Change-Point Methods for Multivariate Autoregressive Models and Multiple Structural Breaks in the Mean

Zur Erlangung des akademischen Grades einer
DOKTORIN DER NATURWISSENSCHAFTEN
von der Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte
DISSERTATION
von
Dipl.-Math. Birte Chantal Simone Muhsal

Tag der mündlichen Prüfung: 08.05.2013
Referentin: JProf. Dr. Claudia Kirch
Korreferent: Prof. Dr. Jürgen Franke
Korreferent: Prof. Dr. Norbert Henze

Acknowledgements

My most sincere and deepest gratitude belongs to my supervisor JProf. Dr. Claudia Kirch for introducing me to the field of change-point analysis, giving valuable hints for my thesis, a constant readiness to discuss mathematical problems, and especially for a most pleasant working atmosphere.

I would like to thank Prof. Dr. Jürgen Franke and Prof. Dr. Norbert Henze for reviewing my thesis. Professor Henze not only found a number of typing errors but also provided valuable comments and help concerning my style of writing. I am deeply grateful to Prof. Dr. Hernando Ombao for a most welcoming and productive research stay at Brown University. I would also like to express my thanks to the whole Institute of Stochastics for a pleasant and inspiring working experience. My special thanks goes to my former and current office mates Julia Hörrmann, Jonathan Ott, Robin Pfeiffer, Anton Popp, Ines Türk and Sebastian Urban.

Further, I am deeply indebted to my family. Without their recognition, their support of my studies, and their encouragement throughout the years, writing this thesis would have been impossible. Finally, my warmest thanks goes to my favourite Frank.

This work was supported by DFG grant KI 1443/2-1. My research stay at the Brown University (Providence, USA) was funded by the Karlsruhe House of Young Scientists (KHYS).

Contents

Introduction **1**

I. Detection of Structural Breaks in Autoregressive Time Series **3**

1. Motivation and Outline . . . 5
2. Test and Estimation Procedure . . . 7
   - 2.1. Univariate Models and Methods . . . 7
     - 2.1.1. At-Most-One-Change and Epidemic Model . . . 7
     - 2.1.2. General Statistics and Null Asymptotics . . . 10
     - 2.1.3. Behaviour under Alternatives . . . 23
     - 2.1.4. Change-Point Estimators . . . 29
     - 2.1.5. Specific Test Statistics and Estimators . . . 31
   - 2.2. Multivariate Models and Methods . . . 34
     - 2.2.1. At-Most-One-Change and Epidemic Model . . . 34
     - 2.2.2. General Statistics and Null Asymptotics . . . 35
     - 2.2.3. Behaviour under Alternatives . . . 38
     - 2.2.4. Change-Point Estimators . . . 38
     - 2.2.5. Specific Test Statistics and Estimators . . . 39
3. Simulation Study . . . 41
   - 3.1. Specification of the Test Statistics . . . 41
   - 3.2. Bootstrap . . . 45
   - 3.3. Univariate Models . . . 46
   - 3.4. Multivariate Models . . . 54
4. Data Analysis . . . 57
   - 4.1. Data Description . . . 57
   - 4.2. Order of the Linear Autoregressive Model . . . 58
   - 4.3. Data Evaluation . . . 60

II. Change-Point Methods for Multiple Structural Breaks and Regime-Switching Models **65**

5. Basics . . . 67
   - 5.1. Literature . . . 68
     - 5.1.1. Multiple Change-Point Analysis . . . 68
     - 5.1.2. Regime-Switching Models . . . 69
   - 5.2. Models for Multiple Changes in Location . . . 71
     - 5.2.1. Classical Multiple Change-Point Location Model . . . 71
     - 5.2.2. Regime-Switching Location Model . . . 72
   - 5.3. Test and Estimation Procedure . . . 73
     - 5.3.1. Moving-Sum Statistic . . . 74
     - 5.3.2. Estimation of Number and Locations of Change Points . . . 78
     - 5.3.3. Estimation of the Long-Run Variance . . . 79

6. Theoretical Results 83
6.1. Consistency of the Estimators 83
6.1.1. Consistency of the Estimator for the Number of Change Points 83
6.1.2. Consistency of the Change-Point Estimators 90
6.1.3. Joint Asymptotic Distribution of the Change-Point Estimators 108
6.2. Asymptotic Results for the Long-Run Variance Estimator 112
6.2.1. Convergence of the Variance Estimator 112
6.2.2. Convergence of the Long-Run Variance Estimator 117
6.3. Examples for the Error Distribution 129
7. Simulation Study 135
7.1. Change Analysis of the Location Model with I.I.D. Errors 135
7.2. Change Analysis of the Location Model with Variance Varying Errors 137

III. Appendix **141**
A. Assumptions 143
B. Theorems from Probability Theory 145

Bibliography **150**

Introduction

In many fields of research time-related data is recorded: birth rates and population growth in demography, stock prices in the finance sector, water levels in geography, CPU loads in computer science, weather data in meteorology, just to mention a few. Scientists use this data to accumulate knowledge, understand interrelationships, explore influences, explain issues, or to discover new phenomena in order to understand our environment, cure deseases, improve techniques or develop new methods.

If the data is ordered in time, it is called a time series and the mathematical modelling and analysis of it is called time-series analysis. Often the question arises whether the data follows a stationary model or exhibits some kind of change(s). By ‘change’ we generally mean a change in the distribution and call it a structural break. The time points at which a structural break occures are called change points. In this work we introduce, explain and analyse statistical methods that are designed to detect and estimate change points. We conduct the step of detection by a statistical test whose construction depends on the model we assume and type of alternative we expect. The change-point estimators are based on these statistics.

The first part of this dissertation was realised in collaboration with Prof. Hernando Ombao (University of California at Irvine) and is motivated by a multi-channel electroencephalography (EEG) data set. EEG is a non-invasive measure of the brain’s electrical activity. These EEGs were recorded in a visual-motor experiment, where a participant is visually instructed to move a joystick either to the right or to the left. The reason to conduct this experiment is to understand the changes in the dynamics of brain processing during a visual-motor task to develop some mapping between EEGs and the intention to move. Eventually, these results can be used to direct the movement of a paralysed patient’s wheelchair or of a patient’s artifical limb.

We are interested in identifying changes in the brain activity during the course of trial. Each trial started in a relaxed state until at a known fixed time point a visual stimulus was given to move the joystick either to the left or right. The first goal is to identify the delay between stimulus shown and hand reaction. To this end we assume an at-most-one-change (AMOC) autoregressive model and consider maximum- and sum-type statistics based on weighted partial sums of residuals. These residuals are estimated by approximating the regression function by a linear regression function. We derive the null asymptotics of the test statistics and show the consistency of the tests under a large class of alternatives. Further, we prove that the change-point estimators are consistent in all cases where the test has asymptotic power one.

The results of the data analysis based on this model indicated that there is probably a second change point, which is related to the end of hand movement. Therefore, we considered additionally an epidemic change-point model, which separates the data into a relaxed state, a state containing the signal of movement, and again a relaxed state. For the epidemic change-point tests we adapted the AMOC test statistics and derived the corresponding asymptotic results.

We want to point out that in each trial the electrical activity was measured in 12 different channels and since we can expect the change(s) at least approximately at the same time point(s) we can pool the information of the data set by using multivariate statistics. Nevertheless, we first consider univariate test and estimation procedures and afterwards generalise them to multivariate methods. To analyse the performance of the introduced methods for small sample sizes we conduct a simulation study.

In the second part of this work we consider a change-point model which allows for several structural breaks. While a large amount of literature is available for AMOC models, less has been published for multiple change-point problems. Usually, computational intensive methods of model selection are proposed which further require to fix an upper bound for the number of structural breaks beforehand. The only well established test and estimation procedure based on hypothesis testing is the binary segmentation procedure of Vostrikova (1981). It tests and estimates the change points sequentially in a way that the overall significance level cannot be controlled.

A not so well-known method is proposed in Antoch et al. (2000). It tests for multiple structural breaks with an asymptotic significance level α and simultaneously estimates the number and locations of change points. It is based on moving-sum (MOSUM) statistics and has the advantages of being computationally fast and applicable to a broad range of models. As a start we consider in this work a location model which allows for changes in the expectation in an otherwise strictly stationary sequence of random variables. We use a maximum-type statistic based on moving-sums of estimated residuals introduced and analysed in the context of changes in the expectation in an otherwise independent and identically distributed random sequence by Hušková and Slabỳ (2001). We generalise their result concerning the asymptotic null distribution to the stationary case and show the consistency of the estimator for the number of structural breaks. Further, we obtain uniform rates of consistency for the change-point estimators and derive the joint asymptotic distribution. To obtain these results, we do not require that the number of changes is known or estimated correctly. Finally, we introduce and analyse variance and long-run variance estimators which are especially suitable for the multiple change-point problem.

The classical setup for change-point models assumes deterministic change points and a bounded number of them. Another more recent approach to model structural breaks are regime-switching models, which allow for random change points and a random, unbounded number of them for a growing number of observations. Since data generated by one of the two models looks the same, methods of classical change-point analysis will give reasonable results for data generated by a regime-switching model as well. But as the modeling of the data is very different, it is not at all obvious that the theoretical results will hold for regime-switching models. We analyse this problem in detail and prove consistency results for the regime-switching model as well. At the end we conduct a simulation study to analyse the small sample behaviour and discuss some issues arising in applications.

The third part is the Appendix and includes an overview of the most important assumptions in this work and some theorems from probability theory.

Part I.

Detection of Structural Breaks in Autoregressive Time Series

1. Motivation and Outline

We consider autoregressive change-point models with an AMOC or epidemic alternative and analyse test and estimation methods based on partial sums of weighted residuals. To obtain the residuals we approximate the observations by a linear autoregressive model of order p.

The case where the observations indeed follow a linear autoregressive model has already been considered in the literature. In this context Davis et al. (1995) derived and analysed the pseudo likelihood-ratio test statistic. Various versions of this statistic were introduced by Hušková et al. (2007). Based on their ideas we discuss a more general statistic, which includes their statistic as a special case. The advantage of our statistic is that we can take prior information about the type of change into account.

The above references considered AR(p) models whereas we use the linear autoregressive model to approximate general autoregressive processes. This has also been done in Davis et al. (2006) in the context of model selection. Moreover, Kirch and Kamgaing (2012) considered nonlinear time series as well, but they used neural networks for approximation. These methods are designed for univariate models whereas we further generalise our univariate methods to the multivariate situation.

Since we consider a very general method there is the natural question how to optimise it for our data analysis. We already give some explanations to this end throughout the following sections and hence introduce here the relevant information about the data set. A full description can be found in Section 4.1. To obtain the EEG data twelve electrodes were applied at different parts of the participants scalp. On a monitor, the participant was shown two types of stimuli: one was an arrow pointing to the left and the other an arrow pointing to the right. When a right (left) arrow was presented the participant was supposed to move the joystick to the right (left). This procedure was repeated $N = N_1 + N_2$ times, where $N_1 = 118$ corresponds to seeing the left arrow and $N_2 = 134$ corresponds to seeing the right arrow. Each movement was done correctly. The data set is split into two parts, the data for movements to the left and the data for movements to the right. Each trial took about 1 second, and the measurements took place at $T = 512$ time points over the 1-second interval. The stimulus was presented at $t = 250$ so that we do not expect a change before that. We use this fact in the following analysis by choosing an appropriate weight function. Moreover, we know from experience that a change is to be expected in the first two autocorrelations. This information is included in the specification of the general test statistic.

The next chapter introduces the AMOC and the epidemic autoregressive model and presents the test and estimation methods. We begin in Section 2.1 with the univariate models and methods and generalise them in Section 2.2 to the multivariate case. Thereby, both sections have the same structure. First, we explain the models. Subsequently we introduce the test statistics, derive the asymptotic null distributions and further show the consistency of the test under a large class of alternatives. That follows the introduction of the change-point estimator(s) and the proof of their consistency. Then, we discuss specific test statistics and estimators which are suitable for our data analysis.

In Chapter 3 we present the results of a simulation study to analyse the behaviour of the test statistics for smaller sample sizes. We further introduce and investigate the performance of a bootstrap procedure. Finally, Chapter 4 includes the data analysis. It begins with a description of the data and discusses the order of the linear autoregressive model we use for approximation. At the end it presents the results of the data analysis.

2. Test and Estimation Procedure

This chapter presents the models and theoretical results. Section 2.1 starts with the univariate case. We introduce the models and test statistics, derive the asymptotic null distributions and further show the consistency of the test and the change-point estimator(s) under a large class of alternatives. Moreover, we discuss specific test statistics and estimators which are suitable for our data analysis. Section 2.2 explains the correspondent multivariate models and presents the proofs for the multivariate situation.

2.1. Univariate Models and Methods

We consider univariate AMOC and epidemic autoregressive models. While under the alternative of the AMOC model only one change point is present, the epidemic alternative involves a second change point, where the process changes back to the original regime. We discuss these models in Section 2.1.1 in detail. Section 2.1.2 deals with the test statistics and their asymptotic null distributions. Section 2.1.3 shows that the tests are consistent under a large class of alternatives. Afterwards, in Section 2.1.4 we introduce the change-point estimator(s) and prove their consistency. Section 2.2.5 concludes with a discussion of specific test statistics suitable for our data analysis.

2.1.1. At-Most-One-Change and Epidemic Model

We consider a time-series model with observed time points $i = 1, \ldots, n$, and a change after an unknown time point $\tilde{k}$:

$$Y_i = \begin{cases} Y_i^{(1)}, & 1 \leq i \leq \tilde{k}, \\ Y_i^{(2)}, & \tilde{k} < i \leq n, \end{cases} \tag{2.1}$$

where $1 < \tilde{k} \leq n$ depends on the number of observations n, i.e. $\tilde{k} = \lfloor \lambda n \rfloor$ with $0 < \lambda \leq 1$. We assume that $\{Y_i^{(1)}\}$ and $\{Y_i^{(2)}\}$ are stationary processes and differ in distribution. Thereby we have in mind that the observations $Y_1, \ldots, Y_n$ follow some autoregressive model

$$Y_i = \begin{cases} g(Y_{i-1}, \ldots) + \xi_i, & 1 \leq i \leq \tilde{k}, \\ h(Y_{i-1}, \ldots) + \xi_i, & \tilde{k} < i \leq n, \end{cases} \tag{2.2}$$

with regression functions g and h, and i.i.d. errors $\xi_1, \ldots, \xi_n$. In this model the process after the change is not stationary since it has starting values from the stationary distribution of the time series before the change point. But stationarity in model (2.1) is assumed to avoid further technical difficulties under the alternative, though the proofs also hold for models as in (2.2) under certain conditions. Refer for more details to Remark 2.2.

The problem is to test the null hypothesis H_0 of no change against the alternative H_1 of one structural break, i.e.

$$H_0: \ \tilde{k} = n \qquad \text{vs.} \qquad H_1: \ \tilde{k} < n.$$

The test statistics will be based on estimated residuals. To obtain these, we use the fact that a large class of processes can already be well approximated by a linear autoregressive model such as

$$Y_i \;=\; \begin{cases} \mathbb{Y}_i^{(1)\top}\mathbf{a}_1 + \varepsilon_i, & 1 < i \le \tilde{k}, \\ \mathbb{Y}_i^{(2)\top}\mathbf{a}_2 + \varepsilon_i, & \tilde{k} < i \le n, \end{cases} \tag{2.3}$$

with i.i.d. errors $\varepsilon_1, \ldots, \varepsilon_n$,

$$\mathbb{Y}_i = (Y_{i-1}, \ldots, Y_{i-p})^\top, \qquad i = 1, \ldots, n,$$

and parameters

$$\mathbf{a}_1 = \left(a_1^{(1)}, \ldots, a_p^{(1)}\right)^\top, \qquad \mathbf{a}_2 = \left(a_1^{(2)}, \ldots, a_p^{(2)}\right)^\top,$$

such that the roots of the characteristic polynomials $1 - a_1^{(1)}t - \ldots - a_p^{(1)}t^p$ and $1 - a_1^{(2)}t - \ldots - a_p^{(2)}t^p$ are outside the unit circle. This ensures the causality of the two processes $\{Y_i^{(1)} : 1 \le i \le n\}$ and $\{Y_i^{(2)} : 1 \le i \le n\}$.

We approximate the observations in (2.1) by a linear autoregressive model such that

$$Y_i \;=\; \begin{cases} \mathbb{Y}_i^{(1)\top}\mathbf{a}_1 + \varepsilon_i^{(1)}, & 1 < i \le \tilde{k}, \\ \mathbb{Y}_i^{(2)\top}\mathbf{a}_2 + \varepsilon_i^{(2)}, & \tilde{k} < i \le n, \end{cases}$$

with errors

$$\begin{aligned} \varepsilon_i^{(1)} \;&:=\; Y_i^{(1)} - \mathbb{Y}_i^{(1)\top}\mathbf{a}_1 \qquad \text{for } i = 1, \ldots, \tilde{k}, \\ \varepsilon_i^{(2)} \;&:=\; Y_i^{(2)} - \mathbb{Y}_i^{(2)\top}\mathbf{a}_2 \qquad \text{for } i = \tilde{k}+1, \ldots, n, \end{aligned} \tag{2.4}$$

and parameters

$$\mathbf{a}_1 := \mathbf{C}_1^{-1}\boldsymbol{\gamma}_1 \quad \text{and} \quad \mathbf{a}_2 := \mathbf{C}_2^{-1}\boldsymbol{\gamma}_2 \tag{2.5}$$

where $\mathbf{C}_1$, $\mathbf{C}_2$ are the autocovariance matrices of $\{Y_i^{(1)}\}$, $\{Y_i^{(2)}\}$, i.e.

$$\mathbf{C}_1 = E\left[\mathbb{Y}_i^{(1)}\mathbb{Y}_i^{(1)\top}\right] \quad \text{and} \quad \mathbf{C}_2 = E\left[\mathbb{Y}_i^{(2)}\mathbb{Y}_i^{(2)\top}\right],$$

which are assumed to be invertible. Moreover, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ are the vectors of the first p autocovariances of $\{Y_i^{(1)}\}$, $\{Y_i^{(2)}\}$, i.e.

$$\boldsymbol{\gamma}_1 = E\left[\mathbb{Y}_i^{(1)}Y_i^{(1)}\right] \quad \text{and} \quad \boldsymbol{\gamma}_2 = E\left[\mathbb{Y}_i^{(2)}Y_i^{(2)}\right].$$

In the sequel we call time series correctly specified if they follow indeed a linear autoregressive model of order p as in (2.3). Otherwise, we call them misspecified. In the latter case we have dependent error sequences $\{\varepsilon_i^{(1)} : 1 \le i \le n\}$ and $\{\varepsilon_i^{(2)} : 1 \le i \le n\}$.

The parameters $\mathbf{a}_1$ and $\mathbf{a}_2$ as in (2.5) are best approximating in the sense that

$$\mathbf{a}_1 = \arg\min_{\mathbf{a}} E\left[\left(Y_i^{(1)} - \mathbb{Y}_i^{(1)\top}\mathbf{a}\right)^2\right] \quad \text{and} \quad \mathbf{a}_2 = \arg\min_{\mathbf{a}} E\left[\left(Y_i^{(2)} - \mathbb{Y}_i^{(2)\top}\mathbf{a}\right)^2\right].$$

This can be seen by setting the first derivative of the expected values in (2.5), given by

$$\begin{aligned} E\left[-2\left(Y_i^{(1)} - \mathbb{Y}_i^{(1)\top}\mathbf{a}\right)\mathbb{Y}_i^{(1)}\right] &= -2E\left[Y_i^{(1)}\mathbb{Y}_i^{(1)} - \mathbb{Y}_i^{(1)\top}\mathbb{Y}_i^{(1)}\mathbf{a}\right] &= -2\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\mathbf{a}\right), \\ E\left[-2\left(Y_i^{(2)} - \mathbb{Y}_i^{(2)\top}\mathbf{a}\right)\mathbb{Y}_i^{(2)}\right] &= -2E\left[Y_i^{(2)}\mathbb{Y}_i^{(2)} - \mathbb{Y}_i^{(2)\top}\mathbb{Y}_i^{(2)}\mathbf{a}\right] &= -2\left(\boldsymbol{\gamma}_2 - \mathbf{C}_2\mathbf{a}\right), \end{aligned}$$

equal to zero. Then, we obtain the Yule-Walker equations in (2.5).

Further, we define the parameter $\widetilde{\mathbf{a}}$ by

$$\widetilde{\mathbf{a}} := \mathbf{Q}^{-1}\boldsymbol{\gamma} \tag{2.6}$$

with

$$\mathbf{Q} = \lambda\mathbf{C}_1 + (1-\lambda)\mathbf{C}_2 \quad \text{and} \quad \boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}_1 + (1-\lambda)\boldsymbol{\gamma}_2$$

and $\mathbf{Q}$ is assumed to be invertible. We can show that

$$\widetilde{\mathbf{a}} = \arg\min_{\mathbf{a}} E_{\mathbf{a}}$$

with

$$E_{\mathbf{a}} = \lambda E\left[\left(Y_i^{(1)} - \mathbb{Y}_i^{(1)\top}\mathbf{a}\right)^2\right] + (1-\lambda)E\left[\left(Y_i^{(2)} - \mathbb{Y}_i^{(2)\top}\mathbf{a}\right)^2\right].$$

The first derivative of $E_{\mathbf{a}}$ is

$$\begin{aligned} -2\left(\lambda\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right) + (1-\lambda)\left(\boldsymbol{\gamma}_2 - \mathbf{C}_2\widetilde{\mathbf{a}}\right)\right) &= -2\left(\left(\lambda\boldsymbol{\gamma}_1 + (1-\lambda)\boldsymbol{\gamma}_2\right) - \left(\lambda\mathbf{C}_1 + (1-\lambda)\mathbf{C}_2\right)\widetilde{\mathbf{a}}\right) \\ &= -2\left(\boldsymbol{\gamma} - \mathbf{Q}\widetilde{\mathbf{a}}\right) \end{aligned}$$

and equals zero for $\widetilde{\mathbf{a}}$ as in (2.6). Under the null hypothesis it holds $\mathbf{Q} = \mathbf{C}_1$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1$ and $\widetilde{\mathbf{a}} = \mathbf{a}_1$.

Under the epidemic model we have a second change point at which the model changes back to the original regime. This model is given by

$$Y_i \;=\; \begin{cases} Y_i^{(1)}, & 1 \le i \le \tilde{k}_1, \\ Y_i^{(2)}, & \tilde{k}_1 < i \le \tilde{k}_2, \\ Y_i^{(1)}, & \tilde{k}_2 < i \le n, \end{cases}$$

with $1 < \tilde{k}_1 = \lfloor \lambda_1 n \rfloor \le \tilde{k}_2 = \lfloor \lambda_2 n \rfloor \le n$ and $0 < \lambda_1 \le \lambda_2 \le 1$. We follow the same ideas as under the AMOC model and obtain the same results, but here we define

$$E_{\mathbf{a}} = (1-(\lambda_2-\lambda_1))E\left[\left(Y_i^{(1)} - \mathbb{Y}_i^{(1)\top}\mathbf{a}\right)^2\right] + (\lambda_2-\lambda_1)E\left[\left(Y_i^{(2)} - \mathbb{Y}_i^{(2)\top}\mathbf{a}\right)^2\right],$$

and

$$\mathbf{Q} = (1-(\lambda_2-\lambda_1))\,\mathbf{C}_1 + (\lambda_2-\lambda_1)\,\mathbf{C}_2 \quad \text{and} \quad \boldsymbol{\gamma} = (1-(\lambda_2-\lambda_1))\lambda\boldsymbol{\gamma}_1 + (\lambda_2-\lambda_1)\boldsymbol{\gamma}_2.$$

2.1.2. General Statistics and Null Asymptotics

We assume that $\{Y_i^{(1)}\}$ and $\{Y_i^{(2)}\}$ fulfill a strong mixing condition given in **(Y)**. In this section we consider the univariate case $d = 1$, but in view of the multivariate situation in Section 2.2, we prefer the following multivariate formulation of the assumption:

(Y) Let $\{\mathbf{Y}_i\}$ be a $\mathbb{R}^d$-valued strictly stationary sequence of random vectors with

$$E[\mathbf{Y}_1] = 0, \qquad E\|\mathbf{Y}_1\|^{4+\nu} < \infty \text{ for some } \nu > 0,$$

satisfying a strong mixing condition with mixing rate

$$\alpha(n) = O\left(n^{-\beta}\right) \qquad \text{for some } \beta > \max(3, (4+\nu)/\nu).$$

Strong mixing is a classical assumption in time series analysis. Recently other concepts of dependency, such as weak dependency (refer to Doukhan and Louhichi (1999)), have been introduced to avoid some drawbacks of mixing.

However, for two reasons we assume a strong mixing condition. First, all results we need are available in the literature. These results include among others a central limit theorem, which can be obtained by an invariance principle of Kuelbs and Philipp (1980), or a Hájek-Rényi-type inequality. Secondly, these results have to hold for certain functionals of $\{Y_i^{(1)}\}$ and $\{Y_i^{(2)}\}$. Under the mixing condition these are easily obtained, since the mixing condition carries over to measurable functions of mixing processes. Our results remain true under different dependency concepts if certain assumptions hold. We discuss this issue in more detail in Remark 2.2.

To calculate the residuals for the test statistic, we use the least squares estimator $\widehat{\mathbf{a}}_n$, which minimises

$$L_n(\mathbf{a}) := \sum_{i=1}^{n} (Y_i - \mathbb{Y}_i^\top \mathbf{a})^2$$

and is given by

$$\widehat{\mathbf{a}}_n = \mathbf{C}_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}_i Y_i \qquad \text{with} \qquad \mathbf{C}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top .$$

The next theorem shows that $\widehat{\mathbf{a}}_n$ is strongly consistent with respect to the parameter $\widetilde{\mathbf{a}}$ and converges stochastically with rate $O_p\left(n^{-\frac{1}{2}}\right)$ under H_0 as well as under H_1.

Theorem 2.1. Let $\{Y_i^{(1)}\}$ and $\{Y_i^{(2)}\}$ fulfill **(Y)**. Then, under H_0, and both alternatives, it holds

(a) $\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}} \; = \; o_p(1)$,

(b) $\sqrt{n}\,(\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) \; = \; O_p(1)$.

Proof of Theorem 2.1. Under the null hypothesis part (a) follows directly from the strong law of large numbers implied by the uniform strong law of large numbers in Theorem 6 in Rao (1962) (refer to Theorem B.1 in the Appendix), since $\{Y_i^2\}$, $\{Y_i Y_{i-1}\}, \ldots, \{Y_i Y_{i-p}\}$ are strongly mixing and hence ergodic. Under the AMOC alternative we have

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}_i Y_i \;&=\; \frac{\lfloor \lambda n \rfloor}{n} \frac{1}{\lfloor \lambda n \rfloor} \sum_{i=1}^{\lfloor \lambda n \rfloor} \mathbb{Y}_i^{(1)} Y_i^{(1)} + \frac{p}{n} \frac{1}{p} \sum_{i=\lfloor \lambda n \rfloor+1}^{\lfloor \lambda n \rfloor+p} \mathbb{Y}_i Y_i^{(2)} \\
&\quad + \frac{n - \lfloor \lambda n \rfloor - p}{n} \frac{1}{n - \lfloor \lambda n \rfloor - p} \sum_{i=\lfloor \lambda n \rfloor+p+1}^{n} \mathbb{Y}_i^{(2)} Y_i^{(2)} \\
&=\; \lambda \boldsymbol{\gamma}_1 + (1-\lambda)\boldsymbol{\gamma}_2 + o_p(1),
\end{aligned}$$

since the second sum is asymptotically neglible and $Y_i^{(2)}$ is by assumption, as $Y_i^{(1)}$, a stationary, strongly mixing processes and hence the strong law of large numbers can be applied. Analogously we obtain

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \;=\; \lambda \mathbf{C}_1 + (1-\lambda)\mathbf{C}_2 + o_p(1) \;=\; \mathbf{Q} + o_p(1).$$

In case of the epidemic alternative we separate the sums at the time points $\lfloor \lambda_1 n \rfloor$, $\lfloor \lambda_1 n \rfloor + p$, $\lfloor \lambda_2 n \rfloor$ and $\lfloor \lambda_2 n \rfloor + p$ and receive the assertion with the same arguments.

Part (b) follows by the central limit theorem, which is induced by the invariance principle of Kuelbs and Philipp (1980) (refer to Theorem B.2 in the Appendix). Under the null hypothesis it holds

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \boldsymbol{\gamma}_1\right) \;=\; O_p(1) \qquad \text{and} \qquad \sqrt{n}\left(\mathbf{C}_n - \mathbf{C}_1\right) = O_p(1).$$

Under the alternatives we split the sums at the change points and apply the central limit theorem separately. This implies

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \boldsymbol{\gamma}\right) \;=\; O_p(1) \qquad \text{and} \qquad \sqrt{n}\left(\mathbf{C}_n - \mathbf{Q}\right) = O_p(1).$$

Then, it follows

$$\begin{aligned}
&\sqrt{n}\left(\mathbf{C}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \mathbf{Q}^{-1}\boldsymbol{\gamma}\right)\\
&= \sqrt{n}\left(\mathbf{C}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \mathbf{Q}^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \left(\mathbf{Q}^{-1}\boldsymbol{\gamma} - \mathbf{Q}^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i\right)\right)\\
&= \sqrt{n}\left(\mathbf{C}_n^{-1}\mathbf{Q}\mathbf{Q}^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \mathbf{C}_n^{-1}\mathbf{C}_n\mathbf{Q}^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \mathbf{Q}^{-1}\left(\boldsymbol{\gamma} - \frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i\right)\right)\\
&= \sqrt{n}\left(\mathbf{C}_n^{-1}\left(\mathbf{Q} - \mathbf{C}_n\right)\mathbf{Q}^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i + \mathbf{Q}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i Y_i - \boldsymbol{\gamma}\right)\right) \;=\; O_p(1).
\end{aligned}$$

This is the assertion. □

The results of Theorem 2.1 were obtained under condition **(Y)** which assumes strict stationarity and a mixing condition for $\{Y_i^{(1)}\}$ and $\{Y_i^{(2)}\}$. We mentioned above that the assumption of strict stationarity is a simplifying assumption, which does not hold under models as in (2.2), since the process $\{Y_i^{(2)}\}$ is started with values from $\{Y_i^{(1)}\}$. Remark 2.2 points out which conditions have to be fulfilled to let the results under these models and other dependency concepts than mixing hold.

Remark 2.2. To prove Theorem 2.1 certain functionals of the process $\{Y_i^{(1)}\}$, such as $\{Y_i^{(1)}Y_i^{(1)}\}$, $\{Y_i^{(1)}Y_{i-1}^{(1)}\}, \ldots, \{Y_i^{(1)}Y_{i-p}^{(1)}\}$, need to fulfill a strong law of large numbers and a central limit theorem. If $\{Y_i^{(1)}\}$ is mixing, these functionals are also mixing with the same rate which implies both the law of large numbers as well as the central limit theorem. If other dependency concepts are used, the validity of these laws need to be checked.

Under the alternative we have to verify the strong law of large numbers and the central limit theorem for the process $\{Y_i^{(2)}Y_i^{(2)}\}$, $\{Y_i^{(2)}Y_{i-1}^{(2)}\}, \ldots, \{Y_i^{(2)}Y_{i-p}^{(2)}\}$ as well. We used here the stationarity of the process, but the stationarity is not needed as long as the law of large numbers and the central limit theorem hold. Concerning the validity of central limit theorems and strong laws of large numbers for time series which do not have starting values obtained from the stationary distribution can be found in Meyn et al. (2009), chapter 17, and Jensen and Rahbek (2007). In case the time series follows a linear autoregressive model, Hušková et al. (2007) derive an explicit formula for the difference of a time series started with values from the the stationary distribution and started with values from the process $\{Y_i^{(1)}\}$. This formula shows that the starting values are irrelevant for the validity of the law of large numbers and the central limit theorem. ■

For the AMOC situation we consider the test statistics

$$T_n^{(1)} = \max_{1 \leq k \leqslant n} \frac{w^2(k/n)}{n} \mathbf{S}_k^\top \mathbf{A} \mathbf{S}_k,$$
$$T_n^{(2)} = \frac{1}{n} \sum_{k=1}^{n} \frac{w^2(k/n)}{n} \mathbf{S}_k^\top \mathbf{A} \mathbf{S}_k$$

with

$$\mathbf{S}_k = \sum_{i=1}^{k} \mathbb{Y}_i \hat{\varepsilon}_i, \qquad \hat{\varepsilon}_i = Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_n$$

and a suitable symmetric, positive semidefinite $p \times p$ matrix $\mathbf{A}$, which can also be replaced by a consistent estimator. These test statistics follow an idea of Hušková et al. (2007), who considered $T_n^{(1)}$ in case of a correctly specified model with $\mathbf{A}$ replaced by $\mathbf{C}_n^{-1}/\hat{\sigma}^2$, which is a suitable estimator for the inverse of the asymptotic variance under H_0. Section 2.1.5 discusses the matrix $\mathbf{A}$ and its estimation in more detail. Hušková et al. (2007) obtained the statistic as a variation of the maximum likelihood statistic introduced by Davis et al. (1995).

For the weight function $w(\cdot)$ we assume:

(W) Let the weight function $w : [0,1] \to \mathbb{R}$ be a left continuous function with existing right limits and a finite number of disconuities $a_1, \ldots, a_K$ and fulfill the regularity conditions

$$\lim_{t \to 0} t^\alpha w(t) < \infty, \quad \lim_{t \to 1} (1-t)^\alpha w(t) < \infty, \qquad \text{for some } 0 \leqslant \alpha < 1/2,$$
$$\sup_{\eta \leqslant t \leqslant 1-\eta} w(t) < \infty, \qquad \text{for all } 0 < \eta \leqslant \frac{1}{2}.$$

Remark 2.3. Assumption **(W)** does not include the continuity of the weight function $w(\cdot)$, but assumes it is left continuous with existing right limits and a finite number of disconuities. We can equivalently assume that $w(\cdot)$ is a right continuous function with existing left limits and a finite number of disconuities or further that $w(\cdot)$ has a finite number of disconuities $a_1, \ldots, a_K$, existing left and right limits and $w(a_j) \leq \max(w(a_j^-), w(a_j^+))$, $j = 1, \ldots, K$. These considerations are relevant in the proof of Theorem 2.6, which gives the asymptotic distributions of the test statistics (refer to equation (2.12)). It is reasonable to use weight functions with discontinuties in situations where the change point is to be expected in certain regions. Regions of no interest are weighted with zero. ■

Typical symmetric weight functions are given by

$w_1(t) = (t(1-t))^{-1/2} \mathcal{I}_{\varepsilon < t < (1-\varepsilon)}(t) \quad$ for some $\varepsilon < 1/2$,
$w_2(t) = (t(1-t))^{-\gamma} \quad$ for some $0 \leqslant \gamma < 1/2$.

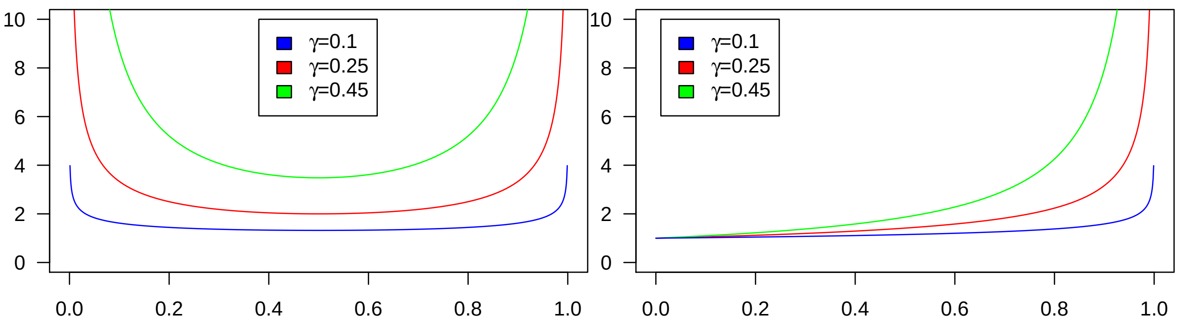

Figure 2.1.: Weight functions $w(t) = (t(1-t))^{-\gamma}$ and $w(t) = (1-t)^{-\gamma}$

Both are closely related to the weight function $w(t) = (t(1-t))^{-1/2}$, which is obtained for the Maximum Likelihood (ML) statistic. However, this weight function does not fulfill assumption **(W)**. Distributional convergence for the test statistics with the ML weight function can be obtained by means of extreme-value behaviour but the convergence is usually rather slow and hence making use of the above related weight functions is more attractive in many practical situations.

The weight function essentially determines where we look closest for a change, e.g. a γ close to $1/2$ will find early or late changes more easily while a γ close to 0 prefers changes in the middle of the time series. This is illustrated in the first diagram of Figure 2.1. In some situations it can also make sense to use non-symmetric weight function, e.g. if one has priori information about the location of the change. In our data example this will be the case as we expect changes only to be in the second half of the data. An example for a non-symmetric weight function is

$$w_3(t) = (1-t)^{-\gamma} \quad \text{for some } 0 \leqslant \gamma < 1/2, \tag{2.7}$$

which is illustrated in the second diagram of Figure 2.1.

Before we derive the asymptotic distributions of the test statistics, the next lemma introduces a functional central limit theorem, which is essential for the proofs of the asymptotic results following in the sequel.

Lemma 2.4. Let $\{Y_i\}$ be a process fulfilling **(Y)** and $\{\varepsilon_i\}$ is defined as in (2.4). Then,

$$(\mathbf{\Sigma} n)^{-\frac{1}{2}} \sum_{i=p+1}^{\lfloor nt \rfloor} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \quad \overset{\mathcal{D}[0,1]}{\longrightarrow} \quad \mathbf{W}(t),$$

where $\{\mathbf{W}(t) = (W_1(t), \ldots, W_p(t))^\top : t \in [0,1]\}$ denotes a p-dimensional standard Wiener process and $\mathbf{\Sigma}$ the long-run autocovariance matrix of $\{\mathbb{Y}_i \varepsilon_i\}$.

Proof of Lemma 2.4. By assumption $\{Y_i\}$ is strongly mixing and by definition of strong mixing so is $\{\mathbb{Y}_i \varepsilon_i\}$, since $\{\varepsilon_i\}$ is defined as a function of $\{Y_i\}$ (refer to (2.4)). Hence the functional central limit theorem follows by the invariance principle of Kuelbs and Philipp (1980). □

Remark 2.5. If the model is correctly specified, we observe a causal AR(p) process with i.i.d. errors $\varepsilon_1, \ldots, \varepsilon_n$ and $\mathbb{Y}_i$ and ε_i are independent, so that we receive under H_0

$$\begin{aligned}
\mathbf{\Sigma} &= \lim_{n\to\infty} \frac{1}{n} E\left[\sum_{i=1}^{n} (\mathbb{Y}_i\varepsilon_i - E[\mathbb{Y}_i]E[\varepsilon_i]) \sum_{j=1}^{n} (\mathbb{Y}_j\varepsilon_j - E[\mathbb{Y}_j]E[\varepsilon_j])^\top\right] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left[\mathbb{Y}_i\mathbb{Y}_i^\top\right] E\left[\varepsilon_i\varepsilon_i\right] + \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{i-1} E\left[\mathbb{Y}_i\mathbb{Y}_j^\top \varepsilon_j\right] E\left[\varepsilon_i\right] \\
&\quad + \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n}\sum_{j=i+1}^{n} E\left[\mathbb{Y}_i\mathbb{Y}_j^\top \varepsilon_i\right] E\left[\varepsilon_j\right] \\
&= \mathbf{C}_1\sigma_1^2.
\end{aligned}$$

■

Theorem 2.6. Let the process $\{Y_i\}$ fulfill **(Y)** and $\{\varepsilon_i\}$ is defined as in (2.4). Moreover let for the weight function **(W)** hold. Then, under H_0,

$$\text{(a)}\quad T_n^{(1)} = \max_{1\le k\le n} \frac{w^2(k/n)}{n}\mathbf{S}_k^\top\mathbf{A}\mathbf{S}_k \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} w^2(t)\sum_{j=1}^{p} B_j^2(t),$$

$$\text{(b)}\quad T_n^{(2)} = \frac{1}{n}\sum_{1\leqslant k\leqslant n} \frac{w^2(k/n)}{n}\mathbf{S}_k^\top\mathbf{A}\mathbf{S}_k \xrightarrow{\mathcal{D}} \int_0^1 w^2(t)\sum_{j=1}^{p} B_j^2(t)\,dt,$$

where $\{\mathbf{B}(t) = (B_1(t), \ldots, B_p(t))^\top : t \in [0,1]\}$ denotes a p-dimensional Brownian bridge with covariance matrix $\mathbf{A}^{\frac{1}{2}}\mathbf{\Sigma}\mathbf{A}^{\frac{1}{2}}$. $\mathbf{\Sigma}$ is the long-run autocovariance matrix of $\{\mathbb{Y}_i\varepsilon_i\}$ and $\mathbf{A}$ is a symmetric, positive semidefinite matrix which can be replaced by a consistent estimator.

Proof of Theorem 2.6. We use the Euclidean norm $\|\cdot\|$ to obtain

$$\max_{1\le k\le n} \frac{w^2(k/n)}{n}\mathbf{S}_k^\top\mathbf{A}\mathbf{S}_k = \max_{1\le k\le n} \frac{w^2(k/n)}{n}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{S}_k\right)^\top \mathbf{A}^{\frac{1}{2}}\mathbf{S}_k = \max_{1\le k\le n} \frac{w^2(k/n)}{n}\left\|\mathbf{A}^{\frac{1}{2}}\mathbf{S}_k\right\|^2.$$

With the definition of the residuals we receive

$$\mathbb{Y}_i\hat{\varepsilon}_i = \mathbb{Y}_i\left(\hat{\varepsilon}_i - \varepsilon_i + \varepsilon_i\right) = \mathbb{Y}_i\varepsilon_i - \mathbb{Y}_i\left(\mathbb{Y}_i^\top\widehat{\mathbf{a}}_n - \mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right) = \mathbb{Y}_i\varepsilon_i - \mathbb{Y}_i\mathbb{Y}_i^\top\left(\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}\right), \tag{2.8}$$

where $\widetilde{\mathbf{a}} = \mathbf{a}_1$ is defined in (2.6). Further, we obtain by the definition of $\widehat{\mathbf{a}}_n$

$$\mathbf{S}_n = \sum_{i=1}^{n}\mathbb{Y}_i\hat{\varepsilon}_i = \sum_{i=1}^{n}\mathbb{Y}_i(Y_i - \mathbb{Y}_i^\top\widehat{\mathbf{a}}_n) = \sum_{i=1}^{n}\mathbb{Y}_iY_i - n\mathbf{C}_n\widehat{\mathbf{a}}_n = 0.$$

Then, it follows

$$
\begin{aligned}
\mathbf{S}_k &= \sum_{i=1}^{k} \left(\mathbb{Y}_i \hat{\varepsilon}_i - n^{-1} \mathbf{S}_n \right) = \sum_{i=1}^{k} \left(\mathbb{Y}_i \hat{\varepsilon}_i - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \hat{\varepsilon}_i \right) \\
&= \sum_{i=1}^{k} \left(\mathbb{Y}_i \varepsilon_i - \mathbb{Y}_i \mathbb{Y}_i^\top (\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) - n^{-1} \sum_{i=1}^{n} \left(\mathbb{Y}_i \varepsilon_i - \mathbb{Y}_i \mathbb{Y}_i^\top (\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) \right) \right) \\
&= \sum_{i=1}^{k} \left(\mathbb{Y}_i \varepsilon_i - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \varepsilon_i \right) - \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top (\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top (\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) \right) \\
&= \sum_{i=1}^{k} \left(\mathbb{Y}_i \varepsilon_i - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \varepsilon_i \right) - \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right) (\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) . \qquad (2.9)
\end{aligned}
$$

Next, we show

$$
\begin{aligned}
& \max_{1 \le k \le n} \frac{w^2(k/n)}{n} \left\| \mathbf{A}^{\frac{1}{2}} \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right) (\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) \right\|^2 \\
&\le \left\| \mathbf{A}^{\frac{1}{2}} \right\|^2 \max_{1 \le k \le n} \frac{w^2(k/n)}{n} \left\| \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right) \right\|^2 \|(\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}})\|^2 = o_p(1).
\end{aligned}
$$

By Theorem 2.1 part (b) we know

$$
\|(\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}})\|^2 = O_p\left(n^{-1}\right).
$$

By adding and subtrating the covariance matrix $\mathbf{C}_1$ we receive

$$
\begin{aligned}
\sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right) &= \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - \mathbf{C}_1 - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top + \mathbf{C}_1 \right) \\
&= \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - \mathbf{C}_1 \right) - \frac{k}{n} \sum_{i=1}^{n} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - \mathbf{C}_1 \right) \qquad (2.10)
\end{aligned}
$$

and obtain with $\varepsilon < 1/2$

$$
\begin{aligned}
& \max_{\varepsilon n \le k \le (1-\varepsilon) n} w^2(k/n) \frac{1}{n^2} \left\| \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - \mathbf{C}_1 \right) \right\|^2 \\
&\le \sup_{\varepsilon \le t \le (1-\varepsilon)} w^2(t) \max_{\varepsilon n \le k \le (1-\varepsilon) n} \left\| \frac{1}{n} \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - \mathbf{C}_1 \right) \right\|^2 \\
&= o(1) \text{ a.s.,}
\end{aligned}
$$

since the supremum is bounded by assumption **(W)** and the maximum converges a.s. to zero by the strong law of large numbers. Moreover, we obtain with $0 \leq \alpha < 1/2$

$$
\begin{aligned}
&\max_{1\leq k\leq \varepsilon n} w^2(k/n)\frac{1}{n^2}\left\|\sum_{i=1}^{k}\left(\mathbb{Y}_i\mathbb{Y}_i^\top - \mathbf{C}_1\right)\right\|^2 \\
\leq\ & \max_{1\leq k\leq \varepsilon n} w^2(k/n)\left(\frac{k}{n}\right)^{2\alpha}\max_{1\leq k\leq \varepsilon n}\left(\frac{n}{k}\right)^{2\alpha}\frac{1}{n^2}\left\|\sum_{i=1}^{k}\left(\mathbb{Y}_i\mathbb{Y}_i^\top - \mathbf{C}_1\right)\right\|^2 \\
\leq\ & \left(\sup_{0\leq t\leq \varepsilon} w(t)t^\alpha\right)^2 n^{2\alpha-2}\max_{1\leq k\leq \varepsilon n}\left\|\sum_{i=1}^{k}\left(\mathbb{Y}_i\mathbb{Y}_i^\top - \mathbf{C}_1\right)\right\|^2 \\
\leq\ & \left(\sup_{0\leq t\leq \varepsilon} w(t)t^\alpha\right)^2 n^{2\alpha-1}\max_{1\leq k\leq \varepsilon n}\left\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\mathbb{Y}_i\mathbb{Y}_i^\top - \mathbf{C}_1\right)\right\|^2 \\
=\ & o_p(1),
\end{aligned}
$$

since we have by assumption **(W)** the rate $O(1)$ for the supremum and for the maximum we obtain the rate $O_p\left((\log(n))^{1/(2+\nu)}\right)$:

$$
\begin{aligned}
&P\left(\max_{1\leq k\leq \varepsilon n}\left\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\mathbb{Y}_i\mathbb{Y}_i^\top - \mathbf{C}_1\right)\right\|^2 > \delta\right) \\
=\ & P\left(\max_{1\leq k\leq \varepsilon n}\sum_{j=1}^{p}\sum_{l=1}^{p}\left(k^{-\frac{1}{2}}\sum_{i=1}^{k}(Y_{i-j}Y_{i-l} - E[Y_{i-j}Y_{i-l}])\right)^2 > \delta\right) \\
\leq\ & \sum_{j=1}^{p}\sum_{l=1}^{p} P\left(\max_{1\leq k\leq \varepsilon n}\left(k^{-\frac{1}{2}}\sum_{i=1}^{k}(Y_{i-j}Y_{i-l} - E[Y_{i-j}Y_{i-l}])\right)^2 > \delta/p^2\right).
\end{aligned}
$$

We define $Z_i = Y_{i-j}Y_{i-l} - E[Y_{i-j}Y_{i-l}]$ and deduce a Hájek-Rényi-type inequality by Markov's inequality and Theorem B.3 in Kirch (2006) (refer to Theorem B.3 in the Appendix) for $j = 1, \ldots, p$, $l = 1, \ldots, p$. Since $4 + \nu$ moments exist, the assumptions of Theorem B.3 are fulfilled (refer to Example 6.22 with $\Delta = 2 + \nu - \delta$ and δ small enough). With $\gamma = 2 + \nu$ and $\varphi = \gamma/2$ we have

$$
\begin{aligned}
&P\left(\max_{1\leq k\leq \varepsilon n} k^{-\frac{1}{2}}\sum_{i=1}^{k} Z_i > \delta/p^2\right) \ \leq\ P\left(\max_{1\leq k\leq n} k^{-\frac{1}{2}}\sum_{i=1}^{k} Z_i > \delta/p^2\right) \\
\leq\ & \frac{1}{(\delta/p^2)^\gamma}E\left[\left|\max_{1\leq k\leq n}\frac{1}{\sqrt{k}}\sum_{i=1}^{k} Z_i\right|^\gamma\right] \ \leq\ O(1)\frac{1}{\delta^\gamma}\sum_{k=1}^{n} k^{-\frac{\gamma}{2}}k^{\varphi-1} \\
=\ & O(1)\frac{1}{\delta^\gamma}\sum_{k=1}^{n} k^{-1} \ =\ O(1)\frac{1}{\delta^\gamma}\int_1^n x^{-1}\,dx \\
=\ & O(1)\frac{1}{\delta^\gamma}\log(n).
\end{aligned}
$$

We receive the analogous result for the maximum over the intervall $((1-\varepsilon)n, n]$. These rates hold for the second sum in (2.10) as well. Hence, we receive

$$\max_{1\le k\le n} \frac{w^2(k/n)}{n} \left\| \sum_{i=1}^{k} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right) (\hat{\mathbf{a}}_n - \mathbf{a}_0) \right\|^2 = o_p(1), \tag{2.11}$$

which implies

$$\begin{aligned} &\max_{1\le k\le n} \frac{w^2(k/n)}{n} \left\| \mathbf{S}_k^\top \mathbf{A} \mathbf{S}_k \right\|^2 \\ &= \max_{1\le k\le n} \frac{w^2(k/n)}{n} \left\| \mathbf{A}^{\frac{1}{2}} \sum_{i=1}^{k} \left(\mathbb{Y}_i \varepsilon_i - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \varepsilon_i \right) \right\|^2 + o_p(1). \end{aligned}$$

Moreover we obtain

$$\begin{aligned} &\max_{1\le k\le n} \frac{w^2(k/n)}{n} \left\| \mathbf{A}^{\frac{1}{2}} \sum_{i=1}^{k} \left(\mathbb{Y}_i \varepsilon_i - n^{-1} \sum_{i=1}^{n} \mathbb{Y}_i \varepsilon_i \right) \right\|^2 \\ &= \max_{1\le k\le n} w^2(k/n) \left\| \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{k} \left((\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) - n^{-1} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right) \right\|^2 \\ &= \max_{1\le k\le n} w^2(k/n) \left\| \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{k} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) - \frac{k}{n} \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2 \\ &= \sup_{0\le t\le 1} w^2(\lfloor tn \rfloor / n) \left\| \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor nt \rfloor} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) - \frac{\lfloor nt \rfloor}{n} \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2 . \end{aligned}$$

By the invariance principle in Lemma 2.4, we have

$$\left\{ (\mathbf{\Sigma} n)^{-\frac{1}{2}} \sum_{i=1}^{\lfloor nt \rfloor} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) : t \in [0,1] \right\} \overset{\mathcal{D}[0,1]}{\longrightarrow} \{ \mathbf{W}(t) : t \in [0,1] \} .$$

This implies that for

$$Z_{\lfloor tn \rfloor / n} := \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor nt \rfloor} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) - \frac{\lfloor nt \rfloor}{n} \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i])$$

holds

$$\left\{ \frac{1}{\sqrt{n}} Z_{\lfloor tn \rfloor / n} : t \in [0,1] \right\} \overset{\mathcal{D}[0,1]}{\longrightarrow} \{ \mathbf{B}(t) : t \in [0,1] \} ,$$

where the p-dimensional Brownian bridge has the covariance structure $\mathbf{A}^{\frac{1}{2}} \mathbf{\Sigma} \mathbf{A}^{\frac{1}{2}}$.

By assumption the weight function $w(\cdot)$ does not have to be continuous but is allowed to have K points of discontinuity $a_1, \ldots, a_K$. However, $w(\cdot)$ is continuous on each subinterval with existing limits and hence $\sup_{a_j<t<a_{j+1}} |w^2(\lfloor tn \rfloor/n) - w(t)| = o_p(1)$, $j = 1, \ldots, K$. Then, we receive with $\varepsilon < a_1$, $1 - \varepsilon > a_K$ and the continuous mapping theorem

$$
\begin{aligned}
&\sup_{\varepsilon<t<1-\varepsilon} w^2(\lfloor tn \rfloor/n) \left\| Z_{\lfloor tn \rfloor/n} \right\|^2 \qquad (2.12)\\
= \quad & \max\Bigg(\sup_{\varepsilon<t<a_1} w^2(\lfloor tn \rfloor/n) \left\| Z_{\lfloor tn \rfloor/n} \right\|^2, w^2(\lfloor a_1 n \rfloor/n) \left\| Z_{\lfloor a_1 n \rfloor/n} \right\|^2,\\
& \sup_{a_1<t<a_2} w^2(\lfloor tn \rfloor/n) \left\| Z_{\lfloor tn \rfloor/n} \right\|^2, w^2(\lfloor a_2 n \rfloor/n) \left\| Z_{\lfloor a_2 n \rfloor/n} \right\|^2,\\
& \ldots, w^2(\lfloor a_K n \rfloor/n) \left\| Z_{\lfloor tn \rfloor/n} \right\|^2, \sup_{a_K<t<1-\varepsilon} w^2(\lfloor tn \rfloor/n) \left\| Z_{\lfloor tn \rfloor/n} \right\|^2 \Bigg)\\
\overset{\mathcal{D}[0,1]}{\longrightarrow} \quad & \max\Bigg(\sup_{\varepsilon<t<a_1} w^2(t) \left\| \mathbf{B}(t) \right\|^2, w^2(a_1^-) \left\| \mathbf{B}(a_1) \right\|^2, \sup_{a_1<t<a_2} w^2(t) \left\| \mathbf{B}(t) \right\|^2,\\
& w^2(a_2^-) \left\| \mathbf{B}(a_2) \right\|^2, \ldots, w^2(a_K^-) \left\| \mathbf{B}(a_K) \right\|^2, \sup_{a_K<t<1-\varepsilon} w^2(t) \left\| \mathbf{B}(t) \right\|^2 \Bigg)\\
\overset{a.s.}{=} \quad & \max\Bigg(\sup_{\varepsilon<t\leq a_1} w^2(t) \left\| \mathbf{B}(t) \right\|^2, \sup_{a_1<t\leq a_2} w^2(t) \left\| \mathbf{B}(t) \right\|^2, \ldots, \sup_{a_K<t<1-\varepsilon} w^2(t) \left\| \mathbf{B}(t) \right\|^2 \Bigg)\\
\overset{a.s.}{=} \quad & \sup_{\varepsilon<t<1-\varepsilon} w^2(t) \left\| \mathbf{B}(t) \right\|^2 \overset{\mathcal{D}}{=} \sup_{\varepsilon<t<1-\varepsilon} w^2(t) \sum_{j=1}^{p} B_j^2(t)
\end{aligned}
$$

by the almost sure continuity of the Brownian bridge. Furthermore, we have

$$
\begin{aligned}
&\sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn \rfloor/n) \left\| \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor nt \rfloor} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) - \frac{\lfloor nt \rfloor}{n} \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2\\
\leq \quad & \sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn \rfloor/n) \left\| \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor nt \rfloor} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2\\
& + \sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn \rfloor/n) \left\| \frac{\lfloor nt \rfloor}{n} \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2. \qquad (2.13)
\end{aligned}
$$

For the second supremum we obtain

$$
\begin{aligned}
&\sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn \rfloor/n) \left\| \frac{\lfloor nt \rfloor}{n} \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2\\
= \quad & \left\| \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2 \sup_{0\leq t\leq\varepsilon} \left(\frac{\lfloor nt \rfloor}{n} \right)^2 w^2(t)\\
\leq \quad & \left\| \mathbf{A}^{\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbb{Y}_i \varepsilon_i - E[\mathbb{Y}_i \varepsilon_i]) \right\|^2 \varepsilon^{2(1-\alpha)} \sup_{0\leq t\leq\varepsilon} t^{2\alpha} w^2(t)
\end{aligned}
$$

and receive by Lemma 2.4

$$\left\|\mathbf{A}^{\frac{1}{2}}n^{-\frac{1}{2}}\sum_{i=1}^{n}\left(\mathbb{Y}_i\varepsilon_i - E[\mathbb{Y}_i\varepsilon_i]\right)\right\|^2 \quad = \quad O_p(1)$$

and by assumption **(W)** with $\alpha < 1/2$

$$\lim_{\varepsilon\to 0}\sup_{0\leq t\leq\varepsilon} t^{2\alpha}w^2(t) \quad = \quad \left(\lim_{\varepsilon\to 0}\sup_{0\leq t\leq\varepsilon} t^{\alpha}w(t)\right)^2 \quad = \quad \left(\lim_{t\to 0} t^{\alpha}w(t)\right)^2 < \infty \tag{2.14}$$

and eventually for all $\delta > 0$

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} P\left(\sup_{0\leq t\leq\varepsilon} w^2(t)\left\|\frac{\lfloor nt\rfloor}{n}\mathbf{A}^{\frac{1}{2}}n^{-\frac{1}{2}}\sum_{i=1}^{n}\left(\mathbb{Y}_i\varepsilon_i - E[\mathbb{Y}_i\varepsilon_i]\right)\right\|^2 > \delta\right) = 0.$$

For the first supremum in (2.13) we have

$$\begin{aligned}
&\sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn\rfloor/n)\left\|\mathbf{A}^{\frac{1}{2}}n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor nt\rfloor}\left(\mathbb{Y}_i\varepsilon_i - E[\mathbb{Y}_i\varepsilon_i]\right)\right\|^2 \\
&\quad\leq \quad \|\mathbf{A}\|\sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn\rfloor/n)\left\|n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor nt\rfloor}\left(\mathbb{Y}_i\varepsilon_i - E[\mathbb{Y}_i\varepsilon_i]\right)\right\|^2
\end{aligned}$$

and obtain

$$\begin{aligned}
&P\left(\sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn\rfloor/n)\left\|n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor nt\rfloor}\left(\mathbb{Y}_i\varepsilon_i - E[\mathbb{Y}_i\varepsilon_i]\right)\right\|^2 > \delta\right) \\
&\quad\leq \quad P\left(\sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn\rfloor/n)\sum_{j=1}^{p}\left(n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor nt\rfloor}\left(Y_{i-j}\varepsilon_i - E[Y_{i-j}\varepsilon_i]\right)\right)^2 > \delta\right) \\
&\quad\leq \quad \sum_{j=1}^{p} P\left(\sup_{0\leq t\leq\varepsilon} w^2(\lfloor tn\rfloor/n)\left(n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor nt\rfloor}\left(Y_{i-j}\varepsilon_i - E[Y_{i-j}\varepsilon_i]\right)\right)^2 > \delta/p\right).
\end{aligned}$$

We define $Z_i = Y_{i-j}\varepsilon_i - E[Y_{i-j}\varepsilon_i]$ and deduce a Hájek-Rényi-type inequality by Markov's inequality and Theorem B.3 in Kirch (2006) for $j = 1, \ldots, p$. To confirm that the assumptions in Theorem B.3 are fulfilled refer to Example 6.22. With $\gamma = 2 + \nu$ and $\varphi = \gamma/2$ we have

$$\begin{aligned}
P\left(\sup_{0\le t\le\varepsilon} w(\lfloor tn\rfloor/n)n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor nt\rfloor} Z_i > \delta/p\right) &= P\left(\max_{1\le k\le\varepsilon n} w(k/n)n^{-\frac{1}{2}}\sum_{i=1}^{k} Z_i > \delta/p\right)\\
\le \frac{1}{(\delta/p)^\gamma}E\left[\left|\max_{1\le k\le\varepsilon n} w(k/n)\frac{1}{\sqrt{n}}\sum_{i=1}^{k} Z_i\right|^\gamma\right] &\le O(1)\frac{1}{\delta^\gamma}\sum_{k=1}^{\varepsilon n}\left(w(k/n)\frac{1}{\sqrt{n}}\right)^\gamma k^{\varphi-1}\\
= O(1)\frac{1}{\delta^\gamma}\frac{1}{n}\sum_{k=1}^{\varepsilon n} w^\gamma(k/n)\left(\frac{k}{n}\right)^{\frac{\gamma}{2}-1} &= O(1)\frac{1}{\delta^\gamma}\left(\int_0^\varepsilon w^\gamma(t)t^{\frac{\gamma}{2}-1}\,dt + o(1)\right),
\end{aligned}$$

where we used in the last step the uniform continuity of $w(\cdot)$ in $[0,\varepsilon]$ for ε small enough. Moreover we obtain

$$\begin{aligned}
\int_0^\varepsilon w^\gamma(t)t^{\frac{\gamma}{2}-1}\,dt + o(1) &= \left(\sup_{0\le t\le\varepsilon} w(t)t^\alpha\right)^\gamma \int_0^\varepsilon t^{\frac{\gamma}{2}-\alpha\gamma-1}\,dt + o(1)\\
= \left(\sup_{0\le t\le\varepsilon} w(t)t^\alpha\right)^\gamma \int_0^\varepsilon t^{-((\alpha-\frac{1}{2})\gamma+1)}\,dt + o(1).
\end{aligned}$$

Since $\left(\alpha - \frac{1}{2}\right)\gamma + 1 < 1$ the integral converges to zero as $\varepsilon \to 0$ and we receive with (2.14)

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} P\left(\sup_{0\le t\le\varepsilon} w^2(t)\left\|\mathbf{A}^{\frac{1}{2}}n^{-\frac{1}{2}}\sum_{i=1}^{nt}(\mathbb{Y}_i\varepsilon_i - E[\mathbb{Y}_i\varepsilon_i])\right\|^2 > \delta\right) = 0.$$

Moreover, we have from Csörgő and Horváth (1993), page 181,

$$\lim_{\varepsilon\to 0}\sup_{0\le t\le\varepsilon} w^2(t)\,\|\mathbb{B}(t)\|^2 = o_p(1), \tag{2.15}$$

with a p-dimensional standard Brownian bridge $\{\mathbb{B}(t)\}$. This clearly holds for $\{\mathbf{B}(t)\}$ as well. We receive analogous results for the maximum over the intervall $((1-\varepsilon)n, n]$ and hence have shown assertion (a). Part (b) follows analogously, where we make repeated use of the fact that for $1 \le a < b \le n$

$$\frac{1}{n}\sum_{k=a+1}^{b}\frac{w^2(k/n)}{n}\mathbf{S}_k^\top\mathbf{A}\mathbf{S}_k \le \max_{a<k\leqslant b}\frac{w^2(k/n)}{n}\mathbf{S}_k^\top\mathbf{A}\mathbf{S}_k.$$

□

Under the epidemic model we use the test statistics

$$
\begin{aligned}
T_n^{(3)} &= \max_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) \frac{1}{n} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1})^\top \mathbf{A} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1}), \\
T_n^{(4)} &= \frac{1}{n^3} \sum_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) (\mathbf{S}_{k_2} - \mathbf{S}_{k_1})^\top \mathbf{A} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1})
\end{aligned}
$$

with

$$
w(t_1, t_2) = \mathcal{I}_{l_1 \leqslant t_1 < t_2 \leqslant l_2}(t_1, t_2) \qquad \text{for } (t_1, t_2) \in [0,1]^2 \qquad \text{and } \ 0 \le l_1 < l_2 \le 1. \tag{2.16}
$$

In contrast to the AMOC test statistics we use here a specific weight function to keep the proofs more simple. More information about epidemic change-point tests can be found in Csörgő and Horváth (1997), Chapter 2.8.4. Theorem 2.7 gives the asymptotic null distributions.

Theorem 2.7. Let the process $\{Y_i\}$ fulfill **(Y)** and $\{\varepsilon_i\}$ is defined as in (2.4). Moreover, let the weight function $w(k_1/n, k_2/n)$ be defined as in (2.16). Then, under H_0,

(a) $T_n^{(3)} \xrightarrow{\mathcal{D}} \sup_{0 \leqslant t_1 < t_2 \leqslant 1} w(t_1, t_2) \sum_{j=1}^{p} (B_j(t_2) - B_j(t_1))^2,$

(b) $T_n^{(4)} \xrightarrow{\mathcal{D}} \iint_{0 \le t_1 < t_2 \le 1} w(t_1, t_2) \sum_{j=1}^{p} (B_j(t_2) - B_j(t_1))^2 \, dt_1 \, dt_2.$

where $\{\mathbf{B}(t) = (B_1(t), \ldots, B_p(t))^\top : t \in [0,1]\}$ denotes a p-dimensional Brownian bridge with covariance matrix $\mathbf{A}^{\frac{1}{2}} \mathbf{\Sigma} \mathbf{A}^{\frac{1}{2}}$, $\mathbf{\Sigma}$ is the long-run autocovariance matrix of $\{\mathbb{Y}_i \varepsilon_i\}$ and $\mathbf{A}$ is a symmetric, positive semidefinite matrix which can be replaced by a consistent estimator.

Proof of Theorem 2.7. For part (a) we obtain with $\mathbf{S}_n = 0$ and (2.8)

$$
\begin{aligned}
&\max_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) \frac{1}{n} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1})^\top \mathbf{A} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1}) \\
&= \max_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) \frac{1}{n} \left\| \mathbf{A}^{\frac{1}{2}} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1}) \right\|^2 \\
&= \max_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) \frac{1}{n} \left\| \mathbf{A}^{\frac{1}{2}} \sum_{i=k_1+1}^{k_2} \left(\mathbb{Y}_i \varepsilon_i - \frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}_i \varepsilon_i \right) \right\|^2 + o_p(1),
\end{aligned}
$$

since

$$
\begin{aligned}
&\max_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1, k_2) \frac{1}{n} \left\| \mathbf{A}^{\frac{1}{2}} \sum_{i=k_1+1}^{k_2} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - \frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right) (\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}) \right\|^2 \\
&\leq \left\| \mathbf{A}^{\frac{1}{2}} \right\| \max_{1 \leqslant k_1 < k_2 \leqslant n} \frac{1}{n} \left\| \sum_{i=k_1+1}^{k_2} \left(\mathbb{Y}_i \mathbb{Y}_i^\top - \frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right) \right\|^2 \|\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}\|^2 = o_p(1),
\end{aligned} \tag{2.17}
$$

by part (b) of Theorem 2.1 and

$$\begin{aligned}
&\max_{1\leqslant k_1<k_2\leqslant n}\left\|\frac{1}{\sqrt{n}}\sum_{i=k_1+1}^{k_2}\left(\mathbb{Y}_i\mathbb{Y}_i^\top-\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i\mathbb{Y}_i^\top\right)\right\|^2\\
&= \max_{1\leqslant k_1<k_2\leqslant n}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{k_2}\left(\mathbb{Y}_i\mathbb{Y}_i^\top-\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i\mathbb{Y}_i^\top\right)-\frac{1}{n}\sum_{i=1}^{k_1}\left(\mathbb{Y}_i\mathbb{Y}_i^\top-\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i\mathbb{Y}_i^\top\right)\right\|^2\\
&\leq 2\max_{1\leqslant k\leqslant n}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\left(\mathbb{Y}_i\mathbb{Y}_i^\top-\frac{1}{n}\sum_{i=1}^{n}\mathbb{Y}_i\mathbb{Y}_i^\top\right)\right\|^2\\
&\leq 2\max_{1\leqslant k\leqslant n}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\left(\mathbb{Y}_i\mathbb{Y}_i^\top-\mathbf{C}_1\right)\right\|^2+2\max_{1\leqslant k\leqslant n}\left\|\frac{1}{\sqrt{n}}\frac{k}{n}\sum_{i=1}^{n}\left(\mathbb{Y}_i\mathbb{Y}_i^\top-\mathbf{C}_1\right)\right\|^2\\
&= o_p(n)
\end{aligned}$$

by the strong law of large numbers. Since $[t_1,t_2]^2$ can be separated into the two sets $A=\{(t_1,t_2): l_1\leq t_1<t_2\leq l_2\}$ and A^c the assertion follows similar to (2.12) with Lemma 2.4, the continous mapping theorem and the almost sure continuity of the Brownian bridge. Part (b) follows analogously. □

2.1.3. Behaviour under Alternatives

Before we prove the consistency of the tests, the following lemmas are obtained.

Lemma 2.8. With $\widetilde{\mathbf{a}}$ defined as in (2.6) it holds

(a) under the AMOC alternative

$$\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}} = -\frac{1-\lambda}{\lambda}\left(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\right),$$

(b) under the epidemic alternative

$$\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}} = -\frac{\lambda_2-\lambda_1}{1-(\lambda_2-\lambda_1)}\left(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\right).$$

Proof of Lemma 2.8. For part (a) we use $\widetilde{\mathbf{a}}=\mathbf{Q}^{-1}\boldsymbol{\gamma}$ to obtain

$$\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}} = \boldsymbol{\gamma}_1-\mathbf{C}_1\mathbf{Q}^{-1}\boldsymbol{\gamma} = \boldsymbol{\gamma}_1-\frac{1}{\lambda}\left(\lambda\mathbf{C}_1+(1-\lambda)\mathbf{C}_2\right)\mathbf{Q}^{-1}\boldsymbol{\gamma}+\frac{1-\lambda}{\lambda}\mathbf{C}_2\mathbf{Q}^{-1}\boldsymbol{\gamma}$$

and further $\mathbf{Q} = \lambda \mathbf{C}_1 + (1-\lambda)\mathbf{C}_2$ and $\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}_1 + (1-\lambda)\boldsymbol{\gamma}_2$ to receive

$$\begin{aligned}\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}} &= \boldsymbol{\gamma}_1 - \frac{1}{\lambda}\boldsymbol{\gamma} + \frac{1-\lambda}{\lambda}\mathbf{C}_2\mathbf{Q}^{-1}\boldsymbol{\gamma} \;=\; \boldsymbol{\gamma}_1 - \frac{1}{\lambda}\left(\lambda\boldsymbol{\gamma}_1 + (1-\lambda)\boldsymbol{\gamma}_2\right) + \frac{1-\lambda}{\lambda}\mathbf{C}_2\mathbf{Q}^{-1}\boldsymbol{\gamma} \\ &= -\frac{1-\lambda}{\lambda}\boldsymbol{\gamma}_2 + \frac{1-\lambda}{\lambda}\mathbf{C}_2\mathbf{Q}^{-1}\boldsymbol{\gamma} \;=\; -\frac{1-\lambda}{\lambda}\left(\boldsymbol{\gamma}_2 - \mathbf{C}_2\mathbf{Q}^{-1}\boldsymbol{\gamma}\right).\end{aligned}$$

This is assertion (a). Part (b) follows analogously with $\mathbf{Q} = (1-(\lambda_2-\lambda_1))\,\mathbf{C}_1 + (\lambda_2-\lambda_1)\,\mathbf{C}_2$ and $\boldsymbol{\gamma} = (1-(\lambda_2-\lambda_1))\,\boldsymbol{\gamma}_1 + (\lambda_2-\lambda_1)\,\boldsymbol{\gamma}_2$. □

Lemma 2.9. Let the processes $\{Y_i^{(1)}\}$, $\{Y_i^{(2)}\}$ fulfill **(Y)**.

(a) Then, under the AMOC alternative,

$$\sup_{0\le t\le 1}\left\|\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor} - \mathbf{g}(t)\right\|^2 = o_p(1)$$

with

$$\mathbf{g}(t) = \begin{cases} t\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & t\le\lambda, \\ (t-1)\left(\boldsymbol{\gamma}_2 - \mathbf{C}_2\widetilde{\mathbf{a}}\right), & t>\lambda. \end{cases} \tag{2.18}$$

(b) Then, under the epidemic alternative,

$$\sup_{0\le t_1<t_2\le 1}\left\|\frac{1}{n}\left(\mathbf{S}_{\lfloor t_2 n\rfloor} - \mathbf{S}_{\lfloor t_1 n\rfloor}\right) - \mathbf{g}(t_1,t_2)\right\|^2 = o_p(1)$$

with $\mathbf{g}(t_1,t_2) = \mathbf{g}_e(t_2) - \mathbf{g}_e(t_1)$ and

$$\mathbf{g}_e(t) = \begin{cases} t\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & t\le\lambda_1, \\ \left(t - \frac{\lambda_1}{1-(\lambda_2-\lambda_1)}\right)\left(\boldsymbol{\gamma}_2 - \mathbf{C}_2\widetilde{\mathbf{a}}\right), & \lambda_1 < t\le\lambda_2, \\ (t-1)\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & t>\lambda_2. \end{cases}$$

Proof of Lemma 2.9. We have

$$\begin{aligned}\mathbf{S}_{\lfloor tn\rfloor} &= \sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\hat{\varepsilon}_i = \sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(Y_i - \mathbb{Y}_i^\top\widehat{\mathbf{a}}_n\right) \\ &= \sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(Y_i - \mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right) - \sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(\mathbb{Y}_i^\top\widehat{\mathbf{a}}_n - \mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right) \\ &= \sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(Y_i - \mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right) - \sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\mathbb{Y}_i^\top\left(\widehat{\mathbf{a}}_n - \widetilde{\mathbf{a}}\right)\end{aligned} \tag{2.19}$$

and receive under the AMOC alternative

$$\sup_{0\le t\le 1}\left\|\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor}-\mathbf{g}(t)\right\|^2 = \sup_{0\le t\le 1}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)-\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\mathbb{Y}_i^\top\left(\widehat{\mathbf{a}}_n-\widetilde{\mathbf{a}}\right)-\mathbf{g}(t)\right\|^2$$

$$= \sup_{0\le t\le 1}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)-\mathbf{g}(t)\right\|^2+o_p(1),$$

since

$$\sup_{0\le t\le 1}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\mathbb{Y}_i^\top\left(\widehat{\mathbf{a}}_n-\widetilde{\mathbf{a}}\right)\right\|^2 \le \max_{0\le t\le 1}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\mathbb{Y}_i^\top\right\|^2\left\|\widehat{\mathbf{a}}_n-\widetilde{\mathbf{a}}\right\|^2 = o_p(1)$$

by part (a) of Theorem 2.1, and the maximum is stochastically bounded by the strong law of large numbers. Since

$$E\left[\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)\right] = \boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}},\qquad \text{for } i\le\lfloor\lambda n\rfloor,$$

we obtain with the strong law of large numbers

$$\sup_{0\le t\le\lambda}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)-\mathbf{g}(t)\right\|^2$$

$$= \sup_{0\le t\le\lambda}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\left(\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)-E\left[\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)\right]\right)\right\|^2 = o(1)\text{ a.s.}$$

With Lemma 2.8 part (a) we receive for $t>\lambda$

$$\begin{aligned}\mathbf{g}(t) &= (t-1)\left(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\right) = (t-\lambda-(1-\lambda))\left(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\right)\\ &= (t-\lambda)\left(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\right)-(1-\lambda)\left(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\right)\\ &= (t-\lambda)\left(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\right)+\lambda\left(\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}}\right)\end{aligned}$$

and further it holds

$$E\left[\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)\right] = \boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}},\qquad \text{for } i>\lfloor\lambda n\rfloor+p.$$

Hence, we obtain with the strong law of large numbers

$$\sup_{\lambda<t\le 1}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor tn\rfloor}\mathbb{Y}_i\left(Y_i-\mathbb{Y}_i^\top\widetilde{\mathbf{a}}\right)-\mathbf{g}(t)\right\|^2$$

$$
\begin{aligned}
&\leq \left\| \frac{1}{n} \sum_{i=1}^{\lfloor \lambda n \rfloor} \left(\mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) - E\left[\mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) \right] \right) \right\|^2 \\
&\quad + \sup_{\lambda < t \leq 1} \left\| \frac{1}{n} \sum_{i=\lfloor \lambda n \rfloor + 1}^{\lfloor tn \rfloor} \left(\mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) - E\left[\mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) \right] \right) \right\|^2 + o_p(1) \;=\; o_p(1),
\end{aligned}
$$

since the mixed terms are asymptotically neglible and the second process is assumed to be stationary. For (b) we have the same decomposition as in (2.19) and receive

$$
\begin{aligned}
&\sup_{0 \leq t_1 < t_2 \leq 1} \left\| \left(\frac{1}{n} \left(\mathbf{S}_{\lfloor t_2 n \rfloor} - \mathbf{S}_{\lfloor t_1 n \rfloor} \right) - \mathbf{g}(t_1, t_2) \right) \right\|^2 \\
&= \sup_{0 \leq t_1 < t_2 \leq 1} \left\| \left(\frac{1}{n} \left(\sum_{i=1}^{\lfloor t_2 n \rfloor} \mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) - \sum_{i=1}^{\lfloor t_1 n \rfloor} \mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) \right) - \mathbf{g}(t_1, t_2) \right) \right\|^2 + o_p(1).
\end{aligned}
$$

Since $\mathbf{g}(t_1, t_2) = \mathbf{g}_e(t_2) - \mathbf{g}_e(t_1)$ we have

$$
\begin{aligned}
&\sup_{0 \leq t_1 < t_2 \leq 1} \left\| \left(\frac{1}{n} \left(\sum_{i=1}^{\lfloor t_2 n \rfloor} \mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) - \sum_{i=1}^{\lfloor t_1 n \rfloor} \mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) \right) - \mathbf{g}(t_1, t_2) \right) \right\|^2 \\
&\leq 2 \sup_{0 \leq t \leq 1} \left\| \frac{1}{n} \sum_{i=1}^{\lfloor tn \rfloor} \mathbb{Y}_i \left(Y_i - \mathbb{Y}_i^\top \widetilde{\mathbf{a}} \right) - \mathbf{g}_e(t) \right\|^2 .
\end{aligned}
$$

By Lemma 2.8 part (b) we receive for $\lambda_1 < t \leq \lambda_2$

$$
\begin{aligned}
\mathbf{g}_e(t) &= \left(t - \frac{\lambda_1}{1 - (\lambda_2 - \lambda_1)} \right) (\boldsymbol{\gamma}_2 - \mathbf{C}_2 \widetilde{\mathbf{a}}) \\
&= \left(t - \frac{\lambda_1 \left(1 - (\lambda_2 - \lambda_1) + (\lambda_2 - \lambda_1)\right)}{1 - (\lambda_2 - \lambda_1)} \right) (\boldsymbol{\gamma}_2 - \mathbf{C}_2 \widetilde{\mathbf{a}}) \\
&= \left(t - \lambda_1 - \frac{\lambda_1 (\lambda_2 - \lambda_1)}{1 - (\lambda_2 - \lambda_1)} \right) (\boldsymbol{\gamma}_2 - \mathbf{C}_2 \widetilde{\mathbf{a}}) \\
&= \lambda_1 (\boldsymbol{\gamma}_1 - \mathbf{C}_1 \widetilde{\mathbf{a}}) + (t - \lambda_1) (\boldsymbol{\gamma}_2 - \mathbf{C}_2 \widetilde{\mathbf{a}})
\end{aligned}
$$

and for $t > \lambda_2$

$$
\begin{aligned}
\mathbf{g}_e(t) &= (t-1) (\boldsymbol{\gamma}_1 - \mathbf{C}_1 \widetilde{\mathbf{a}}) \;=\; (t - (\lambda_2 - \lambda_1) - (1 - (\lambda_2 - \lambda_1)) (\boldsymbol{\gamma}_1 - \mathbf{C}_1 \widetilde{\mathbf{a}}) \\
&= (t - (\lambda_2 - \lambda_1)) (\boldsymbol{\gamma}_1 - \mathbf{C}_1 \widetilde{\mathbf{a}}) - (1 - (\lambda_2 - \lambda_1)) (\boldsymbol{\gamma}_1 - \mathbf{C}_1 \widetilde{\mathbf{a}}) \\
&= (t - (\lambda_2 - \lambda_1)) (\boldsymbol{\gamma}_1 - \mathbf{C}_1 \widetilde{\mathbf{a}}) + (\lambda_2 - \lambda_1) (\boldsymbol{\gamma}_2 - \mathbf{C}_2 \widetilde{\mathbf{a}}) .
\end{aligned}
$$

Then, we obtain assertion (b) with the same arguments as above, but split the supremum into three parts. □

Theorem 2.10. Let $\{Y_i^{(1)}\}$, $\{Y_i^{(2)}\}$ fulfill **(Y)**. Further, we assume

$$\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}) \neq \mathbf{0}, \tag{2.20}$$

where the symmetric, positive semidefinite matrix $\mathbf{A}$ can be replaced by an estimator $\widehat{\mathbf{A}}_n$ with $\|\widehat{\mathbf{A}}_n^{\frac{1}{2}}(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}})\| \geq \varepsilon > 0$ and $\|\widehat{\mathbf{A}}_n^{\frac{1}{2}}\| \leq C < \infty$ for all n.

(a) Moreover, let the weight function $w(\cdot)$ fulfill **(W)** and $w \not\equiv 0$. Then, under the AMOC alternative,

$$T_n^{(1)} \xrightarrow{P} \infty \quad \text{and} \quad T_n^{(2)} \xrightarrow{P} \infty.$$

(b) Let the weight function $w(\cdot,\cdot)$ be defined as in (2.16). Then, under the epidemic alternative,

$$T_n^{(3)} \xrightarrow{P} \infty \quad \text{and} \quad T_n^{(4)} \xrightarrow{P} \infty.$$

Proof of Theorem 2.10. We begin to show part (a). It exists a continuity point t_0 with $w(t_0) > 0$. Then, we obtain a lower bound for the test statistic by looking only at the time point $k_0 = \lfloor t_0 n \rfloor$ rather than at the maximum over all time points:

$$\begin{aligned}
\max_{1\leq k\leq n} \frac{w^2(k/n)}{n}\mathbf{S}_k^\top \mathbf{A}\mathbf{S}_k &= \max_{1\leq k\leq n} \frac{w^2(k/n)}{n}\left\|\mathbf{A}^{\frac{1}{2}}\mathbf{S}_k\right\|^2 \geq \frac{w^2(k_0/n)}{n}\left\|\mathbf{A}^{\frac{1}{2}}\mathbf{S}_{k_0}\right\|^2 \\
&= n\,w^2(k_0/n)\left\|\mathbf{A}^{\frac{1}{2}}\frac{1}{n}\mathbf{S}_{k_0}\right\|^2 = n\,w^2(t_0)\left\|\mathbf{A}^{\frac{1}{2}}\mathbf{g}(t_0)\right\|^2 + o_p(n) \\
&= n\,w^2(t_0)\tilde{g}^2(t_0)\left\|\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}})\right\|^2 + o_p(n)
\end{aligned}$$

by Lemma 2.9 (a) and Lemma 2.8 (a), which implies with

$$\tilde{g}(t) = \begin{cases} t, & t \leq \lambda, \\ \frac{(1-t)\lambda}{1-\lambda}, & t > \lambda, \end{cases}$$

that $\mathbf{g}(t) = \tilde{g}(t)(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}})$. Since $\left\|\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}})\right\|^2 > 0$ and $w^2(t_0) > 0$ by assumption the convergence to infinity is ensured. The assertion for $T^{(2)}$ follows by Lemma 2.9 (a) as well:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{w^2(k/n)}{n^2}\mathbf{S}_k^\top \mathbf{A}\mathbf{S}_k &= \sum_{k=1}^{n} w^2(k/n)\left\|\mathbf{A}^{\frac{1}{2}}\frac{1}{n}\mathbf{S}_k\right\|^2 \\
&= n\int_0^1 w^2(t)\tilde{g}^2(t)\left\|\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}})\right\|^2 dt + o_p(n),
\end{aligned}$$

since by assumption $\int_0^1 w^2(t)\tilde{g}^2(t)\left\|\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}})\right\|^2 dt>0$. We receive part (b) similar to part (a): By Lemma 2.8 (b) it follows $\mathbf{g}_e(t)=\tilde{g}_e(t)(\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}})$ with

$$
\tilde{g}_e(t) \quad = \quad \begin{cases} t, & t\leq\lambda_1, \\ \frac{\lambda_1-t(1-(\lambda_2-\lambda_1))}{\lambda_2-\lambda_1}, & \lambda_1<t\leq\lambda_2 \\ t-1, & t>\lambda_2. \end{cases}
$$

Hence, it exists $k_3=\lfloor t_1 n\rfloor$, $k_4=\lfloor t_2 n\rfloor$ with $l_1<t_1<t_2<l_2$ and $\tilde{g}_e(t_2)-\tilde{g}_e(t_1)\neq 0$. Then, we use Lemma 2.9 (b) to obtain:

$$
\begin{aligned}
&\max_{1\leqslant k_1<k_2\leqslant n} w(k_1/n,k_2/n)\frac{1}{n}(\mathbf{S}_{k_2}-\mathbf{S}_{k_1})^\top\mathbf{A}(\mathbf{S}_{k_2}-\mathbf{S}_{k_1}) \\
&\geq \; w(k_3/n,k_4/n)\frac{1}{n}\left\|\mathbf{A}^{\frac{1}{2}}(\mathbf{S}_{k_4}-\mathbf{S}_{k_3})\right\|^2 \\
&= \; n\left\|\mathbf{A}^{\frac{1}{2}}\mathbf{g}(t_1,t_2)\right\|^2+o_p(n) \\
&= \; n\left(\tilde{g}_e(t_2)-\tilde{g}_e(t_1)\right)^2\left\|\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}})\right\|^2+o_p(n).
\end{aligned}
$$

The result for $T^{(4)}$ follows as above. All considerations hold for $\widehat{\mathbf{A}}_n$ as well. □

Remark 2.11. (a) By Lemma 2.8 the assumption $\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}})\neq\mathbf{0}$ in (2.20) can be replaced by the assumption $\mathbf{A}^{\frac{1}{2}}(\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}})\neq\mathbf{0}$.

(b) If $\mathbf{A}$ is positive definite, assumption (2.20) is fulfilled if

$$
\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}}\neq\mathbf{0} \quad \text{or} \quad \boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}\neq\mathbf{0}. \tag{2.21}
$$

This holds if

$$
\mathbf{C}_1^{-1}\boldsymbol{\gamma}_1 \;\neq\; \mathbf{C}_2^{-1}\boldsymbol{\gamma}_2. \tag{2.22}
$$

We prove that (2.22) implies (2.21) by contradiction. To this end let $\boldsymbol{\gamma}_1-\mathbf{C}_1\widetilde{\mathbf{a}}=0$ and $\boldsymbol{\gamma}_2-\mathbf{C}_2\widetilde{\mathbf{a}}=0$ hold. This leads to $\widetilde{\mathbf{a}}=\mathbf{C}_1^{-1}\gamma_1$ and $\widetilde{\mathbf{a}}=\mathbf{C}_2^{-1}\gamma_2$, which is a contradiction to (2.22). We note that $\mathbf{C}_1^{-1}\boldsymbol{\gamma}_1$ and $\mathbf{C}_2^{-1}\boldsymbol{\gamma}_2$ correspond to $\mathbf{a}_1$ and $\mathbf{a}_2$ as in (2.5).

(c) If $\mathbf{A}$ consists of a positive definite $p'\times p'$ block matrix $\mathbf{A}^*$ and is otherwise 0, i.e.

$$
\mathbf{A} \quad = \quad \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},
$$

$\mathbf{A}$ is positive semidefinite. Then, assumption (2.20) holds if $\widetilde{\mathbf{a}}$ does not fulfill all of the first p' Yule-Walker equations $\boldsymbol{\gamma}_1-\mathbf{C}_1\mathbf{a}=0$ (refer to (2.21)). ■

2.1.4. Change-Point Estimators

Under the AMOC model we estimate the change point by

$$\hat{k}_n := \arg\max_{1<k<n} w^2(k/n)\mathbf{S}_k^\top \mathbf{A}\mathbf{S}_k$$

and under the epidemic model we define

$$\left(\hat{k}_{1,n}, \hat{k}_{2,n}\right) := \arg\max_{1\leqslant k_1<k_2\leqslant n} w(k_1/n, k_2/n)(\mathbf{S}_{k_2} - \mathbf{S}_{k_1})^\top \mathbf{A}(\mathbf{S}_{k_2} - \mathbf{S}_{k_1}).$$

Theorem 2.12 shows that these estimators are consistent in all cases where the test has asymptotic power one.

Theorem 2.12. Let the assumptions of Theorem 2.10 be fulfilled.

(a) Moreover, let the weight function be of the form

$$w^2(t) = \omega^2(t)\mathcal{I}_{l_1\leq t\leq l_2}(t) \quad \text{for} \quad t \in [0,1] \quad \text{and} \quad 0 \leq l_1 < l_2 \leq 1$$

with $\omega(\cdot)$ continuous and bounded on $[l_1, l_2]$. Further, let $w(t)\,\|\mathbf{g}(t)\|$ have a unique maximum at λ with $\mathbf{g}(\cdot)$ as in (2.18). Then, under the AMOC change alternative,

$$\frac{\hat{k}_n}{n} \xrightarrow{P} \lambda.$$

(b) Moreover, let $[l_1, l_2]^2$ include the change point (λ_1, λ_2). Then, under the epidemic alternative,

$$\left(\frac{\hat{k}_{1,n}}{n}, \frac{\hat{k}_{2,n}}{n}\right) \xrightarrow{P} (\lambda_1, \lambda_2)\,.$$

Remark 2.13. The function $\mathbf{g}(\cdot)$ as in (2.18) is by Lemma 2.8 (a) equivalent to $\mathbf{g}(t) = \tilde{g}(t)(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}})$ with

$$\tilde{g}(t) = \begin{cases} t, & t \leq \lambda, \\ \frac{(1-t)\lambda}{1-\lambda}, & t > \lambda. \end{cases} \tag{2.23}$$

Hence $\|\mathbf{g}(\cdot)\|$ is a continuous function with a unique maximum at λ. Thus, the assertion "$w(t)\,\|\mathbf{g}(t)\|$ has a unique maximum in λ" in part (a) of Theorem 2.12 is not a strong assumption. If the interval $[l_1, l_2]$ includes the change point, this assertion is fulfilled for reasonable weight functions, e.g. with $\omega(\cdot)$ as in (2.7) (this can directly be seen by differentiation of $\tilde{g}(t)\omega(t)$).

Proof of Theorem 2.12. For part (a) we have from Lemma 2.9 that

$$
\begin{aligned}
&\sup_{t\in[0,1]}\left| w^2\left(\lfloor tn\rfloor/n\right)\left\|\mathbf{A}^{\frac{1}{2}}\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor}\right\|^2 - w^2(t)\left\|\mathbf{A}^{\frac{1}{2}}\mathbf{g}(t)\right\|^2\right| \\
\leq\;& \sup_{t\in[0,1]}\left\| w\left(\lfloor tn\rfloor/n\right)\mathbf{A}^{\frac{1}{2}}\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor} - w(t)\mathbf{A}^{\frac{1}{2}}\mathbf{g}(t)\right\|^2 \\
\leq\;& \|\mathbf{A}^{\frac{1}{2}}\|^2 \sup_{t\in[0,1]}\left\| w\left(\lfloor tn\rfloor/n\right)\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor} - w(t)\mathbf{g}(t)\right\|^2 \\
=\;& \|\mathbf{A}^{\frac{1}{2}}\|^2 \sup_{t\in[0,1]}\left\| w\left(\lfloor tn\rfloor/n\right)\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor} - w\left(\lfloor tn\rfloor/n\right)\mathbf{g}(t)\right\|^2 + o(1) \\
\leq\;& \|\mathbf{A}^{\frac{1}{2}}\|^2 \sup_{t\in[0,1]} w\left(\lfloor tn\rfloor/n\right)\sup_{t\in[0,1]}\left\|\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor} - \mathbf{g}(t)\right\|^2 + o(1) \\
=\;& o_p(1),
\end{aligned}
$$

Since $\omega(\cdot)\,\|\mathbf{g}(\cdot)\|$ is continuous and has a unique maximum at λ, Theorem B.4 yields the assertion.

For part (b) we have from Lemma 2.9

$$
\begin{aligned}
&\sup_{0\leq t_1<t_2\leq 1}\left|\left\| w\left(\lfloor t_1 n\rfloor/n, \lfloor t_2 n\rfloor/n\right)\mathbf{A}^{\frac{1}{2}}\frac{1}{n}\left(\mathbf{S}_{\lfloor t_2 n\rfloor} - \mathbf{S}_{\lfloor t_1 n\rfloor}\right)\right\|^2 - \left\| w(t_1,t_2)\mathbf{A}^{\frac{1}{2}}\mathbf{g}(t_1,t_2)\right\|^2\right| \\
\leq\;& \|\mathbf{A}^{\frac{1}{2}}\|^2 \sup_{0\leq t_1<t_2\leq 1}\left\| w\left(\lfloor t_1 n\rfloor/n, \lfloor t_2 n\rfloor/n\right)\frac{1}{n}\left(\mathbf{S}_{\lfloor t_2 n\rfloor} - \mathbf{S}_{\lfloor t_1 n\rfloor}\right) - w(t_1,t_2)\mathbf{g}(t_1,t_2)\right\|^2 \\
\leq\;& 2\|\mathbf{A}^{\frac{1}{2}}\|^2 \sup_{0\leq t\leq 1}\left\|\frac{1}{n}\mathbf{S}_{\lfloor tn\rfloor} - \mathbf{g}_e(t)\right\|^2 + o(1) \;=\; o_p(1)
\end{aligned}
$$

with $\mathbf{g}(t_1,t_2) = \mathbf{g}_e(t_2) - \mathbf{g}_e(t_1)$ and

$$
\mathbf{g}_e(t) = \begin{cases} t\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & t\leq\lambda_1, \\ \left(t - \frac{\lambda_1}{1-(\lambda_2-\lambda_1)}\right)\left(\boldsymbol{\gamma}_2 - \mathbf{C}_2\widetilde{\mathbf{a}}\right), & \lambda_1 < t\leq\lambda_2, \\ (t-1)\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & t>\lambda_2. \end{cases}
$$

With Lemma 2.8 we obtain

$$
\mathbf{g}_e(t) = \begin{cases} t\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & t\leq\lambda_1, \\ \left(\frac{\lambda_1}{\lambda_2-\lambda_1} - \frac{t(1-(\lambda_2-\lambda_1))}{\lambda_2-\lambda_1}\right)\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & \lambda_1 < t\leq\lambda_2, \\ (t-1)\left(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}}\right), & t>\lambda_2. \end{cases}
$$

Since g_e is strictly increasing on $[0,\lambda_1)$, strictly decreasing on $(\lambda_1,\lambda_2]$ and again strictly increasing on $(\lambda_2,1]$, $\mathbf{g}$ has a unique maximum at (λ_1,λ_2). Then, Lemma B.4 yields assertion (b). □

2.1.5. Specific Test Statistics and Estimators

The test statistics are based on partial sums $\mathbf{S}_k$, which are weighted by the matrix $\mathbf{A}$. In the literature the common choice of $\mathbf{A}$ is $\mathbf{A} = \mathbf{\Sigma}^{-1}$. In situations where changes in $\{Y_i\}$ are in some components of $\mathbf{S}_k$ more visible than in others, a statistic with a matrix $\mathbf{A}$ giving larger weight to components of $\mathbf{S}_k$, which reflect these changes better, will detect these changes more likely in small sample sizes. Hence, it will have a better power behaviour, but at the expensive of detecting changes in $\{Y_i\}$, which are visible in other components of $\mathbf{S}_k$, less likely.

In our data example we expect changes to be in the first $p' = 2$ autocorrelations and propose for the $p \times p$ matrix $\mathbf{A}$ a matrix $\widetilde{\mathbf{A}}^{(p')}$ which consists of the positive definite $p' \times p'$ block matrix $(\mathbf{\Sigma}[p'])^{-1}$ and is otherwise 0:

$$\widetilde{\mathbf{A}}^{(p')} = \begin{pmatrix} (\mathbf{\Sigma}[p'])^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \qquad \text{with} \qquad \mathbf{\Sigma}[p'] = \begin{pmatrix} \Sigma_{11} & \dots & \Sigma_{1p'} \\ \vdots & & \vdots \\ \Sigma_{p'1} & \dots & \Sigma_{p'p'} \end{pmatrix} \tag{2.24}$$

and $\Sigma_{i,j}$, $i, j = 1, \dots, p$, are the entries of $\mathbf{\Sigma}$. In the following $\mathbf{M}[s]$ is defined as the first $s \times s$ block matrix of a matrix $\mathbf{M}$. This gives the following limit distributions of the test statistics

$$\begin{aligned}
T_n^{(1)} &\xrightarrow{\mathcal{D}} \sup_{0 \le t \le 1} w^2(t) \sum_{j=1}^{p'} B_j^2(t), \\
T_n^{(2)} &\xrightarrow{\mathcal{D}} \int_0^1 w^2(t) \sum_{j=1}^{p'} B_j^2(t)\, dt, \\
T_n^{(3)} &\xrightarrow{\mathcal{D}} \sup_{0 \leqslant t_1 < t_2 \leqslant 1} \sum_{j=1}^{p'} (B_j(t_2) - B_j(t_1))^2, \\
T_n^{(4)} &\xrightarrow{\mathcal{D}} \iint_{0 \le t_1 < t_2 \le 1} \sum_{j=1}^{p'} (B_j(t_2) - B_j(t_1))^2\, dt_1\, dt_2,
\end{aligned}$$

where $\{B_j(t) : t \in [0, 1]\}$, $j = 1, \dots, p'$, are independent Brownian bridges. For information of the power behaviour we refer to Remark 2.11 (c).

In applications $\mathbf{\Sigma}$ has to be estimated, since it is not known. We use two different ideas for estimation. The first one is based on taking the alternative into account and the second one uses information about the data set.

In the correctly specified model it holds $\mathbf{\Sigma} = \mathbf{C}_1 \sigma^2$ (refer to Remark 2.5). The first idea is explained for the AMOC model and takes its alternative into account by estimating a possible change point beforehand. The variance estimator is defined by

$$\hat{\sigma}_{H1}^2 = \frac{1}{n} \left(\sum_{i=1}^{\hat{k}} (Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_{\hat{k}})^2 + \sum_{i=\hat{k}+1}^{n} (Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_{\hat{k}}^0)^2 \right),$$

where $\widehat{\mathbf{a}}_{\hat{k}}$ and $\widehat{\mathbf{a}}^0_{\hat{k}}$ are the least squares estimators based on $Y_1, \ldots, Y_{\hat{k}}$ and $Y_{\hat{k}+1}, \ldots, Y_n$ and $\hat{k}$ is a preliminary estimator of the change point given by

$$\hat{k} \quad = \quad \arg\max_{1 \le k \le n} w(k/n) \mathbf{S}_k^\top \begin{pmatrix} (\mathbf{C}_n[p'])^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{S}_k. \tag{2.25}$$

For the estimation of $\mathbf{C}_1$ we use

$$\mathbf{C}_{H1} \quad := \quad \frac{1}{2}\left(\mathbf{C}_{\hat{k}} + \frac{1}{n-\hat{k}} \sum_{i=\hat{k}+1}^{n} \mathbb{Y}_i \mathbb{Y}_i^\top \right)$$

and define an estimator for $\widetilde{\mathbf{A}}^{(p')}$ by

$$\widehat{\mathbf{C}}^{(p')}_{H1} \quad = \quad \hat{\sigma}^{-2}_{H1} \begin{pmatrix} (\mathbf{C}_{H1}[p'])^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

For the epidemic model it is possible to use a similar estimator, but in view of the data analysis we propose another estimator, which yields better results if more change points are present than the model assumes.

The data collection was designed in a way that no change is suspected before $i = 250$. We exploit this information by using only the first 250 observations $Y_1, \ldots, Y_{250}$ for estimation. Then, the variance estimator is given by

$$\hat{\sigma}^2_{250} \quad := \quad \frac{1}{250} \sum_{i=1}^{250} (Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_{250})^2,$$

where $\widehat{\mathbf{a}}_{250}$ is the least squares estimators based on $Y_1, \ldots, Y_{250}$. Moreover, we estimate $\mathbf{C}_1$ by $\mathbf{C}_{250}$. Then, we estimate $\widetilde{\mathbf{A}}^{(p')}$ by

$$\widehat{\mathbf{C}}^{(p')}_{250} \quad = \quad \hat{\sigma}^{-2}_{250} \begin{pmatrix} (\mathbf{C}_{250}[(p')])^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

In the misspecified AMOC model $\mathbf{\Sigma}$ can be estimated by

$$\mathbf{\Sigma}_{H1} \quad = \quad \widehat{\mathbf{\Gamma}}(0) + 2 \sum_{h=1}^{\Lambda_n} w_c(h/\Lambda_n) \widehat{\mathbf{\Gamma}}(h)$$

with

$$\widehat{\mathbf{\Gamma}}(h) \quad = \quad \frac{1}{n} \sum_{t=1}^{n-h} (\widetilde{\mathbf{x}}_t - \mu_n)(\widetilde{\mathbf{x}}_{t+h} - \mu_n)^\top$$

and

$$\widetilde{\mathbf{x}}_i = (Y_{i-1}\hat{\varepsilon}_i, \ldots, Y_{i-p}\hat{\varepsilon}_i)^\top, \qquad i = 1, \ldots, n,$$
$$\hat{\varepsilon}_i = \begin{cases} Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_{\hat{k}}, & i \le \hat{k}, \\ Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_{\hat{k}}^0, & i > \hat{k}, \end{cases}$$
$$\mu_n = \frac{1}{n}\sum_{i=1}^n \widetilde{\mathbf{x}}_i$$

and

$$\hat{k} = \arg\max_{1\le k\le n} w(k/n)\mathbf{S}_k^\top \begin{pmatrix} (\mathbf{\Sigma}_n[p'])^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{S}_k \tag{2.26}$$

with $\mathbf{\Sigma}_n$ as $\mathbf{\Sigma}_{H1}$, but based on residuals $\hat{\varepsilon}_i = Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_n$, $i = 1, \ldots, n$. Furthermore, we use the flat-top kernel

$$w_c(x) = \begin{cases} 1, & |x| \le 1/2, \\ 2(1 - |x|), & 1/2 < |x| < 1, \\ 0, & |x| > 1, \end{cases}$$

and bandwidth $\Lambda_n = 2 \max_{l,k=1,2} b_{l,k}$, where $b_{l,k}$ is the smallest positive integer such that

$$\left| \widehat{\Gamma}_{l,k}(b_{l,k} + j)/\sqrt{\widehat{\Gamma}_{l,l}(0)\widehat{\Gamma}_{k,k}(0)} \right| < 1.4\sqrt{\log_{10} n/n}, \qquad \text{for } j = 1, \ldots, 3.$$

This adaptive selection of the bandwidth is a variation of a procedure from Politis (2003). Then, we define the estimator for $\widetilde{\mathbf{A}}^{(p')}$ as

$$\widehat{\mathbf{S}}_{H1}^{(p')} = \begin{pmatrix} \left(\mathbf{\Sigma}_{H1}^{(p')}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

The other possibility to estimate $\mathbf{\Sigma}$ is again to use only observations $Y_1, \ldots, Y_{250}$. Then, we receive the estimator

$$\mathbf{\Sigma}_{250} = \widehat{\mathbf{\Gamma}}(0) + 2\sum_{h=1}^{\Lambda_n} w_c(h/\Lambda_n)\widehat{\mathbf{\Gamma}}(h)$$

with

$$\widehat{\mathbf{\Gamma}}(h) = \frac{1}{250}\sum_{t=1}^{250-h} (\widetilde{\mathbf{x}}_t - \mu_{250})(\widetilde{\mathbf{x}}_t - \mu_{250})^\top,$$

where we now estimate the errors by

$\hat{\varepsilon}_i = Y_i - \mathbb{Y}_i^\top \widehat{\mathbf{a}}_{250}, \qquad i = 1, \ldots, 250.$

Then, we define

$$\widehat{\mathbf{S}}_{250}^{(p')} = \begin{pmatrix} \left(\mathbf{\Sigma}_{250}^{(p')}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

2.2. Multivariate Models and Methods

In the multivariate setting more than one time series is observed. If (almost) all of them include a change at the (at least approximately) same time point, the information can be pooled and the power of the test increased.

Section 2.2.1 introduces the multivariate models. Moreover, in Section 2.2.2 we define the appended test statistics and derive their asymptotic null distributions. Section 2.2.3 proves the consistency of the tests under a large class of alternatives. Afterwards, in Section 2.2.4, we introduce the change-point estimator(s) and show their consistency. Section 2.2.5 includes some discussion of specific test statistics.

2.2.1. At-Most-One-Change and Epidemic Model

We consider the following multivariate AMOC model, where $i = -p+1, \ldots, n$ are the observed time points and $l = 1, \ldots, d$ denote the components of the multivariate model:

$$Y_i(l) = \begin{cases} Y_i^{(1)}(l), & 1 \le i \le \tilde{k}, \\ Y_i^{(2)}(l), & \tilde{k} < i \le n, \end{cases}$$

and $\{Y_i^{(1)}(l)\}$, $\{Y_i^{(2)}(l)\}$ are stationary processes and differ at least in one component in distribution. As in the univariate case we approximate the observations $\{Y_i(l)\}$ by a linear autoregressive model of order p:

$$Y_i(l) = \begin{cases} \mathbb{Y}_i^{(1)\top}(l)\mathbf{a}_1(l) + \varepsilon_i^{(1)}(l), & 1 < i \le \tilde{k}, \\ \mathbb{Y}_i^{(2)\top}(l)\mathbf{a}_2(l) + \varepsilon_i^{(2)}(l), & \tilde{k} < i \le n, \end{cases} \tag{2.27}$$

with

$$\mathbb{Y}_i(l) = (Y_{i-1}(l), \ldots, Y_{i-p}(l))^\top, \qquad i = 1, \ldots, n,$$

errors

$$\begin{aligned} \varepsilon_i^{(1)}(l) &:= Y_i^{(1)}(l) - \mathbb{Y}_i^{(1)\top}(l)\mathbf{a}_1(l) \qquad \text{for } i = 1, \ldots, \tilde{k}, \\ \varepsilon_i^{(2)}(l) &:= Y_i^{(2)}(l) - \mathbb{Y}_i^{(2)\top}(l)\mathbf{a}_2(l) \qquad \text{for } i = \tilde{k}+1, \ldots, n, \end{aligned} \tag{2.28}$$

parameters

$$\mathbf{a}_1(l) := \mathbf{C}_1^{-1}(l)\boldsymbol{\gamma}_1(l) \quad \text{and} \quad \mathbf{a}_2(l) := \mathbf{C}_2^{-1}(l)\boldsymbol{\gamma}_2(l),$$

where

$$\mathbf{C}_1(l) = E\left[\mathbb{Y}_i^{(1)}(l)\mathbb{Y}_i^{(1)\top}(l)\right] \quad \text{and} \quad \mathbf{C}_2(l) = E\left[\mathbb{Y}_i^{(2)}(l)\mathbb{Y}_i^{(2)\top}(l)\right]$$

are assumed to be invertible, and

$$\boldsymbol{\gamma}_1(l) = E\left[\mathbb{Y}_i^{(1)}(l)Y_i^{(1)}(l)\right] \quad \text{and} \quad \boldsymbol{\gamma}_2(l) = E\left[\mathbb{Y}_i^{(2)}(l)Y_i^{(2)}(l)\right].$$

The change point $\tilde{k} = \lfloor \lambda n \rfloor$ with $0 < \lambda \leq 1$ is unknown and does not depend on the component $l = 1, \ldots, d$.

To facilitate the readability we define

$$\mathbf{Y}_i \quad := \quad (Y_i(1), \ldots, Y_i(d))^\top.$$

The multivariate epidemic change-point model is given by

$$Y_i(l) \quad = \quad \begin{cases} Y_i^{(1)}(l), & 1 \leq i \leq \tilde{k}_1, \\ Y_i^{(2)}(l), & \tilde{k}_1 < i \leq \tilde{k}_2, \\ Y_i^{(1)}(l), & \tilde{k}_2 < i \leq n, \end{cases}$$

for $l = 1, \ldots, d$ and change points $\tilde{k}_1 = \lfloor \lambda_1 n \rfloor$, $\tilde{k}_2 = \lfloor \lambda_2 n \rfloor$ with $0 < \lambda_1 \leq \lambda_2 \leq 1$. The processes $\{Y_i^{(1)}(l)\}$ and $\{Y_i^{(2)}(l)\}$ are stationary processes and differ at least in one component in distribution. Then, we approximate $\{Y_i(l)\}$ by a linear autoregressive model

$$Y_i(l) \quad = \quad \begin{cases} \mathbb{Y}_i^{(1)\top}(l)\mathbf{a}_1(l) + \varepsilon_i^{(1)}(l), & 1 < i \leq \tilde{k}_1, \\ \mathbb{Y}_i^{(2)\top}(l)\mathbf{a}_2(l) + \varepsilon_i^{(2)}(l), & \tilde{k}_1 < i \leq \tilde{k}_2, \\ \mathbb{Y}_i^{(1)\top}(l)\mathbf{a}_1(l) + \varepsilon_i^{(1)}(l), & \tilde{k}_2 < i \leq n, \end{cases}$$

using the same notation and interpretation as above.

2.2.2. General Statistics and Null Asymptotics

To test under the AMOC model the null hypothesis H_0 of no change against the alternative H_1 of exactly one change we use the statistics

$$M_n^{(1)} \quad = \quad \max_{1 \leqslant k \leqslant n} \frac{w^2(k/n)}{n} \mathbf{Z}_k^\top \mathbf{H} \mathbf{Z}_k,$$

$$M_n^{(2)} \quad = \quad \sum_{1 \leqslant k \leqslant n} \frac{w^2(k/n)}{n^2} \mathbf{Z}_k^\top \mathbf{H} \mathbf{Z}_k$$

with

$$\begin{aligned}
\mathbf{Z}_k^\top &:= \left(S_k^\top(1), \ldots, S_k^\top(d)\right), \qquad S_k(l) := \sum_{i=1}^{k} \mathbb{Y}_i(l)\hat{\varepsilon}_i(l),\\
\hat{\varepsilon}_i(l) &:= Y_i(l) - \mathbb{Y}_i^\top(l)\widehat{\mathbf{a}}_n(l),\\
\widehat{\mathbf{a}}_n(l) &:= \mathbf{C}_n^{-1}(l)\sum_{i=1}^{n} \mathbb{Y}_i(l)Y_i(l), \qquad \mathbf{C}_n(l) := \sum_{t=1}^{n} \mathbb{Y}_i(l)\mathbb{Y}_i^\top(l),
\end{aligned}$$

and $\mathbf{H}$ is a symmetric, positive semidefinite $pd \times pd$ matrix. The next theorem gives the asymptotic null distributions of the test statistics $M_n^{(1)}$ and $M_n^{(2)}$.

Theorem 2.14. Let the process $\{\mathbf{Y}_i\}$ fulfill **(Y)** and $\{\varepsilon_i(l)\}$, $l = 1, \ldots, d$, be defined as in (2.28). Moreover, let the weight function $w(\cdot)$ fulfill **(W)**. Then, under H_0,

(a) $$M_n^{(1)} \xrightarrow{\mathcal{D}} \sup_{0 \le t \le 1} w^2(t) \sum_{j=1}^{pd} B_j^2(t),$$

(b) $$M_n^{(2)} \xrightarrow{\mathcal{D}} \int_0^1 w^2(t) \sum_{j=1}^{pd} B_j^2(t)\, dt,$$

where $\{\mathbf{B}(t) = (B_1(t), \ldots, B_{dp}(t))^\top : t \in [0,1]\}$ denotes a pd-dimensional Brownian bridge with covariance matrix $\mathbf{H}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{H}^{\frac{1}{2}}$. $\mathbf{\Omega}$ is the long-run autocovariance matrix of $\{\left(\mathbb{Y}_i^\top(1)\varepsilon_i(1), \ldots, \mathbb{Y}_i^\top(d)\varepsilon_i(d)\right)^\top\}$ and $\mathbf{H}$ is a symmetric, positive semidefinite matrix, which can be replaced by a consistent estimator.

Proof of Theorem 2.14. The proof is a component-by-component application of the proof of Theorem 2.6, since $\mathbf{Z}_k^\top = \left(S_k^\top(1), \ldots, S_k^\top(d)\right)$. It holds

$$M_n^{(1)} = \max_{1 \le k \le n} \frac{w^2(k/n)}{n} \left\| \mathbf{H}^{\frac{1}{2}} \mathbf{Z}_k \right\|^2$$

and we receive for each $S_k(j)$, $j = 1, \ldots, d$, the same decomposition as in (2.9). Hence, we obtain with $\widetilde{\mathbf{Z}}_k = \left(\sum_{i=1}^{k} \mathbb{Y}_i^\top(1)\varepsilon_i(1), \ldots, \sum_{i=1}^{k} \mathbb{Y}_i^\top(d)\varepsilon_i(d)\right)^\top$ that

$$M_n^{(1)} = \max_{1 \le k \le n} \frac{w^2(k/n)}{n} \left\| \mathbf{H}^{\frac{1}{2}} \left(\widetilde{\mathbf{Z}}_k - \frac{k}{n}\widetilde{\mathbf{Z}}_n \right) \right\|^2 + o_p(1),$$

since (2.11) holds for each component $j = 1, \ldots, d$. Then, we can apply the invariance principle of Kuelbs and Philipp (1980) and the continuous mapping theorem as in (2.12) and conduct component-by-component the further calculations as in the proof of Theorem 2.6 to receive the asymptotic distribution. □

For the epidemic model we define the test statistics

$$
\begin{aligned}
M_n^{(3)} &:= \max_{1 \leq k_1 < k_2 \leq n} w(k_1/n, k_2/n) \frac{1}{n} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1})^\top \mathbf{H} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1}), \\
M_n^{(4)} &:= \sum_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) \frac{1}{n^3} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1})^\top \mathbf{H} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1})
\end{aligned}
$$

with weight function $w(\cdot, \cdot)$ as in (2.16). In the next theorem we obtain the asympotic null distributions.

Theorem 2.15. Let the process $\{\mathbf{Y}_i\}$ fulfill **(Y)** and $\{\varepsilon_i(l)\}$, $l = 1, \dots, d$, be defined as in (2.28). Then, under H_0,

(a) $M_n^{(3)} \xrightarrow{\mathcal{D}} \sup_{0 \leq t_1 < t_2 \leq 1} w(t_1, t_2) \sum_{j=1}^{dp} (B_j(t_2) - B_j(t_1))^2,$

(b) $M_n^{(4)} \xrightarrow{\mathcal{D}} \iint_{0 \leq t_1 < t_2 \leq 1} w(t_1, t_2) \sum_{j=1}^{dp} (B_j(t_2) - B_j(t_1))^2 dt_1\, dt_2,$

where $\{\mathbf{B}(t) = (B_1(t), \dots, B_{dp}(t))^\top : t \in [0,1]\}$ denotes a pd-dimensional Brownian bridge with covariance matrix $\mathbf{H}^{\frac{1}{2}} \mathbf{\Omega} \mathbf{H}^{\frac{1}{2}}$. $\mathbf{\Omega}$ is the long-run autocovariance matrix of $\{\left(\mathbb{Y}_i^\top(1)\varepsilon_i(1), \dots, \mathbb{Y}_i^\top(d)\varepsilon_i(d)\right)^\top\}$ and $\mathbf{H}$ is a symmetric, positive semidefinite matrix, which can be replaced by a consistent estimator.

Proof of Theorem 2.15. The proof follows from the results in the proof of Theorem 2.7. With $\widetilde{\mathbf{Z}}_k$ defined as in the proof of Theorem 2.14 we receive

$$
\begin{aligned}
&\max_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) \frac{1}{n} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1})^\top \mathbf{H} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1}) \\
&= \max_{1 \leqslant k_1 < k_2 \leqslant n} w(k_1/n, k_2/n) \frac{1}{n} \left\| \mathbf{H}^{\frac{1}{2}} \left(\widetilde{\mathbf{Z}}_{k_2} - \frac{k_2}{n} \widetilde{\mathbf{Z}}_n - \left(\widetilde{\mathbf{Z}}_{k_1} - \frac{k_1}{n} \widetilde{\mathbf{Z}}_n \right) \right) \right\|^2 + o_p(1),
\end{aligned}
$$

since (2.17) holds for each component $j = 1, \dots, d$. To receive the asymptotic distribution we use Lemma 2.4, the continous mapping theorem and the a.s. continuity of the Brownian bridge. Part (b) follows analogously. □

2.2.3. Behaviour under Alternatives

As in the univariate case we show the consistency of the tests.

Theorem 2.16. Let $\{\mathbf{Y}_i^{(1)}\}$, $\{\mathbf{Y}_i^{(2)}\}$ fulfill **(Y)**. Further, we assume

$$\mathbf{H}^{\frac{1}{2}} \begin{pmatrix} \gamma_1(1) - \mathbf{C}_1(1)\widetilde{\mathbf{a}}(1) \\ \vdots \\ \gamma_1(d) - \mathbf{C}_1(d)\widetilde{\mathbf{a}}(d) \end{pmatrix} \neq \mathbf{0}$$

with a symmetric, positive semidefinite matrix $\mathbf{H}$, which can also be replaced by an estimator $\widehat{\mathbf{H}}_n$ with $\|\widehat{\mathbf{H}}_n^{\frac{1}{2}}(\boldsymbol{\gamma}_1 - \mathbf{C}_1\widetilde{\mathbf{a}})\| \geq \varepsilon > 0$ and $\|\widehat{\mathbf{H}}_n^{\frac{1}{2}}\| \leq C < \infty$ for all n.

(a) Moreover, let the weight function $w(\cdot)$ fulfill **(W)** and $w \not\equiv 0$. Then, under the AMOC alternative,

$$M_n^{(1)} \xrightarrow{P} \infty \quad \text{and} \quad M_n^{(2)} \xrightarrow{P} \infty.$$

(b) Let the weight function $w(\cdot,\cdot)$ be defined as in (2.16). Then, under the epidemic change alternative,

$$M_n^{(3)} \xrightarrow{P} \infty \quad \text{and} \quad M_n^{(4)} \xrightarrow{P} \infty.$$

Proof of Theorem 2.16. The proofs follow as in Theorem 2.10. It exists a continuity point t_0 with $w(t_0) > 0$ and $k_0 = \lfloor t_0 n \rfloor$. Then, we obtain with Lemma 2.9 (a)

$$\begin{aligned} \max_{1\leq k\leq n} \frac{w^2(k/n)}{n}\mathbf{Z}_k^\top \mathbf{H}\mathbf{Z}_k &= \max_{1\leq k\leq n} \frac{w^2(k/n)}{n} \left\|\mathbf{H}^{\frac{1}{2}}\mathbf{Z}_k\right\|^2 \geq \frac{w^2(k_0/n)}{n} \left\|\mathbf{H}^{\frac{1}{2}}\mathbf{Z}_{k_0}\right\|^2 \\ &= n w^2(t_0)\tilde{g}^2(t_0) \left\| \mathbf{H}^{\frac{1}{2}} \begin{pmatrix} (\boldsymbol{\gamma}_1(1) - \mathbf{C}_1(1)\widetilde{\mathbf{a}}(1)) \\ \vdots \\ (\boldsymbol{\gamma}_1(d) - \mathbf{C}_1(d)\widetilde{\mathbf{a}}(d)) \end{pmatrix} \right\|^2 + o_p(n) \end{aligned}$$

with $\tilde{g}^2(t)$ defined as in (2.23). Then, the result for $M_n^{(1)}$ follows. The other results are obtained with the same arguments as in Theorem 2.10. □

2.2.4. Change-Point Estimators

For the AMOC model we define the change-point estimator as

$$\hat{k}_n := \arg\max_{1<k<n} w^2(k/n)\mathbf{Z}_k^\top \mathbf{H}\mathbf{Z}_k \tag{2.29}$$

and in the epidemic model we define

$$(\hat{k}_{1,n}, \hat{k}_{2,n}) := \arg\max_{1\leq k_1<k_2\leq n} w(k_1/n, k_2/n)(\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1})^\top \mathbf{H}(\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1}).$$

In the next theorem we obtain the consistency of these estimators.

Theorem 2.17. Let the assumptions of Theorem 2.16 hold.

(a) Moreover, let the weight function be of the form

$$w^2(t) = \omega^2(t)\mathcal{I}_{l_1 \le t \le l_2}(t) \quad \text{for} \quad t \in [0,1] \quad \text{and} \quad 0 \le l_1 < l_2 \le 1$$

with $\omega(\cdot)$ continuous and bounded on $[l_1, l_2]$. Further, let $w(t)\tilde{g}(t)$ have a unique maximum at λ with $\tilde{g}(\cdot)$ as in (2.23). Then, under the AMOC change alternative,

$$\frac{\hat{k}_n}{n} \xrightarrow{P} \lambda.$$

(b) Moreover, let $[l_1, l_2]^2$ include the change point (λ_1, λ_2). Then, under the epidemic alternative,

$$\left(\frac{\hat{k}_{1,n}}{n}, \frac{\hat{k}_{2,n}}{n}\right) \xrightarrow{P} (\lambda_1, \lambda_2).$$

Proof of Theorem 2.17. The proof follows from Lemma 2.9 as in the proof of Theorem 2.12. □

2.2.5. Specific Test Statistics and Estimators

The matrix $\mathbf{\Omega}$ is the long-run autocovariance matrix of $\{\left(\mathbb{Y}_i^\top(1)\varepsilon_i(1), \ldots, \mathbb{Y}_i^\top(d)\varepsilon_i(d)\right)^\top\}$ and the choice $\mathbf{\Omega}^{-1}$ for the matrix $\mathbf{H}$ leads to independent Brownian bridges in the asymptotic distribution. If the components of the observations are independent, the matrix $\mathbf{\Omega}$ reduces to a block diagonal matrix with d blocks $\mathbf{\Sigma}(l)$, $l = 1, \ldots, d$, which correspond to the long-run autocovariance matrices of $\{\mathbb{Y}_i(l)\varepsilon_i(l)\}$, $l = 1, \ldots, d$, respectively. Hence, if the channels are independent the multivariate statistics add up the univariate ones.

As in the univariate case (refer to Section 2.1.5) we consider matrices

$$\widetilde{\mathbf{A}}^{(p')}(l) = \begin{pmatrix} (\mathbf{\Sigma}(l)[p'])^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{with} \quad \mathbf{\Sigma}(l)[p'] = \begin{pmatrix} \Sigma_{11}(l) & \ldots & \Sigma_{1p'}(l) \\ \vdots & & \vdots \\ \Sigma_{p'1}(l) & \ldots & \Sigma_{p'p'}(l) \end{pmatrix}$$

and $\Sigma_{i,j}(l)$, $i,j = 1, \ldots, p$, are the entries of $\mathbf{\Sigma}(l)$. Then, we use for $\mathbf{H}$ the matrix $\widetilde{\mathbf{H}}^{(p')} = diag(\widetilde{\mathbf{A}}^{(p')}(1), \ldots, \widetilde{\mathbf{A}}^{(p')}(d))$, which leads to a sum of dp' independent Brownian bridges in the limit. To estimate the block matrices we can use the estimators introduced in Section 2.1.5: $\widehat{\mathbf{C}}_{H1}^{(p')}$, $\widehat{\mathbf{C}}_{250}^{(p')}$ and $\widehat{\mathbf{\Sigma}}_{H1}^{(p')}$, $\widehat{\mathbf{\Sigma}}_{250}^{(p')}$.

In the situation of dependent components, $\mathbf{\Omega}$ is very difficult to estimate and the estimation errors might be large, in particular considering that we then need the inverse. Hence, in our data analysis we only use an approximation by the diagonal block matrix $\widetilde{\mathbf{H}}$. But then the Brownian bridges in the limit are no longer independent but inherit the covariance structure between the components. Since we can not estimate this covariance structure very well (otherwise we would have estimated it in the first place), we can no longer use the correct asymptotic distribution to obtain critical values. Therefore we must use bootstrap methods as described in Section 3.2.

Using the multivariate statistics with a diagonal block matrix **H** means essentially that we do take into account the dependence between the d tests for one component, but decide that in comparison the dependence between components is rather weak and can to some extend be neglected. If we believe it is not completely negligible, we use an asymptotic correction by the bootstrap.

3. Simulation Study

The simulation study analyses the behaviour of the test statistics for smaller sample sizes. In particular it considers the actual size and power of the tests. In view of the data analysis we consider a similar setting and discuss the size and power behaviour of the test statistics in context of the AMOC and epidemic model in the correctly and misspecified situation. We compare the behaviour of the maximum- and sum-type statistics and analyse the influence of the in Section 2.1.5 introduced estimators $\widehat{\mathbf{C}}_{H1}^{(p')}$, $\widehat{\mathbf{C}}_{250}^{(p')}$, $\widehat{\mathbf{S}}_{H1}^{(p')}$ and $\widehat{\mathbf{S}}_{250}^{(p')}$.

To analyse the size and power behaviour we use Size Power Curves (SPCs) which are plots showing the actual size and power on the y-axis and the nominal size on the x-axis. These are created by simulating a certain number of time series under the null hypothesis or respectively under the alternative, calculating the p-values regarding their asymptotic or bootstrap distribution and then plotting the empirical distribution of the p-values. In situations where the nominal and actual size differ a lot it is difficult to compare the power of the test statistics in these plots. Hence, we further plot the actual power against the actual size.

Regarding the data set we have information of the type of change, i.e. the change is expected to be in the first two autocorrelations. Hence, in Section 3.1 we discuss the specification of the test statistic to receive a good power under this type of alternative. Sometimes bootstrap procedures can improve the performance of the tests in the sense that the level is better hold or the power is improved. Section 3.2 introduces a bootstrap and we discuss it further in Section 3.3 together with the univariate asymptotic tests. Section 3.4 shortly discusses some simulations in context of multivariate models.

3.1. Specification of the Test Statistics

We construct the simulation study similar to the general conditions of the data set and use information about the data to optimise our methods. Since the data consists of time series with $n = 512$ observations, we generate samples of size $n = 512$. By the information about how the data set is obtained no change before observation 250 is suspected. Hence, we use a non symmetrical weight function and reduce the interval where we look for a change to $[250, 490]$. With $\gamma = 0.25$ and weight function

$$w(t) \quad = \quad \mathcal{I}_{\left[\frac{250}{512},\frac{490}{512}\right]}(t)(1-t)^{-\gamma}, \qquad t \in [0,1].$$

we consider under the AMOC model the test statistics

$$T^{(1)} \quad = \quad \max_{250 \leq k \leq 490} \frac{(1-k/n)^{-2\gamma}}{n} \mathbf{S}_k^\top \widetilde{\mathbf{A}}^{(p')} \mathbf{S}_k \xrightarrow{\mathcal{D}} \sup_{250/512 \leq t \leq 490/512} (1-t)^{-2\gamma} \sum_{j=1}^{p'} B_j^2(t),$$

$$T^{(2)} \quad = \quad \frac{1}{n} \sum_{250 \leqslant k \leqslant 490} \frac{(1-k/n)^{-2\gamma}}{n} \mathbf{S}_k^\top \widetilde{\mathbf{A}}^{(p')} \mathbf{S}_k \xrightarrow{\mathcal{D}} \int_{250/512}^{490/512} (1-t)^{-2\gamma} \sum_{j=1}^{p'} B_j^2(t),$$

with $\{\mathbf{B}_j(t) : t \in [0,1]\}$, $j = 1, \ldots, p'$, are independent Brownian bridges for $\widetilde{\mathbf{A}}^{(p')}$ defined as in (2.24).

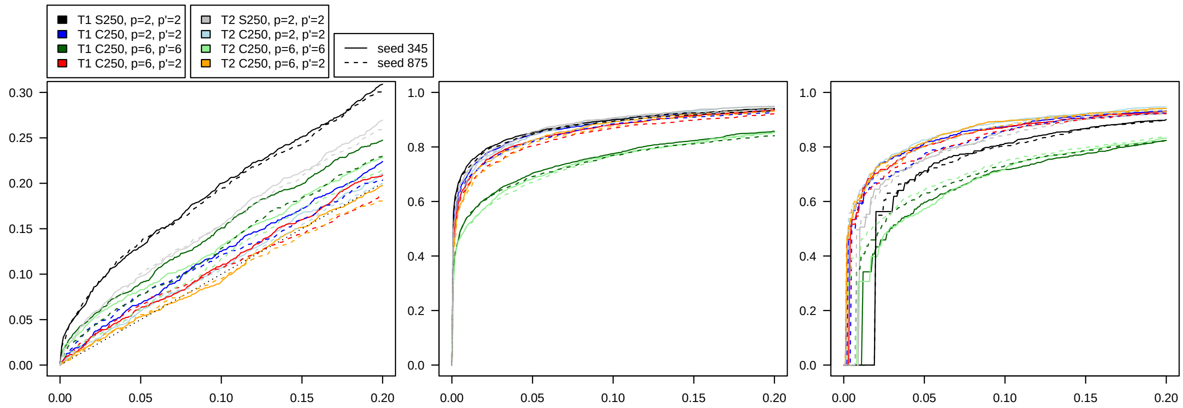

Figure 3.1.: SPCs under the AR1 model.

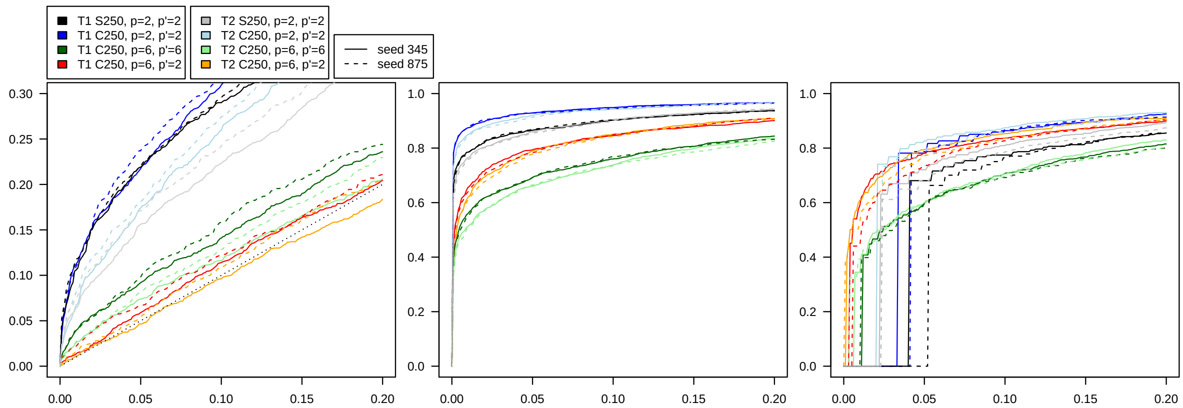

Figure 3.2.: SPCs under the AR2 model.

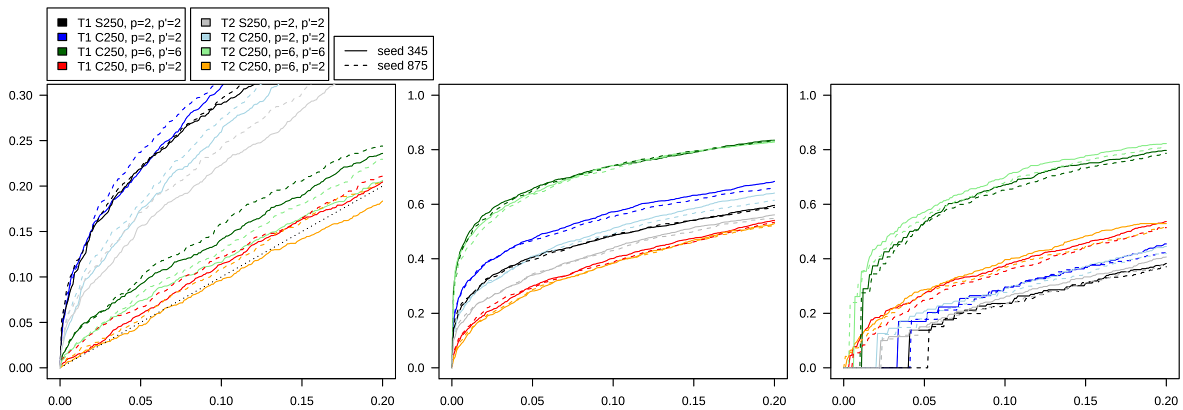

Figure 3.3.: SPCs under the AR3 model.

In the data set the changes are expected to be in the first two autocorrelations, and we want to apply a test which has a high power under this kind of alternative. We consider three options: first we fit an AR(2) model and test with the matrix $\widetilde{\mathbf{A}}^{(2)}$ ($p = 2$, $p' = 2$), second we fit an AR(6) model and test with a matrix $\widetilde{\mathbf{A}}^{(6)}$ ($p = 6$, $p' = 6$) and third we fit an AR(6) model and test with $\widetilde{\mathbf{A}}^{(2)}$ ($p = 6$, $p' = 2$).

These tests are considered under different AR(6) models:

$$Y_i \quad = \quad \begin{cases} \mathbb{Y}_i^\top \boldsymbol{\varphi}_1 + \varepsilon_i, & i \leq \tilde{k}, \\ \mathbb{Y}_i^\top \boldsymbol{\varphi}_2 + \varepsilon_i, & i > \tilde{k}, \end{cases} \tag{3.1}$$

with a change at $\tilde{k} = 330$, standard normal distributed errors $\varepsilon_1, \ldots, \varepsilon_n$ and parameters

- AR1: $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$, $\boldsymbol{\varphi}_2 = (-0.4, 0.1, 0, 0, -0.1, -0.1)^\top$,
- AR2: $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0.2, -0.3, -0.2)^\top$,$\boldsymbol{\varphi}_2 = (-0.2, 0.3, 0, 0.2, -0.3, -0.2)^\top$,
- AR3: $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0.2, -0.3, -0.2)^\top$, $\boldsymbol{\varphi}_2 = (-0.1, 0.2, 0, 0.4, -0.1, -0.2)^\top$.

In case we fit an AR(6) model, we are in the correctly specified situation and use estimators $\widehat{\mathbf{C}}_{250}^{(6)}$ and $\widehat{\mathbf{C}}_{250}^{(2)}$. If we fit an AR(2) model the time series is misspecified and we use estimator $\widehat{\mathbf{S}}_{250}^{(2)}$ but as well $\widehat{\mathbf{C}}_{250}^{(2)}$, since the estimation error of $\widehat{\mathbf{S}}_{250}$ might be larger than the error we make by approximating $\widetilde{\mathbf{A}}^{(2)}$ by $\widehat{\mathbf{C}}_{250}^{(2)}$. Throughout the section the legends follow the notation: T1, T2 correspond to the statistics $T^{(1)}$, $T^{(2)}$ and CH1, C250 to the estimators $\widehat{\mathbf{C}}_{H1}^{(p')}$, $\widehat{\mathbf{C}}_{250}^{(p')}$ and SH1, S250 to the estimators $\widehat{\mathbf{S}}_{H1}^{(p')}$, $\widehat{\mathbf{S}}_{250}^{(p')}$.

Figures 3.1 to 3.3 include the SPCs under models AR1 to AR3. These are computed by $N = 2000$ repetions. We start to consider the size curves, which are in the first plot, respectively. In all three situations the test with $p = 6$ and $p' = 2$ held the level best. The test with $p = 6$ and $p' = 6$ works not as well since we use the full estimated autocovariance matrix and hence have more estimation errors included. The test statistic with $p = 2$ and $p' = 2$ works fine if the level of misspecification is low as in the AR1 model, but if the misspecification is high it fails to hold the level. In theory it should work if estimator $\widehat{\mathbf{S}}_{250}^{(2)}$ is used, but the long-run autocovariance is in general hard to estimate and hence the estimation errors are large, which leads to the poor performance of the test with $p = 2$ and $p' = 2$.

The power behaviour of the tests is presented in plot two and three of Figures 3.1 to 3.3. In the second plot the actual power against the nominal size is plotted and in the third plot the actual power against the actual size. In the latter the power can better be compared if the actual sizes differ a lot. In the first two models the change is present in the first two coefficients and in the third the change mainly takes place in the higher coefficients. Whereas in the first two situations the power is clearly higher for the test statistics with $p' = 2$ the opposite holds for the last situation. Hence, if we suspect a change in the first two coefficients a test with p' detects it more likely. But if we use such a test and the changes are somewhere else we loose power. The tests with p' have a similar power, but since the test with $p = 2$ is rather unstable, we can see the test with $p = 6$ as a robustification of it and apply this one. For the data set we have to figure out which order p for the linear autoregressive model to take. But this example shows that the order should not be too small.

In the sequel, we concentrate on test statistics with matrix $\widetilde{\mathbf{A}}^{(2)}$. For the univariate case we consider $T^{(1)}$, $T^{(2)}$ under the AMOC model and $T^{(3)}$, $T^{(4)}$ under the epidemic model:

$$T^{(1)} := \max_{250\leq k\leq 490} \frac{(1-k/n)^{-\gamma}}{n} \mathbf{S}_k^\top \widetilde{\mathbf{A}}^{(2)} \mathbf{S}_k \xrightarrow{\mathcal{D}} \sup_{\frac{250}{512}\leq t\leq \frac{490}{512}} (1-t)^{-\gamma} \sum_{j=1}^{2} B_j^2(t), \tag{3.2}$$

$$T^{(2)} := \frac{1}{n} \sum_{250\leqslant k\leqslant 490} \frac{(1-k/n)^{-\gamma}}{n} \mathbf{S}_k^\top \widetilde{\mathbf{A}}^{(2)} \mathbf{S}_k \xrightarrow{\mathcal{D}} \int_{\frac{250}{512}}^{\frac{490}{512}} (1-t)^{-\gamma} \sum_{j=1}^{2} B_j^2(t), \tag{3.3}$$

$$T^{(3)} := \max_{250\leqslant k_1\leq k_2\leq 490} \frac{1}{n} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1})^\top \widetilde{\mathbf{A}}^{(2)} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1}) \tag{3.4}$$

$$\xrightarrow{\mathcal{D}} \sup_{\frac{250}{512}\leq t_1 < t_2 \leq \frac{490}{512}} \sum_{j=1}^{2} (B_j(t_2) - B_j(t_1))^2,$$

$$T^{(4)} := \frac{1}{n^3} \sum_{250\leq k_1 < k_2 \leq 490} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1})^\top \widetilde{\mathbf{A}}^{(2)} (\mathbf{S}_{k_2} - \mathbf{S}_{k_1}) \tag{3.5}$$

$$\xrightarrow{\mathcal{D}} \int\int_{\frac{250}{512}\leq t_1 < t_2 \leq \frac{490}{512}} \sum_{j=1}^{2} (B_j(t_2) - B_j(t_1))^2 \, dt_1 \, dt_2,$$

where $\{\mathbf{B}_j(t) : t \in [0,1]\}$, $j = 1, 2$, are independent Brownian bridges.

Further, we consider multivariate models with dimension $d = 12$ and use the test statistics $M^{(1)}$, $M^{(2)}$ under the AMOC model and $M^{(3)}$, $M^{(4)}$ under the epidemic model:

$$M^{(1)} := \max_{250\leqslant k\leqslant 490} \frac{(1-k/n)^{-\gamma}}{n} \mathbf{Z}_k^\top \widetilde{\mathbf{H}}^{(2)} \mathbf{Z}_k \xrightarrow{\mathcal{D}} \sup_{\frac{250}{512}\leq t\leq \frac{490}{512}} w^2(t) \sum_{j=1}^{24} B_j^2(t), \tag{3.6}$$

$$M^{(2)} := \sum_{250\leqslant k\leqslant 490} \frac{(1-k/n)^{-\gamma}}{n^2} \mathbf{Z}_k^\top \widetilde{\mathbf{H}}^{(2)} \mathbf{Z}_k \xrightarrow{\mathcal{D}} \int_{\frac{250}{512}}^{\frac{490}{512}} w^2(t) \sum_{j=1}^{24} B_j^2(t) \, dt, \tag{3.7}$$

$$M^{(3)} := \max_{250\leqslant k_1\leq k_2\leq 490} \frac{1}{n} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1})^\top \widetilde{\mathbf{H}}^{(2)} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1}) \tag{3.8}$$

$$\xrightarrow{\mathcal{D}} \sup_{\frac{250}{512}\leq t_1 < t_2 \leq \frac{490}{512}} \sum_{j=1}^{24} (B_j(t_2) - B_j(t_1))^2,$$

$$M^{(4)} := \sum_{250\leqslant k_1\leq k_2\leq 490} \frac{1}{n^3} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1})^\top \widetilde{\mathbf{H}}^{(2)} (\mathbf{Z}_{k_2} - \mathbf{Z}_{k_1}) \tag{3.9}$$

$$\xrightarrow{\mathcal{D}} \int\int_{\frac{250}{512}\leq t_1 < t_2 \leq \frac{490}{512}} \sum_{j=1}^{24} (B_j(t_2) - B_j(t_1))^2 \, dt_1 \, dt_2,$$

where $\{\mathbf{B}(t) = (B_1(t), \ldots, B_{24}(t))^\top : t \in [0,1]\}$ denotes a 24-dimensional Brownian bridge and $\widetilde{\mathbf{H}}^{(2)} = \operatorname{diag}\left(\widetilde{\mathbf{A}}^{(2)}(1), \ldots, \widetilde{\mathbf{A}}^{(2)}(12)\right)$ is defined in Section (2.2.5). If the components are independent, the Brownian bridges are independent.

Since the distributions of maximums and sums of Brownian bridges are not explicitly known, we approximate them by 10000 simulated samples.

3.2. Bootstrap

For smaller sample sizes the distribution of the test statistic is sometimes not well approximated by the asymptotic distribution and the use of bootstrap procedures might lead to a better result. We discuss a version of the pair bootstrap from Hušková et al. (2008). First, we give the general algorithm :

Choose a block length K such that $n = KL$.

(1) Define $\widetilde{\mathbb{Y}}_i(j) = \big(Y_{i-1}(j), \ldots, Y_{i-p'}(j)\big)^\top$, $i = 1, \ldots, n$, $j = 1, \ldots, d$, and calculate residuals $\hat{\varepsilon}_1(j), \ldots, \hat{\varepsilon}_{\tilde{n}}(j)$, $i = 1, \ldots, \tilde{n}$, by one of the methods explained below and $\tilde{n}$ depends on the estimation method.

(2) Draw L i.i.d. random variables $U_1, \ldots, U_L$ such that $P(U_1 = i) = 1/\tilde{n}$, $i = 1, \ldots, \tilde{n} - 1$.

(3) Let for $k = 1, \ldots, K$, $l = 1, \ldots, L$, $j = 1, \ldots, d$

$$\begin{aligned} \mathbb{Y}^*_{K(l-1)+k}(j) &:= \widetilde{\mathbb{Y}}_{U(l)+k}(j), \qquad \text{where } \mathbb{Y}^*_i = \widetilde{\mathbb{Y}}_{i-\tilde{n}} \text{ if } i > \tilde{n}, \\ \varepsilon^*_{K(l-1)+k}(j) &:= \hat{\varepsilon}_{U(l)+k}(j), \qquad \text{where } \varepsilon^*_i = \hat{\varepsilon}_{i-\tilde{n}} \text{ if } i > \tilde{n}, \end{aligned}$$

and $\mathbb{X}^*_i := \big(\mathbb{Y}^{*\top}_i(1)\varepsilon^*_i(1), \ldots, \mathbb{Y}^{*\top}_i(d)\varepsilon^*_i(d)\big)^\top$, $i = 1, \ldots, \tilde{n}$.

(4) Calculate the multivariate statistics in the same way as above, but with $\mathbf{Z}_k$ replaced by $\mathbf{Z}^*_k$ and $\widetilde{\mathbf{H}}^{(p')}$ by $\boldsymbol{\Omega}^*_n$ where

$$\mathbf{Z}^*_k = \sum_{i=1}^{k} \left(\mathbb{X}^*_i - \frac{1}{n} \sum_{i=1}^{n} \mathbb{X}^*_i \right)$$

and $\boldsymbol{\Omega}^*_n$ is a block diagonal matrix with $j = 1, \ldots, d$ blocks $(\mathbf{C}^*_n(j))^{-1}$ or $(\mathbf{S}^*_n(j))^{-1}$, where

$$\mathbf{C}^*_n(j) = \frac{1}{n} \sum_{l=1}^{n} \varepsilon^{*2}_l(j) \frac{1}{n} \sum_{i=1}^{n} \mathbb{Y}^*_i(j) \mathbb{Y}^{*\top}_i(j)$$

or

$$\mathbf{S}^*_n(j) = \frac{1}{n} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} \mathbb{Y}^*_{lK+k}(j) \varepsilon^*_{lK+k}(j) \right) \left(\sum_{k=1}^{K} \mathbb{Y}^*_{lK+k}(j) \varepsilon^*_{lK+k}(j) \right)^\top .$$

(5) Repeat steps (2)-(4) M times (e.g. $M = 1000$).

(6) The critical value $c^*(\alpha)$ is obtained as the upper α-quantile of the M statistics.

(7) Reject H_0 if the statistic based on the original sample exceeds the critical value $c^*(\alpha)$.

We note that in step (2) we draw a random sample $U_1, \ldots, U_L$ independent of the component $j = 1, \ldots, d$ or in other words we draw one sample for all components. Hence, this bootstrap procedure captures the dependence of components. This has to be further analysed, but in this work we concentrate on the bootstrap for the univariate situation.

In step (1) we have to calculate the residuals. The method depends on the estimator we use for $\widetilde{\mathbf{A}}^{(2)}$. If we use $\widehat{\mathbf{C}}_{H1}^{(2)}$ or $\widehat{\mathbf{S}}_{H1}^{(2)}$ we estimate a preliminary change point by $\hat{k}$ given in (2.25), respectively 2.26, and calculate for $j = 1, \ldots, d$ and $i = 1, \ldots, n$

$$\hat{\varepsilon}_i(j) \quad = \quad \begin{cases} Y_i(j) - \mathbf{x}_i^\top(j)\widehat{\mathbf{a}}_{\hat{k}}(j), & p < i \leq \hat{k}, \\ Y_i(j) - \mathbf{x}_i^\top(j)\widehat{\mathbf{a}}_{\hat{k}}^0(j), & \hat{k} < i \leq n, \end{cases}$$

with $\widehat{\mathbf{a}}_{\hat{k}}(j)$, $\widehat{\mathbf{a}}_{\hat{k}}^0(j)$ are the least squares estimators based on $Y_1(j), \ldots, Y_{\hat{k}}(j)$, respectively $Y_{\hat{k}+1}(j), \ldots, Y_n(j)$.

If we use $\widehat{\mathbf{C}}_{250}^{(2)}$ or $\widehat{\mathbf{S}}_{250}^{(2)}$ we calculate for $j = 1, \ldots, d$ and $i = 1, \ldots, 250$

$$\hat{\varepsilon}_i(j) \quad = \quad Y_i(j) - \mathbf{x}_i^\top(j)\widehat{\mathbf{a}}_{250}(j), \qquad p < i \leq 250,$$

with $\widehat{\mathbf{a}}_{250}(j)$ is the least squares estimator based on $Y_1(j), \ldots, Y_{250}(j)$.

In step (4) we have to decide about the computation of $\mathbf{\Omega}_n^*$. We choose $\mathbf{C}_n^*$ if we use $\widehat{\mathbf{C}}_{H1}^{(2)}$ or $\widehat{\mathbf{C}}_{250}^{(2)}$, and $\mathbf{S}_n^*$ if we use $\widehat{\mathbf{S}}_{H1}^{(2)}$ or $\widehat{\mathbf{S}}_{250}^{(2)}$.

The choice of the block length K influences the performance of the bootstrap. In the correctly specified situation we set $K = 1$ and $L = n$. The more misspecifed the data is the larger should be the block length K. We investigate this in the next section.

In case we use the bootstrap to approximate the distribution of the statistic, we conduct the bootstrap with 1000 samples.

3.3. Univariate Models

We begin to consider three AR(6) models as in (3.1) with parameters

- AR4: $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$,$\boldsymbol{\varphi}_2 = (-0.4, 0.1, 0, 0, -0.1, -0.1)^\top$,
- AR5: $\boldsymbol{\varphi}_1 = (0.5, -0.2, 0.1)^\top$, $\boldsymbol{\varphi}_2 = (0.4, -0.1, 0, 0, 0, 0.6)^\top$,
- AR6: $\boldsymbol{\varphi}_1 = (0.2, 0.1, 0.3, 0, 0, -0.2)^\top$, $\boldsymbol{\varphi}_2 = (0.3, 0.2, 0.3, 0, 0, -0.2)^\top$.

Figure 3.4 includes sample paths of time series under the alternative of model AR4 to AR6 and appended statistics. In the AR1 model we have a change in the first parameter, in the AR2 model in four parameters and in the AR3 model the first two parameters change.

Figure 3.5 – 3.7 presents SPCs for the statistics $T^{(1)}$, $T^{(2)}$ under model AR4 to AR6, respectively. The power and size is calculated for the asymptotic test by 1000 repetitions and the same holds for the bootstrap test. In the first columns we plotted the actual size against nominal size, in the second columns actual power against the nominal size and in the third actual power against the actual size . In each Figure the first row inlcudes the results for the tests using estimator $\widehat{\mathbf{C}}_{H1}$ and the second row includes the results for the tests using estimator $\widehat{\mathbf{C}}_{250}$. Each plot presents the curves for the asymptotic test and the bootstrap test. Since we

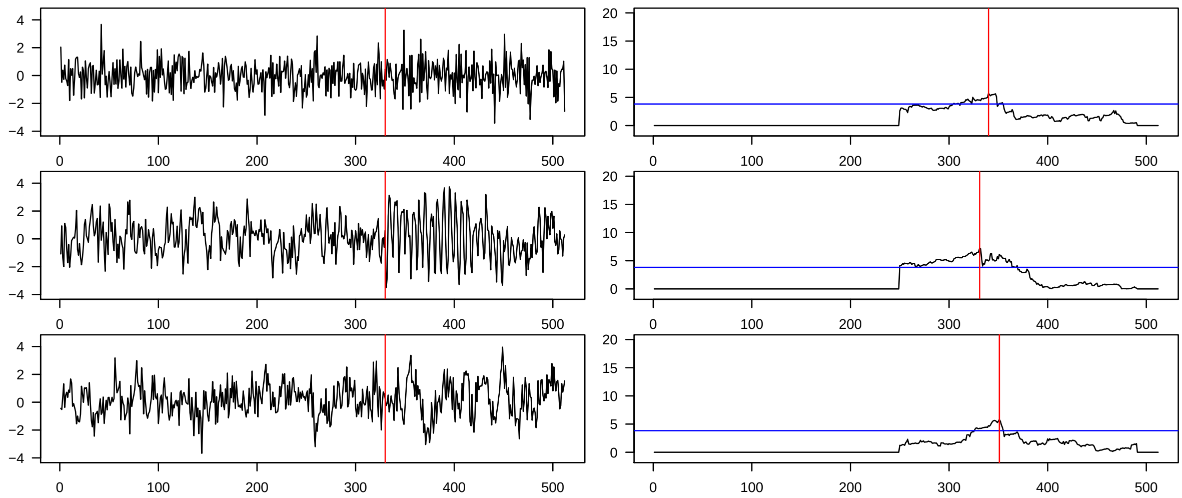

Figure 3.4.: Sample paths and CUSUM statistics under the alternative of models AR4, AR5 and AR6.

consider only estimators $\widehat{\mathbf{C}}_{250}^{(2)}$ and $\widehat{\mathbf{C}}_{H1}^{(2)}$ we suppress the dependence on p' in the sequel to ease the notation.

To begin with, we discuss the empirical size. In the plots the black dotted line gives the nominal size. Thus, under the null hypothesis the curves should be close to this line. We observe that this is more or less fulfilled under all three models. In general the bootstrap reduces the empirical size – more for statistic $T^{(1)}$ than $T^{(2)}$. Further, a test with statistic $T^{(2)}$ has a lower size than the appended statistic $T^{(1)}$ and a test using estimator $\widehat{\mathbf{C}}_{H1}$ has a lower size than a test using estimator $\widehat{\mathbf{C}}_{250}$.

Under the AR4 model all tests perform well. Under the AR5 model the sizes are too small for the test statistics using estimator $\widehat{\mathbf{C}}_{H1}$. Since the size of the asymptotic test is higher, it leads here to better results. A better performance is obtained by the test statistics using estimator $\widehat{\mathbf{C}}_{250}$. In case of $T^{(2)}$ the size is hold well and the difference between the bootstrap and the asymptotic performance is small. For test statistic $T^{(1)}$ the asymptotic size is a little too large and the bootstrap size is a little too low. Under the AR6 model the tests using estimator $\widehat{\mathbf{C}}_{H1}$ hold the size better. As under model AR5 the asymptotic test is preferable. The test statistics using estimator $\widehat{\mathbf{C}}_{250}$ yield under the AR6 model a size higher than the nominal size. Since here, the bootstrap leads to a test with a lower level, it yields the better results.

For the power analysis we mention that the steeper the power curves are the better is the power behaviour. The difference between the power behaviour of $T^{(1)}$ and $T^{(2)}$ is very small in the AR1 and AR3 model, but not in the AR2 model. Here, the power of $T^{(1)}$ is clearly higher. Further, we do not find a significant difference between the power of the asymptotic and bootstrap test.

We conclude that a test with estimator $\widehat{\mathbf{C}}_{H1}$ has rather the tendency to fall below the nominal size and the bootstrap further amplifies this. Hence, here the asymptotic test yields the more reliable results. In contrast, a test with estimator $\widehat{\mathbf{C}}_{250}$ has the tendency to exceed the nominal level and the bootstrap procedure reduces the size and hence improves the performance.

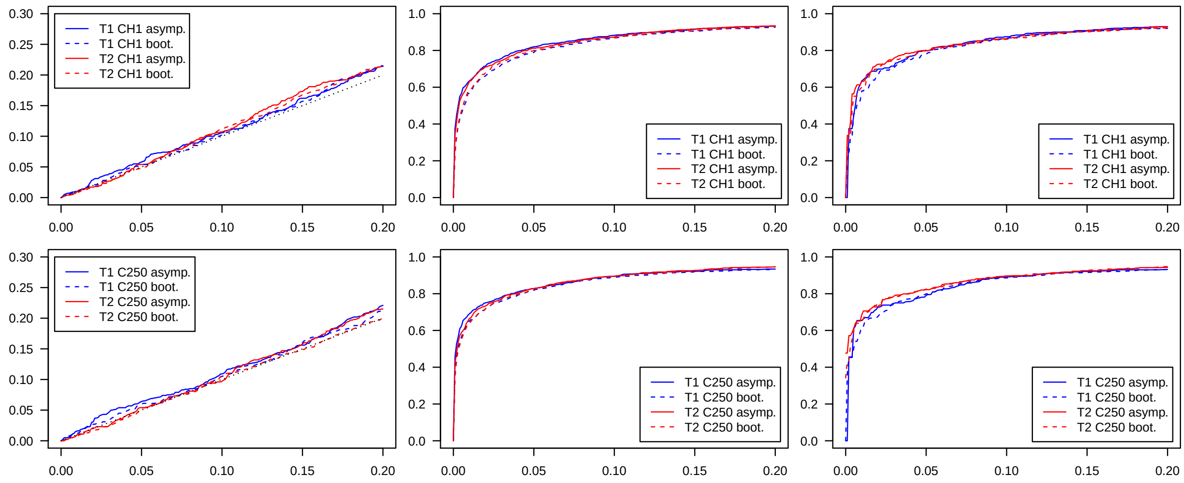

Figure 3.5.: SPCs under the AR4 model.

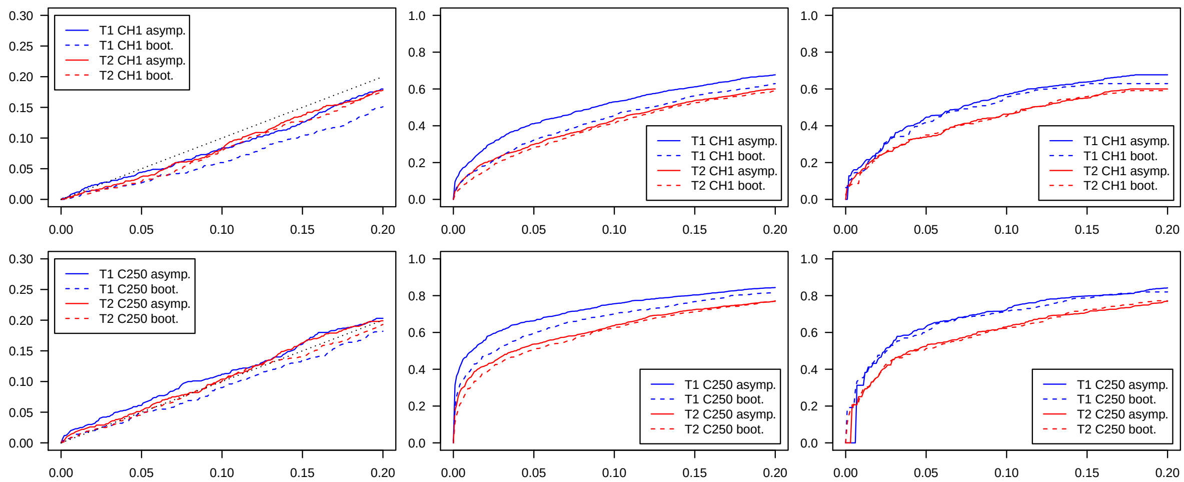

Figure 3.6.: SPCs under the AR5 model.

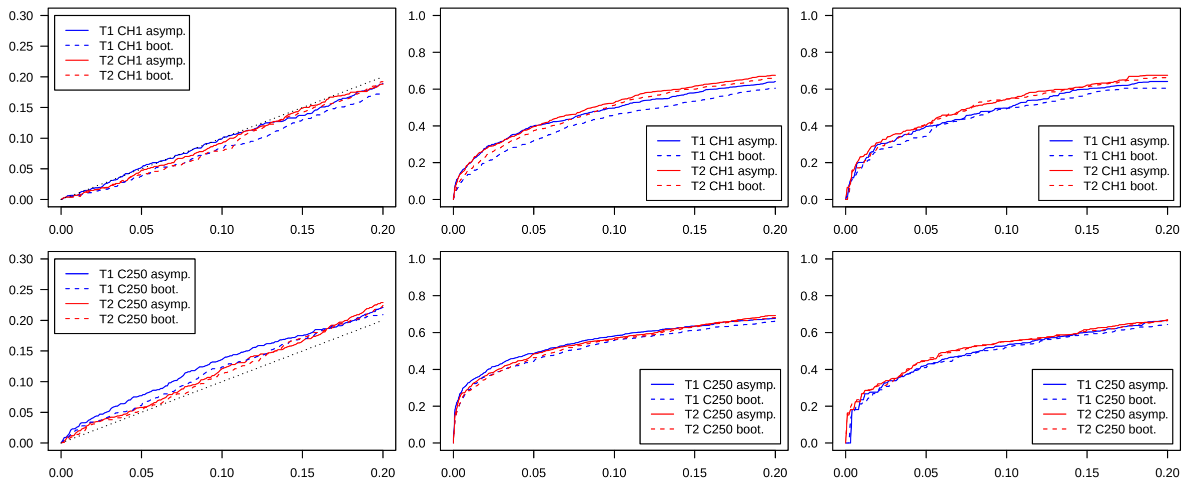

Figure 3.7.: SPCs under the AR6 model.

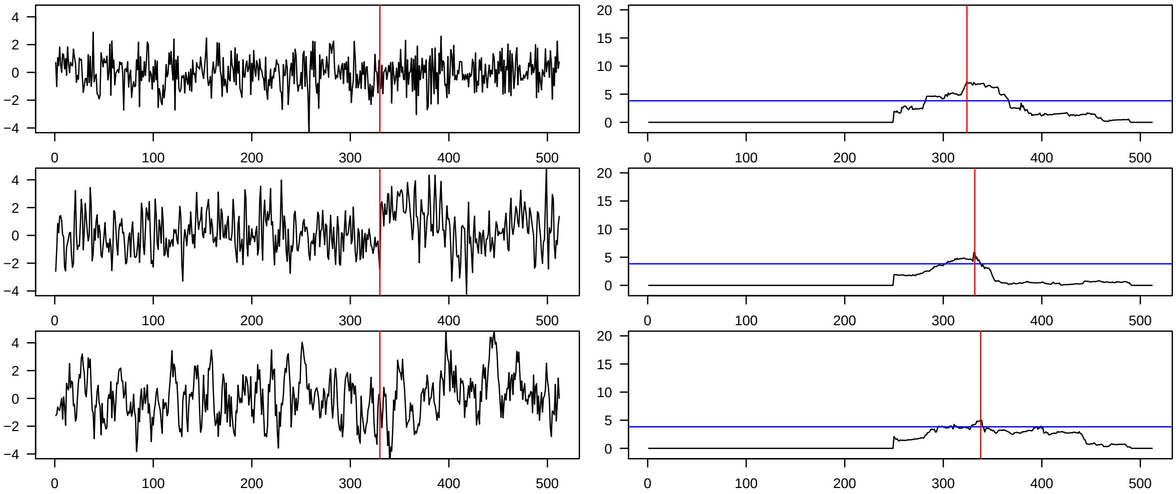

Figure 3.8.: Sample paths and CUSUM statistics under the alternative of models ARMA1, ARMA2 and ARMA3.

In the misspecified situation we consider three ARMA(p,q) models and one TAR(1) model. The ARMA(p,q) model with a change at $\tilde{k} = 330$ is given by:

$$Y_i = \begin{cases} \mathbb{Y}_i^\top \boldsymbol{\varphi}_1 + (\varepsilon_{i-1}, \ldots, \varepsilon_{i-q})\boldsymbol{\rho} + \varepsilon_i, & i \le \tilde{k}, \\ \mathbb{Y}_i^\top \boldsymbol{\varphi}_2 + (\varepsilon_{i-1}, \ldots, \varepsilon_{i-q})\boldsymbol{\rho} + \varepsilon_i, & i > \tilde{k}, \end{cases}$$

with standard normal distributed errors and parameters $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, $\boldsymbol{\rho}$. We consider the parameter constellations

- ARMA1: $\boldsymbol{\rho} = (0.2, 0.1)^\top$, $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$,
  $\boldsymbol{\varphi}_2 = (-0.4, 0.1, 0, 0, -0.1, -0.1)^\top$,
- ARMA2: $\boldsymbol{\rho} = (0.4, -0.3)^\top$, $\boldsymbol{\varphi}_1 = (0.5, -0.2, 0.1)^\top$,
  $\boldsymbol{\varphi}_2 = (0.4, -0.1, 0, 0, 0, 0.6)^\top$,
- ARMA3: $\boldsymbol{\rho} = (0.6, 0.3, 0.2, -0.5)^\top$, $\boldsymbol{\varphi}_1 = (0.2, 0.1, 0.3, 0, 0, -0.2)^\top$,
  $\boldsymbol{\varphi}_2 = (0.3, 0.2, 0.3, 0, 0, -0.3)^\top$.

From model one to model three we enlarged the moving average part, which determines the level of misspecification. In this context we use estimators $\widehat{\mathbf{C}}_{H1}$ and $\widehat{\mathbf{C}}_{250}$ as well as $\widehat{\mathbf{S}}_{H1}$ and $\widehat{\mathbf{S}}_{250}$. A sample path and an appended CUSUM statistic for each model can be found in Figure 3.8.

The bootstrap for test statistics with estimator $\widehat{\mathbf{S}}_{250}$ or $\widehat{\mathbf{S}}_{H1}$ depends on the block length K. We investigate the influence of different block length $K = 1, 2, 5, 8$ on the power and size behaviour. The results are presented in Figures 3.9–3.11. We directly note that the level is not well hold. As above, a test with statistic $T^{(2)}$ has a lower size than a test with $T^{(1)}$ and a test using estimator $\widehat{\mathbf{S}}_{H1}$ has a lower size than a test using estimator $\widehat{\mathbf{S}}_{250}$. Moreover, the choice of the block length K has a higher influence on the test procedure using estimator $\widehat{\mathbf{S}}_{250}$.

Under model ARMA1 the misspecification is low and a block length of $K = 1$ yields the best results. However, under the ARMA2 model the tests using block length $K = 5$ hold the level

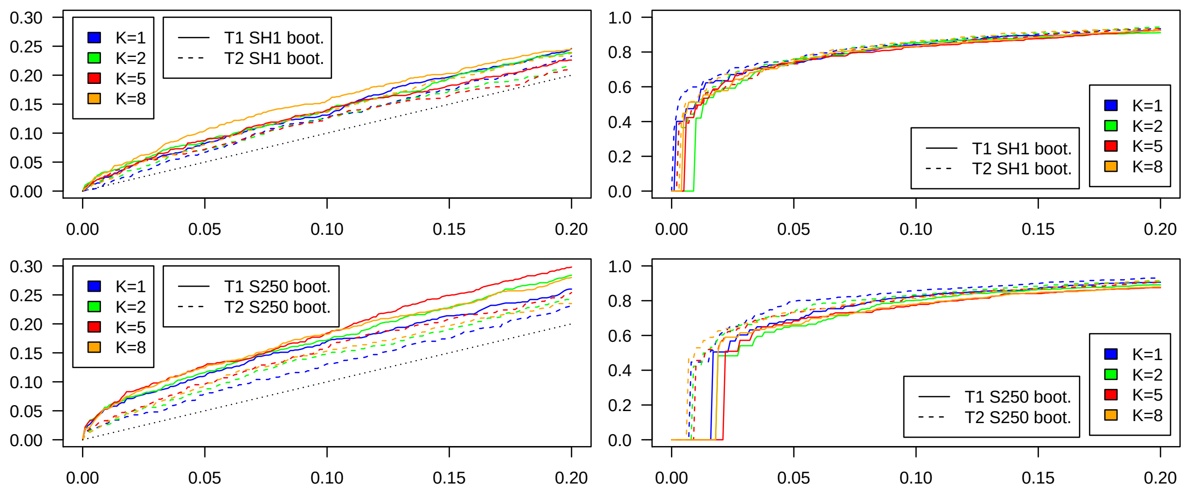

Figure 3.9.: SPCs under the ARMA1 model.

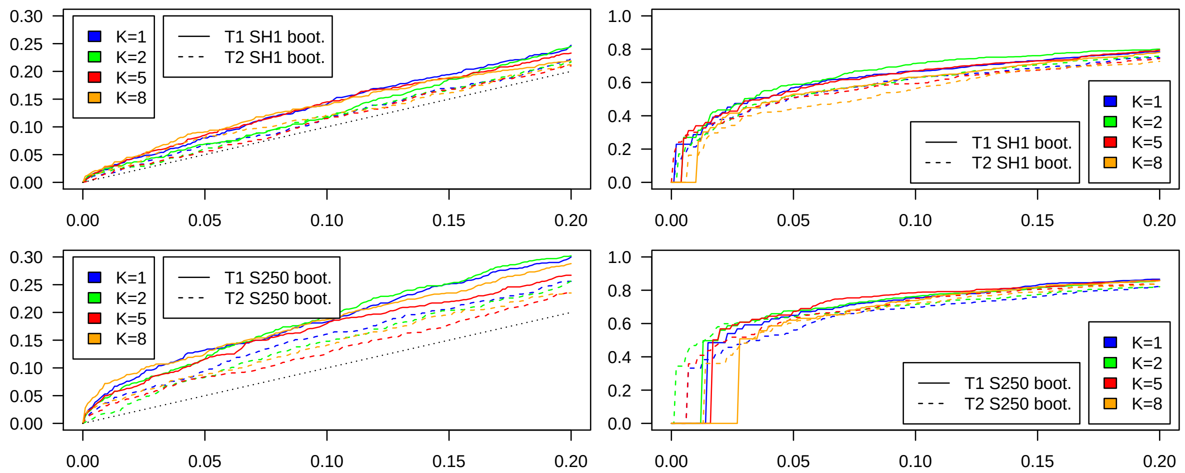

Figure 3.10.: SPCs under the ARMA2 model.

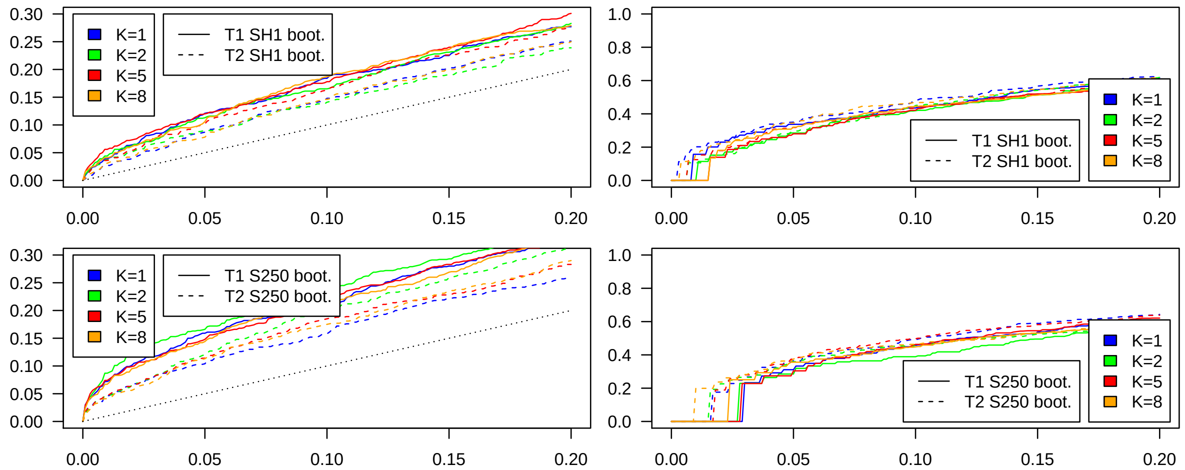

Figure 3.11.: SPCs under the ARMA3 model.

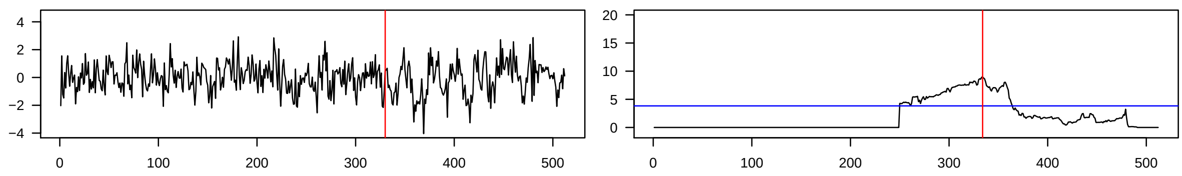

Figure 3.12.: Sample paths and CUSUM statistic under the alternative of model TAR1.

best, but in case of $T^{(1)}$ with estimator $\widehat{\mathbf{S}}_{H1}$. Here, the best choice is $K = 2$. Under model ARMA3 the statistic $T^{(1)}$ with estimator $\widehat{\mathbf{S}}_{H1}$ performs similar for all block length. For the test statistic $T^{(2)}$ with estimator $\widehat{\mathbf{S}}_{H1}$ block length $K = 1, 2, 8$ yield a similar size behaviour. For the test statistic $T^{(1)}$ with estimator $\widehat{\mathbf{S}}_{250}$ block length $K = 8$ holds the level best and for the test statistic $T^{(2)}$ block length $K = 1, 5, 8$ yield a similar size behaviour. The power behaviour is in all cases very similar.

We mentioned before that the long-run autocovariance is very difficult to estimate and the poor performance of the tests using estimators $\widehat{\mathbf{S}}_{H1}$ and $\widehat{\mathbf{S}}_{250}$ is due to the large estimation errors. Hence, we consider the estimators $\widehat{\mathbf{C}}_{H1}$, $\widehat{\mathbf{C}}_{250}$ in the misspecified situation as well.

To this end we refer to Figures 3.13–3.15. Comparing the performances of estimators $\widehat{\mathbf{C}}_{H1}$, $\widehat{\mathbf{C}}_{250}$ and $\widehat{\mathbf{\Sigma}}_{H1}$, $\widehat{\mathbf{\Sigma}}_{250}$ the statistics with estimators $\widehat{\mathbf{C}}_{H1}$, $\widehat{\mathbf{C}}_{250}$ hold the size better while their counter part with $\widehat{\mathbf{\Sigma}}_{H1}$, $\widehat{\mathbf{\Sigma}}_{250}$ exceed it considerably. There is one exception: under the ARMA2 model $T^{(2)}$ with $\widehat{\mathbf{S}}_{H1}$ holds the level better than $T^{(2)}$ with $\widehat{\mathbf{C}}_{H1}$. Further, we observe a slight power advantage for estimators $\widehat{\mathbf{C}}_{H1}$, $\widehat{\mathbf{C}}_{250}$.

Next, we consider a TAR(1) model with a change at $\tilde{k} = 330$:

$$Y_i \quad = \quad \begin{cases} \mathbf{x}_i^\top \boldsymbol{\varphi}_{u1} \mathcal{I}\{Y_{i-1} \leq \lambda\} + \mathbf{x}_i^\top \boldsymbol{\varphi}_{a1} \mathcal{I}\{Y_{i-1} > \lambda\} + \varepsilon_i, & i \leq \tilde{k}, \\ \mathbf{x}_i^\top \boldsymbol{\varphi}_{u2} \mathcal{I}\{Y_{i-1} \leq \lambda\} + \mathbf{x}_i^\top \boldsymbol{\varphi}_{a2} \mathcal{I}\{Y_{i-1} > \lambda\} + \varepsilon_i, & i > \tilde{k}, \end{cases}$$

with standard normal distributed errors and parameters λ, $\boldsymbol{\varphi}_{u1}$, $\boldsymbol{\varphi}_{u2}$, $\boldsymbol{\varphi}_{a1}$, $\boldsymbol{\varphi}_{a2}$:

- TAR1: $\lambda = 0$, $\boldsymbol{\varphi}_{u1} = (-0.1, 0)^\top$, $\boldsymbol{\varphi}_{a1} = (-0.2, 0.1)^\top$, $\boldsymbol{\varphi}_{u2} = (-0.1, -0.1)^\top$, $\boldsymbol{\varphi}_{a2} = (0.2, 0.3)^\top$.

Figure 3.12 inlcudes a sample path and the performance of an appendend CUSUM statistic under the TAR 1 model. We concentrate on the estimators $\widehat{\mathbf{C}}_{H1}$, $\widehat{\mathbf{C}}_{250}$, since we do not expect a better performance than before of tests with estimators $\widehat{\mathbf{S}}_{H1}$, $\widehat{\mathbf{S}}_{250}$.

Under the TAR1 model the performance of the various tests is very similar and the size is held well. As before we observe that a test with estimator $\widehat{\mathbf{C}}_{H1}$ rather has a lower size and a test with estimator $\widehat{\mathbf{C}}_{250}$ exceeds the nominal size. Further, we can not observe a significant difference in the power behaviour.

We summarize the results. Comparing the use of estimators $\widehat{\mathbf{C}}_{H1}$, $\widehat{\mathbf{C}}_{250}$ to the use of $\widehat{\mathbf{\Sigma}}_{H1}$, $\widehat{\mathbf{\Sigma}}_{250}$ the previous ones yield the better results. The problem with the latter ones is the difficult estimation of the long-run autocovariance. The estimation error is larger than the error we make if we use $\widehat{\mathbf{C}}_{H1}$, $\widehat{\mathbf{C}}_{250}$ as an approximation.

Usually the asymptotic statistic $T^{(1)}$ has a higher size than the asymptotic statistic $T^{(2)}$ and as in the correctly specified case the bootstrap reduces the size, more for $T^{(1)}$ than for $T^{(2)}$.

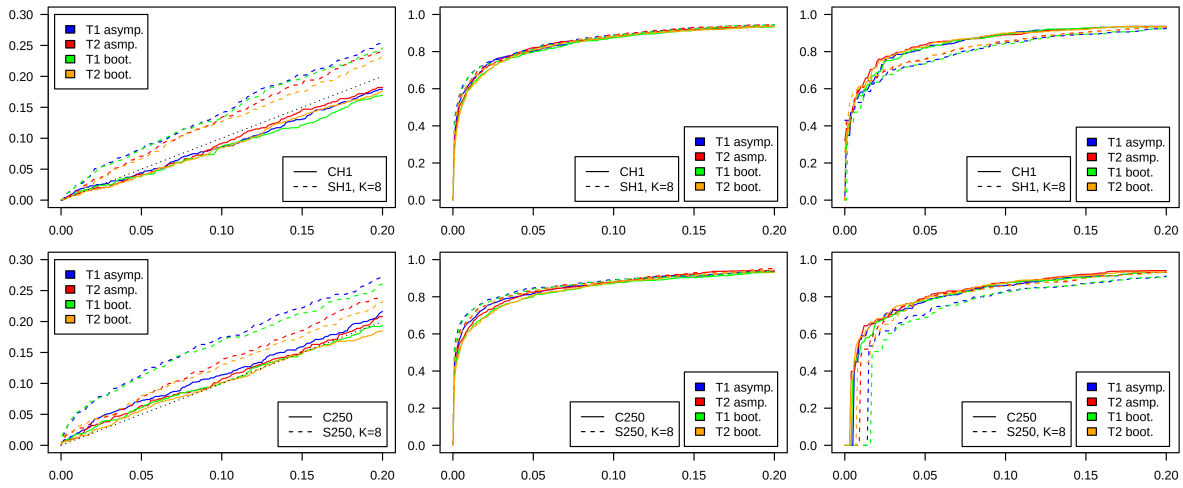

Figure 3.13.: SPCs under the ARMA1 model.

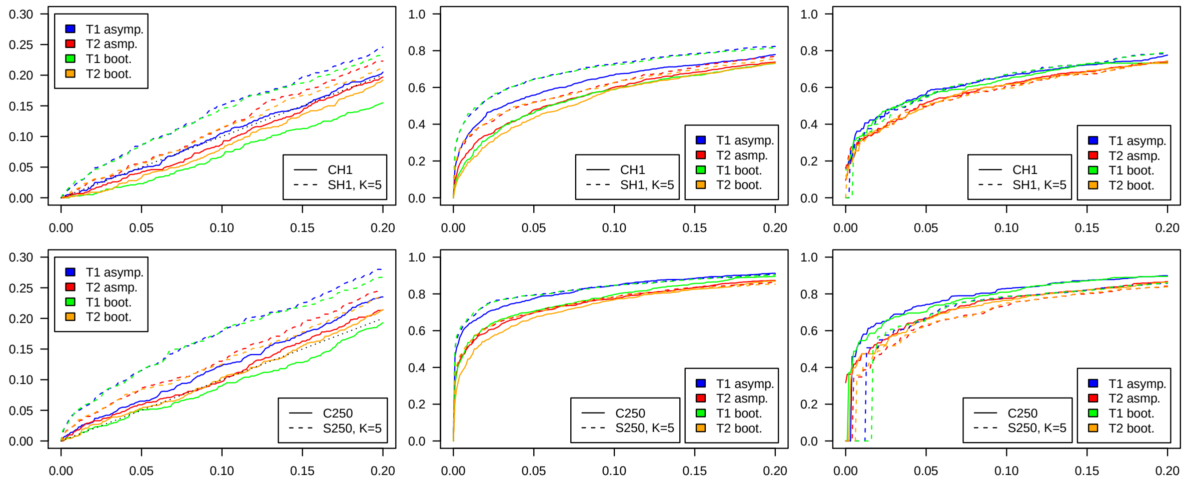

Figure 3.14.: SPCs under the ARMA2 model.

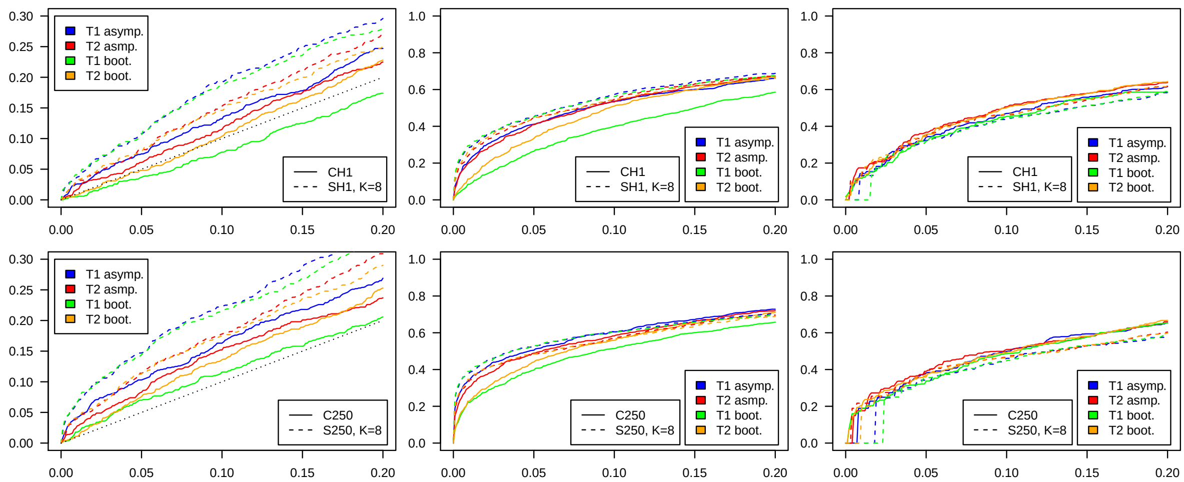

Figure 3.15.: SPCs under the ARMA3 model.

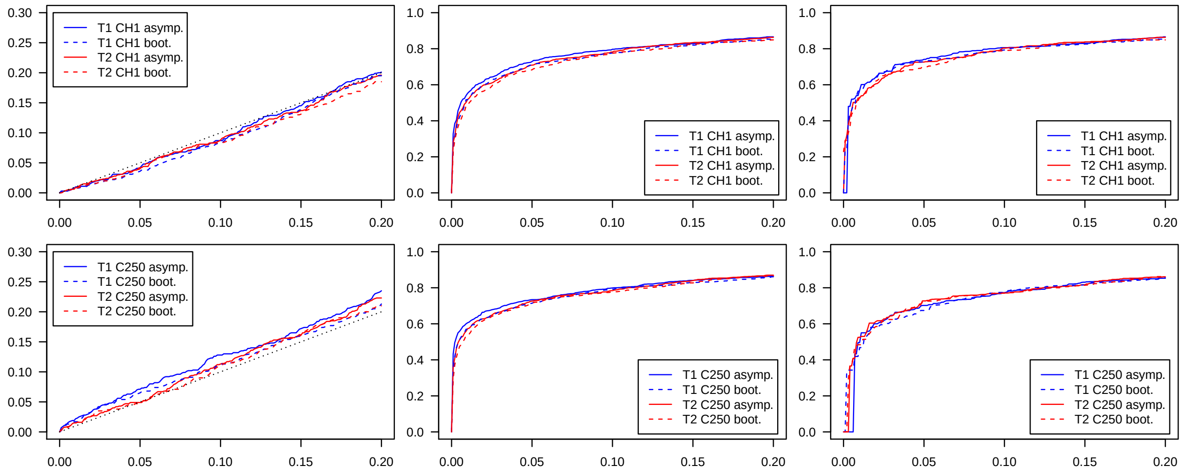

Figure 3.16.: SPCs under the TAR1 model.

Hence, usually the asymptotic statistic $T^{(1)}$ has the highest size, followed by the asymptotic statistic $T^{(2)}$ and the bootstrap statistic $T^{(2)}$, and the lowest size has the bootstrap statistic $T^{(1)}$.

As in the correctly specified case the tests with estimator $\widehat{\mathbf{C}}_{H1}$ have a rather low size. Since the bootstrap further amplifies this, an asymptotic test is in situations of less misspecification preferable. Under a higher level of misspecification the sizes grow and hence, the bootstrap tests work better.

In contrast, a test with estimator $\widehat{\mathbf{C}}_{250}$ has the tendency to exceed the nominal level and since the bootstrap reduces the size it improves the performance.

For the univariate epidemic model we use the statistics $T^{(3)}$, $T^{(4)}$ and focus on the estimator $\widehat{\mathbf{C}}_{250}$, since the performance of $\widehat{\boldsymbol{\Sigma}}_{250}$ does not change under the epidemic alternative. In the following, we restrict our attention to asymptotic tests, due to time restrictions and the high computational time of the bootstrap.

We consider an AR(6), an ARMA(6, 2) and a TAR(1) model with an epidemic change between $\tilde{k}_1 = 330$ and $\tilde{k}_2 = 380$ and parameters

- AR7: $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$,
  $\boldsymbol{\varphi}_2 = (-0.5, 0.2, 0, 0, -0.1, -0.1)^\top$,
- ARMA4: $\boldsymbol{\rho} = (0.4, -0.3)^\top$, $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$,
  $\boldsymbol{\varphi}_2 = (-0.5, 0.2, 0, 0, -0.1, -0.1)^\top$,
- TAR2: $\lambda = 0$, $\boldsymbol{\varphi}_{u1} = (-0.1, 0)^\top$, $\boldsymbol{\varphi}_{a1} = (-0.2, 0.1)^\top$, $\boldsymbol{\varphi}_{u2} = (-0.4, -0.1)^\top$,
  $\boldsymbol{\varphi}_{a2} = (-0.5, 0.3)^\top$.

Compared to the parameters we used for the AMOC situation we increased the amount of change, since the amount of observations with parameters differing from the original parameters is smaller and hence the change is more difficult to detect.

The SPCs for the epidemic models are plotted in Figure 3.17. To get an impression of the stability of the SPCs we plotted two curves for each model. Further, we fixed the seed for the random generator to compare the difference in size and power between the AR and the

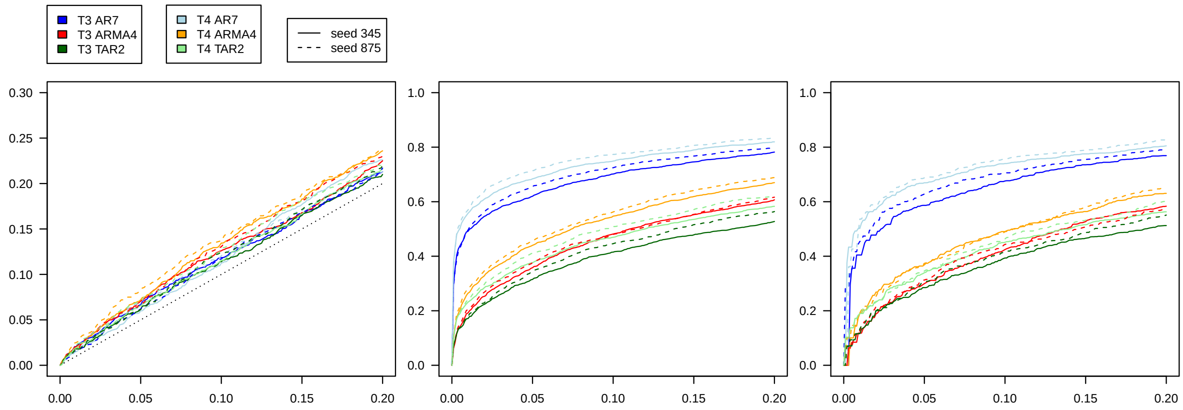

Figure 3.17.: SPCs under univariate epidemic models.

ARMA model. We observe that the size is a bit higher under the ARMA model but not much. On the contrary the power is reduced a lot not only a bit. Further, we note that in contrast to the AMOC statistics the sum-type statistic $T^{(4)}$ has the higher size and as well the higher power.

3.4. Multivariate Models

We expect the behaviour of the multivariate statistics to be similar to the behaviour of the univariate statistics, since the multivariate statistics add up the univariate ones. In the multivariate context it is of interest how the behaviour of the tests differ between independent and dependent components. If the components are dependent the correct asymptotic distribution is not known, since we do not know the covariance structure between the components and can not estimate it correctly. Hence, if the dependence is weak the asymptotic distribution with the independent Brownian bridges might still give reasonable approximations. In case of strong dependencies a bootstrap might improve the performance. This has further to be investigated and we begin here with the first step.

In the multivariate AMOC setting we consider time series with $d = 12$ components, where each component follows an AR(6) model with parameters

- AR8: $\boldsymbol{\varphi}_1 = (0.5, -0.2, 0.1, 0, 0, 0.2)^\top$, $\boldsymbol{\varphi}_2 = (0.5, -0.3, 0.1, 0, 0, 0.2)^\top$

and further an ARMA(6, 2) model with parameters

- ARMA5: $\boldsymbol{\rho} = (0.4, -0.3)^\top$, $\boldsymbol{\varphi}_1 = (0.5, -0.2, 0.1, 0, 0, 0.2)^\top$,
  $\boldsymbol{\varphi}_2 = (0.5, -0.3, 0.1, 0, 0, 0.2)^\top$.

Moreover, we consider the situation of dependent components by adding an inference Z_i to each component with $Z_i \sim \mathcal{N}(0, 0.3)$, i.e.

$$Y_i(j) \quad = \quad Z_i + \begin{cases} \mathbb{Y}_i^\top(j)\boldsymbol{\varphi}_1 + (\varepsilon_{i-1}, \ldots, \varepsilon_{i-q}(j))\boldsymbol{\rho} + \varepsilon_i(j), & i \leq \tilde{k}, \\ \mathbb{Y}_i^\top(j)\boldsymbol{\varphi}_2 + (\varepsilon_{i-1}(j), \ldots, \varepsilon_{i-q}(j))\boldsymbol{\rho} + \varepsilon_i(j), & i > \tilde{k}, \end{cases}$$

with parameters

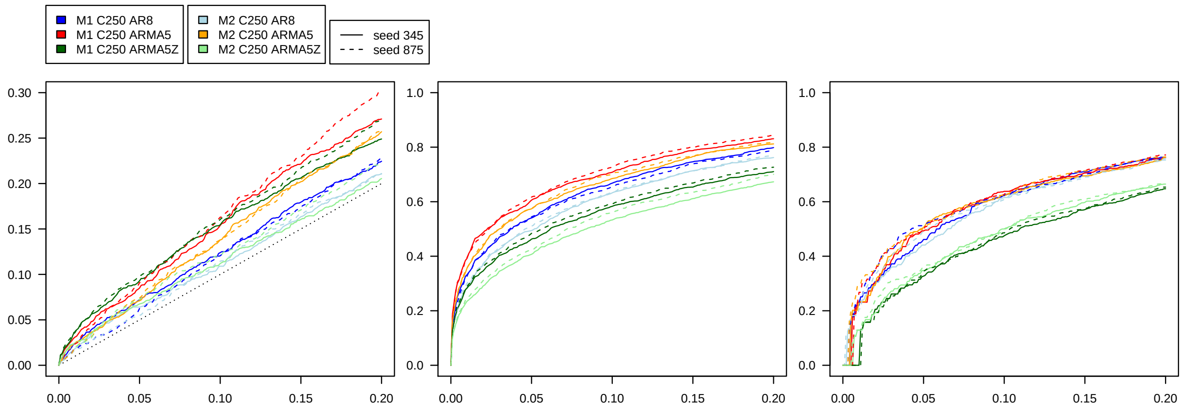

Figure 3.18.: SPCs under multivariate AMOC models.

- ARMAZ5: $\boldsymbol{\rho} = (0.4, -0.3)^\top$, $\boldsymbol{\varphi}_1 = (0.5, -0.2, 0.1, 0, 0, 0.2)^\top$,
  $\boldsymbol{\varphi}_2 = (0.5, -0.3, 0.1, 0, 0, 0.2)^\top$.

The change is as before at $\tilde{k} = 330$.

As in the univariate case the statistic with estimator $\widehat{\mathbf{C}}_{250}$ tends to exceed the nominal size and the maximum-type statistic has again a higher size than the sum-type statistic. Comparing the size of the correctly specified situation with the misspecified situation, we observe the actual size growing. These are the things we would have expected. We now compare the situation of independent and dependent components. Since we only consider one situation of dependence this is rather preliminary. In case of $M^{(1)}$ the size does not really change, but the size with statistic $M^{(2)}$ drops down. This is somehow strange and we have to investigate this in detail. Further, we observe that the power is lowered a lot.

In the multivariate epidemic case we consider a similar setting with time series of the dimension $d = 12$ in context of an AR(6) model, an ARMA(6, 2) model and in the situation of dependent components as above, but with parameters

- AR9: $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$, $\boldsymbol{\varphi}_2 = (-0.3, 0.2, 0, 0, -0.1, -0.1)^\top$,

- ARMA6: $\boldsymbol{\rho} = (0.4, -0.3)^\top$, $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$,
  $\boldsymbol{\varphi}_2 = (-0.3, 0.2, 0, 0, -0.1, -0.1)^\top$,

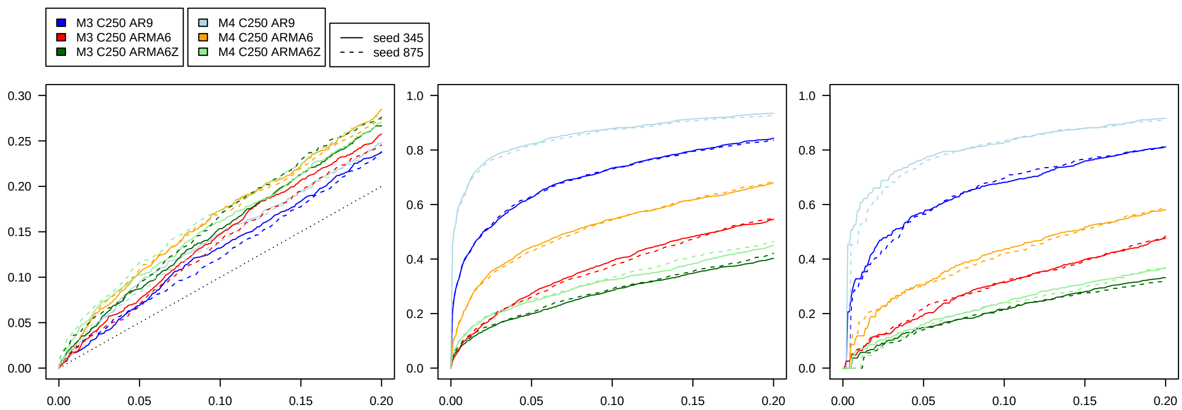

Figure 3.19.: SPCs under multivariate epidemic models.

- ARMA6Z: $\boldsymbol{\rho} = (0.4, -0.3)^\top$, $\boldsymbol{\varphi}_1 = (-0.1, 0.1, 0, 0, -0.1, -0.1)^\top$,
  $\boldsymbol{\varphi}_2 = (-0.3, 0.2, 0, 0, -0.1, -0.1)^\top$

and the change takes place between $(\tilde{k}_1, \tilde{k}_2) = (330, 380)$.

In the univariate epidemic case the sum-type statistic has the higher size and this holds here as well. We further observe that in the misspecified situation the level is higher. For statistic $T^{(1)}$ the size is further increased in the situation of dependent components and for statistic $T^{(2)}$ the size is in one case (seed 875) increased and in the other case (seed 345) decreased. The power results seem to be rather stable and we observe a strong decrease in power from the correctly to the misspecified situation and again from independent to dependent components. In the first two situations the difference between the power of $T^{(1)}$ and $T^{(2)}$ are large. This is reduced under dependent components.

4. Data Analysis

We analyse a twelve-channel EEG data set which was recorded in a visual-motor experiment, where a participant is visually instructed to move a joystick either to the right or to the left. The data is splitted into two parts, the data including the movement to the left side and the data including the movement to the right side. Since the instruction to move is given at a known fixed time point, the first goal is to identify the delay between stimulus shown and hand reaction. To this end we assume a multivariate AMOC model and use the multivariate test and estimation methods. In a second step we analyse the data per channel to identify those channels with a significant behaviour. If later on the test procedure includes only the relevant channels, the power of the test can be increased and the functionality of a device to move an artifical limp be improved.

The data for the left and right side is analysed separately, because we do not only want to detect the intention to move, but further decide in which direction the arm should be moved. In this context the univariate analysis of each channel is important too, since probably some parts of the brain are more involved with right, respectively left, movements. In view of the twelve univariate tests we apply the false discovery rate (FDR) of Benjamini and Hochberg (1995) to correct for the multiple testing.

The results for the AMOC model, more specifically the form of the density estimates of the change point, indicate a second change point which is probably connected to the end of hand movement. Hence, we consider further an epidemic change-point model to divide the time series into a relaxed state and a state containing the signal to move.

Section 4.1 gives detailed information about the data set. In Section 4.2 we discuss the order of the linear model we use for approximation of the data. Finally, Section 4.3 includes the data evaluation.

4.1. Data Description

On a monitor, a right-handed participant was shown two types of stimuli: an arrow pointing to the left side or an arrow pointing to the right side. When a right (left) arrow was presented the participant was supposed to move the joystick to the right (left) side. This procedure was repeated $N = N_1 + N_2$ times, where $N_1 = 118$ corresponds to seeing the left arrow and $N_2 = 134$ corresponds to seeing the right arrow. Each movement was done correctly. The EEGs were recorded across these N trials at $P = 12$ electrodes placed on the scalp. Each trial took about one second, and the recordings were taken at $T = 512$ time points over the one-second interval.

The data is taken from a representative participant of a group of 11 healthy, young adults between 20-35 years. In this experiment 64 electrodes were placed on the scalp. From these a subset of 12 electrodes presumed to have involvement in neural processes engaged in visual-motor actions (refer to Marconi et al. (2001)) were selected. The selected channels were FC3, FC5, C3, P3, and O1 over the left hemisphere; FC4, FC6, C4, P4 and O2 over the right hemisphere; and Cz and Oz over the midline (refer to Figure 4.1). The frontal (FC) electrodes were placed over the prefrontal cortex, regions previously shown to have involvement in premotor processing. The central (C) electrodes were placed over structures involved in motor performance, while the parietal (P) and occipital (O) electrodes were placed over

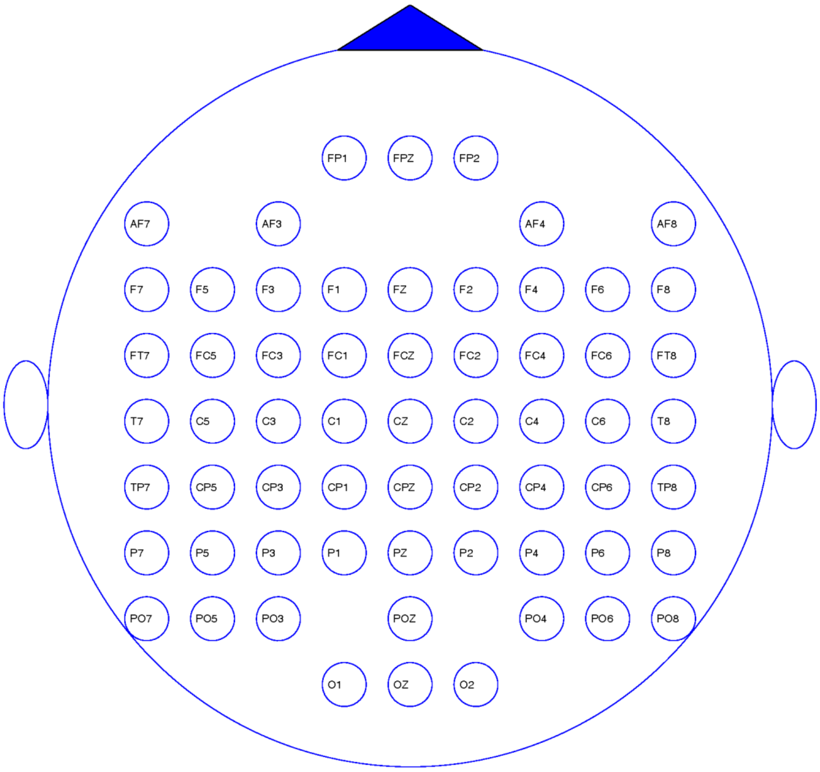

Figure 4.1.: Position of the 64 EEG channels.

structures involved in visual sensation and visual-motor transformations (refer to Marconi et al. (2001)).

This data was already discussed in a number of papers to demonstrate the applicability of novel statistical methods. In Fiecas and Ombao (2011) it was utilised to demonstrate a method for estimating partial coherence between the channels. Böhm et al. (2010) illustrated a method for classification and discrimination of multivariate time series on this data.

Channel	1	2	3	4	5	6	7	8	9	10	11	12
	FC5	FC3	FC4	FC6	C3	CZ	C4	P3	P4	O1	OZ	O2

Table 4.1.: Channels considered in the data analysis.

4.2. Order of the Linear Autoregressive Model

We have to decide about the order of the autoregressive model, which we use to approximate the data. To this end, we do not use the whole data set but constrain us to the data obtained up to time point 250, since until here we consider it stationary. We fit models with order 1 to 17. The Akaike information criterion (AIC) as well as the Bayesian information criterion (BIC) recommend orders between 9 and 17 without any pattern within all channels and trials. Since orders higher than 9 seem to be unreasonably complicated, we fit linear autoregressive models of order 9 for each trial (even if a model with order higher than 9 was suggested). In Figure 4.2 we present two plots of autocorrelation functions (ACF) of residuals in trials

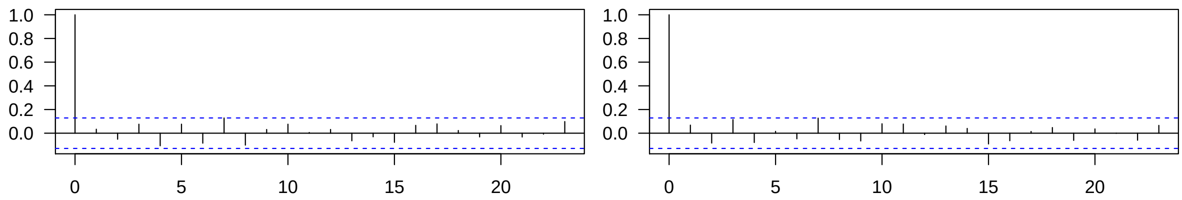

Figure 4.2.: ACF of residuals obtained by fitting an AR(9) model to trials where the AIC recommends order 9.

for which the AIC suggested a model of order 9. We see that the residuals appear to be independent. In Figure 4.3 we present plots of autocorrelation functions of trials for which the AIC a model of higher order than 9 suggested. In the first six plots models of order 10 to 15 and in the last four plots models of order 16 and 17 were suggested. We can see that even here the residuals appear to be white noise or close to white noise.

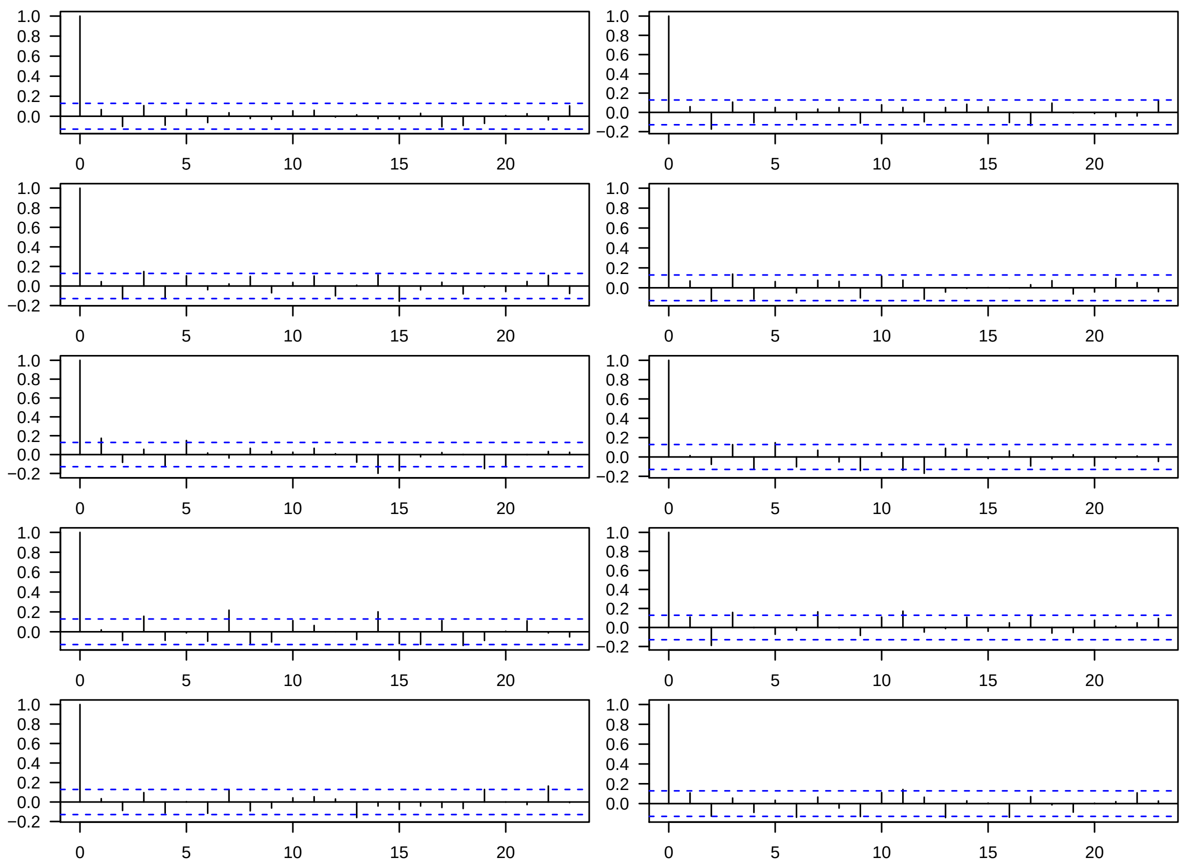

Figure 4.3.: ACF of residuals obtained by fitting an AR(9) model to trials where the AIC recommends orders higher 9.

4.3. Data Evaluation

At the beginning, a multivariate AMOC model is considered. We use the statistics $M^{(1)}$ and $M^{(2)}$, defined in (3.6) and (3.7), with estimator $\widehat{\mathbf{C}}_{250}^{(2)}$. We conduct the tests at a 5% level and compute critical values by the asymptotic distribution and the bootstrap procedure, where we use 500 bootstrap samples. First simulation results on the multivariate bootstrap showed that these are not enough samples to receive a stable approximation. Hence, the bootstrap results here should be viewed as a first impression.

Table 4.2 gives the relative number of estimated change points for the data including movements to the left side and the data including movements to the right side. The results between the asymptotic and the bootstrap test differ a lot. The asymptotic tests detect in almost all, respectively in three quarters, of the channels a change, while the bootstrap tests detect in about half of the channels a change. Based on the respectively identified change points we obtain kernel density estimates presented in Figure 4.4. Though the number of changes included in the estimation is rather different, the shape of the densities look rather similar. The mode of the densities is around 330 for the data set including movements to the left side and around 345 for the data set including movements to the right side. This makes sense, since the participant is right-handed and can give pressure to the left side more easily.

In a next step we want to find out if all channels are equally involved in hand movements or if differences exist, especially between left and right side movements. Hence, we test in each channel separately for a change. To this end we use the univariate test statistics $T^{(1)}$ and $T^{(2)}$, defined in (3.2) and (3.3), with estimator $\widehat{\mathbf{C}}_{250}^{(2)}$. In Table 4.3 and 4.4 the relative number of estimated change points are given. In the first column the relative number of estimated change points is given without correcting for the multiple tests and in the second column we applied the FDR of Benjamini and Hochberg (1995).

For both sides exist strong differences between the channels with respect to the relative number of estimated change points. The results indicate that channel 1, 4, 10 and probably 8 should not be included in a further detection method. Strongly significant is channel 9 followed by 5 and 12. These observations hold for both sides. For further discussion we consider Figure 4.5 which includes density estimates and information on how many estimated change points the estimation is based. On the left (right) the absolute number of significant changes for movements to the left (right) are given. Directly notable is the fact that for channels 8 to 12, which are involved with visual sensation, the density looks similar for left and right, but not for channels 1 to 7, which are involved with premotor and motor performances. The more changes are detected the clearer is the mode of the density. If less changes are detected

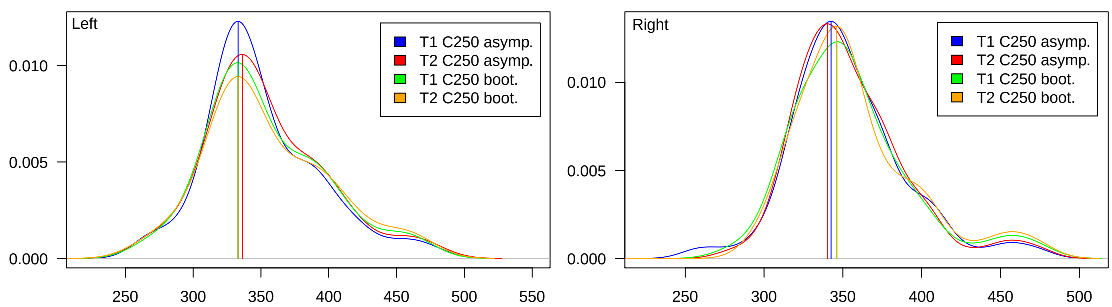

Figure 4.4.: Estimated densities of the change point under the multivariate AMOC model.

it seems that there might be a second mode around 400. Probably there is a strong change connected to the beginning of hand movement and a second change connected to the end of hand movement. The first one might be stronger due to the signal to move immediately and the second is weaker or more spread since you can move faster or slower.

Hence, we further consider the multivariate and univariate epidemic model and test statistics $M^{(3)}$, $M^{(4)}$, $T^{(3)}$, $T^{(4)}$ defined in (3.8), (3.9), (3.4), (3.5) with estimator $\widehat{\mathbf{C}}_{250}^{(2)}$. Compared to the AMOC tests, the multivariate epidemic tests lead to a higher relative number of estimated change points. Concerning the significance of the channels we obtain in one way the same result as in the AMOC case: the channels 1, 4, 8 and 10 are the less significant ones and probably 8 and 10 should be removed, but not channels 1 and 4. They could be used to identify the direction of the movement, since for movements to the left channel 4 is more involved and for movements to the right channel 1.

In Figure 4.6 we plotted the estimated joint density of the first change point and the duration of change. The density has a clear mode at $(328, 60)$ for the left, and for the right at $(333, 42)$. As in the univariate model the reaction time for the movement to the left is in general faster but not so clear as before. Interesting is the longer duration of change for movements to the left. An explanation might be that turning the wrist to the right is more uncomfortable than turning it to the left. Figure 4.7 shows contour plots of the joint densities. Notable is the more compact form of the plot for the right hand (which might be explained by the reason mentioned directly above) and that a late first change point is accompenied by shorter duration times. Probably the participant is more in a hurry if he starts late.

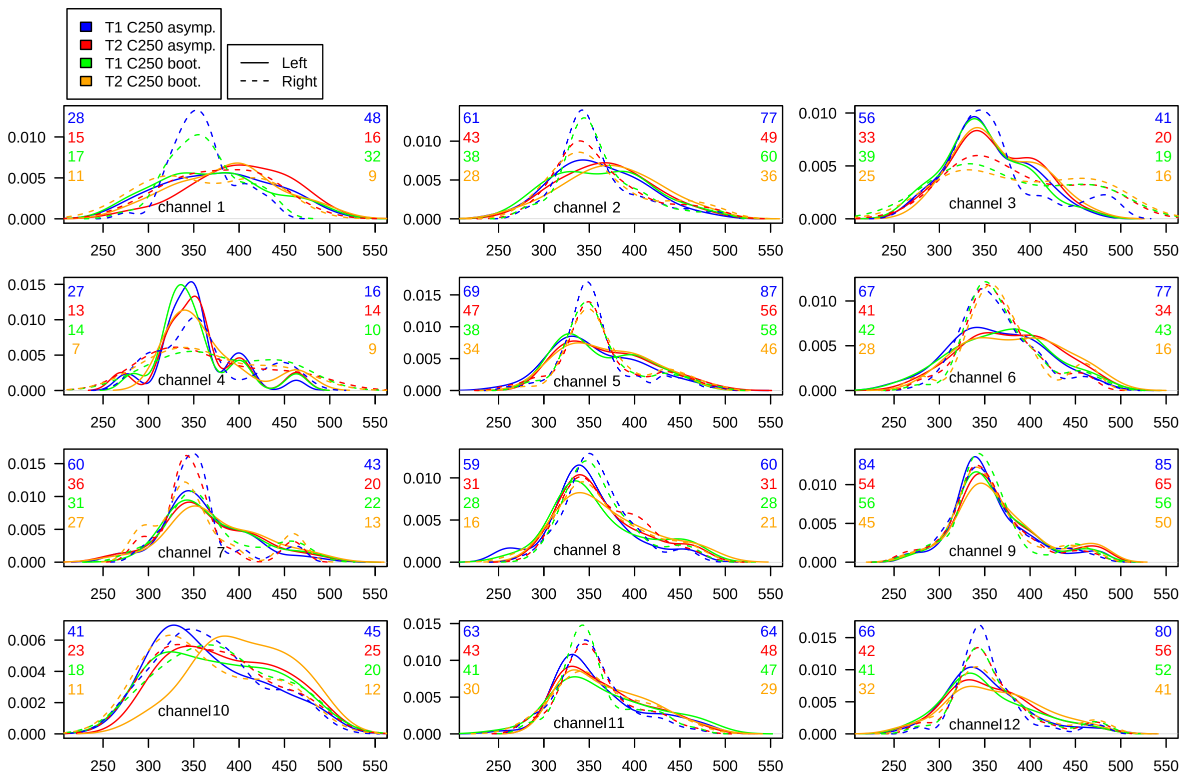

Figure 4.5.: Number of estimated change points and density estimates of the change point under the univariate AMOC model.

	$M^{(1)}$asymp.	$M^{(2)}$asymp.	$M^{(1)}$boot.	$M^{(2)}$boot.
Left	0.92	0.75	0.56	0.49
Right	0.90	0.77	0.57	0.50

Table 4.2.: Relative number of estimated change points under the multivariate AMOC model.

	$T^{(1)}$ asymp.		$T^{(2)}$asymp.		$T^{(1)}$ boot.		$T^{(2)}$ boot.	
channel	uncor.	FDR	uncor.	FDR	uncor.	FDR	uncor.	FDR
1	0.25	0.24	0.15	0.13	0.15	0.14	0.12	0.09
2	0.55	0.52	0.41	0.36	0.37	0.32	0.29	0.24
3	0.56	0.47	0.36	0.28	0.40	0.33	0.25	0.21
4	0.26	0.23	0.14	0.11	0.14	0.12	0.08	0.06
5	0.67	0.58	0.48	0.40	0.40	0.32	0.36	0.29
6	0.67	0.57	0.42	0.35	0.45	0.36	0.31	0.24
7	0.57	0.51	0.38	0.31	0.32	0.26	0.25	0.23
8	0.55	0.50	0.35	0.26	0.30	0.24	0.21	0.14
9	0.75	0.71	0.51	0.46	0.53	0.47	0.40	0.38
10	0.36	0.35	0.25	0.19	0.22	0.15	0.11	0.09
11	0.58	0.53	0.45	0.36	0.38	0.35	0.28	0.25
12	0.64	0.56	0.43	0.36	0.42	0.35	0.31	0.27

Table 4.3.: Relative number of estimated change points for the data including movements to left side under the univariate AMOC model.

	$T^{(1)}$ asymp.		$T^{(2)}$asymp.		$T^{(1)}$ boot.		$T^{(2)}$ boot.	
channel	uncor.	FDR	uncor.	FDR	uncor.	FDR	uncor.	FDR
1	0.40	0.36	0.16	0.12	0.26	0.24	0.10	0.07
2	0.68	0.57	0.46	0.37	0.52	0.45	0.35	0.27
3	0.34	0.31	0.21	0.15	0.17	0.14	0.15	0.12
4	0.13	0.12	0.12	0.10	0.08	0.07	0.07	0.07
5	0.69	0.65	0.51	0.42	0.51	0.43	0.41	0.34
6	0.63	0.57	0.39	0.25	0.39	0.32	0.22	0.12
7	0.35	0.32	0.22	0.15	0.21	0.16	0.13	0.10
8	0.51	0.45	0.34	0.23	0.28	0.21	0.19	0.16
9	0.66	0.63	0.53	0.49	0.48	0.42	0.42	0.37
10	0.36	0.34	0.24	0.19	0.18	0.15	0.13	0.09
11	0.53	0.48	0.41	0.36	0.40	0.35	0.29	0.22
12	0.64	0.60	0.50	0.42	0.45	0.39	0.35	0.31

Table 4.4.: Relative number of estimated change points for the data including movements to the right side under the univariate AMOC model.

	$M^{(3)}$asymp.	$M^{(4)}$asymp.	$M^{(3)}$boot.	$M^{(4)}$boot.
Left	0.98	0.96	0.63	0.72
Right	0.97	0.92	0.68	0.79

Table 4.5.: Relative number of estimated change points under the multivariate epidemic model.

	$T^{(3)}$ asymp.		$T^{(4)}$ asymp.		$T^{(3)}$ boot.		$T^{(4)}$ boot.	
channel	uncor.	FDR	uncor.	FDR	uncor.	FDR	uncor.	FDR
1	0.28	0.28	0.30	0.28	0.19	0.18	0.22	0.19
2	0.74	0.66	0.75	0.71	0.53	0.43	0.66	0.55
3	0.72	0.69	0.72	0.71	0.55	0.50	0.66	0.63
4	0.48	0.46	0.54	0.54	0.32	0.27	0.46	0.41
5	0.74	0.72	0.77	0.76	0.58	0.54	0.73	0.62
6	0.80	0.74	0.83	0.80	0.64	0.53	0.81	0.72
7	0.72	0.67	0.75	0.74	0.55	0.48	0.68	0.61
8	0.71	0.66	0.73	0.69	0.45	0.42	0.64	0.54
9	0.86	0.81	0.87	0.84	0.69	0.64	0.78	0.75
10	0.45	0.43	0.58	0.56	0.25	0.22	0.41	0.34
11	0.77	0.72	0.81	0.75	0.52	0.45	0.66	0.58
12	0.77	0.73	0.85	0.84	0.64	0.54	0.75	0.69

Table 4.6.: Relative number of estimated change points for the data including movements to the left side under the univariate epidemic model.

	$T^{(3)}$ asymp.		$T^{(4)}$ asymp.		$T^{(3)}$ boot.		$T^{(4)}$ boot.	
channel	uncor.	FDR	uncor.	FDR	uncor.	FDR	uncor.	FDR
1	0.55	0.53	0.56	0.53	0.46	0.40	0.50	0.46
2	0.78	0.77	0.78	0.75	0.63	0.57	0.72	0.68
3	0.48	0.44	0.52	0.47	0.27	0.21	0.41	0.37
4	0.15	0.15	0.14	0.13	0.11	0.10	0.12	0.11
5	0.75	0.74	0.76	0.75	0.64	0.58	0.73	0.69
6	0.81	0.74	0.77	0.72	0.61	0.54	0.72	0.63
7	0.57	0.54	0.63	0.57	0.37	0.31	0.56	0.49
8	0.68	0.63	0.68	0.63	0.51	0.38	0.59	0.51
9	0.78	0.73	0.79	0.78	0.63	0.60	0.77	0.71
10	0.52	0.49	0.53	0.51	0.32	0.25	0.42	0.37
11	0.74	0.67	0.75	0.72	0.53	0.48	0.65	0.60
12	0.74	0.69	0.79	0.72	0.61	0.58	0.69	0.62

Table 4.7.: Relative number of estimated change points for the data including movements to the right side under the univariate epidemic model.

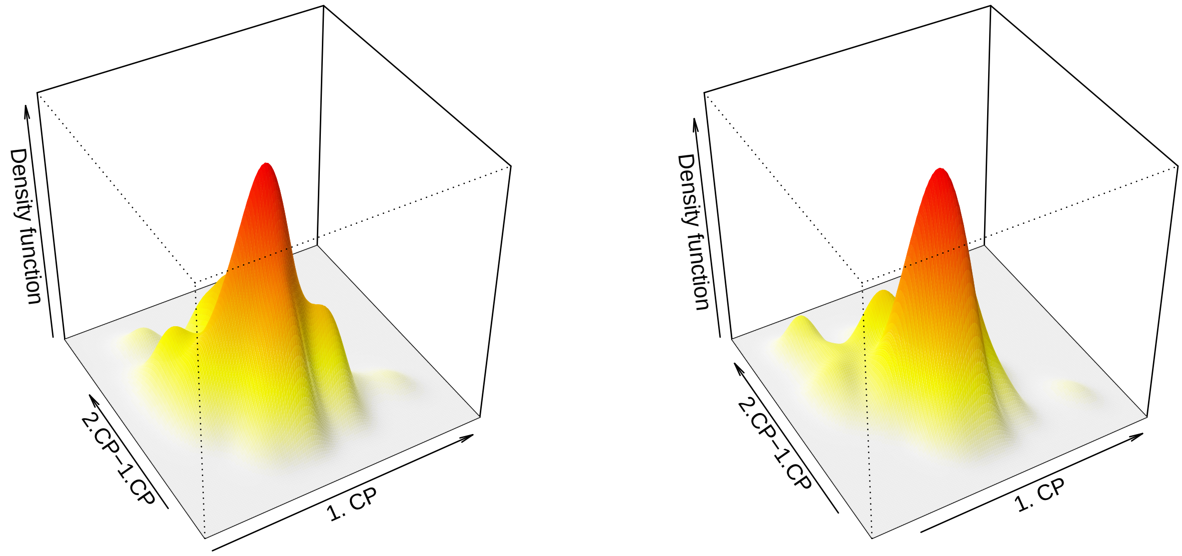

Figure 4.6.: Contour plot of the joint density of 1.CP and 2.CP-1.CP

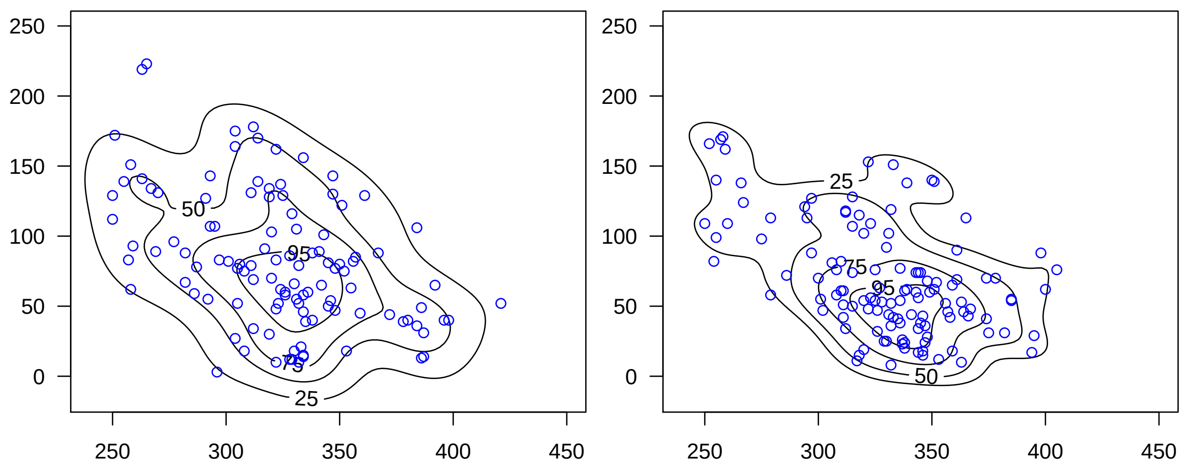

Figure 4.7.: Estimated joint density of 1.CP and 2.CP-1.CP

Part II.

Change-Point Methods for Multiple Structural Breaks and Regime-Switching Models

5. Basics

This part considers multiple change-point location models which allow for several changes in the expectation of an otherwise stationary random sequence. Thereby, we do not restrict our attention to the classical change-point situation with deterministic change points and a bounded number of them, but further consider regime-switching models which allow for random change points and a random as well as unbounded number of structural breaks as the number of observations grows.

This chapter includes the basic informations. Section 5.1 discusses the relevant literature concerning multiple change-point analysis and gives a brief general introduction to regime-switching models. Section 5.2 introduces the classical as well as the regime-switching model for changes in location and highlights the differences between these models. Section 5.3 discusses the on moving sums based test statistic, which was considered by Hušková and Slabý (2001) in case of mean changes in an otherwise independent and identically distributed random sequence. They derived the asymptotic null distribution of the test statistic. We generalise their result to the case of dependent observations. Further, we explain the estimation method for the number and locations of change points proposed, but not mathematically analysed, by Antoch et al. (2000). In the test and estimation procedure the long-run variance of the observations is involved. Since it is unknown in most applications it has to be estimated. The common long-run variance estimators as well as the modified estimators in context of AMOC models are not appropriate for the multiple change-point problem. Hence, we propose long-run variance estimators which are especially suitable in this context.

In Chapter 6 we present and explain our theoretical results. Section 6.1 discusses the consistency results of the change-point estimators with a particular focus on the comparison of the classical and the regime-switching model. We always begin with the result for the classical model and continue with the result for the regime-switching model. To begin with, we show the consistency of the estimator for the number of change points. Whereas this result holds for both models, the uniform rates of consistency for the change-point estimators differ. Under the regime-switching model the rate of convergence depends on the growth rate of the number of change points. The same rate as in the classical model can be obtained if the number of change points is stochastically bounded. Finally, we derive the joint distribution of the change-point estimators under the classical model in case of mean changes in an otherwise independent and identically distributed random sequence. The results concerning the uniform consistency and the joint asymptotic distribution of the change point estimators do not require that the number of change points is known or has been estimated correctly beforehand.

We prove the consistency results above for general long-run variance estimators fulfilling certain assumptions. These assumptions include convergence with a certain rate under the null hypothesis and stochastic boundedness under the alternative. Section 6.2 shows the consistency of the proposed variance and long-run variance estimators under the null hypothesis. The stochastic boundedness under the alternative holds for the variance estimators but not for the long-run variance estimators. However, we can show that the consistency results concerning the change points for the long-run variance estimators still hold with certain modifications on the assumptions.

Since the asymptotic results are obtained under certain assumptions on the model, we demonstrate in Section 6.3 that these assumptions are not too restrictive by presenting several models fulfilling these assumptions.

In Chapter 7 we discuss the results of a simulation study to analyse the performance of the test and estimation procedure for smaller sample sizes and to call the attention to practical issues as bandwidth selection or variance estimation.

5.1. Literature

This section gives an overview of the literature concerned with multiple change-point methods and regime-switching models. Since we are interested in the multiple change-point setting we do not discuss the vast amount of literature for AMOC models. For an introduction to the literature for AMOC models we recommend Csörgő and Horváth (1997). The literature mentioned for regime-switching models are introductory books, since we need essentially a general understanding of regime-switching models, but no extensive knowledge about specific computational aspects.

5.1.1. Multiple Change-Point Analysis

There are basically two different kind of approaches for multiple change-point problems: model selection and hypothesis testing. The first approach is based on information criteria, which measure how well a statistical model fits a given data set. Thereby, a penalty term includes the complexity of the model, i.e. the number of parameters, otherwise the model with most parameters will always fit best.

Yao (1988) was the first who used information criteria in the field of change point analysis. He related the number of change points to the complexity of the model and applied Schwarz's criterion (Schwarz (1978)) for estimating the number of changes in the mean in an otherwise independent and normally distributed sequence of random variables. The general approach is to fix an upper bound R for the number of change points r and to estimate the change points $k_1(r), \ldots, k_r(r)$ as the minimisers of a target function. Then, an information criterion is applied to select one of the R models or rather to estimate the number of change points. The target function can for example be based on the method of least squares (refer to Yao and Au (1989) in context of mean changes in an otherwise i.i.d. sequence of random variables or Liu et al. (1997) in context of multivariate regression models) or on the method of least absolute deviations (see i.g. Braun et al. (2000) for mean and simultaneous variance changes in an otherwise i.i.d. sequence of random variables or Bai (1997) in context of linear regression models). Another possibility is to base the target function on the minimum description length (refer to Rissanen (1989)), which has recently been proposed by Davis et al. (2006) in context of linear autoregressive models.

Chen et al. (2006) refined existing information criteria as the Schwarz criterion by making the model complexity not only a function of the number of change points, but also a function of the location of the change points. This was motivated by the un-necessary complexity of the model, when structural breaks appear close to each other, at the beginning or at the end of the data set. While this work was designed for a single change in a parameter of an i.i.d. sequence (i.e. the number of change points can be 0 or 1), Pan and Chen (2006) generalised this idea for multiple structural breaks. Moreover, they proposed a test based on their information criterion to test the null hypothesis of no change against the alternative of a fixed number of changes in a sequence of independent random variables.

In fact, the literature concerened with hypothesis testing in context of multiple change-point models is very sparse. Some examples are Lombard (1987), Meelis et al. (1991) and Gombay (1994) in situations with at most three change points.

Moreover, there exists some literature concerning hypothesis testing in context of multiple change-point problems which show some resemblence to the method of model selection regarding the construction of the test statistics and the computational effort. Bai and Perron (1998) introduce F-type tests for the null hypothesis of no change against the alternative of a fixed, or unknown but bounded, number of changes in context of linear regression models. Further, they introduced a test for the null hypothesis of a fixed number of changes against the alternative of an additional change, which can be used to sequentially determine the number of changes. These F-type statistics were generalized to M-type statistics by Marušiaková (2009) and also discussed in context of autoregressive models. She also derived the null asympotics of F-type statistics in context of location models with dependent errors fulfilling an α-mixing condition. More tests concerning multiple structural breaks in context of linear regression can be found in Andrews et al. (1996) where a class of finite-sample optimal tests is proposed, and Bai (1999) where a likelihood-ratio-type test is introduced.

Venter and Steel (1996) introduce a computationally intensive nonparametric test based on p-values for the case of a multiple location model, where an upper bound for the number of change points has to be fixed beforehand.

A way to consider multiple change-point problems with no restrictions on the number of change points, is an iterative method, which estimates one structural break after another and was introduced by Vostrikova (1981). She used it to detect changes in a linear trend function. As a start she proposed to test for one structural break in the whole sample. If a change point is identified, the sample is divided at the estimated change point into two subsamples. Now the same hypothesis test is applied on the first subsample and separately on the second sample. The procedure is repeated if more change points are detected and stoped otherwise. Venkatraman (1993) studied it in the context of location models, and Al Ibrahim et al. (2003) derived a test statistic from Schwarz's criterion and combined it with binary segmentation procedure in context of autoregressive models. A drawback of the binary segmentation procedure is the inability to control the overall significance level.

The method of Antoch et al. (2000), we will discuss in the following, does not require to fix an upper bound for the number of changes beforehand, is not computationally intensive and further the overall significance level can be controlled.

5.1.2. Regime-Switching Models

Regime-switching models are a very flexible method to describe non-stationary time series. These models consist of two processes: one observable process $\{X_i : i \in \mathbb{N}\}$ and a non-observable process $\{Q_i : i \in \mathbb{N}\}$, which are linked in a way that Q_i governs the distribution of X_i. For illustrational purposes we consider the example of

$$X_i = \sum_{k=1}^{K} \mathcal{I}\{Q_i = k\}\mu_k + \varepsilon_i, \qquad i = 1, \dots, n,$$

with $\mu_1, \dots, \mu_K \in \mathbb{R}$, a process $\{Q_i\}$ with values in $\{1, \dots, K\}$ and an error sequence $\{\varepsilon_i\}$ of independent standard normally distributed random variables. This model illustrates that X_i

is conditionally normal distributed and the so-called regime or state Q_i determines the mean of X_i. The above model can further be written as

$$X_i = \mu_{Q_i} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where the dependence of the mean of X_i of the regime Q_i becomes directly obvious. We note that the obervations $X_1, \ldots, X_n$ are independent when we choose $\{Q_i\}$ as an independent sequence of random variables, but dropping the assumption of independence and assuming instead $\{Q_i\}$ being a Markov chain yields correlated observations $X_1, \ldots, X_n$. The assumption of a Markov chain as a non-observable process in a regime-switching model is most common and known under Markov-switching models or hidden-Markov models. An applied book in this context is the one from MacDonald and Zucchini (1997), which deals with discrete-valued observations, whereas Cappé et al. (2005) have a more general look at hidden-Markov models. The concept of regime-switching models can be broadened by allowing X_i not only to depend on Q_i but also on past observations. An example for such a model is taken from the book of Hamilton (1994) and is given by a first-order autoregression in which both the constant term and the autoregressive coefficient are allowed to change:

$$X_i = \mu_{Q_i} + \varphi_{Q_i} X_{i-1} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

with a $\{1, \ldots, K\}$-valued non-observable process $\{Q_i\}$, errors $\{\varepsilon_i\}$ and coefficients $\mu_1, \ldots, \mu_K$, $\varphi_1, \ldots, \varphi_K$. Compared to the previous example the distribution of X_i in this model does not only depend on the Markov chain but also on the last observation X_{i-1}. Markov-switching models, which also depend on past observations, are sometimes called Markov-jump systems. Rather complex systems can be defined by letting X_i also depend on past regimes and the regimes again on past observations. Some considerations can be found in Frühwirth-Schnatter (2006), who also gives a broad introduction into regime-switching models and a lot of references to much more detailed literature.

The last example we want to mention is the CHARME (Conditional Heteroscedastic Autoregressive Mixture-of-Experts) model introduced by Stockis et al. (2010). They use a Markov chain with values in $\{1, \ldots, K\}$ and define the observable process as

$$X_i = \sum_{k=1}^{K} S_{ik}(m_k(X_{i-1}, \ldots, X_{i-p}) + \sigma_k(X_{i-1}, \ldots, X_{i-p})\varepsilon_i)$$

with

$$S_{ik} = \begin{cases} 1, & Q_i = k \\ 0, & \text{otherwise}, \end{cases}$$

unknown functions m_k, σ_k, $k = 1, \ldots, K$, and an i.i.d. sequence $\{\varepsilon_i\}$. This is a pretty general model, and we mention it since one goal for future research can be to analyse multiple change-point methods for this type of model.

In the above examples we always defined the hidden-Markov chain with a finite state space, but it is also possible to have continous state spaces. These models are also called state space models. A detailed analysis of these models can be found in Cappé et al. (2005).

In the literature, the name of regime/Markov-switching models differs a lot among research

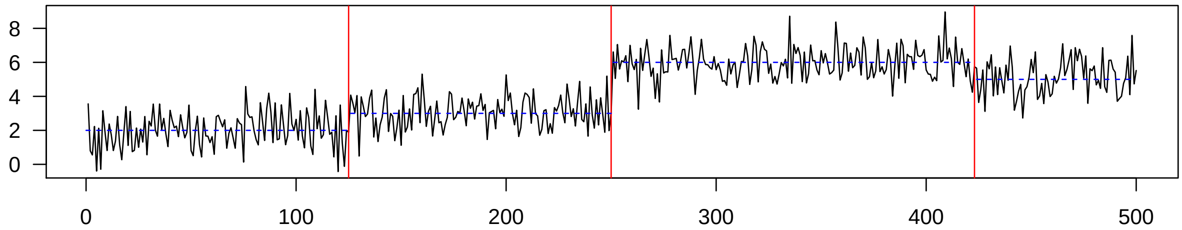

Figure 5.1.: Time series with three changes in the mean.

fields. Econometricians use Markov-switching models, biologists work with Markov-mixture models, engineers talk about hidden-Markov models and mixture-of-expert models are applied in machine-learning literature.

In context of regime-switching models a change point obviously appears when the non-observable process $\{Q_i\}$ switches to another state. Therefore regime-switching models have, in contrast to classical change-point models, random and not deterministic change points. Thus, the regime-switching model is much more flexible and additionally allows meaningful forecasting. As an example we consider a financial dataset, i.e. a stock index. We assume that during the recording of the data an event took place, which surely had an impact on the data. Examples are a national bankruptcy, a war or a natural catastrophe. If we want to make forecasts about the future developement of the index, we cannot exclude the possibility that such an event occures again. Therefore we choose a regime-switching model, and not a classical change-point model, to include the possible occurence of this event.

5.2. Models for Multiple Changes in Location

We introduce two types of models for multiple changes in the mean. Figure 5.1 shows a time series which could have been generated by each of the two models. The red vertical lines mark the change points and at each change point there is a shift in the mean, illustrated by the blue horizontal lines. First, we discuss the classical change-point model and continue with the regime-switching model.

5.2.1. Classical Multiple Change-Point Location Model

The classical change-point model, which allows for multiple changes in the mean, is defined by

$$X_i \;=\; \sum_{j=1}^{q+1} d_j \mathcal{I}\{k_{j-1} < i \le k_j\} + \varepsilon_i, \qquad i = 1, ..., n, \tag{CCM}$$

where the number of structural breaks $q \in \mathbb{N}$, the change points $k_1, \ldots, k_q$ with

$$0 = k_0 < k_1 = \lfloor \vartheta_1 n \rfloor \le ... \le k_q = \lfloor \vartheta_q n \rfloor \le k_{q+1} = n \qquad \text{and} \qquad 0 < \vartheta_1 \le \ldots \vartheta_q \le 1,$$

as well as the expected values $d_1, \ldots, d_{q+1} \in \mathbb{R}$ with $d_j \neq d_{j+1}$, $j = 1, \ldots, q$, are unknown and the stochastic error sequence $\varepsilon_1, \ldots, \varepsilon_n$ fulfills conditions specified below.

In the definition of the model the change points $k_1, \ldots, k_q$, and hence the number of observations between the change points as well, grow with an increasing sample size n. This is a necessary assumption to show asymptotic properties of the change-point methods under the alternative of at least one structural break. In comparison to the change points, the number of change points q does not depend on n. However, it would be no problem if the number of change points q depends on n as long as $q = q_n$ is bounded as n goes to infinity.

For the time being we simply assume for the error sequence $\{\varepsilon_i : 1 \le i \le n\}$:

(E1) Let the errors $\{\varepsilon_i : 1 \le i \le n\}$ be a strictly stationary sequence with

$$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E|\varepsilon_1|^{2+\nu} < \infty \text{ for some } \nu > 0,$$
$$\sum_{h\ge 0} |\gamma(h)| < \infty, \text{ where } \gamma(h) = \operatorname{cov}(\varepsilon_1, \varepsilon_{1+h}),$$

and long-run variance

$$\tau^2 := \sigma^2 + 2\sum_{h>0} \gamma(h) > 0.$$

In the sequel we need further conditions on the errors, but we will introduce and explain them when they are needed. An overview of all assumptions can be found in Appendix A.

Besides the common error properties as $E\varepsilon_1 = 0$ and $0 < \tau^2 < \infty$, where the existence of τ^2 follows from the existence of the sum of the absolute autocovariances, the conditions stated in **(E1)** are assumed to ensure the validness of the invariance principle in assumption **(E2)** and a further moment condition in **(E3)**. Basically it would be enough to assume **(E2)** and **(E3)**, but for all well-known error sequences, such as independent and identically distributed, α-mixing or Markov sequences, **(E1)** nevertheless has to be assumed to ensure **(E2)** and **(E3)**. We refer for details to Section 6.3.

5.2.2. Regime-Switching Location Model

A regime-switching model, which describes changes in the mean, is given by

$$X_i \;=\; d^*_{Q_i} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

with potential expectations $d_1^*, \ldots, d_K^* \in \mathbb{R}$, where $d_i^* \ne d_j^*$ for $i, j = 1, \ldots, K$, $i \ne j$, and errors $\varepsilon_1, \ldots, \varepsilon_n$ fulfilling **(E1)**. The expectation of observation X_i is random and determined by a non-observable $\{1, ..., K\}$-valued stationary process $\{Q_i : 1 \le i \le n\}$, $K \in \mathbb{N}$. Is Q_i for example in state 3, observation X_i has expectation d_3^*.
For our purposes the key features of the non-observable process $\{Q_i\}$ are long duration times, since in the regime-switching model a change point appears, when the non-observable process $\{Q_i\}$ switches to another state. Since we do not want to have a different expectation in each new observation, we do not choose an i.i.d. sequence for the non-observable process $\{Q_i\}$, because here the probability would be high to switch at each time point to another state. A good approach is to choose for the non-observable process $\{Q_i\}$ a Markov chain with an adequate transition matrix, i.e. the probablity to stay in the same state should be high.

For the classical model **(CCM)** we assumed that the distance between two adjacent change points goes to infinity as n goes to infinity in order to obtain asymptotic results under the alternative. To receive a similar setting in the regime-switching model the duration times have to go to infinity as n goes to infinity. To ensure this, the non-observable process $\{Q_i\}$ has to depend on the number of observations n. Hence, we alter the regime-switching model into a triangular scheme:

$$X_i^{(n)} = d^*_{Q_i^{(n)}} + \varepsilon_i^{(n)}$$

with a non-observable $\{1, ..., K\}$-valued stationary process $\{Q_i^{(n)} : 1 \leq i \leq n\}$, $K \in \mathbb{N}$. Thus, we are able to analyse the regime-switching model asymptotically as well. This model is similar to other situations in statistics, e.g. the asymptotic framework by Dahlhaus (1997) to analyse locally stationary processes.

In the example with the Markov chain as non-observable process, the transition matrix of the Markov Chain $\{Q_i^{(n)}\}$ has to depend on n in a way that the probability to stay in the same state tends to 1 as n tends to infinity.

We note that in contrast to the classical model the regime-switching model has random expectations, random change points, and a random number of them, since $\{Q_i^{(n)}\}$ is a random process. Additionally the number of change points depends on n, which we stress by the notation q_n, and is allowed to tend to infinity as n goes to infinity. Although we want to point out the differences we want to illustrate the similarity between the classical and the regime-switching model as well. Thus, we write the regime-switching model in terms of the classical location model **(CCM)**:

$$X_i = \sum_{j=1}^{q_n+1} d_j \mathcal{I}\{k_{j-1} < j \leq k_j\} + \varepsilon_i, \qquad i = 1, ..., n, \qquad \text{and} \quad d_j \in \{d_1^*, \ldots, d_K^*\} \qquad \textbf{(RSM)}$$

where the number of change points q_n, the change points $k_1, ..., k_{q_n}$ as well as $d_1, \ldots, d_{q_n+1}$ are random variables.

5.3. Test and Estimation Procedure

The last section introduced the classical and the regime-switching model which allow for several changes in the expectation in an otherwise stationary sequence of random variables. In Section 5.3.1 we are interested in whether changes occured or not. For this problem we consider a moving-sums based test statistic, which was in detail investigated by Hušková and Slabý (2001) in case of a location model with i.i.d. errors, and test the null hypothesis of no change against the alternative of at least one structural break. If the null hypothesis is rejected we want to find out how many structural breaks are present and where they are. To this end Section 5.3.2 considers estimators introduced, but not mathematically analysed, in Antoch et al. (2000). In the test and estimation procedure the long-run variance of the errors is involved. Since in most applications the long-run variance is unknown Section 5.3.3 introduces variance and long-run variance estimators for the multiple change-point problem.

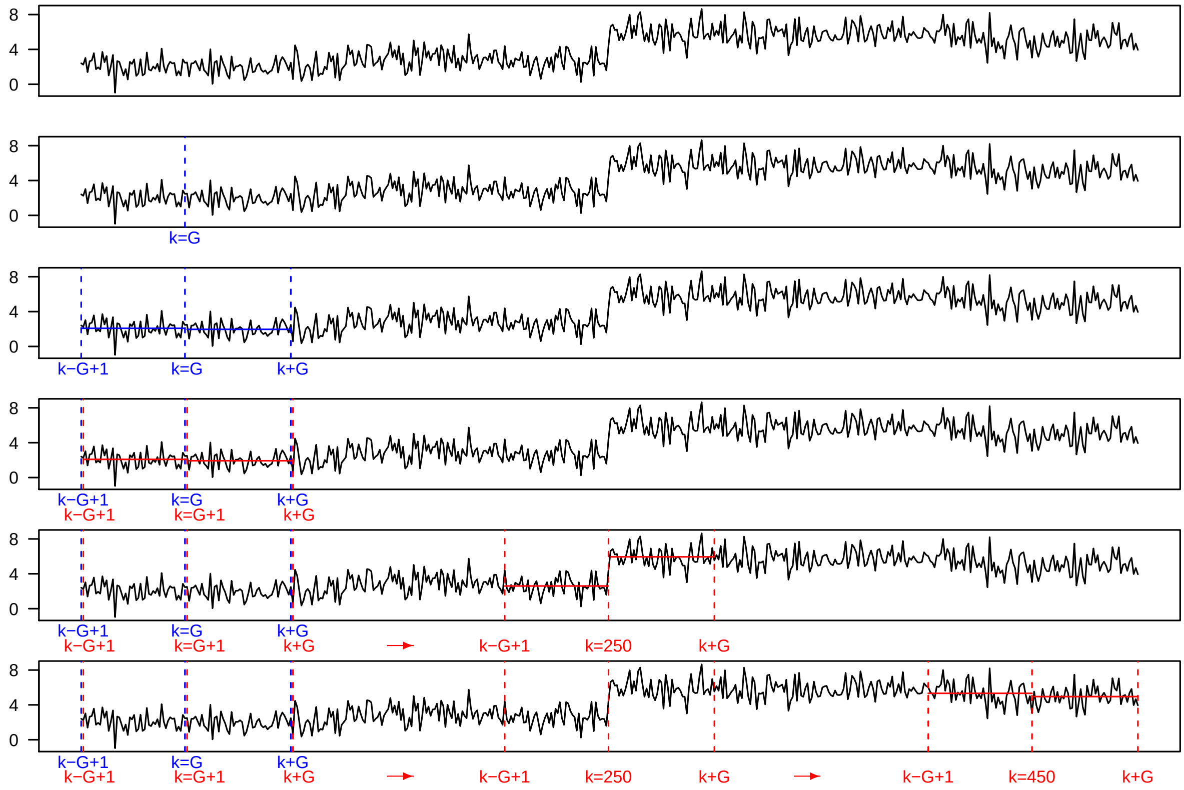

Figure 5.2.: Illustration of the MOSUM procedure.

5.3.1. Moving-Sum Statistic

We are interested in testing the null hypothesis of no change

$$H_0: \qquad k_1 = \ldots = k_q = n$$

against the alternative of at least one structural break

$$H_1: \qquad k_1 \leq \ldots \leq k_q, \text{ where at least one inequality is strict.}$$

The MOSUM statistic discussed in Hušková and Slabý (2001) is given by

$$\max_{G \leq k \leq n-G} \frac{|T_{k,n}(G)|}{\tau} \quad \text{with} \quad T_{k,n}(G) := \frac{1}{\sqrt{2G}} \left(\sum_{i=k+1}^{k+G} X_i - \sum_{i=k-G+1}^{k} X_i \right), \tag{5.1}$$

and bandwidth $G = G(n)$. This test statistic compares for every time point $G \leq k \leq n - G$, the estimated mean of the first subsample $X_{k-G}, \ldots, X_k$ to the mean of the second subsample $X_{k+1}, \ldots, X_{k+G}$. A significant difference in the mean for some $G \leq k \leq n - G$ indicates a change. This is illustrated in Figure 5.2. We start at time point $k = G$ and define a window of size $2G$, which includes the observations $X_1, \ldots, X_{2G}$. Comparing the mean of $X_1, \ldots, X_G$ to the mean of $X_{G+1}, \ldots, X_{2G}$ yields no big differences. Then, we move the window to the next time point $k + 1$ and repeat the comparison with the sample $X_2, \ldots, X_{2G+1}$. Subsequently the window of size $2G$ is moved in terms of k along the sample. Therefore, this statistic is called a moving-sum (MOSUM) statistic. The class of MOSUM statistics has been developed and analysed for many different kind of AMOC models, e.g. in Bauer and Hack (1980), Chu et al. (1995) or Hušková (1990).

To obtain asympotic results we let the following assumption on the bandwidth G hold:

(G) For $n \to \infty$ let

$$\frac{n}{G} \longrightarrow \infty \qquad \text{and} \qquad \frac{n^{\frac{2}{2+\nu}} \log n}{G} \longrightarrow 0.$$

This ensures that the bandwidth G goes to infinity slower than n, but faster than $n^{\frac{2}{2+\nu}} \log n$. The ν in the above expression is the same ν as in the moment condition of the errors in **(E1)**. Hence, there is a connection between the number of moments existing and the strength of the bandwidth assumption. The higher moments exist the weaker is the bandwidth assumption.

To construct an asymptotic test with level α we need the asymptotic distribution under the null hypothesis, which was obtained under the assumption of i.i.d. errors and a known variance σ^2 by Hušková and Slabý (2001). To use their result we assume for the errors an invariance principle, which ensures that the asymptotic behaviour of the test statistic does not depend on the error distribution. Hence, we assume **(E2)**, which is given by

(E2) There exists a standard Wiener process $\{W(k): 1 \le k \le n\}$ such that

$$\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \frac{1}{\tau} \sum_{i=k+1}^{k+G} \varepsilon_i - (W(k+G) - W(k)) \right| = o_p\left((\log(n/G))^{-\frac{1}{2}} \right).$$

Since in most applications the long-run variance τ^2 is unknown, we use an estimator $\hat{\tau}^2_{k,n}$ and define the test statistic $T_n(G)$ as

$$T_n(G) \quad := \quad \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}}. \tag{5.2}$$

We note that in the literature usually variance estimators $\hat{\tau}^2_n$ are suggested, but these are not appropriate for the multiple change-point setting (we refer for more details to Section 5.3.3). Instead, we estimate the long-run variance τ^2 at each time point $G \le k \le n - G$ by an estimator $\hat{\tau}^2_{k,n} = \hat{\tau}^2_{k,n}(G)$ depending on the bandwidth G in a way that we only use observations $X_{k-G}, \ldots, X_{k+G}$. The next section introduces a specific estimator $\hat{\tau}^2_{k,n}$, but to derive the asymptotic distribution of $T_n(G)$ the estimator $\hat{\tau}^2_{k,n}$ only need to fulfill **(L0)** given by

(L0) Let the long-run variance estimators $\hat{\tau}^2_{k,n}$ be based only on observations $X_{k-G+1}, \ldots, X_{k+G}$ and fulfill

$$\max_{G \le k \le n-G} |\hat{\tau}^2_{k,n} - \tau^2| \quad = \quad o_p\left((\log(n/G))^{-1} \right) \text{ under } H_0.$$

The following theorem gives the asymptotic distribution of $T_n(G)$ under the null hypothesis.

Theorem 5.1. Let $X_1, ..., X_n$ follow the classical model **(CCM)** or the regime switching model **(RSM)** and let the assumptions **(E1)**, **(E2)** on errors, **(G)** on the bandwidth G and **(L0)** on the long-run variance estimator $\hat{\tau}^2_{k,n}$ hold. Then, under H_0,

$$a(n/G)\, T_n(G) - b(n/G) \quad \xrightarrow{\mathcal{D}} \quad \Gamma,$$

where the random variable Γ follows a Gumbel extreme value distribution, i.e. $P(\Gamma \leq x) = \exp(-2\exp(-x))$ and

$$a(x) = \sqrt{2\log x},\ b(x) = 2\log(x) + \frac{1}{2}\log\log x + \log(3/2) - \frac{1}{2}\log\pi.$$

Proof of Theorem 5.1. By Theorem 2.1 in Hušková and Slabý (2001) it holds for i.i.d. errors

$$a(n/G) \max_{G\leq k\leq n-G} \frac{|T_{k,n}(G)|}{\tau} - b(n/G) \quad \xrightarrow{\mathcal{D}} \quad \Gamma,$$

under bandwidth assumptions $\frac{G}{n}\log(n/G) \to 0$ and $\frac{n^{\frac{2}{2+\nu}}}{G} \to 0$. These assumptions hold under **(G)**, since

$$\frac{G}{n}\log(n/G) \quad = \quad \frac{\log(n/G)}{n/G} \longrightarrow 0 \qquad \text{and} \qquad \frac{n^{\frac{2}{2+\nu}}}{G} \quad \leq \quad \frac{n^{\frac{2}{2+\nu}}\log n}{G} \quad \longrightarrow \quad 0.$$

In **(E2)** is assumed that the not necessarily independent errors obey a strong invariance principle, which ensures that asymptotically the behaviour of the standardised test statistic does not depend on the error distribution. This follows by

$$\begin{aligned}
&\max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G\tau^2}} \left| \sum_{i=k+1}^{k+G} X_i - \sum_{i=k-G+1}^{k} X_i \right| \quad = \quad \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G\tau^2}} \left| \sum_{i=k+1}^{k+G} \varepsilon_i - \sum_{i=k-G+1}^{k} \varepsilon_i \right| \\
&= \quad \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G\tau^2}} \Bigg| (W(k+G) - W(k)) + \left(\sum_{i=k+1}^{k+G} \varepsilon_i - (W(k+G) - W(k)) \right) \\
&\qquad - (W(k) - W(k-G)) - \left(\sum_{i=k-G+1}^{k} \varepsilon_i - (W(k) - W(k-G)) \right) \Bigg| \\
&= \quad \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G\tau^2}} \left| (W(k+G) - W(k)) - (W(k) - W(k-G)) \right| + o_p\left((\log(n/G))^{-\frac{1}{2}} \right),
\end{aligned}$$

where the last step is obtained by the reverse triangle inequality and assumption **(E2)**. Moreover the increments of the Wiener process can be written as

$$\tau^{-1}\left(W(k+G) - W(k) - (W(k) - W(k-G))\right) \quad = \quad \sum_{i=k+1}^{k+G} \tilde{\varepsilon}_i - \sum_{i=k-G+1}^{k} \tilde{\varepsilon}_i$$

with standard normal distributed random variables $\tilde{\varepsilon}_i$, $i = k - G + 1, \ldots, k$. Since

$$a(n/G) o_p\left((\log(n/G))^{-\frac{1}{2}}\right) \quad = \quad o_p(1)$$

it is enough to derive the asymptotic distribution for independent standard normally distributed errors, which is done in the proof of Hušková and Slabý (2001). Furthermore we have to show that the limit distribution does not change if the long-run variance τ^2 is estimated by $\hat{\tau}^2_{k,n}$. We receive

$$\begin{aligned}
& a(n/G) \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} - b(n/G) \\
&= a(n/G) \left(\max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} - \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\tau} \right) + a(n/G) \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\tau} \\
&\quad - b(n/G) \\
&= a(n/G) \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\tau} - b(n/G) + o_p(1) \quad \xrightarrow{\mathcal{D}} \quad \Gamma
\end{aligned}$$

by Theorem 2.1 in Hušková and Slabý (2001) and

$$\begin{aligned}
& a(n/G) \left| \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} - \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\tau} \right| \\
&\le a(n/G) \max_{G \le k \le n-G} \left| \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} - \frac{|T_{k,n}(G)|}{\tau} \right| \\
&= a(n/G) \max_{G \le k \le n-G} \left| \frac{1}{\hat{\tau}_{k,n}} - \frac{1}{\tau} \right| |T_{k,n}(G)| \\
&\le \sqrt{2\log(n/G)} \max_{G \le k \le n-G} \left| \frac{1}{\hat{\tau}_{k,n}} - \frac{1}{\tau} \right| \max_{G \le k \le n-G} |T_{k,n}(G)| \\
&= o_p(1),
\end{aligned}$$

since

$$a(n/G) \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\tau} - b(n/G) \quad \xrightarrow{\mathcal{D}} \quad \Gamma$$

yields

$$\begin{aligned}
\max_{G \le k \le n-G} |T_{k,n}(G)| &= \tau \max_{G \le k \le n-G} \frac{|T_{k,n}(G)|}{\tau} = \tau \left(\frac{b(n/G)}{a(n/G)} + O_p\left(\frac{1}{a(n/G)}\right)\right) \\
&= O_p\left(\frac{b(n/G)}{a(n/G)}\right) = O_p(\sqrt{\log(n/G)}) \qquad (5.3)
\end{aligned}$$

and by assumption **(L0)**

$$\max_{G\le k\le n-G}\left|\frac{1}{\hat\tau_{k,n}}-\frac{1}{\tau}\right| = \max_{G\le k\le n-G}\frac{|\hat\tau_{k,n}-\tau|}{\tau^2+o_p(1)} = \max_{G\le k\le n-G}\frac{|\hat\tau^2_{k,n}-\tau^2|}{|\hat\tau_{k,n}+\tau|}\frac{1}{\tau^2+o_p(1)}$$
$$= o_p\left((\log(n/G))^{-1}\right).$$

Then, the assertion is shown. □

We can now construct a test with asymptotic level α, by rejecting the null hypothesis if $T_n(G) > D_n(G;\alpha)$ with critical value

$$D_n(G;\alpha) = \frac{b(n/G)-\log\log\frac{1}{\sqrt{1-\alpha}}}{a(n/G)}. \tag{5.4}$$

5.3.2. Estimation of Number and Locations of Change Points

In case the null hypothesis is rejected, we estimate the number of change points q and the change points $k_1,\ldots,k_q$ by a method proposed, but not mathematically investigated, by Antoch et al. (2000). To illustrate the idea of the estimators Figure 5.3 has been included. The upper picture shows a time series with three structural breaks marked by red vertical lines. The lower picture presents the performance of the appended statistic $\hat\tau^{-1}_{k,n}|T_{k,n}(G)|$, where the estimator $\hat\tau^2_{k,n}$ is defined later on in (5.12). The horizontal line indicates the critical value $D_n(G;\alpha)$. The test statistic exceeds the critical value in intervals around the true change points. Therefore it seems to be a good idea to use those time points k where $\hat\tau^{-1}_{k,n}|T_{k,n}(G)|$ exceeds the critical value $D_n(G;\alpha)$ to estimate the change points $k_1,...,k_q$. Hence, we choose all pairs of indices (v_j,w_j) such that

$$\hat\tau^{-1}_{k,n}|T_{k,n}(G)| \ge D_n(G;\alpha_n) \quad \text{for} \quad k=v_j,...,w_j, \tag{5.5}$$
$$\hat\tau^{-1}_{k,n}|T_{k,n}(G)| < D_n(G;\alpha_n) \quad \text{for} \quad k=v_j-1,w_j+1, \tag{5.6}$$
$$w_j-v_j \ge \varepsilon G \quad \text{with } 0<\varepsilon<1/2 \text{ arbitrary but fixed.} \tag{5.7}$$

In Figure 5.3 the statistic takes its local maxima approximately at the true change points. Hence, we define

$$\hat k_j := \arg\max_{v_j\le k\le w_j}\frac{|T_{k,n}(G)|}{\hat\tau_{k,n}} \tag{5.8}$$

as an estimator for a change point and

$$\hat q_n \mathrel{\widehat{=}} \text{the number of pairs } (v_j,w_j) \tag{5.9}$$

as the estimator of the number of change points q. A minimum length εG of the interval $[v_j,w_j]$ is assumed to exclude the case of an additionally estimated change point if the test

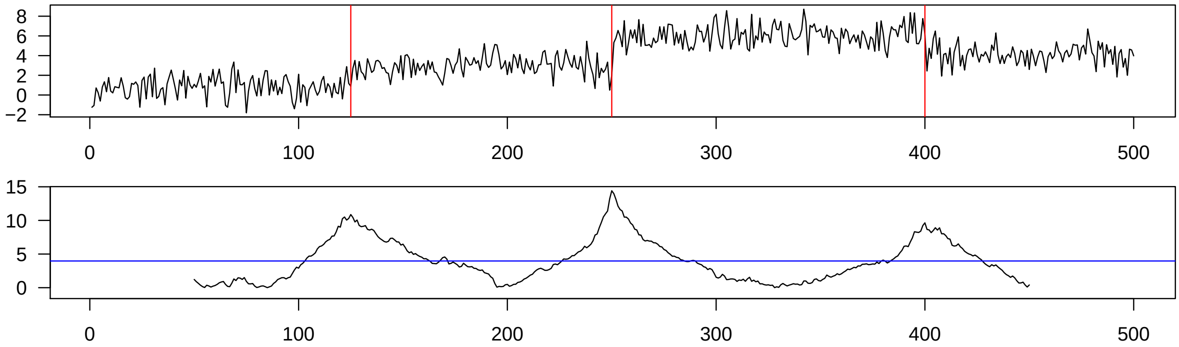

Figure 5.3.: The upper picture shows a time series with three change points marked by the red vertical lines. Below the behaviour of the MOSUM statistic $\hat{\tau}_{k,n}^{-1}|T_{k,n}(G)|$ is illustrated and the horizontal blue line presents the critical value $D_n(G;\alpha)$.

statistic does not increase monotonously after exceeding the critical value, but falls below the critical value again for a short time period. Respectively, if the test statistic does not decrease monotonously after falling below the critival value, but exceeds the critical value again for a short time period. We refer for illustration to Figure 5.3. The condition is also necessary for the proof of consistency of the estimator for the number of change points $\hat{q}_n$ (refer to the proof of Theorem 6.1). Due to the definition of the minimum length of the intervals, in applications the situation can hypothetically occur that $\max_{G\leq k\leq n-G} \hat{\tau}_{k,n}^{-1}|T_{k,n}(G)| > D_n(G;\alpha)$ while at the same time no change points are estimated. In this case the test nevertheless rejects the null hypothesis.

5.3.3. Estimation of the Long-Run Variance

We start with the case of independent and identically distributed errors. The standard variance estimator

$$\hat{\sigma}_n^2 \quad = \quad \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \tag{5.10}$$

is not suitable for change-point problems, since it does not take possible structural breaks into account and overestimates the error variance (refer to Figure 5.4). To overcome this problem

$$\frac{1}{n}\left(\sum_{i=1}^{\hat{k}}(X_i - \overline{X}_{1,\hat{k}})^2 + \sum_{i=\hat{k}+1}^{n}(X_i - \overline{X}_{\hat{k}+1,n})^2\right)$$

with $\hat{k} := \arg\max_{G\leq k<n-G} T_{k,n}(G)$ and $\overline{X}_{j,l}$ denoting the mean of observations $X_j,\ldots,X_l$, $j,l = 1,\ldots,n$, was proposed as variance estimator in the context of AMOC models (refer to Hušková et al. (2007)). Here, a potential change point $\hat{k}$ is taken into account. This idea can not be

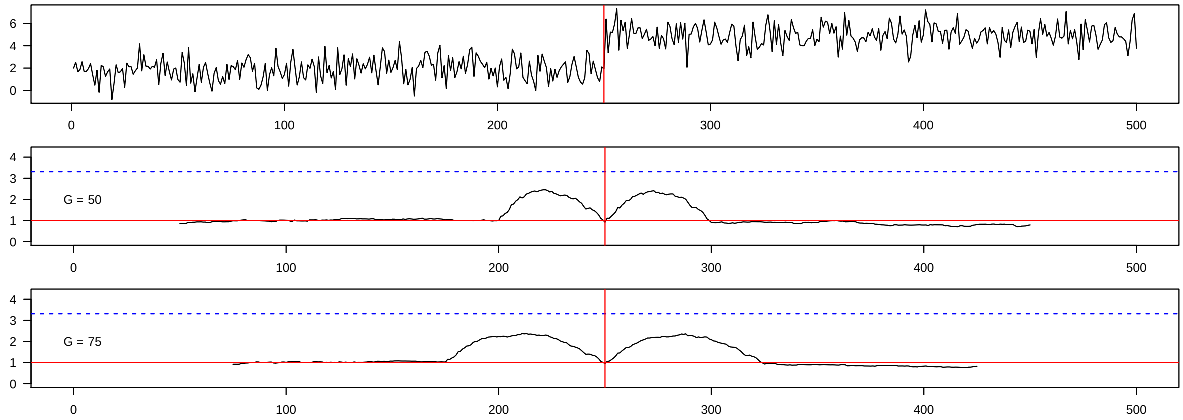

Figure 5.4.: Performance of the Variance estimators $\hat{\sigma}^2_{k,n}$ (black line) and $\hat{\sigma}^2_n$ (blue dotted line) in comparison to the true variance (the red line).

generalised to multiple change-point models, since we do not know how many change points we have to take into account. Instead we will use the estimator

$$\hat{\sigma}^2_{k,n} \quad := \quad \frac{1}{2G}\left(\sum_{i=k-G+1}^{k}(X_i - \overline{X}_{k-G+1,k})^2 + \sum_{i=k+1}^{k+G}(X_i - \overline{X}_{k+1,k+G})^2\right) \tag{5.11}$$

as variance estimator. To illustrate the idea and performance of this estimator Figure 5.4 is included. The upper picture shows a time series with one change point, which is marked by the red vertical line. In the two lower pictures the red line is the true variance σ^2 and the blue dashed line is the standard variance estimator $\hat{\sigma}^2_n$, which clearly overestimates the variance σ^2. The black line is $\hat{\sigma}^2_{k,n}$ for bandwidth $G = 50$ (upper picture) and $G = 75$ (lower picture). The estimator $\hat{\sigma}^2_{k,n}$ performes well far away from the change point and directly at the change point, but overestimates the variance near the change point. This is due to the definition of $\hat{\sigma}^2_{k,n}$. For each time point k a window of length $2G$ is defined and the variance of the samples on the left side of the window is added to the variance of the samples on the right side of the window. Hence, the variance is estimated correctly at the change point $k = 250$, but overestimated near the change point, since e.g. at $k = 220$ the right side of the window includes the change point. The window does not include a change point if the distance between k and a change point is larger than G. Hence, for these time points the variance is estimated correctly.

In case of dependent errors we have to estimate the long-run variance τ^2. This is already a difficult task if we are not in a change-point setup, but have a stationary sequence of random variables. The standard long-run variance estimator is the Bartlett estimator

$$\hat{\tau}^2_n \quad := \quad \hat{\gamma}_n(0) + 2\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\hat{\gamma}_n(h)$$

with autocovariance estimator

$$\hat{\gamma}_n(h) \quad := \quad \frac{1}{n}\sum_{i=1}^{n-h}(X_i - \overline{X}_{1,n})(X_{i+h} - \overline{X}_{1,n}),$$

Bartlett weights

$$\omega(h/\Lambda_n) \quad := \quad (1 - h/\Lambda_n) I\{|h| \leq \Lambda_n\}$$

and bandwidth Λ_n. Berkes et al. (2005) have shown the almost sure convergence of this estimator under weak and strong dependence assumptions. The performance of this estimator is improved by using flat-top kernels instead of Bartlett weights, see e.g. Politis and Romano (1995). The flat-top kernels are defined by

$$w(t) = \begin{cases} 1, & |t| \leq 1/2, \\ 2(1 - |t|), & 1/2 < |t| < 1, \\ 0, & |t| \geq 1. \end{cases}$$

An adapted version of $\hat{\tau}_n^2$ for the AMOC model was discussed in Hušková and Kirch (2010), who have also shown the consistency of their estimator under H_0 and H_1.

For the multiple change-point setting we propose the estimator

$$\hat{\tau}_{k,n}^2 \quad := \quad \hat{\gamma}_k(0) + 2 \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \hat{\gamma}_k(h) \tag{5.12}$$

with autocovariance estimator

$$\begin{aligned} \hat{\gamma}_k(h) \quad := \quad & \frac{1}{2G} \sum_{i=k-G+1}^{k-h} (X_i - \overline{X}_{k-G+1,k})(X_{i+h} - \overline{X}_{k-G+1,k}) \\ & + \frac{1}{2G} \sum_{i=k+1}^{k+G-h} (X_i - \overline{X}_{k+1,k+G})(X_{i+h} - \overline{X}_{k+1,k+G}), \end{aligned}$$

suitable weights ω and bandwidth Λ_n.

For the asymptotic results of the change-point estimators we do not need the specific form of the estimator $\hat{\tau}_{k,n}^2$, but general properties including rates of uniform convergence of $\hat{\tau}_{k,n}^2$ under H_0 and H_1 specified in **(L0)** and

(L1) Let the long-run variance estimator $\hat{\tau}_{k,n}^2$ be translation invariant, i.e.

$$\hat{\tau}_{k,n}(X_{k-G}, \ldots, X_{k+G}) = \hat{\tau}_{k,n}(X_{k-G} + c, \ldots, X_{k+G} + c), \qquad \forall c \in \mathbb{R},$$

and fulfill

$$\max_{G \leq k \leq n-G} \hat{\tau}_{k,n}^2 \quad = \quad O_p(1) \ \text{ under the alternative}$$

as well as

$$P\left(\min_{G \leq k \leq n-G} \hat{\tau}_{k,n} > 0 \right) \quad \longrightarrow \quad 1.$$

The assumption of translation invariance is not a strong assumption, since all reasonable long-run variance estimators we know have this property.

Remark 5.2. In some situations it can be helpful to use slightly modified variance estimators

$$\begin{aligned}\hat{\sigma}^2_{k,n,l} &:= \frac{1}{G}\sum_{i=k-G+1}^{k}(X_i-\overline{X}_{k-G+1,k})^2,\\ \hat{\sigma}^2_{k,n,r} &:= \frac{1}{G}\sum_{i=k+1}^{k+G}(X_i-\overline{X}_{k+1,k+G})^2,\\ \hat{\sigma}^2_{k,n,m} &:= \min(\hat{\sigma}^2_{k,n,l},\hat{\sigma}^2_{k,n,r}).\end{aligned}$$

This is for example the case if the observations are allowed to change simultaneously in mean and variance. Clearly this situation requires an error sequence which does not fulfill the assumption of strict stationarity, but often occures in practise. We will see in Section 7 why in this situation ist is reasonable to use a modified variance estimator. The modified estimators are not as stable as the variance estimator $\hat{\sigma}^2_{k,n}$, since they are based on G instead of $2G$ observations. Thus, they should only be used with large bandwidth G.

In the dependent case we define long-run variance estimators $\hat{\tau}^2_{k,n,l}$ and $\hat{\tau}^2_{k,n,r}$ as in equation (5.12) but with autocovariance estimators

$$\begin{aligned}\hat{\gamma}_{k,l}(h) &:= \frac{1}{G}\sum_{i=k-G+1}^{k-h}(X_i-\overline{X}_{k-G+1,k})(X_{i+h}-\overline{X}_{k-G+1,k}),\\ \hat{\gamma}_{k,r}(h) &:= \frac{1}{G}\sum_{i=k+1}^{k+G-h}(X_i-\overline{X}_{k+1,k+G})(X_{i+h}-\overline{X}_{k+1,k+G}).\end{aligned}$$

Moreover, we define $\hat{\tau}^2_{k,n,m} := \min(\hat{\tau}^2_{k,n,l},\hat{\tau}^2_{k,n,r})$. Since we already need large bandwidths G to achieve a relatively good performance of the estimator $\hat{\tau}^2_{k,n}$, the bandwidth G has to be huge to obtain acceptable results by the modified long-run variance estimators. ■

6. Theoretical Results

We have seen that heuristically the MOSUM procedure is a reasonable instrument to detect and estimate structural changes. However, Section 6.1 considers the MOSUM procedure in a theoretical way, i.e. it analyses the consistency of the estimators for the number and locations of structural breaks and obtains the joint asymptotic distribution of the change-point estimators. To obtain the asymptotic results we have to distinguish carefully between the classical and the regime-switching model. Section 6.2 considers the asymptotic behaviour of the long-run variance estimators proposed in Section 5.3.3 under the null hypothesis and the alternative. All these results are obtained under certain conditions on the error sequence. To demonstrate that these are not too restrictive, Section 6.3 includes some examples of error sequences which fulfill the conditions.

6.1. Consistency of the Estimators

For the asymptotic analysis of the change-point estimators it is important to recall the differences between the classical and the regime-switching model: first the classical model assumes deterministic change points whereas the regime-switching model has random change points, secondly in contrast to the deterministic and bounded number of changes in the classical model the regime-switching model has a random and potentially unbounded number, and third the expectations in the regime-switching model are not deterministic as in the classical model but randomly.

Section 6.1.1 proves the consistency of the estimator for the number of change points in case of the classical as well as the regime-switching model. Then, in Section 6.1.2 we obtain a uniform rate of consistency for the change-point estimators under the classical and the regime-switching model, respectively. The rates differ, since the uniform rate of convergence under the regime-switching model depends on the growth rate of the number of change points. But if the number of changes is stochastically bounded we obtain the same rate. Based on the uniform consistency result we derive in Section 6.1.3 the joint asymptotic distribution of the change-point estimators under the classical model.

6.1.1. Consistency of the Estimator for the Number of Change Points

To begin with, we prove the consistency of the estimator for the number of change points $\hat{q}_n$ under the classical model **(CCM)**. Thereafter, we discuss the alterations we conduct to obtain the consistency result for the regime-switching model **(RSM)**.

We recall that the number of change points is estimated by the number of intervals, where the statistic $\hat{\tau}_{k,n}^{-1} T_{k,n}(G)$ exceeds the critical value $D_n(G; \alpha_n)$ (refer to (5.9)). The performance of the statistic depends on the window length $2G$, since the statistic compares the mean of the subsample on the left side of the window to the mean of the subsample on the right side. For the asymptotic results we need that this window asymptotically does not include more than one change point. In model **(CCM)** we already assumed the distance between two adjacent change points to be tending to infinity as n goes to infinity, but we further need the minimum distance between two adjacent change points to be asymptotically larger than $2G$. Hence, we assume

(Cd) Let for the minimum distance between two adjacent change points hold that

$$\limsup_{n\to\infty} d_0(n)/G = C > 2 \qquad \text{with} \qquad d_0(n) := \min_{0\le j\le q} |k_{j+1} - k_j|.$$

Further, we consider fixed as well as local changes. Thus, we let the expectations $d_1, \ldots, d_{q+1}$ depend on n, i.e. we consider $d_{1,n}, \ldots, d_{q+1,n}$, and let the following assumption hold:

(D) Let $\tilde{d}_{j,n} := d_{j+1,n} - d_{j,n}$, $j = 1, \ldots, q$, fulfill

$$\min_{1\le j\le q} \tilde{d}^2_{j,n} G(\log(n/G))^{-1} \longrightarrow \infty.$$

To obtain results under the alternative H_1, we let the level $\alpha = \alpha_n$ depend on n and fulfill

(A) Let $\{\alpha_n : n \in \mathbb{N}\}$ satisfy

$$\alpha_n \longrightarrow 0 \qquad \text{and} \qquad \frac{\log\log \frac{1}{\sqrt{1-\alpha_n}}}{a(n/G)} = O(1).$$

The next theorem shows that $\hat{q}_n$ defined as in (5.9) is a consistent estimator of the true number of change points q.

Theorem 6.1. Let $X_1, ..., X_n$ follow the classical model **(CCM)**. Further, let the assumptions **(E1)**, **(E2)** on the errors, **(G)** on the bandwidth G, **(A)** on the level α_n, **(Cd)** on the change points and **(D)** on the expectations hold. Moreover, assume **(L1)** for the long-run variance estimator $\hat{\tau}^2_{k,n}$ and let the error sequence fulfill **(L0)**. Then, under H_1,

$$P(\hat{q}_n = q) \quad \longrightarrow \quad 1 \qquad \text{as } n \to \infty.$$

Proof of Theorem 6.1. To estimate the correct number of change points q, the number of intervals $[v_j, w_j]$ has to be equal to q, but $[v_j, w_j]$, $j = 1, \ldots, q$, do not necessarily have to cover the change points k_j, $j = 1, \ldots, q$. But the proof will show that the desired covering of the change points by these intervals will be indeed the case.
To detect a change point the statistic $\hat{\tau}^{-1}_{k,n} T_{k,n}(G)$ has to exceed the critical value $D_n(G; \alpha_n)$ in an interval larger than εG with $\varepsilon < 1/2$ (refer to (5.7)). Defining the set

$$B_{G,q} \quad := \quad \{k \in \{1, \ldots, n\} : \exists k_j \in \{k_1, \ldots, k_q\} \text{ with } 0 \le |k - k_j| < (1-\varepsilon)G\} \tag{6.1}$$

and demanding $\hat{\tau}^{-1}_{k,n} T_{k,n}(G) > D_n(G; \alpha_n)$ for all $k \in B_{G,q}$ ensures the estimation of at least q change points. However, we do not want to estimate more structural breaks than q, thus the test statistic $\hat{\tau}^{-1}_{k,n} T_{k,n}(G)$ has to be smaller than the critical value $D_n(G, \alpha_n)$ for all $k \in A_{G,q}$ with

$$A_{G,q} \quad := \quad \{k \in \{1, \ldots, n\} : |k_j - k| \ge G, \ \forall\, k_1, \ldots, k_q\}. \tag{6.2}$$

Choosing $A_{G,q}$ and $B_{G,q}$ as above omits the indices $(1-\varepsilon)G \le |k - k_j| < G$, $j = 1, \ldots, q$. This is done, because the asymptotic behaviour of $\hat{\tau}^{-1}_{k,n} T_{k,n}(G)$ can be determined for $k \in A_{G,q}$ and $B_{G,q}$, but not for the omitted indices. But by requiring (5.7) it is ensured that it does not matter how the test statistic behaves in the omitted intervals, because in these intervals no additional change point can be estimated since the intervals are too small. This leads to

$$
\begin{aligned}
&P(\hat{q}_n = q) && (6.3)\\
&\geq\; P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n),\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n)\right)\\
&\geq\; 1 - P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n)\right) - P\left(\min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)\right),
\end{aligned}
$$

where the last line follows from the inequality

$$P(A\cap B) \;=\; 1 - P(A^C \cup B^C) \geq 1 - P(A^C) - P(B^C).$$

Thus, it is sufficient to show

$$P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n)\right) \;\longrightarrow\; 0 \tag{6.4}$$

and

$$P\left(\min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)\right) \;\longrightarrow\; 0. \tag{6.5}$$

To prove (6.4) and (6.5) $T_{k,n}(G)$ is split into two parts. Therefore we define

$$T^0_{k,n}(G) := \frac{1}{\sqrt{2G}}\left(\sum_{i=k+1}^{k+G} \varepsilon_i - \sum_{i=k-G+1}^{k} \varepsilon_i\right) \tag{6.6}$$

and receive by the definition of the observations

$$
\begin{aligned}
T_{k,n}(G) \;&=\; T^0_{k,n}(G) + R_{k,n}(G) \qquad \text{with}\\
R_{k,n}(G) \;&:=\; \frac{1}{\sqrt{2G}}\left(\sum_{i=k+1}^{k+G}\sum_{j=1}^{q+1} d_{j,n} I\{k_{j-1} < i \leq k_j\} - \sum_{i=k-G+1}^{k}\sum_{j=1}^{q+1} d_{j,n} I\{k_{j-1} < i \leq k_j\}\right).
\end{aligned}
$$

To prove (6.4), we note that by definition of $A_{G,q}$ the observations $X_{k-G+1},\ldots,X_{k+G}$ have the same mean for $k\in A_{G,q}$. Hence, we receive $R_{k,n}(G) = 0$ for $k\in A_{G,q}$. Further, we know that $\hat{\tau}_{k,n}$ includes only observations with the same mean for $k\in A_{G,q}$. We denote $\hat{\tau}_{k,n}$ including only observations $X_{k-G+1},\ldots,X_{k+G}$ with the same mean as $\hat{\tau}^0_{k,n}$. The 0 indicates that $\hat{\tau}^0_{k,n}$ behaves like a long-run variance estimator under the null hypothesis. This behaviour is due to the assumption of translation invariance. Further, we note that $\max_{G\leq k\leq n-G} \left|T^0_{k,n}(G)\right|/\hat{\tau}^0_{k,n}$ has the same asymptotic behaviour as the test statistic under H_0 (refer to Theorem 5.1).

Then, we obtain

$$
\begin{aligned}
&P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n)\right)\\
&= P\left(\max_{k\in A_{G,q}} \frac{|T^0_{k,n}(G)|}{\hat{\tau}^0_{k,n}} \geq D_n(G,\alpha_n)\right)\\
&\leq P\left(\max_{G\leq k\leq n-G} \frac{|T^0_{k,n}(G)|}{\hat{\tau}^0_{k,n}} \geq D_n(G,\alpha_n)\right)\\
&= P\left(a(n/G)\max_{G\leq k\leq n-G} \frac{|T^0_{k,n}(G)|}{\hat{\tau}^0_{k,n}} - b(n/G) \geq a(n/G)D_n(G,\alpha_n) - b(n/G)\right)\\
&= P\left(a(n/G)\max_{G\leq k\leq n-G} \frac{|T^0_{k,n}(G)|}{\hat{\tau}^0_{k,n}} - b(n/G) \geq \log\log\sqrt{1-\alpha_n}\right)
\end{aligned}
$$

by the definition of the critical value $D_n(G;\alpha_n)$ in (5.4). Further, Theorem 5.1 together with assumption **(A)** yield (6.4):

$$
P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n)\right) \quad = \quad \alpha_n + o(1) \longrightarrow 0.
$$

We note that in Theorem 5.1 the convergence is pointwise, but here a uniform convergence of the distribution function is needed. However, the uniform convergence is given, since the limit distribution function is continuous.

To prove (6.5) we analyse $R_{k,n}(G)$ for $k \in B_{G,q}$, i.e. $|k-k_j| < (1-\varepsilon)G$ for $j = 1,\ldots,q$. By assumption **(Cd)** the sets $\{k : |k-k_j| < (1-\varepsilon)G\}$ include asymptotically only one change point. Hence, we have for n large

$$
\begin{aligned}
&R_{k,n}(G)\\
&= \begin{cases} \frac{1}{\sqrt{2G}}\left(\sum_{i=k+1}^{k_j} d_{j,n} + \sum_{i=k_j+1}^{k+G} d_{j+1,n} - \sum_{i=k-G+1}^{k} d_{j,n}\right), & \text{if } k_j-(1-\varepsilon)G \leq k \leq k_j \\ \frac{1}{\sqrt{2G}}\left(\sum_{i=k+1}^{k+G} d_{j+1} - \sum_{i=k-G+1}^{k_j} d_{j,n} - \sum_{i=k_j+1}^{k} d_{j+1,n}\right), & \text{if } k_j \leq k < k_j+(1-\varepsilon)G \end{cases}\\
&= \begin{cases} \frac{d_{j+1,n}-d_{j,n}}{\sqrt{2G}}(G+k-k_j), & \text{if } k_j-(1-\varepsilon)G \leq k \leq k_j \\ \frac{d_{j+1,n}-d_{j,n}}{\sqrt{2G}}(k_j-k+G), & \text{if } k_j \leq k \leq k_j+(1-\varepsilon)G \end{cases}\\
&= \frac{\tilde{d}_{j,n}}{\sqrt{2G}}(G-|k-k_j|).
\end{aligned} \tag{6.7}
$$

A lower bound for $\min\limits_{k\in B_{n,G}} |T_{k,n}(G)|$ is obtained by

$$
\begin{aligned}
&\min_{k\in B_{n,G}} |T_{k,n}(G)| = \min_{k\in B_{n,G}} |T^0_{k,n}(G) + R_{k,n}(G)|\\
&\geq \min_{k\in B_{n,G}} |R_{k,n}(G)| - \max_{G\leq k\leq n-G} |T^0_{k,n}(G)|\\
&\geq \frac{\min_{1\leq j\leq q} |\tilde{d}_{j,n}|}{\sqrt{2G}}(G-(1-\varepsilon)G) - \max_{G\leq k\leq n-G} |T^0_{k,n}(G)|
\end{aligned}
$$

$$\geq \quad \frac{\min_{1\leq j\leq q}|\tilde{d}_{j,n}|}{\sqrt{2}}\varepsilon\sqrt{G} - \max_{G\leq k\leq n-G}|T^0_{k,n}(G)|.$$

By assumption **(A)** it holds

$$\begin{aligned} D_n(G;\alpha_n) &= \frac{b(n/G)}{a(n/G)} - \frac{\log\log\frac{1}{\sqrt{1-\alpha_n}}}{a(n/G)} = \sqrt{\log(n/G)}(1+o(1)) + O(1) \\ &= O(\sqrt{\log(n/G)}). \end{aligned}$$

We recall that $\max_{G\leq k\leq n-G}\hat{\tau}_{k,n} = O_p(1)$ by assumption **(L1)**,

$$\max_{G\leq k\leq n-G}|T^0_{k,n}(G)| = O_p(\sqrt{\log(n/G)})$$

(refer to (5.3)) to receive

$$\begin{aligned} &P\left(\min_{k\in B_{G,q}}\frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)\right) \\ &\leq P\left(\min_{k\in B_{G,q}}|T_{k,n}(G)| < D_n(G,\alpha_n)\max_{G\leq k\leq n-G}\hat{\tau}_{k,n}\right) \\ &\leq P\left(\sqrt{G}\min_{1\leq j\leq q}|\tilde{d}_{j,n}|\frac{\varepsilon}{\sqrt{2}} - \max_{G\leq k\leq n-G}|T^0_{k,n}(G)| < D_n(G,\alpha_n)\max_{G\leq k\leq n-G}\hat{\tau}_{k,n}\right) \\ &= P\left(\sqrt{G}\min_{1\leq j\leq q}|\tilde{d}_{j,n}|\frac{\varepsilon}{\sqrt{2}} < \max_{G\leq k\leq n-G}|T^0_{k,n}(G)| + D_n(G,\alpha_n)\max_{G\leq k\leq n-G}\hat{\tau}_{k,n}\right) \\ &= P\left(\sqrt{G}\frac{\varepsilon}{\sqrt{2}} < \left(\min_{1\leq j\leq q}|\tilde{d}_{j,n}|\right)^{-1}\left(\max_{G\leq k\leq n-G}|T^0_{k,n}(G)| + D_n(G,\alpha_n)\max_{G\leq k\leq n-G}\hat{\tau}_{k,n}\right)\right) \\ &\longrightarrow 0 \end{aligned} \tag{6.8}$$

by assumption **(D)**. □

To show the consistency of $\hat{q}_n$ under the regime-switching model, we have to reconsider assumption **(Cd)** on the minimum distance between the change points, since we have random change points in the regime-switching model. However, **(Cd)** can be altered into the appropriate stochastic formulation that the probability of the distance between two adjacent change points being larger than $2G$ tends to 0:

(Cr) Let for the minimum distance between two adjacent change points hold that

$$\lim_{n\to\infty}P(d_0(n) > 2G) = 1 \qquad \text{with} \qquad d_0(n) := \min_{0\leq j\leq q_n}|k_{j+1}-k_j|.$$

Further, in the regime-switching model **(RSM)** we have random expectations $d_1,\ldots,d_{q_n+1}$, which can take values in $\{d^*_1,\ldots,d^*_K\}$. We let them depend on n and consider $d^*_{1,n},\ldots,d^*_{K,n}$ and $d_{1,n},\ldots,d_{q_n+1,n}$. We assume

(Dr) Let $\tilde{d}^*_{i,j,n} := d^*_{j,n} - d^*_{i,n}$, $1 \le i < j \le n$, fulfill

$$\min_{1\le i<j\le K} \tilde{d}^{*2}_{i,j,n} G(\log(n/G))^{-1} \longrightarrow \infty.$$

With this assumptions we can show the consistency of $\hat{q}_n$ for the regime-switching model as well. The approach is to generalise the proof of Theorem 6.1 to the case of the regime-switching model.

Theorem 6.2. Let $X_1, ..., X_n$ follow the regime-switching model **(RSM)**. Further, let the assumptions **(E1)**, **(E2)** on the errors, **(G)** on the bandwidth G, **(A)** on the level α_n, **(Cr)** on the change points and **(Dr)** on the expectations hold. Moreover, assume **(L1)** for the long-run variance estimator $\hat{\tau}^2_{k,n}$ and let the error sequence fulfill **(L0)**. Then, under H_1,

$$P(\hat{q}_n = q_n) \quad \longrightarrow \quad 1 \qquad \text{as } n \to \infty.$$

Proof of Theorem 6.2. We start in the same way as in Theorem 6.1 and define A_{G,q_n} and B_{G,q_n} as in (6.2) and (6.1). We note that the sets are random in this setting, since the number and locations of change points are random. As before we obtain

$$\begin{aligned} &P(\hat{q}_n = q_n) \\ &\ge \; 1 - P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)\right) - P\left(\min_{k\in B_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)\right) \end{aligned}$$

and it is left to show that

$$P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)\right) \quad \longrightarrow \quad 0 \tag{6.9}$$

and

$$P\left(\min_{k\in B_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)\right) \quad \longrightarrow \quad 0. \tag{6.10}$$

In the proof of Theorem 6.1 the test statistic was splitted into two parts $T^0_{k,n}(G)$ and $R_{k,n}(G)$. For $k \in A_{G,q}$ the part $R_{k,n}(G)$ is equal to 0 by definition of $A_{G,q}$. To show that $R_{k,n}(G)$ has the form in (6.7) for $k \in B_{G,q}$ as n is large, the exact location of the sets did not matter, only the distance between two adjacent change points being larger than $2G$ for n large was required. Under assumption **(Cr)** the probability that $d_0(n) > 2G$ holds is asymptotically one, but the occurence of a smaller distance has to be considered. Therefore the probabilities in (6.9) and (6.10) are divided by the law of total probability into the cases $d_0(n) \le 2G$ and $d_0(n) > 2G$. We receive

$$\begin{aligned} &P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)\right) \\ &= \; P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n),\ d_0(n) \le 2G\right) \\ &\quad + P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n),\ d_0(n) > 2G\right) \end{aligned}$$

and with assumption **(Cr)**

$$P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n),\ d_0(n)\leq 2G\right) \quad \leq \quad P(d_0(n)\leq 2G) \quad = \quad o(1).$$

This can be analogously done for (6.10) such that it is left to show

$$\begin{aligned}
P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n),\ d_0(n) > 2G\right) &\longrightarrow 0,\\
P\left(\max_{k\in B_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n),\ d_0(n) > 2G\right) &\longrightarrow 0.
\end{aligned}$$

Since we are now in the situation of $d_0(n) > 2G$ it holds

$$\begin{aligned}
&P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n),\ d_0(n) > 2G\right)\\
&\quad = \quad P\left(\max_{k\in A_{G,q_n}} \frac{|T^0_{k,n}(G)|}{\hat{\tau}^0_{k,n}} \geq D_n(G,\alpha_n),\ d_0(n) > 2G\right)\\
&\quad \leq \quad P\left(\max_{G\leq k\leq n-G} \frac{|T^0_{k,n}(G)|}{\hat{\tau}^0_{k,n}} \geq D_n(G,\alpha_n),\ d_0(n) > 2G\right)\\
&\quad \leq \quad P\left(\max_{G\leq k\leq n-G} \frac{|T^0_{k,n}(G)|}{\hat{\tau}^0_{k,n}} \geq D_n(G,\alpha_n)\right) \quad \longrightarrow \quad 0.
\end{aligned}$$

For $k\in B_{G,q}$, i.e. $|k-k_j| < (1-\varepsilon)G$, and $d_0(n) > 2G$ we receive analogously to the calculations in (6.7)

$$R_{k,n}(G) \quad = \quad \frac{\tilde{d}_{j,n}}{\sqrt{2G}}(G - |k-k_j|)$$

where $R_{k,n}(G)$ is random but not the lower bound obtained by

$$\min_{k\in B_{n,G}} |R_{k,n}(G)| \quad = \quad \frac{\min_{1\leq j\leq q_n} |\tilde{d}_{j,n}|}{\sqrt{2G}}(G-(1-\varepsilon)G) \quad \geq \quad \frac{\min_{1\leq i<j\leq K} |\tilde{d}^*_{i,j,n}|}{\sqrt{2}}\varepsilon\sqrt{G}.$$

As in the calculations to (6.8) we obtain

$$\begin{aligned}
&P\left(\min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n),\ d_0(n) > 2G\right)\\
&\quad \leq \quad P\left(\sqrt{G}\min_{1\leq i<j\leq K} |\tilde{d}^*_{i,j,n}|\frac{\varepsilon}{\sqrt{2}} < \max_{G\leq k\leq n-G} |T^0_{k,n}(G)| + D_n(G,\alpha_n)\max_{G\leq k\leq n-G} \hat{\tau}_{k,n}\right)\\
&\qquad \longrightarrow 0,
\end{aligned}$$

since $\sqrt{G}$ converges to infinity faster than $\sqrt{\log(n/G)}$ by bandwidth assumption **(G)**. Hence, we have

$$P(\hat{q}_n = q_n) \quad = \quad 1 + o(1) \longrightarrow 1,$$

which is the assertion. □

6.1.2. Consistency of the Change-Point Estimators

In the last Section we proved that in general the number of change points is estimated asymptotically correctly by $\hat{q}_n$. For smaller samples sizes this is not necessarily the case. If the estimated number of change points is smaller than the true number of change points, $\hat{k}_{\hat{q}_n+1}, \ldots, \hat{k}_q$ are not defined and if the estimated number of change points is larger than the true number of change points, $k_{q+1} \ldots, k_{\hat{q}_n}$ do not exist. Hence, we define

$$\overline{q}_n := \min(q, \hat{q}_n) \tag{6.11}$$

to ensure the existence of the change points and change-point estimators in the following corollary.

Corollary 6.3. Let $X_1, ..., X_n$ follow the classical model **(CCM)**. Under the assumptions of Theorem 6.1 we have, under H_1,

$$P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G\right) \longrightarrow 0.$$

Proof of Corollary 6.3. The proof is based on conclusions from Theorem 6.1 and equations (6.4) and (6.5). We define the sets $A_{G,q}$ and $B_{G,q}$ as in (6.2) and (6.1). First, we distinguish between the cases of a correctly and an uncorrectly estimated number of change points:

$$\begin{aligned} &P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G\right) \\ &\quad = \; P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q = \hat{q}_n\right) + P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q \ne \hat{q}_n\right). \end{aligned}$$

Since we have by Theorem 6.1

$$P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q \ne \hat{q}_n\right) \quad \le \quad P\left(q \ne \hat{q}_n\right) \quad = \quad o(1),$$

we receive

$$P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G\right) \quad = \quad P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q = \hat{q}_n\right) + o(1).$$

Estimating the correct number of change points means having exactly q intervals, where the test statistic exceeds the critical value $D_n(G;\alpha_n)$. From the proof of Theorem 6.1 we know that in general these intervals asymptotically include the change points as well as symmetric intervals of length $2G(1-\varepsilon)$ around the change points. With this in mind we distinguish between the case that in this symmetric intervalls the critical value $D_n(G;\alpha_n)$ is exceeded or not and receive

$$\begin{aligned}
&P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G,\ q = \hat{q}_n\right)\\
&= P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G,\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n),\ q = \hat{q}_n\right)\\
&\quad + P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G,\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G;\alpha_n),\ q = \hat{q}_n\right)\\
&= P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G,\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n),\ q = \hat{q}_n\right) + o(1)
\end{aligned}$$

since by (6.5)

$$\begin{aligned}
&P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G,\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G;\alpha_n),\ q = \hat{q}_n\right)\\
&\leq P\left(\min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G;\alpha_n)\right) = o(1).
\end{aligned}$$

This yields

$$\begin{aligned}
&P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G\right)\\
&= P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G,\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n),\ q = \hat{q}_n\right) + o(1).
\end{aligned}$$

If in this situation the distance between a change-point estimator and a true change point should be larger than G, at least for one $k \in A_{G,k}$ the test statistic $T_{k,n}(G)$ has to exceed $D_n(G,\alpha_n)$. This implies

$$\begin{aligned}
&P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| \geq G,\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n),\ q = \hat{q}_n\right)\\
&= P\left(\exists j \in \{1,\ldots,q\} :\ |\hat{k}_j - k_j| \geq G,\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n),\ q = \hat{q}_n\right)\\
&\leq P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n),\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n),\ q = \hat{q}_n\right)\\
&\leq P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G;\alpha_n)\right) \longrightarrow 0
\end{aligned}$$

by equation (6.4). This gives the assertion. □

We continue with the analogous result for the regime-switching model. Since in the regime-switching model the number of change points depends on n we have a slight change in notation and define

$$\overline{q}_n \quad := \quad \min(\hat{q}_n, q_n).$$

Corollary 6.4. Let $X_1, ..., X_n$ follow the regime-switching model **(RSM)**. Under the assumptions of Theorem 6.2 we have, under H_1,

$$P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G\right) \longrightarrow 0.$$

Proof of Corollary 6.4. In principle the statement can be obtained along the lines of Corollary 6.3 using the assumptions and results for the regime-switching model. We briefly recall the proof and mention the necessary alterations. We analogously start with distinguishing between the cases of a correctly and an uncorrectly estimated number of change points:

$$\begin{aligned} &P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G\right) \\ &\quad = \; P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q_n = \hat{q}_n\right) + P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q_n \ne \hat{q}_n\right). \end{aligned}$$

Then, we use Theorem 6.2 instead of Theorem 6.1 to obtain

$$P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q_n \ne \hat{q}_n\right) \quad \le \quad P\left(q_n \ne \hat{q}_n\right) \quad = \quad o(1).$$

Since the change points are random, we have to consider the cases $d_0(n) > 2G$ and $d_0(n) \le 2G$ and obtain with assumption **(Cr)**:

$$P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G\right) \quad = \quad P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ q_n = \hat{q}_n,\ d_0(n) > 2G\right) + o(1).$$

Further we use (6.10) to obtain

$$\begin{aligned} &P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G\right) \\ &\quad = \; P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ \min_{k\in B_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G;\alpha_n),\ q_n = \hat{q}_n,\ d_0(n) > 2G\right) \\ &\qquad + o(1). \end{aligned}$$

And with the same ideas as in the proof of Corollary 6.3

$$\begin{aligned} &P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge G,\ \min_{k\in B_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G;\alpha_n),\ q = \hat{q}_n,\ d_0(n) > 2G\right) \\ &\quad \le \; P\left(\max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G;\alpha_n)\right) \quad \longrightarrow \quad 0 \end{aligned}$$

by equation (6.9). □

Corollaries 6.3 and 6.4 indicate that the estimation errors of change-point estimators $\hat{k}_j$, $j = 1, \ldots \overline{q}_n$, are uniformly stochastically bounded by G. However, in Theorem 6.6 and 6.10 below we can even show a better rate of uniform convergence. To obtain these results we have to make a further condition on the errors $\varepsilon_1, \ldots, \varepsilon_n$. This is a Hájek-Rényi-type moment condition and given by

(E3) For some $\gamma \geq 1$, $\varphi > 1$ and some constant $C > 0$ it holds

$$E\left|\sum_{k=i}^{j} \varepsilon_k\right|^{\gamma} \leq C|j-i+1|^{\varphi}.$$

With this assumption we can apply the following Hájek-Rényi-inequality in the proofs of the uniform consistency rates.

Lemma 6.5. Under assumption **(E1)**, **(E3)** it holds for any positive and non-increasing sequence $c_1 \geq c_2 \geq \ldots \geq c_n > 0$ and any ξ_n with $1 \leq \xi_n \leq n$ that a constant $A(\varphi, \gamma) \geq 8$ (only depending on φ and γ) exists such that for every $\delta > 0$

$$\delta^{\gamma} P\left(\max_{\xi_n \leq k \leq n} c_k \left|\sum_{j=1}^{k} \varepsilon_j\right| > \delta\right) \quad \leq \quad CA(\varphi, \gamma)\left(c_{\xi_n}^{\gamma} \xi_n^{\varphi} + \sum_{k=\xi_n+1}^{n} c_k^{\gamma} k^{\varphi-1}\right),$$

where C is as in **(E3)**.

Proof of Lemma 6.5. To begin with we split $\sum_{j=1}^{k} \varepsilon_j$ into the two sums $\sum_{j=1}^{\xi_n} \varepsilon_j$ and $\sum_{j=\xi_n+1}^{k} \varepsilon_j$ and use the monotony of the sequence $c_1, \ldots, c_n$ to receive

$$\begin{aligned}
P\left(\max_{\xi_n \leq k \leq n} c_k \left|\sum_{j=1}^{k} \varepsilon_j\right| > \delta\right) \quad &= \quad P\left(\max_{\xi_n \leq k \leq n} c_k \left|\sum_{j=1}^{\xi_n} \varepsilon_j + \sum_{j=\xi_n+1}^{k} \varepsilon_j\right| > \delta\right) \\
\leq \quad & P\left(c_{\xi_n}\left|\sum_{j=1}^{\xi_n} \varepsilon_j\right| + \max_{\xi_n \leq k \leq n} c_k \left|\sum_{j=\xi_n+1}^{k} \varepsilon_j\right| > \delta\right) \\
\leq \quad & P\left(c_{\xi_n}\left|\sum_{j=1}^{\xi_n} \varepsilon_j\right| > \delta/2\right) + P\left(\max_{\xi_n \leq k \leq n} c_k \left|\sum_{j=\xi_n}^{k} \varepsilon_j\right| > \delta/2\right)
\end{aligned}$$

and obtain by Markov's inequality

$$\delta^{\gamma} P\left(c_{\xi_n}\left|\sum_{j=1}^{\xi_n} \varepsilon_j\right| > \delta/2\right) \quad \leq \quad \delta^{\gamma}\left(\frac{\delta}{2}\right)^{-\gamma} c_{\xi_n}^{\gamma} E\left|\sum_{j=1}^{\xi_n} \varepsilon_j\right|^{\gamma} \quad \leq \quad 2^{\gamma} C c_{\xi_n}^{\gamma} \xi_n^{\varphi}$$

and

$$\delta^\gamma P\left(\max_{\xi_n\le k\le n} c_k\left|\sum_{j=\xi_n+1}^{k}\varepsilon_j\right|>\delta/2\right) \;\le\; 2^\gamma E\left(\max_{\xi_n\le k\le n} c_k\left|\sum_{j=\xi_n+1}^{k}\varepsilon_j\right|\right)^\gamma .$$

We define $c_k^* := c_{\xi_n+k}$, use the stationarity of $\varepsilon_1,\ldots,\varepsilon_n$ and apply Theorem B.3. in Kirch (2006) (refer to Theorem B.3 in the appendix) to receive with $l := k-\xi_n$

$$\begin{aligned}
E\left(\max_{\xi_n\le k\le n} c_k\left|\sum_{j=\xi_n+1}^{k}\varepsilon_j\right|\right)^\gamma &= E\left(\max_{1\le l\le n-\xi_n} c_l^*\left|\sum_{j=1}^{l}\varepsilon_j\right|\right)^\gamma \;\le\; CA_1(\varphi,\gamma)\sum_{l=1}^{n-\xi_n} c_l^{*\gamma} l^{\varphi-1}\\
&= CA_1(\varphi,\gamma)\sum_{l=1}^{n-\xi_n} c_{\xi_n+l}^{\gamma} l^{\varphi-1} \;=\; CA_1(\varphi,\gamma)\sum_{k=\xi_n+1}^{n} c_k{}^{\gamma}(k-\xi_n)^{\varphi-1}\\
&\le CA_1(\varphi,\gamma)\sum_{k=\xi_n+1}^{n} c_k^{\gamma} k^{\varphi-1}.
\end{aligned}$$

With $A(\varphi,\gamma) := 2^\gamma A_1(\varphi,\gamma)\ge 8$ follows the assertion. □

Usually we have $\gamma = 2+\nu$ and $\varphi=\gamma/2$ (refer to section 6.3). The next theorem gives the uniform stochastic boundedness of the change-point estimation errors weighted by the amount of change.

Theorem 6.6. Let $X_1,...,X_n$ follow the classical model **(CCM)**. Further let the assumptions **(E1)**, **(E2)**, **(E3)** on the errors, **(G)** on the bandwidth G, **(A)** on the level α_n, **(Cd)** on the change points and **(D)** on the expectations hold. Moreover, assume **(L1)** for the long-run variance estimator $\hat{\tau}_{k,n}^2$ and let the error sequence fulfill **(L0)**. Then, under H_1,

$$\forall\, \varepsilon>0\;\exists M>0:\; P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M\right)\le\varepsilon \qquad \text{for } n \text{ large.}$$

Remark 6.7. From Theorem 6.6 it follows

$$\forall\, \varepsilon>0\;\exists M>0:\; P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j-k_j|>M\left(\min_{j=1,\ldots,q}\tilde{d}_{j,n}^2\right)^{-1}\right)\le\varepsilon \qquad \text{for } n \text{ large}$$

since

$$\begin{aligned}
P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j-k_j|>M\left(\min_{j=1,\ldots,q}\tilde{d}_{j,n}^2\right)^{-1}\right) &= P\left(\min_{j=1,\ldots,q}\tilde{d}_{j,n}^2\max_{1\le j\le \overline{q}_n}|\hat{k}_j-k_j|>M\right)\\
&\le P\left(\max_{1\le j\le \overline{q}_n}\tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M\right).
\end{aligned}$$

Hence, we have $\max_{1\le j\le \bar{q}_n} |\hat{k}_j - k_j| = O_p\left(\left(\min_{j=1,\ldots,q} \tilde{d}_{j,n}^2\right)^{-1}\right)$ and in case of fixed changes $\max_{1\le j\le \bar{q}_n} |\hat{k}_j - k_j| = O_p(1)$. Since stochastic boundedness is the best rate, which can be obtained for change-point estimators, we speak of consistent change-point estimators. ■

Before we can prove Theorem 6.6 we have to consider two other lemmas.

Lemma 6.8. Under assumptions **(E1)**, **(E3)** for $T^0_{k,n}(G)$ defined as in (6.6) and $\beta_n > 0$ it holds with $1 \le \xi_n \le G$

$$\text{(a)}\quad P\left(\max_{k_j-G\le k\le k_j-\xi_n} (k-k_j)^{-1}\left|T^0_{k_j,n}(G) - T^0_{k,n}(G)\right| > \beta_n\right) \;=\; \beta_n^{-\gamma} G^{-\frac{\gamma}{2}} \xi_n^{-\frac{\gamma}{2}} O(1),$$

$$\text{(b)}\quad P\left(\max_{k_j-G\le k\le k_j-\xi_n} (k-k_j)^{-\frac{1}{2}}\left|T^0_{k_j,n}(G) - T^0_{k,n}(G)\right| > \beta_n\right) \;=\; \beta_n^{-\gamma} G^{-\frac{\gamma}{2}} \log(G) O(1),$$

$$\text{(c)}\quad P\left(\max_{k_j-G\le k\le k_j-\xi_n} \left|T^0_{k_j,n}(G) - T^0_{k,n}(G)\right| > \beta_n\right) \;=\; \beta_n^{-\gamma} O(1),$$

$$\text{(d)}\quad P\left(\max_{k_j-G\le k\le k_j-\xi_n} \left|T^0_{k,n}(G)\right| > \beta_n\right) \;=\; \beta_n^{-\gamma} O(1),$$

where the constants does not depend on j.

Proof of Lemma 6.8. For $k_j - G \le k \le k_j - \xi_n$ we consider the decompositions

$$\sum_{i=k+1}^{k+G} \varepsilon_i = \sum_{i=k+1}^{k_j} \varepsilon_i + \sum_{i=k_j+1}^{k+G} \varepsilon_i \quad \text{and} \quad \sum_{i=k-G+1}^{k} \varepsilon_i = \sum_{i=k-G+1}^{k_j-G} \varepsilon_i + \sum_{i=k_j-G+1}^{k} \varepsilon_i \tag{6.12}$$

to receive

$$\begin{aligned}
&T^0_{k_j,n}(G) - T^0_{k,n}(G)\\
&= \frac{1}{\sqrt{2G}}\left(\sum_{i=k_j+1}^{k_j+G} \varepsilon_i - \sum_{i=k_j-G+1}^{k_j} \varepsilon_i - \sum_{i=k+1}^{k+G} \varepsilon_i + \sum_{i=k-G+1}^{k} \varepsilon_i\right)\\
&= \frac{1}{\sqrt{2G}}\left(\sum_{i=k_j+1}^{k_j+G} \varepsilon_i - \sum_{i=k_j-G+1}^{k_j} \varepsilon_i - \sum_{i=k+1}^{k_j} \varepsilon_i - \sum_{i=k_j+1}^{k+G} \varepsilon_i + \sum_{i=k-G+1}^{k_j-G} \varepsilon_i + \sum_{i=k_j-G+1}^{k} \varepsilon_i\right)\\
&= \frac{1}{\sqrt{2G}}\left(\sum_{i=k+G+1}^{k_j+G} \varepsilon_i - \sum_{i=k+1}^{k_j} \varepsilon_i - \sum_{i=k+1}^{k_j} \varepsilon_i + \sum_{i=k-G+1}^{k_j-G} \varepsilon_i\right)\\
&= \frac{1}{\sqrt{2G}}\left(\sum_{i=k+G+1}^{k_j+G} \varepsilon_i + \sum_{i=k-G+1}^{k_j-G} \varepsilon_i - 2\sum_{i=k+1}^{k_j} \varepsilon_i\right).
\end{aligned} \tag{6.13}$$

We define $a_k := k - k_j$ and $\beta_n^* := \beta_n\sqrt{2G}$ and obtain

$$
\begin{aligned}
&P\left(\max_{k_j-G\le k\le k_j-\xi_n}(k-k_j)^{-1}\left|T^0_{k_j,n}(G)-T^0_{k,n}(G)\right|>\beta_n\right)\\
=\;& P\left(\max_{k_j-G\le k\le k_j-\xi_n} a_k^{-1}\left|\sum_{i=k+G+1}^{k_j+G}\varepsilon_i+\sum_{i=k-G+1}^{k_j-G}\varepsilon_i-2\sum_{i=k+1}^{k_j}\varepsilon_i\right|>\beta_n^*\right)\\
\le\;& P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|a_k^{-1}\sum_{i=k+G+1}^{k_j+G}\varepsilon_i\right|+\max_{k_j-G\le k\le k_j-\xi_n}\left|a_k^{-1}\sum_{i=k-G+1}^{k_j-G}\varepsilon_i\right|\right.\\
&\qquad\left.+\max_{k_j-G\le k\le k_j-\xi_n}\left|2a_k^{-1}\sum_{i=k+1}^{k_j}\varepsilon_i\right|>\beta_n^*\right)\\
\le\;& P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|a_k^{-1}\sum_{i=k+G+1}^{k_j+G}\varepsilon_i\right|>\beta_n^*/3\right)\\
&+P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|a_k^{-1}\sum_{i=k-G+1}^{k_j-G}\varepsilon_i\right|>\beta_n/3^*\right)\\
&+P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|a_k^{-1}\sum_{i=k+1}^{k_j}\varepsilon_i\right|>\beta_n^*/3\right)\\
\le\;& 3P\left(\max_{\xi_n\le l\le G}\left|\frac{1}{l}\sum_{i=1}^{l}\varepsilon_{-i}\right|>\beta_n^*/3\right),
\end{aligned}
$$

because the error sequence is strictly stationary. Since the assumptions of Lemma 6.5 are fulfilled for the inverse error sequence as well, we receive with $c_k = k^{-1}$ and $\varphi = \gamma/2$

$$
\begin{aligned}
P\left(\max_{\xi_n\le k\le G}k^{-1}\left|\sum_{i=1}^{k}\varepsilon_{-i}\right|>\beta_n^*\right)\;&\le\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(\xi_n^{-\gamma+\varphi}+\sum_{i=\xi_n+1}^{G}k^{-\gamma+\varphi-1}\right)\\
\le\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(\xi_n^{-\gamma+\varphi}+\int_{\xi_n}^{G}x^{-\gamma+\varphi-1}\,dx\right)\;&\le\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(\xi_n^{-\frac{\gamma}{2}}+\int_{\xi_n}^{\infty}x^{-\frac{\gamma}{2}-1}\,dx\right)\\
\le\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\xi_n^{-\frac{\gamma}{2}}\left(1+\frac{2}{\gamma}\right)&
\end{aligned}
$$

for a suitable constant $A(\gamma,\varphi)$. This yields part (a). The proofs of part (b) and (c) follow analogously: with $c_k = k^{-\frac{1}{2}}$ we have

$$
\begin{aligned}
P\left(\max_{\xi_n\le k\le G}k^{-\frac{1}{2}}\left|\sum_{i=1}^{k}\varepsilon_{-i}\right|>\beta_n^*\right)\;&\le\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(\xi_n^{-\frac{\gamma}{2}}\xi_n^{\varphi}+\sum_{i=\xi_n+1}^{G}k^{-\frac{\gamma}{2}+\varphi-1}\right)\\
\le\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(1+\int_{\xi_n}^{G}x^{-\frac{\gamma}{2}+\varphi-1}\,dx\right)\;&\le\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(1+\int_{1}^{G}x^{-1}\,dx\right)\\
=\;\frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(1+\log(G)\right)\;&=\;\beta_n^{*-\gamma}\log(G)O\left(1\right)
\end{aligned}
$$

and for $c_k = 1$

$$
\begin{aligned}
P\left(\max_{\xi_n \le k \le G}\left|\sum_{i=1}^{k} \varepsilon_{-i}\right| > \beta_n^*\right) \quad &\le \quad \frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(\xi_n^{\varphi} + \sum_{i=\xi_n+1}^{G} k^{\varphi-1}\right) \\
\le \quad \frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(\xi_n^{\frac{\gamma}{2}} + \int_0^G x^{\varphi-1}\,dx\right) \quad &= \quad \frac{CA(\gamma,\varphi)}{\beta_n^{*\gamma}}\left(\xi_n^{\frac{\gamma}{2}} + G^{\frac{\gamma}{2}}\right) \\
= \quad \beta_n^{*-\gamma} G^{\frac{\gamma}{2}} O\left(1\right).
\end{aligned}
$$

To obtain (d) we use the decompositions in (6.12) and receive with Lemma 6.5

$$
\begin{aligned}
&P\left(\max_{k_j - G \le k \le k_j - \xi_n} \left|T_{k,n}^0(G)\right| > \beta_n\right) \\
&= \; P\left(\max_{k_j - G \le k \le k_j - \xi_n}\left|\sum_{i=k+1}^{k_j} \varepsilon_i\right| > \beta_n^*/4\right) + P\left(\max_{k_j - G \le k \le k_j - \xi_n}\left|\sum_{i=k-G+1}^{k_j - G} \varepsilon_i\right| > \beta_n^*/4\right) \\
&\quad + P\left(\max_{k_j - G \le k \le k_j - \xi_n}\left|\sum_{i=k_j+1}^{k+G} \varepsilon_i\right| > \beta_n^*/4\right) + P\left(\max_{k_j - G \le k \le k_j - \xi_n}\left|\sum_{i=k_j-G+1}^{k} \varepsilon_i\right| > \beta_n^*/4\right) \\
&= \; O(1)\left(P\left(\max_{k_j - G \le k \le k_j - \xi_n}\left|\sum_{i=k+1}^{k_j} \varepsilon_i\right| > \beta_n^*/4\right) + P\left(\max_{k_j - G \le k \le k_j - \xi_n}\left|\sum_{i=k_j+1}^{k+G} \varepsilon_i\right| > \beta_n^*/4\right)\right) \\
&= \; O(1)\left(P\left(\max_{1 \le l \le G}\left|\sum_{i=1}^{l} \varepsilon_{-i}\right| > \beta_n^*/4\right) + P\left(\max_{1 \le l \le G}\left|\sum_{i=1}^{l} \varepsilon_i\right| > \beta_n^*/4\right)\right) \\
&= \; \beta_n^{*-\gamma} G^{\frac{\gamma}{2}} O(1).
\end{aligned}
$$

This completes the proof. □

Lemma 6.9. Let $X_1, ..., X_n$ follow the classical model **(CCM)**. With the assumptions of Theorem 6.6 we have, under H_1, $\forall\, \varepsilon > 0\; \exists M > 0:$

$$
P\left(\tilde{d}_{j,n}^2 |\hat{k}_j - k_j| > M,\; M(q,G)\right) \le \varepsilon + o(1),
$$

where

$$
\begin{aligned}
M(q,G) \quad := \quad &\left\{\hat{q}_n = q,\; \max_{k \in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n),\, \min_{k \in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n),\right. \\
&\left.\min_{G \le k \le n-G} \hat{\tau}_{k,n} > 0\right\}
\end{aligned}
$$

with $A_{G,q}$ and $B_{G,q}$ as in (6.2) and (6.1), and $T_{k,n}(G)$ is defined in (5.1).

Proof of Lemma 6.9. For reasons of simplicity we define $C_n = M\tilde{d}_{j,n}^{-2}$. We obtain by the definition of the absolute value and furthermore by the definition of $\hat{k}_j$

$$\begin{aligned}
&P\left(|\hat{k}_j - k_j| > C_n,\ M(q,G)\right)\\
&= P\left(\hat{k}_j - k_j > C_n,\ M(q,G)\right) + P\left(\hat{k}_j - k_j < -C_n,\ M(q,G)\right)\\
&= P\left(\hat{k}_j > k_j + C_n,\ M(q,G)\right) + P\left(\hat{k}_j < k_j - C_n,\ M(q,G)\right)\\
&= P\left(\arg\max_{v_j\le k\le w_j} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} > k_j + C_n,\ M(q,G)\right)\\
&\quad + P\left(\arg\max_{v_j\le k\le w_j} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < k_j - C_n,\ M(q,G)\right).
\end{aligned}$$

To receive the best possible rate of convergence we investigate $\arg\max_{v_j\le k\le w_j} V_{k,n}(G)$ instead of $\arg\max_{v_j\le k\le w_j} |T_{k,n}(G)|$ with

$$V_{k,n}(G) \;:=\; \left(T_{k,n}(G)\right)^2 - \left(T_{k_j,n}(G)\right)^2. \tag{6.14}$$

Noting the equality $\arg\max_{v_j\le k\le w_j} |T_{k,n}(G)| = \arg\max_{v_j\le k\le w_j} V_{k,n}(G)$ leads to

$$\begin{aligned}
&P\left(|\hat{k}_j - k_j| > C_n,\ M(q,G)\right)\\
&= P\left(\arg\max_{v_j\le k\le w_j} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} > k_j + C_n,\ M(q,G)\right)\\
&\quad + P\left(\arg\max_{v_j\le k\le w_j} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} < k_j - C_n,\ M(q,G)\right)
\end{aligned}$$

and using the definition of arg max to

$$\begin{aligned}
&P\left(|\hat{k}_j - k_j| > C_n,\ M(q,G)\right)\\
&= P\left(\max_{k_j+C_n<k\le w_j} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} > \max_{v_j\le k\le k_j+C_n} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}},\ M(q,G)\right)\\
&\quad + P\left(\max_{v_j\le k<k_j-C_n} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} \ge \max_{k_j-C_n\le k\le w_j} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}},\ M(q,G)\right).
\end{aligned}$$

Since $\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)$ (refer to the definition of $M(q,G)$) it holds $|k_j - v_j| < G$ and $|k_j - w_j| < G$. This implies

$$\begin{aligned}
&P\left(|\hat{k}_j - k_j| > C_n,\ M(q,G)\right)\\
&\le P\left(\max_{k_j+C_n<k\le k_j+G} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} \ge \max_{v_j\le k\le k_j+C_n} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}},\ M(q,G)\right)\\
&\quad + P\left(\max_{k_j-G\le k<k_j-C_n} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} \ge \max_{k_j-C_n\le k\le w_j} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}},\ M(q,G)\right).
\end{aligned}$$

Further, $V_{k_j,n}(G) = 0$ and $\min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n)$ implies $v_j < k_j$ and $w_j > k_j$ such that

$$\begin{aligned}
&P\left(|\hat{k}_j - k_j| > C_n,\ M(q,G)\right) \\
&\quad\leq\ P\left(\max_{k_j+C_n\leq k\leq k_j+G} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} \geq 0,\ \min_{G\leq k\leq n-G} \hat{\tau}_{k,n} > 0\right) \\
&\qquad + P\left(\max_{k_j-G\leq k\leq k_j-C_n} \frac{V_{k,n}(G)}{\hat{\tau}_{k,n}} \geq 0,\ \min_{G\leq k\leq n-G} \hat{\tau}_{k,n} > 0\right) \\
&\quad\leq\ P\left(\max_{k_j+C_n\leq k\leq k_j+G} V_{k,n}(G) \geq 0\right) + P\left(\max_{k_j-G\leq k\leq k_j-C_n} V_{k,n}(G) \geq 0\right). \qquad (6.15)
\end{aligned}$$

We focus on the second term in (6.15) and deduce a decomposition of $V_{k,n}(G)$ for all $k_j - G \leq k \leq k_j - C_n$ and n large. Results for the first term follow analogously. We have

$$\begin{aligned}
V_{k,n}(G) &= (T_{k,n}(G))^2 - (T_{k_j,n}(G))^2 = (T_{k,n}(G))^2 - \left((T_{k_j,n}(G) - T_{k,n}(G)) + T_{k,n}(G)\right)^2 \\
&= -(T_{k_j,n}(G) - T_{k,n}(G))^2 - 2(T_{k_j,n}(G) - T_{k,n}(G))T_{k,n}(G) \\
&= -(T_{k_j,n}(G) - T_{k,n}(G))\left((T_{k_j,n}(G) - T_{k,n}(G)) + 2T_{k,n}(G)\right).
\end{aligned}$$

Due to model **(CCM)** we receive for $k_j - G \leq k \leq k_j - C_n$

$$\sum_{i=k+1}^{k+G} X_i = \sum_{i=k+1}^{k_j} X_i + \sum_{i=k_j+1}^{k+G} X_i = \sum_{i=k+1}^{k_j} (\varepsilon_i + d_{j,n}) + \sum_{i=k_j+1}^{k+G} (\varepsilon_i + d_{j+1,n}).$$

This implies with $\tilde{d}_{j,n} = d_{j+1,n} - d_{j,n}$ and $T^0_{k,n}(G) := \frac{1}{\sqrt{2G}}\left(\sum_{i=k+1}^{k+G} \varepsilon_i - \sum_{i=k-G+1}^{k} \varepsilon_i\right)$

$$T_{k_j,n}(G) - T_{k,n}(G) = (2G)^{-\frac{1}{2}} a_k \tilde{d}_{j,n} + \left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right)$$

and

$$(T_{k_j,n}(G) - T_{k,n}(G)) + 2T_{k,n}(G) = (2G)^{-\frac{1}{2}} b_k \tilde{d}_{j,n} + \left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right) + 2T^0_{k,n}(G),$$

where $a_k := k_j - k$ and $b_k := k + 2G - k_j$ are constants with $C_n \leq a_k \leq G$ and $G \leq b_k \leq 2G - C_n$.

This yields

$$\begin{aligned}
V_{k,n}(G) &= -\left(T_{k_j,n}(G) - T_{k,n}(G)\right)\left(\left(T_{k_j,n}(G) - T_{k,n}(G)\right) + 2T_{k,n}(G)\right) \\
&= -\left(\left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right) + a_k(2G)^{-\frac{1}{2}}\tilde{d}_{j,n}\right) \\
&\qquad \left(\left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right) + 2T^0_{k,n}(G) + b_k(2G)^{-\frac{1}{2}}\tilde{d}_{j,n}\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2}G^{-1}a_k b_k \tilde{d}_{j,n}^2 - a_k(2G)^{-\frac{1}{2}}\tilde{d}_{j,n}\left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right) - \sqrt{2}a_k G^{-\frac{1}{2}}\tilde{d}_{j,n}T^0_{k,n}(G) \\
&\quad - \left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right)^2 - 2\left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right)T^0_{k,n}(G) \\
&\quad - \left(T^0_{k_j,n}(G) - T^0_{k,n}(G)\right) b_k(2G)^{-\frac{1}{2}}\tilde{d}_{j,n} \\
&=: K_1 + T_1 + T_2 + T_3 + T_4 + T_5. \qquad (6.16)
\end{aligned}
$$

We obtain by using Hájek-Rényi inequalities (a)-(d) from Lemma 6.8 with $\xi_n = C_n = M\tilde{d}_{j,n}^{-2}$

$$
\begin{aligned}
\max_{k_j-G\le k\le k_j-C_n} \frac{1}{|k-k_j|}\left|T^0_{k_j,n}(G) - T^0_{k,n}(G)\right| &= O_p(1)G^{-\frac{1}{2}}|\tilde{d}_{j,n}|, &(6.17)\\
\max_{k_j-G\le k\le k_j-C_n} \frac{1}{\sqrt{|k-k_j|}}\left|T^0_{k_j,n}(G) - T^0_{k,n}(G)\right| &= O_p(1)\,G^{-\frac{1}{2}}\log(G)^{\frac{1}{\gamma}}, &(6.18)\\
\max_{k_j-G\le k\le k_j-C_n} \left|T^0_{k_j,n}(G) - T^0_{k,n}(G)\right| &= O_p(1), &(6.19)\\
\max_{k_j-G\le k\le k_j-C_n} \left|T^0_{k,n}(G)\right| &= O_p(1). &(6.20)
\end{aligned}
$$

Using (6.17)-(6.20), $a_k = k - k_j \ge C_n$ and $b_k = k + 2G - k_j \ge G$ we conclude

$$
\begin{aligned}
\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_1}{K_1}\right| &= \sqrt{2G}\tilde{d}_{j,n}^{-1}\max_{k_j-G\le k\le k_j-C_n}\left|b_k^{-1}(T^0_{k_j,n}(G) - T^0_{k,n}(G))\right| \\
&= O_p(1)|\tilde{d}_{j,n}|^{-1}G^{-\frac{1}{2}}\max_{k_j-G\le k\le k_j-C_n}\left|(T^0_{k_j,n}(G) - T^0_{k,n}(G))\right| \\
&= O_p(1)\left(\min_{j=1,\dots,q}|\tilde{d}_{j,n}|\right)^{-1}G^{-\frac{1}{2}} = o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_2}{K_1}\right| &= 2^{\frac{3}{2}}G^{\frac{1}{2}}|\tilde{d}_{j,n}|^{-1}\max_{k_j-G\le k\le k_j-C_n} b_k^{-1}\left|T^0_{k,n}(G)\right| \\
&= O_p(1)G^{-\frac{1}{2}}|\tilde{d}_{j,n}|^{-1}\max_{k_j-G\le k\le k_j-C_n}\left|T^0_{k,n}(G)\right| \\
&= O_p(1)\left(\min_{j=1,\dots,q}|\tilde{d}_{j,n}|\right)^{-1}G^{-\frac{1}{2}} = o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_3}{K_1}\right| &= 2G|\tilde{d}_{j,n}|^{-2}\max_{k_j-G\le k\le k_j-C_n} a_k^{-1}b_k^{-1}\left|\left((T^0_{k_j,n}(G) - T^0_{k,n}(G))\right)^2\right| \\
&= O_p(1)|\tilde{d}_{j,n}|^{-2}\max_{k_j-G\le k\le k_j-C_n}\left|\left((k-k_j)^{-\frac{1}{2}}(T^0_{k_j,n}(G) - T^0_{k,n}(G))\right)^2\right| \\
&= O_p(1)\left(\min_{j=1,\dots,q}\tilde{d}_{j,n}^2\right)^{-1}G^{-1}(\log G)^{\frac{2}{\gamma}} = o_p(1),
\end{aligned}
$$

and

$$
\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_4}{K_1}\right| = 4G|\tilde{d}_{j,n}|^{-2}\max_{k_j-G\le k\le k_j-C_n} a_k^{-1}b_k^{-1}\left|(T^0_{k_j,n}(G) - T^0_{k,n}(G))T^0_{k,n}(G)\right|
$$

$$= O_p(1)\tilde{d}_{j,n}^{-2} \max_{k_j-G\le k\le k_j-C_n} \left|(k-k_j)^{-1}(T^0_{k_j,n}(G)-T^0_{k,n}(G))\right| \max_{k_j-G\le k\le k_j-C_n} \left|T^0_{k,n}(G)\right|$$
$$= O_p(1)\left(\min_{j=1,\ldots,q} |\tilde{d}_{j,n}|\right)^{-1} G^{-\frac{1}{2}} = o_p(1)$$

by Assumption **(D)**. We define

$$Z^* := \max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_1}{K_1}\right| + \max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_2}{K_1}\right| + \max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_3}{K_1}\right| + \max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_4}{K_1}\right|$$

with $Z^* = o_p(1)$ as proven above and use $K_1 \le -\frac{1}{2}C_n\tilde{d}_{j,n}^2 < 0$ to obtain

$$\begin{aligned}
&P\left(\max_{k_j-G\le k\le k_j-C_n} V_{k,n}(G)\ge 0\right)\\
&= P\left(\max_{k_j-G\le k\le k_j-C_n} K_1\left(1+\frac{T_1}{K_1}+\frac{T_2}{K_1}+\frac{T_3}{K_1}+\frac{T_4}{K_1}+\frac{T_5}{K_1}\right)\ge 0\right)\\
&= P\left(\min_{k_j-G\le k\le k_j-C}\left(1+\frac{T_1}{K_1}+\frac{T_2}{K_1}+\frac{T_3}{K_1}+\frac{T_4}{K_1}+\frac{T_5}{K_1}\right)\le 0\right)\\
&\le P\left(\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_1}{K_1}+\frac{T_2}{K_1}+\frac{T_3}{K_1}+\frac{T_4}{K_1}+\frac{T_5}{K_1}\right|\ge 1\right)\\
&\le P\left(\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_5}{K_1}\right| + Z^*\ge 1\right)\\
&= P\left(\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_5}{K_1}\right|\ge 1-Z^*, Z^*\le 1/2\right)\\
&\quad + P\left(\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_5}{K_1}\right|\ge 1-Z^*, Z^*> 1/2\right)\\
&\le P\left(\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_5}{K_1}\right|\ge 1/2\right) + P\left(Z^*\ge 1/2\right)\\
&\le P\left(\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_5}{K_1}\right|\ge 1/2\right) + o(1).
\end{aligned}$$

Since by part (a) of Lemma 6.8 for all $\varepsilon > 0$ exists a $M > 0$ such that

$$\begin{aligned}
&P\left(\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_5}{K_1}\right|\ge 1/2\right)\\
&= P\left(\max_{k_j-G\le k\le k_j-C_n}\left|a_k^{-1}\left(T^0_{k_j,n}(G)-T^0_{k,n}(G)\right)(2G)^{\frac{1}{2}}\tilde{d}_{j,n}^{-1}\right|\ge 1/2\right)\\
&= P\left(\max_{k_j-G\le k\le k_j-C_n} a_k^{-1}\left|\left(T^0_{k_j,n}(G)-T^0_{k,n}(G)\right)\right|\ge (2G)^{-\frac{1}{2}}|\tilde{d}_{j,n}|/2\right)\\
&< \varepsilon \quad \text{for } n \text{ large enough,}
\end{aligned}$$

we have proved the assertion. □

We are now able to prove Theorem 6.6.

Proof of Theorem 6.6. To begin with, we distinguish between the cases of a correctly and uncorrectly estimated number of change points:

$$
\begin{aligned}
&P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M\right)\\
&\quad = \; P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M,\ \hat{q}_n = q\right) + P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M,\ \hat{q}_n \ne q\right)
\end{aligned}
$$

and use Theorem 6.1 to show

$$
P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M,\ \hat{q}_n \ne q\right) \;\le\; P\left(\hat{q}_n \ne q\right) \;=\; o(1).
$$

Further, we distinguish between the cases $\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)$, $\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)$ and obtain

$$
\begin{aligned}
&P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M, \hat{q}_n = q\right)\\
&\quad = \; P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M,\ \hat{q}_n = q,\ \max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)\right)\\
&\qquad + P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M,\ \hat{q}_n = q,\ \max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)\right),
\end{aligned}
$$

where we know from (6.4) that

$$
\begin{aligned}
&P\left(\tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M,\ \hat{q}_n = q,\ \max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)\right)\\
&\quad \le \; P\left(\max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)\right) \;=\; o(1).
\end{aligned}
$$

Analogously we continue to distinguish the cases $\min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G,\alpha_n)$ and $\min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n)$ as well as $\min_{G\le k\le n-G} \hat{\tau}_{k,n} > 0$ and $\min_{G\le k\le n-G} \hat{\tau}_{k,n} \le 0$ and receive by (6.5) and assumption **(L1)**

$$
P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M\right) \;=\; P\left(\max_{1\le j\le \overline{q}_n} \tilde{d}^2_{j,n}|\hat{k}_j - k_j| > M,\ M(G,q)\right) + o(1)
$$

with

$$M(q,G) \;:=\; \Big\{\hat{q}_n = q,\ \max_{k\in A_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G,\alpha_n),\ \min_{k\in B_{G,q}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G,\alpha_n),$$
$$\min_{G\leq k\leq n-G} \hat{\tau}_{k,n} > 0\Big\}.$$

Therefrom, we obtain

$$\begin{aligned} P\left(\max_{1\leq j\leq \overline{q}_n} \tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M\right) &= P\left(\max_{1\leq j\leq \overline{q}_n} \tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M, M(G,q)\right)+o(1)\\ &= P\left(\bigcup_{j=1}^{q}\left\{\tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M,\ M(G,q)\right\}\right)+o(1)\\ &\leq \sum_{j=1}^{q} P\left(\tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M,\ M(G,q)\right)+o(1). \end{aligned} \tag{6.21}$$

In Lemma 6.9 we have showen: For all $\varepsilon^*>0$ exists a constant $M>0$:

$$P\left(\tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M,\ M(G,q)\right) \;\leq\; \varepsilon^*+o(1)$$

for $j=1,\ldots,q$. This implies that we can choose for all $\varepsilon>0$ an appropriate ε^*, i.e. $\varepsilon^*<\varepsilon/q$, such that it exists a constant $M>0$:

$$\begin{aligned} P\left(\max_{1\leq j\leq \overline{q}_n} \tilde{d}_{j,n}^2|\hat{k}_j-k_j|>M\right) &\leq q(\varepsilon^*+o(1))+o(1)=q\varepsilon^*+o(1)\\ &< \varepsilon \qquad \text{for } n \text{ large enough, i.e. } o(1)<\varepsilon-q\varepsilon^*. \end{aligned} \tag{6.22}$$

This yields the assertion. □

For the regime-switching model the uniform rate of convergence for the errors of the change-point estimators depends on the growth rate of the number of change points. To make an assumption on the growth rate of q_n we define a sequence γ_n, which bounds the number of change points in a stochastic way. Namely

(Q) Let $\{\gamma_n : n\in\mathbb{N}\}$ be a sequence such that

$$P(q_n>\gamma_n)\longrightarrow 0.$$

Theorem 6.10 gives a uniform rate of convergence for the change-point estimators under the regime-switching model. In the proof we see why assumption **(Q)** is necessary and why the proof and hence the result of Theorem 6.6 is in general not valid for the regime-switching model.

Theorem 6.10. Let $X_1, ..., X_n$ follow the regime switching model **(RSM)**. Let the assumptions **(E1)**, **(E2)**, **(E3)** on the errors, **(G)** on the bandwidth G, **(A)** on the level $\{\alpha_n\}$, **(Cr)** on the change points and **(Dr)** on the expectations hold. Moreover, assume **(L1)** for the long-run variance estimator $\hat{\tau}^2_{k,n}$ and let the error sequence fulfill **(L0)**. Further, let $\{\gamma_n : n \in \mathbb{N}\}$ be a sequence fulfilling **(Q)** and $\{\xi_n : n \in \mathbb{N}\}$ be a strictly positive sequence. Then, under H_1,

$$P\left(\max_{1\leq j\leq \bar{q}_n} |\hat{k}_j - k_j| \geq \xi_n\right) = \gamma_n \left(\min_{1\leq i<j\leq K} \left|\tilde{d}^*_{i,j,n}\right|\right)^{-\gamma} \left(\xi_n^{-\frac{\gamma}{2}} + G^{-\frac{\gamma}{2}} \log G\right) O(1) + o(1).$$

Proof of Theorem 6.10. We define the set

$$\begin{aligned} M^*(G, q_n) \quad := \quad & \left\{ \max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G, \alpha_n), \min_{k\in B_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \geq D_n(G, \alpha_n), \right. \\ & \left. \min_{G\leq k\leq n-G} \hat{\tau}_{k,n} > 0,\ d_0(n) > G \right\} \end{aligned}$$

with A_{G,q_n}, B_{G,q_n} as in (6.2), (6.1). This is the correspondend definition to $M(G, q)$ from the proof of Theorem 6.6 added by the event $d_0(n) > G$ to take the random change points into account and reduced by the event $\hat{q}_n = q_n$ to have the oppertunity to consider it in more detail. With Theorem 6.2, equations (6.9), (6.10) and assumptions **(L1)**, **(Cr)** we receive

$$\begin{aligned} P\left(\max_{1\leq j\leq \bar{q}_n} |\hat{k}_j - k_j| > \xi_n\right) \quad &= \quad P\left(\max_{1\leq j\leq \bar{q}_n} |\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ M^*(G, q_n)\right) + o(1) \\ = \quad P & \left(\bigcup_{j=1}^{\bar{q}_n} \left\{|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ M^*(G, q_n)\right\}\right) + o(1). \end{aligned}$$

We can not conduct the next step as before (refer to (6.21)), since q_n is a random variable. Additionally the last conclusion in (6.22) does not hold, because q_n is allowed to tend to infinity. Therefore, we introduced the sequence $\{\gamma_n\}$ to proceed as follows: We distinguish between the case $q_n \leq \gamma_n$ and $q_n > \gamma_n$ to obtain with assumption **(Q)**

$$\begin{aligned} & P\left(\max_{1\leq j\leq \bar{q}_n} |\hat{k}_j - k_j| > \xi_n\right) \\ & = \quad P\left(\bigcup_{j=1}^{\bar{q}_n} \left\{|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ q_n \leq \gamma_n,\ M^*(G, q_n)\right\}\right) + o(1). \end{aligned}$$

In case of $q_n \leq \gamma_n$ the number of change points q_n can take values between 1 and γ_n so that we receive

$$P\left(\bigcup_{j=1}^{\bar{q}_n} \left\{|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ q_n \leq \gamma_n,\ M^*(G, q_n)\right\}\right)$$

$$
\begin{aligned}
&\leq \sum_{l=1}^{\gamma_n} P\left(\bigcup_{j=1}^{\overline{q}_n}\left\{|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ q_n = l,\ M^*(G,q_n)\right\}\right)\\
&= \sum_{l=1}^{\gamma_n} P\left(\bigcup_{j=1}^{l}\left\{|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ q_n = l,\ M^*(G,q_n)\right\}\right)\\
&\leq \sum_{l=1}^{\gamma_n}\sum_{j=1}^{l} P\left(|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ q_n = l,\ M^*(G,q_n)\right)\\
&= \sum_{j=1}^{\gamma_n}\sum_{l=j}^{\gamma_n} P\left(|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ q_n = l,\ M^*(G,q_n)\right)\\
&= \sum_{j=1}^{\gamma_n} P\left(|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ j \leq q_n \leq \gamma_n,\ M^*(G,q_n)\right).
\end{aligned}
$$

Since $\{\gamma_n\}$ is allowed to tend to infinity the above sum will in general not be bounded. To determine how fast the sum grows we need to know how fast the probabilities go to zero. These rates are obtained in Lemma 6.12 below and yield

$$
\begin{aligned}
&P\left(\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| > \xi_n\right)\\
&= \sum_{j=1}^{\gamma_n} P\left(|\hat{k}_j - k_j| > \xi_n,\ \hat{q}_n = q_n,\ j \leq q_n \leq \gamma_n,\ M^*(G,q_n)\right) + o(1)\\
&= \gamma_n \left(\min_{1\leq i<j\leq K}\left|\tilde{d}^*_{i,j,n}\right|\right)^{-\gamma}\left(\xi_n^{-\frac{\gamma}{2}} + G^{-\frac{\gamma}{2}}\log G\right) O(1) + o(1).
\end{aligned}
$$

This gives the assertion. □

Remark 6.11. (a) If $\{\gamma_n\}$ is constant, i.e. the number of changes is stochastically bounded, we obtain

$$
\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| = O_p\left(\left(\min_{1\leq i<j\leq K}\left|\tilde{d}^*_{i,j,n}\right|\right)^{-2}\right).
$$

(b) If we additionally to the assumptions of Theorem 6.10 assume that the sequence $\{\gamma_n\}$ satisfies

$$
\gamma_n \log(G) G^{-\frac{\gamma}{2}}\left(\min_{1\leq i<j\leq K}\left|\tilde{d}^*_{i,j,n}\right|\right)^{-\gamma} \longrightarrow 0, \tag{6.23}
$$

i.e. the number of changes is allowed to tend to infinity but not too fast, we obtain

$$
\max_{1\leq j\leq \overline{q}_n} |\hat{k}_j - k_j| = O_p\left(\gamma_n^{\frac{2}{\gamma}}\left(\min_{1\leq i<j\leq K}\left|\tilde{d}^*_{i,j,n}\right|\right)^{-2}\right).
$$

(c) If we additionally to the assumptions of Theorem 6.10 assume that the sequence $\{\gamma_n\}$ satisfies (6.23) and the sequence $\{\xi_n\}$ fulfills

$$\gamma_n \xi_n^{-\frac{\gamma}{2}} \left(\min_{1\le i<j\le K} \left| \tilde{d}^*_{i,j,n} \right| \right)^{-\gamma} \quad \longrightarrow \quad 0,$$

we obtain

$$P\left(\max_{1\le j\le \overline{q}_n} |\hat{k}_j - k_j| \ge \xi_n \right) \quad \longrightarrow \quad 0. \qquad \blacksquare$$

It is left to show the result we used in the proof of Theorem 6.10, which is concerned with the individual rates of convergence for the errors of the change-point estimators in case of the regime-switching model.

Lemma 6.12. Let $X_1, ..., X_n$ follow the regime-switching model **(RSM)** and the assumptions of Theorem 6.10 hold. Further let $\{\xi_n : n \in \mathbb{N}\}$ be a strictly positive sequence. Then we have, under H_1,

$$P\left(|\hat{k}_j - k_j| > \xi_n,\ M(G, q_n) \right) \quad = \quad \left(\min_{1\le i<j\le K} \left| \tilde{d}^*_{i,j,n} \right| \right)^{-\gamma} \left(\xi_n^{-\frac{\gamma}{2}} + G^{-\frac{\gamma}{2}} \log(G) \right) O(1)$$

uniformly in $j = 1, ..., q_n$ and

$$M(G; q_n) \quad := \quad \Bigg\{ \hat{q}_n = q_n,\ \max_{k\in A_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} < D_n(G, \alpha_n),\ \min_{k\in B_{G,q_n}} \frac{|T_{k,n}(G)|}{\hat{\tau}_{k,n}} \ge D_n(G, \alpha_n),$$
$$\min_{G\le k\le n-G} \hat{\tau}_{k,n} > 0,\ d_0(n) > 2G \Bigg\}$$

with A_{G,q_n} and B_{G,q_n} as in (6.2) and (6.1).

Proof of Lemma 6.12. We obtain along the lines of the proof of Lemma 6.9

$$\begin{aligned} &P\left(|\hat{k}_j - k_j| > \xi_n,\ M(G, q_n) \right) \\ &\quad \le \quad P\left(\max_{k_j+\xi_n\le k\le k_j+G} V_{k,n}(G) \ge 0,\ d_0(n) > 2G \right) \\ &\qquad + P\left(\max_{k_j-G\le k\le k_j-\xi_n} V_{k,n}(G) \ge 0,\ d_0(n) > 2G \right). \end{aligned}$$

The further calculations are exactly the same up to (6.16), because we are in the situation with $d_0(n) > 2G$. Since we need rates of convergence for the probabilities to compensate for the increasing number of change points we obtain by Lemma 6.8 (a)-(d) with $a_k := k - k_j \ge \xi_n$ and $b_k := k + 2G - k_j \ge G$

$$\begin{aligned} &P\left(\max_{k_j-G\le k\le k_j-\xi_n} \left| \frac{T_1}{K_1} \right| > \frac{1}{2} \right) \\ &\quad \le \quad P\left(\max_{k_j-G\le k\le k_j-\xi_n} \left| (T^0_{k_j,n}(G) - T^0_{k,n}(G)) \right| > \min_{1\le i<j\le K} \left| \tilde{d}^*_{i,j,n} \right| 2^{-\frac{3}{2}} G^{\frac{1}{2}} \right) \end{aligned}$$

$$= O(1)G^{-\frac{\gamma}{2}}\left(\min_{1\le i<j\le K}\left|\tilde{d}^*_{i,j,n}\right|\right)^{-\gamma},$$

$$P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|\frac{T_2}{K_1}\right|>\frac{1}{2}\right) = P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|T^0_{k,n}(G)\right|>G^{\frac{1}{2}}2^{-\frac{5}{2}}\min_{1\le i<j\le K}\left|\tilde{d}^*_{i,j,n}\right|\right)$$
$$= O(1)G^{-\frac{\gamma}{2}}\left(\min_{1\le i<j\le K}\left|\tilde{d}^*_{i,j,n}\right|\right)^{-\gamma}$$

and

$$P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|\frac{T_3}{K_1}\right|>\frac{1}{2}\right)$$
$$\le P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left(|k-k_j|^{-\frac{1}{2}}(T^0_{k_j,n}(G)-T^0_{k,n}(G))\right)^2>\min_{1\le i<j\le K}\tilde{d}^{*2}_{i,j,n}\right)$$
$$= O(1)G^{-\frac{\gamma}{2}}\log(G)\left(\min_{1\le i<j\le K}\left(\tilde{d}^*_{i,j,n}\right)\right)^{-\gamma}.$$

Then, we define

$$Z^* := \max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_1}{K_1}\right|+\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_2}{K_1}\right|+\max_{k_j-G\le k\le k_j-C_n}\left|\frac{T_3}{K_1}\right|$$

and have shown

$$P\left(Z^*\ge 1/2\right) = \left(\min_{1\le i<j\le K}\left(\tilde{d}^*_{i,j,n}\right)\right)^{-\gamma}G^{-\frac{\gamma}{2}}\log(G)O(1).$$

Furthermore we use $P(AB>\varepsilon)\le P(A>\varepsilon^{\frac{1}{2}})+P(B>\varepsilon^{\frac{1}{2}})$ for $A,B\ge 0$ to obtain

$$P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|\frac{T_4}{K_1}\right|>\frac{1}{4}\right)$$
$$= P\left(4G|\tilde{d}_{j,n}|^{-2}\max_{k_j-G\le k\le k_j-C_n}a_k^{-1}b_k^{-1}\left|(T^0_{k_j,n}(G)-T^0_{k,n}(G))T^0_{k,n}(G)\right|>\frac{1}{4}\right)$$
$$= P\left(4|\tilde{d}_{j,n}|^{-2}\max_{k_j-G\le k\le k_j-C_n}a_k^{-1}\left|(T^0_{k_j,n}(G)-T^0_{k,n}(G))T^0_{k,n}(G)\right|>\frac{1}{4}\right)$$
$$= P\left(\max_{k_j-G\le k\le k_j-\xi_n}a_k^{-1}\left|(T^0_{k_j,n}(G)-T^0_{k,n}(G))\right|>2^{-1}\min_{1\le i<j\le K}\left|\tilde{d}^*_{i,j,n}\right|\frac{1}{4}G^{-\frac{1}{2}}\right)$$
$$+P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|T^0_{k,n}(G)\right|>2^{-1}\min_{1\le i<j\le K}\left|\tilde{d}^*_{i,j,n}\right|\frac{1}{4}G^{\frac{1}{2}}\right)$$
$$= O(1)\left(\min_{1\le i<j\le K}\left|\tilde{d}^*_{i,j,n}\right|\right)^{-\gamma}\left(\xi_n^{-\frac{\gamma}{2}}+G^{-\frac{\gamma}{2}}\right).$$

Moreover we have

$$
\begin{aligned}
&P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|\frac{T_5}{K_1}\right|>\frac{1}{4}\right)\\
&= P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|\sqrt{2G}\tilde{d}_{j,n}^{-1}a_k^{-1}(T_{k_j,n}^0(G)-T_{k,n}^0(G))\right|>\frac{1}{4}\right)\\
&\le P\left(\max_{k_j-G\le k\le k_j-\xi_n}a_k^{-1}\left|(T_{k_j,n}^0(G)-T_{k,n}^0(G))\right|>\frac{\min_{1\le i<j\le K}\left(\tilde{d}_{i,j,n}^*\right)}{5\sqrt{2}}G^{-\frac{1}{2}}\right)\\
&= O(1)\xi_n^{-\frac{\gamma}{2}}\left(\min_{1\le i<j\le K}\left|\tilde{d}_{i,j,n}^*\right|\right)^{-\gamma}.
\end{aligned}
$$

We use $K_1\le-\frac{1}{2}\xi_n<0$ to obtain similar to the proof of Lemma 6.9

$$
\begin{aligned}
&P\left(\max_{k_j-G\le k\le k_j-\xi_n}V_k(G)\ge 0,\ d_0(n)>2G\right)\\
&\le P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|\frac{T_5}{K_1}\right|\ge 1/4\right)+P\left(\max_{k_j-G\le k\le k_j-\xi_n}\left|\frac{T_4}{K_1}\right|\ge 1/4\right)+P\left(Z^*\ge 1/2\right)\\
&\le \left(\min_{1\le i<j\le K}\left|\tilde{d}_{i,j,n}^*\right|\right)^{-\gamma}\left(\xi_n^{-\frac{\gamma}{2}}+G^{-\frac{\gamma}{2}}\log(G)\right)O(1).
\end{aligned}
$$

This completes the proof. □

6.1.3. Joint Asymptotic Distribution of the Change-Point Estimators

We derive the joint asymptotic distribution of the normalised change-point estimators under the classical model with local alternatives. We do this under the assumption of an independent and identically distributed error sequence and known variance σ^2.

Theorem 6.13. Let $X_1,...,X_n$ follow model **(CCM)** with an error sequence of independent and identically distributed random variables. Furthermore, let the assumptions **(E1)**, **(E2)**, **(E3)** on the error sequence, **(G)** on the bandwidth G, **(A)** on level α_n and **(Cd)** on the change points hold. Moreover, let the expectations fulfill **(D)** and

$$\max_{j=1...,q}|\tilde{d}_{j,n}|\longrightarrow 0.$$

Then, under H_1,

$$\sigma^{-2}\left(\tilde{d}_{1,n}^2(\hat{k}_1-k_1),\dots,\tilde{d}_{\overline{q}_n,n}^2(\hat{k}_{\overline{q}_n}-k_{\overline{q}_n})\right)\ \xrightarrow[n\to\infty]{\mathcal{D}}\ (S_1,\dots,S_q),$$

where $\overline{q}_n$ is defined as in (6.11),

$$S_i\ :=\ \arg\max\{W_s^{(i)}-|s|/\sqrt{6}:\ s\in\mathbb{R}\}$$

and $\{W_s^{(i)}:s\in\mathbb{R}\}$, $i=1,\dots,q$, are mutually independent standard Wiener processes.

Proof of Theorem 6.13. With $c > 0$ we obtain

$$
\begin{aligned}
&P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1 - k_1}{\sigma^2}, \ldots, \tilde{d}_{\overline{q}_n,n}^2 \frac{\hat{k}_{\overline{q}_n} - k_{\overline{q}_n}}{\sigma^2}\right) \leq (x_1, \ldots, x_{\overline{q}_n})\right) \\
&\quad = \; P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1 - k_1}{\sigma^2}, \ldots, \tilde{d}_{\overline{q}_n,n}^2 \frac{\hat{k}_{\overline{q}_n} - k_{\overline{q}_n}}{\sigma^2}\right) \leq (x_1, \ldots, x_{\overline{q}_n}) \,, \; \max_{1 \leq j \leq \overline{q}_n} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} \leq c\right) \\
&\qquad + o(1),
\end{aligned}
$$

since we have by Theorem 6.6

$$
\begin{aligned}
&P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1 - k_1}{\sigma^2}, \ldots, \tilde{d}_{\overline{q}_n,n}^2 \frac{\hat{k}_{\overline{q}_n} - k_{\overline{q}_n}}{\sigma^2}\right) \leq (x_1, \ldots, x_{\overline{q}_n}) \,, \; \max_{1 \leq j \leq \overline{q}_n} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} > c\right) \\
&\quad \leq \; P\left(\max_{1 \leq j \leq \overline{q}_n} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} > c\right) \; = \; o(1) \qquad \text{for } c \longrightarrow \infty.
\end{aligned}
$$

Further, we receive by Theorem 6.1

$$
\begin{aligned}
&P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1 - k_1}{\sigma^2}, \ldots, \tilde{d}_{\overline{q}_n,n}^2 \frac{\hat{k}_{\overline{q}_n} - k_{\overline{q}_n}}{\sigma^2}\right) \leq (x_1, \ldots, x_{\overline{q}_n}) \,, \; \max_{1 \leq j \leq \overline{q}_n} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} \leq c\right) \\
&\quad = \; P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1 - k_1}{\sigma^2}, \ldots, \tilde{d}_{n,q}^2 \frac{\hat{k}_q - k_q}{\sigma^2}\right) \leq (x_1, \ldots, x_q) \,, \; \max_{1 \leq j \leq q} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} \leq c, \; \hat{q}_n = q\right) \\
&\qquad + o(1).
\end{aligned}
$$

In this situation each change-point estimator $\hat{k}_j$ lies in a symmetric intervall of length $c\sigma^2 \tilde{d}_{j,n}^{-2}$ around k_j and we define

$$
\hat{k}_j^* := \operatorname*{arg\,max}_{|k - k_j| \leq c\sigma^2 \tilde{d}_{j,n}^{-2}} T_{k,n}(G), \qquad j = 1, \ldots, q.
$$

Under $\max_{1 \leq j \leq q} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} \leq c$ it holds that $\hat{k}_j = \hat{k}_j^*$. Thus, we receive

$$
\begin{aligned}
&P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1 - k_1}{\sigma^2}, \ldots, \tilde{d}_{q,n}^2 \frac{\hat{k}_q - k_q}{\sigma^2}\right) \leq (x_1, \ldots, x_q) \,, \; \max_{1 \leq j \leq q} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} \leq c, \; \hat{q}_n = q\right) \\
&\quad = \; P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1^* - k_1}{\sigma^2}, \ldots, \tilde{d}_{q,n}^2 \frac{\hat{k}_q^* - k_q}{\sigma^2}\right) \leq (x_1, \ldots, x_q) \,, \; \max_{1 \leq j \leq q} \tilde{d}_{j,n}^2 \frac{|\hat{k}_j - k_j|}{\sigma^2} \leq c, \; \hat{q}_n = q\right) \\
&\quad = \; P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1^* - k_1}{\sigma^2}, \ldots, \tilde{d}_{q,n}^2 \frac{\hat{k}_q^* - k_q}{\sigma^2}\right) \leq (x_1, \ldots, x_q)\right) + o(1).
\end{aligned}
$$

The advantage of the estimators $\hat{k}_1^*, \ldots, \hat{k}_q^*$ is their independence as n is large. This follows by the following considerations: $\hat{k}_1^*, \ldots, \hat{k}_q^*$ are independent if each of them only includes observations the other estimators do not include. The estimators $\hat{k}_j^*$, $j = 1, \ldots, q$, are respectively based on observations

$$\left\{X_l : k_j - c\sigma^2 \tilde{d}_{j,n}^{-2} - G \leq l \leq k_j + c\sigma^2 \tilde{d}_{j,n}^{-2} + G\right\}, \qquad j = 1, \ldots, q.$$

For an arbitrary $\varepsilon > 0$ and n large it follows by **(D)** that these sets are subsets of

$$\{X_l : k_j - (1+\varepsilon)G \leq l \leq k_j + (1+\varepsilon)G\}, \qquad j = 1, \ldots, q,$$

respectively. By **(Cd)** we can choose an $\varepsilon > 0$ such that these sets are disjunct. Hence, it holds

$$\begin{aligned} & P\left(\left(\tilde{d}_{1,n}^2 \frac{\hat{k}_1^* - k_1}{\sigma^2}, \ldots, \tilde{d}_{q,n}^2 \frac{\hat{k}_q^* - k_q}{\sigma^2}\right) \leq (x_1, \ldots, x_q)\right) \\ = & \prod_{j=1}^q P\left(\tilde{d}_{j,n}^2 \frac{\hat{k}_j^* - k_j}{\sigma^2} \leq x_j\right) + o(1). \end{aligned}$$

Thus, it is sufficient to derive the marginal distributions. For $-c \leq x_j \leq c$ we obtain by the definition of $\hat{k}_j^*$ and with $V_{k,n}(G)$ defined as in (6.14)

$$\begin{aligned} P\left(\tilde{d}_{j,n}^2 \frac{\hat{k}_j^* - k_j}{\sigma^2} \leq x_j\right) &= P\left(-c \leq \tilde{d}_{j,n}^2 \frac{\hat{k}_j^* - k_j}{\sigma^2} \leq x_j\right) \\ = P\left(\max_{-c \leq \tilde{d}_{j,n}^2 \frac{k-k_j}{\sigma^2} \leq x_j} V_{k,n}(G) \geq \max_{x_j < \tilde{d}_{j,n}^2 \frac{k-k_j}{\sigma^2} \leq c} V_{k,n}(G)\right). \end{aligned}$$

It holds that $V_{k,n}(G) = 0$ for $k = k_j$. For $0 < k_j - k \leq c\sigma^2 \tilde{d}_{j,n}^{-2}$ we receive with decomposition (6.16) and the results in the proof of Lemma 6.9 that $K_1 = O(1)$ and further

$$\begin{aligned} V_{k,n}(G) &= K_1 + T_5 + o_p(K_1) \\ &= -\frac{1}{2} G^{-1} |k_j - k| (2G - |k_j - k|) \tilde{d}_{j,n}^2 - (T_{k_j,n}^0(G) - T_{k,n}^0(G))(2G - |k_j - k|) \frac{1}{\sqrt{2G}} \tilde{d}_{j,n} \\ &\quad + o_p(1) \\ &= -\frac{1}{2} G^{-1} |k_j - k| (2G - |k_j - k|) \tilde{d}_{j,n}^2 \\ &\quad - \left(\sum_{i=k+G+1}^{k_j+G} \varepsilon_i + \sum_{i=k-G+1}^{k_j-G} \varepsilon_i - 2 \sum_{i=k+1}^{k_j} \varepsilon_i\right) (2G - |k_j - k|) \frac{1}{2G} \tilde{d}_{j,n} + o_p(1) \\ &= \left(-|k_j - k| \tilde{d}_{j,n}^2 + \tilde{d}_{j,n} \left(\sum_{i=k+G+1}^{k_j+G} \varepsilon_i + \sum_{i=k-G+1}^{k_j-G} \varepsilon_i - 2 \sum_{i=k+1}^{k_j} \varepsilon_i\right)\right) \left(1 - \frac{1}{2G} |k_j - k|\right) \\ &\quad + o_p(1) \end{aligned}$$

$$= \left(-|k_j-k|\tilde{d}_{j,n}^2 + \tilde{d}_{j,n}\left(\sum_{i=k+G+1}^{k_j+G}\varepsilon_i + \sum_{i=k-G+1}^{k_j-G}\varepsilon_i - 2\sum_{i=k+1}^{k_j}\varepsilon_i\right)\right)\left(1+O\left(\frac{\tilde{d}_{j,n}^{-2}}{G}\right)\right)$$
$$+ o_p(1)$$
$$= \left(-|k_j-k|\tilde{d}_{j,n}^2 + \tilde{d}_{j,n}\left(\sum_{i=k+G+1}^{k_j+G}\varepsilon_i + \sum_{i=k-G+1}^{k_j-G}\varepsilon_i - 2\sum_{i=k+1}^{k_j}\varepsilon_i\right)\right)(1+o(1)) + o_p(1)$$

by assumption **(D)**. For $0 < k - k_j \leq c\sigma^2\tilde{d}_{j,n}^{-2}$ we receive

$$V_{k,n}(G)$$
$$= \left(-|k_j-k|\tilde{d}_{j,n}^2 - \tilde{d}_{j,n}\left(\sum_{i=k_j+G+1}^{k+G}\varepsilon_i + \sum_{i=k_j-G+1}^{k-G}\varepsilon_i - 2\sum_{i=k_j+1}^{k}\varepsilon_i\right)\right)(1+o(1)) + o_p(1).$$

With $\tilde{s}_{j,n} := s\sigma^2\tilde{d}_{j,n}^{-2}$ we define the process $\{Z_s(G)\}$ as

$$Z_s(G) \;=\; \begin{cases} -|s|\sigma^2 + \tilde{d}_{j,n}\left(\sum_{i=\lfloor\tilde{s}_{j,n}\rfloor+k_j+G+1}^{k_j+G}\varepsilon_i + \sum_{i=\lfloor\tilde{s}_{j,n}\rfloor+k_j-G+1}^{k_j-G}\varepsilon_i - 2\sum_{i=\lfloor\tilde{s}_{j,n}\rfloor+k_j+1}^{k_j}\varepsilon_i\right),\ s<0 \\ -|s|\sigma^2 - \tilde{d}_{j,n}\left(\sum_{i=k_j+G+1}^{\lfloor\tilde{s}_{j,n}\rfloor+k_j+G}\varepsilon_i + \sum_{i=k_j-G+1}^{\lfloor\tilde{s}_{j,n}\rfloor+k_j-G}\varepsilon_i - 2\sum_{i=k_j+1}^{\lfloor\tilde{s}_{j,n}\rfloor+k_j}\varepsilon_i\right),\ s>0 \\ 0,\ s=0. \end{cases}$$

Since $\max_{-c\leq\tilde{d}_{j,n}^2\frac{k-k_j}{\sigma^2}\leq c} V_{k,n}(G) = \max_{-c\leq s\leq c} Z_s(G)/\sigma^2$ we receive

$$P\left(\max_{-c\leq\tilde{d}_{j,n}^2\frac{k-k_j}{\sigma^2}\leq x_j} V_{k,n}(G) \geq \max_{x_j<\tilde{d}_{j,n}^2\frac{k-k_j}{\sigma^2}\leq c} V_{k,n}(G)\right)$$
$$= P\left(\max_{-c\leq s\leq x_j} Z_s(G) \geq \max_{x_j<s\leq c} Z_s(G)\right).$$

The functional central limit theorem (refer to Billingsley (1968), Theorem 16.1) yields

$$\frac{\tilde{d}_{j,n}}{\sigma^2}\left(\sum_{i=k_j+G+1}^{\tilde{s}_{j,n}+k_j+G}\varepsilon_i + \sum_{i=k_j-G+1}^{\tilde{s}_{j,n}+k_j-G}\varepsilon_i - 2\sum_{i=k_j+1}^{\tilde{s}_{j,n}+k_j}\varepsilon_i\right)$$
$$= \frac{1}{\sqrt{\tilde{d}_{j,n}^{-2}\sigma^2}}\left(\sum_{i=k_j+G+1}^{\tilde{s}_{j,n}+k_j+G}\frac{\varepsilon_i}{\sigma} + \sum_{i=k_j-G+1}^{\tilde{s}_{j,n}+k_j-G}\frac{\varepsilon_i}{\sigma} - 2\sum_{i=k_j+1}^{\tilde{s}_{j,n}+k_j}\frac{\varepsilon_i}{\sigma}\right)$$
$$\xrightarrow[n\to\infty]{\mathcal{D}} W_1(s) + W_2(s) - 2W_3(s),$$

where $\{W_1\}$, $\{W_2\}$ and $\{W_3\}$ are independent standard Wiener processes, since in each sum are different errors ε_i involved. Moreover, the functional central limit theorem holds for the sums above though they are shifted by G, since they are in distribution equal to the sums without the shift G, e.g. the first sum is in distribution equal to $\sum_{i=k_j+1}^{\tilde{s}_{j,n}+k_j} \frac{\varepsilon_i}{\sigma}$. Since the maximum is a measurable function it follows by Corollary 1 to Theorem 5.1 in Billingsley (1968)

$$
\begin{aligned}
P\left(-c \leq \tilde{d}_{j,n}^2 \frac{\hat{k}_j - k_j}{\sigma^2} \leq x_j\right) &= P\left(\max_{-c\leq s\leq x_j} \frac{1}{\sigma^2} Z_s(G) \geq \max_{x_j<s\leq c} \frac{1}{\sigma^2} Z_s(G)\right) \\
\xrightarrow[n\to\infty]{} \quad & P\Bigg(\max_{-c\leq s\leq x_j} (-|s| - (W_1(s) + W_2(s) - 2W_3(s))) \geq \\
& \quad \max_{x_j<s\leq c} (-|s| - (W_1(s) + W_2(s) - 2W_3(s)))\Bigg) \\
= \quad & P(-c \leq \arg\max_{-c\leq s\leq c} (-|s| - (W_1(s) + W_2(s) - 2W_3(s))) \leq x) \xrightarrow[c\to\infty]{} P(S \leq x),
\end{aligned}
$$

since $\{W_1(s)+W_2(s)-2W_3(s)\} \stackrel{\mathcal{D}}{=} \{-\sqrt{6}W(s)\}$, where $\{W(s)\}$ is a standard Wiener process, and a division with $\sqrt{6}$ does not change the arg max. □

6.2. Asymptotic Results for the Long-Run Variance Estimator

In the last section we derived uniform rates of consistency for the change-point estimators under the classical and the regime-witching model. These results were obtained under assumptions **(L0)** and **(L1)**, which include the uniform rate $o_p\left((\log(n/G))^{-1}\right)$ for the long-run variance estimator under the null hypothesis and the uniform stochastic boundedness under the alternative, respectively. In this section we analyse whether the long-run variance estimators, introduced in Section 5.3.3, fulfill these rates.

If the errors are an i.i.d. sequence the long-run variance is equal to the variance. Hence, it is enough just to estimate the variance. For this situation we introduced the estimator $\hat{\sigma}_{k,n}^2$ and some modified versions $\hat{\sigma}_{k,n,l}^2$, $\hat{\sigma}_{k,n,r}^2$ and $\hat{\sigma}_{k,n,m}^2$ in Section 5.3.3. In Section 6.2.1 we show that these estimators converge with the rates specified in assumptions **(L0)** and **(L1)**.

In case of dependend errors we proposed estimators $\hat{\tau}_{k,n}^2$, $\hat{\tau}_{k,n,l}^2$, $\hat{\tau}_{k,n,r}^2$ and $\hat{\tau}_{k,n,m}^2$. For these estimators the rate in assumption **(L0)** holds under certain conditions on bandwidths G and Λ_n. The stochastic boundedness in assumption **(L1)** under the alternative is not fulfilled, but the consistency results for the change-point estimators still hold if we alter condition **(D)**, respectively **(Dr)**, appropriately.

6.2.1. Convergence of the Variance Estimator

We consider the special case of an independent and identically distributed error sequence $\varepsilon_1, \ldots, \varepsilon_n$. To estimate the variance σ^2, we first consider the in (5.11) proposed estimator

$$
\hat{\sigma}_{k,n}^2 \quad := \quad \frac{1}{2G}\left(\sum_{i=k-G+1}^{k} (X_i - \overline{X}_{k-G+1,k})^2 + \sum_{i=k+1}^{k+G} (X_i - \overline{X}_{k+1,k+G})^2\right).
$$

This estimator depends on the bandwidth G and G has to fulfill assumption **(G)**, i.e. G tends to infinity as n goes to infinity but not faster than $n^{\frac{2}{2+\nu}} \log n$, where ν is the same ν as in the moment condition of the errors in **(E1)**. We choose the ν as large as possible to get a weak bandwidth assumption. If $\nu = 2$ is the largest ν for which **(E1)** holds we alter assumption **(G)** to (6.24) to show that the estimator $\hat{\sigma}^2_{k,n}$ converges uniformly in k with rate $o_p\left((\log(n/G))^{-1}\right)$ to σ^2 under the null hypothesis.

Theorem 6.14. Let $X_1, ..., X_n$ follow either model **(CCM)** or **(RSM)** and the errors be an i.i.d. sequence fulfilling **(E1)**. If $\nu = 2$ is the largest ν for which **(E1)** holds, assume for the bandwidth G

$$\frac{n}{G} \longrightarrow \infty \quad \text{and} \quad \frac{n^{\frac{2}{2+\nu}} \log n \sqrt{\log\log n}}{G} \longrightarrow 0. \tag{6.24}$$

Otherwise let bandwidth assumption **(G)** hold. Then, under H_0,

$$\max_{G \leq k \leq n-G} |\hat{\sigma}^2_{k,n} - \sigma^2| \quad = \quad o_p\left((\log(n/G))^{-1}\right).$$

Proof of Theorem 6.14. Under the null hypothesis we have $X_i - \overline{X}_{k+1,k+G} = \varepsilon_i - \overline{\varepsilon}_{k+1,k+G}$ and hence

$$\begin{aligned}
&\max_{G \leq k \leq n-G} |\hat{\sigma}^2_{k,n} - \sigma^2| \\
&= \max_{G \leq k \leq n-G} \left| \frac{1}{2G} \left(\sum_{i=k+1}^{k+G} \left(\varepsilon_i^2 - \sigma^2 - 2\varepsilon_i \overline{\varepsilon}_{k+1,k+G} + \overline{\varepsilon}^2_{k+1,k+G}\right) \right.\right. \\
&\qquad \left.\left. + \sum_{i=k-G+1}^{k} \left(\varepsilon_i^2 - \sigma^2 - 2\varepsilon_i \overline{\varepsilon}_{k-G+1,k} + \overline{\varepsilon}^2_{k-G+1,k}\right)\right)\right| \\
&\leq \max_{G \leq k \leq n-G} \frac{1}{2G} \left| \sum_{i=k+1}^{k+G} (\varepsilon_i^2 - \sigma^2) \right| + \max_{G \leq k \leq n-G} \frac{1}{2} \left|\overline{\varepsilon}^2_{k+1,k+G}\right| \qquad (6.25) \\
&\qquad + \max_{G \leq k \leq n-G} \frac{1}{2G} \left| \sum_{i=k-G+1}^{k} (\varepsilon_i^2 - \sigma^2) \right| + \max_{G \leq k \leq n-G} \frac{1}{2} \left|\overline{\varepsilon}^2_{k-G+1,k}\right|.
\end{aligned}$$

We use a result in Antoch et al. (2000) on page 24, which gives the asymptotic distribution of a properly standardised MOSUM statistic $\max_{G \leq k \leq n} \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} \varepsilon_i \right|$ under H_0, error assumption **(E1)** and bandwidths assumption **(G)**. This result yields

$$\max_{G \leq k \leq n} \frac{1}{\sqrt{2G}} \left| \sum_{i=k-G+1}^{k} \varepsilon_i \right| \quad = \quad O_p(\sqrt{\log(n/G)}) \tag{6.26}$$

and implies

$$\max_{G \leq k \leq n-G} |\overline{\varepsilon}_{k-G+1,k}|^2 \quad = \quad O_p\left(\frac{\log(n/G)}{G}\right) = o_p\left((\log(n/G))^{-1}\right).$$

If $\nu > 2$ holds, then moments of higher order than 4 exist and we can also apply (6.26) on

$$\max_{G\leq k\leq n}\frac{1}{2G}\left|\sum_{i=k-G+1}^{k}(\varepsilon_i^2-\sigma^2)\right| \;=\; O_p\left(\left(\frac{\log(n/G)}{G}\right)^{\frac{1}{2}}\right)=o_p\left((\log(n/G))^{-1}\right).$$

If $\nu = 2$ we obtain with the law of iterated logarithm that

$$\begin{aligned}\max_{G\leq k\leq n-G}\frac{1}{2G}\left|\sum_{i=k+1}^{k+G}(\varepsilon_i^2-\sigma^2)\right| \;&\leq\; \max_{G\leq k\leq n-G}\frac{1}{2G}\left|\sum_{i=1}^{k+G}(\varepsilon_i^2-\sigma^2)\right|+\max_{G\leq k\leq n-G}\frac{1}{2G}\left|\sum_{i=1}^{k}(\varepsilon_i^2-\sigma^2)\right|\\ &=\; O\left(\frac{\sqrt{n\log\log n}}{G}\right)=o\left((\log(n/G))^{-1}\right)\text{ a.s.,}\end{aligned}$$

because $2/(2+\nu) = 1/2$ and $\frac{\sqrt{n\log\log n}}{G}\log(n/G) \to 0$ by bandwidth assumption (6.24).

If $0 < \nu < 2$ we can use the Marcinkiewicz-Zygmund strong law of large numbers (refer to Theorem B.6 in the Appendix) with $p = (2+\nu)/2 \in (1, 2)$ and $E\left[\left|\varepsilon^2\right|^p\right] = E\left[|\varepsilon|^{2+\nu}\right] < \infty$ to show

$$\begin{aligned}\max_{G\leq k\leq n-G}\frac{1}{2G}\left|\sum_{i=k+1}^{k+G}(\varepsilon_i^2-\sigma^2)\right| \;&\leq\; \max_{G\leq k\leq n-G}\frac{1}{2G}\left|\sum_{i=1}^{k+G}(\varepsilon_i^2-\sigma^2)\right|+\max_{G\leq k\leq n-G}\frac{1}{2G}\left|\sum_{i=1}^{k}(\varepsilon_i^2-\sigma^2)\right|\\ &=\; O\left(\frac{n^{\frac{2}{2+\nu}}}{G}\right)=o\left((\log(n/G))^{-1}\right)\text{ a.s.}\end{aligned}\tag{6.27}$$

Analogous calculations yield the rate $o_p\left((\log(n/G))^{-1}\right)$ for the maxima in (6.25). □

The next theorem shows the uniform stochastic boundedness of $\hat{\sigma}_{k,n}^2$ under the alternative.

Theorem 6.15. Let the same assumptions as in Theorem 6.14 hold and assume additionally **(Cd)** and $\max_{1\leq j\leq q}\tilde{d}_{j,n}^2 = O(1)$, respectively **(Cr)** and $\max_{1\leq i<j\leq K}\tilde{d}_{i,j,n}^{*2} = O(1)$, on the change points. Then, we have under H_1

$$\max_{G\leq k\leq n-G}\hat{\sigma}_{k,n}^2 \;=\; O_p(1).$$

Proof of Theorem 6.15. To begin with, we consider the classical model **(CCM)**. Defining the subsets

$$\begin{aligned}A \;&:=\; \{k\in\{G,\ldots,n-G\}:\ |k_j-k|\geq G\ \forall\, k_1,\ldots,k_q\},\\ B \;&:=\; \{k\in\{G,\ldots,n-G\}:\ \exists k_j\in\{k_1,\ldots,k_q\}\text{ with } 0<k_j-k<G\},\\ C \;&:=\; \{k\in\{G,\ldots,n-G\}:\ \exists k_j\in\{k_1,\ldots,k_q\}\text{ with } 0\leq k-k_j<G\},\end{aligned}$$

yields

$$\max_{G\leq k\leq n-G}\hat{\sigma}_{k,n}^2\leq\max_{k\in A}\hat{\sigma}_{k,n}^2+\max_{k\in B}\hat{\sigma}_{k,n}^2+\max_{k\in C}\hat{\sigma}_{k,n}^2.$$

We obtain for n large

$$\max_{k\in A}\hat{\sigma}_{k,n}^2 = \max_{k\in A}\frac{1}{2G}\left|\sum_{i=k-G+1}^{k}(\varepsilon_i-\overline{\varepsilon}_{k-G+1,k})^2+\sum_{i=k+1}^{k+G}(\varepsilon_i-\overline{\varepsilon}_{k+1,k+G})^2\right|$$
$$\leq \max_{G\leq k\leq n-G}\frac{1}{2G}\left|\sum_{i=k-G+1}^{k}(\varepsilon_i-\overline{\varepsilon}_{k-G+1,k})^2+\sum_{i=k+1}^{k+G}(\varepsilon_i-\overline{\varepsilon}_{k+1,k+G})^2\right| = O_p(1)$$

by Lemma 6.14. Further we deal with the case of $k\in B$ and n large. If n is large enough assumption **(Cd)** ensures that B is the disjunct union of the sets $\{k: 0<k_j-k<G\}$, $j=1,\dots,q$. For k in an arbitrary but fixed set $\{k:0<k_j-k<G\}$ we obtain

$$\overline{X}_{k-G+1,k} = \overline{\varepsilon}_{k-G+1,k}+d_{j,n} \quad \text{and} \tag{6.28}$$
$$\overline{X}_{k+1,k+G} = \overline{\varepsilon}_{k+1,k+G}+\frac{(k_j-k)d_{j,n}+(k+G-k_j)d_{j+1,n}}{G}. \tag{6.29}$$

Hence, we receive

$$\begin{aligned}
\hat{\sigma}_{k,n}^2 &= \frac{1}{2G}\Bigg(\sum_{i=k-G+1}^{k}(\varepsilon_i-\overline{\varepsilon}_{k-G+1,k})^2 \\
&\quad+\sum_{i=k+1}^{k_j}\left(\varepsilon_i+d_{j,n}-\overline{\varepsilon}_{k+1,k+G}-\frac{(k_j-k)d_{j,n}+(k+G-k_j)d_{j+1,n}}{G}\right)^2 \\
&\quad+\sum_{i=k_j+1}^{k+G}\left(\varepsilon_i+d_{j+1,n}-\overline{\varepsilon}_{k+1,k+G}-\frac{(k_j-k)d_{j,n}+(k+G-k_j)d_{j+1,n}}{G}\right)^2\Bigg) \\
&= \frac{1}{2G}\Bigg(\sum_{i=k-G+1}^{k}(\varepsilon_i-\overline{\varepsilon}_{k-G+1,k})^2+\sum_{i=k+1}^{k_j}\left(\varepsilon_i-\overline{\varepsilon}_{k+1,k+G}-\frac{(k+G-k_j)\tilde{d}_{j,n}}{G}\right)^2 \\
&\quad+\sum_{i=k_j+1}^{k+G}\left(\varepsilon_i-\overline{\varepsilon}_{k+1,k+G}-\frac{(k-k_j)\tilde{d}_{j,n}}{G}\right)^2\Bigg) \\
&\leq \frac{1}{G}\left(\sum_{i=k-G+1}^{k}(\varepsilon_i-\overline{\varepsilon}_{k-G+1,k})^2+\sum_{i=k+1}^{k+G}(\varepsilon_i-\overline{\varepsilon}_{k+1,k+G})^2\right) \\
&\quad+\frac{1}{G}\left(\frac{(k_j-k)(k+G-k_j)^2\tilde{d}_{j,n}^2}{G^2}+\frac{(k_j-k)^2(k+G-k_j)\tilde{d}_{j,n}^2}{G^2}\right),
\end{aligned}$$

since the inequality $(a+b)^2\leq 2(a^2+b^2)$ for $a,b\in\mathbb{R}$ holds. Applying Lemma 6.14 yields

$$\max_{k\in B}\frac{1}{2G}\left|\sum_{i=k-G+1}^{k}(\varepsilon_i-\overline{\varepsilon}_{k-G+1,k})^2+\sum_{i=k+1}^{k+G}(\varepsilon_i-\overline{\varepsilon}_{k+1,k+G})^2\right| = O_p(1).$$

Moreover, we receive by $k_j - k < G$

$$\begin{aligned}
&\max_{0<k_j-k<G} \frac{1}{G}\left|\frac{(k_j-k)(k+G-k_j)^2\tilde{d}_{j,n}^2}{G^2}+\frac{(k_j-k)^2(k+G-k_j)\tilde{d}_{j,n}^2}{G^2}\right| \\
&= \max_{0<k_j-k<G}\left|\frac{1}{G}\frac{(k_j-k)(k+G-k_j)G\tilde{d}_{j,n}^2}{G^2}\right| \le \max_{1\le j\le q}\tilde{d}_{j,n}^2 = O(1).
\end{aligned}$$

These calculations hold for every set $\{k : 0 < k_j - k < G\}$, $j = 1, \ldots, q$. Hence, we can conclude that $\max_{k\in B}\hat{\sigma}_{k,n}^2 = O_p(1)$. With similar calculations we receive $\max_{k\in C}\hat{\sigma}_{k,n}^2 = O_p(1)$ and obtain the assertion under the classical model. For the calculations it is only important that the distance between two adjacent change points is larger than $2G$ for n large. Under the regime-switching model we can use the law of total probability and assumption **(Cr)** to conclude

$$\begin{aligned}
&P\left(\max_{G\le k\le n-G}\hat{\sigma}_{k,n}^2 > \varepsilon\right) \\
&= P\left(\max_{G\le k\le n-G}\hat{\sigma}_{k,n}^2 > \varepsilon,\ d_0(n) > 2G\right) + P\left(\max_{G\le k\le n-G}\hat{\sigma}_{k,n}^2 > \varepsilon,\ d_0(n) \le 2G\right) \\
&= P\left(\max_{G\le k\le n-G}\hat{\sigma}_{k,n}^2 > \varepsilon,\ d_0(n) > 2G\right) + o(1).
\end{aligned}$$

In this situation the arguments for the classical model remain true and we obtain the assertion. □

In Remark 5.2 we introduced modified variance estimators for which we obtain in the next corollary the same uniform consistency rates.

Corollary 6.16. Let the assumptions of Theorem 6.14 hold.

(a) Then, under H_0

$$\begin{aligned}
\max_{G\le k\le n-G}|\hat{\sigma}_{k,n,l}^2-\sigma^2| &= o_p\left((\log(n/G))^{-1}\right), \\
\max_{G\le k\le n-G}|\hat{\sigma}_{k,n,r}^2-\sigma^2| &= o_p\left((\log(n/G))^{-1}\right), \\
\max_{G\le k\le n-G}|\hat{\sigma}_{k,n,m}^2-\sigma^2| &= o_p\left((\log(n/G))^{-1}\right).
\end{aligned}$$

(b) Assume additionally **(Cd)** on the change points. Then, under H_1

$$\begin{aligned}
\max_{G\le k\le n-G}\hat{\sigma}_{k,n,l}^2 &= O_p(1), \\
\max_{G\le k\le n-G}\hat{\sigma}_{k,n,r}^2 &= O_p(1), \\
\max_{G\le k\le n-G}\hat{\sigma}_{k,n,m}^2 &= O_p(1).
\end{aligned}$$

Proof of Corollary 6.16. In part (a) the results for variance estimators $\hat{\sigma}^2_{k,n,l}$ and $\hat{\sigma}^2_{k,n,r}$ are shown in the proof of Theorem 6.14. For the third estimator $\hat{\sigma}^2_{k,n,m}$ we obtain

$$\begin{aligned}\max_{G\le k\le n-G} |\hat{\sigma}^2_{k,n,m} - \sigma^2| &= \max_{G\le k\le n-G} |\min(\hat{\sigma}^2_{k,n,l}, \hat{\sigma}^2_{k,n,r}) - \sigma^2| \\ \le \max\left(\max_{G\le k\le n-G} |\hat{\sigma}^2_{k,n,l} - \sigma^2|, \max_{G\le k\le n-G} |\hat{\sigma}^2_{k,n,r} - \sigma^2|\right) &= o_p\left((\log(n/G))^{-1}\right)\end{aligned}$$

by the first two assertions. In part (b) the results for $\hat{\sigma}^2_{k,n,l}$ and $\hat{\sigma}^2_{k,n,r}$ are shown in Theorem 6.15 and the assertion for $\hat{\sigma}^2_{k,n,m}$ is obtained as above. □

6.2.2. Convergence of the Long-Run Variance Estimator

We consider the long-run variance estimator

$$\hat{\tau}^2_{k,n} := \hat{\gamma}_k(0) + 2\sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n)\hat{\gamma}_k(h)$$

with autocovariance estimator

$$\begin{aligned}\hat{\gamma}_k(h) := \; & \frac{1}{2G}\sum_{i=k-G+1}^{k-h} (X_i - \overline{X}_{k-G+1,k})(X_{i+h} - \overline{X}_{k-G+1,k}) \\ & + \frac{1}{2G}\sum_{i=k+1}^{k+G-h} (X_i - \overline{X}_{k+1,k+G})(X_{i+h} - \overline{X}_{k+1,k+G}),\end{aligned}$$

Bartlett weights $\omega(x) := (1-x)I\{|x| \le 1\}$ and bandwidth Λ_n. To obtain rates of convergence we make an assumption on the fourth order cumulant of the error sequence as in Berkes et al. (2005), who have shown the almost sure convergence of the common Bartlett estimator under weak and strong dependence assumptions. The fourth order cumulant $\kappa(h, r, s)$ is given by

$$\kappa(h, r, s) = E\varepsilon_i\varepsilon_{i+h}\varepsilon_{i+r}\varepsilon_{i+s} - \gamma(h)\gamma(r-s) - \gamma(r)\gamma(h-s) - \gamma(s)\gamma(h-r)$$

and we define the quantity

$$\begin{aligned}\nu(h, r, s) &= \operatorname{cov}(\varepsilon_i\varepsilon_{i+h}, \varepsilon_{i+r}\varepsilon_{i+s}) = E\varepsilon_i\varepsilon_{i+h}\varepsilon_{i+r}\varepsilon_{i+s} - \gamma(h)\gamma(r-s) \\ &= \kappa(h, r, s) + \gamma(r)\gamma(h-s) + \gamma(s)\gamma(h-r).\end{aligned}$$

In Theorem 6.17 we obtain the uniform consistency of the long-run variance estimator $\hat{\tau}^2_{k,n}$ with rate $o_p\left((\log(n/G))^{-1}\right)$. To receive this result we have to replace bandwidth assumption **(G)** by (6.30). This assumption prevents that the assumptions on the bandwidth can be weakend by the existence of higher momements than 4.

Theorem 6.17. Let $X_1, ..., X_n$ follow either model **(CCM)** or **(RSM)** and the errors fulfill **(E1)**, **(E3)**. Let $0 < \Delta < 1/2$ such that the two following assumptions hold: for the bandwidth G let

$$\frac{n}{G} \longrightarrow \infty \qquad \text{and} \qquad \frac{n^{\frac{1}{2}+\Delta}\log(n/G)}{G} \longrightarrow 0 \tag{6.30}$$

and for Λ_n

$$\Lambda_n \longrightarrow \infty, \qquad \frac{\Lambda_n}{G} \longrightarrow 0, \qquad \frac{(\log(n/G))^2}{\Lambda_n} \longrightarrow 0 \qquad \text{and} \qquad \frac{\Lambda_n}{n^\Delta} = O(1). \tag{6.31}$$

Suppose additionally that

$$\sum_{h=1}^{\infty} \sqrt{h}\,|\gamma(h)| < \infty \tag{6.32}$$

and

$$\sup_{-\infty \le h \le \infty} \sum_{r,s=-\infty}^{\infty} |\kappa(h,r,s)| < \infty. \tag{6.33}$$

Then, under H_0,

$$\max_{G \le k \le n-G} |\hat{\tau}_{k,n}^2 - \tau^2| \;=\; o_p\left((\log(n/G))^{-1}\right).$$

Proof to Theorem 6.17. By the definition of the estimator $\hat{\tau}_{k,n}^2$ and the triangular inequality it follows

$$\begin{aligned}
\max_{G \le k \le n-G} |\hat{\tau}_{k,n}^2 - \tau^2| &= \max_{G \le k \le n-G} \left| \hat{\sigma}_{k,n}^2 + 2\sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n)\hat{\gamma}_k(h) - \sigma^2 - 2\sum_{h>0}\gamma(h) \right| \\
= \quad & \max_{G \le k \le n-G} \left| \hat{\sigma}_{k,n}^2 - \sigma^2 + \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G} \sum_{i=k-G+1}^{k-h} (\varepsilon_i - \overline{\varepsilon}_{k-G+1,k})(\varepsilon_{i+h} - \overline{\varepsilon}_{k-G+1,k}) \right. \\
& \left. + \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G} \sum_{i=k+1}^{k+G-h} (\varepsilon_i - \overline{\varepsilon}_{k+1,k+G})(\varepsilon_{i+h} - \overline{\varepsilon}_{k+1,k+G}) - 2\sum_{h>0}\gamma(h) \right| \\
\le \quad & \max_{G \le k \le n-G} \left|\hat{\sigma}_{k,n}^2 - \sigma^2\right| && (6.34) \\
& + \max_{G \le k \le n-G} \left| \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G} \sum_{i=k-G+1}^{k-h} (\varepsilon_i - \overline{\varepsilon}_{k-G+1,k})(\varepsilon_{i+h} - \overline{\varepsilon}_{k-G+1,k}) - \sum_{h>0}\gamma(h) \right| && (6.35) \\
& + \max_{G \le k \le n-G} \left| \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G} \sum_{i=k+1}^{k+G-h} (\varepsilon_i - \overline{\varepsilon}_{k+1,k+G})(\varepsilon_{i+h} - \overline{\varepsilon}_{k+1,k+G}) - \sum_{h>0}\gamma(h) \right|. && (6.36)
\end{aligned}$$

To begin with, we show that (6.35) is of order $o_p((\log(n/G))^{-1})$. Therefore, we define

$$\bar{\varepsilon}_{k,l} := \frac{1}{G}\sum_{i=k}^{l}\varepsilon_i,$$

consider

$$\begin{aligned}
&\frac{1}{G}\sum_{i=k-G+1}^{k-h}(\varepsilon_i-\bar{\varepsilon}_{k-G+1,k})(\varepsilon_{i+h}-\bar{\varepsilon}_{k-G+1,k})\\
&= \frac{1}{G}\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\varepsilon_i\bar{\varepsilon}_{k-G+1,k}-\bar{\varepsilon}_{k-G+1,k}\varepsilon_{i+h}+\bar{\varepsilon}^2_{k-G+1,k})\\
&= \frac{1}{G}\sum_{i=k-G+1}^{k-h}\varepsilon_i\varepsilon_{i+h}-\bar{\varepsilon}_{k-G+1,k-h}\bar{\varepsilon}_{k-G+1,k}-\bar{\varepsilon}_{k-G+1,k}\bar{\varepsilon}_{k-G+h+1,k}+\frac{G-h}{G}\bar{\varepsilon}^2_{k-G+1,k}\\
&= \frac{1}{G}\sum_{i=k-G+1}^{k-h}\varepsilon_i\varepsilon_{i+h}-\bar{\varepsilon}_{k-G+1,k}\left(\bar{\varepsilon}_{k-G+1,k-h}+\bar{\varepsilon}_{k-G+h+1,k}\right)+\frac{G-h}{G}\bar{\varepsilon}^2_{k-G+1,k}\\
&= \frac{1}{G}\sum_{i=k-G+1}^{k-h}\varepsilon_i\varepsilon_{i+h}-\bar{\varepsilon}_{k-G+1,k}\left(\bar{\varepsilon}_{k-G+1,k}+\bar{\varepsilon}_{k-G+h+1,k-h}\right)+\frac{G-h}{G}\bar{\varepsilon}^2_{k-G+1,k}\\
&= \frac{1}{G}\sum_{i=k-G+1}^{k-h}\varepsilon_i\varepsilon_{i+h}-\bar{\varepsilon}^2_{k-G+1,k}-\bar{\varepsilon}_{k-G+1,k}\bar{\varepsilon}_{k-G+h+1,k-h}+\frac{G-h}{G}\bar{\varepsilon}^2_{k-G+1,k}\\
&= \frac{1}{G}\sum_{i=k-G+1}^{k-h}\varepsilon_i\varepsilon_{i+h}-\bar{\varepsilon}_{k-G+1,k}\bar{\varepsilon}_{k-G+h+1,k-h}-\frac{h}{G}\bar{\varepsilon}^2_{k-G+1,k}
\end{aligned}$$

and show successively

$$\max_{G\le k\le n-G}\left|\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}\varepsilon_i\varepsilon_{i+h}-\sum_{h>0}\gamma(h)\right| = o_p((\log(n/G))^{-1}), \tag{6.37}$$

$$\max_{G\le k\le n-G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\bar{\varepsilon}_{k-G+1,k}\bar{\varepsilon}_{k-G+h+1,k-h}\right| = o_p((\log(n/G))^{-1}), \tag{6.38}$$

$$\max_{G\le k\le n-G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\frac{h}{G}\bar{\varepsilon}^2_{k-G+1,k}\right| = o_p((\log(n/G))^{-1}). \tag{6.39}$$

To prove (6.37) we decompose

$$\begin{aligned}
&\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}\varepsilon_i\varepsilon_{i+h}\\
&= \frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))+\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)(G-h)\gamma(h). \qquad (6.40)
\end{aligned}$$

We use the specific form of the Bartlett weights to receive

$$
\begin{aligned}
\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)(G-h)\gamma(h) &= \frac{1}{G}\sum_{h=1}^{\Lambda_n}\left(1-\frac{h}{\Lambda_n}\right)(G-h)\gamma(h) \qquad (6.41)\\
&= \sum_{h=1}^{\Lambda_n}\gamma(h)-\sum_{h=1}^{\Lambda_n}\frac{h}{\Lambda_n}\gamma(h)-\frac{1}{G}\sum_{h=1}^{\Lambda_n}h\gamma(h)+\frac{1}{G}\sum_{h=1}^{\Lambda_n}\frac{h^2}{\Lambda_n}\gamma(h)\\
&= \sum_{h=1}^{\Lambda_n}\gamma(h)+o_p\left((\log(n/G))^{-1}\right),
\end{aligned}
$$

since assumptions (6.31) and (6.32) of this theorem yield

$$
\sum_{h=1}^{\Lambda_n}\frac{h}{\Lambda_n}\gamma(h) \;\leq\; \Lambda_n^{-\frac{1}{2}}\sum_{h=1}^{\Lambda_n}\sqrt{h}\,|\gamma(h)| \;=\; o_p\left((\log(n/G))^{-1}\right),
$$

$$
\frac{1}{G}\sum_{h=1}^{\Lambda_n}h\gamma(h) \;\leq\; \frac{\sqrt{\Lambda_n}}{G}\sum_{h=1}^{\Lambda_n}\sqrt{h}\,|\gamma(h)| \;=\; o_p\left((\log(n/G))^{-1}\right)
$$

and

$$
\frac{1}{G}\sum_{h=1}^{\Lambda_n}\frac{h^2}{\Lambda_n}\gamma(h) \;\leq\; \frac{\sqrt{\Lambda_n}}{G}\sum_{h=1}^{\Lambda_n}\sqrt{h}\,|\gamma(h)| \;=\; o_p\left((\log(n/G))^{-1}\right).
$$

This implies

$$
\begin{aligned}
\left|\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)(G-h)\gamma(h)-\sum_{h>0}\gamma(h)\right| &= \left|\sum_{h=1}^{\Lambda_n}\gamma(h)-\sum_{h>0}\gamma(h)\right|+o_p((\log(n/G))^{-1})\\
&= \left|\sum_{h=\Lambda_n+1}^{\infty}\gamma(h)\right|+o_p((\log(n/G))^{-1}) \;=\; o_p((\log(n/G))^{-1}),
\end{aligned}
$$

where the last step is obtained by

$$
\left|\sum_{h=\Lambda_n+1}^{\infty}\gamma(h)\right| \;\leq\; \Lambda_n^{-\frac{1}{2}}\sum_{h=1}^{\infty}\sqrt{h}\,|\gamma(h)| \;=\; o_p\left((\log(n/G))^{-1}\right).
$$

For the first term in (6.40) we prove the inequality $\forall\varepsilon>0$

$$
P\left(\max_{G\leq k\leq n-G}\left|\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right|>\varepsilon\right) \;=\; nG^{-2}\Lambda_n^2\varepsilon^{-2}C, \qquad (6.42)
$$

where $C := \sup_{1\le h\le \Lambda_n} \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} |\nu(h,k,l)| > 0$. First, we obtain

$$
\begin{aligned}
& P\left(\max_{G\le k\le n-G}\left|\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right|>\varepsilon\right)\\
&= P\left(\max_{G\le k\le n-G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right|>G\varepsilon\right)\\
&\le \sum_{k=G}^{n-G} P\left(\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right|>G\varepsilon\right).
\end{aligned}
$$

Furthermore we apply the Markov inequality and use the stationarity of the error sequence to conclude

$$
\begin{aligned}
& \sum_{k=G}^{n-G} P\left(\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right|>G\varepsilon\right)\\
&\le G^{-2}\varepsilon^{-2}\sum_{k=G}^{n-G} E\left(\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right)^2\\
&= G^{-2}\varepsilon^{-2}\sum_{k=G}^{n-G} E\left(\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=1}^{G-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right)^2\\
&\le nG^{-2}\varepsilon^{-2} E\left(\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=1}^{G-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\sum_{b=1}^{\Lambda_n}\omega(b/\Lambda_n)\sum_{j=1}^{G-b}(\varepsilon_j\varepsilon_{j+h}-\gamma(h))\right)\\
&= nG^{-2}\varepsilon^{-2}\sum_{h=1}^{\Lambda_n}\sum_{b=1}^{\Lambda_n}\omega(h/\Lambda_n)\omega(b/\Lambda_n)\sum_{i=1}^{G-h}\sum_{j=1}^{G-b} E\left((\varepsilon_i\varepsilon_{i+h}-\gamma(h))(\varepsilon_j\varepsilon_{j+b}-\gamma(b))\right).
\end{aligned}
$$

The boundedness of the weights by 1 and

$$
E\left((\varepsilon_i\varepsilon_{i+h}-\gamma(h))(\varepsilon_j\varepsilon_{j+b}-\gamma(b))\right) = \operatorname{cov}(\varepsilon_i\varepsilon_{i+h},\varepsilon_j\varepsilon_{j+b}) = \nu(h,j-i,j-i+b)
$$

yield

$$
\begin{aligned}
& \sum_{h=1}^{\Lambda_n}\sum_{b=1}^{\Lambda_n}\omega(h/\Lambda_n)\omega(b/\Lambda_n)\sum_{i=1}^{G-h}\sum_{j=1}^{G-b} E\left((\varepsilon_i\varepsilon_{i+h}-\gamma(h))(\varepsilon_j\varepsilon_{j+b}-\gamma(b))\right)\\
&\le \sum_{h=1}^{\Lambda_n}\sum_{b=1}^{\Lambda_n}\sum_{i=1}^{G-h}\sum_{j=1}^{G-b}|\nu(h,j-i,j-i+b)|\\
&\le \sum_{h=1}^{\Lambda_n}\sum_{b=1}^{\Lambda_n}\sup_{1\le h\le\Lambda_n}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\nu(h,k,l)|\\
&= \Lambda_n^2 O(1),
\end{aligned}
$$

where the last step follows by assumption (6.33). Hence, the proof of (6.42) is complete. By (6.42), bandwidth assumptions (6.30) and (6.31) we have

$$\max_{G\le k\le n-G}\left|\frac{1}{G}\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\sum_{i=k-G+1}^{k-h}(\varepsilon_i\varepsilon_{i+h}-\gamma(h))\right| = O_p\left(G^{-1}n^{\frac{1}{2}}\Lambda_n\right)=o_p((\log(n/G))^{-1}).$$

To obtain (6.38) we use Lemma 6.5 to receive

$$\begin{aligned} P\left(\max_{G\le k\le n-G}\left|\sum_{i=1}^{k}\varepsilon_i\right|>\beta_n\right) &= CA(\varphi,\gamma)\beta_n^{-\gamma}\left(G^{\varphi}+\sum_{k=G+1}^{n}k^{\varphi-1}\right)\\ &= O(1)\beta_n^{-\gamma}n^{\frac{\gamma}{2}}\end{aligned}$$

and hence

$$\max_{G\le k\le n-G}\left|\sum_{i=k-G+1}^{k}\varepsilon_i\right| = O_p(\sqrt{n}). \tag{6.43}$$

This implies

$$\begin{aligned}&\max_{G\le k\le n-G}|\overline{\varepsilon}_{k-G+h+1,k-h}| \le \max_{G\le k\le n-G}|\overline{\varepsilon}_{1,k-h}|+\max_{G\le k\le n-G}|\overline{\varepsilon}_{1,k-G+h}|\\ &\le 2\max_{1\le k\le n}|\overline{\varepsilon}_{1,k}| = n^{\frac{1}{2}}G^{-1}O_p(1),\end{aligned}$$

which yields together with bandwidth assumptions (6.30) and (6.31) assertion (6.38):

$$\begin{aligned}&\max_{G\le k\le n-G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\overline{\varepsilon}_{k-G+1,k}\overline{\varepsilon}_{k-G+h+1,k-h}\right|\\ &\le \max_{G\le k\le n-G}|\overline{\varepsilon}_{k-G+1,k}|\sum_{h=1}^{\Lambda_n}\max_{G\le k\le n-G}|\overline{\varepsilon}_{k-G+h+1,k-h}|\\ &= O_p(1)\Lambda_nG^{-2}n = o_p(\log(n/G)^{-1}),\end{aligned}$$

since

$$\frac{n\log(n/G)\Lambda_n}{G^2} = \frac{n^{1+\Delta}\log(n/G)}{G^2}\frac{\Lambda_n}{n^{\Delta}}\longrightarrow 0.$$

For (6.39) we have with the boundedness of the weights by 1, (6.43) and bandwidth assumptions (6.30) and (6.31)

$$\begin{aligned}\max_{G\le k\le n-G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\frac{h}{G}\bar{\varepsilon}^2_{k-G+1,k}\right| &= G^{-1}\left(\max_{G\le k\le n-G}\left|\bar{\varepsilon}_{k-G+1,k}\right|\right)^2\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)h\\ \le G^{-1}\left(\max_{G\le k\le n-G}\left|\bar{\varepsilon}_{k-G+1,k}\right|\right)^2\sum_{h=1}^{\Lambda_n}h &= O(1)G^{-1}G^{-2}n\Lambda_n^2 = o_p(\log(n/G)^{-1}).\end{aligned}$$

Analogously the results for (6.36) can be obtained. From (6.25) in the proof of Lemma 6.14 we have for (6.34)

$$\begin{aligned}&\max_{G\le k\le n-G}\left|\hat{\sigma}^2_{k,n}-\sigma^2\right|\\ &\le \max_{G\le k\le n-G}\frac{1}{2G}\left|\sum_{i=k+1}^{k+G}(\varepsilon_i^2-\sigma^2)\right| + \max_{G\le k\le n-G}\frac{1}{2}\left|\bar{\varepsilon}^2_{k+1,k+G}\right|\\ &\quad + \max_{G\le k\le n-G}\frac{1}{2G}\left|\sum_{i=k-G+1}^{k}(\varepsilon_i^2-\sigma^2)\right| + \max_{G\le k\le n-G}\frac{1}{2}\left|\bar{\varepsilon}^2_{k-G+1,k}\right|.\end{aligned}$$

By (6.43) and bandwidth assumptions (6.30) as well as (6.31) we have

$$\max_{G\le k\le n-G}\frac{1}{2}\left|\bar{\varepsilon}^2_{k-G+1,k}\right| = O_p\left(G^{-2}n\right) = o_p(\log(n/G)^{-1}).$$

Analogously to (6.42) we obtain

$$P\left(\max_{G\le k\le n-G}\left|\frac{1}{G}\sum_{i=k-G+1}^{k}\left(\varepsilon_i^2-\sigma^2\right)\right| > \varepsilon\right) = nG^{-2}\varepsilon^{-2}O(1)$$

and therefore

$$\max_{G\le k\le n-G}\left|\frac{1}{G}\sum_{i=k-G+1}^{k}\left(\varepsilon_i^2-\sigma^2\right)\right| = o_p((\log(n/G))^{-1}).$$

This yields the assertion. □

In Theorem 6.18 we can only show that the rate $O_p(\Lambda_n)$ holds under the alternative. But the consistency results for the change-point estimators remain true if we modify assumption **(D)** to

$$\min_{1\le j\le q}\tilde{d}^2_{j,n}\frac{G}{\Lambda_n\log(n/G)} \longrightarrow \infty,$$

since then (6.8) still holds. This condition is fulfilled for fixed changes due to assumptions (6.30) and (6.31) on Λ_n. The same holds for the regime-switching model with the appropriate modification of **(Dr)**.

Theorem 6.18. Let the assumptions of Theorem 6.17 hold and assume additionally **(Cd)** and $\max_{1\le j\le q} \tilde{d}_{j,n}^2 = O(1)$, respectively **(Cr)** and $\max_{1\le i<j\le K} \tilde{d}_{i,j,n}^{*2} = O(1)$, on the change points. Then, under H_1,

$$\max_{G\le k\le n-G} \left|\hat{\tau}_{k,n}^2 - \tau^2\right| = O_p\left(\Lambda_n\right).$$

Proof of Theorem 6.18. We define the subsets

$$\begin{aligned}
A &:= \{k \in \{G,\dots,n-G\} :\ |k_j - k| \ge G \ \forall\, k_1,\dots,k_q\},\\
B &:= \{k \in \{G,\dots,n-G\} :\ \exists k_j \in \{k_1,\dots,k_q\} \text{ with } 0 \le k_j - k < G\} \quad \text{and}\\
C &:= \{k \in \{G,\dots,n-G\} :\ \exists k_j \in \{k_1,\dots,k_q\} \text{ with } 0 < k - k_j < G\}
\end{aligned}$$

of $\{G,\dots,n-G\}$ and show that the maximum of $\hat{\tau}_{k,n}^2$ for k in A, B and C is of order Λ_n. For $k \in A$ and n large enough due to assumption **(Cd)** no change point is included in the estimation of $\hat{\tau}_{k,n}$, i.e.

$$\hat{\tau}_{k,n}^2 = \hat{\tau}_{k,n,0}^2, \qquad k \in A,$$

where $\hat{\tau}_{k,n,0}^2$ is the long-run variance estimator applied on the errors:

$$\begin{aligned}
\hat{\tau}_{k,n,0}^2 := \sum_{h=0}^{\Lambda_n} \omega(h/\Lambda_n)\frac{1}{G}\Bigg(&\sum_{i=k-G+1}^{k-h} (\varepsilon_i - \overline{\varepsilon}_{k-G+1,k})(\varepsilon_{i+h} - \overline{\varepsilon}_{k-G+1,k})\\
&+ \sum_{i=k+1}^{k+G-h} (\varepsilon_i - \overline{\varepsilon}_{k+1,k+G})(\varepsilon_{i+h} - \overline{\varepsilon}_{k+1,k+G})\Bigg).
\end{aligned}$$

With Theorem 6.17 we obtain

$$\max_{k\in A} \hat{\tau}_{k,n}^2 = \max_{k\in A} \hat{\tau}_{k,n,0}^2 \le \max_{G\le k\le n-G} \hat{\tau}_{k,n,0}^2 = O_p(1).$$

Next we deal with the case of $k \in B$. Since B ist the union of the sets $\{k : 0 \le k_j - k < G\}$, $j = 1,\dots,q$, we consider the maximum over an arbitrary but fixed set $\{k : 0 \le k_j - k < G\}$ and show that

$$\max_{0\le k_j-k<G} \hat{\tau}_{k,n}^2 = O_p(\Lambda_n), \tag{6.44}$$

where the rate is uniformly in $j = 1,\dots,q$. Hence, we receive

$$\max_{k\in B} \hat{\tau}_{k,n}^2 = O_p(\Lambda_n).$$

The calculation of $\max\limits_{0<k_j-k<G} \hat{\tau}^2_{k,n}$ includes the observations X_i, $i = k_j - 2G + 1, \ldots, k_j + G$, with

$$X_i = \varepsilon_i + d_{j,n} I\{i \leq k_j\} + d_{j+1,n} I\{i > k_j\}$$

as soon as n is large enough such that $k_j - k_{j-1} > 2G$ as well as $k_{j+1} - k_j > G$. Then the empirical means $\overline{X}_{k-G+1,k}$, $\overline{X}_{k+1,k+G}$, $k = k_j - G + 1, \ldots, k_j$, are given by (6.28) and (6.29) as

$$\begin{aligned}
\overline{X}_{k-G+1,k} &= \overline{\varepsilon}_{k-G+1,k} + d_{j,n} \qquad \text{and} \\
\overline{X}_{k+1,k+G} &= \overline{\varepsilon}_{k+1,k+G} + \frac{(k_j - k) d_{j,n} + (k + G - k_j) d_{j+1,n}}{G}.
\end{aligned}$$

Therefrom, we obtain for $k_j - G < k \leq k_j$

$$\sum_{i=k-G+1}^{k-h} (X_i - \overline{X}_{k-G+1,k})(X_{i+h} - \overline{X}_{k-G+1,k}) = \sum_{i=k-G+1}^{k-h} (\varepsilon_i - \overline{\varepsilon}_{k-G+1,k})(\varepsilon_{i+h} - \overline{\varepsilon}_{k-G+1,k})$$

and

$$\begin{aligned}
&\sum_{i=k+1}^{k+G-h} (X_i - \overline{X}_{k+1,k+G})(X_{i+h} - \overline{X}_{k+1,k+G}) \\
&= \sum_{i=k+1}^{k+G-h} \left(\varepsilon_i - \overline{\varepsilon}_{k+1,k+G} + D_{i,j,k}\right) \left(\varepsilon_{i+h} - \overline{\varepsilon}_{k+1,k+G} + D_{i+h,j,k}\right)
\end{aligned}$$

with

$$\begin{aligned}
D_{i,j,k} &:= d_{j,n} I\{k < i \leq k_j\} + d_{j+1,n} I\{k_j < i \leq k + G - h\} - \frac{(k_j - k) d_{j,n} + (k + G - k_j) d_{j+1,n}}{G} \\
&= \frac{(G I\{k < i \leq k_j\} - (k_j - k)) d_{j,n} - (G - (k_j - k) - G I\{k_j < i \leq k + G - h\}) d_{j+1,n}}{G} \\
&= \frac{(G I\{k < i \leq k_j\} - (k_j - k))(d_{j,n} - d_{j+1,n})}{G} \\
&= \frac{(G I\{k < i \leq k_j\} - (k_j - k)) \tilde{d}_{j,n}}{G}
\end{aligned}$$

and receive

$$\begin{aligned}
&\sum_{i=k+1}^{k+G-h} (X_i - \overline{X}_{k-G+1,k})(X_{i+h} - \overline{X}_{k-G+1,k}) \\
&= \sum_{i=k+1}^{k+G-h} \left(\varepsilon_i - \overline{\varepsilon}_{k+1,k+G} + D_{i,j,k}\right) \left(\varepsilon_{i+h} - \overline{\varepsilon}_{k+1,k+G} + D_{i+h,j,k}\right) \\
&= \sum_{i=k+1}^{k+G-h} \left(\varepsilon_i - \overline{\varepsilon}_{k+1,k+G}\right) \left(\varepsilon_{i+h} - \overline{\varepsilon}_{k+1,k+G}\right) + \sum_{i=k+1}^{k+G-h} R(i, k, j, h)
\end{aligned}$$

with

$$R(i,k,j,h) \quad := \quad (\varepsilon_i - \overline{\varepsilon}_{k+1,k+G})\, D_{i+h,j,k} + D_{i,j,k}\,(\varepsilon_{i+h} - \overline{\varepsilon}_{k+1,k+G}) + D_{i,j,k}D_{i+h,j,k}.$$

This yields for $k = k_j - G + 1, \ldots, k_j$ and n large

$$\hat{\tau}^2_{k,n} \quad = \quad \hat{\tau}^2_{k,n,0} + \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G} \sum_{i=k+1}^{k+G-h} R(i,k,j,h).$$

Using Theorem 6.17 again shows

$$\max_{0 \le k_j - k < G} \hat{\tau}^2_{k,n,0} \quad \le \quad \max_{G \le k \le n-G} \hat{\tau}^2_{k,n,0} \quad = \quad O_p(1).$$

For the remaining part

$$\max_{0 \le k_j - k < G} \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G} \sum_{i=k+1}^{k+G-h} R(i,k,j,h),$$

we start to examine

$$\begin{aligned}
&\max_{0 \le k_j - k < G} \left| \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G} \sum_{i=k+1}^{k+G-h} (\varepsilon_i - \overline{\varepsilon}_{k+1,k+G})\, D_{i+h,j,k} \right| \\
&= \quad \max_{0 \le k_j - k < G} \left| \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \frac{1}{G^2} \sum_{i=k+1}^{k+G-h} (\varepsilon_i - \overline{\varepsilon}_{k+1,k+G})\,(GI\{i+h \le k_j\} - (k_j - k))\tilde{d}_{j,n} \right| \\
&\le \quad \frac{1}{G^2} \max_{1 \le j \le q} |\tilde{d}_{j,n}| \sum_{h=1}^{\Lambda_n} \omega(h/\Lambda_n) \max_{0 \le k_j - k < G} \left| \sum_{i=k+1}^{k+G-h} \varepsilon_i (GI\{i+h \le k_j\} - (k_j - k)) \right| \\
&\quad + \Lambda_n \max_{1 \le j \le q} |\tilde{d}_{j,n}| \max_{0 \le k_j - k < G} |\overline{\varepsilon}_{k+1,k+G}| \,.
\end{aligned}$$

We split the set $\{k : 0 \le k_j - k < G\}$ into the disjunct sets $\{k : 0 \le k_j - k < h\}$ and $\{k : h \le k_j - k < G\}$ and obtain

$$\begin{aligned}
&\max_{0 \le k_j - k < h} \left| \sum_{i=k+1}^{k+G-h} \varepsilon_i (GI\{i+h \le k_j\} - (k_j - k)) \right| \quad = \quad \max_{0 \le k_j - k < h} \left| \sum_{i=k+1}^{k+G-h} \varepsilon_i (k_j - k) \right| \\
&\le \quad h \max_{0 \le k_j - k < h} \left| \sum_{i=k+1}^{k+G-h} \varepsilon_i \right| \quad = \quad \Lambda_n O(1) \max_{1 < k \le h} \left| \sum_{i=k+1}^{k+G-h} \varepsilon_i \right| \\
&= \quad \Lambda_n O(1) \left(\max_{1 < k \le h} \left| \sum_{i=1}^{k+G-h} \varepsilon_i \right| + \max_{1 < k \le h} \left| \sum_{i=1}^{k} \varepsilon_i \right| \right) \quad = \quad \Lambda_n O(1) \max_{1 \le k \le G} \left| \sum_{i=1}^{k} \varepsilon_i \right|
\end{aligned}$$

as well as

$$\begin{aligned}
&\max_{h\le k_j-k<G}\left|\sum_{i=k+1}^{k+G-h}\varepsilon_i(GI\{i+h\le k_j\}-(k_j-k))\right| \\
&= \max_{h\le k_j-k<G}\left|\sum_{i=k+1}^{k_j-h}\varepsilon_i(G-(k_j-k))-\sum_{i=k_j-h+1}^{k+G-h}\varepsilon_i(k_j-k)\right| \\
&\le (G-h)\max_{h\le k_j-k<G}\left|\sum_{i=k+1}^{k_j-h}\varepsilon_i\right|+G\max_{h\le k_j-k<G}\left|\sum_{i=k_j-h+1}^{k+G-h}\varepsilon_i\right| \\
&= GO(1)\max_{1\le k\le G}\left|\sum_{i=1}^{k}\varepsilon_i\right|.
\end{aligned}$$

Similarly, we receive

$$\max_{1\le k\le 2G}\left|\sum_{i=1}^{k}\varepsilon_i\right| = O_p\left(\sqrt{G}\right)$$

and

$$\max_{0\le k_j-k<G}|\overline{\varepsilon}_{k+1,k+G}| = G^{-1}O(1)\max_{1\le k\le 2G}\left|\sum_{i=1}^{k}\varepsilon_i\right| = O_p\left(G^{-\frac{1}{2}}\right)$$

by Lemma 6.5. Therefrom, we conclude by using that the weights are bounded by 1

$$\begin{aligned}
&\max_{0\le k_j-k<G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\frac{1}{G}\sum_{i=k+1}^{k+G-h}(\varepsilon_i-\overline{\varepsilon}_{k+1,k+G})\,D_{i+h,j,k}\right| \\
&= O_p(1)\max_{j=1,\dots,q}|\tilde{d}_{j,n}|G^{-\frac{1}{2}}\Lambda_n = o_p(1),
\end{aligned}$$

since by (6.30) and (6.31)

$$G^{-\frac{1}{2}}\Lambda_n = \frac{\Lambda_n}{n^{\Delta}}\frac{n^{\frac{1}{2}+\Delta}}{G}\frac{\sqrt{G}}{\sqrt{n}} \longrightarrow 0.$$

Similarly, we obtain

$$\begin{aligned}
&\max_{0\le k_j-k<G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\frac{1}{G}\sum_{i=k+1}^{k+G-h}D_{i,j,k}\,(\varepsilon_{i+h}-\overline{\varepsilon}_{k+1,k+G})\right| = O_p(1)\max_{j=1,\dots,q}|\tilde{d}_{j,n}|G^{-\frac{1}{2}}\Lambda_n \\
&= o_p(1).
\end{aligned}$$

For the last term, we have

$$
\begin{aligned}
&\max_{0\le k_j-k<G}\left|\sum_{h=1}^{\Lambda_n}\omega(h/\Lambda_n)\frac{1}{G}\sum_{i=k+1}^{k+G-h}D_{i,j,k}D_{i+h,j,k}\right| \\
\le\;& \frac{1}{G^3}\max_{j=1,\dots,q}\tilde{d}^2_{j,n} \\
&\sum_{h=1}^{\Lambda_n}\max_{0\le k_j-k<G}\left|\sum_{i=k+1}^{k+G-h}(GI\{i\le k_j\}-(k_j-k))(GI\{i+h\le k_j\}-(k_j-k))\right| \\
=\;& O_p(1)\max_{j=1,\dots,q}\tilde{d}^2_{j,n}\Lambda_n,
\end{aligned}
$$

since

$$
\begin{aligned}
&\max_{0\le k_j-k<G}\left|\sum_{i=k+1}^{k+G-h}(GI\{i\le k_j\}-(k_j-k))(GI\{i+h\le k_j\}-(k_j-k))\right| \\
\le\;& \max_{0\le k_j-k<G}\left|\sum_{i=k+1}^{k_j-h}(GI\{i\le k_j\}-(k_j-k))(GI\{i+h\le k_j\}-(k_j-k))\right| \\
&+\max_{0\le k_j-k<G}\left|\sum_{i=k_j-h+1}^{k_j}(GI\{i\le k_j\}-(k_j-k))(GI\{i+h\le k_j\}-(k_j-k))\right| \\
&+\max_{0\le k_j-k<G}\left|\sum_{i=k_j+1}^{k+G-h}(GI\{i\le k_j\}-(k_j-k))(GI\{i+h\le k_j\}-(k_j-k))\right| \\
=\;& \max_{0\le k_j-k<G}\sum_{i=k+1}^{k_j-h}(G-(k_j-k))^2 \\
&+\max_{0\le k_j-k<G}\left|\sum_{i=k_j-h+1}^{k_j}(G-(k_j-k))(k_j-k)\right| \\
&+\max_{0\le k_j-k<G}\sum_{i=k_j+1}^{k+G-h}(k_j-k)^2 \\
=\;& O_p(1)(G^3+\Lambda_nG^2+G^3).
\end{aligned}
$$

Similar calculations yield $\max_{k\in C}\hat{\tau}^2_{k,n}=O_p(\Lambda_n)$. This yields the assertion under the classical model. Since we have a uniform rate in equation (6.44), the assertion further holds under the regime-switching model with the same argument as in the proof of Theorem 6.15. □

Remark 6.19. Theorem 6.17 and 6.18 remain true for arbitrary weights $\omega(h/\Lambda_n)$ fulfilling

$$
\begin{aligned}
\omega(h/\Lambda_n)&=0, && h>\Lambda_n,\\
0\le\omega(h/\Lambda_n)&\le 1, && h\le\Lambda_n,
\end{aligned}
$$

and

$$\sum_{h=1}^{\Lambda_n}(\omega(h/\Lambda)-1)\gamma(h)=o_p\left((\log(n/G))^{-1}\right).$$

Proof of Remark 6.19. The results remain true, since we only used the first two properties of the weights, but in (6.41). However, this equation holds by the last assumption. □

As in Corollary 6.16, we obtain that the rates hold for the modified long-run variance estimators as well.

Corollary 6.20. Let the assumptions of Theorem 6.17 hold.

(a) Then, under H_0,

$$\begin{aligned}
\max_{G\le k\le n-G}|\hat{\tau}^2_{k,n,l}-\tau^2| &= o_p\left((\log(n/G))^{-1}\right),\\
\max_{G\le k\le n-G}|\hat{\tau}^2_{k,n,r}-\tau^2| &= o_p\left((\log(n/G))^{-1}\right),\\
\max_{G\le k\le n-G}|\hat{\tau}^2_{k,n,m}-\tau^2| &= o_p\left((\log(n/G))^{-1}\right).
\end{aligned}$$

(b) Assume additionally **(Cd)**, respectively **(Cr)**, on the change points. Then, under H_1,

$$\begin{aligned}
\max_{G\le k\le n-G}\hat{\tau}^2_{k,n,l} &= O_p(\Lambda_n),\\
\max_{G\le k\le n-G}\hat{\tau}^2_{k,n,r} &= O_p(\Lambda_n),\\
\max_{G\le k\le n-G}\hat{\tau}^2_{k,n,m} &= O_p(\Lambda_n).
\end{aligned}$$

Proof of Corollary 6.20. The argument is analogous to the argument in the proof of Corollary 6.16, aside from using Theorem 6.17 and Theorem 6.18 instead of Theorem 6.14 and Theorem 6.15. □

6.3. Examples for the Error Distribution

In Section 6.1 we obtained rates of consistency for the change-point estimators and derived the joint asymptotic distribution. Further, we obtained in Section 6.2 consistency results for the long-run variance estimators. These results were obtained under conditions **(E1)**, **(E2)** and **(E3)** on the error sequence. To see that these assumptions are not too restrictive, we discuss some examples of error sequences, which fulfill these conditions. We introduced condition

(E1) Let the errors $\{\varepsilon_i\}$ be a strictly stationary sequence with

$$E\varepsilon_1=0,\quad 0<\sigma^2=E\varepsilon_1^2<\infty,\quad E\left|\varepsilon_1\right|^{2+\nu}<\infty \text{ for some } \nu>0,$$

$$\sum_{h\ge 0}|\gamma(h)|<\infty,\ \text{ where } \gamma(h)=\operatorname{cov}(\varepsilon_0,\varepsilon_h),$$

and long-run variance

$$\tau^2:=\sigma^2+2\sum_{h>0}\gamma(h)>0.$$

to ensure the validness of the invariance principle

(E2) It exists a standard Wiener process $\{W(k): 1 \le k \le n\}$ such that

$$\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \frac{1}{\tau} \sum_{i=k+1}^{k+G} \varepsilon_i - (W(k+G) - W(k)) \right| = o_p\left((\log(n/G))^{-\frac{1}{2}} \right).$$

and the Hájek-Rényi-type moment condition

(E3) For some $\gamma \ge 1$, $\varphi > 1$ and some constant $C > 0$ it holds

$$E \left| \sum_{k=i}^{j} \varepsilon_k \right|^{\gamma} \le C|j-i+1|^{\varphi}.$$

We show that the assumptions **(E2)** and **(E3)** hold for an i.i.d. sequence and further for various dependent sequences.

Example 6.21. (i.i.d. sequences) Let $\{\varepsilon_i : i \ge 1\}$ be a sequence of i.i.d. random variables with

$$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E\left|\varepsilon_1\right|^{2+\nu} \le C < \infty \text{ for some } \nu > 0,\ C > 0.$$

To apply the invariance principle of Komlós et al. (1975) and Komlós et al. (1976) we have to ensure that the probability space of the errors is rich enough to let a Wiener process $\{W(i) : i \ge 0\}$ exist on it. For example the probability space of $Bin(1,p)$ distributed errors would be too small. Hence, we assume that the probability space is rich enough or we have enlarged it sufficiently. Then we receive

$$\begin{aligned}
&\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=k+1}^{k+G} \varepsilon_i - \sigma(W(k+G) - W(k)) \right| \\
&= \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=1}^{k+G} \varepsilon_i - \sum_{i=1}^{k} \varepsilon_i - \sigma(W(k+G) - W(k)) \right| \\
&\le \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=1}^{k+G} \varepsilon_i - \sigma W(k+G) \right| + \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=1}^{k} \varepsilon_i - \sigma W(k) \right|.
\end{aligned}$$

We consider the first maximum and the analogous results hold for the second term. We obtain

$$\begin{aligned}
&\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=1}^{k+G} \varepsilon_i - \sigma W(k+G) \right| \\
&= O_p(1) G^{-\frac{1}{2}} \max_{2G \le k \le n} \left| \sum_{i=1}^{k} \varepsilon_i - \sigma W(k) \right| \\
&= O_p(1) G^{-\frac{1}{2}} n^{\frac{1}{2+\nu}} \max_{2G \le k \le n} k^{-\frac{1}{2+\nu}} \left| \sum_{i=1}^{k} \varepsilon_i - \sigma W(k) \right|
\end{aligned}$$

and receive by the invariance principle of Komlós et al. (1975) and Komlós et al. (1976) that a Wiener process $\{W(i) : i \geq 0\}$ exists with

$$n^{-\frac{1}{2+\nu}} \left| \sum_{i=1}^{n} \varepsilon_i - \sigma W(n) \right| = o(1) \qquad \text{a.s.}$$

and hence

$$\max_{2G \leq k \leq n} k^{-\frac{1}{2+\nu}} \left| \sum_{i=1}^{k} \varepsilon_i - \sigma W(k) \right| = O(1).$$

This implies with bandwidth assumption **(G)** the validness of **(E2)**:

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=1}^{k+G} \varepsilon_i - \sigma W(k+G) \right| = O_p\left(n^{\frac{1}{2+\nu}} G^{-\frac{1}{2}}\right) = o_p\left((\log(n/G))^{-\frac{1}{2}}\right).$$

To receive **(E3)** we can apply Theorem 3.7.8 in Stout (1974) (refer to Theorem B.5 in the Appendix) with $\gamma = 2 + \nu$ and $\varphi = \gamma/2$. Then we have $E \left| \sum_{i=1}^{n} \varepsilon_i \right|^{2+\nu} \leq D n^{\frac{2+\nu}{2}}$, where D only depends on C and ν.

Example 6.22. (strongly mixing sequences) Let $\{\varepsilon_i : i \geq 1\}$ be a strictly stationary sequence with

$$E\varepsilon_1 = 0 \qquad \text{and} \qquad E\left|\varepsilon_1\right|^{2+\delta+\Delta} < \infty \text{ for some } \delta > 0,\ \Delta > 0,$$

and fulfill a strong mixing condition with mixing coefficient α_n, i.e. it exists a nonincreasing sequence $\{\alpha_n : n \geq 1\}$ with $\alpha_n \to 0$ such that

$$|P(AB) - P(A)P(B)| \leq \alpha_n$$

for all $n,\ k \geq 1$, all $A \in \mathcal{M}_1^k$ and $B \in \mathcal{M}_{k+n}^{\infty}$, where $\mathcal{M}_j^l$ is the σ-Algebra generated by $\{\varepsilon_i :\ j \leq i \leq l\}$. Under the assumption of

$$\alpha_n = O\left(n^{-\beta}\right)$$

with some $\beta > \max(3, (2+\delta+\Delta)/(\delta+\Delta))$, we receive from Theorem B.2 that on a possibly larger probability space a Wiener process $\{W(i) : i \geq 0\}$ exists such that

$$\left| \sum_{i=1}^{n} \varepsilon_i - \tau W(n) \right| = O_p\left(n^{\frac{1}{2}-\theta}\right)$$

with some $\theta > 0$. Analogously to the calculations in Example 6.21 we receive from this invariance principle

$$\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=1}^{k+G} \varepsilon_i - \sigma W(k+G) \right| = O_p\left(n^{\frac{1}{2}-\theta} G^{-\frac{1}{2}} \right). \tag{6.45}$$

With a slightly different bandwidth assumption than **(G)**, namely let

$$\frac{n}{G} \longrightarrow \infty \qquad \text{and} \qquad \frac{n^{1-2\theta} \log n}{G} \longrightarrow 0, \tag{6.46}$$

we receive

$$\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left| \sum_{i=1}^{k+G} \varepsilon_i - \sigma W(k+G) \right| = o_p\left((\log(n/G))^{-\frac{1}{2}} \right).$$

If a sequence $\{\tilde{\alpha}_n\}$ with $\alpha_n \le \tilde{\alpha}_n$, $n \in \mathbb{N}$, exists, such that

$$\sum_{i=0}^{\infty} (i+1)^{\frac{\delta+\Delta}{2}-1} \tilde{\alpha}_i^{\frac{\delta+\Delta}{2+\delta+\Delta}} < \infty, \tag{6.47}$$

follows with Theorem 1 in Yokoyama (1980) assumption **(E3)**.

Example 6.23. (linear process) Let $\{\varepsilon_i : i \ge 1\}$ be a linear process

$$\varepsilon_i = \sum_{s \ge 0} \omega_s e_{i-s}, \qquad i \ge 1,$$

where $\{e_i : i \in \mathbb{Z}\}$ is an i.i.d. sequence with

$$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E\left|\varepsilon_1\right|^{2+\nu} < \infty \text{ for some } \nu > 0$$

and a smooth density function f satisfying

$$\sup_{-\infty < s < \infty} \frac{1}{s} \int_{-\infty}^{\infty} |f(t+s) - f(t)| \, dt < \infty.$$

Let the weights $\{\omega_s : s \ge 1\}$ satisfy

$$\omega_s = O(s^{-\psi}) \quad \text{as } s \to \infty \quad \text{for some } \psi > 3/2.$$

Furthermore let

$$g(z) = \sum_{s \ge 0} \omega_s z^s, \; z \in \mathbb{C},$$

and assume

$$g(z) \neq 0 \quad \text{for all } |z| \leq 1.$$

In the proofs of Lemma 2.1 and Lemma 2.2 in Horváth (1997) was pointed out that under these assumptions a linear process is strongly mixing with rate $O\left(n^{-\beta}\right)$, $\beta > 0$, and fulfills the assumptions of Theorem 4 in Kuelbs and Philipp (1980) such that (6.45) holds. If further (6.47) holds, **(E3)** follows.

Example 6.24. (m-dependent sequences) Let $\{\varepsilon_i : i \geq 1\}$ be a strictly stationary sequence of random variables with

$$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E\left|\varepsilon_1\right|^{2+\nu} < \infty \text{ for some } \nu > 0.$$

Moreover, we let $\{\varepsilon_i\}$ be m-dependent, i.e. ε_i and ε_j are independent whenever $|i-j| > m$. Since m-dependent sequences are α-mixing with any rate $\beta > 0$, the results of Example 6.22 hold.

Example 6.25. (near-epoch dependent (NED) sequences) NED is a concept which goes back to Billingsley (1968) and was further developed by McLeish (1975). Ling (2007) established a strong invariance principle, which we use to show the validness of the assumption **(E2)**. The idea of NED is to consider the error sequence $\{\varepsilon_i : -\infty < i < \infty\}$ and relate it to an additional process $\{\xi_i : -\infty < i < \infty\}$. Then the dynamic dependence structure of the error sequence $\{\varepsilon_i\}$ is described with respect to $\{\xi_i\}$. As the name of NED indicates the dependence structure of ε_i should essentially be described by $\xi_i, \ldots, \xi_{i-m}$ that are close in time.

We assume that $\{\xi_i\}$ is an i.i.d sequence and ε_i is $\mathcal{F}_{-\infty}^i$-measurable, where $\mathcal{F}_k^l$ denotes the σ-algebra generated by $\xi_k, \ldots, \xi_l$. Then a process $\{\varepsilon_i\}$ is called $L_p(\lambda)$-near-epoch-dependent on $\{\xi_i\}$ if

$$\sup_{-\infty<i<\infty} E|\varepsilon_i|^p < \infty \qquad \text{and} \qquad \sup_{-\infty<i<\infty} \left(E\left|\varepsilon_i - E[\varepsilon_i|\mathcal{F}_{i-k}^i]\right|^p\right)^{\frac{1}{p}} = O(k^{-\lambda}) \;\; (k \to \infty),$$

where $p \geq 1$ and $\lambda > 0$. To apply the invariance principle from Ling (2007), we choose $p = 2 + \nu$ with $\nu > 0$ and $\lambda > 2$. Moreover, we let $\{\varepsilon_i\}$ be a martingale difference in terms of $\mathcal{F}_i$ with long-run variance τ^2. Then there exists, on a probably enlarged probability space, a Wiener process $\{W_i : i \geq 0\}$ and a constant $\theta > 0$, such that

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left|\sum_{i=1}^{k+G} \varepsilon_i - \tau W(k+G)\right| \;=\; O_p\left(n^{\frac{1}{2}-\theta} G^{-\frac{1}{2}}\right) \;=\; o_p\left((\log(n/G))^{-\frac{1}{2}}\right)$$

with bandwidth assumption (6.46). To confirm assumption **(E3)** we again apply Theorem 3.7.8 in Stout (1974), which states that the moment condition holds for martingale differences. Since we have the stronger assumptions that $\{\varepsilon_i\}$ is a martingale difference in terms of $\mathcal{F}_{-\infty}^i$ assumption **(E3)** holds.

7. Simulation Study

This section analyses the performance of the change-point estimators for smaller sample sizes. We discuss how the variance estimator as well as the bandwidth influence the performance. Due to the in general poor behaviour of the long-run variance estimators, we concentrate on mean changes in an otherwise i.i.d. random sequence. Since in many applications models are used which allow for simultaneous changes in the mean and in the variance, we further discuss the performance of the MOSUM procedure in this context and illustrate how the refined variance estimators introduced in Remark 5.2 can improve the performance.

7.1. Change Analysis of the Location Model with I.I.D. Errors

To get an overview of the performance of the MOSUM procedure, we have to consider several different change-point constellations, i.e. we have to vary in the amount of change and in the distance between change points. Since in the regime-switching model the waiting time until particular change-point constellations occur can be very long, we use the classical model with the deterministic change points as generating system.

We consider the situation of independent and identically standard normally distributed random errors and simulate a random sample $X_1, ..., X_n$ of size $n = 500$. The test statistic $T_{k,n}(G)$ is calculated for four different bandwidths $G = 25, 40, 50, 75$. Further, we set $\varepsilon = 0.15$ (refer to (5.7)), choose the level $\alpha = 0.05$ and compute the critical value $D_n(G; \alpha)$ along (5.4).

The results of this investigation are displayed in Figure 7.1 and 7.2. In each figure are three blocks consisting of five pictures. In each block the topmost picture shows the same time series with its change points marked by red vertical lines. Below, in each column the performance of $\sigma^{-1}T_{k,n}(G)$, $\hat{\sigma}_n^{-1}T_{k,n}(G)$ and $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ is displayed with bandwidths $G = 25,\ 40,\ 50,\ 75$. The critical value $D_n(G; \alpha)$ is the blue horizontal line.

To begin with, we consider Figure 7.1 and have a look at the first two columns, where the performances of $\sigma^{-1}T_{k,n}(G)$ and $\hat{\sigma}_n^{-1}T_{k,n}(G)$ are presented. The performance of $\sigma^{-1}T_{k,n}(G)$ is good, whereas $\hat{\sigma}_n^{-1}T_{k,n}(G)$ performs quite poorly, since in case of the smaller bandwidths only one change point is estimated and in case of the biggest bandwidth at least two. However, we also note that the peaks of the statistic $\hat{\sigma}_n^{-1}T_{k,n}(G)$ are at the right place, but they are just too small to detect the change. This is due to the overestimation of the variance (refer to Section 5.3.3). The behaviour of the statistic $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ is displayed in the third column, where the use of variance estimator $\hat{\sigma}_{k,n}^2$ clearly improves the performance, since for the smallest bandwidths at least two change points are detected and for the bigger bandwidths all three change points are estimated almost correctly. If we compare the performance of the statistic $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ to the performance of the statistic $\sigma^{-1}T_{k,n}(G)$, we discover that they yield the same results. However, the performance of $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ is even slightly better, since the peaks are tighter. This is due to the overestimation of the variance near the change point.

Further, we observe that in general the bigger changes are detected better and larger bandwidths are needed to detect smaller changes. However, we will see in the example of Figure 7.2 that a bigger bandwidth G does not always achieve a better performance. In Figure 7.2 the MOSUM procedure is analysed in the situation of smaller distances between the change points. We have $(k_1,\ k_2,\ k_3) = (200,\ 250,\ 320)$.

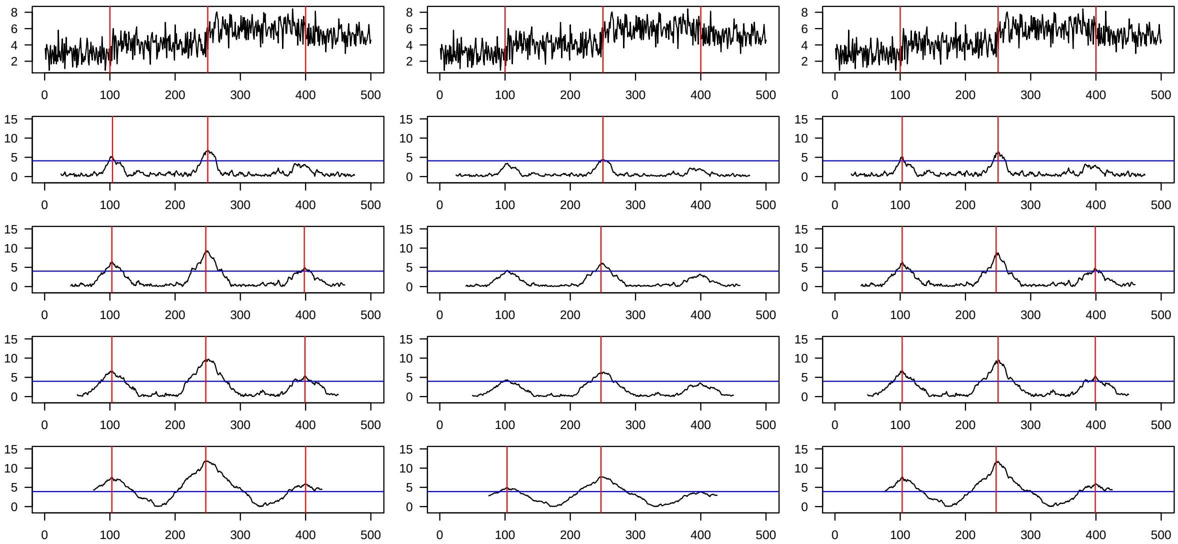

Figure 7.1.: Sample path and performance of $\sigma^{-1}T_{k,n}(G)$ (first column), $\hat{\sigma}_n^{-1}T_{k,n}(G)$ (second column) and $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ (third column) with bandwidths $G = 25,\ 40,\ 50,\ 75$ for independent normal distributed errors with variance 1.

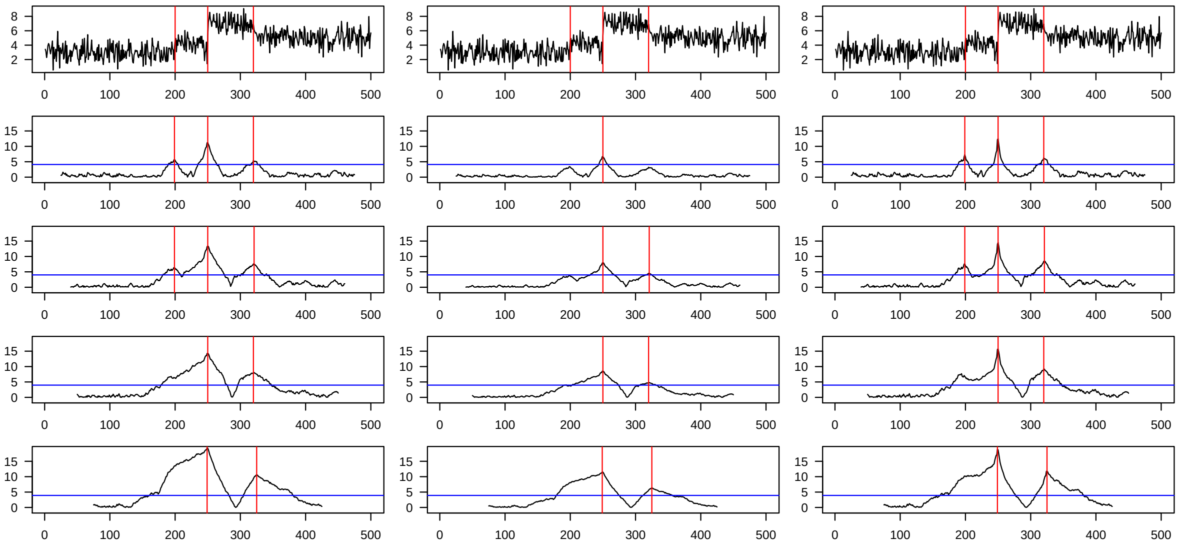

Figure 7.2.: Sample path and performance of $\sigma^{-1}T_{k,n}(G)$ (first column), $\hat{\sigma}_n^{-1}T_{k,n}(G)$ (second column) and $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ (third column) with bandwidths $G = 25,\ 40,\ 50,\ 75$ for independent normal distributed errors with variance 1.

The test statistics with the smaller bandwidths $G = 25,\ 50$ have three clear peaks, which indicate the existence of three change points. Whereas with a growing bandwidth, the first two peaks merge to one big peak. This behaviour is called oversmoothing and appears if the moving window includes more than one change point.

In Figure 7.3 we consider the situation of three change points at $(k_1,\ k_2,\ k_3) = (200,\ 250,\ 320)$ and independent t-distributed errors with three degrees of freedom. In general the performance of the MOSUM procedure will be better for normal distributed errors, since Hušková and Slabý (2001) derived the distribution of the test statistic for normal distributed observations. In this example the change points are estimated correctly under the use of variance estimator $\hat{\sigma}^2_{k,n}$ with an appropriate bandwidth, i.e. $G = 40$. Further, we mention that the choice of the parameter ε influences the performance. For bandwidth $G = 25$ and statistic $\hat{\sigma}^{-1}_{k,n} T_{k,n}(G)$ the correct number of change points would have been estimated if ε has been smaller, but for bandwidth $G = 40$ and statistic $\hat{\sigma}^{-1}_{n} T_{k,n}(G)$ a bigger ε would have improved the performance.

In general, we recommend to apply the MOSUM procedure with different bandwidths to get an impression of an adequate bandwidth. The bandwidth should be chosen as large as possible, under the restriction that the resulting window does not include more than one change point.

7.2. Change Analysis of the Location Model with Variance Varying Errors

In practise a mean change is often accompanied by a change in variance. Therefore, we investigate in this section the performance of the MOSUM procedure for simultaneous changes in mean and variance and discuss how and why the in Section 5.3.3 modified variance estimators $\hat{\sigma}^2_{k,n,l}$, $\hat{\sigma}^2_{k,n,r}$ as well as $\hat{\sigma}^2_{k,n,m}$ alter, respectively improve, the performance.

Figure 7.4 includes in the upper row three different time series with a change in the mean from two to four. Whereas the first time series has a simultaneous change in the variance from one to five, the second one has a change in the variance from five to one and the last one has no change. In the row below the performance of the test statistic $\hat{\sigma}^{-1}_{k,n} T_{k,n}(G)$ is presented. We find that in presence of a variance change the change point is not detected. Rows three and four include the behaviour of $\hat{\sigma}^{-1}_{k,n,l} T_{k,n}(G)$ and $\hat{\sigma}^{-1}_{k,n,r} T_{k,n}(G)$, where the first one detects the mean change in the first case and the second one in the second case. The last row illustrates the performance of $\hat{\sigma}^{-1}_{k,n,m} T_{k,n}(G)$, which detects the change point in all cases.

To explain these results Figure 7.5 has been included and presents the performance of the four different variance estimators in the three different situations. In the second row the estimator $\hat{\sigma}^2_{k,n}$ does not estimate the variance correctly near the change point and also not directly at the change point in presence of a change in the variance. At the change point we receive the mean of the two different variances as an estimate. The second variance estimator $\hat{\sigma}^2_{k,n,l}$, which uses only the samples on the left side of the window, estimates the variance correctly before the change point and underestimates it close after the change point if the variance gets larger, and overestimates it close after the change point if the variance gets smaller. We observe the opposite behaviour for the variance estimator $\hat{\sigma}^2_{k,n,r}$, which uses only the samples on the right side of the window. Since both estimators work only in one of the two situations well, we prefer the minimum of them, which is given by variance estimator $\hat{\sigma}^2_{k,n,m}$. This estimator gives a good performance in both cases, as well as in the case of no variance change. In the last case the variance is estimated correctly for all time points, though here the use of the estimator $\hat{\sigma}^2_{k,n}$ is preferable, since it is more stable and makes the peaks of the test statistic tighter.

Finally, we discuss the simulations in Figure 7.6 to confirm the good performance of $\hat{\sigma}^2_{k,n,m}$ in the test and estimation procedure. We consider three different situations with three change points, respectively. We begin with expectations $(d_1, d_2, d_3, d_4) = (1, 3, 5, 3)$ and variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (3, 1, 3, 1)$, then we alter the variances to $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (5, 1, 5, 1)$ and in the last case we also alter the expectations to $(d_1, d_2, d_3, d_4) = (1, 3, 4, 3)$.
In the first case, where the variance changes are small, the original test statistic $\hat{\sigma}^{-1}_{k,n} T_{k,n}(G)$ obtains the three change points, but only by a small margin compared to $\hat{\sigma}^{-1}_{k,n,m} T_{k,n}(G)$. However, in case of bigger changes in the variance the estimation procedure with $\hat{\sigma}^2_{k,n}$ fails, whereas the procedure with $\hat{\sigma}^2_{k,n,m}$ still yields good results. Even if the changes in the means are smaller, as it is in the third case, all change points are detected.

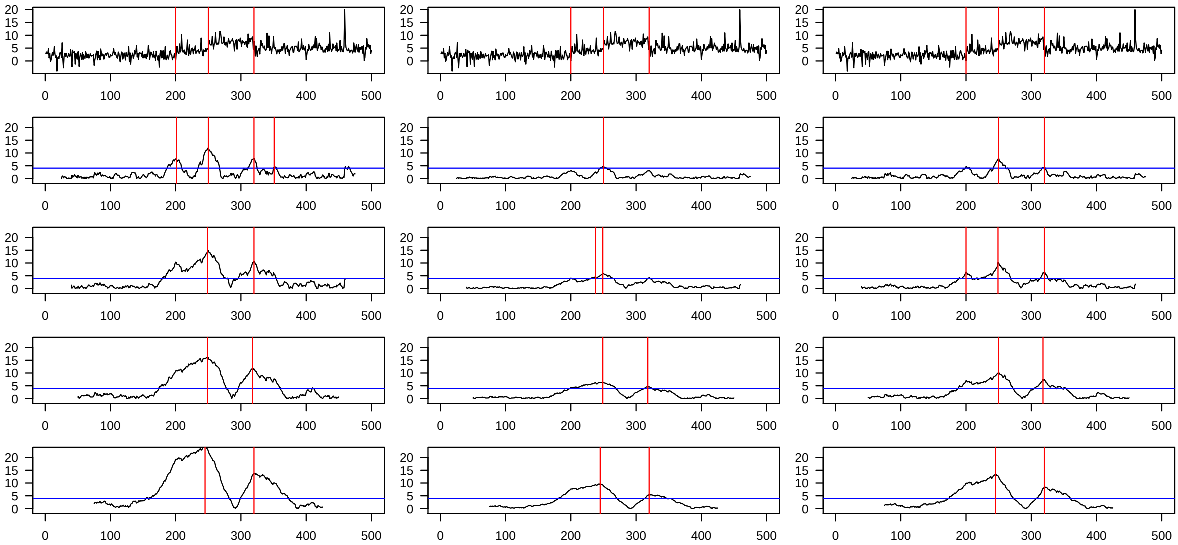

Figure 7.3.: Sample path and performance of $\sigma^{-1}T_{k,n}(G)$ (first column), $\hat{\sigma}_n^{-1}T_{k,n}(G)$ (second column) and $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ (third column) with bandwidths $G = 25$, 40, 50, 75 for independent t-distributed errors with 3 degrees of freedom.

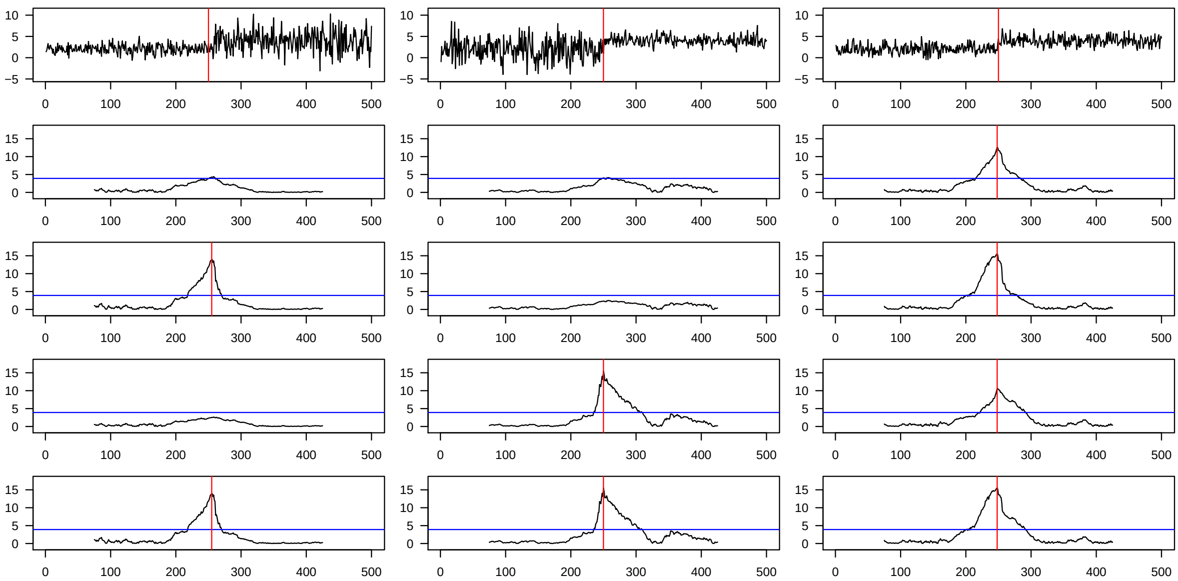

Figure 7.4.: Sample paths and performances of $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ (second row), $\hat{\sigma}_{k,n,l}^{-1}T_{k,n}(G)$ (third row), $\hat{\sigma}_{k,n,r}^{-1}T_{k,n}(G)$ (fourth row) as well as $\hat{\sigma}_{k,n,m}^{-1}T_{k,n}(G)$ (fifth row).

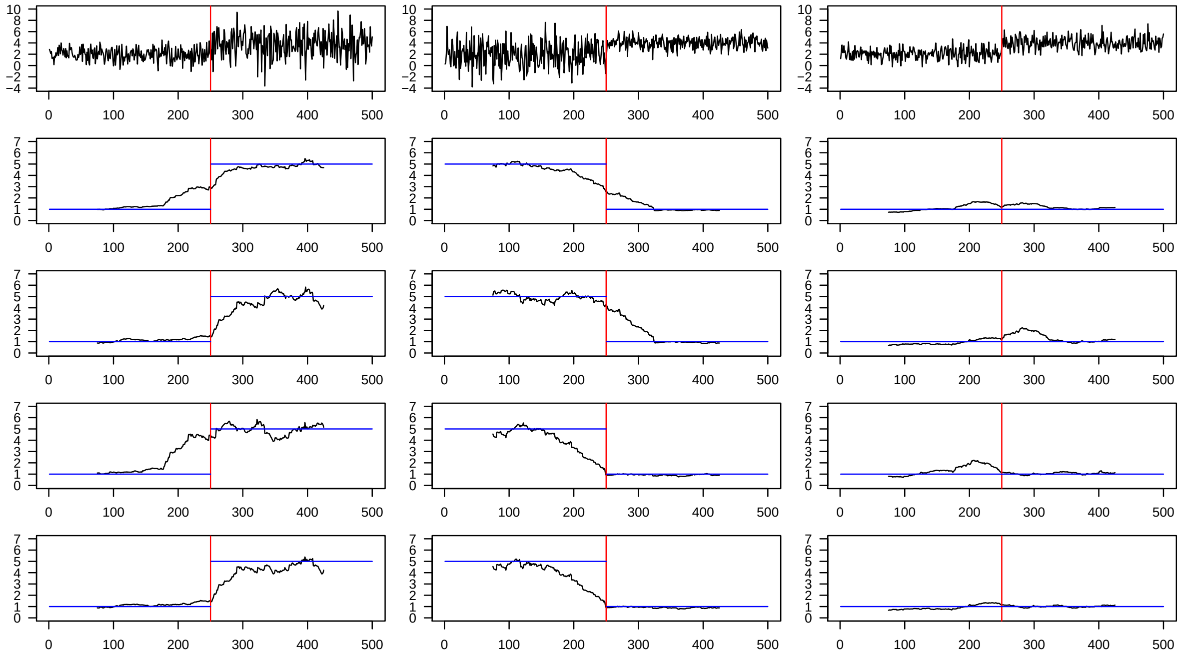

Figure 7.5.: Sample paths and performances of $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ (second row), $\hat{\sigma}_{k,n,l}^{-1}T_{k,n}(G)$ (third row), $\hat{\sigma}_{k,n,r}^{-1}T_{k,n}(G)$ (fourth row) as well as $\hat{\sigma}_{k,n,m}^{-1}T_{k,n}(G)$ (fifth row).

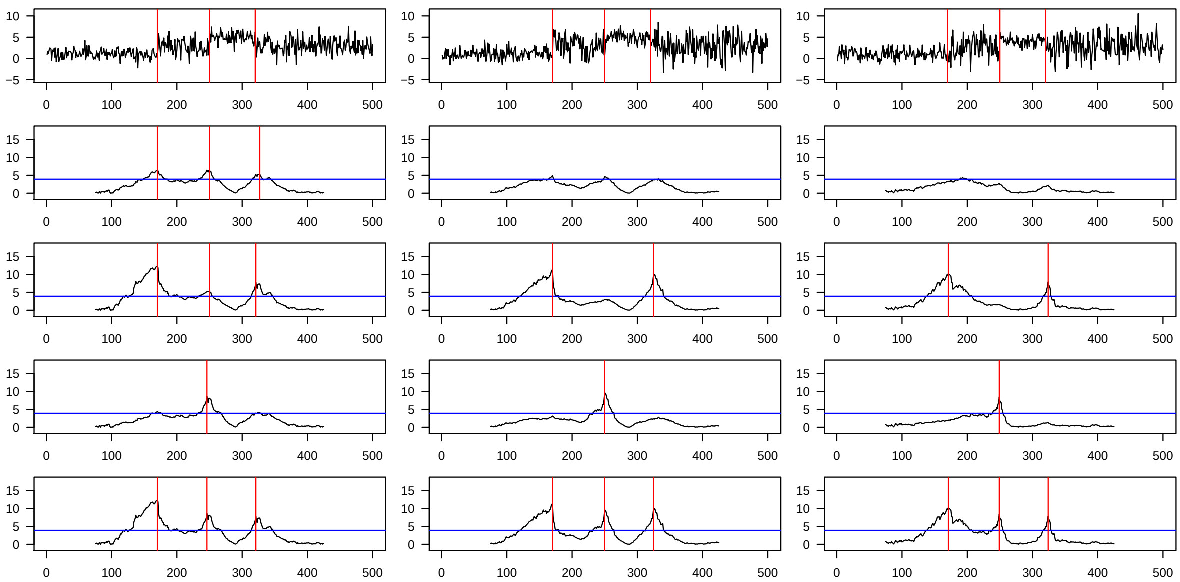

Figure 7.6.: Sample paths and performances of $\hat{\sigma}_{k,n}^{-1}T_{k,n}(G)$ (second row), $\hat{\sigma}_{k,n,l}^{-1}T_{k,n}(G)$ (third row), $\hat{\sigma}_{k,n,r}^{-1}T_{k,n}(G)$ (fourth row) as well as $\hat{\sigma}_{k,n,m}^{-1}T_{k,n}(G)$ (fifth row).

Part III.

Appendix

A. Assumptions

(Y) Let $\{\mathbf{Y}_i\}$ be a $\mathbb{R}^d$-valued strictly stationary sequence of random vectors with

$$E[\mathbf{Y}_1] = 0, \qquad E\|\mathbf{Y}_1\|^{4+\nu} < \infty \text{ for some } \nu > 0,$$

satisfying a strong mixing condition with mixing rate

$$\alpha(n) = O\left(n^{-\beta}\right) \qquad \text{for some } \beta > \max(3, (4+\nu)/\nu).$$

(W) Let the weight function $w : [0,1] \to \mathbb{R}$ be a left continuous function with existing right limits and a finite number of disconuities $a_1, \ldots, a_K$ and fulfill the regularity conditions

$$\lim_{t\to 0} t^{\alpha} w(t) < \infty, \quad \lim_{t\to 1} (1-t)^{\alpha} w(t) < \infty, \qquad \text{for some } 0 \leqslant \alpha < 1/2,$$
$$\sup_{\eta \leqslant t \leqslant 1-\eta} w(t) < \infty, \qquad \text{for all } 0 < \eta \leqslant \frac{1}{2}.$$

(E1) Let the errors $\{\varepsilon_i : 1 \leq i \leq n\}$ be a strictly stationary sequence with

$$E\varepsilon_1 = 0, \quad 0 < \sigma^2 = E\varepsilon_1^2 < \infty, \quad E\left|\varepsilon_1\right|^{2+\nu} < \infty \text{ for some } \nu > 0,$$
$$\sum_{h\geq 0} |\gamma(h)| < \infty, \text{ where } \gamma(h) = \operatorname{cov}(\varepsilon_1, \varepsilon_{1+h}),$$

and long-run variance

$$\tau^2 := \sigma^2 + 2\sum_{h>0} \gamma(h) > 0.$$

(E2) There exists a standard Wiener process $\{W(k) : \ 1 \leq k \leq n\}$ such that

$$\max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \left| \frac{1}{\tau} \sum_{i=k+1}^{k+G} \varepsilon_i - (W(k+G) - W(k)) \right| = o_p\left((\log(n/G))^{-\frac{1}{2}} \right).$$

(E3) For some $\gamma \geq 1$, $\varphi > 1$ and some constant $C > 0$ it holds

$$E\left|\sum_{k=i}^{j} \varepsilon_k\right|^{\gamma} \leq C|j-i+1|^{\varphi}.$$

(G) For $n \to \infty$ let

$$\frac{n}{G} \longrightarrow \infty \qquad \text{and} \qquad \frac{n^{\frac{2}{2+\nu}} \log n}{G} \longrightarrow 0.$$

(L0) Let the long-run variance estimators $\hat{\tau}_{k,n}^2$ be based only on observations $X_{k-G+1}, \ldots, X_{k+G}$ and fulfill

$$\max_{G\leq k\leq n-G} |\hat{\tau}_{k,n}^2 - \tau^2| \quad = \quad o_p\left((\log(n/G))^{-1} \right) \text{ under } H_0.$$

(L1) Let the long-run variance estimator $\hat{\tau}^2_{k,n}$ be translation invariant, i.e.

$$\hat{\tau}_{k,n}(X_{k-G}, \ldots, X_{k+G}) = \hat{\tau}_{k,n}(X_{k-G} + c, \ldots, X_{k+G} + c), \qquad \forall c \in \mathbb{R},$$

and fulfill

$$\max_{G \le k \le n-G} \hat{\tau}^2_{k,n} \;=\; O_p(1) \text{ under the alternative}$$

as well as

$$P\left(\min_{G \le k \le n-G} \hat{\tau}_{k,n} > 0\right) \;\longrightarrow\; 1.$$

(A) Let $\{\alpha_n : n \in \mathbb{N}\}$ satisfy

$$\alpha_n \longrightarrow 0 \qquad \text{and} \qquad \frac{\log\log \frac{1}{\sqrt{1-\alpha_n}}}{a(n/G)} = O(1).$$

(Cd) Let for the minimum distance between two adjacent change points hold that

$$\limsup_{n\to\infty} d_0(n)/G = C > 2 \qquad \text{with} \qquad d_0(n) := \min_{0 \le j \le q} |k_{j+1} - k_j|.$$

(Cr) Let for the minimum distance between two adjacent change points hold that

$$\lim_{n\to\infty} P(d_0(n) > 2G) = 1 \qquad \text{with} \qquad d_0(n) := \min_{0 \le j \le q_n} |k_{j+1} - k_j|.$$

(D) Let $\tilde{d}_{j,n} := d_{j+1,n} - d_{j,n}$, $j = 1, \ldots, q$, fulfill

$$\min_{1 \le j \le q} \tilde{d}^2_{j,n} G(\log(n/G))^{-1} \longrightarrow \infty.$$

(Dr) Let $\tilde{d}^*_{i,j,n} := d^*_{j,n} - d^*_{i,n}$, $1 \le i < j \le n$, fulfill

$$\min_{1 \le i < j \le K} \tilde{d}^{*2}_{i,j,n} G(\log(n/G))^{-1} \longrightarrow \infty.$$

(Q) Let $\{\gamma_n : n \in \mathbb{N}\}$ be a sequence such that

$$P(q_n > \gamma_n) \longrightarrow 0.$$

B. Theorems from Probability Theory

Theorem B.1. (Uniform law of large numbers, Theorem 6.5. in Rao (1962)) Let $\|\cdot\|$ be any norm on $\mathbb{R}^d$ and $v_i(\mathbf{a})$ be a stationary ergodic random sequence with values in $\mathbb{C}(K, \mathbb{R}^d)$ satisfying

$$E \sup_{\mathbf{a}\in K} \|v_1(\mathbf{a})\| < \infty,$$

then

$$\sup_{\mathbf{a}\in K} \left\| \frac{1}{n} \sum_{i=1}^{n} v_i(\mathbf{a}) - E v_1(\mathbf{a}) \right\| \longrightarrow 0 \qquad \text{as} \qquad n \longrightarrow \infty.$$

Theorem B.2. (Invariance principle of Kuelbs and Philipp (1980)) Let $\{\boldsymbol{\xi}_n : n \geq 1\}$ be a weak sense stationary sequence of $\mathbb{R}^d$-valued random vectors, centered at expectations and having $(2+\delta)$th moments with $0 < \delta \leq 1$, uniformly bounded by 1. Suppose that $\{\boldsymbol{\xi}_n : n \geq 1\}$ is α-mixing with

$$\alpha(n) = O\left(n^{-(1+\epsilon)(1+2/\delta)}\right), \qquad \epsilon > 0.$$

Write $\boldsymbol{\xi}_n = (\xi_{n1}, \ldots, \xi_{nd})$. Then, the two series in

$$\gamma_{i,j} = E\xi_{1i}\xi_{1j} + \sum_{k\geq 2} E\xi_{1i}\xi_{kj} + \sum_{k\geq 2} E\xi_{ki}\xi_{1j}$$

converge absolutely. Denote the matrix $(\gamma_{ij})_{(1\leq i,j\leq d)}$ by $\boldsymbol{\Gamma}$. Then, we can define the sequence $\{\boldsymbol{\xi}_n, n \geq 1\}$ on a new probability space together with a Brownian motion $\mathbf{X}_t$ with covariance matrix $\boldsymbol{\Gamma}$ such that

$$\sum_{n\leq t} \boldsymbol{\xi}_n - \mathbf{X}_t \quad = \quad O\left(t^{\frac{1}{2}-\lambda}\right) \qquad \text{a.s.}$$

for some $\lambda > 0$ depending on ϵ, δ and d only.

Theorem B.3. (Theorem B.3 in Kirch (2006)) Let $\{Y_i : i \in \mathbb{N}\}$ be a sequence of random variables satisfying

$$E \left| \sum_{k=i}^{j} Y_k \right|^{\gamma} \leq C |j - i + 1|^{\varphi}$$

for some $\gamma \geq 1$, $\varphi > 1$ and some constant $C > 0$. Then, for any positive and non-increasing sequence $b_1 \geq b_2 \geq \ldots, \geq b_n > 0$, there exists a constant $A(\varphi, \gamma) \geq 4$ (only depending on φ and γ) with

$$E \left(\max_{1\leq k\leq n} c_k \left| \sum_{j=1}^{k} Y_j \right| \right)^{\gamma} \quad \leq \quad C A(\varphi, \gamma) \sum_{k=1}^{n} c_k^{\gamma} k^{\varphi - 1}, \quad \text{where } C \text{ is as above.}$$

Theorem B.4. Let $K \subset \mathbb{R}^p$, $p \in \mathbb{N}$, be a compact set, $f : K \to \mathbb{R}$ a continous function and $\mathbf{x}_0$ a unique maximizer of f, i.e. $\mathbf{x}_0 = \arg\max_{\mathbf{x}\in K} f(\mathbf{x})$. Furthermore, let f_n be a sequence of stochastic functions with $\max_{\mathbf{x}\in K} |f_n(\mathbf{x}) - f(\mathbf{x})| \xrightarrow{P} 0$ and $\widehat{\mathbf{x}}_n = \arg\max_{\mathbf{x}\in K} f_n(\mathbf{x})$. Then, $\widehat{\mathbf{x}}_n \xrightarrow{P} \mathbf{x}_0$.

Proof of Theorem B.4. We prove the assertion in a deterministic setting, since then the assertion follows in the almost sure sense and hence stochastically, too. We suppose that $\widehat{\mathbf{x}}_n$ does not converge to $\mathbf{x}_0$. Since K is compact, there exists a subsequence $\widehat{\mathbf{x}}_{m_n}$ with $\widehat{\mathbf{x}}_{m_n} \xrightarrow{P} \mathbf{x}_1 \neq \mathbf{x}_0$. Then, we receive

$$\begin{aligned}
|f_n(\widehat{\mathbf{x}}_{m_n}) - f(x_1)| \quad &\leq \quad |f_n(\widehat{\mathbf{x}}_{m_n}) - f(\widehat{\mathbf{x}}_{m_n})| + |f(\mathbf{x}_1) - f(\widehat{\mathbf{x}}_{m_n})| \\
\leq \quad \max_{\mathbf{x}\in K} |f_n(\mathbf{x}) - f(\mathbf{x})| + |f(\mathbf{x}_1) - f(\widehat{\mathbf{x}}_{m_n})| \quad &= \quad o(1),
\end{aligned}$$

since the first term converges to 0 by assumption and the second one by the continuity of f. Further, we have

$$|f_n(\widehat{\mathbf{x}}_n) - f(\mathbf{x}_0)| \quad = \quad \left| \max_{\mathbf{x}\in K} f_n(\mathbf{x}) - \max_{\mathbf{x}\in K} f(\mathbf{x}) \right| \leq \max_{\mathbf{x}\in K} |f_n(\mathbf{x}) - f(\mathbf{x})| = o(1).$$

Since $\mathbf{x}_0$ is the unique maximizer of f it holds $f(\mathbf{x}_0) > f(\mathbf{x}_1)$, implying that the subsequence and the original sequence have a different limit behaviour. This is a contradiction. $\square$

Theorem B.5. (Theorem 3.7.8 in Stout (1974)) Let $\{X_i : i \geq 1\}$ satisfy $E\left[|X_i|^\gamma\right] \leq C$ for all $i \geq 1$, some $\gamma > 2$, and some $C < \infty$. Then,

$$E\left[\left|\sum_{i=a+1}^{a+n} X_i\right|^\gamma\right] \quad \leq \quad K n^\varphi$$

for all $a \geq 0$, all $n \geq 1$, and some $K < \infty$ holds if either

(i) $\{X_i : i \geq 1\}$ is a martingale difference sequence;

(ii) $\{X_i : i \geq 1\}$ is stationary and ϕ-mixing with $E[X_1] = 0$ and

$$E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \quad \leq \quad Kn$$

for some $K < \infty$ and all $n \geq 1$.

Theorem B.6. (Marcinkiewicz-Zygmund strong law of large numbers, Theorem 4.23 in Kallenberg (2002)) Let $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables, and fix any $p \in (0, 2)$. Then, $n^{-1/p} \sum_{k=1}^{n} \varepsilon_k$ converges a.s. if $E\left[|\varepsilon|^p\right] < \infty$ and either $p \leq 1$ or $E\varepsilon = 0$. In that case the limit equals $E\varepsilon$ for $p = 1$ and is otherwise 0.

Bibliography

Al Ibrahim, A., Ahmed, M., and BuHamra, S. (2003). Testing for multiple change-points in an autoregressive model using SIC criterion. *Focus On Applied Statistics*, pages 37–51.

Andrews, D. W., Lee, I., and Ploberger, W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, 70(1):9–38.

Antoch, J., Hušková, M., and Jarušková, D. (2000). Change point detection. *5th ERS IASC Summer School*.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, 13(03):315–352.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, 91(2):299–323.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Bauer, P. and Hack, P. (1980). An extension of the MOSUM technique for quality control. *Technometrics*, 22(1):1–7.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(2):289–300.

Berkes, I., Horváth, L., Kokoszka, P., and Shao, Q.-m. (2005). Almost sure convergence of the bartlett estimator. *Periodica Mathematica Hungarica*, 51(1):11–25.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley.

Böhm, H., Ombao, H., von Sachs, R., and Sanes, J. (2010). Classification of multivariate non-stationary signals: The slex-shrinkage approach. *Journal of Statistical Planning and Inference*, 140(12):3754–3763.

Braun, J. V., Braun, R., and Müller, H.-G. (2000). Multiple changepoint fitting via quasi-likelihood, with application to DNA sequence segmentation. *Biometrika*, 87(2):301–314.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer.

Chen, J., Gupta, A., and Pan, J. (2006). Information criterion and change point problem for regular models. *Sankhyā: The Indian Journal of Statistics*, 68(2):252–282.

Chu, C.-S., Hornik, K., and Kaun, C.-M. (1995). MOSUM tests for parameter constancy. *Biometrika*, 82(3):603–617.

Csörgő, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics.* Wiley.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis.* Wiley.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37.

Davis, R. A., Huang, D., and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, 23(1):282–304.

Davis, R. A., Lee, T. C., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84(2):313–342.

Fiecas, M. and Ombao, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics*, 5(2A):1102–1125.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models.* Springer.

Gombay, E. (1994). Testing for change-points with rank and sign statistics. *Statistics & Probability Letters*, 20(1):49–55.

Hamilton, J. D. (1994). *Time series analysis.* Cambridge University Press.

Horváth, L. (1997). Detection of changes in linear sequences. *Annals of the Institute of Statistical Mathematics*, 49(2):271–283.

Hušková, M. (1990). Asymptotics for robust MOSUM. *Commentationes Mathematicae Universitatis Carolinae*, 31(2):345–356.

Hušková, M. and Kirch, C. (2010). A note on studentized confidence intervals for the change-point. *Computational Statistics*, 25(2):269–289.

Hušková, M., Kirch, C., Prášková, Z., and Steinebach, J. (2008). On the detection of changes in autoregressive time series, II. Resampling procedures. *Journal of Statistical Planning and Inference*, 138(6):1697–1721.

Hušková, M., Prášková, Z., and Steinebach, J. (2007). On the detection of changes in autoregressive time series I. Asymptotics. *Journal of statistical planning and inference*, 137(4):1243–1259.

Hušková, M. and Slabý, A. (2001). Permutation tests for multiple changes. *Kybernetika*, 37(5):605–622.

Jensen, S. T. and Rahbek, A. (2007). Notes and problems on the law of large numbers for (geometrically) ergodic markov chains. *Econometric Theory*, 23:761–766.

Kallenberg, O. (2002). *Foundations of modern probability.* Springer.

Kirch, C. (2006). Resampling methods for the change analysis of dependent data. *PhD Thesis, University of Cologne, Cologne 2006.*

Kirch, C. and Kamgaing, J. T. (2012). Testing for parameter stability in nonlinear autoregressive models. *Journal of Time Series Analysis*, 33(3):365–385.

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent rv'-s, and the sample df. I. *Probability Theory and Related Fields*, 32(1):111–131.

Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent rv's, and the sample df. II. *Probability Theory and Related Fields*, 34(1):33–58.

Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing b-valued random variables. *The Annals of Probability*, pages 1003–1036.

Ling, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *The Annals of Statistics*, 35(3):1213–1237.

Liu, J., Wu, S., and Zidek, J. V. (1997). On segmented multivariate regression. *Statistica Sinica*, 7:497–526.

Lombard, F. (1987). Rank tests for changepoint problems. *Biometrika*, 74(3):615–624.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series: A Practical Introduction using R.* Chapman & Hall.

Marconi, B., Genovesio, A., Battaglia-Mayer, A., Ferraina, S., Squatrito, S., Molinari, M., Lacquaniti, F., and Caminiti, R. (2001). Eye–hand coordination during reaching. I. Anatomical relationships between parietal and frontal cortex. *Cerebral Cortex*, 11(6):513–527.

Marušiaková, M. (2009). Tests for multiple changes in linear regression models. *PhD Thesis, Charles University, Prague.*

McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *The Annals of probability*, 3(5):829–839.

Meelis, E., Bressers, M., and Haccou, P. (1991). Non-parametric testing for the number of change points in a sequence of independent random variables. *Journal of statistical computation and simulation*, 39(3):129–137.

Meyn, S. P., Tweedie, R. L., and Glynn, P. W. (2009). *Markov chains and stochastic stability.* Cambridge University Press Cambridge.

Pan, J. and Chen, J. (2006). Application of modified information criterion to multiple change point problems. *Journal of multivariate analysis*, 97(10):2221–2241.

Politis, D. N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 15(4-5):517–533.

Politis, D. N. and Romano, J. P. (1995). Bias-corrected nonparametric spectral estimation. *Journal of time series analysis*, 16(1):67–103.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33(2):659–680.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry theory*. World Scientific Publishing.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Stockis, J.-P., Franke, J., and Kamgaing, J. T. (2010). On geometric ergodicity of charme models. *Journal of Time Series Analysis*, 31(3):141–152.

Stout, W. F. (1974). *Almost sure convergence*. Academic Press New York.

Venkatraman, E. S. (1993). Consitency results in multiple change-point problems.

Venter, J. and Steel, S. (1996). Finding multiple abrupt change points. *Computational statistics & data analysis*, 22(5):481–504.

Vostrikova, L. J. (1981). Detecting disorder in multidimensional random processes. *Soviet Mathematics Doklady*, 24:55–59.

Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz'criterion. *Statistics & Probability Letters*, 6(3):181–189.

Yao, Y.-C. and Au, S. (1989). Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics*, A:370–381.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Probability Theory and Related Fields*, 52(1):45–57.