Sequential Change Point Procedures Based on U-Statistics and the Detection of Covariance Changes in Functional Data

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von Christina Stöhr
geb. am 12.05.1989 in Kaiserslautern

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Claudia Kirch
Prof. Dr. Alexander Aue

Eingereicht am: 13.12.2018
Verteidigung am: 15.03.2019

Abstract

Change point analysis is concerned with the detection of changes in the underlying model of a sequence of observations. There are two different approaches in the context of change point analysis. In the classical a-posteriori approach, the completely observed data set is available when starting the testing procedure. In sequential change point analysis, data is monitored by testing for a structural break after each new observation.

In the first and more theoretical part of this thesis we propose a general framework of sequential testing procedures based on U-statistics which, as an example, yields a robust sequential change point procedure related to a Wilcoxon-type test statistic. The critical values can be obtained from the derived limit distribution of the test statistic under the null hypothesis and we show that the proposed tests have asymptotic power one. Furthermore, we consider monitoring schemes that are adapted to late changes. We derive the respective asymptotics under the null hypothesis as well as under the alternative. Sequential change point procedures naturally involve a certain detection delay as some data needs to be collected after the change to obtain statistical significance. The speed of detection is of particular importance in sequential change point analysis as, for example, monitoring patient or machine data requires an intervention as soon as possible after a structural break occurred. Therefore, we derive the asymptotic distribution of the stopping time. In a simulation study we assess the finite sample performance of the testing procedures as well as the stopping time.

In the second part of this work we develop a-posteriori change point procedures for the evaluation of covariance stationarity in functional data where the focus is on the application to functional magnetic resonance imaging (fMRI) data. Such scans provide a large amount of information for analyzing activities in the brain and in particular the interactions between brain regions. Resting state fMRI data is widely used for inferring connectivities in the brain which are not due to external factors. As such analyses strongly rely on stationarity, change point procedures can be applied in order to detect possible deviations from this crucial assumption. We model fMRI data as functional time series and develop tools for the detection of deviations from covariance stationarity via change point alternatives. We propose a nonparametric procedure which is based on dimension reduction techniques. However, as the projection of the functional time series onto a finite and rather low-dimensional subspace involves the risk of missing changes which are orthogonal to the projection space, we also consider two test statistics which take the full functional structure into account. The proposed methods are compared in a simulation study and applied to more than 100 resting state fMRI data sets.

Zusammenfassung

Die Changepoint Analyse befasst sich mit der Erkennung von Änderungen in dem Modell, welches einer Folge von Beobachtungen zugrunde liegt. In diesem Gebiet gibt es zwei unterschiedliche Ansätze. Im klassischen a-posteriori Ansatz wird das Testverfahren auf den komplett beobachteten Datensatz angewandt. In der sequentiellen Changepoint Analyse hingegen werden Daten dadurch überwacht, dass nach jeder neuen Beobachtung ein Test auf einen Strukturbruch durchgeführt wird.

Im ersten und eher theoretischen Teil dieser Arbeit führen wir eine allgemeine Klasse von sequentiellen Testverfahren ein, welche auf U-Statistiken basieren. Damit ist es unter andem möglich, basierend auf einer Wilcoxon-Teststatistik ein robustes sequentielles Verfahren zur Erkennung von Strukturbrüchen zu erhalten. Die kritischen Werte können mit Hilfe der hergeleiteten Grenzverteilung der Teststatistik unter der Nullhypothese bestimmt werden und wir zeigen, dass die Testverfahren asymptotische Güte eins besitzen. Außerdem betrachten wir alternative Teststatistiken, welche insbesondere auf die bessere Erkennung von späten Änderungen abzielen. Auch für diese Verfahren leiten wir die Grenzverteilung unter der Nullhypothese her und zeigen, dass sie asymptotische Güte eins haben. Sequentielle Testverfahren beinhalten immer eine gewisse Verzögerung in der Erkennung eines Strukturbruches, da zunächst einige Beobachtungen nach der Änderung gesammelt werden müssen, um statistische Signifikanz zu erhalten. Die Schnelligkeit der Strukturbrucherkennung ist von besonderer Relevanz in der sequentiellen Changepoint Analyse, da beispielsweise die Überwachung von Partienten- oder Maschinendaten einen unmittelbaren Eingriff erfordert, nachdem eine Änderung eingetreten ist. Dazu leiten wir die asymptotische Verteilung der Stoppzeit her. In einer Simulationsstudie betrachten wir das Verhalten der Testverfahren sowie der Stoppzeiten für endlichen Stichprobenumfang.

Im zweiten Teil dieser Arbeit entwickeln wir a-posteriori Verfahren zur Evaluierung der Kovarianzstationarität in funktionalen Daten. Dabei liegt der Fokus auf der Anwendung auf Daten der funktionalen Magnetresonanztomographie (fMRT). Solche Aufnahmen liefern eine riesige Menge an Informationen zur Analyse von Gehirnaktivitäten und insbesondere der Interaktionen zwischen verschiedenen Gehirnregionen. Im Ruhezustand aufgenommene fMRT Daten werden häufig genutzt, um Konnektivitäten im Gehirn abzuleiten, welche nicht durch externe Faktoren verursacht werden. Da solche Analysen stark auf die Stationaritätsannahme angewiesen sind, können Verfahren zur Strukturbrucherkennung eingesetzt werden, um mögliche Abweichungen von dieser entscheidenden Annahme zu detektieren. Wir modellieren fMRT Daten als funktionale Zeitreihen und enwickeln Methoden zur Erkennung einer Abweichung von der Kovarianzstationarität mittels Changepoint Alternativen. Wir schlagen zunächst ein Verfahren vor, welches auf Dimensionsreduktion basiert. Allerdings geht die Projektion der funktionalen Zeitreihe auf einen niedrigdimensionalen Unterraum mit dem Risiko einher, dass Änderungen, welche orthogonal zum Projektionsraum sind, nicht erkannt werden. Daher betrachten wir zwei alternative Teststatistiken, welche die volle funktionale Struktur der Daten berücksichtigen. Die vorgeschlagenen Methoden werden in einer Simulationsstudie verglichen und auf mehr als 100 Ruhezustand fMRT Datensätze angewandt.

Acknowledgements

First of all, I would like to express my most sincere gratitude to my supervisor Prof. Dr. Claudia Kirch for her constant support which already started with my Diploma thesis and has continued ever since. I deeply appreciate the friendly and motivating working atmosphere, the helpful discussions, the valuable input whenever needed as well as the thorough corrections and comments on every new part of this work. Her deep insights and intuition have helped and taught me a lot.
I would like to thank Prof. Dr. John Aston for the most pleasant and instructive research stays at the University of Cambrige from which the second part of this work results. I would also like to thank Prof. Dr. Alexander Aue for reviewing this thesis. I thank the Institute of Stochastics at the KIT, where I spent the first 6 months of my PhD, for the pleasant working atmosphere. Furthermore, I thank all members of the Institute for Mathematical Stochastics at the Otto von Guericke University Magdeburg for the family atmosphere which made the change of university so easy and every day in the office so enjoyable. In particular, I would like to thank Kerstin Altenkirch for taking care of everyone and everything at the institute, Alexander Meier for the nice atmosphere in our working group and the helpful comments on my manuscript, and last but not least Kerstin Reckrühm who went through all the stages of finishing a thesis with me at the same time - thank you for daily scientific and non-scientific conversations but most of all for being such a good friend. My heartfelt thanks go to my friends Ariane, Jennifer, Josefine, Maria and Sarah not only for many motivating learning sessions and everlasting memories of our studies but also for staying so close and always being there when needed although life has spread us apart geographically. I'm deeply grateful to my family, especially my parents and my sister who believe in me and support me throughout my whole life. My special thanks belong to Gregor for his unconditional love and support. Thank you for always being there for me.

My research stay at the University of Cambridge was partly founded by a DAAD scholarship. The support of Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 314838170, GRK 2297 MathCoRe is greatly acknowledged.

Contents

List of Figures

List of Tables

Introduction

Change Point Analysis

Change point analysis provides powerful statistical tools for evaluating structural stability of data which is a problem that first arose in the context of quality control and is now of great interest in numerous and diverse fields such as economics, medicine, technology and meteorology. More precisely, it deals with the detection of structural breaks in time series which are sequences of observations that are ordered in time. A structural break, which is called a change point, is a point in time at which the model of the observed time series changes. In addition to the obvious practical relevance for the detection of critical changes in data, change point procedures are also of great interest for the validation of applicability of other statistical methods which are often based on the assumption that the observed time series is stationary which means its behavior is stable over time. Change point analysis is not only concerned with deciding whether there is a change or not (hypothesis testing) but also with estimating the number and the location of the change points if the null hypothesis of structural stability has been rejected. In this work, we focus on the testing problem.

With roots dating back to the 1950's (Page (1954), Page (1955a), Page (1955b)), change point analysis is a well-established research area. In recent years, it has experienced a boom due to increasing amounts of automatically collected data which simultaneously become more and more complex and thus require new statistical methods. At most one change (AMOC) in the mean of independent observations is the simplest and most extensively studied change point problem which has been and still is a good basis for the development of change point procedures for various models and dependency structures in order to tackle the challenges of modern data streams. Change point procedures can be broadly divided into two main classes. On the one hand, a- posteriori (offline) procedures use the completely observed data set to test for a change point. On the other hand, sequential (online) procedures are applied to decide whether a change has occurred while collecting the data. Sequential procedures are of great interest wherever data needs to be monitored and changes need to be detected very fast after they have occurred. For example, monitoring medical data of patients requires quick interventions in case of critical changes. In finance, a steady control of accounting processes enables a fast detection of fraudulent activities in order to avert further damage.

In the a-posteriori approach, the decision if there is a change or not is based on one single hypothesis test where the test statistic is calculated using the whole observed data set, whereas a sequential procedure tests for a structural break after each new observation. The a-posteriori procedure detects a change if the test statistic exceeds a critical value. In the sequential approach, the monitoring continues as long as no change is detected such that the monitoring horizon is random and even possibly infinite. Therefore, sequential procedures use weighted critical values called critical curves. The sequential testing procedure stops and detects a change as soon as a suitable monitoring statistic exceeds the critical curve. For both approaches, critical values are typically determined based on the limit distribution of the test statistic under the null hypothesis. Other properties, such as the power of the procedure, are also obtained asymptotically. For a-posteriori procedures, such results are established by letting the length of the observed data set increase to infinity. This is not possible for sequential procedures as the number of observations is not prespecified. Following the approach of Chu *et al.* (1996) we assume that there exists a historic data set without a change. Asymptotic considerations can be conducted with respect to the length of this data set increasing to infinity. Due to the open-ended character, sequential procedures require different asymptotic methods than a-posteriori procedures.

Contributions

This thesis provides both theoretical and practical contributions to change point analysis, in particular with a view to challenges that occur with increasingly large and complex data. The growing desire to continuously monitor data and quickly detect changes requires new developments in sequential testing which are achieved in the first part of this thesis in a mainly theoretical way. In the second part, we develop a-posteriori methods where we focus on the application to functional magnetic resonance imaging (fMRI) data which is noisy, complex and of huge dimension and thus imposes methodological as well as computational challenges that many other modern applications share. Throughout this work, we consider nonparametric approaches which means we do not impose any assumptions on the distribution of the data. Furthermore, we allow for time dependencies which is in most applications more realistic than assuming the observations to be independent.

In the first part, we propose a general framework of sequential change point procedures based on U-statistics. The sequential CUSUM (cumulative sum) procedure for the detection of a change in the mean is very well studied but sensitive to outliers as it is based on the arithmetic mean. As many other statistics, the CUSUM statistic can be represented in the form of a U-statistic. Hence, the proposed framework covers the sequential CUSUM procedure but also allows, for example, to construct a more robust Wilcoxon-type testing procedure. However, as the respective monitoring statistic takes all new observations into account, it shares one drawback of the classical CUSUM monitoring scheme, namely a long detection delay for late changes. Therefore, the modified MOSUM (moving sum) and the Page-CUSUM have been introduced in the sequential literature. Those monitoring schemes are more stable with respect to the time of the change and we extend them to the framework of sequential tests based on U-statistics. The motivation of the modified MOSUM and the Page-CUSUM indicates that the stopping time, i.e. the time when the procedure detects a change, is of particular interest for sequential procedures. The asymptotic distribution of the stopping time for the classical sequential CUSUM as well as for the Page-CUSUM has been derived by Aue *et al.* (2009a) and Fremdt (2014) for relatively early changes. We generalize those results for the CUSUM monitoring scheme to the general framework considered in this work. To the best of our knowledge, up to now, there are no results on the stopping time of sequential procedures for late changes available and the results for early changes cannot be extended straightforwardly. We extend the existing literature in this regard where we first identify the relevant issues that occur for late changes but not for early ones and then find the necessary steps to handle them such than we can finally assess the asymptotic behavior of the stopping time for late changes. The finite sample performance of the testing procedures as well as the stopping time is investigated in a simulation study.

In the second part, we develop a-posteriori change point procedures for the detection of deviations from covariance stationarity with focus on the application to functional magnetic resonance imaging (fMRI) data. An fMRI data set consists of a sequence of three dimensional images of the brain that have been recorded every few seconds. Such scans provide a huge amount of information for analyzing activities in the brain and in particular the interactions between brain regions. In this work, we focus on resting state fMRI data which is widely used to infer connectivities in the brain excluding external factors. As such analyses strongly rely on stationarity, change point procedures can be applied in order to detect possible deviations from this crucial assumption. In this context, Aston & Kirch (2012b) already considered the evaluation of mean stationarity. However, not only deviations from mean stationarity but also deviations from covariance stationarity can contaminate connectivity studies and this may (and usually will) lead to wrong conclusions. To this end, we develop tools for the detection of deviations from covariance stationarity via change point alternatives. We will model fMRI data as functional time series which means we assume that each observation, in this case each image per point in time, can be represented as a function. The statistical analysis of functional data is currently a rapidly progressing field of research as an increasing number of applications provides data which can be modeled in such a way. The nonparametric methodology developed in this work is widely applicable beyond the considered application of fMRI data, hence also of independent interest in functional data analysis in general. First, we introduce a multivariate procedure which is based on dimension reduction techniques. In order to obtain a pivotal limit distribution of the test statistic, which is needed to determine the asymptotic critical values, the long-run covariance has to be estimated. This is statistically challenging and usually leads to an unstable testing procedure but can be avoided by using resampling procedures. We apply a circular block bootstrap to obtain the critical values for an adapted test statistic where we only correct for the diagonal elements of the long-run covariance. Furthermore, the projection of the functional time series on a finite and rather low-dimensional subspace involves the risk of missing changes that are orthogonal to the projection space. Inspired by the methods proposed in Bucchia & Wendler (2017) and Aue *et al.* (2018) for the mean change problem we consider a test statistic which takes the full functional structure into account but, in contrast to the multivariate procedure, does not correct for different variances in the components. Without such a correction, small changes in components with small variances are hard to detect. Therefore, we propose a weighted version of the functional test statistic which combines the advantages of the multivariate and the unweighted functional statistic and is thus a very promising approach for the detection of change points in functional data analysis, not only for detecting changes in the covariance as considered in this work but, in an analogous version, also for the mean change problem. The proposed testing procedures are compared in a simulation study and applied to more than 100 resting state fMRI data sets. In practice, the calculation of the functional test statistics for fMRI data is computationally challenging. We develop an algorithm which makes use of certain characteristics of the data in order to provide a way of calculating the test statistics in a reasonable time.

The Appendix provides an overview of the most important assumptions of the first part. Furthermore, it includes some theoretical results that have been useful in this work.

Part I.

Sequential Tests Based on U-Statistics

1. Introduction to Sequential Testing

Sequential testing procedures have been introduced by Wald in 1943 (cf. for example Wald (1947)). Based on the assumption that collecting data is costly, Wald's aim was to obtain a procedure which, on average, requires fewer observations than a procedure which uses a predetermined number of observations and provides the same reliability in terms of controlling the possible errors. After each new observation it is to be decided whether the null hypothesis is kept, rejected, or the next observation should be considered. The procedure stops as soon as one of the first two decisions is made. However, as nowadays data is collected steadily and at virtually no cost, there is no need to stop monitoring as long as the system is stable or, to put it another way, a continuous monitoring is even desired. Based on this argument, Chu *et al.* (1996) modified Wald's approach and allow for an infinite monitoring horizon. The main assumption is the existence of a stationary historic data set such that asymptotic results can be established by letting the length of the historic data set tend to infinity. This approach allows to construct sequential procedures with asymptotic power one where additionally the asymptotic size is controlled. The work of Chu gave raise to many further developments in sequential testing such as Horváth *et al.* (2004) where sequential CUSUM procedures based on ordinary and recursive residuals for linear models with independent and identically distributed errors are proposed followed by Aue *et al.* (2006) who relaxed the assumptions on the innovations and added a new method based on the squared prediction errors. A further extension can be found in Hušková & Koubková (2005) where procedures based on quadratic forms of partial sums of weighted residuals enable the detection of a larger class of changes. Kirch & Tadjuidje Kamgaing (2015) introduced sequential change point procedures based on estimating functions under very general regularity conditions. This framework unifies existing sequential change point tests and provides monitoring procedures in time series that have not been considered in the literature before such as non-linear autoregressive time series and count time series. A further extension of this setup can be found in Kirch & Weber (2018) where monitoring schemes adapted to late changes are considered. Instead of controlling the type-1 error asymptotically, another branch of sequential procedures is constructed based on optimizing the mean delay for detection and the mean time between false alarms. Lorden *et al.* (1971) proposed appropriate optimality criteria based on the average runlength. More details on corresponding methodologies can be found in Tartakovsky *et al.* (2014).

In this work, we follow the sequential approach of Chu *et al.* (1996) which is based on the so-called 'non-contamination assumption' which means we assume that there exists a historic data set $X_1, \ldots, X_m$ without a change. This assumption is realistic as usually some data has to be collected before any statistical inference can be made. As the distribution of the observed time series is unknown, the historic observations can be used to estimate unknown parameters consistently. Asymptotic results can be obtained by letting the length of the historic data set increase to infinity. Subsequent to the historic observations we start monitoring new incoming data by testing for a structural break after each new observation $X_{m+k}, k \geq 1$, where k denotes the monitoring time. Those tests are based on a monitoring statistic $\Gamma(m, k)$. Since the monitoring continues as long as no change is detected, the monitoring horizon might be infinite such that a weight function $w(m, k)$ is required in order to control the asymptotic size of the procedures. The null hypothesis is rejected as soon as

$$w(m,k)\,|\Gamma(m,k)| > c_\alpha,$$

where the critical value c_α is chosen such that the testing procedure holds the level α asymptotically. If $w(m, k) \neq 0$, an equivalent formulation is given by

$$|\Gamma(m,k)| > \frac{c_\alpha}{w(m,k)}, \tag{1.1}$$

where the weighted critical value $\frac{c_\alpha}{w(m,k)}$ is called critical curve. As long as

$$w(m,k)\,|\Gamma(m,k)| \leq c_\alpha,$$

we continue monitoring. The stopping time τ_m is defined as follows:

$$\tau_m = \begin{cases} \inf\{k \geq 1 : w(m,k)\,|\Gamma(m,k)| > c_\alpha\}, \\ \infty, \quad \text{if } w(m,k)\,|\Gamma(m,k)| \leq c_\alpha \text{ for all k.} \end{cases} \tag{1.2}$$

It denotes the time when the absolute value of the weighted monitoring statistic exceeds the critical value, i.e. the time when we reject the null hypothesis. If this never happens, the stopping time is set to infinity. With this, we clearly see that the monitoring horizon of the sequential testing procedure is random and even possibly infinite. For a given level α, we aim at determining the critical value c_α such that the testing procedure has asymptotic size α with respect to the length of the historic data set increasing to infinity. We test the null hypothesis of stationarity (H_0) against the alternative (H_1) that there occurs a change at some point in the monitoring period. As a finite stopping time means that we reject the null hypothesis at some point, the test has asymptotic size α if

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

and asymptotic power 1 if

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

By the definition of the stopping time it holds that

$$P_{H_0}(\tau_m < \infty) = P_{H_0}\left(w(m,k) \sup_{k\geq 1} |\Gamma(m,k)| > c_\alpha \right).$$

Hence, c_α is actually an asymptotic critical value in the classical sense given by the $1-\alpha$ quantile of the limit distribution of $\sup_{k\geq 1} w(m,k)\,|\Gamma(m,k)|$. Throughout this first part of the thesis, we always call $\Gamma(m, k)$ the monitoring statistic and $\sup_{k\geq 1} w(m,k)\,|\Gamma(m,k)|$ the test statistic.

Sequential CUSUM procedure The classical sequential CUSUM test for mean changes has been studied, among others, by Chu *et al.* (1996), Horváth *et al.* (2004), Aue *et al.* (2006) and Hušková & Koubková (2005) for linear regression models. Consider the problem of detecting a mean change by the following model:

$$X_i = Y_i + 1_{\{i>k^*+m\}} d_m, \quad i \geq 1,$$

where $\{Y_i\}_{i\in\mathbb{Z}}$ is a stationary time series with mean μ and the change d_m is allowed to depend on m. In particular, this includes fixed mean changes ($d_m = d \neq 0$ for all m) as well as local mean changes ($d_m \to 0$ as $m \to \infty$). The respective testing problem is given by $H_0 : d_m = 0$ against $H_1 : d_m \neq 0$. The corresponding sequential CUSUM procedure has been studied in Aue (2003) under weak invariance principles. The CUSUM-type monitoring statistic has the following form

$$\Gamma_C(m,k) = \sum_{j=m+1}^{m+k} \left(\overline{X}_m - X_j\right) = k\left(\overline{X}_m - \frac{1}{k}\sum_{j=m+1}^{m+k} X_j\right). \tag{1.3}$$

It is the cumulative sum of the deviations of the new observations from the mean of the historic observations. The second representation shows that it compares the mean of the new observations with the historic mean in order to detect a level shift. More precisely, after each new observation the arithmetic mean of the historic observations and the arithmetic mean of the new observations up to the current monitoring time point are compared. If the weighted difference of those means is too large, this indicates a change and the null hypothesis is rejected. Due to the fact that this approach is based on the arithmetic mean it is not robust against extreme observations caused by, for example, outliers or heavy tailed distributions. This is also revealed by the simulation study in Chapter 6.

1.1. Outline

In Chapter 2 we introduce a general framework of sequential procedures based on U-statistics which allows us, for example, to derive a sequential procedure with a Wilcoxon-type monitoring statistic which is more robust than the classical CUSUM statistic. We study the asymptotic behavior of the proposed class of sequential procedures in Chapter 3. The limit distribution under the null hypothesis, which is derived in Section 3.2, allows us to determine the critical value such that the testing procedure has asymptotic size α. Furthermore, in Section 3.3, we provide a condition on the type and size of the change such that it is detected asymptotically with probability one. In Chapter 4 we embed the idea of the modified MOSUM and the Page-CUSUM in the framework of sequential change point procedures based on U-statistics and derive the respective asymptotics. Chapter 5 deals with the asymptotic behavior of the stopping time of the procedures introduced in Chapter 2. We generalize existing results on the limit distribution of the stopping time of the sequential CUSUM procedure for early changes in Section 5.1. Extensions to later changes are provided in Section 5.2 and 5.3. The behavior of the proposed procedures and the standardized stopping times for finite historic data sets are assessed in a simulation study in Chapter 6.

2. Sequential Tests Based on U-Statistics

In this chapter, we propose a general framework of sequential change point procedures based on U-statistics which allows us, for example, to construct a robust sequential test based on a Wilcoxon-type monitoring statistic. Robust procedures in change point analysis are of high importance in order to reliably detect structural changes even if the data contains extreme observations. Such extreme observations can be caused, for example, by skewed distributions, heavy tails or outliers. In particular for sequential procedures robustness is very important as their application is mostly associated with the need of quick interventions if there is evidence for a structural change. Monitoring patient or machine data are examples for such applications. If the underlying procedure is not robust, outliers in the data can easily cause false alarms and thus unnecessary efforts and costs. A posteriori tests based on U-statistics have been studied by Csörgő & Horváth (1988) for independent data and by Dehling *et al.* (2015b) for dependent data.

2.1. Monitoring Statistic

In order to detect a structural break in the monitoring period we use two-sample U-statistics with bivariate kernels.

Definition 2.1. *Consider two samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ and let $h : \mathbb{R}^2 \mapsto \mathbb{R}$ be a measurable function . Then,*

$$U_{n_1,n_2} = \frac{1}{n_1 n_2} \sum_{1 \le i \le n_1} \sum_{1 \le j \le n_2} h(X_i, Y_j)$$

is a two-sample U-statistic with kernel h.

For independent observations, a U-statistic constitutes an unbiased estimator for the mean of the kernel function. We refer to Korolyuk & Borovskich (1994) for more details on U-statistics. In order to compare the sample of historic observations $X_1, \ldots, X_m$ with the new observations $X_{m+1}, \ldots, X_{m+k}$ that have been collected up to the current monitoring time point k we propose the following monitoring statistic:

$$\Gamma(m, k) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=m+1}^{m+k} (h(X_i, X_j) - \theta), \tag{2.1}$$

where the kernel $h : \mathbb{R}^2 \to \mathbb{R}$ is a measurable function. Except for the scaling factor,

$$\frac{1}{m} \sum_{i=1}^{m} \sum_{j=m+1}^{m+k} h(X_i, X_j)$$

is a two-sample U-statistic. Under the null hypothesis, the new observations have the same distribution as the historic ones. Therefore, we center the monitoring statistic with $\theta = \mathrm{E}(h(X, Y))$, where X and Y are independent random variables with the same distribution as X_1. We will allow for dependency structures but only in such a way that the observations are approximately independent if they are far enough from each other. Hence, θ still approximates $\mathrm{E}(h(X_i, X_j))$ with increasing lag $j - i$.

The monitoring statistic in (2.1) sums up the differences of the kernel function and its expected value under the null hypothesis over all possible tuples of historic and new observations. Thus, it takes rather small values under the null hypothesis of stationarity and we expect it to increase under the alternative if the change in the new observations yields a mean change in the kernel function.

Example 2.2. *(i) Using the kernel function*

$$h_C(x_1, x_2) = x_1 - x_2$$

we can represent the classical CUSUM monitoring statistic from (1.3) in the form of (2.1):

$$\Gamma_C(m, k) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=m+1}^{m+k} (X_i - X_j) = \sum_{j=m+1}^{m+k} \left(\overline{X}_m - X_j\right) \tag{2.2}$$

with

$$\theta^C = \mathrm{E}\, h_C(X, \tilde{X}) = \mathrm{E}\left(X - \tilde{X}\right) = \mathrm{E}\, X - \mathrm{E}\, \tilde{X} = 0, \tag{2.3}$$

where $\tilde{X}$ is an independent copy of X.

(ii) The kernel function

$$h_W(x_1, x_2) = 1_{\{x_1 < x_2\}}$$

yields a Wilcoxon-type monitoring statistic. Let X have a continuous distribution with density f and distribution function F. It holds

$$\begin{aligned} \theta^W &= \mathrm{E}\, h_W(X, \tilde{X}) = P(X < \tilde{X}) \\ &= \int_{-\infty}^{\infty} f(x) P(X < x) dx = \int_{-\infty}^{\infty} f(x) F(x) dx = \mathrm{E}(F(X)) = \frac{1}{2} \end{aligned} \tag{2.4}$$

as $F(X) \sim U(0, 1)$, where $\tilde{X}$ is an independent copy of X. Hence, we obtain the monitoring statistic

$$\Gamma_W(m, k) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=m+1}^{m+k} \left(1_{\{X_i < X_j\}} - 1/2\right).$$

2.2. Change Point Model

The monitoring statistic in (2.1) is constructed such that it detects structural breaks in the underlying time series which lead to a mean change in the kernel function h. We assume that the underlying time series is stationary before and after the change which is expressed by the following model:

$$X_{i,m} = 1_{\{1 \leq i \leq k^*+m\}} Y_i + 1_{\{i > k^*+m\}} Z_{i,m}, \quad i \geq 1, \tag{2.5}$$

where $\{Y_i\}_{i\in\mathbb{Z}}$ and $\{Z_{i,m}\}_{i\in\mathbb{Z}}$ are suitable stationary time series not necessarily with mean 0 with certain properties that will be specified in the respective sections. The distribution of the time series after the change and thus the change itself is allowed to depend on m. Let $Y \stackrel{\mathcal{D}}{=} Y_1$ and $Z_m \stackrel{\mathcal{D}}{=} Z_{1,m}$. Before the change, the kernel function has the same mean as under the null hypothesis which is given by

$$\theta = \mathrm{E}\, h(Y, \tilde{Y}),$$

where $\tilde{Y}$ is an independent copy of Y. After the structural break, the mean of the kernel function changes to

$$\theta_m^* = \mathrm{E}\, h(Y, \tilde{Z}_m)$$

for a copy $\tilde{Z}_m$ of Z_m which is independent of Y. The mean change in the kernel function is thus given by

$$\Delta_m := \theta_m^* - \theta. \tag{2.6}$$

A change in the observed time series is not visible for the proposed procedure if $\Delta_m = 0$. As will be seen in Theorem 3.14, a change is asymptotically detected with probability 1 if

$$\sqrt{m}\, |\Delta_m| \to \infty. \tag{2.7}$$

We consider the following change point model where $\tilde{X}_{i,m}$ is a copy of $X_{i,m}$ and is independent of $Y \stackrel{D}{=} Y_1$:

H_0 : $\mathrm{E}\, h(Y, \tilde{X}_{i,m}) = \theta$ for all $i \geq 1$.
H_1 : There exists a $k^* \geq 1$ such that

$$\begin{aligned} &\mathrm{E}\, h(Y, \tilde{X}_{i,m}) = \theta && \text{for } 1 \leq i \leq k^* + m; \\ &\mathrm{E}\, h(Y, \tilde{X}_{i,m}) = \theta + \Delta_m =: \theta_m^* && \text{for } i > k^* + m. \end{aligned}$$

As usual in change point problems, the parameter θ before the change, the size Δ_m of the change as well as the time of the change k^* are unknown.

The above model covers a wide range of alternatives, namely any change in the observed time series which implies a mean change in the kernel function. As we allow the time series after the change to depend on m, this does not only include fixed alternatives but also local ones where Δ_m is allowed to decay to zero but not too fast in the sense of restriction 2.7.

Example 2.3. *Let us consider a mean change according to the model*

$$X_{i,m} = Y_i + 1_{\{i>k^*+m\}} d_m, \quad d_m \neq 0, \tag{2.8}$$

where $\{Y_i\}_{i\geq 1}$ is a stationary time series with mean μ. The change in the mean is given by d_m and is allowed to depend on m. Note that in the general model (2.5) this corresponds to $Z_{i,m} = Y_i + d_m$. Recall the kernel functions of the CUSUM and Wilcoxon statistic as given in Example 2.2. Let $\tilde{Z}_m = \tilde{Y} + d_m$ where $\tilde{Y}$ is an independent copy of $Y \overset{\mathcal{D}}{=} Y_1$.

(i) For the CUSUM statistic we obtain

$$\theta_m^{*C} = \mathrm{E}\, h_C(Y, \tilde{Z}_m) = \mathrm{E}\left(Y - \tilde{Z}_m\right) = \mathrm{E}\, Y - \mathrm{E}\, \tilde{Z}_m = -d_m \neq 0.$$

Hence, with (2.3), the change in the kernel function is given by

$$\Delta_m^C = \theta_m^{*C} - \theta^C = -d_m \neq 0. \tag{2.9}$$

Consequently, fixed mean changes fulfill condition (2.7) and the decay of local mean changes is restricted to satisfy $\sqrt{m}|d_m| \to \infty$ in order to be detected asymptotically with probability one.

(ii) For the Wilcoxon statistic it holds for $d_m > 0$

$$\begin{aligned}\theta_m^{*W} &= \mathrm{E}\, h_W(Y, \tilde{Z}_m) = P(Y < \tilde{Y} + d_m) \\ &= P(Y < \tilde{Y} + d_m, Y < \tilde{Y}) + P(Y < \tilde{Y} + d_m, Y \geq \tilde{Y}) \\ &= P(Y < \tilde{Y}) + P(\tilde{Y} \leq Y < \tilde{Y} + d_m) = \theta^W + P(\tilde{Y} \leq Y < \tilde{Y} + d_m)\end{aligned}$$

and thus, with (2.4),

$$\begin{aligned}\Delta_m^W &= \theta_m^{*W} - \theta^W = P(\tilde{Y} \leq Y < \tilde{Y} + d_m) \\ &= \int_{-\infty}^{\infty} f_Y(z) P(z \leq Y < z + d_m) dz = \int_{-\infty}^{\infty} f_Y(z) \left(F_Y(z + d_m) - F_Y(z)\right) dz.\end{aligned} \tag{2.10}$$

There exist $z_1 \in \mathbb{R}$ and $0 < \delta_m < d_m$ such that $f(z) > 0$ holds for all $z \in [z_1, z_1 + \delta_m]$ as otherwise f_Y would be equal to zero except from countably many points such that f_Y would integrate to zero instead of one. As $F_Y(z + \delta_m/2) - F_Y(z) > 0$ holds for all $z \in (z_1, z_1 + \delta_m/2)$ it follows

$$\begin{aligned}&\int_{-\infty}^{\infty} f_Y(z) \left(F_Y(z + d_m) - F_Y(z)\right) dz \\ \geq &\int_{z_1}^{z_1+\delta_m/2} f_Y(z) \left(F_Y(z + \delta_m/2) - F_Y(z)\right) dz > 0.\end{aligned}$$

Hence, a fixed mean change satisfies condition (2.7) for the Wilcoxon-type testing procedure as $\delta_m = \delta$ is also fixed. For local mean changes condition (2.7) is fulfilled if there exists a $z_1 \in \mathbb{R}$ with $\sqrt{m} \int_{z_1}^{z_1+\delta_m/2} f_Y(z) \left(F_Y(z + \delta_m/2) - F_Y(z)\right) dz \to \infty$. An analogous result is obtained for $d_m < 0$.

3. Asymptotics

In this chapter, we analyze the asymptotic behavior of the proposed sequential procedures with respect to the length of the historic data set increasing to infinity. First, we introduce Hoeffding's decomposition which is a very useful tool for studying U-statistics. In Section 3.2 we derive the limit distribution of the test statistic under the null hypothesis which allows us to determine the asymptotic critical values. We continue with showing that the procedure has asymptotic power one in Section 3.3.

3.1. Hoeffding's Decomposition

As proposed by Hoeffding (1948) we consider the following decomposition of the kernel function

$$h(x_1, x_2) = \theta + h_1(x_1) + h_2(x_2) + r(x_1, x_2), \tag{3.1}$$

with

$$\begin{aligned} \theta &= \mathrm{E}(h(X_1, X_2)) \\ h_1(x_1) &= \mathrm{E}(h(x_1, X_2)) - \theta, \\ h_2(x_2) &= \mathrm{E}(h(X_1, x_2)) - \theta, \\ r(x_1, x_2) &= h(x_1, x_2) - h_1(x_1) - h_2(x_2) - \theta, \end{aligned}$$

where X_1 and X_2 are independent random variables with distribution functions F_{X_1} and F_{X_2}. With Fubini's Theorem it holds

$$\begin{aligned} \mathrm{E}\left(h_1(X_1)\right) &= \int h_1(x_1) dF_{X_1}(x_1) = \int\int h(x_1, x_2) dF_{X_2}(x_2) dF_{X_1}(x_1) - \theta \\ &= \int\int h(x_1, x_2) dF_{X_1}(x_1) dF_{X_2}(x_2) - \theta = \mathrm{E}\left(h_2(X_2)\right). \end{aligned}$$

Hence, $h_1(X_1)$ and $h_2(X_2)$ are centered as

$$\begin{aligned} \mathrm{E}\left(h_1(X_1)\right) = \mathrm{E}\left(h_2(X_2)\right) &= \int\int h(x_1, x_2) dF_{X_1}(x_1) dF_{X_2}(x_2) - \theta \\ &= \mathrm{E}(h(X_1, X_2)) - \theta = 0 \end{aligned} \tag{3.2}$$

due to the independence of X_1 and X_2.
The kernel $r(x_1, x_2)$ is degenerate which means

$$\mathrm{E}\left(r(x_1, X_2)\right) = \mathrm{E}\left(r(X_1, x_2)\right) = 0.$$

With (3.1) we obtain the following decomposition of the monitoring statistic:

$$\Gamma(m,k) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k} r(X_i,X_j) + \sum_{j=m+1}^{m+k} h_2(X_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(X_i). \tag{3.3}$$

Hence, Hoeffding's decomposition allows us to decompose the monitoring statistic in two linear and centered parts and a remainder term. The remainder term is not necessarily centered if the observations are not independent as

$$\mathrm{E}(r(X_i,X_j)) = \mathrm{E}(h(X_i,X_j)) - \theta \tag{3.4}$$

and θ is the expected value of the kernel function for independent random variables. Asymptotic results related to U-statistic are usually obtained by first showing that the remainder term is asymptotically negligible.

Example 3.1. *We derive the explicit form of Hoeffding's decomposition for the CUSUM and the Wilcoxon kernel (see Example 2.2) under the null hypothesis, i.e. we observe a stationary time series* $\{Y_i\}_{i\geq 1}$*. Let* $Y \overset{\mathcal{D}}{=} Y_1$ *and* $\mathrm{E}\,Y = \mu$*.*

(i) Hoeffding's decomposition of the CUSUM kernel is given by

$$\begin{aligned} h_1^C(y) &= \mathrm{E}\left(h_C(y,Y)\right) = y - E(Y) = y - \mu \\ h_2^C(y) &= \mathrm{E}\left(h_C(Y,y)\right) = \mathrm{E}(Y) - y = \mu - y = -h_1^C(y) \\ r^C(y_1,y_2) &= y_1 - y_2 - (y_1-\mu) - (\mu - y_2) = 0. \end{aligned}$$

(ii) Hoeffding's decomposition of the Wilcoxon kernel for a continuous random variable is given by

$$\begin{aligned} h_1^W(y) &= \mathrm{E}\left(h_W(y,Y)\right) - \frac{1}{2} = \mathrm{E}(1_{\{y<Y\}}) - \frac{1}{2} = P(y<Y) - \frac{1}{2} = \frac{1}{2} - F_Y(y) \\ h_2^W(y) &= \mathrm{E}\left(h_W(Y,y)\right) - \frac{1}{2} = \mathrm{E}(1_{\{Y<y\}}) - \frac{1}{2} = P(Y<y) - \frac{1}{2} = F_Y(y) - \frac{1}{2}, \\ r^W(y_1,y_2) &= 1_{\{y_1<y_2\}} - \left(\frac{1}{2} - F_Y(y_1)\right) - \left(F_Y(y_2) - \frac{1}{2}\right) - \frac{1}{2} \\ &= 1_{\{y_1<y_2\}} + F_Y(y_1) - F_Y(y_2) - \frac{1}{2}. \end{aligned}$$

3.2. Asymptotics under the Null Hypothesis

Our aim is to choose the critical value c_α such that the testing procedure has asymptotic size alpha, i.e. we can control the type-1 error asymptotically. Therefore we need to derive the limit distribution of the supremum over the weighted monitoring statistic under the null hypothesis.

Kirch & Tadjuidje Kamgaing (2015) identified the necessary properties that the weight function needs to possess such that the type-1 error can be controlled asymptotically. We also derive the asymptotic results for this general class of weight functions with the following regularity conditions:

Assumption 3.2. *Let the weight function satisfy*

(i) $w(m,k) = m^{-1/2}\tilde{w}(m,k)$, *where* $\tilde{w}(m,k) = \rho\left(\frac{k}{m}\right)$ *for* $k > l_m$ *with* $\frac{l_m}{m} \to 0$ *and* $\tilde{w}(m,k) = 0$ *for* $k \le l_m$. *The function* $\rho : [0,\infty] \to \mathbb{R}^+$ *is positive and continuous.*

(ii) $\lim_{t\to 0} t^\gamma \rho(t) < \infty$ *for some* $0 \le \gamma < \frac{1}{2}$.

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

Part (i) allows to start the monitoring only after some observations have been collected in order to avoid false alarms at the very beginning of the monitoring period which can easily be caused as the first values of the monitoring statistic are quite volatile due to the small monitoring sample. By (i) and (ii) the behavior of the weight function is controlled at the beginning and the infinite end of the monitoring period. The weight function

$$w(m,k) = m^{-1/2}\left(1+\frac{k}{m}\right)^{-1}\left(\frac{k}{m+k}\right)^{-\gamma}, \quad 0 \le \gamma < \frac{1}{2}, \tag{3.5}$$

i.e. $\rho(t) = (1+t)^{-1}\left(\frac{t}{1+t}\right)^{-\gamma}$ fulfills the above assumptions and is often used in the literature as it leads to nice limit distributions such as in Corollary 3.7.

All asymptotic results will be obtained based on Hoeffing's decomposition such that the regularity conditions on the time series and the kernel function are established with respect to the terms resulting from this decomposition. Therefore, recall that under the null hypothesis it holds with (3.3)

$$\Gamma(m,k) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k} r(Y_i,Y_j) + \sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i),$$

where $\{Y_i\}_{i\ge 1}$ is a stationary time series. The following regularity conditions are very general such that the proposed procedures can be applied for various dependency structures and kernel functions as long as Assumption 3.3 is fulfilled. Furthermore, the asymptotics for other monitoring schemes can be derived based on the same set of regularity conditions as will be seen in Chaper 4.

Assumption 3.3. *Let* $\{Y_i\}_{i\in\mathbb{Z}}$ *be a stationary time series that fulfills the following assumptions for a given kernel function* h.

(i) $\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r(Y_i,Y_j)\right|^2\right) \le u(m)(k_2-k_1+1)$ *for all* $m+1 \le k_1 \le k_2$ *with* $\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \to 0$ *and* γ *as in Assumption 3.2.*

(ii) The following functional central limit theorem holds for any $T > 0$

$$\left\{\frac{1}{\sqrt{m}}\sum_{i=1}^{[mt]}(h_1(Y_i),h_2(Y_i)) : 0 < t \le T\right\} \xrightarrow{D} \left\{\left(\tilde{W}_1(t),\tilde{W}_2(t)\right) : 0 < t \le T\right\},$$

where $\left\{\left(\tilde{W}_1(t), \tilde{W}_2(t)\right) : 0 < t \leq T\right\}$ *is a bivariate Wiener process with mean zero and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$$

with

$$\sigma_1^2 = \sum_{h \in \mathbb{Z}} Cov(h_1(Y_0), h_1(Y_h)), \quad \sigma_2^2 = \sum_{h \in \mathbb{Z}} Cov(h_2(Y_0), h_2(Y_h)). \tag{3.6}$$

(iii) For all $0 \leq \alpha < \frac{1}{2}$ *the following Hájek-Rényi-type inequality holds*

$$\sup_{1 \leq k \leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \left| \sum_{j=1}^{k} h_2(Y_j) \right| = O_P(1) \qquad \text{as } m \to \infty.$$

(iv) The following Hájek-Rényi-type inequality holds uniformly in m *for any sequence* $k_m > 0$

$$\sup_{k \geq k_m} \frac{1}{k} \left| \sum_{j=1}^{k} h_2(Y_j) \right| = O_P\left(\frac{1}{\sqrt{k_m}}\right) \qquad \text{as } k_m \to \infty.$$

The asymptotic negligibility of the remainder term can be established based on (i) where the second moment is equal to the variance for independent observations but not necessarily for dependent ones due to (3.4) and the respective explanation. The actual limit distribution is determined by the two linear terms. We assume that the respective partial sum processes fulfill a joint functional central limit theorem as stated in (ii). In addition, (iii) and (iv) are required to control the behavior at the beginning and at the infinite end of the monitoring period. In Section 3.2.1 and 3.2.2 we examine Assumption 3.3 for independent observations and functionals of mixing processes.

Under the above assumptions on the weight function, the time series and the kernel function, we now derive the asymptotic distribution of the test statistic

$$\sup_{k \geq 1} w(m,k) \left|\Gamma(m,k)\right|$$

under the null hypothesis. We first show that the remainder term is uniformly asymptotically negligible. The following Lemma states the key implication of Assumption 3.3 (i) which will be useful for dealing with the remainder term throughout this work.

Lemma 3.4. *Let* $\{Y_i\}_{i \geq 1}$ *and* $\{Y'_{i,m}\}_{i \geq 1}$ *be sequences of random variables and* $g_m : \mathbb{R}^2 \to \mathbb{R}$ *a function such that*

$$\mathrm{E}\left(\left|\sum_{i=1}^{m} \sum_{j=k_1}^{k_2} g_m(Y_i, Y'_{j,m})\right|^2\right) \leq u(m)(k_2 - k_1 + 1) \quad \text{for all } 1 \leq k_1 \leq k_2.$$

Then, it holds for any $\epsilon > 0$

$$P\left(\max_{1 \leq k \leq n} \left|\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| > \epsilon\right) \leq \frac{1}{\epsilon^2} \frac{u(m)}{m^2} (\log_2(2n))^2 n.$$

Proof. Let the process $\{G_m(k) : 1 \le k \le m\}$ be defined by

$$G_m(k) := \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m}).$$

We consider the increments of the process $G_m(k), k \ge 1$ as random variables

$$\xi_k = G_m(k) - G_m(k-1), \quad G_m(0) = 0.$$

Let $S_l = \xi_1 + \xi_2 + \ldots + \xi_l$ such that $S_l = G_m(l)$. It holds for $1 \le k_1 \le k_2$

$$\begin{aligned} \mathrm{E}\left(|S_{k_2} - S_{k_1+1}|^2\right) &= \mathrm{E}\left(|G_m(k_2) - G_m(k_1+1)|^2\right) \\ &= \frac{1}{m^2} \mathrm{E}\left(\left|\sum_{i=1}^{m} \sum_{j=k_1}^{k_2} g_m(Y_i, Y'_{j,m})\right|^2\right) \le \frac{u(m)}{m^2}(k_2 - k_1). \end{aligned} \tag{3.7}$$

Hence, we obtain with Theorem C.6 and Markov's inequality

$$P\left(\max_{1\le k\le n} |G_m(k)| \ge \epsilon\right) = P\left(\max_{1\le k\le n} |S_k| \ge \epsilon\right) \le \frac{u(m)}{\epsilon^2 m^2}(\log_2(2n))^2 n.$$

□

Lemma 3.5. *Let $\{Y_i\}_{i\ge1}$ and $\{Y'_{i,m}\}_{i\ge1}$ be sequences of random variables. Let Assumption 3.2 be fulfilled for the weight function. Assume that for $g_m : \mathbb{R}^2 \to \mathbb{R}$ it holds*

$$\mathrm{E}\left(\left|\sum_{i=1}^{m} \sum_{j=k_1}^{k_2} g_m(Y_i, Y'_{j,m})\right|^2\right) \le u(m)(k_2 - k_1 + 1) \quad \textit{for all } 1 \le k_1 \le k_2 \tag{3.8}$$

with $\frac{u(m)}{m^{2-2\gamma}} \log(m)^2 \to 0$ for all $\delta > 0$. Then, it follows

$$\sup_{k\ge1} w(m,k) \left|\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| = o_P(1), \quad \textit{as } m \to \infty.$$

Proof. With Lemma 3.4 we get for all $\epsilon > 0$

$$P\left(\left(\sqrt{\frac{u(m)}{m}} \log_2(2m)\right)^{-1} \max_{1\le k\le m} \left|\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| \ge \epsilon\right) \le \frac{1}{\epsilon^2}$$

and thus

$$\max_{1\le k\le m} \left|\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| = O_P\left(\sqrt{\frac{u(m)}{m}} \log_2(2m)\right) = O_P\left(\sqrt{\frac{u(m)}{m}} \log(m)\right). \tag{3.9}$$

Furthermore, it holds

$$
\begin{aligned}
&\max_{1\le k\le m} k^{\gamma} w(m,k)\sqrt{\frac{u(m)}{m}}\log(m)\\
=&\max_{l_m<k\le m}\sqrt{\frac{u(m)}{m}}\log(m)k^{\gamma}\rho\left(\frac{k}{m}\right)\\
=&\sqrt{\frac{u(m)}{m^{2-2\gamma}}\log(m)^2}\max_{l_m<k\le m}\left(\frac{k}{m}\right)^{\gamma}\rho\left(\frac{k}{m}\right)\\
\le&\sqrt{\frac{u(m)}{m^{2-2\gamma}}\log(m)^2}\sup_{0<t\le 1} t^{\gamma}\rho\left(t\right)=o_P(1)
\end{aligned}
\tag{3.10}
$$

as $t^{\gamma}\rho(t)$ is bounded on $(0,1]$ by Assumption 3.2 (ii). Combining this with (3.9) we obtain

$$
\max_{1\le k\le m} k^{\gamma} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i,Y'_{j,m})\right|=o_P(1) \tag{3.11}
$$

In particular, (3.11) implies

$$
\max_{1\le k\le m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i,Y'_{j,m})\right|=o_P(1). \tag{3.12}
$$

It remains to prove that

$$
\sup_{k>m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i,Y'_{j,m})\right|=o_P(1).
$$

Let $S_{l_1,l_2}:=\sum_{j=l_1}^{l_2}\xi_j$ with $\xi_j=\sum_{i=1}^{m} r_m(Y_i,Y'_{j,m})$. It holds with (3.8)

$$
\begin{aligned}
\mathrm{E}\left|S_{l_1,l_2}\right|^2&=\mathrm{E}\left(\sum_{j=l_1}^{l_2}\sum_{i=1}^{m} g_m(Y_i,Y'_{j,m})\right)^2\\
&=\mathrm{E}\left(\sum_{i=1}^{m}\sum_{j=l_1}^{l_2} g_m(Y_i,Y'_{j,m})\right)^2\le u(m)(l_2-l_1+1).
\end{aligned}
$$

We obtain with Corollary C.8 for $q_l\ge 2p_l-2$

$$
\begin{aligned}
&\mathrm{E}\left(\max_{p_l\le k\le q_l}\frac{1}{k\sqrt{m}}|S_{1,k}|\right)^2\le\frac{1}{p_l^2 m}\,\mathrm{E}\left(\max_{p_l\le k\le q_l}|S_{1,k}|\right)^2\\
\le&\frac{4u(m)}{p_l^2 m}\left(\log_2\left(4(q_l-p_l+1)\right)\right)^2(q_l-p_l+1).
\end{aligned}
\tag{3.13}
$$

If we choose p_l, q_l such that $q_l \geq 2p_l - 2$ and $\sum_{l\geq 0}[p_l, q_l) = [m, \infty)$, we get with (3.13) and Markov's inequality

$$\begin{aligned}
&P\left(\sup_{k>m}\left|\frac{1}{k\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| > \epsilon\right) \\
=&P\left(\sup_{k>m}\left|\frac{1}{k\sqrt{m}}\sum_{j=1}^{k}\sum_{i=1}^{m} g_m(Y_i, Y'_{j,m})\right| > \epsilon\right) \\
=&P\left(\sup_{k>m}\frac{1}{k\sqrt{m}}|S_{1,k}| > \epsilon\right) \\
\leq&P\left(\sup_{l\geq 0}\max_{p_l\leq k\leq q_l}\frac{1}{k\sqrt{m}}|S_{1,k}| > \epsilon\right) \\
\leq&\sum_{l\geq 0} P\left(\max_{p_l\leq k\leq q_l}\frac{1}{k\sqrt{m}}|S_{1,k}| > \epsilon\right) \\
\leq&\sum_{l\geq 0}\frac{4u(m)}{p_l^2 m\epsilon^2}\left(\log_2\left(4(q_l - p_l + 1)\right)\right)^2(q_l - p_l + 1).
\end{aligned}$$

With $p_l = m2^l, q_l = p_{l+1} = m2^{l+1}$, we obtain with (3.8)

$$\begin{aligned}
&P\left(\sup_{k>m}\left|\frac{1}{k\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| > \epsilon\right) \\
\leq&\sum_{l\geq 0}\frac{4u(m)}{p_l^2 m\epsilon^2}\left(\log_2\left(4(q_l - p_l + 1)\right)\right)^2(q_l - p_l + 1) \\
=&\sum_{l\geq 0}\frac{4u(m)}{2^{2l}m^3\epsilon^2}\left(\log_2\left(4(m2^l + 1)\right)\right)^2(m2^l + 1) \\
\leq&\sum_{l\geq 0}\frac{4u(m)}{m^2\epsilon^2}(\log_2(m2^{l+3}))^2\left(\frac{1}{2^l} + \frac{1}{m2^{2l}}\right) \\
\leq&\frac{8u(m)}{m^2\epsilon^2}\sum_{l\geq 0}\frac{(\log_2(m2^{l+3}))^2}{2^l} \\
=&\frac{8u(m)}{m^2\epsilon^2}\left(\sum_{l\geq 0}\frac{\log_2(m)^2}{2^l} + 2\sum_{l\geq 0}\frac{\log_2(m)(l+3)}{2^l} + \sum_{l\geq 0}\frac{(l+3)^2}{2^l}\right) \\
=&\frac{8}{\epsilon^2}\frac{u(m)}{m^{2-2\gamma}}\log(m)^2\frac{1}{m^{2\gamma}}O(1) \to 0 \quad \text{as } m \to \infty,
\end{aligned}$$

as $\sum_{l\geq 0}\frac{1}{2^l} < \infty, 2\sum_{l\geq 0}\frac{l+3}{2^l} < \infty$ and $C_3 = \sum_{l\geq 0}\frac{(l+3)^2}{2^l} < \infty$. Hence, it holds

$$\sup_{k>m}\left|\frac{1}{k\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| = o_P(1) \quad \text{as } m \to \infty \tag{3.14}$$

and with

$$\sup_{k>m} w(m,k)\frac{k}{\sqrt{m}} \leq \sup_{k>m}\frac{k}{m}\rho\left(\frac{k}{m}\right) \leq \sup_{t>1} t\rho\left(t\right) = O_P(1) \quad \text{as } m \to \infty \tag{3.15}$$

it follows

$$\sup_{k>m} w(m,k) \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m}) \right| = o_P(1) \quad \text{as } m \to \infty \tag{3.16}$$

Together with (3.12) we obtain

$$\begin{aligned}
&\sup_{k\geq 1} w(m,k) \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m}) \right| \\
\leq &\max_{1\leq k\leq m} w(m,k) \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m}) \right| + \sup_{k>m} w(m,k) \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} g_m(Y_i, Y'_{j,m}) \right| \\
= &o_P(1) \quad \text{as } m \to \infty
\end{aligned}$$

and thus the assertion. □

Having shown that the remainder term is uniformly asymptotically negligible, we can now derive the limit distribution based on the linear parts.

Theorem 3.6. *Let the regularity conditions given in Assumption 3.2 and 3.3 be fulfilled. Then, it holds under H_0, as $m \to \infty$,*

$$\sup_{k\geq 1} w(m,k) \left| \Gamma(m,k) \right| \xrightarrow{\mathcal{D}} \sup_{t>0} \rho(t) \left| \sigma_2 W_2(t) + t\sigma_1 W_1(1) \right|$$

for two independent standard Wiener processes $\{W_1(t) : t > 0\}$ and $\{W_2(t) : t > 0\}$.

Proof. With Lemma 3.5 it holds

$$\begin{aligned}
&\sup_{k\geq 1} \left| w(m,k)\Gamma(m,k) - w(m,k) \sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m} \sum_{i=1}^{m} h_1(Y_i) \right| \\
= &\sup_{k\geq 1} w(m,k) \frac{1}{m} \left| \sum_{i=1}^{m} \sum_{j=m+1}^{m+k} r(Y_i, Y_j) \right| \\
= &\sup_{k\geq 1} w(m,k) \frac{1}{m} \left| \sum_{i=1}^{m} \sum_{j=1}^{k} r(Y_i, Y_{m+j}) \right| = o_P(1), \quad m \to \infty
\end{aligned}$$

due to Assumption 3.3 (i). According to Lemma B.1 it remains to be shown that

$$\sup_{k\geq 1} w(m,k) \left| \sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m} \sum_{i=1}^{m} h_1(Y_i) \right| \xrightarrow{\mathcal{D}} \sup_{t>0} \rho(t) \left| \sigma_2 W_2(t) + t\sigma_1 W_1(1) \right|$$

for two independent standard Wiener processes $\{W_1(t) : t > 0\}$ and $\{W_2(t) : t > 0\}$. With Assumptions 3.2 and 3.3 (ii) we get that for any fixed $\tau, T > 0$ and $m \to \infty$

$$\begin{aligned}
&\sup_{t>\tau} \rho(t) \left| \frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
\xrightarrow{\mathcal{D}} &\sup_{t>\tau} \rho(t) \left| \tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1) + t\tilde{W}_1(1) \right|.
\end{aligned} \tag{3.17}$$

As $\frac{l_m}{m} \to 0$, there exists an m_τ such that $\frac{l_m}{m} < \tau$ for all $m \geq m_\tau$. Hence, it holds for $m \geq m_\tau$

$$\begin{aligned}
&\left| \sup_{t>\tau} \rho(t) \left| \frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \right. \\
&\left. - \sup_{k>\tau m} w(m,k) \left| \sum_{j=m+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k}{m} \sum_{i=1}^{m} h_1(Y_i) \right| \right| \\
= &\left| \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \rho(t) \left| \frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \right. \\
&\left. - \sup_{\frac{k}{m}>\tau} \rho\left(\frac{k}{m}\right) \left| \frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k}{m}\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \right| \\
\leq &\sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| \rho(t) \left(\frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \right. \\
&\left. -\rho\left(\frac{k}{m}\right) \left(\frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k}{m}\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \right| \\
\leq &\sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right) \right| \sup_{t>\tau} \left| \frac{1}{t}\frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&+ \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| \left(\frac{m}{k} - \frac{1}{t}\right) \frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) \right| \\
\leq &\sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right) \right| \sup_{\frac{k}{m}>\tau} \left| \frac{1}{t}\frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&+ \frac{1}{m\tau^2} \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| \frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+[m\min(t,T)]} h_2(Y_j) \right| \\
= &\sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right) \right| O_P(1) + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) o_P(1) \quad \text{as } m \to \infty
\end{aligned}$$

with Assumption 3.3 (ii). Due to Assumption 3.2 (iii) it holds

$$\sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) < \sup_{t>\tau} t\rho(t) < \infty. \tag{3.18}$$

Let $\epsilon > 0$. By Assumption 3.2 (iii) there exists a $T_\epsilon \geq \tau$ such that $\left|t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| < \epsilon$ for $t, \frac{k}{m} > T_\epsilon$. Hence, we obtain

$$\sup_{\frac{k}{m}>T_\epsilon} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right) \right| < \epsilon. \tag{3.19}$$

Furthermore, as $t\rho(t)$ is uniformly continuous on $[\tau, T_\epsilon]$, there exists a $\delta > 0$ such that

$$\left|t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| < \epsilon \quad \text{for all } t, \frac{k}{m} \in [\tau, T_\epsilon] \text{ with } \left|t - \frac{k}{m}\right| < \delta.$$

With $m_0 := \lceil \frac{1}{\delta} \rceil$ it holds $\left|t - \frac{k}{m}\right| < \frac{1}{m} \leq \delta$ for all $m \geq m_0, t \in \left[\frac{k}{m}, \frac{k+1}{m}\right)$. It follows

$$\sup_{\frac{k}{m}\in[\tau,T_\epsilon]} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left|t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| < \epsilon \quad \text{for all } m \geq m_0$$

and consequently

$$\sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left|t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| \to 0 \quad \text{as } m \to \infty. \tag{3.20}$$

It follows

$$\left|\sup_{t>\tau}\rho(t)\left|\frac{1}{\sqrt{m}}\sum_{j=m+1}^{m+\min([mt],mT)} h_2(Y_j) + \frac{t}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right|\right.$$
$$\left. - \sup_{k>\tau m} w(m,k)\left|\sum_{j=m+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right|\right| = o_P(1) \quad \text{as } m \to \infty$$

and with (3.17) we get

$$\sup_{k>\tau m} w(m,k)\left|\sum_{j=m+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right|$$
$$\xrightarrow{\mathcal{D}} \sup_{t>\tau}\rho(t)\left|\tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1) + t\tilde{W}_1(1)\right|. \tag{3.21}$$

By Assumption 3.3 (iv) and the stationarity we obtain

$$\begin{aligned}
&\sup_{k>mT} w(m,k)\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
=&\sup_{k>mT} \frac{kw(m,k)}{\sqrt{mT}} \cdot \frac{\sqrt{mT}}{k}\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
\overset{\mathcal{D}}{=}&\sup_{k>\max(mT,l_m)} \frac{k\rho\left(\frac{k}{m}\right)}{m\sqrt{T}} \cdot \frac{\sqrt{mT}}{k}\left|\sum_{j=1}^{k} h_2(Y_j)\right| \\
\leq&\frac{1}{\sqrt{T}}\sup_{k>mT}\frac{k}{m}\rho\left(\frac{k}{m}\right)\sup_{k>mT}\frac{\sqrt{mT}}{k}\left|\sum_{j=1}^{k} h_2(Y_j)\right| \\
\leq&\frac{1}{\sqrt{T}}\sup_{k>mT}\frac{k}{m}\rho\left(\frac{k}{m}\right)O_P(1) \\
\leq&\frac{1}{\sqrt{T}}\sup_{t>T} t\rho(t)O_P(1) \\
=&\frac{1}{\sqrt{T}}O_P(1) \xrightarrow{P} 0 \quad \text{as } T \to \infty \text{ uniformly in m}
\end{aligned} \tag{3.22}$$

as $\lim_{t\to\infty} t\rho(t) < \infty$. Hence, we get

$$\begin{aligned}
&\sup_{k>\tau m}\Bigg| w(m,k)\left(\sum_{j=m+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right) \\
&\qquad - w(m,k)\left(\sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right)\Bigg| \\
&= \sup_{k>\tau m}\left| w(m,k)\sum_{j=m+1}^{m+\min(k,mT)} h_2(Y_j) - w(m,k)\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\le \sup_{\tau m<k\le mT}\left| w(m,k)\sum_{j=m+1}^{m+k} h_2(Y_j) - w(m,k)\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\qquad + \sup_{k>mT}\left| w(m,k)\sum_{j=m+1}^{m+mT} h_2(Y_j) - w(m,k)\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&= \sup_{k>mT}\left| w(m,k)\sum_{j=m+1}^{m+mT} h_2(Y_j) - w(m,k)\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\le \sup_{k>mT} w(m,k)\left|\sum_{j=m+1}^{m+mT} h_2(Y_j)\right| + \sup_{k>mT} w(m,k)\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\le \sup_{k>mT} \frac{k}{m}\rho\left(\frac{k}{m}\right)\sqrt{\frac{m}{k}}\frac{1}{\sqrt{k}}\left|\sum_{j=m+1}^{m+mT} h_2(Y_j)\right| + \sup_{k>mT} w(m,k)\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\le \sqrt{\frac{1}{T}}\sup_{t>T} t\rho\left(t\right) O_P(1) + o_P(1) = o_P(1) \quad \text{as } T\to\infty \text{ uniformly in m} \qquad (3.23)
\end{aligned}$$

noting that Assumption 3.2 (iii) and the continuity of ρ yield that $\rho(t)$ as well as $t\rho(t)$ are bounded on (T,∞) for all $T \ge 1$ such that

$$\sup_{t>T}\rho(t) = O(1) \quad \text{and} \quad \sup_{t>T} t\rho(t) = O(1). \qquad (3.24)$$

For $\gamma < \alpha < \frac{1}{2}$ it holds

$$\begin{aligned}
&\sup_{1\le k\le\tau m} w(m,k)\left|\sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right| \\
&\le \sup_{1\le k\le\tau m} w(m,k)\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| + \sup_{1\le k\le\tau m} w(m,k)\left|\frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right| \\
&= \sup_{1\le k\le\tau m} m^{\frac{1}{2}-\alpha}k^{\alpha} w(m,k)\left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\quad + \sup_{1\le k\le\tau m} \frac{k}{m} w(m,k)\sqrt{m}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right|
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{l_m<k\leq \tau m} \left(\frac{k}{m}\right)^{\alpha} \rho\left(\frac{k}{m}\right) \left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \sum_{j=m+1}^{m+k} h_2(Y_j)\right| + \sup_{l_m<k\leq \tau m} \frac{k}{m}\rho\left(\frac{k}{m}\right) \left|\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i)\right| \\
&\leq \sup_{1\leq k\leq \tau m} \left(\frac{k}{m}\right)^{\alpha} \rho\left(\frac{k}{m}\right) \sup_{1\leq k\leq \tau m} \left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\quad + \sup_{1\leq k\leq \tau m} \frac{k}{m}\rho\left(\frac{k}{m}\right) \left|\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i)\right|. \qquad (3.25)
\end{aligned}
$$

Assumption 3.2 (ii) yields that $t^{\gamma}\rho(t)$ is bounded on $(0,\tau]$ such that

$$
\sup_{1\leq k\leq \tau m} \left(\frac{k}{m}\right)^{\alpha} \rho\left(\frac{k}{m}\right) \leq \sup_{0<t\leq \tau} t^{\alpha}\rho\left(t\right) \leq \tau^{\alpha-\gamma} \sup_{0<t\leq \tau} t^{\gamma}\rho\left(t\right) \to 0 \quad \text{as } \tau\to 0 \qquad (3.26)
$$

and

$$
\sup_{1\leq k\leq \tau m} \frac{k}{m}\rho\left(\frac{k}{m}\right) \leq \sup_{0<t\leq \tau} t\rho\left(t\right) \leq \tau^{1-\gamma} \sup_{0<t\leq \tau} t^{\gamma}\rho\left(t\right) \to 0 \quad \text{as } \tau\to 0. \qquad (3.27)
$$

With Assumption 3.3 (iii) and the stationarity it holds for $\tau<1$

$$
\begin{aligned}
\sup_{1\leq k\leq \tau m} \left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \sum_{j=m+1}^{m+k} h_2(Y_j)\right| &\leq \sup_{1\leq k\leq m} \left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\overset{\mathcal{D}}{=} \sup_{1\leq k\leq m} \left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \sum_{j=1}^{k} h_2(Y_j)\right| = O_P(1) \quad \text{as } m\to\infty
\end{aligned}
$$

and with (ii) it holds

$$
\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) = O_P(1) \quad \text{as } m\to\infty.
$$

Hence, (3.25) yields

$$
\sup_{1\leq k\leq \tau m} w(m,k) \left|\sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right| = o_P(1) \quad \text{as } \tau\to 0 \text{ uniformly in m.} \qquad (3.28)
$$

With the law of the iterated logarithm for Wiener Processes (see Csörgő & Révész (1981)) it holds

$$
\begin{aligned}
&\sup_{t>T} \frac{\left|\tilde{W}_2(1+t)\right|}{1+t} \\
&= \sup_{t>T} \frac{\sqrt{(1+t)\log\log(1+t)}}{1+t} \frac{\left|\tilde{W}_2(1+t)\right|}{\sqrt{(1+t)\log\log(1+t)}}
\end{aligned}
$$

$$\leq \sup_{t>T} \frac{\sqrt{(1+t)\log\log(1+t)}}{1+t} \sup_{t>T} \frac{\left|\tilde{W}_2(1+t)\right|}{\sqrt{(1+t)\log\log(1+t)}}$$

$$\leq \sup_{t>T} \sqrt{\frac{\log\log(1+t)}{1+t}} \sup_{t>T} \frac{\left|\tilde{W}_2(1+t)\right|}{\sqrt{(1+t)\log\log(1+t)}} = o(1) \quad \text{a.s} \quad \text{as } T \to \infty \tag{3.29}$$

and thus

$$\begin{aligned}
&\sup_{t>T} \rho(t) \left|\tilde{W}_2(1+t)\right| \\
=&\sup_{t>T} (1+t)\rho(t) \frac{\left|\tilde{W}_2(1+t)\right|}{1+t} \\
\leq&\sup_{t>T} (1+t)\rho(t) \sup_{t>T} \frac{\left|\tilde{W}_2(1+t)\right|}{1+t} \\
\leq&(\sup_{t>T} \rho(t) + \sup_{t>T} t\rho(t)) \sup_{t>T} \frac{\left|\tilde{W}_2(1+t)\right|}{1+t} = o(1) \quad \text{a.s} \quad \text{as } T \to \infty.
\end{aligned} \tag{3.30}$$

Furthermore, it holds

$$\sup_{t>T} \left|\rho(t)\tilde{W}_2(1+T)\right| \tag{3.31}$$

$$\leq \sup_{t>T} t\rho(t) \left|\frac{1}{T}\tilde{W}_2(1+T)\right| \tag{3.32}$$

$$= \sup_{t>T} t\rho(t) \frac{1+T}{T} \sqrt{\frac{\log\log(1+T)}{1+T}} \left|\frac{\tilde{W}_2(1+T)}{\sqrt{(1+T)\log\log(1+T)}}\right| = o(1) \text{ a.s.} \quad \text{as } T \to \infty. \tag{3.33}$$

Hence, we obtain

$$\begin{aligned}
&\sup_{t>\tau} \left|\rho(t)\left(\tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1) + t\tilde{W}_1(1)\right) - \rho(t)\left(\tilde{W}_2(1+t) - \tilde{W}_2(1) + t\tilde{W}_1(1)\right)\right| \\
=&\sup_{t>\tau} \left|\rho(t)\tilde{W}_2(1+\min(t,T)) - \rho(t)\tilde{W}_2(1+t)\right| \\
\leq& \sup_{\tau<t\leq T} \left|\rho(t)\tilde{W}_2(1+t) - \rho(t)\tilde{W}_2(1+t)\right| + \sup_{t>T} \left|\rho(t)\tilde{W}_2(1+T) - \rho(t)\tilde{W}_2(1+t)\right| \\
=&\sup_{t>T} \left|\rho(t)\tilde{W}_2(1+T) - \rho(t)\tilde{W}_2(1+t)\right| \\
\leq&\sup_{t>T} \left|\rho(t)\tilde{W}_2(1+T)\right| + \sup_{t>T} \left|\rho(t)\tilde{W}_2(1+t)\right| = o(1) \text{ a.s.} \quad \text{as } T \to \infty.
\end{aligned} \tag{3.34}$$

Let $\{W_2(t) : 0 < t \leq T\} := \{\frac{1}{\sigma_2}(\tilde{W}_2(1+t) - \tilde{W}_2(1)) : 0 < t \leq T\}$ and $\{W_1(t) : 0 < t \leq T\} := \{\frac{1}{\sigma_2}\tilde{W}_1(t) : 0 < t \leq T\}$ which are independent standard Wiener Processes. It holds

$$\tilde{W}_2(1+t) - \tilde{W}_2(1) + t\tilde{W}_1(1) = \sigma_2 W_2(t) + t\sigma_1 W_1(1), \quad 0 < t \leq T. \tag{3.35}$$

Note that

$$\begin{aligned}
&\sup_{0<t\le\tau} \rho(t)\left|\sigma_2 W_2(t) + t\sigma_1 W_1(1)\right| \\
\le &\sup_{0<t\le\tau} \rho(t)\left|\sigma_2 W_2(t)\right| + \sup_{0<t\le\tau} t\rho(t)\left|\sigma_1 W_1(1)\right| \\
\le &\sup_{0<t\le\tau} \rho(t)\left|\sigma_2 W_2(t)\right| + \tau^{1-\gamma} \sup_{0<t\le\tau} t^{\gamma}\rho\left(t\right)\left|\sigma_1 W_1(1)\right|.
\end{aligned}$$

With the self-similarity of the Wiener Process and the law of the iterated logarithm we obtain

$$\begin{aligned}
&\sup_{0<t\le\tau} \rho(t)\left|\sigma_2 W_2(t)\right| \\
=&\sigma_2 \sup_{s\ge\frac{1}{\tau}} \rho\left(\frac{1}{s}\right)\left|W_2\left(\frac{1}{s}\right)\right| \\
\overset{\mathcal{D}}{=}&\sigma_2 \sup_{s\ge\frac{1}{\tau}} \left|W_2(s)\right| \frac{1}{s}\rho\left(\frac{1}{s}\right) \\
=&\sigma_2 \sup_{s\ge\frac{1}{\tau}} \frac{\sqrt{s\log\log s}}{s}\rho\left(\frac{1}{s}\right)\frac{\left|W_2(s)\right|}{\sqrt{s\log\log s}} \\
\le&\sigma_2 \sup_{s\ge\frac{1}{\tau}} \frac{\sqrt{\log\log s}}{s^{1-\gamma-\frac{1}{2}}} \sup_{s\ge\frac{1}{\tau}} \left(\frac{1}{s}\right)^{\gamma}\rho\left(\frac{1}{s}\right) \sup_{s\ge\frac{1}{\tau}} \frac{\left|W_2(s)\right|}{\sqrt{s\log\log s}} \\
=&o_P(1) \quad \text{as } \tau\to 0.
\end{aligned} \tag{3.36}$$

Together with (3.35) it follows

$$\sup_{0<t\le\tau} \rho(t)\left|\tilde{W}_2(1+t) - \tilde{W}_2(1) + t\tilde{W}_1(1)\right| = o_P(1) \quad \text{as } \tau\to 0. \tag{3.37}$$

Based on Lemma B.2 we can combine (3.21), (3.23), (3.28), (3.34), (3.35) and (3.37) such that we obtain

$$\sup_{k\ge1} w(m,k)\left|\sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i)\right| \xrightarrow{\mathcal{D}} \sup_{t>0} \rho(t)\left|\sigma_2 W_2(t) + t\sigma_1 W_1(1)\right| \tag{3.38}$$

and thus the assertion. □

However, the critical values cannot be determined based on the limit distribution in Theorem 3.6 as it contains the unknown parameters σ_1 and σ_2. Nevertheless, this result can be used to derive pivotal limit distributions as stated in the following Corollary. The unknown parameters can be estimated consistently based on the historic data set.

Corollary 3.7. *(i) If $\sigma_1 = \sigma_2 =: \sigma$ the limit distribution in Theorem 3.6 reduces to $\sup_{0<s<1} \rho\left(\frac{t}{1-t}\right)\frac{|W(t)|}{1-t}$, where $\{W(t): t\ge 0\}$ is a standard Wiener Process, such that*

$$\frac{1}{\hat{\sigma}_m}\sup_{k\ge1}\frac{|\Gamma(m,k)|}{g(m,k)} \xrightarrow{\mathcal{D}} \sup_{0<s<1}\rho\left(\frac{s}{1-s}\right)\frac{|W(s)|}{1-s}$$

if $\hat{\sigma}_m \xrightarrow{P} \sigma$. For $w(m,k) = m^{-1/2}\left(1+\frac{k}{m}\right)^{-1}\left(\frac{k}{m+k}\right)^{-\gamma}$, i.e. $\rho(t) = (1+t)^{-1}\left(\frac{t}{1+t}\right)^{-\gamma}$, the above limit even simplifies to $\sup_{0<t<1}\frac{|W(t)|}{t^{\gamma}}$.

(ii) *For $\sigma_1 \neq \sigma_2$ consider $\tilde{w}(m,k) = m^{-1/2}\sigma_1\left(\sigma_2^2 + \sigma_1^2\frac{k}{m}\right)^{-1}$. Then, it holds*

$$\sup_{k\geq 1} \tilde{w}(m,k)\left|\Gamma(m,k)\right| \xrightarrow{\mathcal{D}} \sup_{0<t<1} \left|W(t)\right|,$$

where $\{W(t) : t \geq 0\}$ is a standard Wiener Process, such that

$$\sup_{k\geq 1} \frac{\hat{\sigma}_{1,m}^2}{\sqrt{m}\left(\hat{\sigma}_{2,m}^2 + \hat{\sigma}_{1,m}^2\frac{k}{m}\right)}\left|\Gamma(m,k)\right| \xrightarrow{\mathcal{D}} \sup_{0<t<1}\left|W(t)\right|$$

if $\hat{\sigma}_{1,m} \xrightarrow{P} \sigma_1$ and $\hat{\sigma}_{2,m} \xrightarrow{P} \sigma_2$.

Proof. (i) For $\sigma_1 = \sigma_2 =: \sigma$ we obtain with Theorem 3.6

$$\sup_{k\geq 1} w(m,k)\left|\Gamma(m,k)\right| \xrightarrow{\mathcal{D}} \sup_{t>0} \sigma\rho(t)\left|W_2(t) + tW_1(1)\right|.$$

In order to get a simpler representation of the limit distribution we consider the process

$$\left\{(1+t)W\left(\frac{t}{1+t}\right) : t>0\right\},$$

where $\{W(t) : t \geq 0\}$ is a standard Wiener Process. As $\{W_2(t) + tW_1(1) : t \geq 0\}$ and $\left\{(1+t)W\left(\frac{t}{1+t}\right) : t \geq 0\right\}$ are Gaussian Processes with almost sure continuous sample paths, their distribution is determined by their mean and covariance structure. Both of the processes have mean zero and it holds

$$\begin{aligned}
&\operatorname{Cov}\left((1+t)W\left(\frac{t}{1+t}\right), (1+s)W\left(\frac{s}{1+s}\right)\right) \\
=&(1+t)(1+s)\min\left(\frac{t}{1+t}, \frac{s}{1+s}\right) \\
=&\min(t(1+s), s(1+t)) = \min(t,s) + ts = \operatorname{Cov}\left(W_2(t) + tW_1(1), W_2(s) + sW_1(1)\right).
\end{aligned}$$

Thus, we get

$$\{W_2(t) + tW_1(1) : t \geq 0\} \stackrel{\mathcal{D}}{=} \left\{(1+t)W\left(\frac{t}{1+t}\right) : t > 0\right\} \tag{3.39}$$

such that

$$\sup_{t>0}\rho(t)\left|W_2(t) + tW_1(1)\right| \stackrel{\mathcal{D}}{=} \sup_{t>0}\rho(t)(1+t)\left|W\left(\frac{t}{1+t}\right)\right| \stackrel{\mathcal{D}}{=} \sup_{0<s<1} \rho\left(\frac{s}{1-s}\right)\frac{|W(s)|}{1-s}.$$

Hence,

$$\sup_{k\geq 1} w(m,k)\left|\Gamma(m,k)\right| \xrightarrow{\mathcal{D}} \sigma \sup_{0<s<1} \rho\left(\frac{s}{1-s}\right)\frac{|W(s)|}{1-s}.$$

As

$$\left|\frac{1}{\hat{\sigma}_m} - \frac{1}{\sigma}\right| = \left|\frac{\hat{\sigma}_m - \sigma}{\hat{\sigma}_m\sigma}\right| = o_P(1),$$

it follows

$$\frac{1}{\hat{\sigma}_m}\sup_{k\geq 1} w(m,k)\left|\Gamma(m,k)\right| \xrightarrow{\mathcal{D}} \sup_{0<s<1} \rho\left(\frac{s}{1-s}\right)\frac{|W(s)|}{1-s}.$$

(ii) With Theorem 3.6 it holds

$$\sup_{k\geq 1} \tilde{w}(m,k)\,|\Gamma(m,k)| \xrightarrow{\mathcal{D}} \sup_{t>0} \frac{|\sigma_1^2(\sigma_2 W_2(t) + t\sigma_1 W_1(1))|}{\sigma_2^2 + \sigma_1^2 t}.$$

Comparing the covariance structures we get

$$\{\sigma_1^2\left(\sigma_2 W_2(t) + t\sigma_1 W_1(1)\right) : t > 0\} \stackrel{\mathcal{D}}{=} \left\{(\sigma_2^2 + \sigma_1^2 t) W\left(\frac{\sigma_1^2 t}{\sigma_2^2 + \sigma_1^2 t}\right) : t > 0\right\}.$$

as

$$\begin{aligned}
&\mathrm{Cov}\left((\sigma_2^2 + \sigma_1^2 t) W\left(\frac{\sigma_1^2 t}{\sigma_2^2 + \sigma_1^2 t}\right), (\sigma_2^2 + \sigma_1^2 s) W\left(\frac{\sigma_1^2 s}{\sigma_2^2 + \sigma_1^2 s}\right)\right)\\
=&(\sigma_2^2 + \sigma_1^2 t)(\sigma_2^2 + \sigma_1^2 s) \min\left(\frac{\sigma_1^2 t}{\sigma_2^2 + \sigma_1^2 t}, \frac{\sigma_1^2 s}{\sigma_2^2 + \sigma_1^2 s}\right)\\
=&\sigma_1^2 \min\left(t(\sigma_2^2 + \sigma_1^2 s), s(\sigma_2^2 + \sigma_1^2 t)\right) = \sigma_1^2 \min\left(\sigma_2^2 t + \sigma_1^2 st, \sigma_2^2 s + \sigma_1^2 st\right)\\
=&\sigma_1^2\left(\sigma_2^2 \min(s,t) + \sigma_1^2 st\right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}\left(\sigma_1\left(\sigma_2 W_2(t) + t\sigma_1 W_1(1)\right), \sigma_1\left(\sigma_2 W_2(s) + s\sigma_1 W_1(1)\right)\right)\\
=&\sigma_1^2\left(\sigma_2^2 \min(s,t) + \sigma_1^2 st\right).
\end{aligned}$$

Thus, it holds

$$\sup_{t>0} \frac{|\sigma_1(\sigma_2 W_2(t) + t\sigma_1 W_1(1))|}{\sigma_2^2 + \sigma_1^2 t} \stackrel{\mathcal{D}}{=} \sup_{t>0}\left|W\left(\frac{\sigma_1^2 t}{\sigma_2^2 + \sigma_1^2 t}\right)\right| = \sup_{0<s<1} |W(s)|.$$

If $\hat{\sigma}_{1,m} \xrightarrow{P} \sigma_1$ and $\hat{\sigma}_{2,m} \xrightarrow{P} \sigma_2$, we obtain

$$\begin{aligned}
&\left|\sup_{k\geq 1} \frac{\frac{\sigma_2^2}{\sigma_1} + \sigma_1 \frac{k}{m}}{\frac{\hat{\sigma}_{m,2}^2}{\hat{\sigma}_{m,1}} + \hat{\sigma}_{1,m}\frac{k}{m}} - 1\right|\\
\leq&\frac{\left|\frac{\sigma_2^2}{\sigma_1} - \frac{\hat{\sigma}_{m,2}^2}{\hat{\sigma}_{m,1}}\right|}{\frac{\hat{\sigma}_{m,2}^2}{\hat{\sigma}_{m,1}}} + \sup_{1\leq k\leq m} \frac{\frac{k}{m}\left|\sigma_1 - \hat{\sigma}_{m,1}\right|}{\frac{\hat{\sigma}_{2,m}^2}{\hat{\sigma}_{m,1}}} + \sup_{k>m} \frac{\frac{k}{m}\left|\sigma_1 - \frac{\sigma_1}{\hat{\sigma}_{m,1}^2}\right|}{\frac{\hat{\sigma}_{2,m}^2}{\hat{\sigma}_{m,1}} + \hat{\sigma}_{1,m}\frac{k}{m}}\\
\leq&\frac{\left|\frac{\sigma_2^2}{\sigma_1} - \frac{\hat{\sigma}_{m,2}^2}{\hat{\sigma}_{m,1}}\right|}{\frac{\hat{\sigma}_{m,2}^2}{\hat{\sigma}_{m,1}}} + \frac{\left|\sigma_1 - \hat{\sigma}_{m,1}\right|}{\frac{\hat{\sigma}_{2,m}^2}{\hat{\sigma}_{m,1}}} + \frac{\left|\sigma_1 - \frac{\sigma_1}{\hat{\sigma}_{m,1}^2}\right|}{\hat{\sigma}_{1,m}} = o_P(1)
\end{aligned}$$

and thus

$$\begin{aligned}
&\left|\sup_{k\geq 1} \frac{\hat{\sigma}_{m,1}}{\sqrt{m}\left(\hat{\sigma}_{m,2}^2 + \hat{\sigma}_{m,1}^2 \frac{k}{m}\right)} |\Gamma(m,k)| - \sup_{k\geq 1} \frac{\sigma_1}{\sqrt{m}\left(\sigma_2^2 + \sigma_1^2 \frac{k}{m}\right)} |\Gamma(m,k)|\right|\\
\leq&\left|\sup_{k\geq 1} \frac{\frac{\sigma_2^2}{\sigma_1} + \sigma_1 \frac{k}{m}}{\frac{\hat{\sigma}_{m,2}^2}{\hat{\sigma}_{m,1}} + \hat{\sigma}_{1,m}\frac{k}{m}} - 1\right| \sup_{k\geq 1} \tilde{w}(m,k)\,|\Gamma(m,k)| = o_P(1).
\end{aligned}$$

□

Corollary 3.7 (i) holds obviously for all kernels with $h_2(y) = -h_1(y)$ or $h_2(y) = h_1(y)$ which in particular includes symmetric or anti-symmetric kernels. According to Example 3.1, the CUSUM kernel is antisymmetric and the Wilcoxon kernel fulfills $h_2(y) = -h_1(y)$.

3.2.1. Independent and Identically Distributed Random Variables

Let $Y_1, Y_2, \dots$ be i.i.d. random variables with distribution function F and let $h : \mathbb{R}^2 \to \mathbb{R}$ be a kernel with

$$Eh^2(Y_1, Y_2) < \infty. \tag{3.40}$$

We assume there exists a $\nu > 2$ such that

$$\begin{aligned} &0 < \mathrm{E}\,|h_1(Y_1)|^\nu < \infty \\ &0 < \mathrm{E}\,|h_2(Y_1)|^\nu < \infty. \end{aligned} \tag{3.41}$$

Then, Assumption 3.3 is fulfilled:

(i) For $i_1, i_2 \in \{1, \dots, m\}$, $i_1 \neq i_2$, $j \in \{m+1, \dots, m+k\}$ it holds with Fubini's Theorem, (3.2) and the independence of Y_{i_1}, Y_{i_2} and Y_j

$$\begin{aligned} \mathrm{E}\left(h(Y_{i_1}, Y_j)h(Y_{i_2}, Y_j)\right) &= \int\int\int h(y_{i_1}, y_j)h(y_{i_2}, y_j)dF(y_{i_1})dF(y_{i_2})dF(y_j) \\ &= \int\int h(y_{i_1}, y_j)\int h(y_{i_2}, y_j)dF(y_{i_2})dF(y_{i_1})dF(y_j) \\ &= \int\int h(y_{i_1}, y_j)(h_2(y_j)+\theta)dF(y_{i_1})dF(y_j) \\ &= \int (h_2(y_j)+\theta)\int h(y_{i_1}, y_j)dF(y_{i_1})dF(y_j) \\ &= \int (h_2(y_j)+\theta)^2 dF(y_j) \\ &= \mathrm{E}((h_2(Y_j)+\theta)^2) = \mathrm{E}(h_2(Y_j)^2) + 2\theta\,\mathrm{E}(h_2(Y_j)) + \theta^2 \\ &= \mathrm{E}(h_2(Y_j)^2) + \theta^2, \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left(h(Y_{i_1}, Y_j)h_2(Y_j)\right) &= \int\int h(y_{i_1}, y_j)h_2(y_j)dF(y_{i_1})dF(y_j) \\ &= \int h_2(y_j)\int h(y_{i_1}, y_j)dF(y_{i_1})dF(y_j) \\ &= \int h_2(y_j)^2 dF(y_j) + \theta\int h_2(y_j)dF(y_j) \\ &= \mathrm{E}(h_2(Y_j)^2). \end{aligned}$$

It follows by the independence of Y_{i_1}, Y_{i_2} and Y_j

$$
\begin{aligned}
&\mathrm{Cov}\left(r(Y_{i_1},Y_j), r(Y_{i_2},Y_j)\right)\\
=&\mathrm{Cov}\left(h(Y_{i_1},Y_j)-h_1(Y_{i_1})-h_2(Y_j)-\theta, h(Y_{i_2},Y_j)-h_1(Y_{i_2})-h_2(Y_j)-\theta\right)\\
=&\mathrm{Cov}\left(h(Y_{i_1},Y_j), h(Y_{i_2},Y_j)\right)-\mathrm{Cov}\left(h(Y_{i_1},Y_j), h_2(Y_j)\right)-\mathrm{Cov}\left(h_2(Y_j), h(Y_{i_2},Y_j)\right)\\
&+\mathrm{Var}\left(h_2(Y_j)\right)\\
=&\mathrm{E}\left(h(Y_{i_1},Y_j)h(Y_{i_2},Y_j)\right)-\theta^2-\mathrm{E}\left(h(Y_{i_1},Y_j)h_2(Y_j)\right)-\mathrm{E}\left(h_2(Y_j)h(Y_{i_2},Y_j)\right)\\
&+\mathrm{E}\left(h_2(Y_j)^2\right)\\
=&\mathrm{E}(h_2(Y_j)^2)+\theta^2-\theta^2-\mathrm{E}(h_2(Y_j)^2)-\mathrm{E}(h_2(Y_j)^2)+\mathrm{E}(h_2(Y_j)^2)=0.
\end{aligned}
$$

For $j_1, j_2 \in \{m+1,\dots,m+k\}$, $j_1 \neq j_2$, $i \in \{1,\dots,m\}$ it follows analogously that $r(Y_i, Y_{j_1})$ and $r(Y_i, Y_{j_2})$ are uncorrelated. In the case of $i_1 \neq i_2 \neq j_1 \neq j_2$ $r(Y_{i_1}, Y_{j_1})$ and $r(Y_{i_2}, Y_{j_2})$ are independent. Hence, the summands of $\sum_{i=1}^m \sum_{j=k_1}^{k_2} r(Y_i, Y_j)$ are uncorrelated and centered for $0 \le k_1 \le k_2$ such that we obtain

$$
\mathrm{E}\left(\left|\sum_{i=1}^m \sum_{j=k_1}^{k_2} r(Y_i,Y_j)\right|^2\right) = \mathrm{Var}\left(\sum_{i=1}^m \sum_{j=k_1}^{k_2} r(Y_i,Y_j)\right) = \sigma_r^2 m(k_2-k_1+1)
$$

with $\sigma_r^2 = \mathrm{Var}\left(r(Y_1,Y_2)\right) < \infty$ due to (3.40) and (3.41). Now, (i) follows as

$$
\frac{\log_2(m)^2}{m^{1-2\gamma}} \to 0 \quad \text{as } m \to \infty \tag{3.42}
$$

because $\frac{\log_2(x)^2}{x} \to 0$ for all $a > 0$.

(ii) Follows with the 2-dimensional version of Donsker's Theorem (see Theorem 1.1. in Einmahl (2009)).

(iii),(iv) With Theorem C.2 we obtain for every $\epsilon > 0$

$$
\begin{aligned}
&P\left(\sup_{1\le k\le m} \frac{1}{m^{\frac{1}{2}-\alpha}k^\alpha}\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| > C\right) \le \frac{\sigma_2^2}{C^2} m^{2\alpha-1}\sum_{k=1}^m \frac{1}{k^{2\alpha}}\\
\le&\frac{\sigma_2^2}{C^2} m^{2\alpha-1}\int_0^m \frac{1}{x^{2\alpha}}dx = \frac{\sigma_2^2}{C^2}\frac{1}{1-2\alpha} = \epsilon \quad \text{for all } m \ge 1
\end{aligned}
$$

with $C = C(\epsilon) = \frac{\sigma_2}{\sqrt{\epsilon(1-2\alpha)}}$ and $\sigma_2^2 = \mathrm{Var}(h_2(Y_1))$. Furthermore, it holds

$$
\begin{aligned}
&P\left(\sup_{k>k_m} \frac{\sqrt{k_m}}{k}\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| > C\right) \le \frac{\sigma_2^2}{C^2}\left(1+k_m\sum_{k=k_m+1}^{\infty}\frac{1}{k^2}\right)\\
\le&\frac{\sigma_2^2}{C^2}\left(1+k_m\int_{k_m}^{\infty}\frac{1}{x^2}dx\right) = 2\frac{\sigma_2^2}{C^2} = \epsilon
\end{aligned}
$$

with $C = C(\epsilon) = \sqrt{2}\frac{\sigma_2}{\sqrt{\epsilon}}$.

3.2.2. Functionals of Mixing Processes

In the following we consider weakly dependent observations in the form of functionals of absolutely regular processes. The work of Borovkova *et al.* (2001) provides useful limit theorems for functionals of mixing processes, in particular related to U-statistics. Absolute regularity has been introduced by Volkonskii & Rozanov (1959).

Definition 3.8. *Let $\mathcal{A}_{i_1}^{i_2} = \sigma(Z_{i_1}, Z_{i_1+1} \ldots, Z_{i_2})$. A stochastic process $\{Z_i : i \in \mathbb{Z}\}$ is called absolutely regular if*

$$\begin{aligned}\beta(k) &= \sup_{i\geq 1}\left\{\mathrm{E}\left(\sup_{A\in\mathcal{A}_{i+k}^{\infty}} |P(A|\mathcal{A}_{-\infty}^{i}) - P(A)|\right)\right\}\\ &= \frac{1}{2}\sup_{i\geq 1}\left\{\sup\sum_{j=1}^{J}\sum_{l=1}^{L}|P(A_j\cap B_l) - P(A_j)P(B_l)|\right\} \to 0 \quad (k\to\infty)\end{aligned}$$

where the inner supremum in the second representation is taken over all finite $\mathcal{A}_{-\infty}^{i}-$measurable partitions $(A_1, \ldots, A_J)$ and all finite $\mathcal{A}_{i+k}^{\infty}-$ measurable partitions $(B_1, \ldots, B_L)$, J, L arbitrary. The equality between the two representations has been shown in Volkonskii & Rozanov (1959) based on the product measure on $\mathcal{A}_{-\infty}^{i} \otimes \mathcal{A}_{i+k}^{\infty}$.

Other well-known concepts of weak dependence are related to the mixing coefficients

$$\begin{aligned}\alpha(k) &= \sup_{i}\sup_{A\in\mathcal{A}_{-\infty}^{i}}\sup_{B\in\mathcal{A}_{i+k}^{\infty}}|P(A\cap B) - P(A)P(B)|,\\ \varphi(k) &= \sup_{i}\sup_{A\in\mathcal{A}_{-\infty}^{i}}\sup_{B\in\mathcal{A}_{i+k}^{\infty}}|P(B|A) - P(B)|,\\ \Psi(k) &= \sup_{i}\sup_{A\in\mathcal{A}_{-\infty}^{i}}\sup_{B\in\mathcal{A}_{i+k}^{\infty}}\left|\frac{P(A\cap B)}{P(A)P(B)} - 1\right|.\end{aligned}$$

For an extensive introduction and overview on mixing conditions we refer to Bradely (2007). It holds (see Proposition 3.11 in Bradely (2007)) $\frac{1}{2}\Psi(k) \geq \varphi(k) \geq \beta(k) \geq 2\alpha(k)$. Hence, absolute regularity implies the strong ($\alpha-$)mixing condition. However, there are popular processes that do not satisfy a mixing condition. A prominent example given in Andrews (1984) is an $AR(1)$ time series where the innovations follow a discrete distribution. Many of such processes which are not mixing but fulfill the classical central limit theorems can be represented as functionals of mixing processes. This concept has already been studied in Billingsley (1999).

Definition 3.9. *(Borovkova* et al. *, 2001, Definition 1.1. and 1.4.) Let $\{Z_i : i \in \mathbb{Z}\}$ be a stationary stochastic process.*

(i) A sequence $\{Y_i : i \geq 1\}$ is called a (two-sided) functional of $\{Z_i : i \in \mathbb{Z}\}$ if there exists a measurable function $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ such that

$$Y_i = f((Z_{i+n})_{n\in\mathbb{Z}}).$$

In particular, $\{Y_i : i \geq 1\}$ is also stationary.

(ii) *A sequence $\{Y_i : i \geq 1\}$ is called an r-approximating functional with approximating constants $\{a_k\}_{k \geq 0}$ of $\{Z_i : i \in \mathbb{Z}\}$ if*

$$\mathrm{E}\,|Y_0 - \mathrm{E}(Y_0|Z_{-k}, \ldots, Z_k)|^r \leq a_k$$

with $a_k \to 0$ as $k \to \infty$.

We assume that the observed time series is a 1-approximating functional of an absolutely regular process. The 1-approximating property is also called L_1-near-epoch-dependence. Additionally, we need to impose a continuity assumption on the kernel such that the 1-approximating property is preserved when applying the kernel function or the functions h_1 and h_2 of Hoeffdings decomposition.

Definition 3.10. *Let $\{Y_i : i \geq 1\}$ be a stationary stochastic process.*

(i) *A measurable function $u : \mathbb{R} \to \mathbb{R}$ is called 1-continuous, if there exists a function $\Phi : (0, \infty) \to (0, \infty)$ with $\Phi(\epsilon) \to 0$ as $\epsilon \to 0$ such that for all $\epsilon > 0$*

$$\mathrm{E}\left(|u(Y') - u(Y)| 1_{\{|Y - Y'| < \epsilon\}}\right) \leq \Phi(\epsilon)$$

for all random variables Y and Y' having the same distribution as Y_1 (see Definition 2.10. in Borovkova et al. *(2001)).*

(ii) *A kernel $h : \mathbb{R}^2 \to \mathbb{R}$ is called 1-continuous, if there exists a function $\Phi : (0, \infty) \to (0, \infty)$ with $\Phi(\epsilon) \to 0$ as $\epsilon \to 0$ such that for all $\epsilon > 0$*

$$\mathrm{E}\left(|h(X', Y) - h(X, Y)| 1_{\{|X - X'| < \epsilon\}}\right) \leq \Phi(\epsilon)$$
$$\mathrm{E}\left(|h(X, Y') - h(X, Y)| 1_{\{|Y - Y'| < \epsilon\}}\right) \leq \Phi(\epsilon)$$

for all random variables X, X', Y and Y' having the same marginal distribution as Y_1, and such that X, Y are either independent or have joint distribution $P_{(Y_1, Y_k)}$ for some integer k (see Definition 2 in Dehling et al. *(2015b)).*

The 1-continuity of the Wilcoxon-kernel $h_W(x, y) = 1_{\{x<y\}}$ follows similarly to Example 2.2 in Borovkova *et al.* (2001) by first noting that

$$|1_{\{X'<Y\}} - 1_{\{X<Y\}}| 1_{\{|X - X'| < \epsilon\}} \leq 1_{\{|X - Y| < \epsilon\}}$$

with X, X', Y and Y' as in Definition 3.10 (ii). Hence, we obtain

$$\mathrm{E}\left(|1_{\{X'<Y\}} - 1_{\{X<Y\}}| 1_{\{|X - X'| < \epsilon\}}\right) \leq P(|X - Y| < \epsilon) \leq \Phi(\epsilon)$$

with $\Phi(\epsilon) = \max\left(\sup_{k \geq 1} P(|Y_1 - Y_k| < \epsilon), P(|X - Y| < \epsilon)\right)$. Let F_k be the distribution function of $Y_1 - Y_k$ and assume that the functions $F_k, k \geq 1$, are equicontinuous in zero. Then, it holds

$$\sup_{k \geq 1} P(|Y_1 - Y_k| < \epsilon) = \sup_{k \geq 1} (F_k(\epsilon) - F_k(-\epsilon)) \to 0 \quad (\epsilon \to 0).$$

Furthermore, we get with Lebesgue's dominated convergence theorem

$$\lim_{\epsilon \to 0} P(|X - Y| < \epsilon) = \int_{-\infty}^{\infty} f_{Y_1}(y) \lim_{\epsilon \to 0} (F_{Y_1}(y + \epsilon) - F_{Y_1}(y - \epsilon)) dy = 0$$

if F_{Y_1} is continuous in zero. Consequently, we obtain $\Phi(\epsilon) \to 0$ as $\epsilon \to 0$.

Let $\{Y_i : i \geq 1\}$ be a 1-approximating functional with approximating constants $\{a_k\}_{k\geq 0}$ of an absolutely regular process with mixing coefficients $\{\beta(k)\}_{k\geq 0}$ and let $h(x, y)$ be a bounded 1-continuous kernel. In the following we show that Assumption 3.3 is fulfilled if

$$\sum_{k\geq 1} k^2 \left(\beta(k) + \sqrt{a_k} + \Phi(\sqrt{a_k})\right) < \infty. \tag{3.43}$$

(i) Analogously to Lemma 1 in Dehling *et al.* (2015b) it can be shown that there exists a constant C such that

$$\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r(Y_i, Y_j)\right|^2\right) \leq Cm(k_2 - k_1 + 1).$$

The assumption now follows with (3.42).

(ii) Lemma D.3 yields that $h_1(\cdot)$ and $h_2(\cdot)$ are also 1-continuous functions. Furthermore, they are bounded and centered such that with (3.43) the functional central limit theorem is obtained by Proposition D.4.

(iii),(iv) With Lemma D.2,$\{h_2(Y_i) : i \in \mathbb{Z}\}$ is also a 1-approximating functional of an absolutely regular process with approximating constants

$$a'_k = \Phi(\sqrt{2a_k}) + C\sqrt{2a_k}.$$

Hence, by (3.43) the assumption a) of Lemma D.1 is fulfilled for $\{h_2(Y_i) : i \in \mathbb{Z}\}$. Consequently, it follows with Theorem B.3. in Kirch (2006) and the stationarity that there exists a constant $A \geq 4$ such that

$$\mathrm{E}\left(\sup_{1\leq k\leq m} \frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right|^4\right) = \mathrm{E}\left(\sup_{1\leq k\leq m} \frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=1}^{k} h_2(Y_j)\right|^4\right)$$
$$\leq CA\frac{1}{m^{2-4\alpha}}\sum_{k=1}^{m} k^{1-4\alpha} \leq CA\frac{1}{m^{2-4\alpha}}\int_0^m x^{1-4\alpha}dx = \frac{1}{2-4\alpha}CA$$

and with

$$b_k = \begin{cases} \frac{1}{\sqrt{k_m}}, & k = 1, \dots, k_m \\ \frac{\sqrt{k_m}}{k}, & k > k_m \end{cases}$$

we get

$$\mathrm{E}\left(\sup_{k>k_m} \frac{\sqrt{k_m}}{k}\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right|^4\right) = \mathrm{E}\left(\sup_{k>k_m} \frac{\sqrt{k_m}}{k}\left|\sum_{j=1}^{k} h_2(Y_j)\right|^4\right)$$
$$\leq \mathrm{E}\left(\sup_{k\geq 1} b_k\left|\sum_{j=1}^{k} h_2(Y_j)\right|^4\right) \leq CA\sum_{k\geq 1} b_k^4 k = CA\left(\sum_{k=1}^{k_m}\frac{k}{k_m^2} + \sum_{k=k_m+1}^{\infty}\frac{k_m^2}{k^3}\right)$$
$$\leq CA\left(1 + k_m^2\int_{k_m}^{\infty} x^{-3}dx\right) = \frac{3}{2}CA.$$

The CUSUM-kernel $h_C(x, y) = x - y$ is not bounded but nevertheless Assumption 3.3 is fulfilled for many weak dependency concepts. First note that (i) can be omitted as the remainder term is zero. Furthermore, it holds that $h_1(Y_i) = Y_i - \mu = -h_2(x)$ such that in (3.17) in the proof of Theorem 3.6 we only need a functional central limit theorem for the partial sum process $\frac{1}{\sqrt{m}} \sum_{i=1}^{[mt]} (Y_i - \mu)$. Such results have been shown for a wide range of weakly dependent data, as for example in Billingsley (1999) for functionals of mixing processes. Furthermore, weak invariance principles do not only imply the functional central limit theorem but also the Hájek-Rényi-type inequalities in (iii) and (iv). The latter is obtained by first applying the weak invariance principle and then the standard Hájek-Rényi inequality in C.2 to the independent increments of the Wiener process. Then, (iii) and (iv) follow as in Section 3.2.1. Examples for processes that satisfy such weak invariance principles are given in Aue (2003).

3.3. Asymptotics under the Alternative

In the following we analyze the asymptotic behavior of the proposed procedures under alternatives according to the change point model described in Section 2.2. More precisely, we will show that they have asymptotic power one which means that if there is a change it will be detected at some point with probability tending to one as the length m of the historic data set goes to infinity. First, note that it holds for $k > k^*$

$$\begin{aligned}\Gamma(m,k) =& \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*}(h(Y_i,Y_j)-\theta) + \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k}(h(Y_i,Z_{j,m})-\theta)\\ =&\Gamma(m,k^*) + \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k}(h(Y_i,Z_{j,m})-\theta_m^*) + (k-k^*)\Delta_m,\end{aligned}$$

where $\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k} h(Y_i,Z_{j,m})$ is a two-sample U-statistic for which the two samples are no longer generated by the same distribution. Consider the following version of Hoeffding's decomposition:

$$h(x_1,x_2) = \theta_m^* + h_{1,m}^*(x_1) + h_{2,m}^*(x_2) + r_m^*(x_1,x_2) \tag{3.44}$$

with

$$\begin{aligned}\theta_m^* &= \mathrm{E}(h(Y,Z_m))\\ h_{1,m}^*(x_1) &= \mathrm{E}(h(x_1,Z_m)-\theta_m^*)\\ h_{2,m}^*(x_2) &= \mathrm{E}(h(Y,x_2)-\theta_m^*) = h_2(x_2)-\Delta_m,\\ r_m^*(x_1,x_2) &= h(x_1,x_2) - h_{1,m}^*(x_1) - h_{2,m}^*(x_2) - \theta_m^*,\\ \Delta_m &= \theta_m^* - \theta\end{aligned}$$

for independent random variables $Y \overset{\mathcal{D}}{=} Y_1$ and $Z_m \overset{\mathcal{D}}{=} Z_{m,1}$. $r_m^*(x_1,x_2)$ is degenerate and analogously to (3.2) $\mathrm{E}(h_{1,m}^*(Y)) = \mathrm{E}(h_{2,m}^*(Z_m)) = 0$. Based on (3.44) we get the following representation of the monitoring statistic under the alternative for $k > k^*$

$$\begin{aligned}\Gamma(m,k) =& \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*} r(Y_i,Y_j) + \sum_{j=m+1}^{m+k^*} h_2(Y_j) + \frac{k^*}{m}\sum_{i=1}^{m} h_1(Y_i)\\ &+\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k} r_m^*(Y_i,Z_{j,m}) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})\\ &+\frac{k-k^*}{m}\sum_{i=1}^{m} h_{1,m}^*(Y_i) + (k-k^*)\Delta_m.\end{aligned} \tag{3.45}$$

Example 3.11. *We derive the explicit form of Hoeffding's decomposition for the CUSUM and the Wilcoxon kernel (see Example 2.2) for a mean change as given in (2.8). Let* $\tilde{Z}_m = \tilde{Y} + d_m$ *where* $\tilde{Y}$ *is an independent copy of* $Y \overset{\mathcal{D}}{=} Y_1$ *and recall the results of Example 2.3.*

(i) *Hoeffding's decomposition of the CUSUM kernel is given by*

$$h_{1,m}^{*C}(y) = \mathrm{E}\left(h_C(y, \tilde{Z}_m)\right) - \theta_m^{*C} = \mathrm{E}(y - (\tilde{Y} + d_m)) + d_m = y - \mu$$
$$h_{2,m}^{*C}(z) = \mathrm{E}\left(h_C(Y, z)\right) - \theta_m^{*C} = \mathrm{E}(Y - z) + d_m = \mu - z + d_m$$

$$r^{*C}(y, z) = y - z - (y - \mu) - (\mu - z + d_m) + d_m = 0.$$

(ii) *Hoeffding's decomposition of the Wilcoxon kernel is given by*

$$\begin{aligned}
h_{1,m}^{*W}(y) &= \mathrm{E}\left(h_W(y, \tilde{Z}_m)\right) - \theta_m^{*W} = \mathrm{E}(1_{\{y<\tilde{Y}+d_m\}}) - \left(\Delta_m^W + \frac{1}{2}\right) \\
&= \frac{1}{2} - F_Y(y - d_m) - \Delta_m^W \\
h_{2,m}^{*W}(z) &= \mathrm{E}\left(h_W(Y, z)\right) - \theta_m^{*W} = \mathrm{E}(1_{\{Y<z\}}) - \left(\Delta_m^W + \frac{1}{2}\right) \\
&= F_Y(z) - \Delta_m^W + \frac{1}{2} \\
r^{*W}(y, z) &= 1_{\{y<z\}} - \left(\frac{1}{2} - F_Y(y - d_m) - \Delta_m^W\right) - \left(F_Y(z) - \Delta_m^W + \frac{1}{2}\right) \\
&\quad - \left(\Delta_m^W + \frac{1}{2}\right) \\
&= 1_{\{y<z\}} + F_Y(y - d_m) - F_Y(z) + \Delta_m^W - \frac{3}{2}
\end{aligned}$$

Under the alternative, we need to impose some additional but still mild conditions on the weight function:

Assumption 3.12. (i) *If* $\frac{k^*}{m} \to \infty$, *assume that* $\liminf_{t\to\infty} t\rho(t) > 0$.

(ii) *If* $\frac{k^*}{m} = O(1)$, *i.e.* $\frac{k^*}{m} < \nu$ *for all* $m \geq 1$ *for some* $\nu > 0$, *assume that there exist* $t_0 > \nu, \epsilon > 0$ *such that* $\rho(t) > 0$ *for all* $t \in (t_0 - \epsilon, t_0 + \epsilon)$.

We assume that the time series before the change fulfills the regularity conditions under the null hypothesis given in Assumption 3.3. Regarding the terms that additionally appear in the above decomposition of the monitoring statistic after the change, we impose the following regularity conditions:

Assumption 3.13. *Let* $\{Y_i\}_{i\in\mathbb{Z}}$ *and* $\{Z_{i,m}\}_{i\in\mathbb{Z}}$ *be stationary time series that fulfill the following assumptions for a given kernel function* h.

(i) $\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r_m^*(Y_i, Z_{j,m})\right|^2\right) \leq u(m)(k_2 - k_1 + 1)$ *for all* $m + k^* + 1 \leq k_1 \leq k_2$ *with* $\frac{u(m)}{m^{2-2\gamma}} \log(m)^2 \to 0$ *and* γ *as in Assumption 3.2.*

(ii) $\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_{1,m}^*(Y_i) = O_p(1)$ *as* $m \to \infty$

(iii) $\frac{1}{\sqrt{k_m}} \sum_{j=m+k^*+1}^{m+k^*+k_m} h_{2,m}^*(Z_{j,m}) = O_p(1)$ *as* $k_m \to \infty$.

The following consistency result holds without any restriction on the time of the change.

Theorem 3.14. *Let the regularity conditions given in Assumptions 3.2, 3.3, 3.12 and 3.13 be fulfilled. Furthermore assume that $\sqrt{m}|\Delta_m| \to \infty$. Then, it holds under the alternative*

$$\sup_{k\geq 1} w(m,k)\,|\Gamma(m,k)| \xrightarrow{P} \infty.$$

Proof. For $\tilde{k} > k^*$ it holds with (3.45)

$$\begin{aligned}\Gamma\left(m,\tilde{k}\right) =& \Gamma(m,k^*) + \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+\tilde{k}} r_m^*(Y_i,Z_{j,m}) \\ &+ \sum_{j=m+k^*+1}^{m+\tilde{k}} h_{2,m}^*(Z_{j,m}) + \frac{\tilde{k}-k^*}{m}\sum_{i=1}^{m} h_1^*(Y_i) + (\tilde{k}-k^*)\Delta_m. \end{aligned} \tag{3.46}$$

Let us first consider late changes with $\frac{k^*}{m} \to \infty$. As $\rho(t) = (1+t)^{-1}$ fulfills Assumption 3.2, we obtain with Theorem 3.6

$$\frac{\sqrt{m}}{k^*}\,|\Gamma(m,k^*)| = \left(\frac{m}{k^*}+1\right) O_P(1) = O_P(1) \quad \text{as } m\to\infty. \tag{3.47}$$

Furthermore, Lemma 3.5 yields

$$\begin{aligned}&\frac{\sqrt{m}}{k^*}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+\tilde{k}} r_m^*(Y_i,Z_{j,m})\right| \\ =& \left(\frac{m}{k^*}+1\right) o_P(1) = o_P(1) \quad \text{as } m\to\infty. \end{aligned} \tag{3.48}$$

Now, consider $\tilde{k} = 2k^*$. With Assumption 3.13 (ii), (iii) we get

$$\begin{aligned}&\frac{\sqrt{m}}{k^*}\left|\sum_{j=m+k^*+1}^{m+\tilde{k}} h_{2,m}^*(Z_{j,m}) + \frac{\tilde{k}-k^*}{m}\sum_{i=1}^{m} h_1^*(Y_i)\right| \\ \leq& \frac{\sqrt{m}}{k^*}\left(\sqrt{\tilde{k}}\left|\frac{1}{\sqrt{\tilde{k}}}\sum_{j=m+k^*+1}^{m+\tilde{k}} h_{2,m}^*(Z_{j,m})\right| + \frac{\tilde{k}-k^*}{\sqrt{m}}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1^*(Y_i)\right|\right) \\ =& \frac{\sqrt{m}}{k^*}\left(\sqrt{2k^*} + \frac{k^*}{\sqrt{m}}\right) O_P(1) \\ =& \left(\sqrt{\frac{2m}{k^*}}+1\right) O_P(1) = O_P(1) \quad \text{as } m\to\infty. \end{aligned} \tag{3.49}$$

As $\frac{l_m}{k^*} = \frac{m}{k^*}\frac{l_m}{m} \to 0$ as $m\to\infty$ there exists an $m_0 \in \mathbb{N}$ such that $\tilde{k} > k^* > l_m$ for all $m \geq m_0$. Thus, for those $m \geq m_0$, based on the representation of $\Gamma(m,\tilde{k})$ as given in (3.46) it follows with (3.48) and (3.49)

$$w(m,\tilde{k})|\Gamma(m,\tilde{k}) - \Gamma(m,k^*) - (\tilde{k}-k^*)\Delta_m| = O_P(1) \tag{3.50}$$

as

$$\frac{k^*}{\sqrt{m}} w(m, 2k^*) = \frac{1}{2}\left(2\frac{k^*}{m}\rho\left(2\frac{k^*}{m}\right)\right) = O(1)$$

with Assumption 3.2 (iii). Hence, (3.47) yields

$$w(m, \tilde{k})\Gamma\left(m, \tilde{k}\right) = w(m, \tilde{k})(\tilde{k} - k^*)\Delta_m + O_P(1). \tag{3.51}$$

Furthermore, it holds

$$\begin{aligned}
&w\left(m, \tilde{k}\right)(\tilde{k} - k^*)\left|\Delta_m\right| \\
=&\frac{k^*}{\sqrt{m}}\rho\left(2\frac{k^*}{m}\right)\left|\Delta_m\right| = \frac{1}{2}\left(2\frac{k^*}{m}\rho\left(2\frac{k^*}{m}\right)\right)\sqrt{m}\left|\Delta_m\right| \to \infty \quad \text{as } m \to \infty
\end{aligned} \tag{3.52}$$

with Assumption 4.4 (i) and $\sqrt{m}|\Delta_m| \to \infty$. Now, it follows

$$\sup_{k \geq 1} w(m, k)\left|\Gamma(m, k)\right| \geq w\left(m, \tilde{k}\right)\left|\Gamma\left(m, \tilde{k}\right)\right| \xrightarrow{P} \infty \quad \text{as } m \to \infty. \tag{3.53}$$

For $\frac{k^*}{m} = O(1)$, let us consider a weight function $\tilde{w}$ according to Assumption 3.2 with $\tilde{l}_m = 0$ and $\rho \equiv 1$ on $[0, \nu]$. Then, it holds with Theorem 3.6 for $\tilde{k} = [t_0 m]$ with t_0 as in Assumption 3.12

$$\begin{aligned}
w\left(m, \tilde{k}\right)\left|\Gamma(m, k^*)\right| &\leq \rho\left(\frac{\tilde{k}}{m}\right)\left|\Gamma(m, k^*)\right| = \rho\left(\frac{[t_0 m]}{m}\right) O_P(1) \\
&= \rho\left(t_0 + o(1)\right) O_P(1) = O_P(1) \quad \text{as } m \to \infty
\end{aligned} \tag{3.54}$$

as ρ is bounded on any compact interval due to its continuity. Lemma 3.5 yields

$$\begin{aligned}
&w\left(m, \tilde{k}\right)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+\tilde{k}} r_m^*(Y_i, Z_{j,m})\right| = w\left(m, \tilde{k}\right)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{\tilde{k}-k^*} r_m^*(Y_i, Z_{m+k^*+j,m})\right| \\
\leq&\rho\left(\frac{[t_0 m]}{m}\right)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[t_0 m]-k^*} r_m^*(Y_i, Z_{m+k^*+j,m})\right| \\
=&\rho\left(t_0 + o(1)\right) o_P(1) = o_P(1) \quad \text{as } m \to \infty.
\end{aligned} \tag{3.55}$$

as ρ is bounded around t_0. With Assumption 3.13 (ii), (iii) we get

$$\begin{aligned}
&w\left(m, \tilde{k}\right)\left|\sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m}) + \frac{\tilde{k} - k^*}{m}\sum_{i=1}^{m} h_1^*(Y_i)\right| \\
\leq&w\left(m, \tilde{k}\right)\left(\sqrt{\tilde{k}}\left|\frac{1}{\sqrt{\tilde{k}}}\sum_{j=m+k^*+1}^{m+\tilde{k}} h_{2,m}^*(Z_{j,m})\right| + \frac{\tilde{k} - k^*}{\sqrt{m}}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1^*(Y_i)\right|\right) \\
=&w\left(m, \tilde{k}\right)\left(\sqrt{\tilde{k}} + \frac{\tilde{k} - k^*}{\sqrt{m}}\right) O_P(1) \\
\leq&\left(\sqrt{t_0} + t_0\right)\rho\left(t_0 + o(1)\right) O_P(1) = O_P(1) \quad \text{as } m \to \infty.
\end{aligned} \tag{3.56}$$

Hence, for the representation of $\Gamma(m,\tilde{k})$ as given in (3.46) it follows with (3.55) and (3.56)

$$w(m,\tilde{k})\left|\Gamma\left(m,\tilde{k}\right)-\Gamma(m,k^*)-(\tilde{k}-k^*)\Delta_m\right|=O_P(1) \tag{3.57}$$

such that we obtain with (3.54)

$$w(m,\tilde{k})\Gamma\left(m,\tilde{k}\right)=w(m,\tilde{k})(\tilde{k}-k^*)|\Delta_m|+O_P(1). \tag{3.58}$$

As $\frac{l_m}{m}\to 0$ for $m\to\infty$ there exists an $m_0\in\mathbb{N}$ such that $m>l_m$ for all $m\geq m_0$. For those $m\geq m_0$ it holds

$$\begin{aligned}
&w\left(m,\tilde{k}\right)(\tilde{k}-k^*)|\Delta_m|=\rho\left(\frac{[mt_0]}{m}\right)([mt_0]-k^*)|\Delta_m|\\
\geq&\rho\left(t_0+o(1)\right)\left(\frac{[mt_0]}{m}-\frac{[m\nu]}{m}\right)\sqrt{m}|\Delta_m|\\
\geq&\rho\left(t_0+o(1)\right)\left(t_0-\nu+o(1)\right)\sqrt{m}|\Delta_m|\xrightarrow{P}\infty\quad\text{as } m\to\infty
\end{aligned} \tag{3.59}$$

as $\sqrt{m}|\Delta_m|\to\infty$ with Assumption 3.12 (ii). Now, it follows with (3.58) and (3.59)

$$\sup_{k\geq 1}w(m,k)\left|\Gamma(m,k)\right|\geq w\left(m,\tilde{k}\right)\left|\Gamma\left(m,\tilde{k}\right)\right|\xrightarrow{P}\infty\quad\text{as } m\to\infty. \tag{3.60}$$

□

4. Modified MOSUM and Page-CUSUM

In the previous chapters we have considered the monitoring statistic as given in (2.1) which is a generalization of the classical sequential CUSUM procedure. However, as it compares the historic data set with all observations that have been collected during the monitoring period so far, it shares the drawback of the classical CUSUM monitoring scheme that it takes a rather long time to detect changes that occur rather late in the monitoring period. The reason behind that is quite obvious: the later the change occurs the more data that follows the distribution under the null hypothesis is used in the monitoring statistic for the CUSUM scheme such that correspondingly more observations need to be collected after the change before the monitoring statistic indicates a significant difference to the historic data set. Therefore, several adaptations of the monitoring scheme have been proposed in the literature. They all aim at being less dependent on the time of the change by only taking the most recent observations into account. The selection of the time window for which the observations are included differs between the procedures. The standard MOSUM procedure has been considered in Horváth *et al.* (2008) and Aue *et al.* (2012) for the mean change model. The modified MOSUM has been proposed in Chen & Tian (2010) and the Page-CUSUM in Fremdt (2015), both for the linear model. All of those monitoring schemes have been considered in Kirch & Weber (2018) in the framework of sequential change point tests based on estimating functions. Based on a simulation study they conclude that the MOSUM procedure has noticeable problems with the detection of changes. Although its detection delay is very small for those changes that are detected, we are interested in a reliable testing procedure in the first place and thus only consider the modified MOSUM and the Page-CUSUM. According to those monitoring schemes we obtain the following monitoring statistics in our framework:

Modified MOSUM:

$$\begin{aligned}\Gamma_2(m,k) &= \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+[kh]+1}^{m+k}(h(X_i,X_j)-\theta)\\ &= \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+[kh]+1}^{m+k} r(X_i,X_j) + \sum_{j=m+[kh]+1}^{m+k} h_2(X_j) + \frac{k-[kh]}{m}\sum_{i=1}^{m} h_1(X_i),\end{aligned}$$

where $h \in (0,1)$ is a given tuning parameter.

Page-CUSUM:

$$\begin{aligned}
\Gamma_3(m,k) &= \sup_{1\le l\le k} |\Gamma(m,k)-\Gamma(m,l)| \\
&= \sup_{1\le l\le k} \left| \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+l+1}^{m+k} (h(X_i,X_j)-\theta)\right| \\
&= \sup_{1\le l\le k} \left| \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+l+1}^{m+k} r(X_i,X_j) + \sum_{j=m+l+1}^{m+k} h_2(X_j) + \frac{k-l}{m}\sum_{i=1}^{m} h_1(X_i)\right|.
\end{aligned}$$

The fixed parameter h in the modified MOSUM determines the percentage of the earlier observations that are discarded, whereas the Page-CUSUM does not require an a priori choice of a parameter.

4.1. Asymptotics Under the Null Hypothesis

In order to obtain the critical values for the modified MOSUM and the Page-CUSUM procedure we derive the limit distribution of the respective test statistic under the null hypothesis. All asymptotic results can be established based on the regularity conditions introduced in Chapter 3. As before we start with showing that the remainder term is uniformly asymptotically negligible.

Lemma 4.1. *Let $\{Y_i\}_{i\in\mathbb{Z}}$ and $\{Y'_{i,m}\}_{i\in\mathbb{Z}}$ be sequences of random variables. Let Assumption 3.2 be fulfilled for the weight function. Assume that for $g_m : \mathbb{R}^2 \to \mathbb{R}$ it holds*

$$\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} g_m(Y_i,Y'_{j,m})\right|^2\right) \le u(m)(k_2-k_1+1) \quad \textit{for all } 0\le k_1\le k_2 \tag{4.1}$$

with $\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \to 0$ for all $\delta>0$ and γ as in Assumption 3.2. Then, it holds as $m\to\infty$

(i) $\sup_{k\ge 1} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[kh]+1}^{k} g_m(Y_i,Y'_{j,m})\right| = o_P(1)$.

(ii) $\sup_{k\ge 1} w(m,k)\sup_{1\le l\le k}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=l+1}^{k} g_m(Y_i,Y'_{j,m})\right| = o_P(1)$.

Proof. (i) By (3.9) and (3.10) we get

$$\begin{aligned}
&\max_{1\le k\le m} k^\gamma w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i,Y'_{j,m})\right| \\
\le &\max_{1\le k\le m} k^\gamma w(m,k) \max_{1\le [kh]\le [mh]}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i,Y'_{j,m})\right| \\
\le &\max_{1\le k\le m} k^\gamma w(m,k) \max_{1\le l\le m}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{l} g_m(Y_i,Y'_{j,m})\right| = o_P(1).
\end{aligned}$$

In particular, this implies

$$\max_{1\le k\le m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i, Y'_{j,m})\right| = o_P(1).$$

Combining this with (3.12) we obtain

$$\begin{aligned}
&\max_{1\le k\le m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[kh]+1}^{k} g_m(Y_i, Y'_{j,m})\right| \\
\le &\max_{1\le k\le m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i, Y'_{j,m})\right| + \max_{1\le k\le m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i, Y'_{j,m})\right| \\
=&o_P(1) \quad \text{as } m\to\infty.
\end{aligned} \tag{4.2}$$

With (3.14) and (3.9) it holds

$$\begin{aligned}
&\sup_{k>m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i, Y'_{j,m})\right| \\
=&\sup_{k>\max(m,l_m)} \frac{1}{\sqrt{m}}\rho\left(\frac{k}{m}\right)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i, Y'_{j,m})\right| \\
\le &\sup_{k>m} \frac{[kh]}{k}\frac{k}{m}\rho\left(\frac{k}{m}\right)\left|\frac{1}{[kh]\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i, Y'_{j,m})\right| \\
\le &h\sup_{t>1} t\rho\left(t\right)\sup_{k>m}\left|\frac{1}{[kh]\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i, Y'_{j,m})\right| \\
\le &h\sup_{t>1} t\rho\left(t\right)\left(\sup_{[mh]\le l\le m}\left|\frac{1}{l\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{l} g_m(Y_i, Y'_{j,m})\right| + \sup_{l>m}\left|\frac{1}{l\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{l} g_m(Y_i, Y'_{j,m})\right|\right) \\
\le &h\sup_{t>1} t\rho\left(t\right)\left(\sup_{1\le l\le m}\left|\frac{1}{[mh]\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{l} g_m(Y_i, Y'_{j,m})\right| + o_P(1)\right) \\
=&h\sup_{t>1} t\rho\left(t\right)\left(\sqrt{\frac{u(m)}{m^{2-2\gamma}}log(m)^2 m^{-\gamma}}\frac{m}{[mh]}o_P(1) + o_P(1)\right) = o_P(1) \quad \text{as } m\to\infty
\end{aligned}$$

as by Assumption 3.2 (iii) $t\rho(t)$ is bounded on $(1,\infty)$.

Hence, we get with (3.16)

$$\begin{aligned}
&\sup_{k>m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[kh]+1}^{k} g_m(Y_i,Y'_{j,m})\right| \\
\leq &\sup_{k>m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} rg_m(Y_i,Y'_{j,m})\right| + \sup_{k>m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{[kh]} g_m(Y_i,Y'_{j,m})\right| \\
= &o_P(1) \quad \text{as } m\to\infty.
\end{aligned}$$

Together with (4.2) it follows

$$\begin{aligned}
&\sup_{k\geq 1} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[kh]+1}^{k} g_m(Y_i,Y'_{j,m})\right| \\
\leq &\max_{1\leq k\leq m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[kh]+1}^{k} r(Y_i,Y_j)\right| + \sup_{k>m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[kh]+1}^{k} g_m(Y_i,Y'_{j,m})\right| \\
= &o_P(1) \quad \text{as } m\to\infty
\end{aligned}$$

and thus assertion (i).

(ii) With (3.9) and (3.10) we get

$$\max_{1\leq k\leq m} k^{\gamma} w(m,k)\max_{1\leq l\leq k}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{l} g_m(Y_i,Y'_{j,m})\right| = o_P(1). \tag{4.3}$$

In particular, this implies

$$\max_{1\leq k\leq m} w(m,k)\max_{1\leq l\leq k}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{l} g_m(Y_i,Y'_{j,m})\right| = o_P(1).$$

Combining this with (3.12) we obtain

$$\begin{aligned}
&\max_{1\leq k\leq m} w(m,k)\max_{1\leq l\leq k}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=l+1}^{k} g_m(Y_i,Y'_{j,m})\right| \\
\leq &\max_{1\leq k\leq m} w(m,k)\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} g_m(Y_i,Y'_{j,m})\right| \\
&+\max_{1\leq k\leq m} w(m,k)\max_{1\leq l\leq k}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{l} g_m(Y_i,Y'_{j,m})\right| \\
= &o_P(1) \quad \text{as } m\to\infty.
\end{aligned} \tag{4.4}$$

With (3.9), (3.14) and (3.15) and it holds

$$
\begin{aligned}
&\sup_{k>m} w(m,k) \max_{1\le l\le k} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{l} g_m(Y_i, Y'_{j,m}) \right| \\
\le &\sup_{k>m} w(m,k) \frac{k}{\sqrt{m}} \sup_{k>m} \max_{1\le l\le k} \left| \frac{1}{k\sqrt{m}} \sum_{i=1}^{m} \sum_{j=1}^{l} g_m(Y_i, Y'_{j,m}) \right| \\
\le &\sup_{k>m} w(m,k) \frac{k}{\sqrt{m}} \left(\max_{1\le l\le m} \left| \frac{1}{m^{3/2}} \sum_{i=1}^{m} \sum_{j=1}^{l} g_m(Y_i, Y'_{j,m}) \right| \right. \\
&\left. + \sup_{l>m} \left| \frac{1}{l\sqrt{m}} \sum_{i=1}^{m} \sum_{j=1}^{l} g_m(Y_i, Y'_{j,m}) \right| \right) = o_P(1) \quad \text{as } m \to \infty.
\end{aligned}
$$

Hence, we get with (3.16)

$$
\begin{aligned}
&\sup_{k>m} w(m,k) \max_{1\le l\le k} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=l+1}^{k} g_m(Y_i, Y'_{j,m}) \right| \\
\le &\sup_{k>m} w(m,k) \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} r(Y_i, Y_j) \right| + \sup_{k>m} w(m,k) \max_{1\le l\le k} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{l} g_m(Y_i, Y'_{j,m}) \right| \\
= &o_P(1) \quad \text{as } m \to \infty.
\end{aligned}
$$

Together with (4.4) it follows

$$
\begin{aligned}
&\sup_{k\ge 1} w(m,k) \max_{1\le l\le k} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=l+1}^{k} g_m(Y_i, Y'_{j,m}) \right| \\
\le &\max_{1\le k\le m} w(m,k) \max_{1\le l\le k} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=l+1}^{k} g_m(Y_i, Y'_{j,m}) \right| \\
&+ \sup_{k>m} w(m,k) \max_{1\le l\le k} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=l+1}^{k} g_m(Y_i, Y'_{j,m}) \right| \\
= &o_P(1) \quad \text{as } m \to \infty
\end{aligned}
$$

and thus assertion (ii).

□

The following theorem states the limit distribution of the test statistic for the Page-CUSUM as well as the modified MOSUM.

Theorem 4.2. *Let the regularity conditions given in Assumption 3.2 and 3.3 be fulfilled. Then, there exist two independent standard Wiener processes $\{W_1(t)\}$ and $\{W_2(t)\}$ such that, as $m \to \infty$,*

(i) $\sup_{k\ge 1} w(m,k) \, |\Gamma_2(m,k)| \xrightarrow{\mathcal{D}} \sup_{t>0} \rho(t) \, |\sigma_2(W_2(t) - W_2(th)) + t(1-h)\sigma_1 W_1(1)|$,

(ii) $\sup_{k\geq 1} w(m,k)\,|\Gamma_3(m,k)| \xrightarrow{\mathcal{D}} \sup_{t>0}\rho(t)\sup_{0<s\leq t}|\sigma_2(W_2(t)-W_2(s))+(t-s)\sigma_1 W_1(1)|$,

where σ_1 and σ_2 are as in Assumption 3.3 (ii).

Proof. (i) With Lemma 4.1 it holds

$$\begin{aligned}
&\sup_{k\geq 1}\left| w(m,k)\Gamma_2(m,k) - w(m,k)\left(\sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) + \frac{k-[kh]}{m}\sum_{i=1}^{m} h_1(Y_i)\right)\right| \\
&\leq \sup_{k\geq 1} w(m,k)\frac{1}{m}\left|\sum_{i=1}^{m}\sum_{j=m+[kh]+1}^{m+k} r(Y_i,Y_j)\right| = o_P(1), \quad m\to\infty
\end{aligned}$$

due to Assumption 3.3 (i). According to Lemma B.1 it remains to show that there exist two independent standard Wiener processes $\{W_1(t)\}$ and $\{W_2(t)\}$ such that, as $m\to\infty$,

$$\begin{aligned}
&\sup_{k\geq 1} w(m,k)\left|\sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) + \frac{k-[kh]}{m}\sum_{i=1}^{m} h_1(Y_i)\right| \\
&\xrightarrow{\mathcal{D}} \sup_{t>0}\rho(t)\left|\sigma_2(W_2(t)-W_2(th)) + t(1-h)\sigma_1 W_1(1)\right|.
\end{aligned}$$

With Assumption 3.3 (ii) we get that for any fixed $\tau, T>0$ and $m\to\infty$

$$\begin{aligned}
&\sup_{t>\tau}\rho(t)\left|\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t(1-h)}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right| \\
&\xrightarrow{\mathcal{D}} \sup_{t>\tau}\rho(t)\left|\tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1+\min(t,T)h) + t(1-h)\tilde{W}_1(1)\right|. \qquad (4.5)
\end{aligned}$$

As $\frac{l_m}{m}\to 0$, there exists an m_τ such that $\frac{l_m}{m}<\tau$ for all $m\geq m_\tau$. Hence, it holds for $m\geq m_\tau$

$$\begin{aligned}
&\left|\sup_{t>\tau}\rho(t)\left|\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t(1-h)}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right|\right. \\
&\left. -\sup_{k>\tau m} w(m,k)\left|\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-[kh]}{m}\sum_{i=1}^{m} h_1(Y_i)\right|\right| \\
=&\left|\sup_{\frac{k}{m}>\tau}\ \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}}\rho(t)\left|\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t(1-h)}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right|\right. \\
&\left. -\sup_{\frac{k}{m}>\tau}\rho\left(\frac{k}{m}\right)\left|\frac{1}{\sqrt{m}}\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-[kh]}{m}\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right|\right|
\end{aligned}$$

$$
\begin{aligned}
\leq \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \Bigg| & \rho(t) \left(\frac{1}{\sqrt{m}} \sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t(1-h)}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \\
& -\rho\left(\frac{k}{m}\right) \left(\frac{1}{\sqrt{m}} \sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-[kh]}{m} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \Bigg| \\
\leq \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \Bigg| & \left(t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right) \left(\frac{1}{t}\frac{1}{\sqrt{m}} \sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) \right. \\
& \left. +(1-h)\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \Bigg| \\
& + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \rho\left(\frac{k}{m}\right) \Bigg| \frac{1}{\sqrt{m}} \sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)} h_2(Y_j) \\
& -\frac{k}{mt}\frac{1}{\sqrt{m}} \sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) \Bigg| \\
& + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \left| \frac{k-[kh]}{k}\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) - \frac{1-h}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
\leq \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} & \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| \left(\frac{1}{\tau} \sup_{t>\tau} \left| \frac{1}{\sqrt{m}} \sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) \right| \right. \\
& \left. +(1-h) \left| \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \right) \\
& + \sup_{\frac{k}{m}>\tau} \rho\left(\frac{k}{m}\right) \left(\sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| \frac{1}{\sqrt{m}} \sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)} h_2(Y_j) \right. \right. \\
& \left. -\frac{1}{\sqrt{m}} \sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) \right| \\
& \left. + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| 1 - \frac{k}{mt}\frac{1}{t} \right| \sup_{t>\tau} \left| \frac{1}{\sqrt{m}} \sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]} h_2(Y_j) \right| \right) \\
& + \sup_{\frac{k}{m}>\tau} \frac{k}{m}\rho\left(\frac{k}{m}\right) \sup_{\frac{k}{m}>\tau} \left| h - \frac{[kh]}{k} \right| \left| \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right|
\end{aligned}
$$

$$
\begin{aligned}
=&o_P(1)\left(\frac{1}{\tau}\sup_{t>\tau}\left|\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]}h_2(Y_j)\right|+(1-h)\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m}h_1(Y_i)\right|\right)\\
&+O(1)\sup_{\frac{k}{m}>\tau}\sup_{\frac{k}{m}\le t<\frac{k+1}{m}}\left|\frac{1}{\sqrt{m}}\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)}h_2(Y_j)-\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]}h_2(Y_j)\right|\\
&+O(1)\sup_{\frac{k}{m}>\tau}\sup_{\frac{k}{m}\le t<\frac{k+1}{m}}\left|1-\frac{k}{mt}\right|\sup_{t>\tau}\left|\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]}h_2(Y_j)\right|\Bigg)\\
&+O(1)\sup_{\frac{k}{m}>\tau}\left|h-\frac{[kh]}{k}\right|\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m}h_1(Y_i)\right| && (4.6)\\
& && (4.7)
\end{aligned}
$$

with (3.18), (3.20) and Assumption 3.2 (iii) as $\sup_{\frac{k}{m}>\tau}\frac{k}{m}\rho\left(\frac{k}{m}\right)\le\sup_{t>\tau}t\rho(t)<\infty$. Assumption 3.3 (ii) yields

$$\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]}h_2(Y_j)=O_P(1) \tag{4.8}$$

as well as

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}h_1(Y_i)=O_P(1). \tag{4.9}$$

Furthermore, it holds

$$\sup_{\frac{k}{m}>\tau}\left|h-\frac{[kh]}{k}\right|=\sup_{\frac{k}{m}>\tau}\left|\frac{kh-[kh]}{k}\right|\le\frac{1}{\tau m}=o(1)\quad\text{as } m\to\infty \tag{4.10}$$

and

$$\sup_{\frac{k}{m}>\tau}\sup_{\frac{k}{m}\le t<\frac{k+1}{m}}\left|1-\frac{k}{mt}\right|\le\sup_{\frac{k}{m}>\tau}\frac{1}{k+1}\le\frac{1}{\tau m}=o(1)\quad\text{as } m\to\infty. \tag{4.11}$$

Now, consider

$$
\begin{aligned}
&\sup_{\frac{k}{m}>\tau}\sup_{\frac{k}{m}\le t<\frac{k+1}{m}}\left|\frac{1}{\sqrt{m}}\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)}h_2(Y_j)-\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]}h_2(Y_j)\right|\\
=&\sup_{\tau<\frac{k}{m}\le T}\sup_{\frac{k}{m}\le t<\frac{k+1}{m}}\left|\frac{1}{\sqrt{m}}\sum_{j=1}^{[mth]}h_2(Y_j)-\frac{1}{\sqrt{m}}\sum_{j=1}^{[kh]}h_2(Y_j)\right|\\
\le&\sup_{\tau<t\le T}|X(t)-X(t-)|, && (4.12)
\end{aligned}
$$

where $X(t):=\frac{1}{\sqrt{m}}\sum_{j=1}^{[mth]}h_2(Y_j)$ and $X(t-)=\lim_{\epsilon\to0}X(t-\epsilon)$. Let $F:D[0,T]\to\mathbb{R}$ with $x\mapsto\sup_{t>\tau}|x(t)-x(t-)|$, where $D[0,T]$ denotes the set of càdlàg functions on $[0,T]$. Let $\delta > 0$. Then, it holds for all $x, y \in D[0,T]$ with $\sup_{\tau<t\leq T}|x(t) - y(t)| < \frac{\delta}{2}$

$$\begin{aligned}
&\left|\sup_{\tau<t\leq T}|x(t)-x(t-)| - \sup_{\tau<t\leq T}|y(t)-y(t-)|\right| \\
\leq &\sup_{\tau<t\leq T}|(x(t)-y(t)-(x(t-)-y(t-))| \\
\leq &\sup_{\tau<t\leq T}|(x(t)-y(t)| + \sup_{\tau<t\leq T}|(x(t-)-y(t-))| \\
\leq &2\sup_{\tau<t\leq T}|(x(t)-y(t)| < \delta.
\end{aligned}$$

Hence, the functional F is continuous such that it follows with (4.12) and Assumption 3.2 (iii)

$$\sup_{\frac{k}{m}>\tau}\sup_{\frac{k}{m}\leq t<\frac{k+1}{m}}\left|\frac{m}{k}\frac{1}{\sqrt{m}}\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)}h_2(Y_j) - \frac{1}{t}\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]}h_2(Y_j)\right| = o_P(1). \tag{4.13}$$

Combining (4.6), (4.8), (4.9), (4.10), (4.11) and (4.13) we obtain

$$\begin{aligned}
&\left|\sup_{t>\tau}\rho(t)\left|\frac{1}{\sqrt{m}}\sum_{j=m+[m\min(t,T)h]+1}^{m+[m\min(t,T)]}h_2(Y_j) + \frac{t(1-h)}{\sqrt{m}}\sum_{i=1}^{m}h_1(Y_i)\right|\right. \\
&\left.-\sup_{k>\tau m}w(m,k)\left|\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)}h_2(Y_j) + \frac{k-[kh]}{m}\sum_{i=1}^{m}h_1(Y_i)\right|\right| = o_P(1) \quad \text{as } m\to\infty
\end{aligned}$$

and with (4.5) we get

$$\begin{aligned}
&\sup_{k>\tau m}w(m,k)\left|\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)}h_2(Y_j) + \frac{k-[kh]}{m}\sum_{i=1}^{m}h_1(Y_i)\right| \\
\xrightarrow{\mathcal{D}}&\sup_{t>\tau}\rho(t)\left|\tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1+\min(t,T)h) + t(1-h)\tilde{W}_1(1)\right|.
\end{aligned} \tag{4.14}$$

By Assumption 3.3 (iv) and the stationarity we obtain

$$\begin{aligned}
&\sup_{k>k_m}\frac{\sqrt{k_m}}{k}\left|\sum_{j=m+1}^{m+[kh]}h_2(Y_j)\right| \leq \sup_{k>k_m}\frac{h\sqrt{k_m}}{[kh]}\left|\sum_{j=m+1}^{m+[kh]}h_2(Y_j)\right| \\
\leq&\sqrt{h\frac{k_m h}{[k_m h]}}\sup_{l>[k_m h]}\frac{\sqrt{[k_m h]}}{l}\left|\sum_{j=m+1}^{m+l}h_2(Y_j)\right| \\
\leq&\sqrt{h\frac{1}{1-\frac{1}{k_m h}}}\sup_{l>[k_m h]}\frac{\sqrt{[k_m h]}}{l}\left|\sum_{j=m+1}^{m+l}h_2(Y_j)\right| \\
\stackrel{\mathcal{D}}{=}&\sqrt{h\frac{1}{1-\frac{1}{k_m h}}}\sup_{l>[k_m h]}\frac{\sqrt{[k_m h]}}{l}\left|\sum_{j=1}^{l}h_2(Y_j)\right| = O_P(1) \quad \text{as } k_m\to\infty
\end{aligned}$$

uniformly in m and thus

$$\sup_{k>k_m} \frac{\sqrt{k_m}}{k} \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right|$$
$$\leq \sup_{k>k_m} \frac{\sqrt{k_m}}{k} \left| \sum_{j=m+1}^{m+k} h_2(Y_j) \right| + \sup_{k>k_m} \frac{\sqrt{k_m}}{k} \left| \sum_{j=m+1}^{m+[kh]} h_2(Y_j) \right| = O_P(1) \quad \text{as } k_m \to \infty$$

uniformly in m. This yields

$$\begin{aligned}
&\sup_{k>mT} w(m,k) \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
=& \sup_{k>mT} \frac{k w(m,k)}{\sqrt{mT}} \cdot \frac{\sqrt{mT}}{k} \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
=& \sup_{k>\max(mT,l_m)} \frac{k\rho\left(\frac{k}{m}\right)}{m\sqrt{T}} \cdot \frac{\sqrt{mT}}{k} \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
\leq& \frac{1}{\sqrt{T}} \sup_{k>mT} \frac{k}{m}\rho\left(\frac{k}{m}\right) \sup_{k>mT} \frac{\sqrt{mT}}{k} \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
\leq& \frac{1}{\sqrt{T}} \sup_{k>mT} \frac{k}{m}\rho\left(\frac{k}{m}\right) O_P(1) \\
\leq& \frac{1}{\sqrt{T}} \sup_{t>T} t\rho(t) O_P(1) \\
=& \frac{1}{\sqrt{T}} O_P(1) \xrightarrow{P} 0 \quad \text{as } T \to \infty \text{ uniformly in m}
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{k>mT} w(m,k) \left| \sum_{j=m+[mTh]+1}^{m+mT} h_2(Y_j) \right| \\
\leq& \sup_{k>mT} \frac{k}{m}\rho\left(\frac{k}{m}\right) \left| \frac{1}{T\sqrt{m}} \sum_{j=m+[mTh]+1}^{m+mT} h_2(Y_j) \right| \\
\leq& \sup_{t>T} t\rho(t) O_P(1) \xrightarrow{P} 0 \quad \text{as } T \to \infty \text{ uniformly in m.}
\end{aligned}$$

Hence, we get

$$\begin{aligned}
\sup_{k>\tau m} \Bigg| & w(m,k) \left(\sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-[kh]}{m} \sum_{i=1}^{m} h_1(Y_i) \right) \\
& -w(m,k) \left(\sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) + \frac{k-[kh]}{m} \sum_{i=1}^{m} h_1(Y_i) \right) \Bigg|
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{k>\tau m} \left| w(m,k) \sum_{j=m+[\min(k,mT)h]+1}^{m+\min(k,mT)} h_2(Y_j) - w(m,k) \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&\leq \sup_{\tau m<k\leq mT} \left| w(m,k) \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) - w(m,k) \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&\quad + \sup_{k>mT} \left| w(m,k) \sum_{j=m+[mTh]+1}^{m+mT} h_2(Y_j) - w(m,k) \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&= \sup_{k>mT} \left| w(m,k) \sum_{j=m+[mTh]+1}^{m+mT} h_2(Y_j) - w(m,k) \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&\leq \sup_{k>mT} w(m,k) \left| \sum_{j=m+[mTh]+1}^{m+mT} h_2(Y_j) \right| + \sup_{k>mT} w(m,k) \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&= o_P(1) \quad \text{as } T \to \infty \text{ uniformly in m.}
\end{aligned} \tag{4.15}
$$

For $\gamma < \alpha < \frac{1}{2}$ it holds

$$
\begin{aligned}
&\sup_{1\leq k\leq\tau m} w(m,k) \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) + \frac{k-[kh]}{m} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&\leq \sup_{1\leq k\leq\tau m} w(m,k) \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| + \sup_{1\leq k\leq\tau m} w(m,k) \left| \frac{k-[kh]}{m} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&= \sup_{1\leq k\leq\tau m} m^{\frac{1}{2}-\alpha} k^{\alpha} w(m,k) \left| \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&\quad + \sup_{1\leq k\leq\tau m} \frac{k-[kh]}{m} w(m,k)\sqrt{m} \left| \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&\leq \sup_{l_m<k\leq\tau m} \left(\frac{k}{m}\right)^{\alpha} \rho\left(\frac{k}{m}\right) \left| \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&\quad + \sup_{l_m<k\leq\tau m} \frac{k}{m} \rho\left(\frac{k}{m}\right) \left| \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&\leq \sup_{1\leq k\leq\tau m} \left(\frac{k}{m}\right)^{\alpha} \rho\left(\frac{k}{m}\right) \sup_{1\leq k\leq\tau m} \left| \frac{1}{m^{\frac{1}{2}-\alpha} k^{\alpha}} \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
&\quad + \sup_{1\leq k\leq\tau m} \frac{k}{m} \rho\left(\frac{k}{m}\right) \left| \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right|.
\end{aligned}
$$

With Assumption 3.3 (iii) and the stationarity it holds for all $0 < \alpha < \frac{1}{2}, \tau < 1$

$$
\begin{aligned}
&\sup_{1\le k\le \tau m} \frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
\le & \sup_{1\le k\le m} \frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) \right| \\
\le & \sup_{1\le k\le m} \frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \left| \sum_{j=m+1}^{m+k} h_2(Y_j) \right| + \sup_{1\le k\le m} \frac{h^{\alpha}}{m^{\frac{1}{2}-\alpha}([kh])^{\alpha}} \left| \sum_{j=m+1}^{m+[kh]} h_2(Y_j) \right| \\
\overset{\mathcal{D}}{=} & \sup_{1\le k\le m} \frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}} \left| \sum_{j=1}^{k} h_2(Y_j) \right| + \sup_{1\le k\le m} \frac{h^{\alpha}}{m^{\frac{1}{2}-\alpha}([kh])^{\alpha}} \left| \sum_{j=1}^{[kh]} h_2(Y_j) \right| \\
= & O_P(1) \quad \text{as } m \to \infty.
\end{aligned}
$$

and

$$\left| \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| = O_P(1) \quad \text{as } m \to \infty$$

holds by Assumption 3.3 (ii). With (3.26) and (3.27) we obtain

$$\sup_{1\le k\le \tau m} w(m,k) \left| \sum_{j=m+[kh]+1}^{m+k} h_2(Y_j) + \frac{k-[kh]}{m} \sum_{i=1}^{m} h_1(Y_i) \right| = o_P(1) \quad \text{as } \tau \to 0 \tag{4.16}$$

uniformly in m. With (3.24) and (3.29) we obtain

$$
\begin{aligned}
&\sup_{t>T} \rho(t) \left| \tilde{W}_2(1+th) \right| \\
=& \sup_{t>T} (1+th)\rho(t) \frac{\left| \tilde{W}_2(1+th) \right|}{1+th} \\
\le& \sup_{t>T} (1+th)\rho(t) \sup_{t>T} \frac{\left| \tilde{W}_2(1+th) \right|}{1+th} \\
\le& (\sup_{t>T} \rho(t) + h \sup_{t>T} t\rho(t)) \sup_{s>Th} \frac{\left| \tilde{W}_2(1+s) \right|}{1+s} = o(1) \quad \text{a.s.} \quad \text{as } T \to \infty
\end{aligned}
$$

such that it follows with (3.30)

$$
\begin{aligned}
&\sup_{t>T} \rho(t) \left| \tilde{W}_2(1+t) - \tilde{W}_2(1+th) \right| \\
\le& \sup_{t>T} \rho(t) \left| \tilde{W}_2(1+t) \right| + \sup_{t>T} \rho(t) \left| \tilde{W}_2(1+th) \right| = o(1) \quad \text{a.s.} \quad \text{as } T \to \infty.
\end{aligned}
$$

Furthermore, it holds

$$\begin{aligned}
&\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+Th)\right| \\
\le&\sup_{t>T}(1+th)\rho(t)\frac{\left|\tilde{W}_2(1+Th)\right|}{1+Th} \\
\le&(\sup_{t>T}\rho(t)+h\sup_{t>T}t\rho(t))\frac{\left|\tilde{W}_2(1+Th)\right|}{1+Th}=o(1)\quad\text{a.s.}\quad\text{as } T\to\infty
\end{aligned}$$

and with (3.31) it follows

$$\begin{aligned}
&\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+T)-\tilde{W}_2(1+Th)\right| \\
\le&\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+T)\right|+\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+Th)\right|=o(1)\quad\text{a.s.}\quad\text{as } T\to\infty.
\end{aligned}$$

Consequently, we get

$$\begin{aligned}
&\sup_{t>\tau}\Big|\rho(t)\left(\tilde{W}_2(1+\min(t,T))-\tilde{W}_2(1+\min(t,T)h)+t(1-h)\tilde{W}_1(1)\right) \\
&-\rho(t)\left(\tilde{W}_2(1+t)-\tilde{W}_2(1+th)+t(1-h)\tilde{W}_1(1)\right)\Big| \\
=&\sup_{t>\tau}\rho(t)\left|\tilde{W}_2(1+\min(t,T))-\tilde{W}_2(1+\min(t,T)h)-\left(\tilde{W}_2(1+t)-\tilde{W}_2(1+th)\right)\right| \\
\le&\sup_{\tau<t\le T}\rho(t)\left|\tilde{W}_2(1+t)-\tilde{W}_2(1+th)-\left(\tilde{W}_2(1+t)-\tilde{W}_2(1+th)\right)\right| \\
&+\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+T)-\tilde{W}_2(1+Th)-\left(\tilde{W}_2(1+t)-\tilde{W}_2(1+th)\right)\right| \\
=&\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+T)-\tilde{W}_2(1+Th)-\left(\tilde{W}_2(1+t)-\tilde{W}_2(1+th)\right)\right| \\
\le&\rho(t)\left|\tilde{W}_2(1+T)-\tilde{W}_2(1+Th)\right|+\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+t)-\tilde{W}_2(1+th)\right| \\
\le&2\sup_{t>T}\rho(t)\left|\tilde{W}_2(1+t)-\tilde{W}_2(1+th)\right|=o(1)\text{ a.s.}\quad\text{as } T\to\infty.
\end{aligned}\tag{4.17}$$

Let $\{W_2(t):0<t\le T\}:=\{\frac{1}{\sigma_2}(\tilde{W}_2(1+t)-\tilde{W}_2(1)):0<t\le T\}$ and $\{W_1(t):0<t\le T\}:=\{\frac{1}{\sigma_2}\tilde{W}_1(t):0<t\le T\}$ which are independent standard Wiener Processes. Then, it holds for $0<t\le T$

$$\tilde{W}_2(1+t)-\tilde{W}_2(1+th)+t(1-h)\tilde{W}_1(1)=\sigma_2\left(W_2(t)-W_2(th)\right)+t(1-h)\sigma_1W_1(1).\tag{4.18}$$

Hence, we obtain

$$\begin{aligned}
&\sup_{0<t\le\tau}\rho(t)\left|\tilde{W}_2(1+t)-\tilde{W}_2(1+th)+t(1-h)\tilde{W}_1(1)\right| \\
=&\sup_{0<t\le\tau}\rho(t)\left|\sigma_2\left(W_2(t)-W_2(th)\right)+t(1-h)\sigma_1W_1(1)\right| \\
\le&\sup_{0<t\le\tau}\rho(t)\left|\sigma_2W_2(t)\right|+\sup_{0<t\le\tau}\rho(t)\left|\sigma_2W_2(th)\right|+(1-h)\sup_{0<t\le\tau}t\rho(t)\left|\sigma_1W_1(1)\right| \\
\le&\sup_{0<t\le\tau}\rho(t)\left|\sigma_2W_2(t)\right|+\sup_{0<t\le\tau}\rho(t)\left|\sigma_2W_2(th)\right|+(1-h)\tau^{1-\gamma}\sup_{0<t\le\tau}t^{\gamma}\rho\left(t\right)\left|\sigma_1W_1(1)\right|,
\end{aligned}$$

With the self-similarity of the Wiener Process and the law of the iterated logarithm we obtain

$$
\begin{aligned}
&\sup_{0<t\le\tau} \rho(t)\left|\sigma_2 W_2(th)\right| \\
=&\sigma_2 \sup_{s\ge\frac{1}{h\tau}} \rho\left(\frac{1}{hs}\right)\left|W_2\left(\frac{1}{s}\right)\right| \\
\overset{\mathcal{D}}{=}&\sigma_2 \sup_{s\ge\frac{h}{\tau}} \left|W_2(s)\right| \frac{1}{s}\rho\left(\frac{1}{hs}\right) \\
=&\sigma_2 \sup_{s\ge\frac{1}{h\tau}} \frac{\sqrt{s\log\log s}}{s}\rho\left(\frac{1}{hs}\right)\frac{|W_2(s)|}{\sqrt{s\log\log s}} \\
\le&\sigma_2 h^\gamma \sup_{s\ge\frac{1}{h\tau}} \frac{\sqrt{\log\log s}}{s^{1-\gamma-\frac{1}{2}}} \sup_{s\ge\frac{h}{\tau}}\left(\frac{1}{hs}\right)^{\gamma}\rho\left(\frac{1}{hs}\right)\sup_{s\ge\frac{h}{\tau}}\frac{|W_2(s)|}{\sqrt{s\log\log s}} \\
=&o_P(1) \quad \text{as } \tau\to 0
\end{aligned}
$$

and with (3.36) it follows

$$
\sup_{0<t\le\tau} \rho(t)\left|\tilde{W}_2(1+t)-\tilde{W}_2(1+th)+t(1-h)\tilde{W}_1(1)\right| = o_P(1) \quad \text{as } \tau\to 0. \tag{4.19}
$$

Based on Lemma B.2 we can combine (4.14), (4.15), (4.16), (4.17), (4.18) and (4.19) such that we obtain

$$
\begin{aligned}
&\sup_{k\ge1} w(m,k)\left|\sum_{j=m+[kh]+1}^{m+k} h_2(Y_j)+\frac{k-[kh]}{m}\sum_{i=1}^{m}h_1(Y_i)\right| \\
\xrightarrow{\mathcal{D}}&\sup_{t>0}\rho(t)\left|\sigma_2\left(W_2(t)-W_2(th)\right)+t(1-h)\sigma_1 W_1(1)\right|
\end{aligned}
$$

and thus the assertion.

(ii) With Lemma 4.1 it holds

$$
\begin{aligned}
&\sup_{k\ge1}\left|w(m,k)|\Gamma_3(m,k)|-w(m,k)\sup_{1\le l\le k}\left|\sum_{j=m+l+1}^{m+k}h_2(Y_j)+\frac{k-l}{m}\sum_{i=1}^{m}h_1(Y_i)\right|\right| \\
\le&\sup_{k\ge1}w(m,k)\sup_{1\le l\le k}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+l+1}^{m+k}r(Y_i,Y_j)\right| = o_P(1), \quad m\to\infty
\end{aligned}
$$

due to Assumption 3.3 (i). According to Lemma B.1 it remains to show that, as $m\to\infty$,

$$
\begin{aligned}
&\sup_{k\ge1}w(m,k)\sup_{1\le l\le k}\left|\sum_{j=m+l+1}^{m+k}h_2(Y_j)+\frac{k-l}{m}\sum_{i=1}^{m}h_1(Y_i)\right| \\
\xrightarrow{\mathcal{D}}&\sup_{t>0}\rho(t)\sup_{0<s\le t}\left|\sigma_2(W_2(t)-W_2(s))+(t-s)\sigma_1 W_1(1)\right|,
\end{aligned}
$$

where $\{W_1(t)\}$ and $\{W_2(t)\}$ are wo independent standard Wiener processes. With Assumption 3.3 (ii) we get that for any fixed $\tau, T > 0$ and $m \to \infty$

$$\begin{aligned}
&\sup_{t>\tau} \rho(t) \sup_{0<s\leq t} \left| \frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t-s}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&\xrightarrow{\mathcal{D}} \sup_{t>\tau} \rho(t) \sup_{0<s\leq t} \left| \tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1+\min(s,T)) + (t-s)\tilde{W}_1(1) \right|.
\end{aligned} \tag{4.20}$$

As $\frac{l_m}{m} \to 0$, there exists an m_τ such that $\frac{l_m}{m} < \tau$ for all $m \geq m_\tau$. Hence, it holds for $m \geq m_\tau$

$$\begin{aligned}
&\Bigg| \sup_{t>\tau} \rho(t) \sup_{0<s\leq t} \left| \frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t-s}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&- \sup_{k>\tau m} w(m,k) \sup_{1\leq l\leq k} \left| \sum_{j=m+\min(l,mT)+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-l}{m} \sum_{i=1}^{m} h_1(Y_i) \right| \Bigg| \\
=& \Bigg| \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \rho(t) \sup_{0<s\leq t} \left| \frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t-s}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \\
&- \sup_{\frac{k}{m}>\tau} \rho\left(\frac{k}{m}\right) \sup_{1\leq l\leq k} \left| \frac{1}{\sqrt{m}} \sum_{j=m+\min(l,mT)+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-l}{m} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right| \Bigg| \\
\leq& \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \Bigg| \rho(t) \sup_{0<s\leq t} \left(\frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \frac{t-s}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \\
&- \rho\left(\frac{k}{m}\right) \sup_{1\leq l\leq k} \left(\frac{1}{\sqrt{m}} \sum_{j=m+\min(l,mT)+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-l}{m} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \Bigg| \\
\leq& \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right) \right| \sup_{t>\tau} \sup_{0<s\leq t} \Bigg| \frac{1}{t} \frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j) \\
&+ \left(1 - \frac{s}{t}\right) \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \Bigg| \\
&+ \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\leq t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \Bigg| \sup_{1\leq l\leq k} \Bigg(\frac{m}{k} \frac{1}{\sqrt{m}} \sum_{j=m+\min(l,mT)+1}^{m+\min(k,mT)} h_2(Y_j) \\
&+ \left(1 - \frac{l}{k}\right) \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \Bigg) \\
&- \sup_{0<s\leq t} \left(\frac{1}{t} \frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j) + \left(1 - \frac{s}{t}\right) \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \right) \Bigg|
\end{aligned}$$

$$
\begin{aligned}
&\le \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| \sup_{t>\tau} \sup_{0<s\le t} \left| \frac{1}{t}\frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j) \right. \\
&\quad \left. + \left(1-\frac{s}{t}\right)\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right| \\
&\quad + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \cdot \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \sup_{0<\frac{l}{m}\le\frac{k}{m}} \sup_{\frac{l}{m}<s\le\frac{l+1}{m}} \left| \left(\frac{s}{t}-\frac{l}{k}\right)\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right| \\
&\quad + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \cdot \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \sup_{0<s\le t} \left| \left(\frac{m}{k}-\frac{1}{t}\right)\frac{1}{\sqrt{m}} \sum_{j=m+[m\min(s,T)]+1}^{m+[m\min(t,T)]} h_2(Y_j)\right| \\
&\le \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| O_P(1) \\
&\quad + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \cdot \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \sup_{0<\frac{l}{m}\le\frac{k}{m}} \sup_{\frac{l}{m}<s\le\frac{l+1}{m}} \left|\frac{s}{t}-\frac{l}{k}\right| O_P(1) \\
&\quad + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right) \cdot \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \sup_{0<s\le t} \left|\frac{m}{k}-\frac{1}{t}\right| O_P(1) \\
&\le \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \left| t\rho(t) - \frac{k}{m}\rho\left(\frac{k}{m}\right)\right| O_P(1) \\
&\quad + \sup_{\frac{k}{m}>\tau} \sup_{\frac{k}{m}\le t<\frac{k+1}{m}} \frac{k}{m}\rho\left(\frac{k}{m}\right)\left(\frac{1}{\tau m}O_P(1) + \frac{1}{m}O_P(1)\right) = o_P(1) \quad \text{as } m\to\infty
\end{aligned}
$$

with Assumption 3.3 (ii) as well as (3.18) and (3.20). Together with (4.20) it follows

$$
\begin{aligned}
&\sup_{k>\tau m} w(m,k) \sup_{1\le l\le k} \left| \sum_{j=m+\min(l,mT)+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-l}{m}\sum_{i=1}^{m} h_1(Y_i)\right| \\
&\xrightarrow{\mathcal{D}} \sup_{t>\tau} \rho(t) \sup_{0<s\le t} \left| \tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1+\min(s,T)) + (t-s)\tilde{W}_1(1)\right|.
\end{aligned} \tag{4.21}
$$

It holds

$$
\begin{aligned}
&\sup_{k>mT} \frac{\sqrt{m}}{k} \sup_{1\le l\le k} \left| \sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\le \sup_{k>mT} \sup_{1\le l\le m} \frac{\sqrt{m}}{k} \left| \sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \sup_{k>mT} \sup_{m\le l\le m\sqrt{T}} \frac{\sqrt{m}}{k} \left| \sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\quad + \sup_{k>mT} \sup_{m\sqrt{T}\le l\le mT} \frac{\sqrt{m}}{k} \left| \sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \sup_{k>mT} \sup_{mT<l\le k} \frac{\sqrt{m}}{k} \left| \sum_{j=m+1}^{m+l} h_2(Y_j)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{1\leq l\leq m} \frac{\sqrt{m}}{Tm}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \sup_{m\leq l\leq m\sqrt{T}} \frac{\sqrt{m}}{Tm}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\quad + \sup_{m\sqrt{T}\leq l\leq mT} \frac{\sqrt{m}}{Tm}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \sup_{l>mT} \frac{\sqrt{m}}{l}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&=\frac{1}{T}\sup_{1\leq l\leq m} \frac{1}{\sqrt{m}}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \frac{1}{\sqrt{T}}\sup_{m\leq l\leq m\sqrt{T}} \frac{\sqrt{m}}{\sqrt{T}m}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\quad + \frac{1}{T^{1/4}}\sup_{m\sqrt{T}\leq l\leq mT} \frac{T^{1/4}\sqrt{m}}{Tm}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \frac{1}{\sqrt{T}}\sup_{l>mT} \frac{\sqrt{Tm}}{l}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\leq\frac{1}{T}\sup_{1\leq l\leq m} \frac{1}{\sqrt{m}}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \frac{1}{\sqrt{T}}\sup_{l\geq m} \frac{\sqrt{m}}{l}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\quad + \frac{1}{T^{1/4}}\sup_{l\geq m\sqrt{T}} \frac{T^{1/4}\sqrt{m}}{l}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| + \frac{1}{\sqrt{T}}\sup_{l>mT} \frac{\sqrt{Tm}}{l}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\overset{\mathcal{D}}{=}\frac{1}{T}\sup_{1\leq l\leq m} \frac{1}{\sqrt{m}}\left|\sum_{j=1}^{l} h_2(Y_j)\right| + \frac{1}{\sqrt{T}}\sup_{l\geq m} \frac{\sqrt{m}}{l}\left|\sum_{j=1}^{l} h_2(Y_j)\right| \\
&\quad + \frac{1}{T^{1/4}}\sup_{l\geq m\sqrt{T}} \frac{T^{1/4}\sqrt{m}}{l}\left|\sum_{j=1}^{l} h_2(Y_j)\right| + \frac{1}{\sqrt{T}}\sup_{l>mT} \frac{\sqrt{Tm}}{l}\left|\sum_{j=1}^{l} h_2(Y_j)\right| \\
&=o_P(1) \quad \text{as } T\to\infty \text{ uniformly in m,}
\end{aligned}
$$

where we consider Assumption 3.3 (ii) for the convergence of the first summand and Assumption 3.3 (iv) for the remaining summands. It follows

$$
\begin{aligned}
&\sup_{k>mT} w(m,k) \sup_{1\leq l<k}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&= \sup_{k>mT} \frac{kw(m,k)}{\sqrt{m}} \sup_{1\leq l<k} \frac{\sqrt{m}}{k}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&= \sup_{k>\max(mT,l_m)} \frac{k\rho\left(\frac{k}{m}\right)}{m} \sup_{1\leq l<k} \frac{\sqrt{m}}{k}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\leq \sup_{k>mT} \frac{k}{m}\rho\left(\frac{k}{m}\right) \sup_{1\leq l<k} \frac{\sqrt{m}}{k}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&\leq \sup_{t>T} t\rho(t) \sup_{1\leq l<k} \frac{\sqrt{m}}{k}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
&=o_P(1) \quad \text{as } T\to\infty \text{ uniformly in m}
\end{aligned}
$$

as $\lim_{t\to\infty} t\rho(t) < \infty$. By Assumption 3.3 (iv) and the stationarity we obtain

$$\begin{aligned}
&\sup_{k>mT} w(m,k)\left|\sum_{j=m+1}^{m+mT} h_2(Y_j)\right| \\
\leq &\frac{1}{\sqrt{mT}}\left|\sum_{j=m+1}^{m+[mT]} h_2(Y_j)\right| \sup_{k>mT}\frac{k}{\sqrt{mT}} w(m,k) \\
\leq &O_P(1)\frac{1}{\sqrt{T}}\sup_{k>mT}\frac{k}{m}\rho\left(\frac{k}{m}\right) \\
\leq &O_P(1)\frac{1}{\sqrt{T}}\sup_{t>T} t\rho(t) \\
= &\frac{1}{\sqrt{T}}O_P(1) \xrightarrow{P} 0 \quad \text{as } T\to\infty \text{ uniformly in m}
\end{aligned}$$

as $\lim_{t\to\infty} t\rho(t) < \infty$. Together with (3.22) we get

$$\begin{aligned}
&\sup_{k>\tau m}\Bigg| w(m,k)\sup_{1\leq l\leq k}\left(\sum_{j=m+\min(l,mT)+1}^{m+\min(k,mT)} h_2(Y_j) + \frac{k-l}{m}\sum_{i=1}^{m} h_1(Y_i)\right) \\
&\quad - w(m,k)\sup_{1\leq l\leq k}\left(\sum_{j=m+l+1}^{m+k} h_2(Y_j) + \frac{k-l}{m}\sum_{i=1}^{m} h_1(Y_i)\right)\Bigg| \\
\leq &\sup_{k>\tau m} w(m,k)\sup_{1\leq l\leq k}\left|\sum_{j=m+\min(l,mT)+1}^{m+\min(k,mT)} h_2(Y_j) - \sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
\leq &\sup_{\tau m<k\leq mT} w(m,k)\sup_{1\leq l\leq k}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j) - \sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
&\quad + \sup_{k>mT} w(m,k)\sup_{1\leq l<mT}\left|\sum_{j=m+l+1}^{m+mT} h_2(Y_j) - \sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
&\quad + \sup_{k>mT} w(m,k)\sup_{mT\leq l\leq k}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
= &\sup_{k>mT} w(m,k)\left|\sum_{j=m+k+1}^{m+mT} h_2(Y_j)\right| + \sup_{k>mT} w(m,k)\sup_{mT\leq l\leq k}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
\leq &\sup_{k>mT} w(m,k)\left|\sum_{j=m+k+1}^{m+mT} h_2(Y_j)\right| + \sup_{k>mT} w(m,k)\sup_{mT\leq l\leq k}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
\leq &\sup_{k>mT} w(m,k)\left|\sum_{j=m+1}^{m+mT} h_2(Y_j)\right| + 2\sup_{k>mT} w(m,k)\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \\
&\quad + \sup_{k>mT} w(m,k)\sup_{1\leq l\leq k}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| = o_P(1) \quad \text{as } T\to\infty \text{ uniformly in m.}
\end{aligned} \tag{4.22}$$

For $\gamma < \alpha < \frac{1}{2}$ it holds

$$\begin{aligned}
&\sup_{1\le k\le \tau m} w(m,k) \sup_{1\le l\le k}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j) + \frac{k-l}{m}\sum_{i=1}^{m} h_1(Y_i)\right| \\
\le& \sup_{1\le k\le \tau m} w(m,k) \sup_{1\le l\le k}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| + \sup_{1\le k\le \tau m} w(m,k) \sup_{1\le l\le k}\left|\frac{k-l}{m}\sum_{i=1}^{m} h_1(Y_i)\right| \\
\le& \sup_{1\le k\le \tau m} m^{\frac{1}{2}-\alpha}k^{\alpha} w(m,k) \sup_{1\le l\le k}\left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
&+ \sup_{1\le k\le \tau m} \frac{k}{m} w(m,k)\sqrt{m}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right| \\
=& \sup_{l_m< k\le \tau m} \left(\frac{k}{m}\right)^{\alpha}\rho\left(\frac{k}{m}\right) \sup_{1\le l\le k}\left|\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
&+ \sup_{l_m< k\le \tau m} \frac{k}{m}\rho\left(\frac{k}{m}\right)\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right| \\
\le& \sup_{1\le k\le \tau m} \left(\frac{k}{m}\right)^{\alpha}\rho\left(\frac{k}{m}\right) \sup_{1\le k\le \tau m}\sup_{1\le l\le k}\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
&+ \sup_{1\le k\le \tau m} \frac{k}{m}\rho\left(\frac{k}{m}\right)\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right|.
\end{aligned}$$

With Assumption 3.3 (iii) and the stationarity it holds for all $0<\alpha<\frac{1}{2}, \tau<1$

$$\begin{aligned}
&\sup_{1\le k\le \tau m}\sup_{1\le l\le k}\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
\le& \sup_{1\le k\le m}\sup_{1\le l\le k}\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j)\right| \\
\le& \sup_{1\le k\le m}\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| + \sup_{1\le k\le m}\sup_{1\le l\le k}\frac{1}{m^{\frac{1}{2}-\alpha}l^{\alpha}}\left|\sum_{j=m+1}^{m+l} h_2(Y_j)\right| \\
\le& 2\sup_{1\le k\le m}\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right| \overset{\mathcal{D}}{=} 2\sup_{1\le k\le m}\frac{1}{m^{\frac{1}{2}-\alpha}k^{\alpha}}\left|\sum_{j=1}^{k} h_2(Y_j)\right| \\
=& O_P(1) \quad (m\to\infty).
\end{aligned}$$

It holds

$$\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_1(Y_i)\right| = O_P(1)$$

as $m\to\infty$ by Assumption 3.3 (ii). With (3.26) und (3.27) we obtain

$$\sup_{1\le k\le \tau m} w(m,k) \sup_{1\le l\le k}\left|\sum_{j=m+l+1}^{m+k} h_2(Y_j) + \frac{k-l}{m}\sum_{i=1}^{m} h_1(Y_i)\right| = o_P(1) \tag{4.23}$$

as $\tau \to 0$ uniformly in m.

With (3.24) and (3.29) we obtain

$$\begin{aligned}
&\sup_{t>T} \rho(t) \left|\tilde{W}_2(1+T)\right| \leq \sup_{t>T} \rho(t)(1+T) \frac{\left|\tilde{W}_2(1+T)\right|}{1+T} \\
\leq &(\sup_{t>T} \rho(t) + \sup_{t>T} t\rho(t)) \sup_{t>T} \frac{\left|\tilde{W}_2(1+t)\right|}{1+t} = o(1) \quad \text{a.s} \quad \text{as } T \to \infty,
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{t>T} \rho(t) \sup_{T<s\leq t} \left|\tilde{W}_2(1+s)\right| \leq \sup_{t>T} (1+t)\rho(t) \sup_{s>T} \frac{\left|\tilde{W}_2(1+s)\right|}{1+s} \\
\leq &(\sup_{t>T} \rho(t) + \sup_{t>T} t\rho(t)) o(1) = o(1) \quad \text{a.s} \quad \text{as } T \to \infty.
\end{aligned}$$

It follows with (3.30)

$$\begin{aligned}
&\sup_{t>\tau} \Bigg| \rho(t) \sup_{0<s\leq t} \left|\tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1+\min(s,T)) + (t-s)\tilde{W}_1(1)\right| \\
&-\rho(t) \sup_{0<s\leq t} \left|\tilde{W}_2(1+t) - \tilde{W}_2(1+s) + (t-s)\tilde{W}_1(1)\right| \Bigg| \\
\leq &\sup_{t>\tau} \rho(t) \sup_{0<s\leq t} \left|\tilde{W}_2(1+\min(t,T)) - \tilde{W}_2(1+\min(s,T)) - \Big(\tilde{W}_2(1+t) - \tilde{W}_2(1+s)\Big)\right| \\
\leq &\sup_{\tau<t\leq T} \rho(t) \sup_{0<s\leq t} \left|\tilde{W}_2(1+t) - \tilde{W}_2(1+s) - \Big(\tilde{W}_2(1+t) - \tilde{W}_2(1+s)\Big)\right| \\
&+ \sup_{t>T} \rho(t) \sup_{0<s\leq T} \left|\tilde{W}_2(1+T) - \tilde{W}_2(1+s) - \Big(\tilde{W}_2(1+t) - \tilde{W}_2(1+s)\Big)\right| \\
&+ \sup_{t>T} \rho(t) \sup_{T<s\leq t} \left|\tilde{W}_2(1+T) - \tilde{W}_2(1+T) - \Big(\tilde{W}_2(1+t) - \tilde{W}_2(1+s)\Big)\right| \\
= &\sup_{t>T} \rho(t) \left|\tilde{W}_2(1+T) - \tilde{W}_2(1+t)\right| + \sup_{t>T} \rho(t) \sup_{T<s\leq t} \left|\tilde{W}_2(1+t) - \tilde{W}_2(1+s)\right| \\
\leq &\sup_{t>T} \rho(t) \left|\tilde{W}_2(1+T)\right| + 2\sup_{t>T} \rho(t) \left|\tilde{W}_2(1+t)\right| + \sup_{t>T} \rho(t) \sup_{T<s\leq t} \left|\tilde{W}_2(1+s)\right| \\
= &o(1) \text{ a.s.} \quad \text{as } T \to \infty.
\end{aligned} \tag{4.24}$$

Let $\{W_2(t) : 0 < t \leq T\} := \{\frac{1}{\sigma_2}(\tilde{W}_2(1+t) - \tilde{W}_2(1)) : 0 < t \leq T\}$ and $\{W_1(t) : 0 < t \leq T\} := \{\frac{1}{\sigma_2}\tilde{W}_1(t) : 0 < t \leq T\}$ which are again independent standard Wiener Processes. Then, it holds for $0 < t \leq T, 0 < s \leq t$

$$\begin{aligned}
&\tilde{W}_2(1+t) - \tilde{W}_2(1+s) + (t-s)\tilde{W}_1(1) \\
= &\sigma_2 \left(W_2(t) - W_2(s)\right) + (t-s)\sigma_1 W_1(1).
\end{aligned} \tag{4.25}$$

Hence, we obtain

$$
\begin{aligned}
&\sup_{0<t\leq\tau}\rho(t)\sup_{0<s\leq t}\left|\tilde{W}_2(1+t)-\tilde{W}_2(1+s)+(t-s)\tilde{W}_1(1)\right|\\
=&\sup_{0<t\leq\tau}\rho(t)\sup_{0<s\leq t}\left|\sigma_2\left(W_2(t)-W_2(s)\right)+(t-s)\sigma_1 W_1(1)\right|\\
\leq&\sup_{0<t\leq\tau}\rho(t)\left|\sigma_2 W_2(t)\right|+\sup_{0<t\leq\tau}\rho(t)\sup_{0<s\leq t}\left|\sigma_2 W_2(s)\right|\\
&+\sup_{0<t\leq\tau}t\rho(t)\left|\sigma_1 W_1(1)\right|\\
\leq&\sup_{0<t\leq\tau}\rho(t)\left|\sigma_2 W_2(t)\right|+\sup_{0<t\leq\tau}\rho(t)\sup_{0<s\leq t}\left|\sigma_2 W_2(s)\right|+\sup_{0<t\leq\tau}t\rho(t)\left|\sigma_1 W_1(1)\right|\\
\leq&\sup_{0<t\leq\tau}\rho(t)\left|\sigma_2 W_2(t)\right|+\sup_{0<t\leq\tau}\rho(t)\sup_{0<s\leq t}\left|\sigma_2 W_2(s)\right|+\tau^{1-\gamma}\sup_{0<t\leq\tau}t^{\gamma}\rho\left(t\right)\left|\sigma_1 W_1(1)\right|,
\end{aligned}
$$

where $\{W_2(t) : 0 < t \leq T\} := \{\frac{1}{\sigma_2}\tilde{W}_2(1+t)-\tilde{W}_2(1) : 0 < t \leq T\}$, $\{W_1(t) : 0 < t \leq T\} := \{\frac{1}{\sigma_2}\tilde{W}_1(t) : 0 < t \leq T\}$ are independent standard Wiener Processes. With the self-similarity of the Wiener Process and the law of the iterated logarithm we obtain

$$
\begin{aligned}
&\sup_{0<t\leq\tau}\rho(t)\sup_{0<s\leq t}\left|\sigma_2 W_2(s)\right|\\
=&\sigma_2\sup_{0<t\leq\tau}\rho(t)\sup_{\tilde{s}\geq\frac{1}{t}}\left|W_2\left(\frac{1}{\tilde{s}}\right)\right|\\
\stackrel{\mathcal{D}}{=}&\sigma_2\sup_{0<t\leq\tau}\rho(t)\sup_{\tilde{s}\geq\frac{1}{t}}\frac{\left|W_2\left(\tilde{s}\right)\right|}{\tilde{s}}\\
=&\sigma_2\sup_{0<t\leq\tau}\rho(t)\sup_{\tilde{s}\geq\frac{1}{t}}\frac{\sqrt{\tilde{s}\log\log\tilde{s}}}{\tilde{s}}\frac{\left|W_2\left(\tilde{s}\right)\right|}{\sqrt{\tilde{s}\log\log\tilde{s}}}\\
\leq&\sigma_2\sup_{0<t\leq\tau}t^{\gamma}\rho(t)\sup_{\tilde{s}\geq\frac{1}{t}}\frac{\sqrt{\log\log\tilde{s}}}{\tilde{s}^{1-\gamma-\frac{1}{2}}}\sup_{\tilde{s}\geq\frac{1}{t}}\frac{\left|W_2\left(\tilde{s}\right)\right|}{\sqrt{\tilde{s}\log\log\tilde{s}}}\\
=&o_P(1)\quad\text{as }\tau\to 0
\end{aligned}
$$

and with (3.36) we get

$$
\sup_{0<t\leq\tau}\rho(t)\sup_{0<s\leq t}\left|\tilde{W}_2(1+t)-\tilde{W}_2(1+s)+(t-s)\tilde{W}_1(1)\right|=o_P(1)\quad\text{as }\tau\to 0. \tag{4.26}
$$

Due to Lemma B.2 we can combine (4.21), (4.22), (4.23), (4.24), (4.25) and (4.26) such that we obtain

$$
\begin{aligned}
&\sup_{k\geq 1}w(m,k)\sup_{1\leq l\leq k}\left|\sum_{j=m+l+1}^{m+k}h_2(Y_j)+\frac{k-l}{m}\sum_{i=1}^{m}h_1(Y_i)\right|\\
\stackrel{\mathcal{D}}{\to}&\sup_{t>0}\rho(t)\sup_{0<s\leq t}\left|\sigma_2\left(W_2(t)-W_2(s)\right)+(t-s)\sigma_1 W_1(1)\right|
\end{aligned}
$$

and thus the assertion.

□

Corollary 4.3. *a) If $\sigma_1 = \sigma_2 =: \sigma$ the limit distributions in Theorem 4.2 reduce to*

(i) $\sup_{k\geq 1} w(m,k) \left|\Gamma_2(m,k)\right| \xrightarrow{\mathcal{D}} \sigma \sup_{0<t<1} \rho\left(\frac{t}{1-t}\right) \left|\frac{W(t)}{1-t} - (1-t(1-h))\frac{W\left(\frac{th}{1-t(1-h)}\right)}{1-t}\right|$

(ii) $\sup_{k\geq 1} w(m,k) \left|\Gamma_3(m,k)\right| \xrightarrow{\mathcal{D}} \sigma \sup_{0<t<1} \rho\left(\frac{t}{1-t}\right) \sup_{0<s\leq t} \left|\frac{W(t)}{1-t} - \frac{W(s)}{1-s}\right|,$

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

b) In this case, we obtain for the weight function in (3.5)

(i) $\sup_{k\geq 1} w(m,k) \left|\Gamma_2(m,k)\right| \xrightarrow{\mathcal{D}} \sigma \sup_{0<t<1} t^{-\gamma} \left|W(t) - (1-t(1-h))W\left(\frac{th}{1-t(1-h)}\right)\right|$

(ii) $\sup_{k\geq 1} w(m,k) \left|\Gamma_3(m,k)\right| \xrightarrow{\mathcal{D}} \sigma \sup_{0<t<1} t^{-\gamma} \sup_{0<s\leq t} \left|W\left(t\right) - \frac{1-t}{1-s} W\left(s\right)\right|$

Proof. a) With (3.39) we obtain

$$\begin{aligned}
&\sup_{t>0} \rho(t) \left|W_2(t) - W_2(th) + t(1-h)W_1(1)\right| \\
=&\sup_{t>0} \rho(t) \left|W_2(t) + tW_1(1) - (W_2(th) + thW_1(1))\right| \\
\stackrel{\mathcal{D}}{=}&\sup_{t>0} \rho(t) \left|(1+t)W\left(\frac{t}{1+t}\right) - (1+th)W\left(\frac{th}{1+th}\right)\right| \\
=&\sup_{0<s<1} \rho\left(\frac{s}{1-s}\right) \left|\frac{W(s)}{1-s} - (1-s(1-h))\frac{W\left(\frac{sh}{1-s(1-h)}\right)}{1-s}\right|
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{t>0} \rho(t) \sup_{0<s\leq t} \left|W_2(t) - W_2(s) + (t-s)W_1(1)\right| \\
=&\sup_{t>0} \rho(t) \sup_{0<s\leq t} \left|W_2(t) + tW_1(1) - (W_2(s) + sW_1(1))\right| \\
\stackrel{\mathcal{D}}{=}&\sup_{t>0} \rho(t) \sup_{0<s\leq t} \left|(1+t)W\left(\frac{t}{1+t}\right) - (1+s)W\left(\frac{s}{1+s}\right)\right| \\
\stackrel{\mathcal{D}}{=}&\sup_{0<\tilde{t}<1} \rho\left(\frac{\tilde{t}}{1-\tilde{t}}\right) \sup_{0<\tilde{s}\leq\tilde{t}} \left|\frac{W\left(\tilde{t}\right)}{1-\tilde{t}} - \frac{W\left(\tilde{s}\right)}{1-\tilde{s}}\right|
\end{aligned}$$

For $\sigma_1 = \sigma_2 =: \sigma$ the assertions now follow with Theorem 4.2.

b) The assertion is obtained with $\rho(t) = (1+t)^{-1}\left(\frac{t}{1+t}\right)^{-\gamma}$ in a). Hence $\rho\left(\frac{t}{1-t}\right) = (1-t)t^{-\gamma}$.

□

4.2. Asymptotics Under the Alternative

In the following we will show that the Page-MOSUM as well as the modified MOSUM procedure have asymptotic power one. For the Page-MOSUM we use the following adapted assumption on the weight function.

Assumption 4.4. *(i) If* $\frac{k^*}{m} \to \infty$, *assume that* $\liminf_{t\to\infty} t\rho(t) > 0$.

(ii) If $\frac{k^*}{m} = O(1)$, *i.e.* $\frac{k^*}{m} < \nu$ *for all* $m \geq 1$ *for some* $\nu > 0$, *assume that there exist* $t_0 > \nu, \epsilon > 0$ *such that* $\rho(t) > 0$ *for all* $t \in \left(\frac{t_0}{h} - \epsilon, \frac{t_0}{h} + \epsilon\right)$.

Theorem 4.5. *Let the regularity conditions given in Assumption 3.2 , and 3.13 be fulfilled. Furthermore assume that* $\sqrt{m}\Delta_m \to \infty$. *Then it holds under the alternative*

(i) $\sup_{k\geq 1} w(m,k)\,|\Gamma_2(m,k)| \xrightarrow{P} \infty$ *if Assumption 4.4 is fulfilled.*

(ii) $\sup_{k\geq 1} w(m,k)\,|\Gamma_3(m,k)| \xrightarrow{P} \infty$ *if Assumption 3.12 is fulfilled.*

Proof. (i) For $\tilde{k} > \frac{k^*}{h}$ Hoeffding's decomposition in (3.44) yields

$$\begin{aligned}\Gamma_2(m,k) =&\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+[\tilde{k}h]+1}^{m+\tilde{k}} (h(Y_i, Z_{j,m}) - \theta)\\ =&\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+[\tilde{k}h]+1}^{m+k} r_m^*(Y_i, Z_{j,m}) + \sum_{j=m+[\tilde{k}h]+1}^{m+\tilde{k}} h_{2,m}^*(Z_{j,m})\\ &+\frac{\tilde{k}-[\tilde{k}h]}{m}\sum_{i=1}^{m} h_{1,m}^*(Y_i) + (\tilde{k}-[\tilde{k}h])\Delta_m.\end{aligned}$$

We first consider late changes with $\frac{k^*}{m} \to \infty$ and $\tilde{k} = \left[2\frac{k^*}{h}\right]$. Let $k' = \tilde{k} - k^*$ and $h' = \frac{\tilde{k}h - k^*}{\tilde{k}-k^*} \in (0,1)$. As $\rho(t) = (1+t)^{-1}$ fulfills Assumption 3.2, we obtain with Lemma 4.1

$$\begin{aligned}&\frac{\sqrt{m}}{k^*}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+[\tilde{k}h]+1}^{m+\tilde{k}} r_m^*(Y_i, Z_{j,m})\right| = \frac{\sqrt{m}}{k^*}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[k'h']+1}^{k'} r_m^*(Y_i, Z_{m+k^*+j,m})\right|\\ =&\left(\frac{m}{k^*} + \frac{k'}{k^*}\right) o_P(1) \leq \left(\frac{m}{k^*} + \frac{2}{h} - 1\right) o_P(1) = o_P(1) \quad m\to\infty.\end{aligned}$$

With Assumption 3.13 (iii) we get

$$\begin{aligned}&\left|\sum_{j=m+[\tilde{k}h]+1}^{m+\tilde{k}} h_{2,m}^*(Z_{j,m}) + \frac{\tilde{k}-[\tilde{k}h]}{m}\sum_{i=1}^{m} h_{1,m}^*(Y_i)\right|\\ \leq&\sqrt{[\tilde{k}h]}\left|\frac{1}{\sqrt{[k'h]}}\sum_{j=m+k^*+1}^{m+k^*+[k'h]} h_{2,m}^*(Z_{j,m})\right| + \sqrt{\tilde{k}}\left|\frac{1}{\sqrt{k'}}\sum_{j=m+k^*+1}^{m+k^*+k'} h_{2,m}^*(Z_{j,m})\right|\\ &+\frac{\tilde{k}-[\tilde{k}h]}{\sqrt{m}}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_{1,m}^*(Y_i)\right|\\ =&\sqrt{[\tilde{k}h]} + \sqrt{\tilde{k}} + \frac{\tilde{k}-[\tilde{k}h]}{\sqrt{m}} O_P(1).\end{aligned} \tag{4.27}$$

Hence, it follows

$$
\begin{aligned}
&\frac{\sqrt{m}}{k^*}\left|\sum_{j=m+[\tilde{k}h]+1}^{m+\tilde{k}} h^*_{2,m}(Z_{j,m})+\frac{k-[\tilde{k}h]}{m}\sum_{i=1}^{m} h^*_{1,m}(Y_i)\right| \\
\leq&\frac{\sqrt{m}}{k^*}\left(\sqrt{[\tilde{k}h]}+\sqrt{\tilde{k}}+\frac{\tilde{k}-[\tilde{k}h]}{\sqrt{m}}\right)O_P(1) \\
=&\frac{\tilde{k}}{k^*}\left(\sqrt{\frac{m}{\tilde{k}}}\sqrt{\frac{[\tilde{k}h]}{\tilde{k}}}+\sqrt{\frac{m}{\tilde{k}}}+\frac{\tilde{k}-[\tilde{k}h]}{\tilde{k}}\right)O_P(1) \\
=&\left(\frac{2}{h}+o(1)\right)\left(\sqrt{\frac{m}{\tilde{k}}}\sqrt{\frac{[\tilde{k}h]}{\tilde{k}}}+\sqrt{\frac{m}{\tilde{k}}}+1-\frac{[\tilde{k}h]}{\tilde{k}}\right)O_P(1)=O_P(1)
\end{aligned}
\tag{4.28}
$$

as $\frac{m}{\tilde{k}}<h\frac{m}{k^*}\to 0$. Due to $\frac{l_m}{\tilde{k}}\leq\frac{l_m}{k^*}=\frac{m}{k^*}\frac{l_m}{m}\to 0$ there exists an $m_0\in\mathbb{N}$ such that $\tilde{k}>k^*>l_m$ for all $m\geq m_0$. Hence, it holds for $m\geq m_0$

$$
\begin{aligned}
&\sup_{k\geq 1} w(m,k)\left|\Gamma_2(m,k)\right|\geq w(m,\tilde{k})\left|\Gamma_2(m,\tilde{k})\right| \\
=&\frac{k^*}{\sqrt{m}}w(m,\tilde{k})\left(\left(\frac{\tilde{k}-[\tilde{k}h]}{k^*}\right)\sqrt{m}\Delta_m+O_P(1)\right) \\
=&\frac{k^*}{\tilde{k}}\frac{\tilde{k}}{m}\rho\left(\frac{\tilde{k}}{m}\right)\left(\left(\frac{\tilde{k}}{k^*}-\frac{[\tilde{k}h]}{k^*}\right)\sqrt{m}|\Delta_m|+O_P(1)\right) \\
\geq&\left(\frac{2}{h}+o(1)\right)\frac{\tilde{k}}{m}\rho\left(\frac{\tilde{k}}{m}\right)\left((1-h)\left(\frac{2}{h}+o(1)\right)\sqrt{m}|\Delta_m|+O_P(1)\right)\to\infty
\end{aligned}
$$

with Assumption 3.12 (i) and $\sqrt{m}|\Delta_m|\to\infty$.

For $\frac{k^*}{m}=O(1)$ we choose $\tilde{k}=\left[m\frac{t_0}{h}\right]$ with t_0 as in Assumption 4.4. Let $k'=\tilde{k}-k^*$ and $h'=\frac{\tilde{k}h-k^*}{\tilde{k}-k^*}\in(0,1)$. We consider a weight function $\tilde{w}$ according to Assumption 3.2 with $\tilde{l}_m=0$ and $\rho\equiv 1$ on $[0,\nu]$. Lemma 4.1 yields

$$
\begin{aligned}
&w(m,\tilde{k})\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+[\tilde{k}h]+1}^{m+\tilde{k}} r^*_m(Y_i,Z_{j,m})\right|=w(m,\tilde{k})\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=[k'h']+1}^{k'} r^*_m(Y_i,Z_{m+k^*+j,m})\right| \\
\leq&\rho\left(\frac{[m\frac{t_0}{h}]}{m}\right)o_P(1)=\rho\left(\frac{t_0}{h}+o(1)\right)O_P(1)=o_P(1)\quad m\to\infty.
\end{aligned}
$$

as ρ is bounded around $\frac{t_0}{h}$. Furthermore, we obtain with (4.27)

$$
\begin{aligned}
&w(m,\tilde{k})\left|\sum_{j=m+[\tilde{k}h]+1}^{m+\tilde{k}} h_{2,m}^*(Z_{j,m}) + \frac{\tilde{k}-[\tilde{k}h]}{m}\sum_{i=1}^{m} h_{1,m}^*(Y_i)\right| \\
\leq &w(m,\tilde{k})\left(\sqrt{[\tilde{k}h]} + \sqrt{\tilde{k}} + \frac{\tilde{k}-[\tilde{k}h]}{\sqrt{m}}\right) O_P(1) \\
\leq &\rho\left(\frac{\left[m\frac{t_0}{h}\right]}{m}\right)\left(\sqrt{t_0} + \sqrt{\frac{t_0}{h}} + t_0\left(h^{-1}-t\right) + \frac{h+1}{m}\right) O_P(1) \\
= &\rho\left(\frac{t_0}{h} + o(1)\right) O_P(1) = O_P(1).
\end{aligned}
$$

It holds $\frac{l_m}{\tilde{k}} = \frac{l_m}{m}O(1) \to 0$ as $m \to \infty$. Hence, there exists an $m_0 \in \mathbb{N}$ such that $\tilde{k} > l_m$ for all $m \geq m_0$. For such m we obtain

$$
\begin{aligned}
&\sup_{k\geq 1} w(m,k)\left|\Gamma_2(m,k)\right| \geq w(m,\tilde{k})\left|\Gamma_2(m,\tilde{k})\right| \\
= &\frac{1}{\sqrt{m}}\rho\left(\frac{\tilde{k}}{m}\right)(\tilde{k}-[\tilde{k}h])|\Delta_m| + O_P(1) \\
\geq &\rho\left(\frac{t_0}{h} + o(1)\right) x_0\left(1-h\right)\sqrt{m}|\Delta_m| + O_P(1) \xrightarrow{P} \infty
\end{aligned}
$$

with $\sqrt{m}|\Delta_m| \to \infty$ and Assumption 4.4 (ii).

(ii) For $\tilde{k} > k^*$ it holds

$$
\begin{aligned}
\left|\Gamma_3(m,\tilde{k})\right| &= \sup_{1\leq i\leq \tilde{k}}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+i+1}^{m+\tilde{k}}(h(X_i,X_j)-\theta)\right| \\
\geq \left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+\tilde{k}}(h(Y_i,Z_{j,m})-\theta)\right| &= \left|\Gamma(m,\tilde{k})-\Gamma(m,k^*)\right|.
\end{aligned}
$$

Hence, we obtain

$$
\sup_{k\geq 1} w(m,k)\left|\Gamma_3(m,k)\right| \geq w\left(m,\tilde{k}\right)\left|\Gamma(m,\tilde{k})-\Gamma(m,k^*)\right| \xrightarrow{P} \infty \quad \text{as } m \to \infty
$$

for $\frac{k^*}{m} \to \infty$ with $\tilde{k} = 2k^*$ and (3.50) as well as (3.52) and for $\frac{k^*}{m} = O(1)$ with $\tilde{k} = [x_0 m]$ and (3.57) as well as (3.59).

□

5. Stopping Time

Sequential procedures naturally involve a detection delay as some data has to be collected after the change to obtain a significant difference to the historic data set. Hence, the behavior of the stopping time of sequential procedures is of great interest, in particular for the comparison of different sequential procedures. In the literature, the speed of detection of sequential tests is mostly evaluated based on the average run length which is the expected value of the stopping time. Particularly such procedures that, in contrast to the approach adopted in this work, stop asymptotically with probability one even under structural stability are constructed referring to the mean detection delay and the mean time between false alarms as optimality criteria. For more details on such procedures and the analysis of the respective average run length see, for example, Lorden *et al.* (1971), Basseville *et al.* (1993) and Siegmund & Venkatraman (1995). In the framework of sequential testing as introduced in Chu *et al.* (1996), however, it is possible to derive the precise limit distribution of the stopping time which obviously provides more information than the average run length. This has first been achieved in Aue & Horváth (2004) for the CUSUM test for local mean changes that occur very early after the monitoring period has started. Fremdt (2014) extended those results to the larger class of bounded changes and relaxed the assumption on the time of the change which is, however, still restricted to growing slower than m. Furthermore, this work also provides the asymptotic distribution of the stopping time for the Page CUSUM procedure which allows the comparison with the ordinary CUSUM monitoring scheme and shows that the Page CUSUM is indeed superior for later changes as intended by construction. In Aue *et al.* (2009b), the limit distribution of the stopping time is derived for a CUSUM procedure when testing for a parameter change in a multiple time series regression model. The asymptotic behavior of the stopping time related to the procedure proposed in Hušková & Koubková (2005) for the detection of changes in linear models is analyzed in Černíková *et al.* (2013).

In Section 5.1 we generalize the results provided in Aue & Horváth (2004) and Fremdt (2014) for the CUSUM procedure to the class of sequential tests based on U-statistics for bounded changes assuming that the time of the change is sublinear in m. However, by the latter assumption, this result is still restricted to relatively early changes in the sense that the time of the change grows slower than m. To the best of our knowledge, even for the CUSUM kernel, the asymptotic behavior of the stopping time for later changes has not been assessed so far. We extend the existing literature by deriving the limit distribution of the stopping time for change points that are superlinear in m in Section 5.2 followed by linear changes in Section 5.3. In both cases, the analysis of the asymptotic behavior of the stopping time involves difficulties that do not occur for early changes but can be handled by conditioning on functionals of the historic observations and by letting the critical value increase to infinity.

In order to assess the asymptotic behavior of the stopping time τ_m we need to find normalizing sequences a_m and b_m such that we can derive the limit distribution of

$$\frac{\tau_m - a_m}{b_m}.$$

We follow the main idea of Aue & Horváth (2004) which is to establish a duality between the standardized stopping time and the monitoring statistic. This can be obtained by finding an appropriate sequence $N := N(m, x)$ such that

$$\lim_{m\to\infty} P\left(\frac{\tau_m - a_m}{b_m} \leq x\right) = 1 - \lim_{m\to\infty} P\left(\tau_m > N\right) \tag{5.1}$$

as the definition of τ_m in (1.2) implies

$$P\left(\tau_m > N\right) = P\left(\sup_{1\leq k\leq N} w(m,k)\left|\Gamma(m,k)\right| \leq c\right),$$

where c is the critical value. It follows

$$\lim_{m\to\infty} P\left(\frac{\tau_m - a_m}{b_m} \leq x\right) = 1 - \lim_{m\to\infty} P\left(\sup_{1\leq k\leq N} w(m,k)\left|\Gamma(m,k)\right| \leq c\right). \tag{5.2}$$

Hence, this duality enables us to derive the limit distribution of interest if $N = N(m, x)$ is chosen such that

$$1 - \lim_{m\to\infty} P\left(\sup_{1\leq k\leq N} w(m,k)\left|\Gamma(m,k)\right| \leq c\right) \tag{5.3}$$

defines a distribution function in x. To this end, it is often useful to center and scale $\sup_{1\leq k\leq N} w(m,k)\left|\Gamma(m,k)\right|$ in the following way:

$$\begin{aligned} &P\left(\sup_{1\leq k\leq N} w(m,k)\left|\Gamma(m,k)\right| \leq c\right) \\ =&P\left(e_m\left(\sup_{1\leq k\leq N} w(m,k)\left|\Gamma(m,k)\right| - d_m\right) \leq e_m\left(c - d_m\right)\right), \quad e_m > 0. \end{aligned}$$

The centering sequence d_m should capture the diverging part of the test statistic and e_m is needed to expand the centered test statistic in such a way that it converges in distribution. Let

$$\Psi(z) := \lim_{m\to\infty} P\left(e_m\left(\sup_{1\leq k\leq N} w(m,k)\left|\Gamma(m,k)\right| - d_m\right) \leq z\right).$$

Then, we get

$$\lim_{m\to\infty} P\left(\frac{\tau_m - a_m}{b_m} \leq x\right) = 1 - \Psi(-x)$$

if

$$e_m\left(c - d_m\right) \to -x \quad \text{as } m \to \infty. \tag{5.4}$$

An obvious choice of N fulfilling (5.1) is

$$N := N(m, x) = xb_m + a_m \tag{5.5}$$

as this yields

$$\begin{aligned} &P\left(\frac{\tau_m - a_m}{b_m} \leq x\right) = P\left(\tau_m \leq xb_m + a_m\right) \\ =&1 - P\left(\tau_m > xb_m + a_m\right) = 1 - P\left(\tau_m > N\right). \end{aligned} \tag{5.6}$$

However, based on this choice we do not necessarily obtain a distribution function in x in (5.3). In the following we consider the weight function as in (3.5). For sublinear changes in Section 5.1 we can build on existing literature and allow for arbitrary $\gamma \in [0, \frac{1}{2})$. In this case, N will not be chosen as in (5.5). As we are not aware of any results on later changes, we start with deriving the asymptotic distribution of the respective stopping times for $\gamma = 0$ which anyway mostly provides the best performance as can be seen in the simulation study in Chapter 6. An extension to $\gamma \in [0, \frac{1}{2})$ is probably possible using similar methods but beyond the scope of this work.

5.1. Sublinear Changes

In the following we derive the limit distribution of sequential change point procedures based on U-statistics for change-points that grow slower than m building on the work of Aue & Horváth (2004) and Fremdt (2014).

We consider the change-point model as described in Section 2.2 and the weight function as in (3.5). For ease of notation we use $g(m, k) := w(m, k)^{-1}$. We impose the following conditions on the time and the size of the change:

Assumption 5.1.

(i) $\Delta_m = O(1)$.

(ii) $\sqrt{m}|\Delta_m| \to \infty$.

(iii) There exists a $\lambda > 0$ *such that* $k^* = [\lambda m^\beta], 0 \leq \beta < 1$. *This can be divided into the following cases:*

- *(I)* $m^{\beta(1-\gamma)-1/2+\gamma}|\Delta_m| \to 0$,
- *(II)* $m^{\beta(1-\gamma)-1/2+\gamma}|\Delta_m| \to C_1\lambda^{\gamma-1} \in (0, \infty)$,
- *(III)* $m^{\beta(1-\gamma)-1/2+\gamma}|\Delta_m| \to \infty$.

By Assumption 5.1 (i) and (ii) we allow for fixed as well as local alternatives where the size of the change is allowed to decay to zero but slower than $\sqrt{m}$ such that it is detected with probability tending to one by Theorem 3.14. The division in part (iii) relates to how early the change appears in relation to the size of the change. We will see that the limit distribution depends on that. For the time series before the change we require Assumption 3.3 to be fulfilled. We impose the following conditions on the time series after the change:

Assumption 5.2. *Let $\{Y_i\}_{i\in\mathbb{Z}}$ and $\{Z_{i,m}\}_{i\in\mathbb{Z}}$ be stationary time series that fulfill the following assumptions for a given kernel function h.*

(i) $\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r_m^*(Y_i, Z_{j,m})\right|^2\right) \le u(m)(k_2-k_1+1)$ *for all* $m+k^*+1 \le k_1 \le k_2$ *with* $\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \to 0$ *and* γ *as in Assumption 3.2.*

(ii) $\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_{1,m}^*(Y_i)\right| = O_P(1)$.

(iii) For all $0 \le \alpha < \frac{1}{2}$ *the following Hajek-Renyi-type inequality holds*

$$\sup_{1\le l\le l_m} \frac{1}{m^{\frac{1}{2}-\alpha} l^{\alpha}} \left|\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right| = O_P(1) \quad \text{as } l_m \to \infty.$$

(iv) The following functional central limit theorem is satisfied for $k_m \to \infty$

$$\left\{\frac{1}{\sqrt{k_m}}\sum_{j=1}^{[k_m t]} (h_2(Y_{m+j}), h_{2,m}^*(Z_{m+k^*+j,m})) : 0 < t \le 1\right\} \xrightarrow{D} \{(W(t), W^*(t)) : 0 < t \le 1\},$$

where $\{(W(t), W^*(t)) : 0 < t \le T\}$ *is a bivariate Wiener process with mean zero and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma^2 & \tilde{\rho} \\ \tilde{\rho} & \sigma^{*2} \end{pmatrix}$$

with $\sigma^2 = \sum_{h\in\mathbb{Z}} Cov(h_2(Y_0), h_2(Y_h)), \sigma^{*2} = \sum_{h\in\mathbb{Z}} Cov(h_{2,m}^*(Z_{0,m}), h_{2,m}^*(Z_{h,m}))$.

The joint functional central limit theorem in (iv) is only needed for case (I) and (II) of Assumption 5.1 (iii), whereas for case (III) a central limit theorem for $h_2(Y_j)$ and assuming the supremum over the partial sum process of $h_{2,m}^*(Z_{j,m})$ to be stochastically bounded as in Assumption 5.15 (ii) and(iii) would be sufficient.

Recall the main approach that has been described at the beginning of this chapter. Based on the work of Aue & Horváth (2004) and Fremdt (2014), a_m and b_m will be chosen such that

$$\lim_{m\to\infty} P\left(\frac{a_m^{\gamma}}{1-\gamma}\cdot\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m} \le x\right) = \lim_{m\to\infty} P\left(\frac{\tau_m - a_m}{b_m} \le x\right). \tag{5.7}$$

Hence,

$$N := N(m,x) = \left(a_m^{1-\gamma} + x\frac{b_m}{a_m^{\gamma}}(1-\gamma)\right)^{\frac{1}{1-\gamma}} \tag{5.8}$$

satisfies (5.1). The diverging part of the monitoring statistic as given in (3.45) comes from the signal $|\Delta_m|$ and can be captured by $(N-k^*)|\Delta_m|$. Taking an asymptotic simplification of the weight function into account, d_m is defined by

$$d_m = \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}. \tag{5.9}$$

The scaling sequence is chosen such that it cancels out the asymptotic simplification of the weight function and induces the factor $\frac{1}{\sqrt{N}}$ for the functional central limit theorem:

$$e_m = \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}.$$

Having defined the centering and the scaling sequences, a_m and b_m are derived such that, with N as in (5.8), condition (5.4) is fulfilled. Fremdt (2014) showed that this is the case for the sequences that are used in the following theorem which states the asymptotic normality of the standardized stopping time. The variance of the limit distribution depends on the scenarios in Assumption 5.1 (iii).

Theorem 5.3. *Let Assumptions 3.3,5.1 and 5.2 be satisfied. Then, it holds under the alternative for all $x \in \mathbb{R}$*

$$\lim_{m\to\infty} P\left(\frac{\tau_m - a_m}{b_m} \leq x\right) = \Psi(x),$$

where Ψ is the distribution function of

$$Z \sim \begin{cases} N(0,\sigma^{*2}) & \textit{under } (I) \\ N(0,\delta_1\sigma^2 + (1-\delta_1)\sigma^{*2}) & \textit{under } (II) \textit{ with } \delta_1 = 1 - \frac{c}{C_1}\delta_1^{1-\gamma} \\ N(0,\sigma^2) & \textit{under } (III), \end{cases}$$

where $\sigma^2 = \sum_{h\in\mathbb{Z}} Cov(h_2(Y_0), h_2(Y_h)), \sigma^{*2} = \sum_{h\in\mathbb{Z}} Cov(h^*_{2,m}(Z_{0,m}), h^*_{2,m}(Z_{h,m}))$. The standardizing sequences are given as follows: a_m is the solution of

$$a_m = \left(\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|} + \frac{k^*}{a_m^\gamma}\right)^{\frac{1}{1-\gamma}}, \tag{5.10}$$

where $c := c_\alpha$ is the asymptotic critical value which is given by the $1-\alpha$ quantile of the limit distribution in Theorem 3.6,

$$b_m = \sqrt{a_m}|\Delta_m|^{-1}\left(1-\gamma\left(1-\frac{k^*}{a_m}\right)\right)^{-1}. \tag{5.11}$$

As indicated by the variances of the limit distribution for the different cases, the behavior of the stopping time is asymptotically dominated by the observations after the change in case (I) as the change occurs immediately after the monitoring has started. In case (III), the change occurs sufficiently late such that the observations before the change dominate asymptotically. Case (II) is the transition between those two states. Using the CUSUM kernel in the mean change model as in (2.8), we obtain

$$h_2^C(Y_i) = h_{2,m}^{*C}(Z_{i,m}) = \mu - Y_i \tag{5.12}$$

with h_2^C and $h_{2,m}^{*C}$ given in Example 3.1 and 3.11. Hence, the limit distribution is the same in all three cases wich conforms to Theorem 2.1 b) in Fremdt (2014).

In the following, we prove Theorem 5.3 stepwise. First, we consider the crucial relations between the sequences a_m and N and the size as well as the time of the change. According to (5.8) we obtain with a_m and b_m as in (5.10) and (5.11)

$$N = N(m,x) = \left(\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|} + \frac{k^*}{a_m^\gamma} + x \frac{a_m^{\frac{1}{2}-\gamma}(1-\gamma)}{|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))} \right)^{\frac{1}{1-\gamma}}, \quad x \in \mathbb{R}. \tag{5.13}$$

Lemma 5.4. *(cf. Fremdt (2014)[Lemma A.2]) Under Assumption 5.1 it holds*

a) *(i)* $\frac{a_m}{m} \to 0$

(ii) $\sqrt{a_m}|\Delta_m| \to \infty$

(iii) $\frac{k^*}{m} \to 0$

(iv) $\frac{k^*}{a_m} \to \begin{cases} 0 & \textit{under } (I), \\ \delta_1 \in (0,1) & \textit{under } (II) \textit{ with } \delta_1 = 1 - \frac{c}{C_1}\delta_1^{1-\gamma} \\ 1 & \textit{under } (III). \end{cases}$

for all $x \in \mathbb{R}$ as $m \to \infty$. In particular it holds for all cases that $\frac{k^*}{a_m} = O(1)$.

b) $\frac{N}{a_m} \to 1$ *such that part a) is still valid when replacing a_m by N.*

c) $\lim_{m\to\infty} \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \left(c - \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma} \right) = -x$ *for all $x \in \mathbb{R}$.*

Proof. The following proof is very similar to the proof of Lemma A.2 in Fremdt (2014). First, note that the definition of a_m is equivalent to

$$a_m = \frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|} a_m^\gamma + k^* \tag{5.14}$$

and similarly to Proposition A.1 in Fremdt (2014) it holds

$$a_m = (1+o(1)) \begin{cases} \left(\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|}\right)^{\frac{1}{1-\gamma}} & \text{under } (I), \\ \delta_2 k^* & \text{under } (II) \text{ with } \delta_2 = \left(\frac{c}{C_1} + \delta_1^\gamma\right)^{\frac{1}{1-\gamma}}, \\ k^* & \text{under } (III) \end{cases} \tag{5.15}$$

which can be proven based on (5.14) as follows. Under (I), we obtain

$$a_m = \frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|} a_m^\gamma \left(1 + \frac{|\Delta_m| {k^*}^{1-\gamma}}{cm^{\frac{1}{2}-\gamma}} \left(\frac{k^*}{a_m}\right)^\gamma \right) = \frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|} a_m^\gamma (1+o(1))$$

as $a_m \geq k^*$ and thus

$$a_m = \left(\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|}\right)^{\frac{1}{1-\gamma}} (1+o(1)).$$

Under (II), note that it holds

$$\frac{a_m}{k^*} = \frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|{k^*}^{1-\gamma}} \left(\frac{a_m}{k^*}\right)^\gamma + 1 = \frac{c}{C_1}\left(\frac{a_m}{k^*}\right)^\gamma + 1 + o(1).$$

Consider the function $f(x) = x - \frac{c}{C_1}x^\gamma - 1$ which satisfies $f(1) = -\frac{c}{C_1} < 0$ and $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} x(1 - \frac{c}{C_1}x^{\gamma-1} - x^{-1}) = \infty$. Furthermore, the derivative shows that f takes its minimum at $\left(\frac{C_1}{\gamma c}\right)^{\frac{1}{1-\gamma}} > 0$ and is strictly decreasing before that point and strictly increasing after. Hence, as f is continuous, there exists a unique $\delta_1^{-1} \in (1,\infty)$ such that $f(\delta_1^{-1}) = 0$. We conclude that $\lim_{m\to\infty} \frac{a_m}{k^*} = \delta_1^{-1}$ and thus

$$\lim_{m\to\infty} \frac{k^*}{a_m} = \delta_1 \in (0,1), \tag{5.16}$$

where δ_1 is the unique solution of $\delta_1 = 1 - \frac{c}{C_1}\delta_1^{1-\gamma}$. Now, with (5.14), we obtain

$$\begin{aligned} a_m &= k^*\left(c|\Delta_m|^{-1}m^{\frac{1}{2}-\gamma}{k^*}^{\gamma-1} + \left(\frac{k^*}{a_m}\right)^\gamma\right)^{\frac{1}{1-\gamma}} = k^*\left(\frac{c}{C_1} + \delta_1^\gamma\right)^{\frac{1}{1-\gamma}}(1+o(1)) \\ &= \delta_2 k^*(1+o(1)) \end{aligned}$$

with $\delta_2 = \left(\frac{c}{C_1} + \delta_1^\gamma\right)^{\frac{1}{1-\gamma}}$. Under (III), (5.10) yields

$$\frac{a_m}{k^*} = \left(\left(\frac{k^*}{a_m}\right)^\gamma + c|\Delta_m|^{-1}m^{\frac{1}{2}-\gamma}{k^*}^{\gamma-1}\right)^{\frac{1}{1-\gamma}} = O(1)$$

which implies

$$a_m^{1-\gamma} = \frac{k^*}{a_m^\gamma}\left(1 + \frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|{k^*}^{1-\gamma}}\left(\frac{k^*}{a_m}\right)^\gamma\right) = \frac{k^*}{a_m^\gamma}(1+o(1)).$$

a) First, it is to mention that (iii) follows directly with $\beta < 1$ in Assumption 5.1 (iii).

(i) Under (I), we get with Assumption 5.1 (ii) and (5.15)

$$\frac{a_m}{m} = \left(\frac{c}{\sqrt{m}|\Delta_m|}\right)^{\frac{1}{1-\gamma}}(1+o(1)) \to 0.$$

For (II) and (III), the assertion follows with (iii) and (5.15) as

$$\frac{a_m}{m} = (1+o(1))\begin{cases} \delta_2\frac{k^*}{m}, & \text{under } (II), \\ \frac{k^*}{m}, & \text{under } (III). \end{cases}$$

(ii) As $a_m \geq \left(c|\Delta_m|^{-1}m^{\frac{1}{2}-\gamma}\right)^{\frac{1}{1-\gamma}}$ Assumption 5.1 (ii) yields

$$\sqrt{a_m}|\Delta_m| \geq \left(c|\Delta_m|^{-1}m^{\frac{1}{2}-\gamma}|\Delta_m|^{2(1-\gamma)}\right)^{\frac{1}{2(1-\gamma)}} = c^{\frac{1}{2(1-\gamma)}}(\sqrt{m}|\Delta_m|)^{\frac{\frac{1}{2}-\gamma}{2(1-\gamma)}} \to \infty.$$

(iv) Under (I) we obtain with (5.15)

$$\frac{a_m}{k^*} = (1+o(1))\left(cm^{\frac{1}{2}-\gamma}{k^*}^{\gamma-1}|\Delta_m|^{-1}\right)^{\frac{1}{1-\gamma}} \to \infty.$$

For (II) the assertion follows directly with (5.15) and for (III) it is obtained by (5.16).

b) With (5.8) and (5.11) we obtain

$$\frac{N^{1-\gamma}}{a_m^{1-\gamma}} = 1 + x\frac{b_m}{a_m}(1-\gamma) = 1 + x\frac{1-\gamma}{\sqrt{a_m}|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))}. \tag{5.17}$$

It holds

$$1-\gamma\left(1-\frac{k^*}{a_m}\right) \to \begin{cases} 1-\gamma > 0 & \text{under } (I),\\ 1-\gamma(1-\delta_1) > 0 & \text{under } (II),\\ 1 & \text{under } (III) \end{cases} \tag{5.18}$$

such that it follows with a) (ii)

$$x\frac{1-\gamma}{\sqrt{a_m}|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))} \to 0 \quad \text{as } m \to \infty. \tag{5.19}$$

Hence, we get $\left(\frac{N}{a_m}\right)^{1-\gamma} \to 1$ which implies $\frac{N}{a_m} \to 1$.

c)

$$\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|} + \frac{k^*}{a_m^\gamma} + x\frac{a_m^{\frac{1}{2}-\gamma}(1-\gamma)}{|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))}$$

First, we replace N in $N^{1-\gamma}$ by (5.13) such that we obtain

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(c - \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\\
=&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(c - |\Delta_m|m^{\gamma-\frac{1}{2}}\left(N^{1-\gamma} - \frac{k^*}{N^\gamma}\right)\right)\\
=&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(|\Delta_m|m^{\gamma-\frac{1}{2}}\left(\frac{k^*}{N^\gamma} - \frac{k^*}{a_m^\gamma}\right) - x\left(\frac{a_m}{m}\right)^{\frac{1}{2}-\gamma}\frac{1-\gamma}{1-\gamma(1-\frac{k^*}{a_m})}\right)\\
=&\frac{|\Delta_m|k^*}{\sqrt{N}}\left(1-\left(\frac{N}{a_m}\right)^\gamma\right) - x\left(\frac{a_m}{N}\right)^{\frac{1}{2}-\gamma}\frac{1-\gamma}{1-\gamma(1-\frac{k^*}{a_m})}\\
=&\frac{|\Delta_m|k^*}{\sqrt{N}}\left(1^{\frac{\gamma}{1-\gamma}} - \left(\left(\frac{N}{a_m}\right)^{1-\gamma}\right)^{\frac{\gamma}{1-\gamma}}\right) - x\left(\frac{a_m}{N}\right)^{\frac{1}{2}-\gamma}\frac{1-\gamma}{1-\gamma(1-\frac{k^*}{a_m})}.
\end{aligned} \tag{5.20}$$

By the mean value theorem there exists a $\xi_m \in (1, \frac{N}{a_m}^{1-\gamma})$ with $\xi_m \to 1$ due to b) such that

$$
\begin{aligned}
&1^{\frac{\gamma}{1-\gamma}} - \left(\left(\frac{N}{a_m}\right)^{1-\gamma}\right)\frac{\gamma}{1-\gamma} = \frac{\gamma}{1-\gamma}\xi_m^{\frac{2\gamma-1}{1-\gamma}}\left(1-\left(\frac{N}{a_m}\right)^{1-\gamma}\right)\\
=&-\frac{\gamma}{1-\gamma}\xi_m^{\frac{2\gamma-1}{1-\gamma}}x\frac{b_m}{a_m}(1-\gamma)\\
=&-\xi_m^{\frac{2\gamma-1}{1-\gamma}}x\frac{1}{\sqrt{a_m}|\Delta_m|}\frac{1-\gamma}{1-\gamma(1-\frac{k^*}{a_m})}
\end{aligned}
$$

by (5.8). Hence, with (5.20), we get

$$
\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(c-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\\
=&-x\left(\frac{a_m}{N}\right)^{\frac{1}{2}-\gamma}\left(1+\frac{\gamma}{1-\gamma}\frac{k^*}{a_m}\frac{a_m^{\gamma}}{N^{\gamma}}\xi_m^{\frac{2\gamma-1}{1-\gamma}}\right)\frac{1-\gamma}{1-\gamma(1-\frac{k^*}{a_m})}\\
=&-x\left(\frac{a_m}{N}\right)^{\frac{1}{2}-\gamma}\left(1+\frac{\gamma}{1-\gamma}\frac{k^*}{a_m}\frac{a_m^{\gamma}}{N^{\gamma}}\xi_m^{\frac{2\gamma-1}{1-\gamma}}\right)\left(1-\frac{\gamma}{1-\gamma}\frac{k^*}{a_m}\right)^{-1} \to -x \quad \text{as } m\to\infty.
\end{aligned}
$$

□

The following Lemma shows that the behavior of the monitoring statistic before the change is negligible for the asymptotic distribution of the stopping time. In Section 5.2 we will see that this is not the case for later changes which is one of the reasons why the analysis of the stopping time for $\beta < 1$ cannot be extended straightforwardly to $\beta \geq 1$.

Lemma 5.5. *Let Assumption 3.3 and 5.1 be fulfilled. Then, it holds*

$$
\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1\le k\le k^*}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\xrightarrow{P} -\infty \quad \textit{as } m\to\infty.
$$

Proof. For $k < k^*$ it holds with (3.3)

$$
\Gamma(m,k) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k} r(Y_i,Y_j) + \sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i). \tag{5.21}
$$

With Lemma 3.5 we get

$$
\sup_{1\le k\le k^*}\frac{1}{g(m,k)}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k} r(Y_i,Y_j)\right| \le \sup_{k\ge 1}\frac{1}{g(m,k)}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k} r(Y_i,Y_j)\right| = o_P(1) \tag{5.22}
$$

as $m \to \infty$. With Lemma 5.4 a) (iii) and Assumption 3.3 (iii) it holds for $\gamma < \alpha < \frac{1}{2}$ and m large enough such that $k^* \leq m$

$$
\begin{aligned}
&\sup_{1\leq k\leq k^*} \frac{1}{g(m,k)} \left| \sum_{j=m+1}^{m+k} h_2(Y_j) \right| \\
\leq &\sup_{1\leq k\leq k^*} \frac{m^{\frac{1}{2}-\alpha}k^\alpha}{g(m,k)} \sup_{1\leq k\leq m} \frac{1}{m^{\frac{1}{2}-\alpha}k^\alpha} \left| \sum_{j=m+1}^{m+k} h_2(Y_j) \right| \\
= &\sup_{1\leq k\leq k^*} \frac{m^{\frac{1}{2}-\alpha}k^\alpha}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma} O_P(1) \\
= &\sup_{1\leq k\leq k^*} \left(\frac{k}{m}\right)^{\alpha-\gamma} \frac{\left(\frac{k}{m}\right)^\gamma}{\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma} O_P(1) \\
= &\sup_{1\leq k\leq k^*} \left(\frac{k}{m}\right)^{\alpha-\gamma} \left(1+\frac{k}{m}\right)^{\gamma-1} O_P(1) \\
\leq &\sup_{1\leq k\leq k^*} \left(\frac{k}{m}\right)^{\alpha-\gamma} O_P(1) \\
= &\left(\frac{k^*}{m}\right)^{\alpha-\gamma} O_P(1) = o_P(1) \quad \text{as } m \to \infty.
\end{aligned} \tag{5.23}
$$

Assumption 3.3 (ii) yields

$$
\begin{aligned}
&\sup_{1\leq k\leq k^*} \frac{1}{g(m,k)} \left| \frac{k}{m} \sum_{i=1}^{m} h_1(Y_i) \right| \\
= &\sup_{1\leq k<k^*} \frac{\sqrt{m}\frac{k}{m}}{g(m,k)} O_P(1) \\
= &\sup_{1\leq k<k^*} \left(\frac{k}{m}\right)^{1-\gamma} \frac{\left(\frac{k}{m}\right)^\gamma}{\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma} O_P(1) \\
= &\sup_{1\leq k<k^*} \left(\frac{k}{m}\right)^{1-\gamma} \left(1+\frac{k}{m}\right)^{\gamma-1} O_P(1) \\
= &\left(\frac{k^*}{m}\right)^{1-\gamma} O_P(1) = o_P(1) \quad \text{as } m \to \infty
\end{aligned} \tag{5.24}
$$

with Lemma 5.4 a) (iii). By combining (5.21),(5.22),(5.23) and (5.24) we obtain

$$
\sup_{1\leq k\leq k^*} \frac{|\Gamma(m,k)|}{g(m,k)} = o_P(1) \quad \text{as } m \to \infty. \tag{5.25}
$$

As shown in the proof of Lemma A.3 in Fremdt (2014) the deterministic part diverges, i.e.

$$
\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \frac{|\Delta_m|(N-k^*)}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma} \to \infty \quad \text{as } m \to \infty \tag{5.26}
$$

such that the assertion follows together with (5.25).

□

The parts of the monitoring statistic which are asymptotically relevant are identified in the following. It should be noted that, in this context, not only the remainder term but also the terms which involve the historic observations are asymptotically negligible which is, again, different for later changes and requires further adaption of the approach.

Lemma 5.6. *Let Assumption 5.1 be satisfied as well as Assumption 5.2 (i) and (ii). Then, it holds*

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{\left|\Gamma(m,k) - \left(\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m}) + (k-k^*)\Delta_m\right)\right|}{g(m,k)}$$
$$= o_P(1) \quad \textit{as } m\to\infty.$$

Proof. Using the representation of the test statistic as given in (3.45) we obtain

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{\left|\Gamma(m,k) - \left(\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m}) + (k-k^*)\Delta_m\right)\right|}{g(m,k)}\\
\leq& \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*} r(Y_i,Y_j)\right|\\
&+ \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left|\frac{k^*}{m}\sum_{i=1}^{m} h_1(Y_i)\right|\\
&+ \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k} r^*_m(Y_i,Z_{j,m})\right|\\
&+ \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left|\frac{k-k^*}{m}\sum_{i=1}^{m} h^*_{1,m}(Y_i)\right|. \qquad (5.27)
\end{aligned}$$

It holds

$$\begin{aligned}
\sup_{k^*<k\leq N} \frac{\left(\frac{k}{m}\right)^\gamma}{g(m,k)} &= \frac{1}{\sqrt{m}} \sup_{k^*<k\leq N} \frac{\left(\frac{k}{m}\right)^\gamma}{\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma}\\
&= \frac{1}{\sqrt{m}} \sup_{k^*<k\leq N} \left(1+\frac{k}{m}\right)^{\gamma-1}\\
&= \frac{1}{\sqrt{m}} \left(1+\frac{k^*}{m}\right)^{\gamma-1} = O\left(\frac{1}{\sqrt{m}}\right) \quad \text{as } m\to\infty \qquad (5.28)
\end{aligned}$$

with Lemma 5.4 a) (iii). In particular, this implies

$$\sup_{k^*<k\leq N} \frac{1}{g(m,k)} \leq \left(\frac{k^*}{m}\right)^{-\gamma} \sup_{k^*<k\leq N} \frac{\left(\frac{k}{m}\right)^\gamma}{g(m,k)} \leq \frac{1}{\sqrt{m}} \left(\frac{k^*}{m}\right)^{-\gamma} O(1) \quad \text{as } m\to\infty. \quad (5.29)$$

With Assumption 3.3 (i) and Markov's inequality we obtain

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{k^*}m}\left|\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*} r(Y_i,Y_j)\right| \geq \epsilon\right) = P\left(\left|\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*} r(Y_i,Y_j)\right| \geq \sqrt{k^*}m\epsilon\right)\\
\leq& \frac{u(m)k^*}{k^*m^2\epsilon^2} = \frac{1}{\epsilon^2}\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \frac{1}{m^{2\gamma}\log(m)^2} \to 0 \quad \text{as } m\to\infty
\end{aligned}$$

for all $\epsilon > 0$ and thus

$$\left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=m+1}^{m+k^*} r(Y_i, Y_j) \right| = o_P \left(\sqrt{k^*} \right) \quad \text{as } m \to \infty. \tag{5.30}$$

This yields

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=m+1}^{m+k^*} r(Y_i, Y_j) \right| \\
=&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{\sqrt{k^*}}{g(m,k)} o_P(1) \\
=&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \left(\frac{k^*}{m}\right)^{\frac{1}{2}-\gamma} o_P(1) \\
=&\left(\frac{k^*}{N}\right)^{\frac{1}{2}-\gamma} o_P(1) = o_P(1) \quad \text{as } m \to \infty
\end{aligned} \tag{5.31}$$

with Lemma 5.4 a) (iv) and b). Due to Assumption 3.3 (ii) it holds

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) = O_P(1).$$

Hence, we get

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left| \frac{k^*}{m} \sum_{i=1}^{m} h_1(Y_i) \right| = \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{k^*}{\sqrt{m} g(m,k)} O_P(1) \\
\leq&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \left(\frac{k^*}{m}\right)^{1-\gamma} O_P(1) \\
=&O_P\left(\sqrt{\frac{k^*}{m}} \left(\frac{k^*}{N}\right)^{\frac{1}{2}-\gamma} \right) = o_P(1) \quad \text{as } m \to \infty
\end{aligned}$$

with (5.29) and Lemma 5.4 a) (iii), (iv) and b). We obtain with Lemma 3.4 for m large enough such that $N \leq m$

$$\begin{aligned}
&P\left(\frac{1}{m^{-\gamma} N^{\frac{1}{2}}} \sup_{k^*<k\leq N} \frac{1}{m} \left| \sum_{i=1}^{m} \sum_{j=m+k^*+1}^{m+k} r_m^*(Y_i, Z_{j,m}) \right| > \epsilon \right) \\
\leq&P\left(\frac{1}{m^{-\gamma} N^{\frac{1}{2}}} \sup_{1\leq k\leq N} \frac{1}{m} \left| \sum_{i=1}^{m} \sum_{j=1}^{k} r_m^*(Y_i, Z_{m+k^*+j,m}) \right| > \epsilon \right) \\
\leq&\frac{1}{\epsilon^2} \frac{u(m)}{m^{2-2\gamma}} (\log_2(2N))^2 = \frac{1}{\epsilon^2} \frac{u(m)}{m^{2-2\gamma}} \log(m)^2 O(1) = o(1) \quad \text{as } m \to \infty
\end{aligned} \tag{5.32}$$

for any $\epsilon > 0$. Hence, it follows

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left| \frac{1}{m} \sum_{i=1}^{m} \sum_{j=m+k^*+1}^{m+k} r_m^*(Y_i, Z_{j,m}) \right| \\
&\leq \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} N^{\frac{1}{2}} m^{-\gamma} m^{-\frac{1}{2}} \left(\frac{k^*}{m}\right)^{-\gamma} o_P(1) \\
&= \left(\frac{N}{m}\right)^{\gamma} \frac{1}{k^{*\gamma}} o_P(1) = o_P(1) \quad \text{as } m \to \infty. \qquad (5.33)
\end{aligned}$$

It holds

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left| \frac{k-k^*}{m} \sum_{i=1}^{m} h_{1,m}^*(Y_i) \right| \\
&\leq \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{1}{g(m,k)} \left| \frac{k}{m} \sum_{i=1}^{m} h_{1,m}^*(Y_i) \right| \\
&= \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{k}{\sqrt{m} g(m,k)} O_P(1) \\
&\leq \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \frac{N^{1-\gamma}}{m^{\frac{1}{2}-\gamma}} \sup_{k^*<k\leq N} \frac{\left(\frac{k}{m}\right)^{\gamma}}{g(m,k)} O_P(1) \\
&= \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \frac{N^{1-\gamma}}{m^{\frac{1}{2}-\gamma}} \frac{1}{\sqrt{m}} O_P(1) \\
&= \left(\frac{N}{m}\right)^{\frac{1}{2}} O_P(1) = o_P(1) \quad \text{as } m \to \infty
\end{aligned}$$

with (5.28), (5.29) and Lemma 5.4 a) (i),(iii), (iv) and b). The assertion now follows as we showed that all the summands in (5.27) converge to zero in probability as m tends to infinity. □

The following Lemma shows that for the remaining considerations the weight function can asymptotically be replaced by a simpler term.

Lemma 5.7. *Given that Assumption 5.1 and 5.2 (iii) and (iv) are fulfilled, it holds*

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \left| \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m}) + (k-k^*)\Delta_m}{g(m,k)} \right| \\
&= \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \left| \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m}) + (k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}} \right| + o_p(1).
\end{aligned}$$

Proof. The reverse triangle inequality yields

$$\begin{aligned}
&\Bigg|\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\left|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m}) + (k-k^*)\Delta_m}{g(m,k)}\right| \\
&- \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\left|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m}) + (k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\right|\Bigg| \\
\leq &\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\Bigg|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m}) + (k-k^*)\Delta_m}{g(m,k)} \\
&- \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m}) + (k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\Bigg| \\
\leq &\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\Bigg|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})}{g(m,k)} \\
&- \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\Bigg| \\
&+ \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\left|\frac{(k-k^*)\Delta_m}{g(m,k)} - \frac{(k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\right|.
\end{aligned} \tag{5.34}$$

For the first summand it holds

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\Bigg|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})}{g(m,k)} \\
&- \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\Bigg| \\
\leq &\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\left|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\right| \sup_{k^*<k\leq N}\left|\frac{1}{\left(1+\frac{k}{m}\right)^{1-\gamma}} - 1\right| \\
= &\left(1-\left(\frac{1}{1+\frac{N}{m}}\right)^{1-\gamma}\right)\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\left|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\right| \\
=&o(1)\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N}\left|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\right|.
\end{aligned} \tag{5.35}$$

Note that with (5.26) we get

$$|\Delta_m| N^{-\frac{1}{2}}(N-k^*) \to \infty \quad \text{as } m \to \infty. \tag{5.36}$$

As $\Delta_m = O(1)$, this implies

$$N - k^* \to \infty \quad \text{as } m \to \infty. \tag{5.37}$$

Assumption 5.2 (iii) and (iv) as well as Lemma 5.4 a) (iv) and b) yield

$$
\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})\right|}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}} \\
\leq &\left(\frac{k^*}{N}\right)^{\frac{1}{2}-\gamma} \left|\frac{1}{\sqrt{k^*}} \sum_{j=m+1}^{m+k^*} h_2(Y_j)\right| + \sup_{1\leq l\leq N-k^*} N^{\gamma-\frac{1}{2}}(l+k^*)^{-\gamma} \left|\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})\right| \\
\leq &\left(\frac{k^*}{N}\right)^{\frac{1}{2}-\gamma} \left|\frac{1}{\sqrt{k^*}} \sum_{j=m+1}^{m+k^*} h_2(Y_j)\right| + \sup_{1\leq l\leq N-k^*} \left(1-\frac{N}{k^*}\right)^{\frac{1}{2}-\gamma} (N-k^*)^{\gamma-\frac{1}{2}} l^{-\gamma} \left|\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})\right| \\
= &O_P(1) \quad \text{as } m\to\infty.
\end{aligned}
$$

Hence, with (5.35), we get

$$
\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \left| \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})}{g(m,k)} \right. \\
&\left. - \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}} \right| = o_p(1) \quad \text{as } m\to\infty.
\end{aligned}
$$

For the second summand we obtain

$$
\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \left| \frac{(k-k^*)\Delta_m}{g(m,k)} - \frac{(k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}} \right| \\
= &\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \left| \frac{(k-k^*)\Delta_m}{\sqrt{m}\left(1+\frac{k}{m}\right)^{1-\gamma}\left(\frac{k}{m}\right)^{\gamma}} - \frac{(k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}} \right| \\
= &|\Delta_m| N^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} (k-k^*)k^{-\gamma} \left| \left(\frac{1}{1+\frac{k}{m}}\right)^{1-\gamma} - 1 \right| \\
\leq &|\Delta_m| N^{\gamma-\frac{1}{2}} \sup_{k^*<k\leq N} \frac{k-k^*}{k^{\gamma}} \left(1-\left(\frac{1}{1+\frac{N}{m}}\right)^{1-\gamma}\right) \\
= &|\Delta_m| N^{\gamma-\frac{1}{2}} \frac{N-k^*}{N^{\gamma}} \left(1-\left(1+\frac{N}{m}\right)^{\gamma-1}\right) \\
= &|\Delta_m| N^{-\frac{1}{2}} (N-k^*) \left(1-\left(1+\frac{N}{m}\right)^{\gamma-1}\right)
\end{aligned}
\tag{5.38}
$$

as $\frac{k-k^*}{k^{\gamma}}$ is strictly increasing in k. Consider the function f with $f(x) = (1+x)^{\gamma-1}$. By the mean value theorem, there exists a $z_m \in (0, \frac{N}{m})$ such that

$$
1-\left(1+\frac{N}{m}\right)^{\gamma-1} = \frac{N}{m}(1-\gamma)(1+z_m)^{\gamma-2} = \frac{N}{m}O(1) \quad \text{as } m\to\infty. \tag{5.39}
$$

Furthermore, the definition of N given in (5.13) yields

$$\frac{|\Delta_m|\sqrt{N}(N-k^*)}{m} = \frac{|\Delta_m|N^{\frac{1}{2}+\gamma}}{m}\left(N^{1-\gamma} - \frac{k^*}{N^\gamma}\right)$$
$$=\frac{|\Delta_m|N^{\frac{1}{2}+\gamma}}{m}\left(\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|} + x\frac{a_m^{\frac{1}{2}-\gamma}(1-\gamma)}{|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))} + \frac{k^*}{a_m^\gamma} - \frac{k^*}{N^\gamma}\right)$$
$$=c\left(\frac{N}{m}\right)^{\frac{1}{2}+\gamma} + x\frac{N}{m}\left(\frac{a_m}{N}\right)^{\frac{1}{2}-\gamma}\frac{1-\gamma}{1-\gamma(1-\frac{k^*}{a_m})} + |\Delta_m|\sqrt{N}\frac{k^*}{m}\left(\left(\frac{N}{a_m}\right)^\gamma - 1\right). \quad (5.40)$$

Consider the function $f(x) = x^{\frac{\gamma}{1-\gamma}}$ with $f'(x) = \frac{\gamma}{1-\gamma}x^{\frac{-1+2\gamma}{1-\gamma}}$. By the mean value theorem there exists a $z_m \in \left\{\min\left(1, \left(\frac{N}{a_m}\right)^{1-\gamma}\right), \max\left(1, \left(\frac{N}{a_m}\right)^{1-\gamma}\right)\right\}$ such that

$$\left(\frac{N}{a_m}\right)^\gamma - 1 = \left(\left(\frac{N}{a_m}\right)^{1-\gamma}\right)^{\frac{\gamma}{1-\gamma}} - 1 = \frac{\gamma}{1-\gamma}z_m^{\frac{-1+2\gamma}{1-\gamma}}\left(\left(\frac{N}{a_m}\right)^{1-\gamma} - 1\right)$$
$$=\left(\left(\frac{N}{a_m}\right)^{1-\gamma} - 1\right)O(1) = \frac{1}{\sqrt{a_m}|\Delta_m|}O(1)$$

by Lemma 5.4 b), (5.17) and (5.18). Hence, (5.40) yields

$$\frac{|\Delta_m|\sqrt{N}(N-k^*)}{m} = c\left(\frac{N}{m}\right)^{\frac{1}{2}+\gamma} + x\frac{N}{m}\left(\frac{a_m}{N}\right)^{\frac{1}{2}-\gamma}\frac{1-\gamma}{1-\gamma(1-\frac{k^*}{a_m})} + \sqrt{\frac{N}{a_m}}\frac{k^*}{m}O(1)$$
$$=o(1) \quad \text{as } m \to \infty \quad (5.41)$$

with Lemma 5.4 a) (i) and b) and (5.18). Now, combining (5.38), (5.39) and (5.41) we obtain

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{k^*<k\le N}\left|\frac{(k-k^*)\Delta_m}{g(m,k)} - \frac{(k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^\gamma}\right|$$
$$\le|\Delta_m|N^{-\frac{1}{2}}(N-k^*)\left(1-\left(1+\frac{N}{m}\right)^{\gamma-1}\right)$$
$$=\frac{|\Delta_m|\sqrt{N}(N-k^*)}{m}O(1) = o(1) \quad \text{as } m \to \infty.$$

The assertion follows as both summands in (5.34) converge to zero as m tends to infinity. □

In order to assess the supremum over k on $\{k^* < k \le N\}$ we shift the index and consider the supremum over l on $\{1 < l \le N - k^*\}$ which we split at $(1-\delta)(N-k^*)$ for a fixed $\delta \in (0, 1)$.

Lemma 5.8. *Let Assumption 5.1 and 5.2 (iii) be satisfied. Then, for $\delta \in (0,1)$ fixed, it holds*

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \left(\sup_{1<l<(1-\delta)(N-k^*)} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}} \right.$$
$$\left. - \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}} \right) \xrightarrow{P} -\infty$$

as $m \to \infty$.

Proof. We get with Assumption 5.2 (iv) that

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{1\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}} \le \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \frac{\left|\frac{1}{\sqrt{k^*}}\sum_{j=m+1}^{m+k^*} h_2(Y_j)\right|}{\sqrt{\frac{m}{k^*}}\left(\frac{k^*}{m}\right)^{\gamma}}$$
$$= \left(\frac{k^*}{N}\right)^{\frac{1}{2}-\gamma} O_P(1) = O_P(1) \quad \text{as } m \to \infty. \tag{5.42}$$

Furthermore, choosing $\gamma < \alpha < \frac{1}{2}$ in Assumption 5.2 (iii) yields with (5.37)

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{1\le l\le N-k^*} \frac{\left|\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}} \\
\le &\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{1\le l\le N-k^*} \frac{N^{\frac{1}{2}-\alpha} l^{\alpha}}{m^{\frac{1}{2}-\gamma}(l+k^*)^{\gamma}} O_P(1) \\
= &O_P(1) N^{\gamma-\alpha} \sup_{1\le l\le N-k^*} l^{\alpha-\gamma} \sup_{1\le l\le N-k^*} \left(\frac{l}{l+k^*}\right)^{\gamma} \\
= &O_P(1) \left(\frac{N-k^*}{N}\right)^{\alpha-\gamma} \sup_{1\le l\le N-k^*} \left(\frac{1}{1+\frac{k^*}{l}}\right)^{\gamma} \\
= &O_P(1) \left(\frac{N-k^*}{N}\right)^{\alpha-\gamma} \left(\frac{1}{1+\frac{k^*}{N-k^*}}\right)^{\gamma} \\
= &O_P(1) \left(1 - \frac{k^*}{N}\right)^{\alpha} = O_P(1) \quad \text{as } m \to \infty
\end{aligned} \tag{5.43}$$

as $1 - \frac{k^*}{N} = O(1)$ with Lemma 5.4 a) (iv) and b). Additionally, it holds

$$\sup_{1<l<(1-\delta)(N-k^*)} \frac{l|\Delta_m|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}} = |\Delta_m| m^{\gamma-\frac{1}{2}} \sup_{1<l<(1-\delta)(N-k^*)} \frac{l}{(l+k^*)^{\gamma}}$$
$$= |\Delta_m| m^{\gamma-\frac{1}{2}} \sup_{1<l<(1-\delta)(N-k^*)} l^{1-\gamma} \left(\frac{1}{1+\frac{k^*}{l}}\right)^{\gamma} = |\Delta_m| m^{\gamma-\frac{1}{2}} \frac{(1-\delta)(N-k^*)}{((1-\delta)N+\delta k^*)^{\gamma}}. \tag{5.44}$$

Now, by combining (5.42), (5.43), (5.44) we obtain

$$
\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1<l<(1-\delta)(N-k^*)}\frac{\left|\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h^*_{2,m}(Z_{j,m})+l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right.\\
&\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\\
\leq&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1\leq l\leq N-k^*}\frac{\left|\sum_{j=m+1}^{m+k^*}h_2(Y_j)\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}+\sup_{1\leq l\leq N-k^*}\frac{\left|\sum_{j=m+k^*+1}^{m+k^*+l}h^*_{2,m}(Z_{j,m})\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right.\\
&\left.+\sup_{1<l<(1-\delta)(N-k^*)}\frac{l|\Delta_m|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\\
=&O_P(1)+|\Delta_m|N^{-\frac{1}{2}}(N-k^*)\left((1-\delta)\left(\frac{1}{1-\delta\left(1-\frac{k^*}{N}\right)}\right)^\gamma-1\right).
\end{aligned}
$$

With Lemma 5.4 a) (iv) and b) we get

$$
(1-\delta)\left(\frac{1}{1-\delta\left(1-\frac{k^*}{N}\right)}\right)^\gamma\to\begin{cases}(1-\delta)^{1-\gamma} & \text{under } (I),\\ \frac{1-\delta}{(1-\delta(1-\delta_1))^\gamma}\in(0,1) & \text{under } (II),\\ 1-\delta & \text{under } (III).\end{cases}
$$

which is smaller than 1 in all cases, such that the assertion follows with (5.36). □

Lemma 5.9. *Let Assumption 5.1 and 5.2 (iii), (iv) be fulfilled. Then, it holds for all* $z\in\mathbb{R}$

$$
\begin{aligned}
P&\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{k^*<k\leq N}\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k}h^*_{2,m}(Z_{j,m})+(k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^\gamma}\right.\right.\\
&\left.\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\leq z\right)\to\Psi(z)\quad \text{as } m\to\infty,
\end{aligned}
$$

where Ψ *is the distribution function of*

$$
Z\sim\begin{cases}N(0,\sigma^{*2}) & \text{under } (I)\\ N(0,\delta_1\sigma^2+(1-\delta_1)\sigma^{*2}) & \text{under } (II)\\ N(0,\sigma^2) & \text{under } (III).\end{cases}
$$

Proof of Lemma 5.9. It holds for any $z \in \mathbb{R}, \delta > 0$

$$
\begin{aligned}
&P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{k^*<k\leq N}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})+(k-k^*)\Delta_m\right|}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\right.\right.\\
&\qquad\left.\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\leq z\right)\\
&=P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1<l\leq N-k^*}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})+l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right.\right.\\
&\qquad\left.\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\leq z\right)\\
&=P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1<l<(1-\delta)(N-k^*)}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})+l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right.\right.\\
&\qquad\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\leq z,\\
&\qquad\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})+l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right.\\
&\qquad\left.\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\leq z\right)
\end{aligned}
$$

such that we get with Lemma 5.8 and Lemma B.3 (i)

$$
\begin{aligned}
&\lim_{m\to\infty}P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{k^*<k\leq N}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k} h_{2,m}^*(Z_{j,m})+(k-k^*)\Delta_m\right|}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\right.\right.\\
&\qquad\left.\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\leq z\right)\\
&=\lim_{m\to\infty}P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})+l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right.\right.\\
&\qquad\left.\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\leq z\right).
\end{aligned}
\tag{5.45}
$$

We get with Assumption 5.2 (iv) that

$$\begin{aligned}
&\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right|}{l|\Delta_m|} \\
&\le \sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\sqrt{k^*}}{l|\Delta_m|} \left|\frac{1}{\sqrt{k^*}} \sum_{j=m+1}^{m+k^*} h_2(Y_j)\right| \\
&\quad + \sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\sqrt{N-k^*}}{l|\Delta_m|} \left|\frac{1}{\sqrt{N-k^*}} \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right| \\
&\le \left(\frac{\sqrt{k^*}}{(1-\delta)(N-k^*)|\Delta_m|} + \frac{1}{(1-\delta)\sqrt{N-k^*}|\Delta_m|}\right) O_P(1). \qquad (5.46)
\end{aligned}$$

With (5.36) and $\frac{k^*}{N} = \frac{k^*}{a_m}\frac{a_m}{N} = O(1)$ by Lemma 5.4 it holds

$$\frac{\sqrt{k^*}}{(1-\delta)(N-k^*)|\Delta_m|} = \frac{\sqrt{\frac{k^*}{N}}}{(1-\delta)N^{-\frac{1}{2}}(N-k^*)|\Delta_m|} = o(1) \quad \text{as } m \to \infty.$$

Additionally, (5.36) implies

$$\begin{aligned}
|\Delta_m|\sqrt{N-k^*} &= N^{-\frac{1}{2}}(N-k^*)|\Delta_m|\sqrt{\frac{N}{N-k^*}} \\
&\ge N^{-\frac{1}{2}}(N-k^*)|\Delta_m| \to \infty \quad \text{as } m \to \infty.
\end{aligned}$$

Consequently, (5.46) yields

$$\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right|}{l|\Delta_m|} = o_P(1) \quad \text{as } m \to \infty.$$

This shows that the absolute value of the deterministic part exceeds the absolute value of the stochastic part for m large enough and thus determines the sign. Hence, it follows $\lim_{m\to\infty} \mathrm{P}(B_m) = 1$ where B_m is the event that

$$\inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \operatorname{sign}(\Delta_m) \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma} \ge 0.$$

Let

$$z_m := z\left(\frac{N}{m}\right)^{\frac{1}{2}-\gamma} + \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}.$$

By Lemma B.3 (i) it follows

$$\begin{aligned}
\mathrm{P}\Bigg(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \Bigg(&\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma} \\
&- \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\Bigg) \le z\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{P}\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma} \le z_m \right)\\
&= \mathrm{P}\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m\right|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma} \le z_m, B_m \right)\\
&\quad + o(1)\\
&= \mathrm{P}\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma} \le z_m, B_m \right)\\
&\quad + o(1)\\
&= \mathrm{P}\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma} \right.\right.\\
&\quad \left.\left. - \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right) \le z \right) + o(1) \quad \text{as } m\to\infty, \qquad (5.47)
\end{aligned}
$$

It holds with Assumption 5.1 (i)

$$
\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \left(\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma} \right.\right.\\
&\qquad \left.\left. + \frac{l|\Delta_m|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right) - \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\\
&\ge \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \left(\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+\lfloor N-k^*\rfloor} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{\lfloor N-k^*\rfloor+k^*}{m}\right)^\gamma} \right.\\
&\qquad \left. + \frac{\lfloor N-k^*\rfloor |\Delta_m|}{\sqrt{m}\left(\frac{\lfloor N-k^*\rfloor+k^*}{m}\right)^\gamma} - \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\\
&= \left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^\gamma \operatorname{sign}(\Delta_m)\left(\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \frac{1}{\sqrt{N}}\sum_{j=m+k^*+1}^{m+\lfloor N\rfloor} h_{2,m}^*(Z_{j,m})\right)\\
&\qquad + \left(\left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^\gamma \frac{\lfloor N-k^*\rfloor}{\sqrt{N}} - \frac{N-k^*}{\sqrt{N}}\right)|\Delta_m|\\
&= (1+o(1))\operatorname{sign}(\Delta_m)\left(\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \frac{1}{\sqrt{N}}\sum_{j=m+k^*+1}^{m+\lfloor N\rfloor} h_{2,m}^*(Z_{j,m})\right) + o(1)
\end{aligned}
$$

as $m\to\infty$ noting that the mean value theorem yields

$$
\begin{aligned}
&\left|\left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^\gamma - 1\right| \le \left(\frac{N}{N-1}\right)^\gamma - 1 = \frac{N^\gamma-(N-1)^\gamma}{(N-1)^\gamma} \le \frac{\gamma(N-1)^{\gamma-1}}{(N-1)^\gamma}\\
&=\gamma\frac{1}{N-1} = O\left(\frac{1}{N}\right)
\end{aligned}
$$

such that

$$\begin{aligned}
&\left|\left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^{\gamma}\frac{\lfloor N-k^*\rfloor}{\sqrt{N}}-\frac{N-k^*}{\sqrt{N}}\right|\\
\leq&\left|\left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^{\gamma}\frac{\lfloor N-k^*\rfloor}{\sqrt{N}}-\left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^{\gamma}\frac{N-k^*}{\sqrt{N}}\right|\\
&+\left|\left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^{\gamma}\frac{N-k^*}{\sqrt{N}}-\frac{N-k^*}{\sqrt{N}}\right|\\
\leq&O\left(\frac{1}{N}\right)\frac{N-k^*}{\sqrt{N}}+\left(\frac{N}{\lfloor N-k^*\rfloor+k^*}\right)^{\gamma}\frac{1}{\sqrt{N}}\\
\leq&O\left(\frac{1}{\sqrt{N}}\right)\left(1-\frac{k^*}{N}\right)+\left(\frac{N}{N-1}\right)^{\gamma}\frac{1}{\sqrt{N}}=O\left(\frac{1}{\sqrt{N}}\right)=o(1)
\end{aligned}$$

as $\left(1-\frac{k^*}{N}\right)=O(1)$ for all cases in Lemma 5.4 a) (iv). Because of the stationarity we get with Assumption 5.2 (iv)

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\frac{1}{\sqrt{N}}\sum_{j=m+k^*+1}^{m+\lfloor N\rfloor}h^*_{2,m}(Z_{j,m})\\
\overset{\mathcal{D}}{=}&\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*}h_2(Y_j)+\frac{1}{\sqrt{N}}\sum_{j=k^*+1}^{\lfloor N\rfloor}h^*_{2,m}(Z_{j,m})\\
=&\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*}h_2(Y_j)+\frac{1}{\sqrt{N}}\sum_{j=1}^{\lfloor N\rfloor}h^*_{2,m}(Z_{j,m})-\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*}h^*_{2,m}(Z_{j,m})\\
\overset{\mathcal{D}}{\to}&Z:=\begin{cases}W(0)+W^*(1)-W^*(0)=W^*(1) & \text{under } (I)\\ W(\delta_1)+W^*(1)-W^*(\delta_1) & \text{under } (II)\\ W(1) & \text{under } (III)\end{cases}
\end{aligned}$$

by Lemma 5.4 a) (iv). It holds

$$Z\sim\begin{cases}N(0,\sigma^{*2}) & \text{under } (I)\\ N(0,\delta_1\sigma^2+(1-\delta_1)\sigma^{*2}) & \text{under } (II)\\ N(0,\sigma^2) & \text{under } (III)\end{cases}$$

as $W^*(1)-W^*(\delta_1)\sim N\left(0,\sigma^{*2}(1-\delta_1)\right)$ and $W(\delta_1)\sim N(0,\sigma^2)$ are independent increments of the two dimensional Wiener Process in Assumption 5.2 (iv). Due to the symmetry of the Gaussian distribution, the above limit distribution still holds when multiplying with $\operatorname{sign}(\Delta_m)\in\{-1,1\}$. For $\lim_{m\to\infty}\operatorname{sign}(\Delta_m)=s_\Delta\in\{-1,1\}$ this is obvious. Otherwise, $\{\Delta_m\}_{m\geq1}$ can be decomposed in two subsequences with $\lim_{n\to\infty}\operatorname{sign}(\Delta_{m_n})=-1$ and $\lim_{n\to\infty}\operatorname{sign}(\Delta_{m'_n})=1$ which both lead to the above limit distribution. Hence, we get with Slutsky's theorem

$$\begin{aligned}
&\lim_{m\to\infty} P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\left(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right.\right.\right.\\
&\qquad\left.\left.\left.+\frac{\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})+l\Delta_m}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right)-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\le z\right)\\
&\le \lim_{m\to\infty} P\left(\operatorname{sign}(\Delta_m)\left(\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\frac{1}{\sqrt{N}}\sum_{j=m+k^*+1}^{m+\lfloor N\rfloor} h^*_{2,m}(Z_{j,m})\right)\le z\right)\\
&=P(Z\le z).
\end{aligned}\tag{5.48}$$

In addition, it holds

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})+l\Delta_m}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right.\\
&\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\\
&\le\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\right.\\
&\left.+\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\frac{l|\Delta_m|}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\\
&=\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}
\end{aligned}$$

as $\frac{l}{(l+k^*)^\gamma}=l^{1-\gamma}\left(\frac{1}{1+\frac{k^*}{l}}\right)^{\gamma}$ is increasing in l. Assumption 5.2 (iv) yields

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\\
&\overset{\mathcal{D}}{=}\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\frac{\sum_{j=1}^{k^*} h_2(Y_j)+\sum_{j=k^*+1}^{k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\\
&=\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\frac{\sum_{j=1}^{k^*} h_2(Y_j)+\sum_{j=1}^{k^*+l} h^*_{2,m}(Z_{j,m})-\sum_{j=1}^{k^*} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^{\gamma}}\\
&=\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\frac{\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*} h_2(Y_j)+\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*+l} h^*_{2,m}(Z_{j,m})-\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*} h^*_{2,m}(Z_{j,m})}{\left(\frac{l+k^*}{N}\right)^{\gamma}}\\
&=\sup_{(1-\delta)+\delta\frac{k^*}{N}\le\frac{l}{N}\le 1}\frac{\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*} h_2(Y_j)+\frac{1}{\sqrt{N}}\sum_{j=1}^{l} h^*_{2,m}(Z_{j,m})-\frac{1}{\sqrt{N}}\sum_{j=1}^{k^*} h^*_{2,m}(Z_{j,m})}{\left(\frac{l}{N}\right)^{\gamma}}\\
&\overset{\mathcal{D}}{\to}\begin{cases}\sup_{(1-\delta)\le t\le 1}\frac{W^*(t)}{t^\gamma} & \text{under } (I)\\ \sup_{(1-\delta(1-\delta_1))\le t\le 1}\frac{W(\delta_1)+W^*(t)-W^*(\delta_1)}{t^\gamma} & \text{under } (II)\\ W(1) & \text{under } (III)\end{cases}\qquad \text{as } m\to\infty.
\end{aligned}$$

As $\{(-W(t), -W^*(t))\}$ is again a Wiener processes with mean zero and the same covariance structure as $\{(W(t), W^*(t))\}$, we obtain the same limit distribution for

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*} \operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}.$$

Due to the almost sure continuity of the Wiener process we get for $\delta \to 0$

$$\begin{aligned}
&\left|\sup_{1-\delta\leq t\leq 1} \frac{W^*(t)}{t^\gamma} - W^*(1)\right| \\
\leq &\left|\sup_{1-\delta\leq t\leq 1}\left(\frac{W^*(t)}{t^\gamma} - W^*(t)\right)\right| + \left|\sup_{1-\delta\leq t\leq 1} W^*(t) - W^*(1)\right| \\
\leq &\sup_{1-\delta\leq t\leq 1} |W^*(t) - W^*(1)| + \left|\frac{1}{(1-\delta)^\gamma} - 1\right| \sup_{1-\delta\leq t\leq 1} |W^*(t)| = o_P(1)
\end{aligned}$$

and

$$\begin{aligned}
&\left|\sup_{(1-\delta(1-\delta_1))\leq t\leq 1} \frac{W(\delta_1) + W^*(t) - W^*(\delta_1)}{t^\gamma} - (W(\delta_1) + W^*(1) - W^*(\delta_1))\right| \\
\leq &\left|\sup_{(1-\delta(1-\delta_1))\leq t\leq 1}\left(\frac{W(\delta_1) + W^*(t) - W^*(\delta_1)}{t^\gamma} - (W(\delta_1) + W^*(t) - W^*(\delta_1))\right)\right| \\
&+\left|\sup_{(1-\delta(1-\delta_1))\leq t\leq 1} (W(\delta_1) + W^*(t) - W^*(\delta_1)) - (W(\delta_1) + W^*(1) - W^*(\delta_1))\right| \\
\leq &\sup_{(1-\delta(1-\delta_1))\leq t\leq 1} |W^*(t) - W^*(1)| \\
&+\left|\frac{1}{(1-\delta(1-\delta_1))^\gamma} - 1\right| \sup_{1-\delta(1-\delta_1)\leq t\leq 1} |W(\delta_1) + W^*(t) - W^*(\delta_1)| = o_P(1).
\end{aligned}$$

We conclude that

$$\begin{aligned}
&\lim_{m\to\infty} P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*} \operatorname{sign}(\Delta_m)\left(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right.\right.\right. \\
&\qquad\left.\left.\left.+\frac{\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) + l\Delta_m}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right) - \frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right) \leq z\right) \\
\geq &\lim_{m\to\infty} P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*} \operatorname{sign}(\Delta_m)\left(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right.\right. \\
&\qquad\left.\left.+\frac{\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right) \leq z\right) \\
= &P(Z\leq z) + a_\delta,
\end{aligned} \tag{5.49}$$

where $a_\delta \to 0$ for $\delta \to 0$.

Combining (5.48)and (5.49) we get

$$\left|\lim_{m\to\infty} P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \operatorname{sign}(\Delta_m)\left(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right.\right.\right.\right.$$
$$\left.\left.\left.\left.+\frac{\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})+l\Delta_m}{\sqrt{m}\left(\frac{l+k^*}{m}\right)^\gamma}\right)-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\le z\right)-\Psi(z)\right|\le a_\delta,$$

hence by (5.45) and (5.47)

$$\left|\lim_{m\to\infty} P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{k^*<k\le N}\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})+(k-k^*)\Delta_m}{\sqrt{m}\left(\frac{k}{m}\right)^\gamma}\right.\right.\right.$$
$$\left.\left.\left.-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\le z\right)-\Psi(z)\right|\le a_\delta\to 0\quad\text{as }\delta\to 0.$$

The assertion now follows as the left-hand side does not depend on δ and δ was arbitrary. □

Lemma 5.10. *Let Assumption 3.3, 5.1 and 5.2 be satisfied. Then, it holds*

$$P\left\{\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1\le k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{|\Delta_m|(N-k^*)}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\le z\right\}\to\Psi(z)\quad \textit{as } m\to\infty.$$

Proof. With Lemma 5.5 and B.3 it holds for any $z\in\mathbb{R}$

$$\begin{aligned}
&\lim_{m\to\infty} P\left\{\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1\le k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\le z\right\}\\
&=\lim_{m\to\infty} P\left\{\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1\le k\le k^*}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\le z,\right.\\
&\qquad\left.\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{k^*<k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\le z\right\}\\
&=\lim_{m\to\infty} P\left\{\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{k^*<k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^\gamma}\right)\le z\right\}.
\end{aligned}$$

Together with the reverse triangle inequality, Lemma 5.6 implies

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{k^*<k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}\\
&=\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{k^*<k\le N}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})+(k-k^*)|\Delta_m|\right|}{g(m,k)}+o_P(1)
\end{aligned}$$

as $m\to\infty$. This can be further simplified to

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{k^*<k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}\\
&=\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sup_{k^*<k\le N}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})+(k-k^*)|\Delta_m|\right|}{\sqrt{m}\left(\frac{k}{m}\right)^\gamma}+o_P(1)
\end{aligned}$$

as $m \to \infty$ with Lemma 5.7. Hence, Slutsky's Theorem and Lemma 5.9 yield

$$\lim_{m\to\infty} P\left\{\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1\le k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\le z\right\}=\Psi(z).$$

□

Proof of Theorem 5.3. As Ψ is continuous it holds with Lemma 5.4 c) and Lemma 5.10

$$\begin{aligned}
&\lim_{m\to\infty} P(\tau_m > N)\\
=&\lim_{m\to\infty} P\left(\sup_{1\le k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}\le c\right)\\
=&\lim_{m\to\infty} P\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\sup_{1\le k\le N}\frac{|\Gamma(m,k)|}{g(m,k)}-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\right.\\
&\qquad\left.\le\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(c-\frac{(N-k^*)|\Delta_m|}{\sqrt{m}\left(\frac{N}{m}\right)^{\gamma}}\right)\right)\\
=&\Psi(-x)
\end{aligned}\tag{5.50}$$

for all real x. Furthermore, it holds by (5.13)

$$\begin{aligned}
&1-\lim_{m\to\infty} P\left(\tau_m > N\right)\\
=&\lim_{m\to\infty} P\left(\tau_m \le N\right)\\
=&\lim_{m\to\infty} P\left(\tau_m^{1-\gamma} \le N^{1-\gamma}\right)\\
=&\lim_{m\to\infty} P\left(\tau_m^{1-\gamma}\le\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|}+\frac{k^*}{a_m^{\gamma}}+x\frac{a_m^{\frac{1}{2}-\gamma}(1-\gamma)}{|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))}\right)\\
=&\lim_{m\to\infty} P\left(\tau_m^{1-\gamma}-a_m^{1-\gamma}\le x\frac{a_m^{\frac{1}{2}-\gamma}(1-\gamma)}{|\Delta_m|\left(1-\gamma(1-\frac{k^*}{a_m})\right)}\right)\\
=&\lim_{m\to\infty} P\left(\frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma}-a_m^{1-\gamma}}{b_m}\le x\frac{\sqrt{a_m}}{|\Delta_m|\left(1-\gamma(1-\frac{k^*}{a_m})\right)}b_m^{-1}\right)\\
=&\lim_{m\to\infty} P\left(\frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma}-a_m^{1-\gamma}}{b_m}\le x\right)
\end{aligned}\tag{5.51}$$

as the definitions of a_m and b_m in (5.10) and (5.11) imply

$$\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|}+\frac{k^*}{a_m^{\gamma}}=a_m^{1-\gamma}$$

and

$$\frac{\sqrt{a_m}}{|\Delta_m|\left(1-\gamma(1-\frac{k^*}{a_m})\right)}b_m^{-1}=1.$$

With (5.17) it holds

$$N^{1-\gamma} = a_m^{1-\gamma}\left(1 + x\frac{1-\gamma}{\sqrt{a_m}|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))}\right) = a_m^{1-\gamma}(1+\delta_m)$$

with $0 < \delta_m = x\frac{1-\gamma}{\sqrt{a_m}|\Delta_m|(1-\gamma(1-\frac{k^*}{a_m}))} \to 0$ by (5.19) as $m \to \infty$. Hence, we get for any $\epsilon > 0$ and m large enough

$$\begin{aligned}P\left(\tau_m > N\right) &= P\left(\tau_m^{1-\gamma} > N^{1-\gamma}\right) = P\left(\tau_m^{1-\gamma} > a_m^{1-\gamma}(1+\delta_m)\right)\\ &= P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 > \delta_m\right) \geq P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 > \epsilon\right).\end{aligned}$$

Using (5.50) this yields

$$\lim_{m\to\infty} P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 > \epsilon\right) \leq \Psi(-x) \quad \text{for all } x \in \mathbb{R}$$

such that, letting x tend to ∞, we get

$$\lim_{m\to\infty} P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 > \epsilon\right) = 0.$$

Analogously, it holds

$$\begin{aligned}P\left(\tau_m \leq N\right) &= P\left(\tau_m^{1-\gamma} \leq N^{1-\gamma}\right) = P\left(\tau_m^{1-\gamma} \leq a_m^{1-\gamma}(1+\delta_m)\right)\\ &= P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 \leq \delta_m\right) \geq P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 < -\epsilon\right)\end{aligned}$$

for m large enough and thus

$$\lim_{m\to\infty} P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 < -\epsilon\right) \leq 1 - \Psi(-x) \quad \text{for all } x \in \mathbb{R}.$$

Now, letting x tend to $-\infty$, we obtain

$$\lim_{m\to\infty} P\left(\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} - 1 < -\epsilon\right) = 0.$$

In total, those considerations show that

$$\left(\frac{\tau_m}{a_m}\right)^{1-\gamma} \xrightarrow{P} 1 \quad \text{as } m \to \infty. \tag{5.52}$$

With the mean value theorem it holds

$$\begin{aligned}\frac{\tau_m - a_m}{b_m} &= \frac{(\tau_m^{1-\gamma})^{\frac{1}{1-\gamma}} - (a_m^{1-\gamma})^{\frac{1}{1-\gamma}}}{b_m}\\ &= \frac{1}{1-\gamma} z_m^{\frac{1}{1-\gamma}-1} \frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m},\end{aligned}$$

where z_m lies between $a_m^{1-\gamma}$ and $\tau_m^{1-\gamma}$. Considering (5.52) it also has to hold that $z_m = a_m^{1-\gamma}(1 + o_P(1))$. Hence, we get

$$\begin{aligned}\frac{\tau_m - a_m}{b_m} &= \frac{1}{1-\gamma}\left(a_m^{1-\gamma}(1+o_P(1))\right)^{\frac{1}{1-\gamma}-1}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m}\\ &= \frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m}(1+o_P(1))\\ &= \frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m} + o_P(1)\end{aligned}$$

as $\frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m} = O_P(1)$ by (5.51). Now, Slutzky's lemma yields that

$$\frac{\tau_m - a_m}{b_m} \quad \text{and} \quad \frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m}$$

have the same limit distribution such that (5.7) is satisfied. Combining this with (5.51), we obtain (5.1). Now, it follows with (5.50)

$$\lim_{m\to\infty} P\left(\frac{\tau_m - a_m}{b_m}\right) = 1 - \Psi(-x) = \Psi(x)$$

as Ψ is symmetric. □

5.1.1. Comparison of CUSUM and Wilcoxon procedure

Based on Theorem 5.3 we are now able to compare the expected stopping time of the CUSUM and Wilcoxon procedure which can be approximated by a_m^C resp. a_m^W. By (5.10) it holds for $\gamma = 0$

$$a_m^{W/C} = \frac{c_\alpha^{W/C}\sqrt{m}}{\Delta_m^{W/C}} + k^*, \tag{5.53}$$

where $c_\alpha^{W/C}$ is the $1-\alpha$ quantile of $\sup_{0<t<1}\sigma_{W/C}|W(t)|$ with

$$\sigma_{W/C}^2 = \sum_{h\in\mathbb{Z}} Cov(h_2^{W/C}(Y_0), h_2^{W/C}(Y_h))$$

as Corollary 3.7 (i) holds for the CUSUM as well as for the Wilcoxon kernel. Let $\tilde{c}_\alpha$ be the $1-\alpha$-quantile of $\sup_{0<t<1}|W(t)|$ such that $c_\alpha^{W/C} = \sigma_{W/C}\tilde{c}_\alpha$. Then, it holds

$$a_m^W - a_m^C = \left(\tfrac{\sigma_W}{|\Delta_m^W|} - \tfrac{\sigma_C}{|\Delta_m^C|}\right)\tilde{c}_\alpha\sqrt{m}.$$

Hence, depending on the size of change and the distribution of the underlying time series in terms of $|\Delta_m|$ and σ, the factor $\left(\tfrac{\sigma_W}{|\Delta_m^W|} - \tfrac{\sigma_C}{|\Delta_m^C|}\right)$ determines whether the CUSUM or the Wilcoxon procedure is expected to be faster. In the following, we examine this factor exemplarily for a fixed mean change in i.i.d. data. Therefore, consider Example

	N(0,1)	Laplace(0,1)	t(3)
$\frac{a_m^W - a_m^C}{\tilde{c}_\alpha \sqrt{m}}$	0.109	-0.126	-0.379

Table 5.1.: Scaled differences of expected stopping times.

2.3 with $\{Y_i\}_{i\geq 1}$ i.i.d. with a continuous distribution and $d_m = 1$. With h_1^C and h_1^W as in Example 3.1 we obtain

$$\sigma_C^2 = \mathrm{Var}(h_2^C(Y_1)) = \mathrm{Var}(Y_1), \tag{5.54}$$

$$\sigma_W^2 = \mathrm{Var}(h_2^W(Y_1)) = \mathrm{Var}\left(F_Y(Y_1)\right) = \frac{1}{12}. \tag{5.55}$$

The change in the CUSUM kernel is given by $\Delta^C = -1$ as derived in (2.9). According to (2.10) it holds $\Delta^W = P(Y_1' \leq Y_1 < Y_1' + 1)$.

Table 5.1 shows the scaled differences of the approximations of the expected stopping times for particular distributions. It can be observed that the CUSUM procedure is expected to detect a change faster than the Wilcoxon procedure for normally distributed random variables, while the Wilcoxon procedure is expected to have a shorter detection delay for heavy tailed distributions such as the Laplace distribution or the t distribution with 3 degrees of freedom. This is not surprising as the Wilcoxon kernel can better deal with extreme observations due to its robustness.

5.2. Superlinear Changes

In this section, we analyze the asymptotic behavior of the stopping time for superlinear changes in the sense that $k^* = [\lambda m^\beta]$ with $\beta > 1$. We consider the change point model as described in Section 2.2 and the weight function as in (3.5) for $\gamma = 0$.

The stopping time for change points that are linear in m, i.e. for $\beta = 1$, is considered in Section 5.3 below where the main approach is the same as in this section but the proofs differ in some details. To the best of our knowledge, up to now, there does not even exist a counterpart of Theorem 5.3 for $\beta \geq 1$ for the CUSUM procedure. The analysis of the stopping time for late changes requires a different approach for the following reasons: the probability of rejecting the null hypothesis before the change occurs is not negligible when using a fixed critical value and the behavior of the stopping time strongly depends on what we have observed until we start monitoring. The positive probability of an early rejection would contaminate the limit distribution of the stopping time. However, it is not clear how this limit distribution, which one would expect to be bimodal, could be derived and furthermore we are mainly interested in investigating the detection delay related to the actual change rather than stopping times due to false positives. Lemma 5.11 shows that, for late changes, we can obtain a procedure for which the probability of falsely rejecting the null hypothesis before the change occurs tends to zero by letting the critical value increase to infinity. The same Lemma also reveals that this is not necessary for early changes as the respective probability of rejecting too early is already negligible for any fixed critical value. This relates to the assertion of Lemma 5.5.

Lemma 5.11. *Let Assumptions 3.3 and 5.2 be satisfied and let* $k^* = [\lambda m^\beta]$ *and* $S_{1,m} := \frac{1}{\sqrt{m}} \sum_{i=1}^m h_1(Y_i)$.

(i) For $\beta < 1$ *and* $\frac{1}{c_m} = O(1)$ *it holds*

$$P\left(\sup_{1\le k\le k^*} \frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} > c_m \right) \to 0. \tag{5.56}$$

(ii) For $\beta = 1$, *(5.56) holds if* $c_m \to \infty$ *and* $\limsup_{m\to\infty} \frac{|S_{1,m}|}{c_m} < 1 + \frac{1}{\lambda}$ *a.s..*

(iii) For $\beta > 1$ *(5.56) holds if* $c_m \to \infty$ *and* $\limsup_{m\to\infty} \frac{|S_{1,m}|}{c_m} < 1$ *a.s..*

Proof. For $1 \le k \le k^*$ consider the representation of the monitoring statistic as given in (3.3):

$$\Gamma(m,k) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k} r(Y_i,Y_j) + \sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^{m} h_1(Y_i).$$

With Lemma 3.5 it holds

$$\begin{aligned}
&\sup_{1\le k\le k^*} \left| \frac{\Gamma(m,k)}{\sqrt{m}\left(1+\frac{k}{m}\right)} - \frac{\sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^m h_1(Y_i)}{\sqrt{m}\left(1+\frac{k}{m}\right)} \right| \\
\le &\sup_{k\ge 1} \frac{\frac{1}{m}\left|\sum_{i=1}^m \sum_{j=m+1}^{m+k} r(Y_i,Y_j)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} = o_P(1), \quad m\to\infty.
\end{aligned}$$

Now, observe that

$$\begin{aligned}
&\sup_{1\le k\le k^*} \left| \frac{\sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{m}\sum_{i=1}^m h_1(Y_i)}{\sqrt{m}\left(1+\frac{k}{m}\right)} \right| = \sup_{1\le k\le k^*} \left| \frac{\sum_{j=m+1}^{m+k} h_2(Y_j) + \frac{k}{\sqrt{m}} S_{1,m}}{\sqrt{m}\left(1+\frac{k}{m}\right)} \right| \\
\le &\sup_{1\le k\le k^*} \frac{\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} + \sup_{1\le k\le k^*} \frac{\left|\frac{k}{\sqrt{m}} S_{1,m}\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} \\
\le &\sup_{1\le k\le k^*} \frac{\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} + \sup_{1\le k\le k^*} \frac{\frac{k}{m}}{1+\frac{k}{m}} |S_{1,m}| \\
\le &\sup_{1\le k\le k^*} \frac{\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} + \frac{\frac{k^*}{m}}{1+\frac{k^*}{m}} |S_{1,m}|
\end{aligned} \tag{5.57}$$

with Assumption 3.3 (ii). For $\beta < 1$, there exists an $m_0 \in \mathbb{N}$ such that $\frac{k^*}{m} < 1$ for all $m \ge m_0$. Hence, (5.57) implies for m large enough

$$\sup_{1\le k\le k^*} \frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} \le \frac{\sqrt{\frac{k^*}{m}}}{1+\frac{k^*}{m}} O_P(1) + \frac{\frac{k^*}{m}}{1+\frac{k^*}{m}} |S_{1,m}| + o_P(1) = o_P(1)$$

as $\frac{k^*}{m} \to 0$. For $\beta > 1$, we obtain with (5.57) and Assumption 3.3 (iv) as well as the stationarity

$$\begin{aligned}
&\frac{1}{c_m} \sup_{1\le k\le k^*} \frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} \\
\le&\frac{1}{c_m}\left(\sup_{1\le k\le m} \frac{\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} + \sup_{m<k\le k^*} \frac{\left|\sum_{j=m+1}^{m+k} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} + |S_{1,m}| + o_P(1)\right) \\
=&\frac{1}{c_m}\left(\sup_{1\le k\le m} \frac{1}{1+\frac{k}{m}} \left|\frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+k} h_2(Y_j)\right| + \sup_{m<k\le k^*} \frac{\frac{k}{m}}{1+\frac{k}{m}} \left|\frac{\sqrt{m}}{k} \sum_{j=m+1}^{m+k} h_2(Y_j)\right|\right. \\
&\qquad \left. + |S_{1,m}| + o_P(1)\right) \\
\le&\frac{1}{c_m}\left(\sup_{1\le k\le m} \left|\frac{1}{\sqrt{m}} \sum_{j=m+1}^{m+k} h_2(Y_j)\right| + \sup_{m<k\le k^*} \left|\frac{\sqrt{m}}{k} \sum_{j=m+1}^{m+k} h_2(Y_j)\right| + |S_{1,m}| + o_P(1)\right) \\
\overset{\mathcal{D}}{=}&\frac{1}{c_m}\left(\sup_{1\le k\le m} \left|\frac{1}{\sqrt{m}} \sum_{j=1}^{k} h_2(Y_j)\right| + \sup_{m<k\le k^*} \left|\frac{\sqrt{m}}{k} \sum_{j=1}^{k} h_2(Y_j)\right| + |S_{1,m}| + o_P(1)\right) \\
=&O_P\left(\frac{1}{c_m}\right) + \frac{|S_{1,m}|}{c_m} + o_P\left(\frac{1}{c_m}\right).
\end{aligned}$$

For $\beta = 1$ it holds

$$\begin{aligned}
\frac{1}{c_m} \sup_{1\le k\le k^*} \frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} &\le \frac{\sqrt{\frac{k^*}{m}}}{1+\frac{k^*}{m}} O_P\left(\frac{1}{c_m}\right) + \frac{\frac{k^*}{m}}{1+\frac{k^*}{m}} \frac{|S_{1,m}|}{c_m} + o_P\left(\frac{1}{c_m}\right) \\
&= O_P\left(\frac{1}{c_m}\right) + \frac{\frac{k^*}{m}}{1+\frac{k^*}{m}} \frac{|S_{1,m}|}{c_m} + o_P\left(\frac{1}{c_m}\right)
\end{aligned}$$

with $\frac{k^*}{m} \to \lambda$. □

In the following, we choose the critical value according to Lemma 5.11 such that the stopping time asymptotically coincides with

$$\tilde{\tau}_m := \inf\left\{k > k^* : \frac{|\Gamma(m,k)|}{g(m,k)} > c_m\right\}. \tag{5.58}$$

We solve the issue of the stopping time depending on the situation at the beginning of the monitoring period by conditioning on $S_{1,m} = s_{1,m}$ and $S^*_{1,m} = s^*_{1,m}$ with

$$S_{1,m} = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h_1(Y_i) \quad \text{and} \quad S^*_{1,m} = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} h^*_{1,m}(Y_i).$$

The conditioning random variables only involve the historic data set which has been observed before the monitoring starts. However, it should be mentioned that the functions h_1 and $h^*_{1,m}$ are not necessarily known as they depend on the unknown distribution of the time series before and after the change.

Example 5.12. *Based on Hoeffding's decomposition in Example 3.11 for a mean change, we obtain the following conditioning random variables for the CUSUM and the Wilcoxon kernel:*

$$\begin{aligned}
S_{1,m}^{C} = S_{1,m}^{*C} &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}(Y_i-\mu),\\
S_{1,m}^{W} &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{1}{2}-F(Y_i)\right),\\
S_{1,m}^{*W} &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{1}{2}-F(Y_i-d_m)-\Delta_m^W\right).
\end{aligned}$$

We impose the following assumptions on the time and the size of the change.

Assumption 5.13.

(i) *There exists a* $\lambda > 0$ *such that* $k^* = [\lambda m^\beta]$ *with some* $\beta > 1$.

(ii) $\Delta_m = O(1)$.

(iii) $\lim_{m\to\infty} \frac{c_m}{\sqrt{m}|\Delta_m|} = 0$.

As we let the critical value c_m increase to infinity, (iii) is needed for the detectability of the change. With (3.50)-(3.53) we can see that the procedure has asymptotic power one under this condition. Furthermore, we require the conditioning sequences to fulfill Assumption 5.14. Part (iii) is derived from Lemma 5.11 and ensures that the probability of rejecting too early tends to zero. In Remark 5.19 we will see that if $S_{1,m}$ and $S_{1,m}^*$ fulfill the law of the iterated logarithm, c_m can be chosen such that almost all realizations of the conditioning random variables satisfy the following assumptions.

Assumption 5.14.

(i) $c_m \to \infty$ *and* $\limsup_{m\to\infty}\left|\frac{s_{1,m}}{c_m}\right| < 1$.

(ii) $\frac{s_{1,m}^*}{s_{1,m}} = O(1)$ *and* $\frac{1}{s_{1,m}} = O(1)$.

(iii) $\lim_{m\to\infty}\frac{|s_{1,m}|}{\sqrt{m}|\Delta_m|} = \lim_{m\to\infty}\frac{|s_{1,m}^*|}{\sqrt{m}|\Delta_m|} = 0$.

Regarding the stochastic terms we get by with weaker assumptions than in the sublinear case.

Assumption 5.15. *Let* $\{Y_i\}_{i\in\mathbb{Z}}$ *and* $\{Z_{i,m}\}_{i\in\mathbb{Z}}$ *be stationary time series that fulfill the following assumptions for a given kernel function* h.

(i) $\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r_m^*(Y_i, Z_{j,m})\right|^2\right) \le u(m)(k_2-k_1+1)$ *for all* $m+k^*+1 \le k_1 \le k_2$ *with* $\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \to 0$ *and* γ *as in Assumption 3.2.*

(ii) $\sup_{1\le l\le l_m} \frac{1}{\sqrt{l_m}} \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m}) = O_P(1)$ *as* $l_m \to \infty$.

(iii) $\frac{1}{\sqrt{k_m}} \sum_{j=1}^{k_m} h_2(Y_j) \xrightarrow{\mathcal{D}} N(0,\sigma^2)$ *as* $k_m \to \infty$.

Part (ii) is very general and can be obtained, for example, by a functional central limit theorem or by Kolmogorov's inequality.

In order to assess the asymptotic behavior of the stopping time we need to find normalizing sequences a_m and b_m such that we can derive the conditional limit distribution of

$$\frac{\tilde{\tau}_m - a_m}{b_m}.$$

In contrast to the sublinear case in the previous section, due to the conditioning, a_m and b_m can and will depend on $s_{1,m}$ and $s^*_{1,m}$. We follow the general approach that has been described at the beginning of this chapter where the respective probabilities now appear with the condition on $\{S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\}$ and by the definition of $\tilde{\tau}_m$ we consider the supremum only over $k^* < k \le N$. Consequently, we need to find a centering sequence d_m and a scaling sequence e_m such that we can derive the conditional limit distribution

$$\Psi(z) := \lim_{m\to\infty} P\left(e_m \left(\sup_{k^*<k\le N} \frac{|\Gamma(m,k)|}{g(m,k)} - d_m \right) < x \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m} \right).$$

Then, we get

$$\lim_{m\to\infty} P\left(\frac{\tilde{\tau}_m - a_m}{b_m} \le x \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m} \right) = 1 - \Psi(-x)$$

if

$$e_m\,(c_m - d_m) \to -x \quad \text{as } m \to \infty. \tag{5.59}$$

In the following we consider the weight function as in (3.5) with $\gamma = 0$. In order to determine d_m, we consider the representation of the monitoring statistic as given in (3.45). For sublinear changes, it is sufficient to subtract the signal part (see (5.9)) as $\frac{k^*}{m}\sum_{i=1}^m h_1(Y_i) = \frac{k^*}{\sqrt{m}} S_{1,m}$ and $\frac{k-k^*}{m}\sum_{i=1}^m h^*_{1,m}(Y_i) = \frac{N-k^*}{\sqrt{m}} S^*_{1,m}$ are asymptotically negligible. However, this is not the case for late changes such that we need to additionally subtract those terms as follows:

$$d_m = \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\,\mathrm{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\,\mathrm{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}. \tag{5.60}$$

The different signs are needed in order to take into account whether $s_{1,m}$ and $s^*_{1,m}$ go in the same or opposite direction of the change. With the central limit theorem in mind we suggest to scale with

$$e_m = \sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right) \tag{5.61}$$

which basically cancels out the weight function and induces the factor $\frac{1}{\sqrt{N}}$ for the central limit theorem. With N as in (5.5) it holds

$$\begin{aligned}
&e_m\left(c_m - d_m\right)\\
=&c_m\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right) - \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{N}}\\
=&c_m\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right) - x\frac{b_m|\Delta_m|}{\sqrt{N}}\\
&- \frac{(a_m-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{N}}.
\end{aligned} \tag{5.62}$$

Hence, we obtain (5.59) if we choose a_m and b_m such that

$$b_m\frac{|\Delta_m|}{\sqrt{N}} = 1 + o(1) \quad \text{as } m \to \infty \tag{5.63}$$

and

$$c_m\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right) - \frac{(a_m-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{N}} = o(1). \tag{5.64}$$

The latter is satisfied if

$$a_m - k^* = \frac{c_m\sqrt{m}}{|\Delta_m|}\left(1+\frac{N}{m}\right) - \frac{k^*}{\sqrt{m}\Delta_m}s_{1,m} - \frac{N-k^*}{\sqrt{m}\Delta_m}s^*_{1,m} + o\left(\frac{\sqrt{N}}{|\Delta_m|}\right). \tag{5.65}$$

For N as in (5.5) and b_m fulfilling (5.63), this is satisfied by the solution a_m of

$$a_m = \frac{c_m\sqrt{m}}{|\Delta_m|} + \frac{a_m}{\sqrt{m}\,|\Delta_m|}\left(c_m - s^*_{1,m}\operatorname{sign}(\Delta_m)\right) + k^* + \frac{k^*}{\sqrt{m}\Delta_m}(s^*_{1,m} - s_{1,m}) \tag{5.66}$$

by Assumption 5.13 (iii). In particular, it holds for this solution and $k^* = [\lambda m^\beta], \beta \geq 1$

$$a_m(1+o(1)) = k^*(1+o(1)) + \frac{c_m}{\sqrt{m}|\Delta_m|}\cdot m = \lambda m^\beta(1+o(1)) \tag{5.67}$$

by Assumption 5.13 (iii) and Assumption 5.14 (iii). As $c_m \to \infty$, Assumption 5.13 (iii) implies in particular that

$$\sqrt{m}|\Delta_m| \to \infty \tag{5.68}$$

such that it follows with (5.67)

$$\sqrt{a_m}|\Delta_m| \to \infty \quad \text{as } m \to \infty. \tag{5.69}$$

From this we see that the choice

$$b_m = \frac{\sqrt{a_m}}{|\Delta_m|} \tag{5.70}$$

satisfies (5.63) as it holds

$$\frac{N}{a_m} = 1 + x\frac{b_m}{a_m} = 1 + x\frac{1}{\sqrt{a_m}|\Delta_m|} \to 1 \quad \text{as } m \to \infty. \tag{5.71}$$

Having defined the standardizing sequences a_m and b_m we get with (5.5)

$$N = \frac{\sqrt{a_m}}{|\Delta_m|}x + \frac{c_m\sqrt{m}}{|\Delta_m|} + \frac{a_m}{\sqrt{m}\,|\Delta_m|}\left(c_m - s^*_{1,m}\,\text{sign}(\Delta_m)\right) + \frac{k^*}{\sqrt{m}\Delta_m}(s^*_{1,m} - s_{1,m}) + k^*. \tag{5.72}$$

We start with analyzing the asymptotic behavior of the deterministic sequences a_m and N where, in particular, their interplay with the time and the size of the change is important.

Lemma 5.16. *Under Assumption 5.13 and Assumption 5.14 (iii) it holds*

a) *(i)* $\frac{k^*}{m} \to \infty$

(ii) $\sqrt{a_m}\,|\Delta_m| \to \infty$

(iii) $\frac{a_m}{m} \to \infty$

(iv) $\frac{k^*}{a_m} \to 1$

b) $\frac{N}{a_m} \to 1$ *such that part a) is still valid when replacing* a_m *by* N.

c) $\lim_{m\to\infty} e_m\,(c_m - d_m) = -x$ *for all* $x \in \mathbb{R}$.

Proof. Assertion a) (i) follows immediately from Assumption 5.13 (i). The remaining assertions in a) are direct implications of (5.67), where (ii) is already given in (5.69). Furthermore, b) is given in (5.71) and c) holds by (5.62), (5.63) and (5.64). □

By the following Theorem we obtain asymptotic normality of the standardized stopping time of the CUSUM kernel (see Example 2.2 (i)) given the realizations $s_{1,m}$ and $s^*_{1,m}$. Let

$$\tilde{\tau}^C_m := \inf\left\{k > k^* : \frac{|\Gamma_C(m,k)|}{g(m,k)}\; c_m\right\}$$

with Γ_C as in (2.2).

Theorem 5.17. *Let Assumptions 3.3 (ii), 5.13 and 5.15 be satisfied. Furthermore, assume that* $\{X_1, \ldots, X_m\}$ *are independent of* $\{X_{m+j} : j \geq 1\}$. *Then, for all sequences* $s_{1,m}$ *and* $s^*_{1,m}$ *that satisfy Assumption 5.14 it holds under the alternative for the CUSUM kernel*

$$\lim_{m\to\infty} P\left(\frac{\tilde{\tau}^C_m - a^C_m}{b^C_m} \leq x \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\right) = \Phi\left(\frac{x}{\sigma}\right) \quad \text{for all } x \in \mathbb{R},$$

where Φ *is the distribution function of the standard Gaussian distribution,* $\sigma^2 = \sum_{h\in\mathbb{Z}} Cov(Y_0, Y_h)$, $a^C_m = a^C_m(s_{1,m}, s^*_{1,m})$ *is the unique solution of (5.66) and* $b^C_m = b^C_m(s_{1,m}, s^*_{1,m})$ *as in (5.70).*

More generally, the proof of Theorem 5.17 holds for any monitoring statistic obtained by a kernel function for which the remainder term in Hoeffding's decomposition is equal to zero. Assuming that the remainder term is equal to zero means that the representation of the monitoring statistic in (3.45) for $k > k^*$ reduces to

$$\begin{aligned}\tilde{\Gamma}(m,k) := &\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m}) + (k-k^*)\Delta_m \\ &+ \frac{k^*}{m}\sum_{i=1}^{m} h_1(Y_i) + \frac{k-k^*}{m}\sum_{i=1}^{m} h^*_{1,m}(Y_i). \qquad (5.73)\end{aligned}$$

In Example 3.1 we have seen that this is true for the CUSUM kernel but not for the Wilcoxon kernel. Hence, Theorem 5.17 excludes the Wilcoxon kernel such as many others. Nevertheless, it provides the essential basis for Theorem 5.20 which is not restricted to kernels without remainder term. The assumption of independence between the new observations and the historic data is only needed to obtain unconditional probabilities as in Lemma 5.18. This can probably be relaxed to the usual concepts of asymptotic independence but goes beyond the scope of this work.

Lemma 5.18. *Assume that $\{X_1,\ldots,X_m\}$ are independent of $\{X_{m+j} : j \geq 1\}$. Then, it holds for $k > k^*$*

$$\begin{aligned}&P\left(e_m\left(\sup_{p_m\leq k<q_m}\frac{\left|\tilde{\Gamma}(m,k)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right) \leq z \,\middle|\, S_{1,m}=s_{1,m}, S^*_{1,m}=s^*_{1,m}\right) \\ &= P\left(e_m\left(\sup_{p_m\leq k<q_m}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})\right.}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right.\right. \\ &\qquad \left.\left. + \frac{\left.(k-k^*)\Delta_m + \frac{k^*}{\sqrt{m}}s_{1,m} + \frac{k-k^*}{\sqrt{m}}s^*_{1,m}\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right) \leq z\right)\end{aligned}$$

for any deterministic sequences e_m, d_m, p_m, q_m with $p_m < q_m$ and $\tilde{\Gamma}$ as in (5.73).

Proof. With (5.73) it holds

$$\begin{aligned}&P\left(e_m\left(\sup_{p_m\leq k<q_m}\frac{\left|\tilde{\Gamma}(m,k)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right) \leq z \,\middle|\, S_{1,m}=s_{1,m}, S^*_{1,m}=s^*_{1,m}\right) \\ &= P\left(e_m\left(\sup_{p_m\leq k<q_m}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k} h^*_{2,m}(Z_{j,m})\right.}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right.\right. \\ &\qquad \left.\left. + \frac{\left.(k-k^*)\Delta_m + \frac{k^*}{\sqrt{m}}s_{1,m} + \frac{k-k^*}{\sqrt{m}}s^*_{1,m}\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right) \leq z \,\middle|\, S_{1,m}=s_{1,m}, S^*_{1,m}=s^*_{1,m}\right).\end{aligned}$$

As $S_{1,m}$ and $S^*_{1,m}$ only involve the historic observations $\{X_1,\ldots,X_m\}$ which are assumed to be independent of $\{X_{m+j} : j \geq 1\}$, the assertion now follows with Corollary 2 in Section 7.1. in Chow & Teicher (1997). □

Remark 5.19. *Under Assumption 5.13 (iii), Assumption 5.14 is satisfied if for some* $\rho, \rho^* > 0$

$$\limsup_{m\to\infty} \frac{s_{1,m}}{\sqrt{2\varrho^2 \log\log m}} = \frac{s^*_{1,m}}{\sqrt{2\varrho^{*2} \log\log m}} = 1 \tag{5.74}$$

and

$$c_m = \left(\sqrt{2\max(\varrho^2, \varrho^{*2})} + \epsilon\right) \sqrt{\log\log m}. \tag{5.75}$$

The above remark indicates that, based on Theorem 5.17, we can obtain almost sure convergence when conditioning on the random variables $S_{1,m}$ and $S^*_{1,m}$ which fulfill the law of the iterated logarithm such that almost every realization satisfies the assumptions of Theorem 5.17. In this case, we can even handle the remainder term such that the following result is not restricted to the CUSUM kernel.

Theorem 5.20. *Let Assumptions 3.3 (i) and (ii), 5.13 as well as 5.15 be satisfied. Furthermore assume that* $\{X_1, \ldots, X_m\}$ *are independent of* $\{X_{m+j} : j \geq 1\}$ *and that* $S_{1,m}$ *as well as* $S^*_{1,m}$ *fulfill the law of the iterated logarithm, i.e.*

$$\limsup_{m\to\infty} \frac{|S_{1,m}|}{\sqrt{2\varrho^2 \log\log m}} = 1 \quad \text{and} \quad \limsup_{m\to\infty} \frac{\left|S^*_{1,m}\right|}{\sqrt{2\varrho^{*2} \log\log m}} = 1 \quad a.s., \tag{5.76}$$

with $\varrho^2 = \sum_{h\in\mathbb{Z}} Cov(h_1(Y_0), h_1(Y_h))$ *and* $\varrho^{*2} = \sum_{h\in\mathbb{Z}} Cov(h^*_{1,m}(Z_{0,m}), h^*_{1,m}(Z_{h,m}))$. *Then, if* c_m *is chosen as in (5.75), it holds under the alternative*

$$P\left(\frac{\tilde{\tau}_m - a_m(S_{1,m}, S^*_{1,m})}{b_m(S_{1,m}, S^*_{1,m})} \leq x \,\middle|\, S_{1,m}, S^*_{1,m}\right) \to \Phi\left(\frac{x}{\sigma}\right) \quad a.s.,$$

where Φ is the distribution function of the standard Gaussian distribution, $\sigma^2 = \sum_{h\in\mathbb{Z}} Cov(h_2(Y_0), h_2(Y_h))$, $a_m = a_m(s_{1,m}, s^*_{1,m})$ is the unique solution of (5.66) and $b_m = b_m(s_{1,m}, s^*_{1,m})$ as in (5.70).

Corollary 5.21. *(i) If* $S_{1,m}$ *and* $S^*_{1,m}$ *fulfill a weak invariance principle with the usual rates, (5.76) is satisfied in a P-stochastic sense. By the subsequence principle this gives the result of Theorem 5.20 but with convergence in probability.*

(ii) With Lebesgue's dominated convergence theorem we obtain an unconditional result by

$$\begin{aligned} &P\left(\frac{\tilde{\tau}_m - a_m(S_{1,m}, S^*_{1,m})}{b_m(S_{1,m}, S^*_{1,m})} \leq x\right) \\ =&\mathrm{E}\left(P\left(\frac{\tilde{\tau}_m - a_m(S_{1,m}, S^*_{1,m})}{b_m(S_{1,m}, S^*_{1,m})} \leq x \,\middle|\, S_{1,m}, S^*_{1,m}\right)\right) \to \Phi\left(\frac{x}{\sigma}\right). \end{aligned}$$

The proofs of Theorem 5.17 and 5.20 are divided into several lemmas. In order to assess the supremum over k on $\{k^* < k \leq N\}$ we shift the index and consider the supremum over l on $\{1 < l \leq N - k^*\}$ which we split at $(1-\delta)(N-k^*)$ for a fixed $\delta \in (0,1)$.

Lemma 5.22. *Let the assumptions of Theorem 5.17 be satisfied. Then, for $\delta \in (0, \frac{1}{2})$ fixed, it holds for all $z \in \mathbb{R}$, as $m \to \infty$,*

$$\lim_{m\to\infty} P\left(e_m \left(\sup_{1\le l<(1-\delta)(N-k^*)} \frac{\left|\tilde{\Gamma}(m, k^*+l)\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} - d_m \right) \le z \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m} \right) = 1$$

with d_m and e_m as in (5.60) *and* (5.61) *and $\tilde{\Gamma}$ as in* (5.73).

Proof. As $\{X_1, \ldots, X_m\}$ are independent of $\{X_{m+j} : j \ge 1\}$ we obtain with Lemma 5.18 and d_m and e_m as in (5.60) and (5.61)

$$\begin{aligned}
&P\left(e_m \left(\sup_{1\le l<(1-\delta)(N-k^*)} \frac{|\Gamma(m, k^*+l)|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} - d_m \right) \le z \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m} \right) \\
&= P\left(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right) \left(\sup_{1\le l<(1-\delta)(N-k^*)} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})\right.}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \right.\right. \\
&\quad + \frac{\left. l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{l}{\sqrt{m}} s^*_{1,m} \right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad \left.\left. - \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\,\mathrm{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\,\mathrm{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)} \right) \le z \right).
\end{aligned} \tag{5.77}$$

The proof of this Lemma is based on showing that the above random variable is dominated by $-(N-k^*)|\Delta_m|$ which diverges to $-\infty$ as stated by the following considerations. With (5.72) we obtain

$$\begin{aligned}
&N - k^* \\
&= \frac{\sqrt{a_m}}{|\Delta_m|}x + \frac{c_m\sqrt{m}}{|\Delta_m|} + \frac{a_m}{\sqrt{m}\,|\Delta_m|}\left(c_m - s^*_{1,m}\,\mathrm{sign}(\Delta_m)\right) + \frac{k^*}{\sqrt{m}\Delta_m}(s^*_{1,m} - s_{1,m}) \\
&= \frac{c_m\sqrt{m}}{|\Delta_m|}\left(1+\frac{a_m}{m}\right)\left(1 + \frac{x\sqrt{a_m}}{c_m\sqrt{m}\left(1+\frac{a_m}{m}\right)} - \frac{s^*_{1,m}\,\mathrm{sign}(\Delta_m)}{c_m}\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}} \right. \\
&\quad \left. + \frac{\mathrm{sign}(\Delta_m)(s^*_{1,m} - s_{1,m})}{c_m}\frac{\frac{k^*}{m}}{1+\frac{a_m}{m}} \right) \\
&= \frac{c_m\sqrt{m}}{|\Delta_m|}\left(1+\frac{a_m}{m}\right)\left(1 + \frac{x}{c_m}\sqrt{\frac{m}{a_m}}\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}} + \frac{s^*_{1,m}\,\mathrm{sign}(\Delta_m)}{c_m}\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}}\left(\frac{k^*}{a_m} - 1\right) \right. \\
&\quad \left. - \frac{\mathrm{sign}(\Delta_m)s_{1,m}}{c_m}\frac{k^*}{a_m}\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}} \right)
\end{aligned} \tag{5.78}$$

It follows with Lemma 5.16 a) (iii) and (iv) as well as Assumption 5.14 (i) and (ii) that

$$N - k^* = \frac{c_m\sqrt{m}}{|\Delta_m|}\left(1+\frac{a_m}{m}\right)\left(1 - \frac{s_{1,m}}{c_m}\,\mathrm{sign}(\Delta_m) + o(1)\right) \tag{5.79}$$

as $\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}} = \frac{1}{\frac{m}{a_m}+1} \to 1$. Hence, with Assumption 5.13 (ii) and Assumption 5.14 (i), it holds in particular

$$N - k^* \to \infty \tag{5.80}$$

and

$$(N-k^*)|\Delta_m| = c_m\sqrt{m}\left(1+\frac{a_m}{m}\right)\left(1-\frac{s_{1,m}}{c_m}\operatorname{sign}(\Delta_m)+o(1)\right) \to \infty \tag{5.81}$$

and with Assumption 5.13 (iii)

$$\sqrt{N-k^*}\,|\Delta_m| = \sqrt{c_m\sqrt{m}\,|\Delta_m|\left(1+\frac{a_m}{m}\right)}\sqrt{1-\frac{s_{1,m}}{c_m}\operatorname{sign}(\Delta_m)+o(1)} \to \infty \tag{5.82}$$

as $m \to \infty$. Furthermore, it holds with Assumption 5.14 (i) and (5.79)

$$\begin{aligned} N^{-\frac{1}{2}}(N-k^*)\,|\Delta_m| &= c_m\sqrt{\frac{m}{N}}\left(1+\frac{a_m}{m}\right)\left(1-\frac{s_{1,m}}{c_m}\operatorname{sign}(\Delta_m)+o(1)\right) \\ =& c_m\left(\sqrt{\frac{m}{N}}+\sqrt{\frac{N}{m}}\frac{a_m}{N}\right)\left(1-\frac{s_{1,m}}{c_m}\operatorname{sign}(\Delta_m)+o(1)\right) \to \infty \quad \text{as } m\to\infty. \end{aligned} \tag{5.83}$$

With (5.81), Assumption 5.14 (i) and Lemma 5.16 a) (iii) and b) we obtain

$$\frac{\sqrt{m}\left(1+\frac{N}{m}\right)}{(N-k^*)|\Delta_m|} = \frac{1}{c_m}\frac{1+\frac{N}{m}}{1+\frac{a_m}{m}}\left(1-\frac{s_{1,m}}{c_m}\operatorname{sign}(\Delta_m)+o(1)\right)^{-1} = O\left(\frac{1}{c_m}\right) \tag{5.84}$$

such that it holds

$$\frac{\sqrt{m}\left(1+\frac{N}{m}\right)}{(N-k^*)|\Delta_m|}\sup_{1\le l<(1-\delta)(N-k^*)}\frac{\left|\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} = o_P(1) \tag{5.85}$$

if

$$\sup_{1\le l<(1-\delta)(N-k^*)}\frac{\left|\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} = O_P(1). \tag{5.86}$$

Assumption 5.15 (ii) with $k_m = N-k^* \to \infty$ by (5.80) as well as Lemma 5.16 a) (i), (iii), (iv) and b) yield

$$\begin{aligned} &\sup_{1\le l\le N-k^*}\frac{\left|\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\ \le& \sqrt{\frac{N-k^*}{m}}\frac{1}{1+\frac{k^*}{m}}O_P(1) \le \frac{\sqrt{\frac{N}{m}}}{1+\frac{k^*}{m}}O_P(1) = \sqrt{\frac{N}{k^*}}\frac{\sqrt{\frac{k^*}{m}}}{1+\frac{k^*}{m}}O_P(1) = o_P(1). \end{aligned} \tag{5.87}$$

Furthermore, we get with Assumption 5.15 (iii), Lemma 5.16 (i) and the stationarity of $\{Y_j\}$

$$\sup_{1\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \overset{\mathcal{D}}{=} \sup_{1\le l\le N-k^*} \frac{\left|\sum_{j=1}^{k^*} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}$$
$$=\frac{\sqrt{\frac{k^*}{m}}}{1+\frac{k^*}{m}} O_P(1) = \sqrt{\frac{m}{k^*}}\frac{\frac{k^*}{m}}{1+\frac{k^*}{m}} O_P(1) = o_P(1). \tag{5.88}$$

Combining (5.87) and (5.88) yields (5.86), where we even obtain $o_P(1)$, and thus we get (5.85) as mentioned above. Furthermore, it holds with (5.72), Assumption 5.14 (ii) and Lemma 5.16 a) (i) and (iv)

$$\frac{(N-k^*)\sqrt{m}|\Delta_m|}{s_{1,m}k^*\operatorname{sign}(\Delta_m)}$$
$$=\frac{x}{s_{1,m}}\sqrt{\frac{a_m}{k^*}}\sqrt{\frac{m}{k^*}}\operatorname{sign}(\Delta_m) + \frac{c_m}{s_{1,m}}\operatorname{sign}(\Delta_m)\left(\frac{m}{k^*}+\frac{a_m}{k^*}\right) + \frac{s_{1,m}^*}{s_{1,m}}\left(1-\frac{a_m}{k^*}\right) - 1$$
$$=\frac{c_m}{s_{1,m}}\operatorname{sign}(\Delta_m)\left(\frac{m}{k^*}+\frac{a_m}{k^*}\right) - 1 + o(1) = q_m - 1 + o(1), \tag{5.89}$$

where

$$q_m := \frac{c_m}{s_{1,m}}\operatorname{sign}(\Delta_m)\left(\frac{m}{k^*}+\frac{a_m}{k^*}\right). \tag{5.90}$$

With Lemma 5.16 a) (i), (iv) and Assumption 5.14 (i) it holds

$$\liminf_{m\to\infty} |q_m| > 1 \tag{5.91}$$

and thus

$$\liminf_{m\to\infty} (q_m-1)^{-1} > -\frac{1}{2} \tag{5.92}$$

as well as

$$\liminf_{m\to\infty} |q_m-1| > 0. \tag{5.93}$$

Hence, (5.89) implies

$$\frac{s_{1,m}k^*\operatorname{sign}(\Delta_m)}{(N-k^*)\sqrt{m}|\Delta_m|} = (q_m-1)^{-1} + o(1) \tag{5.94}$$

such that we obtain with Assumption 5.14 (iii)

$$(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s_{1,m}^*\operatorname{sign}(\Delta_m)$$
$$=(N-k^*)|\Delta_m|\left(1+\frac{s_{1,m}k^*\operatorname{sign}(\Delta_m)}{(N-k^*)\sqrt{m}|\Delta_m|} + \frac{s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}|\Delta_m|}\right)$$
$$=(N-k^*)|\Delta_m|\left(1+(q_m-1)^{-1}+o(1)\right) \tag{5.95}$$

and

$$\begin{aligned}
&\sup_{1\le l<(1-\delta)(N-k^*)} \left| l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{l}{\sqrt{m}} s_{1,m}^* \right| \\
=&(N-k^*)\left|\Delta_m\right| \sup_{1\le l<(1-\delta)(N-k^*)} \left| \frac{l}{N-k^*} + \frac{s_{1,m}k^*}{(N-k^*)\sqrt{m}\Delta_m} + \frac{l}{N-k^*}\frac{s_{1,m}^*}{\sqrt{m}\Delta_m} \right| \\
=&(N-k^*)\left|\Delta_m\right| \left(\sup_{1\le l<(1-\delta)(N-k^*)} \left| \frac{l}{N-k^*} + (q_m-1)^{-1} \right| + o(1) \right). \qquad (5.96)
\end{aligned}$$

The latter implies

$$\begin{aligned}
&\frac{\sqrt{m}\left(1+\frac{N}{m}\right)}{(N-k^*)|\Delta_m|} \sup_{1\le l<(1-\delta)(N-k^*)} \frac{\left| l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{l}{\sqrt{m}} s_{1,m}^* \right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
\le&\frac{1+\frac{N}{m}}{1+\frac{k^*+1}{m}} \sup_{1\le l<(1-\delta)(N-k^*)} \frac{\left| l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{l}{\sqrt{m}} s_{1,m}^* \right|}{(N-k^*)|\Delta_m|} \\
\le&\frac{\frac{m}{N}+1}{\frac{m}{N}+\frac{k^*}{N}} \left(\sup_{1\le l<(1-\delta)(N-k^*)} \left| \frac{l}{N-k^*} + (q_m-1)^{-1} \right| + o(1) \right) \\
=&(1+o(1)) \left(\sup_{1\le l<(1-\delta)(N-k^*)} \left| \frac{l}{N-k^*} + (q_m-1)^{-1} \right| + o(1) \right) \\
=&(1+o(1)) \left(\max\left\{ 1-\delta+(q_m-1)^{-1}, -\frac{1}{N-k^*} - (q_m-1)^{-1} \right\} + o(1) \right) \\
=&\max\left\{ 1-\delta+(q_m-1)^{-1}, -\frac{1}{N-k^*} - (q_m-1)^{-1} \right\} + o(1) \qquad (5.97)
\end{aligned}$$

as

$$\frac{\frac{m}{N}+1}{\frac{m}{N}+\frac{k^*}{N}} \to 1 \quad \text{as } m\to\infty \qquad (5.98)$$

with Lemma 5.16 a) (iii), (iv) and b) as well as

$$\max\left\{ 1-\delta+(q_m-1)^{-1}, -\frac{1}{N-k^*} - (q_m-1)^{-1} \right\} = O(1)$$

by (5.80) and (5.92). Now, we obtain with (5.85), (5.95) and (5.97)

$$\begin{aligned}
&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right) \left(\sup_{1\le l<(1-\delta)(N-k^*)} \frac{\left| \sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) \right.}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \right. \\
&+ \frac{\left. l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s_{1,m}^* \right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\left. - \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)} \right)
\end{aligned}$$

$$
\begin{aligned}
\leq & N^{-\frac{1}{2}}(N-k^*)|\Delta_m|\left(\max\left\{1-\delta+(q_m-1)^{-1},-\frac{1}{N-k^*}-(q_m-1)^{-1}\right\}\right.\\
&\left.-1-(q_m-1)^{-1}+o_P(1)\right)\\
=& N^{-\frac{1}{2}}(N-k^*)|\Delta_m|\left(\max\left\{-\delta,-1-2(q_m-1)^{-1}\right\}+o_P(1)\right). \qquad (5.99)
\end{aligned}
$$

As

$$
\limsup_{m\to\infty}\left(-1-2(q_m-1)^{-1}\right)<0
$$

by (5.92), the assertion follows with (5.99) and (5.83). □

Lemma 5.23. *Under the assumptions of Theorem 5.17 it holds for all $z\in\mathbb{R}$*

$$
\lim_{m\to\infty} P\left(e_m\left(\sup_{k^*<k<N}\frac{\left|\tilde{\Gamma}(m,k)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)\leq z\,\middle|\, S_{1,m}=s_{1,m},S^*_{1,m}=s^*_{1,m}\right)=\Phi\left(\frac{z}{\sigma}\right)
$$

with d_m and e_m as in (5.60) and (5.61) and $\tilde{\Gamma}$ as in (5.73).

Proof. We obtain with the Lemmas B.3 (i), 5.18 and 5.22

$$
\begin{aligned}
&P\left(e_m\left(\sup_{k^*<k\leq N}\frac{\left|\tilde{\Gamma}(m,k)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)\leq z\,\middle|\, S_{1,m}=s_{1,m},S^*_{1,m}=s^*_{1,m}\right)\\
=&P\left(e_m\left(\sup_{k^*<k\leq N}\frac{\left|\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k}h^*_{2,m}(Z_{j,m})\right.}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right.\right.\\
&\left.\left.+\frac{\left.(k-k^*)\Delta_m+\frac{k^*}{\sqrt{m}}s_{1,m}+\frac{k-k^*}{\sqrt{m}}s^*_{1,m}\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)\leq z\right)\\
=&P\left(e_m\left(\sup_{1\leq l\leq N-k^*}\frac{\left|\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h^*_{2,m}(Z_{j,m})\right.}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\right.\right.\\
&\left.\left.+\frac{\left.(k-k^*)\Delta_m+\frac{k^*}{\sqrt{m}}s_{1,m}+\frac{k-k^*}{\sqrt{m}}s^*_{1,m}\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)\leq z\right)
\end{aligned}
$$

$$
\begin{aligned}
= P\Bigg(& e_m \Bigg(\sup_{1\le l<(1-\delta)(N-k^*)} \frac{\Big|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
& + \frac{(k-k^*)\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{k-k^*}{\sqrt{m}} s^*_{1,m}\Big|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m \Bigg) \le z, \\
& e_m \Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\Big|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
& + \frac{(k-k^*)\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{k-k^*}{\sqrt{m}} s^*_{1,m}\Big|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m \Bigg) \le z, \Bigg) \\
= P\Bigg(& \sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right) \Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\Big|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
& + \frac{l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\Big|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
& - \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\,\mathrm{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\,\mathrm{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)} \Bigg) \le z \Bigg) + o(1).
\end{aligned}
\tag{5.100}
$$

We continue with showing that, for m large enough, the sign of $\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m}) + l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}$ is determined by the sign of Δ_m for all $(1-\delta)(N-k^*) \le l \le N-k^*$. We get with Assumption 5.15, Lemma 5.16 a) (iv) and b) as well as (5.80), (5.68), (5.82) and (5.83) that

$$
\begin{aligned}
& \sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \left| \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{l\Delta_m} \right| \\
\le & \sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\sqrt{k^*}}{|l\Delta_m|} \left| \frac{1}{\sqrt{k^*}} \sum_{j=m+1}^{m+k^*} h_2(Y_j) \right| \\
& + \sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\sqrt{N-k^*}}{l\,|\Delta_m|} \left| \frac{1}{\sqrt{N-k^*}} \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m}) \right| \\
= & \left(\frac{\sqrt{k^*}}{(N-k^*)\,|\Delta_m|} + \frac{1}{\sqrt{N-k^*}\,|\Delta_m|} \right) O_P(1) \\
= & \left(\frac{\sqrt{\frac{k^*}{N}}}{N^{-\frac{1}{2}}(N-k^*)\,|\Delta_m|} + \frac{1}{\sqrt{N-k^*}\,|\Delta_m|} \right) O_P(1) = o_P(1).
\end{aligned}
\tag{5.101}
$$

It holds with Assumption 5.14 (iii) and (5.94)

$$\begin{aligned}
&\inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\frac{k^*}{\sqrt{m}}s_{1,m} + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}}{l\Delta_m} \\
&= \inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{N-k^*}{l}\left(\frac{k^* s_{1,m}}{(N-k^*)\Delta_m\sqrt{m}} + \frac{s^*_{1,m}}{\sqrt{m}\Delta_m}\right) \\
&= \inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \left(\frac{N-k^*}{l}(q_m-1)^{-1} + o(1)\right) \\
&> -\frac{1}{2(1-\delta)} + o(1) \qquad (5.102)
\end{aligned}$$

for $m \ge m_0$ with (5.92). Together with (5.101) and Lemma B.3 (ii) we obtain for $(1-\delta)(N-k^*) \le l \le N-k^*$, $\epsilon < 1 - \frac{1}{2(1-\delta)}$, $m \ge m_0$

$$\begin{aligned}
&P\left(\inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \operatorname{sign}(\Delta_m)\left(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\right.\right. \\
&\quad \left.\left.+\frac{l\Delta_m + \frac{k^*}{\sqrt{m}}s_{1,m} + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\right) \ge 0\right) \\
&= P\left(\inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{l|\Delta_m|}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\left(1 + \frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{l\Delta_m}\right.\right. \\
&\quad \left.\left.+\frac{\frac{k^*}{\sqrt{m}}s_{1,m} + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}}{l\Delta_m}\right) \ge 0\right) \\
&\ge P\left(\inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{l|\Delta_m|}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\left(1 - \sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\left|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)}{l\Delta_m}\right.\right.\right. \\
&\quad \left.+\frac{\sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{l\Delta_m}\right| + \inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \left.\frac{\frac{k^*}{\sqrt{m}}s_{1,m} + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}}{l\Delta_m}\right) \ge 0 \Bigg| \\
&\quad \left.\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\left|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{l\Delta_m}\right| < \epsilon\right) + o(1) \\
&\ge P\left(\inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{l|\Delta_m|}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\left(1-\epsilon-\frac{1}{2(1-\delta)}+o(1)\right) \ge 0\right) + o(1) \to 1. \\
&\qquad (5.103)
\end{aligned}$$

as $m \to \infty$. Let B_m be the event that

$$\begin{aligned}
&\inf_{(1-\delta)(N-k^*)\le l\le N-k^*} \operatorname{sign}(\Delta_m)\left(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m}) + l\Delta_m}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\right. \\
&\quad \left.+\frac{\frac{k^*}{\sqrt{m}}s_{1,m} + \frac{N-k^*}{\sqrt{m}}s^*_{1,m}}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\right) \ge 0.
\end{aligned}$$

With (5.103) it holds

$$P(B_m) \to 1.$$

Let

$$z_m := z\left(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\right)^{-1} + \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}.$$

By Lemma B.3 (i) it follows

$$\begin{aligned}
&P\Bigg(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})\right.}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad + \frac{\left. l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad - \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\Bigg) \le z\Bigg) \\
=&P\Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\Bigg|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad + \frac{l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\Bigg| \le z_m\Bigg) \\
=&P\Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\Bigg|\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad + \frac{l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\Bigg| \le z_m, B_m\Bigg) + o(1) \\
=&P\Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\Bigg(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad + \frac{l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\Bigg) \le z_m, B_m\Bigg) + o(1) \\
=&P\Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\Bigg(\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad + \frac{l\Delta_m + \frac{k^*}{\sqrt{m}} s_{1,m} + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\Bigg) \le z_m\Bigg) + o(1) \\
=&P\Bigg(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\Bigg(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j) + \sum_{j=m+k^*+1}^{m+k^*+l} h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad + \frac{l|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
&\quad - \frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\Bigg) \le z\Bigg) + o(1). \quad (5.104)
\end{aligned}$$

In the following we will prove that

$$\begin{aligned}
&\Bigg|\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\Bigg(\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&+\frac{l|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&-\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\Bigg)\\
&-\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\Bigg|\\
=&o(1). \qquad (5.105)
\end{aligned}$$

Therefore, we find upper bounds $D_{1,m}$ and $D_{2,m}$ such that

$$\begin{aligned}
&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\Bigg(\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&+\frac{l|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&-\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\Bigg)\\
&-\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
\leq& D_{1,m}\xrightarrow{P}0 \qquad (5.106)
\end{aligned}$$

and

$$\begin{aligned}
&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&-\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\Bigg(\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&+\frac{l|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&-\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\Bigg)\leq D_{2,m}\xrightarrow{P}0. \qquad (5.107)
\end{aligned}$$

We obtain (5.106) as there exists an $m_0 \in \mathbb{N}$ such that

$$
\begin{aligned}
&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\right.\\
&+\frac{l|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&\left.-\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\right)\\
&-\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h^*_{2,m}(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
\le&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\frac{l|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\right.\\
&\left.-\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\right)=0\quad\text{for all } m\ge m_0
\end{aligned}
$$

as the supremum over

$$
\frac{l|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \tag{5.108}
$$

is obtained for $l = N-k^*$ as the following considerations show:

For $q_{1,m} = \frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)$ and $q_{2,m}=\sqrt{m}+\frac{k^*}{\sqrt{m}}$ (5.108) can be represented by

$$
\begin{aligned}
\frac{l|\Delta_m|+q_{1,m}}{\frac{l}{\sqrt{m}}+q_{2,m}}&=\sqrt{m}\frac{l|\Delta_m|+\sqrt{m}q_{2,m}|\Delta_m|-\sqrt{m}q_{2,m}|\Delta_m|+q_{1,m}}{l+\sqrt{m}q_{2,m}}\\
=&\sqrt{m}|\Delta_m|-\sqrt{m}\frac{\sqrt{m}q_{2,m}|\Delta_m|-q_{1,m}}{l+\sqrt{m}q_{2,m}}.
\end{aligned} \tag{5.109}
$$

By Assumption 5.14 (iii) and Lemma 5.16 a) (i), (iv) and b), there exists an $m_0 \in \mathbb{N}$ such that

$$
\begin{aligned}
\sqrt{m}q_{2,m}|\Delta_m|-q_{1,m}&=(m+k^*)|\Delta_m|-\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)-\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)\\
=&m|\Delta_m|\left(1+\frac{k^*}{m}-\frac{s_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}|\Delta_m|}\frac{k^*}{m}-\frac{N-k^*}{m}\frac{s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}|\Delta_m|}\right)\\
=&m|\Delta_m|\left(1+\frac{k^*}{m}\left(1-\frac{s_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}|\Delta_m|}-\frac{N}{k^*}\frac{s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}|\Delta_m|}+\frac{s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}|\Delta_m|}\right)\right)\\
=&m|\Delta_m|\left(1+\frac{k^*}{m}(1+o(1))\right)>0\quad\text{for all } m\ge m_0.
\end{aligned} \tag{5.110}
$$

Hence, (5.109) is increasing in l for m large enough and thus

$$\begin{aligned}
&\sup_{(1-\delta)(N-k^*)\le l\le N-k^*} \frac{l|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{l+k^*}{m}\right)}\\
=&\frac{(N-k^*)|\Delta_m| + \frac{k^*}{\sqrt{m}} s_{1,m}\operatorname{sign}(\Delta_m) + \frac{N-k^*}{\sqrt{m}} s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)} \quad \text{for all } m\ge m_0.
\end{aligned}$$

We continue with showing (5.107). First, observe that

$$\begin{aligned}
&\Bigg|\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*} h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&\quad -\operatorname{sign}(\Delta_m)\left(\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*} h_2(Y_j)\right)\Bigg|\\
\le&\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\left|\frac{1+\frac{N}{m}}{1+\frac{l+k^*}{m}}\left(\frac{1}{\sqrt{N}}\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right)\right|\\
&\quad+\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\left|\frac{1+\frac{N}{m}}{1+\frac{l+k^*}{m}}-1\right|\sqrt{\frac{k^*}{N}}O_P(1)\\
\le&\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\left|\frac{1+\frac{N}{m}}{1+\frac{l+k^*}{m}}\left(\frac{1}{\sqrt{N}}\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right)\right|\\
&\quad+\left|\frac{1+\frac{N}{m}}{1+(1-\delta)\frac{N}{m}+\delta\frac{k^*}{m}}-1\right|\sqrt{\frac{k^*}{N}}O_P(1)\\
\overset{\mathcal{D}}{=}&\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\left|\frac{1+\frac{N}{m}}{1+\frac{l+k^*}{m}}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{l} h_{2,m}^*(Z_{j,m})\right)\right|+o_P(1)\\
\le&\frac{1+\frac{N}{m}}{1+\frac{k^*}{m}}\sqrt{\frac{N-k^*}{N}}O_P(1)+o_P(1)=\frac{1+\frac{N}{m}}{1+\frac{k^*}{m}}\sqrt{1-\frac{k^*}{N}}O_P(1)+o_P(1)=o_P(1)
\end{aligned} \tag{5.111}$$

with Assumption 5.15 (iii) and (iv), (5.98) and Lemma 5.16 a) (i), (iii), (iv) and b) noting that

$$\begin{aligned}
&\left|\frac{1+\frac{N}{m}}{1+(1-\delta)\frac{N}{m}+\delta\frac{k^*}{m}}-1\right|=\frac{\delta\frac{N-k^*}{m}}{1+(1-\delta)\frac{N}{m}+\delta\frac{k^*}{m}}\\
\le&\frac{\delta\frac{N-k^*}{m}}{1+(1-\delta)\frac{N}{m}}=\frac{\delta\left(1-\frac{k^*}{N}\right)}{\frac{m}{N}+1-\delta}=o(1).
\end{aligned}$$

Hence, we obtain with Assumption 5.14 (iii) and Assumption 5.15 (ii) and (iii) by replacing the second supremum with the value at $l = [N - k^*]$

$$\begin{aligned}
&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&-\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\left(\sup_{(1-\delta)(N-k^*)\le l\le N-k^*}\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+l}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\right.\\
&+\frac{l|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\\
&\left.-\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\right)\\
\le&\operatorname{sign}(\Delta_m)\left(\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*}h_2(Y_j)\right)+o_P(1)\\
&-\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\left(\operatorname{sign}(\Delta_m)\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)+\sum_{j=m+k^*+1}^{m+k^*+\lfloor N-k^*\rfloor}h_{2,m}^*(Z_{j,m})}{\sqrt{m}\left(1+\frac{k^*+\lfloor N-k^*\rfloor}{m}\right)}\right.\\
&+\frac{\lfloor N-k^*\rfloor|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{k^*+\lfloor N-k^*\rfloor}{m}\right)}\\
&\left.-\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s_{1,m}^*\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\right)\\
=&\operatorname{sign}(\Delta_m)\sqrt{\frac{k^*}{N}}\left(1-\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right)O_P(1)\\
&-\operatorname{sign}(\Delta_m)\left(\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right)\sqrt{\frac{\lfloor N-k^*\rfloor}{N}}O_P(1)\\
&+\left(\frac{(N-k^*)\Delta_m}{\sqrt{N}}-\frac{[N-k^*]\Delta_m}{\sqrt{N}}\right)\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\\
&+\frac{|\Delta_m|}{\sqrt{N}}N\left(1-\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right)\left(1-\frac{k^*}{N}+\frac{k^*}{N}\frac{s_{1,m}}{|\Delta_m|\sqrt{m}}\operatorname{sign}(\Delta_m)\right.\\
&\left.+\left(1-\frac{k^*}{N}\right)\frac{s_{1,m}^*}{\sqrt{m}|\Delta_m|}\operatorname{sign}(\Delta_m)\right)\\
\le&\sqrt{\frac{k^*}{N}}\left|1-\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right|O_P(1)+\left|\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right|\sqrt{1-\frac{k^*}{N}}O_P(1)\\
&+\frac{|\Delta_m|}{\sqrt{N}}\left|\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right|+\frac{|\Delta_m|}{\sqrt{N}}N\left|1-\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right|O(1)=o_P(1)
\end{aligned}$$

with Lemma 5.16 a) (iii), (iv) and b) noting that

$$\left|\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right| \leq \frac{1+\frac{N}{m}}{1+\frac{N}{m}-\frac{1}{m}} = \frac{\frac{m}{N}+1}{\frac{m}{N}+1-\frac{1}{N}} = O(1)$$

and

$$\begin{aligned}
&\left|1-\frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}\right| = \frac{1+\frac{N}{m}}{1+\frac{\lfloor N-k^*\rfloor+k^*}{m}}-1\\
\leq&\frac{1+\frac{N}{m}}{1+\frac{N}{m}-\frac{1}{m}}-1 = \frac{\frac{1}{m}}{1+\frac{N}{m}-\frac{1}{m}} = \frac{1}{N}\frac{1}{\frac{m}{N}+1-\frac{1}{N}} = O\left(\frac{1}{N}\right) = o(1).
\end{aligned}$$

Having shown that (5.105) holds, it follows with (5.111)

$$\begin{aligned}
&\lim_{m\to\infty} P\left(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\left(\sup_{(1-\delta)(N-k^*)\leq l\leq N-k^*}\operatorname{sign}(\Delta_m)\left(\frac{\sum_{j=m+1}^{m+k^*}h_2(Y_j)}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\right.\right.\right.\\
&+\left.\frac{\sum_{j=m+k^*+1}^{m+k^*+l}h^*_{2,m}(Z_{j,m})+l\Delta_m+\frac{k^*}{\sqrt{m}}s_{1,m}+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)}\right)\\
&-\left.\left.\frac{(N-k^*)|\Delta_m|+\frac{k^*}{\sqrt{m}}s_{1,m}\operatorname{sign}(\Delta_m)+\frac{N-k^*}{\sqrt{m}}s^*_{1,m}\operatorname{sign}(\Delta_m)}{\sqrt{m}\left(1+\frac{N}{m}\right)}\right)\leq z\right)\\
=&\lim_{m\to\infty} P\left(\operatorname{sign}(\Delta_m)\left(\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*}h_2(Y_j)\right)\leq z\right).
\end{aligned}$$

With Assumption 5.15 (iii) and the stationarity as well as Lemma 5.13 a) (iv) and b) we get

$$\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*}h_2(Y_j) \stackrel{\mathcal{D}}{=} \sqrt{\frac{k^*}{N}}\frac{1}{\sqrt{k^*}}\sum_{j=m+1}^{m+k^*}h_2(Y_j) \xrightarrow{\mathcal{D}} N(0,\sigma^2).$$

Due to the symmetry of the Gaussian distribution, this still holds when multiplying with $\operatorname{sign}(\Delta_m)\in\{-1,1\}$. For $\lim_{m\to\infty}\operatorname{sign}(\Delta_m)=s_\Delta\in\{-1,1\}$ this is obvious. Otherwise, $\{\Delta_m\}_{m\geq 1}$ can be decomposed in two subsequences with $\lim_{n\to\infty}\operatorname{sign}(\Delta_{m_n})=-1$ and $\lim_{n\to\infty}\operatorname{sign}(\Delta_{m'_n})=1$ which both lead to the above limit distribution. Now, the assertion follows with (5.100) as

$$\lim_{m\to\infty} P\left(\operatorname{sign}(\Delta_m)\left(\frac{1}{\sqrt{N}}\sum_{j=m+1}^{m+k^*}h_2(Y_j)\right)\leq z\right) = \Phi\left(\frac{x}{\sigma}\right).$$

□

Proof of Theorem 5.17. In Example 3.1 we have seen that the remainder term of the CUSUM kernel is equal to zero such that the monitoring statistic can be represented as in (5.73). Hence, it holds for $\Gamma = \Gamma^C$, $\tilde{\tau}_m = \tilde{\tau}_m^C$, $a_m = a_m^C$ and $b_m = b_m^C$

$$
\begin{aligned}
&\lim_{m\to\infty} P\left(\frac{\tilde{\tau}_m - a_m}{b_m} \le x \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\right)\\
=&\lim_{m\to\infty} P\left(\tilde{\tau}_m \le x b_m + a_m \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\right)\\
=&1 - \lim_{m\to\infty} P\left(\tilde{\tau}_m > x b_m + a_m \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\right)\\
=&1 - \lim_{m\to\infty} P\left(\tilde{\tau}_m > N \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\right)\\
=&1 - \lim_{m\to\infty} P\left(\sup_{k^*<k\le N} \frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} \le c_m \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\right)\\
=&1 - \lim_{m\to\infty} P\left(e_m\left(\sup_{k^*<k\le N} \frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right) < e_m\left(c_m - d_m\right) \,\middle|\, S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m}\right)\\
=&1 - \Phi\left(-\frac{x}{\sigma}\right) = \Phi\left(\frac{x}{\sigma}\right)
\end{aligned}
\tag{5.112}
$$

with Lemma 5.23 and Lemma 5.16 c), where d_m and e_m are given by □

For the proof of Theorem 5.20 we additionally need the following result on the negligibility of the remainer term.

Lemma 5.24. *Let Assumption 3.3 (i), 5.15 (i) as well as 5.13 be satisfied. Then it holds, as $m \to \infty$,*

$$\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\le N}\frac{1}{\sqrt{m}\left(1+\frac{k}{m}\right)}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*} r(Y_i,Y_j)\right| = o_P(1),$$

$$\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\le N}\frac{1}{\sqrt{m}\left(1+\frac{k}{m}\right)}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k} r^*_m(Y_i,Z_{j,m})\right| = o_P(1).$$

Proof. It holds

$$\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\le N}\frac{1}{\sqrt{m}\left(1+\frac{k}{m}\right)} = \frac{1}{\sqrt{N}}\frac{1+\frac{N}{m}}{1+\frac{k^*}{m}} = \frac{1}{\sqrt{N}}\frac{\frac{m}{N}+1}{\frac{m}{N}+\frac{k^*}{N}} = O\left(\frac{1}{\sqrt{N}}\right) \tag{5.113}$$

as $m \to \infty$ with Lemma 5.16 a) (iii),(iv) and b). Hence, we obtain with (5.30)

$$\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\le N}\frac{1}{\sqrt{m}\left(1+\frac{k}{m}\right)}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*} r(Y_i,Y_j)\right| = \sqrt{\frac{k^*}{N}}o_P(1) = o_P(1) \tag{5.114}$$

as $m \to \infty$ with Lemma 5.16 a) (iv). Using (5.32), we obtain with Lemma 5.16 a) (i),

(iv) and b)

$$\begin{aligned}
&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\frac{1}{\sqrt{m}\left(1+\frac{k}{m}\right)}\left|\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k}r_m^*(Y_i,Z_{j,m})\right|\\
=&\frac{1}{\sqrt{N}}\left(1+\frac{N}{m}\right)m^{-\gamma}N^{\frac{1}{2}}\frac{1}{1+\frac{k^*}{m}}o_P(1)\\
=&\left(\frac{m}{k^*}+\frac{N}{k^*}\right)m^{-\gamma}o_P(1)=o_P(1)\quad\text{as } m\to\infty. \qquad (5.115)
\end{aligned}$$

□

Proof of Theorem 5.20. Consider $\tilde{\Gamma}$ as in (5.73). We obtain with Lemma 5.23 and (5.76) for all $z\in\mathbb{R}$

$$\lim_{m\to\infty}P\left(e_m\left(\sup_{k^*<k<N}\frac{\left|\tilde{\Gamma}(m,k)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)\leq z\,\middle|\,S_{1,m},S_{1,m}^*\right)=\Phi\left(\frac{z}{\sigma}\right)\quad\text{a.s.}. \qquad (5.116)$$

With the reverse triangle inequality and (3.45) it holds

$$\begin{aligned}
&\left|e_m\left(\sup_{k^*<k\leq N}\frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)-e_m\left(\sup_{k^*<k\leq N}\frac{\left|\tilde{\Gamma}(m,k)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)\right|\\
\leq&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\left|\frac{\Gamma(m,k)-\tilde{\Gamma}(m,k)}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right|\\
=&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\left|\frac{\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*}r(Y_i,Y_j)+\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k}r_m^*(Y_i,Z_{j,m})}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right|\\
\leq&\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\left|\frac{\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*}r(Y_i,Y_j)}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right|\\
&+\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\left|\frac{\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k}r_m^*(Y_i,Z_{j,m})}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right|. \qquad (5.117)
\end{aligned}$$

We obtain with Lemma 5.24 and the law of iterated expectations that

$$\begin{aligned}
&\mathrm{E}\left(P\left(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\left|\frac{\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*}r(Y_i,Y_j)}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right|>\epsilon\,\middle|\,S_{1,m},S_{1,m}^*\right)\right)\\
=&P\left(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\left|\frac{\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*}r(Y_i,Y_j)}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right|>\epsilon\right)\to 0\quad\text{as } m\to\infty
\end{aligned}$$

for all $\epsilon>0$ such that it follows with Markov's inequality

$$0\leq P\left(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\leq N}\left|\frac{\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+1}^{m+k^*}r(Y_i,Y_j)}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right|>\epsilon\,\middle|\,S_{1,m},S_{1,m}^*\right)\xrightarrow{P}0$$

as $m \to \infty$ and analogously

$$P\left(\sqrt{\frac{m}{N}}\left(1+\frac{N}{m}\right)\sup_{k^*<k\le N}\left|\frac{\frac{1}{m}\sum_{i=1}^{m}\sum_{j=m+k^*+1}^{m+k} r_m^*(Y_i, Z_{j,m})}{\sqrt{m}\left(1+\frac{k}{m}\right)}\right| > \epsilon \,\middle|\, S_{1,m}, S_{1,m}^*\right) \xrightarrow{P} 0.$$

Hence, we get with (5.117) for all $\epsilon > 0$

$$P\left(\left|e_m\left(\sup_{k^*<k\le N}\frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right) - e_m\left(\sup_{k^*<k\le N}\frac{\left|\tilde{\Gamma}(m,k)\right|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right)\right| > \epsilon \,\middle|\, S_{1,m}, S_{1,m}^*\right)$$
$$\xrightarrow{P} 0 \quad \text{as } m \to \infty. \tag{5.118}$$

Combining (5.116) and (5.118) we obtain with Slutsky's theorem and the subsequence principle that

$$P\left(e_m\left(\sup_{k^*<k<N}\frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)} - d_m\right) \le z \,\middle|\, S_{1,m}, S_{1,m}^*\right) \xrightarrow{P} \Phi\left(\frac{z}{\sigma}\right) \quad \text{as } m \to \infty. \tag{5.119}$$

Now, the assertion follows analogously to (5.112). □

5.3. Linear Changes

In this section we consider the asymptotic behavior of the delay time for change points that are linear in m where the approach is the same as in the previous section. We refer to the change point model as described in Section 2.2 and use the weight function in (3.5) for $\gamma = 0$.

Assumption 5.25.

(i) There exists a $\lambda > 0$ such that $k^* = [\lambda m]$.

(ii) $\Delta_m = O(1)$.

(iii) $\lim_{m\to\infty} \frac{c_m}{\sqrt{m}|\Delta_m|} = 0$.

Assumption 5.26.

(i) $c_m \to \infty$ *and* $\limsup_{m\to\infty} \left|\frac{s_{1,m}}{c_m}\right| < 1 + \frac{1}{\lambda}$.

(ii) $\frac{s_{1,m}^*}{s_{1,m}} = O(1)$ *and* $\frac{1}{s_{1,m}} = O(1)$.

(iii) $\lim_{m\to\infty} \frac{|s_{1,m}|}{\sqrt{m}|\Delta_m|} = \lim_{m\to\infty} \frac{|s_{1,m}^*|}{\sqrt{m}|\Delta_m|} = 0$.

Theorem 5.27. *Let Assumptions 3.3 (ii), 5.15 and 5.25 be satisfied. Furthermore assume that $\{X_1, \ldots, X_m\}$ are independent of $\{X_{m+j} : j \geq 1\}$. Then, for all sequences $s_{1,m}$ and $s^*_{1,m}$ that fulfill Assumption 5.26 it holds under the alternative for the CUSUM kernel*

$$\lim_{m\to\infty} P\left(\frac{\tilde{\tau}^C_m - a^C_m}{b^C_m} \leq x \middle| S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m} \right) = \Phi\left(\frac{x}{\sigma}\right) \quad \text{for all } x \in \mathbb{R},$$

where Φ is the distribution function of the standard Gaussian distribution, $\sigma^2 = \sum_{h\in\mathbb{Z}} Cov(Y_0, Y_h)$, $a^C_m = a^C_m(s_{1,m}, s^*_{1,m})$ is the unique solution of (5.66) and $b^C_m = b^C_m(s_{1,m}, s^*_{1,m})$ as in (5.70).

Remark 5.28. *Assumption 5.26 is satisfied if*

$$\limsup_{m\to\infty} \frac{s_{1,m}}{\sqrt{2\tau^2 \log\log m}} = \frac{s^*_{1,m}}{\sqrt{2\tilde{\tau}^2 \log\log m}} = 1 \tag{5.120}$$

$$c_m = \left(\sqrt{2\max(\tau^2, \tilde{\tau}^2)\frac{\lambda}{\lambda+1}} + \epsilon \right) \sqrt{\log\log m} \tag{5.121}$$

As the asymptotic negligibility of the remainder term in Lemma 5.24 follows analogously for linear changes, Theorem 5.20 and Corollary 5.21 also hold for linear changes if c_m is chosen as in (5.121). In the following we refer to the deterministic sequences as defined in the previous Section.

Lemma 5.29. *Under Assumption 5.1 with $\beta = 1$ it holds*

a) *(i)* $\frac{k^*}{m} \to \lambda$
   (ii) $\sqrt{a_m}\,|\Delta_m| \to \infty$
   (iii) $\frac{a_m}{m} \to \lambda$
   (iv) $\frac{k^*}{a_m} \to 1$

b) $\frac{N}{a_m} \to 1$ *such that part a) is still valid when replacing a_m by N.*

c) $\lim_{m\to\infty} e_m\,(c_m - d_m) = -x$ *for all $x \in \mathbb{R}$.*

Proof. Part a) (i) follows directly with Assumption 5.25 (i). Furthermore, note that (5.67)-(5.72) also hold for $\beta = 1$ and immediately imply the remaining assertions in a), where (ii) is already given in (5.69). Furthermore, b) is given in (5.71) and c) holds as the sequences are derived under exactly this condition. □

Lemma 5.30. *Let the assumptions of Theorem 5.27 be satisfied. Then, for $\delta \in (0, \frac{1}{2})$ fixed, it holds for all $z \in \mathbb{R}$, as $m \to \infty$,*

$$\lim_{m\to\infty} P\left(e_m \left(\sup_{1\leq l<(1-\delta)(N-k^*)} \frac{|\Gamma(m, k^*+l)|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} - d_m \right) \leq z \middle| S_{1,m} = s_{1,m}, S^*_{1,m} = s^*_{1,m} \right) = 1.$$

Proof. Most of this proof is analogous to the proof of Lemma 5.22. In the following we will point out the relevant differences. First, we obtain with (5.78), Assumption 5.26 (i) and (ii) as well as Lemma 5.29 a) (iii) and (iv)

$$\begin{aligned}
&N - k^* \\
=&\frac{c_m\sqrt{m}}{|\Delta_m|}\left(1+\frac{a_m}{m}\right)\left(1+\frac{x}{c_m}\sqrt{\frac{m}{a_m}}\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}}+\frac{s_{1,m}^*\operatorname{sign}(\Delta_m)}{c_m}\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}}\left(\frac{k^*}{a_m}-1\right)\right. \\
&\left.-\frac{\operatorname{sign}(\Delta_m)s_{1,m}}{c_m}\frac{k^*}{a_m}\frac{\frac{a_m}{m}}{1+\frac{a_m}{m}}\right) \\
=&\frac{c_m\sqrt{m}}{|\Delta_m|}\left(1+\lambda+o(1)\right)\left(1-\frac{s_{1,m}}{c_m}\frac{\lambda}{1+\lambda}\operatorname{sign}(\Delta_m)+o(1)\right) \qquad (5.122)
\end{aligned}$$

such that it follows with Assumption 5.26 (i), as $m \to \infty$,

$$\sqrt{N-k^*}\,|\Delta_m| \to \infty \qquad (5.123)$$

and

$$N^{-\frac{1}{2}}(N-k^*)\,|\Delta_m| \to \infty. \qquad (5.124)$$

Furthermore, note that for q_m as in (5.90) it holds

$$|q_m| = \left|\frac{c_m}{s_{1,m}}\right|\left(1+\frac{1}{\lambda}+o(1)\right).$$

With Assumption 5.14 (i) there exist $C < 1, m_0 \in \mathbb{N}$ such that

$$|q_m|^{-1} \le C < 1 \quad \text{for all } m \ge m_0$$

and thus

$$|q_m| \ge \frac{1}{C} > 1 \quad \text{for all } m \ge m_0.$$

Now, (5.92) follows as before. As (5.84) holds with Assumption 5.14 (i) and Lemma 5.29 a) (iii) and b), the stochastic terms need to be controlled in terms of (5.86) which is obtained as follows. Assumption 5.15 (ii) with $k_m = N - k^* \to \infty$ by (5.80) as well as Lemma 5.29 a) (i), (iv) and b) yield

$$\begin{aligned}
&\sup_{1\le l\le N-k^*}\frac{\left|\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \\
\le&\sqrt{\frac{N-k^*}{m}}\frac{1}{1+\frac{k^*}{m}}O_P(1) = \sqrt{\frac{N}{m}-\frac{k^*}{m}}\frac{1}{1+\lambda+o(1)}O_P(1) = o_P(1). \qquad (5.125)
\end{aligned}$$

Furthermore, we get with Assumption 5.15 (iii), Lemma 5.29 (i) and the stationarity of $\{Y_j\}$

$$\sup_{1\le l\le N-k^*}\frac{\left|\sum_{j=m+1}^{m+k^*} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} \overset{\mathcal{D}}{=} \sup_{1\le l\le N-k^*}\frac{\left|\sum_{j=1}^{k^*} h_2(Y_j)\right|}{\sqrt{m}\left(1+\frac{k^*+l}{m}\right)} = \frac{\sqrt{\frac{k^*}{m}}}{1+\frac{k^*}{m}}O_P(1) = O_P(1). \qquad (5.126)$$

□

Lemma 5.31. *Under the assumptions of Theorem 5.27 it holds for all* $z \in \mathbb{R}$

$$\lim_{m\to\infty} P\left(e_m\left(\sup_{k^*<k<N}\frac{|\Gamma(m,k)|}{\sqrt{m}\left(1+\frac{k}{m}\right)}-d_m\right)\leq z\,\middle|\, S_{1,m}=s_{1,m}, S^*_{1,m}=s^*_{1,m}\right)=\Phi\left(\frac{z}{\sigma}\right).$$

Proof. This proof is analogous to the proof of Lemma 5.23. In the following we give some remarks on those steps that have to be checked for the linear case. (5.101) follows with Lemma 5.29 a) (iv) and b) as well as (5.123), (5.124). (5.102) is obtained by (5.92) which also holds in the linear case as shown in the proof of Lemma 5.30. (5.110) holds with Assumption 5.26 (iii) and Lemma 5.29 a) (i), (iv) and b). Now, (5.105) and finally the assertion are obtained in the same way as in the proof of Lemma 5.23 by using Assumption 5.15 (i) and (ii) as well as Lemma 5.29. □

Proof of Theorem 5.27. With Lemma 5.31 and Lemma 5.29 c) the assertion follows analogously to (5.112). □

6. Simulation Study

In this chapter, we assess the finite sample behavior of the proposed sequential procedures for the CUSUM and the Wilcoxon kernel. The empirical size, power and stopping time are analyzed in Section 6.1 for all monitoring schemes that have been considered in this work. The theoretical results on the asymptotic behavior of the stopping time are validated empirically in Section 6.2. Throughout this chapter we consider the mean change model as in (2.8) where the size of the change is fixed. More precisely, we generate time series of length $m + Nm$ by

$$X_i = Y_i + 1_{\{i>k^*+m\}}d, \quad , i = 1, \ldots, m + Nm, \tag{6.1}$$

where $\{Y_i\}_{i\geq 1}$ is stationary with mean zero. Under the alternative, we insert a mean change at $k^* < Nm$, whereas $k^* = Nm$ under the null hypothesis. As historic data set we use $X_1, \ldots, X_m$. The remaining observations $X_{m+1}, \ldots, X_{m+Nm}$ are used for monitoring. In particular, this means that we stop monitoring after Nm observations even if no change has been detected. Hence, the supremum in the test statistics is approximated by a maximum over Nm observations. We consider the weight function in (3.5).

6.1. Comparison of Monitoring Schemes and Kernels

In the following, we compare the performance of the sequential procedures for different monitoring schemes as well as for different kernels via their empirical size, power and stopping time. For the CUSUM kernel, the monitoring schemes have already been compared in a simulation study in Kirch & Weber (2018). The present simulation study does not only add the comparison of the monitoring schemes for the Wilcoxon kernel but also allows to compare the CUSUM and the Wilcoxon kernel in various scenarios, in particular with respect to their robustness.

The results in this section are based on simulations of independent observations $\{Y_i\}_{i\geq 1}$. We use historic data sets of length $m = 100$ and a monitoring horizon of 2000, i.e. $N = 20$. For the simulations of the alternative we add a mean change of size $d = 0.5$ at $k^* = m^\beta$, $\beta = 0.5, 1, 1.4$. The asymptotic critical values are obtained based on 50 000 realizations of the limit distributions in Corollary 3.7 (i) and Corollary 4.3 b) where the Wiener processes are approximated on a grid of 10 000 points. σ_C^2 and σ_W^2 for independent data are given in (5.54) and (5.55). For the Wilcoxon kernel, the variance $\sigma_W^2 = Var(F_{Y_1}(Y_1))$ is known to be $\frac{1}{12}$ for any continuous distribution. In order to obtain a fair comparison we also use the true variance $\sigma_C^2 = \mathrm{Var}(Y_1)$. The empirical results are obtained based on 10 000 replications for each scenario. The simulations in Kirch & Weber (2018) have shown that in particular the modified MOSUM

Monitoring scheme	**CUSUM kernel** $\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$	**Wilcoxon kernel** $\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$
CUSUM	4.50	4.54	3.24	4.26	4.40	3.13
Page-CUSUM	4.40	4.27	2.59	4.25	4.18	2.52
mMOSUM						
$h = 0.1$	4.46	4.55	3.19	4.35	4.48	2.91
$h = 0.4$	4.44	4.34	2.61	4.84	4.31	2.25
$h = 0.9$	3.46	3.81	2.88	2.09	0.86	0.03

Table 6.1.: Empirical size (in %) for a nominal level of $\alpha = 5\%$ for independent $N(0,1)$-distributed observations.

with large h and $\gamma \neq 0$ raises a lot of false alarms at the beginning of the monitoring period. The reason for that is that the first values of monitoring statistic are based on very few observations (even only one for the first 10 monitoring time points with $h = 0.9$) and, additionally, those values are heavily weighted for $\gamma \neq 0$. In order to avoid this effect, we wait for a_m observations before we start monitoring which is allowed by Assumption 3.2 (i). Although false alarms at the beginning of the monitoring period occur less often for the other monitoring schemes and $\gamma = 0$, it is still beneficial to discard the first values of the monitoring statistic in order to obtain a more stable procedure. Hence, we choose $a_m = \sqrt{m} = 10$ in all settings. As the empirical size usually differs from the nominal level, comparing different procedures based on their empirical power and stopping time obtained for a given nominal level is inappropriate. For example, the empirical power is underestimated if the empirical size is smaller than the nominal level. Hence, we consider the size-corrected power and the size-corrected stopping time which are obtained by fixing the empirical level instead of the nominal level. Additionally, we correct the estimated density plots of the stopping time in such a way that the superiority of procedures with higher empirical power is visible. Such a correction is obtained by scaling the estimated density such that it integrates to the size-corrected power instead of one which corresponds to accounting for those changes that are not detected by setting the respective stopping time to infinity.

Table 6.1 shows the empirical size for independent standard normally distributed observations for the CUSUM as well as for the Wilcoxon kernel using different monitoring schemes and $\gamma = 0, 0.25, 0.45$ in the weight function. It is first of all apparent that the nominal level of 5% is maintained by all procedures. Except for the modified MOSUM (mMOSUM) with $h = 0.9$, the empirical size obtained with $\gamma = 0$ is close to the nominal level. In most of the cases, the empirical size decreases and in particular moves away from the nominal level for increasing γ. In this regard, recall that the null hypothesis is rejected as soon as the monitoring statistic exceeds the critical curve (see (1.1)). The critical curves are illustrated in Figure 6.1 for the CUSUM monitoring scheme but look similar for the others as only the critical values differ. Taking into account that we consider a monitoring horizon of 2000, the critical curves cross relatively early and afterwards the critical curve for $\gamma = 0$ is always below the others which leads to more rejections in the long term but still in a conservative way under the null hypothesis. The critical curve for $\gamma = 0.45$ clearly exceeds the others already

Monitoring scheme	CUSUM kernel $\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$	Wilcoxon kernel $\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$
CUSUM	4.37	4.65	4.18	4.39	4.36	3.12
Page-CUSUM	4.29	4.64	3.68	4.27	4.09	2.37
mMOSUM						
$h = 0.1$	4.55	5.02	4.13	4.51	4.34	2.78
$h = 0.4$	5.45	5.77	4.46	4.46	3.80	1.94
$h = 0.9$	18.47	20.99	16.78	2.30	0.98	0.03

Table 6.2.: Empirical size (in %) for a nominal level of $\alpha = 5\%$ for independent $t(3)$-distributed observations.

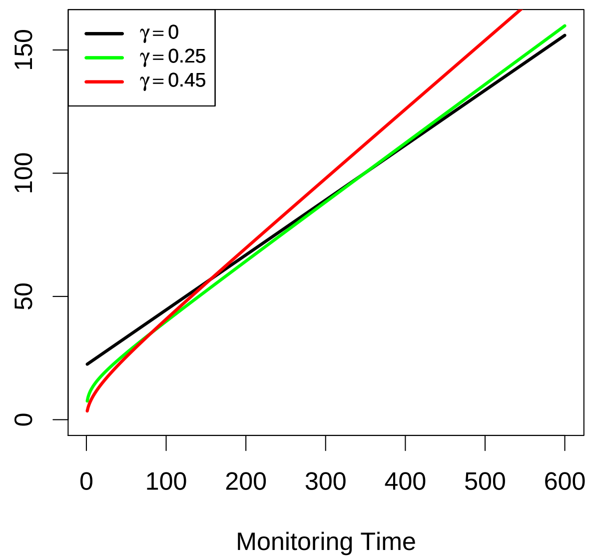


Figure 6.1.: Critical curve for the CUSUM monitoring scheme at a level of $\alpha = 5\%$.

at the beginning of the monitoring period and thus yields the smallest empirical size. For $\gamma = 0.25$, the critical curve crosses somewhat later such that the empirical size is closer to the one for $\gamma = 0$ and even slightly higher in a few cases. The empirical size of the modified MOSUM is decreasing in h in most of the settings. As we wait for $a_m = 10$ observations, high rejection rates for $h = 0.9$ are avoided. The Wilcoxon kernel yields a more conservative testing procedure than the CUSUM kernel in almost all considered cases. The modified MOSUM in combination with the Wilcoxon kernel has an extremely low empirical size for $h = 0.9$ and $\gamma \neq 0$. In this context, it should be mentioned that a larger monitoring horizon might be needed for large h in order to exploit the full size. Using the CUSUM kernel when the observations follow a $t(3)$ distribution (see table 6.2, the empirical size of the mMOSUM exceeds the nominal level slightly in some situations for $h = 0.1$ and $h = 0.4$ and drastically for $h = 0.9$. However, the Wilcoxon kernel still yields conservative procedures.

CUSUM kernel

Monitoring scheme	$\boldsymbol{\beta = 0.25}$ $(\boldsymbol{k^* = 3})$			$\boldsymbol{\beta = 1}$ $(\boldsymbol{k^* = 100})$			$\boldsymbol{\beta = 1.4}$ $(\boldsymbol{k^* = 630})$		
	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = 0.25}$	$\boldsymbol{\gamma = 0.45}$	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = 0.25}$	$\boldsymbol{\gamma = 0.45}$	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = 0.25}$	$\boldsymbol{\gamma = 0.45}$
CUSUM	99.82	99.71	99.37	99.37	99.16	98.33	88.45	85.61	78.04
Page-CUSUM	99.80	99.74	99.34	99.58	99.43	98.83	96.33	93.81	85.77
mMOSUM									
$h = 0.1$	99.77	99.68	99.16	99.64	99.55	99.00	93.64	91.24	83.53
$h = 0.4$	99.71	99.46	98.18	99.65	99.30	97.72	99.58	98.87	95.08
$h = 0.9$	93.49	79.20	41.65	92.47	72.70	20.65	78.84	38.11	5.81

Wilcoxon kernel

Monitoring scheme	$\boldsymbol{\beta = 0.25}$ $(\boldsymbol{k^* = 3})$			$\boldsymbol{\beta = 1}$ $(\boldsymbol{k^* = 100})$			$\boldsymbol{\beta = 1.4}$ $(\boldsymbol{k^* = 630})$		
	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = 0.25}$	$\boldsymbol{\gamma = 0.45}$	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = 0.25}$	$\boldsymbol{\gamma = 0.45}$	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = 0.25}$	$\boldsymbol{\gamma = 0.45}$
CUSUM	99.68	99.59	99.00	99.10	98.84	97.79	86.52	83.59	74.78
Page-CUSUM	99.66	99.55	99.02	99.39	99.13	98.20	94.68	91.65	82.25
mMOSUM									
$h = 0.1$	99.64	99.52	98.86	99.54	99.33	98.48	92.40	89.30	80.67
$h = 0.4$	99.45	99.10	97.62	99.39	98.92	96.80	99.21	98.16	93.47
$h = 0.9$	91.58	80.55	52.72	91.00	76.98	38.45	76.50	44.73	9.57

Table 6.3.: Size corrected power (in %) for independent $N(0,1)$-distributed observations.

The size-corrected power for independent standard normally observations is reported in Table 6.3. $\beta = 0.25$, i.e. $k^* = 3$ indicates an extremely early change. Recall that we start monitoring with a delay of 10 observations such that this change occurs so early that the monitoring sample does not even contain null observations.

First of all, we would like to point out that $\gamma = 0$ performs best in all situations. This can again be explained by the interplay of the weight functions as presented in Figure 6.1. The superiority of $\gamma = 0$ is particularly strong for late changes ($\beta = 1.4$) which is not surprising as the respective critical curve is already considerably below the others when the change occurs. More surprisingly, it also has the best power for early changes which occur at a time where the critical curve for $\gamma = 0$ is still above the others. This indicates that early changes are detected either very quickly or not at all for $\gamma \neq 0$, whereas for $\gamma = 0$ it might take a bit longer but the change will be detected at some point with a very high reliability. The modified MOSUM with $h = 0.9$ has a very poor power when using $\gamma = 0.45$, in particular for late changes. The modified MOSUM with $h = 0.4$ has the most stable power with respect to the time of the change for both kernels. In most cases, the Wilcoxon kernel yields a slightly smaller power than the CUSUM kernel which is nevertheless close to one except for the just mentioned problems of the modified MOSUM for large h and γ. However, the Wilcoxon kernel is advantageous if the observations are generated by a heavy-tailed distribution. This can clearly be seen in Table 6.4 which shows the size corrected power for $t(3)$-distributed observations. Whereas the Wilcoxon kernel has size corrected power close to one in most of the situations, the CUSUM kernel yields a rather poor power.
As for sequential procedures not only the power but also the speed of detection is of interest, we additionally consider the size-corrected stopping time which is presented

CUSUM kernel

Monitoring scheme	$\beta = 0.25$ ($k^* = 3$)			$\beta = 1$ ($k^* = 100$)			$\beta = 1.4$ ($k^* = 630$)		
	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$
CUSUM	86.75	81.59	64.35	79.74	71.96	50.17	30.47	22.18	9.30
Page-CUSUM	86.28	79.53	61.40	80.68	70.94	48.32	30.82	20.52	8.22
mMOSUM									
$h = 0.1$	84.46	76.66	55.25	83.84	74.88	51.12	38.47	26.57	10.47
$h = 0.4$	69.41	50.54	19.00	69.19	49.75	17.53	61.78	40.25	11.02
$h = 0.9$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Wilcoxon kernel

Monitoring scheme	$\beta = 0.25$ ($k^* = 3$)			$\beta = 1$ ($k^* = 100$)			$\beta = 1.4$ ($k^* = 630$)		
	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$
CUSUM	99.67	99.61	99.02	99.07	98.87	97.81	86.39	83.73	75.00
Page-CUSUM	99.66	99.55	99.09	99.39	99.13	98.36	94.61	91.63	83.24
mMOSUM									
$h = 0.1$	99.64	99.54	98.94	99.54	99.35	98.59	91.76	89.44	81.38
$h = 0.4$	99.49	99.20	97.77	99.43	99.05	97.02	99.26	98.46	93.84
$h = 0.9$	91.60	80.98	53.06	91.01	77.43	38.74	76.53	45.20	9.71

Table 6.4.: Size corrected power (in %) for independent $t(3)$-distributed observations.

in Figure 6.2 for independent standard normally distributed observations. We have seen that $\gamma = 0$ leads to the best size-corrected power in all situations. The estimated density plots show the only advantage of $\gamma = 0.45$ namely the shorter detection delay for extremely early changes ($\beta = 0.25$). However, the speed of detection for $\gamma = 0$ is also entirely reasonable in this case and already for $\beta = 0.75$, which is still a rather early change, it is only slightly slower than for $\gamma = 0.45$, in particular for the modified MOSUM with $h = 0.4$. For later changes, $\gamma = 0$ yields shorter detection delays and $\gamma = 0.45$ causes a higher amount of false alarms before the actual change occurs. We will now have a closer look at the stopping times for $\gamma = 0$. For both kernels, as intended by construction, the Page-CUSUM and the modified MOSUM outperform the CUSUM monitoring scheme for late changes ($\beta = 1$ and $\beta = 1.4$), the Page-CUSUM and the modified MOSUM even for earlier changes ($\beta = 0.75$). Regarding the parameters of the modified MOSUM, small values of h yield a shorter detection delay for early changes and large values of h for late changes. Hence, if one has a rough expectation on the time of the change, the parameter h can be chosen accordingly. However, it should be noted that using large h for a quicker detection of late changes comes with a considerable loss of power. In all situations, $h = 0.4$ yields good results and seems to be a good compromise. Using the Wilcoxon kernel, changes are detected somewhat slower than using the CUSUM kernel which is, however, due to the Gaussian distribution. For $t(3)$-distributed random variables, the Wilcoxon kernel is clearly superior to the CUSUM kernel as can be seen in Figure E.1.

6.1.1. Robustness

In order to assess the robustness of the testing procedures we randomly replaced 1% of the simulated observations by independent realizations of a gamma distribution with shape parameter 5 and scale parameter 10. Table 6.5 reports the resulting empirical

Monitoring scheme	CUSUM kernel			Wilcoxon kernel		
	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.45$
CUSUM	99.92	99.88	99.76	4.44	4.49	3.38
Page-CUSUM	99.99	99.97	99.94	4.36	4.32	2.72
$h = 0.1$	99.95	99.94	99.82	4.32	4.55	3.28
$h = 0.4$	100.00	100.00	99.97	4.91	4.52	2.45
$h = 0.9$	100.00	100.00	100.00	2.48	1.03	0.03

Table 6.5.: Empirical size (in %) for a nominal level of $\alpha = 5\%$ with outliers.

size for all considered monitoring schemes and kernels. The superiority of the Wilcoxon kernel in terms of its robustness against outliers is clearly visible here. The null hypothesis is rejected for almost all realizations when using the CUSUM kernel whereas the Wilcoxon kernel keeps the nominal level in all situations and is even still conservative. This is also reflected in the density estimates of the stopping times which is shown in Figure 6.3. Even if the change occurs at the beginning of the monitoring period ($\beta = 0.75$), the procedures based on the CUSUM kernel reject the null hypothesis in many cases before the change occurs and only the second and, except for the modified MOSUM with $h = 0.4$, smaller bump of the density estimates relates to the actual change. For later changes, the null hypothesis is rejected in most of the cases before the change even occurs when using the CUSUM kernel. In contrast to that, the stopping times for the Wilcoxon kernel are not significantly affected by the outliers.

6.2. Stopping Time

In Chapter 5, we have shown that the stopping time converges to a Gaussian distribution if standardized appropriately. In the following, we consider the standardized stopping times for finite historic data sets for the CUSUM as well as the Wilcoxon kernel. Therefore, we simulate a mean change as in (6.1) with $d = 1$. We use the weight function in (3.5) with $\gamma = 0$ which turned out to be the best choice in the previous Section.

6.2.1. Sublinear Changes

First, we consider sublinear changes for $\beta = 0.25, 0.5, 0.75$ which cover all three cases of the limit distribution in Theorem 5.3. We use historic data sets of length $m = 100, 1000, 10000$ and a monitoring horizon of $20m$. The stopping times are centered with $a_m^{W/C}$ as in (5.53). Due to Corollary 3.7 (i), the critical value can by obtained by $c_\alpha^{W/C} = \sigma^{W/C}\tilde{c}_\alpha$, where $\tilde{c}_\alpha$ is the $(1-\alpha)$-quantile of $\sup_{0<t<1} |W(t)|$ which is determined in the same way as in the previous section. In order to obtain a basis for comparing different kernels as well as different distributions of the underlying time series, we scale the stopping time with

$$\tilde{b}_m^{W/C} = \begin{cases} \sigma^* b_m^{W/C} & \text{under } (I) \\ (\delta_1\sigma + (1-\delta_1)\sigma^*) b_m^{W/C} & \text{under } (II) \\ \sigma b_m^{W/C} & \text{under } (III), \end{cases}$$

$b_m^{W/C}$ as in (5.11), such that the limit distribution is standard Gaussian in all situations. With (5.12) it holds

$$\left(\sigma^C\right)^2 = \left(\sigma^{*C}\right)^2 = \sum_{h\in\mathbb{Z}} Cov(Y_0, Y_h).$$

For the Wilcoxon kernel we obtain with Example 3.1 and 3.11

$$\left(\sigma^W\right)^2 = \sum_{h\in\mathbb{Z}} Cov(F_Y(Y_0), F_Y(Y_h)),$$
$$\left(\sigma^{*W}\right)^2 = \sum_{h\in\mathbb{Z}} Cov(F_Y(Y_0+d), F_Y(Y_h+d)).$$

Δ^C and Δ^W are given by (2.9) and (2.10), where the latter is determined numerically. For the simulations of independent data, we use the true variances $\sigma^C = \sigma^{*C} = \mathrm{Var}(Y_1)$ and $\sigma^W = \mathrm{Var}(F_{Y_1}(Y_1)) = \frac{1}{12}$. The variance $\sigma^{*W} = \mathrm{Var}(F_{Y_1}(Y_1+1))$ is estimated based on the empirical distribution function of the historic data set. We also simulate dependent data in form of AR(1) time series with coefficient $a = 0.2$ and independent standard normally distributed errors. For the estimation of the long run covariance matrix we use a Bartlett kernel estimator.

Figure 6.4 shows the estimated densities of the standardized stopping times for different lengths of the historic data sets for independent standard normally distributed observations as well as for independent observations with a t distribution with 3 degrees of freedom. In all cases considered, the estimated densities converge from the right to the limit distribution with increasing length of the historic data set. Hence, the empirical stopping times tend to be larger than predicted by the asymptotic distribution. For both kernels, the convergence is somewhat slower for the t distribution. In particular for small historic data sets it turns out that for normally distributed observations the convergence is slightly faster when using the CUSUM kernel, whereas the Wilcoxon kernel yields a slightly faster convergence for the t distribution. With respect to the time of the change it can be observed that the estimated density drifts away from the limit distribution for increasing β which indicates that the asymptotic distribution does not hold for $\beta \geq 1$. The standardized stopping times for AR(1) time series with normally distributed errors (see Figure 6.5) behave very similar to those obtained for independent normally distributed observations.

In Section 5.1.1 we have calculated the scaled differences of the approximations of the expected stopping times

$$\frac{a_m^W - a_m^C}{\tilde{c}_\alpha \sqrt{m}}$$

for independent observations of certain distributions. Figure 6.6 shows the estimated densities based on the realizations of $\frac{\tau_m^W - \tau_m^C}{\tilde{c}_\alpha \sqrt{m}}$, i.e. the scaled differences of the stopping times, for independent and identically distributed data. For the standard normal distribution, the estimated densities are well concentrated around the theoretical value in Table 5.1 which is positive and thus indicates that with the CUSUM kernel changes are detected more quickly than with the Wilcoxon kernel. Opposite behavior can be observed for the heavy-tailed t distribution as predicted by the negative value in Table

5.1. The estimated densities are less concentrated around this theoretical value but are still maximal around this point and most of the mass is on the left of zero which means the Wilcoxon-type procedure is faster. The respective plots for AR(1) time series (see Figure 6.7) with normally distributed errors conform again with the observations for independent and normally distributed data. We did not calculate a theoretical value in this case. However, it can be seen that the CUSUM kernel leads to a faster detection of the change as most of the mass of the estimated densities is right of zero.

6.2.2. Linear and Superlinear Changes

We conclude the simulation study with empirical results related to the asymptotic distribution of the stopping time for superlinear and linear changes. We consider $\beta = 1$ as linear change and $\beta = 1.4$ as superlinear change. The monitoring horizon is extended to $30m$. The simulations refer to Corollary 5.21 (ii). For each realization we calculate $s_{1,m}$ and $s^*_{1,m}$. Based on their values we standardize the respective stopping time using $a_m = a_m(s_{1,m}, s^*_{1,m})$ given in (5.66) and $\tilde{b}_m = \sigma^{W/C} b_m(s_{1,m}, s^*_{1,m})$ with $b_m(s_{1,m}, s^*_{1,m})$ as in (5.70) such that the limit distribution is standard normal for both kernels. $\sigma^{W/C}$ are determined in the same way as described in the previous section. We consider independent and standard normally distributed observations as well as $AR(1)$ time series with coefficient $a = 0.2$. Figures 6.8 and 6.9 show that, in contrast to the sublinear case, the estimated densities converge in a centered way to the limit distribution. The approximation via the limit distribution seems to be reasonable already for small historic data sets ($m = 100$) and quite good for larger historic data sets.

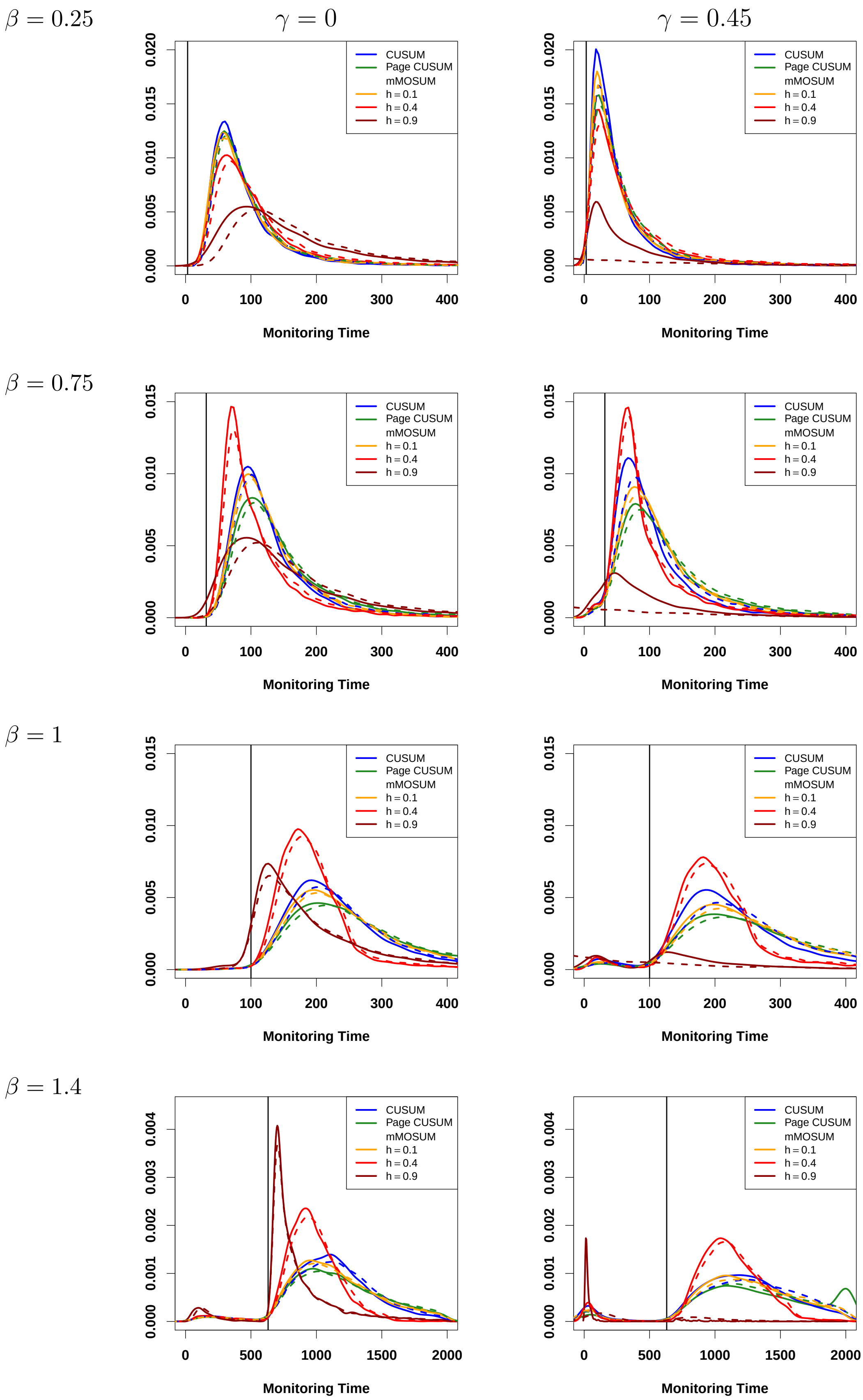


Figure 6.2.: Estimated densities of the stopping time for the CUSUM kernel (solid lines) and the Wilcoxon kernel (dashed lines) for independent $N(0, 1)$-distributed observations.

with outliers

$\beta = 0.75$

$\beta = 1$

$\beta = 1.4$

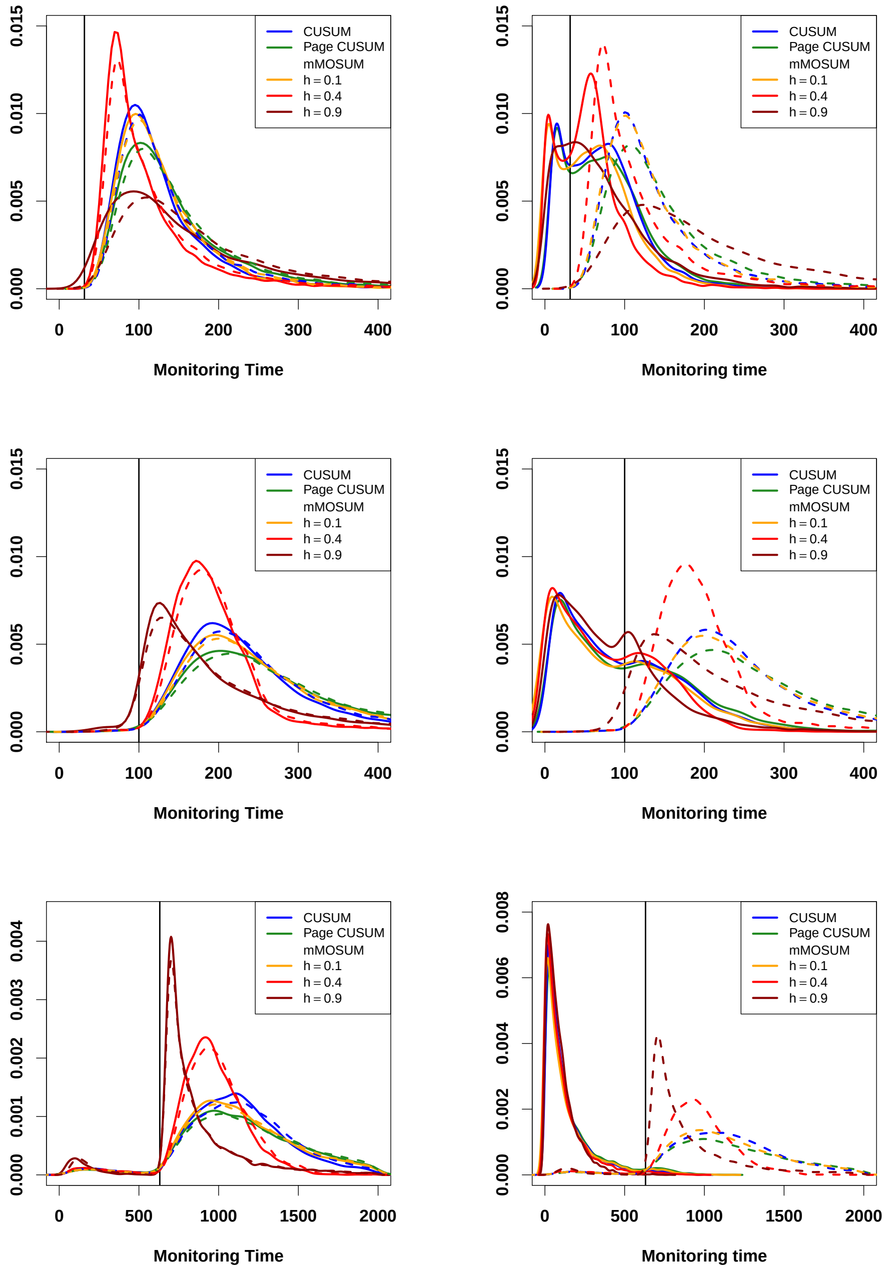


Figure 6.3.: Estimated densities of the stopping time with outliers for the CUSUM kernel (solid lines) and the Wilcoxon kernel (dashed lines) for independent observations with $Y_1 \sim N(0, 1)$.

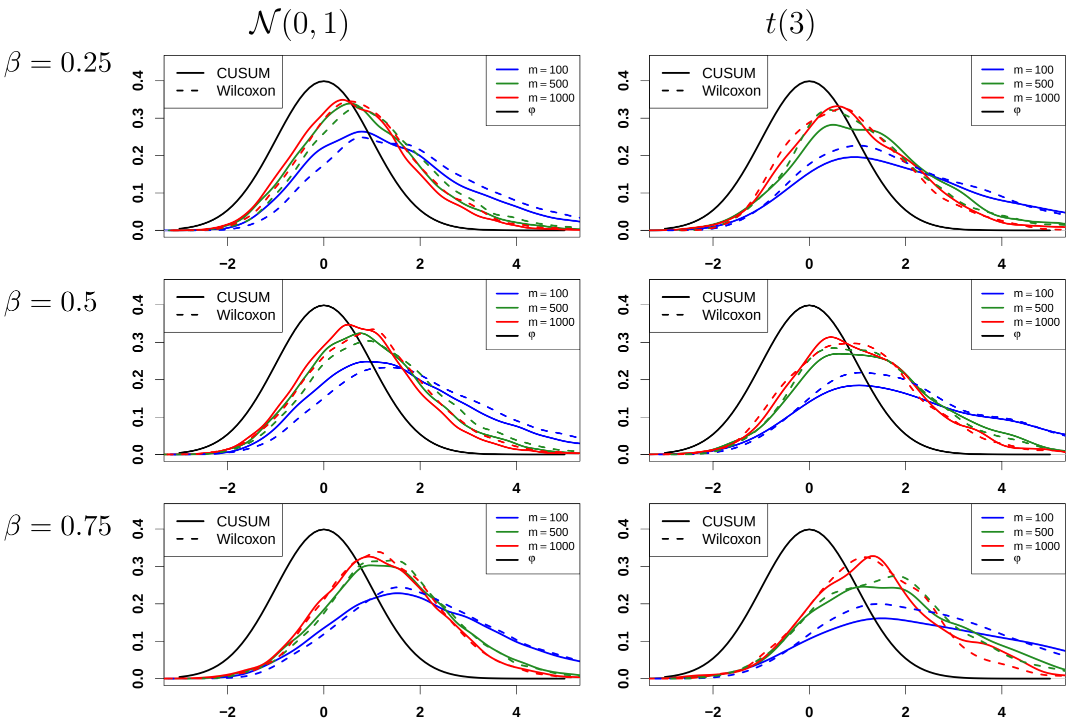


Figure 6.4.: Estimated densities of the standardized stopping time for i.i.d. data for early changes.

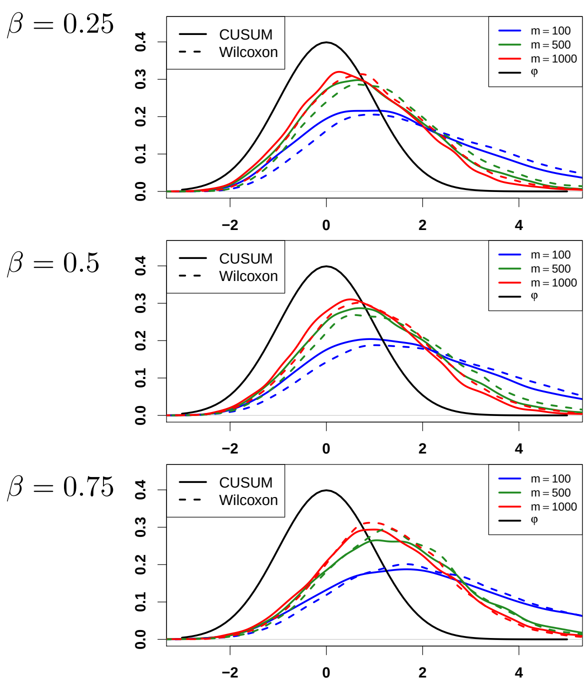


Figure 6.5.: Estimated densities of the standardized stopping time for AR(1) time series with i.i.d. standard Gaussian errors and coefficient $a = 0.2$ for early changes.

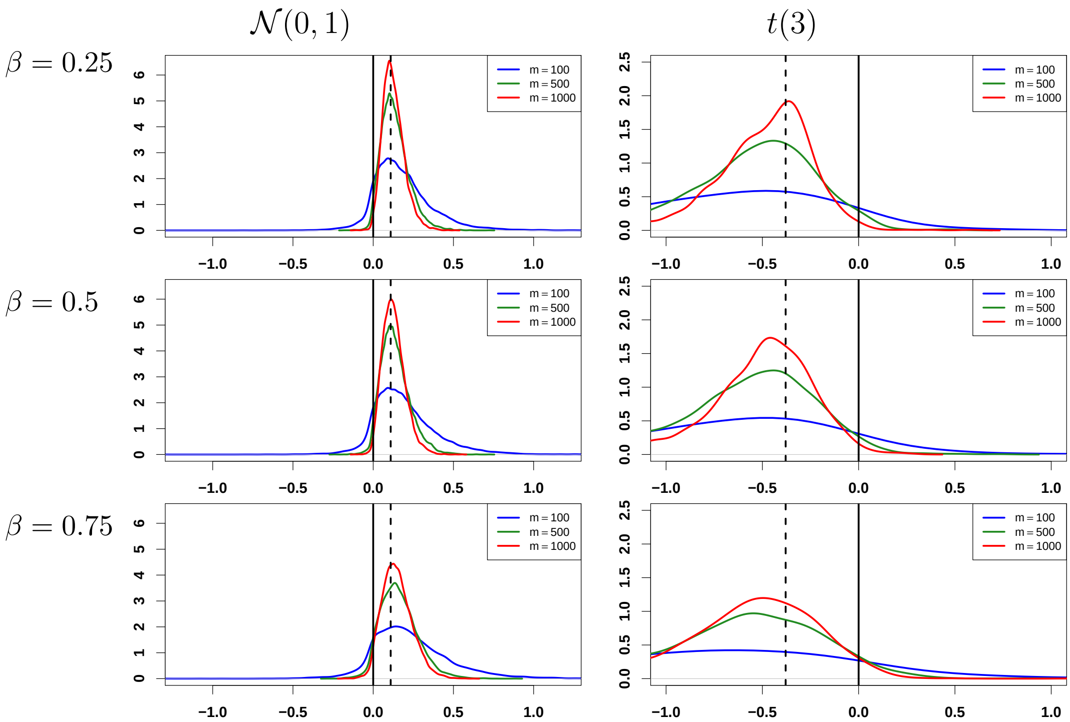


Figure 6.6.: Estimated densities of the scaled differences for i.i.d. data for early changes. The dashed line indicates the theoretical value from Table 5.1.

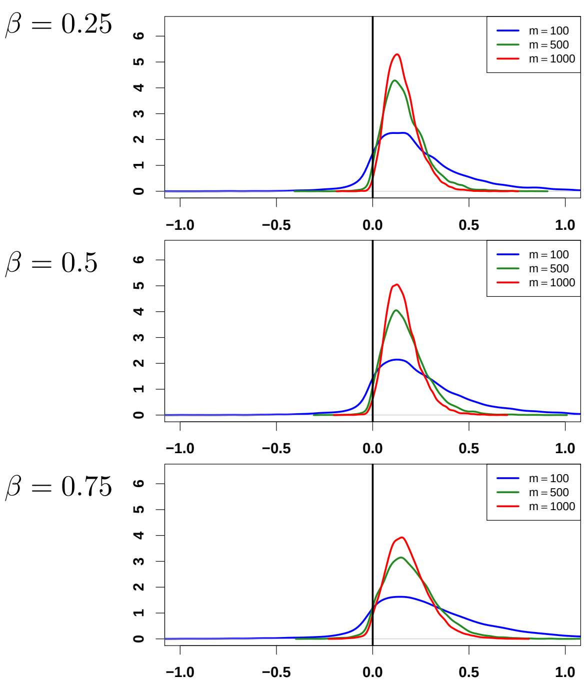


Figure 6.7.: Estimated densities of the scaled differences for AR(1) time series with i.i.d. standard Gaussian errors and coefficient $a = 0.2$ for early changes.

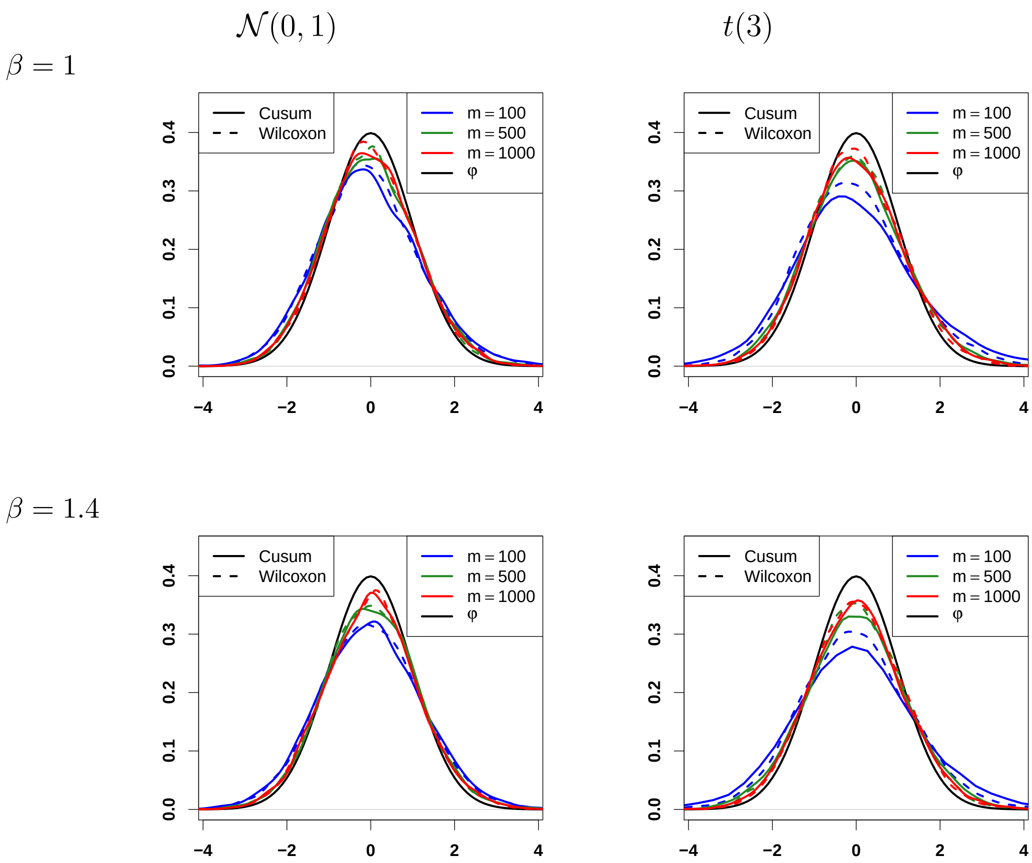
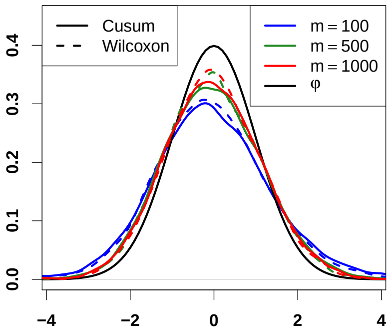


Figure 6.8.: Estimated densities of the standardized stopping time for linear and superlinear changes for independent observations.

$\beta = 1$



$\beta = 1.4$

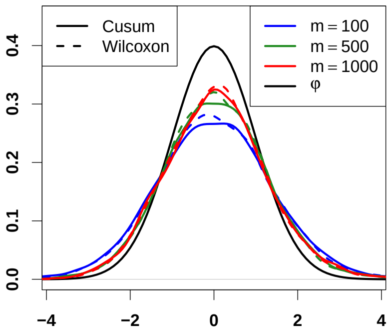


Figure 6.9.: Estimated densities of the standardized stopping time for AR(1) time series with i.i.d. standard Gaussian errors and coefficient $a = 0.2$ for linear and superlinear changes.

7. Conclusions

In the first part of this thesis we have provided a general framework of sequential testing procedures based on U-statistics. The asymptotic results under the null hypothesis as well as under the alternative are obtained based on very general assumptions and allow to derive sequential testing procedures for various kernel functions and dependency structures. For example, we obtain a robust sequential procedure when using a Wilcoxon-type kernel. The expected advantages of the Wilcoxon kernel for heavy tailed distributions and even more for data which contains strong outliers have been verified in the simulation study. Furthermore, we extended the proposed class of procedures to monitoring schemes adapted to late changes. The speed of detection is of particular interest in sequential testing. Therefore, we derived the limit distribution of the stopping time with appropriate standardization for the weight function in (3.5). The results for early changes, ie. where the time of the change is sublinear in m, are generalizations of existing results for the classical sequential CUSUM procedure. Furthermore, we derived the limit distribution of the stopping time for linear and superlinear changes by conditioning on functionals of the historic observations. To the best of our knowledge, even for the classical sequential CUSUM procedure, the stopping time for such late changes has not been considered before. The stopping time for late changes has only been provided for $\gamma = 0$ as the simulation study in Section 6 has shown that this is the best choice in the vast majority of situations. It would be of future interest to also derive the limit distribution for $\gamma \neq 0$ in order to prove the superiority of $\gamma = 0$ theoretically. The simulation study has also revealed that the Page-CUSUM and the modified MOSUM outperform the CUSUM monitoring scheme for both kernels under consideration if the change does not occur at the beginning of the monitoring period. For the CUSUM kernel, the superiority of the Page-CUSUM over the CUSUM monitoring scheme has been shown in Fremdt (2014) for sublinear changes that do not occur immediately after the monitoring starts. This could be extended by deriving the limit distribution of the stopping time for the Page-CUSUM as well as the modified MOSUM for the framework considered in this thesis. The superiority of the adapted monitoring schemes is expected to be even stronger for later changes such that the respective limit distributions of the stopping times for linear and superlinear changes would be of particular interest.

Part II.

Detecting Changes in the Covariance Structure of Functional Time Series With FMRI Data in View

8. Motivation

Functional Magnetic Resonance Imaging (fMRI) is a widely used technique to capture brain activity. The blood flow and thus the demand of oxygen increases in the activated areas of the brain. This results in an increasing ratio of blood oxygenation and deoxygenated hemoglobin in the respective areas which can can be measured based on the blood oxygenation level-dependent (BOLD) contrast Ogawa *et al.* (1990). An fMRI data set consists of a sequence of three-dimensional images that have been recorded every few seconds. fMRI facilitates a noninvasive real time functional brain mapping with a high spatial resolution and thus yields large amounts of data requiring the development of appropriate statistical methodologies. Such scans can be obtained related to a task or in a resting state where the person is told to go through the scanning procedure without thinking of anything while not falling asleep. Resting state data is used to analyze brain activities excluding external factors where the examination of the covariance structure between brain regions is of particular interest as it is associated with neural connectivity. Such analyses strongly rely on the assumption that resting state data is first and second order stationary. This assumption is by no means guaranteed as it might happen, for example, that during the scan the person suddenly remembers something such that the mean activities deviate from their resting state baseline in some areas of the brain. If such a scan is then used for analyzing connectivities without taking a possible change into account, the results will be contaminated by the mean change leading to wrong conclusions. Therefore, Aston & Kirch (2012b) developed testing procedures to detect deviations from mean stationarity. However, it is not only deviations from mean stationarity but also deviations from covariance stationarity that will contaminate the analysis and ultimately the conclusions. Therefore, in this part of the thesis, we develop tools to test for deviations from covariance stationarity in fMRI data which will be modeled as functional time series. This means that each observation of the time series, in this case each 3-d image of the brain, can be viewed as a function. Indeed, taking into consideration that the brain works as a single unit with spatial dependencies, it is a natural approach to model each image as a discretized observation of a functional response. In contrast, a voxelwise approach requires a difficult adaption for multiple testing and may miss signals that are very small in any voxel but considerably large if information across voxels is used. Dependencies in time, i.e. between subsequent images, which are also present in fMRI data, can be captured by a time series structure. Lifting the multivariate observations to a functional space makes them mathematically easier to handle as one can exploit functional properties, such as smoothness, making use of many well established statistical techniques.

We adapt a nonparametric approach where we tackle the problem by means of a change point procedure without assuming any parametric spatial or temporal correlation structure. Such nonparametric methods become more and more refined in the analysis of functional data (cf. Ferraty & Vieu (2000) and Horváth & Kokoszka (2012)). Nonparametric tests for at most one change (AMOC) in the mean function have been considered for independent observations in Aue *et al.* (2009c) and Berkes *et al.* (2009) as well as for weakly dependent data in Hörmann & Kokoszka (2010). Aston & Kirch (2012a) extend these results to a more general class of dependency structures and also consider epidemic changes where the mean function returns to its original state after some time. The analysis of functional connectivity data is a very active field of research in neuroimaging. The detection of change points in the observed data without assuming the specifications of the experiment to be known is of particular interest. In this context, Cribben *et al.* (2012) propose a data-driven approach, the so called Dynamic Connectivity Regression (DCR), for detecting changes in the functional connectivity between a set of brain regions and estimate a connectivity graph for each temporal interval between the change points. They use resampling methods in order to decide whether a change is significant. With a view to single-subject data, DCR is further developed in Cribben *et al.* (2013). In this part of the thesis, we develop statistical procedures for the detection of deviations from covariance stationarity in functional time series that can be applied to fMRI data without being restricted to predefined regions of interest.

8.1. Outline

In Chapter 9 we propose a procedure based on dimension reduction techniques such as principal component analysis, to detect deviations from covariance stationarity. The test statistics and their asymptotic behavior are investigated in Chapter 10. The proposed procedures require the estimation of the long-run covariance which is statistically unstable. Using a missspecified estimator is a possible solution but leads to an unknown limit distribution such that resampling procedures, as described in Chapter11, are unavoidable. Alternative test statistics which take the full functional structure into account without reducing the dimension are discussed in Chapter 12. The different procedures proposed in this work are compared in a simulation study in Chapter 13. The application to fMRI data is presented in Chapter 14.

9. Testing for Changes in the Covariance Structure of Functional Data

We assume that the observations are obtained from a functional time series with the respective mean function being constant over time, i.e.

$$X_t(s) = \mu(s) + Y_t(s), \quad 1 \leq t \leq n,$$

where t denotes the time point and s a spatial coordinate in a compact set $\mathcal{Z}$. The constant mean function is given by $\mu(\cdot)$ while the random fluctuations are represented by $Y_t(\cdot)$ with $\mathrm{E}(Y_t(s)) = 0$ which is not necessarily stationary but can have a time-dependent covariance structure as detailed in Chapter 9.1. $\mu(\cdot)$ as well as all elements of $\{Y_t(\cdot) : 1 \leq t \leq n\}$ are assumed to be square integrable on $\mathcal{Z}$. The mean stationarity can be checked previously as described in Aston & Kirch (2012b).

The covariance structure of a functional time series is determined by the covariance operator respectively the covariance kernel as given in the following definition.

Definition 9.1. *Let $\{X_t(\cdot) : 1 \leq t \leq n\} \in \mathcal{L}^2(\mathcal{Z})$ be a functional time series, where $\mathcal{Z}$ is a compact set. The square integrable covariance operator $C_t : \mathcal{L}^2(\mathcal{Z}) \mapsto \mathcal{L}^2(\mathcal{Z})$ is defined by*

$$C_t(z) = \int c_t(\cdot, s) z(s) ds,$$

where $c_t(u, s) = \mathrm{Cov}\left(X_t(u), X_t(s)\right)$ is the covariance kernel of $X_t(\cdot)$.

9.1. Change Point Model

First consider the at most one change (AMOC) alternative given by

$$Y_t(s) = Y_t^{(1)}(s) 1_{\{1 \leq t \leq \theta n\}} + Y_t^{(2)}(s) 1_{\{\theta n < t \leq n\}}, \quad 1 \leq t \leq n \tag{9.1}$$

with $\mathrm{Cov}\left(Y_t^{(1)}(u), Y_t^{(1)}(s)\right) = c(u, s)$ and $\mathrm{Cov}\left(Y_t^{(2)}(u), Y_t^{(2)}(s)\right) = c(u, s) + \delta(u, s)$ for some $0 < \theta < 1$ and $c(u, s), \delta(u, s) \in \mathcal{L}^2(\mathcal{Z} \times \mathcal{Z})$. According to this model, the covariance change occurs at the unknown time point $[\theta n]$. The covariance kernel $c(u, s)$ before the change as well as the change in covariance $\delta(u, s) \neq 0$ are both unknown.

Definition 9.2. *(Hörmann & Kokoszka, 2010, Definition 2.1) Let $\{Y_t(\cdot) : t \geq 1\} \in \mathcal{L}^2(\mathcal{Z})$ be a functional time series with $\mathrm{E}\left\|Y_t(\cdot)\right\|^4 = \int \mathrm{E}\left[(Y_t(s))^4\right] ds < \infty$ which can be represented by*

$$Y_t = f(\epsilon_t, \epsilon_{t-1}, \ldots),$$

where $\{\epsilon_i : i \in \mathbb{Z}\}$ *are i.i.d. random variables with values in a measurable space* S *and* f *is a measurable function* $f : S^\infty \to H$*. Then,* $\{Y_t(\cdot)\}$ *is called* L^4_m-*approximable if*

$$\sum_{m=1}^{\infty} E\|Y_t(\cdot)\|^4 = \int \mathrm{E}\left[\left(Y_t(s) - Y_t^{(m)}(s)\right)^4\right] ds < \infty$$

with $Y_t^{(m)} = f(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-m+1}, \epsilon'_{t-m}, \epsilon'_{t-m-1}, \ldots)$*, where* ϵ'_i *is an independent copy of* ϵ_i*.*

L^4_m-approximability is a nonparametric concept of dependence which provides the necessary mathematical tools for the proofs and is satisfied for a large class of time series. Full details can be found in Hörmann & Kokoszka (2010).

Assumption 9.3. *Assume that for* $\{Y_t(\cdot)\}$ *as in* (9.1) *it holds*

(i) $\{Y_t^{(1)}(\cdot)\} \in \mathcal{L}^2(\mathcal{Z})$ *with*

$$\mathrm{E}\, Y_1^{(1)}(s) = 0 \quad \textit{and} \quad \mathrm{E}\, \|Y_1^{(1)}(\cdot)\|^4 = \int \mathrm{E}\left[\left(Y_1^{(1)}(s)\right)^4\right] ds < \infty$$

is L^4_m-*approximable and hence, in particular, stationary and ergodic.*

(ii) $\{Y_t^{(2)}(\cdot)\} \in \mathcal{L}^2(\mathcal{Z})$ *is ergodic with*

$$\mathrm{E}\, Y_1^{(2)}(s) = 0 \quad \textit{and} \quad \mathrm{E}\, \|Y_1^{(2)}(\cdot)\|^2 = \int \mathrm{E}\left[\left(Y_1^{(2)}(s)\right)^2\right] ds < \infty.$$

As we do not assume $Y_t^{(2)}$ to be stationary, the time series after the change is allowed to have starting values from a different distribution.

Testing for covariance stationarity against the AMOC alternative can be described by the following hypotheses:

$$H_0 : \theta = 1 \quad \text{against} \quad H_1 : \quad 0 < \theta < 1.$$

In order to obtain a test for more general alternatives of nonstationarities in the covariance, we consider the following epidemic alternative:

$$Y_t(s) = Y_t^{(1)}(s) 1_{\{1 \le t \le \theta_1 n\}} + Y_t^{(2)}(s) 1_{\{\theta_1 n < t \le \theta_2 n\}} + Y_t^{(1)}(s) 1_{\{\theta_2 n < t \le n\}}, \quad 1 \le t \le n$$

with $\mathrm{Cov}\left(Y_t^{(1)}(u), Y_t^{(1)}(s)\right) = c(u,s)$ and $\mathrm{Cov}\left(Y_t^{(2)}(u), Y_t^{(2)}(s)\right) = c(u,s) + \delta(u,s)$ for some $0 < \theta_1 < \theta_2 < 1$. It would also be possible to allow for contaminated starting values in the time series after the change. This alternative can be viewed as a better approximation to the expected kind of deviation from covariance stationarity.

9.2. Dimension Reduction Techniques

A common approach in functional data analysis is the transition to a multivariate setting by projecting the data into a d-dimensional space spanned by an orthonormal basis $\{v_k(\cdot) : k = 1, \ldots, d\}$. In this case, the projection scores are obtained by

$$\langle X_t, v_l \rangle = \int X_t(s) v_l(s) ds, \quad t = 1, \ldots, n, l = 1, \ldots, d. \tag{9.2}$$

As we aim at assessing the above functional testing problem by applying a multivariate testing procedure to the projection scores we first need to verify if a change in the covariance structure of the observed functional time series implies a change in the covariance of the scores. To this end, we observe

$$\begin{aligned}&\mathrm{Cov}(\langle X_t, v_{l_1} \rangle, \langle X_t, v_{l_2} \rangle)\\ &= \int\int c(u,s) v_{l_1}(u) v_{l_2}(s) du\, ds + 1_{\{\theta n < t \leq n\}} \int\int \delta(u,s) v_{l_1}(u) v_{l_2}(s) du\, ds.\end{aligned}$$

Thus, a necessary condition for the covariance change to be visible in the projection scores is

$$\int\int \delta(u,s) v_{l_1}(u) v_{l_2}(s) du\, ds \neq 0 \quad \text{for some } l_1, l_2 = 1, \ldots, d. \tag{9.3}$$

In contrast to other applications we do not require the dimension reduction technique to explain a large amount of the variation of the data but to yield a good signal-to-noise ratio where the signal is determined by $\int\int \delta(u,s) v_{l_1}(u) v_{l_2}(s) du\, ds$.

9.2.1. Principal component analysis

Principal component analysis (PCA) is a widely used data driven dimension reduction technique which projects the functional data on the subspace spanned by the first d principal components explaining the most variance of any subspace of size d. Let $\{\lambda_l : l \geq 1\}$ be the non-negative decreasing sequence of eigenvalues of the covariance operator and $\{v_l(\cdot) : l \geq 1\}$ a set of corresponding orthonormal eigenfunctions defined by

$$\int c(u,s) v_l(s) ds = \lambda_l v_l(u), \quad l = 1, 2, \ldots; u \in \mathcal{Z}.$$

By Mercer's Lemma, see Lemma 1.3 in Bosq (2006), the covariance kernel can be expressed as

$$c(u,s) = \sum_{l=1}^{\infty} \lambda_k v_l(u) v_l(s)$$

and the Karhunen-Loève expansion, see Theorem 1.5 in Bosq (2006), yields

$$X_t(s) - \mu(s) = \sum_{l=1}^{\infty} \eta_{t,l} v_l(s), \tag{9.4}$$

where the scores $\{\eta_{t,l} : l = 1, 2, \ldots\}$ given by $\eta_{t,l} = \int (X_t(s) - \mu(s))\, v_l(s)ds$ are uncorrelated and centered with variance λ_l. As the covariance kernel is unknown, PCA is usually conducted based on the empirical covariance function

$$\hat{c}_n(u,s) = \frac{1}{n}\sum_{t=1}^{n} \left(X_t(u) - \overline{X}_n(u)\right)\left(X_t(s) - \overline{X}_n(s)\right),$$

where $\overline{X}_n(s) = \frac{1}{n}\sum_{t=1}^{n} X_t(s)$. Under the null hypothesis, the empirical covariance function estimates the actual covariance kernel $c(u,s)$ whereas under the alternative it often converges to a contaminated limit $k(u,s)$ as stated in (10.8) below. As projection basis we determine the eigenfunctions $\{\hat{v}_l(\cdot) : l = 1, \ldots, d\}$ of $\hat{c}_n$ belonging to the d largest eigenvalues and obtain the projection scores by

$$\hat{\eta}_{t,l} = \int \left(X_t(s) - \overline{X}_n(s)\right)\hat{v}_l(s)ds = \langle X_t, \hat{v}_l\rangle - \overline{\langle X, \hat{v}_l\rangle}_n, \quad t = 1, \ldots, n, l = 1, \ldots, d$$

with $\overline{\langle X, \hat{v}_l\rangle}_n = \frac{1}{n}\sum_{t=1}^{n}\langle X_t, \hat{v}_l\rangle$. For more details on functional principal component analysis, in particular for the consistency of the empirical eigenvalues and eigenfunctions, see, for example, Horváth & Kokoszka (2012).

Lemma 9.4. *Let $\hat{v}_l(\cdot)$ be orthonormal eigenfunctions of $\hat{c}_n(u,s)$ and $\tilde{v}_l(\cdot)$ be orthonormal eigenfunctions of $\tilde{c}(u,s)$, where both sets of eigenfunctions are arranged according to the respective eigenvalues in decreasing order. Furthermore, assume that the eigenvalues of $\tilde{c}(u,s)$ are separated, i.e. $\tilde{\lambda}_1 > \tilde{\lambda}_2 > \ldots > \tilde{\lambda}_d > \tilde{\lambda}_{d+1}$.*

a) If $\int\int(\hat{c}_n(u,s) - \tilde{c}(u,s))^2 du\, ds = o_P(1)$, it holds for $l_1, l_2 = 1, \ldots, d$

$$\int\int\left(\tilde{g}_{l_1}\tilde{g}_{l_2}\hat{v}_{l_1}(u)\hat{v}_{l_2}(s) - \tilde{v}_{l_1}(u)\tilde{v}_{l_2}(s)\right)^2 du\, ds = o_P(1),$$

where $\tilde{g}_l = sgn\left(\int \tilde{v}_l(s)\hat{v}_l(s)ds\right)$.

b) If $\int\int(\hat{c}_n(u,s) - \tilde{c}(u,s))^2 du\, ds = O_P(n^{-1})$, it holds for $l_1, l_2 = 1, \ldots, d$

$$\int\int\left(\tilde{g}_{l_1}\tilde{g}_{l_2}\hat{v}_{l_1}(u)\hat{v}_{l_2}(s) - \tilde{v}_{l_1}(u)\tilde{v}_{l_2}(s)\right)^2 du\, ds = O_P(n^{-1}).$$

Proof. Observing that

$$\begin{aligned}
&\int\int\left(\tilde{g}_{l_1}\tilde{g}_{l_2}\hat{v}_{l_1}(u)\hat{v}_{l_2}(s) - \tilde{v}_{l_1}(u)\tilde{v}_{l_2}(s)\right)^2 du\, ds \\
=&\int\int\Big((\tilde{g}_{l_1}\hat{v}_{l_1}(u) - \tilde{v}_{l_1}(u))(\tilde{g}_{l_2}\hat{v}_{l_2}(s) - \tilde{v}_{l_2}(s)) + \tilde{v}_{l_1}(u)(\tilde{g}_{l_2}\hat{v}_{l_2}(s) - \tilde{v}_{l_2}(s)) \\
&\quad + \tilde{v}_{l_2}(s)(\tilde{g}_{l_1}\hat{v}_{l_1}(u) - \tilde{v}_{l_1}(u))\Big)^2 du\, ds \\
\leq& C\left(\int(\tilde{g}_{l_1}\hat{v}_{l_1}(u) - \tilde{v}_{l_1}(u))^2 du \int(\tilde{g}_{l_2}\hat{v}_{l_2}(s) - \tilde{v}_{l_2}(s))^2 ds + \int\tilde{v}_{l_1}^2(u)du\int(\tilde{g}_{l_2}\hat{v}_{l_2}(s) - \tilde{v}_{l_2}(s))^2 ds\right. \\
&\quad\left. + \int\tilde{v}_{l_2}(s)^2 ds\int(\tilde{g}_{l_1}\hat{v}_{l_1}(u) - \tilde{v}_{l_1}(u))^2 du\right) \\
=& C\left(\int(\tilde{g}_{l_1}\hat{v}_{l_1}(u) - \tilde{v}_{l_1}(u))^2 du\int(\tilde{g}_{l_2}\hat{v}_{l_2}(s) - \tilde{v}_{l_2}(s))^2 ds + \int(\tilde{g}_{l_2}\hat{v}_{l_2}(s) - \tilde{v}_{l_2}(s))^2 ds\right. \\
&\quad\left. + \int(\tilde{g}_{l_1}\hat{v}_{l_1}(u) - \tilde{v}_{l_1}(u))^2 du\right)
\end{aligned}$$

the assertions follow with Theorem 2.1 in Aston & Kirch (2012a). □

Separable covariance structure

As fMRI data is collected voxelwise ($\sim M := 10^5$ voxels), using the empirical covariance function requires the calculation and storage of an $M \times M$-dimensional matrix in addition to the respective eigenanalysis. While this is computationally infeasible, one can show that there is a one-to-one correspondence between the eigenvalues and eigenvectors of the spatial covariance matrix ($M \times M$) and that of the time domain ($n \times n$). As $M \gg n$ the eigenanalysis in the time domain requires less computational effort. However, this relationship also reveals that the number of nonzero eigenvalues is limited by the sample size and hence indicates a considerable loss of precision when using the nonparametric covariance estimator. Based on those considerations, Aston & Kirch (2012b) suggest to use a separable covariance structure in the estimation procedure given by

$$c\left((u_1, u_2, u_3), (s_1, s_2, s_3)\right) = c_1\left(u_1, s_1\right) c_2\left(u_2, s_2\right) c_3\left(u_3, s_3\right).$$

In this work, we adopt this approach of estimating the covariance matrix separately in each direction (64×64 resp. 64×64 resp. 33×33) and calculate the respective eigenvalues and eigenfunctions. The projection basis can then be obtained by the tensor product of the first d eigenfunctions of each direction. Even if the actual covariance structure is not separable we obtain a valid projection such that the proposed dimension reduction can be applied for our purposes. While this is an obvious simplification, most smoothing techniques in fMRI make use of tensor based formulations leading to very similar implicit assumptions.

10. Test Statistic and Statistical Properties

We assess the functional testing problem by testing for a change in the covariance structure of the d-dimensional estimated score vectors. As proposed by Aue *et al.* (2009a) we construct the test statistics based on the following version of the traditional CUSUM-statistic for the AMOC alternative:

$$S_k = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{k} \mathrm{vech}[\hat{\eta}_t \hat{\eta}_t^T] - \frac{k}{n} \sum_{t=1}^{n} \mathrm{vech}[\hat{\eta}_t \hat{\eta}_t^T] \right). \tag{10.1}$$

We consider the following sum-type and max-type test statistics:

$$\Omega_n = \frac{1}{n} \sum_{k=1}^{n} S_k^T \hat{\Sigma}_n^{-1} S_k \quad \text{and} \quad \Lambda_n = \max_{1 \leq k \leq n} S_k^T \hat{\Sigma}_n^{-1} S_k$$

where $\hat{\Sigma}_n$ is an estimator for the long-run covariance

$$\Sigma_0 = \sum_{t \in \mathbb{Z}} \mathrm{Cov}\left(\mathrm{vech}[\eta_0 \eta_0^T], \mathrm{vech}[\eta_t \eta_t^T]\right)$$

under H_0. We assume that $\hat{\Sigma}_n$ is consistent under the null hypothesis and

$$|\hat{\Sigma}_n - \Sigma_1| = o_p(1) \quad \text{under } H_1,$$

where Σ_1 is some positive-definite matrix which can differ from Σ_0. For the epidemic change point alternative we propose the test statistics

$$\Omega_n^{ep} = \frac{1}{n} \sum_{1 \leq k_1 < k_2 \leq n} S_{k_1,k_2}^T \hat{\Sigma}_n^{-1} S_{k_1,k_2} \quad \text{and} \quad \Lambda_n^{ep} = \max_{1 \leq k_1 < k_2 \leq n} S_{k_1,k_2}^T \hat{\Sigma}_n^{-1} S_{k_1,k_2}$$

with $S_{k_1,k_2} = S_{k_2} - S_{k_1}$.

As large values of the weighted partial sum process indicate a change, the point where this process takes it maximum is usually a good estimator for the location of the change. More precisely, we estimate the change point by

$$\hat{k}^* = \mathrm{argmax}_{1 \leq k \leq n} S_k^T \hat{\Sigma}_n^{-1} S_k$$

(see (2.12) in Aue *et al.* (2009a)) respectively

$$(\hat{k}_1^*, \hat{k}_2^*) = \mathrm{argmax}\left(S_{k_1,k_2}^T \hat{\Sigma}_n^{-1} S_{k_1,k_2} : 1 \leq k_1 < k_2 \leq n\right),$$

where $(\hat{x}, \hat{y}) = \mathrm{argmax}((Z(x,y) : 1 \leq x < y \leq n)$ if and only if $\hat{x} = \min(1 \leq x < n : Z(x,y) = \max_{1 \leq k_1 < k_2 \leq n} Z(k_1,k_2)$ for some $y)$ and $\hat{y} = \max(\hat{x} < y \leq n : Z(\hat{x},y) = \max_{1 \leq k_1 < k_2 \leq n} Z(k_1,k_2))$ (see (4.4) in Aston & Kirch (2012b)).

10.1. Behavior Under the Null Hypothesis

We allow for a weak dependency structure by assuming the observed functional time series to be L^4_m-approximable. Effectively, this means the time series can be well approximated (in an L^4_m-sense) by an m-dependent one (see Definition 2.1 in Hörmann & Kokoszka (2010)) - a property that many of the usual time series models possess. The following Theorem provides the limit distributions of the proposed test statistics. Choosing the critical value as the $(1-\alpha)$-quantile of the respective limit distribution yields a procedure with asymptotic size α.

Theorem 10.1. *Let Assumption 9.3 (i) be satisfied. Additionally, we assume that the first $d+1$ eigenvalues of $c(u,s)$ are separated, i.e. $\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1}$. Then, the following asymptotics hold under the null hypothesis if $\hat{\Sigma}$ is a consistent estimator for the long-run covariance Σ.*

$$\Omega_n \xrightarrow{\mathcal{D}} \sum_{l=1}^{\mathfrak{d}} \int_0^1 B_l^2(x)dx \quad \text{and} \quad \Omega_n^{ep} \xrightarrow{\mathcal{D}} \sum_{l=1}^{\mathfrak{d}} \int_0^1 \int_0^y \left(B_l(y) - B_l(x)\right)^2 dx\, dy$$

as well as

$$\Lambda_n \xrightarrow{\mathcal{D}} \sup_{0\le x\le 1} \sum_{l=1}^{\mathfrak{d}} B_l^2(x) \quad \text{and} \quad \Lambda_n^{ep} \xrightarrow{\mathcal{D}} \sup_{0\le x<y\le 1} \sum_{l=1}^{\mathfrak{d}} \left(B_l(y) - B_l(x)\right)^2,$$

where $\mathfrak{d} = d(d+1)/2$ and $(B_l(x) : x \in [0,1], 1 \le l \le \mathfrak{d})$ are independent standard Brownian bridges.

Proof. We first show that the L^p_m-approximability is passed on to the projection scores. Let $Y_t^{(m)}$ be the m-approximations for an L^p_m-approximable sequence Y_t. The sequence $\eta_t^{(m)}$ with components $\eta_{t,l}^{(m)} = \int Y_t^{(m)}(s)v_l(s)ds$ is m-dependent as $Y_t^{(m)}$ is m-dependent. Furthermore, it holds with the Cauchy-Schwarz inequality, similarly to the proof of Theorem 5.1 in Hörmann & Kokoszka (2010),

$$\begin{aligned}
&\sum_{m\ge1}\left(\mathrm{E}\left[\left|\eta_t - \eta_t^{(m)}\right|^p\right]\right)^{1/p} = \sum_{m\ge1}\left(\mathrm{E}\left[\left(\sum_{l=1}^{d}\left(\eta_{t,l} - \eta_{t,l}^{(m)}\right)^2\right)^{p/2}\right]\right)^{1/p}\\
=&\sum_{m\ge1}\left(\mathrm{E}\left[\left(\sum_{l=1}^{d}\left(\int\left(Y_t(s) - Y_t^{(m)}(s)\right)v_l(s)ds\right)^2\right)^{p/2}\right]\right)^{1/p}\\
\le&\sum_{m\ge1}\left(\mathrm{E}\left[\left(\sum_{l=1}^{d}\int\left(Y_t(s) - Y_t^{(m)}(s)\right)^2 ds\int v_l^2(s)ds\right)^{p/2}\right]\right)^{1/p}\\
=&\sqrt{d}\sum_{m\ge1}\left(\mathrm{E}\left[\left(\int\left(Y_t(s) - Y_t^{(m)}(s)\right)^2 ds\right)^{p/2}\right]\right)^{1/p}\\
=&\sqrt{d}\sum_{m\ge1}\left(\mathrm{E}\left[\left\|Y_t - Y_t^{(m)}\right\|^p\right]\right)^{1/p} < \infty,
\end{aligned}$$

where $|\cdot|$ denotes the Euclidean norm. Thus, the score vectors $\{\eta_t\}$with $\eta_{t,l} = \int Y_t(s)v_l(s)ds$ are L_m^4-approximable and it holds with Theorem A.2 in Aue *et al.* (2009a) and the continuous mapping theorem

$$\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{[nx]}\mathrm{vech}[\eta_t\eta_t^T]-\frac{[nx]}{n}\sum_{t=1}^{n}\mathrm{vech}[\eta_t\eta_t^T]\right)\overset{D^{\mathfrak{d}}[0,1]}{\longrightarrow} B_\Sigma(x), \tag{10.2}$$

where $\{B_\Sigma(x) : 0 \le x \le 1\}$ is a $\mathfrak{d}$-dimensional centered Gaussian process with covariance function $\mathrm{Cov}\left(B_\Sigma(x), B_\Sigma(y)\right) = \Sigma(\min\{x, y\} - xy)$. The convergence in (10.2) still holds true if the projection basis is obtained based on the empirical covariance kernel. More precisely, it holds for $\check{\eta}_{t,l} = \int Y_t(s)\hat{v}_l(s)ds$ and $g_l = sgn\left(\int v_l(s)\hat{v}_l(s)ds\right)$ with the Cauchy-Schwarz inequality

$$\begin{aligned}
&\sup_{0\le x\le 1}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]} g_{l_1}g_{l_2}\left(\check{\eta}_{t,l_1}\check{\eta}_{t,l_2}-\overline{\check{\eta}_{l_1}\check{\eta}_{l_2}}\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\left(\eta_{t,l_1}\eta_{t,l_2}-\overline{\eta_{l_1}\eta_{l_2}}\right)\right|\\
&=\sup_{0\le x\le 1}\left|\int\int\left(\frac{1}{n}\sum_{t=1}^{[nx]}\left(Y_t(u)Y_t(s)-\overline{Y(u)Y(s)}\right)\right)\sqrt{n}\left(g_{l_1}g_{l_2}\hat{v}_{l_1}(u)\hat{v}_{l_2}(s)-v_{l_1}(u)v_{l_2}(s)\right)du\,ds\right|\\
&\le\sup_{0\le x\le 1}\left(\int\int\left(\frac{1}{n}\sum_{t=1}^{[nx]}\left(Y_t(u)Y_t(s)-\overline{Y(u)Y(s)}\right)\right)^2du\,ds\right)^{\frac{1}{2}}\\
&\qquad\cdot\left(n\int\int\left(g_{l_1}g_{l_2}\hat{v}_{l_1}(u)\hat{v}_{l_2}(s)-v_{l_1}(u)v_{l_2}(s)\right)^2du\,ds\right)^{\frac{1}{2}}.
\end{aligned}\tag{10.3}$$

By Lemma 2.3 b) in Aston & Kirch (2012a) and the separation of the eigenvalues of $c(u, s)$ the assumptions of Theorem 9.4 b) are fulfilled such that we obtain

$$\left(n\int\int\left(g_{l_1}g_{l_2}\hat{v}_{l_1}(u)\hat{v}_{l_2}(s)-v_{l_1}(u)v_{l_2}(s)\right)^2du\,ds\right)^{\frac{1}{2}}=O_P(1).$$

Lemma 2.1 in Hörmann & Kokoszka (2010) yields that $Z_t(u, s) = Y_t(u)Y_t(s)$ is L_m^4-approximable and with the invariance principle in Berkes *et al.* (2013) we obtain

$$\sup_{0\le x\le 1}\int\int\left(\frac{1}{n}\sum_{t=1}^{[nx]}\left[Y_t(u)Y_t(s)-\mathrm{E}(Y_1(u)Y_1(s))\right]\right)^2du\,ds=O_P\left(n^{-1}\right)=o_P(1). \tag{10.4}$$

It follows that

$$
\begin{aligned}
&\sup_{0\le x\le 1}\int\int\left(\frac{1}{n}\sum_{t=1}^{[nx]}\left(Y_t(u)Y_t(s)-\overline{Y(u)Y(s)}\right)\right)^2 du\,ds\\
=&\sup_{0\le x\le 1}\int\int\left(\left(\frac{1}{n}\sum_{t=1}^{[nx]}\left[Y_t(u)Y_t(s)-E\left(Y_1(u)Y_1(s)\right)\right]\right)\right.\\
&\left.-\frac{[nx]}{n}\left(\frac{1}{n}\sum_{t=1}^{n}\left[Y_t(u)Y_t(s)-E\left(Y_1(u)Y_1(s)\right)\right]\right)\right)^2 du\,ds\\
\le& C\sup_{0\le x\le 1}\int\int\left(\frac{1}{n}\sum_{t=1}^{[nx]}\left[Y_t(u)Y_t(s)-E\left(Y_1(u)Y_1(s)\right)\right]\right)^2 du\,ds\\
&+C\int\int\left(\frac{1}{n}\sum_{t=1}^{n}\left[Y_t(u)Y_t(s)-E\left(Y_1(u)Y_1(s)\right)\right]\right)^2 du\,ds\\
\le& 2C\sup_{0\le x\le 1}\int\int\left(\frac{1}{n}\sum_{t=1}^{[nx]}\left[Y_t(u)Y_t(s)-E\left(Y_1(u)Y_1(s)\right)\right]\right)^2 du\,ds=o_P(1).
\end{aligned}
$$

Hence, (10.3) yields

$$
\sup_{0\le x\le 1}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}g_{l_1}g_{l_2}\left(\check{\eta}_{t,l_1}\check{\eta}_{t,l_2}-\overline{\check{\eta}_{l_1}\check{\eta}_{l_2}}\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\left(\eta_{t,l_1}\eta_{t,l_2}-\overline{\eta_{l_1}\eta_{l_2}}\right)\right|=o_P(1). \tag{10.5}
$$

We obtain the same limit distribution if we replace $\check{\eta}_{t,l}$ by $\hat{\eta}_{t,l}=\int\left(X_t(s)-\overline{X}_n(s)\right)\hat{v}_l(s)ds$ as in our statistics. Indeed, with the notations $\tilde{Y}_t:=Y_t-\overline{Y}_n$, $\overline{\tilde{Y}(u)\tilde{Y}(s)}=\frac{1}{n}\sum_{t=1}^{n}\tilde{Y}_t(u)\tilde{Y}_t(s)$, $\overline{Y}_k(u)=\frac{1}{n}\sum_{t=1}^{k}Y_t(u)$ we obtain

$$
\begin{aligned}
&\sup_{0\le x\le 1}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\left(\check{\eta}_{t,l_1}\check{\eta}_{t,l_2}-\overline{\check{\eta}_{l_1}\check{\eta}_{l_2}}\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\left(\hat{\eta}_{t,l_1}\hat{\eta}_{t,l_2}-\overline{\hat{\eta}_{l_1}\hat{\eta}_{l_2}}\right)\right|\\
=&\sup_{0\le x\le 1}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\int\int\left(Y_t(u)Y_t(s)-\overline{Y(u)Y(s)}\right)\hat{v}_{l_1}(u)\hat{v}_{l_2}(s)du\,ds\right.\\
&\left.-\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\int\int\left(\tilde{Y}_t(u)\tilde{Y}_t(s)-\overline{\tilde{Y}(u)\tilde{Y}(s)}\right)\hat{v}_{l_1}(u)\hat{v}_{l_2}(s)du\,ds\right|\\
=&\sup_{0\le x\le 1}\left|\int\int\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\left(Y_t(u)\overline{Y}_n(s)+\overline{Y}_n(u)Y_t(s)-2\overline{Y}_n(u)\overline{Y}_n(s)\right)\right)\hat{v}_{l_1}(u)\hat{v}_{l_2}(s)du\,ds\right|\\
=&\sup_{0\le x\le 1}\left|\int\overline{Y}_{[nx]}(u)\hat{v}_{l_1}(u)du\int\sqrt{n}\,\overline{Y}_n(s)\hat{v}_{l_2}(s)ds+\int\sqrt{n}\,\overline{Y}_n(u)\hat{v}_{l_1}(u)du\int\overline{Y}_{[nx]}(s)\hat{v}_{l_2}(s)ds\right.\\
&\left.-2\frac{[nx]}{n}\int\sqrt{n}\,\overline{Y}_n(u)\hat{v}_{l_1}(u)du\int\overline{Y}_n(s)\hat{v}_{l_2}(s)ds\right|
\end{aligned}
$$

$$
\begin{aligned}
&\le \sup_{0\le x\le 1}\Bigg(\left|\int \overline{Y}_{[nx]}(u)\hat{v}_{l_1}(u)du\int \sqrt{n}\,\overline{Y}_n(s)\hat{v}_{l_2}(s)ds\right| + \left|\int \sqrt{n}\,\overline{Y}_n(u)\hat{v}_{l_1}(u)du\int \overline{Y}_{[nx]}(s)\hat{v}_{l_2}(s)ds\right|\\
&\qquad + \left|2\frac{[nx]}{n}\int \sqrt{n}\,\overline{Y}_n(u)\hat{v}_{l_1}(u)du\int \overline{Y}_n(s)\hat{v}_{l_2}(s)ds\right|\Bigg)\\
&\le \sup_{0\le x\le 1}\Bigg[\left(\int \overline{Y}^2_{[nx]}(u)du\right)^{\frac{1}{2}}\left(\int \hat{v}^2_{l_1}(u)du\right)^{\frac{1}{2}}\left(\int \left(\sqrt{n}\,\overline{Y}_n(s)\right)^2ds\right)^{\frac{1}{2}}\left(\int \hat{v}^2_{l_2}(s)ds\right)^{\frac{1}{2}}\\
&\qquad + \left(\int \left(\sqrt{n}\,\overline{Y}_n(u)\right)^2du\right)^{\frac{1}{2}}\left(\int \hat{v}^2_{l_1}(u)du\right)^{\frac{1}{2}}\left(\int \overline{Y}^2_{[nx]}(s)ds\right)^{\frac{1}{2}}\left(\int \hat{v}^2_{l_2}(s)ds\right)^{\frac{1}{2}}\\
&\qquad +2\frac{[nx]}{n}\left(\int \left(\sqrt{n}\,\overline{Y}_n(u)\right)^2du\right)^{\frac{1}{2}}\left(\int \hat{v}^2_{l_1}(u)du\right)^{\frac{1}{2}}\left(\int \overline{Y}^2_{n}(s)ds\right)^{\frac{1}{2}}\left(\int \hat{v}^2_{l_2}(s)ds\right)^{\frac{1}{2}}\Bigg]\\
&= \sup_{0\le x\le 1} 2\left(\int \overline{Y}^2_{[nx]}(u)du\right)^{\frac{1}{2}}\left(\int \left(\sqrt{n}\,\overline{Y}_n(s)\right)^2ds\right)^{\frac{1}{2}}\\
&\qquad + 2\sup_{0\le x\le 1}\frac{[nx]}{n}\left(\int \left(\sqrt{n}\,\overline{Y}_n(u)\right)^2du\right)^{\frac{1}{2}}\left(\int \overline{Y}^2_{n}(s)ds\right)^{\frac{1}{2}}\\
&\le 2\left(\int \left(\sqrt{n}\,\overline{Y}_n(s)\right)^2ds\right)^{\frac{1}{2}}\left(\left(\sup_{0\le x\le 1}\int \overline{Y}^2_{[nx]}(s)ds\right)^{\frac{1}{2}} + \left(\int \overline{Y}^2_{n}(s)ds\right)^{\frac{1}{2}}\right) = o_P(1)
\end{aligned}
\tag{10.6}
$$

as it holds with the ergodic theorem (see, for example, Ranga Rao (1962))

$$
\int \overline{Y}^2_n(s)ds = o_P(1). \tag{10.7}
$$

Combining (10.2), (10.5) and (10.6) we obtain

$$
S_{[nx]} \stackrel{D^{\mathfrak{d}}[0,1]}{\to} B_\Sigma(x).
$$

and the assertions follow by the continuous mapping theorem. □

Based on this result we can now determine the critical value as $(1-\alpha)$-quantile of the respective limit distribution. This can be done by using Monte Carlo simulations. However, it is notoriously difficult to estimate the long-run covariance (see discussion in Aston & Kirch (2012b)). In this case, i.e. if $\hat{\Sigma}$ is not consistent or the convergence too slow to be appropriate for small samples, the limit distributions in Theorem 10.1 are no longer good approximations.

10.2. Behavior Under the Alternative

Condition (9.3) is examined for two exemplary alternatives, where the projection basis is determined based on principal component analysis. The following Lemma states that, under the alternative, the empirical covariance function converges to a contaminated limit $k(u,s)$.

Lemma 10.2. *Under Assumption 9.3 it holds*

$$\int\int (\hat{c}_n(u,s) - k(u,s))^2 \, du \, ds = o_P(1),$$

where $k(u,s) = c(u,s) + (1-\theta)\delta(u,s)$.

Proof. We split the empirical covariance as follows:

$$\begin{aligned}
\hat{c}_n(u,s) =& \frac{1}{n}\sum_{t=1}^{n} \left(X_t(u) - \overline{X}_n(u)\right)\left(X_t(s) - \overline{X}_n(s)\right) \\
=& \frac{1}{n}\sum_{t=1}^{n} \left(Y_t(u) - \overline{Y}_n(u)\right)\left(Y_t(s) - \overline{Y}_n(s)\right) \\
=& \frac{1}{n}\sum_{t=1}^{[\theta n]} \left(Y_t^{(1)}(u) - \overline{Y}_n(u)\right)\left(Y_t^{(1)}(s) - \overline{Y}_n(s)\right) + \frac{1}{n}\sum_{t=[\theta n]+1}^{n} \left(Y_t^{(2)}(u) - \overline{Y}_n(u)\right)\left(Y_t^{(2)}(s) - \overline{Y}_n(s)\right).
\end{aligned}$$

Now, observe that

$$\begin{aligned}
&\int\int \left(\frac{1}{n}\sum_{t=1}^{[\theta n]} \left[\left(Y_t^{(1)}(u) - \overline{Y}_n(u)\right)\left(Y_t^{(1)}(s) - \overline{Y}_n(s)\right) - c(u,s)\right]\right)^2 du \, ds \\
\leq & C \int\int \left(\frac{1}{n}\sum_{t=1}^{[\theta n]} \left(Y_t^{(1)}(u) Y_t^{(1)}(s) - E(Y_1(u)Y_1(s))\right)\right)^2 du \, ds \\
&+ C \int\int \left((\theta + o(1))\overline{Y}_n(u)\overline{Y}_n(s) - \overline{Y}_{[\theta n]}(u)\overline{Y}_n(s) - \overline{Y}_n(u)\overline{Y}_{[\theta n]}(s)\right)^2 du \, ds.
\end{aligned}$$

Furthermore, it holds

$$\begin{aligned}
&\int\int \left((\theta + o(1))\overline{Y}_n(u)\overline{Y}_n(s) - \overline{Y}_{[\theta n]}(u)\overline{Y}_n(s) - \overline{Y}_n(u)\overline{Y}_{[\theta n]}(s)\right)^2 du \, ds \\
\leq & C \left((\theta + o(1)) \int \overline{Y}_n^2(u) du \int \overline{Y}_n^2(s) ds + \int \overline{Y}_{[\theta n]}^2(u) du \int \overline{Y}_n^2(s) ds \right. \\
&\left. + \int \overline{Y}_n^2(u) du \int \overline{Y}_{[\theta n]}^2(s) ds \right) = o_P(1)
\end{aligned}$$

by (10.7), where one needs to note that this assertion remains true under the alternative which can easily be seen by splitting the time series at the change point. By the ergodic theorem it holds

$$\begin{aligned}
&\int\int \left(\frac{1}{n}\sum_{t=1}^{[\theta n]} \left(Y_t^{(1)}(u) Y_t^{(1)}(s) - E(Y_1(u)Y_1(s))\right)\right)^2 du \, ds \\
=& \left(\frac{[\theta n]}{n}\right)^2 \int\int \left(\frac{1}{[\theta n]}\sum_{t=1}^{[\theta n]} \left(Y_t^{(1)}(u) Y_t^{(1)}(s) - E(Y_1(u)Y_1(s))\right)\right)^2 du \, ds = o_P(1).
\end{aligned}$$

Hence, we obtain

$$\int\int \left(\frac{1}{n} \sum_{t=1}^{[\theta n]} \left[\left(Y_t^{(1)}(u) - \overline{Y}_n(u)\right) \left(Y_t^{(1)}(s) - \overline{Y}_n(s)\right) - c(u,s) \right] \right)^2 du\, ds = o_P(1)$$

and analogously

$$\int\int \left(\frac{1}{n} \sum_{t=[\theta n]+1}^{n} \left[\left(Y_t^{(2)}(u) - \overline{Y}_n(u)\right) \left(Y_t^{(2)}(s) - \overline{Y}_n(s)\right) - (c(u,s) + \delta(u,s)) \right] \right)^2 du\, ds = o_P(1).$$

As

$$\int\int \left(\frac{[\theta n]}{n} c(u,s) + \frac{n - [\theta n]}{n} \left(c(u,s) + \delta(u,s)\right) - k(u,s) \right)^2 du\, ds = o_P(1),$$

where $k(u,s) = \theta c(u,s) + (1-\theta)\left(c(u,s) + \delta(u,s)\right) = c(u,s) + (1-\theta)\delta(u,s)$, it follows that

$$\int\int \left(\hat{c}_n(u,s) - k(u,s)\right)^2 du\, ds = o_P(1). \tag{10.8}$$

□

Example 10.3 (Change does not affect eigenfunctions)**.** *We consider a covariance change that does not affect the eigenfunctions, i.e. the covariance kernel after the change has the same eigenfunctions $v_l(\cdot)$ as the covariance kernel before the change:*

$$\int \left(c(u,s) + \delta(u,s)\right) v_l(s) ds = \tilde{\lambda}_l v_l(u),$$

where $v_l(\cdot)$ and λ_l are the eigenfunctions and eigenvalues of $c(u,s)$ and $\tilde{\lambda}_l = \lambda_l + \delta_l$ with $\delta_l \neq 0$ for some $l = 1, \dots, d$.

Condition (9.3) is fulfilled as it holds

$$\int \delta(u,s) v_l(s) ds = \int \left(c(u,s) + \delta(u,s)\right) v_l(s) ds - \int c(u,s) v_l(s) ds = \delta_l v_l(u) \tag{10.9}$$

and thus

$$\int\int \delta(u,s) v_{l_1}(u) v_{l_2}(s) du\, ds = \delta_{l_1} \int v_{l_1}(u) v_{l_2}(s) du = \begin{cases} 0, & l_1 \neq l_2 \\ \delta_{l_1}, & l_1 = l_2. \end{cases} \tag{10.10}$$

Assuming that the eigenvalues of $k(u,s)$ are separated, the change is still detectable if the eigendirections are estimated based on the empirical covariance function as shown in the following.

By (10.9) each v_l is an eigenfunction of $k(u, s)$ with eigenvalue $\lambda_l + \theta\delta_l$. It follows with the Cauchy-Schwarz inequality

$$\begin{aligned}
&\left| g_{l_1} g_{l_2} \int\int \delta(u,s) v_{l_1}(u) v_{l_2}(s) du\, ds - \int\int \delta(u,s) \hat{v}_{l_1}(u) \hat{v}_{l_2}(s) du\, ds \right| \\
=&\left| \int\int \delta(u,s) \left(g_{l_1} g_{l_2} v_{l_1}(u) v_{l_2}(s) - \hat{v}_{l_1}(u) \hat{v}_{l_2}(s) \right) du\, ds \right| \\
\leq& \left(\int\int \delta^2(u,s) du\, ds \right)^{\frac{1}{2}} \left(\int\int \left(g_{l_1} g_{l_2} v_{l_1}(u) v_{l_2}(s) - \hat{v}_{l_1}(u) \hat{v}_{l_2}(s) \right)^2 du\, ds \right)^{\frac{1}{2}} = o_p(1)
\end{aligned} \tag{10.11}$$

with Theorem 9.4 a) and $\delta(u, s) \in \mathcal{L}^2(\mathcal{Z})$. Hence, we get

$$\begin{aligned}
&\int\int \delta(u,s) \hat{v}_{l_1}(u) \hat{v}_{l_2}(s) du\, ds \\
=& g_{l_1} g_{l_2} \delta_{l_1} \int v_{l_1}(u) v_{l_2}(u) du + o_P(1) = \begin{cases} o_P(1), & l_1 \neq l_2 \\ g_{l_1} g_{l_2} \delta_{l_1} + o_P(1), & l_1 = l_2. \end{cases}
\end{aligned} \tag{10.12}$$

Example 10.4 (Additive noise term)**.** *In this example, a covariance change in the functional time series occurs due to an additive noise term in the scores of the first m leading eigendirections. More precisely, it holds $X_t(s) - \mu(s) = \sum_{l=1}^{\infty} \tilde{\eta}_{t,l} v_l(s)$ with*

$$\tilde{\eta}_{t,l} = \eta_{t,l} + 1_{\{\theta n < t \leq n, 1 \leq l \leq m\}} \epsilon_{t,l},$$

where $\epsilon_1, \ldots, \epsilon_n$ with $\epsilon_t = (\epsilon_{t,1}, \ldots, \epsilon_{t,m})$ are independent and identically distributed with mean 0 and $\operatorname{Cov}(\epsilon_{t,l_1}, \epsilon_{t,l_2}) = \sigma_{l_1,l_2}$ and independent of η. In this setting, it holds

$$\int\int \delta(u,s) v_{l_1}(u) v_{l_2}(s) du\, ds = \sigma_{l_1,l_2}$$

for $l_1, l_2 \in \{1, \ldots, m\}$. Hence, condition (9.3) is fulfilled if $\sigma_{l_1,l_2} \neq 0$ for some $l_1, l_2 \in \{1, \ldots, m\}$. According to (10.14) the change can be detected by projecting on the subspace spanned by the first d eigendirections of the empirical covariance kernel if $\sum_{k,l=1}^{m} \sigma_{k,l} \left(\int v_k(u) \tilde{v}_{l_1}(u) du \int v_l(s) \tilde{v}_{l_2}(s) ds \right) \neq 0$ for at least one pair $l_1, l_2 \in \{1, \ldots, \min\{d, m\}\}$, where $\{\tilde{v}_l(\cdot) : l \geq 1\}$ are the eigenfunctions of $k(u, s)$.

Proof. First observe that, as $\epsilon_{t,l}$ is independent of $\eta_{t,l}$ and as the score components are uncorrelated,

$$\operatorname{Cov}(\eta_{t,k} + \epsilon_{t,k}, \eta_{t,l} + \epsilon_{t,l}) = \operatorname{Cov}(\eta_{t,k}, \eta_{t,l}) + \operatorname{Cov}(\epsilon_{t,k}, \epsilon_{t,l}) = \begin{cases} \lambda_k + \sigma_{k,k}, & k = l, \\ \sigma_{k,l}, & k \neq l. \end{cases}$$

Hence, it holds with (9.4) for $t > \theta n$

$$\operatorname{Cov}(X_t(u), X_t(s)) = \sum_{k,l=1}^{\infty} v_k(u)v_l(s)\operatorname{Cov}(\tilde{\eta}_{t,k}, \tilde{\eta}_{t,l})$$

$$= \sum_{l=m+1}^{\infty} \lambda_l v_l(u)v_l(s) + \sum_{k,l=1}^{m} v_k(u)v_l(s)\operatorname{Cov}(\eta_{t,k} + \epsilon_{t,k}, \eta_{t,l} + \epsilon_{t,l})$$

$$= \sum_{l=m+1}^{\infty} \lambda_l v_l(u)v_l(s) + \sum_{l=1}^{m}(\lambda_l + \sigma_{l,l})v_l(u)v_l(s) + \sum_{k,l=1,k\neq l}^{m} \sigma_{k,l}v_k(u)v_l(s)$$

$$= \sum_{l=1}^{\infty} \lambda_l v_l(u)v_l(s) + \sigma_{k,l}\sum_{k,l=1}^{m} v_k(u)v_l(s)$$

$$= c(u,s) + 1_{\{\theta n < t \leq n\}}\sum_{k,l=1}^{m} \sigma_{k,l}v_k(u)v_l(s)$$

such that the change in the covariance kernel is given by

$$\delta(u,s) = \sum_{k,l=1}^{m} \sigma_{k,l}v_k(u)v_l(s). \tag{10.13}$$

For $l_1, l_2 \in \{1, \ldots, m\}$ it holds

$$\int\int \delta(u,s)v_{l_1}(u)v_{l_2}(s)du\,ds = \sum_{k,l=1}^{m} \sigma_{k,l}\int\int v_k(u)v_l(s)v_{l_1}(u)v_{l_2}(s)du\,ds$$

$$= \sum_{l,k=1}^{m} \sigma_{k,l}\left(\int v_k(u)v_{l_1}(u)du \int v_l(s)v_{l_2}(s)ds\right) = \sigma_{l_1,l_2}.$$

Hence, condition (9.3) is fufilled. Analogously to (10.11) we obtain

$$\int\int \delta(u,s)\hat{v}_{l_1}(u)\hat{v}_{l_2}(s)du\,ds = \tilde{g}_l\tilde{g}_k\sum_{k,l=1}^{m} \sigma_{k,l}\left(\int v_k(u)\tilde{v}_{l_1}(u)du \int v_l(s)\tilde{v}_{l_2}(s)ds\right) + o_P(1), \tag{10.14}$$

showing that the change is detectable if the eigendirections are estimated based on the empirical covariance function if $\sum_{k,l=1}^{m} \sigma_{k,l}\left(\int v_k(u)\tilde{v}_{l_1}(u)du \int v_l(s)\tilde{v}_{l_2}(s)ds\right) \neq 0$ for at least one pair $l_1, l_2 \in \{1, \ldots, \min\{d, m\}\}$. □

10.3. Estimation of the Long-Run Covariance

The estimation of the long-run covariance matrix is a challenging issue in change point analysis. In the case where Y_j are independent under H_0 the long-run covariance reduces to the covariance, i.e.

$$\Sigma_0 = \operatorname{Cov}\left(\operatorname{vech}[\eta_0\eta_0^T]\right) = \operatorname{E}\left(\operatorname{vech}[\eta_0\eta_0^T]\operatorname{vech}[\eta_0\eta_0^T]^T\right) - \operatorname{E}\left(\operatorname{vech}[\eta_0\eta_0^T]\right)\operatorname{E}\left(\operatorname{vech}[\eta_0\eta_0^T]\right)^T.$$

The components of the scores are known to be uncorrelated. However, this does not necessarily imply a diagonal long-run covariance as, in general, the squared components are not uncorrelated. By additionally assuming that the scores are Gaussian we get a diagonal long-run covariance depending only on the eigenvalues of the covariance kernel which can be estimated by the eigenvalues of the estimated covariance kernel. More precisely, it holds

$$\Sigma_0 = \operatorname{Cov}\left(\operatorname{vech}[\eta_0\eta_0^T]\right) = \operatorname{diag}(2\lambda_1^2, \lambda_1\lambda_2, \ldots, 2\lambda_2^2, \lambda_2\lambda_3, \ldots, 2\lambda_d^2). \tag{10.15}$$

Proof. Assuming a normal distribution, the components $\{\eta_{t,l} : l = 1, \ldots, d\}$ of the score vectors are independent. This leads to

$$\begin{aligned}
\operatorname{Cov}(\eta_{t,l_1}\eta_{t,l_2}, \eta_{t,l_3}\eta_{t,l_4}) &= \begin{cases} \operatorname{E}(\eta_{t,l_1}^4) - \operatorname{E}(\eta_{t,l_1}^2)^2, & l_1 = l_2 = l_3 = l_4 \\ \operatorname{E}(\eta_{t,l_1}^2)\operatorname{E}(\eta_{t,l_3}^2) - \operatorname{E}(\eta_{t,l_1}^2)\operatorname{E}(\eta_{t,l_3}^2), & l_1 = l_2 \neq l_3 = l_4, \\ \operatorname{E}(\eta_{t,l_1}^2)\operatorname{E}(\eta_{t,l_2}^2), & l_1 = l_3 \neq l_2 = l_4, \\ \operatorname{E}(\eta_{t,l_1}^2)\operatorname{E}(\eta_{t,l_2}^2), & l_1 = l_4 \neq l_2 = l_3, \\ 0, & \text{otherwise}, \end{cases} \\
&= \begin{cases} 3\lambda_{l_1}^2 - \lambda_{l_1}^2, & l_1 = l_2 = l_3 = l_4 \\ \lambda_{l_1}\lambda_{l_2}, & l_1 = l_3 \neq l_2 = l_4, \\ \lambda_{l_1}\lambda_{l_2}, & l_1 = l_4 \neq l_2 = l_3, \\ 0, & \text{otherwise}. \end{cases}
\end{aligned}$$

With $\operatorname{vech}[\eta_0\eta_0^T] = (\eta_{0,1}^2, \eta_{0,1}\eta_{0,2}, \ldots, \eta_{0,2}^2, \eta_{0,2}\eta_{0,3}, \ldots, \eta_{0,d}^2)$, we obtain

$$\Sigma_0 = \operatorname{Cov}\left(\operatorname{vech}[\eta_0\eta_0^T]\right) = \operatorname{diag}(2\lambda_1^2, \lambda_1\lambda_2, \ldots, 2\lambda_2^2, \lambda_2\lambda_3, \ldots, 2\lambda_d^2).$$

□

However, when dealing with a time series structure and non-Gaussian structure one has to estimate the full long-run covariance. Usual estimators, such as the Bartlett estimator, lead to problems, in particular if the dimension is large compared to the sample size (see Aston & Kirch (2012b)). Aston & Kirch (2012b) conclude that the change point procedure becomes more stable and conservative if one only corrects for the long-run variance, i.e. the diagonal of the long-run covariance matrix. In our case, this approach leads to the following test statistic:

$$\tilde{\Omega}_n = \frac{1}{n}\sum_{k=1}^{n} S_k^T \hat{D}_n^{-1} S_k, \tag{10.16}$$

where $\hat{D}_n^{-1}$ is an estimator for the inverse of the diagonal matrix given by the diagonal elements of Σ. The respective versions for the max-type statistic and the test statistics adapted to epidemic changes are obtained analogously. The test statistic in (10.16) is not pivotal in the sense that the asymptotic critical value depends on the unknown correlation structure. As a consequence, this approach requires resampling procedures. As detailed in the next chapter we apply a circular block bootstrap where we estimate the long-run variance of the bootstrap samples by the block sample variance given in

(11.3). The estimator $\hat{D}_n$ for the test statistic has to be chosen carefully with respect to its interaction with the estimator used for the bootstrap statistic. We decide to estimate the long-run variance for the test statistic with the block estimator in (11.2) as, based on simulations, this seems to yield the most stable size in comparison to, for example, the flat-top kernel estimator introduced in Politis (2011) with automatic bandwidth selection.

11. Resampling Procedures

The critical values of change point procedures are usually chosen based on the limit distribution of the test statistic under the null hypothesis. Resampling methods can be applied to get a better small sample performance but cannot be avoided if the limit distribution is non-pivotal and cannot be estimated otherwise, as is the case in our example. Previous work on resampling procedures for functional time series include McMurry & Politis (2011) for independent data and Politis & Romano (1994) as well as Dehling *et al.* (2015a) for dependent Hilbert space-valued random variables. Recently, Paparoditis (2018) introduced a sieve-type bootstrap procedure for functional time series based on a vector autoregressive representation of the scores.

In order to prove the validity of a bootstrap procedure it has to be shown that, given the observations, the bootstrap test statistic has the same limit distribution as the actual test statistic under the null hypothesis and thus leads to the same asymptotic critical values. For a good power behavior under alternatives, it is important to take into account that the underlying observations may contain a change. Ideally, the respective limit distribution holds under the null hypothesis as well as under the alternative showing that the bootstrap test is asymptotically equivalent to the asymptotic test. Theoretical justifications for the bootstrap procedure providing better small sample behavior are mainly available for simple test statistics such as the mean (see for example Singh (1981)). Therefore, simulation studies are usually performed in order to assess the size and power of a bootstrap procedure. In this work, we apply the bootstrap to the projections as resampling the functional observations would require the estimation of the covariance kernel for each bootstrap sample which is computationally infeasible. Whether this leads to theoretically justifiable bootstrap procedures remains to be seen in future work.

11.1. Circular Block Bootstrap

As discussed above, due to the non-pivotal limit distribution, resampling procedures are required to obtain critical values for our test. Aston & Kirch (2012b) obtained reasonable results by applying multivariate block bootstrap procedures for the corresponding mean change procedure. We apply a circular block bootstrap to the $\mathfrak{d} := d(d+1)/2-$dimensional sequence of the score products. In order to correct the data for a possible change we first estimate the change point in each component $i = 1, \ldots, \mathfrak{d}$ as follows:

$$\hat{k}_i^* = \operatorname{argmax}_{1\leq k\leq n} \left(\sum_{t=1}^{k} \hat{q}_i(t) - \frac{k}{n} \sum_{t=1}^{n} \hat{q}_i(t) \right), \quad \text{where } \hat{q}(t) := \operatorname{vech}[\hat{\eta}_t \hat{\eta}_t^T].$$

Thus, we can estimate the uncontaminated data by

$$\tilde{q}_i(t) = \hat{q}_i(t) - \begin{cases} \overline{\hat{q}_i^0}, & 1 \le t \le \hat{k}_i^*, \\ \overline{\hat{q}_i^1}, & t > \hat{k}_i^*, \end{cases} \tag{11.1}$$

where $\overline{\hat{q}_i^0} = \frac{1}{\hat{k}_i^*} \sum_{t=1}^{\hat{k}_i^*} \hat{q}_i(t)$ and $\overline{\hat{q}_i^1} = \frac{1}{n-\hat{k}_i^*} \sum_{t=\hat{k}_i^*+1}^{n} \hat{q}_i(t)$. We estimate the long-run variance of the original test statistic by

$$\hat{D}_n(i,i) = \frac{1}{n} \sum_{j=0}^{L-1} \left(\sum_{k=1}^{K} \tilde{q}_i(Kj+k) \right)^2, \quad \hat{D}_n(i,j) = 0 \quad \text{for } i \neq j, \tag{11.2}$$

where we use the same blocklength K as in the following bootstrap procedure. We split the whole sequence of length n circularly into overlapping subsequences of length K and repeat the following steps B times to obtain the bootstrap statistics $\tilde{\Omega}_n^{*(b)}, b = 1, \ldots, B$:

(1) Draw the starting points of the blocks as realizations of

$$U(0), \ldots, U(L) \overset{i.i.d.}{\sim} U(\{0, \ldots, n-1\})$$

with $L := \left\lfloor \frac{n}{K} \right\rfloor$.

(2) Generate a bootstrap sample by

$$q_i^*(Kj+k) := \tilde{q}_i(U(j)+k), \quad j = 0, \ldots, L, \quad k = 1, \ldots, K, \quad i = 1, \ldots, \mathfrak{d},$$

where $\tilde{q}_i(t) = \tilde{q}_i(t-n)$ if $t > n$.

(3) Calculate residuals $\tilde{q}_i^*(t)$ of the bootstrap sample of length n analogously to (11.1).

(4) Calculate D_n^* by

$$D_n^*(i,i) = \frac{1}{n} \sum_{j=0}^{L-1} \left(\sum_{k=1}^{K} \tilde{q}_i^*(Kj+k) \right)^2, \quad D_n^*(i,j) = 0 \quad \text{for } i \neq j. \tag{11.3}$$

(5) Calculate the bootstrap statistic by

$$\tilde{\Omega}_n^* = \frac{1}{n} \sum_{k=1}^{n} S_k^{*T} D_n^{*-1} S_k^*,$$

with $S_k^* = (S_k^*(1), \ldots, S_k^{*T}(\mathfrak{d}))$, $S_k^*(i) = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{k} (q_i^*(t) - \overline{q}_{n,i}^*) \right)$, $\overline{q}_{n,i}^* = \frac{1}{n} \sum_{t=1}^{n} q_i^*(t)$.

We obtain the critical values as the upper α-quantiles of the B realizations $\tilde{\Omega}_n^{*(b)}$, $b = 1, \ldots, B$. Step (3) might seem surprising at first glance as the bootstrap sample is obtained based on the residuals $\tilde{q}_i$ and thus there is no need to correct for a possible change. However, it is important to note that (11.1) does not only estimate the uncontaminated data but also leads to a reduced variance estimate even if it does not contain a change. Let us clarify this comment with the simple example of the empirical variance, i.e. $K = 1$ in (11.3). Consider a sample $x_1, \dots, x_k$, $k \in \mathbb{N}$. The arithmetic mean $\frac{1}{k}\sum_{i=1}^{k} x_i$ minimizes the function

$$f(c) = \sum_{t=1}^{k}(x_t - c)^2$$

such that it follows

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left(x_t - \frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 &= \frac{1}{n}\sum_{t=1}^{k}\left(x_t - \frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 + \frac{1}{n}\sum_{t=k+1}^{n}\left(x_t - \frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 \\ \geq &\frac{1}{n}\sum_{t=1}^{k}\left(x_t - \frac{1}{k}\sum_{i=1}^{k}x_i\right)^2 + \frac{1}{n}\sum_{t=k+1}^{n}\left(x_t - \frac{1}{n-k}\sum_{i=k+1}^{n}x_i\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t^2\end{aligned}$$

with

$$\tilde{x}_t = x_t - \begin{cases}\frac{1}{k}\sum_{i=1}^{k}x_i, & 1 \leq t \leq k \\ \frac{1}{n-k}\sum_{i=k+1}^{n}x_i, & t > k\end{cases}$$

for any $1 \leq k \leq n$. The last term is equal to the empirical variance of the sample $\tilde{x}_1, \dots, \tilde{x}_n$ as $\frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t = 0$. Hence, if we split a sample at any point k and center each of the two subsamples with the respective mean as in (11.1), the empirical variance of the resulting sample is less or equal to the empirical variance of the original sample. For the test statistic, the variance is always estimated based on the residuals in order to prevent a contamination by a possible change. If a sample is obtained under the null hypothesis, this step leads to a reduced variance estimation. The bootstrap sample of the residuals is just another sample under the null hypothesis such that without step (3) the bootstrap statistic would be corrected by systematically larger variances than the test statistic. This would lead to size problems which have been seen in simulations.

The validity of the corresponding multivariate block bootstrap has been shown in Weber (2017) Part 2 taking possible changes into account. In the functional setting this should carry over as long as the eigenvalues are well separated but a detailed theoretic analysis is beyond the scope of this work.

12. Some Alternative Test Statistics

The main drawback of change point procedures based on dimension reduction techniques is their inability to detect changes which are orthogonal to the projection space as given by condition (9.3) for the covariance change. Furthermore, the asymptotic distributions do not yield reasonable small sample approximations if the dimension of the projection space is chosen too large. This is particularly problematic when testing for a covariance change as the procedure is based on the $d(d+1)/2$-dimensional product vector when projecting on a d-dimensional subspace. Even if we only use the first two leading eigenfunctions of each direction in the separable dimension reduction and thus risk missing possible changes which do not occur in this very limited number of eigendirections we project on a 8-dimensional subspace and obtain 36-dimensional product vectors. Taking 3 eigenfunctions in each direction results in a 378-dimensional product vector which is considerably larger than the sample size and thus problematic for the multivariate procedure. This motivates us to consider fully functional test statistics. Recall that after reducing the dimension, the test statistic as given in (10.16) is based on

$$T_k = S_k^T D_n^{-1} S_k = \frac{1}{n} \sum_{l_1=1}^{d} \sum_{l_2=l_1}^{d} \frac{1}{\hat{\gamma}^2_{l_1,l_2}} \left(\sum_{t=1}^{k} \left(\hat{\eta}_{t,l_1} \hat{\eta}_{t,l_2} - \overline{\hat{\eta}_{l_1} \hat{\eta}_{l_2}} \right) \right)^2 \tag{12.1}$$

with S_k as given in (10.1), $\overline{\hat{\eta}_{l_1} \hat{\eta}_{l_2}} = \frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_{t,l_1} \hat{\eta}_{t,l_2}$ and $\hat{\gamma}^2_{l_1,l_2}$ is an estimator for $\gamma^2_{l_1,l_2} = \sum_{t \in \mathbb{Z}} \mathrm{Cov}\left(\hat{\eta}_{0,l_1} \hat{\eta}_{0,l_2}, \hat{\eta}_{t,l_1} \hat{\eta}_{t,l_2}\right)$. The weight $\frac{1}{\hat{\gamma}^2_{l_1,l_2}}$ corrects for different variances in the time series of the score products making smaller changes in components with smaller variances better visible for the test statistic. This approach is related to the likelihood ratio statistic in the multivariate case. However, the price to pay is that changes - even big ones - in score components other than the first d will not be detected at all. This seems quite unnatural. Therefore, we consider alternative test statistics related to the procedures proposed in Bucchia & Wendler (2017) and Aue *et al.* (2018) for the mean change problem which take the full functional structure into account. An obvious and well defined alternative to reducing the dimension is

$$T_k^F = \frac{1}{n} \sum_{l_1=1}^{\infty} \sum_{l_2=l_1}^{\infty} \left(\sum_{t=1}^{k} (\eta_{t,l_1} \eta_{t,l_2} - \overline{\eta_{l_1} \eta_{l_2}}) \right)^2 \tag{12.2}$$

which takes all scores of the basis expansion into account but without correcting for different variances as the multivariate test statistic does. Due to the squared summability of the eigenvalues, this infinite sum is well defined. In order to keep the advantage of T_k in terms of the weights improving the visibility of changes in components with smaller variances while not risking to miss a change due to dimension reduction we suggest the following weighting

$$T_k^W = \frac{1}{n}\sum_{l_1=1}^{\infty}\sum_{l_2=l_1}^{\infty}\frac{1}{s_{1,1}^2+\hat{\gamma}_{l_1,l_2}^2}\left(\sum_{t=1}^{k}(\hat{\eta}_{t,l_1}\hat{\eta}_{t,l_2}-\overline{\hat{\eta}_{l_1}\hat{\eta}_{l_2}})\right)^2, \tag{12.3}$$

where $s_{1,1}^2$ is the estimated variance of the first squared score component. This additive constant is needed for bounding the denominator of the weights away from zero and is chosen such that the test statistic is scale invariant. By (10.15) for independent Gaussian scores the variance of the first squared score component is given by $2\lambda_1^2$ which is the largest element in the long-run covariance matrix.

We calculate the critical values analogously to the bootstrap procedure described in Section 11. For the weighted functional procedure the long-run variances are estimated for each bootstrap sample with the block estimator as in step (4) whereas we keep the variance of the first squared score component fixed. Analogously to the multivariate procedure we also use the block estimator (11.3) for estimating the long-run covariance for the test statistics.

Remark 12.1. *T_k^F is related to the statistic $\|S_k^F\|^2$, where*

$$S_k^F(u,s) = \frac{1}{\sqrt{n}}\sum_{t=1}^{k}\left(X_t(u)X_t(s)-\overline{X(u)X(s)}\right),$$

with $\overline{X(u)X(s)} = \frac{1}{n}\sum_{t=1}^{n}X_t(u)X_t(s)$. More precisely, it holds

$$\begin{aligned}
&\|S_k^F\|^2\\
=&\frac{1}{n}\int\int\sum_{t_1,t_2=1}^{k}\left(\left(X_{t_1}(u)X_{t_1}(s)-\overline{X(u)X(s)}\right)\right)\left(X_{t_2}(u)X_{t_2}(s)-\overline{X(u)X(s)}\right)du\,ds\\
=&\frac{1}{n}\sum_{t_1,t_2=1}^{k}\sum_{l_1,l_2,l_3,l_4=1}^{\infty}(\eta_{t_1,l_1}\eta_{t_1,l_2}-\overline{\eta_{l_1}\eta_{l_2}})(\eta_{t_2,l_3}\eta_{t_2,l_4}-\overline{\eta_{l_3}\eta_{l_4}})\int v_{l_1}(u)v_{l_3}(u)du\int v_{l_2}(s)v_{l_4}(s)\,ds\\
=&\frac{1}{n}\sum_{t_1,t_2=1}^{k}\sum_{l_1,l_2=1}^{\infty}(\eta_{t_1,l_1}\eta_{t_1,l_2}-\overline{\eta_{l_1}\eta_{l_2}})(\eta_{t_2,l_1}\eta_{t_2,l_2}-\overline{\eta_{l_1}\eta_{l_2}}) = \frac{1}{n}\sum_{l_1,l_2=1}^{\infty}\left(\sum_{t=1}^{k}(\eta_{t,l_1}\eta_{t,l_2}-\overline{\eta_{l_1}\eta_{l_2}})\right)^2.
\end{aligned} \tag{12.4}$$

In contrast to T_k^F, this statistic contains all combinations of $l_1 \neq l_2$ twice such that the cross-covariances have double weights compared to the variances. This is an artefact when dealing with a bivariate symmetric function which does not occur in the mean change problem. In accordance with T_k we construct the functional statistic T_k^F such that each combination is only contained once.

13. Simulation Study

In the following simulation study we assess the empirical size and power of the proposed procedures. As there are no mathematical justifications for the bootstrap procedures for the functional test statistics available yet, the simulation study is of particular interest to evaluate their performance. Independent innovations $e_t(s) = \sum_{l=1}^{D} \eta_{t,l} v_l(s), t = 1, \ldots, n$, of length $n = 200$ are generated using a Fourier basis with $D = 55$ basis functions $\{v_1, \ldots, v_{55}\}$ on $[0, 1]$, where $v_1(s) \equiv 1$ followed by pairs of $\sin(i \cdot s)$ and $\cos(i \cdot s)$ for $i = 2, \ldots, 27$. The scores $\{\eta_{t,l} : l = 1, \ldots, 55\}$ are independent and normally distributed with standard deviations $\{\sigma_l : l = 1, \ldots, 55\}$. Following the simulation study in Aue *et al.* (2018) we consider the following settings:

Setting 1 (small number of nonzero eigenvalues): $\sigma_l = 1$ for $l = 1, \ldots, 8$ and $\sigma_l = 0$ for $l = 9, \ldots, 55$.

Setting 2 (fast decay of eigenvalues using): $\sigma_l = 3^{-l}, l = 1, \ldots, 55$.

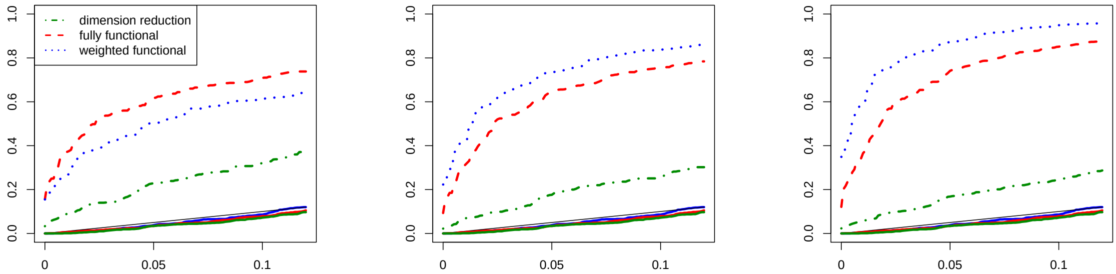
Setting 3 (slow decay of eigenvalues using): $\sigma = l^{-1}, l = 1, \ldots, 55$.

Functional autoregressive time series $X_t = \Psi(X_{t-i}) + e_t$ are simulated where the linear operator Ψ can be represented as a $D \times D$-matrix that is applied to the coefficients of the basis representation via $\{v_1, \ldots, v_{55}\}$ (for further details see Aue *et al.* (2015)). In this simulation study we use the operator with 0.4 on the diagonal and 0.1 on the superdiagonal and the subdiagonal which has infinity norm 0.6 such that the resulting functional autoregressive time series is stationary. A covariance change at the time point $0.5n$ is inserted in the first m leading eigendirections for $m = 2, 25, 50$ by adding a common additive noise term $\epsilon_{t,l} = \epsilon_t$ with variance $\sigma_{l_1,l_2} = \frac{\sigma_\epsilon^2}{m}$ according to Example 10.4. The variance of the noise term is chosen such that $\int\int \delta^2(u, s) du\, ds = 1$ for all m. In view of the application to fMRI data in Chapter 14 the multivariate procedure is applied to the projections on the subspace spanned by the first 8 eigendirections of the empirical covariance function. The empirical results are obtained based on 1000 repetitions with 1000 bootstrap iterations each.

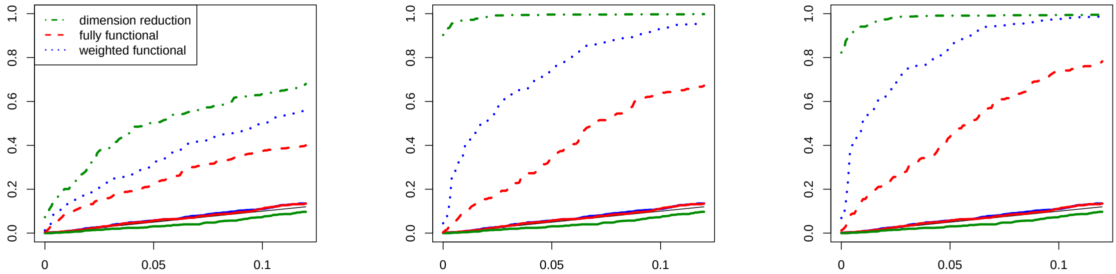
The plots in Figure 13.1 show the empirical size and the size corrected power for the different procedures. The multivariate procedure is very conservative in all settings whereas the size of the functional procedures is mostly larger but closer to the nominal level except for setting 1. However, it should be mentioned that for independent data (see Figure 13.1), using Efron's Bootstrap to obtain the critical values, all procedures keep the level very well in all settings. As expected by construction, the procedure based on PCA fails to detect the change in setting 1 for increasing m as most of the change is orthogonal to the first 8 eigendirections which are still dominating the contaminated covariance kernel. The advantage of the procedures which take the full functional structure into account is clearly visible here. The opposite power behavior can be observed for the fast decay of eigenvalues in setting 2, where the procedure based on PCA is superior to the functional procedures. In particular the unweighted functional procedure has problems to detect the change in this setting. In setting 3, the functional procedures have good power for all choices of m. In applications where one aims to explain a large amount of the variability of the data via PCA, a slow decay of eigenvalues as in setting 3 usually leads to a bad performance. However, this is not true when PCA is applied for change point detection if the change leads to an increased variability in the affected directions which is true for the alternative in this simulation study. Hence, directions which are affected by the change but orthogonal to the uncontaminated subspace are more likely to be chosen by PCA if the eigenvalues are flat. This effect can be observed when comparing the power of the multivariate procedure for $m = 50$ in setting 2 and 3. For $m = 2$ the power of the multivariate procedure is slightly better in setting 2 than in setting 3 as for the fast decay of eigenvalues the change occurs in those eigendirections which already clearly dominate in the uncontaminated subspace. Across all situations considered in this simulation study except for setting 1 with $m = 2$ the weighted functional procedure outperforms the unweighted functional procedure. Hence, the weighted functional procedure behaves not only as a compromise between the other two procedures but even more as a promising improvement of the unweighted functional approach.

Discussion

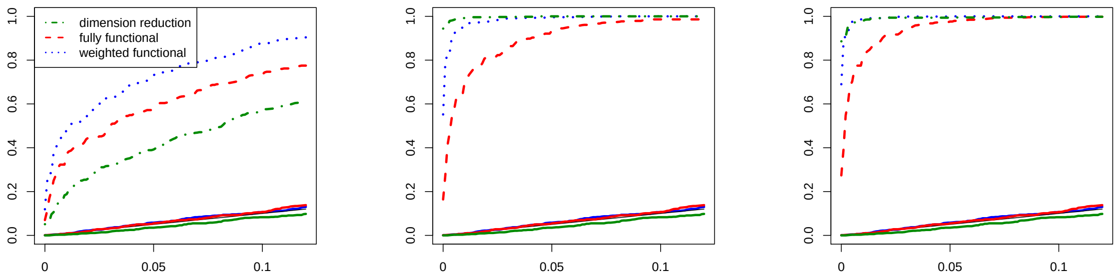
The above simulation study reveals that the functional test statistics with critical values obtained by the block bootstrap described in Chapter 11 can be liberal for dependent data. This is mostly a small sample effect which did not occur in simulations of longer time series (T=500). However, even for the sample size in the present simulations, the size is reasonable up to a nominal level of 5% and the procedures seem to be suitable for the purpose of the application in this work. We do not expect them to cause too many false rejections and we are in particular interested in a good power behavior in order to avoid nonstationarities to contaminate subsequent analyses. For future work, it would be of great interest to investigate the mathematical validity of this bootstrap approach as well as develop procedures that improve the behavior of the functional procedures for dependent data and can also deal with stronger dependency structures.



(a) Setting 1,
$\sigma_\epsilon = 1, m = 2, 25, 50$

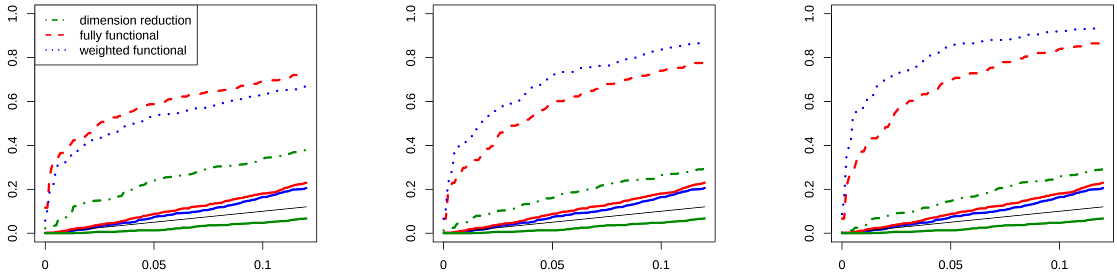


(b) Setting 2,
$\sigma_\epsilon = 0.2, m = 2, 25, 50$

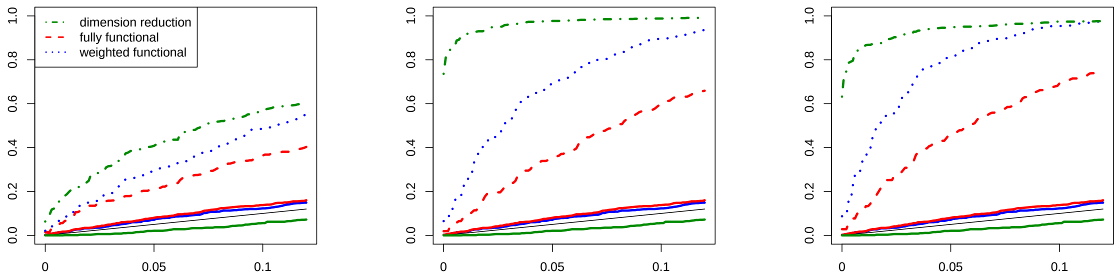


(c) Setting 3,
$\sigma_\epsilon = 0.8, m = 2, 25, 50$

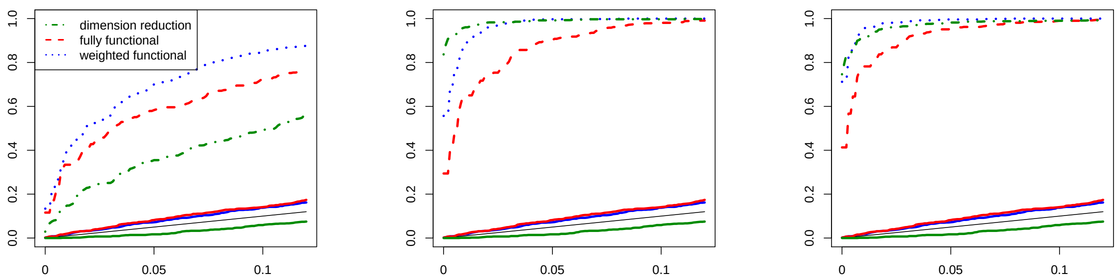
Figure 13.1.: Empirical size (solid lines) and size corrected power (dashed lines) of the proposed procedures for independent data using the multivariate procedure after dimension reduction based on (12.1) (green), the fully functional procedure based on (12.2) (red) and the weighted functional procedure based on (12.3) (blue).



(a) Setting 1,
$\sigma_\epsilon = 1, m = 2, 25, 50$



(b) Setting 2,
$\sigma_\epsilon = 0.2, m = 2, 25, 50$



(c) Setting 3,
$\sigma_\epsilon = 0.8, m = 2, 25, 50$

Figure 13.2.: Empirical size (solid lines) and size corrected power (dashed lines) of the proposed procedures with $K = 6$ for functional autoregressive time series using the multivariate procedure after dimension reduction based on (12.1) (green), the fully functional procedure based on (12.2) (red) and the weighted functional procedure based on (12.3) (blue).

14. Application to Resting State FMRI Data

We consider the publicly available *1000 Connectome Resting State Data* which consists of 1200 resting state data sets. In Aston & Kirch (2012b) a subset of 198 scans which have all been recorded at the same location (Beijing, China) are tested for an epidemic mean change. We test for deviations from covariance stationarity in those 118 data sets among these where no epidemic mean change was detected at a level of 5% in the previously mentioned work. Each scan consists of a three-dimensional image of size $64 \times 64 \times 33$ ($\sim 10^5$ voxels) recorded every 2 seconds at 225 time points. Each data set is preprocessed by voxelwise removing a polynomial trend of order 3 to correct for technical effects as for example scanner drift. We apply the separable covariance estimation and for the multivariate procedure we reduce the dimension by projecting on the 8-dimensional subspace obtained by taking the first two eigenfunctions in each direction.

14.1. Implementation of the Functional Procedures

In practice, the sums in (12.2) and (12.3) are finite as we cut after the number N of strictly positive eigenvalues obtained by principal component analysis. In the above simulation study we obtained $N \approx 100$ but for the fMRI data sets the separable covariance estimation yields $N \approx 10^5$. For the functional test statistics all combinations of the score components have to be taken into account. As the number of those combinations is of order 10^{10} this is computationally infeasible, in particular with regard to the fact that the test statistic also has to be calculated for every single bootstrap sample. However, as the variances of the score products rapidly decrease, most of the score products only have a negligible influence on the value of the test statistic in comparison to those with large variances and can thus be omitted. Figure 14.1 shows the 200 largest variances of the score products after correcting for a possible change in decreasing order exemplarily for one subject. The variance of the first squared score component is approximately 10 times larger than the second one and is thus omitted in this plot for a better visibility. It can clearly be seen that the variances strongly decrease and quickly level off at a magnitude which is only a small fraction of the larger variances. We make use of this observation to solve the computational problem discussed above where the main idea is to only consider those score products that have a sufficiently large variance compared to the variance of the first squared score component. However, estimating this ratio by calculating the empirical variance of the residuals for each of the 10^{10} combinations is still very time consuming. Therefore, we use a preselection step where we predict which combinations could possibly exceed a

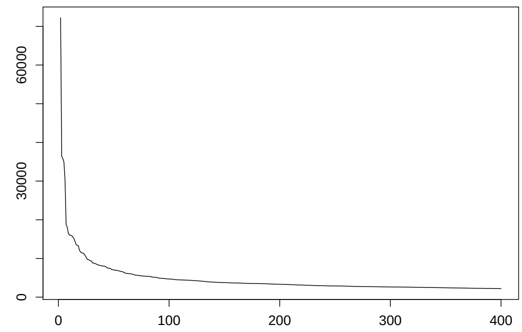


Figure 14.1.: sub06880: 2nd to 200 largest variance of score products in decreasing order.

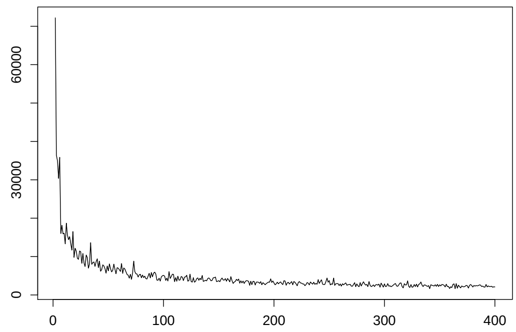


Figure 14.2.: sub06880: 2nd to 200 largest variance of score products ordered according to their approximations given by the products of the variances of the single components.

certain threshold based on the variances of the single score components. More precisely, we proceed as follows:

(1) For each $l_1, l_2 = 1, \ldots, N$ calculate

$$\hat{r}_{l_1,l_2} = \begin{cases} \frac{s_{l_1} s_{l_2}}{2 s_1^2}, & l_1 \neq l_2, \\ \frac{s_{l_1}^2}{s_1^2}, & l_1 = l_2 \end{cases} \quad \text{with} \quad s_l^2 = \frac{1}{n-1} \sum_{t=1}^{n} \tilde{\eta}_l(t)^2,$$

where $\tilde{\eta}_l(t)$ is the estimated residual of $\hat{\eta}_{t,l}$ obtained as in (11.1). Determine for $\epsilon_1 = 0.0005$

$$P := \{(l_1, l_2) : l_1, l_2 = 1, \ldots, N, \hat{r}_{l_1,l_2} \geq \epsilon_1\}.$$

The above estimation of the ratio is based on the Gaussian approximation as given in (10.15). While this is only correct in the Gaussian case and if the separability assumption is correct, according to some preliminary analysis (see Figure 14.2) it at least approximates the order of magnitude in the misspecified case. Figure 14.2 shows the variances of the score products ordered according to their approximations given by the products of the variances of the single components.

(2) Perform the following steps for each $(l_1, l_2) \in P$:

(2.1) Estimate the ratio nonparametrically (without relying on Gaussanity or the separability assumption) by

$$r_{l_1,l_2} = \frac{s_{l_1,l_2}^2}{s_{1,1}^2} \quad \text{with} \quad s_{l_1,l_2}^2 = \frac{1}{n-1} \sum_{t=1}^{n} \left(\widetilde{\eta_{l_1}\eta_{l_2}}(t)\right)^2,$$

where $\widetilde{\eta_{l_1}\eta_{l_2}}(t)$ is the estimated residual of the product $\hat{\eta}_{t,l_1}\hat{\eta}_{t,l_2}$ obtained analogously to (11.1).

(2.2) If $r_{l_1,l_2} \geq \epsilon_2 = 0.0025$ continue with step (2.3), else skip this combination and continue with step (2.1) for the next combination.

(2.3) Update

$$T_k^W = T_k^W + \frac{1}{s_{1,1}^2 + \hat{\gamma}_{l_1,l_2}^2} \left(\sum_{t=1}^{k} (\hat{\eta}_{t,l_1}\hat{\eta}_{t,l_2} - \overline{\hat{\eta}_{l_1}\hat{\eta}_{l_2}}) \right)^2, \quad k = 1, \ldots, n$$

$$T_k^F = T_k^F + \left(\sum_{t=1}^{k} (\hat{\eta}_{t,l_1}\hat{\eta}_{t,l_2} - \overline{\hat{\eta}_{l_1}\hat{\eta}_{l_2}}) \right)^2, \quad k = 1, \ldots, n.$$

with $\hat{\gamma}_{l_1,l_2}^2 = \frac{1}{n} \sum_{j=0}^{L-1} \left(\sum_{k=1}^{K} \widetilde{\eta_{l_1}\eta_{l_2}}(Kj + k) \right)^2$, where K is the block length of the respective bootstrap procedure and $L := \left\lfloor \frac{n}{K} \right\rfloor$.

(3) Calculate the test statistics: $\Omega_n^W = \frac{1}{n} \sum_{k=1}^{n} T_k^W$ and $\Omega_n^F = \frac{1}{n} \sum_{k=1}^{n} T_k^F$.

We additionally applied the procedure with $\epsilon_2 = 0.005$ to the resting state fMRI data and the results are similar to those obtained for $\epsilon_2 = 0.0025$. Hence, there is no need to further reduce the threshold as there is already no considerable loss of information when reducing it from 0.005 to 0.0025. In the preselection step we find those combinations for which $\hat{r}_{l_1,l_2} \geq \epsilon_1$ with a very conservative threshold $\epsilon_1 = 0.0005$. In the above example the predicted ratio $\hat{r}_{l_1,l_2}$ is at most 1.2 times larger than the actual ratio such than $\epsilon_1 = 0.0005$ is indeed very conservative. We calculate the critical values analogously to the bootstrap procedure described in Chapter 11. For the weighted procedure the long-run variances are estimated for each bootstrap sample with the block estimator as in step (4) whereas we keep the variance of the first squared score component fixed.

14.2. Results

In this section, we describe the results of the analysis of the *1000 Connectome Resting State Data*. We refer to the p-values obtained for $\epsilon_2 = 0.0025$ and a blocklength of $K = \sqrt[3]{225} \approx 6$. The results of the data analysis are illustrated exemplary by the score products of certain subjects as a change in the covariance structure is visible as a mean change in the products which is indicated by the black line in the plots. However, as the functional procedures are, on average, based on around 10000 score products, the plots are limited to the 64 most significant products in the sense of having the smallest p-values which are obtained by componentwise calculating the p-values for the weighted functional statistic based on the respective bootstrap components. The main findings of the data analysis can be summarized as follows:

- When testing for the AMOC alternative at a level of 5%, the null hypothesis of covariance stationarity is rejected for 43% of the data sets by the multivariate procedure, for 39% by the unweighted functional procedure and for 36% by the weighted functional procedure. As an example, in sub12220 a covariance change is detected by all considered procedures with p-values of at most 0.001. Figure 14.3 shows the 64 most significant components of the score products for the weighted functional procedure. The estimated global change point is $\hat{k}^* = 57$.

The p-values obtained by the two functional procedures are consistent meaning that in most of the cases they imply the same test decision and if they lead to different test decisions at a certain level α the p-values are nevertheless of the same magnitude, i.e. for one procedure the p-value is slightly below α and for the other procedure it slightly exceeds α. In some cases, the test decision based on the multivariate procedure differs from test decisions based on the functional procedures. Those deviations occur in both directions. On the one hand, the multivariate procedure is not able to detect changes which are orthogonal to the projection subspace. On the other hand, false alarms can occur as the few components which are considered after reducing the dimension might contain some irregularities which lead to a rejection of the null hypothesis but are not significant when considering the full functional structure. This can be observed, for example, when analyzing sub34943. The multivariate procedure detects a deviation from covariance stationarity in the 8-dimensional time series of the scores but the null hypothesis is not rejected by the functional procedures. Figure 14.4 shows the 36 score products which are considered in the multivariate procedure. Calculating the componentwise p-values of the weighted functional procedure which includes 1083 score products it turns out that more than one third of the 36 components considered in the multivariate procedure belong to the 100 smallest p-values of the weighted functional procedure.

- There are some data sets with epidemic changes. For example, sub08816 is not significant when testing for the AMOC alternative with a p-value of 0.11 for the multivariate procedure and at least 0.36 for the functional procedures whereas the test for the epidemic alternative yields p-values which are smaller than 0.04 for the functional procedures. Figure 14.5 shows the 64 most significant components of the score products for the epidemic alternative. The respective plots for the AMOC alternative can be found in Figure 14.6. A visual inspection of those two figures suggests that the epidemic model is indeed more suitable in this case. Furthermore, the small p-values are reasoned by the fact that the epidemic changes in the single components tend to be aligned.

- Some data sets contain outliers which cause the rejection of the null hypothesis. For example, testing for an epidemic change in sub08992 yields p-values smaller than 0.05 for all considered procedures. Figure 14.7 reveals that the procedure picks the outlier as epidemic change in form of a very small interval. The mean of this interval is obviously much larger than the mean of the remaining observations and additionally always at the same position determined by the outlier such that the test for an epidemic change is significant. Although, in this case, the rejection of the null hypothesis is not due to an actual change in the covariance structure, an outlier constitutes a deviation from stationarity which contaminates the subsequent analyses if they are not robust. On the other hand, if the data is only involved in analyses which require stationarity but are robust against outliers, it would be of interest to have robust change point procedures such as in Dehling *et al.* (2015c) for the univariate mean change problem. At this point it should be mentioned that even though for the AMOC alternative the null hypothesis is not rejected for sub08992 (see Figure 14.8) the procedures proposed in this work are not robust against outliers as all of them are based on the empirical covariance. For another setting, for example if the outlier occurs rather at the beginning of the observations, the null hypothesis of stationarity might also be rejected for the AMOC alternative which is the case for sub08455 (see Figure 14.9).

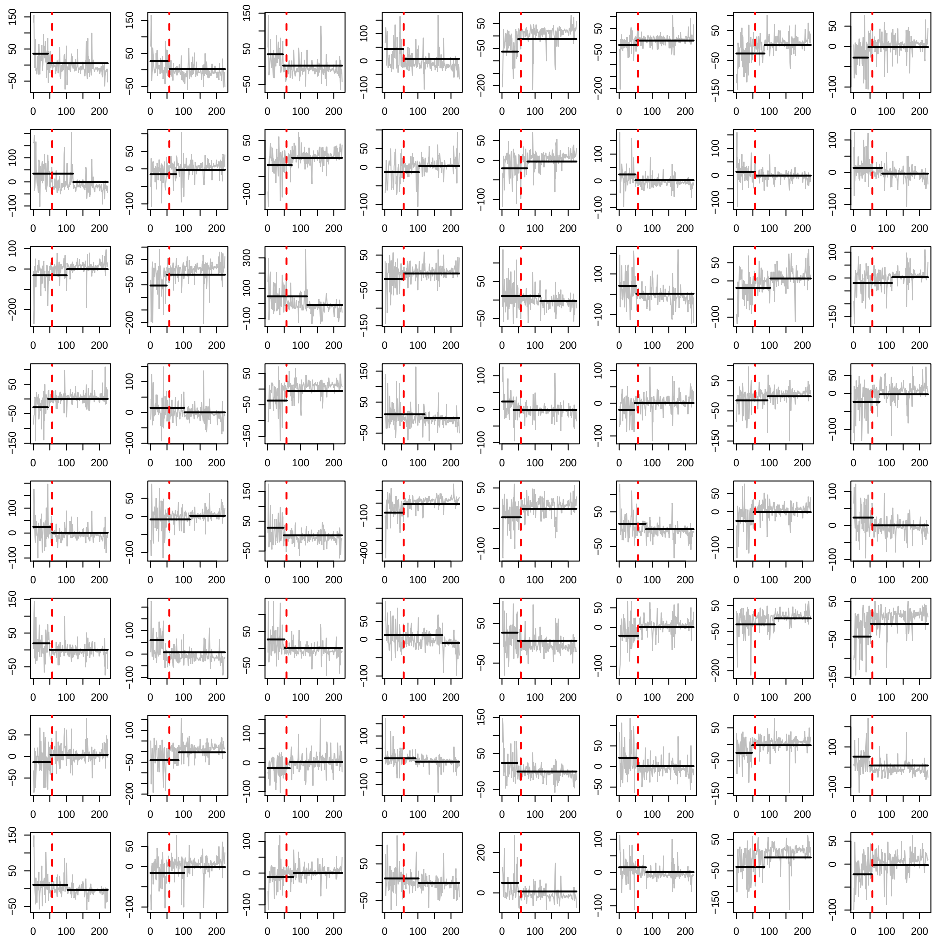


Figure 14.3.: sub12220: 64 score products with the smallest p-values for the weighted functional statistic when testing for the AMOC alternative. The global estimated change is $\hat{k}^* = 57$ (dashed line).

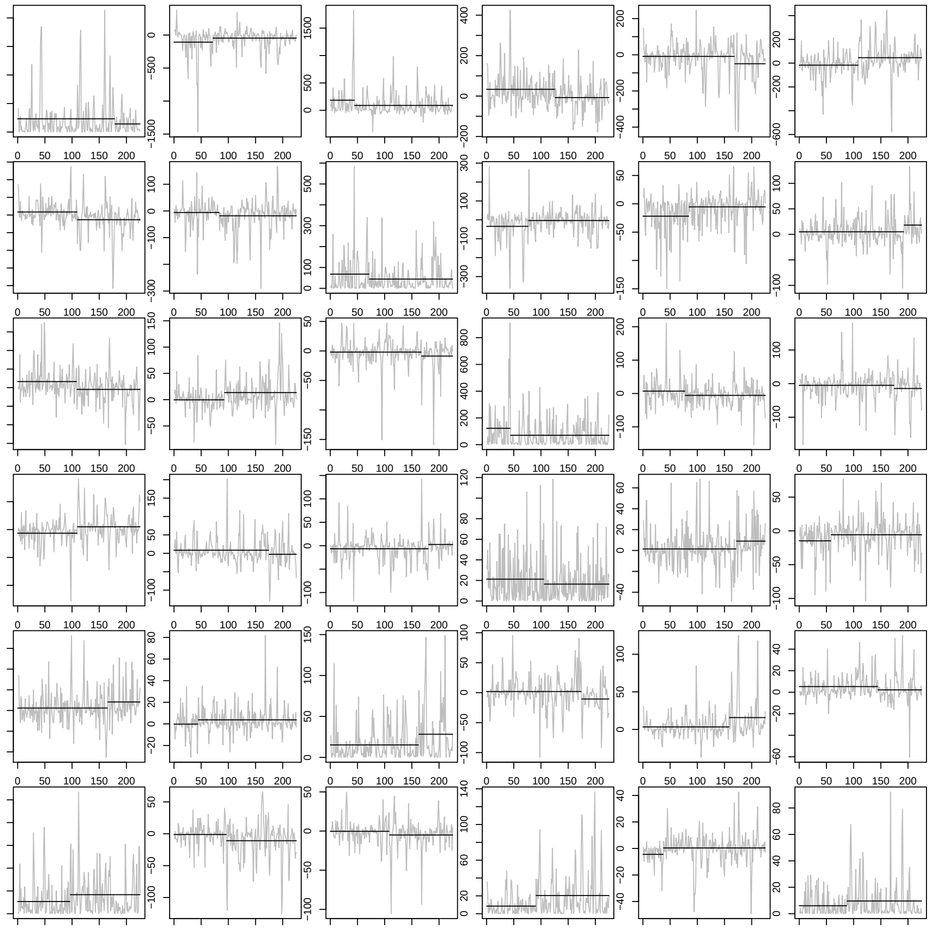


Figure 14.4.: sub34943: All 36 score products obtained by dimension reduction.

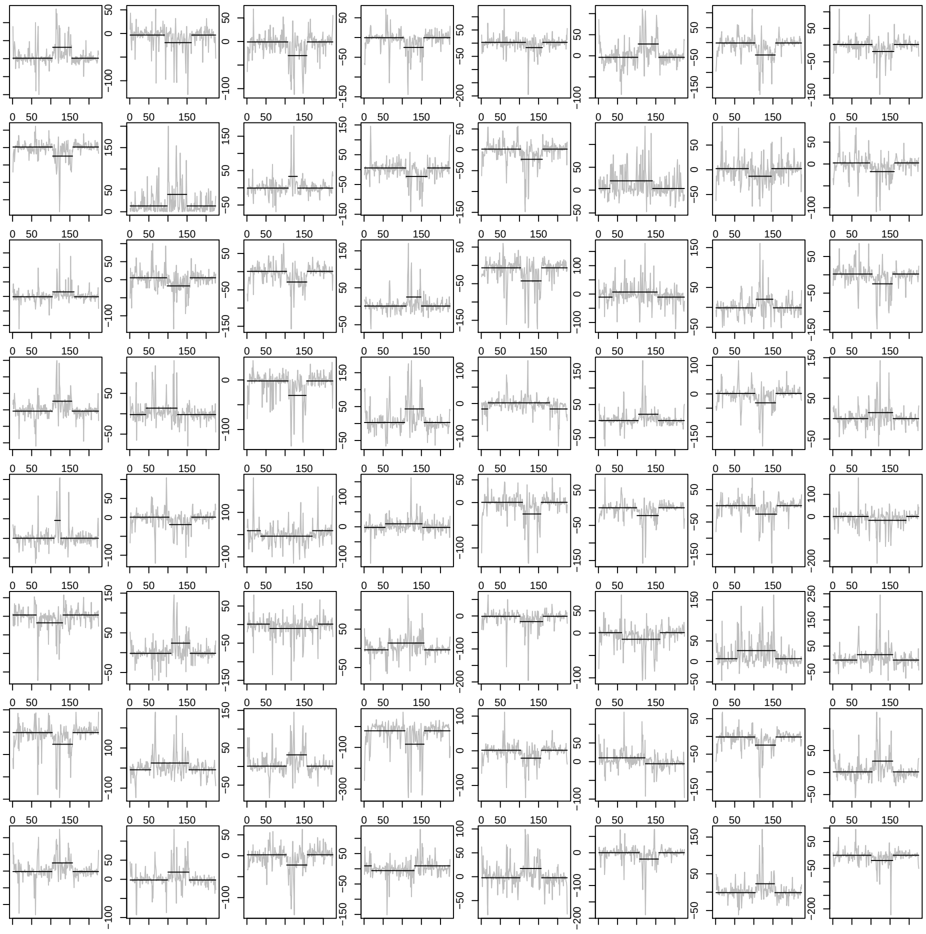


Figure 14.5.: sub08816: 64 score products with the smallest p-values for the weighted functional statistic when testing for the epidemic alternative.

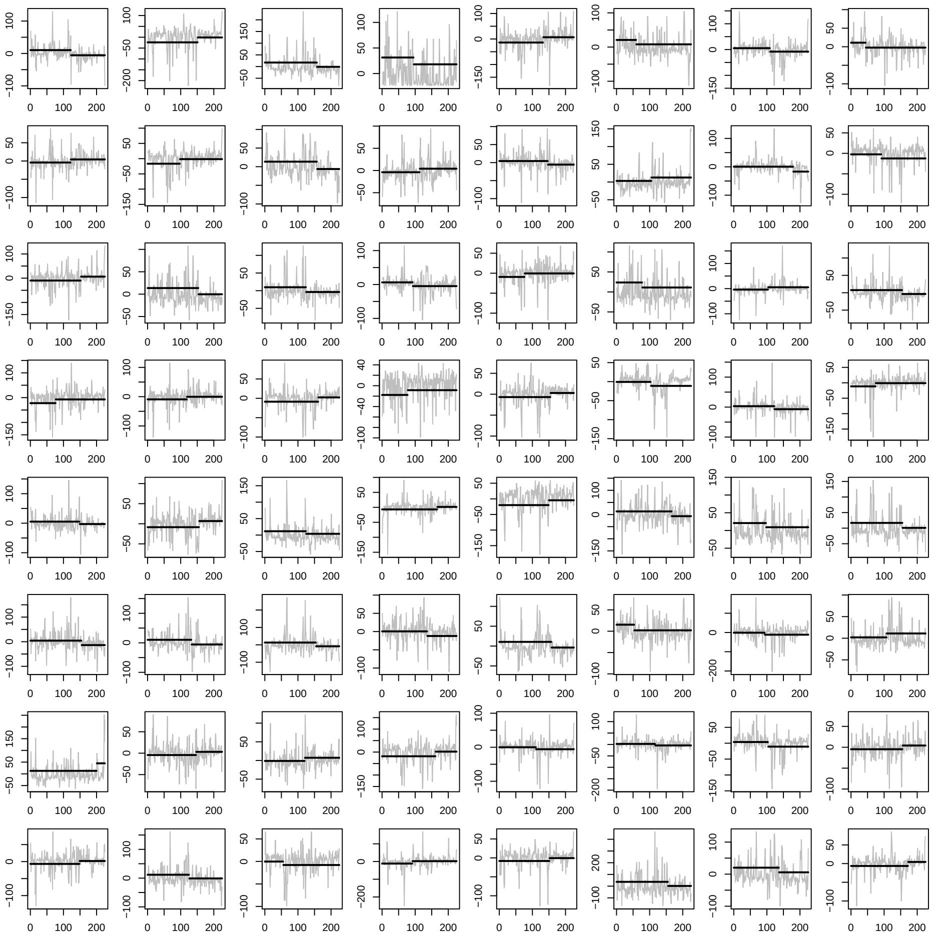


Figure 14.6.: sub08816: 64 score products with the smallest p-values for the weighted functional statistic when testing for the AMOC alternative.

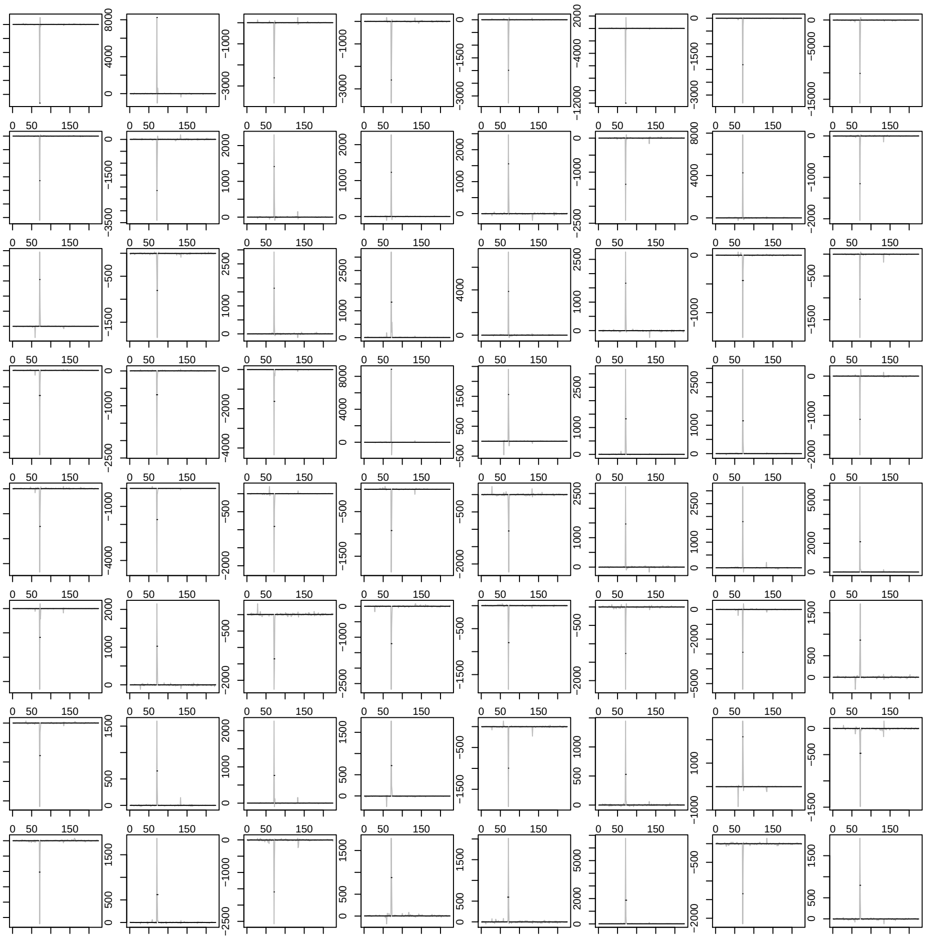


Figure 14.7.: sub08992: 64 score products with the smallest p-values for the weighted functional statistic when testing for the epidemic alternative.

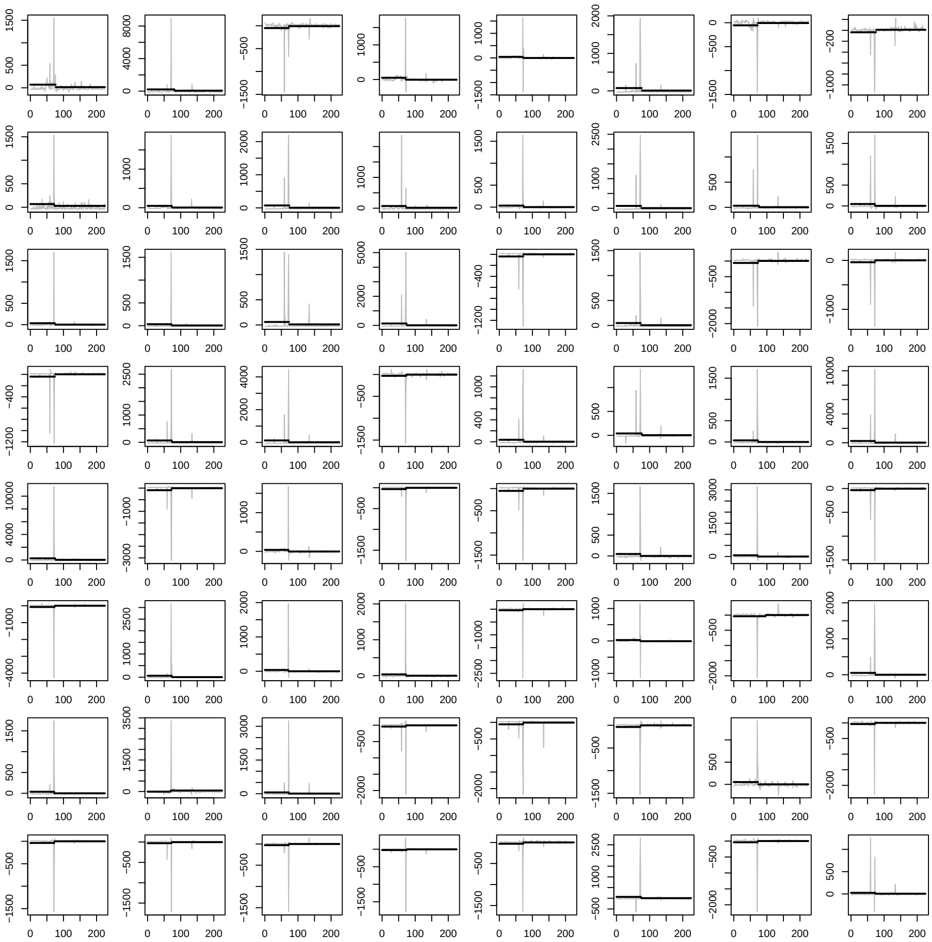


Figure 14.8.: sub08992: 64 score products with the smallest p-values for the weighted functional statistic when testing for the AMOC alternative.

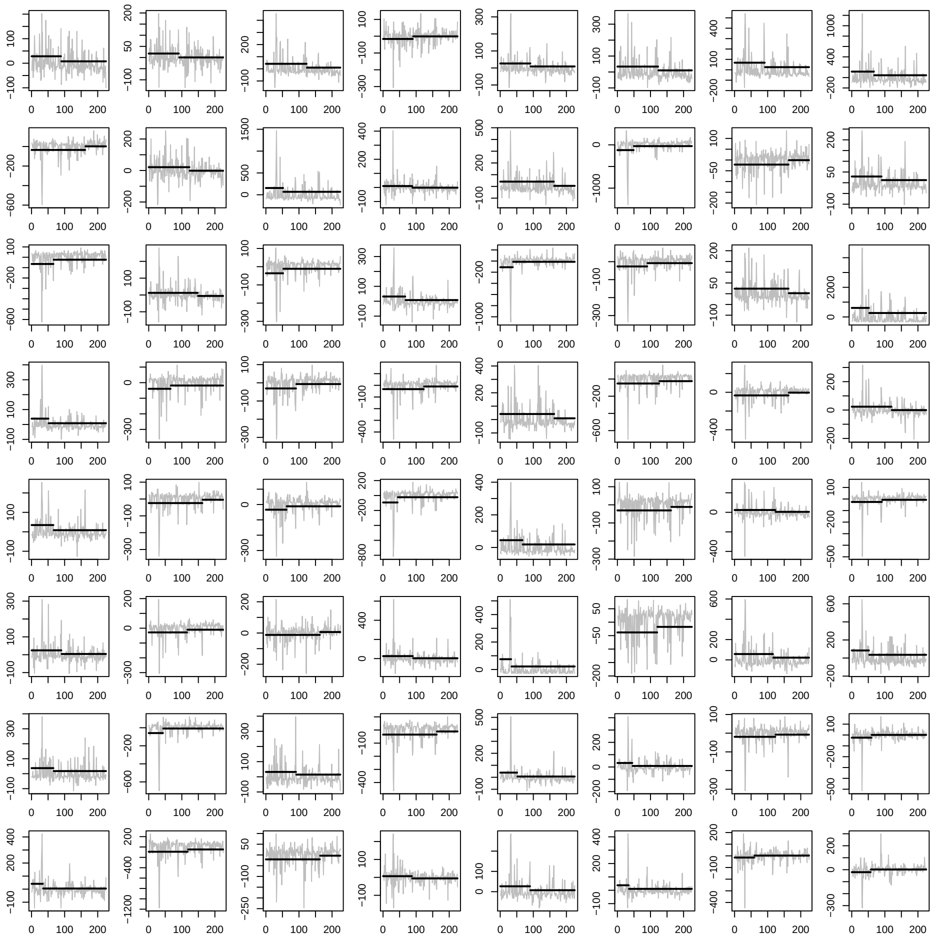


Figure 14.9.: sub08455: 64 score products with the smallest p-values for the weighted functional statistic when testing for the AMOC alternative.

15. Conclusions

In this part, different methods for detecting deviations from covariance stationarity in functional time series have been introduced and investigated with the main focus on applications to fMRI data. Dimension reduction via projections is a very common approach in functional time series analysis and enables the application of a multivariate change point procedure. We derived the asymptotic distribution of the test statistic based on the projection scores for the AMOC alternative as well as for the epidemic alternative. This asymptotic procedure requires the estimation of the long-run covariance which is statistically unstable but can be avoided by using resampling procedures. We applied a circular block bootstrap to obtain the critical values for an adapted test statistic where we only correct for the diagonal elements of the long-run covariance. This gave us a reasonable approach for detecting changes in the covariance structure of fMRI data which, however, comes with the risk of missing changes that are orthogonal to the projection subspace. As alternative solution we provided two test statistics which both take the full functional structure into account and differ with respect to the weighting. The unweighted functional test statistic has been derived from the L^2-norm of the functional partial sum process without additional weights. In contrast to that, the weights in the multivariate procedure correct for different variances of the components. We incorporate this idea into the functional approach by proposing the weighted functional test statistic. Simulations confirmed that this statistic indeed improves the unweighted functional procedure in different situations and is thus a very promising approach for the detection of change points in functional data analysis, not only for detecting changes in the covariance as considered in this work but, in an analogous version, also for the mean change problem. A mathematical investigation of this test statistic, as for example deriving the asymptotic distribution, will be of future interest. While the validity of the multivariate block bootstrap has been proven in Weber (2017), it still has to be shown for the functional procedures. However, the simulation study already indicates their reasonable performance. The application of the proposed methods to resting state fMRI data has shown that taking possible nonstationarities in the covariance structure into account is crucial. Although we only considered data sets where no mean change was detected the null hypothesis of covariance stationarity was still rejected in more than one third of the cases. Many of those nonstationarities have been detected when testing for the AMOC alternative while in some cases the epidemic alternative seemed to be more appropriate. For some data sets, the null hypothesis was rejected due to outliers, so that the development of more robust methods is of future interest.

Appendix

A. Assumptions

A.1. Asymptotics Under H_0 and H_1

Assumption 3.2. *Let the weight function satisfy*

(i) $w(m,k) = m^{-1/2}\tilde{w}(m,k)$, *where* $\tilde{w}(m,k) = \rho\left(\frac{k}{m}\right)$ *for* $k > l_m$ *with* $\frac{l_m}{m} \to 0$ *and* $\tilde{w}(m,k) = 0$ *for* $k \leq l_m$. *The function* $\rho : [0,\infty] \to \mathbb{R}^+$ *is positive and continuous.*

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

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

Assumption 3.3. *Let* $\{Y_i\}_{i\in\mathbb{Z}}$ *be a stationary time series that fulfills the following assumptions for a given kernel function* h.

(i) $\mathrm{E}\left(\left|\sum_{i=1}^m \sum_{j=k_1}^{k_2} r(Y_i,Y_j)\right|^2\right) \leq u(m)(k_2 - k_1 + 1)$ *for all* $m+1 \leq k_1 \leq k_2$ *with* $\frac{u(m)}{m^{2-2\gamma}} \log(m)^2 \to 0$ *and* γ *as in Assumption 3.2.*

(ii) The following functional central limit theorem holds for any $T > 0$

$$\left\{\frac{1}{\sqrt{m}} \sum_{i=1}^{[mt]} (h_1(Y_i), h_2(Y_i)) : 0 < t \leq T\right\} \xrightarrow{D} \left\{\left(\tilde{W}_1(t), \tilde{W}_2(t)\right) : 0 < t \leq T\right\},$$

where $\left\{\left(\tilde{W}_1(t), \tilde{W}_2(t)\right) : 0 < t \leq T\right\}$ *is a bivariate Wiener process with mean zero and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$$

with

$$\sigma_1^2 = \sum_{h\in\mathbb{Z}} Cov(h_1(Y_0), h_1(Y_h)), \quad \sigma_2^2 = \sum_{h\in\mathbb{Z}} Cov(h_2(Y_0), h_2(Y_h)). \tag{3.6}$$

(iii) For all $0 \leq \alpha < \frac{1}{2}$ *the following Hájek-Rényi-type inequality holds*

$$\sup_{1\leq k\leq m} \frac{1}{m^{\frac{1}{2}-\alpha} k^\alpha} \left|\sum_{j=1}^k h_2(Y_j)\right| = O_P(1) \qquad \textit{as } m \to \infty.$$

(iv) The following Hájek-Rényi-type inequality holds uniformly in m for any sequence $k_m > 0$

$$\sup_{k \geq k_m} \frac{1}{k} \left| \sum_{j=1}^{k} h_2(Y_j) \right| = O_P\left(\frac{1}{\sqrt{k_m}}\right) \quad \text{as } k_m \to \infty.$$

Assumption 3.12. *(i) If $\frac{k^*}{m} \to \infty$, assume that $\liminf_{t\to\infty} t\rho(t) > 0$.*

(ii) If $\frac{k^*}{m} = O(1)$, i.e. $\frac{k^*}{m} < \nu$ for all $m \geq 1$ for some $\nu > 0$, assume that there exist $t_0 > \nu, \epsilon > 0$ such that $\rho(t) > 0$ for all $t \in (t_0 - \epsilon, t_0 + \epsilon)$.

Assumption 3.13. *Let $\{Y_i\}_{i\in\mathbb{Z}}$ and $\{Z_{i,m}\}_{i\in\mathbb{Z}}$ be stationary time series that fulfill the following assumptions for a given kernel function h.*

(i) $\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r_m^*(Y_i, Z_{j,m})\right|^2\right) \leq u(m)(k_2-k_1+1)$ for all $m+k^*+1 \leq k_1 \leq k_2$ with $\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \to 0$ and γ as in Assumption 3.2.

(ii) $\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_{1,m}^*(Y_i) = O_p(1)$ as $m \to \infty$

(iii) $\frac{1}{\sqrt{k_m}}\sum_{j=m+k^*+1}^{m+k^*+k_m} h_{2,m}^*(Z_{j,m}) = O_p(1)$ as $k_m \to \infty$.

Assumption 4.4. *(i) If $\frac{k^*}{m} \to \infty$, assume that $\liminf_{t\to\infty} t\rho(t) > 0$.*

(ii) If $\frac{k^*}{m} = O(1)$, i.e. $\frac{k^*}{m} < \nu$ for all $m \geq 1$ for some $\nu > 0$, assume that there exist $t_0 > \nu, \epsilon > 0$ such that $\rho(t) > 0$ for all $t \in \left(\frac{t_0}{h} - \epsilon, \frac{t_0}{h} + \epsilon\right)$.

A.2. Stopping Time

Assumption 5.1.

(i) $\Delta_m = O(1)$.

(ii) $\sqrt{m}|\Delta_m| \to \infty$.

(iii) There exists a $\lambda > 0$ such that $k^* = [\lambda m^\beta], 0 \leq \beta < 1$. This can be divided into the following cases:

(I) $m^{\beta(1-\gamma)-1/2+\gamma}|\Delta_m| \to 0$,

(II) $m^{\beta(1-\gamma)-1/2+\gamma}|\Delta_m| \to C_1\lambda^{\gamma-1} \in (0,\infty)$,

(III) $m^{\beta(1-\gamma)-1/2+\gamma}|\Delta_m| \to \infty$.

Assumption 5.2. *Let $\{Y_i\}_{i\in\mathbb{Z}}$ and $\{Z_{i,m}\}_{i\in\mathbb{Z}}$ be stationary time series that fulfill the following assumptions for a given kernel function h.*

(i) $\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r_m^*(Y_i, Z_{j,m})\right|^2\right) \leq u(m)(k_2-k_1+1)$ *for all* $m+k^*+1 \leq k_1 \leq k_2$ *with* $\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \to 0$ *and* γ *as in Assumption 3.2.*

(ii) $\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m} h_{1,m}^*(Y_i)\right| = O_P(1)$.

(iii) *For all* $0 \leq \alpha < \frac{1}{2}$ *the following Hajek-Renyi-type inequality holds*

$$\sup_{1\leq l\leq l_m} \frac{1}{m^{\frac{1}{2}-\alpha} l^{\alpha}} \left|\sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m})\right| = O_P(1) \quad \text{as } l_m \to \infty.$$

(iv) *The following functional central limit theorem is satisfied for* $k_m \to \infty$

$$\left\{\frac{1}{\sqrt{k_m}}\sum_{j=1}^{[k_m t]}(h_2(Y_{m+j}), h_{2,m}^*(Z_{m+k^*+j,m})) : 0 < t \leq 1\right\} \xrightarrow{D} \{(W(t), W^*(t)) : 0 < t \leq 1\},$$

where $\{(W(t), W^*(t)) : 0 < t \leq T\}$ *is a bivariate Wiener process with mean zero and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma^2 & \tilde{\rho} \\ \tilde{\rho} & \sigma^{*2} \end{pmatrix}$$

with $\sigma^2 = \sum_{h\in\mathbb{Z}} Cov(h_2(Y_0), h_2(Y_h)), \sigma^{*2} = \sum_{h\in\mathbb{Z}} Cov(h_{2,m}^*(Z_{0,m}), h_{2,m}^*(Z_{h,m}))$.

Assumption 5.13.

(i) *There exists a* $\lambda > 0$ *such that* $k^* = [\lambda m^{\beta}]$ *with some* $\beta > 1$.

(ii) $\Delta_m = O(1)$.

(iii) $\lim_{m\to\infty} \frac{c_m}{\sqrt{m}|\Delta_m|} = 0$.

Assumption 5.14.

(i) $c_m \to \infty$ *and* $\limsup_{m\to\infty}\left|\frac{s_{1,m}}{c_m}\right| < 1$.

(ii) $\frac{s_{1,m}^*}{s_{1,m}} = O(1)$ *and* $\frac{1}{s_{1,m}} = O(1)$.

(iii) $\lim_{m\to\infty}\frac{|s_{1,m}|}{\sqrt{m}|\Delta_m|} = \lim_{m\to\infty}\frac{|s_{1,m}^*|}{\sqrt{m}|\Delta_m|} = 0$.

Assumption 5.15. *Let* $\{Y_i\}_{i\in\mathbb{Z}}$ *and* $\{Z_{i,m}\}_{i\in\mathbb{Z}}$ *be stationary time series that fulfill the following assumptions for a given kernel function* h.

(i) $\mathrm{E}\left(\left|\sum_{i=1}^{m}\sum_{j=k_1}^{k_2} r_m^*(Y_i, Z_{j,m})\right|^2\right) \leq u(m)(k_2-k_1+1)$ *for all* $m+k^*+1 \leq k_1 \leq k_2$ *with* $\frac{u(m)}{m^{2-2\gamma}}\log(m)^2 \to 0$ *and* γ *as in Assumption 3.2.*

(ii) $\sup_{1\le l\le l_m} \frac{1}{\sqrt{l_m}} \sum_{j=m+k^*+1}^{m+k^*+l} h_{2,m}^*(Z_{j,m}) = O_P(1)$ *as* $l_m \to \infty$.

(iii) $\frac{1}{\sqrt{k_m}} \sum_{j=1}^{k_m} h_2(Y_j) \xrightarrow{\mathcal{D}} N(0,\sigma^2)$ *as* $k_m \to \infty$.

Assumption 5.25.

(i) There exists a $\lambda > 0$ *such that* $k^* = [\lambda m]$.

(ii) $\Delta_m = O(1)$.

(iii) $\lim_{m\to\infty} \frac{c_m}{\sqrt{m}|\Delta_m|} = 0$.

Assumption 5.26.

(i) $c_m \to \infty$ *and* $\limsup_{m\to\infty} \left|\frac{s_{1,m}}{c_m}\right| < 1 + \frac{1}{\lambda}$.

(ii) $\frac{s_{1,m}^*}{s_{1,m}} = O(1)$ *and* $\frac{1}{s_{1,m}} = O(1)$.

(iii) $\lim_{m\to\infty} \frac{|s_{1,m}|}{\sqrt{m}|\Delta_m|} = \lim_{m\to\infty} \frac{|s_{1,m}^*|}{\sqrt{m}|\Delta_m|} = 0$.

B. Results from Probability Theory

Lemma B.1. *Let $\{X(m,k)\}$ and $\{Y(m,k)\}$ be sequences of random variables with*

$$\sup_{k\geq 1}|X(m,k)-Y(m,k)| \xrightarrow{P} 0$$

and

$$\sup_{k\geq 1}|Y(m,k)| \xrightarrow{\mathcal{D}} Y$$

as $m\to\infty$ for a random variable Y. Then, it holds

$$\sup_{k\geq 1}|X(m,k)| \xrightarrow{\mathcal{D}} Y \quad \text{as } m\to\infty$$

Proof. With the reverse triangle inequality it holds

$$\left|\sup_{k\geq 1}|X(m,k)| - \sup_{k\geq 1}|Y(m,k)|\right| \leq \sup_{k\geq 1}|X(m,k)-Y(m,k)| = o_P(1).$$

The assertion follows with Slutsky's Theorem. □

Lemma B.2. *Assume that it holds for any $\tau > 0$*

(i) $\sup_{k>\tau m}\left|\tilde{X}(m,T,k)-X(m,k)\right| = o_P(1)$ *as $T\to\infty$ uniformly in m,*

(ii) $\sup_{t>\tau}\left|\tilde{Y}(t,T)-Y(t)\right| = o_P(1)$ *as $T\to\infty$,*

(iii) $\sup_{k>\tau m}\tilde{X}(m,T,k) \xrightarrow{\mathcal{D}} \sup_{t>\tau}\tilde{Y}(t,T)$ *as $m\to\infty$ for T fixed*

as well as

(iv) $\sup_{1\leq k\leq \tau m}X(m,k) = o_P(1)$ *as $\tau\to 0$ uniformly in m,*

(v) $\sup_{0<t\leq\tau}Y(t) = o_P(1)$ *as $\tau\to 0$,*

where $Y(t)$ and $X(m,k)$ are positive. Given that the distribution function of $\sup_{t>\tau}\tilde{Y}(t,T)$ is continuous for all $T\in\mathbb{N},\tau>0$ it holds under the above assumptions that

$$\sup_{k\geq 1}X(m,k) \xrightarrow{\mathcal{D}} \sup_{t>0}Y(t) \text{ as } m\to\infty.$$

Proof. Consider $\epsilon > 0$ arbitrary but fixed.
By (i) and (ii) we get that for fixed $\tau, \delta > 0$ there exist $T_1 = T_1(\epsilon, \delta, \tau), T_2 = T_2(\epsilon, \delta, \tau) \in \mathbb{N}$ such that

$$\begin{aligned} &P\left(\left|\sup_{k>\tau m} \tilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) \\ \leq &P\left(\sup_{k>\tau m} \left|\tilde{X}(m,T,k) - X(m,k)\right| > \delta\right) < \epsilon \quad \text{for all } T \geq T_1, m \in \mathbb{N}, \end{aligned} \tag{B.1}$$

$$\begin{aligned} &P\left(\left|\sup_{t>\tau} \tilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right) \\ \leq &P\left(\sup_{t>\tau} \left|\tilde{Y}(t,T) - Y(t)\right| > \delta\right) < \epsilon \quad \text{for all } T \geq T_2, m \in \mathbb{N}. \end{aligned} \tag{B.2}$$

By (iii) and the continuity of the distribution function of $\sup_{t>\tau} \tilde{Y}(t,T)$ we get that for fixed $T \in \mathbb{N}, \tau > 0$ there exists an $M_0 = M_0(T, \epsilon, \tau) \in \mathbb{N}$

$$\sup_{z\in\mathbb{R}} \left| P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \leq z\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z\right)\right| < \epsilon \quad \text{for all } m \geq M_0. \tag{B.3}$$

If we now choose $T \geq \max(T_1, T_2)$, (B.1) and (B.2) hold for all $m \in \mathbb{N}$. Given that the distribution function of $\sup_{t>\tau} \tilde{Y}(t,T)$ is continuous, there exists a $\delta_0 > 0$ with

$$\left| P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z + \delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z - \delta\right)\right| < \epsilon \quad \text{for all } \delta \leq \delta_0, z \in \mathbb{R}. \tag{B.4}$$

We choose a $\delta \leq \delta_0$. According to (B.3), for the chosen T there exists an $M_0 \in \mathbb{N}$ such that it holds

$$\begin{aligned} &\left| P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \leq z + \delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z - \delta\right)\right| \\ \leq &\left| P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \leq z + \delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z + \delta\right)\right| \\ &+ \left| P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z + \delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z - \delta\right)\right| \\ < &2\epsilon \quad \text{for all } m \geq M_0, \end{aligned} \tag{B.5}$$

$$\begin{aligned} &\left| P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \leq z - \delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z + \delta\right)\right| \\ \leq &\left| P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \leq z - \delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z - \delta\right)\right| \\ &+ \left| P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z + \delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \leq z - \delta\right)\right| \\ < &2\epsilon \quad \text{for all } m \geq M_0. \end{aligned} \tag{B.6}$$

Moreover, it holds

$$\begin{aligned}
&P\left(\sup_{k>\tau m} X(m,k) \le z\right) \\
=&P\left(\sup_{k>\tau m} X(m,k) \le z, \left|\sup_{k>\tau m} X(m,k) - \sup_{k>\tau m} \tilde{X}(m,T,k)\right| \le \delta\right) \\
&+ P\left(\sup_{k>\tau m} X(m,k) \le z, \left|\sup_{k>\tau m} X(m,k) - \sup_{k>\tau m} \tilde{X}(m,T,k)\right| > \delta\right) \\
\le&P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \le z+\delta\right) + P\left(\left|\sup_{k>\tau m} X(m,k) - \sup_{k>\tau m} \tilde{X}(m,T,k)\right| > \delta\right).
\end{aligned}$$

Switching the roles of $\sup_{k>\tau m} X(m,k)$ and $\sup_{k>\tau m} \tilde{X}(m,T,k)$ we get

$$\begin{aligned}
&P\left(\sup_{k>\tau m} X(m,k) \le z\right) \\
\ge&P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \le z-\delta\right) - P\left(\left|\sup_{k>\tau m} X(m,k) - \sup_{k>\tau m} \tilde{X}(m,T,k)\right| > \delta\right).
\end{aligned}$$

Lower and upper bounds for $P\left(\sup_{t>\tau} Y(t) \le z\right)$ with respect to $\tilde{Y}(t,T)$ can be obtained analogously. Together with (B.4),(B.5) and (B.6), this leads to

$$\begin{aligned}
&P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{t>\tau} Y(t) \le z\right) \\
\le&P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \le z+\delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \le z-\delta\right) \\
&+ P\left(\left|\sup_{k>\tau m} \tilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) + P\left(\left|\sup_{t>\tau} \tilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right) \\
<&4\epsilon \quad \text{for all } m \ge M_0
\end{aligned} \tag{B.7}$$

and

$$\begin{aligned}
&P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{t>\tau} Y(t) \le z\right) \\
\ge&P\left(\sup_{k>\tau m} \tilde{X}(m,T,k) \le z-\delta\right) - P\left(\sup_{t>\tau} \tilde{Y}(t,T) \le z+\delta\right) \\
&- P\left(\left|\sup_{k>\tau m} \tilde{X}(m,T,k) - \sup_{k>\tau m} X(m,k)\right| > \delta\right) - P\left(\left|\sup_{t>\tau} \tilde{Y}(t,T) - \sup_{t>\tau} Y(t)\right| > \delta\right) \\
>&-4\epsilon \quad \text{for all } m \ge M_0.
\end{aligned} \tag{B.8}$$

Hence, we get

$$\left|P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{t>\tau} Y(t) \le z\right)\right| < 4\epsilon \quad \text{for all } m \ge M_0 \tag{B.9}$$

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

$$P\left(\sup_{1\le k\le \tau m} X(m,k) > z\right) < \epsilon \quad \text{for all } \tau \le \tau_1, m \in \mathbb{N}, \tag{B.10}$$

$$P\left(\sup_{0<t\le \tau} Y(t) > z\right) < \epsilon \quad \text{for all } \tau \le \tau_2, m \in \mathbb{N}, \tag{B.11}$$

We choose a $\tau \le \min(\tau_1, \tau_2)$. Equation (B.10) and (B.11) are fulfilled for this fixed τ and all $m \in \mathbb{N}$. For this chosen τ we can find an $M_0 \in \mathbb{N}$ as described above such that (B.9) holds. Hence, we obtain

$$\begin{aligned}
&\left|P\left(\sup_{k\ge 1} X(m,k) \le z\right) - P\left(\sup_{t>0} Y(t) \le z\right)\right| \\
=&\left|P\left(\max\left(\sup_{1\le k\le \tau m} X(m,k), \sup_{k>\tau m} X(m,k)\right) \le z\right) - P\left(\max\left(\sup_{0<t\le\tau} Y(t), \sup_{t>\tau} Y(t)\right) \le z\right)\right| \\
=&\left|P\left(\sup_{1\le k\le \tau m} X(m,k) \le z, \sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{0<t\le\tau} Y(t) \le z, \sup_{t>\tau} Y(t) \le z\right)\right| \\
=&\left|P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{1\le k\le \tau m} X(m,k) > z, \sup_{k>\tau m} X(m,k) \le z\right)\right. \\
&\left. - P\left(\sup_{t>\tau} Y(t) \le z\right) + P\left(\sup_{0<t\le\tau} Y(t) > z, \sup_{t>\tau} Y(t) \le z\right)\right| \\
\le&\left|P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{t>\tau} Y(t) \le z\right)\right| \\
&+ P\left(\sup_{1\le k\le \tau m} X(m,k) > z, \sup_{k>\tau m} X(m,k) \le z\right) + P\left(\sup_{0<t\le\tau} Y(t) > z, \sup_{t>\tau} Y(t) \le z\right) \\
\le&\left|P\left(\sup_{k>\tau m} X(m,k) \le z\right) - P\left(\sup_{t>\tau} Y(t) \le z\right)\right| \\
&+ P\left(\sup_{1\le k\le \tau m} X(m,k) > z\right) + P\left(\sup_{0<t\le\tau} Y(t) > z\right) < 6\epsilon \quad \text{for all } m \ge M_0
\end{aligned} \tag{B.12}$$

and thus the assertion. ☐

Lemma B.3. *For two sequences of events A_m and B_m with $P(B_m) \to 1$ it holds*

(i) $P(A_m) = P(A_m \cap B_m) + o(1)$.

(ii) $P(A_m) = P(A_m | B_m) + o(1)$

Proof. It holds with the rule of total probability

$$P(A_m) = P(A_m \cap B_m) + P(A_m \cap \overline{B_m}).$$

(i) With $P(\overline{B_m}) \to 0$ as $m \to \infty$ it follows $P(A_m \cap \overline{B_m}) \le P(\overline{B_m}) \to 0$ as $m \to \infty$ and thus the assertion.

(ii) The assertion follows with (i) and $P(A_m \cap B_m) = P(A_m|B_m)P(B_m) = P(A_m|B_m)(1 + o(1)) = P(A_m|B_m) + o(1)$.

☐

C. Useful Inequalities

Theorem C.1. *(Billingsley, 1999, Theorem 10.2)*
Let $Y_1, \ldots, Y_n$ be random variables and $S_k = Y_1 + \ldots + Y_k$ with $S_0 = 0$. Suppose that $\alpha > \frac{1}{2}$ and $\beta \geq 0$ and that $u_1, \ldots, u_n$ are nonnegative numbers such that

$$P\left(|S_j - S_i| \geq \epsilon\right) \leq \frac{1}{\epsilon^{4\beta}} \left(\sum_{i<l\leq j} u_l \right)^{2\alpha}, \quad 0 \leq i \leq j \leq n,$$

for $\epsilon > 0$. Then

$$P\left(\max_{0\leq k\leq n} |S_k| \geq \epsilon \right) \leq \frac{K}{\epsilon^{4\beta}} \left(\sum_{0<l\leq n} u_l \right)^{2\alpha}$$

for $\epsilon > 0$, where $K = K(\alpha, \beta)$ depends only on α and β.

Theorem C.2. *Hájek & Rényi (1955)*
Let $\{Y_j : j \geq 1\}$ be a sequence of mutually independent random variables with $\mathrm{E}(Y_k) = 0$ and finite variances $\mathrm{E}(Y_k^2) = \sigma_k^2$ for $k = 1, 2, \ldots$ and $S_k = Y_1 + \ldots + Y_k$. Then, it holds for a positive and non-increasing sequence $\{c_k\}_{k=1\geq 1}$ that for any $\epsilon > 0$ and any positive integers n_1 and n_2, $n_1 < n_2$,

$$P\left(\max_{n_1\leq k\leq n_2} c_k |S_k| \geq \epsilon \right) \leq \frac{1}{\epsilon^2} \left(c_{n_1}^2 \sum_{k=1}^{n_1} \sigma_k^2 + \sum_{k=n_1+1}^{n_2} c_k^2 \sigma_k^2 \right).$$

The following theorems are generalizations of the results in Appendix B in Kirch (2006).

Theorem C.3. *Let $\{Y_j : j \geq 1\}$ be a sequence of random variables such that for all $l_1, l_2 \geq 1, l_1 \leq l_2$,*

$$\mathrm{E}\,|S_{l_1,l_2}|^\gamma \leq C|l_2 - l_1 + 1|^\varphi$$

for some $\gamma \geq 1, \varphi > 1$ and some constant C>0, where $S_{l_1,l_2} = \sum_{j=l_1}^{l_2} Y_j$. Then, for any positive and non-decreasing sequence $0 < b_p \leq b_{p+1} \leq \ldots \leq b_q$ where $p, q \geq 1, q \geq p$, there exists a constant with $A(\varphi, \gamma) \geq 1$ with

$$\mathrm{E}\left(\max_{p\leq k\leq q} \frac{|S_{1,k}|}{b_k} \right)^\gamma \leq CA(\varphi,\gamma) \left(\frac{p-1}{q-p+1}(p-1)^{\varphi-1} + (q-p+1)^{\varphi-1} \right) \sum_{k=p}^{q} b_k^{-\gamma}.$$

Proof. For $p \leq k \leq q$ we consider the following decomposition

$$S_{1,k} = S_{1,p-1} + S_{p,k}$$

and with Minkowski's inequality we get

$$\left(\mathrm{E}\left(\max_{p\le k\le q}\frac{|S_{1,k}|}{b_k}\right)^{\gamma}\right)^{\frac{1}{\gamma}} \le \left(\mathrm{E}\left(\left(\frac{|S_{1,p-1}|}{b_p}\right)^{\gamma}\right)\right)^{\frac{1}{\gamma}} + \left(\mathrm{E}\left(\max_{p\le k\le q}\frac{|S_{p,k}|}{b_k}\right)^{\gamma}\right)^{\frac{1}{\gamma}} \tag{C.1}$$

It holds

$$\mathrm{E}\left(\frac{|S_{1,p-1}|}{b_p}\right)^{\gamma} \le C b_p^{-\gamma}(p-1)^{\varphi}. \tag{C.2}$$

Theorem B.1. in Kirch (2006) and the stationarity yield

$$\begin{aligned}
&\mathrm{E}\left(\max_{p\le k\le q}\frac{|S_{p,k}|}{b_k}\right)^{\gamma} = \mathrm{E}\left(\max_{1\le l\le q-p+1}\frac{|S_{p,p+l-1}|}{b_{p+l-1}}\right)^{\gamma} = \mathrm{E}\left(\max_{1\le l\le q-p+1}\frac{|S_{1,l}|}{b_{p+l-1}}\right)^{\gamma}\\
\le & CA(\varphi,\gamma)(q-p+1)^{\varphi-1}\sum_{l=1}^{q-p+1} b_{p+l-1}^{-\gamma} = CA(\varphi,\gamma)(q-p+1)^{\varphi-1}\sum_{k=p}^{q} b_k^{-\gamma}.
\end{aligned} \tag{C.3}$$

with a constant $A(\varphi,\gamma)\ge 1$. Combining (C.1) with (C.2) and (C.3) we obtain

$$\begin{aligned}
&\left(\mathrm{E}\left(\max_{p\le k\le q}|S_{1,k}|\right)^{\gamma}\right)^{\frac{1}{\gamma}}\\
\le & \left(Cb_p^{-\gamma}(p-1)^{\varphi}\right)^{\frac{1}{\gamma}} + \left(CA(\varphi,\gamma)(q-p+1)^{\varphi-1}\sum_{k=p}^{q} b_k^{-\gamma}\right)^{\frac{1}{\gamma}}\\
\le & \left(C(p-1)^{\varphi-1}\frac{p-1}{q-p+1}\sum_{k=p}^{q} b_k^{-\gamma}\right)^{\frac{1}{\gamma}} + \left(CA(\varphi,\gamma)(q-p+1)^{\varphi-1}\sum_{k=p}^{q} b_k^{-\gamma}\right)^{\frac{1}{\gamma}}\\
\le & 2\left(CA(\varphi,\gamma)\left(\frac{p-1}{q-p+1}(p-1)^{\varphi-1} + (q-p+1)^{\varphi-1}\right)\sum_{k=p}^{q} b_k^{-\gamma}\right)^{\frac{1}{\gamma}}
\end{aligned}$$

and thus the assertion. □

Theorem C.4. *Let $\{Y_j : j\ge 1\}$ be a sequence of random variables. Assume that there exist non-negative numbers $a_p,\dots,a_q$ and a fixed $\gamma>0$ such that for all $p\le\tilde q\le q$*

$$\mathrm{E}\left(\max_{p\le k\le\tilde q}|S_k|\right)^{\gamma} \le \sum_{k=p}^{\tilde q} a_k.$$

Then it holds for any positive and non-decreasing sequence $0<b_p\le b_{p+1}\le\ldots\le b_q$

$$\mathrm{E}\left(\max_{p\le k\le q}\frac{|S_k|}{b_k}\right)^{\gamma} \le 4\sum_{k=p}^{q}\frac{a_k}{b_k^{\gamma}}.$$

Proof. This proof is similar to the proof of Fazekas & Klesov (2000)[Theorem 1.1]. Without loss of generality we can assume that $b_p=1$. For $c=2^{\frac{1}{\gamma}}$ consider the sets

$$A_i = \left\{k : c^i \le b_k < c^{i+1}\right\}, \quad i=0,1,2,\dots.$$

The index of the last nonempty A_i is given by $i(q) = \max\{i : A_i \neq \emptyset\}$. For $i = 0, 1, 2, \ldots$ we define

$$k(i) = \begin{cases} \max\{k : k \in A_i\}, & \text{if } A_i \neq \emptyset, \\ k(i-1) & \text{if } A_i = \emptyset \end{cases}$$

and set $k(-1) = 0$. Let

$$\delta_l = \begin{cases} \sum_{j=k(l-1)+1}^{k(l)} a_j, & \text{if } A_i \neq \emptyset, \\ 0, & \text{if } A_i = \emptyset \end{cases}$$

for $i = 0, 1, 2, \ldots$. Hence, we get

$$\begin{aligned}
\mathrm{E}\left(\max_{p\le k\le q}\frac{|S_k|}{b_k}\right)^\gamma &\le \sum_{i=0}^{i(q)} \mathrm{E}\left(\max_{l\in A_i}\frac{|S_l|}{b_l}\right)^\gamma \le \sum_{i=0}^{i(q)} c^{-i\gamma}\,\mathrm{E}\left(\max_{l\in A_i}|S_l|\right)^\gamma \\
&\le \sum_{i=0}^{i(q)} c^{-i\gamma}\,\mathrm{E}\left(\max_{p\le k\le k(i)}|S_k|\right)^\gamma \le \sum_{i=0}^{i(q)} c^{-i\gamma}\sum_{k=p}^{k(i)} a_k \\
&= \sum_{i=0}^{i(q)} c^{-i\gamma}\sum_{l=0}^{i}\delta_l = \sum_{l=0}^{i(q)}\delta_l\sum_{i=l}^{i(q)} c^{-i\gamma} \le \sum_{l=0}^{i(q)}\delta_l\sum_{i=l}^{\infty} c^{-i\gamma} \\
&= \sum_{l=0}^{i(q)}\delta_l\sum_{i=0}^{\infty} c^{-(i+l)\gamma} = \frac{1}{1-c^{-\gamma}}\sum_{l=0}^{i(q)}\delta_l c^{-l\gamma} \\
&= \frac{1}{1-c^{-\gamma}}\sum_{l=0}^{i(q)} c^{-l\gamma}\sum_{k=k(l-1)+1}^{k(l)} a_k \\
&\le \frac{c^\gamma}{1-c^{-\gamma}}\sum_{l=0}^{i(q)}\sum_{k=k(l-1)+1}^{k(l)}\frac{a_k}{b_k^\gamma} = 4\sum_{k=p}^{q}\frac{a_k}{b_k^\gamma}.
\end{aligned}$$

□

Theorem C.5. *Let $\{Y_j : j \ge 1\}$ be a sequence of random variables such that for all $l_1, l_2 \ge 1, l_1 \le l_2$,*

$$\mathrm{E}\,|S_{l_1,l_2}|^\gamma \le C|l_2 - l_1 + 1|^\varphi$$

for some $\gamma \ge 1, \varphi > 1$ and some constant C>0, where $S_{l_1,l_2} = \sum_{j=l_1}^{l_2} Y_j$. Then, for any positive and and non-decreasing sequence $0 < b_p \le b_{p+1} \le \ldots \le b_q$ where $p, q \ge 1, q \ge p$, there exists a constant with $A(\varphi,\gamma) \ge 4$ with

$$\mathrm{E}\left(\max_{p\le k\le q}\frac{|S_{1,k}|}{b_k}\right)^\gamma \le CA(\varphi,\gamma)\sum_{k=p}^{q}\frac{(p-1)^\varphi + (k-p+1)^{\varphi-1}}{b_k}.$$

Proof. With $b_k \equiv 1$ in Theorem C.3 we obtain

$$\mathrm{E}\left(\max_{p\le k\le \tilde q}|S_{1,k}|\right)^\gamma \le C\tilde A(\varphi,\gamma)\left((p-1)^\varphi + (\tilde q - p + 1)^\varphi\right)$$

$$\le C\varphi\tilde A(\varphi,\gamma)\sum_{k=1}^{\tilde q - p + 1}\left((p-1)^\varphi + k^{\varphi-1}\right) = C\varphi\tilde A(\varphi,\gamma)\sum_{l=p}^{\tilde q}\left((p-1)^\varphi + (l-p+1)^{\varphi-1}\right)$$

as $\sum_{k=1}^{\tilde{q}-p+1} k^{\varphi-1} \geq \int_0^{\tilde{q}-p+1} x^{\varphi-1}dx = \frac{1}{\varphi}(\tilde{q}-p+1)^{\varphi}$ and $(p-1)^{\varphi} \leq \sum_{k=1}^{\tilde{q}-p+1}(p-1)^{\varphi}$. Now, we get with Theorem C.4

$$\mathrm{E}\left(\max_{p\leq k\leq q}\frac{|S_{1,k}|}{b_k}\right)^{\gamma} \leq 4C\varphi\tilde{A}(\varphi,\gamma)\sum_{k=p}^{q}\frac{(p-1)^{\varphi}+(k-p+1)^{\varphi-1}}{b_k}.$$

□

Theorem C.6. *Let $\{Y_j : j \geq 1\}$ be a sequence of random variables such that for all $l_1, l_2 \geq 1, l_1 \leq l_2$,*

$$\mathrm{E}\,|S_{l_1,l_2}|^{\gamma} \leq C(l_2 - l_1 + 1)$$

for some $\gamma > 1$ and some constant C>0, where $S_{l_1,l_2} = \sum_{j=l_1}^{l_2} Y_j$. Then it holds for all $p, q \geq 1, p \leq q$,

$$\mathrm{E}\left(\max_{p\leq k\leq q}|S_{1,k}|\right)^{\gamma} \leq C\left(\log_2\left(2^{\left(\frac{p-1}{q-p+1}\right)^{\gamma}+1}(q-p+1)\right)\right)^{\gamma}(q-p+1).$$

Proof. It holds

$$\mathrm{E}\,|S_{1,p-1}|^{\gamma} \leq C(p-1). \tag{C.4}$$

Theorem 3 in Móricz (1976) yields

$$\mathrm{E}\left(\max_{p\leq k\leq q}|S_{p,k}|\right)^{\gamma} = \mathrm{E}\left(\max_{1\leq l\leq q-p+1}|S_{p,p-1+l}|\right)^{\gamma} \leq C(\log_2(2(q-p+1)))^{\gamma}(q-p+1) \tag{C.5}$$

Combining (C.1) with (C.4) and (C.5) we obtain

$$\begin{aligned}\left(\mathrm{E}\left(\max_{p\leq k\leq q}|S_{1,k}|\right)^{\gamma}\right)^{\frac{1}{\gamma}} &\leq (C(p-1))^{\frac{1}{\gamma}} + \log_2(2(q-p+1))\,(C(q-p+1))^{\frac{1}{\gamma}}\\ &= \left(\left(\frac{p-1}{q-p+1}\right)^{\frac{1}{\gamma}} + \log_2(2(q-p+1))\right)(C(q-p+1))^{\frac{1}{\gamma}}\\ &= \log_2\left(2^{\left(\frac{p-1}{q-p+1}\right)^{\gamma}+1}(q-p+1)\right)(C(q-p+1))^{\frac{1}{\gamma}}\end{aligned}$$

and thus the assertion. □

Theorem C.7. *Let $\{Y_j : j \geq 1\}$ be a sequence of random variables such that for all $l_1, l_2 \geq 1, l_1 \leq l_2$,*

$$\mathrm{E}\,|S_{l_1,l_2}|^{\gamma} \leq C(l_2 - l_1 + 1)$$

for some $\gamma \geq 1$ and some constant C>0, where $S_{l_1,l_2} = \sum_{j=l_1}^{l_2} Y_j$. Then for any positive and non-increasing sequence $b_p \geq b_{p+1} \geq \ldots \geq b_q > 0$ where $p, q \geq 1, q \geq p$, it holds

$$\mathrm{E}\left(\max_{p\leq k\leq q}\frac{|S_{1,k}|}{b_k}\right)^{\gamma} \leq 4C\left(\log_2\left(2^{\left(\frac{p-1}{q-p+1}\right)^{\gamma}+1}(q-p+1)\right)\right)^{\gamma}\sum_{j=p}^{q}b_j^{-\gamma}.$$

Proof. The assertion follows by first applying Theorem C.6 and then Theorem C.4 with $a_k = C\left(\log_2\left(2^{\left(\frac{p-1}{q-p+1}\right)^{\gamma}+1}(q-p+1)\right)\right)^{\gamma}, \quad k = p, \ldots, q.$ □

Corollary C.8. *If $q \geq 2p-2$ Theorem C.7 implies*

$$\mathrm{E}\left(\max_{p\leq k\leq q}\frac{|S_{1,k}|}{b_k}\right)^{\gamma} \leq 4C\left(\log_2\left(4(q-p+1)\right)\right)^{\gamma}\sum_{j=p}^{q}b_j^{-\gamma}.$$

D. Functionals of Mixing Processes

Lemma D.1. *(Borovkova* et al. *, 2001, Lemma 2.24) Let $\{Y_i\}_{i\in\mathbb{Z}}$ be a centered 1-approximating functional with approximating constants $\{a_k\}_{k\geq 0}$ of an absolutely regular process with mixing coefficients $\{\beta(k)\}_{k\geq 0}$. Suppose that one of the following two conditions holds:*

a) Y_0 is bounded a.s. and $\sum_{k=0}^{\infty} k^2(a_k + \beta(k)) < \infty$

b) $E|Y_0|^{4+\delta} < \infty$ and $\sum_{k=0}^{\infty} k^2 \left(a_k^{\frac{\delta}{3+\delta}} + \beta^{\frac{\delta}{4+\delta}}(k) \right) < \infty$.

Then, there exists a constant C such that it holds for $S_k = Y_1 + \ldots + Y_k$

$$\mathrm{E}(S_k^4) \leq Ck^2.$$

Proposition D.2. *(Borovkova* et al. *, 2001, Proposition 2.11.) Let $\{Y_i\}_{i\in\mathbb{Z}}$ be a centered 1-approximating functional of $\{Z_i\}_{i\in\mathbb{Z}}$ with approximating constants $\{a_k\}_{k\geq 0}$. Then $\{u(Y_i)\}_{i\in\mathbb{Z}}$ is also a 1-approximating functional of $\{Z_i\}_{i\in\mathbb{Z}}$ with approximating constants*

$$a_k' = \Phi(\sqrt{2a_k}) + 2\|u(Y_0)\|_{2+\delta}(2a_k)^{\frac{1+\delta}{4+2\delta}},$$

if $\|u(Y_0)\|_{2+\delta} < \infty$, for some $\delta > 0$. If u is bounded, the same holds with

$$a_k' = \Phi(\sqrt{2a_k}) + 2\|u(Y_0)\|_{2+\delta}\sqrt{2a_k}.$$

Lemma D.3. *(Borovkova* et al. *, 2001, Lemma 2.15.) Let h be a p-continuous kernel and define*

$$h_1(x) = \int_{\mathbb{R}} h(x,y)dF(y).$$

Then, h_1 is also p-continuous.

Proposition D.4. *Let $\{Y_i : i \geq 1\}$ be a 1-approximating functional with approximating constants $\{a_k\}_{k\geq 0}$ of an absolutely regular process with mixing coefficients $\{\beta(k)\}_{k\geq 0}$. Furthermore, let $h_1(\cdot), h_2(\cdot)$ be bounded 1-continuous functions with*

$$\mathrm{E}(h_1(Y_1)) = \mathrm{E}(h_2(Y_1)) = 0$$

such that

$$\sum_{k\geq 1} k^2 \left(\beta(k) + \sqrt{a_k} + \phi(\sqrt{a_k})\right) < \infty. \tag{D.1}$$

Then, the functional central limit theorem in Assumption 3.3 (ii) holds.

Proof. Proposition 2 in Dehling *et al.* (2015b) can be obtained in the same way for $\{0 \leq t \leq T\}, T > 0$. Regarding the summability condition it should be noted that in the proof of Proposition 2 in Dehling *et al.* (2015b) Lemma D.1 a) is applied to $\{h(Y_i) : i \geq 1\}$ which, with Proposition D.2, has approximation constants $a'_k = \Phi(\sqrt{2a_k}) + C\sqrt{a_k}$. Hence, a stronger assumption than (53) in Dehling *et al.* (2015b) is needed which is given in (D.1). □

E. Further Simulations

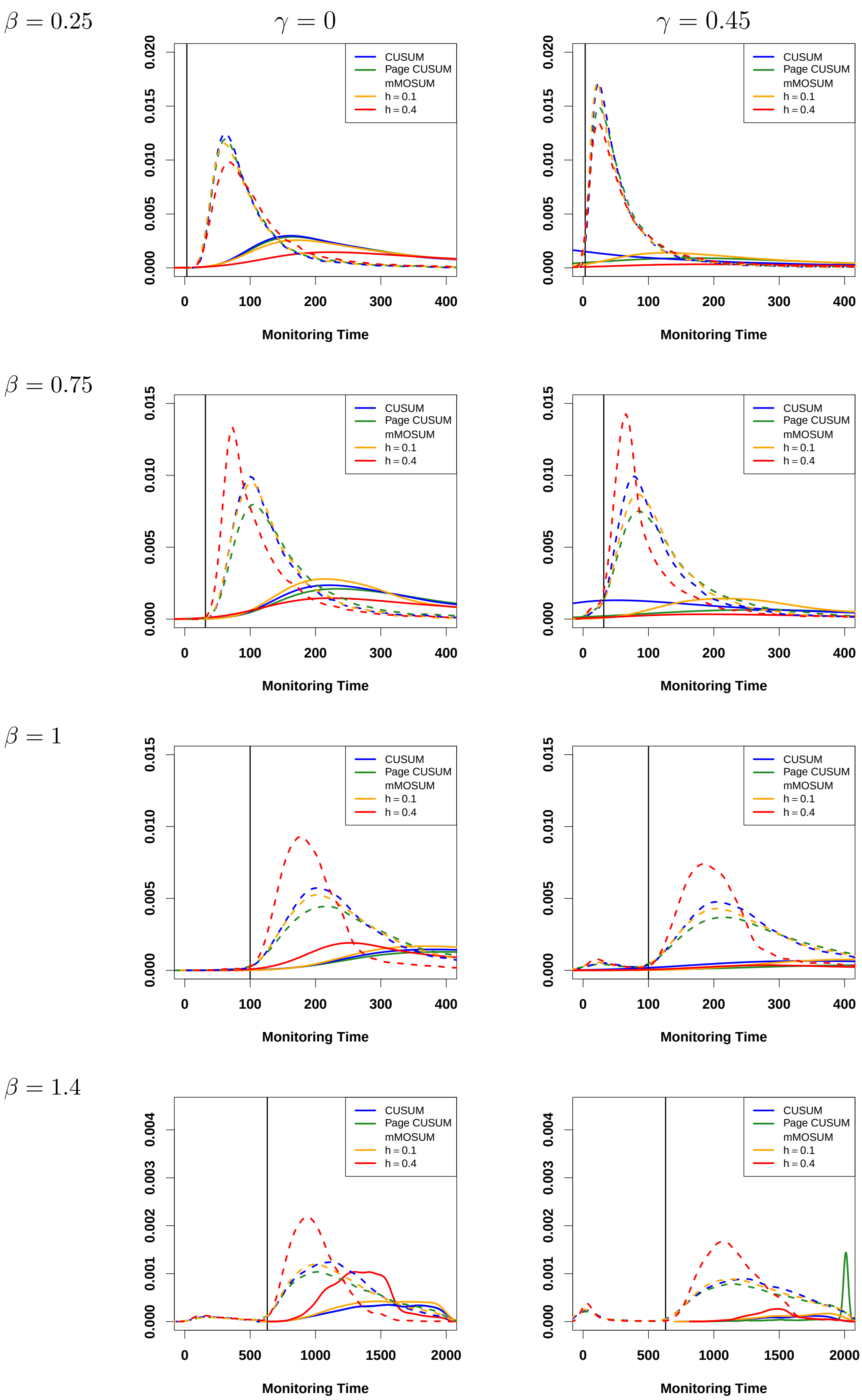


Figure E.1.: Estimated densities of the stopping time for the CUSUM kernel (solid lines) and the Wilcoxon kernel (dashed lines) for independent $t(3)$-distributed observations.

Bibliography

Andrews, D.W. K. 1984. Non-strong mixing autoregressive processes. *Journal of Applied Probability*, **21**(4), 930–934.

Aston, J. A. D., & Kirch, C. 2012a. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, **109**, 204–220.

Aston, J. A. D., & Kirch, C. 2012b. Evaluating stationarity via change-point alternatives with applications to fMRI data. *The Annals of Applied Statistics*, **6**(4), 1906–1948.

Aue, A. 2003. *Sequential change-point analysis based on invariance principles.* dissertation, Universität zu Köln.

Aue, A., & Horváth, L. 2004. Delay time in sequential detection of change. *Statistics & Probability Letters*, **67**(3), 221–231.

Aue, A., Horváth, L., Hušková, M., & Kokoszka, P. 2006. Change-point monitoring in linear models. *The Econometrics Journal*, **9**(3), 373–403.

Aue, A., Hörmann, S., & Horváth, L. 2009a. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, **37**, 4046–4087.

Aue, A., Horváth, L., & Reimherr, M.L. 2009b. Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics*, **149**(2), 174–190.

Aue, A., Gabrys, R., Horváth, L., & Kokoszka, P. 2009c. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, **100**, 2254–2269.

Aue, A., Horváth, L., Kühn, M., & Steinebach, J. 2012. On the reaction time of moving sum detectors. *Journal of Statistical Planning and Inference*, **142**(8), 2271–2288.

Aue, A., Norinho, D.D., & Hörmann, S. 2015. On the prediction of stationary functional time series. *Journal of the American Statistical Association*, **110**(509), 378–392.

Aue, A., Rice, G., & Sönmez, O. 2018. Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**(3), 509–529.

Basseville, M., Nikiforov, I.V., *et al.* . 1993. *Detection of abrupt changes: theory and application.* Vol. 104. Prentice Hall Englewood Cliffs.

Berkes, I., Gabrys, R., Horváth, L., & Kokoszka, P. 2009. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 927–946.

Berkes, I., Horváth, L., & Rice, G. 2013. Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications*, **123**, 385–403.

Billingsley, P. 1999. *Convergence of probability measures*. 2 edn. Wiley series in probability und mathematical statistics. New York: Wiley.

Borovkova, S., Burton, R., & Dehling, H. 2001. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Transactions of the American Mathematical Society*, **353**(11), 4261–4318.

Bosq, D. 2006. *Linear processes in function spaces*. New York: Springer.

Bradely, R.C. 2007. *Introduction to strong mixing conditions*. Vol. 1-3. Kendrick Press.

Bucchia, B., & Wendler, M. 2017. Change-point detection and bootstrap for Hilbert space valued random fields. *Journal of Multivariate Analysis*, **155**, 344–368.

Černíková, A., Hušková, M., Prášková, Z., & Steinebach, J. 2013. Delay time in monitoring jump changes in linear models. *Statistics*, **47**(1), 1–25.

Chen, Z., & Tian, Z. 2010. Modified procedures for change point monitoring in linear models. *Mathematics and computers in simulation*, **81**(1), 62–75.

Chow, Y.S., & Teicher, H. 1997. *Probability theory: independence, interchangeability, Martingales*. 3 edn. New York: Springer.

Chu, C.-S.J., Stinchcombe, M., & White, H. 1996. Monitoring structural change. *Econometrica*, **64**, 1045–1065.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., & Lindquist, M. A. 2012. Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage*, **61**, 907–920.

Cribben, I., Wager, T. D., & Lindquist, M. A. 2013. Detecting functional connectivity change points for single-subject fMRI data. *Frontiers in Computational Neuroscience*, 7–143.

Csörgő, M., & Horváth, L. 1988. Invariance principles for changepoint problems. *Journal of Multivariate Analysis*, **27**, 151–168.

Csörgő, M, & Révész, P. 1981. *Strong approximations in probability and statistics*.

Dehling, H., Sharipov, O.S., & Wendler, M. 2015a. Bootstrap for dependent Hilbert space-valued random variables with application to von Mises statistics. *Journal of Multivariate Analysis*, **133**, 200–215.

Dehling, H., Fried, R., & Wendler, M. 2015b. Change-point detection under dependence based on two-sample U-statistics. *Dawson, Kulik, Ould Haye, Szyszkowicz, Zhao (Eds.) Asymptotic Laws and Methods in Stochastics – A Volume in Honour of Miklos Csörgö, Springer, New York.*

Dehling, H., Fried, R., & Wendler, M. 2015c. A robust method for shift detection in time series. *arXiv preprint arXiv:1506.03345.*

Einmahl, U. 2009. A new strong invariance principle for sums of independent random vectors. *Journal of Mathematical Sciences*, **163**(4), 311–327.

Fazekas, I., & Klesov, O. 2000. A general approach to the strong law of large numbers. *Theory of Probability and its Applications*, **45**, 436–448.

Ferraty, F., & Vieu, P. 2000. *Nonparametric functional data analysis: theory and practice.* New York: Springer.

Fremdt, S. 2014. Asymptotic distribution of the delay time in Page's sequential procedure. *Journal of Statistical Planning and Inference*, **145**, 74–91.

Fremdt, S. 2015. Page's sequential procedure for change-point detection in time series regression. *Statistics*, **49**(1), 128–155.

Hájek, J, & Rényi, A. 1955. Generalization of an inequality of Kolmogorov. *Acta Mathematica Hungarica*, **6**(3-4), 281–283.

Hoeffding, W. 1948. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, **19**, 293–325.

Hörmann, S., & Kokoszka, P. 2010. Weakly dependent functional data. *The Annals of Statistics*, **38**(3), 1845–1884.

Horváth, L., & Kokoszka, P. 2012. *Inference for functional data with applications.* Springer Series in Statistics. New York: Springer.

Horváth, L., Hušková, M., Kokoszka, P., & Steinebach, J. 2004. Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, **126**, 225–251.

Horváth, L., Kühn, M., & Steinebach, J. 2008. On the performance of the fluctuation test for structural change. *Sequential Analysis*, **27**(2), 126–140.

Hušková, M., & Koubková, A. 2005. Monitoring jump changes in linear models. *Journal of Statistical Research*, **39**(2), 51–70.

Kirch, C. 2006. *Resampling methods for the change analysis of dependent data.* dissertation, Universität zu Köln.

Kirch, C., & Tadjuidje Kamgaing, J. 2015. On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference*, **161**, 25–49.

Kirch, C., & Weber, S. 2018. Modified sequential change point procedures based on estimating functions. *Electronic Journal of Statistics*, **12**(1), 1579–1613.

Korolyuk, V.S., & Borovskich, Y.V. 1994. *Theory of U-statistics*. Dordrecht: Kluwer Acad.

Lorden, Gary, *et al.* . 1971. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, **42**(6), 1897–1908.

McMurry, T, & Politis, D.N. 2011. Resampling methods for functional data. *The Oxford Handbook of Functional Data Analysis*, 189–209.

Móricz, F. 1976. Moment inequalities and the strong laws of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **35**, 299–314.

Ogawa, S., Lee, T.M., Kay, A.R., & Tank, D.W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA*, **87**, 9868–9872.

Page, E.S. 1954. Continuous inspection schemes. *Biometrika*, **41**, 100–115.

Page, E.S. 1955a. Control charts with warning lines. *Biometrika*, **42**, 243–257.

Page, E.S. 1955b. A test for a change in a parameter occurring at an unknown point. *Biometrika*, **43**, 523–527.

Paparoditis, E. 2018. Sieve bootstrap for functional time series. *The Annals of Statistics*, **46**(6B), 3510–3538.

Politis, D. N., & Romano, J. P. 1994. Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*, 461–476.

Politis, D.N. 2011. Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, **27**(4), 703–744.

Ranga Rao, R. 1962. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 659–680.

Siegmund, D., & Venkatraman, E.S. 1995. Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics*, 255–271.

Singh, K. 1981. On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, 1187–1195.

Tartakovsky, Alexander, Nikiforov, Igor, & Basseville, Michele. 2014. *Sequential analysis: Hypothesis testing and changepoint detection*. Chapman and Hall/CRC.

Van der Vaart, A.W. 2000. *Asymptotic statistics*. Vol. 3. Cambridge university press.

Volkonskii, V.A., & Rozanov, Y.A. 1959. Some limit theorems for random functions. I. *Theory of Probability & Its Applications*, **4**(2), 178–197.

Wald, A. 1947. *Sequential analysis*. New York: Wiley.

Weber, S. 2017. *Change-point procedures for multivariate dependent data*. Ph.D. thesis, Karlsruhe Institute of Technology (KIT). URN: urn:nbn:de:swb:90-689812.

Notation

Probability Theory

$a.s.$	almost surely
$i.i.d.$	independent and identically distributed
$\xrightarrow{\mathcal{D}}$	convergence in distribution
$\xrightarrow{D^{\mathfrak{d}}[0,1]}$	weak convergence in the $\mathfrak{d}$-dimensional Skorohod space $D^{\mathfrak{d}}[0,1]$
$\xrightarrow{P}$	convergence in probability
o_P, O_P	Landau symbolds (see Van der Vaart (2000))

Sequential Testing

m	length of the historic data set
k	monitoring time
k^*	change point
$\Gamma(m,k)$	monitoring statistic
$w(m,k)$	weight function
$\sup_{k\geq 1} w(m,k)\,\lvert\Gamma(m,k)\rvert$	test statistic
τ_m	stopping time
$c,\ c_\alpha,\ c_m$	critical value
h	kernel function
$h_1, h_2\ (h_1^*, h_{2,m}^*)$	functions of Hoeffding's decomposition under H_0 (after the change)
$r\ (r_m^*)$	remainder in Hoeffding's decomposition under H_0 (after the change)
$\theta\ (\theta_m^*)$	expected value of the kernel under H_0 (after the change)
Δ_m	change in the kernel function

Functional Data

n	sample size
$\mathcal{Z}$	compact set
$\int$	integral on $\mathcal{Z}$
$\mathcal{L}^2(\mathcal{Z})$	set of square integrable functions on $\mathcal{Z}$
$\lVert\cdot\rVert^2$	norm of $\mathcal{L}^2(\mathcal{Z})$
$<\cdot,\cdot>$	inner product of $\mathcal{L}^2(\mathcal{Z})$
$\lvert\cdot\rvert$	Euclidean norm
C_t	covariance operator
c_t	covariance kernel
$\hat{c}_n$	empirical covariance function
δ	change in the covariance kernel
Ω_n, Λ_n	test statistics
Σ_0	long-run covariance matrix under H_0