}} \rho_M^2(t) \left\| \frac{1}{\sqrt{h}} \sum_{i=\lfloor h\left(\frac{m}{h}+t-1\right)\rfloor+1}^{\lfloor h\left(\frac{m}{h}+t\right)\rfloor} H(\boldsymbol{X}_i,\theta_0) - \frac{h}{m}\frac{1}{\sqrt{h}}\boldsymbol{B}(\theta_0)\sum_{j=1}^{h\frac{m}{h}} G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2$$

$$\xrightarrow{D} \sup_{0<t\leq \frac{N}{\beta}} \rho_M^2(t) \left\| \boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right)\right\|_{\boldsymbol{A}}^2$$

where $\{\boldsymbol{W}_1(t) : t \geq 0\}$ and $\{\boldsymbol{W}_2(t) : t \geq 0\}$ are dependent Wiener processes with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as in Assumption 3.3a).

b) First we get by Assumption 3.3a)

$$\sup_{k\geq 1} \left\| w_M(h,\min(k,mT+h)) \sum_{j=m+\lfloor \min(k,mT+h)\rfloor-h+1}^{m+\min(k,mT+h)} H(\boldsymbol{X}_j,\theta_0) - w_M(h,k)\frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2$$

$$\xrightarrow{D} \sup_{t>0} \left\| \rho_M\left(\min\left(t,\frac{T}{\beta}+1\right)\right)\left(\boldsymbol{W}_1\left(\frac{1}{\beta}+\min\left(t,\frac{T}{\beta}+1\right)\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+\min\left(t,\frac{T}{\beta}+1\right)-1\right)\right)\right.$$

$$\left. -\beta\rho_M(t)\boldsymbol{W}_2(1)\right\|_{\boldsymbol{A}}^2 .$$

Then we can bound the first term of the previous statement according to

$$\sup_{k\geq 1} \left\| w_M(h,\min(k,mT+h)) \sum_{j=m+\lfloor \min(k,mT+h)\rfloor-h+1}^{m+\min(k,mT+h)} H(\boldsymbol{X}_j,\theta_0) - w_M(h,k)\frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right.$$

$$\left. - \left(w_M(h,k) \sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - w_M(h,k)\frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j,\theta_0)\right)\right\|_{\boldsymbol{A}}^2$$

$$\leq \sup_{1\leq k<Tm+h} \left\| w_M(h,k) \sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j,\theta_0) - w_M(h,k) \sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2$$

$$+ \sup_{k\geq Tm+h} \left\| w_M(h,mT+h) \sum_{j=m+mT+1}^{m+mT+h} H(\boldsymbol{X}_j,\theta_0) - w_M(h,k) \sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2$$

$$\leq \left\| w_M(h,mT+h) \sum_{j=m+mT+1}^{m+mT+h} H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2 + \sup_{k\geq Tm+h} \left\| w_M(h,k) \sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2$$

$$\leq 2 \sup_{k\geq Tm+h} \left\| w_M(h,k) \sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j,\theta_0)\right\|_{\boldsymbol{A}}^2$$

$$
\begin{aligned}
&= 2 \sup_{t \geq \frac{Tm}{h}+1} t^2 \rho^2(t) \frac{h}{Tm} \sup_{t \geq \frac{Tm}{h}+1} \frac{\frac{Tm}{h}}{ht^2} \left\| \sum_{j=m+\lfloor ht \rfloor - h+1}^{m+\lfloor ht \rfloor} H(\boldsymbol{X}_j, \theta_0) \right\|_{\boldsymbol{A}}^2 \\
&= o_P(1) \quad \text{as } T \to \infty, \text{ uniformly in } m \text{ by Assumption 3.4b) and Proposition 3.11.}
\end{aligned}
$$

And we obtain the same result for the limiting Wiener process with Proposition 3.12:

$$
\begin{aligned}
&\sup_{t>0} \left\| \rho_M \left(\min\left(t, \frac{T}{\beta}+1\right)\right) \left(\boldsymbol{W}_1 \left(\frac{1}{\beta} + \min\left(t, \frac{T}{\beta}+1\right)\right) - \boldsymbol{W}_1 \left(\frac{1}{\beta} + \min\left(t, \frac{T}{\beta}+1\right) - 1\right)\right) \right. \\
&\quad \left. -\beta \rho_M(t) \boldsymbol{W}_2(1) - \left(\rho_M(t) \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right) - \beta \rho_M(t) \boldsymbol{W}_2(1)\right) \right\|_{\boldsymbol{A}}^2 \\
&= \sup_{t>0} \left\| \rho_M \left(\min\left(t, \frac{T}{\beta}+1\right)\right) \left(\boldsymbol{W}_1 \left(\frac{1}{\beta} + \min\left(t, \frac{T}{\beta}+1\right)\right) \right.\right. \\
&\quad \left.\left. - \boldsymbol{W}_1 \left(\frac{1}{\beta} + \min\left(t, \frac{T}{\beta}+1\right) - 1\right)\right) - \rho_M(t) \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right) \right\|_{\boldsymbol{A}}^2 \\
&\leq \sup_{t \geq \frac{T}{\beta}+1} \left\| \rho_M \left(\frac{T}{\beta}+1\right) \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+\frac{T}{\beta}+1\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+\frac{T}{\beta}\right)\right) \right. \\
&\quad \left. - \rho_M(t) \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right) \right\|_{\boldsymbol{A}}^2 \\
&\leq \left\| \rho_M \left(\frac{T}{\beta}+1\right) \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+\frac{T}{\beta}+1\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+\frac{T}{\beta}\right)\right) \right\|_{\boldsymbol{A}}^2 \\
&\quad + \sup_{t \geq \frac{T}{\beta}+1} \left\| \rho_M(t) \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right) \right\|_{\boldsymbol{A}}^2 \\
&\leq 2 \sup_{t \geq \frac{T}{\beta}+1} \left\| \rho_M(t) \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right) \right\|_{\boldsymbol{A}}^2 \\
&= 2 \sup_{t \geq \frac{T}{\beta}+1} t^2 \rho_M^2(t) \sup_{t \geq \frac{T}{\beta}+1} \frac{1}{t^2} \left\| \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right) \right\|_{\boldsymbol{A}}^2 \\
&\leq 2 \sup_{t \geq \frac{T}{\beta}+1} t^2 \rho_M^2(t) \sup_{t \geq T} \frac{1}{t^2} \left\| \left(\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right) \right\|_{\boldsymbol{A}}^2 \\
&= o_P(1), \quad T \to \infty.
\end{aligned}
$$

Consequently the assertion follows analogously to the proof of Theorem 3.1.

□

Corollary 3.3

Assume $h \to \infty$ as $m \to \infty$ and $\frac{h}{m} \to \beta$ for some $\beta \in (0, 1]$.
If the supremum of the statistic is only taken over $k \geq h$ (which is equivalent to $\rho(t) = 0$ for

$t < 1$), then the limit distribution for a possibly infinite time horizon simplifies to

$$\sup_{h \le k < N(m)} w_M^2(h,k) \left\| \boldsymbol{S}_3(m,k,h) \right\|_{\boldsymbol{A}}^2$$
$$\xrightarrow{D} \sup_{1 \le t < \frac{N(m)}{h}} \rho_M^2(t) \left\| \boldsymbol{W}_1(t) - \boldsymbol{W}_1(t-1) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right\|_{\boldsymbol{A}}^2,$$

where $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$ are independent Wiener processes with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively.

Proof. Let $\{\boldsymbol{W}_1(\cdot)\}$ and $\{\boldsymbol{W}_1(\cdot)\}$ be Wiener processes as defined in Assumption 3.3a). Then we obtain

$$\sup_{1 \le t < N(m)} \rho_M^2(t) \left\| \boldsymbol{W}_1\left(\frac{1}{\beta} + t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta} + t - 1\right) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right\|_{\boldsymbol{A}}^2$$
$$= \sup_{1 \le t < N(m)} \rho_M^2(t) \left\| \boldsymbol{W}_1\left(\frac{1}{\beta} + t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}\right) - \left(\boldsymbol{W}_1\left(\frac{1}{\beta} + t - 1\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}\right)\right) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right\|_{\boldsymbol{A}}^2$$
$$\stackrel{D}{=} \sup_{1 \le t < N(m)} \rho_M^2(t) \left\| \widetilde{\boldsymbol{W}}_1(t) - \widetilde{\boldsymbol{W}}_1(t-1) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right\|_{\boldsymbol{A}}^2,$$

where $\{\widetilde{\boldsymbol{W}}_1(\cdot)\}$ and $\{\boldsymbol{W}_2(\cdot)\}$ are independent Wiener processes with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively. □

If the block length h is much smaller than the length of the training period, i.e. if $\frac{h}{m} \xrightarrow{m \to \infty} 0$, we need some further assumptions. First the rate of convergence of the weight function specified in condition 3.4b) can be weaken to a lower rate of convergence for the boundary function as in Assumption 3.4c). So instead of Assumption 3.2 with $\gamma = 0$, we require Assumption 3.5 below.

As we see in Theorem 3.6, which states the limit distribution of the MOSUM for the case $\frac{h}{m} \to 0$, the limiting process is the same for the closed-end and the open-end procedure. More precisely, the supremum in the limit distribution is taken over all positive integers even in the closed-end method. Thus the statistic maximizing over the time points after Nm, is negligible if the time horizon grows to infinity. Therefore, we need the strong invariance principle as stated in Assumption 3.6. The details are shown in the proof of Theorem 3.6.

Assumption 3.5

The following approximation holds under H_0, where the observation horizon $N(m)$ can be infinite:

$$\sup_{1 \le k < N(m)} \frac{1}{\sqrt{h}} \min\left(1, \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right) \left\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \hat{\theta}_m) \right.$$

$$- \left(\sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| = o_P(1),$$

for some θ_0 and ν as in Assumption 3.4c).

Assumption 3.6

There is a Wiener process $\{\boldsymbol{W}(t), 0 \leq t < \infty\}$ with covariance matrix $\boldsymbol{\Sigma}_1$ such that, as $k \to \infty$

$$\sum_{t=1}^{k} H(\boldsymbol{X}_t, \theta_0) - \boldsymbol{W}(k) = O\left(k^{\frac{1}{\nu}}\right) \quad a.s., \quad \text{with } \nu \text{ as in Assumption 3.4c).}$$

Proposition 3.13

Under the Assumptions 3.4a) and c) and Assumption 3.5, we have

$$\sup_{1 \leq k < N(m)} w_M(h,k) \Bigg\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \hat{\theta}_m) - \left(\sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| = o_P(1).$$

Proof. By the Assumptions 3.4a) and c) as well as 3.5, it follows that

$$\begin{aligned}
&\sup_{1 \leq k < N(m)} w_M(h,k) \Bigg\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&= \sup_{1 \leq k < N(m)} \frac{1}{\sqrt{h}} \rho_M\left(\frac{k}{h}\right) \Bigg\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\| \\
&= \sup_{1 \leq k < N(m)} \frac{1}{\sqrt{h}} \min\left(1, \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right) \max\left(1, \left(\frac{k}{h}\right)^{\frac{1}{\nu}}\right) \rho_M\left(\frac{k}{h}\right) \\
&\qquad \Bigg\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{h}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \Bigg\|
\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{1\leq k<N(m)} \max\left(1, \left(\frac{k}{h}\right)^{\frac{1}{\nu}}\right)\rho_M\left(\frac{k}{h}\right) \\
&\sup_{1\leq k<N(m)} \frac{1}{\sqrt{h}} \min\left(1, \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right)\Bigg\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) \\
&\qquad - \left(\sum_{j=m+k-h+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0)\right)\Bigg\| \\
&= o_P(1).
\end{aligned}
$$

□

Theorem 3.6

Assume $h \to \infty$ as $m \to \infty$ and $\frac{h}{m} \to 0$.
Let Assumptions 3.3a), 3.4a), 3.5, 3.4c) and 3.6 hold. Then, under the null hypothesis and a possibly infinite time horizon $N(m)$, we have

$$
\sup_{1\leq k<N(m)} w_M^2(h,k)\, \|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 \xrightarrow{D} \sup_{0<t<\infty} \rho_M^2(t)\, \|(\boldsymbol{W}(t+1) - \boldsymbol{W}(t))\|_{\boldsymbol{A}}^2\,,
$$

where $\{\boldsymbol{W}(t) : t \geq 0\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$.
If we replace the matrix $\boldsymbol{A}$ by a consistent estimator, the assertion still holds true.

Proof. Proposition 3.13 gives

$$
\begin{aligned}
&\sup_{1\leq k<N(m)} w_M^2(h,k)\, \|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 \\
&= \sup_{1\leq k<N(m)} w_M^2(h,k) \left\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \theta_0) - \frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0)\right\|_{\boldsymbol{A}}^2 + o_P(1).
\end{aligned}
$$

If $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$, ρ is bounded and Assumption 3.3a) holds, then the stationarity of $\boldsymbol{X}_t$ under H_0 gives

$$
\begin{aligned}
&\sup_{1\leq k\leq Th} w_M^2(h,k) \left\| \sum_{i=m+k-h+1}^{m+k} H(\boldsymbol{X}_i, \theta_0) - \frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0)\right\|_{\boldsymbol{A}}^2 \\
&= \sup_{\frac{1}{h}\leq t\leq T} \rho_M^2(t) \left\| \frac{1}{\sqrt{h}} \sum_{i=m+\lfloor ht\rfloor-h+1}^{m+\lfloor ht\rfloor} H(\boldsymbol{X}_i, \theta_0) - \sqrt{\frac{h}{m}}\frac{1}{\sqrt{m}}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0)\right\|_{\boldsymbol{A}}^2 \\
&= \sup_{\frac{1}{h}\leq t\leq T} \rho_M^2(t) \left\| \frac{1}{\sqrt{h}} \sum_{i=m+\lfloor ht\rfloor-h+1}^{m+\lfloor ht\rfloor} H(\boldsymbol{X}_i, \theta_0)\right\|_{\boldsymbol{A}}^2 + o_P(1) \\
&\overset{D}{=} \sup_{\frac{1}{h}\leq t\leq T} \rho_M^2(t) \left\| \frac{1}{\sqrt{h}} \sum_{i=\lfloor ht\rfloor+1}^{\lfloor ht\rfloor+h} H(\boldsymbol{X}_i, \theta_0)\right\|_{\boldsymbol{A}}^2 + o_P(1)
\end{aligned}
$$

$$= \sup_{\frac{1}{h}\le t\le T} \rho_M^2(t) \left\| \frac{1}{\sqrt{h}} \sum_{i=\lfloor ht\rfloor+1}^{\lfloor h(t+1)\rfloor} H(\boldsymbol{X}_i,\theta_0) \right\|_{\boldsymbol{A}}^2 + o_P(1),$$

and

$$\sup_{1\le k\le Th} w_M^2(h,k)\left\|\boldsymbol{S}_3(m,k,h)\right\|_{\boldsymbol{A}}^2 \overset{D}{\longrightarrow} \sup_{0<t\le T} \rho_M^2(t)\left\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\right\|_{\boldsymbol{A}}^2, \tag{3.24}$$

where $\{\boldsymbol{W}(t): t\ge 0\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$.

Furthermore, we obtain that the limit distribution after time point T is negligible if $T\to\infty$, since

$$\begin{aligned}&\sup_{t\ge T}\rho_M^2(t)\left\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\right\|_{\boldsymbol{A}}^2\\ &\le \sup_{t\ge T} t^{\frac{2}{\nu}}\rho_M^2(t) \sup_{t\ge T} t^{-\frac{2}{\nu}}\left\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\right\|_{\boldsymbol{A}}^2 .\end{aligned} \tag{3.25}$$

By Assumption 3.4c) the first term is bounded. And by Theorem 1.2.1 in Csörgő and Révész (1981) the second term is asymptotically negligible, since

$$\begin{aligned}&\sup_{t\ge T}\left\|\boldsymbol{W}_1(t+1)-\boldsymbol{W}_1(t)\right\| t^{-\frac{1}{\nu}} \le \sup_{t\ge T}\sup_{0\le s\le 1}\left\|\boldsymbol{W}_1(t+s)-\boldsymbol{W}_1(t)\right\| t^{-\frac{1}{\nu}}\\ &\le \sup_{t\ge T}\sup_{0\le s\le 1}\frac{\left\|\boldsymbol{W}_1(t+s)-\boldsymbol{W}_1(t)\right\|}{\sqrt{2(\log(t)+\log\log(t))}} \sup_{t\ge T}\frac{\sqrt{2(\log(t)+\log\log(t))}}{t^{\frac{1}{\nu}}}\\ &= o_P(1), \quad T\to\infty,\end{aligned} \tag{3.26}$$

because of

$$\sup_{t\ge T}\sup_{0\le s\le 1}\frac{\left\|\boldsymbol{W}_1(t+s)-\boldsymbol{W}_1(t)\right\|}{\sqrt{2(\log(t)+\log\log(t))}} \overset{T\to\infty}{\longrightarrow} 1 \quad a.s.$$

and

$$\sup_{t\ge T}\frac{\sqrt{2(\log(t)+\log\log(t))}}{t^{\frac{1}{\nu}}} = o(1), \quad T\to\infty.$$

Next we have to prove that the statistic with using the data after Th are asymptotically ignored. Therefore we use Assumption 3.6, to obtain

$$\begin{aligned}&\sup_{k>Th}\rho_M^2\left(\frac{k}{h}\right)\left\|\frac{1}{\sqrt{h}}\sum_{i=k+1}^{k+h} H(X_i,\theta_0)\right\|_{\boldsymbol{A}}^2\\ &= \sup_{k>Th}\rho_M^2\left(\frac{k}{h}\right)\left\|\frac{1}{\sqrt{h}}\left(\sum_{i=1}^{k+h} H(X_i,\theta_0)-\sum_{i=1}^{k} H(X_i,\theta_0)\right)\right\|_{\boldsymbol{A}}^2\\ &= \sup_{k>Th}\rho_M^2\left(\frac{k}{h}\right)\left(\left\|\frac{1}{\sqrt{h}}\left(\boldsymbol{W}_1(k+h)-\boldsymbol{W}_1(k)\right)\right\|_{\boldsymbol{A}}^2 + O_P\left(\frac{(k+h)^{\frac{2}{\nu}}}{h}+\frac{k^{\frac{2}{\nu}}}{h}\right)\right)\\ &\le \sup_{k>Th}\rho_M^2\left(\frac{k}{h}\right)\left\|\boldsymbol{W}_1\left(\frac{k}{h}+1\right)-\boldsymbol{W}_1\left(\frac{k}{h}\right)\right\|_{\boldsymbol{A}}^2\end{aligned}$$

$$+ O_P(1) \sup_{k>Th} \rho_M^2\left(\frac{k}{h}\right)\left(\frac{(k+h)^{\frac{2}{\nu}}}{h}+\frac{k^{\frac{2}{\nu}}}{h}\right). \tag{3.27}$$

The first term is in $o_P(1)$ by (3.25) and (3.26). For the second term, Assumptions 3.4a) and c) yield

$$\begin{aligned}
&O_P(1) \sup_{k>Th}\left(\frac{k}{h}\right)^{\frac{2}{\nu}} \rho_M^2\left(\frac{k}{h}\right) \sup_{k>Tm}\left(\frac{k}{h}\right)^{-\frac{2}{\nu}}\left(\frac{(k+h)^{\frac{2}{\nu}}}{h}+\frac{k^{\frac{2}{\nu}}}{h}\right)\\
&= O_P(1)O(1) \sup_{k>Th} \frac{1}{h^{1-\frac{2}{\nu}}}\left(\left(\frac{k+h}{k}\right)^{\frac{2}{\nu}}+1\right)\\
&= O_P(1)O(1)\frac{1}{h^{1-\frac{2}{\nu}}} \sup_{k>Th}\left(\left(1+\frac{h}{k}\right)^{\frac{2}{\nu}}+1\right)\\
&\leq O_P(1)O(1)\frac{1}{h^{1-\frac{2}{\nu}}}\left(\left(1+\frac{1}{T}\right)^{\frac{2}{\nu}}+1\right)\\
&= O_P(1)O(1)o(1)O(1) = o_P(1) \quad \text{as} \quad T\to\infty \quad \text{uniformly in } m.
\end{aligned} \tag{3.28}$$

Consequently for the open-end procedure the assertion follows in the same way as in the proof of Theorem 3.1. For the closed-end procedure with time horizon Nm we can finish the proof with (3.24)-(3.28) from which we obtain

$$\begin{aligned}
&\left|\sup_{1\leq k\leq Nm} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 - \sup_{0<t<\infty} \rho_M^2(t)\,\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\|_{\boldsymbol{A}}^2\right|\\
&= \Bigg|\sup_{1\leq k\leq Nh} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 + \sup_{Nh<k\leq Nm} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2\\
&\quad - \sup_{0<t\leq N} \rho_M^2(t)\,\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\|_{\boldsymbol{A}}^2 - \sup_{N<t<\infty} \rho_M^2(t)\,\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\|_{\boldsymbol{A}}^2\Bigg|\\
&\leq \left|\sup_{1\leq k\leq Nh} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 - \sup_{0<t\leq N} \rho_M^2(t)\,\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\|_{\boldsymbol{A}}^2\right|\\
&\quad + \left|\sup_{Nh<k\leq Nm} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 - \sup_{N<t<\infty} \rho_M^2(t)\,\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\|_{\boldsymbol{A}}^2\right|\\
&\leq \left|\sup_{1\leq k\leq Nh} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 - \sup_{0<t\leq N} \rho_M^2(t)\,\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\|_{\boldsymbol{A}}^2\right|\\
&\quad + \sup_{Nh<k\leq Nm} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 + \sup_{N<t<\infty} \rho_M^2(t)\,\|\boldsymbol{W}(t+1)-\boldsymbol{W}(t)\|_{\boldsymbol{A}}^2\\
&= o_P(1).
\end{aligned}$$

□

To calculate critical values more precisely we can use another stopping time such that we obtain a limit distribution with a supremum over a finite interval such as in the closed-end procedure.

The new stopping time is

$$\tau_m^{Nh} = \begin{cases} \min\{k : 1\leq k\leq Nh,\ w_M^2(h,k)\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 \geq c\},\\ \infty, \quad w_M^2(h,k)\|\boldsymbol{S}_2(m,k,h)\|_{\boldsymbol{A}}^2 < c \quad \forall\, k\in[1,Nh]. \end{cases}$$

Then we obtain the following limit distributions:

Corollary 3.4

Suppose that Assumptions 3.3a) and 3.4a) hold. Then under the null hypothesis with time horizon Nh and for any symmetric positive semi-definite matrix $\boldsymbol{A}$

a) if $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta, \quad \beta \in (0,1]$
and additionally Assumption 3.2 with $\gamma = 0$ with time horizon Nh is fulfilled, we obtain

$$\sup_{1\le k\le Nh} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2$$
$$\overset{D}{\longrightarrow} \sup_{0<t\le N} \rho_M^2(t) \left\| \boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right\|_{\boldsymbol{A}}^2,$$

where $\{\boldsymbol{W}_1(t) : t \ge 0\}$ and $\{\boldsymbol{W}_2(t) : t \ge 0\}$ are dependent Wiener processes defined as in Assumption 3.3a).

In the situation of starting to search for a change after h observations, so $k \ge h$, the limit distribution simplifies analogous to Corollary 3.3 to

$$\sup_{h\le k\le Nh} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2$$
$$\overset{D}{\longrightarrow} \sup_{1\le t\le N} \rho_M^2(t) \left\| \boldsymbol{W}_1(t) - \boldsymbol{W}_1(t-1) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right\|_{\boldsymbol{A}}^2. \tag{3.29}$$

b) If $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$
and additionally Assumption 3.5 with $N(m) = Nh$ is fulfilled, we obtain

$$\sup_{1\le k\le Nh} w_M^2(h,k)\,\|\boldsymbol{S}_3(m,k,h)\|_{\boldsymbol{A}}^2 \overset{D}{\longrightarrow} \sup_{0<t\le N} \rho_M^2(t)\,\|\boldsymbol{W}(t+1) - \boldsymbol{W}(t)\|_{\boldsymbol{A}}^2, \tag{3.30}$$

where $\{\boldsymbol{W}(t) : t \ge 0\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$.

The matrix $\boldsymbol{A}$ can be replaced by a consistent estimator.

If $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta,\ \beta \in (0,1]$ and $\boldsymbol{A} = \boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_2^{-1}$ as well as $\boldsymbol{W}_1(\cdot)$ and $\boldsymbol{W}_2(\cdot)$ are independent, then we obtain a pivotal limit.

As well as if $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$, we obtain a pivotal limit by choosing the matrix $\boldsymbol{A} = \boldsymbol{\Sigma_1}^{-1}$.

Corollary 3.5

Let the assumptions of Corollary 3.4 hold and additionally $\boldsymbol{A} = \boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_2^{-1}$ be the covariance matrix of the independent Wiener processes $\boldsymbol{W}_1(\cdot)$ and $\boldsymbol{W}_2(\cdot)$, then

a) if $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} \beta, \quad \beta \in (0,1]$

$$\sup_{1\le k\le Nh} w_M^2(h,k) \left\| \boldsymbol{S}_3(m,k,h) \right\|^2_{\boldsymbol{\Sigma}_1^{-1}} \overset{D}{\longrightarrow} \sup_{0<t\le N} \rho_M^2(t) \left\| \boldsymbol{W}_1(t+1) - \boldsymbol{W}_1(t) - \boldsymbol{W}_2(\beta) \right\|^2,$$

where $\{\boldsymbol{W}_1(t) : t \geq 0\}$ and $\{\boldsymbol{W}_2(t) : t \geq 0\}$ are independent standard Wiener processes. In the situation of starting to search for a change after $m+h$ observations, so $k \geq h$, the limit distribution simplifies analogously to Corollary 3.3 to

$$\sup_{h\le k\le Nh} w_M^2(h,k) \left\| \boldsymbol{S}_3(m,k,h) \right\|^2_{\boldsymbol{\Sigma}_1^{-1}} \overset{D}{\longrightarrow} \sup_{1\le t\le N} \rho_M^2(t) \left\| \boldsymbol{W}_1(t) - \boldsymbol{W}_1(t-1) - \boldsymbol{W}_2(\beta) \right\|^2,$$

where the Wiener processes are defined as above.

b) If $\frac{h}{m} \overset{m\to\infty}{\longrightarrow} 0$ then

$$\sup_{1\le k\le Nh} w_M^2(h,k) \left\| \boldsymbol{S}_3(m,k,h) \right\|^2_{\boldsymbol{\Sigma}_1^{-1}} \overset{D}{\longrightarrow} \sup_{0<t\le N} \rho_M^2(t) \left\| \boldsymbol{W}(t+1) - \boldsymbol{W}(t) \right\|^2,$$

where $\{\boldsymbol{W}(t) : t \geq 0\}$ is a standard Wiener process.

The norm $\|\cdot\|$ is the l_2-norm, and the matrix $\boldsymbol{A}$ can be replaced by a consistent estimator.

Proof. a) First the following simple calculation holds for independent Wiener processes $\boldsymbol{W}_1(\cdot)$ and $\boldsymbol{W}_2(\cdot)$

$$\begin{aligned}
&\sup_{0<t\le N} \rho_M^2(t) \left\| \boldsymbol{W}_1\left(\frac{1}{\beta}+t\right) - \boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right) - \beta \boldsymbol{W}_2\left(\frac{1}{\beta}\right) \right\|^2_{\boldsymbol{\Sigma}_1^{-1}} \\
&\overset{D}{=} \sup_{0<t\le N} \rho_M^2(t) \left\| \boldsymbol{\Sigma}_1^{\frac{1}{2}} \left(\widetilde{\boldsymbol{W}}_1\left(\frac{1}{\beta}+t\right) - \widetilde{\boldsymbol{W}}_1\left(\frac{1}{\beta}+t-1\right) - \beta \widetilde{\boldsymbol{W}}_2\left(\frac{1}{\beta}\right) \right) \right\|^2_{\boldsymbol{\Sigma}_1^{-1}} \\
&= \sup_{0<t\le N} \rho_M^2(t) \left\| \widetilde{\boldsymbol{W}}_1\left(\frac{1}{\beta}+t\right) - \widetilde{\boldsymbol{W}}_1\left(\frac{1}{\beta}+t-1\right) - \beta \widetilde{\boldsymbol{W}}_2\left(\frac{1}{\beta}\right) \right\|^2 \\
&\overset{D}{=} \sup_{0<t\le N} \rho_M^2(t) \left\| \widetilde{\boldsymbol{W}}_1(t+1) - \widetilde{\boldsymbol{W}}_1(t) - \widetilde{\boldsymbol{W}}_2(\beta) \right\|^2.
\end{aligned}$$

Here, $\widetilde{\boldsymbol{W}}_1(\cdot)$ and $\widetilde{\boldsymbol{W}}_2(\cdot)$ are independent standard Wiener processes. If we start monitoring after h observations, the proof is analogous.

b) If $\frac{h}{m}$ converges to zero, we get with $\boldsymbol{\Sigma}_1$ being the covariance matrix of $\boldsymbol{W}(\cdot)$

$$\begin{aligned}
&\sup_{0<t\le N} \rho_M^2(t) \left\| \boldsymbol{W}(t+1) - \boldsymbol{W}(t) \right\|^2_{\boldsymbol{\Sigma}_1^{-1}} \\
&\overset{D}{=} \sup_{0<t\le N} \rho_M^2(t) \left\| \boldsymbol{\Sigma}_1^{-\frac{1}{2}} \left(\widetilde{\boldsymbol{W}}(t+1) - \widetilde{\boldsymbol{W}}(t) \right) \right\|^2_{\boldsymbol{\Sigma}_1^{-1}}
\end{aligned}$$

$$= \sup_{0<t\leq N} \rho_M^2(t) \left\| \widetilde{\boldsymbol{W}}(t+1) - \widetilde{\boldsymbol{W}}(t) \right\|^2,$$

where $\widetilde{\boldsymbol{W}}(\cdot)$ is a standard Wiener process.

□

4. Asymptotics under the Alternative Hypothesis

In this chapter we will show that under some mild conditions all monitoring schemes have asymptotic power one if the length m of the historical data set converges to infinity. This means, if a change occurs within the time series the procedure will stop at a finite point in time with probability approaching one. If a procedure has asymptotic power one, it is often called consistent.

The null hypothesis holds for the time series as long as no change occurs, so the time series before the change has to fulfill the assumptions under the null hypothesis which are stated for each statistic in part a) of the following assumptions. Since m converges to infinity the change point which is at some point in time after m, also converges to infinity. Thus we distinguish between the cases that the change-point k^* is of the same or lower order as m or the change-point grows faster to infinity than m. Note that the second case can only take place in the open-end procedure. Since in case of the MOSUM statistic the constant bandwidth h converges to infinity with m growing to infinity, m is replaced by h in Assumptions 4.3 about the MOSUM. The required assumptions under the alternative for the two mentioned cases of the behaviour of k^* are denoted in part b) respectively c) in the assumptions for each statistic.

4.1. Consistency of the Modified MOSUM

Assumption 4.1

a) The time series before the change fulfills the assumptions under the null hypothesis.

b) The change-point is of the form $k^* \leq \lfloor m\nu \rfloor$ for some $0 < \nu < N$ (in case of the open-end procedure $N = \infty$). Furthermore, there exists a ball $U(x_0)$ around x_0 with $N > x_0 > \nu$ and $\rho(x) \geq c > 0$ for $x \in U(x_0)$ as well as

$$\frac{1}{m} \left\| \sum_{j=m+\max(k^*,\lfloor\lfloor x_0 m\rfloor h\rfloor)+1}^{m+\lfloor x_0 m\rfloor} \Big(H(\boldsymbol{X}_j, \widehat{\theta}_m) - \boldsymbol{E}_H \Big) \right\| = o_P(1)$$

for some $\boldsymbol{E}_H$.

c) In the open-end procedure for an arbitrarily late change k^* and if $\liminf_{x\to\infty} x\rho(x) > 0$, then we have for $l \to \infty$

$$\frac{1}{l}\left\|\sum_{j=m+\lfloor\lceil\frac{k^*}{h}\rceil h\rfloor+1}^{m+\lfloor\lceil\frac{k^*}{h}\rceil h\rfloor+l}\Big(H(\boldsymbol{X}_j,\widehat{\theta}_m)-\boldsymbol{E}_H\Big)\right\| = o_P(1),$$

for some $\boldsymbol{E}_H$.

The $\boldsymbol{E}_H$, used in the previous assumption as well as in the following assumptions in this section concerning the other statistics, is usually equal to $E(H(\boldsymbol{X}_t,\theta_0)),\ t > k^*$ which is not zero. By the assumptions, the monitoring function inserted with the time series after the change and the estimated parameter can be asymptotically replaced by $\boldsymbol{E}_H$. So $\boldsymbol{E}_H$ influences the alternatives which are asymptotically always detected by the method. More precisely, if $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H \neq 0$ the procedure has asymptotic power one (see the next Theorem about the consistency).

Theorem 4.1

Let the alternative hypothesis hold.
Under Assumptions 4.1a)-b) and $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H \neq 0$, the procedure with a possibly infinite time horizon has asymptotic power one:

$$\sup_{1\le k<N(m)} w^2(m,k)\boldsymbol{S}_2(m,k,h)^T\boldsymbol{A}\boldsymbol{S}_2(m,k,h) \xrightarrow{P} \infty \tag{4.1}$$

If $N(m) = \infty$ and the change-point is arbitrarily late, we require the additional Assumption 4.1c). The matrix $\boldsymbol{A}$ can be replaced by a consistent estimator.

Proof. Since we have to show equation (4.1) it is enough to prove

$$w^2(m,\tilde{k})\boldsymbol{S}_2(m,\tilde{k},h)^T\boldsymbol{A}\boldsymbol{S}_2(m,\tilde{k},h) \xrightarrow{P} \infty$$

for a point $\tilde{k}$ in time later than the change-point k^*. However, in the closed-end procedure $\tilde{k}$ has to be earlier than the end of the observation horizon.

a) Closend-end procedure:
The change-point is $k^* = \lfloor \nu m\rfloor$ and we choose $\tilde{k} = \lfloor x_0 m\rfloor$. So by $x_0 > \nu$ we have $\tilde{k} > k^*$. But then the time point $\lfloor \tilde{k}h\rfloor$, including in the lower bound of the sum of the modified MOSUM, can be earlier or later than the change-point k^*. So we have to distinguish the following two cases:

(i) If $k^* > \lfloor \tilde{k}h\rfloor$:
We split the time series into the observations before and after the change-point. First we look at the observations prior to the change. Assumption 4.1a) and an application of Theorem 3.1 in Kirch and Kamgaing (2015) implies

$$\frac{1}{m}\sum_{t=m+\lfloor\tilde{k}h\rfloor+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m) = o_P(1).$$

The explanation for the previous equality is that first of all

$$\begin{aligned}
&\frac{1}{m}\sum_{t=m+\lfloor\tilde{k}h\rfloor+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m)\\
&=\frac{1}{m}\left(\sum_{t=m+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m)-\sum_{t=m+1}^{m+\lfloor\tilde{k}h\rfloor} H(\boldsymbol{X}_t,\hat{\theta}_m)\right).
\end{aligned}\tag{4.2}$$

Then Theorem 3.1 in Kirch and Kamgaing (2015) implies

$$\begin{aligned}
O_P(1)&=\sup_{1\le k\le k^*} w^2(m,k)\left\|\sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\hat{\theta}_m)\right\|_{\boldsymbol{A}}^2\\
&\ge\frac{1}{m}\rho^2\left(\frac{k^*}{m}\right)\left\|\sum_{t=m+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m)\right\|_{\boldsymbol{A}}^2.
\end{aligned}$$

Since ρ is bounded due to the fact that k^* and m have the same order, we know

$$\frac{1}{m}\left\|\sum_{t=m+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m)\right\|_{\boldsymbol{A}}^2=O_P(1)$$

and hence

$$\frac{1}{m^2}\left\|\sum_{t=m+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m)\right\|_{\boldsymbol{A}}^2=\left\|\frac{1}{m}\sum_{t=m+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m)\right\|_{\boldsymbol{A}}^2=o(1)O_P(1)=o_P(1).$$

The second sum figuring in equation (4.2) can be treated in the same way.

Next we focus on the observations between the change-point k^* and $\tilde{k}$. By Assumption 4.1b) we obtain

$$\frac{1}{m}\sum_{t=m+k^*+1}^{m+\lfloor mx_0\rfloor} H(\boldsymbol{X}_t,\hat{\theta}_m)=\frac{1}{m}\left(\lfloor mx_0\rfloor-k^*\right)\boldsymbol{E}_H+o_P(1)=\left(x_0-\frac{k^*}{m}\right)\boldsymbol{E}_H+o_P(1).$$

Further we combine

$$\begin{aligned}
\sum_{t=m+\lfloor\tilde{k}h\rfloor+1}^{m+\tilde{k}} H(\boldsymbol{X}_t,\hat{\theta}_m)&=\sum_{t=m+\lfloor\tilde{k}h\rfloor+1}^{m+k^*} H(\boldsymbol{X}_t,\hat{\theta}_m)+\sum_{t=m+k^*+1}^{m+\tilde{k}} H(\boldsymbol{X}_t,\hat{\theta}_m)\\
&=m\left(\left(x_0-\frac{k^*}{m}\right)\boldsymbol{E}_H+o_P(1)\right).
\end{aligned}$$

Finally we get with $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H\neq 0$

$$\begin{aligned}
w^2(m,\tilde{k})\left\|\boldsymbol{S}_2(m,\tilde{k},h)\right\|_{\boldsymbol{A}}^2&=\frac{1}{m}\rho^2\left(\frac{\tilde{k}}{m}\right)\left\|m((x_0-\frac{k^*}{m})\boldsymbol{E}_H+o_P(1))\right\|_{\boldsymbol{A}}^2\\
&=m\rho^2\left(\frac{\lfloor x_0 m\rfloor}{m}\right)\left(x_0-\frac{k^*}{m}\right)^2\left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2+o_P(1)\right)
\end{aligned}$$

$$
\begin{aligned}
&= m\rho^2 (x_0 + o(1)) \left(x_0 - \frac{k^*}{m} \right)^2 \left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1) \right) \\
&\geq m\rho^2 (x_0 + o(1)) \left(x_0 - \frac{\lfloor m\nu \rfloor}{m} \right)^2 \left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1) \right) \\
&\geq m\rho^2 (x_0 + o(1)) (x_0 - \nu)^2 \left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1) \right) \\
&\xrightarrow{P} \infty.
\end{aligned}
$$

(ii) If $k^* \leq \lfloor \tilde{k}h \rfloor$:
In this case Assumption 4.1b) implies

$$
\frac{1}{m} \left\| \sum_{j=m+\lfloor \lfloor x_0 m \rfloor h \rfloor + 1}^{m+\lfloor x_0 m \rfloor} \left(H(\boldsymbol{X}_j, \widehat{\theta}_m) - \boldsymbol{E}_H \right) \right\| = o_P(1)
$$

and consequently

$$
\begin{aligned}
\frac{1}{m} \sum_{j=m+\lfloor \lfloor x_0 m \rfloor h \rfloor + 1}^{m+\lfloor x_0 m \rfloor} H(\boldsymbol{X}_j, \widehat{\theta}_m) &= \frac{1}{m} \left(\lfloor m x_0 \rfloor - \lfloor \lfloor x_0 m \rfloor h \rfloor \right) \boldsymbol{E}_H + o_P(1) \\
&= x_0 (1-h) \boldsymbol{E}_H + o_P(1).
\end{aligned}
$$

Finishing the proof we obtain from $\boldsymbol{A}^{\frac{1}{2}} \boldsymbol{E}_H \neq 0$

$$
\begin{aligned}
w^2(m, \tilde{k}) \left\| \boldsymbol{S}_2(m, \tilde{k}, h) \right\|_{\boldsymbol{A}}^2 &= \frac{1}{m} \rho^2 \left(\frac{\tilde{k}}{m} \right) \| m (x_0 (1-h) \boldsymbol{E}_H + o_P(1)) \|_{\boldsymbol{A}}^2 \\
&= m\rho^2 \left(\frac{\lfloor x_0 m \rfloor}{m} \right) x_0^2 (1-h)^2 \left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1) \right) \\
&= m\rho^2 (x_0 + o(1)) x_0^2 (1-h)^2 \left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1) \right) \xrightarrow{P} \infty.
\end{aligned}
$$

b) Open-end procedure:

- $k^* \leq \lfloor m\nu \rfloor$ with $N > \nu < x_0$:
  The proof is analogous to the closed-end case.
- $k^*/m \to \infty$:
  In this case we choose $\tilde{k} = \left\lceil \frac{k^*}{h} \right\rceil$. This choice automatically results in the second case of part a), where $k^* \leq \lfloor \tilde{k}h \rfloor$.

  From Assumption 4.1c) it follows that

$$
\frac{1}{\left\lceil \frac{k^*}{h} \right\rceil - \left\lfloor \left\lceil \frac{k^*}{h} \right\rceil h \right\rfloor} \sum_{m+\left\lfloor \left\lceil \frac{k^*}{h} \right\rceil h \right\rfloor + 1}^{m+\left\lceil \frac{k^*}{h} \right\rceil} H(\boldsymbol{X}_t, \widehat{\theta}_m) = \boldsymbol{E}_H + o_P(1).
$$

  Further we obtain

$$
\sup_{k \geq 1} w^2(m, k) \left\| \sum_{m+\lfloor kh \rfloor + 1}^{m+k} H(\boldsymbol{X}_t, \widehat{\theta}_m) \right\|_{\boldsymbol{A}}^2 \geq w^2(m, \tilde{k}) \left\| \sum_{m+\lfloor \tilde{k}h \rfloor + 1}^{m+\tilde{k}} H(\boldsymbol{X}_t, \widehat{\theta}_m) \right\|_{\boldsymbol{A}}^2
$$

$$= \frac{1}{m}\rho^2\left(\frac{\tilde{k}}{m}\right)\left\|\left(\left\lceil\frac{k^*}{h}\right\rceil - \left\lfloor\left\lceil\frac{k^*}{h}\right\rceil h\right\rfloor\right)(\boldsymbol{E}_H + o_P(1))\right\|_{\boldsymbol{A}}^2$$
$$= \rho^2\left(\frac{\tilde{k}}{m}\right)\left(\frac{\left(\lceil\frac{k^*}{h}\rceil - \lfloor\lceil\frac{k^*}{h}\rceil h\rfloor\right)^2}{m}\right)\left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1)\right).$$

Eventually, using $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H \neq 0$ and

$$\frac{\left(\lceil\frac{k^*}{h}\rceil - \lfloor\lceil\frac{k^*}{h}\rceil h\rfloor\right)^2}{m} \geq \frac{\left(\lceil\frac{k^*}{h}\rceil(1-h)\right)^2}{m} = \left(\frac{\widetilde{k}}{m}\right)^2(1-h)^2 m$$

together with $\liminf_{x\to\infty} x\rho(x) > 0$, the assertion follows.

□

4.2. Consistency of the Page-CUSUM

To guarantee asymptotic power one of the monitoring procedure that uses the Page-CUSUM, we need the same assumption as in the case of employing the standard CUSUM statistic. The reason is that the Page-CUSUM can be traced back to the CUSUM statistic for proving consistency. Details are shown in the appropriate Theorem 4.2.

Since the CUSUM statistic has already been considered in Kirch and Kamgaing (2015), Assumption 4.2 is equivalent to Assumption A.4 in this publication.

Assumption 4.2

a) The time series before the change fulfills the assumptions under the null hypothesis.

b) The change-point is of the form $k^* \leq \lfloor m\nu \rfloor$ for some $0 < \nu < N$ (in case of the open-end procedure $N = \infty$). Furthermore, there exists a ball $U(x_0)$ around x_0 with $N > x_0 > \nu$ and $\rho(x) \geq c > 0$ for $x \in U(x_0)$ as well as

$$\frac{1}{m}\left\|\sum_{j=m+k^*+1}^{m+\lfloor x_0 m\rfloor}\left(H(\boldsymbol{X}_j, \widehat{\theta}_m) - \boldsymbol{E}_H\right)\right\| = o_P(1).$$

c) In the open-end procedure for an arbitrarily late change k^* and if $\liminf_{x\to\infty} x\rho(x) > 0$ as well as for $l \to \infty$, we have

$$\frac{1}{l}\left\|\sum_{j=m+k^*+1}^{m+k^*+l}\left(H(\boldsymbol{X}_j, \widehat{\theta}_m) - \boldsymbol{E}_H\right)\right\| = o_P(1),$$

for some $\boldsymbol{E}_H$.

Theorem 4.2

Let the alternative hypothesis hold.
Under Assumptions 4.2a)-b) and $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H \neq 0$, the procedure with a possibly infinite time horizon has asymptotic power one:

$$\sup_{1\leq k\leq Nm} w^2(m,k) \max_{0\leq i\leq k} \boldsymbol{S}_4(m,k,i)^T \boldsymbol{A}\boldsymbol{S}_4(m,k,i) \xrightarrow{P} \infty$$

If $N(m) = \infty$ and the change-point is arbitrarily late, the additional Assumption 4.2c) is required. The matrix $\boldsymbol{A}$ can be replaced by a consistent estimator.

Proof. First we have

$$\begin{aligned}
&\sup_{1\leq k<N(m)} w^2(m,k) \max_{0\leq i\leq k} \boldsymbol{S}_4(m,k,i)^T \boldsymbol{A}\boldsymbol{S}_4(m,k,i)\\
&= \sup_{1\leq k<N(m)} w^2(m,k) \max_{0\leq i\leq k} \left\| \sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t,\widehat{\theta}_m) \right\|_{\boldsymbol{A}}^2\\
&\geq \sup_{1\leq k<N(m)} w^2(m,k) \left\| \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t,\widehat{\theta}_m) \right\|_{\boldsymbol{A}}^2
\end{aligned}$$

This term is equal to the standard CUSUM-statistic. So to finish the proof of showing that the Page-CUSUM has asymptotic power one, we refer to Theorem 4.1 in Kirch and Kamgaing (2015). □

4.3. Consistency of the MOSUM

As explained in the beginning of the chapter, we distinguish between the cases that the change-point k^* and h are of the same or lower order and k^* grows faster than h to prove the consistency for the MOSUM statistic.

However in Assumption 4.3c) for the open-end procedure we need a rate for the speed of convergence to infinity if k^*/h approaches infinity.

Assumption 4.3

a) The time series before the change fulfills the assumptions under the null hypothesis.

b) The change point is of the form $k^* \leq \lfloor h\nu \rfloor$ for some $0 < \nu < N\frac{m}{h}$ (in case of the open-end procedure $N = \infty$). Furthermore, there exists a ball $U(x_0)$ around x_0 with $N\frac{m}{h} - 1 > x_0 > \nu$ and $\rho(x+1) \geq c > 0$ for $x \in U(x_0)$ as well as

$$\frac{1}{h}\left\| \sum_{j=m+\lfloor x_0 h\rfloor+1}^{m+\lfloor x_0 h\rfloor+h} \Big(H(\boldsymbol{X}_j,\widehat{\theta}_m) - \boldsymbol{E}_H \Big) \right\| = o_P(1).$$

c) In the open-end procedure and if the change point fulfills $k^* = o\left(h^{1+\frac{\nu}{2}}\right), \nu > 0$ as well as if $\liminf_{x\to\infty} x^{\frac{1}{\nu}}\rho(x) > 0,\ \nu > 2$, then for $h \to \infty$ we have

$$\frac{1}{h}\left\|\sum_{j=m+k^*+1}^{m+k^*+h}\left(H(\boldsymbol{X}_j,\widehat{\theta}_m) - \boldsymbol{E}_H\right)\right\| = o_P(1),$$

for some $\boldsymbol{E}_H$.

Theorem 4.3

Let the alternative hypothesis hold.
Under Assumptions 4.3a)-b) and $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H \neq 0$, the procedure has asymptotic power one:

$$\sup_{1\leq k<N(m)} w_M^2(h,k)\boldsymbol{S}_3(m,k,h)^T\boldsymbol{A}\boldsymbol{S}_3(m,k,h) \xrightarrow{P} \infty \tag{4.3}$$

If $N(m) = \infty$ and the change-point is arbitrarily late, then Assumption 4.3c) is additionally needed. The matrix $\boldsymbol{A}$ can be replaced by a consistent estimator.

Proof. The proof is analogous to the one that shows the consistency of the modified MOSUM.
Since we have to show (4.3) it is enough to prove

$$w_M^2(h,\tilde{k})\boldsymbol{S}_3(m,\tilde{k},h)^T\boldsymbol{A}\boldsymbol{S}_3(m,\tilde{k},h) \xrightarrow{P} \infty$$

for a point in time $\tilde{k}$ later than the change-point k^* and prior to the time horizon in the closed-end case.

a) Closed-end procedure:
It holds $k^* \leq \lfloor \nu h\rfloor$ and we choose $\tilde{k} = \lfloor x_0 h\rfloor + h$. Then it holds $k^* \leq \tilde{k} - h$. With $\tilde{k} = \lfloor hx_0\rfloor + h$ and Assumption 4.3b) we know

$$\frac{1}{h}\left\|\sum_{j=m+\lfloor x_0h\rfloor+1}^{m+\lfloor x_0h\rfloor+h}\left(H(\boldsymbol{X}_j,\widehat{\theta}_m) - \boldsymbol{E}_H\right)\right\| = o_P(1)$$

and conclude

$$\begin{aligned}\frac{1}{h}\sum_{j=m+\lfloor x_0h\rfloor+1}^{m+\lfloor x_0h\rfloor+h} H(\boldsymbol{X}_j,\widehat{\theta}_m) &= \frac{1}{h}(\lfloor x_0h\rfloor + h - \lfloor x_0h\rfloor)\boldsymbol{E}_H + o_P(1)\\ &= \boldsymbol{E}_H + o_P(1).\end{aligned}$$

Now, $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H \neq 0$ entails

$$w^2(h,\tilde{k})\left\|\boldsymbol{S}_3(m,\tilde{k},h)\right\|_{\boldsymbol{A}}^2 = \frac{1}{h}\rho_M^2\left(\frac{\tilde{k}}{h}\right)\left\|h\left(\boldsymbol{E}_H + o_P(1)\right)\right\|_{\boldsymbol{A}}^2$$

$$= h\rho_M^2\left(\frac{\lfloor x_0 h\rfloor}{h}\right)\left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1)\right)$$
$$= h\rho_M^2\left(x_0 + 1 + o(1)\right)\left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1)\right) \xrightarrow{P} \infty.$$

b) Open-end procedure:

- $k^* \le \lfloor h\nu\rfloor$ with $Nm/h > x_0 > \nu$:
  The proof is analogous to that of the closed-end case.
- For $k^*/h \to \infty$ and $k^* = o\left(h^{1+\frac{\nu}{2}}\right)$:
  We choose $\tilde{k} = k^* + h$ which means $k^* = \tilde{k} - h$.
  From Assumption 4.3c) it follows that

$$\frac{1}{h}\sum_{m+\tilde{k}-h+1}^{m+\tilde{k}} H(\boldsymbol{X}_t, \widehat{\theta}_m) = \frac{1}{h}\sum_{m+k^*+1}^{m+k^*+h} H(\boldsymbol{X}_t, \widehat{\theta}_m) = \boldsymbol{E}_H + o_P(1).$$

We further have

$$\sup_{k\ge 1} w_M^2(h,k)\left\|\sum_{m+k-h+1}^{m+k} H(\boldsymbol{X}_t, \widehat{\theta}_m)\right\|_{\boldsymbol{A}}^2 \ge w_M^2(h,\tilde{k})\left\|\sum_{m+\tilde{k}-h+1}^{m+\tilde{k}} H(\boldsymbol{X}_t, \widehat{\theta}_m)\right\|_{\boldsymbol{A}}^2$$
$$= \frac{1}{h}\rho_M^2\left(\frac{\tilde{k}}{h}\right)\|h(\boldsymbol{E}_H + o_P(1))\|_{\boldsymbol{A}}^2$$
$$= \rho^2\left(\frac{\tilde{k}}{h}\right) h\left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1)\right)$$
$$= \left(\frac{\widetilde{k}}{h}\right)^{\frac{2}{\nu}}\rho^2\left(\frac{\tilde{k}}{h}\right) h^{1+\frac{2}{\nu}}\widetilde{k}^{-\frac{2}{\nu}}\left(\|\boldsymbol{E}_H\|_{\boldsymbol{A}}^2 + o_P(1)\right)$$

and finally

$$h^{1+\frac{2}{\nu}}\tilde{k}^{-\frac{2}{\nu}} = h\left(\frac{h}{\tilde{k}}\right)^{\frac{2}{\nu}} = h\left(\frac{h}{k^*+h}\right)^{\frac{2}{\nu}} = h\left(\frac{1}{\frac{k^*}{h}+1}\right)^{\frac{2}{\nu}} = h\left(\frac{h}{k^*+h}\right)^{\frac{2}{\nu}}$$
$$= h\left(\frac{\frac{h}{h^{1+\frac{\nu}{2}}}}{\frac{k^*}{h^{1+\frac{\nu}{2}}} + h^{-\frac{\nu}{2}}}\right)^{\frac{2}{\nu}} = h\left(\frac{h^{-\frac{\nu}{2}}}{\frac{k^*}{h^{1+\frac{\nu}{2}}} + h^{-\frac{\nu}{2}}}\right)^{\frac{2}{\nu}} = \left(\frac{1}{\frac{k^*}{h^{1+\frac{\nu}{2}}} + h^{-\frac{\nu}{2}}}\right)^{\frac{2}{\nu}}$$
$$= \left(\frac{1}{o(1) + h^{-\frac{\nu}{2}}}\right)^{\frac{2}{\nu}} = \frac{1}{o(1)} \to \infty.$$

Since $\liminf_{x\to\infty} x^{\frac{1}{\nu}}\rho(x) > 0$ and $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{E}_H \ne 0$, the assertion follows.

□

5. Examples

In this chapter we give an overview of the models in which the discussed methods have already been considered in the literature. Moreover, we extend these models to new examples which are included in our very general setup.

5.1. Overview

In the sequential procedure, the modified MOSUM and the Page-CUSUM statistics have, to the best of our knowledge, only been considered in the linear model in Chen and Tian (2010) and Fremdt (2014), respectively. The MOSUM statistic has even only been analysed in the location model in Horváth et al. (2012) and Aue et al. (2008).

In Chen and Tian (2010) the limit distribution and the consistency under the alternative of the modified MOSUM are developed. However, due to an error in their proof of the asymptotics under the null hypothesis, their limit distribution is incorrect.

In the mentioned publication of Kirch and Kamgaing (2015), the theory is developed for the CUSUM statistic in an equivalently general setting. To unify the assumptions for all statistics under the null hypothesis we used the same assumptions as for the CUSUM statistic to prove the corresponding limit distributions for the different types of statistics. Under the alternative there are modified assumptions for each statistic discussed in this work. However, Assumption A.4c) for the CUSUM statistic in Kirch and Kamgaing (2015) is stricter than necessary for our statistics but unifies the assumptions for the CUSUM as well as our statistics.

Therefore, in the mentioned publication about the CUSUM there are many new examples for our statistics, in which the assumptions are fulfilled and have not yet been considered in the literature, i.e. binary models, poisson-autoregressive models and non-linear models. These models will be studied in detail in the next paragraphs where we will give the estimating and monitoring functions to clarify the generality of our method.

Since the statistics dealt with in this work have only been examined in the location model or linear model, these models will be considered in detail in the next two paragraphs, to see how the general setting applies to specified models of time series.

Note that the MOSUM in case of $h/m \to 0,\ m \to \infty$, discussed in Section 3.3, needs a different assumption under H_0 compared to the other statistics as stated in Assumption 3.5. Therefore, in the next paragraphs, we will point out the required conditions on the regressors to fulfill this assumption.

5.2. Linear regression model

The data set has the form

$$X_t = \boldsymbol{x}_t^T \boldsymbol{\beta}_t + \epsilon_t,\ 0 \le t < \infty,$$

where $\boldsymbol{\beta}_t = (\beta_{t,1}, \ldots, \beta_{t,p})^T$ is the unknown parameter and $\boldsymbol{x} = (1, x_{t,2}, \ldots, x_{t,p})$ are the random regressors. Moreover, ϵ_t are the errors with mean zero and variance σ^2. The regressors and the error sequence have to be independent. In the literature the residuals are often supposed to be i.i.d. or uncorrelated with some moment conditions, but the minimal requirement is a functional central limit theorem which follows from an invariance principle as in Assumption 3.6. In the case of i.i.d residuals this follows immediately from Komlós et al. (1975).

The testing procedure is established by

$$\begin{array}{rll} \text{the null hypothesis } H_0: & \boldsymbol{\beta}_t = \boldsymbol{\beta}_0 \quad 0 \le t < \infty & \text{against} \\ \text{the alternative hypothesis } H_A: & \boldsymbol{\beta}_t = \boldsymbol{\beta}_0 \quad m+1 \le t \le k^*, & \\ & \boldsymbol{\beta}_t = \boldsymbol{\beta}_A \quad k^*+1 \le t < \infty. & \end{array}$$

Here, k^* is the change point, and m denotes the length of the historical data set.

Assuming the data follows the linear model, the modified MOSUM is considered in Chen and Tian (2010) and the Page-CUSUM in Fremdt (2014). Both use the least square estimator to obtain an approximation for the unknown parameter $\boldsymbol{\beta}_0$ based on the historical data set. So the estimating function is

$$G(X_t, \boldsymbol{\beta}) = \boldsymbol{x}_t(X_t - \boldsymbol{\beta}^T \boldsymbol{x}_t).$$

And the least squares estimator is obtained as solution of the equation

$$\sum_{t=1}^{m} G(X_t, \widehat{\boldsymbol{\beta}}_m) = 0.$$

If we adopt another estimator we would have to modify the estimating function in a suitable way.

The monitoring function for parameter changes is given by the sequence of the estimated residuals, which can be expressed by

$$H(X_t, \boldsymbol{\beta}) = X_t - \boldsymbol{\beta}^T \boldsymbol{x}_t.$$

Since H is the first row of G, $B(\theta_0) = (1, 0, \ldots, 0)^T$ to garantee that the functions $B(\theta_0)G$ and H have the same dimension.

If we want to search for a change in the variance, the monitoring function is given by

$$H(X_t, \boldsymbol{\beta}) = (X_t - \boldsymbol{\beta}^T \boldsymbol{x}_t)^2 - \frac{1}{m} \sum_{t=1}^{m} (X_t - \boldsymbol{\beta}^T \boldsymbol{x}_t)^2.$$

To justify Assumption 3.2, Chen and Tian (2010) assumed a law of large numbers for the regressors: There exist a positive definite matrix $\boldsymbol{C}$ and $\tau > 0$ such that

$$\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T - \boldsymbol{C} = O(n^{-\tau}), \quad a.s. \tag{5.1}$$

Fremdt (2014) uses a slightly stronger assumption to get Assumption 3.2. The regressors are a stationary series, and there is a vector $d = (d_1, \ldots, d_p)^T$ and constants $K > 0,\ \tau > 0$ such that

$$E\left(\left| \sum_{i=1}^{k} (x_{i,j} - d_j) \right|^{\nu} \right) \le K k^{\frac{\nu}{2}}, \quad 1 \le j \le p, \quad \nu > 2,$$

see Lemma 1 in Fremdt (2014). However, Assumption (5.1) is sufficient.

Since $H(X_t, \boldsymbol{\beta}_0) = \epsilon_t, 0 \leq t < \infty$, Assumption 3.3 are accordingly simplified. The existing literature often uses a much stronger assumption on the error process, the residuals should satisfy a strong invariance principle, namely as in Assumption 3.6 or in Assumption A.3c) in Kirch and Kamgaing (2015), in order to obtain a pivotal limit in an easy way (see Chen and Tian (2010) for the modified MOSUM, Fremdt (2014) for the Page-CUSUM and i.e. Aue et al. (2006b) and Horváth et al. (2004) as well as Kirch and Kamgaing (2015) for the CUSUM statistic). In a dependent setting, this strong invariance principle is harder to obtain than Assumption 3.3 because the sums of the historical residuals and the subsequent residuals have to be asymptotically independent. This strong invariance principle is fulfilled for i.i.d. errors (Komlós et al. (1975), Komlós et al. (1976)). However they are also satisfied for certain weak dependent processes (Schmitz and Steinebach (2010)) as well as for certain martingale differences (Aue et al. (2006a)).

Under the alternative Assumption A4c) in Kirch and Kamgaing (2015), which is the unifying condition for all statistics, is obtained in case of a parameter change with $E_H = E(H(\boldsymbol{X}_t, \beta_0)) = \boldsymbol{c}_1^T(\beta_0 - \beta_A) \neq 0,\ t > k^*$, where $\boldsymbol{c}_1$ is the first column of the matrix $\boldsymbol{C}$ in (5.1).

As mentioned, so far the MOSUM procedure has not been considered in linear regression. However if $h/m = O(1)$ the needed assumptions are identical to those of the CUSUM method, so they are also satisfied with the same conditions on the regressors.

But if $h/m \to 0,\ m \to \infty$, we need Assumption 3.5. So we have to check the necessary conditions to fulfill this assumption.

Lemma 5.1 In case of a parameter change

Let $\frac{h}{m} \to 0$ and the errors $\{e_i\}$ be uncorrelated as well as the errors and the regressors $\{\boldsymbol{x}_i\}$ be independent. Furthermore, let $\boldsymbol{x}_i$ be stationary and satisfy (5.1). As estimator for $\boldsymbol{\beta}$ is used the least squares estimator. Then Assumption 3.5 holds for the closed-end procedure with time horizon Nh.

Proof. In the linear model, Assumption 3.5 simplifies to

$$\sup_{1\leq k<N(m)} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}k^{\frac{1}{\nu}}}\right)\left|\sum_{i=m+k-h+1}^{m+k}\widehat{\epsilon}_i - \left(\sum_{i=m+k-h+1}^{m+k}\epsilon_i - \frac{h}{m}\sum_{i=1}^{m}\epsilon_i\right)\right| = o_P(1).$$

First of all the least squares estimator is

$$\widehat{\boldsymbol{\beta}}_m = \left(\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{x}_i X_i.$$

Furthermore, it holds

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_0 &= \left(\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{x}_i X_i - \boldsymbol{\beta}_0 = \boldsymbol{C}_m^{-1}\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i(\boldsymbol{x}_i^T\beta_0 + \epsilon_i) - \boldsymbol{\beta}_0\\ &= \boldsymbol{C}_m^{-1}\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i\boldsymbol{x}_i^T\boldsymbol{\beta}_0 + \boldsymbol{C}_m^{-1}\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i\epsilon_i - \boldsymbol{\beta}_0\\ &= \boldsymbol{C}_m^{-1}\boldsymbol{C}_m\boldsymbol{\beta}_0 + \boldsymbol{C}_m^{-1}\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i\epsilon_i - \boldsymbol{\beta}_0 = \boldsymbol{C}_m^{-1}\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i\epsilon_i.\end{aligned}$$

The estimated residuals are defined as

$$\widehat{\epsilon}_i = X_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_m$$

and consequently we have

$$\begin{aligned}\sum_{i=m+k-h+1}^{m+k} \widehat{\epsilon}_i &= \sum_{i=m+k-h+1}^{m+k} \left(y_i - x_i^T \widehat{\beta}_m\right) = \sum_{i=m+k-h+1}^{m+k} \left(\epsilon_i - x_i^T (\widehat{\beta}_m - \beta_0)\right) \\ &= \sum_{i=m+k-h+1}^{m+k} \epsilon_i - \left(\sum_{i=m+k-h+1}^{m+k} x_i\right)^T C_m^{-1} \frac{1}{m} \sum_{i=1}^{m} x_i \epsilon_i.\end{aligned}$$

Accordingly, it is sufficient by $\boldsymbol{c}_1^T \boldsymbol{C}^{-1} \sum_{i=1}^m \boldsymbol{x}_i \epsilon_i = \sum_{i=1}^m \epsilon_i$ to show

$$\sup_{1 \le k < N(m)} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}} k^{\frac{1}{\nu}}}\right) \left| \frac{1}{m} \left(\left(\sum_{i=m+k-h+1}^{m+k} \boldsymbol{x}_i \right)^T \boldsymbol{C}_m^{-1} - h \boldsymbol{c}_1^T \boldsymbol{C}^{-1} \right) \sum_{i=1}^{m} \boldsymbol{x}_i \epsilon_i \right| = o_P(1), \tag{5.2}$$

where $\boldsymbol{C}_m = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i^T$. Equation (5.2) can be rewritten as

$$\begin{aligned}&\sup_{1 \le k < N(m)} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}} k^{\frac{1}{\nu}}}\right) \Bigg| \left(\frac{1}{m} \left(\left(\sum_{i=m+k-h+1}^{m+k} \boldsymbol{x}_i \right)^T - h \boldsymbol{c}_1^T \right) \left(\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1} \right) \right. \\ &\left. + \frac{1}{m} \left(\left(\sum_{i=m+k-h+1}^{m+k} \boldsymbol{x}_i \right)^T - h \boldsymbol{c}_1^T \right) \boldsymbol{C}^{-1} - \frac{h}{m} \boldsymbol{c}_1^T \left(\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1} \right) \right) \sum_{i=1}^{m} \boldsymbol{x}_i \epsilon_i \Bigg| \\ &= \sup_{1 \le k < N(m)} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}} k^{\frac{1}{\nu}}}\right) \Bigg| \left(\frac{1}{\sqrt{m}} \left(\left(\sum_{i=m+k-h+1}^{m+k} \boldsymbol{x}_i \right)^T - h \boldsymbol{c}_1^T \right) \left(\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1} \right) \right. \\ &\left. + \frac{1}{\sqrt{m}} \left(\left(\sum_{i=m+k-h+1}^{m+k} \boldsymbol{x}_i \right)^T - h \boldsymbol{c}_1^T \right) \boldsymbol{C}^{-1} - \frac{h}{\sqrt{m}} \boldsymbol{c}_1^T \left(\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1} \right) \right) \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \boldsymbol{x}_i \epsilon_i \Bigg|.\end{aligned} \tag{5.3}$$

Notice that, because $\{\epsilon_i, 1 \le i < \infty\}$ and $\{\boldsymbol{x}_i, 1 \le i < \infty\}$ are independent, then

$$\sum_{i=1}^{m} \boldsymbol{x}_i \epsilon_i = O_P(\sqrt{m}). \tag{5.4}$$

Next we show that the last term in brackets figuring in (5.3) is asymptotically negligible by (5.1) and $\frac{h}{m} \to 0$. To this end, observe that

$$\begin{aligned}&\sup_{1 \le k < N(m)} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}} k^{\frac{1}{\nu}}}\right) \frac{h}{\sqrt{m}} \boldsymbol{c}_1^T \left(\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1} \right) \\ &= \sup_{1 \le k < N(m)} \min\left(1, \frac{1}{h^{-\frac{1}{\nu}} k^{\frac{1}{\nu}}}\right) \sqrt{\frac{h}{m}} \boldsymbol{c}_1^T \left(\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1} \right)\end{aligned}$$

$$\leq \sqrt{\frac{h}{m}} \boldsymbol{c}_1^T \left(\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1} \right)$$
$$= o_P(1).$$

Consequently, by (5.1) and (5.4) it is sufficient that the random regressors $\boldsymbol{x_i}$ satisfy

$$\sup_{1 \leq k < N(m)} \frac{1}{\sqrt{m}} \min \left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}} k^{\frac{1}{\nu}}} \right) \left| \sum_{i=m+k-h+1}^{m+k} \boldsymbol{x}_i - h\boldsymbol{c}_1 \right| = o_P(1). \tag{5.5}$$

Since the time horizon is Nh and the assumption that the regressors x_i satisfy a law of large numbers as in (5.1), we have

$$\begin{aligned}
&\sup_{1 \leq k \leq Nh} \frac{1}{h} \left| \sum_{i=k+1}^{k+h} \boldsymbol{x}_i - h\boldsymbol{c}_1 \right| = \sup_{1 \leq k \leq Nh} \frac{1}{h} \left| \sum_{i=1}^{k+h} \boldsymbol{x}_i - (k+h)\boldsymbol{c}_1 - \left(\sum_{i=1}^{k} \boldsymbol{x}_i - k\boldsymbol{c}_1 \right) \right| \\
&\leq \sup_{1 \leq k \leq Nh} \frac{1}{h} \left| \sum_{i=1}^{k+h} \boldsymbol{x}_i - (k+h)\boldsymbol{c}_1 \right| + \sup_{1 \leq k \leq Nh} \frac{1}{h} \left| \sum_{i=1}^{k} \boldsymbol{x}_i - k\boldsymbol{c}_1 \right| \\
&= \sup_{1 \leq k \leq Nh} \frac{k+h}{h} \frac{1}{k+h} \left| \sum_{i=1}^{k+h} \boldsymbol{x}_i - (k+h)\boldsymbol{c}_1 \right| + \sup_{1 \leq k \leq Nh} \frac{k}{h} \frac{1}{k} \left| \sum_{i=1}^{k} \boldsymbol{x}_i - k\boldsymbol{c}_1 \right| \\
&\leq \sup_{1 \leq k \leq Nh} \frac{k+h}{h} \sup_{1 \leq k \leq Nh} \frac{1}{k+h} \left| \sum_{i=1}^{k+h} \boldsymbol{x}_i - (k+h)\boldsymbol{c}_1 \right| + \sup_{1 \leq k \leq Nh} \frac{k}{h} \sup_{1 \leq k \leq Nh} \frac{1}{k} \left| \sum_{i=1}^{k} \boldsymbol{x}_i - k\boldsymbol{c}_1 \right| \\
&\leq (N+1) \sup_{1 \leq k \leq Nh} \frac{1}{k+h} \left| \sum_{i=1}^{k+h} \boldsymbol{x}_i - (k+h)\boldsymbol{c}_1 \right| + N \sup_{1 \leq k \leq Nh} \frac{1}{k} \left| \sum_{i=1}^{k} \boldsymbol{x}_i - k\boldsymbol{c}_1 \right| = o_P(1).
\end{aligned}$$

By the stationarity of the regressors x_i we obtain

$$\begin{aligned}
&\sup_{1 \leq k \leq Nh} \frac{1}{\sqrt{m}} \min \left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}} k^{\frac{1}{\nu}}} \right) \left| \sum_{i=m+k-h+1}^{m+k} \boldsymbol{x}_i - h\boldsymbol{c}_1 \right| \\
&\stackrel{D}{=} \sup_{1 \leq k \leq Nh} \frac{1}{\sqrt{m}} \min \left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}} k^{\frac{1}{\nu}}} \right) \left| \sum_{i=k+1}^{k+h} \boldsymbol{x}_i - h\boldsymbol{c}_1 \right| \\
&\leq \sup_{1 \leq k \leq Nh} \frac{1}{\sqrt{m}} \min \left(\sqrt{h}, \sqrt{h} \frac{1}{h^{-\frac{1}{\nu}} k^{\frac{1}{\nu}}} \right) \sup_{1 \leq k \leq Nh} \frac{1}{h} \left| \sum_{i=k+1}^{k+h} \boldsymbol{x}_i - h\boldsymbol{c}_1 \right| \\
&\leq \sqrt{\frac{h}{m}} \sup_{1 \leq k \leq Nh} \frac{1}{h} \left| \sum_{i=k+1}^{k+h} \boldsymbol{x}_i - h\boldsymbol{c}_1 \right| \\
&= 2o_P(1) = o_P(1).
\end{aligned}$$

□

However if we use, besides stationarity and the law of large numbers in (5.1), a strong invariance principle analogous to Assumption 3.6 for the regressors $\boldsymbol{x}_i$, we are able to show condition (5.5) in the open-end case.

Lemma 5.2

Let the assumptions of Lemma 5.1 be satisfied and the regressors $\{\boldsymbol{x}_i\}$ fulfill a strong invariance principle

$$\sum_{i=1}^{k}(\boldsymbol{x}_t - \boldsymbol{c}_1) - \boldsymbol{W}(k) = O(k^{\frac{1}{\nu}}) \quad a.s., \quad \nu > 2,$$

where $\boldsymbol{W}$ is a Wiener process with some covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{c_1}$ the first column of the matrix $\boldsymbol{C}$ in (5.1). Then Assumption 3.5 holds for the open-end procedure.

Proof. As we saw in the proof of the previous lemma, it is sufficient to show equation (5.5).

By analogy with the reasoning leading to (3.27) in the proof of Theorem 3.6, we have

$$\begin{aligned}
&\sup_{k>Nh} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}k^{\frac{1}{\nu}}}\right)\left|\sum_{i=k+1}^{k+h}\boldsymbol{x}_i - h\boldsymbol{c}_1\right| \\
&= \sup_{k>Nh} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}k^{\frac{1}{\nu}}}\right)\left|\sum_{i=1}^{k+h}\boldsymbol{x}_i - (k+h)\boldsymbol{c}_1 - \left(\sum_{i=1}^{k}\boldsymbol{x}_i - k\boldsymbol{c}_1\right)\right| \\
&= \sup_{k>Nh} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}k^{\frac{1}{\nu}}}\right)\left(|\boldsymbol{W}_1(k+h) - \boldsymbol{W}_1(k)| + O_P\left((k+h)^{\frac{1}{\nu}} + k^{\frac{1}{\nu}}\right)\right) \\
&\leq \sup_{k>Nh} \min\left(1, \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right)\left|\boldsymbol{W}_1\left(\frac{k}{h}+1\right) - \boldsymbol{W}_1\left(\frac{k}{h}\right)\right| \\
&\quad + O_P(1) \max_{Nh<k<\infty} \min\left(1, \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right)\left(\frac{(k+h)^{\frac{1}{\nu}}}{\sqrt{h}} + \frac{k^{\frac{1}{\nu}}}{\sqrt{h}}\right) \\
&\leq \sup_{k>Nh} \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\left|\boldsymbol{W}_1\left(\frac{k}{h}+1\right) - \boldsymbol{W}_1\left(\frac{k}{h}\right)\right| \\
&\quad + O_P(1) \sup_{k>Nh} \min\left(1, \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right)\left(\frac{(k+h)^{\frac{1}{\nu}}}{\sqrt{h}} + \frac{k^{\frac{1}{\nu}}}{\sqrt{h}}\right).
\end{aligned}$$

The first term is in $o_P(1)$ by (3.26). For the second term, Assumptions 3.4a) and c) yield

$$\begin{aligned}
&O_P(1) \sup_{k>Nh} \min\left(1, \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right)\left(\frac{(k+h)^{\frac{1}{\nu}}}{\sqrt{h}} + \frac{k^{\frac{1}{\nu}}}{\sqrt{h}}\right) \\
&\leq O_P(1) \sup_{k>Nh} \left(\frac{h}{k}\right)^{\frac{1}{\nu}}\left(\frac{(k+h)^{\frac{1}{\nu}}}{\sqrt{h}} + \frac{k^{\frac{1}{\nu}}}{\sqrt{h}}\right) \\
&= O_P(1) \sup_{k>Nh} \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}}\left(\left(\frac{k+h}{k}\right)^{\frac{1}{\nu}} + 1\right) \\
&= O_P(1) \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}} \sup_{k>Nh}\left(\left(1+\frac{h}{k}\right)^{\frac{1}{\nu}} + 1\right) \\
&\leq O_P(1) \frac{1}{h^{1-\frac{2}{\nu}}}\left(\left(1+\frac{1}{N}\right)^{\frac{1}{\nu}} + 1\right)
\end{aligned}$$

$$= O_P(1)o(1)O(1) = o_P(1) \quad \text{as} \quad N \to \infty \quad \text{uniformly in } m.$$

Then condition (5.5) follows

$$\max_{1\le k<N(m)} \frac{1}{\sqrt{m}} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}k^{\frac{1}{\nu}}}\right)\left|\sum_{k+1}^{k+h} \boldsymbol{x}_i - h\boldsymbol{c}_1\right|$$
$$\le \max_{1\le k\le Nh} \frac{1}{\sqrt{m}} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}k^{\frac{1}{\nu}}}\right)\left|\sum_{k+1}^{k+h} \boldsymbol{x}_i - h\boldsymbol{c}_1\right| + \sup_{Nh<k<\infty} \frac{1}{\sqrt{m}} \min\left(\frac{1}{\sqrt{h}}, \frac{1}{h^{\frac{1}{2}-\frac{1}{\nu}}k^{\frac{1}{\nu}}}\right)\left|\sum_{k+1}^{k+h} \boldsymbol{x}_i - h\boldsymbol{c}_1\right|$$
$$= o_P(1) + o_P(1) = o_P(1).$$

□

In summary, for the closed-end procedure condition (5.5) is fulfilled for ergodic regressors, and in case of an infinite time horizon we additionally need a strong invariance principle.

5.3. Mean change model

The observations have the form

$$X_t = \mu + \Delta \mathbb{1}_{\{t>k^*\}} + \epsilon_t, \quad 0 \le t < \infty,$$

where k^* is the change point, $\mu \in \mathbb{R}$ is the mean prior the change, $\Delta \neq 0$, and ϵ_t is the error sequence with mean zero and variance σ^2. The multivariate mean change model is also included in the general setup but for simplicity we concentrate on the univariate model.

If we choose $p = 1$ in the linear regression model, then $\beta_0 = \mu$ and $\beta_A = \mu + \Delta$. So the mean change model, also called location model, is included in the linear model.

The function G to determine an estimator for the unknown mean μ calculated by the historical data set simplifies to

$$G(X_t, \mu) = X_t - \mu.$$

Thus we obtain the arithmetic mean as estimator. Since the monitoring function H is based on the estimated residuals, H simplifies to

$$H(X_t, \widehat{\mu}) = X_t - \overline{X}_m, \quad \overline{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i.$$

$B(\theta_0) = 1$ because the function G and H have both dimension 1. Assumption 3.2 holds exactly in the mean change model as shown in the following calculation:

$$\sum_{i=m+1}^{m+k} H(X_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(X_j, \theta_0) - \frac{k}{m} B(\theta_0) \sum_{j=1}^{m} G(X_j, \theta_0)\right)$$
$$= \sum_{j=m+1}^{m+k} X_j - k\overline{X}_m - \left(\sum_{j=m+1}^{m+k} X_j - k\mu - \frac{k}{m}\sum_{j=1}^{m}(X_j - \mu)\right) = 0. \tag{5.6}$$

In Horváth et al. (2012) which considers the location model for the MOSUM statistic but only in the closed-end procedure, the lines in (5.6) are shown in (A.1). In the mentioned publication and in Aue et al. (2008), they use a general continuous weight function satisfying

$$\inf_{0\le t\le T} g(t) > 0,$$

and this condition matches with Assumption 3.4a). The conditions in Assumptions 3.4b) and c) are not necessary because there is only a focus on the closed-end procedure. So they used the weight functions

$$\rho_1(t) = \left(\max\left(1, t^{\frac{1}{3}}\right)\right)^{-1},$$
$$\rho_2(t) = \left(\max\left(1, \log(1+t)\right)\right)^{-\frac{1}{2}}$$

in their simulation study. These functions fulfill Assumption 3.4a) which controls the behaviour at zero and is necessary for the closed-end procedure. However, the conditions which have to hold for the open-end case are not satisfied. Precisely, ρ_1 satisfies Assumption 3.4c), but does not conform to Assumption 3.4b). The function ρ_2 does not conform to both assumptions.

We recall the fact that in the linear model a parameter change can be detected if $\boldsymbol{c}_1^T(\beta_A - \beta_0) \neq 0$. Hence in case of the location model the consistency can be shown as long as there is a mean change in the time series, so $\Delta \neq 0$.

5.4. Non-linear model

Aue et al. (2006a) suppose a GARCH-sequence and search for changes in its parameters with the CUSUM statistic. They use the log likelihood score function as estimating as well as monitoring function. To obtain the limit distributions for our statistics we need their Lemma 6.4, in which they establish Assumption 3.2 as well as Lemma 3.3, in which Assumption 3.3 is shown.

Ciupera (2013) considers a non-linear regression model

$$X_t = f(\boldsymbol{x}_t, \boldsymbol{\beta}_0) + \epsilon_t,$$

where the residuals are assumed to be i.i.d. and the function f is known.

To estimate the unknown parameter $\boldsymbol{\beta}_0$ based on the historical data set, they use the least squares estimator. Thus the estimating function is

$$G(X_t, \boldsymbol{\beta}) = \nabla f(\boldsymbol{x}_t, \boldsymbol{\beta})(X_t - f(\boldsymbol{x}_t, \boldsymbol{\beta})),$$

and the monitoring function is defined as the estimated errors

$$H(X_t, \boldsymbol{\beta}) = X_t - f(\boldsymbol{x}_t, \boldsymbol{\beta}).$$

Assumption 3.2 is shown in their Lemma A.1. Under the alternative Assumption, A.4c) in Kirch and Kamgaing (2015) is proven in their Theorem 3.2 with $E_H = E(f(\boldsymbol{x}_t, \boldsymbol{\beta}_0)) - E(f(\boldsymbol{x}_t, \boldsymbol{\beta}_1))$, where $\boldsymbol{\beta}_1$ is the parameter after the change has occurred. Since for our statistics Assumptions 4.1c), 4.2c) and 4.3c) follow from the Assumption A.4c) in Kirch and Kamgaing (2015), these assumptions are also fulfilled. Furthermore, Kirch and Kamgaing (2011) proposed a similar approach for non-linear autoregressive time series, but in a nonparametric setup, where they search for a change in the autoregression function g that determines the observation structure

$$X_t = g(X_{t-1}, \ldots, X_{t-p}) + \epsilon_t.$$

5.5. Binary model

Typically, a binary model is defined as

$$X_t|X_{t-1}, X_{t-2}, \dots, Z_{t-1}, Z_{t-2}, \dots \sim Bern(\pi_t(\boldsymbol{\beta})), \quad g(\pi_t(\boldsymbol{\beta})) = \boldsymbol{\beta}^T \boldsymbol{Z}_{t-1},$$

where $\boldsymbol{Z}_{t-1} = (Z_{t-1}, \dots, Z_{t-p})$ is a regressor.

There are a lot of examples for binary time series, i.e. Wilks and Wilby (1999) test whether it has been raining on a specified day. Kauppi and Saikkonen (2008) and Startz (2008) investigate whether a recession has occurred in a specific month. The canonical link function $g(x) = \log(x/(1-x))$ is usually used as function g, and the monitoring function is typically $G((X_t, \boldsymbol{Z}_{t-1}), \boldsymbol{\beta}) = \boldsymbol{Z}_{t-1}(X_t - \pi_t(\boldsymbol{\beta}))$, which comes from the partial likelihood scores.

Kirch and Kamgaing (2015) gives a simulation example as well as a real data example about US recession data for the monitoring scheme with the CUSUM statistic, where the monitoring function H is chosen equal to the estimating function G.

5.6. Poisson autoregressive model

The Poisson autoregressive model is defined as

$$X_t|X_{t-1}, \dots, X_{t-p} \sim Pois(\lambda_t), \quad \lambda_t = f_\theta(\boldsymbol{X}_{t-1}), \boldsymbol{X}_{t-1} = (X_{t-1}, \dots, X_{t-p})^T.$$

There exists a stationary and ergodic solution which is β-mixing with exponential rate (see Neumann (2011)) if $f_\theta(\boldsymbol{x})$ is Lipschitz-continuous in $\boldsymbol{x}$ for all $\theta \in \Theta$ with a Lipschitz constant smaller than 1. For the mentioned solution we obtain Assumption 3.3.

The monitoring function G is again obtained by the partial log likelihood scores and the monitoring function H is often chosen equal to G as in the simulation study for Poisson autoregressive time series in Kirch and Kamgaing (2015).

We refer to Kirch and Kamgaing (2015) for more details about these models.

6. Simulation Study

In this chapter we compare our monitoring schemes including the standard CUSUM statistic in terms of empirical size, power and run length. The run length is the time period from $m+1$ till the procedure stops, so till the stopping time point.

Particularly, we use the size-adjusted power and the normalized density estimation of the size-adjusted run-length for a better comparison of the monitoring schemes. More precisely, a higher empirical size leads to a better power and shorter run-length. Hence we can only compare the statistics in a meaningful way if we fix the empirical size before determining the power and run-length which are the so-called size-adjusted power and size-adjusted run-length. However if we compare the size-adjusted run length of the procedures with the help of the density estimation, we should also take into account the size-adjusted power. The reason is, if there are two procedures with similar run length but one of them has a better size-adjusted power, this advantage for one of the methods has to be visible in the density estimation for comparison. Thus we use the normalized density estimation of the run length where the area under the estimated density function is equal to the size-adjusted power of each procedure rather than to the standard density estimations where the area under the curve is always equal to 1.

We suppose the mean-change model $X_t = \mu + \Delta \mathbb{1}_{\{t>k^*\}} + \epsilon_t,\ t = 1, \cdots, N$, with $\mu = 0$ and $\Delta \neq 0$. If there is a mean change, we have $k^* < N$. Otherwise, $k^* = N$. So the observations first have mean zero and at time k^*+1 the mean possibly jumps up or down to the level Δ. The errors are supposed to be i.i.d. standard normally distributed. The estimating function is $G(X_t, \mu) = X_t - \mu$, and the monitoring function $H(X_t, \widehat{\mu}) = \widehat{\epsilon}_t = X_t - \overline{X}_m$.

The empirical results are based on a training period of length $m = 100$, a monitoring period of $N = 200$, and 2500 repetitions. The simulations of the limit processes for each statistic are based on 10000 repetitions.

We put $1/\sigma$ in front of the sum of the monitoring function and use the true variance of the errors $\sigma = 1$ for the simulations as well as we use for the CUSUM, Page-CUSUM and modified MOSUM statistic the most popular boundary function

$$w(m,k) = m^{-\frac{1}{2}} \left(1 + \frac{k}{m}\right)^{-1} \left(\frac{k}{k+m}\right)^{-\gamma}.$$

Then we get a pivotal limit, i.e. the limit process does not include any unknowns, as stated in the Corollaries 3.1 and 3.2. For the MOSUM we use the weight function

$$w_M(h,k) = \left(2 \max\left(1, \log\left(1 + \frac{k}{h}\right)\right)\right)^{-\frac{1}{2}}.$$

As window size width we choose 10 and 20. So the bandwidth is relatively small compared to $m = 100$. Consequently we simulate the pivotal limit in Corollary 3.5b) with a chosen rescaled time horizon $N'(h)$ depending on the used window size width h such that $N'(h)h = Nm = 200$.

In Table 6.1 the empirical size is stated for all statistics. We notice that the empirical size of the CUSUM, Page-CUSUM and modified MOSUM statistics increases in γ, and the modified MOSUM method additionally in the bandwidth parameter h. The increase in γ comes from its influence on the behaviour of the boundary function especially at the beginning of the monitoring time. More precisely, the higher γ the lower is the boundary function for small k (see Figure 6.1). Thus for γ near to $\frac{1}{2}$, the procedure stops more often already after a few observations than for a γ near to zero. The reason is that at the beginning of the monitoring, and consequently k is small, only a few observations are used for summation, which increases the risk of a false alarm. We see this effect in Figure 6.1. In the left plot the procedure does not stop under the null hypothesis but on the right side it does after a few observations. The empirical size of the MOSUM statistic also increases with the bandwidth h caused by the similar behaviour of the used weight function. It has lower values at the beginning of the monitoring time for greater window size and the other way around. However the nominal level is maintained for all parameter constellations except in

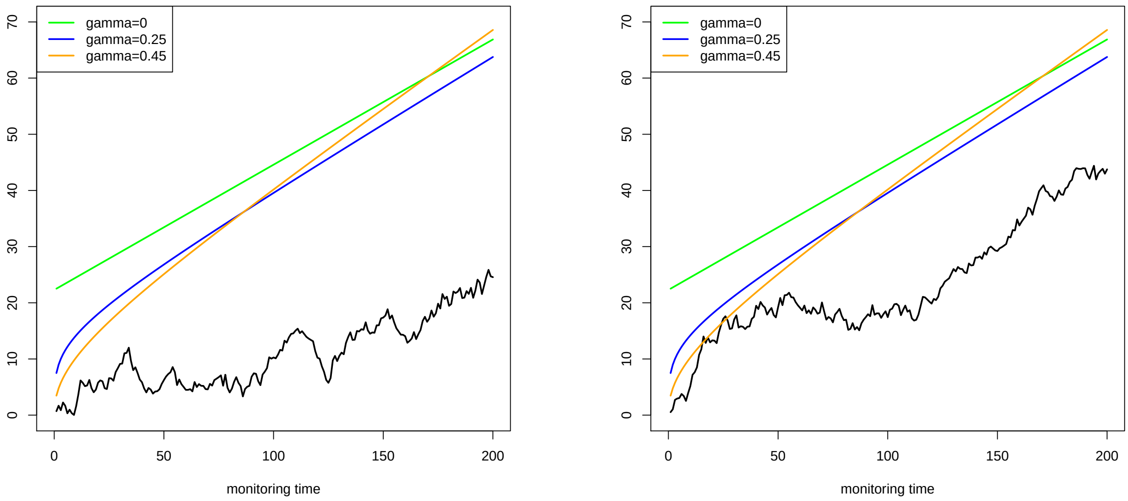


Figure 6.1.: CUSUM under the null hypothesis

case of the modified MOSUM procedure, where γ and h have simultaneously very high values, i.e γ near to $\frac{1}{2}$ and h near to 1. There the mentioned effect is huge as we see in the Table 6.1 by the red numbers. If h is near to 1 there are extremely few observations used to calculate the statistic, because the lower and upper bound of the corresponding sum are very close to each other, particularly if k is small, i.e at the beginning of the monitoring time.

To solve this problem of the modified MOSUM statistic we can wait for a_m observations with $a_m/m \to 0$, $m \to \infty$ and then start monitoring. The null asymptotics still remain the same showed in the proof of Theorem 3.1. In Table 6.2 we use $a_m = \sqrt{m}$ and see that the empirical size is lower or equal for all values of h as before. For $\gamma = 0$ the empirical sizes are unaltered so there is no consequence of waiting for a_m observations. For $\gamma = 0.25$ or 0.45 and $h = 0.1$ or 0.4 the empirical size is somewhat lower than before, however the procedures have already hold the nominal level before. But if $h = 0.9$ and $\gamma = 0.25$ the values are also only slightly lower but now the level is hold. The focus is on the empirical size if the values of h and γ are simultaneously very high, actually the empirical size is much lower than without waiting for a_m observations (see the green numbers in Table 6.2). So we conclude that the major part of the false alarms occurs at the beginning of the monitoring time.

By the comparison of the size-adjusted power in Table 6.3, first notice that the use of $\gamma = 0$ is best in all situations. Furthermore early changes are detected best by the modified MOSUM statistic with $h = 0.1$ and changes in the middle of the monitoring time with $h = 0.4$ and late

	γ	**0**		**0.25**		**0.45**	
	α	**5**	**10**	**5**	**10**	**5**	**10**
CUSUM							
		1.12	2.96	2.32	5.92	3.36	5.96
Page-CUSUM							
		0.96	2.76	1.8	5.32	2.16	3.84
mod. MOSUM							
$h = 0.1$:		0.88	3	1.92	4.48	3.04	6.88
$h = 0.4$:		1.76	3.68	2.88	6.08	4.08	8.84
$h = 0.9$:		4.4	7.76	8.76	14.24	**38.04**	**44.8**
MOSUM	h	**10**		**20**			
		4.24	8.4	5.04	10.56		

Table 6.1.: Empirical size (in %)

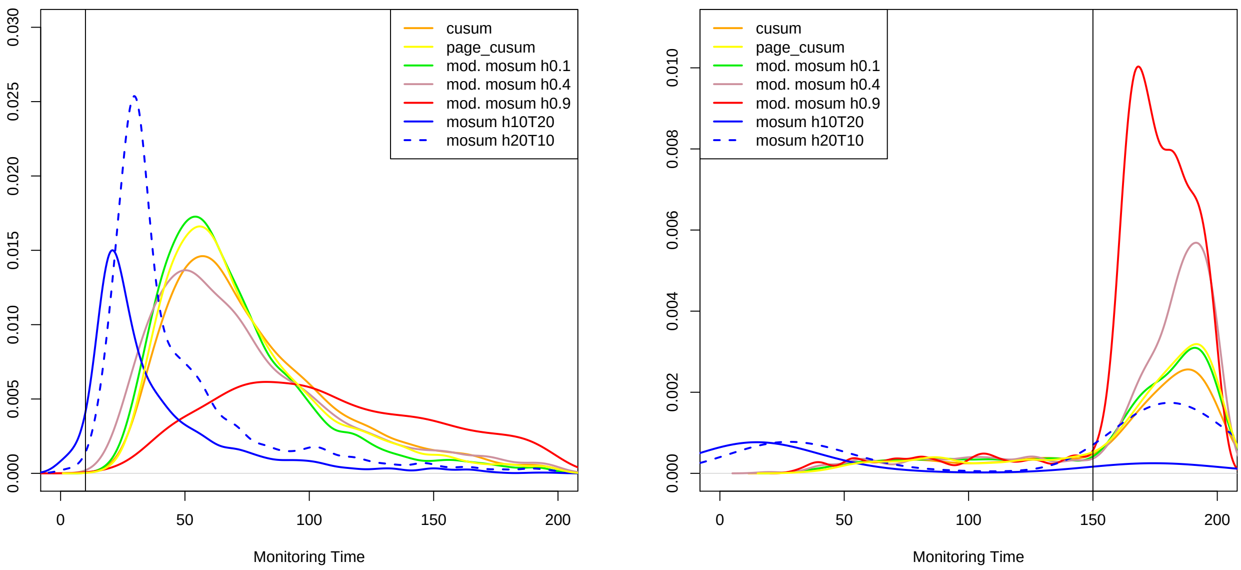
	γ	**0**		**0.25**		**0.45**	
	α	**5**	**10**	**5**	**10**	**5**	**10**
$h = 0.1$:		0.88	3	1.92	4.48	2.32	5.08
$h = 0.4$:		1.76	3.68	2.88	6.08	2.2	5.04
$h = 0.9$:		4.4	7.76	4.76	8.84	**3.6**	**6.36**

Table 6.2.: Empirical size modified MOSUM statistic with $a_m = \sqrt{m}$ (in %)

changes with $h = 0.9$ highlighted by the green numbers in the table. Further note that the very high empirical size for high values of γ and h ruins the power such that it is equal to the nominal level of 5 % (in Table 6.3 the red numbers). But if we start monitoring again after $a_m = \sqrt{m}$ observations, the power increases particularly for $h = 0.9$ (the bold numbers in Table 6.3).

Finally we compare our procedures concerning the size-adjusted run length using the normalized density estimation plots. The MOSUM procedure is very stable at the beginning of the monitoring time because the window width reaches into the training period and does not only use a few observations for the summation, in contrast to the other schemes. This is reflected in Figure 6.2(a) where the MOSUM statistic is the quickest procedure to detect an early change. For a late change the modified MOSUM method with h near to 1 is the quickest procedure (see Figure 6.2(b)). This is due to the fact that the modified MOSUM statistic with a high value for h and in case of a late change includes only some observations in the statistic but not all. With including all observations after the training period as in case of the CUSUM statistic, it would take a longer time to reach the boundary function. On the other, hand the modified MOSUM statistic does not use too few observations where the risk for a false alarm is high. Figure 6.3(a)-

Figure 6.2.: Normalized density estimation of the run length (change=0.5, $\gamma = 0$)



(a) $k^* = 10$ (b) $k^* = 150$

(b) shows the effect of too early detection in case of the modified MOSUM statistic with a high value of h and even already if $\gamma = 0.25$. Directly after starting monitoring there are peaks which means that the procedure gives false alarms and stops before the change has occurred.

However if we wait for a_m observations, the peaks disappear as we see in Figure 6.3(c)-(d).

If we choose $\gamma = 0.45$, the effect of too early detection would already appear for $h = 0.4$ for the modified MOSUM statistic as well as for the Page-CUSUM and CUSUM statistic to a less extent. The effect also slightly appears for the MOSUM statistic as we see in Figure 6.3(a)-(d). But in case of the Page-CUSUM and CUSUM procedures we could also wait for a_m observations and the limit processes would remain the same as noted in the proofs of Theorem 3.3 and Theorem 3.1 in Kirch and Kamgaing (2015). In case of the MOSUM procedure we could start monitoring after the duration of one bandwidth and use the corresponding limit process

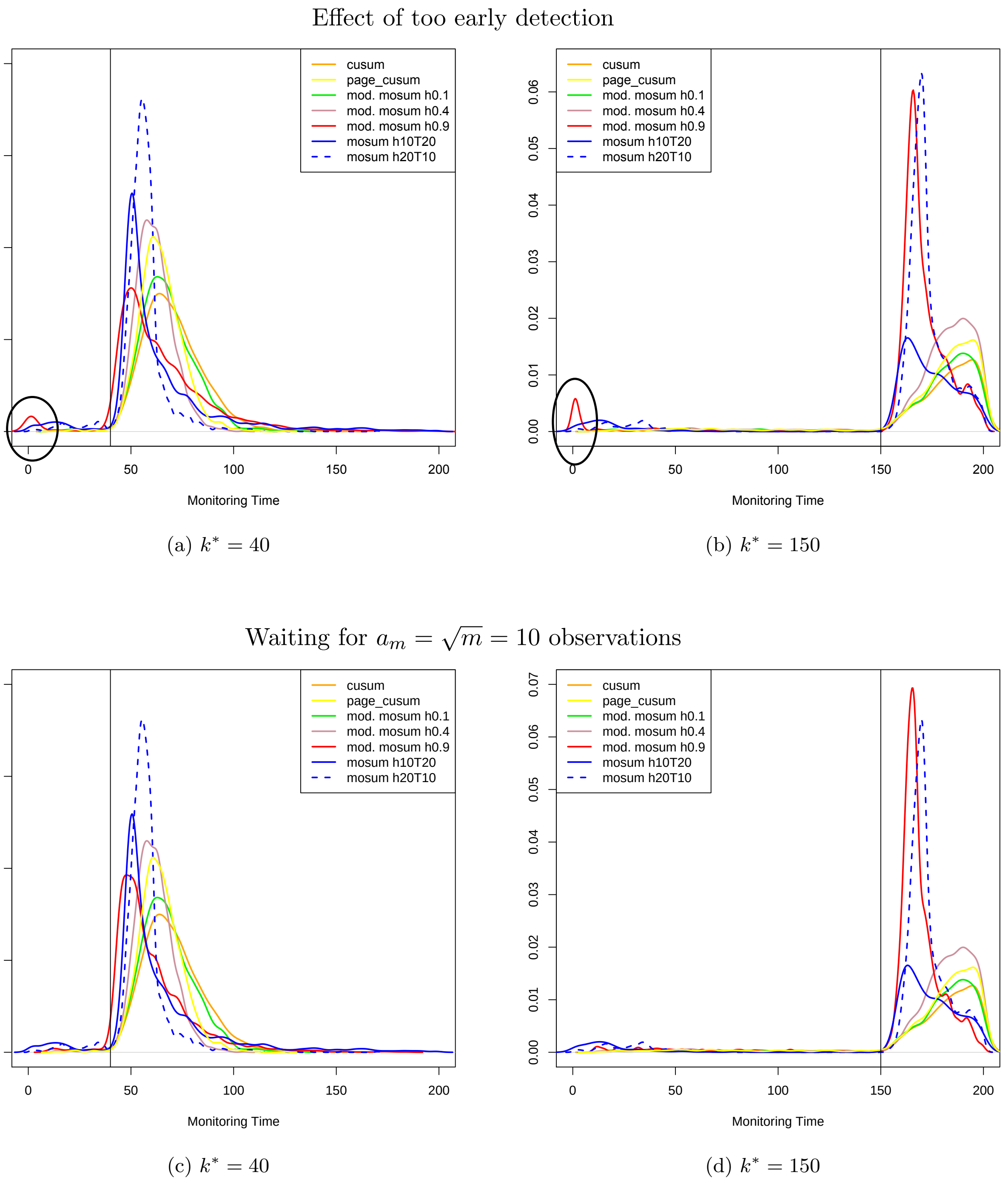
	γ	**0**	**0.25**	**0.45**
CUSUM	**k^*-1=0.05q**	95.84	95.24	92.4
	k^*-1=0.2q	86.36	84.48	76.24
	k^*-1=0.5q	45.32	42.76	31.16
	k^*-1=0.75q	13.24	11.76	8.68
Page-CUSUM	**k^*-1=0.05q**	97.08	96.64	94.8
	k^*-1=0.2q	92.16	90.76	85.36
	k^*-1=0.5q	55.36	50.32	39.76
	k^*-1=0.75q	14.96	12.52	9.72
mod. MOSUM				
$h = 0.1$:				
	k^*-1=0.05q	**97.32**	96.88	93.76
	k^*-1=0.2q	91.92	89.72	82.08
	k^*-1=0.5q	51.48	46.28	34.4
	k^*-1=0.75q	15.04	12.28	9.24
$h = 0.4$:				
	k^*-1=0.05q	96	95.28	87.64
	k^*-1=0.2q	**94.84**	93.16	82.36
	k^*-1=0.5q	**74**	67.08	46.68
	k^*-1=0.75q	21	16.28	9.12
$h = 0.9$:				
	k^*-1=0.05q	69.52	50.36	**5.04**
	k^*-1=0.2q	68.8	47.08	**5.04**
	k^*-1=0.5q	57.6	32.64	**5.04**
	k^*-1=0.75q	**37.36**	18.48	**5.04**
mod. MOSUM with				
$a_m = \sqrt{m} = 10$				
$h = 0.1$:				
	k^*-1=0.05q	**97.32**	96.88	95.16
	k^*-1=0.2q	91.92	89.72	84.2
	k^*-1=0.5q	51.48	46.24	37.08
	k^*-1=0.75q	15	12.28	9.52
$h = 0.4$:				
	k^*-1=0.05q	96	95.28	91.44
	k^*-1=0.2q	**94.84**	93.12	86.6
	k^*-1=0.5q	**74**	67.04	52.6
	k^*-1=0.75q	21	16.24	10.6
$h = 0.9$:				
	k^*-1=0.05q	69.52	62.68	**38.28**
	k^*-1=0.2q	68.8	59.2	**28.76**
	k^*-1=0.5q	57.56	42.48	**14.92**
	k^*-1=0.75q	**37.32**	24.4	**8.32**

MOSUM	h	**10**	**20**
	k^*-1=0.05q	45.92	75.68
	k^*-1=0.2q	24.12	56.96
	k^*-1=0.5q	10.4	30
	k^*-1=0.75q	6.68	15

Table 6.3.: Size-adjusted power (in %)

as stated in (3.29) or (3.30), where the supremum is only taken from the time points 1 to N.

Figure 6.3.: Normalized density estimation of the run length: Effect of too early detection (change=1, $\gamma = 0.25$)



(a) $k^* = 40$ (b) $k^* = 150$

(c) $k^* = 40$ (d) $k^* = 150$

Part II.

Multivariate Bootstrap Methods

7. Introduction and Motivation

We consider the offline change-point procedure where we observe a data set of length N and search for a change in the parameter of the assumed model.

The multivariate observations $\boldsymbol{X}(t) = (X_1(t), X_2(t), ..., X_d(t))^T, t = 1, ..., N$, have possibly a change in the mean and thus follow the model

$$X_i(t) = \mu_i + \Delta_i \mathbb{1}_{\{t>k^*\}} + \epsilon_i(t). \tag{7.1}$$

Here, $\boldsymbol{\mu} = (\mu_1, ..., \mu_d)^T$ is the mean vector (before the change), $\boldsymbol{\Delta} = (\Delta_1, ..., \Delta_d)^T$ the change vector, and k^* the change point. Further, let $\boldsymbol{\epsilon}(t) = (\epsilon_1, ..., \epsilon_d)^T$ be a centered stationary time series with existing second moments. Additionally, we assume the validity of the functional central limit theorem:

$$\left\{ \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Nt \rfloor} \boldsymbol{\epsilon}(j) : 0 \leq t \leq 1 \right\} \overset{D[0,1]}{\longrightarrow} \{\boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{W}(t)\}, \tag{7.2}$$

where $\{\boldsymbol{W}(t) : t \geq 0\}$ is a standard multivariate Wiener process, and $\boldsymbol{\Sigma}$ is the long-run covariance matrix of the multivariate error sequence $\boldsymbol{\epsilon}(t)$, which is defined as

$$\boldsymbol{\Sigma}_{i,j} = \sum_{h \in \mathbb{Z}} \operatorname{Cov}(\epsilon_i(0), \epsilon_j(h)), \quad i, j = 1, \cdots, d.$$

To test the null hypothesis H_0 of no change in the mean μ against the alternative H_1 that there exists a mean change in the time series at a time point k^*, we use the offline CUSUM statistic

$$S_N := \max_{1 \leq t \leq N} \frac{1}{N} \left(\sum_{j=1}^{t} (\boldsymbol{X}(j) - \overline{\boldsymbol{X}}_N) \right)^T \boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^{t} (\boldsymbol{X}(j) - \overline{\boldsymbol{X}}_N) \right).$$

To compute the statistic S_N, the long-run covariance matrix $\boldsymbol{\Sigma}$ must be estimated. A suitable estimator of $\boldsymbol{\Sigma}$ is the flat-top-kernel estimator introduced in Politis (2011). However, if the data are strongly dependent or the dimension d is large, the estimator is not suitable, especially because we need its inverse. The problem is that the estimation error of $\boldsymbol{\Sigma}$ increases with the dimension d and if the inverse is needed, the numerical error of the calculation of the inverse comes on top. Even if the dimension is too high, the inverse can not be calculated. Alternatively, there is a statistic which does not take all dependencies into account. It only respects the dependence in time, but not between the components, by merely using the diagonal elements of $\boldsymbol{\Sigma}$. Consequently, the statistic is defined as

$$T_N := \max_{1 \leq t \leq N} \frac{1}{N} \left(\sum_{j=1}^{t} (\boldsymbol{X}(j) - \overline{\boldsymbol{X}}_N) \right)^T \boldsymbol{\Lambda}^{-1} \left(\sum_{j=1}^{t} (\boldsymbol{X}(j) - \overline{\boldsymbol{X}}_N) \right),$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\Sigma}_{1,1}, \cdots, \boldsymbol{\Sigma}_{d,d})$.

If we want to test for a change in a time series with a relatively low dimension d such as $d \leq 4$, the estimator of the long-run covariance matrix $\boldsymbol{\Sigma}$ performs still well enough, and so we can apply the statistic S_N for testing. We can either do bootstrapping, which often provides better approximations for the critical values if the sample size is small, or we use an asymptotic test where the limit process of the statistic is required for approximating the critical values. If we observe a time series of higher order, the estimator of $\boldsymbol{\Sigma}$ is not stable, thus we have to use the statistic T_N, where we only take some dependencies into account.

In the next Theorem we develop the limit distributions for both statistics under H_0.

Theorem 7.1

Let the observations $\boldsymbol{X}(t) = (X_1(t), X_2(t), ..., X_d(t))^T, t = 1, ..., N$, follow the model in (7.1) and the error sequence fulfills (7.2). Then under H_0, we have

a)

$$S_N \xrightarrow{D} \sup_{0 \leq t \leq 1} \boldsymbol{B}^T(t)\boldsymbol{B}(t) = \sup_{0 \leq t \leq 1} \sum_{i=1}^{d} B_i^2(t) \quad \text{as } N \to \infty,$$

where $\{\boldsymbol{B}(t)\}$ is a d-dimensional standard Brownian bridge and B_i, $i = 1, \cdots, d$, are independent univariate Brownian bridges.

b)

$$T_N \xrightarrow{D} \sup_{0 \leq t \leq 1} \boldsymbol{B}^T(t)\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{B}(t) = \sup_{0 \leq t \leq 1} \sum_{i=1}^{d} \widetilde{B}_i^2(t) \quad \text{as } N \to \infty.$$

where $\{\boldsymbol{B}(t)\}$ is a d-dimensional standard Brownian bridge and $\widetilde{B}_i(t)$, $i = 1, \cdots, d$, are univariate Brownian bridges with covariance matrix $\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-\frac{1}{2}}$.

The assertions continue to hold true if we replace $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ by consistent estimators.

Proof. Under H_0 we have

$$\sum_{j=1}^{t}(\boldsymbol{X}(j) - \overline{\boldsymbol{X}}_N) = \sum_{j=1}^{t}(\boldsymbol{\epsilon}(j) - \overline{\boldsymbol{\epsilon}}_N), \quad \text{where } \overline{\boldsymbol{\epsilon}}_N = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i.$$

a) Let $\widehat{\boldsymbol{\Sigma}}$ be a consistent estimator of $\boldsymbol{\Sigma}$. Then from (7.2) it follows that

$$\begin{aligned}
&\max_{1 \leq t \leq N} \frac{1}{N}\left(\sum_{j=1}^{t}(\boldsymbol{X}(j) - \overline{\boldsymbol{X}}_N)\right)^T \widehat{\boldsymbol{\Sigma}}^{-1}\left(\sum_{j=1}^{t}(\boldsymbol{X}(j) - \overline{\boldsymbol{X}}_N)\right) \\
&= \max_{1 \leq t \leq N} \frac{1}{N}\left(\sum_{j=1}^{t}(\boldsymbol{\epsilon}(j) - \overline{\boldsymbol{\epsilon}}_N)\right)^T \widehat{\boldsymbol{\Sigma}}^{-1}\left(\sum_{j=1}^{t}(\boldsymbol{\epsilon}(j) - \overline{\boldsymbol{\epsilon}}_N)\right) \\
&= \max_{\frac{1}{N} \leq s \leq 1} \frac{1}{N}\left(\sum_{j=1}^{\lfloor Ns \rfloor}(\boldsymbol{\epsilon}(j) - \overline{\boldsymbol{\epsilon}}_N)\right)^T \widehat{\boldsymbol{\Sigma}}^{-1}\left(\sum_{j=1}^{\lfloor Ns \rfloor}(\boldsymbol{\epsilon}(j) - \overline{\boldsymbol{\epsilon}}_N)\right)
\end{aligned}$$

$$\xrightarrow{D} \sup_{0<s\leq 1} \left(\boldsymbol{\Sigma}^{\frac{1}{2}} \left(W(s) - sW(1)\right)\right)^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}} \left(W(s) - sW(1)\right)$$

$$= \sup_{0<s\leq 1} \boldsymbol{B}^T(s)\boldsymbol{B}(s) = \sup_{0\leq s\leq 1} \sum_{i=1}^{d} B_i^2(s).$$

b) The proof is analogous to the proof of a).

□

By the previous theorem we conclude that the limit process of S_N is distribution-free and can be used to compute asymptotic critical values. However, the limit distribution of T_N does not have this property, hence it is not suitable for asymptotic testing. For T_N we definitely need bootstrap methods. More precisely, for data in higher dimensions we cannot use S_N and consequently no asymptotic test. Instead, we have to use T_N with bootstrapping. In this part of the work we focus on higher-dimensional data, thus we use the statistic T_N with bootstrapping.

The bootstrap methods we will use are based on the estimated errors. Notice that we have collected data following the mean change model specified in (7.1) and first have to estimate the unknown errors. Subsequently, the estimated errors can be bootstrapped.

In general, bootstrapping means that many new sequences are generated from a given sequence by sampling with replacement in some way. Bootstrap methods in the change-point analysis are methods where the critical values are not approximated by quantiles of the asymptotic distribution under the null hypothesis, but calculated by bootstrapping the series with the hope of a better approximation of the true distribution of the test statistic. Hence, the critical values are approximated by the quantiles of the bootstrapped sequences. Thus the bootstrap method is valid if the probability function of the statistic based on the bootstrapped sequence conditioned by the real data set still converges in probability or even almost surely to the limit process under H_0 of the statistic based on the original data set, no matter whether the null hypothesis or the alternative holds, while the sample size N grows to infinity.

The first bootstrap considered is the Efron-Bootstrap introduced in Efron (1979). The Efron-Bootstrap is suitable for independent observations, due to the sampling of each single data point with replacement. Every dependency structure in the data would be destroyed by this way of bootstrapping. Then the Efron-Boostrap is advanced for dependent data, where no longer every single data point is sampled with replacement, but the sequence is splitted into blocks which are sampled with replacement. The idea is to preserve the existing dependency structure in the data set within the blocks. This bootstrap method is called the block-bootstrap first suggested in Hall (1985) and Carlstein (1986).

Later, several modifications of the block-bootstrap have been developed. First, there is the so-called moving-block-bootstrap (Künsch (1989) and Liu and Singh (1992)), there the blocks are allowed to be overlapping. To avoid the problem of underrepresented data points at the beginning and the end of the time series, the circular overlapping block-bootstrap was considered by Politis et al. (1992), where the data set is periodically extended and the bootstrapped data points are centered around the sample mean.

Furthermore there are the dependent wild-bootstrap and the AR-sieve-bootstrap developed for dependent series. The first one, proposed by Shao (2010), does not use blocks for bootstrapping, but instead a suitable kernel function with a bandwidth l. The AR-sieve-bootstrap is a parametric bootstrap based on an AR-assumption on the error sequence, which was first introduced by Kreiss (1988). However, these two bootstrap methods have not yet been transformed into the change-point context.

In the next chapter we will introduce the mentioned bootstrap methods in more detail and then focus on the circular overlapping block-bootstrap, where we develop its validity. To show the validity of a bootstrap method, first it has to be shown, that the limit process of the statistic based on scores fulfilling some moment conditions, is the same than that of the original statistic under H_0. Then the scores can be replaced by the estimated residuals, which satisfy the requirements of the scores P-stochastically or almost surely.

In contrast to the other methods, the circular overlapping block-bootstrap has already been analysed in the change-point context by Kirch (2006), but only in the univariate mean change model and with the help of an invariance principle satisfied by the scores as stated in Corollary $D.2$ in the Appendix of Kirch (2006).

However, we will derive functional central limit theorems for scores in Chapters 9 and 10, which are sufficient to obtain the correct limit behaviour of the bootstrapped series given a data set following either H_0 or H_1. First we develop such a theorem in the simplest case, namely the univariate Efron-Bootstrap, and extend it to the univariate circular overlapping block-bootstrap in order to obtain the limit theorems in the multivariate Efron-case and finally for the multivariate circular overlapping block-bootstrap.

In Chapter 11 we compare the mentioned bootstrap methods in a detailed simulation study.

8. Bootstrap Methods

As explained in the previous chapter we definitely require bootstrap methods in combination with the statistic T_N, which we have to use for higher-dimensional data. The investigated bootstrap methods are the circular overlapping block-bootstrap, the dependent wild-bootstrap and the AR-sieve-bootstrap. In this chapter we will introduce these methods in detail, where all bootstrap methods are based on the estimated residuals. They are defined as follows:

$$\widehat{e}_i(t) = X_i(t) - \widehat{\mu}_i - \widehat{\Delta}_i \mathbb{1}_{(t > \widehat{k}_i^*)}, \quad i = 1, \ldots, d; \quad t = 1, \ldots, N;$$

where

$$\widehat{k}_i^* = \arg\max \left\{ 1 \leq t \leq N : \sum_{j=1}^{t} \left(X_i(j) - \overline{X}_{i,N} \right) \right\}, \quad \overline{X}_{i,N} = \sum_{j=1}^{N} X_i(j),$$

$$\widehat{\mu}_i = \frac{1}{\widehat{k}_i^*} \sum_{t=1}^{\widehat{k}_i^*} X_i(t), \quad \widehat{\Delta}_i = \frac{1}{N - \widehat{k}_i^*} \sum_{t=\widehat{k}_i^*+1}^{N} (X_i(t) - \widehat{\mu}_i).$$

8.1. Circular Overlapping Block-Bootstrap

Before using the circular overlapping block-bootstrap we first have to prepare the time series by extending it periodically. Afterwards, the blocks are built and sampled with replacement in the way explained in the following lines.

The block length is defined as K and the number of blocks as L. We construct a bootstrap sequence $\boldsymbol{X}_b^*(1), \ldots, \boldsymbol{X}_b^*(N)$ with $\boldsymbol{X}_b^*(t) = (X_{b,1}^*(t), \ldots, X_{b,d}^*(t))$ from the observations $\boldsymbol{X}(t) = (X_1(t), \ldots, X_d(t))$, $t = 1, \ldots, N$, by putting

$$X_{b,i}^*(Kl + j) = \hat{e}_i(U_N(l) + j), \quad l = 0, \ldots, L = \left\lfloor \frac{N}{K} \right\rfloor, \quad j = 1, \ldots, K.$$

Here, $\hat{e}_i(t), t = 1, \ldots, N$, are the estimated residuals, and $\{U_N(l)\}, l = 0, \ldots, L$, are i.i.d. uniformly distributed in $0, \ldots, N-1$ and independent of $\{\boldsymbol{X}(t)\}$. Consequently, the starting points for the blocks are $0, \ldots, N-1$. Thus we need the above-mentioned periodic extension of the time series $X(t)$ from which we estimate the error sequence according as

$$\hat{e}_i(t) = \hat{e}_i(t - N), \quad t > N.$$

The statistic based on the bootstrapped sequence is:

$$T_N^* = \max_{1 \leq t \leq N} \frac{1}{N} \left(\sum_{j=1}^{t} (\boldsymbol{X}_b^*(j) - \overline{\boldsymbol{X}^*}_{b,N}) \right)^T \widehat{\boldsymbol{\Lambda}^*}^{-1} \left(\sum_{j=1}^{t} (\boldsymbol{X}_b^*(j) - \overline{\boldsymbol{X}^*}_{b,N}) \right),$$

where $\widehat{\boldsymbol{\Lambda}^*}$ is a diagonal matrix with

$$\widehat{\boldsymbol{\Lambda}^*}_{i,i} = \frac{1}{N} \sum_{l=0}^{L} \left(\sum_{k=1}^{K} (X_{b,i}^*(Kl + k) - \overline{X^*}_{b,i,N}) \right)^2, \quad i = 1, \ldots, d. \tag{8.1}$$

8.2. Dependent Wild-Bootstrap

The dependent wild-bootstrap does not use blocks. Instead, it produces the bootstrapped series by using a kernel function. The sequence $\boldsymbol{X}_w^*(1), \ldots, \boldsymbol{X}_w^*(N)$ is generated as follows:

$$\boldsymbol{X}_w^*(t) = (X_{w,1}^*(t), \ldots, X_{w,d}^*(t)), \quad \text{with} \quad X_{w,1}^*(t) = \hat{e}_i(t)Z(t),$$

where $Z(t)$, $t = 1, \ldots, N$, are centered, have variance 1 and are independent of $\boldsymbol{X}(t)$, $t = 1, \ldots, N$. The covariance of $Z(t)$ is $k((t-s)/l), 0 \leq t \leq N, 0 \leq s \leq N$ with a suitable kernel and bandwidth l. Later we will use the Bartlett kernel for simulations

$$k(x) = \begin{cases} 1 - |x|, & |x| < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The statistic based on the bootstrapped sequence with the dependent wild-bootstrap is calculated equivalently on the block-bootstrap according to

$$T_N^* = \max_{1 \leq t \leq N} \frac{1}{N} \left(\sum_{j=1}^{t} (\boldsymbol{X}_w^*(j) - \overline{\boldsymbol{X}^*}_{w,N}) \right)^T \widehat{\boldsymbol{\Lambda}^*}^{-1} \left(\sum_{j=1}^{t} (\boldsymbol{X}_w^*(j) - \overline{\boldsymbol{X}^*}_{w,N}) \right).$$

Here, $\widehat{\boldsymbol{\Lambda}^*}$ is the flat-top kernel estimator as proposed in Politis (2011), only with diagonal elements. In the simulation study we will explain the flat-top kernel estimator in detail.

8.3. Vector-AR-Sieve-Bootstrap

The AR-sieve-bootstrap is a parametric bootstrap with an $AR(\infty)$-assumption on the multivariate errors. The order p of the fitted AR-model of the errors depends on the sample size N and it is assumed that $p(N) \to \infty$ as $N \to \infty$. The bootstrap method can be described as follows:

First calculate the multivariate Yule-Walker estimators $\widehat{\boldsymbol{A}}_1, \ldots, \widehat{\boldsymbol{A}}_p$ for a given order p of the AR(p)-errors by the equation

$$\left(\widehat{\boldsymbol{A}}_1, \ldots, \widehat{\boldsymbol{A}}_p\right) \left(\widehat{\boldsymbol{\Gamma}}_{r,s=1,\ldots,p}\right) = \left(\widehat{\boldsymbol{\Gamma}}(1), \ldots, \widehat{\boldsymbol{\Gamma}}(p)\right),$$

where $\widehat{\boldsymbol{\Gamma}}(h)$ is the empirical autocovariance matrix of the estimated errors $\widehat{\boldsymbol{e}}(t)$, $t = 1, \cdots, N$.

According to the assumption on the errors, they are an AR(p)-sequence. Hence they have the form $\boldsymbol{e}(t) = \boldsymbol{\epsilon(t)} + \sum_{i=1}^{p} \boldsymbol{A}_j \boldsymbol{e}(t-i)$, where $\epsilon(t)$ is the noise term and $\boldsymbol{A}_j$ is the parameter matrix. To calculate the bootstrap sequence we first need the estimated noise term of the AR(p)-sequence

$$\boldsymbol{r}(t) = \widehat{\boldsymbol{e}}(t) - \sum_{j=1}^{p} \widehat{\boldsymbol{A}}_j \widehat{\boldsymbol{e}}(i-j)$$

which we have to center up according to

$$\widetilde{\boldsymbol{r}}(t) = \boldsymbol{r}(t) - \overline{\boldsymbol{r}}_N.$$

Then the bootstrap sequence is constructed according to

$$\boldsymbol{X}_s^*(t) = \sum_{j=1}^{p} \widehat{\boldsymbol{A}}_j \boldsymbol{X}^*(i-j) + \boldsymbol{r}^*(t), \quad \boldsymbol{r}^*(t) = \widetilde{\boldsymbol{r}}(U_l),$$

where U_l is independent uniformly distributed in $1, \ldots, N$. Then we obtain the bootstrapped sequence $\boldsymbol{X}_s^*(1), \ldots, \boldsymbol{X}_s^*(N)$. The statistic is calculated like in the dependent wild-bootstrap case, i.e.,

$$T_N^* = \max_{1 \leq t \leq N} \frac{1}{N} \left(\sum_{j=1}^{t} (\boldsymbol{X}_s^*(j) - \overline{\boldsymbol{X}^*}_{s,N}) \right)^T \widehat{\boldsymbol{\Lambda}^*}^{-1} \left(\sum_{j=1}^{t} (\boldsymbol{X}_s^*(j) - \overline{\boldsymbol{X}^*}_{s,N}) \right),$$

where $\widehat{\boldsymbol{\Lambda}^*}$ is the flat-top kernel estimator.

9. Univariate Bootstrap Methods

In this chapter we prove the validity of the univariate Efron bootstrap and the circular overlapping block-bootstrap.

To investigate the validity of a bootstrap method, we have to ensure that the critical values calculated by the bootstrapped sequences are good approximations for the critical values based on the null asymptotics, which means that the bootstrap test asymptotically hold a given level.

Precisely, we must show that, given the observed data set no matter whether it follows H_0 or H_1, the statistic based on the bootstrapped series has the same limit behaviour as the original statistic under H_0.

As seen in Theorem 7.1 in the previous chapter, if a functional central limit theorem holds for the bootstrapped series under H_0 as well as under H_1, then we easily obtain the same limit process.

The way to get the validity of the bootstrap methods is to show a functional central limit theorem for scores sampled with replacement in the desired way and afterwards replace them by the estimated residuals of the actually observed data, if they satisfy the required conditions of the scores P-stochastically under H_0 and H_1.

To prove a functional central limit theorem, we have to prove the convergence of the finite-dimensional distributions and the tightness of the corresponding partial sum process.

To finally obtain the asymptotics for the multivariate bootstrap methods, we first deal with the univariate methods.

9.1. Univariate Efron-Bootstrap

We first focus on the Efron-bootstrap where the bootstrapped sequences are created by sampling the observed data points of a time series with replacement.

A univariate data set $X(t),\ t = 1, \ldots, N$ is given. The Efron-bootstrap sequence is defined as

$$X^*(t) = \widehat{e}(U_N(t)),\ t = 1, \ldots, N,$$

where $\hat{e}(t), t = 1, \ldots, N$, are the estimated residuals$\{U_N(t)\}, t = 1, \ldots, N$, are i.i.d. uniformly distributed in $1, \ldots, N$ and independent of $\{X(t)\}$.

Theorem 9.1

Functional Central Limit Theorem (FCLT) for univariate Scores by using the Efron-Bootstrap:
Let $a_N(i), i = 1, ..., N$, be scores and $U_N(i), i = 1, ..., N$, independent uniformly distributed random variables in $\{1, 2, ..., N\}$ and independent of $\{X(\cdot)\}$. Suppose there are a $\delta > 0$ and

constants $D_1, D_2 > 0$ such that

$$\frac{1}{N}\sum_{i=1}^{N}|a_N(i)-\overline{a}_N|^{2+\delta} \leq D_1 \tag{9.1}$$

and

$$\sigma_N^2(a) = \text{Var}(a_N(U_N(1))) = \frac{1}{N}\sum_{i=1}^{N}|a_N(i)-\overline{a}_N|^2 \geq D_2. \tag{9.2}$$

Then, as $N \to \infty$

$$\left\{\frac{1}{\sqrt{N}\sigma_N(a)}\sum_{i=1}^{\lfloor Nt\rfloor}(a_N(U_N(i))-\overline{a}_N) : 0 \leq t \leq 1\right\} \overset{D[0,1]}{\longrightarrow} \{W(t) : 0 \leq t \leq 1\},$$

where $\{W(t)\}$ is a standard Wiener process.

Proof. According to Billingsley (1968) Theorem 15.1, it is sufficient to show the tightness and the convergence of the finite-dimensional distributions to get the FCLT.

First of all we have $E(a_N(U_N(i))) = \overline{a}_N$.

1. The convergence of the finite-dimensional distributions:
   For a single time point $0 \leq s \leq 1$ we have

$$\frac{1}{\sqrt{N}\sigma_N(a)}\sum_{i=1}^{\lfloor Ns\rfloor}(a_N(U_N(i))-\overline{a}_N) \overset{D}{\longrightarrow} W(s),$$

   because the Ljapunov-Condition holds due to the conditions (9.1) and (9.2):

$$\frac{1}{\left(N\sigma_N^2(a)\right)^{1+\frac{\delta}{2}}}\sum_{i=1}^{\lfloor Ns\rfloor}E\left(|a_N(U_N(i))-E(a_N(U_N(i)))|^{2+\delta}\right)$$
$$= \frac{\lfloor Ns\rfloor}{N}\frac{\sum_{i=1}^{N}|a_N(i)-\overline{a}_N|^{2+\delta}}{\left(\sum_{i=1}^{N}|a_N(i)-\overline{a}_N|^2\right)^{\frac{2+\delta}{2}}} = \frac{\lfloor Ns\rfloor}{N}\frac{N}{N^{1+\frac{\delta}{2}}}\frac{\frac{1}{N}\sum_{i=1}^{N}|a_N(i)-\overline{a}_N|^{2+\delta}}{\left(\frac{1}{N}\sum_{i=1}^{N}|a_N(i)-\overline{a}_N|^2\right)^{\frac{2+\delta}{2}}} = o(1).$$

   By the central limit theorem we then get

$$\frac{1}{\sqrt{N}\sigma_N(a)}\sum_{i=1}^{\lfloor Ns\rfloor}(a_N(U_N(i))-\overline{a}_N)$$
$$= \frac{1}{\sigma_N(a)}\sqrt{\frac{\lfloor Ns\rfloor}{N}}\frac{1}{\sqrt{\lfloor Ns\rfloor}}\sum_{i=1}^{\lfloor Ns\rfloor}(a_N(U_N(i))-\overline{a}_N) \overset{D}{\longrightarrow} W(s).$$

   For two time points $0 \leq s < t \leq 1$:

   First we define $X_N(s) := \frac{1}{\sqrt{N}\sigma_N(a)}\sum_{i=1}^{\lfloor Ns\rfloor}(a_N(U_N(i))-\overline{a}_N)$.

With the independence of $X_N(s)$ and $X_N(t)-X_N(s)$ and by the central limit theorem

$$c_1X_N(s)+c_2\left(X_N(t)-X_N(s)\right)\stackrel{D}{\longrightarrow} c_1W(s)+c_2(W(t)-W(s)),\quad c_1,c_2\in\mathbb{R}.$$

By the Cramér-Wold device, it thus follows that

$$\left(X_N(s),X_N(t)-X_N(s)\right)\stackrel{D}{\longrightarrow}\left(W(s),W(t)-W(s)\right).$$

By the continuous mapping theorem we obtain the convergence of the 2-dimensional distributions

$$\left(X_N(s),X_N(t)\right))\stackrel{D}{\longrightarrow}\left(W(s),W(t)\right).$$

The proof is analogous with more than two time points.

2. Tightness:
We define again
$S_k:=\sum\limits_{i=1}^{k}\left(a_N(U_N(i))-\overline{a}_N\right),\ X_N(s):=\frac{1}{\sigma_N(a)\sqrt{N}}S_{\lfloor Ns\rfloor},$
From condition (9.1) we get

$$E\left|a_N(U_N(i))-\overline{a}_N\right|^{2+\delta}=\frac{1}{N}\sum_{i=1}^{N}\left|a_N(i)-\overline{a}_N\right|^{2+\delta}\le D_1$$

and, using Theorem B.2 for the $2+\delta$-moment of S_N, we conclude

$$E|S_N|^{2+\delta}\le DN^{\frac{2+\delta}{2}}.$$

To show tightness we use Theorem B.1. But we first apply Lemma B.1 and the Markov inequality. For each $\epsilon>0$ there is a $\lambda>1$ such that

$$\begin{aligned}
&P\left(\max_{i\le N}|S_{k+i}-S_k|\ge\lambda\sigma_N(a)\sqrt{N}\right)=P\left(\max_{i\le N}|S_i|\ge\lambda\sigma_N(a)\sqrt{N}\right)\\
&=2P\left(|S_N|\ge(\lambda-\sqrt{2})\sigma_N(a)\sqrt{N}\right)\le 2P\left(|S_N|\ge\frac{1}{2}\lambda\sigma_N(a)\sqrt{N}\right)\\
&\le 2\frac{E|S_N|^{2+\delta}}{\left(\frac{1}{2}\sigma_N(a)\lambda\sqrt{N}\right)^{2+\delta}}=2\left(\frac{1}{\frac{1}{2}\sigma_N(a)\lambda}\right)^{2+\delta}\frac{E|S_N|^{2+\delta}}{N^{\frac{2+\delta}{2}}}\\
&\le\frac{\epsilon}{\lambda^2}.
\end{aligned}$$

The assertion follows from Theorem B.1.

□

Corollary 9.1

Let the assumption of Theorem 9.1 hold. Then the univariate score statistic has the same limit process as the original statistic under the null hypothesis, i.e., we have

$$\max_{1\le t\le N}\frac{1}{\sigma_N(a)\sqrt{N}}\left|\sum_{i=1}^{t}\left(a_N(U_N(i))-\overline{a}_{N,U}\right)\right|\stackrel{D}{\longrightarrow}\sup_{0<s\le1}|B(s)|,$$

where $B(s)$ is a standard Brownian bridge.

Proof. By Theorem 9.1 we conclude that the score statistic has the same limit process as the original statistic under the null hypothesis, because of

$$\begin{aligned}
&\max_{1\le t\le N}\frac{1}{\sigma_N(a)\sqrt{N}}\left|\sum_{i=1}^{t}\left(a_N(U_N(i))-\overline{a}_{N,U}\right)\right| \\
&=\max_{1\le t\le N}\frac{1}{\sigma_N(a)\sqrt{N}}\left|\sum_{i=1}^{t}\left(a_N(U_N(i))-\overline{a}_N-(\overline{a}_{N,U}-\overline{a}_N)\right)\right| \\
&=\max_{1\le t\le N}\frac{1}{\sigma_N(a)\sqrt{N}}\left|\sum_{i=1}^{t}\left(a_N(U_N(i))-\overline{a}_N\right)-t\left(\frac{1}{N}\sum_{i=1}^{N}\left(a_N(U_n(i))-\overline{a}_N\right)\right)\right| \\
&=\max_{\frac{1}{N}\le s\le 1}\frac{1}{\sigma_N(a)\sqrt{N}}\left|\sum_{i=1}^{\lfloor Ns\rfloor}\left(a_N(U_N(i))-\overline{a}_N\right)-\frac{\lfloor Ns\rfloor}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(a_N(U_N(i))-\overline{a}_N\right)\right)\right| \\
&\xrightarrow{D}\sup_{0<s\le 1}|W(s)-sW(1)|=\sup_{0<s\le 1}|B(s)| \qquad (9.3)
\end{aligned}$$

□

Furthermore, if we can replace the scores by the estimated residuals, we obtain the bootstrapped series. If the estimated residuals satisfy the required conditions for the scores P-stochastically, then the validity of the bootstrap method is demonstrated. The reason is that the critical values, obtained by the bootstrapping and conditioned on the observed time series, would be asymptotically correct no matter whether H_0 or H_1 holds. Thus now the question is, which conditions the estimated residuals replacing the scores and consequently the time series has to fulfill. The answer is stated in the next theorem.

Theorem 9.2

Let $e(t), t=1,...,N$, be centered random variables with existing second moments and $E(e(i))^{2+\delta}<\infty$ for some $\delta>0$. Additionally, let $e(t)$ obey a law of large numbers, and under the alternative, let the change Δ be fixed.

Then, under the null hypothesis as well as under the alternative of $k^*=\lfloor\nu N\rfloor,\ 0<\nu<1$, we have

$$P\left(\max_{1\le t\le N}\frac{1}{\sigma_N\sqrt{N}}\left|\sum_{i=1}^{t}\left(X^*(i)-\overline{X^*}_N\right)\right|\le x|X_1,...,X_N\right)\xrightarrow{P}P\left(\sup_{0<s\le 1}|B(s)|\le x\right),$$

where $B(s)$ is a standard Brownian bridge and

$$\sigma_N^2=\frac{1}{N}\sum_{i=1}^{N}\left|X^*(i)-\overline{X^*}_N\right|^2.$$

Proof. First we define $\mu_1:=\mu+\Delta$ and $\widehat{\mu}_1:=\widehat{\mu}+\widehat{\Delta}$.

Under H_0, we have $\mu = \mu_1$, then

$$\begin{aligned}\widehat{e}(i) &= X(i) - \widehat{\mu}\mathbb{1}_{[1,\widehat{k^*}]}(i) - \widehat{\mu}_1\mathbb{1}_{[\widehat{k^*}+1,N]}(i)\\ &= e(i) + (\mu - \widehat{\mu})\mathbb{1}_{[1,\widehat{k^*}]}(i) + (\mu - \widehat{\mu}_1)\mathbb{1}_{[\widehat{k^*}+1,N]}(i)\\ &= e(i) + (\mu - \widehat{\mu})\mathbb{1}_{[1,\widehat{k^*}]}(i) + (\mu_1 - \widehat{\mu}_1)\mathbb{1}_{[\widehat{k^*}+1,N]}(i).\end{aligned}$$

Furthermore under H_0

$$\begin{aligned}&\widehat{e}(i) - \overline{\widehat{e}}_N\\ &= e(i) - \overline{e}_N + (\mu - \widehat{\mu})\left(\mathbb{1}_{[1,\widehat{k^*}]}(i) - \frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{[1,\widehat{k^*}]}(j)\right) + (\mu_1 - \widehat{\mu}_1)\left(\mathbb{1}_{[\widehat{k^*}+1,N]}(i) - \frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{[\widehat{k^*}+1,N]}(j)\right)\\ &= e(i) - \overline{e}_N + (\mu - \widehat{\mu})\left(\mathbb{1}_{[1,\widehat{k^*}]}(i) - \frac{\widehat{k^*}}{N}\right) + (\mu_1 - \widehat{\mu}_1)\left(\mathbb{1}_{[\widehat{k^*}+1,N]}(i) - \frac{N-\widehat{k^*}}{N}\right)\\ &\leq e(i) - \overline{e}_N + |\mu - \widehat{\mu}|\left|\mathbb{1}_{[1,\widehat{k^*}]}(i) - \frac{\widehat{k^*}}{N}\right| + |\mu_1 - \widehat{\mu}_1|\left|\mathbb{1}_{[\widehat{k^*}+1,N]}(i) - \frac{N-\widehat{k^*}}{N}\right|\\ &= (e(i) - \overline{e}_N) + \frac{1}{\sqrt{N}}O_P(1)\\ &= e(i) - \overline{e}_N + o_P(1). \end{aligned} \tag{9.4}$$

The equality in (9.4) follows from

$$|\mu - \widehat{\mu}| = O_P\left(\frac{1}{\sqrt{N}}\right) \quad \text{and} \quad |\mu_1 - \widehat{\mu}_1| = O_P\left(\frac{1}{\sqrt{N}}\right), \tag{9.5}$$

and it follows

$$\frac{\widehat{k^*}}{N}|\mu - \widehat{\mu}| = O_P\left(\frac{1}{\sqrt{N}}\right) \quad \text{and} \quad \left(1 - \frac{\widehat{k^*}}{N}\right)|\mu_1 - \widehat{\mu}_1|\mathbb{1}_{\widehat{k^*}<N} = O_P\left(\frac{1}{\sqrt{N}}\right).$$

Under H_1 we have

$$\begin{aligned}\widehat{e}(i) &= X(i) - \widehat{\mu}\mathbb{1}_{[1,\widehat{k^*}]}(i) - \widehat{\mu}_1\mathbb{1}_{[\widehat{k^*}+1,N]}(i)\\ &= e(i) + \mu\mathbb{1}_{[1,k^*]}(i) + \mu_1\mathbb{1}_{[k^*+1,N]}(i) - \widehat{\mu}\mathbb{1}_{[1,\widehat{k^*}]}(i) - \widehat{\mu}_1\mathbb{1}_{[\widehat{k^*}+1,N]}(i)\\ &= e(i) + (\mu - \widehat{\mu})\mathbb{1}_{[1,\min(k^*,\widehat{k^*})]}(i) + (\mu_1 - \widehat{\mu}_1)\mathbb{1}_{(\max(k^*,\widehat{k^*}),N]}(i)\\ &\quad + (\mu - \widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i) + (\mu_1 - \widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)\end{aligned}$$

and thus

$$\begin{aligned}&\widehat{e}(i) - \overline{\widehat{e}}_N\\ &= e(i) - \overline{e}_N\\ &\quad + (\mu - \widehat{\mu})\left(\mathbb{1}_{[1,\min(k^*,\widehat{k^*})]}(i) - \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{[1,\min(k^*,\widehat{k^*})]}(i)\right)\end{aligned}$$

$$
\begin{aligned}
&+(\mu_1-\widehat{\mu}_1)\left(\mathbb{1}_{(\max(k^*,\widehat{k^*}),N]}(i)-\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{(\max(k^*,\widehat{k^*}),N]}(i)\right)\\
&+(\mu-\widehat{\mu}_1)\left(\mathbb{1}_{(\widehat{k^*},k^*]}(i)-\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{(\widehat{k^*},k^*]}(i)\right)+(\mu_1-\widehat{\mu})\left(\mathbb{1}_{(k^*,\widehat{k^*}]}(i)-\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{(k^*,\widehat{k^*}]}(i)\right)\\
=\ &e(i)-\overline{e}_N\\
&+(\mu-\widehat{\mu})\left(\mathbb{1}_{[1,\min(k^*,\widehat{k^*})]}(i)-\frac{\min(k^*,\widehat{k^*})}{N}\right)\\
&+(\mu_1-\widehat{\mu}_1)\left(\mathbb{1}_{(\max(k^*,\widehat{k^*}),N]}(i)-\frac{N-\max(k^*,\widehat{k^*})}{N}\right)\\
&+(\mu-\widehat{\mu}_1)\left(\mathbb{1}_{(\widehat{k^*},k^*]}(i)-\frac{k^*-\widehat{k^*}}{N}\right)+(\mu_1-\widehat{\mu})\left(\mathbb{1}_{(k^*,\widehat{k^*}]}(i)-\frac{\widehat{k^*}-k^*}{N}\right)\\
=\ &e(i)-\overline{e}_N\\
&+(\mu-\widehat{\mu})\mathbb{1}_{[1,\min(k^*,\widehat{k^*})]}(i)+(\mu_1-\widehat{\mu}_1)\mathbb{1}_{(\max(k^*,\widehat{k^*}),N]}(i)+(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)\\
&-\left((\mu-\widehat{\mu})\frac{\min(k^*,\widehat{k^*})}{N}+(\mu_1-\widehat{\mu}_1)\frac{N-\max(k^*,\widehat{k^*})}{N}+(\mu-\widehat{\mu}_1)\frac{k^*-\widehat{k^*}}{N}+(\mu_1-\widehat{\mu})\frac{\widehat{k^*}-k^*}{N}\right)\\
=\ &e(i)-\overline{e}_N\\
&+\frac{1}{\sqrt{N}}O_P(1)+(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\quad\text{uniformly in } i. \qquad (9.6)
\end{aligned}
$$

The last equality is established by the fact that the term in bracket is in $o_P(1)$, which is shown in Kirch (2006) in (4.6.23).

Then we are ready to show that the relevant conditions mentioned above based on the estimated residuals, can be simplified to the same conditions plugged in the original error plus an $o_P(1)$ term. We will show this simplification under H_1, because under H_0 it runs analogously if we set the additional term of H_1 equal to zero. By (9.5) and $|\widehat{k^*}-k^*|/N=o_P(1)$, which holds with Lemma 4.6.1(iii) in Kirch (2006), and together with $\delta>0$ we obtain

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left|\widehat{e}_i-\overline{\widehat{e}}_N\right|^{2+\delta}\leq\frac{1}{N}\sum_{i=1}^{N}\left|e(i)-\overline{e}_N+(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^{2+\delta}\\
&\leq 2^{2+\delta}\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^{2+\delta}+2^{2+\delta}\frac{1}{N}\sum_{i=1}^{N}\left|(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^{2+\delta}\\
&\leq 2^{2+\delta}\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^{2+\delta}+2^{2(2+\delta)}\frac{1}{N}\sum_{i=1}^{N}\left|(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)\right|^{2+\delta}\\
&\quad+2^{2(2+\delta)}\frac{1}{N}\sum_{i=1}^{N}\left|(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^{2+\delta}
\end{aligned}
$$

$$\leq 2^{2+\delta}\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^{2+\delta}+2^{2(2+\delta)}\frac{|\widehat{k^*}-k^*|}{N}|\mu-\widehat{\mu}_1|^{2+\delta}+2^{2(2+\delta)}\frac{|\widehat{k^*}-k^*|}{N}|\mu_1-\widehat{\mu}+o_P(1)|^{2+\delta}$$

$$\leq 2^{2+\delta}\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^{2+\delta}+o_P(1).$$

Condition (9.2) holds, since we have by (9.5), $|\widehat{k^*}-k^*|/N=o_P(1)$ and Cauchy-Schwarz inequality

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left|\widehat{e}(i)-\overline{\widehat{e}}_N\right|^2\\
&=\frac{1}{N}\sum_{i=1}^{N}\left|e(i)-\overline{e}_N+(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^2\\
&=\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^2+\frac{1}{N}\sum_{i=1}^{N}\left|(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^2\\
&\quad+2\frac{1}{N}\sum_{i=1}^{N}(e(i)-\overline{e}_N)\left((\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right)\\
&\geq\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^2+\frac{1}{N}\sum_{i=1}^{N}\left|(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^2\\
&\quad-2\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|\left|(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|\\
&\geq\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^2+\frac{1}{N}\sum_{i=1}^{N}\left|(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^2\\
&\quad-2\frac{1}{N}\left(\sum_{i=1}^{N}|e(i)-\overline{e}_N|^2\sum_{i=1}^{N}\left|(\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i)+o_P(1)\right|^2\right)^{\frac{1}{2}}\\
&=\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^2+o_P(1)\\
&=\operatorname{Var}(e(i))+o_P(1). \qquad (9.7)
\end{aligned}$$

The last equality follows by the law of large numbers. Using this law again gives

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}|e(i)-\overline{e}_N|^{2+\delta}\leq 2^{2+\delta}\frac{1}{N}\sum_{i=1}^{N}\left(|e(i)|^{2+\delta}+|\overline{e}_N|^{2+\delta}\right)\\
&=2^{2+\delta}\frac{1}{N}\sum_{i=1}^{N}|e(i)|^{2+\delta}+o_P(1)=2^{2+\delta}E\,|e(i)|^{2+\delta}+o_P(1)\leq C,\quad C>0.
\end{aligned}$$

Theorem 9.1 yields the assertion by an application of the subsequent principle. □

9.2. Univariate circular overlapping Block-Bootstrap

Furthermore we have to deal with the univariate block-bootstrap to be able to prove the asymptotics of the multivariate block-bootstrap. The circular overlapping block-bootstrap involves a periodic extension of the time series, which is portioned into overlapping blocks. By sampling these blocks with replacement, we obtain the bootstrapped sequence, which can be obviously longer that the time horizon N, so the remaining bootstrapped observations are not considered such that the bootstrapped series has the correct length.

The idea of proving the asymptotics of the score statistic using the block-bootstrap is to split the bootstrapped series into independent blocks and the remaining observations caused by the circular overlapping blocks. Then it can be shown that the remainder term is asymptotically irrelevant, and the blocks can be treated analogously to the Efron-bootstrap, see the details in the next theorem.

Theorem 9.3

FCLT for univariate Scores using the Circular Overlapping Block-Bootstrap:
Let $a_N(i), i = 1, ..., N-1+K$, with $a_N(i) = a_N(i-N), i > N$, be scores and $U_N(i), i = 1, 2, 3, ...L$ be independent uniformly distributed random variables in $\{0, 1, 2, ..., N-1\}$ and independent of $\{X(\cdot)\}$. L is the number of blocks and K is the block length such that $N = KL + \widetilde{k}$, where $0 \leq \widetilde{k} < K$ and $\frac{K}{L} = o(1)$. If there exists a $\delta > 0$ and $C, D_1, D_2 > 0$ such that

$$\frac{1}{N}\sum_{i=1}^{N} |a_N(i) - \overline{a}_N| \leq C, \tag{9.8}$$

$$\frac{1}{N}\sum_{i=0}^{N-1} \left| \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (a_N(i+j) - \overline{a}_N) \right|^{2+\delta} \leq D_1 \tag{9.9}$$

and

$$\tau_N^2(a) = \mathrm{Var}\left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} a_N(U_N(1) + j) \right) = \frac{1}{N}\sum_{i=0}^{N-1} \left| \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (a_N(i+j) - \overline{a}_N) \right|^{2} \geq D_2. \tag{9.10}$$

Then, as $N \to \infty$,

$$\left\{ \frac{1}{\sqrt{N}\tau_N(a)} \sum_{i=1}^{\lfloor Nt \rfloor} (a_N^*(i) - \overline{a}_N) \right\} \overset{D[0,1]}{\longrightarrow} \{W(t)\},$$

where $a_N^*(i) = a_N(U(l) + k),\ l = 1, \ldots, L,\ k = 1, \ldots, K$ and $W(\cdot)$ is a standard Wiener process.

Proof. The idea of the proof as mentioned above is to separate the bootstrap term into whole blocks, which can be treated like the Efron-Bootstrap because they are independent, and the remaining bootstrap observations, which are asymptotically negligible.

First we note

$$\frac{1}{\sqrt{N}\tau_N(a)} \sum_{i=1}^{\lfloor Nt \rfloor} (a_N^*(i) - \overline{a}_N)$$

$$= \frac{1}{\sqrt{N}\tau_N(a)} \left(\sum_{i=1}^{K\lfloor Lt \rfloor} (a_N^*(i) - \overline{a}_N) + \sum_{i=K\lfloor Lt \rfloor+1}^{\lfloor Nt \rfloor} (a_N^*(i) - \overline{a}_N) \right).$$

Then we show that the second term converges stochastically to zero by using condition (9.8):

$$\begin{aligned}
&E \left| \frac{1}{\sqrt{N}\tau_N(a)} \sum_{i=K\lfloor Lt \rfloor+1}^{\lfloor Nt \rfloor} (a_N^*(i) - \overline{a}_N) \right| \\
&\leq \frac{1}{\sqrt{N}\tau_N(a)} \sum_{i=K\lfloor Lt \rfloor+1}^{\lfloor Nt \rfloor} E \left| a_N^*(i) - \overline{a}_N \right| \\
&= \frac{1}{\sqrt{N}\tau_N(a)} \sum_{j=1}^{\lfloor Nt \rfloor - K\lfloor Lt \rfloor} E \left| a_N^* \left(K \lfloor Lt \rfloor + j \right) - \overline{a}_N \right| \\
&= \frac{1}{\sqrt{N}\tau_N(a)} \sum_{j=1}^{\lfloor Nt \rfloor - K\lfloor Lt \rfloor} E \left| a_N \left(U_N \left(\lfloor Lt \rfloor \right) + j \right) - \overline{a}_N \right| \\
&= \frac{1}{\sqrt{N}\tau_N(a)} \sum_{j=1}^{\lfloor Nt \rfloor - K\lfloor Lt \rfloor} \frac{1}{N} \sum_{i=0}^{N-1} \left| a_N \left(i + j \right) - \overline{a}_N \right| \\
&= \frac{1}{\sqrt{N}\tau_N(a)} \left(\lfloor Nt \rfloor - K \lfloor Lt \rfloor \right) \frac{1}{N} \sum_{i=1}^{N} \left| a_N(i) - \overline{a}_N \right| \\
&\leq \frac{1}{\sqrt{N}\tau_N(a)} \left(Nt + 1 - K(Lt - 1) \right) \frac{1}{N} \sum_{i=1}^{N} \left| a_N(i) - \overline{a}_N \right| \\
&\leq \frac{1}{\sqrt{N}\tau_N(a)} \left(\widetilde{k}t + 1 + K \right) \frac{1}{N} \sum_{i=1}^{N} \left| a_N(i) - \overline{a}_N \right| \\
&\leq \frac{1}{\sqrt{N}\tau_N(a)} 2K \frac{1}{N} \sum_{i=1}^{N} \left| a_N(i) - \overline{a}_N \right| \leq 2 \frac{K}{\sqrt{N}} \frac{C}{\sqrt{D_2}} \\
&= 2 \frac{K}{\sqrt{KL + \widetilde{k}}} \frac{C}{\sqrt{D_2}} \leq 2 \frac{K}{\sqrt{KL}} \frac{C}{\sqrt{D_2}} = 2 \sqrt{\frac{K}{L}} \frac{C}{\sqrt{D_2}} = 2o(1)O(1) = o(1).
\end{aligned}$$

Consequently we only have to consider the term composed of the blocks, while the second term is asymptotically negligible. We rewrite the relevant term as follows:

$$\begin{aligned}
&\frac{1}{\sqrt{N}\tau_N(a)} \sum_{i=1}^{K\lfloor Lt \rfloor} (a_N^*(i) - \overline{a}_N) \\
&= \frac{1}{\sqrt{L}\tau_N(a)} \sum_{l=0}^{\lfloor Lt \rfloor} \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (a_N^*(Kl + j) - \overline{a}_N) \\
&= \frac{1}{\sqrt{L}\tau_N(a)} \sum_{l=0}^{\lfloor Lt \rfloor} \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (a_N(U_N(l) + j) - \overline{a}_N). \qquad (9.11)
\end{aligned}$$

Since the blocks are independent, we can use Theorem 9.1 on the Efron-bootstrap for the term in (9.11). So the scores in Theorem 9.1 are now set to be the blocks:

$$a_N(i)^{9.1} := \frac{1}{\sqrt{K}} \sum_{j=1}^{K} a_N(i+j), \quad i = 1, \ldots, L.$$

The scores $a_N(i)^{9.1}$ satisfy the conditions (9.1) and (9.2) because they fulfill the conditions (9.9) and (9.10).

Consequently Theorem 9.1 gives a functional central limit theorem for the independent blocks:

$$\frac{1}{\sqrt{L}\tau_N(a)} \sum_{l=0}^{\lfloor Lt \rfloor} \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (a_N^*(Kl+j) - \overline{a}_N) \overset{D[0,1]}{\longrightarrow} W(t) \tag{9.12}$$

The calculations above yields

$$\begin{aligned}
&\frac{1}{\sqrt{N}\tau_N(a)} \sum_{i=1}^{\lfloor Nt \rfloor} (a_N^*(i) - \overline{a}_N) \\
&= \frac{1}{\sqrt{L}\tau_N(a)} \sum_{l=0}^{\lfloor Lt \rfloor} \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (a_N^*(Kl+j) - \overline{a}_N) + \frac{1}{\sqrt{N}\tau_N(a)} \sum_{i=K\lfloor Lt \rfloor+1}^{K\lfloor Lt \rfloor + K\left(Lt + \frac{\tilde{k}}{K} - \lfloor Lt \rfloor\right)} (a_N^*(i) - \overline{a}_N) \\
&= \frac{1}{\sqrt{L}\tau_N(a)} \sum_{l=0}^{\lfloor Lt \rfloor} \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (a_N^*(Kl+j) - \overline{a}_N) + o_P(1). \qquad (9.13)
\end{aligned}$$

Finally, by (9.12) and (9.13), the assertion follows. □

Analogous to the Efron case in (9.3), with the help of the previous theorem we obtain the same limit distribution of the score statistic as for the original statistic under H_0, where $a_N^*(i)$ is the block-bootstrapped sequence.

Corollary 9.2

Let the assumptions of Theorem 9.3 hold. If $a_N^*(i)$ is the block-bootstrapped sequence, then

$$\max_{1 \le t \le N} \frac{1}{\tau_N(a)\sqrt{N}} \left| \sum_{i=1}^{t} \left(a_N^*(i) - \overline{a}_{N,U}^*\right) \right| \overset{D}{\longrightarrow} \sup_{0<s\le 1} |B(s)|,$$

where $B(s)$ is a standard Brownian bridge.

Proof. By Theorem 9.3 we obtain

$$\begin{aligned}
&\max_{1 \le t \le N} \frac{1}{\tau_N(a)\sqrt{N}} \left| \sum_{i=1}^{t} \left(a_N^*(i) - \overline{a}_{N,U}^*\right) \right| = \max_{1 \le t \le N} \frac{1}{\tau_N(a)\sqrt{N}} \left| \sum_{i=1}^{t} \left(a_N^*(i) - \overline{a}_N - \left(\overline{a}_{N,U}^* - \overline{a}_N\right)\right) \right| \\
&= \max_{1 \le t \le N} \frac{1}{\tau_N(a)\sqrt{N}} \left| \sum_{i=1}^{t} (a_N^*(i) - \overline{a}_N) - t \left(\frac{1}{N} \sum_{i=1}^{N} (a_N^*(i) - \overline{a}_N) \right) \right|
\end{aligned}$$

$$= \max_{\frac{1}{N}\leq s\leq 1} \frac{1}{\tau_N(a)\sqrt{N}} \left| \sum_{i=1}^{\lfloor Ns \rfloor} (a_N^*(i) - \overline{a}_N) - \frac{\lfloor Ns \rfloor}{N} \left(\frac{1}{N} \sum_{i=1}^{N} (a_N^*(i) - \overline{a}_N) \right) \right|$$

$$\xrightarrow{D} \sup_{0<s\leq 1} |W(s) - sW(1)| = \sup_{0<s\leq 1} |B(s)| .$$

□

In the next theorem, we prove the asymptotic correctness of the critical values obtained by the block-bootstrap. Therefore, we replace the scores by the estimated residuals, whereby the original error sequence has to fulfill some moment and mixing conditions.

Theorem 9.4

Let $e(t), t = 1, ..., N$, fulfill the assumptions of Theorem 3.5.1 in Kirch (2006), and under the alternative of $k^* = \lfloor N\nu \rfloor,\ 0 < \nu < 1$, let the change Δ be fixed.

Then under the null hypothesis as well as under the alternative we have

$$P\left(\max_{1\leq t\leq N} \frac{1}{\tau_N \sqrt{N}} \left| \sum_{i=1}^{t} \left(X^*(i) - \overline{X^*}_N \right) \right| \leq x | X_1, ..., X_N \right) \xrightarrow{P} P\left(\sup_{0<l\leq 1} |B(l)| \leq x \right),$$

where

$$\tau_N^2 = \frac{1}{N} \sum_{i=0}^{N-1} \left| \frac{1}{\sqrt{K}} \sum_{j=1}^{K} \left(X_N^*(i+j) - \overline{X^*}_N \right) \right|^2 .$$

Proof. We have to verify conditions (9.8), (9.9) and (9.10) for the bootstrapped sequence. Therefore we use the transformation in (9.6) for the estimated residuals under the alternative. Under the null hypothesis the simplification runs correspondingly with (9.4).

So first we define again $\mu_1 := \mu + \Delta$ and $\widehat{\mu}_1 := \widehat{\mu} + \widehat{\Delta}$.

Then we show (9.8) by the help of (9.5) and $|\widehat{k^*} - k^*|/N = o_P(1)$, which holds with Lemma 4.6.1(iii) in Kirch (2006). Indeed, we have

$$\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} \left| \widehat{e}(i) - \overline{\widehat{e}}_N \right| \\
&\leq \frac{1}{N} \sum_{i=1}^{N} \left| e(i) - \overline{e}_N + (\mu - \widehat{\mu}_1) \mathbb{1}_{(\widehat{k^*}, k^*]}(i) + (\mu_1 - \widehat{\mu}) \mathbb{1}_{(k^*, \widehat{k^*}]}(i) + o_P(1) \right| \\
&\leq \frac{1}{N} \sum_{i=1}^{N} |e(i) - \overline{e}_N| + |\mu - \widehat{\mu}_1| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{(\widehat{k^*}, k^*]}(i) + |\mu_1 - \widehat{\mu}| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{(k^*, \widehat{k^*}]}(i) + \frac{1}{N} \sum_{i=1}^{N} o_P(1) \\
&= \frac{1}{N} \sum_{i=1}^{N} |e(i) - \overline{e}_N| + |\mu - \widehat{\mu}_1| \frac{k^* - \widehat{k^*}}{N} + |\mu_1 - \widehat{\mu}| \frac{\widehat{k^*} - k^*}{N} + o_P(1) \\
&= \frac{1}{N} \sum_{i=1}^{N} |e(i) - \overline{e}_N| + o_P(1) \\
&= E|e(1) - Ee(1)| + o_P(1)
\end{aligned}$$

$$= o_P(1).$$

The last equality follows from Theorem B.8. in Kirch (2006).
Further by (9.5) and $|k^* - \widehat{k^*}| = O_P(1)$ and with $\delta > 0$ we get

$$\frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\widehat{e}(i+j) - \overline{\widehat{e}}_N\right)\right|^{2+\delta}$$

$$= \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(e(i+j) - \overline{e}_N + (\mu - \widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j) + (\mu_1 - \widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j) + o_P(1)\right)\right|^{2+\delta}$$

$$\leq O(1)\left(\frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j) - \overline{e}_N)\right|^{2+\delta} + \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\mu - \widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j)\right|^{2+\delta}\right.$$

$$\left. + \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\mu_1 - \widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j)\right|^{2+\delta} + \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}o_P(1)\right|^{2+\delta}\right)$$

$$\leq O(1)\left(\frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j) - \overline{e}_N)\right|^{2+\delta} + |\mu - \widehat{\mu}_1|\left(\frac{(k^* - \widehat{k^*})}{\sqrt{K}}\right)^{2+\delta} + |\mu_1 - \widehat{\mu}|\left(\frac{(\widehat{k^*} - k^*)}{\sqrt{K}}\right)^{2+\delta}\right.$$

$$\left. + \left(\sqrt{\frac{K}{N}}\right)^{2+\delta}\right)$$

$$\leq O(1)\frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j) - \overline{e}_N)\right|^{2+\delta} + o_P(1).$$

That the first term is bounded by a constant greater than zero has already been shown in Theorem 3.5.1. in Kirch (2006). To prove the second condition (9.10), we do an analogous transformation as in (9.7) for the Efron-bootstrap and obtain

$$\frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\widehat{e}(i+j) - \overline{\widehat{e}}_N\right)\right|^{2}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(e(i+j) - \overline{e}_N + (\mu - \widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j) + (\mu_1 - \widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j) + o_P(1)\right)\right|^{2}$$

$$= \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j) - \overline{e}_N)\right|^{2}$$

$$+ \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left((\mu - \widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j) + (\mu_1 - \widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j) + o_P(1)\right)\right|^{2}$$

$$
\begin{aligned}
&+ 2\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j)-\overline{e}_N)\right) \\
&\qquad\qquad \left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j)+o_P(1)\Big)\right) \\
&\geq \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j)-\overline{e}_N)\right|^2 \\
&\quad + \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j)+o_P(1)\Big)\right|^2 \\
&\quad - 2\frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j)-\overline{e}_N)\right| \\
&\qquad\qquad \left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j)+o_P(1)\Big)\right| \\
&\geq \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j)-\overline{e}_N)\right|^2 \\
&\quad + \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j)+o_P(1)\Big)\right|^2 \\
&\quad - 2\frac{1}{N}\left(\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j)-\overline{e}_N)\right|^2\right. \\
&\qquad\qquad \left.\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\mu-\widehat{\mu}_1)\mathbb{1}_{(\widehat{k^*},k^*]}(i+j)+(\mu_1-\widehat{\mu})\mathbb{1}_{(k^*,\widehat{k^*}]}(i+j)+o_P(1)\Big)\right|^2\right)^{\frac{1}{2}} \\
&= \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(e(i+j)-\overline{e}_N)\right|^2 + o_P(1) \qquad (9.14)
\end{aligned}
$$

To show that the term in (9.14) is greater than a constant has already been done in Theorem 3.5.1. in Kirch (2006).

Then by Theorem 9.3 we obtain the assertion. □

10. Multivariate Bootstrap Methods

We are now able to prove the validity of the multivariate bootstrap methods. Analogously to the univariate setting, we first attend to the corresponding FCLT theorems for the scores in case of the Efron-bootstrap and then to the block-bootstrap. Then we replace the scores by the estimated residuals and specify their requirements to fulfill the necessary conditions for the FCLT of the scores P-stochastically.

As usual the proofs of the FCLT theorems are composed of two steps. First we show the convergence of the finite-dimensional distributions and then the tightness of the partial sum process which, however, is now multivariate. To verify these properties we make use of the FCLT's of the univariate scores.

10.1. Multivariate Efron-Bootstrap

Theorem 10.1

FCLT for Multivariate Scores using the Efron-Bootstrap:
Let $\boldsymbol{a}_N(i), i = 1, ..., N$, be d-dimensional score-vectors and $U_N(i), i = 1, ..., N$, be independent uniformly distributed random variables in {1, 2, ..., N} and independent of $\{\boldsymbol{X}(\cdot)\}$. Suppose there is a $\delta > 0$ and a constant $D_1^M > 0$ such that the scores fulfill

$$\frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\|^{2+\delta} \leq D_1^M. \tag{10.1}$$

Then, as $N \to \infty$,

$$\left\{ \frac{1}{\sqrt{N}} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \sum_{i=1}^{\lfloor Nt \rfloor} \left(\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N\right) : 0 \leq t \leq 1 \right\} \stackrel{D[0,1]^d}{\longrightarrow} \{\boldsymbol{W}(t) : 0 \leq t \leq 1\},$$

where $\{\boldsymbol{W}(t)\}$ is a d-dimensional standard Wiener Process and

$$\boldsymbol{\Sigma}_{\boldsymbol{a}} = \frac{1}{N}\sum_{i=1}^{N} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)^T.$$

Proof. Again we have to show the convergence of the finite-dimensional distributions and the tightness of the partial sum process, which is now multivariate.

1. The convergence of the finite-dimensional distributions:
   First we define
   $$\boldsymbol{X}_N(t) := \frac{1}{\sqrt{N}} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \sum_{i=1}^{\lfloor Nt \rfloor} \left(\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N\right).$$

By condition (10.1) the univariate scores $\boldsymbol{\lambda}^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \boldsymbol{a}_N$ with $\boldsymbol{\lambda}^T\boldsymbol{\lambda} > 0$ fulfill (9.1) and (9.2): There are constants $\delta, D_1 > 0$ such that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left|\boldsymbol{\lambda}^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\right|^{2+\delta} \\
&\leq \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{\lambda}\|^{2+\delta} \|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\|^{2+\delta} \\
&= \|\boldsymbol{\lambda}\|^{2+\delta} \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\|^{2+\delta} \leq D_1.
\end{aligned}$$

Moreover, there is a constant $D_2 > 0$ such that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left|\boldsymbol{\lambda}^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\right|^{2} \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\boldsymbol{\lambda}\right) \\
&= \boldsymbol{\lambda}^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)^T\right) \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{a}} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^T\boldsymbol{\lambda} \geq D_2.
\end{aligned}$$

Then, for an arbitrary linear combination $\boldsymbol{\lambda}^T \boldsymbol{X}_N(t)$, the FCLT for the univariate Efron-bootstrap gives

$$\boldsymbol{\lambda}^T \boldsymbol{X}_N(t) \xrightarrow{D} V(t),$$

where $V(t)$ is a univariate Wiener Process with variance $\boldsymbol{\lambda}^T\boldsymbol{\lambda}$.

Now, for an arbitrary $\boldsymbol{\lambda}$ with $\boldsymbol{\lambda}^T\boldsymbol{\lambda} > 0$, we can rewrite $V(t) = \boldsymbol{\lambda}^T \boldsymbol{W}(t)$, where $\boldsymbol{W}(t)$ is a multivariate standard Wiener process. Then we deduce $\boldsymbol{\lambda}^T \boldsymbol{X}_N(t) \xrightarrow{D} \boldsymbol{\lambda}^T \boldsymbol{W}(t)$ and the Cramér-Wold-device yields $\boldsymbol{X}_N(t) \xrightarrow{D} \boldsymbol{W}(t)$.

For two time points $s < t$ we analogously obtain with the univariate FCLT in the Efron case and by the independence of $\boldsymbol{X}_N(s)$ and $\boldsymbol{X}_N(t) - \boldsymbol{X}_N(s)$

$$\boldsymbol{\lambda}^T \boldsymbol{X}_N(s) + \boldsymbol{\lambda}^T \left(\boldsymbol{X}_N(t) - \boldsymbol{X}_N(s)\right) \xrightarrow{D} \boldsymbol{\lambda}^T \boldsymbol{W}(s) + \boldsymbol{\lambda}^T \left(\boldsymbol{W}(t) - \boldsymbol{W}(s)\right)$$

The Cramér-Wold-device implies $\left(\boldsymbol{X}_N(s), \boldsymbol{X}_N(t) - \boldsymbol{X}_N(s)\right) \xrightarrow{D} \left(\boldsymbol{W}(s), \boldsymbol{W}(t) - \boldsymbol{W}(s)\right)$, and the continuous mapping theorem entails $\left(\boldsymbol{X}_N(s), \boldsymbol{X}_N(t)\right) \xrightarrow{D} \left(\boldsymbol{W}(s), \boldsymbol{W}(t)\right)$.

For more than two time points the proof runs analogously.

2. Tightness:
Since $\boldsymbol{X}_N(t)$ is multivariate, it is a random element of the product space $D[0,1]^d = D[0,1] \times \cdots \times D[0,1]$, and each component is in $D[0,1]$. With Lemma B.2 we have to show that the sequence of the component random elements $X_{N,i},\ i = 1, \ldots, d$ of $\boldsymbol{X}_N$, are tight.

We write $X_{N,i} = \boldsymbol{u}_i^T \boldsymbol{X}_N(t),\ i = 1, ..., d$, where $\boldsymbol{u}_i = (0, ..., 0, 1, 0, ..., 0)$ is the i-th unit vector. As shown above, the component random elements $X_{N,i}$ converge in distribution to univariate Wiener processes.

Thus, by Proposition B.1, $X_{N,i}$ is relatively compact for fixed i. Since $D[0, 1]$ is complete and separable under the Skorohod metric, we deduce with Theorem B.3 that the sequence of component random elements $X_{N,i}$ is tight. Lemma B.2 then yields the tightness of $\boldsymbol{X}_N$.

□

The next corollary, shows that the multivariate score statistic when using the Efron-bootstrap has the same limit process as the original multivariate statistic under H_0.

Corollary 10.1

If the assumptions of Theorem 10.1 are satisfied, then

$$\max_{1\le k\le N} \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_{N,U})\right)^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_{N,U})\right) \xrightarrow{D} \sup_{0<s\le 1} \boldsymbol{B}^T(s)\boldsymbol{B}(s),$$

where $\boldsymbol{B}(s)$ is a multivariate standard Brownian bridge.

Proof. By Theorem 10.1 we have

$$\begin{aligned}
&\max_{1\le k\le N} \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_{N,U})\right)^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_{N,U})\right) \\
&= \max_{1\le k\le N} \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N - (\overline{\boldsymbol{a}}_{N,U} - \overline{\boldsymbol{a}}_N))\right)^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N - (\overline{\boldsymbol{a}}_{N,U} - \overline{\boldsymbol{a}}_N))\right) \\
&= \max_{1\le k\le N} \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N) - k\left(\frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N)\right)\right)^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \\
&\qquad \left(\sum_{i=1}^{k} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N) - k\left(\frac{1}{N}\sum_{i=1}^{T} (\boldsymbol{a}_N(U_N(i)) - \overline{\boldsymbol{a}}_N)\right)\right) \\
&\xrightarrow{D} \sup_{0<s\le 1} \left| \left(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{\frac{1}{2}} (\boldsymbol{W}(s) - s\boldsymbol{W}(1))\right)^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \left(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{\frac{1}{2}} (\boldsymbol{W}(s) - s\boldsymbol{W}(1))\right) \right| = \sup_{0<s\le 1} \left|\boldsymbol{B}^T(s)\boldsymbol{B}(s)\right|.
\end{aligned}$$

□

10.2. Multivariate Block-Bootstrap

The proof of the FCLT for scores by using the multivariate block-bootstrap runs along with the same strategy as in the univariate case. We split the bootstrapped sequence into independent blocks and a remainder term which is asymptotically negligible.

Theorem 10.2

FCLT for Multivariate Scores using the Circular Overlapping Block-Bootstrap
Let $\boldsymbol{a}_N(i), i = 1, ..., N$, be d-dimensional score-vectors, with $\boldsymbol{a}_N(i) = \boldsymbol{a}_N(i-N), i > N$, and $U_N(i), i = 1, 2, 3, ...L$, be i.i.d. uniformly distributed random variables in $\{0, 1, 2, ..., \widetilde{N-1}\}$. L is the number of blocks and K is the block length. Put $N = KL + \widetilde{k}$, where $0 \leq \widetilde{k} < K$, and $\frac{K}{L} = o(1)$. Suppose there exist $C^M, D_1^M > 0$ such that

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N\right)\right\| \leq C^M \tag{10.2}$$

and

$$\frac{1}{N}\sum_{i=0}^{N-1}\left\|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{a}_N(i+j) - \overline{\boldsymbol{a}}_N\right)\right\|^{2+\delta} \leq D_1^M, \tag{10.3}$$

where $\boldsymbol{a}_N^*(t) = \boldsymbol{a}_N(U_N(l) + k)$ with $l = 1, ..., L$ and $k = 1, ..., K$. Then, as $N \to \infty$,

$$\left\{\frac{1}{\sqrt{N}}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\sum_{i=1}^{\lfloor Nt\rfloor}\left(\boldsymbol{a}_N^*(i) - \overline{\boldsymbol{a}}_N\right) : 0 \leq t \leq 1\right\} \overset{D[0,1]^d}{\longrightarrow} \{\boldsymbol{W}(t) : 0 \leq t \leq 1\},$$

where $\{\boldsymbol{W}(t)\}$ is a d-dimensional Standard-Wiener Process and
$\boldsymbol{\Sigma}_{\boldsymbol{a}} = \frac{1}{N}\sum_{i=0}^{N-1}\left(\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{a}_N(i+j) - \overline{\boldsymbol{a}}_N\right)\right)\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{a}_N(i+j) - \overline{\boldsymbol{a}}_N\right)\right)^T\right)$.

Proof. The proof follows the same reasoning as the one for the univariate block-bootstrap.

1. Convergence of the finite-dimensional distributions:
   First we define

$$\boldsymbol{X}_N(t) := \frac{1}{\sqrt{N}}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\sum_{i=1}^{\lfloor Nt\rfloor}\left(\boldsymbol{a}_N^*(i) - \overline{\boldsymbol{a}}_N\right)$$

   and split $\boldsymbol{X}_N(t)$ in the same way as in the univariate case: The first part is the sum of the entire blocks

$$\boldsymbol{X}_N(t) = \frac{1}{\sqrt{N}}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\left(\sum_{i=1}^{K\lfloor Lt\rfloor}\left(\boldsymbol{a}_N^*(i) - \overline{\boldsymbol{a}}_N\right) + \sum_{i=K\lfloor Lt\rfloor+1}^{\lfloor Nt\rfloor}\left(\boldsymbol{a}_N^*(i) - \overline{\boldsymbol{a}}_N\right)\right).$$

   Then, analogously to Assumption (10.2), we have

$$\begin{aligned}
&E\left\|\frac{1}{\sqrt{N}}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\sum_{i=K\lfloor Lt\rfloor+1}^{\lfloor Nt\rfloor}\left(\boldsymbol{a}_N^*(i) - \overline{\boldsymbol{a}}_N\right)\right\| \\
&= E\left\|\frac{1}{\sqrt{N}}\sum_{i=K\lfloor Lt\rfloor+1}^{\lfloor Nt\rfloor}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\left(\boldsymbol{a}_N^*(i) - \overline{\boldsymbol{a}}_N\right)\right\|
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{\sqrt{N}} \sum_{i=K\lfloor Lt \rfloor+1}^{\lfloor Nt \rfloor} E \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N^*(i) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&= \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Nt \rfloor - K\lfloor Lt \rfloor} E \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N^* \left(K \lfloor Lt \rfloor + j \right) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&= \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Nt \rfloor - K\lfloor Lt \rfloor} E \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N \left(U_N \left(\lfloor Lt \rfloor \right) + j \right) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&= \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Nt \rfloor - K\lfloor Lt \rfloor} \frac{1}{N} \sum_{i=0}^{N-1} \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N \left(i + j \right) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&\leq \frac{1}{\sqrt{N}} (\lfloor Nt \rfloor - K\lfloor Lt \rfloor) \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&\leq \frac{1}{\sqrt{N}} (Nt + 1 - K(Lt-1)) \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&= \frac{1}{\sqrt{N}} (\widetilde{k}t + 1 + K) \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&\leq \frac{1}{\sqrt{N}} 2K \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N \right) \right\| \leq 2 \frac{K}{\sqrt{N}} C^M \\
&= 2 \frac{K}{\sqrt{KL + \widetilde{k}}} C^M \leq 2 \frac{K}{\sqrt{KL}} C^M = 2 \sqrt{\frac{K}{L}} C^M = 2o(1)O(1) = o(1).
\end{aligned}
$$

Hence we obtain the part consisting of $(\lfloor Lt \rfloor + 1)$ independent blocks, which we define as

$$
\widetilde{\boldsymbol{X}}_N(t) := \frac{1}{\sqrt{L}} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \sum_{l=0}^{\lfloor Lt \rfloor} \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (\boldsymbol{a}_N^*(Kl + j) - \overline{\boldsymbol{a}}_N)
$$

and a remainder term that is asymptotically negligible.

Analogously to the proof of the Efron-bootstrap we get by conditions (10.2) and (10.3) that the univariate scores $\{\lambda^T \boldsymbol{X}_N(t) : 0 \leq t \leq 1\}$ fulfill the required conditions (9.8) and (9.9) of the univariate block-bootstrap:
First we assume $\lambda^T \lambda > 0$.

By (10.2) there is a constant C such that

$$
\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} \left\| \lambda^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&\leq \|\lambda\| \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}} \left(\boldsymbol{a}_N(i) - \overline{\boldsymbol{a}}_N \right) \right\| \\
&\leq C.
\end{aligned}
$$

By (10.3) there are constants $\delta, D_1 > 0$ such that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=0}^{N-1}\left|\lambda^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N)\right|^{2+\delta} \\
&\leq \|\lambda\|^{2+\delta}\frac{1}{N}\sum_{i=0}^{N-1}\left\|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N)\right\|^{2+\delta} \\
&\leq D_1.
\end{aligned}$$

Additionally, there exists a constant $D_2 > 0$ with

$$\begin{aligned}
&\frac{1}{N}\sum_{i=0}^{N-1}\left|\lambda^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N)\right|^{2} \\
&= \frac{1}{N}\sum_{i=0}^{N-1}\left(\lambda^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N)\right)\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N)\right)^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\lambda\right) \\
&= \lambda^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\left(\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N)\right)\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N)\right)^T\right)\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\lambda \\
&= \lambda^T \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\boldsymbol{a}}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\lambda = \lambda^T\lambda \geq D_2.
\end{aligned}$$

Then the proof runs analogously to the lines of the proof on the multivariate Efron-bootstrap by replacing $\boldsymbol{X}_N$ with $\widetilde{\boldsymbol{X}}_N$.

2. Tightness:
   This proof is equivalent to the proof of tightness for the Efron-bootsrap again by replacing $\boldsymbol{X}_N$ with $\widetilde{\boldsymbol{X}}_N$.

We know that the independent blocks satisfy a multivariate functional central limit theorem and the remaining term converges stochastically to zero. Then the assertion follows in the same way as in the univariate case. □

We finally ask, which conditions a time series must satisfy in order to obtain asymptotically correct critical values under the null and alternative hypothesis by using the multivariate bootstrap methods.

The answer is as follows. If the univariate conditions in the theorems about the Efron- and block-bootstrap hold, then there is only one condition missing, such that the multivariate conditions are also satisfied. We show this claim in the next lines for the necessary multivariate conditions 10.1, 10.2 and 10.3:

Let the multivariate scores satisfy conditions (9.1),(9.2) and (9.8),(9.9) in every component and there is a constant $D > 0$ such that $\|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\|_F \leq D < \infty$. Then there are constants $C^M, D^M_{1,efron}, D^M_{1,block} > 0$ such that

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}(\boldsymbol{a}_N(i)-\overline{\boldsymbol{a}}_N)\right\| \leq \|\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-\frac{1}{2}}\|_F\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{a}_N(i)-\overline{\boldsymbol{a}}_N\| \leq C^M,$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\Sigma_a}^{-\frac{1}{2}}\left(\boldsymbol{a}_N(i)-\overline{\boldsymbol{a}}_N\right)\|^{2+\delta} \leq \|\boldsymbol{\Sigma_a}^{-\frac{1}{2}}\|_F \frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{a}_N(i)-\overline{\boldsymbol{a}}_N\|^{2+\delta} \leq D^M_{1,efron},$$

and

$$\frac{1}{N}\sum_{i=0}^{N-1}\left\|\boldsymbol{\Sigma_a}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N\right)\right\|^{2+\delta}$$
$$\leq \|\boldsymbol{\Sigma_a}^{-\frac{1}{2}}\|_F \frac{1}{N}\sum_{i=0}^{N-1}\left\|\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{a}_N(i+j)-\overline{\boldsymbol{a}}_N\right)\right\|^{2+\delta} \leq D^M_{1,block},$$

where $\|\cdot\|_F$ is the Frobenius norm, which is submultiplicative by Banerjee and Roy (2014).

Corollary 10.2

Consequently we get the correct limit behaviour of the bootstrap statistic, given the observed data set for multivariate time series $\boldsymbol{X}(t)$ fulfill the same conditions in every component, which are necessary for the validity of the univariate Efron- and block-bootstrap and if the long-run covariance matrix $\boldsymbol{\Sigma}$ of the multivariate errors $\boldsymbol{e(t)}$ is positive definite.

Proof. First with the Cauchy-Schwarz inequality we obtain

$$\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\widehat{\boldsymbol{e}}(i+j)-\overline{\widehat{\boldsymbol{e}}}_N\right)\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\widehat{\boldsymbol{e}}(i+j)-\overline{\widehat{\boldsymbol{e}}}_N\right)^T\right)$$
$$=\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N+(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{\boldsymbol{k}^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\right)\right)$$
$$\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N+(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{k^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\right)^T\right)$$
$$=\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)+\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left((\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{\boldsymbol{k}^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\right)\right)$$
$$\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)^T+\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left((\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{\boldsymbol{k}^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\right)^T\right)$$
$$=\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)^T\right.$$
$$+\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left((\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{\boldsymbol{k}^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\right)^T$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{\boldsymbol{k}^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\Big)\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)^T\\
&+\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{\boldsymbol{k}^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\Big)\\
&\left.\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big((\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}_1)\mathbb{1}_{(\widehat{\boldsymbol{k}^*},\boldsymbol{k}^*]}(i)+(\boldsymbol{\mu}_1-\widehat{\boldsymbol{\mu}})\mathbb{1}_{(\boldsymbol{k}^*,\widehat{\boldsymbol{k}^*}]}(i)+o_P(1)\Big)^T\right)\\
&=\frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)^T+o_P(1)\\
&=\operatorname{Var}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)\right)+o_P(1)\\
&=\frac{1}{K}\operatorname{Var}\left(\sum_{j=1}^{K}\left(\boldsymbol{e}(i+j)-\overline{\boldsymbol{e}}_N\right)\right)+o_P(1)\\
&=\boldsymbol{\Sigma}+o_P(1).
\end{aligned}
$$

By the assumption the long-run covariance matrix $\boldsymbol{\Sigma}$ of the errors $\boldsymbol{e}(t)$ is positive definite. Consequently its inverse is it too and the square root of the inverse also. So now we have

$$
A:=\boldsymbol{\Sigma}^{-\frac{1}{2}}\text{ is positive definite.}
$$

It holds $\|A\|_F\le\sqrt{\lambda_{max}}$, where λ_{max} is the maximum of the eigenvalues of the matrix A^TA. If the matrix A has the eigenvalues $\sqrt{\lambda_1},\dots,\sqrt{\lambda_d}$, then the matrix A^TA has the strictly positive eigenvalues $\lambda_1,\dots,\lambda_d$. Then it follows that there is a constant D such that $\|A\|_F\le D$ and

$$
\frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big(\widehat{\boldsymbol{e}}(i+j)-\overline{\widehat{\boldsymbol{e}}}_N\Big)\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\Big(\widehat{\boldsymbol{e}}(i+j)-\overline{\widehat{\boldsymbol{e}}}_N\Big)^T\right)\le D+o_P(1).
$$

□

11. Comparison of the Bootstrap Methods in Simulations

In this chapter we compare the bootstrap methods concerning the empirical size and size-adjusted power. The last one is already explained in Chapter 6.

We still focus on the statistic which only takes into account the dependence in time, so we focus on the bootstrap statistic T_N^*. This version can be used for higher dimensions because the long-run covariance matrix is a diagonal matrix and the corresponding estimator for the inverse is stable. But we also want to compare the bootstrap methods in lower dimensions and check the behaviour compared with the asymptotic method because, if the dimension is small, the inverse of the estimator for the whole long-run covariance matrix can still be calculated. Thus the simulation study is done with dimensions $2, 3$ and dimensions 10 and 20.

We use the multivariate mean change model with mean zero

$$X_t = \mathbf{\Delta} \mathbb{1}_{\{t > k^*\}} + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\epsilon}_t$ is a AR(1)-sequence with parameter -0.5 and standard normal distributed innovations, as well as the multivariate error sequence is either independent or dependent between their components. If the errors have dependent components, the covariance matrix is

$$\begin{pmatrix} 1 & 0.3 & \cdots & \cdots & 0.3 \\ 0.3 & 1 & 0.3 & \cdots & 0.3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0.3 & \cdots & 0.3 & 1 & 0.3 \\ 0.3 & \cdots & 0.3 & 0.3 & 1 \end{pmatrix}.$$

The simulations are based on 2500 repetitions and 1000 bootstrap replications. The length of the time series is $N = 300$. For the unknown change point we choose an early, a middle and a late time point, set to $0.05N$, $0.5N$ and $0.75N$, respectively. The size of the change depending on the dimension is given by:

$$\begin{aligned}
\text{Dimension } 2: \quad & \mathbf{\Delta} := \mathbf{\Delta}_{2,1} = 0.3 \cdot (1, 0)^T \\
& \mathbf{\Delta} := \mathbf{\Delta}_{2,2} = 0.3 \cdot (1, 1)^T \\
\text{Dimension } 3: \quad & \mathbf{\Delta} := \mathbf{\Delta}_{3,1} = 0.3 \cdot (1, 0, 0)^T \\
& \mathbf{\Delta} := \mathbf{\Delta}_{3,2} = 0.3 \cdot (1, 1, 0)^T \\
& \mathbf{\Delta} := \mathbf{\Delta}_{3,3} = 0.3 \cdot (1, 1, 1)^T \\
\text{Dimension } 10: \quad & \mathbf{\Delta} := \mathbf{\Delta}_{10,1} = 0.15 \cdot (1, 1, 1, 1, 1, 0, 0, 0, 0, 0)^T \\
& \mathbf{\Delta} := \mathbf{\Delta}_{10,2} = 0.15 \cdot (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)^T \\
& \mathbf{\Delta} := \mathbf{\Delta}_{10,3} = 0.15 \cdot (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)^T \\
\text{Dimension } 20: \quad & \mathbf{\Delta} := \mathbf{\Delta}_{20,1} = 0.1 \cdot (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^T \\
& \mathbf{\Delta} := \mathbf{\Delta}_{20,2} = 0.1 \cdot (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0)^T
\end{aligned}$$

$$\mathbf{\Delta} := \mathbf{\Delta}_{20,3} = 0.1 \cdot (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)^T.$$

The long-run covariance matrix $\mathbf{\Sigma}$, which is used for the asymptotic method, and the diagonal matrix $\mathbf{\Lambda}$ consisting of the long-run variances used for the bootstrap methods are estimated with the flat-top kernel estimator as introduced in Politis (2011). Except in case of the block-bootstrap the long-run covariance matrix can be calculated as stated in (8.1).

The matrix $\mathbf{\Sigma}$ can be estimated by the flat top kernel estimator

$$\widehat{\mathbf{\Sigma}}_{j,k} = \sum_{m=-(N-1)}^{N-1} \lambda\left(\frac{m}{S_{j,k}}\right) \widehat{\Gamma}_{j,k}(m), \quad j, k = 1, \ldots, d.$$

Here, $\widehat{\mathbf{\Gamma}}(\cdot)$ is an estimator for the autocovariance matrix, given by

$$\widehat{\mathbf{\Gamma}}(j) = \frac{1}{N} \sum_{t=1}^{N-j} \boldsymbol{e}_t \boldsymbol{e}_{t+j}^T, \; 0 \le j \le N-1 \quad \text{and} \quad \widehat{\mathbf{\Gamma}}(j) = \widehat{\mathbf{\Gamma}}(-j), \; N-1 \le j \le 0.$$

As flat-top kernel we use a trapezoidal function

$$\lambda(t) = \begin{cases} 1, & |t| \le \frac{1}{2}, \\ 2(1-|t|), & \frac{1}{2} < |t| < 1, \\ 0, & |t| \ge 1. \end{cases}$$

Since we need the inverse of the estimator for the long-run covariance matrix, we have to adapt the matrix estimator. In particular, this is of high interest in case of an estimator for the complete matrix $\mathbf{\Sigma}$. There exist an orthogonal matrix $\widehat{\boldsymbol{U}}$ and a diagonal matrix $\widehat{\boldsymbol{D}}$ such that $\widehat{\mathbf{\Sigma}} = \widehat{\boldsymbol{U}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{U}}^T$. To get a positive definite matrix as estimator, we can take a sequence b_N and determine the diagonal matrix $\widehat{\boldsymbol{D}}^b = \operatorname{diag}(\widehat{\lambda}_1^a, \ldots, \widehat{\lambda}_d^b)$, $\widehat{\lambda}_j^b = \max(\widehat{\lambda}_j, b_N)$. We choose $b_N = \frac{1}{N}$.

The bandwidth $S_{j,k}$, $j, k = 1, \ldots, d$, is estimated by

$$\widehat{S}_{j,k} = \max\left(\left\lceil \frac{\widehat{q}_{j,k}}{c_{ef}} \right\rceil, 1\right),$$

where $\widehat{q}_{j,k}$ is the smallest nonnegative integer such that

$$|\widehat{\rho}_{j,k}(\widehat{q}_{j,k} + m)| < C_0 \sqrt{\frac{\log_{10} N}{N}}, \quad m = 0, 1, \ldots, K_N.$$

The function $\widehat{q}_{j,k}$ is the autocorrelation function

$$\widehat{q}_{j,k} = \widehat{\mathbf{\Gamma}}_{j,k}(m) / \sqrt{\widehat{\mathbf{\Gamma}}_{j,j}(0) \widehat{\mathbf{\Gamma}}_{k,k}(0)}.$$

The constant c_{ef} is defined as the largest number satisfying

$$\lambda(t) \ge 1 - \epsilon \quad \forall\, t \in [-c_{ef}, c_{ef}].$$

If $j = k$, then

$$\widehat{S}_{j,k} = \max\left(\left\lceil \frac{\widehat{q}_{k,k}}{c_{ef}} \right\rceil, 1\right),$$

and if $j \neq k$, then

$$\widehat{S}_{j,k} = \widehat{S}_{k,j} = \max\left(\left\lceil \frac{\widehat{q}}{c_{ef}} \right\rceil, 1\right), \quad \widehat{q} = \max(\widehat{q}_{j,k}, \widehat{q}_{k,j}).$$

We choose the parameters as suggested in Politis (2011), hence $C_0 = 2, K_N = \max\left(5, \sqrt{\log_{10} N}\right)$ and $\epsilon = 0.01$.

In case of the block-bootstrap we use the method where we choose the block length manually equal to 5, 10 and 30. Furthermore, we use an automatic block length calculation and a tapered block bootstrap, where the bootstrap observations at the margins in a block get lesser weight than the ones in the middle. This yields a less biased estimator for the asymptotic variance of the sample mean than without the weighting, see Paparoditis and Politis (2001) for more details.

The automatic bandwidth is calculated as in Politis and White (2004), and therefore we use the function "*b.star*" included in the package "*np*" from the statistical software *R*. More precisely, the automatic bandwidth of the multivariate time series is calculated as the maximum of the automatic bandwidths of the univariate time series, which are calculated by the function "*b.star*".

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_N$ be multivariate observations. Then the tapered bootstrap sequence is defined as

$$\boldsymbol{X}^*_{mb+j} = w\left(\frac{j}{b+1}\right) \frac{b}{\left(\sum\limits_{t=1}^{b} w^2\left(\frac{t}{b+1}\right)\right)^{\frac{1}{2}}} \boldsymbol{X}_{i_m+j-1}, \quad j = 1, \dots, b; \quad m = 0, 1, \dots, k-1,$$

where b is the block length and $k = \lfloor \frac{N}{b} \rfloor$ is the number of blocks. Moreover, $i_0, i_1, \dots, i_{k-1}$ are i.i.d. uniformly distributed on $\{1, 2, \dots, N-b+1\}$. The weight function $w(t)$ determines the weights for each observation number in a block. We use a trapezoidal function

$$w(t) = \begin{cases} \frac{t}{c}, & t \in [0, c] \\ 1, & t \in [c, 1-c] \\ \frac{1-t}{c}, & t \in [1-c, 1], \end{cases}$$

with $c = 0.43$ as suggested in Paparoditis and Politis (2001) and with $c = 0.5$, where the weight function is then triangular. We additionally choose the triangular weight function for reasons of comparison with the dependent wild-bootstrap. For the wild-bootstrap method we choose the bartlett kernel

$$k(x) = \begin{cases} 1 - |x|, & |x| < 1 \\ 0, & \text{otherwise}, \end{cases}$$

which also has a triangular shape. The kernel function determines the covariance of the random variable $Z(t)$ by $k((t-s)/l), 0 \leq t \leq N, 0 \leq s \leq N$ multiplied with the estimated errors in each component. The parameter $l > 0$ is the chosen bandwidth, we use the bandwidth $3, 10$ and 30.

By using the AR-sieve bootstrap we first choose the correct order of the AR(1)-error sequence which is p equal to 1. Furthermore, we use the AIC information criteria to choose the order which is relevant in practice where we do not know the correct order. However, we set the upper bound for the order equal to 4.

In the following figures the empirical size as well as the size-adjusted power are illustrated, first in case of the independent errors and then for dependent errors with covariance matrix stated above. In Tabular 11.1 the legend of the following figures are explained in detail.

Asymp	Asymptotic method
Blockbootstrap_K5	Block-bootstrap with $K = 5$
Blockbootstrap_K5_trapez_tapered	tapered Block-bootstrap with $K = 5$ and $c = 0.43$
Blockbootstrap_K5_triangular_tapered	Block-bootstrap with $K = 5$ and $c = 0.5$
Blockbootstrap_K10	Block-bootstrap with $K = 10$
Blockbootstrap_K10_trapez_tapered	Block-bootstrap with $K = 10$ and $c = 0.43$
Blockbootstrap_K10_triangular_tapered	Block-bootstrap with $K = 10$ and $c = 0.5$
Blockbootstrap_K30	Block-bootstrap with $K = 30$
Blockbootstrap_K30_trapez_tapered	tapered Block-bootstrap with $K = 30$ and $c = 0.43$
Blockbootstrap_K30_triangular_tapered	Block-bootstrap with $K = 30$ and $c = 0.5$
Blockbootstrap_Kauto	Block-bootstrap with automatic block length choice
Blockbootstrap_Kauto_trapez_tapered	Block-bootstrap with automatic block length choice, $c = 0.43$
Blockbootstrap_Kauto_triangular_tapered	Block-bootstrap with automatic block length choice, $c = 0.5$
Wildbootstrap_l3	Wild-bootstrap with $l = 3$
Wildbootstrap_l10	Wild-bootstrap with $l = 10$
Wildbootstrap_l30	Wild-bootstrap with $l = 30$
ARSiebbootstrap_p1	AR-Sieve-bootstrap with order $p = 2$ of the AR-sequence
ARSiebbootstrap_paic	AR-Sieve-bootstrap with AIC-criterion for order p

Table 11.1.: Explanation of the legend of the following figures

First we look at the case of independent components concerning the empirical size in Figure 11.1. In dimensions 2, 3 and 4 the asymptotic method is also illustrated, because the long-run covariance matrix is still calculable. As we see in Figures 11.1(a)-(c), the empirical size of the asymptotic method increases rapidly with growing dimensions. In dimension 2 the empirical size of the asymptotic procedure is still acceptable but already in dimension 3 and 4 the asymptotic method is unfeasible due to the extremely high empirical size. Furthermore, the empirical size of the block-bootstrap with a short block length equal to 5 is higher than the empirical size of the asymptotic method if $d = 2$ and only somewhat lower if $d \in \{3, 4\}$. Even the empirical size of the block-bootstrap with block size 10 is very high for the lower dimensions $d \in \{2, 3, 4\}$. Additionally, the block-bootstrap with block length 5 and 10 performs extremely bad if the dimension is high (see Figures 11.1(d)-(e)), and even with tapering the performance is extremely bad. The bad performance of the block-bootstrap with a short block length is an evidence that the dependence structure of the AR(1)-sequence cannot be represented by blocks of such a short length. The block-bootstrap with automatic choice of the block length has a very high empirical size if the dimension is high, namely $d \in \{10, 20\}$ (Figures 11.1(d)-(e)), however with using tapering the empirical size is getting much better. The other bootstrap methods are very robust against the increasing dimension concerning the empirical size, and they either maintain or almost maintain the nominal level even in high dimensions. The empirical size of the tapered bootstrap method using a trapezoidal and a triangular weight function is very similar, but lower compared with the corresponding bootstrap method without tapering. This also holds for the bootstrap methods where the blocks length is chosen automatically. We suppose that the block length is always chosen between 10 and 30 because the empirical size with automatic bandwidth choice is between the empirical size of the corresponding bootstrap methods with block length 10 and 30 in Figures 11.1(a)-(e).

By using the wild-bootstrap the empirical size decreases with the bandwidth which is equal to 3, 10, 30 and is getting higher with growing dimension.

In case of the AR-Sieve-bootstrap there is almost no difference in the empirical size between using the correct order of the $AR(1)$-sequence or using the AIC criterion for choosing the order of the AR-sequence where we define the maximum order equal to 4. Moreover, this method has the lowest empirical size for all dimensions.

Regrading the empirical size, the methods with using dependencies between the components besides the dependence in time perform similar as without the components dependencies, except that the empirical size is lower in general as without the component dependencies for all methods (see Figure 11.2). Particularly the tapered block-bootstrap with a short block size equal to 5 or 10 has now an acceptable performance concerning the empirical size, even in higher dimensions.

Next we compare the size-adjusted power of the procedures, which is shown in Figures 11.3-11.6 for the different dimensions. First of all the size-adjusted power is best for the change in the middle of the time horizon for all dimensions. For a later change the size-adjusted power is still good, but for a early change the procedures perform very bad concerning the size-adjusted power, particularly for higher dimensions where the power is equal to the nominal level. The last mentioned effect is probably caused by the generally lower power for all time points of the changes, because the changes in each dimension are slightly lower than in case of the lower dimensions.

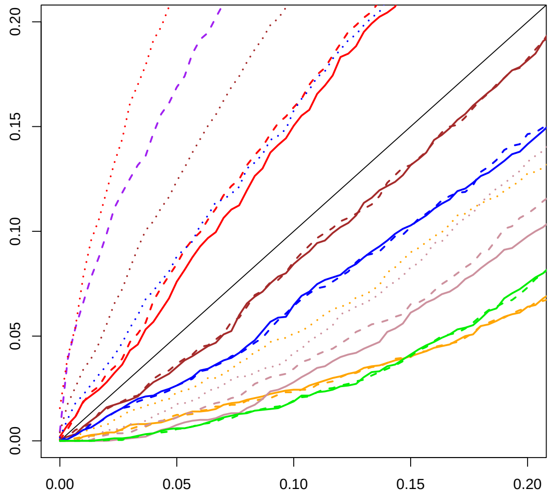
The asymptotic method is included in the comparison for dimension 2 and 3. In case of dimension 2 the size-adjusted power is still similar to the power of the bootstrap methods, except for an early change, where the size-adjusted power of the asymptotic method is very bad, namely equal to the nominal level and this is the same in dimension 3.

However, already for dimension 3, the asymptotic method is getting worse, and the bootstrap procedures have all similar higher size-adjusted power. In dimension 10 and 20 the size-adjusted power is also very similar for all bootstrap methods in all considered scenarios except for one case. In dimension 10 the change point is late and the mean only changes in 5 dimensions of 10., there the Wild-bootstrap and the AR-Sieve-bootstrap are worse than the other methods.

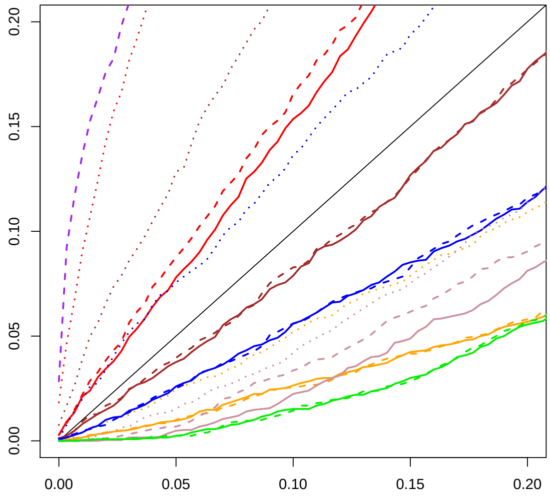
If the components are dependent, the size-adjusted power is only illustrated for the change point in the middle of the time horizon and for the cases that mean changes occur only in 1 component if $d \in \{2, 3\}$ and in 5 components if $d = 10$ and in 10 components if $d = 20$ (see Figure 11.7). The reason is that the behaviour between the different methods still remains the same as with independent components.

As we see in Figure 11.7 the asymptotic methods are still better than the bootstrap methods for dimension 2, but already for dimension 3 it is getting worse and the bootstrap methods perform better concerning the size-adjusted power. In all dimensions the bootstrap methods have similar size-adjusted power.

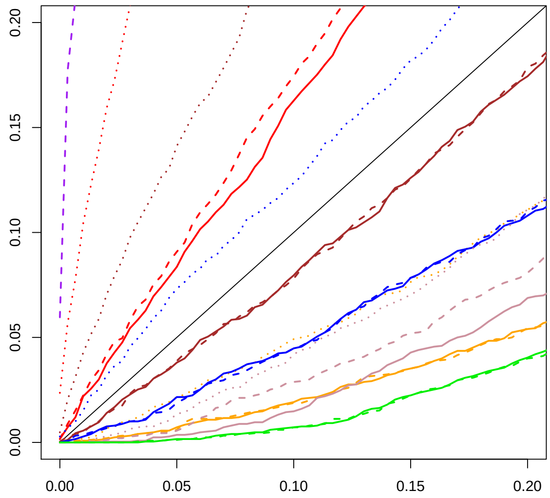
Figure 11.1.: Empirical size of the bootstrap methods with independent components.



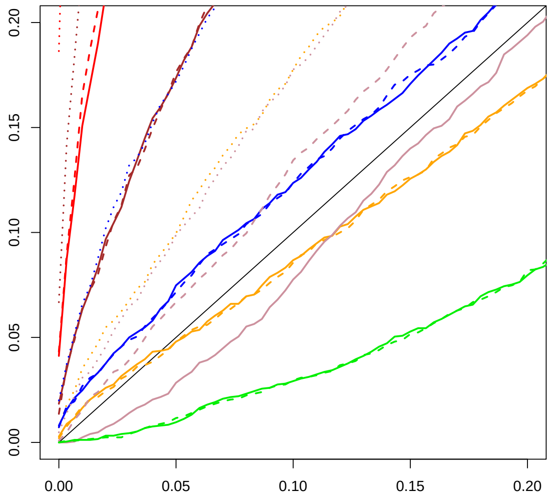
(a) Dimension 2



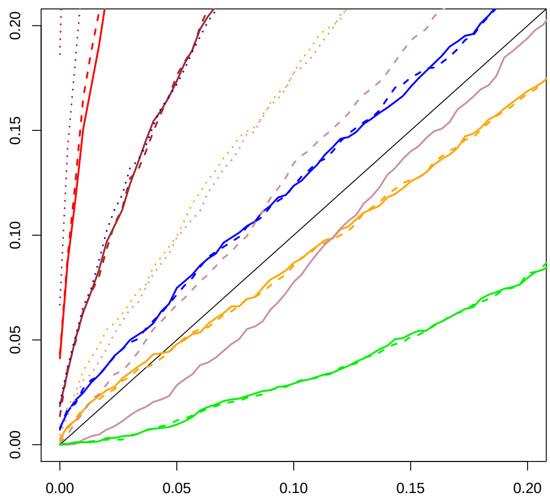
(b) Dimension 3



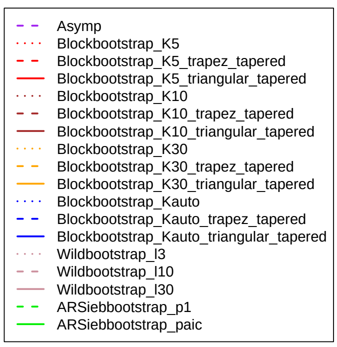
(c) Dimension 4



(d) Dimension 10

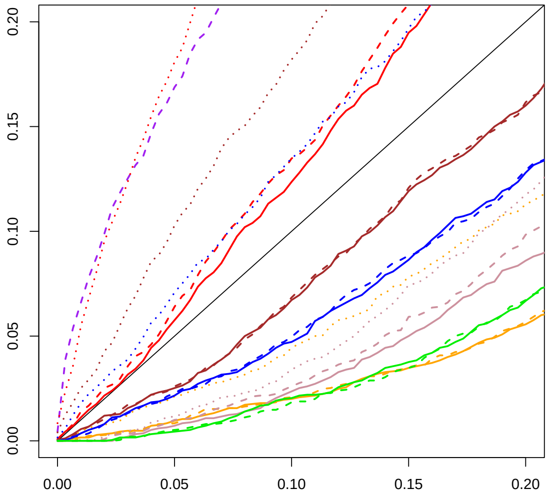


(e) Dimension 20

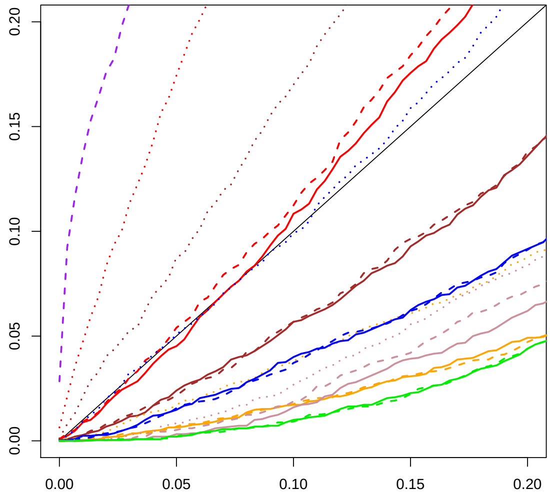


(f) Legend

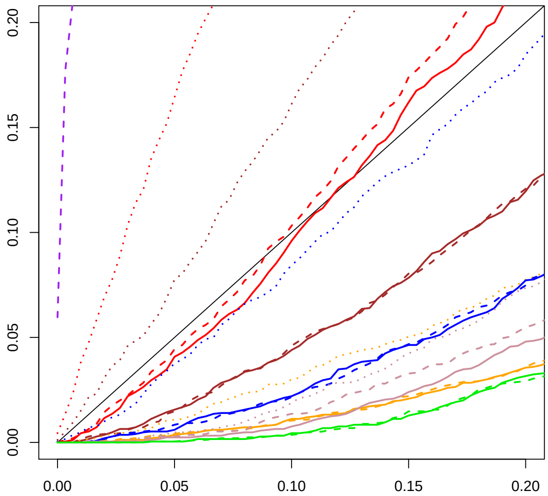
Figure 11.2.: Empirical size of the bootstrap methods with covariance of the components of 0.3.



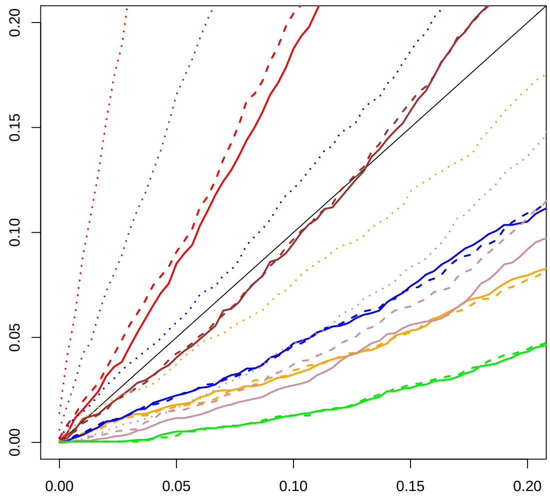
(a) Dimension 2



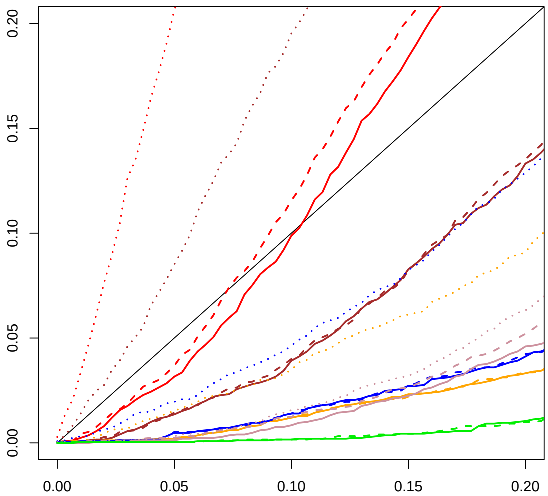
(b) Dimension 3



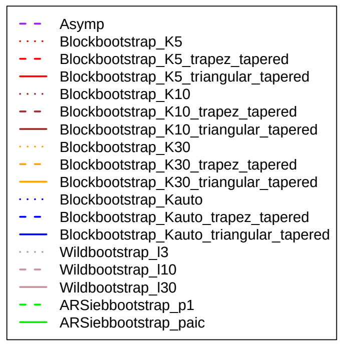
(c) Dimension 4



(d) Dimension 10

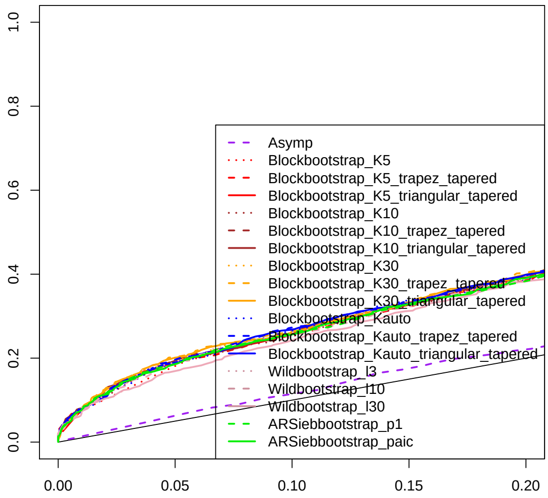


(e) Dimension 20

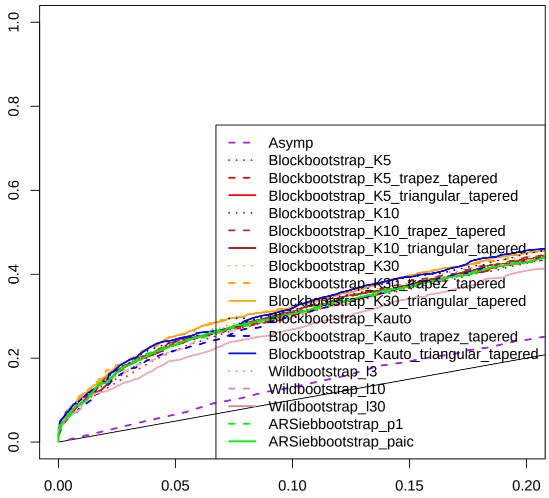


(f) Legend

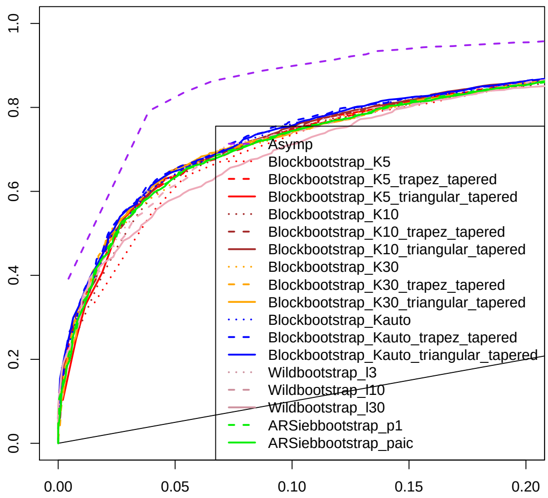
Figure 11.3.: Size-adjusted power of the bootstrap methods for dimension 2 with independent components



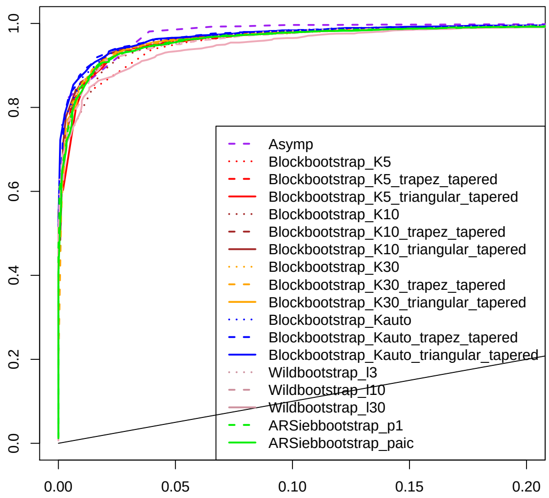
(a) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{2,1}$



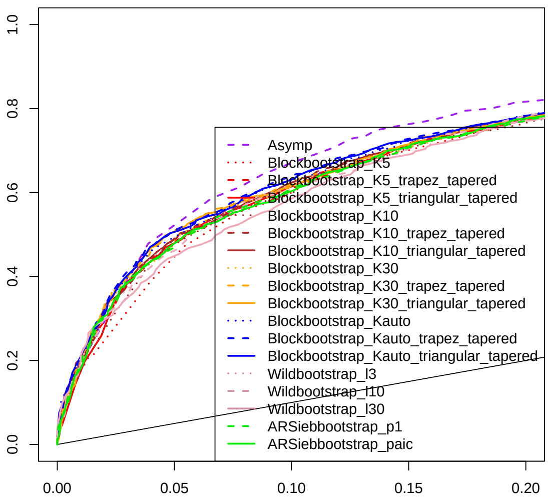
(b) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{2,2}$



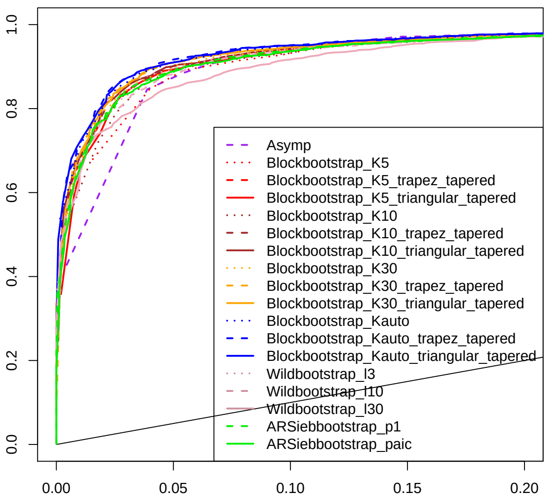
(c) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{2,1}$



(d) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{2,2}$

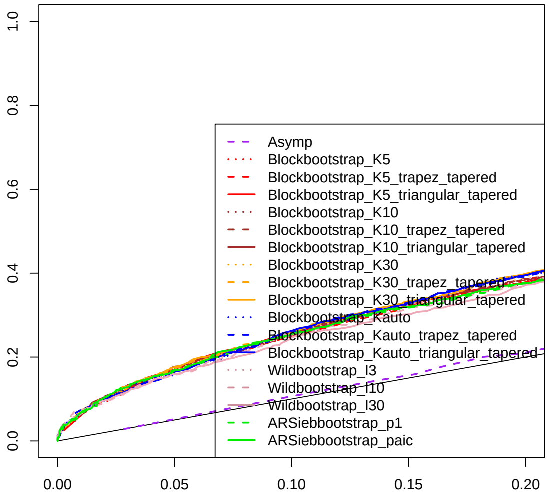


(e) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{2,1}$

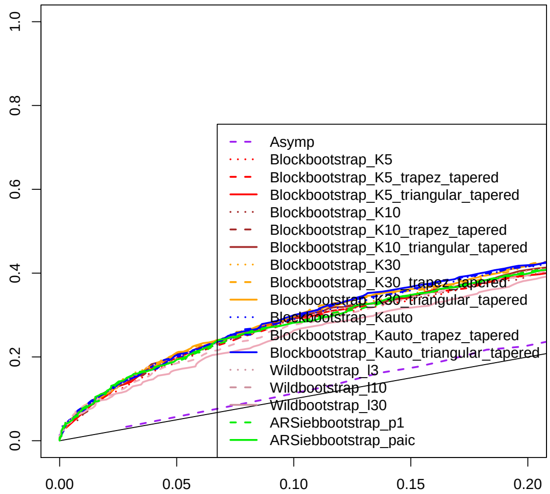


(f) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{2,2}$

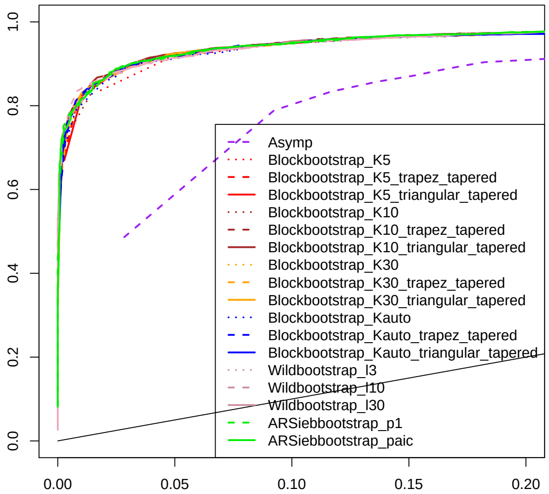
Figure 11.4.: Size-adjusted power of the bootstrap methods for dimension 3 with independent components



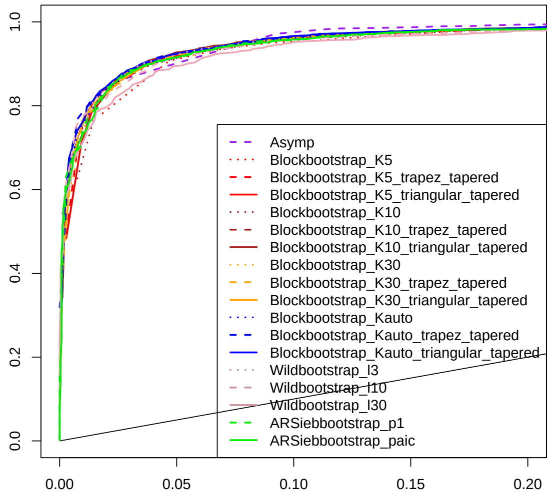
(a) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,1}$



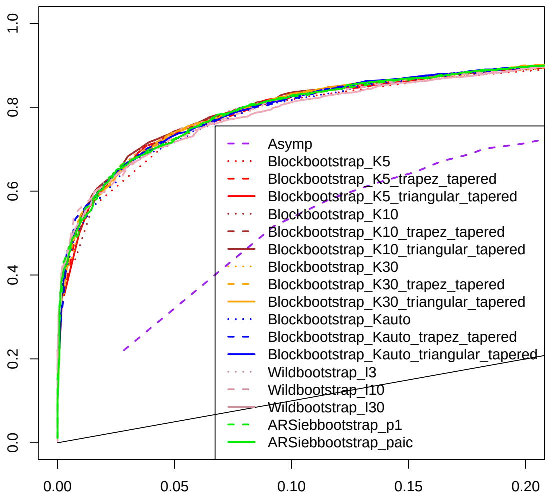
(b) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,2}$



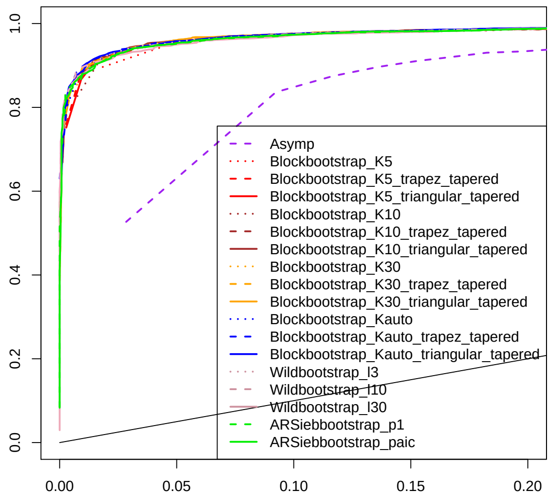
(c) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,1}$



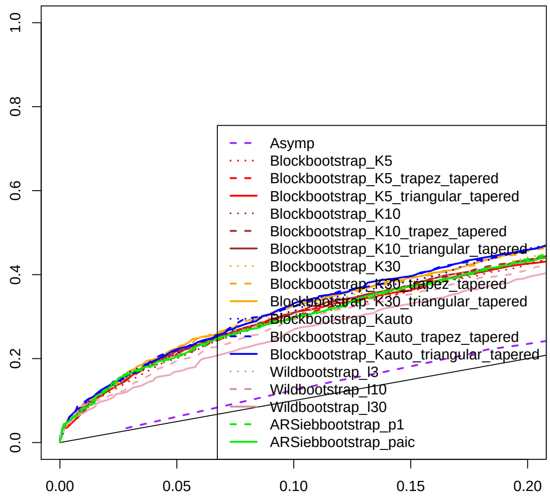
(d) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,2}$



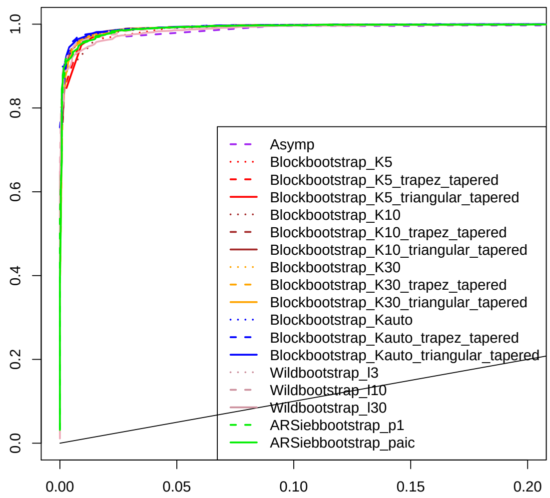
(e) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,1}$



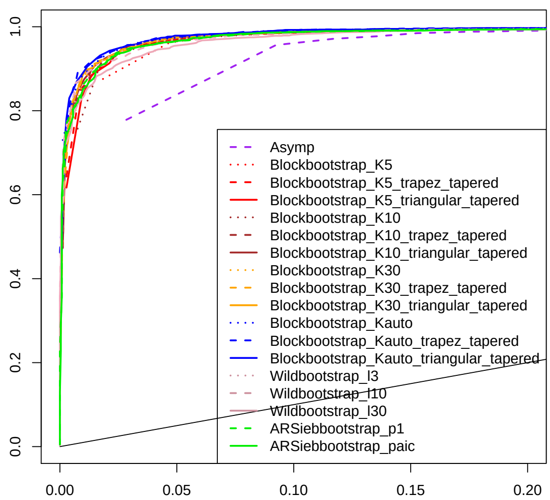
(f) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,2}$



(g) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,3}$

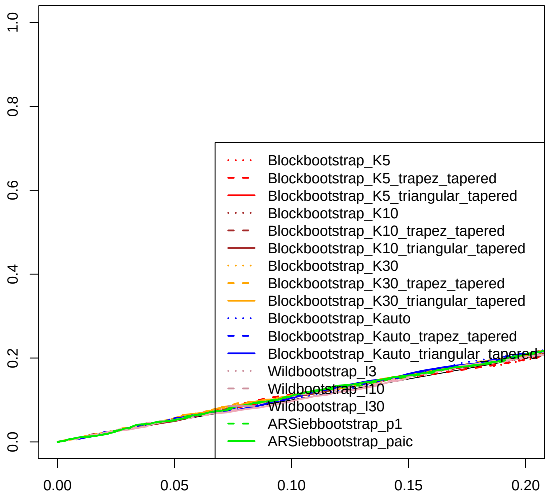


(h) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,3}$

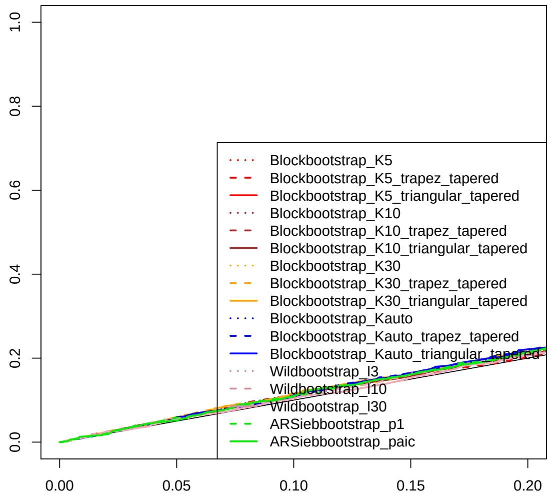


(i) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,3}$

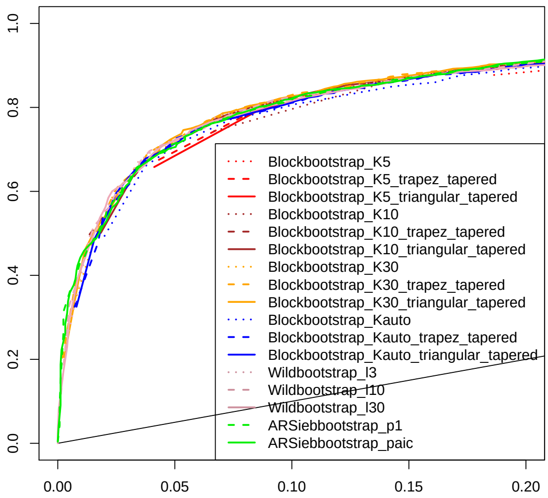
Figure 11.5.: Size-adjusted power of the bootstrap methods for dimension 10 with independent components



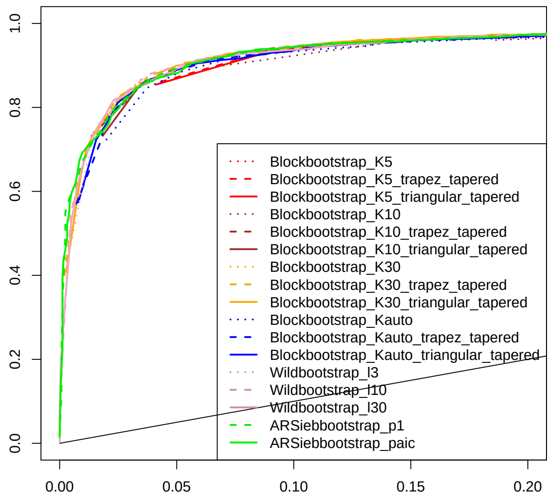
(a) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,1}$



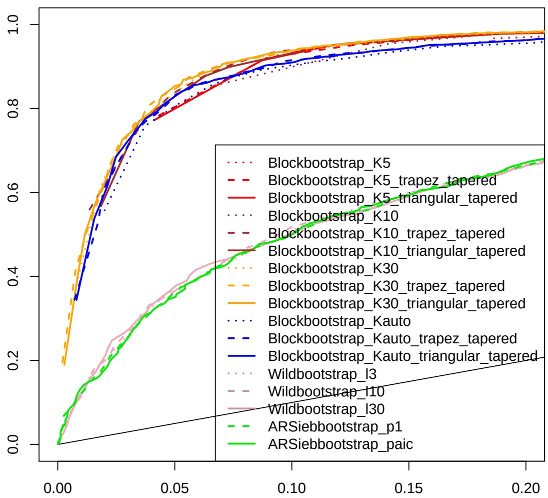
(b) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,2}$



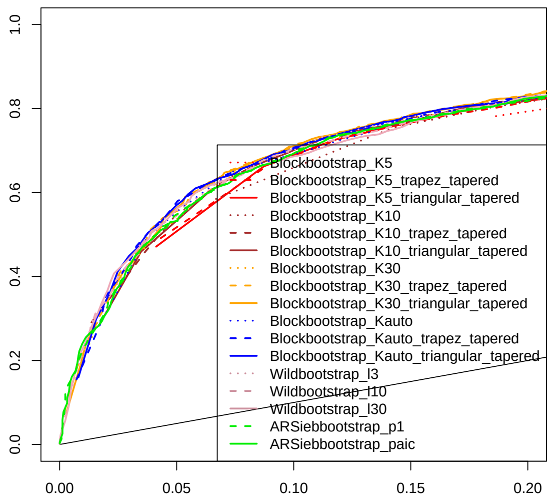
(c) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,1}$



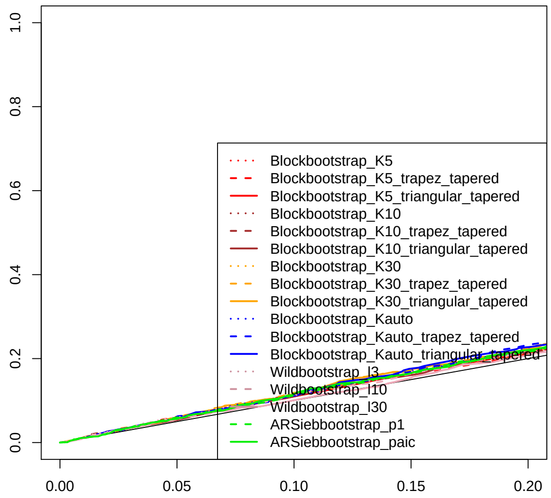
(d) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,2}$



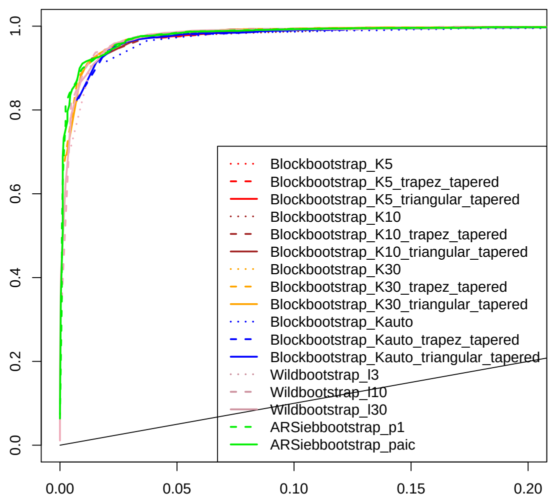
(e) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,1}$



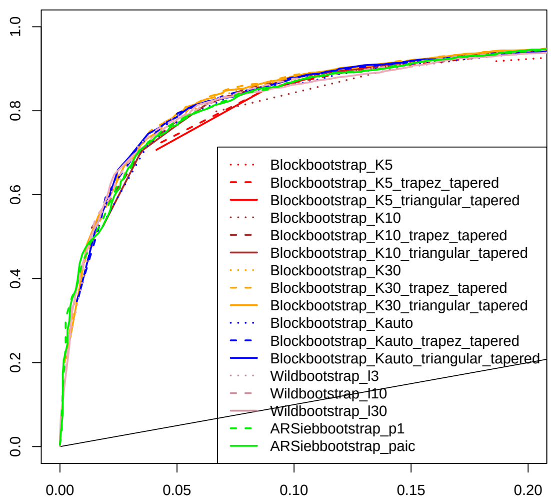
(f) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,2}$



(g) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,3}$

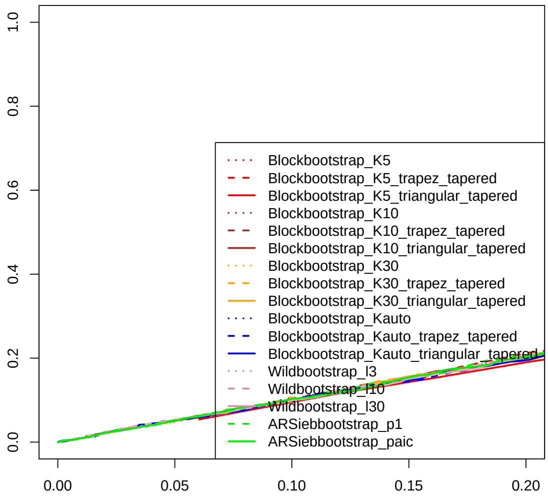


(h) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,3}$

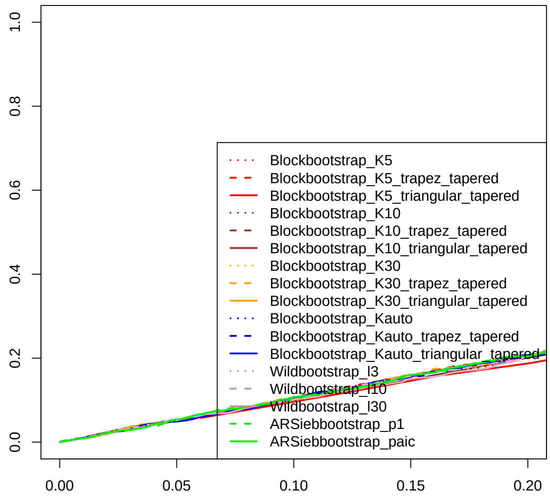


(i) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,3}$

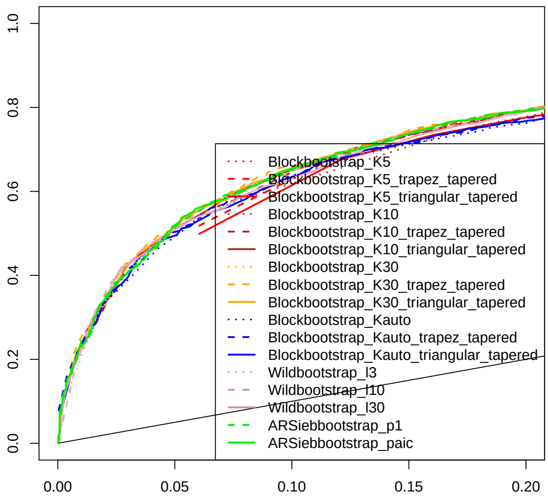
Figure 11.6.: Size-adjusted power of the bootstrap methods for dimension 20 with independent components



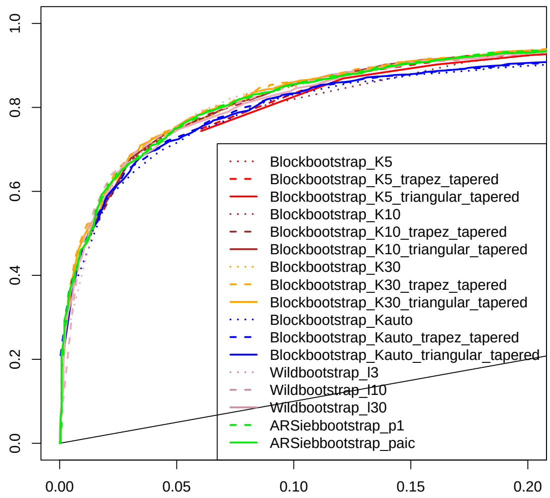
(a) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,1}$



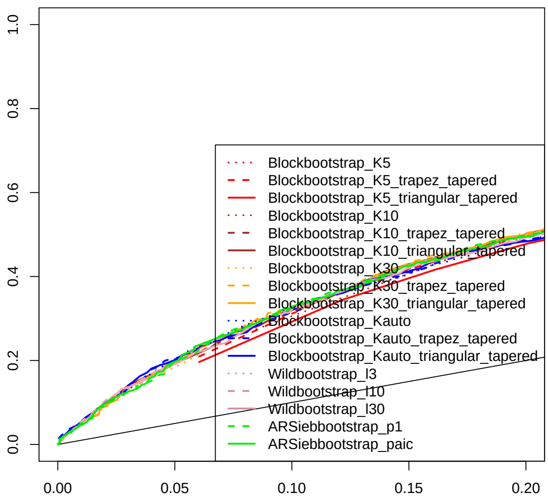
(b) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,2}$



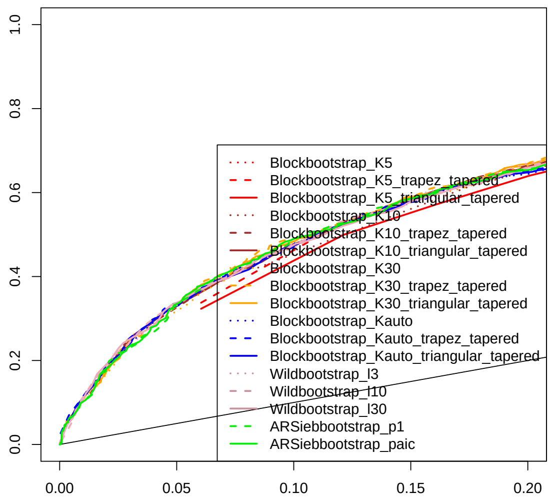
(c) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,1}$



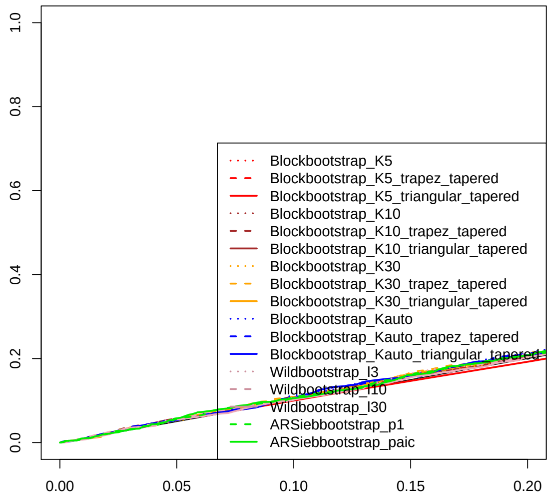
(d) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,2}$



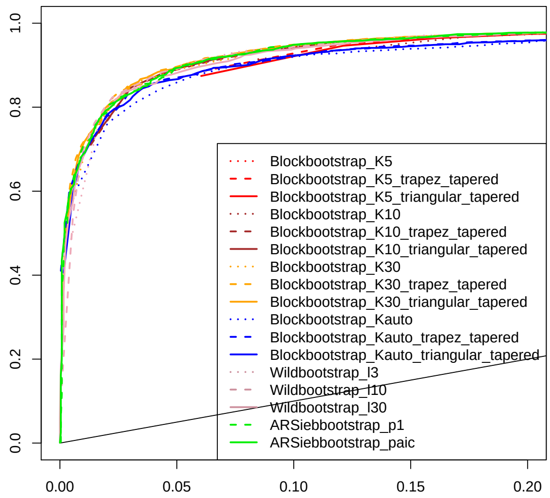
(e) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,1}$



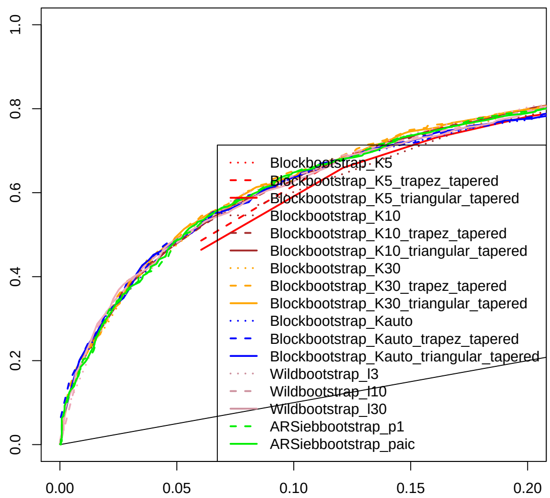
(f) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,2}$



(g) $k^* = 0.05N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,3}$

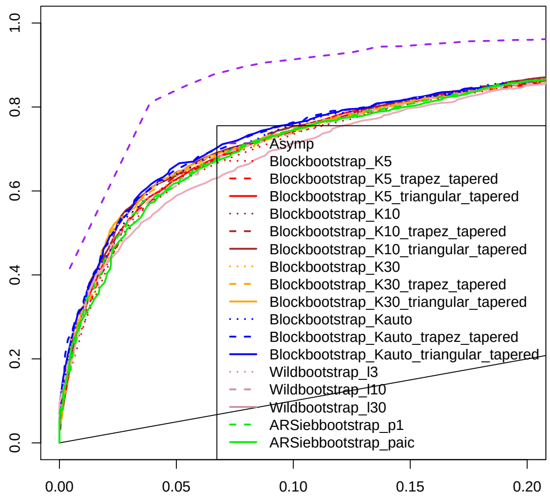


(h) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,3}$

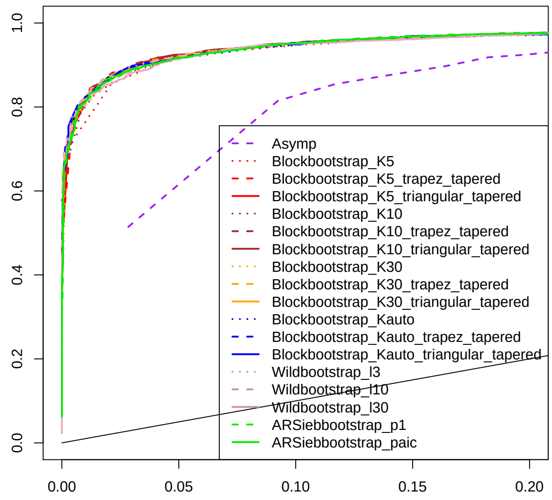


(i) $k^* = 0.75N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,3}$

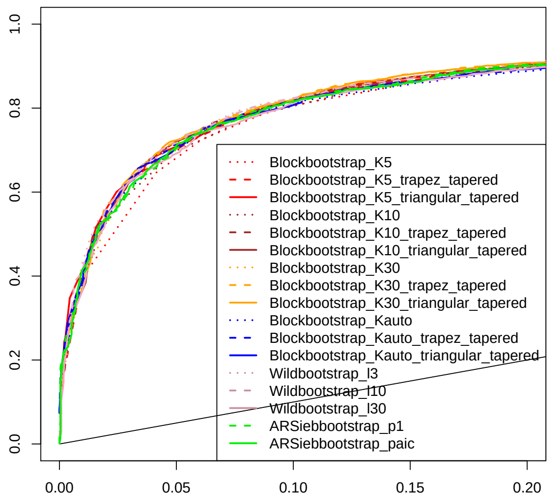
Figure 11.7.: Size-adjusted power of the bootstrap methods with covariance of the components of 0.3 and $k^* = 0.5N$



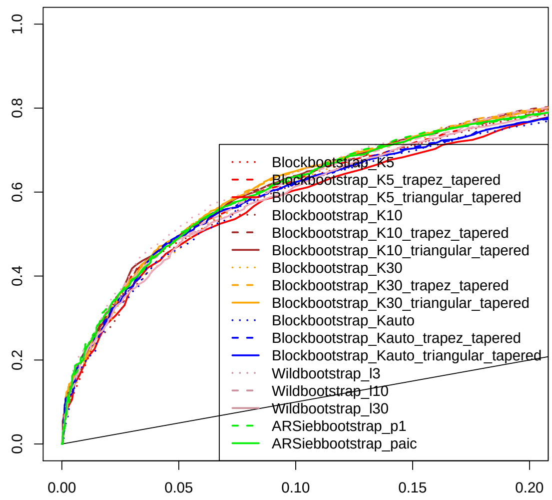
(a) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{2,1}$



(b) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{3,1}$



(c) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{10,1}$



(d) $k^* = 0.5N$ and $\mathbf{\Delta} = \mathbf{\Delta}_{20,1}$

Part III.

Testing in a Multivariate Epidemic Mean Change Model

12. Introduction and an Example of Application

In the third part of this thesis we deal again with the offline change point procedure. We consider a multivariate epidemic mean change model with dependent errors where the mean can change abruptly in each component at a time point t_1 and returns back at a time point t_2. The change points can occur at different time points in each component, however they are not arbitrary but follow a particular pattern generated by the data set.

There is already some literature about time series with epidemic mean changes even in case of multivariate sequences e.g. Aston and Kirch (2012), Muhsal (2013) and Kirch et al. (2015). The first publication is about epidemic changes in functional data, where the observations are represented as functions of compact sets instead of discrete time points. The other publications deal with epidemic mean changes in autoregressive time series. But all of them only allow that the epidemic changes occur at the same time points in all components.

To find the changes in each component by our procedure, the knowledge of the functional relation of the changes depending on the components is necessary. We use the standard multivariate statistic based on the observed multivariate time series which is related to the statistic introduced in Horváth et al. (1999) for the 'at most one change (AMOC)' model as well as to the statistic discussed in Horváth and Hušková (2012) for panel data, where the dimension can be larger than the length of the data set, but only considered in the AMOC model, too. Furthermore, we use the projection statistic where we first project the data to obtain a univariate data set. A detailed introduction of the statistics will follow in the next chapter.

Projected multivariate time series are proposed by Aston and Kirch (2016). They supposed a mean change model, where the shape of the mean changes in each component is generally defined by a function $g(t)$ which is Riemann-integrable. However, the function is independent of the components. Thus the changes occur at the same time points in all components of the multivariate time series. The supposed model includes the usual used change point alternatives as the 'at most one change (AMOC)' model and the epidemic mean change model.

By supposing our model where the changes are allowed to occur at different time points in each component, we consequently obtain a univariate series with a stepwise gradual change in the mean after projecting. The projection statistic, we will use, is related to the statistic introduced in Hušková and Steinebach (2000), who suppose a gradual non-epidemic change and obtained their statistic as likelihood ratio statistic. They developed the extreme-value distribution of the statistic with a given polynomial slope of the change, and in the continuing paper Hušková and Steinebach (2002) derived the limit process of the same statistic.

An application of our setup is to locate a source of gas emission in a huge area. We expect a source of gas emission under the ground of an area, and the aim is to locate the source as near as possible. The source exclaims methane gas in the air in form of a gas plume. Inside the gas plume the methane gas concentration is higher than outside the plume.

We collect the methane gas concentration along a flight path of a plane which flies approximately in constant height during the measurement procedure. If there is a source, the plane flies through the gas plume, and the change points of the methane gas concentration in the collected sequence occur when the plane flies inside and outside the plume. First we have to transform the collected data into a multivariate time series. The detailed transformation method is explained in the real data example in Chapter 18. If we know the shape of the gas plume and the location of the change points, we are able to locate the source. Thus, to locate the source precisely, we first have to detect the changes in each component of the multivariate sequence as precisely as possible. The focus lies in locating the source not on the question if changes have occurred.

Other examples with the focus on locating the change-points, e.g. if it is known that a change has occurred, are data in neoroscience Kirch et al. (2015) or monitoring sports activity Haynes et al. (2016).

Hirst et al. (2013) used a Bayesian inference to locate a source of gas emission. But we will introduce a novel approach to solve this location problem. We will use the same data set for our real data example in Chapter 18 as analysed in Hirst et al. (2013), to compare our results with the results obtained by their method.

In Figure 12.1 the situation is illustrated. The different lines are interpreted as different

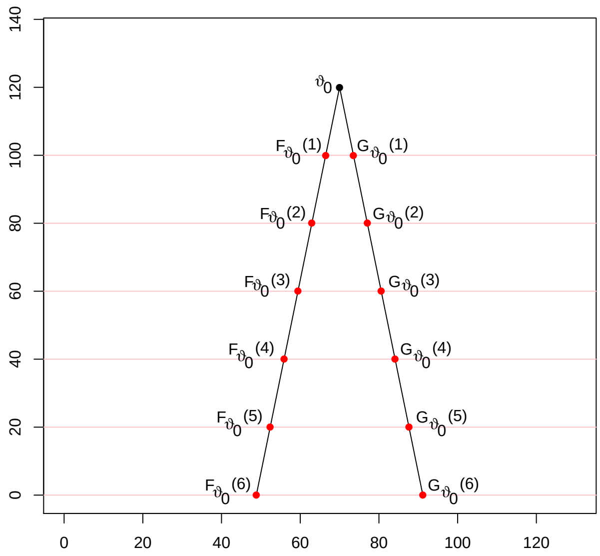


Figure 12.1.: The linear plume.

components, thus we have a multivariate time series. The black lines constitute the gas plume in which we have a higher gas concentration than outside. The change points are the interfaces where the left and right boundary functions of the plume cross the flight path (red points in the figure). In the figure the black lines specify the functional relation between the change points in each component. If we know this functional relationship and identify the change points we can conclude the location of the source (black point, namely ϑ_0, in the figure).

This part of the work is arranged as follows. First, we introduce the testing procedure with the supposed model of the data and a detailed explanation about the statistics.

Then, we develop the limit distributions under the null and the alternative hypothesis for both statistics. Furthermore, we will show the consistency of the estimators resulting from the statistics.

Finally we compare the statistics in a detailed simulation study concerning the empirical size and power where we vary the parameter β, which influences the weight functions (see in the next chapter), the height of the changes and the real location of the source. Especially, we compare the performance of the estimators based on both statistics.

To show that our approach also works well in case of a real data set, we will apply our procedure to the landfill data set which has already been analysed in Hirst et al. (2013).

13. Testing Procedure

13.1. Model of Data

We consider a multivariate data set $\boldsymbol{X}(t) = (X_1(t), \ldots, X_d(t))^T$, where

$$X_i(t) = \mu_i + \widetilde{\Delta}_i \mathbb{1}\left\{F_{\vartheta_0}(i) < \frac{t}{N} \leq G_{\vartheta_0}(i)\right\} + e_i(t), \quad i = 1, \ldots, d; \quad t = 1, \ldots, N. \tag{13.1}$$

The change-points in every component are determined by the tuple $(F_{\vartheta_0}(i), G_{\vartheta_0}(i)),\ i = 1 \ldots, d$ (in rescaled time). The tuple is not arbitrary, but the change-points in the components follow a functional relationship which is parametrized by ϑ_0 and are influenced by given parameters, such as the wind direction and the wind strength as well as the unknown location of the source. Note that the model (13.1) includes the case of a standard multivariate epidemic mean change where the changes occur at the same time point in all components.

We define $\boldsymbol{F}_\vartheta = (F_\vartheta(1), \ldots, F_\vartheta(d))$ and $\boldsymbol{G}_\vartheta = (G_\vartheta(1), \ldots, G_\vartheta(d))$, where ϑ comes from a parameter space Θ and $\vartheta_0 \in \Theta$ is unknown. The mean vector before the changes is defined as $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T$, and $\widetilde{\boldsymbol{\Delta}} = (\widetilde{\Delta}_1, \ldots, \widetilde{\Delta}_d)^T$ is the change vector. The errors $\boldsymbol{e}(t) = (e_1(t), \ldots, e_d(d))^T$ are centered and have variances $\sigma_i^2,\ i = 1, \ldots, d$. They may be dependent but have to satisfy a multivariate functional central limit theorem (FCLT). This is a very weak assumption on the error sequence. For independent errors it follows directly by Donsker's Theorem (see Theorem 16.1 in Billingsley (1968)), e.g. there is literature for weak dependent random variables including mixing sequences Herrndorf (1984a),Herrndorf (1984b) as well as for strongly mixing sequences Doukhan et al. (1994).

In the application of locating the source of gas emission outside of the plume we have the mean $\mu_i,\ i = 1, \ldots, d$, and inside the plume there is a higher level of concentration $\mu_i + \widetilde{\Delta}_i$ (see Figure 13.1). The functions $F_\vartheta(\cdot)$ and $G_\vartheta(\cdot)$ constitute the shape of the plume. Thus they depend only on the known parameters like the components and the wind speed determining the opening angle of the plume. A strong wind leads to a small opening angle, and a weak wind leads to a big opening angle. The unknown location of the source is represented by the parameter ϑ_0. If the opening angle is also unknown, it is also included in the unknown parameter ϑ_0.

We want to test the null hypothesis of no epidemic change in all components against the alternative hypothesis of an epidemic change in at least one component.

The null hypothesis H_0 is	$\widetilde{\Delta}_i = 0\ \forall i = 1, \ldots, d.$
The alternative hypothesis H_1 is	$\widetilde{\Delta}_i \neq 0$ for at least one $i = 1, \ldots, d.$

13.2. Statistics

For testing we use two different types of statistics which we will introduce in this section. The standard multivariate test statistic where we maximize over all possible functional relations between the changes in the components and we sum up the observations between the two possible changes in each component. To use the projection statistic the observed multivariate series is

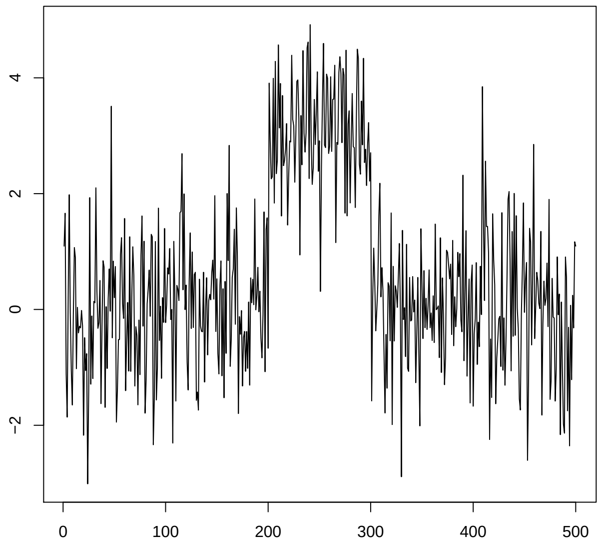


Figure 13.1.: Abrupt epidemic mean change in each component.

first of all projected into a direction not orthogonal to the direction of the change. Hence, we obtain a univariate time series for calculating the test statistic.

13.2.1. Multivariate Statistic

The multivariate test statistic is defined as

$$T^M := \max_{\vartheta \in \Theta} \frac{1}{N} \boldsymbol{S}_\vartheta^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta, \tag{13.2}$$

where

$$\boldsymbol{S}_\vartheta = (S_\vartheta(1), \cdots, S_\vartheta(d))^T, \qquad S_\vartheta(i) = \sum_{t=\lfloor NF_\vartheta(i) \rfloor + 1}^{\lfloor NG_\vartheta(i) \rfloor} \left(X_i(t) - \frac{1}{N} \sum_{j=1}^{N} X_j(t) \right),$$

and

$$\boldsymbol{\Sigma} = \sum_{h \in \mathbb{Z}} \boldsymbol{\Gamma}(h), \quad \boldsymbol{\Gamma}(h) = E\boldsymbol{e}(0)\boldsymbol{e}(h)^T, h \geq 0, \quad \boldsymbol{\Gamma}(h) = -\boldsymbol{\Gamma}(h)^T, h < 0,$$

is the long-run covariance matrix of the errors.

The weighted multivariate test statistic is defined as

$$T^M := \max_{\vartheta \in \Theta} \frac{1}{N} \boldsymbol{S}_\vartheta^{wT} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta^w,$$

where

$$\boldsymbol{S}_\vartheta^w = (S_\vartheta^w(1), \cdots, S_\vartheta^w(d)), \qquad S_\vartheta^w(i) = \sum_{t=\lfloor NF_\vartheta(i) \rfloor + 1}^{\lfloor NG_\vartheta(i) \rfloor} w_{M,i} \left(\frac{\lfloor NF_\vartheta(i) \rfloor}{N}, \frac{\lfloor NG_\vartheta(i) \rfloor}{N} \right) \left(X_i(t) - \frac{1}{N} \sum_{j=1}^{N} X_j(t) \right),$$

$$w_{M,i}(k_1(i), k_2(i)) := \left(\frac{1}{(k_2(i) - k_1(i))(1 - (k_2(i) - k_1(i)))} \right)^{\beta}, \tag{13.3}$$

where $0 \leq k_1(i) < k_2(i) \leq 1,\ i = 1, \cdots, d$ and $\beta \in \left[0, \frac{1}{2}\right]$. Note that if $\beta = 0$, we receive the statistic in (13.2).

13.2.2. Projection Statistic

The data set is first projected into the direction of a vector $\boldsymbol{\Delta}_P$ which is not the null vector. But initially we have to standardize the multivariate observations $\boldsymbol{X}(t)$ and the vector $\boldsymbol{\Delta}_P$ to achieve the best signal-to-noise ratio. Thus with a vector $\boldsymbol{\Delta}_P$ satisfying $\left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \widetilde{\boldsymbol{\Delta}}, \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle \neq 0$ we obtain a univariate time series $Y(t) = \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{X}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle$, where $\boldsymbol{\Delta}_P = (\Delta_{P,1}, \cdots, \Delta_{P,d})$.

The best power gives the projection into the direction of the true change, i.e. if $\boldsymbol{\Delta}_P = c\widetilde{\boldsymbol{\Delta}},\ c \neq 0$, even under misspecification of $\boldsymbol{\Sigma}$ (see more details in Aston and Kirch (2016)).

The structure of the change vector $\widetilde{\boldsymbol{\Delta}}$ in the example of a source of gas emission, which will be used for simulations, is determined by

$$\widetilde{\Delta}_i = \delta h(i), \quad i = 1, \ldots, d,$$

where δ depends on the strength of the source as well as on the distance between the source and $i = 1$. The function h, which is standardized like $\| (h(1), ..., h(d))^T \|_2 = 1$, gives the decay rate of the gas concentration, depending on the distance to the source, which is expressed by the argument i as the component in the function h.

Note that the plume is in reality a 3-D object, and the jump of the gas concentration to a higher level inside the plume is rather a gradual increase. So if the plane only flies into the plume at some distance behind the source, the concentration of the gas first increases and afterwards decreases with the increasing distance to the source. Thus we choose the standard log-normal distribution for the function h. More precisely, the first component is weighted by the value of the log-normal distribution at the x-value 0.09 and the last component at 1.5 , because its values of the log-normal distribution are almost equal. The components between the first and the last one are weighted accordingly to their distance to the source. After standardization we obtain the function h as in the Figure 17.1.

Define $\boldsymbol{\Delta} = (\Delta_1, \cdots, \Delta_d)^T := (h(1), ..., h(d))$. Since δ is a constant it is unimportant for the projection, we project the multivariate observations into the direction of the change direction $\boldsymbol{\Delta}$, thus getting the univariate observations $Y^{\boldsymbol{\Delta}}(t) = \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{X}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta} \right\rangle = \boldsymbol{X}(t)^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta}$. This specified structure of the change vector $\widetilde{\boldsymbol{\Delta}}$ is important for the simulation study in Chapter 17.

For the theory we will use the projection vector $\boldsymbol{\Delta}_P$. The projected univariate observations under the alternative hypothesis have the form

$$Y(t) = \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{e}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle + \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \widetilde{\boldsymbol{\Delta}} \mathbb{1}\left\{ F_{\vartheta_0}(\cdot) < \frac{t}{N} \leq G_{\vartheta_0}(\cdot) \right\}, \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle + \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle$$

and under the null hypothesis

$$Y(t) = \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{e}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle + \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle.$$

We define

$$D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s) := \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \mathbb{1}\{F_\vartheta(\cdot) < s \leq G_\vartheta(\cdot)\}, \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle = \left(\boldsymbol{\Delta}_P \mathbb{1}\{F_\vartheta(\cdot) < s \leq G_\vartheta(\cdot)\}\right)^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta}_P,$$

where $\mathbb{1}\{F_\vartheta(\cdot) < s \leq G_\vartheta(\cdot)\} = (\mathbb{1}\{F_\vartheta(1) < s \leq G_\vartheta(1)\}, \ldots, \mathbb{1}\{F_\vartheta(d) < s \leq G_\vartheta(d)\})^T$. The function $\vartheta \mapsto D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\cdot)$ is continuous and $s \mapsto D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)$ is left-continuous.

The projected errors are

$$e_P(t) = \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{e}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Delta}_P \right\rangle = \boldsymbol{e}(t)^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta}_P.$$

If we set $\boldsymbol{\Sigma}$ equal to the unit matrix which means that the components are uncorrelated and the long-run variances are 1 then

$$D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s) = \langle \boldsymbol{\Delta}_P \mathbb{1}\{F_\vartheta(\cdot) < s \leq G_\vartheta(\cdot)\}, \boldsymbol{\Delta}_P \rangle = \sum_{i=1}^{d} c^2 \widetilde{\Delta}_i^2 \mathbb{1}\{F_\vartheta(i) < s \leq G_\vartheta(i)\}.$$

Now it is obvious that there is no abrupt epidemic mean change anymore, the projected data set $Y(t)$ has a gradual epidemic mean change (see Figure 13.2).

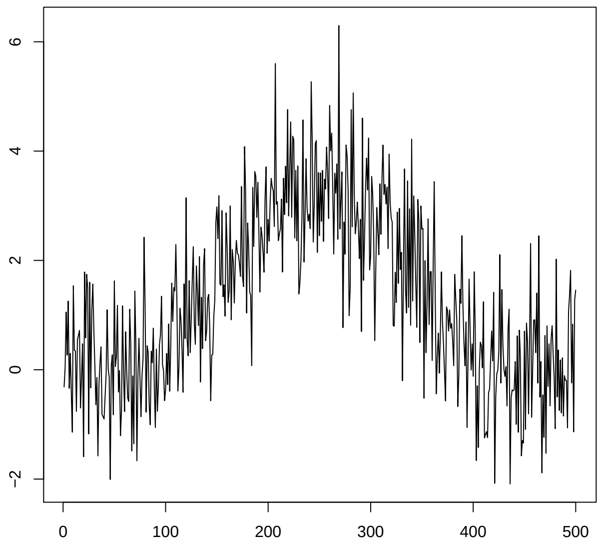


Figure 13.2.: Gradual epidemic mean change.

In case of using the projection statistic, we first have to project the standardized data into the direction of the standardized change vector $\boldsymbol{\Delta}_P$, then we calculate the statistic which is defined as

$$T^P := \frac{1}{\sqrt{N}} \frac{1}{\sigma} \max_{\vartheta \in \Theta} \left(\frac{\left| \sum_{t=1}^{N} \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right) - \frac{1}{N} \sum_{l=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{l}{N}\right) \right) Y(t) \right|}{w_P\left(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta\right)} \right)$$

$$= \frac{1}{\sqrt{N}} \frac{1}{\sigma} \max_{\vartheta \in \Theta} \left(\frac{\left| \sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{t}{N}\right) \left(Y(t) - \overline{Y}_N\right) \right|}{w_P \left(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta\right)} \right),$$

where

$$w_P \left(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta\right) = \left(\frac{1}{N} \sum_{t=1}^{N} \left(D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{t}{N}\right) - \frac{1}{N} \sum_{l=1}^{N} D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{l}{N}\right) \right)^2 \right)^\beta . \qquad (13.4)$$

where $\beta \in \left[0, \frac{1}{2}\right]$ and σ^2 is the long-run variance with $\sigma^2 = \sum_{h \in \mathbb{Z}} \text{Cov}\left(e_P(0), e_P(h)\right)$.

The parameter β figuring in the weight function influences which region of the searching area (the area in which we search for a source) is preferred by the testing method. The details will be discussed in the extensive simulation study in Chapter 17.

Particularly, sources on the upper bound of the searching area can be found most easily because the plume is relatively wide, which results in a long period of change in each component.

13.3. Estimators

We introduce the estimators based on the mentioned statistics. In Chapter 16 we will prove the consistency of the estimators based on both statistics.

The estimator of the unknown parameter ϑ_0 based on the multivariate statistic is defined as

$$\widehat{\vartheta}_M = \arg\max_{\vartheta \in \Theta} \boldsymbol{S}_\vartheta^{wT} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta^{w}.$$

For the estimator based on the projection statistic we have to choose β equal to $\frac{1}{2}$ because only then the estimator is consistent, which will be shown in Theorem 16.2 of Chapter 16. Consequently, the estimator is defined as

$$\widehat{\vartheta}_P = \arg\max_{\vartheta \in \Theta} \frac{\left| \sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{t}{N}\right) \left(Y(t) - \overline{Y}_N\right) \right|}{\left(\frac{1}{N} \sum_{t=1}^{N} \left(D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{t}{N}\right) - \frac{1}{N} \sum_{l=1}^{N} D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{l}{N}\right) \right)^2 \right)^{\frac{1}{2}}}.$$

14. Asymptotics under the Null Hypothesis

To decide between rejection or acceptance of the null hypothesis we need the limit distribution of the statistics to use its quantiles as critical values for controlling the type-I-error.

14.1. Null Asymptotics of the Multivariate Statistic

In the following theorem we develop the limit process of the multivariate statistic. Each component of S_ϑ sums up the observations between two time points. Usually if one uses an epidemic mean change model, where the changes occur at the same time points in all components, one could directly use the given FCLT of the errors by building the difference between the cumulative sum till the later and the earlier time point. Caused by the different change points in each component, we cannot use the FCLT of the errors directly, but the statistic can be built by a suitable function using the partial sum process of the errors as argument. Afterwards we are able to use the FCLT in combination with proving that this suitable function is continuous and conclude the proof using the continuous mapping theorem.

Theorem 14.1

Let the errors $\boldsymbol{e}(t) = (e_1(t), \ldots, e_d(t))^T, t = 1, \ldots, N$, be a time series which fulfills a multivariate functional central limit theorem towards a Wiener process with covariance matrix $\boldsymbol{\Sigma}$. Then, under the null hypothesis,

$$\max_{\vartheta \in \Theta} \frac{1}{N} \boldsymbol{S}_\vartheta^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta \xrightarrow{D} \sup_{\vartheta \in \Theta} \sum_{i=1}^{d} (B_i(G_\vartheta(i)) - B_i(F_\vartheta(i)))^2,$$

where $B_i,\ i = 1, \ldots, d$, are independent standard Brownian bridges. The assertion still holds if we replace $\boldsymbol{\Sigma}$ by a consistent estimator.

Proof. In the following lines let $||\boldsymbol{x}||_{\boldsymbol{\Sigma}}^2 = \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}$.

Under the null hypothesis the test statistic can be written as

$$\max_{\vartheta \in \Theta} \frac{1}{N} \boldsymbol{S}_\vartheta^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta = \max_{\vartheta \in \Theta} \frac{1}{N} \left\| \begin{pmatrix} \sum\limits_{t=\lfloor NF_\vartheta(1) \rfloor + 1}^{\lfloor NG_\vartheta(1) \rfloor} (e_1(t) - \bar{e}_1) \\ \vdots \\ \sum\limits_{t=\lfloor NF_\vartheta(d) \rfloor + 1}^{\lfloor NG_\vartheta(d) \rfloor} (e_d(t) - \bar{e}_d) \end{pmatrix} \right\|_{\boldsymbol{\Sigma}}^2 .$$

We do a further transformation in order to easily understand the main idea of the proof later

$$\begin{pmatrix} \sum_{t=\lfloor NF_\vartheta(1)\rfloor+1}^{\lfloor NG_\vartheta(1)\rfloor} (e_1(t)-\overline{e}_1) \\ \vdots \\ \sum_{t=\lfloor NF_\vartheta(d)\rfloor+1}^{\lfloor NG_\vartheta(1)\rfloor} (e_d(t)-\overline{e}_d) \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{\lfloor NG_\vartheta(1)\rfloor} (e_1(t)-\overline{e}_1) \\ \vdots \\ \sum_{t=1}^{\lfloor NG_\vartheta(d)\rfloor} (e_d(t)-\overline{e}_1) \end{pmatrix} - \begin{pmatrix} \sum_{t=1}^{\lfloor NF_\vartheta(1)\rfloor} (e_1(t)-\overline{e}_1) \\ \vdots \\ \sum_{t=1}^{\lfloor NF_\vartheta(d)\rfloor} (e_d(t)-\overline{e}_1) \end{pmatrix} \quad (14.1)$$

First we define the projection

$$P_{j,t} : D^d[0,1] \to \mathbb{R}^d, \qquad P_{j,t}(\boldsymbol{x}(\cdot)) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_j(t) \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where $\boldsymbol{x}(\cdot) = (x_1(\cdot), \dots, x_d(\cdot))$. With the help of this projection we construct another projection $P_{t_1,\dots,t_d}$ which gives a vector of the components of the vector $\boldsymbol{x}(\cdot)$ for different arguments $t_1, \dots, t_d$. For the projection

$$P_{t_1,\dots,t_d} := P_{1,t_1} + P_{2,t_2} + \cdots + P_{d,t_d},$$

we have

$$P_{t_1,\dots,t_d} : D^d[0,1] \to \mathbb{R}^d,$$

$$P_{t_1,\dots,t_d}(\boldsymbol{x}(\cdot)) = P_{1,t_1}(\boldsymbol{x}(\cdot)) + P_{2,t_2}(\boldsymbol{x}(\cdot)) + \cdots + P_{d,t_d}(\boldsymbol{x}(\cdot)) = \begin{pmatrix} x_1(t_1) \\ x_2(t_2) \\ \vdots \\ x_d(t_d) \end{pmatrix}.$$

Now we define a function $\widetilde{H}$ from $D^d[0,1]$ to $\mathbb{R}^d$ and a function H from $D^d[0,1]$ to $\mathbb{R}$ using the projection $P_{t_1,\dots,t_d}$:

$$\widetilde{H}(\boldsymbol{x}(\cdot)) := P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\boldsymbol{x}(\cdot)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\boldsymbol{x}(\cdot)),$$

$$H(\boldsymbol{x}(\cdot)) := \max_{\vartheta\in\Theta} \left\| \widetilde{H}(\boldsymbol{x}(\cdot)) - \widetilde{H}(\mathrm{id}(\cdot)) P_{1,\dots,1}(\boldsymbol{x}(\cdot)) \right\|_{\boldsymbol{\Sigma}}^2,$$

where $\widetilde{H}(\mathrm{id}(\cdot))$ and $P_{1,\dots,1}(\boldsymbol{x}(\cdot))$ are multiplied componentwise.

If we choose the partial sum process of the errors as $\boldsymbol{x}(\cdot)$, i.e.

$$\boldsymbol{x}\left(\frac{\lfloor Nt\rfloor}{N}\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt\rfloor} \begin{pmatrix} e_1(i) \\ \vdots \\ e_d(i) \end{pmatrix},$$

which is a function in $D^d[0,1]$, then we obtain $H(\boldsymbol{x}(\cdot)) = \max\limits_{\vartheta\in\Theta} \frac{1}{N}\boldsymbol{S}_\vartheta^T\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_\vartheta$ (compare equation (14.1)).

The last step of the proof is to show that the function H is continuous because then the assertion follows from the continuous mapping theorem and the multivariate functional central limit theorem, since

$$\begin{aligned}
&H\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{\lfloor Nt\rfloor}\begin{pmatrix} e_1(i)\\ \vdots\\ e_d(i)\end{pmatrix}\right) \xrightarrow{D} H\left(\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} W_1(t)\\ \vdots\\ W_d(t)\end{pmatrix}\right)\\
&= \max_{\vartheta\in\Theta}\left\| P_{G_\vartheta(1),\dots,G_\vartheta(d)}\left(\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} W_1(t)\\ \vdots\\ W_d(t)\end{pmatrix}\right) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}\left(\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} W_1(t)\\ \vdots\\ W_d(t)\end{pmatrix}\right)\right.\\
&\quad \left. -\left(P_{G_\vartheta(1),\dots,G_\vartheta(d)}\left(\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} t\\ \vdots\\ t\end{pmatrix}\right) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}\left(\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} t\\ \vdots\\ t\end{pmatrix}\right)\right) P_{1,\dots,1}\left(\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} W_1(t)\\ \vdots\\ W_d(t)\end{pmatrix}\right)\right\|_{\boldsymbol{\Sigma}}^2\\
&= \max_{\vartheta\in\Theta}\left\|\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} W_1(G_\vartheta(1)) - W_1(F_\vartheta(1)) - (G_\vartheta(1)-F_\vartheta(1))W_1(1)\\ \vdots\\ W_d(G_\vartheta(d)) - W_d(F_\vartheta(d)) - (G_\vartheta(d)-F_\vartheta(d))W_d(1)\end{pmatrix}\right\|_{\boldsymbol{\Sigma}}^2\\
&= \max_{\vartheta\in\Theta}\left\|\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} W_1(G_\vartheta(1)) - G_\vartheta(1)W_1(1) - (W_1(F_\vartheta(1)) - F_\vartheta(1)W_1(1))\\ \vdots\\ W_d(G_\vartheta(d)) - G_\vartheta(d)W_d(1) - (W_d(F_\vartheta(d)) - F_\vartheta(d)W_d(1))\end{pmatrix}\right\|_{\boldsymbol{\Sigma}}^2\\
&= \max_{\vartheta\in\Theta}\left\|\boldsymbol{\Sigma}^{\frac12}\begin{pmatrix} B_1(G_\vartheta(1)) - B_1(F_\vartheta(1))\\ \vdots\\ B_d(G_\vartheta(d)) - B_d(F_\vartheta(d))\end{pmatrix}\right\|_{\boldsymbol{\Sigma}}^2\\
&= \max_{\vartheta\in\Theta}\left\|\begin{pmatrix} B_1(G_\vartheta(1)) - B_1(F_\vartheta(1))\\ \vdots\\ B_d(G_\vartheta(d)) - B_d(F_\vartheta(d))\end{pmatrix}\right\|^2 = \sup_{\vartheta\in\Theta}\sum_{i=1}^d (B_i(G_\vartheta(i)) - B_i(F_\vartheta(i)))^2.
\end{aligned}$$

To prove the continuity of H we define the vectors

$$\boldsymbol{x}(t) = \begin{pmatrix} x_1(t)\\ \vdots\\ x_d(t)\end{pmatrix}, \qquad \boldsymbol{y}(t) = \begin{pmatrix} y_1(t)\\ \vdots\\ y_d(t)\end{pmatrix}, \quad t\in[0,1].$$

For an arbitrary $\epsilon > 0$ we have

$$\begin{aligned}
&|H(\boldsymbol{x}(t)) - H(\boldsymbol{y}(t))|\\
&= \Big|\max_{\vartheta\in\Theta}\big\|P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\boldsymbol{x}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\boldsymbol{x}(t))\\
&\quad - \left(P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\mathrm{id}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\mathrm{id}(t))\right) P_{1,\dots,1}(\boldsymbol{x}(t))\big\|_{\boldsymbol{\Sigma}}^2\\
&\quad - \max_{\vartheta\in\Theta}\big\|P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\boldsymbol{y}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\boldsymbol{y}(t))
\end{aligned}$$

$$
\begin{aligned}
&\quad - \left(P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\mathrm{id}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\mathrm{id}(t))\right) P_{1,\dots,1}(\boldsymbol{y}(t))\big\|^2_{\boldsymbol{\Sigma}}\Big| \\
&\le \max_{\vartheta\in\Theta} \big\| P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\boldsymbol{x}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\boldsymbol{x}(t)) \\
&\quad - \left(P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\mathrm{id}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\mathrm{id}(t))\right) P_{1,\dots,1}(\boldsymbol{x}(t)) \\
&\quad - \left(P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\boldsymbol{y}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\boldsymbol{y}(t))\right. \\
&\quad - \left(P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\mathrm{id}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\mathrm{id}(t))\right) P_{1,\dots,1}(\boldsymbol{y}(t))\big)\big\|^2_{\boldsymbol{\Sigma}} \\
&\le \max_{\vartheta\in\Theta} \big\| P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\boldsymbol{x}(t)) - P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\boldsymbol{y}(t))\big\|^2_{\boldsymbol{\Sigma}} \\
&\quad + \max_{\vartheta\in\Theta} \big\| \left(P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\boldsymbol{x}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\boldsymbol{y}(t))\right)\big\|^2_{\boldsymbol{\Sigma}} \\
&\quad + \max_{\vartheta\in\Theta} \big\| \left(P_{G_\vartheta(1),\dots,G_\vartheta(d)}(\mathrm{id}(t)) - P_{F_\vartheta(1),\dots,F_\vartheta(d)}(\mathrm{id}(t))\right)\left(P_{1,\dots,1}(\boldsymbol{x}(t)) - P_{1,\dots,1}(\boldsymbol{y}(t))\right)\big\|^2_{\boldsymbol{\Sigma}} \\
&= \max_{\vartheta\in\Theta} \left\| \begin{pmatrix} x_1(G_\vartheta(1)) - y_1(G_\vartheta(1)) \\ \vdots \\ x_d(G_\vartheta(d)) - y_d(G_\vartheta(d)) \end{pmatrix} \right\|^2_{\boldsymbol{\Sigma}} + \max_{\vartheta\in\Theta} \left\| \begin{pmatrix} x_1(F_\vartheta(1)) - y_1(F_\vartheta(1)) \\ \vdots \\ x_d(F_\vartheta(d)) - y_d(F_\vartheta(d)) \end{pmatrix} \right\|^2_{\boldsymbol{\Sigma}} \\
&\quad + \max_{\vartheta\in\Theta} \left\| \begin{pmatrix} G_\vartheta(1) - F_\vartheta(1) \\ \vdots \\ G_\vartheta(d) - F_\vartheta(d) \end{pmatrix} \begin{pmatrix} x_1(1) - y_1(1) \\ \vdots \\ x_d(1) - y_d(1) \end{pmatrix} \right\|^2_{\boldsymbol{\Sigma}} \le \epsilon,
\end{aligned}
$$

if there is a $\delta = \delta(\epsilon)$ with $||\boldsymbol{x}(t) - \boldsymbol{y}(t)||_\infty = \max_{\substack{i=1,\dots,d \\ 0\le t\le 1}} |x_i(t) - y_i(t)| < \delta(\epsilon)$, since $|G_\vartheta(i)| \le 1$ and $|F_\vartheta(i)| \le 1$.

We can replace $\boldsymbol{\Sigma}$ by an consistent estimator $\widehat{\boldsymbol{\Sigma}}_N$ because of

$$
\left|\widehat{\boldsymbol{\Sigma}}_N - \boldsymbol{\Sigma}\right| = o_P(1).
$$

Hence it holds

$$
\begin{aligned}
&\left| \frac{1}{N} \max_{\vartheta\in\theta} \|\boldsymbol{S}_\vartheta\|^2_{\widehat{\boldsymbol{\Sigma}}} - \frac{1}{N} \max_{\vartheta\in\theta} \|\boldsymbol{S}_\vartheta\|^2_{\boldsymbol{\Sigma}} \right| = \left| \frac{1}{N} \max_{\vartheta\in\theta} \left\| \widehat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{S}_\vartheta \right\|^2 - \frac{1}{N} \max_{\vartheta\in\theta} \left\| \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{S}_\vartheta \right\|^2 \right| \\
&\le \frac{1}{N} \max_{\vartheta\in\theta} \left\| \widehat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{S}_\vartheta - \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{S}_\vartheta \right\|^2 = \frac{1}{N} \max_{\vartheta\in\theta} \left\| \left(\widehat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}} - \mathrm{Id} \right) \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{S}_\vartheta \right\|^2 \\
&\le \left\| \widehat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}} - \mathrm{Id} \right\|^2 \frac{1}{N} \max_{\vartheta\in\theta} \left\| \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{S}_\vartheta \right\|^2 = \left\| \widehat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}} - \mathrm{Id} \right\|^2 \max_{\vartheta\in\theta} \frac{1}{N} \boldsymbol{S}_\vartheta^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta = o_P(1).
\end{aligned}
$$

□

Remark 14.1. If we use an estimator $\widehat{\boldsymbol{\Sigma}}_N$ for the long-run covariance matrix satisfying $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, where $\boldsymbol{\Sigma}_A$ is a positive definite matrix, then we obtain the following limit distribution for the multivariate statistic

$$
\max_{\vartheta\in\Theta} \begin{pmatrix} B_1(G_\vartheta(1)) - B_1(F_\vartheta(1)) \\ \vdots \\ B_d(G_\vartheta(d)) - B_d(F_\vartheta(d)) \end{pmatrix}^T \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\Sigma}_A^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}} \begin{pmatrix} B_1(G_\vartheta(1)) - B_1(F_\vartheta(1)) \\ \vdots \\ B_d(G_\vartheta(d)) - B_d(F_\vartheta(d)) \end{pmatrix}.
$$

It depends on the true long-run covariance matrix, thus in this case we need Bootstrap methods.

Remark 14.2. Note that in Theorem 14.1 we use the weight function in (13.3) with $\beta = 0$.

We suppose the assumptions of Theorem 14.1 and use the weight function with $0 < \beta \leq \frac{1}{2}$. Additionally this weight function is assumed to be bounded on Θ, which means that there is a $\epsilon > 0$ such that $\epsilon < G_\vartheta(i) - F_\vartheta(i) < 1 - \epsilon,\ \forall i, \forall \vartheta$. Then if $\mathbf{\Sigma}$ is a diagonal matrix, we get

$$\begin{aligned} &\max_{\vartheta \in \Theta} \frac{1}{N} \boldsymbol{S}_\vartheta^{wT} \mathbf{\Sigma}^{-1} \boldsymbol{S}_\vartheta^w \\ &\xrightarrow{D} \sup_{\vartheta \in \Theta} \sum_{i=1}^{d} w_{M,i}^2(F_\vartheta(i), G_\vartheta(i))(B_i(G_\vartheta(i)) - B_i(F_\vartheta(i)))^2. \end{aligned} \tag{14.2}$$

The way to prove the result is analogous to the way of the case if $\beta = 0$. Therefore we choose the function $\boldsymbol{x}(\cdot)$ as

$$\boldsymbol{x}(t) = \frac{1}{\sqrt{N}} \begin{pmatrix} \boldsymbol{w}_{M,1}\left(\frac{\lfloor NF_\vartheta(1) \rfloor}{N}, \frac{\lfloor NG_\vartheta(1) \rfloor}{N}\right) \\ \vdots \\ \boldsymbol{w}_{M,d}\left(\frac{\lfloor NF_\vartheta(d) \rfloor}{N}, \frac{\lfloor NG_\vartheta(d) \rfloor}{N}\right) \end{pmatrix} \sum_{i=1}^{\lfloor Nt \rfloor} \begin{pmatrix} e_1(i) \\ \vdots \\ e_d(i) \end{pmatrix}.$$

Then we also obtain with the same function H as in the proof of Theorem 14.1 and by its continuity and a continuous transformation of the FCLT for the errors

$$H\left(\frac{1}{\sqrt{N}} \begin{pmatrix} \boldsymbol{w}_{M,1}\left(\frac{\lfloor NF_\vartheta(1) \rfloor}{N}, \frac{\lfloor NG_\vartheta(1) \rfloor}{N}\right) \\ \vdots \\ \boldsymbol{w}_{M,d}\left(\frac{\lfloor NF_\vartheta(d) \rfloor}{N}, \frac{\lfloor NG_\vartheta(d) \rfloor}{N}\right) \end{pmatrix} \sum_{i=1}^{\lfloor Nt \rfloor} \begin{pmatrix} e_1(i) \\ \vdots \\ e_d(i) \end{pmatrix}\right) \tag{14.3}$$

$$\xrightarrow{D} H\left(\frac{1}{\sqrt{N}} \begin{pmatrix} \boldsymbol{w}_{M,1}\left(F_\vartheta(1), G_\vartheta(1)\right) \\ \vdots \\ \boldsymbol{w}_{M,d}\left(F_\vartheta(d), G_\vartheta(d)\right) \end{pmatrix} \sum_{i=1}^{\lfloor Nt \rfloor} \begin{pmatrix} e_1(i) \\ \vdots \\ e_d(i) \end{pmatrix}\right). \tag{14.4}$$

Note that by choosing the function H as in the proof of Theorem 14.1 the term in (14.3) is equal to the weighted multivariate statistic and the term in (14.4) is equal to the sum of the weighted Brownian bridges. So we get (14.2).

If we use a linear plume where the opening angle is known as in the first part of the simulation study in Chapter 17, the weight function is bounded on Θ, if the opening angle is chosen such that $\epsilon < G_\vartheta(i) - F_\vartheta(i) < 1 - \epsilon,\ \forall i = 1, \ldots, d, \quad \forall \vartheta \in \Theta$.

Additionally maximizing over the opening angle as in the second part of the simulations Chapter 17 and in the data example in Chapter 18, leads to the weight function being bounded on Θ only if there is a minimum and maximum bound on the opening angle. The minimum bound could be for example $1°$ and the maximum bound $89°$. Only the condition above has to be satisfied for every opening angle between the upper and lower bound.

Particularly, plumes of which only a part is inside the flight path can also be allowed.

14.2. Null Asymptotics of the Projection Statistic

The next Theorem gives the limit process of the projection statistic.

Since the projection statistic consists the projected errors, which are equal to the projected sequence $Y(t)$ under H_0, we first have to develop a FCLT for the projected errors. Then the statistic can be represented as a function of the partial sum process of the projected errors. And then we can conclude in the same way as in case of the multivariate statistic.

Theorem 14.2

Let the errors $\boldsymbol{e}(t) = (e_1(t), \ldots, e_d(t))^T, t = 1, \ldots, N$, be a time series which fulfills a multivariate functional central limit theorem towards a Wiener process with covariance matrix $\boldsymbol{\Sigma}$. Then, under the null hypothesis, we have

$$\frac{1}{\sqrt{N}} \frac{1}{\sigma} \max_{\vartheta \in \Theta} \left| \sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{t}{N} \right) \left(Y(t) - \overline{Y}_N \right) \right| \xrightarrow{D} \sup_{\vartheta \in \Theta} \left| \sum_{s \in \mathcal{M}_\vartheta} \left(D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s+) - D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s-) \right) B(s) \right|,$$

where $\boldsymbol{F}_\vartheta = (F_\vartheta(1), \ldots, F_\vartheta(d))$ and $\boldsymbol{G}_\vartheta = (G_\vartheta(1), \ldots, G_\vartheta(d))$, $\mathcal{M}_\vartheta = \{0 < s < 1 : D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s+) - D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s) \neq 0\}$ and $\{B(\cdot)\}$ is a standard Brownian bridge.
The assertion still holds true if we replace σ and $\boldsymbol{\Sigma}$ by consistent estimators $\widehat{\sigma}_N$ and $\widehat{\boldsymbol{\Sigma}}_N$.

Proof. With partial summation (confer Knopp (1996)) we derive the equality

$$\begin{aligned} &\frac{1}{\sqrt{N}} \sum_{i=1}^{N} D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{i}{N} \right) \left(Y(i) - \overline{Y}_N \right) \\ &= - \sum_{i=1}^{N-1} \left(D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{i+1}{N} \right) - D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{i}{N} \right) \right) \frac{1}{\sqrt{N}} \sum_{j=1}^{i} \left(Y(j) - \overline{Y}_N \right). \end{aligned} \tag{14.5}$$

First notice that $D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(\cdot)$ is step-wise constant and has a finite number of points of discontinuity, namely the elements of $\mathcal{M}_\vartheta$. Thus with the left-continuity of $D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(\cdot)$, we have

$$\begin{aligned} &- \sum_{i=1}^{N-1} \left(D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{i+1}{N} \right) - D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{i}{N} \right) \right) g \left(\frac{i}{N} \right) \\ &= - \sum_{s \in \mathcal{M}_\vartheta} \left(D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s+) - \boldsymbol{D}_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s) \right) g \left(\frac{\lfloor Ns \rfloor}{N} \right), \end{aligned}$$

for all functions $g : [0,1] \to \mathbb{R}$ and $\mathcal{M}_\vartheta = \{0 < s < 1 : D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s+) - D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}(s) \neq 0\}$.

Consequently, equation (14.5) can be rewritten as follows:

$$\begin{aligned} &- \sum_{i=1}^{N-1} \left(D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{i+1}{N} \right) - D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta} \left(\frac{i}{N} \right) \right) \frac{1}{\sqrt{N}} \sum_{j=1}^{i} \left(Y(j) - \overline{Y}_N \right) \\ &= - \sum_{s \in \mathcal{M}_\vartheta} \left(D_\vartheta(s+) - \boldsymbol{D}_\vartheta(s) \right) \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} \left(Y(j) - \overline{Y}_N \right). \end{aligned} \tag{14.6}$$

Next we have a look at the term including the univariate observations $Y(t)$ which can be replaced by the projected errors under H_0, thus it yields

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} \left(Y(j) - \overline{Y}_N \right) = \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} \left(e_P(j) - \overline{e}_{P,N} \right).$$

Thus if a functional central limit theorem holds for the projected error sequence, we can conclude the proof.

A simple calculation gives

$$\begin{aligned}\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor} e_P(j) &= \frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{e}(j), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P\right\rangle \\ &= \frac{1}{\sigma}\left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\boldsymbol{e}(j), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P\right\rangle = \frac{1}{\sigma}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\boldsymbol{e}(j)\right)^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P.\end{aligned}$$

We define a vector $x(t) = (x_1(t), \dots, x_d(t))^T$ which is a function from $D[0,1]$ to $\mathbb{R}^d$ and a function $H : D[0,1] \to \mathbb{R}$ with $H(x(\cdot)) = \frac{1}{\sigma}x^T(\cdot)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P$. It is easy to see that

$$\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor} e_P(j) = H\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\boldsymbol{e}(j)\right).$$

For an arbitrary $\epsilon > 0$ as well as $\boldsymbol{x}(t) = (x_1(t), \dots, x_d(t))^T$ and $\boldsymbol{y}(t) = (y_1(t), \dots, y_d(t))^T$ we have

$$|H(x(t)) - H(y(t))| = \left|\frac{1}{\sigma}x^T(t)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P - \frac{1}{\sigma}y^T(t)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P\right| = \left|\frac{1}{\sigma}\left(x^T(t) - y^T(t)\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P\right| < \epsilon,$$

if there is a $\delta = \delta(\epsilon)$ with $\|\boldsymbol{x}(t) - \boldsymbol{y}(t)\|_\infty = \max_{\substack{i=1,\dots,d\\ 0\le t\le 1}} |x_i(t) - y_i(t)| < \delta(\epsilon)$.

Since the function H is continuous and the multivariate errors fulfill a FCLT, it holds a FCLT for the univariate projected errors by applying the continuous mapping theorem

$$\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor} e_P(j) = H\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\boldsymbol{e}(j)\right) \overset{D[0,1]}{\longrightarrow} H(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{W}(s)) = \frac{1}{\sigma}\boldsymbol{W}^T(s)\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P,$$

where $\boldsymbol{W}(s)$ is a multivariate standard Wiener process. Now we check the variance of $\boldsymbol{W}(s)^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P$:

$$\begin{aligned}&\operatorname{Var}(\boldsymbol{W}(s)^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P) = \operatorname{Var}\left((\boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{W}(s))^T\right)\\ &= s\boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-\frac{1}{2}})^T = s\boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P\\ &= s\boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P = s\sigma^2,\end{aligned}$$

with

$$\begin{aligned}\sigma^2 &= \sum_{h\in\mathbb{Z}}\operatorname{Cov}\left(e_P(0), e_P(h)\right) \qquad (14.7)\\ &= \sum_{h\in\mathbb{Z}}\operatorname{Cov}\left(\left\langle\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{e}(0), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P\right\rangle, \left\langle\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{e}(h), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P\right\rangle\right)\\ &= \sum_{h\in\mathbb{Z}}\operatorname{Cov}\left(\boldsymbol{e}(0)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P, \boldsymbol{e}(h)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P\right)\\ &= \boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-1}\operatorname{Cov}\left(\boldsymbol{e}(0), \boldsymbol{e}(h)\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P\\ &= \boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P\\ &= \boldsymbol{\Delta}_P^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P.\end{aligned}$$

We conclude that the variance of $\frac{1}{\sigma}\boldsymbol{W}^T(s)\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P$ is s. So we get

$$\left\{\frac{1}{\sigma}\boldsymbol{W}^T(s)\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P : 0 \le s \le 1\right\} \overset{D}{=} \{W(s) : 0 \le s \le 1\},$$

where $W(s)$ is a standard Wiener process, and thus a functional central limit theorem also holds for the projected errors:

$$\left\{\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor} e_P(j) : 0 \le s \le 1\right\} \overset{D[0,1]}{\longrightarrow} \{W(s) : 0 \le s \le 1\}.$$

Hence the pointwise convergence in distribution follows

$$\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor} e_P(j) \overset{D}{\longrightarrow} W(s) \quad \forall\ 0 \le s \le 1.$$

To finish we define the function $V : \mathbb{R} \to \mathbb{R}$ and $x : [0,1] \to \mathbb{R}$ according to

$$V\left(x\left(\frac{\lfloor Ns\rfloor}{N}\right)\right) = \max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - \boldsymbol{D}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\left(x\left(\frac{\lfloor Ns\rfloor}{N}\right) - \mathrm{id}\left(\frac{\lfloor Ns\rfloor}{N}\right)x(1)\right)\right|.$$

If $x(s) = \frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor} e_P(j),\ 0 \le s \le 1$, then $V(x(s))$ is equal to equation (14.6). Let $x(t)$ and $y(t)$ be functions from $[0,1]$ to $\mathbb{R}$, then the function $V(\cdot)$ is continuous because of we have for an arbitrary $\epsilon > 0$ and with $t = \frac{\lfloor Ns\rfloor}{N}$

$$\begin{aligned}
&|V(x(t)) - V(y(t))| \\
&= \Bigg|\max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - \boldsymbol{D}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\left(x(t) - \mathrm{id}(t)x(1)\right)\right| \\
&\quad - \max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - \boldsymbol{D}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\left(y(t) - \mathrm{id}(t)y(1)\right)\right|\Bigg| \\
&\le \max_{\vartheta\in\Theta}\Bigg|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - \boldsymbol{D}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\left(x(t) - \mathrm{id}(t)x(1)\right) \\
&\quad - \sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - \boldsymbol{D}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\left(y(t) - \mathrm{id}(t)y(1)\right)\Bigg| \\
&= \max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - \boldsymbol{D}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\left(x(t) - y(t) - \mathrm{id}(t)(x(1) - y(1))\right)\right| < \epsilon,
\end{aligned}$$

if there is a $\delta = \delta(\epsilon)$ with $||\boldsymbol{x}(t) - \boldsymbol{y}(t)||_\infty = \max_{\substack{i=1,\dots,d \\ 0\le t\le 1}} |x_i(t) - y_i(t)| < \delta(\epsilon)$.

Consequently we get

$$\max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s-)\right)\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(e_P(j)-\bar{e}_{P,N}\right)\right|$$
$$\xrightarrow{D}\max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s-)\right)B(s)\right|,\quad N\to\infty,$$

where $\{B(s)\}$ is a standard Brownian bridge. For consistent estimators $\widehat{\sigma}$ of σ and $\widehat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ and with the functional limit theorem, we have

$$\Bigg|\max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(e_P(j)-\bar{e}_{P,N}\right)\right|$$
$$-\max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}_\vartheta}\left(D^{\widehat{\boldsymbol{\Sigma}}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D^{\widehat{\boldsymbol{\Sigma}}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P\right)\right|\Bigg|$$

$$\leq\max_{\vartheta\in\Theta}\Bigg|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(e_P(j)-\bar{e}_{P,N}\right)$$
$$-\sum_{s\in\mathcal{M}_\vartheta}\left(D^{\widehat{\boldsymbol{\Sigma}}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D^{\widehat{\boldsymbol{\Sigma}}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P\right)\Bigg|$$
$$=\max_{\vartheta\in\Theta}\Bigg|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(\boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P\right)$$

$$-\sum_{s\in\mathcal{M}_\vartheta}\left(D^{\widehat{\boldsymbol{\Sigma}}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D^{\widehat{\boldsymbol{\Sigma}}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P\right)\Bigg|$$
$$=\max_{\vartheta\in\Theta}\Bigg|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\frac{1}{\sigma}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(\boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P\right)$$

$$-\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)+o_P(1)\right)\frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P\right)\Bigg|$$

$$=\max_{\vartheta\in\Theta}\Bigg|\sum_{s\in\mathcal{M}_\vartheta}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)\Bigg(\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(\frac{1}{\sigma}\boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P-\frac{1}{\widehat{\sigma}}\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P\right)$$
$$-\frac{\lfloor Ns\rfloor}{N}\sum_{i=1}^{N}\left(\frac{1}{\sigma}\boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P-\frac{1}{\widehat{\sigma}}\boldsymbol{e}^T(j)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P\right)\Bigg)\Bigg|+O_P(1)$$

$$= \max_{\vartheta\in\Theta} \left| \sum_{s\in\mathcal{M}_\vartheta} \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right) \right.$$

$$\left. \left(\frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns\rfloor} \boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1} \left(\frac{1}{\sigma}\,\mathrm{Id} - \frac{1}{\widehat{\sigma}}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\right) \boldsymbol{\Delta}_P - \frac{\lfloor Ns\rfloor}{N} \sum_{j=1}^{N} \boldsymbol{e}^T(j)\boldsymbol{\Sigma}^{-1} \left(\frac{1}{\sigma}\,\mathrm{Id} - \frac{1}{\widehat{\sigma}}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\right) \boldsymbol{\Delta}_P \right) \right|$$

$$= O_P(1)\left(O_P(1)o_P(1)O_P(1) - O_P(1)O_P(1)o_P(1)O_P(1)\right) = o_P(1), \tag{14.8}$$

where

$$D^{\widehat{\boldsymbol{\Sigma}}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s) := \left(\boldsymbol{\Delta}_P \mathbb{1}\{F_\vartheta(\cdot) < s \le G_\vartheta(\cdot)\}\right)^T \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Delta}_P$$

and

$$D^{\boldsymbol{\Sigma}_A}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s) := \left(\boldsymbol{\Delta}_P \mathbb{1}\{F_\vartheta(\cdot) < s \le G_\vartheta(\cdot)\}\right)^T \boldsymbol{\Sigma}_A^{-1}\boldsymbol{\Delta}_P.$$

□

Remark 14.3. Note that $D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\cdot)$ still depends on the long-run covariance matrix $\boldsymbol{\Sigma}$. Consequently, the limit process is not pivotal and we have to calculate the critical value for each data set separately.

Remark 14.4. The calculation in (14.7) shows that the long-run variance of the projected errors σ is influenced by the long-run covariance matrix $\boldsymbol{\Sigma}$ of the errors $\boldsymbol{e}(t)$. This means that if we use an estimator $\widehat{\boldsymbol{\Sigma}}_N$ for the true long-run covariance matrix $\boldsymbol{\Sigma}$, then we use simultaneously an estimator $\widehat{\sigma}_N$ for the long-run variance of the projected errors. However, the long-run variance of the projected errors is more easily to estimate.

If we use an estimator $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, which is a positive definite matrix and possibly different from $\boldsymbol{\Sigma}$, and a consistent estimator $\widehat{\sigma}_N$ of σ with $\widehat{\sigma}_N \xrightarrow{P} \sigma_0 \neq 0$, possibly different from σ, but σ_0 is the true long-run variance of the projected errors standardized by $\boldsymbol{\Sigma}_A$. Then we are able to develop the limit process under H_0:

First we define $Y^{\boldsymbol{\Sigma}}(t) = \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{X}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P \right\rangle$ and $\overline{Y}^{\boldsymbol{\Sigma}}_N = \frac{1}{N}\sum_{t=1}^N \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{Y}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P \right\rangle$. Analogously, we define $e^{\boldsymbol{\Sigma}}_P(t) = \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{e}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P \right\rangle$ and $\overline{e}^{\boldsymbol{\Sigma}}_{P,N} = \frac{1}{N}\sum_{t=1}^N \left\langle \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{e}(t), \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Delta}_P \right\rangle$ and further

$$D^{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right) := \left(\boldsymbol{\Delta}_P \mathbb{1}\left\{F_\vartheta(\cdot) < \frac{t}{N} \le G_\vartheta(\cdot)\right\}\right)^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_P, \quad t = 1,\dots,N,$$

and

$$\mathcal{M}^{\boldsymbol{\Sigma}}_\vartheta := \{0 < s < 1 : D^{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+) - D^{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s-) \neq 0\}.$$

Note that by analogy with the proof of Theorem 14.2, there is a functional central limit theorem for the projected errors standardized by $\boldsymbol{\Sigma}_A$:

$$\left\{ \frac{1}{\sigma_0}\frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor Ns\rfloor} e^{\boldsymbol{\Sigma}_A}_P(i) : 0 \le s \le 1 \right\} \xrightarrow{D[0,1]} \{W(s) : 0 \le s \le 1\}.$$

Furthermore, we obtain in the same way as in the lines (14.8)

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\frac{1}{\widehat{\sigma}_N}\max_{\vartheta\in\Theta}\left|\sum_{t=1}^{N} D^{\widehat{\boldsymbol{\Sigma}}_N}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y^{\widehat{\boldsymbol{\Sigma}}_N}(t)-\overline{Y}^{\widehat{\boldsymbol{\Sigma}}_N}_N\right)\right| \\
&= \max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}^{\widehat{\boldsymbol{\Sigma}}_N}_\vartheta}\left(D^{\widehat{\boldsymbol{\Sigma}}_N}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D^{\widehat{\boldsymbol{\Sigma}}_N}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s-)\right)\frac{1}{\widehat{\sigma}_N}\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor Ns\rfloor}\left(e^{\widehat{\boldsymbol{\Sigma}}_N}_P(j)-\overline{e}^{\widehat{\boldsymbol{\Sigma}}_N}_{P,N}\right)\right| \\
&\xrightarrow{D}\max_{\vartheta\in\Theta}\left|\sum_{s\in\mathcal{M}^{\boldsymbol{\Sigma}_A}_\vartheta}\left(D^{\boldsymbol{\Sigma}_A}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D^{\boldsymbol{\Sigma}_A}_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s)\right)B(s)\right|,\quad N\to\infty. \qquad (14.9)
\end{aligned}$$

Remark 14.5. Note, that Theorem 14.2 uses the weight function from (13.4) with $\beta = 0$.

For $0 < \beta \leq \frac{1}{2}$, suppose that the assumptions of Theorem 14.2 hold and

$$\sup_{\vartheta\in\Theta}\left(\int_0^1\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z)-\int_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{-\beta} < \infty,$$

then we get

$$\frac{1}{\sqrt{N}}\frac{1}{\sigma}\max_{\vartheta\in\Theta}\frac{\left|\sum\limits_{t=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right|}{w_P\left(\frac{t}{N}\right)} \xrightarrow{D} \sup_{\vartheta\in\Theta}\left|\frac{\sum\limits_{s\in\mathcal{M}_\vartheta}(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s+)-D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(s-))B(s)}{\left(\int\limits_0^1\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z)-\int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^\beta}\right|,$$

where $\{B(\cdot)\}$ is a Brownian bridge.

It is sufficient that condition 14.5 is satisfied, because then the asymptotic weight function in the denominator of the limit distribution is away from zero. Then the asymptotic distribution is well-defined.

In the example of the gas emission source, condition 14.5 is fullfilled if $D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\cdot)$ is not constant, e.g. if we choose a minimum and maximum opening angle for the gas plume such that $\epsilon < G_\vartheta(i) - F_\vartheta(i) < 1-\epsilon$ for at least one $i = 1,\ldots,d$, $\forall\vartheta\in\Theta$. Particularly, the condition is weaker than in case of the multivariate statistic because the mentioned condition has only to be fulfilled for at least one component. Furthermore, plumes on the boundaries of the x-values of the searching area are also allowed, even a lot more plumes compared to the usage of the multivariate statistic. It also holds true if we replace σ and $\boldsymbol{\Sigma}$ by consistent estimators as in Remark 14.4.

Remark 14.6. Note that the gradual change of the projected data set is step-wise constant and has a finite number of points of discontinuity .

If we would have a differentiable function determining the slope of the gradual change, we get an integral of the Brownian bridge weighted by the derivative of that function, analogously to Hušková and Steinebach (2002).

15. Asymptotics under the Alternative

In this chapter we show the consistency of the test based on the multivariate statistic and on the projection statistic. In fact, we have to prove that under the alternative hypothesis of the existence of an epidemic mean change in at least one component, the test statistic converges to infinity in probability. This fact ensures the convergence to one of the probability that the procedure actually finds the changes under the alternative hypothesis, which is called the asymptotic power one or consistency of the test.

15.1. Consistency: Test based on the Multivariate Statistic

To show the consistency of the test based on the multivariate statistic, we first need a statement that allows to replace $S_\vartheta(i)$ by its asymptotic signal part in asymptotical considerations in a stochastic sense under the alternative hypothesis.

Lemma 15.1

Under the assumptions on the errors of Theorem 14.1, we have under the alternative hypothesis

$$\sup_{i=1,\cdots,d}\sup_{\vartheta\in\Theta}\Bigg|\frac{1}{N}\sum_{t=\lfloor N\widetilde{F}_\vartheta(i)\rfloor+1}^{\lfloor N\widetilde{G}_\vartheta(i)\rfloor}\left(X_i(t)-\frac{1}{N}\sum_{l=1}^{N}X_i(l)\right)$$
$$-\widetilde{\Delta}_i(g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(\widetilde{G}_\vartheta(i))-g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(\widetilde{F}_\vartheta(i)))\Bigg|=o_P(1).$$

Here, $\widetilde{F}_\vartheta$ and $\widetilde{G}_\vartheta$ are arbitrary but fixed functions that are possibly different from the true function F_ϑ and G_ϑ, and

$$g_{t_0,t_1}(s)=\begin{cases}-s(t_1-t_0),\ s\leq t_0\\ s(1-(t_1-t_0))-t_0,\ t_0<s\leq t_1,\\ (1-s)(t_1-t_0),\ s>t_1\end{cases}\quad\text{where}\quad 0\leq t_0<t_1\leq 1,\quad 0\leq s\leq 1.$$

Proof. First, notice that for $i=1,\ldots,d$

$$\frac{1}{N}\sum_{t=N\widetilde{F}_\vartheta(i)+1}^{\lfloor N\widetilde{G}_\vartheta(i)\rfloor}\left(X_i(t)-\frac{1}{N}\sum_{l=1}^{N}X_i(l)\right)$$
$$=\frac{1}{N}\sum_{t=1}^{\lfloor N\widetilde{G}_\vartheta(i)\rfloor}\left(X_i(t)-\frac{1}{N}\sum_{l=1}^{N}X_i(l)\right)-\frac{1}{N}\sum_{t=1}^{\lfloor N\widetilde{F}_\vartheta(i)\rfloor}\left(X_i(t)-\frac{1}{N}\sum_{l=1}^{N}X_i(l)\right).$$

So for $0 < \widetilde{G}_\vartheta(i) \leq 1$ and $i = 1, \ldots, d$ we derive

$$\sup_{i=1,\cdots,d} \sup_{\vartheta\in\Theta} \left| \frac{1}{N} \sum_{t=1}^{\lfloor N\widetilde{G}_\vartheta(i)\rfloor} \left(X_i(t) - \frac{1}{N}\sum_{l=1}^{N} X_i(l) \right) \right.$$
$$\left. -\widetilde{\Delta}_i \left(\frac{1}{N} \sum_{t=1}^{\lfloor N\widetilde{G}_\vartheta(i)\rfloor} \mathbb{1}\left\{ F_{\vartheta_0}(i) < \frac{t}{N} \leq G_{\vartheta_0}(i) \right\} - \frac{\lfloor N\widetilde{G}_\vartheta(i)\rfloor}{N^2} \sum_{l=1}^{N} \mathbb{1}\left\{ F_{\vartheta_0}(i) < \frac{l}{N} \leq G_{\vartheta_0}(i) \right\} \right) \right|$$

$$= \frac{1}{\sqrt{N}} \sup_{i=1,\cdots,d} \sup_{\vartheta\in\Theta} \left| \frac{1}{\sqrt{N}} \sum_{t=1}^{\lfloor N\widetilde{G}_\vartheta(i)\rfloor} (e_i(t) - \bar{e}_i) \right| = o_P(1)O_P(1) = o_P(1),$$

by the fact that the errors satisfy a functional central limit theorem. This lines are analogous for $\widetilde{F}_\vartheta$. Finally we get

$$\widetilde{\Delta}_i \left(\frac{1}{N} \sum_{t=1}^{\lfloor N\widetilde{G}_\vartheta(i)\rfloor} \mathbb{1}\left\{ F_{\vartheta_0}(i) < \frac{t}{N} \leq G_{\vartheta_0}(i) \right\} - \frac{\lfloor N\widetilde{G}_\vartheta(i)\rfloor}{N^2} \sum_{l=1}^{N} \mathbb{1}\left\{ F_{\vartheta_0}(i) < \frac{l}{N} \leq G_{\vartheta_0}(i) \right\} \right)$$
$$= \widetilde{\Delta}_i g_{F_{\vartheta_0},G_{\vartheta_0}}(\widetilde{G}_\vartheta(i)) + o_P(1),$$

where function g_{t_0,t_1}, $i = 1, \ldots, d$, is defined as stated above and the same holds if we replace $\widetilde{G}_\vartheta$ by $\widetilde{F}_\vartheta$. □

Now we are able to prove the consistency of the test based on the multivariate statistic by using the previous lemma.

Theorem 15.1

Let the errors fulfill the assumptions of Theorem 14.1. Then, under the alternative hypothesis,

$$\max_{\vartheta\in\Theta} \frac{1}{N} \boldsymbol{S}_\vartheta^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta \xrightarrow{P} \infty.$$

The assertion is even true if we replace $\boldsymbol{\Sigma}$ by an estimator $\widehat{\boldsymbol{\Sigma}}_N$ satisfying $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, where $\boldsymbol{\Sigma}_A$ is a positive definite matrix possibly different to $\boldsymbol{\Sigma}$.

Proof. First, note that

$$\max_{\vartheta\in\Theta} \frac{1}{N} \boldsymbol{S}_\vartheta^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta \geq \frac{1}{N} \boldsymbol{S}_{\vartheta_0}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_{\vartheta_0}.$$

Then, by Lemma 15.1 we conclude

$$\frac{1}{N} \boldsymbol{S}_{\vartheta_0}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_{\vartheta_0} = N \left(\frac{1}{N^2} \boldsymbol{S}_{\vartheta_0}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_{\vartheta_0} \right) = N \left(\boldsymbol{D}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{D} + o_P(1) \right),$$

where $\boldsymbol{D} = (D_1, \ldots, D_d)^T$ with $D_i = \widetilde{\Delta}_i(g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(G_{\vartheta_0}(i)) - g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(F_{\vartheta_0}(i)))$. The function g_{t_0,t_1} is defined as in Lemma 15.1. Due to $\widetilde{\Delta}_i \neq 0$ for at least one $i = 1, \ldots, d$ under the alternative hypothesis, we obtain for at least one $i = 1, \ldots, d$

$$g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(G_{\vartheta_0}(i)) - g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(F_{\vartheta_0}(i))$$

$$
\begin{aligned}
&= G_{\vartheta_0}(i)(1-(G_{\vartheta_0}(i)-F_{\vartheta_0}(i))) - F_{\vartheta_0}(i) + F_{\vartheta_0}(i)(G_{\vartheta_0}(i)-F_{\vartheta_0}(i)) \\
&= (G_{\vartheta_0}(i)-F_{\vartheta_0}(i)) - G_{\vartheta_0}(i)(G_{\vartheta_0}(i)-F_{\vartheta_0}(i)) + F_{\vartheta_0}(i)(G_{\vartheta_0}(i)-F_{\vartheta_0}(i)) \\
&= (G_{\vartheta_0}(i)-F_{\vartheta_0}(i))\left(1-(G_{\vartheta_0}(i)-F_{\vartheta_0}(i))\right) > 0.
\end{aligned}
$$

Hence, since $\boldsymbol{\Sigma}$ is positive definite, it follows that

$$
N\left(\boldsymbol{D}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{D} + o_P(1)\right) \xrightarrow{P} \infty \qquad (N\to\infty).
$$

If we use an estimator $\widehat{\boldsymbol{\Sigma}}_N$ of $\boldsymbol{\Sigma}$ with $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, where $\boldsymbol{\Sigma}_A$ is positive definite, then we complete the proof again with Lemma 15.1, because of

$$
\frac{1}{N}\boldsymbol{S}_{\vartheta_0}^T\widehat{\boldsymbol{\Sigma}}_N^{-1}\boldsymbol{S}_{\vartheta_0} = N\left(\frac{1}{N^2}\boldsymbol{S}_{\vartheta_0}^T\widehat{\boldsymbol{\Sigma}}_N^{-1}\boldsymbol{S}_{\vartheta_0}\right) = N\left(\boldsymbol{D}^T\boldsymbol{\Sigma}_A^{-1}\boldsymbol{D} + o_P(1)\right) \xrightarrow{P} \infty \qquad (N\to\infty).
$$

□

Remark 15.1. Let the assumptions of Remark 14.2 hold. Then due to the boundedness on Θ of the weight function $w_M(\cdot,\cdot)$ with $0<\beta\le\frac{1}{2}$, we have

$$
\begin{aligned}
\sup_{i=1,\cdots,d}\sup_{\vartheta\in\Theta}\Bigg| & w_{M,i}\left(\frac{\lfloor NF_\vartheta(i)\rfloor}{N},\frac{\lfloor NG_\vartheta(i)\rfloor}{N}\right)\frac{1}{N}\sum_{t=NF_\vartheta(i)+1}^{NG_\vartheta(i)}\left(X_i(t)-\frac{1}{N}\sum_{l=1}^{N}X_i(l)\right) \\
& - w_{M,i}\left(F_\vartheta(i),G_\vartheta(i)\right)\widetilde{\Delta}_i(g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(G_\vartheta(i)) - g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(F_\vartheta(i)))\Bigg| = o_P(1), \quad (15.1)
\end{aligned}
$$

where $g_{t_0,t_1}(s) = \begin{cases} -s(t_1-t_0), & s\le t_0 \\ s(1-(t_1-t_0)-t_0, & t_0<s\le t_1 \\ (1-s)(t_1-t_0), & s>t_1\end{cases}$, $i=1,\ldots,d$ with $0\le t_0<t_1\le 1$.

We know that replacing $w_{M,i}\left(\frac{\lfloor NF_{\vartheta_0}(i)\rfloor}{N},\frac{\lfloor NG_{\vartheta_0}(i)\rfloor}{N}\right)$ with $w_{M,i}\left(F_{\vartheta_0}(i),G_{\vartheta_0}(i)\right)$ is asymptotically negligible. The second term in (15.1) is greater than zero for at least one $i=1,\ldots,d$. Hence, the consistency of the test based on the weighted multivariate statistic follows.

Remark 15.2. Let us use the functions $\widetilde{F}_\vartheta(i)$ and $\widetilde{G}_\vartheta(i)$, $i=1,\ldots,d$, which are possibly different from the true functions $F_\vartheta(i)$ and $G_\vartheta(i)$, $i=1,\ldots,d$. Further, suppose the assumptions of Lemma 15.1, where $\beta=0$, and the assumptions of Remark 15.1, where $\beta\in\left(0,\frac{1}{2}\right]$. Additionally we suppose that there is a $\theta_0\in\Theta$ such that

$$
\Delta_i\left(g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(\widetilde{G}_{\theta_0}(i)) - g_{F_{\vartheta_0}(i),G_{\vartheta_0}(i)}(\widetilde{F}_{\theta_0}(i))\right) \neq 0 \quad \text{for at least one } i \qquad (15.2)
$$

where

$$
\begin{aligned}
& g_{s_1,s_2}(t_1) - g_{s_1,s_2}(t_0) \\
&= \begin{cases}
(s_1-s_2)(t_1-t_0), & s_1<s_2\le t_0 \\
s_2-t_0+(t_1-t_0)(s_1-s_2), & s_1\le t_0, t_0<s_2\le t_1 \\
(t_1-t_0)(s_1-s_2)+(t_1-t_0), & s_1\le t_0, s_2>t_1 \\
(s_2-s_1)(1-(t_1-t_0)), & t_0<s_1<s_2\le t_1 \\
(t_1-t_0)(s_1-s_2+1)+t_0-s_1, & t_0<s_1\le t_1, s_2>t_1 \\
(t_1-t_0)(s_1-s_2), & s_2>s_1>t_1
\end{cases}, \quad \text{with} \quad 0\le t_0<t_1\le 1.
\end{aligned}
$$

Then the consistency of the test based on the multivariate statistic still holds true.

Hence, we obtain

$$\max_{\vartheta\in\Theta}\frac{1}{N}\left(w_{M,i}\left(\frac{\lfloor N\widetilde{\boldsymbol{F}}_\vartheta\rfloor}{N},\frac{\lfloor N\widetilde{\boldsymbol{G}}_\vartheta\rfloor}{N}\right)\widetilde{\boldsymbol{S}}_\vartheta\right)^T\boldsymbol{\Sigma}^{-1}\left(w_{M,i}\left(\frac{\lfloor N\widetilde{\boldsymbol{F}}_\vartheta\rfloor}{N},\frac{\lfloor N\widetilde{\boldsymbol{G}}_\vartheta\rfloor}{N}\right)\widetilde{\boldsymbol{S}}_\vartheta\right)\xrightarrow{P}\infty,$$

where

$$\widetilde{\boldsymbol{S}}_\vartheta=\begin{pmatrix}\sum\limits_{t=\lfloor N\widetilde{F}_\vartheta(1)\rfloor+1}^{\lfloor N\widetilde{G}_\vartheta(1)\rfloor}\big(X_1(t)-\overline{X}_1\big)\\ \vdots\\ \sum\limits_{t=\lfloor N\widetilde{F}_\vartheta(d)\rfloor+1}^{\lfloor N\widetilde{G}_\vartheta(d)\rfloor}\big(X_d(t)-\overline{X}_d\big)\end{pmatrix}.$$

The key reason is that Lemma 15.1 and equation (15.1) hold true for arbitrary but fixed functions $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$ which do not have to be necessarily correct.

With the help of equation (15.2), the proof is analogous to the proof of Theorem 15.1.

15.2. Consistency: Test based on the Projection Statistic

Analogous to the multivariate statistic the first step to prove consistency of the test based on the projection statistic is to show the asymptotic equivalency of the statistic and its asymptotic signal part in a stochastic sense. But note, that the function $D_{t_0,t_1}(\cdot),\ 0<t_0\le t_1\le 1$, depends on the long-run covariance matrix $\boldsymbol{\Sigma}$.

Lemma 15.2

Let the errors fulfill the assumptions of Theorem 14.2. Then, under the alternative hypothesis, we get

$$\sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^{N}D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)(Y(t)-\overline{Y}_N)-g(\widetilde{F}_\vartheta(1),\ldots,\widetilde{F}_\vartheta(d),\widetilde{G}_\vartheta(1),\ldots,\widetilde{G}_\vartheta(d))\right|=o_P(1),$$

where $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$ are arbitrary but fixed functions, not necessarily equal to the true functions F_ϑ and G_ϑ, and

$$g(\widetilde{F}_\vartheta(1),\ldots,\widetilde{F}_\vartheta(d),\widetilde{G}_\vartheta(1),\ldots,\widetilde{G}_\vartheta(d))=\int\limits_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)-\int\limits_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)dz.$$

The assertion holds true if we replace the long-run covariance matrix $\boldsymbol{\Sigma}$ by an estimator $\widehat{\boldsymbol{\Sigma}}_N$ satisfying $\widehat{\boldsymbol{\Sigma}}_N\xrightarrow{P}\boldsymbol{\Sigma}_A$, where $\boldsymbol{\Sigma}_A$ is positive definite and possibly different from $\boldsymbol{\Sigma}$, thus we have

$$\sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^{N} D^{\widehat{\boldsymbol{\Sigma}}_N}_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)(Y^{\widehat{\boldsymbol{\Sigma}}_N}(t)-\overline{Y}^{\widehat{\boldsymbol{\Sigma}}_N}_N) - g^{\boldsymbol{\Sigma}_A}(\widetilde{F}_\vartheta(1),\dots,\widetilde{F}_\vartheta(d),\widetilde{G}_\vartheta(1),\dots,\widetilde{G}_\vartheta(d))\right| = o_P(1),$$

and

$$g^{\boldsymbol{\Sigma}_A}(\widetilde{F}_\vartheta(1),\dots,\widetilde{F}_\vartheta(d),\widetilde{G}_\vartheta(1),\dots,\widetilde{G}_\vartheta(d)) = \int_0^1 D^{\boldsymbol{\Sigma}_A}_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)\left(D^{\boldsymbol{\Sigma}_A}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int_0^1 D^{\boldsymbol{\Sigma}_A}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)dz,$$

where $D^{\boldsymbol{\Sigma}_A}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}$ and $Y^{\widehat{\boldsymbol{\Sigma}}_N}(t)$ and $\overline{Y}^{\widehat{\boldsymbol{\Sigma}}_N}_N$ are defined in Remark (14.4).

Proof. First we write out the projected observations

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N} D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)(Y(t)-\overline{Y}_N)\\
&=\frac{1}{N}\sum_{t=1}^{N} D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right) - \overline{D}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}} + e_P(t) - \overline{e}_P\right)\\
&=\frac{1}{N}\sum_{t=1}^{N}\left(D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right) - \overline{D}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\right) + D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)(e_P(t)-\overline{e}_P)\right).
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
&\sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^{N}\left(D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right) - \overline{D}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\right) + D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)(e_P(t)-\overline{e}_P)\right)\right.\\
&\qquad\left. - \int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)dz\right|\\
&\leq \sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^{N}\left(D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right) - \overline{D}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\right)\right)\right.\\
&\qquad\left. - \int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)dz\right|\\
&\quad+\sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^{N} D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)(e_P(t)-\overline{e}_P)\right|.
\end{aligned}$$

The last term is in $o_P(1)$ by Theorem 14.2 because the projected time series $Y(t)$ under the null hypothesis is equal to the projected errors.

We approximate the other term as follows:

$$\sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^{N}\left(D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right) - \overline{D}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\right)\right)\right.$$

$$
\begin{aligned}
&-\int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)-\int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)dz\Bigg|\\
&=\sup_{\vartheta\in\Theta}\Bigg|\frac{1}{N}\sum_{t=1}^N D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right)-\frac{1}{N}\sum_{t=1}^N D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\overline{D}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\\
&-\left(\int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)dz-\int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)dz\int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)\Bigg|\\
&\leq\sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^N D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right)-\int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)dz\right|\\
&+\sup_{\vartheta\in\Theta}\left|\frac{1}{N}\sum_{t=1}^N D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\overline{D}_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}-\int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z)dz\int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right|\\
&=o_P(1).
\end{aligned}
$$

If we use an estimator $\widehat{\boldsymbol{\Sigma}}_N$ satisfying $\widehat{\boldsymbol{\Sigma}}_N\xrightarrow{P}\boldsymbol{\Sigma}_A$, where $\boldsymbol{\Sigma}_A$ is positive definite, for proving we have to use equation (14.9) and

$$
\begin{aligned}
&\max_x\left|D_{t_0,t_1}^{\widehat{\boldsymbol{\Sigma}}_N}(x)-D_{t_0,t_1}^{\boldsymbol{\Sigma_N}}(x)\right|\\
&=\max_x\left|(\boldsymbol{\Delta}_P\mathbb{1}\{F_\vartheta(\cdot)<x\leq G_\vartheta(\cdot)\})^T\,\widehat{\boldsymbol{\Sigma}}_N^{-1}\boldsymbol{\Delta}_P-(\boldsymbol{\Delta}_P\mathbb{1}\{F_\vartheta(\cdot)<x\leq G_\vartheta(\cdot)\})^T\,\boldsymbol{\Sigma}_A{}^{-1}\boldsymbol{\Delta}_P\right|\\
&=\max_x\left|(\boldsymbol{\Delta}_P\mathbb{1}\{F_\vartheta(\cdot)<x\leq G_\vartheta(\cdot)\})^T\left(\widehat{\boldsymbol{\Sigma}}_N^{-1}-\boldsymbol{\Sigma}_A{}^{-1}\right)\boldsymbol{\Delta}_P\right|\\
&\leq\max_x\left\|(\boldsymbol{\Delta}_P\mathbb{1}\{F_\vartheta(\cdot)<x\leq G_\vartheta(\cdot)\})\right\|^2\left\|\left(\widehat{\boldsymbol{\Sigma}}_N^{-1}-\boldsymbol{\Sigma}_A{}^{-1}\right)\boldsymbol{\Delta}_P\right\|^2\\
&=\max_x\left\|(\boldsymbol{\Delta}_P\mathbb{1}\{F_\vartheta(\cdot)<x\leq G_\vartheta(\cdot)\})\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_N^{-1}-\boldsymbol{\Sigma}_A{}^{-1}\right\|^2\left\|\boldsymbol{\Delta}_P\right\|^2\\
&=O_P(1)o_P(1)O_P(1)=o_P(1).
\end{aligned}
$$

□

Next we are able to prove the consistency of the projection statistic.

Theorem 15.2

Let the errors fulfill the assumptions of Theorem 14.2. Then, under the alternative hypothesis, we get

$$
\frac{1}{\sqrt{N}}\frac{1}{\sigma}\max_{\vartheta\in\Theta}\left|\sum_{t=1}^N D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right|\xrightarrow{P}\infty.
$$

This assertion holds true if we replace $\boldsymbol{\Sigma}$ by $\widehat{\boldsymbol{\Sigma}}_N$ with $\widehat{\boldsymbol{\Sigma}}_N\xrightarrow{P}\boldsymbol{\Sigma}_A$ where $\boldsymbol{\Sigma}_A$ is positive definite, and if we replace σ by an estimator $\widehat{\sigma}_N$ satisfying $\widehat{\sigma}_N\xrightarrow{P}\sigma_A$, where $\sigma_A=\sum_{h\in\mathbb{Z}}\operatorname{Cov}(e_p(0),e_P(h))\neq 0$.

Proof. A simple approximation and Lemma 15.2 yield

$$\frac{1}{\sqrt{N}}\frac{1}{\sigma}\max_{\vartheta\in\Theta}\left|\sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right| \geq \frac{1}{\sqrt{N}}\frac{1}{\sigma}\left|\sum_{t=1}^{N} D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right|$$
$$= \sqrt{N}\frac{1}{N}\frac{1}{\sigma}\left|\sum_{t=1}^{N} D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right|$$
$$= \sqrt{N}\frac{1}{\sigma}\left(\left|g(G_{\vartheta_0}(1),\ldots,G_{\vartheta_0}(d),F_{\vartheta_0}(1),\ldots,F_{\vartheta_0}(d))\right| + o_P(1)\right). \tag{15.3}$$

Finally if the absolute value of the function $g(\cdot)$, plugged into the $\boldsymbol{F}_{\vartheta_0}$ and $\boldsymbol{G}_{\vartheta_0}$, is not zero, the assertion follows. Under the alternative we have with the Jensen inequality

$$g(G_{\vartheta_0}(1),\ldots,G_{\vartheta_0}(d),F_{\vartheta_0}(1),\ldots,F_{\vartheta_0}(d))$$
$$= \int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)dz$$
$$= \int_0^1 D^2_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)dz - \left(\int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right)^2 dz \geq 0,$$

and the equality only holds if $D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\cdot)$ is constant, which is not the case under the alternative hypothesis because there is an epidemic mean change in at least one component. Then we conclude

$$\sqrt{N}\left|g(G_{\vartheta_0}(1),\ldots,G_{\vartheta_0}(d),F_{\vartheta_0}(1),\ldots,F_{\vartheta_0}(d))\right| \xrightarrow{P} \infty.$$

If an estimator $\widehat{\sigma}_N$ is used that satisfies $\widehat{\sigma}_N \xrightarrow{P} \sigma_A \neq 0$ is used, the assertion still holds true because σ is replaced by σ_A in equation (15.3) as well as $g(\cdot)$ is replaced by $g^{\boldsymbol{\Sigma}_A}(\cdot)$. The proof is finished since $g^{\boldsymbol{\Sigma}_A}(G_{\vartheta_0}(1),\ldots,G_{\vartheta_0}(d),F_{\vartheta_0}(1),\ldots,F_{\vartheta_0}(d)) \neq 0$ under the alternative hypothesis. □

Remark 15.3. Let the assumptions of Remark 14.5 hold, then an analogon with the weight function $w_P(\cdot)$ with $0 < \beta \leq \frac{1}{2}$ of Theorem 15.2 holds true.

First note, that we have an analogon to Lemma 15.2

$$\sup_{\vartheta\in\Theta}\left|\left(w_P\left(\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta\right)\right)^{-1}\frac{1}{N}\sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right.$$
$$\left. - \left(\int_0^1\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{-\beta} g(F_\vartheta(1),\ldots,F_\vartheta(d),G_\vartheta(1),\ldots,G_\vartheta(d))\right|$$
$$= o_P(1),$$

where $g(\cdot)$ is defined in Lemma 15.2.

Finally, the consistency of the test based on the projection statistic follows, due to the fact that the asymptotic weight function, by plugging in the true parameter ϑ_0, is not equal to zero under the alternative hypothesis of an epidemic mean change in at least one component.

The assertion continues to hold if we use estimators $\widehat{\boldsymbol{\Sigma}}_N$ and $\widehat{\sigma}_N$ that converge stochastically to a positive definite matrix $\boldsymbol{\Sigma}_A$ and $\sigma_A \neq 0$, respectively due to the fact that Lemma 15.2 also holds using these estimators.

Remark 15.4. The consistency of the test based on the projection statistic still even holds if the statistic uses functions $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$ which are possibly different from the true functions $F_\vartheta(\cdot)$ and $G_\vartheta(\cdot)$ as well as there is a $\theta_0 \in \Theta$ such that

$$\int_0^1 D_{\widetilde{\boldsymbol{F}}_{\theta_0},\widetilde{\boldsymbol{G}}_{\theta_0}}(z) \left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega) d\omega \right) \neq 0. \tag{15.4}$$

The explanation is that Lemma 15.2 still holds true for functions $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$ which differ from $F_\vartheta(\cdot)$ and $G_\vartheta(\cdot)$. Furthermore, with condition (15.4) we have

$$\begin{aligned} & g(\widetilde{G}_{\theta_0}(1), \dots, \widetilde{G}_{\theta_0}(d), \widetilde{F}_{\theta_0}(1), \dots, \widetilde{F}_{\theta_0}(d)) \\ &= \int_0^1 D_{\widetilde{\boldsymbol{F}}_{\theta_0},\widetilde{\boldsymbol{G}}_{\theta_0}}(z) \left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega) d\omega \right) dz \neq 0 \end{aligned}$$

under the alternative hypothesis. Theorem 15.2 is proven with arbitrary functions $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$, thus the consistency follows in the misspecified case.

This holds also true for the test based on the weighted version of the projection statistic by the condition (15.4) and because $w_M(\cdot,\cdot) > 0$.

The assertion still remains true even if we use estimators $\widehat{\boldsymbol{\Sigma}}_N$ and $\widehat{\sigma}_N$ which converge stochastically to a positive definite matrix $\boldsymbol{\Sigma}_A$ and $\sigma_A \neq 0$, respectively.

16. Consistency of the Estimators

In Section 13.3 we introduced the estimators based on both the multivariate and projection statistic. In this chapter we prove the consistency of both estimators, i.e the estimators for ϑ_0 converges stochastically to the true parameter ϑ_0. For the proofs of both estimators we need identifiability conditions for all $\vartheta \in \Theta$ which restrict the allowed functions determining the relationship between the components.

In the application example of locating a source of a gas emission, a linear plume fulfills this conditions. We will use such a plume in our simulation study and the real data example.

16.1. Consistency of the Estimator based on the Multivariate Statistic

The estimator for ϑ_0 based on the multivariate statistic is given by

$$\widehat{\vartheta}_M = \arg\max_{\vartheta \in \Theta} \boldsymbol{S}_\vartheta^{wT} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta^w.$$

The first theorem shows the consistency of this estimator for $\beta = 0$, then the estimator simplifies to

$$\widehat{\vartheta}_M = \arg\max_{\vartheta \in \Theta} \boldsymbol{S}_\vartheta^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta.$$

Theorem 16.1

Let $\beta = 0$ and the assumptions on the errors of Theorem 14.1 hold. In addition, let the following condition be fulfilled:

$$\arg\max_{\vartheta \in \Theta} \|\boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{D}_\vartheta\| = \vartheta_0, \text{ where } \vartheta_0 \text{ is unique,} \tag{16.1}$$

where $D_\vartheta(i) = \widetilde{\Delta}_i \left(g_{F_{\vartheta_0}(i), G_{\vartheta_0}(i)}(G_\vartheta(i)) - g_{F_{\vartheta_0}(i), G_{\vartheta_0}(i)}(F_\vartheta(i)) \right)$ and $\boldsymbol{D}_\vartheta = (D_\vartheta(1), \ldots, D_\vartheta(d))^T$, g is defined in Lemma 15.1.

Then under the alternative hypothesis H_1 we get:
The estimator based on the multivariate statistic $\widehat{\vartheta}_M$ is consistent, i.e. $\widehat{\vartheta}_M \xrightarrow{P} \vartheta_0$.
The assertion even holds true if we replace $\boldsymbol{\Sigma}$ by an estimator $\widehat{\boldsymbol{\Sigma}}_N$ satisfying $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, where $\boldsymbol{\Sigma}_A$ is a positive definite matrix and fulfills condition (16.1). $\boldsymbol{\Sigma}_A$ is possibly different from $\boldsymbol{\Sigma}$.

Proof. The estimator $\widehat{\vartheta}_M$ is also determined by

$$\widehat{\vartheta}_M = \arg\max_{\vartheta \in \Theta} \frac{1}{N^2} \boldsymbol{S}_\vartheta^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta. \tag{16.2}$$

First we take a look at the following approximations, in which we use Lemma 15.1 and the assumption $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$ as well as the fact that $D_\vartheta(i),\ \vartheta \in \Theta$, is bounded because $g_{t_0,t_1}(s),\ s \in [0,1], 0 \le t_0 < t_1 \le 1$, is also bounded and thus the same holds for $||\boldsymbol{D}_\vartheta||$, where $\|\cdot\|$ is the l_2-norm:

$$\begin{aligned}
&\max_{\vartheta\in\Theta}\left|\left(\frac{1}{N^2}\boldsymbol{S}_\vartheta^T\widehat{\boldsymbol{\Sigma}}_N^{-1}\boldsymbol{S}_\vartheta\right)^{\frac{1}{2}} - \left(\boldsymbol{D}_\vartheta^T\boldsymbol{\Sigma}_A^{-1}\boldsymbol{D}_\vartheta\right)^{\frac{1}{2}}\right| = \max_{\vartheta\in\Theta}\left|\left\|\frac{1}{N}\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\boldsymbol{S}_\vartheta\right\| - \left\|\boldsymbol{\Sigma}_A^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|\right| \\
&= \max_{\vartheta\in\Theta}\left(\left\|\frac{1}{N}\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\boldsymbol{S}_\vartheta\right\| - \left\|\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\| + \left\|\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\| - \left\|\boldsymbol{\Sigma}_A^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|\right) \\
&\le \max_{\vartheta\in\Theta}\left(\left\|\frac{1}{N}\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\boldsymbol{S}_\vartheta - \widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\| + \left\|\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\boldsymbol{D}_\vartheta - \boldsymbol{\Sigma}_A^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|\right) \\
&\le \max_{\vartheta\in\Theta}\left(\left\|\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\left(\frac{1}{N}\boldsymbol{S}_\vartheta - \boldsymbol{D}_\vartheta\right)\right\| + \left\|\left(\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}} - \boldsymbol{\Sigma}_A^{-\frac{1}{2}}\right)\boldsymbol{D}_\vartheta\right\|\right) \\
&\le \max_{\vartheta\in\Theta}\left(\left\|\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}}\right\|_F\left\|\frac{1}{N}\boldsymbol{S}_\vartheta - \boldsymbol{D}_\vartheta\right\| + \left\|\left(\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}} - \boldsymbol{\Sigma}_A^{-\frac{1}{2}}\right)\right\|_F\|\boldsymbol{D}_\vartheta\|\right) \\
&= \left(\left\|\boldsymbol{\Sigma}_A^{-\frac{1}{2}}\right\|_F + o_P(1)\right)\max_{\vartheta\in\Theta}\left\|\left(\frac{1}{N}\boldsymbol{S}_\vartheta\right) - \boldsymbol{D}_\vartheta\right\| + o_P(1) = o_P(1).
\end{aligned}$$

The norm $\|\cdot\|_F$ is the Frobenius norm which is submultiplicative (see Banerjee and Roy (2014)) and by the continuous-mapping theorem and the continuity of the Frobenius norm, it holds $\left\|\left(\widehat{\boldsymbol{\Sigma}}_N^{-\frac{1}{2}} - \boldsymbol{\Sigma}_A^{-\frac{1}{2}}\right)\right\|_F = o_P(1)$.

If we want to show the consistency of the estimator using the true long-run covariance matrix $\boldsymbol{\Sigma}$, the approximations can be simplified to give

$$\begin{aligned}
&\max_{\vartheta\in\Theta}\left|\left(\frac{1}{N^2}\boldsymbol{S}_\vartheta^T\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_\vartheta\right)^{\frac{1}{2}} - \left(\boldsymbol{D}_\vartheta^T\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_\vartheta\right)^{\frac{1}{2}}\right| = \max_{\vartheta\in\Theta}\left|\left\|\frac{1}{N}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}_\vartheta\right\| - \left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|\right| \\
&\le \max_{\vartheta\in\Theta}\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\frac{1}{N}\boldsymbol{S}_\vartheta - \boldsymbol{D}_\vartheta\right)\right\| \le \left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\right\|_F\max_{\vartheta\in\Theta}\left\|\left(\frac{1}{N}\boldsymbol{S}_\vartheta\right) - \boldsymbol{D}_\vartheta\right\| = o_P(1). \qquad (16.3)
\end{aligned}$$

In view of (16.2) and Theorem B.4, it is enough to show that the term $\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|$ and $\left\|\boldsymbol{\Sigma}_A^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|$, respectively, has a unique maximum at $\vartheta = \vartheta_0$ and is a continuous function of ϑ.

The continuity follows directly from the continuity of the function g. Furthermore, condition (16.1) ensures that $\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|$ has a unique maximum at ϑ_0. Since if we use an estimator $\widehat{\boldsymbol{\Sigma}}_N$ with $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, the condition (16.1) is satisfied for $\boldsymbol{\Sigma}_A$ instead of $\boldsymbol{\Sigma}$, it is also ensures that $\left\|\boldsymbol{\Sigma}_A^{-\frac{1}{2}}\boldsymbol{D}_\vartheta\right\|$ has a unique maximum at ϑ_0. □

Corollary 16.1

Let the assumptions of Theorem 16.1 be fulfilled. In addition, let all $\vartheta \in \Theta$ be identifiable, i.e. if $\vartheta_1, \vartheta_2 \in \Theta$ and $\vartheta_1 \neq \vartheta_2$ then $F_{\vartheta_1}(i) \neq F_{\vartheta_2}(i)$ or $G_{\vartheta_1}(i) \neq G_{\vartheta_2}(i)$ for at least one $i \in \{1, \dots, d\}$

with $\widetilde{\Delta}_i \neq 0$. If $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_A$, respectively, are diagonal matrices, then the estimator $\widehat{\vartheta}_M$ is consistent.

Proof. We have to establish that condition (16.1) holds for the diagonal matrices $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_A$, respectively. Note that the diagonal elements of $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_A$ are strictly positive.

The function $g_{t_0,t_1}(s),\ s \in [0,1], 0 \leq t_0 < t_1 \leq 1$, defined in Lemma 15.1, is piecewise constant, and starts in zero, decreases till t_0 then increases and crosses the x-axis, consequently goes from negative to positive values, till it reaches t_1 on the x-axis, and then again decreases till zero again. Hence $g_{(t_0,t_1)}$ has a maximum at t_1 and a minimum at t_0 and both are unique. Furthermore, ϑ_0 achieve the maximum in each component but not necessarily unique, thus we have

$$|D_\vartheta(i)| \leq |D_{\vartheta_0}(i)|, \quad \forall\, \vartheta \in \Theta,\ i = 1, \dots d.$$

However, due to the identifiability

$$\text{for each } \vartheta \in \Theta \text{ different from } \vartheta_0, \text{ there exists an } i \text{ with } |D_\vartheta(i)| < |D_{\vartheta_0}(i)|. \tag{16.4}$$

Since $\mathbf{\Sigma}_A = \operatorname{diag}(s_1, \dots, s_d),\ s_i > 0$ is a diagonal matrix, we finally get for all $\vartheta \neq \vartheta_0$

$$\left\| \mathbf{\Sigma}_A^{-\frac{1}{2}} \boldsymbol{D}_\vartheta \right\|^2 = \sum_{i=1}^d \frac{1}{s_i} D_\vartheta^2(i) < \sum_{i=1}^d \frac{1}{s_i} D_{\vartheta_0}^2(i) = \left\| \mathbf{\Sigma}_A^{-\frac{1}{2}} \boldsymbol{D}_{\vartheta_0} \right\|^2.$$

Consequently, condition (16.1) is satisfied and thus $\left\| \mathbf{\Sigma}_A^{-\frac{1}{2}} \boldsymbol{D}_\vartheta \right\|$ has a unique maximum at ϑ_0. □

Remark 16.1. Suppose that the assumptions of Theorem 16.1 and Remark 15.1 hold. Thus the weight function $\boldsymbol{w}_M(\cdot,\cdot)$ is bounded on Θ. Furthermore, let

$$\arg\max_{\vartheta \in \Theta} \mathbf{\Sigma}_A^{-\frac{1}{2}} (\boldsymbol{w}_M(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta) \boldsymbol{D}_\vartheta) = \vartheta_0, \text{ where } \vartheta_0 \text{ is unique,} \tag{16.5}$$

$$\boldsymbol{w}_M(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta) = \begin{pmatrix} \boldsymbol{w}_{M,1}\left(\frac{\lfloor N F_\vartheta(1) \rfloor}{N}, \frac{\lfloor N G_\vartheta(1) \rfloor}{N}\right) \\ \vdots \\ \boldsymbol{w}_{M,d}\left(\frac{\lfloor N F_\vartheta(d) \rfloor}{N}, \frac{\lfloor N G_\vartheta(d) \rfloor}{N}\right) \end{pmatrix}$$ and $\boldsymbol{D}_\vartheta$ is defined in Theorem 16.1. They are multiplied component-wise.

Then the estimator based on the weighted multivariate statistic

$$\widehat{\vartheta}_M = \arg\max_{\vartheta \in \Theta} \boldsymbol{S}_\vartheta^{wT} \mathbf{\Sigma}^{-1} \boldsymbol{S}_\vartheta^w$$

is also consistent.

The proof is analogous to the proof of the non-weighted version of the estimator. Using (15.1), we have

$$\begin{aligned} &\max_{\vartheta \in \Theta} \left| \left\| \frac{1}{N} \widehat{\mathbf{\Sigma}}_N^{-\frac{1}{2}} \boldsymbol{S}_\vartheta^w \right\| - \left\| \mathbf{\Sigma}_A^{-\frac{1}{2}} (\boldsymbol{w}_M(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta) \boldsymbol{D}_\vartheta) \right\| \right| \\ &\leq \max_{\vartheta \in \Theta} \left\| \widehat{\mathbf{\Sigma}}_N^{-\frac{1}{2}} \right\|_F \left\| \frac{1}{N} \boldsymbol{S}_\vartheta^w - (\boldsymbol{w}_M(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta) \boldsymbol{D}_\vartheta) \right\| \\ &\quad + \left\| \left(\widehat{\mathbf{\Sigma}}_N^{-\frac{1}{2}} - \mathbf{\Sigma}_A^{-\frac{1}{2}} \right) \right\|_F \|(\boldsymbol{w}_M(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta) \boldsymbol{D}_\vartheta)\| \end{aligned}$$

$$= \left(\left\|\boldsymbol{\Sigma}_A^{-\frac{1}{2}}\right\|_F + o_P(1)\right) \max_{\vartheta\in\Theta} \left\| \left(\frac{1}{N}\boldsymbol{S}_\vartheta^w\right) - (\boldsymbol{w}_M(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta)\boldsymbol{D}_\vartheta)\right\| + o_P(1)$$
$$= o_P(1),$$

where $\|\cdot\|_F$ is again the Frobenius norm. Now, since condition (16.5) holds, we can finish the proof.

To show the consistency of the estimator using the true long-run covariance matrix $\boldsymbol{\Sigma}$, the reasoning is analogous to the lines in the proof of the non-weighted version in (16.3). By condition (16.5), in which we have to replace $\boldsymbol{\Sigma}_A$ by $\boldsymbol{\Sigma}$, we finish the proof.

In the misspecified case, that arises if we use the functions $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$, which are different from the true functions $F_\vartheta(\cdot)$ and $G_\vartheta(\cdot)$, consistency cannot be shown. The reason is that equation (16.4) does not hold anymore by plugging in the misspecified functions.

16.2. Consistency of the Estimator based on the Projection Statistic

The estimator for ϑ_0 obtained by the projection statistic is defined as

$$\widehat{\vartheta}_P = \arg\max_{\vartheta\in\Theta} \frac{\left|\sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right|}{\left(\sum_{t=1}^{N}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right) - \frac{1}{N}\sum_{l=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{l}{N}\right)\right)^2\right)^{\frac{1}{2}}}.$$

Note that the estimator depends on the long-run covariance matrix $\boldsymbol{\Sigma}$, since $D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\cdot)$ depends on $\boldsymbol{\Sigma}$.

Theorem 16.2

Let the assumptions of Remark 14.5 hold. In addition, let all $\vartheta \in \Theta$ identifiable, i.e. if $\vartheta_1, \vartheta_2 \in \Theta$ and $\vartheta_1 \neq \vartheta_2$ then $D_{\boldsymbol{F}_{\vartheta_1},\boldsymbol{G}_{\vartheta_1}}(\cdot) \neq aD_{\boldsymbol{F}_{\vartheta_2},\boldsymbol{G}_{\vartheta_2}}(\cdot) + b$ with $a \neq 0$ and $b \in \mathbb{R}$. Then under the alternative hypothesis H_1 we get:
The estimator $\widehat{\vartheta}_P$ is consistent, i.e. $\widehat{\vartheta}_P \xrightarrow{P} \vartheta_0$.
The assertion still holds true if we replace the long-run covariance matrix by an estimator $\widehat{\boldsymbol{\Sigma}}_N$ satisfying $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, where $\boldsymbol{\Sigma}_A$ is a positive definite matrix, possibly different from $\boldsymbol{\Sigma}$.

Proof. The estimator can also be written in the form

$$\widehat{\vartheta}_P = \arg\max_{\vartheta\in\Theta} \frac{1}{N} \frac{\left|\sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right|}{\left(\sum_{t=1}^{N}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right) - \frac{1}{N}\sum_{l=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{l}{N}\right)\right)^2\right)^{\frac{1}{2}}}.$$

Invoking Remark 15.3, we have

$$\left| \frac{1}{N} \frac{\left|\sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t)-\overline{Y}_N\right)\right|}{\left(\sum_{t=1}^{N}\left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right) - \frac{1}{N}\sum_{l=1}^{N} D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}\left(\frac{l}{N}\right)\right)^2\right)^{\frac{1}{2}}} \right.$$

$$- \frac{g(F_\vartheta(1),\ldots,F_\vartheta(d),G_\vartheta(1),\ldots,G_\vartheta(d))}{\left(\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{\frac{1}{2}}} \Bigg| = o_P(1),$$

where $g(\cdot)$ is defined in Lemma 15.2. According to Theorem B.4 we have to show that the last term has a unique maximum at $\vartheta = \vartheta_0$ and is a continuous function of ϑ. To prove that the maximum at $\vartheta = \vartheta_0$ is unique, we need

$$\begin{aligned}
&\arg\max_{\vartheta\in\Theta} \frac{g(F_\vartheta(1),\ldots,F_\vartheta(d),G_\vartheta(1),\ldots,G_\vartheta(d))}{\left(\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{\frac{1}{2}}} \qquad (16.6)\\
&= \arg\max_{\vartheta\in\Theta} \frac{\int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int\limits_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right) dz}{\left(\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{\frac{1}{2}}}\\
&= \arg\max_{\vartheta\in\Theta} \frac{\int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z)D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)dz - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z)\int\limits_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega dz}{\left(\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{\frac{1}{2}}}\\
&= \arg\max_{\vartheta\in\Theta} \frac{\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right) D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z)dz}{\left(\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{\frac{1}{2}}}\\
&= \arg\max_{\vartheta\in\Theta} \frac{\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int\limits_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right) dz}{\left(\int\limits_0^1 \left(D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{\frac{1}{2}}}\\
&= \arg\max_{\vartheta\in\Theta} \frac{\int\limits_0^1 H_\vartheta(z)H_{\vartheta_0}(z)dz}{\left(\int\limits_0^1 H_\vartheta^2(z)dz\right)^{\frac{1}{2}}}, \qquad (16.7)
\end{aligned}$$

where $H_\vartheta(z) = D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int\limits_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega,\ z \in (0,1]$. This function is continuous in ϑ because $D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\cdot)$ is continuous in ϑ and thus also the function in (16.6). It remains to analyse the term in (16.7) concerning the unique maximum at $\vartheta = \vartheta_0$. The Cauchy-Schwarz inequality gives

$$\frac{\int_0^1 H_\vartheta(z) H_{\vartheta_0}(z)dz}{\left(\int_0^1 H_\vartheta^2(z)dz\right)^{\frac{1}{2}}} \leq \frac{\left(\int_0^1 H_\vartheta^2(z)dz \int_0^1 H_{\vartheta_0}^2(z)dz\right)^{\frac{1}{2}}}{\left(\int_0^1 H_\vartheta^2(z)dz\right)^{\frac{1}{2}}} = \left(\int_0^1 H_{\vartheta_0}^2(z)dz\right)^{\frac{1}{2}}.$$

The maximum value $\left(\int_0^1 H_{\vartheta_0}^2(z)dz\right)^{\frac{1}{2}}$ is achieved if $H_\vartheta(z) = cH_{\vartheta_0}(z), \forall z \in [0,1], c \neq 0$. For this condition, we have

$$\begin{aligned}
&H_\vartheta(z) = cH_{\vartheta_0}(z) \quad \forall z \in [0,1], c \neq 0\\
&\Leftrightarrow D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega = c\left(D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) - \int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega\right) \quad \forall z \in [0,1], c \neq 0\\
&\Leftrightarrow D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - cD_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) = \int_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega - c\int_0^1 D_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(\omega)d\omega \quad \forall z \in [0,1], c \neq 0.
\end{aligned}$$

Thus there is a constant $b \in \mathbb{R}$ such that

$$D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - cD_{\boldsymbol{F}_{\vartheta_0},\boldsymbol{G}_{\vartheta_0}}(z) = b \quad \forall z \in [0,1], c \neq 0.$$

This equation only holds if $\vartheta = \vartheta_0$ by the identifiability condition. We conclude that the maximum of $\frac{\int_0^1 H_\vartheta(z)H_{\vartheta_0}(z)dz}{\left(\int_0^1 H_\vartheta^2 dz\right)^{\frac{1}{2}}}$ is attained at $\vartheta = \vartheta_0$. Finally, by Theorem B.4, we get $\widehat{\vartheta}_P \xrightarrow{P} \vartheta_0$.

If we use $\widehat{\boldsymbol{\Sigma}}_N$ satisfying $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, the proof runs analogously using Lemma 15.2. □

Remark 16.2. If we use the projection statistic with misspecified functions $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$, then let the assumptions of Remark 14.5 hold with $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$. And if $\widetilde{\vartheta}_0$ is the unique maximizer of $\frac{\int_0^1 \widetilde{H}_\vartheta(z)H_{\vartheta_0}(z)dz}{\left(\int_0^1 \widetilde{H}_\vartheta^2 dz\right)^{\frac{1}{2}}}$ with $\widetilde{H}_\vartheta(z) = D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z) - \int_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(\omega)d\omega,\ z \in (0,1]$ and $H_\vartheta(z) = D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(z) - \int_0^1 D_{\boldsymbol{F}_\vartheta,\boldsymbol{G}_\vartheta}(\omega)d\omega,\ z \in (0,1]$, then under the alternative hypothesis H_1, we have

$$\widehat{\vartheta}_P = \underset{\vartheta\in\Theta}{\arg\max}\frac{1}{N}\frac{\left|\sum_{t=1}^N D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t) - \overline{Y}_N\right)\right|}{\left(\sum_{t=1}^N\left(D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{t}{N}\right) - \frac{1}{N}\sum_{l=1}^N D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}\left(\frac{l}{N}\right)\right)^2\right)^{\frac{1}{2}}} \xrightarrow{P} \widetilde{\vartheta}_0.$$

Thus the parameter $\widetilde{\vartheta}_0$ is a best-approximating parameter.

The corresponding proof first uses the fact that Remark 15.3 holds true for arbitrary functions $\widetilde{F}_\vartheta(\cdot)$ and $\widetilde{G}_\vartheta(\cdot)$ which have to satisfy that $D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(t), 0 < t \leq 1$ is not constant. These functions are allowed to be different from the true functions $F_\vartheta(\cdot)$ and $G_\vartheta(\cdot)$. Then we have

$$\arg\max_{\vartheta\in\Theta} \frac{g(\widetilde{F}_\vartheta(1),\ldots,\widetilde{F}_\vartheta(d),\widetilde{G}_\vartheta(1),\ldots,\widetilde{G}_\vartheta(d))}{\left(\int\limits_0^1 \left(D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z) - \int\limits_0^1 D_{\widetilde{\boldsymbol{G}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(\omega)d\omega\right)^2 dz\right)^{\frac{1}{2}}} = \frac{\int\limits_0^1 \widetilde{H}_\vartheta(z)H_{\vartheta_0}(z)dz}{\left(\int\limits_0^1 \widetilde{H}_\vartheta^2(z)dz\right)^{\frac{1}{2}}},$$

where $\widetilde{H}_\vartheta(z) = D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(z) - \int\limits_0^1 D_{\widetilde{\boldsymbol{F}}_\vartheta,\widetilde{\boldsymbol{G}}_\vartheta}(\omega)d\omega,\ z \in (0,1]$ and function $H(\cdot)$ is defined as in the proof of Theorem 16.2.

Since $\widetilde{\vartheta}_0$ is the unique maximizer of 16.2, we conclude with Theorem B.4 that $\widehat{\vartheta}_P \xrightarrow{P} \widetilde{\vartheta}_0$.

If we use $\widehat{\boldsymbol{\Sigma}}_N$ satisfying $\widehat{\boldsymbol{\Sigma}}_N \xrightarrow{P} \boldsymbol{\Sigma}_A$, the proof runs analogously using Lemma 15.2.

17. Simulations

In this chapter, we compare the two statistics concerning the empirical size and the size-adjusted power (see Chapter 6) by means of simulations. Furthermore, we compare the estimators based on both statistics.

We use the setting of a source of gas emission, so we expect a source in an area and we want to locate it as close as possible. The source spreads out gas into the air and the gas is carry away by the wind in its direction. The concentration of the gas in the air can be measured by a plane flying wiggly lines through the gas plume (see Figure 18.1). Inside the plume the concentration of gas is higher than outside. Therefore, we suppose a multivariate epidemic mean change model as in (13.1) with mean zero. The observation has the form $\boldsymbol{X}(t) = (X_1(t), \dots, X_d(t))^T$, where

$$X_i(t) = \delta\Delta_i \mathbb{1}\left\{F_{\vartheta_0}(i) < \frac{t}{N} \leq G_{\vartheta_0}(i)\right\} + e_i(t), \quad i = 1, \dots, d; \quad t = 1, \dots, N,$$

and the parameter δ and $\Delta_1, \dots \Delta_d$ are introduced in Subsection 13.2.2. We choose $d = 6$ and $N = 240$ as well as i.i.d. standard normal distributed errors, and the shape of the plume is assumed to be linear. Thus the covariance matrix $\boldsymbol{\Sigma}$ is equal to the unit matrix and the variance of the projected errors σ^2 is 1. We use this facts for the simulations. The change vector consists of the function h which determines the decay of the concentration of gas with distance. Therefore, as explained in Subsection 13.2.2, we use the log-normal distribution, see its density in Figure 17.1. The first component is weighted by the value of the log-normal density in the x-value 0.09 and the last component in 1.5 because its values of the log-normal density are almost equal. We also project the multivariate data $\boldsymbol{X}(t)$ in direction of the vector resulting of function h.

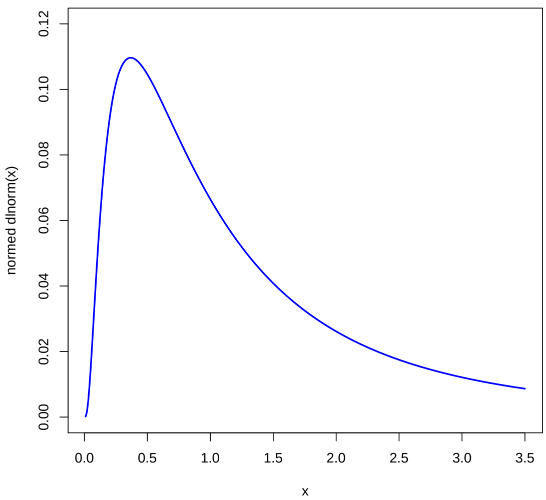


Figure 17.1.: Function h of the decay of the concentration of gas depending on the distance to source.

The limit distributions are approximated by 10000 realisations of time series without changes and the empirical size and size-adjusted power as well as the estimator based on 5000 realisations of multivariate series. The opening angle of the plume is first chosen fixed as 10° and afterwards we do some simulations where we vary the opening angle between a minimum and maximum bound. The unknown parameter ϑ_0 comprised the x- and y-coordinates of the location of source and in addition the opening angle, respectively. The true location of the source has the coordinates $y_s = 120$ and $x_s = 70$. The size of the changes in each component is controlled by the parameter δ, and we choose $\delta = 0.7$, $\delta = 1$ and $\delta = 1.5$. The parameter space Θ is the area in which we search for a source. In the Figures 17.2 and 17.3, illustrating the simulated data set under H_0 and H_1, this area is bounded in y-direction by two black lines with the y-coordinates 105 and 130 and the x-values are between 0 and 130. The parameter β which influences the weight functions is chosen equal to $0, 0.25$ and 0.45 except in case of the estimator based on the projection statistic, where β is equal to $\frac{1}{2}$ anyway caused by the reason of consistency (see Theorem 16.2).

Figure 17.2 shows the time series under the null hypothesis, so there is no source of gas emission in the searching area (the area in which we search for a change), and hence there are no changes in the time series. In Figure 17.3 there is a source from which the gas spreads out in form of a linear plume. Inside the plume the concentration of gas in the air increases and decreases with the distance to source as a log-normal distribution, as shown in Figure 17.1. In Figure 17.3 we used $\delta = 10$ such that the changes are big enough to see them easily.

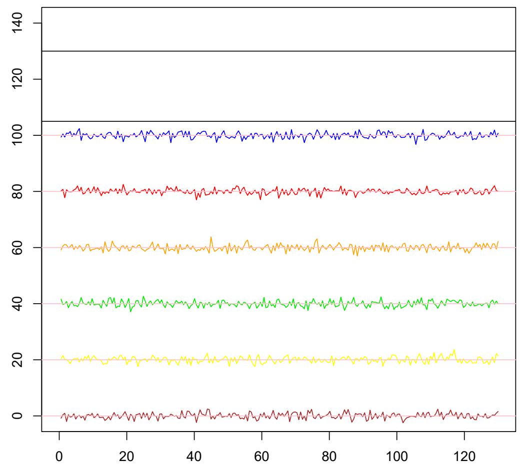


Figure 17.2.: Data set under H_0

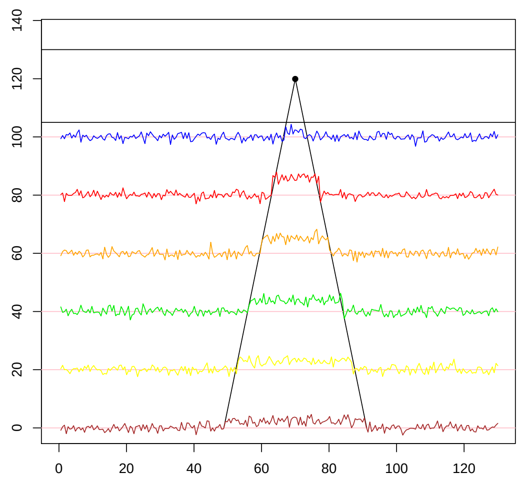


Figure 17.3.: Data set under H_1 with $\delta = 10$

For the procedure with the projection statistic we first suppose the correct specified case, i.e. we use the function h to calculate the statistic which was used to generate the time series. Later we analyse the effect in case of the projection statistic that occurs if we add an error to the changes in the time series while we still use the same function h for the procedure.

17.1. Analysis of empirical Size, Power and Estimators

The empirical size illustrated in Figure 17.4 is very good for both statistics and, no matter which value of β we use, they almost maintain the nominal level. As expectable, the power is higher the bigger the changes are in every component. Moreover, the projection statistic has better power because we use the additional information of employing the correct function h (see Figure 17.5).

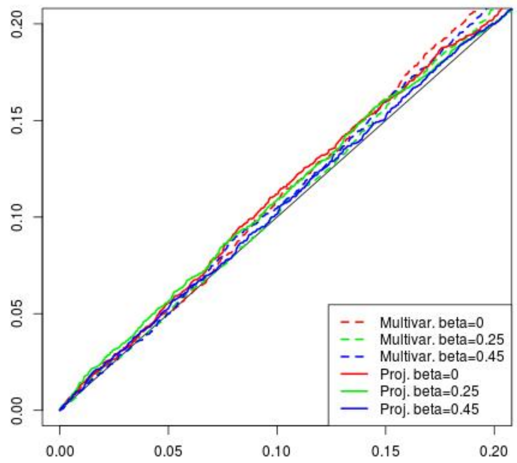


Figure 17.4.: Empirical size

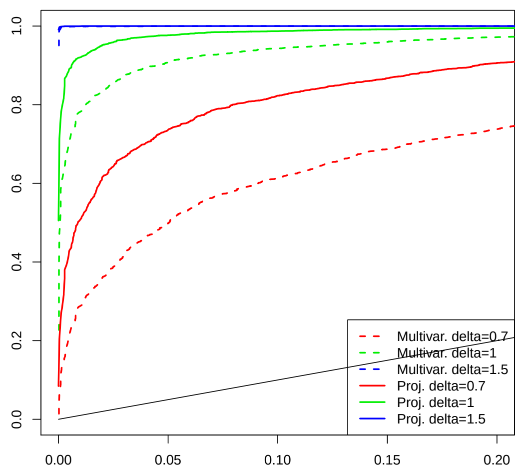


Figure 17.5.: Size-adjusted power with $\beta = 0$ and a source in the middle of the searching area

As seen in Figure 17.6, the power increases from sources near to the lower bound of the searching area to the middle and the upper bound. The reason is that the linear plume is wider for sources located on the upper bound than for sources underneath due to the fixed opening angle. This fact results in a longer period of mean changes in each component, which is more easily to find by the procedure.

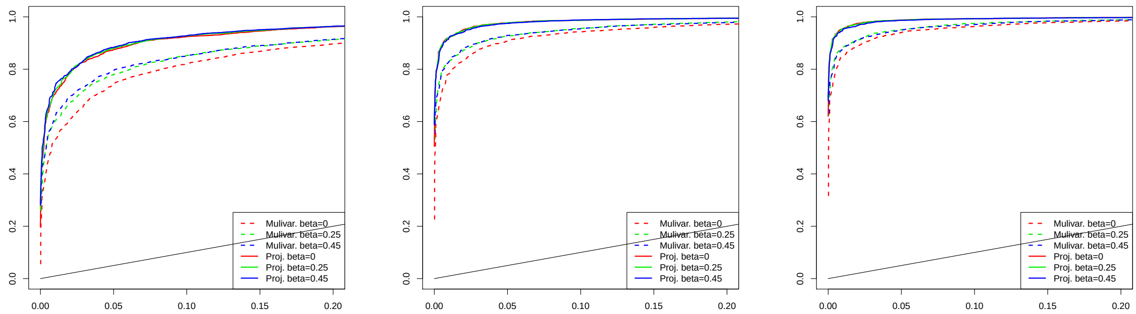
In Figure 17.7 to Figure 17.12 we plotted the estimated plumes of the realisations where the source is detected. Therefore we use the level of 5%. In these figures the black lines and red points correspond to the real plume and to the real location of the source , respectively, and the green lines are the estimated plumes.

Notice that it is not significant to look at the estimated points for the source location, but it is more meaningful to analyse the estimated plumes by using the estimated location of the source. The reason is that a small variation of the plume implicates a relatively great variation in the source location. So we expect that we cannot distinguish between different locations of the sources in y-direction.

Figure 17.7 shows that the estimators based on the multivariate statistic spread over the whole searching area for small δ. With increasing values of δ the estimators get more precise. Based on the projection statistic the variation of estimators are almost the same, independently of the value of δ (see Figure 17.8). In particular, the estimator based on the projection statistic is extremely more precise for small values of δ than the estimator using the multivariate statistic.

The weight functions are influenced by the exponent β. For β near to zero the statistic prefers

Figure 17.6.: Size-adjusted power for different source locations and $\delta = 1$.

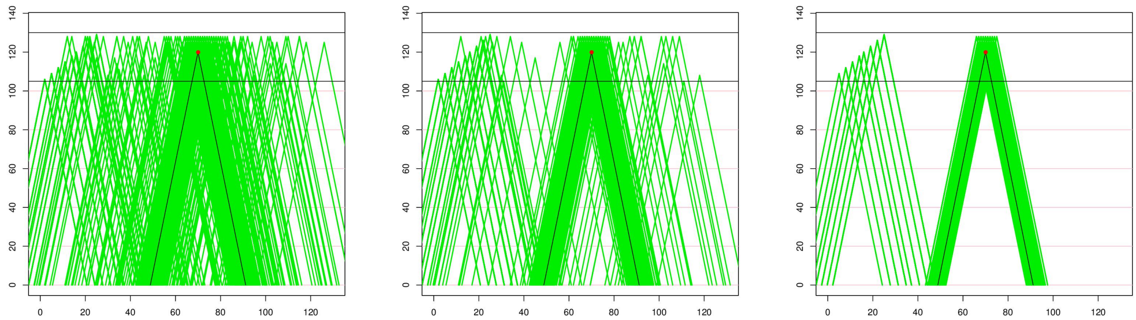


(a) A source near to the lower bound of the searching area.

(b) A source in the middle of the searching area.

(c) A source near to the upper bound of the searching area.

Figure 17.7.: Estimated plumes with multivariate statistic, $\beta = 0$

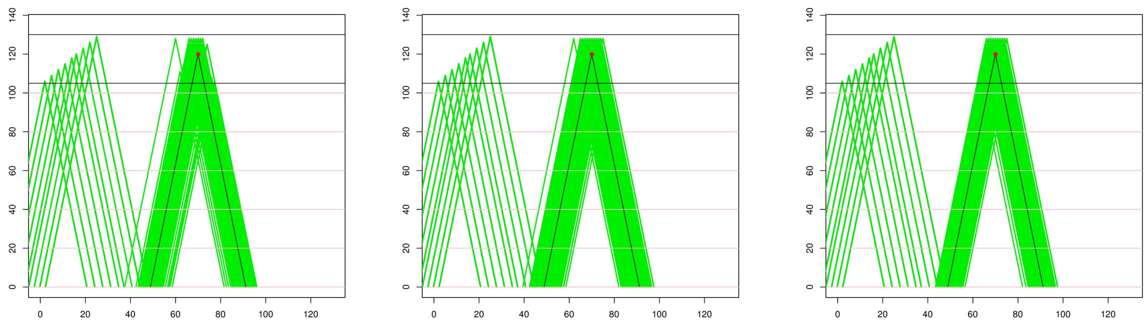


(a) $\delta = 0.7$

(b) $\delta = 1$

(c) $\delta = 1.5$

Figure 17.8.: Estimated sources with projection statistic, $\beta = 0.5$



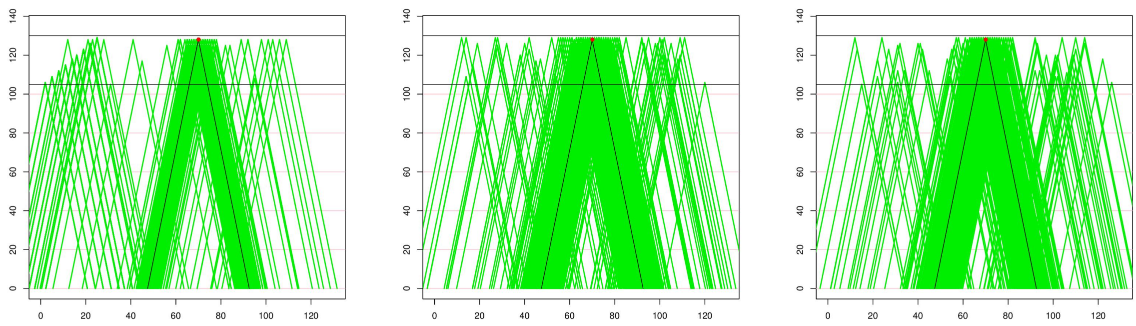
(a) $\delta = 0.7$

(b) $\delta = 1$

(c) $\delta = 1.5$

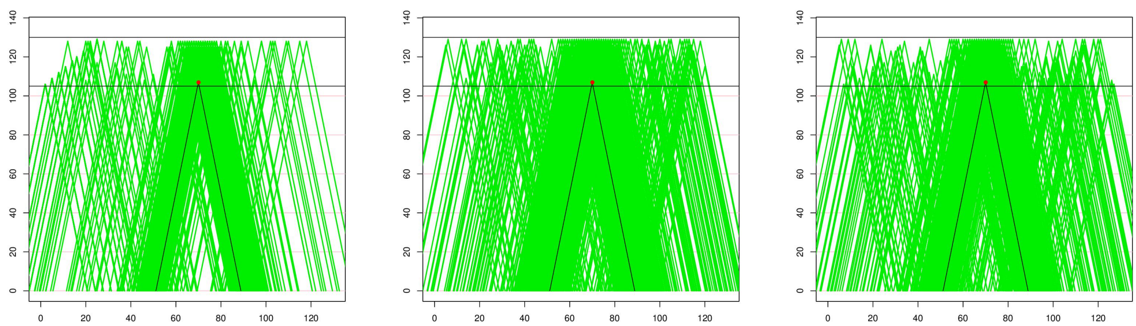
the locations near to the upper bound, because of the resulting wider plume as mentioned above. For an increasing β the locations on the lower bound are preferred. Moreover, the points on the upper region get more weight. However this does not mean that the estimator of the plume is better, as we see in Figures 17.9, 17.10 and 17.11 where the estimators for the plume are best for $\beta = 0$ no matter where the real source is located. Hence it follows that we can actually not distinguish between upper and lower sources as we expected.

Figure 17.9.: Estimators of a source near to the upper bound of the searching area with multivariate statistic ($\delta = 1$)



(a) $\beta = 0$ (b) $\beta = 0.25$ (c) $\beta = 0.45$

Figure 17.10.: Estimators of a source near to the lower bound of the searching area with multivariate statistic ($\delta = 1$)

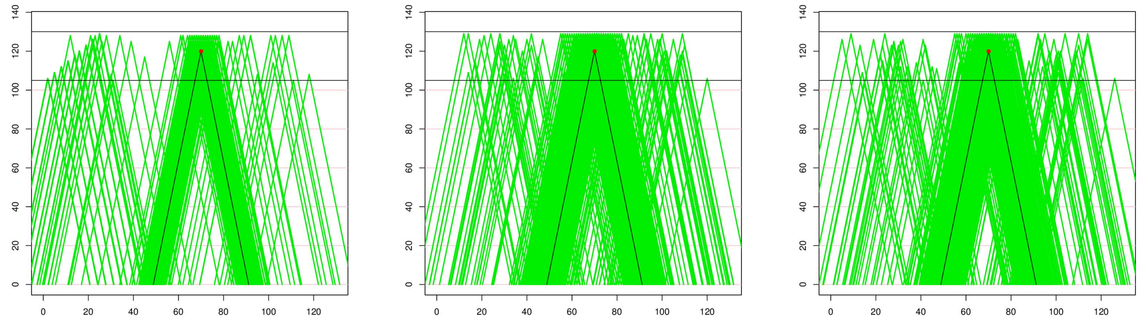


(a) $\beta = 0$ (b) $\beta = 0.25$ (c) $\beta = 0.45$

In case of the projection statistic it is reasonable to use $\beta = 0.5$, because then the estimator is consistent. If we search a source near to the upper bound, the lower bound or in the middle of the searching area, the variation of the estimators of the plume based on the projection statistic is in all cases almost the same (see Figure 17.12).

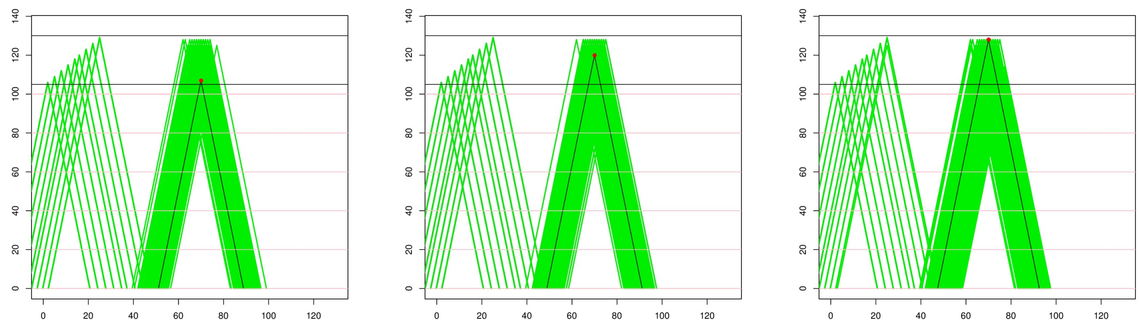
In particular, the estimators of the plumes do not restrict the y-coordinates of the estimated sources because the surface of $\boldsymbol{S}_\vartheta^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_\vartheta$ and $w_P(\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta)^{-1} \left| \sum_{t=1}^{N} D_{\boldsymbol{F}_\vartheta, \boldsymbol{G}_\vartheta}\left(\frac{t}{N}\right)\left(Y(t) - \overline{Y}_N\right)\right|$ with $\beta = 1/2$ is very flat along the y-axis as we see in the heatmaps in the Figures 18.18 and 18.19.

Figure 17.11.: Estimators of a source in the middle of the searching area with multivariate statistic ($\delta = 1$)



(a) $\beta = 0$ (b) $\beta = 0.25$ (c) $\beta = 0.45$

Figure 17.12.: Source estimators with the projection statistic ($\delta = 1, \beta = 0.5$)



(a) Source near to the lower bound of the searching area (b) Source in the middle of the searching area (c) Source near to the upper bound of the searching area

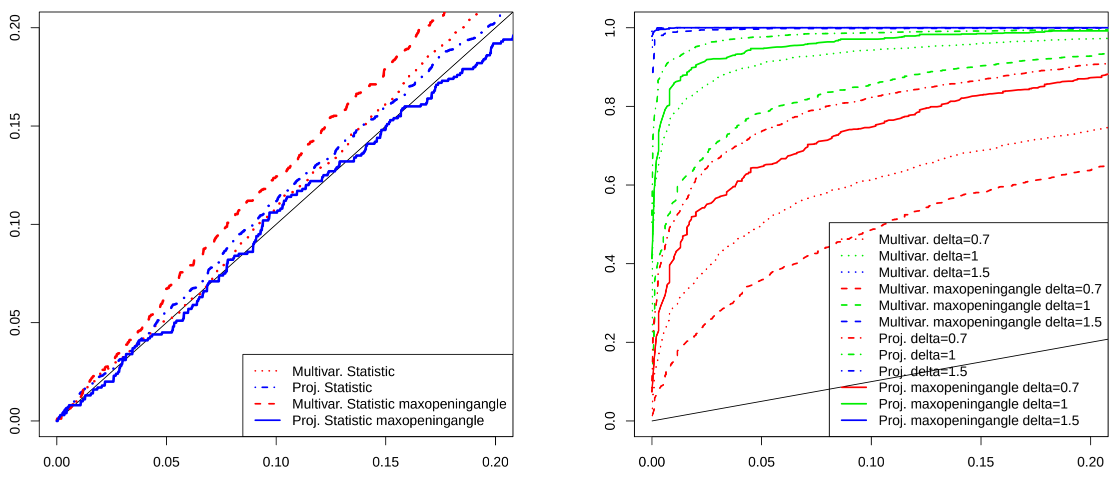
The heatmaps are plotted from all possible estimators of the source location based on the landfill data set, which will be discussed in the next chapter. This flat trend along the y-axis is due to the fact that a slight modification of the plume in even only one component can implicate a relatively great movement in the source location.

So far we analysed the procedure with a linear plume where the opening angle is fixed. But in reality we maybe do not exactly know the opening angle, thus next we allow to vary the opening angle of the plume between $1°$ and $89°$.

The limit distribution of the multivariate statistic is approximated by 5000 replications of generated series without a source, the critical values of the projection statistic are based on 2000 replications. The simulations to calculate the empirical size and size-adjusted power of the multivariate statistic are based on 5000 replications and of the projection statistic based on 1000 replications.

As we see in Figure 17.13(a), the empirical size of both statistics is somewhat worse than without maximizing the opening angle but still very good. The size-adjusted power in Figure 17.13(b) is higher with a fixed opening angle than with maximizing over the angle because, in the second case the method has to estimate an additional parameter. Since by using the multivariate statistic the best choice for β is 0 with the above results, we illustrate the plume estimators with this choice of β. In case of the projection statistic we use again $\beta = 0.5$ to get a consistent estimator. The estimators of the plumes are worse in case of the multivariate statistic with the same choices of δ than with fixed opening angle as we see in Figure 17.14. The estimators based on the projection statistic are more precise but expectably worse than with fixed opening angle, see Figure 17.15. Next we have a closer look at the projection statistic.

Figure 17.13.: Procedures with $\beta = 0$ and maximizing over the opening angle (which is labelled with "maxopeningangle" in the caption)

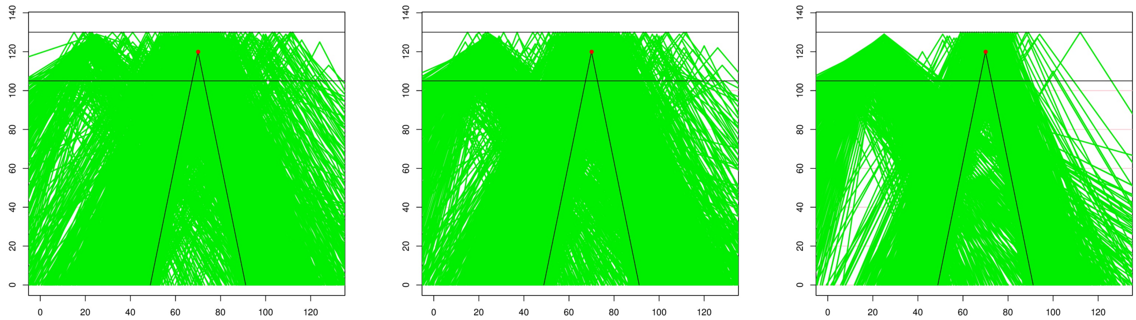


(a) Empirical size

(b) Size-adjusted power ("delta = c" means $\delta = c$)

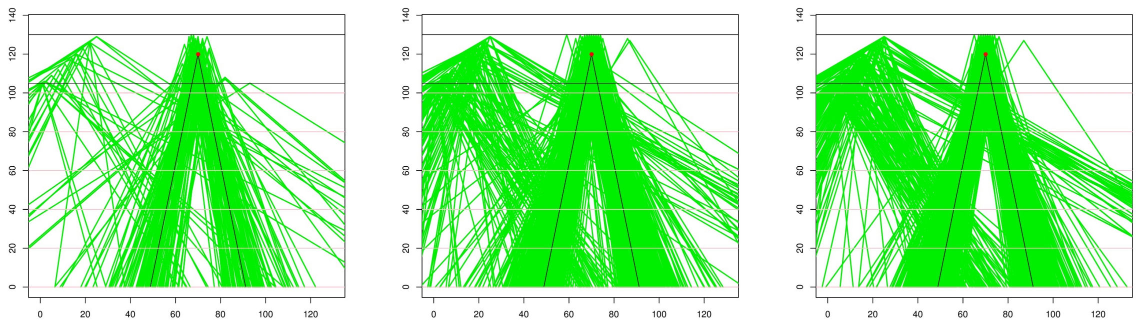
Till now we assumed that we know the function h, which gives the decay of the concentration of gas in the air with distance. But in practice we maybe do not know the function h exactly. Thus we want to clarify, how does the procedure using the projection statistic still performs under misspecification of h. Therefore, we add a normally distributed error with mean zero and

Figure 17.14.: Estimated plumes with multivariate statistic with $\beta = 0$ and maximizing over the opening angle



(a) $\delta = 0.7$ (b) $\delta = 1$ (c) $\delta = 1.5$

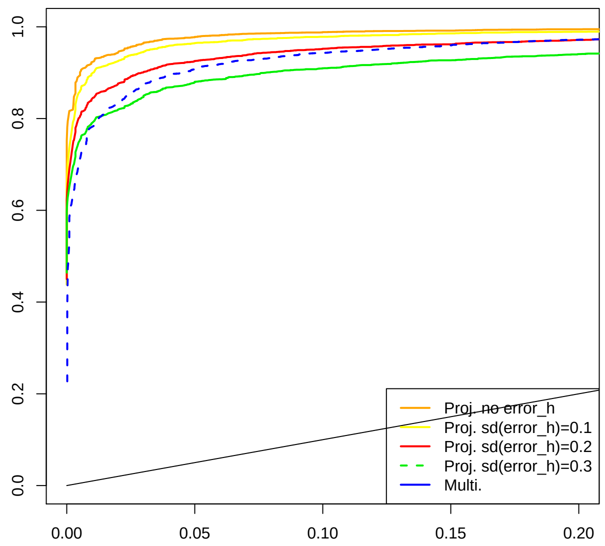
Figure 17.15.: Estimated plumes with projection statistic with $\beta = 0.5$ and maximizing over the opening angle



(a) $\delta = 0.7$ (b) $\delta = 1$ (c) $\delta = 1.5$

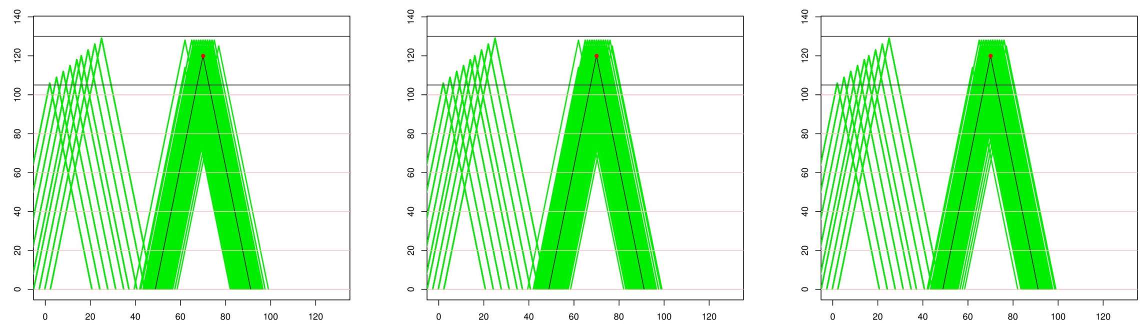
standard deviation $0.1, 0.2$ and 0.3 to the true decay in each component in every simulated time series and still use the function h in Figure 17.1 for the calculation of the projection statistic. The power is illustrated in Figure 17.16. It decreases with growing deviation between the applied function h for calculating the statistic and the true decay of concentration of the data. Till the standard deviation of 0.2, the projection statistic has still higher power than the multivariate statistic. The plume estimators are very stable with respect to the added error (see Figure 17.17). The deviation between the applied function h and the true decay of the concentration has almost no effect on the estimators by comparing Figure 17.8 and Figure 17.17, even the increasing standard deviation of the added error term remains without consequence (see Figure 17.17). In summary, the projection statistic performs better concerning the size-adjusted

Figure 17.16.: Size-adjusted power of the projection statistic with $\beta = 0.5$ and an added error to the changes in each component ($\delta = 1$)
("sd(error_h)" in the caption means the standard deviation of the added error)



power and the estimators due to the use of additional information of knowing the decay of the gas concentration with distance. However, it still performs quite well under misspecification of the function h which determines the decay of the concentration of gas.

Figure 17.17.: Estimated plumes with projection statistic with $\beta = 0.5$ and an added error to the changes in each component ($\delta = 1$)



(a) Standard deviation of the error 0.1

(b) Standard deviation of the error 0.2

(c) Standard deviation of the error 0.3

18. Landfill Data

In this chapter we apply our methods to real data. The data set is composed of measured concentrations of methane gas in the air along a flight path of a plane. The methane gas is emitted by two landfills. The plane flies above the area where the sources of gas emission are expected in wiggly lines vertical to the wind direction, as illustrated in Figure 18.1. So in the situation of Figure 18.1 the wind comes from direction north-east. The concentration of methane gas in the air is measured every second. Because of two existing landfills in the searching area we first have to separate the complete trajectory into the blue trajectory and the red trajectory, as in Figure 18.1, to apply our procedure. In Figure 18.2 we see the left-hand flight path and the corresponding landfill as black pigmented area. Since there is not only a single point as a source of gas emission but an area in which every point emits gas, we have to approximate the arising gas plume such that the landfill area is tangented by the plume, as illustrated in Figure 18.2. Thus the origin of the plume has a somewhat greater y-value than the landfill points themselves, and hence we want to narrow down the area in which the landfill is contained.

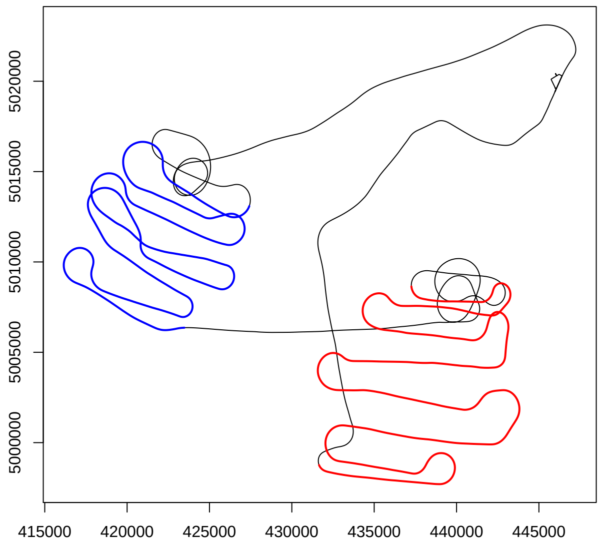


Figure 18.1.: Trajectory

Before we can apply our methods to the data set we have to transform the data of each flight path in the following way in order to obtain a multivariate time series:

1. Only the coordinates of the points from the flight path in Figure 18.1 are available where the wind direction comes from the north-east. Consequently the corresponding trajectory to each landfill area is sloped. Thus we first rotate the trajectory such that the wind direction is parallel to the y-axis.

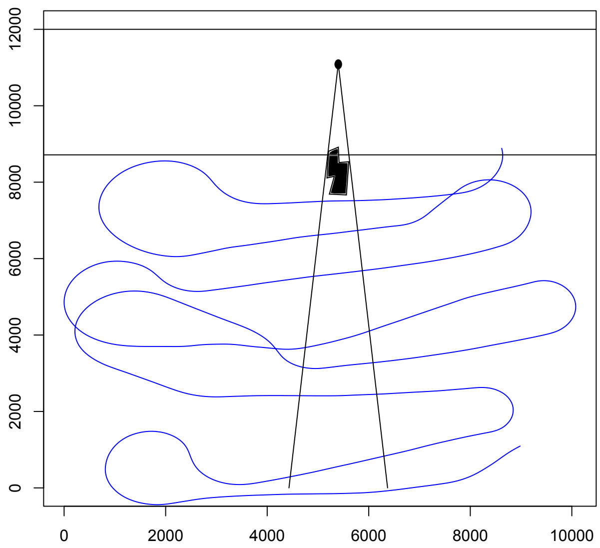


Figure 18.2.: Left-hand trajectory with the landfill area and the corresponding approximated plume

2. Next we split the flight path where the plane changes its course. The result are separate lines in approximately horizontal direction.

3. Since these lines are only approximately horizontal, we have to average the y-values of each resulting line (see Figure 18.3) which represent the components of the multivariate time series.

4. In reality we do not have equidistant measurement points where we could easily connect the points with the same x-values to a multivariate data point of the time series. We have different lags between the measurement points in each component because the plane measures the gas concentration every second and does not fly at a constant speed. Thus we have to transform the x-values as well as the values of the concentration. Additionally, as we see in Figure 18.3, the averaged horizontal lines have different length. Consequently we create a new grid for the trajectory.

   So we take the average lag from all measurement points of the whole trajectory as the lag of the transformed x-values for the multivariate series. The starting point of the multivariate time series is the smallest x-value of all measurement points. Then the next new x-values are obtained by adding the average lag. In Figure 18.4 the black points are the real measurement points, and the brown vertical lines are the transformed x-values of the multivariate series. Note that, between any two vertical lines in the new grid, there may be none, one, or more than one measurement points (see Figure 18.5). Hence the next step is to assign the concentration values to the transformed x-values. We calculate the mean of the points between two lines and allocate the averaged concentration to the next transformed x-value on the right-hand side. If there are no points between the lines, we fill up with the previous value.

Since the horizontal lines, representing the different components of the multivariate series, have different length we fill the missing values with zeros. However, by the zeros we produce artificial changes in the multivariate time series, so we search for a source within the region where the corresponding plumes have no missing values, in order to shorten the run-time of the written code. A way of taking those plumes into account which, at least in one component, includes zeroes due to the missing values, one can sum up the non-zero values between the time points they exist.

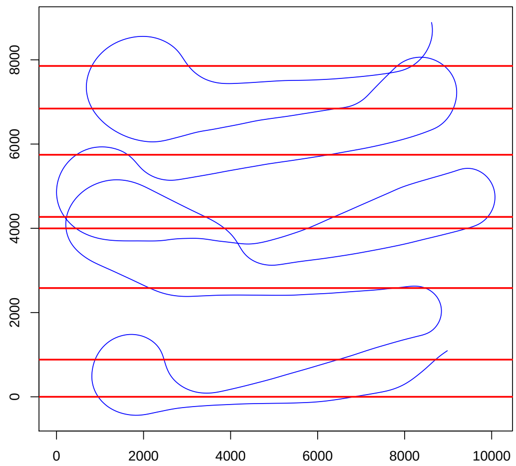


Figure 18.3.: The mean of the y-values of the horizontal lines of the left-hand trajectory

In Figure 18.6 we see the modified data of the left-hand trajectory after the transformation steps explained above. Obviously, in the middle of each line there is a gradual epidemic change in the structure of the gas concentration. We approximate the form of the gas plume as a linear shape, where the concentration of gas increases inside the plume. Besides the coordinates of the source, the opening angle is unknown. So equivalent to the second part of the simulation study we have to maximize over all possible coordinates of the source location and, additionally, of the opening angle.

To calculate the projection statistic we assume the decay of the gas concentration in the air with distance to the source is equal to the function h as in Figure 17.1. Of course, this is only an approximation for the reality which, however, is no problem because we have seen in the simulation study that the method using the projection statistic is very stable against approximation errors for the function h.

Our procedure focuses on an abrupt epidemic mean change in each component and we will see that our procedure works very well, even if we assume an abrupt mean change instead of a gradual mean change or some other kinds of changes like a variance change.

Since we have to estimate the long-run covariance matrix to calculate the statistics, we next analyse the dependence structure of the modified data. In Figures 18.7,18.9,18.11 and 18.12 the estimated residuals and consequently the modified data seem to be independent in their components but dependent in the x-direction (see Figures 18.8,18.10,18.13,18.14). Caused by the dependence in the x-direction and independence in the components, we assume that the long-run covariance matrix is a diagonal matrix and we only have to estimate the long-run

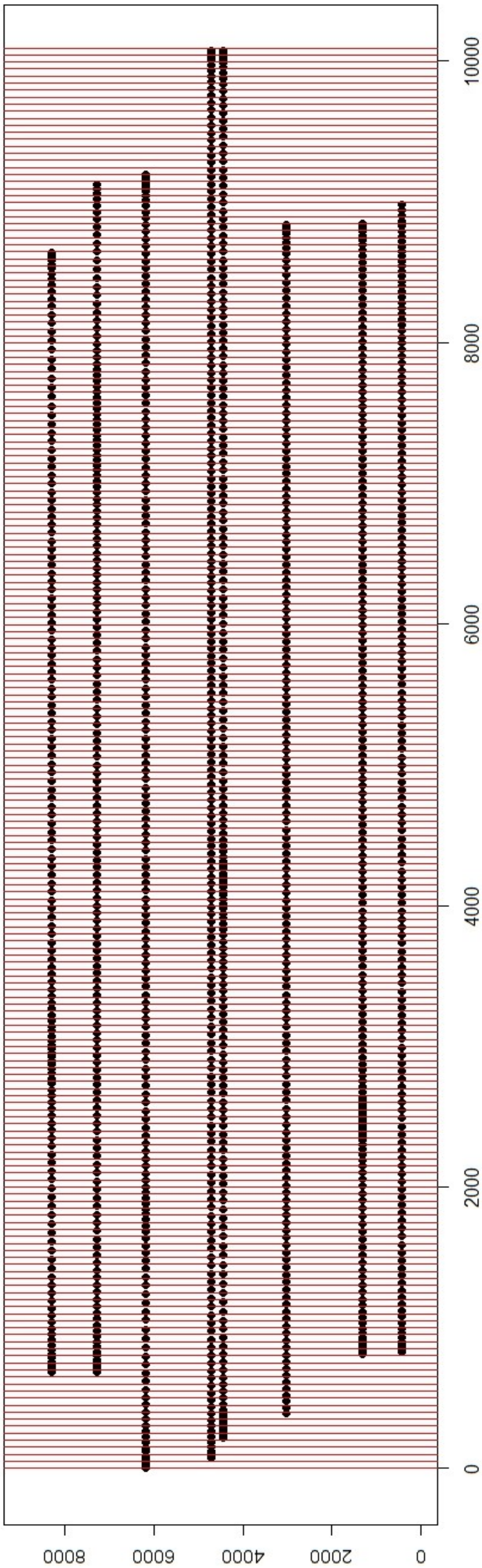


Figure 18.4.: The created grid of the left-hand trajectory.
The vertical equidistant lines are the x-values of the created grid and the points are received by the steps 1.-3.

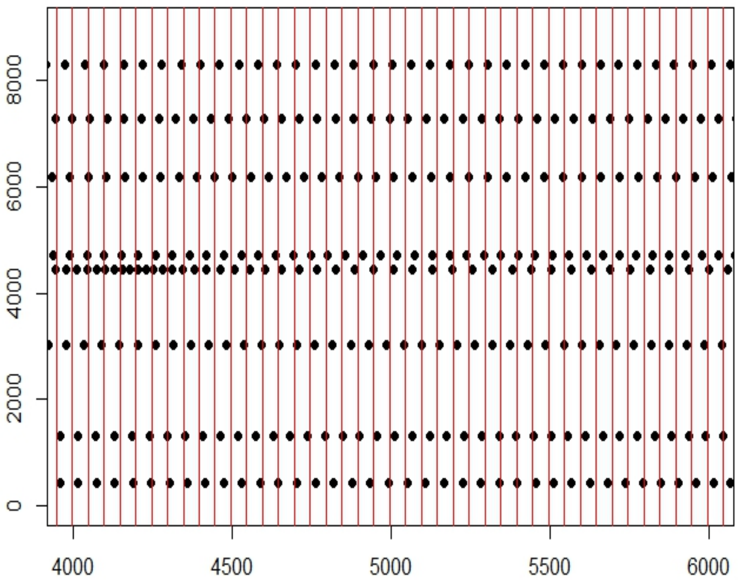


Figure 18.5.: Detail of the figure on the left with x-values between 4000 and 6000

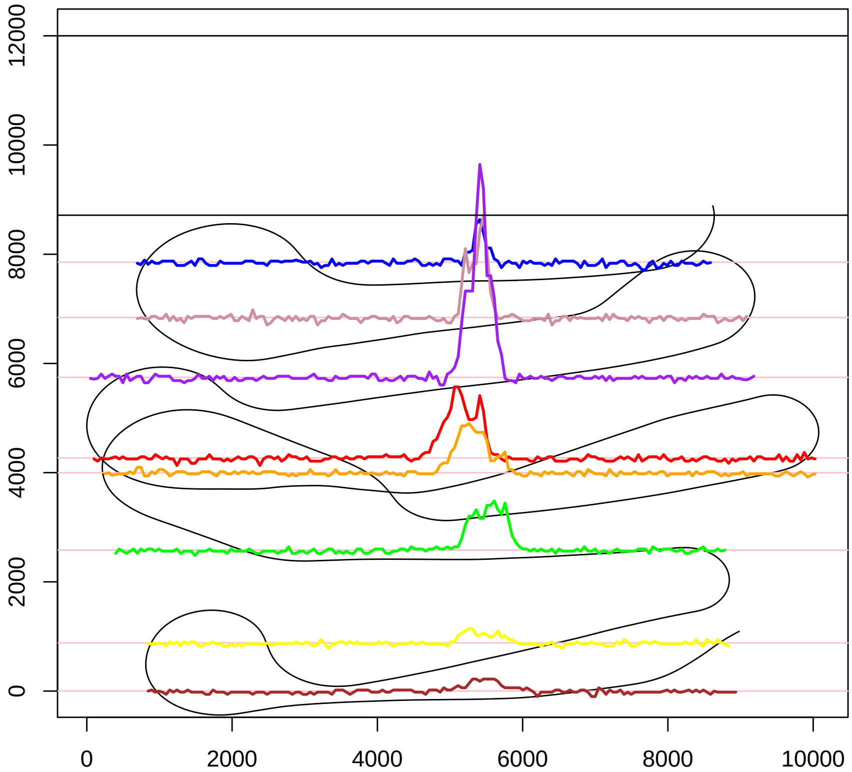


Figure 18.6.: The modified landfill data of the left-hand trajectory

variances. Thus the long-run covariance matrix is $\mathbf{\Sigma} = \mathrm{diag}(\Sigma_{1,1}, \ldots, \Sigma_{d,d})$, where

$$\Sigma_{i,i} = \sum_{h \in \mathbb{Z}} \mathrm{Cov}\left(e_i(0), e_i(h)\right), \quad i = 1, \ldots, d.$$

To estimate the long-run covariance matrix we need the estimated errors. They are calculated as follows:

$$\left(\widehat{f}_i, \widehat{g}_i\right) = \arg\max \left\{ \left| \sum_{t=f_i+1}^{g_i} \left(X_i(t) - \overline{X}_{i,N}\right) \right| : 1 \leq f_i < g_i \leq N \right\},$$

where

$$\widehat{\mu}_i = \frac{1}{\widehat{f}_i + N - \widehat{g}_i} \left(\sum_{t=1}^{\widehat{f}_i} X_i(t) + \sum_{t=\widehat{g}_i+1}^{N} X_i(t) \right)$$

and

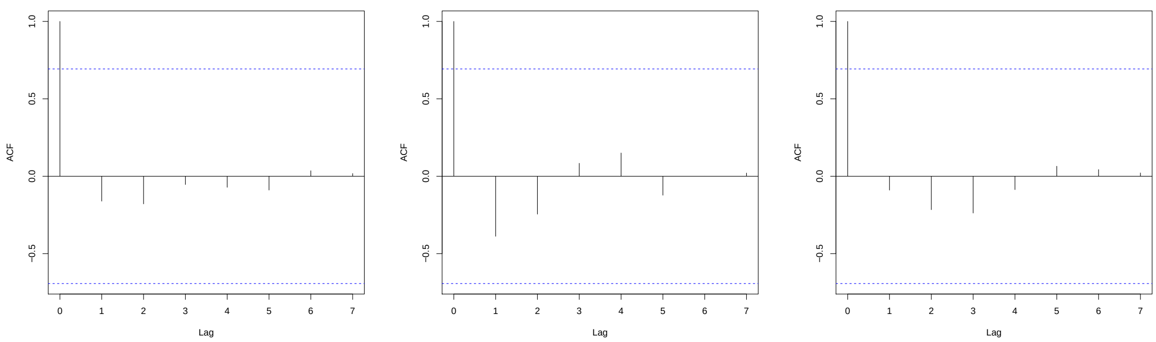
$$\widehat{\Delta}_i = \frac{1}{\widehat{g}_i - \widehat{f}_i} \sum_{t=\widehat{f}_i+1}^{\widehat{g}_i} \left(X_i(t) - \hat{\mu}_i\right).$$

The estimated errors are defined as

$$\hat{e}_i(t) = X_i(t) - \widehat{\mu}_i - \widehat{\Delta}_i \mathbb{1}\left\{ \widehat{f}_i < t \leq \widehat{g}_i \right\}.$$

Then we apply our methods to the data set. The real wind direction is 33°, which is important

Figure 18.7.: ACF's of the columns of the estimated errors of the left-hand trajectory



(a) Estimated errors column 1 of the left-hand trajectory.

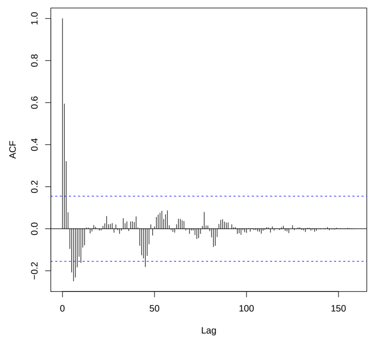
(b) Estimated errors column 100 of the left-hand trajectory

(c) Estimated errors column 200 of the left-hand trajectory

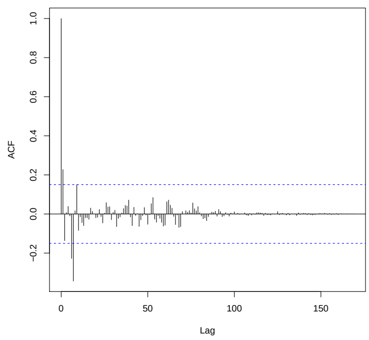
for the first step of the transformation where we rotate the coordinate system about 33°.

First we consider the left-hand trajectory. In Figure 18.15(a) the multivariate statistic is used and we see clearly that the estimator of the plume is very good because the procedure detects the epidemic changes in each component very precisely. If we use the projection statistic the epidemic changes are found well, especially in component 2 and 3 (see 18.15(b)). The reason is that these components are most weighted by the function h, as we see in Figure 17.1. The estimated source of the plume is somewhat too near at the lower bound of the searching area. This effect arises because of the very flat structure of the statistic along the y-axis, as illustrated

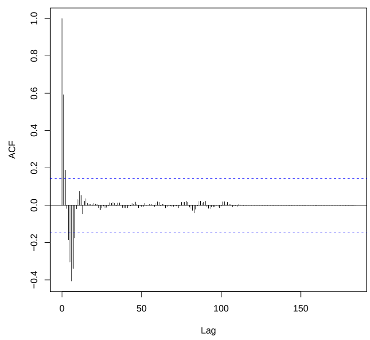
Figure 18.8.: ACF's of the rows of the estimated errors of the left-hand trajectory.



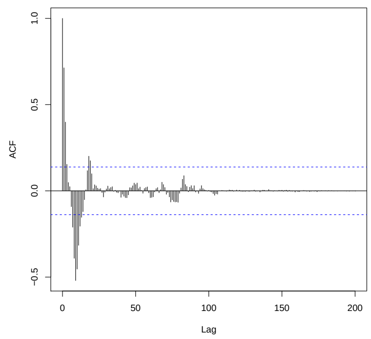
(a) Estimated errors row 1 of the left-hand trajectory



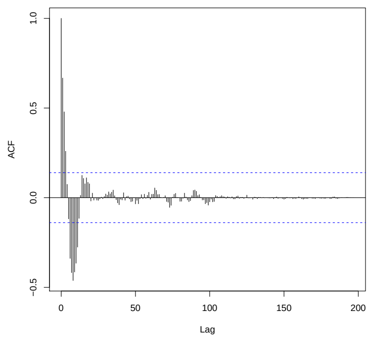
(b) Estimated errors row 2 of the left-hand trajectory



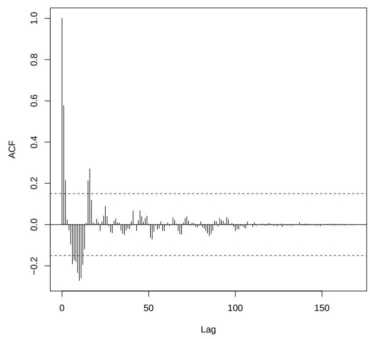
(c) Estimated errors row 3 of the left-hand trajectory



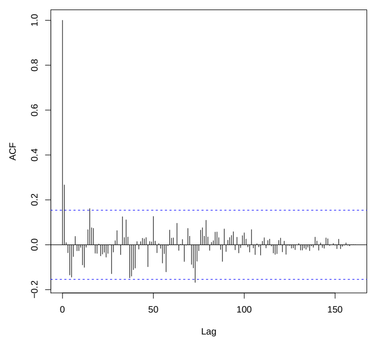
(d) Estimated errors row 4 of the left-hand trajectory



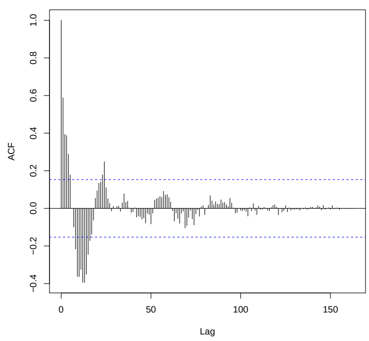
(e) Estimated errors row 5 of the left-hand trajectory



(f) Estimated errors row 6 of the left-hand trajectory

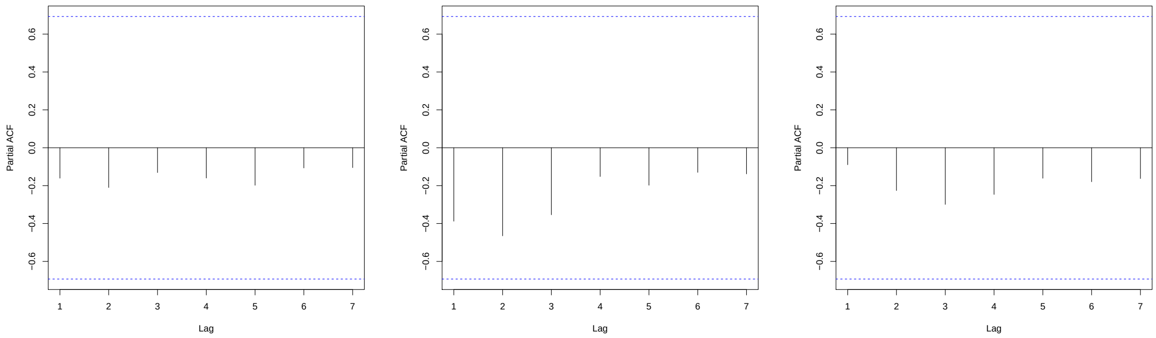


(g) Estimated errors row 7 of the left-hand trajectory



(h) Estimated errors row 8 of the left-hand trajectory

Figure 18.9.: PACF's of the columns of the estimated errors of the left-hand trajectory



(a) Estimated errors column 1 of the left-hand trajectory.

(b) Estimated errors column 100 of the left-hand trajectory

(c) Estimated errors column 200 of the left-hand trajectory

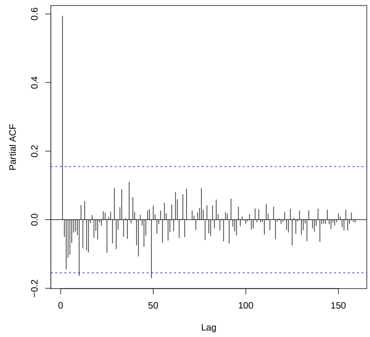
in Figure 18.18 and thus we cannot reasonably distinguish between the sources along the y-axis as explained in the simulation study.

To estimate the corresponding plume of the right-hand trajectory is more complex. In Figure 18.16 first note that the right-hand trajectory is inclined to the other direction as before after rotating about the degrees of 33° according to the wind direction. The reason is that the wind direction has changed during the flight from the left-hand to the right-hand trajectory, as well as even during the flight of the right-hand trajectory again. Consequently, the approximation of the y-values by the averaged horizontal lines is very imprecise and the epidemic changes in the different components are not one below each other. Thus there is no linear plume in wind direction which can find all changes almost correctly. However, as we see in Figure 18.16, the procedure locates the area in which the source is located quite well.

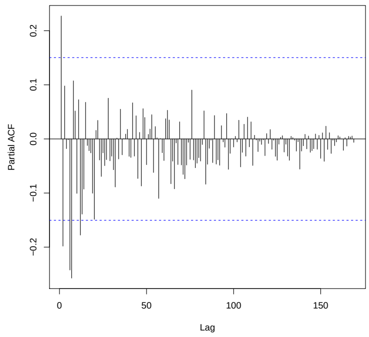
Hirst et al. (2013) apply a Bayesian inference to locate sources of gas emission, and they estimate the emission rate of the sources. The same landfill data set is analysed in this paper. With the Bayesian approach there are two sources found in the right-hand flight path, and the wrongly located source has a higher estimated value for the emission rate. This happens due to the change of the wind direction during the flight path. So even if we assume that Hirst et al. (2013) know that there is only one possible source in the area, they would choose the source with higher emission rate which, however, does not exist in reality.

Usually the rotation angle for the complete flight path corresponding to a certain source is equal to the wind direction, measured in degrees. Due to the change in the wind direction while the plane flies along the right-hand trajectory, we choose another rotation angle. We rotate the first four components about 28° and the other components about 23°. The approximation of the averaged horizontal lines corresponding to the components of the time series are somewhat more precise than before and now the epidemic changes are ordered among each other. Thus there exists a linear plume which is capable to detect the changes in every component nearly correct. The new rotated trajectory with the corresponding estimated plumes is illustrated in Figure 18.17. The data in the right-hand flight path has in addition two peaks during the epidemic change period in the first four components, so both statistics concentrate on the higher peak. The estimated location of the source by the projection statistic is near to the lower bound of the searching area and by the multivariate statistic it is near to the upper bound due to the flat structures of the values of the possible estimators in y-direction (see Figure 18.19).

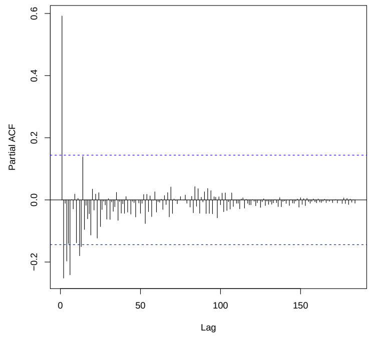
Figure 18.10.: PACF's of the rows of the estimated errors of the left-hand trajectory.



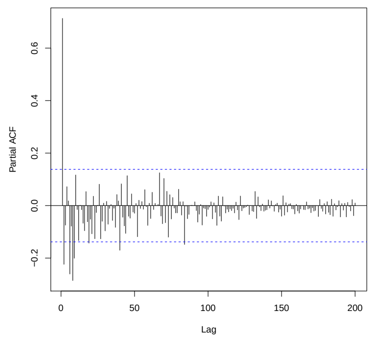
(a) Estimated errors row 1 of the left-hand trajectory



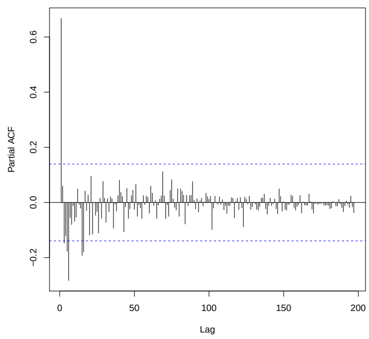
(b) Estimated errors row 2 of the left-hand trajectory



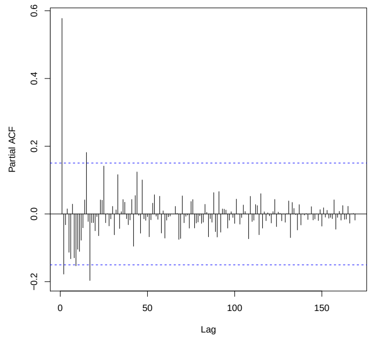
(c) Estimated errors row 3 of the left-hand trajectory



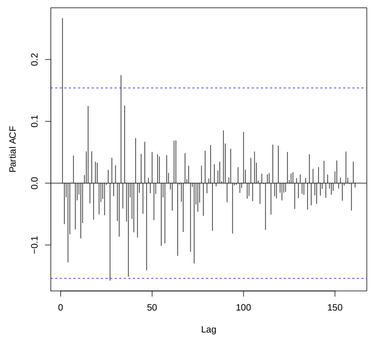
(d) Estimated errors row 4 of the left-hand trajectory



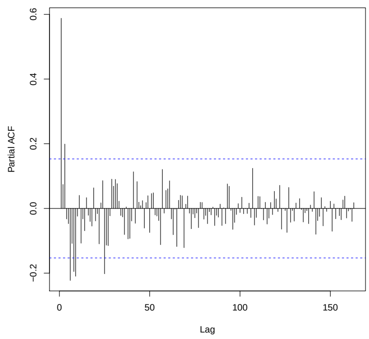
(e) Estimated errors row 5 of the left-hand trajectory



(f) Estimated errors row 6 of the left-hand trajectory

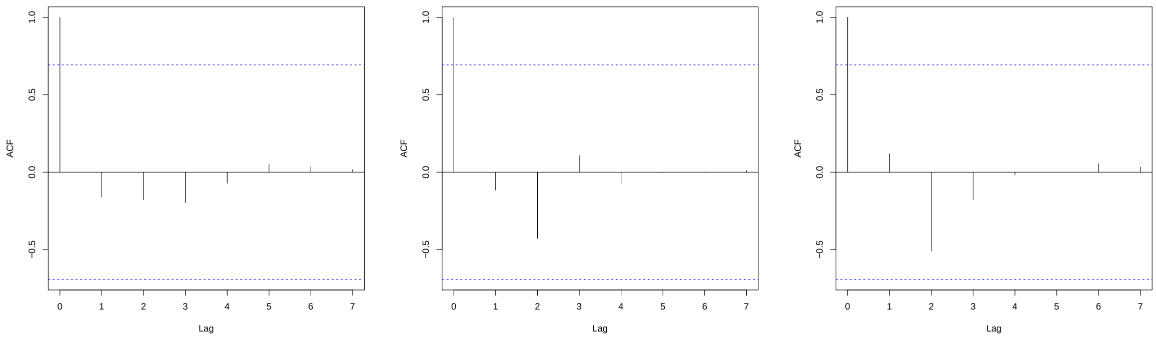


(g) Estimated errors row 7 of the left-hand trajectory



(h) Estimated errors row 8 of the left-hand trajectory

Figure 18.11.: ACF's of the columns of the estimated errors of the right-hand trajectory

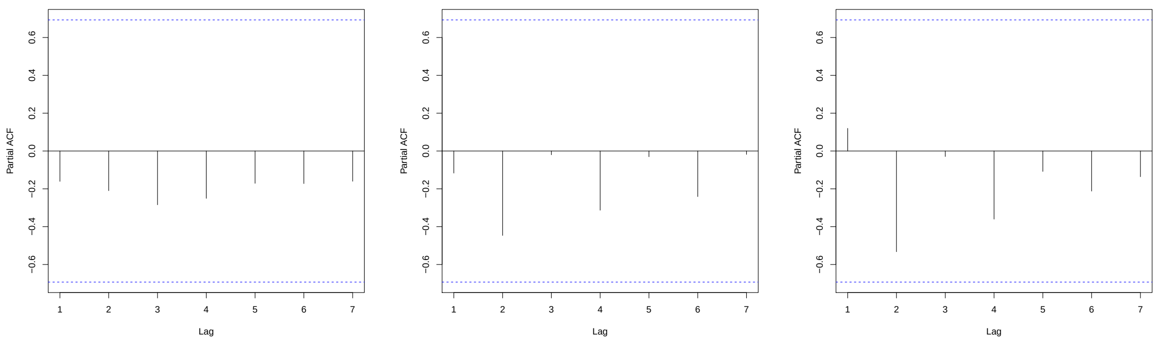


(a) Estimated errors column 1 of the right-hand trajectory.

(b) Estimated errors column 100 of the right-hand trajectory

(c) Estimated errors column 200 of the right-hand trajectory

Figure 18.12.: PACF's of the columns of the estimated errors of the right-hand trajectory

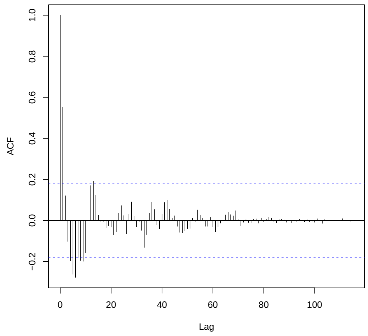


(a) Estimated errors column 1 of the left-hand trajectory.

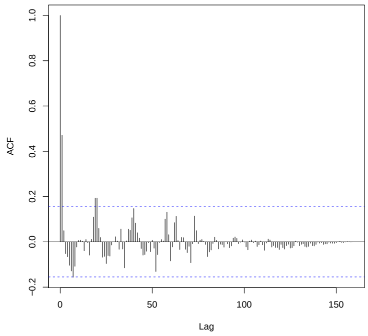
(b) Estimated errors column 100 of the right-hand trajectory

(c) Estimated errors column 200 of the right-hand trajectory

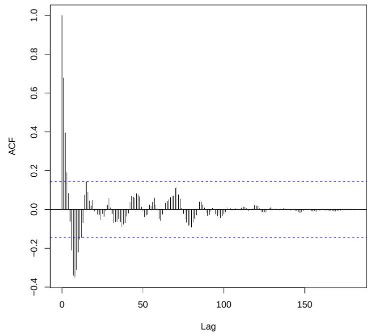
Figure 18.13.: ACF's of the rows of the estimated errors of the right-hand trajectory.



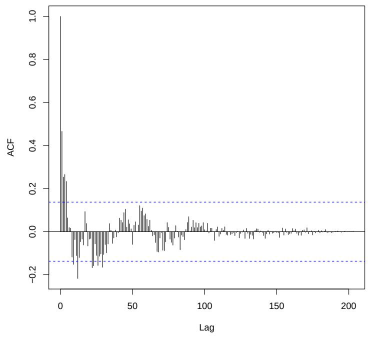
(a) Estimated errors row 1 of the right-hand trajectory



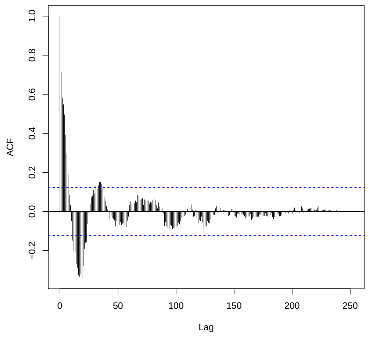
(b) Estimated errors row 2 of the right-hand trajectory



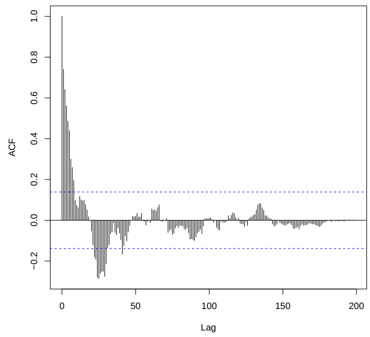
(c) Estimated errors row 3 of the right-hand trajectory



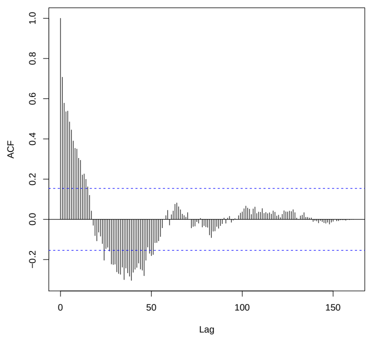
(d) Estimated errors row 4 of the right-hand trajectory



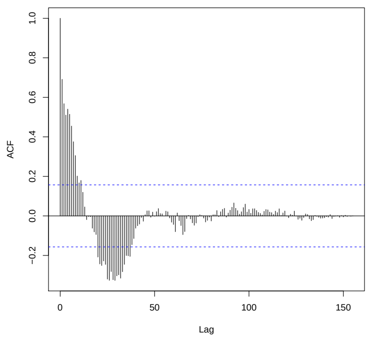
(e) Estimated errors row 5 of the right-hand trajectory



(f) Estimated errors row 6 of the right-hand trajectory

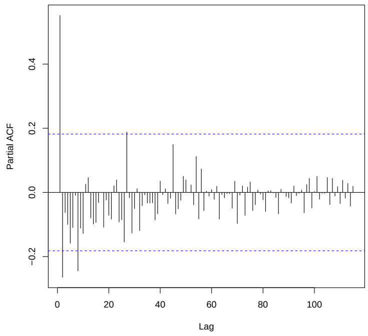


(g) Estimated errors row 7 of the right-hand trajectory

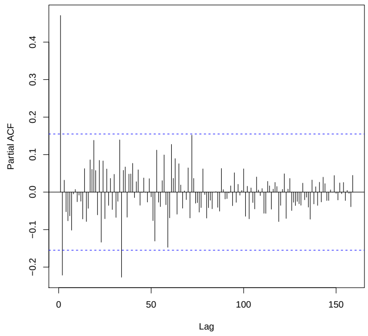


(h) Estimated errors row 8 of the right-hand trajectory

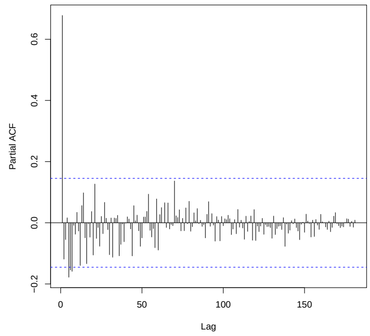
Figure 18.14.: PACF's of the rows of the estimated errors of the right-hand trajectory.



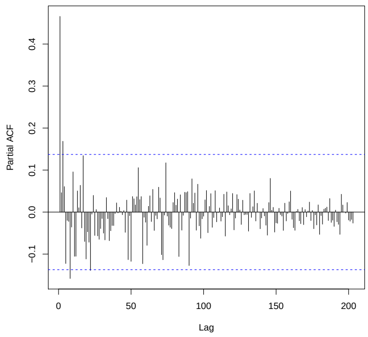
(a) Estimated errors row 1 of the right-hand trajectory



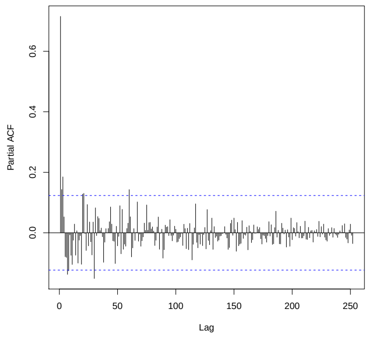
(b) Estimated errors row 2 of the right-hand trajectory



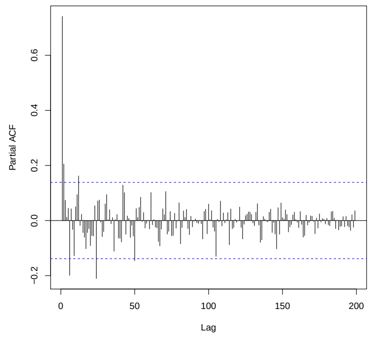
(c) Estimated errors row 3 of the right-hand trajectory



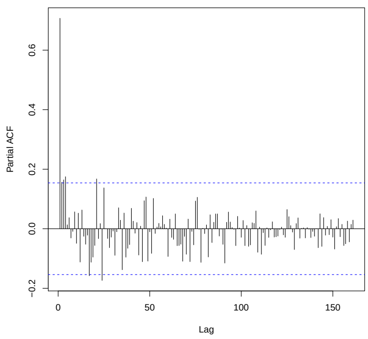
(d) Estimated errors row 4 of the right-hand trajectory



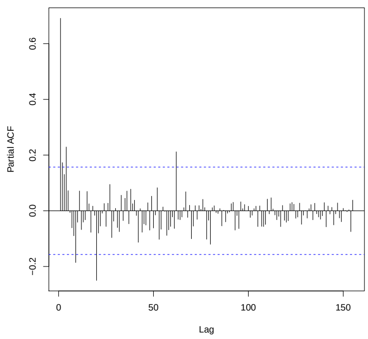
(e) Estimated errors row 5 of the right-hand trajectory



(f) Estimated errors row 6 of the right-hand trajectory

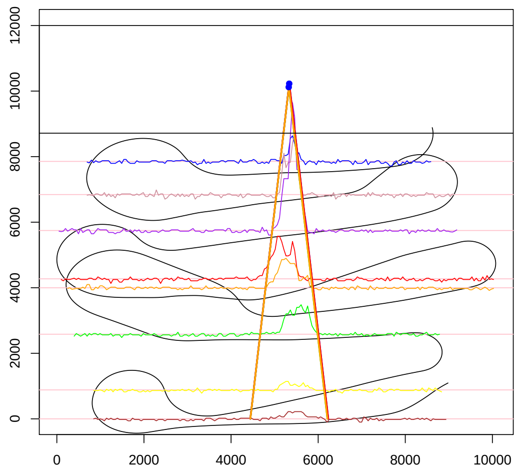


(g) Estimated errors row 7 of the right-hand trajectory

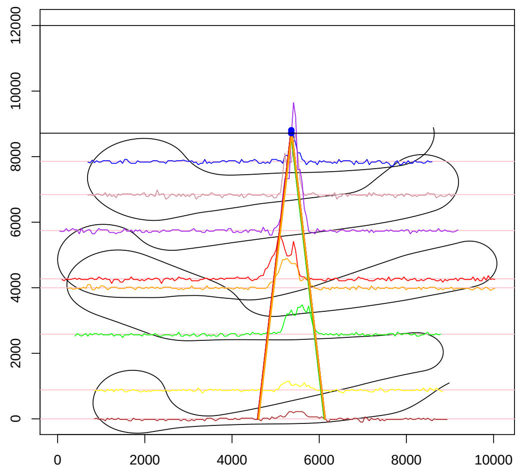


(h) Estimated errors row 8 of the right-hand trajectory

Figure 18.15.: Left-hand trajectory with maximizing the opening angle.

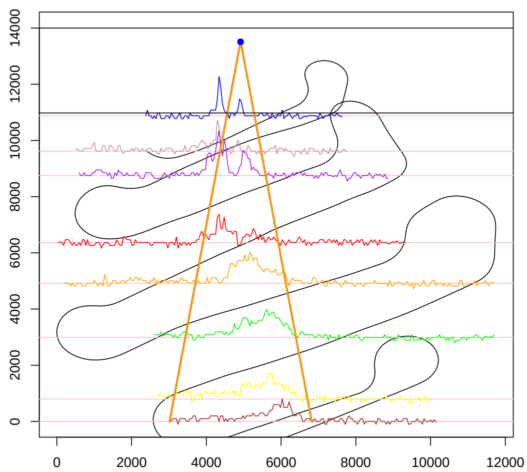


(a) The use of the multivariate statistic

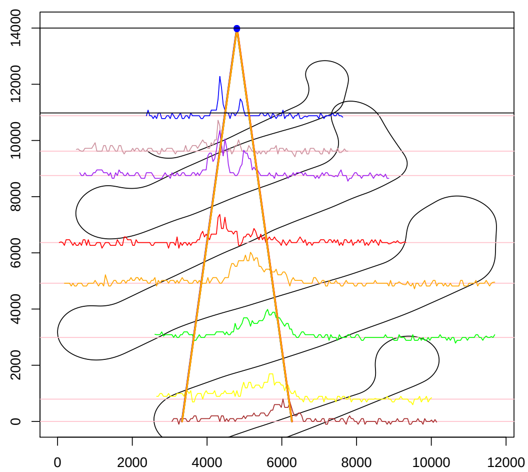


(b) The use of the projection test statistic

Figure 18.16.: Right-hand trajectory with maximizing the opening angle

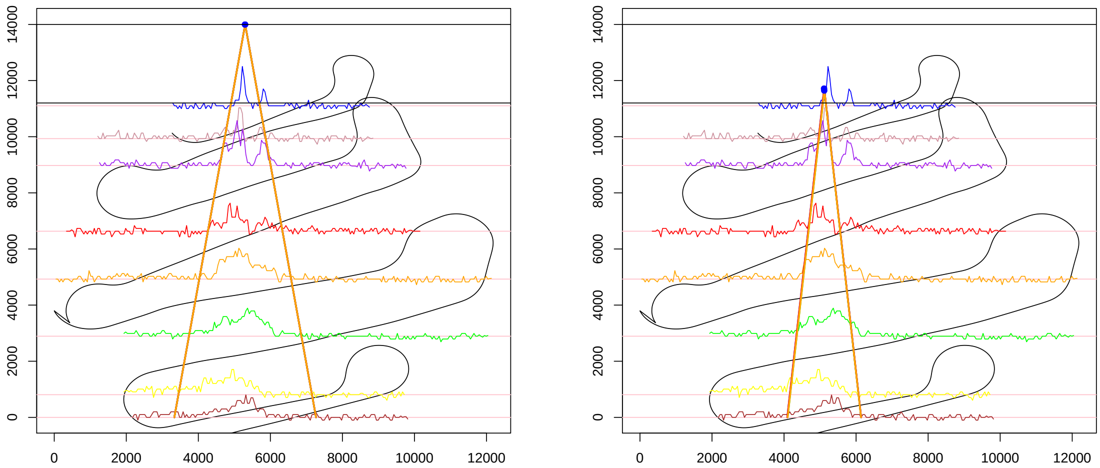


(a) The use of the multivariate test statistic.



(b) The use of the projection test statistic

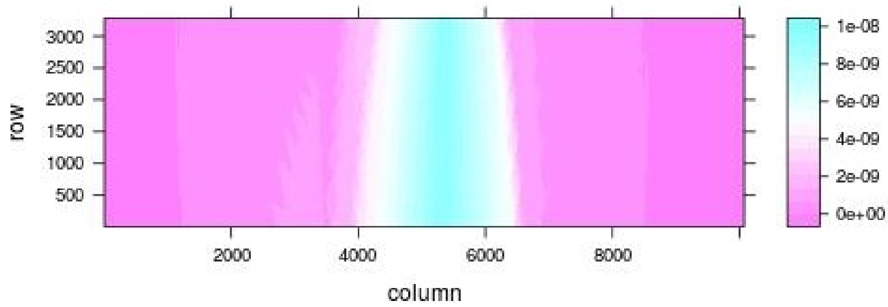
Figure 18.17.: Right-hand trajectory with maximizing the opening angle and rotation about 28° of the first four lines and the second four lines about 23°



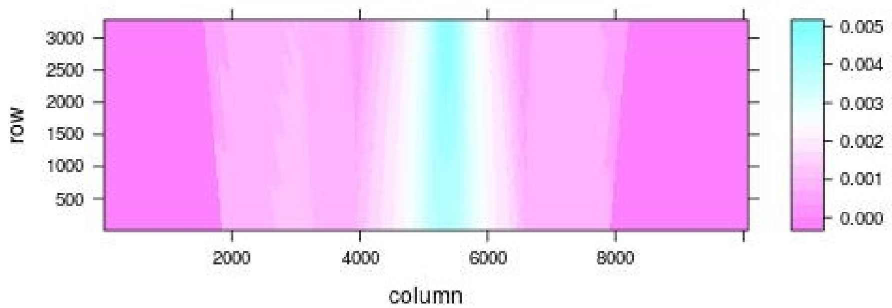
(a) The use of the multivariate test statistic

(b) The use of the projection test statistic

Figure 18.18.: Heatmaps of the left-hand trajectory with maximizing the opening angle (in every point the value is illustrated, which is the maximum of all possible estimators in this point by varying the opening angle)

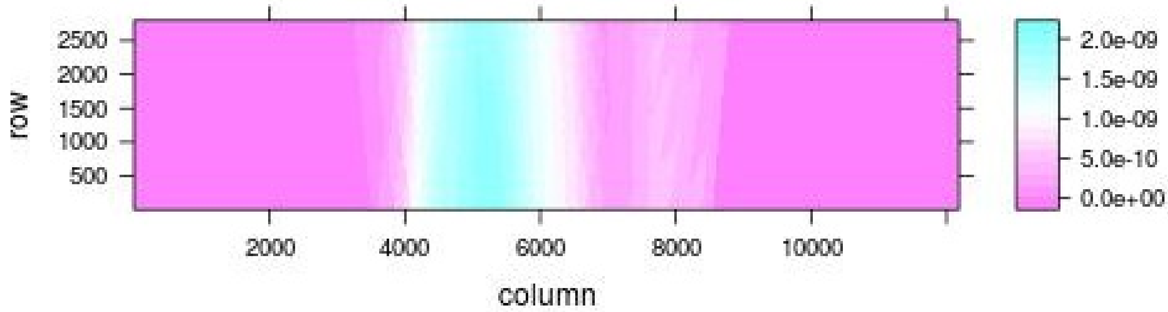


(a) With the multivariate statistic

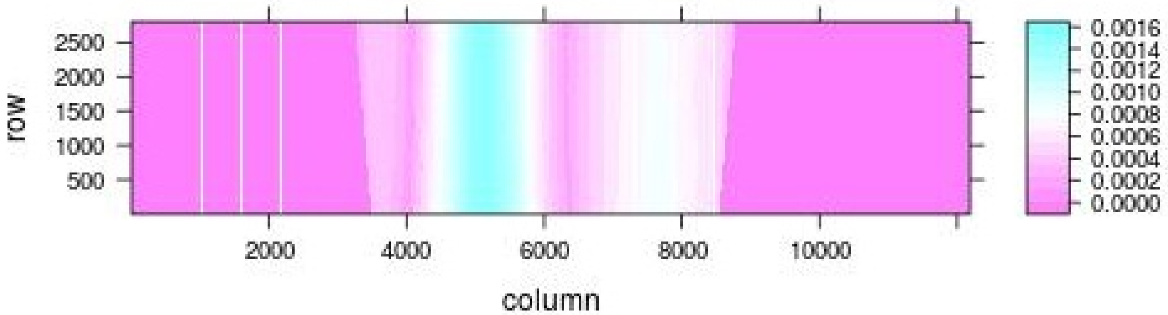


(b) With the projection statistic

Figure 18.19.: Heatmaps of the right-hand trajectory with maximizing the opening angle and rotation about 28° of the first four lines and the second four lines about 23° (in every point the value is illustrated, which is the maximum of all possible estimators in this point by varying the opening angle)



(a) With the multivariate statistic



(b) With the projection statistic

Part IV.

Appendix

A. Assumptions and Propositions under the Null Hypothesis of the Sequential Testing Problem

A.1. Modified MOSUM

Assumption 3.1

a) The weight function has the form

$$w(m,k) = m^{-\frac{1}{2}}\widetilde{w}(m,k)$$

with

$$\widetilde{w}(m,k) = \begin{cases} \rho\left(\frac{k}{m}\right), & k \geq a_m \\ 0, & k < a_m \end{cases}$$

and $\frac{a_m}{m} \to 0$ as $m \to \infty$. In addition, we assume that ρ is continuous and that

$$\lim_{t\to 0} t^{\gamma}\rho(t) < \infty \quad \text{for some} \quad 0 \leq \gamma < \frac{1}{2}.$$

b) For the open-end procedure we additionally assume

$$\lim_{t\to\infty} t\rho(t) < \infty.$$

Assumption 3.2
The following approximation holds under H_0, where $N(m)$ is the observation horizon and can be infinite:

$$\sup_{1\leq k<N(m)} \frac{1}{\sqrt{m}} \min\left(\frac{1}{m^{-\gamma}k^{\gamma}}, \frac{m}{k}\right) \Bigg\| \sum_{i=m+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0)\right) \Bigg\| = o_P(1),$$

for some θ_0 and γ as in Assumption 3.1a).

Assumption 3.3

a) The partial sum process

$$\left\{ \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} (H(\boldsymbol{X}_t, \theta_0), \boldsymbol{B}(\theta_0) G(\boldsymbol{X}_t, \theta_0)) : 1 \leq s \leq T \right\}$$

fulfills a functional limit theorem for any $T > 0$:

$$\left\{ \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} (H(\boldsymbol{X}_t, \theta_0), \boldsymbol{B}(\theta_0) G(\boldsymbol{X}_t, \theta_0)) : 1 \leq s \leq T \right\}$$
$$\xrightarrow{D} \{(\boldsymbol{W}_1(s), \boldsymbol{W}_2(s)) : 1 \leq s \leq T\},$$

where $(\boldsymbol{W}_1(s), \boldsymbol{W}_2(s))$ is a multivariate Wiener process with covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{C} \\ \boldsymbol{C}^T & \boldsymbol{\Sigma}_2 \end{pmatrix}$.

b) The following Hájék-Rényi-type inequality holds for all $0 < \alpha < \frac{1}{2}$:

$$\max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1) \quad (m \to \infty).$$

c) For the open-end procedure the following Hájék-Rényi-type inequality holds for any sequence $k_m > 0$

$$\max_{k \geq k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1) \quad (m \to \infty).$$

Proposition 3.1
Under the null hypothesis let Assumptions 3.1 and 3.2 hold. Then we obtain:

$$\sup_{1 \leq k < N(m)} w(m,k) \left\| \sum_{i=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_i, \hat{\theta}_m) - \left(\sum_{j=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_j, \theta_0) - \frac{k - \lfloor kh \rfloor}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right) \right\| = o_P(1).$$

Proposition 3.2
If Assumption 3.3b) holds, then the following Hájék-Rényi-type inequality is valid for all $0 < \alpha < \frac{1}{2}$:

$$\max_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left\| \sum_{t=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1).$$

Proposition 3.3
If $\{\boldsymbol{W}_1(t) : t \geq 0\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$ then

$$\sup_{0<t\leq\tau} \frac{1}{t^\alpha} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)\| = o_P(1), \quad \tau \to 0,$$

where $0 < \alpha < \frac{1}{2}$.

Proposition 3.4
Let Assumption 3.3c) hold, then a Hájék-Rényi-type inequality for the open-end procedure is fulfilled for any sequence $k_m > 0$:

$$\max_{k\geq k_m} \frac{\sqrt{k_m}}{k} \left\| \sum_{t=m+\lfloor kh \rfloor+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) \right\| = O_P(1).$$

Proposition 3.5
If $\{\boldsymbol{W}_1(t) : t \geq 0\}$ is a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$ then

$$\max_{t\geq T} \frac{1}{t} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+th)\| = o_P(1), \quad T \to \infty.$$

A.2. Page-CUSUM

Proposition 3.6
Under the null hypothesis let Assumptions 3.1 and 3.2 hold. Then

$$\sup_{1\leq k<N(m)} w(m,k) \left| \max_{0\leq i\leq k} \left\| \sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t, \hat{\theta}_m) \right\| \right.$$
$$\left. - \max_{0\leq i\leq k} \left\| \sum_{t=m+i+1}^{m+k} H(\boldsymbol{X}_t, \theta_0) - \frac{k-i}{m} \boldsymbol{B}(\theta_0) \sum_{j=1}^{m} G(\boldsymbol{X}_j, \theta_0) \right\| \right| = o_P(1).$$

Proposition 3.7
If Assumptions 3.3b) and c) hold, we have the Hájék-Rényi-type inequality

$$\sup_{k\geq Tm} \frac{\sqrt{Tm}}{k} \max_{1\leq i\leq k} \left\| \sum_{j=m+1}^{m+i} H(\boldsymbol{X}_j, \theta_0) \right\| = O_P(1).$$

Proposition 3.8
For a Wiener process $\{\boldsymbol{W}_1(\cdot)\}$ with covariance matrix $\boldsymbol{\Sigma}_1$ and $0 < \alpha < \frac{1}{2}$, we have

$$\sup_{0<t\leq\tau} \frac{1}{t^\alpha} \max_{0\leq s\leq t} \|\boldsymbol{W}_1(1+t) - \boldsymbol{W}_1(1+s)\| = o_P(1), \quad \tau \to 0.$$

Proposition 3.9
For a Wiener process $\{\boldsymbol{W}_1(\cdot)\}$ with covariance matrix $\boldsymbol{\Sigma}_1$, we have

$$\sup_{t\geq T}\frac{1}{t}\max_{0\leq s\leq t}\|\boldsymbol{W}_1(1+t)-\boldsymbol{W}_1(1+s)\|=o_P(1),\quad T\to\infty.$$

A.3. MOSUM

Assumption 3.4
Using the MOSUM statistic to test for a structural change in a time series,

a) the weight function has the form

$$w_M(h,k)=h^{-\frac{1}{2}}\rho_M\left(\frac{k}{h}\right),$$

where ρ_M is bounded and continuous and $h=h(m)\to\infty$, as $m\to\infty$.

b) In the open-end procedure if $\frac{h}{m}\to\beta$, as $m\to\infty$, for some $\beta\in(0,1]$, suppose that

$$\lim_{t\to\infty}t\rho_M(t)<\infty.$$

c) If $\frac{h}{m}\to 0$, as $m\to\infty$, we need the following weaker condition:

$$\limsup_{t\to\infty}t^{\frac{1}{\nu}}\rho_M(t)<\infty,\quad \text{for some}\quad \nu>2$$

Assumptions 3.4b) and c) are equivalent to Assumption 3.1b) on the boundary function of the modified MOSUM and the Page-CUSUM statistics.

Proposition 3.10
Let the null hypothesis hold as well as $h(m)\overset{m\to\infty}{\longrightarrow}\infty$ and $\frac{h}{m}\overset{m\to\infty}{\longrightarrow}\beta,\ \beta\in(0,1]$. The boundary function ρ satisfies Assumption 3.2 with $\gamma=0$ and Assumption 3.4a)-b). Then we have

$$\sup_{1\leq k<N(m)}w_M(h,k)\left\|\sum_{i=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\hat{\theta}_m)\right.$$
$$\left.-\left(\sum_{j=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)-\frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right)\right\|=o_P(1).$$

Proposition 3.11
Under Assumption 3.3c) hold, we have a Hájék-Rényi-type inequality in the open-end procedure, namely

$$\max_{k\geq k_m+h}\frac{\sqrt{k_m}}{k}\left\|\sum_{t=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)\right\|=O_P(1)\quad\text{for any sequence } k_m\geq 0.$$

Proposition 3.12
Let $\{\boldsymbol{W}(\cdot)\}$ be a Wiener process with covariance matrix $\boldsymbol{\Sigma}_1$. Then

$$\sup_{t\geq T}\frac{1}{t}\left\|\boldsymbol{W}_1\left(\frac{1}{\beta}+t\right)-\boldsymbol{W}_1\left(\frac{1}{\beta}+t-1\right)\right\|=o_P(1),\quad T\to\infty.$$

Assumption 3.5
The following approximation holds under H_0, where the observation horizon $N(m)$ can be infinite:

$$\sup_{1\leq k<N(m)}\frac{1}{\sqrt{h}}\min\left(1,\left(\frac{h}{k}\right)^{\frac{1}{\nu}}\right)\left\|\sum_{i=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\hat{\theta}_m)\right.$$
$$\left.-\left(\sum_{i=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)-\frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right)\right\|=o_P(1),$$

for some θ_0 and ν as in Assumption 3.4c).

Assumption 3.6
There is a Wiener process $\{\boldsymbol{W}(t),0\leq t<\infty\}$ with covariance matrix $\boldsymbol{\Sigma}_1$ such that, as $k\to\infty$

$$\sum_{t=1}^{k}H(\boldsymbol{X}_t,\theta_0)-\boldsymbol{W}(k)=O\left(k^{\frac{1}{\nu}}\right)\quad a.s.,\quad\text{with }\nu\text{ as in Assumption 3.4c).}$$

Proposition 3.13
Under the Assumptions 3.4a) and c) and Assumption 3.5, we have

$$\sup_{1\leq k<N(m)}w_M(h,k)\left\|\sum_{i=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\hat{\theta}_m)\right.$$
$$\left.-\left(\sum_{j=m+k-h+1}^{m+k}H(\boldsymbol{X}_t,\theta_0)-\frac{h}{m}\boldsymbol{B}(\theta_0)\sum_{j=1}^{m}G(\boldsymbol{X}_j,\theta_0)\right)\right\|=o_P(1).$$

B. Theorems of Probability Theory

Theorem B.1

(Billingsley (1968) Theorem 8.4)
Let $\xi_1, \xi_2, ...$ be random variables and $S_T = \xi_1 + ... + \xi_T$ and

$$X_T(s) = \frac{1}{\sigma\sqrt{T}} S_{\lfloor Ts \rfloor},$$

then $\{X_T\}$ is tight if $\forall \epsilon > 0 \; \exists \lambda > 1$ and $T_0 \in \mathbb{N}$ such that

$$P\left(\max_{i \leq T} |S_{k+i} - S_k| \geq \lambda\sigma\sqrt{T}\right) \leq \frac{\epsilon}{\lambda^2}, \quad T \geq T_0$$

holds $\forall k$.

Lemma B.1

(Billingsley (1968) Lemma p.69)
Let $\xi_1, \xi_2, ..., \xi_T$ be independent random variables with mean 0 and finite variances σ_i^2 and $S_i = \xi_1 + \xi_2 + ... + \xi_i$, $s_i^2 = \sigma_1^2 + \sigma_2^2 + ... + \sigma_i^2$. Then

$$P\left(\max_{i \leq T} |S_i| \geq \lambda s_T\right) \leq 2P\left(|S_T| \geq (\lambda - \sqrt{2})s_T\right).$$

Theorem B.2

(Kirch (2006) Theorem B.7)
Let $\{Y(i) : i \geq 1\}$ be a sequence of independent random variables with $E(Y(i)) = 0, i \geq 1$, satisfying $E|Y(i)|^{2+\delta} \leq C, i \geq 1$, for some $\delta > 0$. Then there is a constant D such that

$$E\left|\sum_{i=1}^{n} Y(i)\right|^{2+\delta} \leq Dn^{\frac{2+\delta}{2}}.$$

Lemma B.2

(Billingsley (1968), Problem 6, p.41)
Probability measures on a product space are tight if all the marginal probability measures are tight on the component spaces.

Proposition B.1

(Billingsley (1968), p.35)
If the sequence $\{P_T\}$ converges weakly to P, then $\{P_T\}$ is relatively compact.

Theorem B.3

(Billingsley (1968), Theorem 6.2)
Let Π be a family of probability measures on $(S, \mathscr{S})$, where S is a metric space and $\mathscr{S}$ is a σ-field of Borel-sets. Suppose S is separable and complete and Π is relatively compact, then it is tight.

Theorem B.4

Let $K \subseteq \mathbb{R}^p, p \in \mathbb{N}$, be a compact set, $f : K \to \mathbb{R}$ a continuous function and x_0 a unique maximizer of f, i.e. $x_0 = \underset{x \in S}{\arg\max} f(x)$. Furthermore, let $\widehat{x}_n = \underset{x \in S}{\arg\max} f_n(x)$, where f_n is a sequence of stochastic functions with $\max_{x \in S} |f_n(x) - f(x)| \xrightarrow{P} 0$. Then, $\widehat{x}_n \xrightarrow{P} x_0$.

Proof. Suppose that $\widehat{x}_n$ does not converge stochastically to x_0. Since S is compact, there is a subsequence $\widehat{x}_{k_n}$ with $\widehat{x}_{k_n} \xrightarrow{P} x_1$ for some $x_1 \neq x_0$. Then

$$\begin{aligned}
|f_{k_n}(\widehat{x}_{k_n}) - f(x_1)| &= |f_{k_n}(\widehat{x}_{k_n}) - f(\widehat{x}_{k_n}) + f(\widehat{x}_{k_n}) - f(x_1)| \\
&\leq |f_{k_n}(\widehat{x}_{k_n}) - f(\widehat{x}_{k_n})| + |f(\widehat{x}_{k_n}) - f(x_1)| \\
&\leq \max_{x \in S} |f_{k_n}(x) - f(x)| + |f(\widehat{x}_{k_n}) - f(x_1)| = o_P(1).
\end{aligned}$$

Thus both terms are $o_P(1)$, the first one by assumption and the second one by the continuity of the function f.

However we also obtain

$$|f_n(\widehat{x}_n) - f(x_0)| = \left| \max_{x \in S} f_n(x) - \max_{x \in S} f(x) \right| \leq \max_{x \in S} |f_n(x) - f(x)| = o_P(1),$$

by assumption. Since x_0 is a unique maximizer of f we have $f(x_1) < f(x_0)$, implying

$$\begin{aligned}
0 < |f(x_1) - f(x_0)| &= |f(x_1) - f_n(\widehat{x}_{k_n}) + f_n(\widehat{x}_{k_n}) - f(x_0)| \\
&\leq |f(x_1) - f_n(\widehat{x}_{k_n})| + |f_n(\widehat{x}_{k_n}) - f(x_0)| \xrightarrow{P} 0,
\end{aligned}$$

which is a contradiction. □

Bibliography

J. Aston and C. Kirch. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220, 2012.

J. Aston and C. Kirch. Efficiency of change point tests in high dimensional settings. *Preprint*, 2016.

A. Aue, I. Berkes, and L. Horváth. Strong approximation for the sums of squares of augmented garch sequences. *Bernoulli*, 12(4):583–608, 2006a.

A. Aue, L. Horváth, M. Hušková, and P. Kokoszka. Change-point monitoring in linear models. *Econometrics Journal*, 9(3):373–403, 2006b.

A. Aue, L. Horváth, M. Kühn, and J. Steinebach. On the reaction time of moving sum detectors. *Journal of Statistical Planning and Inference*, 27(2):126–140, 2008.

S. Banerjee and A. Roy. *Linear algebra and matrix analysis for statistics*, page 493. CRC Press, 2014.

P. Billingsley. *Convergence of Probability Measures.* Wiley and Sons, 1968.

E. Carlstein. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14(3):1171–1179, 1986.

Z. Chen and Z. Tian. Modified procedures for change point monitoring in linear models. *Mathematics and computers in simulation*, 81(1):62–75, 2010.

C.-S.J. Chu, M. Stinchcombe, and H. White. Monitoring structural change. *Econometrica*, 64: 1045–1065, 1996.

G. Ciupera. Two tests for sequential detection of a change-point in a nonlinear model. *Journal of Statistical Planning and Inference*, 143(10):1719–1743, 2013.

G. Ciuperca. Two tests for sequential detection of a change-point in nonlinear model. *Journal of Statistical Planning and Inference*, 143(10):1719–1743, 2013.

M. Csörgő and P. Révész. *Strong Approximations in Probability and Statistics.* Academic Press, 1981.

P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. *Annales de l'Institut Henri Poicaré - Probabilités et Statistiques*, 30(1):63–82, 1994.

B. Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1):1–26, 1979.

S. Fremdt. Page's sequential procedure for change-point detection in time series regression. *Statistics: A Journal of Theoretical and Applied Statistics*, 48(4):1–28, 2014.

P. Hall. Resampling a coverage pattern. *Stochastic Processes and their Applications*, 20(2): 231–246, 1985.

K. Haynes, P. Fearnhead, and I. Eckley. A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing*, pages 1–13, 2016.

N. Herrndorf. A functional central limit theorem for ρ-mixing sequences. *Journal of Multivariate Analysis*, 15(1):141–146, 1984a.

N. Herrndorf. A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 12(1):141–153, 1984b.

B. Hirst, P. Jonathan, F.González del Cueto, D. Randell, and O.Kosut. Locating and quantifying gas emission sources using remotely obtained concentration data. *Atmospheric Environment*, 74:141–158, 2013.

L. Horváth and M. Hušková. Change-point detection in panel data. *Journal of Time Series Analysis*, 33:631–648, 2012.

L. Horváth, P. Kokoszka, and J. Steinebach. Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis*, 68:96–119, 1999.

L. Horváth, M. Hušková, P. Kokoszka, and J. Steinebach. Monitoring changes in linear models. *Journal of statistical and inference*, 126(1):225–251, 2004.

L. Horváth, M. Kühn, and J. Steinebach. On the performance of the fluctuation test for structural change. *Sequential Analysis*, 142:2271–2288, 2012.

M. Hušková and A. Koubková. Monitoring jump changes in linear models. *Journal of the American Statistical Association*, 39(2):1–70, 2005.

M. Hušková and J. Steinebach. Limit theorems for a class of tests of gradual changes. *Journal of Statistical Planning and Inference*, 89:57–77, 2000.

M. Hušková and J. Steinebach. Asymptotic tests for gradual changes. *Statistics & Decisions*, 20:137–151, 2002.

H. Kauppi and P. Saikkonen. Prediction us recessions with dynamic binary response. *The Review of Economics and Statistics*, 90(4):777–791, 2008.

C. Kirch. *Resampling Methods for the Change Analysis of Dependent Data.* PhD thesis, University Cologne, 2006.

C. Kirch and J. Tadjuidje Kamgaing. An online approach to detecting changes in nonlinear autoregressive models. *Preprint*, 2011.

C. Kirch and J. Tadjuidje Kamgaing. On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference*, 161:25–49, 2015.

C. Kirch, B. Muhsal, and H. Ombao. Detection of changes in multivariate time series with application to eeg data. *Journal of the American Statistical Association*, 110:1197–1216, 2015.

K. Knopp. *Theorie und Anwendung der unendlichen Reihen*, page 322. Springer, 1996.

J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rvs and the sample df. 1. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32: 111–131, 1975.

J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rvs and the sample df. 2. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 34: 33–58, 1976.

J.-P. Kreiss. *Asymptotic statistical inference for a class of stochastic processes.* PhD thesis, University Hamburg, 1988.

H.R. Künsch. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3):1217–1241, 1989.

R.Y. Liu and K. Singh. *Moving blocks jackknife and bootstrap capture weak dependence*, pages 225–248. Wiley, 1992.

B. Muhsal. *Change-Point Methods for multivariate autoregressive models and multiple structural breaks in the mean.* PhD thesis, Karlsruhe Institute of Technology (KIT), 2013.

M. Neumann. A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, 17:120–134, 2013.

N.H. Neumann. Absolute regularity and ergodicity of poisson count processes. *Bernoulli*, 17(4): 1268–1284, 2011.

E. Paparoditis and D. Politis. Tapered block bootstrap. *Biometrika*, 88(4):1105–1119, 2001.

D. Politis. Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, 27:703–744, 2011.

D. Politis and H. White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

D.N. Politis, J.P. Romano, and M. Wolf. *A circular block-resampling procedure for stationary data*, pages 263–270. Wiley, 1992.

A. Schmitz and J. Steinebach. A note on the monitoring of changes in linear models with dependent errors. *Dependence in Probability and Statistics*, Volume 200 of the series Lecture Notes in Statistics:159–174, 2010.

X. Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235, 2010.

R. Startz. Binomial autoregressive moving average models with an application to us recession. *Journal of Business and Economic Statistics*, 26(1):1–8, 2008.

D. Wilks and R. Wilby. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, 23(3):329–357, 1999.