Estimating Multiple Structural Breaks in Time Series: A Generalized MOSUM Approach Based on Estimating Functions

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von Kerstin Reckrühm
geb. am 27.08.1987 in Bernburg

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

Gutachter:	Prof. Dr. Claudia Kirch
	Prof. Dr. Roland Fried

Eingereicht am:	18.12.2018
Verteidigung am:	28.03.2019

Abstract

Multiple change point detection is a special area of change point analysis where we are mainly interested in localizing changes in the underlying model of an observed time series. The MOSUM (moving sum) procedure investigated by Eichinger & Kirch (2018) is one of the basic approaches to detect and estimate multiple changes in the classical mean change model. The statistic is constructed by comparing the arithmetic means of subsamples of size G around each time point where G denotes the bandwidth. Hence, a quite natural generalization of this procedure to several parameter change problems would be to use MOSUM Wald-type statistics based on differences of local estimators. However, especially in non-linear models, applying these statistics can lead to high computational effort and large numerical errors. To reduce the complexity in computation we consider MOSUM statistics based on estimating functions (score-type statistics) where only one global estimator of the parameter has to be computed. However, this comes at the cost that a single global estimator may not be able to detect all possible changes. Therefore, we need to repeat the procedure with several estimators and combine the information from all of them.

After an introduction in the first chapter, the second chapter of the thesis focuses on MOSUM score-type statistics. We construct a corresponding test statistic and investigate its asymptotic behavior under the null hypothesis and the alternative. Further, we consider estimators for the number and the locations of the changes and examine their statistical properties. The theoretical results derived in this part enable us to develop a theory for MOSUM Wald-type statistics. In Chapter 3 we investigate the asymptotic properties of a change point test and estimators based on Wald-type statistics. This is followed by some simulation studies for a linear regression model and a Poisson autoregressive model where we compare the performance of the two MOSUM procedures. The simulation results illustrate that the MOSUM Wald-type procedure usually performs better than its score-type counterpart. Further, we observe that, in particular, the score-type procedure strongly depends on the selection of the bandwidth. This bandwidth problem and the problem in detectability can be solved by applying a multiscale method which merges the results obtained from different bandwidths and global estimators in an appropriate way as discussed in Chapter 5. After describing the multiscale procedure for the classical mean change model introduced by Cho & Kirch (2018), we adapt their method to a general setting and the linear regression model and we derive first theoretical results constituting the basis for future work.

Zusammenfassung

Die Detektion multipler Strukturbrüche ist ein spezielles Gebiet der Changepoint Analyse, in dem man hauptsächlich daran interessiert ist, Änderungen in einem einer beobachteten Zeitreihe zugrunde liegendem Modell zu lokalisieren. Das MOSUM-Verfahren (moving sum), welches von Eichinger & Kirch (2018) näher untersucht wurde, repräsentiert eines der grundlegenden Verfahren zur Erkennung und Schätzung multipler Änderungen im klassischen Erwartungswertmodell. Die dazugehörige Statistik ergibt sich aus dem Vergleich der Stichprobenmittelwerte, die auf Teilstichproben der Größe G, auch Bandbreite genannt, vor und nach jedem Zeitpunkt berechnet werden. Ein sehr intuitiver Ansatz zur Verallgemeinerung dieses Verfahrens auf Modelle verschiedener Parameteränderungen wäre daher die Verwendung von Wald-Statistiken, die auf den Differenzen lokaler Parameterschätzer basieren. Ein Nachteil dieser Statistiken ist jedoch, dass sie insbesondere in nicht-linearen Modellen nur mit hohem Rechenaufwand und großem numerischen Fehler bestimmt werden können. Die rechnerische Komplexität kann durch Anwendung von MOSUM-Statistiken basierend auf Schätzfunktionen (Score-Statistiken) deutlich reduziert werden, da zur Berechnung dieser Statistiken lediglich ein globaler Schätzer bestimmt werden muss. Dieses Verfahren hat dennoch den Nachteil, dass ein globaler Schätzer allein nicht zwingend dazu in der Lage ist alle Strukturbrüche zu erkennen, was eine Wiederholung des Verfahrens mit verschiedenen globalen Schätzern mit anschließender Zusammenführung der Ergebnisse erforderlich macht.

Nach einer Einleitung im ersten Kapitel wenden wir uns im zweiten Kapitel dieser Arbeit den MOSUM Score-Statistiken zu. Wir konstruieren eine entsprechende Teststatistik und untersuchen deren asymptotisches Verhalten unter der Nullhypothese und der Alternative. Außerdem werden Schätzer für die Anzahl und die Positionen der Strukturbrüche betrachtet und im Hinblick auf ihre statistischen Eigenschaften analysiert. Die theoretischen Resultate dieses Kapitels ermöglichen es uns, eine Theorie für die Wald-Statistiken zu entwickeln. Im dritten Kapitel werden die asymptotischen Eigenschaften eines Changepoint Tests und von Changepoint Schätzern basierend auf Wald-Statistiken genauer untersucht. Im folgenden Kapitel werden die Simulationsstudien für ein lineares Regressionsmodell und ein Poisson autoregressives Modell beschrieben und die beiden MOSUM-Verfahren werden miteinander verglichen. Die Simulationsergebnisse zeigen, dass das Wald-Verfahren in den meisten Fällen besser abschneidet. Es ist zudem ersichtlich, dass insbesondere die Leistungsfähigkeit des Score-Verfahrens von der Wahl der Bandbreite abhängt. Das Bandbreitenproblem und das Problem in der Detektierbarkeit können beide durch Anwendung eines Multiskalen-Verfahrens, das die Ergebnisse unter Verwendung verschiedener Bandbreiten und globaler Schätzer in geeigneter Weise zusammenführt, gelöst werden. In Kapitel 5 wird zuerst das von Cho & Kirch (2018) für das Erwartungswertmodell eingeführte Multiskalen-Verfahren beschrieben, bevor wir es auf ein allgemeines Modell bzw. ein lineares Regressionsmodell anpassen und erste theoretische Resultate herleiten, die die Grundlage für weiterführende Untersuchungen bilden.

Danksagung

An dieser Stelle möchte ich allen Kollegen, Freunden und meiner Familie, die mich während meiner Promotion unterstützt haben und somit zur Fertigstellung dieser Arbeit in gewisser Weise beigetragen haben, von ganzem Herzen danken.

Mein besonderer Dank gilt meiner Betreuerin Prof. Dr. Claudia Kirch, die mir die Möglichkeit gegeben hat im Bereich der Changepoint Analyse zu forschen und zu promovieren. Vielen Dank für die permanente Unterstützung und Inspiration.
Außerdem möchte ich Prof. Dr. Roland Fried meinen Dank für die Begutachtung meiner Arbeit aussprechen.
Ein herzlicher Dank gilt zudem meinen Kollegen und Freunden am Institut für Mathematische Stochastik. Vielen Dank für die unterhaltsamen Kaffeepausen und die familiäre Arbeitsatmosphäre. Dabei möchte ich mich insbesondere bei Kerstin Alterkirch für die Fürsorge und Herzlichkeit bedanken. Vielen Dank an Alexander Meier für die freundschaftliche Zusammenarbeit und die Unterstützung in der Endphase meiner Promotion. Mein besonderer Dank gilt meiner Doktorschwester Christina Stöhr für die wunderbare gemeinsame Zeit. Vielen Dank für deine Unterstützung und deine Freundschaft.
Ich möchte mich zudem ganz herzlich bei meiner Familie, meinen Eltern, meinem Bruder und meiner Schwägerin, für die Fürsorge und Liebe bedanken.
Mein größter Dank geht an meinen Partner und besten Freund Philipp. Vielen Dank, dass du immer an mich glaubst und mich bei allem unterstützt.

Die Arbeit wurde durch die Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 314838170, GRK 2297 MathCoRe unterstützt. Mein Forschungsaufenthalt an der Macquarie University in Sydney im März 2017 wurde durch das Büro für Gleichstellungsfragen der Otto-von-Guericke-Universität gefördert.

Contents

1. Introduction

Multiple change point detection is a special area of change point analysis with main focus on finding structural breaks in an observed time series within a fixed period of time. A structural break or a change point describes a time point of the observation period at which the underlying model and, consequently, the distributional properties of the observations change. In this work we consider a general parameter change model defined in Section 1.1, in which we allow for multiple changes in the parameter vector $\boldsymbol{\theta}$ specifying the distribution of the observations $X_1, \ldots, X_n$. This general framework covers many different change point problems. Some examples are given in Section 1.1 as well. The number of the change points as well as their locations are unknown and our goal is to find suitable estimators for these values. Therefore, we are going to investigate two general MOSUM (moving sum) procedures where the first one is based on score-type statistics and the second one uses Wald-type statistics. Before explaining these approaches in detail, we want to get an idea of what has been discussed so far in the scientific literature.

Multiple change point detection is a current topic in research and comprises problems in a wide range of fields like finance, quality control, medicine or climate. For example, Braun *et al.* (2000) applied a technique of multiple change point detection on DNA sequences. Aggarwal *et al.* (1999) focussed on finding structural breaks in the volatility of stock market returns. More recently, Killick *et al.* (2010) and Killick *et al.* (2012) gave an interesting application to oceanography by detecting changes in the variance of time series for wave heights.

All the techniques and methods designed for localising multiple structural breaks are more or less inspired by model selection or hypothesis testing. These are the two origins of multiple change point detection which have been developing alongside. Approaches based on model selection rest upon the idea of finding change points by optimizing a target or cost function over all possible change point constellations where overfitting is avoided by setting an upper bound for the number of changes or by introducing a penalty on the complexity of the model. They date back to Yao (1988) who applied Schwarz's criterion to estimate the number of changes in a mean change model under normality assumptions. Since then many different detection procedures have been considered for several models. For instance, Yao & Au (1989) used a least-squares based target function to localise changes in the mean of i.i.d. (independent and identically distributed) observations or, more recently, Davis *et al.* (2006) applied the minimum description length principle in order to detect changes in the regression coefficients of a linear autoregressive process. More references are given in Section 5. Nevertheless, all these methods have in common that the change point estimates are determined by solving a multivariate optimization problem which can be computationally challenging. In order to solve this problem more efficiently several algorithms have been developed. One of these is the PELT (pruned exact linear time) algorithm by Killick *et al.* (2012)

which is based on dynamic programming combined with a pruning step.
In hypothesis testing there are two basic approaches dealing with multiple change points, the binary segmentation procedure and the MOSUM procedure. The first one is an iterative method going back to Vostrikova (1981) which rests upon the idea that tests for AMOC (at most one change) settings still have some power for multiple changes. By conducting an AMOC-test on the whole data sequence, a change point estimate is obtained and the sample is split at this point into two subsamples. Then, the steps are repeated recursively on the subsamples as long as the test rejects the null hypothesis of having no structural break. By introducing a random localisation mechanism for choosing the segments or subsamples on which the test is conducted the wild binary segmentation procedure by Fryzlewicz (2014) solves the power problem of the original procedure for specific change point constellations. Moreover, the concept of binary segmentation has attracted attention in model selection too mainly due to its low computational complexity, see for example Killick & Eckley (2014). However, one main drawback of binary segmentation in comparison to the MOSUM procedure is that it involves multiple testing such that the overall significance level cannot be controlled.
Using MOSUM statistics in hypothesis testing goes back to Bauer & Hackl (1980) who considered MOSUM based test motivated by an application in quality control. Furthermore, Hušková (1990) and Chu *et al.* (1995) investigated a MOSUM-test based on least squares residuals in linear regression models. Whereas the tests mentioned before were constructed for AMOC alternatives, Hušková & Slabý (2001) proposed a MOSUM statistic actually designed for detecting multiple changes. More recently, Eichinger & Kirch (2018) investigated the MOSUM procedure for the classical mean change model and derived consistency for the estimators of the number and the locations of the changes. In this particular example the MOSUM score-type statistic coincides with the MOSUM Wald-type statistic as explained in the following Section 1.2.

1.1. Examples

In this work we investigate MOSUM procedures in order to estimate multiple structural breaks in a very general setting which is described in the following. The observations $X_1, \dots, X_n$ follow a general parameter change model if

$$X_i = \begin{cases} X_i^{(1)}, & \text{if } 1 \leq i \leq k_{1,n} \\ X_i^{(2)}, & \text{if } k_{1,n} < i \leq k_{2,n} \\ \vdots & \\ X_i^{(q+1)}, & \text{if } k_{q,n} < i \leq n \end{cases}, \tag{1.1}$$

where q denotes the number of structural breaks and $k_{1,n}, \dots, k_{q,n}$ are the change points. The sequences $\{X_i^{(j)}\}_{i\geq 1}, j = 1, \dots, q+1$, are assumed to be stationary with a distribution specified by some parameter vector $\boldsymbol{\theta}_j$ such that $\{X_i\}_{i\geq 1}$ is piecewise stationary. Note that the application of the MOSUM procedures is not restricted to correctly specified models and that we allow for misspecification as well which will be explained in the respective sections. The model description above incorporates many different change point models. A few examples are given in the following. One main example is the classical mean change model which has been considered extensively in the change point literature.

Example 1.1.1 (Mean Change Model). *The observations can be described by*

$$X_i = \begin{cases} \mu_1 + \varepsilon_i, & \text{if } 1 \le i \le k_{1,n} \\ \mu_2 + \varepsilon_i, & \text{if } k_{1,n} < i \le k_{2,n} \\ \vdots & \\ \mu_{q+1} + \varepsilon_i, & \text{if } k_{q,n} < i \le n \end{cases},$$

where $\{\varepsilon_i\}_{i\ge1}$ *is a stationary error sequence with an expectation of zero and long-run variance* $0 < \tau^2 < \infty$ *and* $\mu_1, \dots, \mu_{q+1}$ *represent the expected values.*

In this classical example, a time series deviates randomly from a specific value, the expected value, which may change several times.
Another well investigated model in statistics is the linear regression model. A description of that including change points is given in the next example.

Example 1.1.2 (Linear Regression Model with Structural Breaks). *The model equation of a linear regression with two regressors* $\{X_{i,1}\}$ *and* $\{X_{i,2}\}$ *is given by*

$$Y_i = \begin{cases} \boldsymbol{X}_i^T \boldsymbol{\beta}_1 + \varepsilon_i, & \text{if } 1 \le i \le k_{1,n} \\ \boldsymbol{X}_i^T \boldsymbol{\beta}_2 + \varepsilon_i, & \text{if } k_{1,n} < i \le k_{2,n} \\ \vdots & \\ \boldsymbol{X}_i^T \boldsymbol{\beta}_{q+1} + \varepsilon_i, & \text{if } k_{q,n} < i \le n \end{cases},$$

where Y_i *denotes the response variable,* $\boldsymbol{X}_i = (1, X_{i,1}, X_{i,2})^T$, $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{q+1}$ *are the parameter vectors and* $\{\varepsilon_i\}_{i\ge1}$ *represents an error sequence as in Example 1.1.1.*

This linear regression model will be discussed in more detail in Section 4.1.
In the following example we consider an integer-valued time series model, the Poisson autoregressive model of order one, which will be investigated in Section 4.2.

Example 1.1.3 (Poisson Autoregressive Model with Structural Breaks). *A Poisson autoregressive model of order one, also known as INARCH(1) model, with q change points can be described by*

$$Y_i = \begin{cases} Y_i^{(1)}, & \text{if } 1 \le i \le k_{1,n} \\ Y_i^{(2)}, & \text{if } k_{1,n} < i \le k_{2,n} \\ \vdots & \\ Y_i^{(q+1)}, & \text{if } k_{q,n} < i \le n \end{cases},$$

where $\{Y_i^{(j)}\}$ *is an INARCH(1) time series with parameter* $\boldsymbol{\theta}_j = (\theta_{j,1}, \theta_{j,2})^T$, $j = 1, \dots, q+1$, *i.e.*

$$Y_i^{(j)} | \mathcal{F}_{i-1} \sim P(\lambda_i), \quad \text{with } \lambda_i = \theta_{j,1} + \theta_{j,2} Y_{i-1}.$$

In this model the observation Y_i conditioned on the past is Poisson distributed with parameter λ_i which is described by an autoregressive structure of order one whose coefficients may change over time.

1.2. Motivation of the MOSUM Statistics

A classical estimation method in statistics is based on finding the solution $\widehat{\boldsymbol{\theta}}_{1,n}$ of the estimating equation system $\sum_{i=1}^{n} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) \stackrel{!}{=} \mathbf{0}$ for a suitable choice of $\mathbb{X}_i$, see below for some examples. The function $\boldsymbol{H}$ is called estimating function and $\widehat{\boldsymbol{\theta}}_{1,n}$ is an M-estimator or Z-estimator of $\boldsymbol{\theta}$. For more information on that estimation method we refer to Van der Vaart (2007), Chapter 5. Furthermore, note that the estimating function $\boldsymbol{H}$ will be vector-valued for the estimation of multidimensional parameter vectors.
In the classical model in Example 1.1.1 we consider a time series $X_1, \ldots, X_n$ fluctuating randomly around a mean μ. A typical estimator for the expectation is the sample mean $\bar{X}_{1,n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ which is the solution of the estimating equation $\sum_{i=1}^{n}(X_i - \mu) \stackrel{!}{=} 0$. Another M-estimator for the expectation in this example is based on the estimating function $H(X_i, \mu) = \frac{2}{\pi}\arctan(\mu - X_i)$. This estimating function is closely related to the estimating function for the median which is given by the sign-function, but has the advantage that it has nice differentiability properties. Henceforth, it will therefore be called median-like estimator.
In the linear regression model of Example 1.1.2 the classical least squares estimator minimises the sum of squared residuals and thus represents an M-estimator for the parameter vector $\boldsymbol{\beta}$ with vector-valued estimating function $H\left((Y_i, X_{i,1}, X_{i,2})^T, \boldsymbol{\beta}\right) = -2\boldsymbol{X}_i\left(Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}\right)$ so that $\mathbb{X}_i = (Y_i, X_{i,1}, X_{i,2})^T$.
In the Poisson autoregressive model of Example 1.1.3 an M-estimator based on the likelihood approach can be used. Its estimating function is given by $\boldsymbol{H}((Y_i, Y_{i-1})^T, \boldsymbol{\theta}) = -2\boldsymbol{Y}_{i-1}\left(\frac{Y_i}{\boldsymbol{Y}_{i-1}^T\boldsymbol{\theta}} - 1\right)$, where $\boldsymbol{Y}_{i-1} = (1, Y_{i-1})^T$.
In order to explain how change point tests can be constructed based on estimating functions, first consider the at most one change (AMOC) situation.

$\boldsymbol{H_0}$: no structural break vs. $\boldsymbol{H_1}$: at least one structural break

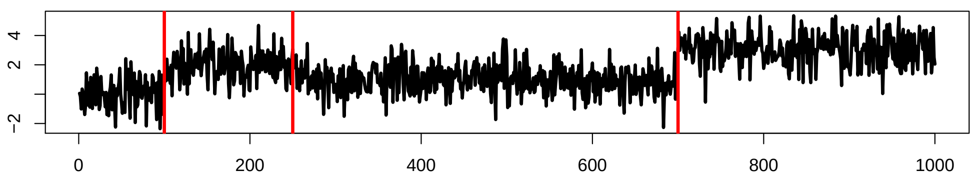
The first plot shows a time series fluctuating around a constant mean value of zero whereas the second plot illustrates a change in mean from 0 to 2 at time point 200, which can easily be seen by eye. However, in many situations changes cannot be found only by visual inspection and statistical tools are needed. Tests based on weighted CUSUM statistics

$$\chi^2 = \max_{1 \le k \le n} \sqrt{\frac{n}{k(n-k)}} \frac{1}{\tau} \left| \sum_{i=1}^{k} (X_i - \bar{X}_{1,n}) \right| \tag{1.2}$$

can be applied in order to detect a change in mean. For further information on the statistic see, for example, Chapter 2 of Csörgö & Horváth (1997). After some transformations we get the following representation of the statistic

$$\chi^2 = \max_{1\le k\le n} \sqrt{\frac{k(n-k)}{n}} \frac{1}{\tau} \left| \bar{X}_{1,k} - \bar{X}_{k+1,n} \right|, \tag{1.3}$$

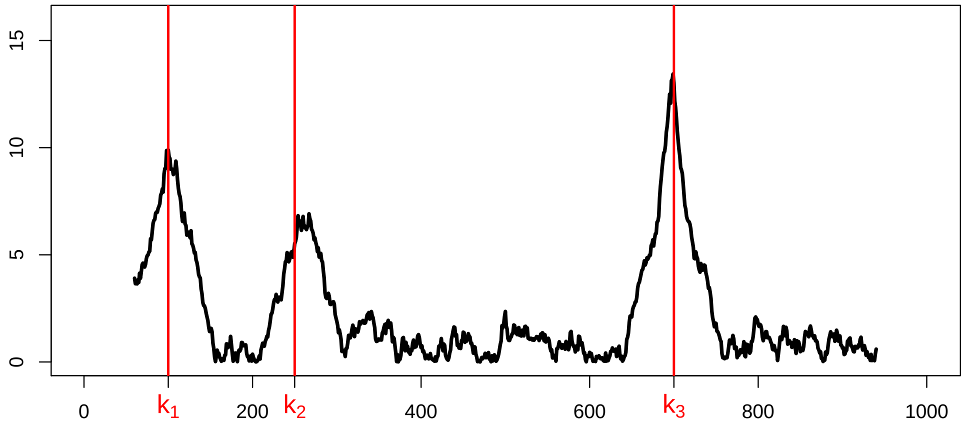
which bases on the absolute difference of the arithmetic mean of the subsample $X_1, \ldots, X_k$ and the arithmetic mean of the of the subsample $X_{k+1}, \ldots, X_n$.
By recalling that the estimating function of the classical mean is given by $H(X_i, \mu) = X_i - \mu$, the version of the weighted CUSUM statistic in (1.2) can be conceived as a special case of score-type statistics based on $\sum_{i=1}^{k} \boldsymbol{H}(X_i, \widehat{\boldsymbol{\theta}}_{1,n})$ in the general paramter change model where $\widehat{\boldsymbol{\theta}}_{1,n}$ denotes the global M-estimator and $\boldsymbol{H}$ its corresponding estimating function. Score-type statistics have already received attention in the literature, e.g. Hušková (1996) examined weighted CUSUM score-type tests and change point estimators under the null hypothesis and local alternatives with general regularity conditions. Furthermore, Hušková *et al.* (2007) detected changes in linear autoregressive time series by applying test statistics based on partial sums of weighted residuals which are specific score-type statistics using the least squares estimating function. More recently, Kirch & Tadjuidje Kamgaing (2016) contructed change point tests using score-type statistics and derived consistency for the tests under the alternative with an application to binary models and Poisson autoregressive models.
The representation of the CUSUM statistic in (1.3) is an example for Wald-type statistics in general resting upon the comparison of the M-estimator computed on the first k observations and the M-estimator of the last $n-k$ observations: $\widehat{\boldsymbol{\theta}}_{1,k} - \widehat{\boldsymbol{\theta}}_{k+1,n}$. For instance, Andrews (1993) considered Wald-type statistics based on GMM (generalized method of moments) estimators in a quite general setting, which constitutes a generalization of the results of Hawkins (1987) investigating Wald-type tests based on ML (maximum likelihood) estimators in the i.i.d. case with known probability density function.
Now we allow for multiple changes in the expectation μ or parameter vector $\boldsymbol{\theta}$ under the alternative and we are interested in estimating the number of structural breaks and their locations, i.e. the time points when the changes occur. In the plot below we see an example of a time series with multiple changes in the mean where the red vertical lines indicate the change points.



In order to detect all structural breaks, it is reasonable to use a moving window statistic like the MOSUM statistic considered by Eichinger & Kirch (2018):

$$T_n(G) = \max_{G\le k\le n-G} \frac{1}{\tau\sqrt{2G}} \left| \sum_{i=k+1}^{k+G} X_i - \sum_{i=k-G+1}^{k} X_i \right|$$

with bandwidth G determining the length of the moving window. The following graph illustrates the behavior of the MOSUM statistic in the example. We can see that it is suitable to detect multiple changes as the statistic gets quite large in intervals around the change points with local maxima close to the true locations of the changes.



Similar to the weighted CUSUM statistic, the MOSUM statistic for mean changes can be regarded as an example of MOSUM Wald-type and score-type statistics since further transformations yield

$$T_n(G) = \max_{G \le k \le n-G} \frac{\sqrt{G}}{\tau\sqrt{2}} \left| \bar{X}_{k+1,k+G} - \bar{X}_{k-G+1,k} \right| .$$

and

$$\begin{aligned} T_n(G) &= \max_{G \le k \le n-G} \frac{1}{\tau\sqrt{2G}} \left| \sum_{i=k+1}^{k+G} (X_i - \bar{X}_{1,n}) - \sum_{i=k-G+1}^{k} (X_i - \bar{X}_{1,n}) \right| \\ &= \max_{G \le k \le n-G} \frac{1}{\tau\sqrt{2G}} \left| \sum_{i=k+1}^{k+G} H(X_i, \bar{X}_{1,n}) - \sum_{i=k-G+1}^{k} H(X_i, \bar{X}_{1,n}) \right| . \end{aligned}$$

In the general parameter change model, MOSUM Wald-type statistics are based on the difference of local M-estimators: $\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}$. For each time point k between G and $n-G$ we compare the estimator computed on the subsample $X_{k+1}, \ldots, X_{k+G}$ with the estimator calculated on $X_{k-G+1}, \ldots, X_k$ whereupon a large difference indicates a change at this point. This constitutes a natural approach to find changes in a parameter vector $\boldsymbol{\theta}$ but it has one main drawback. Calculating two estimates for each time point, i.e. $2(n-2G)$ in total, can be computationally challenging and can lead to high numerical errors especially in non-linear models, for example non-linear AR models or Poisson autoregressive models, where numerical methods are needed to determine the estimates. In order to reduce the complexity in computation we will consider MOSUM score-type statistics for general parameter change problems which are based on the difference of sums of the estimating function where a global estimator is employed:

$$\sum_{i=k+1}^{k+G} \boldsymbol{H}\left(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}_n\right) - \sum_{i=k-G+1}^{k} \boldsymbol{H}\left(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}_n\right) \quad \text{with} \quad \widetilde{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_{1,n}.$$

The basic idea here is to convert a general multiple parameter change problem to a multiple mean change problem of the estimating function. After computing the global estimator, we focus on the transformed sequence $\{\boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}_n)\}_{i\geq 1}$ and try to find changes in the mean of this new series by applying a multivariate version of the classical MOSUM statistic. This approach is computationally very fast as it only requires the calculation of one estimator. However, it can happen that not all structural breaks are detectable by MOSUM score-type statistics since a change in the parameter vector $\boldsymbol{\theta}$ does not necessarily cause a change in the expectation of the transformed series. We will discuss this in detail later.
Both approaches have advantages and disadvantages but they have one problem in common. The performance of the procedures crucially depends on the choice of the bandwidth and in some situations it might be helpful to use more than one bandwidth. For example, the detection of changes, which are large in magnitude and lie close to each other, requires a small bandwidth whereas small changes located far away from each other can be found by a large bandwidth. Hence, a mixture of these scenarios can cause problems which we try to solve by adapting multiscale method. We will explain this later.

1.3. Structure of the Thesis

My thesis is structured as follows. In Chapter 2 we consider MOSUM score-type statistics which are also called MOSUM statistics based on estimating functions. After deriving the asymptotics under the null hypothesis, we construct an asymptotic level α test and prove its consistency under the alternative. In Section 2.1.3 we introduce a MOSUM procedure, which is similar to that of Eichinger & Kirch (2018), to determine estimators for the number and the locations of the changes followed by proofs of consistency. The convergence rates of the estimator sequences can be improved under some modifications and additional assumptions on the time series which is shown in Section 2.2. Since the assumptions used in previous subsections are expressed as generally as possible we prove in Section 2.4 that the assumptions are satisfied by i.i.d. series and stationary and strongly mixing time series under some moment conditions. We complete this chapter by discussing possible problems of the procedure. Chapter 3 focuses on MOSUM Wald-type statistics. In the first section, we consider a general parameter change model for those two examples. After investigating the asymptotic behavior of the statistic under the null hypothesis, we examine the properties of the corresponding test and estimators under the alternative in Section 3.1.2. In the second part of this chapter, MOSUM Wald-type statistics for the linear regression model are considered and statistical properties of the corresponding change point test and estimators are derived. This is followed by a discussion of possible problems of the Wald-type procedure in Section 3.3. In Chapter 4 we describe the results of simulation studies for a linear regression model and a Poisson autoregressive model and compare the performance of the two MOSUM procedures. With the goal to solve the bandwidth problem of both procedures and the problem in detectability of the MOSUM score-type approach, we consider a multiscale method in Chapter 5. The multiscale MOSUM procedure with localized pruning introduced by Cho & Kirch (2018) is described in Section 5.2, before we adapt this method to the linear regression model and a general parameter change model. After providing all the theoretical tools, we prove a first result for the output of the procedure. This constitutes the foundation for future work as described in the outlook concluding the main part of the thesis. In the appendix, we give an overview of the assumptions followed by a summary of several theoretical results which are used throughout this work.

2. MOSUM Score-Type Statistics

We consider a general parameter change model as in (1.1) with observations $X_1, \dots, X_n$. The distribution of these random variables is determined by a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\boldsymbol{\Theta}$ denotes the parameter space being a subset of $\mathbb{R}^p$. The estimating function $\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})$ is a measurable function with respect to $\mathbb{X}_i$ and we assume that it takes values in $\mathbb{R}^p$ as well. All the assumptions on the time series that are required under the null hypothesis and the alternative are introduced in the respective sections and summarized in the Appendix A.1 and A.2.
With the goal to detect and estimate changes in the parameter vector $\boldsymbol{\theta}$ we introduce the following MOSUM statistics based on estimating functions which is also called MOSUM score-type statistic:

Definition 2.0.1. *A MOSUM statistic based on estimating functions is given by*

$$T_{k,n}(G, \widetilde{\boldsymbol{\theta}}) = \frac{1}{\sqrt{2G}} \sqrt{\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}}, \quad \textit{for some } \widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta} \textit{ and } k \in \{G, \dots, n-G\}$$

with $\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} = \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})$ *and* $\boldsymbol{\Sigma}_k$ *as long-run covariance matrix of* $\boldsymbol{H}(\mathbb{X}_k, \widetilde{\boldsymbol{\theta}})$ *which is assumed to be positive definite.*

Note that we use an arbitrary $\widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ instead of the estimator sequence $\widehat{\boldsymbol{\theta}}_{1,n}$ like in Section 1.2. Doing so, we are able to develop a theory which is a little bit more general concerning the choice of the global estimator sequence. Later, we will see that $\widetilde{\boldsymbol{\theta}}$ can be replaced by any estimator sequence $\widetilde{\boldsymbol{\theta}}_n$ satisfying some specific conditions.
The following remark will be helpful in the proofs and we will use it throughout this work without referring to it again.

Remark 2.0.2. *It holds that*

$$T_{k,n}(G, \widetilde{\boldsymbol{\theta}}) = \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|_F = \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|,$$

where $\|\cdot\|_F$ *is the Frobenius norm described in Appendix E.1.1 and* $\|\cdot\|$ *the Euclidean norm.*

Proof. Since the positive definite matrix $\boldsymbol{\Sigma}_k^{-1}$ has a symmetric square root $\boldsymbol{\Sigma}_k^{-1/2}$ (see Roy & Banerjee (2014) page 415f) we obtain

$$\sqrt{\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}} = \sqrt{\operatorname{tr}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}^T \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right)} = \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|_F.$$

Furthermore, by Lemma E.1.4 in the appendix we know that the Frobenius norm of a vector coincides with its Euclidean norm. Hence, we receive $\left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|_F = \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|$. □

Moreover, to simplify the wording in the theoretical results and the calculations in the proofs we use the following notations:

- We denote the maximum of the MOSUM score-type statistic by

$$T_n(G, \widetilde{\boldsymbol{\theta}}) = \max_{G \leq k \leq n-G} T_{k,n}(G, \widetilde{\boldsymbol{\theta}}).$$

- If the long-run covariance matrix $\boldsymbol{\Sigma}_k$ is replaced by an estimator $\widehat{\boldsymbol{\Sigma}}_{k,n}$ we write

$$\widehat{T}_{k,n}(G, \widetilde{\boldsymbol{\theta}}) = \frac{1}{\sqrt{2G}} \sqrt{\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}^T \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}} \text{ and } \widehat{T}_n(G, \widetilde{\boldsymbol{\theta}}) = \max_{G \leq k \leq n-G} \widehat{T}_{k,n}(G, \widetilde{\boldsymbol{\theta}}).$$

- Besides, we use

$$\boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta}) := \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\right).$$

- Furthermore, note that

$$E\left(\boldsymbol{A}_{\boldsymbol{\theta},k}\right) = \sum_{i=k+1}^{k+G} E\left(\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\right) - \sum_{i=k-G+1}^{k} E\left(\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\right).$$

- Moreover, we use $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ for the long-run covariance matrix of $\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})$. We only write $\boldsymbol{\Sigma}$ if it is obvious which $\boldsymbol{\theta}$ it refers to.

2.1. Theoretical Results

In this section we investigate the asymptotic behavior of the statistic under the null hypothesis that no structural break occurs and under the alternative that at least one structural break arises.

2.1.1. Asymptotics Under the Null Hypothesis

In the first theorem we derive the limit distribution under the null hypothesis which enables us to contruct a test and gives a threshold for the estimating procedure under the alternative later on. In order to prove this result, the following assumptions are needed. Note that $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}$ holds for all k under H_0.

A.1.1: *Let the bandwidth G depend on n, i.e. $G = G(n)$. Furthermore, for $\nu > 0$ assume that*

$$\frac{n}{G} \to \infty \quad and \quad \frac{n^{\frac{2}{2+\nu}} \log(n)}{G} \to 0 \quad for \quad n \to \infty.$$

This assumption is very important and used throughout this chapter. It ensures that the bandwidth tends to infinity as n goes to infinity but not too fast or too slow. Note

that the constant value ν comes from the invariance principle in Assumption A.1.3 and is specified there.
The following assumption is essential and describes the main setting under the null hypothesis.

A.1.2 *Let $\{\mathbb{X}_i : i \geq 1\}$ be a stationary series following a distribution determined by $\boldsymbol{\theta}_0$ in a correctly specified model. Under misspecification let $\boldsymbol{\theta}_0$ be the best approximating parameter for $\{\mathbb{X}_i : i \geq 1\}$ in the sense of $E(\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0)) = \mathbf{0}$. Furthermore, we assume that the stationary sequence $\{\boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) : i \geq 1\}$ has a positive definite long-run covariance matrix $\boldsymbol{\Sigma}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}$.*

This additionally shows that we do not only restrict our attention to correctly specified models and allow for misspecification as well. Moreover, note that the stationarity of the transformed sequence $\{\boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\}$ follows immediately from the stationarity of $\{\mathbb{X}_i\}$ and the measurability of $\boldsymbol{H}$ with respect to $\mathbb{X}_i$. This implies that $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$ holds for all k as already used in the assumption above.
The third assumption gives a strong invariance principle for the transformed series.

A.1.3: *Let $\boldsymbol{S}(k, \widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{k} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})$ fulfill a strong invariance principle. So possibly after changing the probability space, there exists a p-dimensional standard Wiener process $\{\boldsymbol{W}(k) : k \geq 0\}$ with identity matrix $\boldsymbol{I}_p$ as covariance matrix and $\nu > 0$ such that*

$$\left\| \boldsymbol{\Sigma}^{-1/2} \left(\boldsymbol{S}(k, \widetilde{\boldsymbol{\theta}}) - E(\boldsymbol{S}(k, \widetilde{\boldsymbol{\theta}})) \right) - \boldsymbol{W}(k) \right\| = O(k^{1/(2+\nu)}) \ \ a.s.$$

as k goes to infinity.

If we want to replace $\widetilde{\boldsymbol{\theta}}$ by an estimator sequence $\widehat{\boldsymbol{\theta}}_n$ in the statistic we have to assume that the sequence satifies the following condition.

A.1.4: *Let*

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n, k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}}, k} \right\| = o_P\left((\log(n/G))^{-1/2} \right)$$

hold for some $\widetilde{\boldsymbol{\theta}}$.

As explained in Section 2.3 below, Assumption A.1.4 holds for i.i.d. and stationary and strongly mixing time series under some moment conditions if the estimating function $\boldsymbol{H}$ is twice continuously differentiable on $\boldsymbol{\Theta}$.
Since the long-run covariance matrix $\boldsymbol{\Sigma}$ is typically unknow it might be reasonable to replace it by an appropriate estimator as well. The following assumption describes the conditions that a matrix-valued estimator sequence for $\boldsymbol{\Sigma}$ has to fulfill for this purpose.

A.1.5 *The estimator $\widehat{\boldsymbol{\Sigma}}_{k,n}$ of the long-run covariance matrix $\boldsymbol{\Sigma}$ satisfies*

$$\max_{G \leq k \leq n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}^{-1/2} \right\|_F = o_P\left(\left(\log(n/G) \right)^{-1} \right)$$

under the null hypothesis.

Before we come to a first main result of this thesis, we need to introduce the following definitions. Let

$$a(x) = \sqrt{2\log(x)} \text{ and} \tag{2.1}$$
$$b(x) = 2\log(x) + \frac{p}{2}\log(\log(x)) - \log\left(\frac{2}{3}\Gamma\left(\frac{p}{2}\right)\right),$$

where p is the dimension of the parameter space and Γ denotes the gamma function.

Theorem 2.1.1 (Limit Distribution). *Let Assumption A.1.1 hold for the bandwidth G and Assumptions A.1.2 and A.1.3 hold for some $\widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$.*

(a) Then, under H_0,

$$a(n/G)T_n(G, \widetilde{\boldsymbol{\theta}}) - b(n/G) \xrightarrow{\mathcal{D}} E$$

with E as Gumbel distributed random variable, i.e. $P(E \leq x) = \exp(-2\exp(-x))$ and with $a(x)$ and $b(x)$ as in (2.1).

(b) Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ be a sequence of estimators fulfilling Assumption A.1.4 for $\widetilde{\boldsymbol{\theta}}$. Then, under H_0,

$$a(n/G)T_n(G, \widehat{\boldsymbol{\theta}}_n) - b(n/G) \xrightarrow{\mathcal{D}} E.$$

(c) Furthermore, the covariance matrix $\boldsymbol{\Sigma}$ can be replaced by an estimator $\widehat{\boldsymbol{\Sigma}}_{k,n}$ satisfying Assumption A.1.5 without changing the results of this theorem.

Proof. (a) The proof of (a) consists of four main parts.

1) Replacing $\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}$ by increments of a p-dimensional standard Wiener process $\{\boldsymbol{W}(t) : t \geq 0\}$:

On noting that $E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) = \mathbf{0}$ holds under H_0 for all time points k, Assumption A.1.3 yields

$$\begin{aligned}
&\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} + 2\boldsymbol{W}(k) - \boldsymbol{W}(k-G) - \boldsymbol{W}(k+G)\right\| \\
&= \left\|\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right) + 2\boldsymbol{W}(k) - \boldsymbol{W}(k-G) - \boldsymbol{W}(k+G)\right\| \\
&= \left\|\boldsymbol{\Sigma}^{-1/2}\left(\sum_{i=k+1}^{k+G}\boldsymbol{H}_0(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k}\boldsymbol{H}_0(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\right)\right. \\
&\qquad \left. + 2\boldsymbol{W}(k) - \boldsymbol{W}(k-G) - \boldsymbol{W}(k+G)\right\| \\
&\leq \left\|\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^{k+G}\boldsymbol{H}_0(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k+G)\right\| + 2\left\|\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^{k}\boldsymbol{H}_0(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k)\right\| \\
&\quad + \left\|\boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^{k-G}\boldsymbol{H}_0(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k-G)\right\|
\end{aligned}$$

$$= O((k+G)^{1/(2+\nu)}) + O(k^{1/(2+\nu)}) + O((k-G)^{1/(2+\nu)}) \ a.s.$$
$$= O(n^{1/(2+\nu)}) \ a.s. \quad \text{uniformly in } k.$$

Hence, by Assumption A.1.1 we obtain

$$\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} + 2\boldsymbol{W}(k) - \boldsymbol{W}(k-G) - \boldsymbol{W}(k+G) \right\|$$
$$= O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{G}} \right) = o_P\left((a(n/G))^{-1} \right).$$

Furthermore, by Lemma E.1.2 in the appendix together with the result above we receive

$$\Bigg| \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) \right\| \Bigg|$$
$$\le \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - \boldsymbol{W}(k-G) + 2\boldsymbol{W}(k) - \boldsymbol{W}(k+G) \right\|$$
$$= o_P\Big(\big(a(n/G) \big)^{-1} \Big),$$

implying that

$$a(n/G) \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G)$$
$$= a(n/G) \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) \right\| - b(n/G) + o_P(1).$$

2) Replacing a discrete maximum by a supremum over real numbers: Applying Lemma E.1.2 in connection with the triangle inequality for the Euclidean norm yields

$$\sup_{r \in [G, n-G]} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(r+G) - 2\boldsymbol{W}(r) + \boldsymbol{W}(r-G) \right\| - \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) \right\|$$
$$= \sup_{t \in [1, \frac{n}{G}-1]} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(tG+G) - 2\boldsymbol{W}(tG) + \boldsymbol{W}(tG-G) \right\| - \sup_{t \in [1, \frac{n}{G}-1]} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(\lfloor tG \rfloor + G) - 2\boldsymbol{W}(\lfloor tG \rfloor) + \boldsymbol{W}(\lfloor tG \rfloor - G) \right\|$$
$$\le \frac{1}{\sqrt{2G}} \sup_{t \in [1, \frac{n}{G}-1]} \Big\| \boldsymbol{W}(tG+G) - 2\boldsymbol{W}(tG) + \boldsymbol{W}(tG-G) - \boldsymbol{W}(\lfloor tG \rfloor + G) + 2\boldsymbol{W}(\lfloor tG \rfloor) - \boldsymbol{W}(\lfloor tG \rfloor - G) \Big\|$$

$$
\begin{aligned}
\leq & \frac{1}{\sqrt{2G}} \left(\sup_{t \in [1, \frac{n}{G}-1]} \| \boldsymbol{W}(tG+G) - \boldsymbol{W}(\lfloor tG \rfloor + G) \| + 2 \sup_{t \in [1, \frac{n}{G}-1]} \| \boldsymbol{W}(tG) - \boldsymbol{W}(\lfloor tG \rfloor) \| \right. \\
& \left. + \sup_{t \in [1, \frac{n}{G}-1]} \| \boldsymbol{W}(tG-G) - \boldsymbol{W}(\lfloor tG \rfloor - G) \| \right) \\
\leq & \frac{4}{\sqrt{2G}} \sup_{0 \leq t \leq n-1} \sup_{0 \leq s \leq 1} \| \boldsymbol{W}(t+s) - \boldsymbol{W}(t) \| \\
= & \frac{4}{\sqrt{2G}} O\left(\sqrt{\log(n)} \right) = \frac{4}{\sqrt{2G}} O_P\left(n^{1/(2+\nu)} \right) = o_P\left((a(n/G))^{-1} \right),
\end{aligned}
$$

where the last line follows from Lemma E.2.2 together with Assumption A.1.1. Hence, we obtain

$$
\begin{aligned}
& a(n/G) \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \| \boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) \| - b(n/G) \\
& = a(n/G) \sup_{r \in [G, n-G]} \frac{1}{\sqrt{2G}} \| \boldsymbol{W}(r+G) - 2\boldsymbol{W}(r) + \boldsymbol{W}(r-G) \| - b(n/G) + o_P(1).
\end{aligned}
$$

3) Adapting the supremum:
By the self-similarity of the Wiener process, e.g. described in Bauer (2001) on page 352, we obtain

$$
\begin{aligned}
& \sup_{r \in [G, n-G]} \frac{1}{\sqrt{2G}} \| \boldsymbol{W}(r+G) - 2\boldsymbol{W}(r) + \boldsymbol{W}(r-G) \| \\
& \overset{D}{=} \sup_{t \in [1, n/G-1]} \frac{1}{\sqrt{2}} \| \boldsymbol{W}(t+1) - 2\boldsymbol{W}(t) + \boldsymbol{W}(t-1) \| \\
& = \sup_{t \in [0, n/G-2]} \frac{1}{\sqrt{2}} \| \boldsymbol{W}(t+2) - 2\boldsymbol{W}(t+1) + \boldsymbol{W}(t) \| \\
& = \sup_{t \in [0, n/G-2]} \frac{1}{\sqrt{2}} \| 2\boldsymbol{W}(t+1) - \boldsymbol{W}(t) - \boldsymbol{W}(t+2) \| = \sup_{t \in [0, n/G-2]} \| \boldsymbol{Z}(t) \|,
\end{aligned}
$$

with $\{\boldsymbol{Z}(t) : t \geq 0\}$ denoting a stochastic process defined by $\boldsymbol{Z}(t) = \frac{1}{\sqrt{2}} (2\boldsymbol{W}(t+1) - \boldsymbol{W}(t) - \boldsymbol{W}(t+2))$.
Now, we want to show that the supremum over the interval $[0, n/G-2]$ can be replaced by the supremum over $[0, n/G]$ without changing the limit distribution. We receive

$$
\begin{aligned}
& \sup_{t \in [n/G-2, n/G]} \| \boldsymbol{Z}(t) \| \\
& \leq \sup_{t \in [n/G-2, n/G]} \frac{1}{\sqrt{2}} \| \boldsymbol{W}(t+1) - \boldsymbol{W}(t) \| + \sup_{t \in [n/G-2, n/G]} \frac{1}{\sqrt{2}} \| \boldsymbol{W}(t+2) - \boldsymbol{W}(t+1) \| \\
& = \sup_{t \in [n/G-2, n/G]} \frac{1}{\sqrt{2}} \| \boldsymbol{W}(t+1) - \boldsymbol{W}(t) \| + \sup_{t \in [n/G-1, n/G+1]} \frac{1}{\sqrt{2}} \| \boldsymbol{W}(t+1) - \boldsymbol{W}(t) \| \\
& \leq \sqrt{2} \sup_{n/G-2 \leq t \leq n/G+1} \sup_{0 \leq s \leq 1} \| \boldsymbol{W}(t+s) - \boldsymbol{W}(t) \| \\
& \overset{D}{=} \sqrt{2} \sup_{0 \leq t \leq 3} \sup_{0 \leq s \leq 1} \| \boldsymbol{W}(t+s) - \boldsymbol{W}(t) \|,
\end{aligned}
$$

where the last line follows from the stationarity of the increments of a Wiener process. Since a Wiener process has continuous paths with probability 1 and the intervals $[0,3]$ and $[0,1]$ are compact, we obtain

$$\sup_{0\le t\le 3}\sup_{0\le s\le 1}\|\boldsymbol{W}(t+s)-\boldsymbol{W}(t)\| = O_P(1),$$

which implies that

$$\sup_{t\in[n/G-2,n/G]}\|\boldsymbol{Z}(t)\| = O_P(1) = o_P\left(\frac{b(n/G)}{a(n/G)}\right)$$

as $\lim_{n\to\infty}\frac{a(n/G)}{b(n/G)} = 0$. On noting that

$$\sup_{t\in[0,n/G]}\|\boldsymbol{Z}(t)\| = \max\left\{\sup_{t\in[0,n/G-2]}\|\boldsymbol{Z}(t)\|, \sup_{t\in[n/G-2,n/G]}\|\boldsymbol{Z}(t)\|\right\}$$

Lemma E.2.3 can be applied to receive that

$$a(n/G)\sup_{t\in[0,n/G-2]}\|\boldsymbol{Z}(t)\| - b(n/G) \quad \text{and} \quad a(n/G)\sup_{t\in[0,n/G]}\|\boldsymbol{Z}(t)\| - b(n/G)$$

have the same limit distribution.

4) Limit distribution:
This part of the proof is similar to the proof of Theorem 2.3 in Steinebach & Eastwood (1996) on page 295.
Under the null hypothesis, the stochastic process $\{\boldsymbol{Z}(t) : t \ge 0\}$ is an $\mathbb{R}^p$-valued separable stationary Gaussian process with independent and standardized component processes. The covariance function r is identical for each component and given by

$$r(h) = \begin{cases} 1-\frac{3}{2}|h|, & \text{for } |h|\le 1 \\ \frac{1}{2}|h|-1, & \text{for } 1<|h|\le 2 \\ 0, & \text{else} \end{cases},$$

which has been shown in Lemma E.2.5. Finally, Lemma E.2.4 can be used to determine the limit distribution where we choose $C=\frac{3}{2}$ and $m=n/G$. This completes the proof of part (a).

(b) For proving b), we replace $\widehat{\boldsymbol{\theta}}_n$ by $\widetilde{\boldsymbol{\theta}}$. We obtain

$$\begin{aligned}
&\left|\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}\right\| - \max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\|\right| \\
&\le \frac{1}{\sqrt{2G}}\max_{G\le k\le n-G}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\| \\
&\le \left\|\boldsymbol{\Sigma}^{-1/2}\right\|_F\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\| = o_P\left((a(n/G))^{-1}\right),
\end{aligned}$$

where the last line follows from Lemma E.1.5 and Assumption A.1.4. This yields

$$a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G)$$
$$= a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} \right\| - b(n/G) + o_P(1).$$

Thus, $\widehat{\boldsymbol{\theta}}_n$ can be replaced by $\widetilde{\boldsymbol{\theta}}$ in the statistic without changing the limit distribution, i.e.

$$a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} \right\| - b(n/G) \xrightarrow{D} E.$$

(c) The convergence in distribution of part (a) leads to

$$a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G) = O_P(1),$$

implying that

$$\max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = O_P(\sqrt{\log(n/G)}).$$

Hence, with Lemma E.1.5 we get

$$\max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|$$
$$\le \left\| \boldsymbol{\Sigma}^{1/2} \right\|_F \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = O_P(\sqrt{\log(n/G)}).$$

Moreover, Lemma E.1.5 combined with Assumption A.1.5 can be used to obtain

$$\left| \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \right|$$
$$\le \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \left(\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}^{-1/2} \right) \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|$$
$$\le \max_{G\le k\le n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}^{-1/2} \right\|_F \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\|$$
$$= o_P\left((\log(n/G))^{-1} \right) O_P(\sqrt{\log(n/G)}) = o_P\left((a(n/G))^{-1} \right)$$

and consequently

$$a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G)$$
$$= a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G) + o_P(1).$$

Thus, we receive

$$a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G) \xrightarrow{D} E.$$

By applying the result of part (b) instead of (a) it can be shown in an analogous manner that

$$a(n/G) \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} \right\| - b(n/G) \xrightarrow{D} E.$$

□

2.1.2. Asymptotic Power of the MOSUM-based Tests

Under the alternative we examine scenarios, in which the parameter vector $\boldsymbol{\theta}$ changes multiple times in the considered time period as in (1.1), specified by the following assumptions.

A.2.1 *Let q be the number of change points, occuring in the time period, which is unknown but fixed. Furthermore, let $k_{1,n} < \ldots < k_{q,n}$ be the change points depending on the sample size n in the following way: $k_{j,n} = \lfloor \lambda_j n \rfloor$ with λ_j as rescaled change point being a constant but unknown value in $(0,1)$, for $j = 1, \ldots, q$.*

More precisely, at each change point $k_{j,n}$ the parameter vector changes from $\boldsymbol{\theta}_j$ to $\boldsymbol{\theta}_{j+1}$. Besides, note that Assumptions A.1.1 and A.2.1 ensure that the following statement holds for the minimal distance between two neighboring structural breaks:

$$\min_{1\le j\le q+1} |k_{j,n} - k_{j-1,n}| > 2G, \text{ for } n \text{ large},$$

with $k_{0,n} = 0$ and $k_{q+1,n} = n$. We will use this in the proofs without mentioning again. Moreover, we consider the following assumption.

A.2.2 *Let $\{\mathbb{X}_i : i \ge 1\}$ be a piecewise stationary series such that*

$$\mathbb{X}_i = \begin{cases} \mathbb{X}_i^{(1)}, & \text{if } 1 \le i \le k_{1,n} \\ \mathbb{X}_i^{(2)}, & \text{if } k_{1,n} < i \le k_{2,n} \\ \vdots & \\ \mathbb{X}_i^{(q+1)}, & \text{if } k_{q,n} < i \le n \end{cases},$$

where $\{\mathbb{X}_i^{(j)} : i \ge 1\}$ is stationary following a distribution determined by $\boldsymbol{\theta}_j$, for $j = 1, \ldots, q+1$, in a correctly specified model. Under misspecification let $\boldsymbol{\theta}_j$ be the best approximating parameter for $\{\mathbb{X}_i^{(j)} : i \ge 1\}$ in the sense of $E\left(\boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right) = \mathbf{0}$. Furthermore, we assume that the stationary sequence $\{\boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) : i \ge 1\}$ has a positive definite long-run covariance matrix $\boldsymbol{\Sigma}_{(j)}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}_{(j)}$, for all $j = 1, \ldots, q+1$.

The stationarity of the transformed sequences $\{\boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\}$ follows immediately from the stationarity of $\{\mathbb{X}_i^{(j)}\}$ and the measurability of $\boldsymbol{H}$ with respect to $\mathbb{X}_i$. Similar to the null hypothesis, replacing the partial sum processes of the transformed sequences by Wiener processes will be an essential part of the proofs and therefore a strong invariance principle, as described in the next assumption, is needed.

A.2.3 *Let $\boldsymbol{S}(j,k,\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{k} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})$ fulfill a strong invariance principle for all $j = 1, \ldots, q+1$. So possibly after changing the probability space there exists a p-dimensional standard Wiener process $\{\boldsymbol{W}(k) : k \ge 0\}$ with identity matrix $\boldsymbol{I}_p$ as covariance matrix and $\nu > 0$ such that*

$$\left\| \boldsymbol{\Sigma}_{(j)}^{-1/2} \left(\boldsymbol{S}(j,k,\widetilde{\boldsymbol{\theta}}) - E(\boldsymbol{S}(j,k,\widetilde{\boldsymbol{\theta}})) \right) - \boldsymbol{W}(k) \right\| = O(k^{1/(2+\nu)}) \ \ a.s., \quad k \to \infty.$$

Furthermore, using an estimator $\widehat{\boldsymbol{\theta}}_n$ instead of $\widetilde{\boldsymbol{\theta}}$ in the statistic requires the following condition to hold.

A.2.4 *Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ be a sequence of estimators fulfilling, for some $\widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$,*

$$\max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = O_P\left(\sqrt{\log(n/G)}\right).$$

As shown in Section 2.3 below, Assumption A.2.4 is satisfied by i.i.d. or stationary and strongly mixing time series under some moment conditions if the estimating function $\boldsymbol{H}$ is twice continuously differentiable on a compact parameter space $\boldsymbol{\Theta}$ and if the estimator sequence is $\sqrt{n}$-consistent for $\widetilde{\boldsymbol{\theta}}$.
The set $A_{n,G}$, which is defined by

$$A_{n,G} := \{k \in \{G, \dots, n-G\} : |k - \lfloor \lambda_j n \rfloor| \ge G \;\forall\, j \in \{1, \dots, q\}\}, \tag{2.2}$$

will often be considered in the proofs. It contains all time points which do not lie in a G-environment of any change such that the statistic is not contaminated by changes and behaves like under the null hypothesis in these points.
Moreover, replacing the long-run covariance matrix $\boldsymbol{\Sigma}_k$ by an appropriate estimator is an important issue here as well. The following assumptions are needed:

A.2.5 *The estimator $\widehat{\boldsymbol{\Sigma}}_{k,n}$ of the long-run covariance matrix $\boldsymbol{\Sigma}_k$ is positive definite and satisfies*

(a) $\max_{G\le k\le n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \right\|_F = O_P(1)$

(b) $\max_{k\in A_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_k^{-1/2} \right\|_F = o_P\left(\log(n/G)^{-1}\right)$ *with $A_{n,G}$ as in (2.2),*

(c) $\max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F = o_P(1)$, *where*

$$B_{n,G} := \{k \in \{G, \dots, n-G\} : \exists\, j \in \{1, \dots, q\} : |k - k_{j,n}| \le G\} \tag{2.3}$$

and $\{\boldsymbol{\Sigma}_{A,k}\}$ is a sequence of positive definite matrices fulfilling $\sup_n \sup_{k\in B_{n,G}} \|\boldsymbol{\Sigma}_{A,k}\|_F < \infty$.

In Assumption A.2.5 (b) another important set of time points has been introduced: $B_{n,G}$ defined in (2.3). It contains all time points lying in a G-environment of a change point so that the statistic is contaminated by a change in these points.
Furthermore, we have to take into consideration that $\boldsymbol{\Sigma}_k$ changes with k under the alternative. By Assumption A.2.2 we receive

$$\boldsymbol{\Sigma}_k = \begin{cases} \boldsymbol{\Sigma}_{(1)}, & \text{if } 1 \le k \le k_{1,n} \\ \boldsymbol{\Sigma}_{(2)}, & \text{if } k_{1,n} < k \le k_{2,n} \\ \vdots & \\ \boldsymbol{\Sigma}_{(q+1)}, & \text{if } k_{q,n} < k \le n \end{cases}.$$

The main goals of this subsection are constructing and investigating a test for the following test problem:

$$\boldsymbol{H_0}\colon\ q = 0, \text{ i.e. no structural break occurs}$$
$$\text{versus}$$
$$\boldsymbol{H_1}\colon\ q > 0, \text{ i.e. at least one structural break occurs}$$

Applying Theorem 2.1.1 yields

$$P\left(a(n/G)T_n(G,\widetilde{\boldsymbol{\theta}}) - b(n/G) > c_\alpha\right) \to \alpha, \text{ under } H_0,$$

where $c_\alpha := -\log\log\frac{1}{\sqrt{1-\alpha}}$ denotes the $(1-\alpha)$-quantile of a Gumbel distribution as in Theorem 2.1.1. Hence, an **asymptotic level $\boldsymbol{\alpha}$ test** for the test problem is given by:

$$\begin{aligned}&\text{Reject } H_0 \text{ if } T_n(G,\widetilde{\boldsymbol{\theta}}) > D_n(G,\alpha),\\ &\text{with } D_n(G,\alpha) = \frac{b(n/G)+c_\alpha}{a(n/G)}.\end{aligned}$$

As already mentioned in the introduction, by using the MOSUM statistic based on estimating functions we convert a general multiple parameter change problem to a multiple mean change problem of the estimating function. Hence, by applying this statistic we can only detect changes in the parameter vector causing changes in the expectation of the estimating function. To obtain consistency of the test it is sufficient to assume:

A.2.6 *For at least one $j \in \{1,\dots,q\}$ it holds that*

$$E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\widetilde{\boldsymbol{\theta}})\right) \neq E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\widetilde{\boldsymbol{\theta}})\right).$$

This assumption says that the signal part of the statistic, which can be separated from the noise, is strictly positive in a G-environment of at least one change point.
Beyond that, decomposing the MOSUM statistic into noise and signal and investigating the behavior of these two parts under alternatives will be an essential part in all of the proofs. By a simple zero expansion we get the decomposition of $\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}$ into noise and signal:

$$\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} = \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) + E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right).$$

Whereas the signal is constant equal to zero under the null hypothesis, it behaves differently under alternative which is shown in the following lemma.

Lemma 2.1.2. *Let Assumptions A.2.1 and A.2.2 hold, then, for large n,*

$$\begin{aligned}E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) &= \sum_{i=k+1}^{k+G} E\left(\boldsymbol{H}(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})\right) - \sum_{i=k-G+1}^{k} E\left(\boldsymbol{H}(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})\right)\\ &= \begin{cases} \boldsymbol{0}, & \text{if } k \in A_{n,G} \text{ as in 2.2}\\ (G-|k-k_{j,n}|)\boldsymbol{d}_j, & \text{if } k \in B_{n,G} \text{ as in 2.3}\end{cases},\end{aligned}$$

with $\boldsymbol{d}_j = E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\widetilde{\boldsymbol{\theta}})\right) - E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\widetilde{\boldsymbol{\theta}})\right)$ for $j = j(k)$ with $j(k)$ being the index of the closest change point to k.

Proof. At first we consider all time points $k \in A_{n,G}$ not lying in a G-environment of any change. We set $\lambda_0 = 0$ and $\lambda_{q+1} = 1$. On noting that there exists exactly one $j^* \in \{1, \ldots, q+1\}$ such that $k_{j^*-1,n} < k < k_{j^*,n}$, we obtain

$$E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) = \sum_{i=k+1}^{k+G} E\left(\boldsymbol{H}(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}})\right) - \sum_{i=k-G+1}^{k} E\left(\boldsymbol{H}(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}})\right) = \boldsymbol{0},$$

which follows from the stationarity of the transformed sequence $\{\boldsymbol{H}\left(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}}\right) : i \geq 1\}$ given by Assumption A.2.2.
Now, we focus on time points $k \in B_{n,G}$. Since the Assumptions A.1.1 and A.2.1 guarantee that the minimum distance between two adjacent structural breaks is greater than $2G$ for large sample sizes n the function $j(k)$ is well defined on $B_{n,G}$ in this sense. For $k \in B_{n,G}$ with $k_{j,n} < k < k_{j,n} + G$, we get by Assumption A.2.2

$$\begin{aligned}
&E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\\
&= \sum_{i=k+1}^{k+G} E\left(\boldsymbol{H}(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right) - \sum_{i=k-G+1}^{k_{j,n}} E\left(\boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right) - \sum_{i=k_{j,n}+1}^{k} E\left(\boldsymbol{H}(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right)\\
&= (G - (k - k_{j,n}))\, E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right) - (G - (k - k_{j,n}))\, E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right)\\
&= (G - |k - k_{j,n}|)\, \boldsymbol{d}_j.
\end{aligned}$$

Similar arguments can be used for $k_{j,n} - G < k \leq k_{j,n}$ □

In Subsection 2.1.3 below, we will consider the signal part of the score-type statistic in more detail and illustrate possible difficulties by some examples.
Before we are able to derive consistency for the test, we have to examine the behavior of the noise under the alternative. Lemma 2.1.4 gives an approximation of the noise, i.e. it shows in what range the statistic fluctuates asymptotically around the signal. In order to prove this result the following lemma is needed.

Lemma 2.1.3. *Let Assumptions A.1.1 on the bandwidth, A.2.1, A.2.2 and A.2.3 hold for some $\widetilde{\boldsymbol{\theta}}$. Then,*

$$a(n/G) \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\| - b(n/G) \xrightarrow{D} E,$$

with E as Gumbel distributed random variable and $a(x)$ and $b(x)$ as in Theorem 2.1.1.

Proof. We consider the sets $A_{n,G}$ and $B_{n,G}$, given in (2.2) and (2.3), separately. The set $B_{n,G}$ is further subdivided into

$$B_{n,G}^{(1)} := \{k \in B_{n,G} : k_{j,n} - G < k < k_{j,n}, j \in \{1, \ldots, q\}\}$$

and

$$B_{n,G}^{(2)} := \{k \in B_{n,G} : k_{j,n} \leq k < k_{j,n} + G, j \in \{1, \ldots, q\}\}.$$

We start with the set $B^{(2)}_{n,G}$. Let $j = j(k)$ with $j(k)$ as in Lemma 2.1.2 which is well defined on the G-environments of the change points for large n by Assumptions A.1.1 and A.2.1. Furthermore, on noting that $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{(j+1)}$ and

$$\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) = \sum_{i=k+1}^{k+G} \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k_{j,n}+1}^{k} \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}})$$

hold for all $k \in B^{(2)}_{n,G}$, applying Lemma E.1.5 in combination with Assumptions A.2.2 and A.2.3 yields

$$\begin{aligned}
&\Bigg|\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\| \\
&- \left\|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k_{j,n}) - \boldsymbol{\Sigma}_{(j+1)}^{-1/2}\boldsymbol{\Sigma}_{(j)}^{1/2}\left(\boldsymbol{W}(k_{j,n}) - \boldsymbol{W}(k-G)\right)\right\|\Bigg| \\
&\leq \left\|\boldsymbol{\Sigma}_{(j+1)}^{-1/2}\sum_{i=1}^{k+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k+G)\right\| + 2\left\|\boldsymbol{\Sigma}_{(j+1)}^{-1/2}\sum_{i=1}^{k}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k)\right\| \\
&\quad + \left\|\boldsymbol{\Sigma}_{(j+1)}^{-1/2}\sum_{i=1}^{k_{j,n}}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k_{j,n})\right\| \\
&\quad + \left\|\boldsymbol{\Sigma}_{(j+1)}^{-1/2}\boldsymbol{\Sigma}_{(j)}^{1/2}\right\|_F\left\|\boldsymbol{\Sigma}_{(j)}^{-1/2}\sum_{i=1}^{k_{j,n}}\boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k_{j,n})\right\| \\
&\quad + \left\|\boldsymbol{\Sigma}_{(j+1)}^{-1/2}\boldsymbol{\Sigma}_{(j)}^{1/2}\right\|_F\left\|\boldsymbol{\Sigma}_{(j)}^{-1/2}\sum_{i=1}^{k-G}\boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k-G)\right\| \\
&= O((k+G)^{1/(2+\nu)}) = O(n^{1/(2+\nu)}) \;\; a.s. \text{ uniformly in } k \in B^{(2)}_{n,G}.
\end{aligned}$$

Similar arguments can be used to receive

$$\begin{aligned}
&\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\| \\
&= \left\|\boldsymbol{W}(k-G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k_{j,n}) + \boldsymbol{\Sigma}_{(j)}^{-1/2}\boldsymbol{\Sigma}_{(j+1)}^{1/2}\left(\boldsymbol{W}(k+G) - \boldsymbol{W}(k_{j,n})\right)\right\| \\
&\qquad + O\left(n^{1/(2+\nu)}\right) \;\; a.s. \;\; \text{uniformly in } k \in B^{(1)}_{n,G}.
\end{aligned}$$

Moreover, note that for each $k \in A_{n,G}$ there exists exactly one $j^* \in \{1, \ldots, q+1\}$ with $k_{j^*-1,n} < k < k_{j^*,n}$ so that

$$\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) = \sum_{i=k+1}^{k+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k}\boldsymbol{H}_0(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}}).$$

Thus, by using Assumption A.2.3 we get

$$\begin{aligned}
&\left|\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\| - \left\|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\right\|\right| \\
&\leq \left\|\boldsymbol{\Sigma}_{(j^*)}^{-1/2}\sum_{i=1}^{k+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k+G)\right\| + 2\left\|\boldsymbol{\Sigma}_{(j^*)}^{-1/2}\sum_{i=1}^{k}\boldsymbol{H}_0(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k)\right\| \\
&\quad + \left\|\boldsymbol{\Sigma}_{(j^*)}^{-1/2}\sum_{i=1}^{k+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j^*)}, \widetilde{\boldsymbol{\theta}}) - \boldsymbol{W}(k+G)\right\|
\end{aligned}$$

$$= O\left(n^{1/(2+\nu)}\right) \quad a.s. \quad \text{uniformly in } k \in A_{n,G}.$$

Thus, the results above together with Assumption A.1.1 lead to

$$\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \left(\boldsymbol{A}_{\tilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\tilde{\boldsymbol{\theta}},k}\right)\right)\right\| \tag{2.4}$$
$$= \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) + \boldsymbol{R}_{G,n}(k) \right\| + o_P\left(a(n/G)^{-1}\right)$$

with

$$\boldsymbol{R}_{G,n}(k) = \begin{cases} \left(\boldsymbol{\Sigma}_{(j)}^{-1/2}\boldsymbol{\Sigma}_{(j+1)}^{1/2} - \boldsymbol{I}_p\right)\left(\boldsymbol{W}(k+G) - \boldsymbol{W}(k_{j,n})\right), & \text{if } k_{j,n} - G < k < k_{j,n} \\ \left(\boldsymbol{I}_p - \boldsymbol{\Sigma}_{(j+1)}^{-1/2}\boldsymbol{\Sigma}_{(j)}^{1/2}\right)\left(\boldsymbol{W}(k_{j,n}) - \boldsymbol{W}(k-G)\right), & \text{if } k_{j,n} \le k < k_{j,n} + G \\ \boldsymbol{0}, & \text{if } k \in A_{n,G} \end{cases}.$$

This implies

$$a(n/G) \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \left(\boldsymbol{A}_{\tilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\tilde{\boldsymbol{\theta}},k}\right)\right)\right\| - b(n/G)$$
$$= a(n/G) \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) + \boldsymbol{R}_{G,n}(k) \right\|$$
$$- b(n/G) + o_P(1).$$

Hence, it is sufficient to consider the limit distribution of the expression on the right side of the equation above.
By investigating the behavior of the remainder term $\boldsymbol{R}_{G,n}(k)$ in the G-environment of a change point, the self-similarity of a Wiener process and the Markov property of increments of a Wiener process can be applied as follows

$$\max_{k_{j,n}-G<k<k_{j,n}+G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{R}_{G,n}(k) \right\|$$
$$\le \left(\left\| \boldsymbol{I}_p - \boldsymbol{\Sigma}_{(j+1)}^{-1/2}\boldsymbol{\Sigma}_{(j)}^{1/2} \right\|_F + \left\| \boldsymbol{I}_p - \boldsymbol{\Sigma}_{(j)}^{-1/2}\boldsymbol{\Sigma}_{(j+1)}^{1/2} \right\|_F \right)$$
$$\sup_{k_{j,n}-G \le t \le k_{j,n}} \sup_{0 \le s \le G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{W}(t+s) - \boldsymbol{W}(t) \right\|$$
$$\overset{D}{=} \left(\left\| \boldsymbol{I}_p - \boldsymbol{\Sigma}_{(j+1)}^{-1/2}\boldsymbol{\Sigma}_{(j)}^{1/2} \right\|_F + \left\| \boldsymbol{I}_p - \boldsymbol{\Sigma}_{(j)}^{-1/2}\boldsymbol{\Sigma}_{(j+1)}^{1/2} \right\|_F \right)$$
$$\sup_{\frac{k_{j,n}}{G}-1 \le t \le \frac{k_{j,n}}{G}} \sup_{0 \le s \le 1} \frac{1}{\sqrt{2}} \left\| \boldsymbol{W}(t+s) - \boldsymbol{W}(t) \right\|$$
$$\overset{D}{=} \left(\left\| \boldsymbol{I}_p - \boldsymbol{\Sigma}_{(j+1)}^{-1/2}\boldsymbol{\Sigma}_{(j)}^{1/2} \right\|_F + \left\| \boldsymbol{I}_p - \boldsymbol{\Sigma}_{(j)}^{-1/2}\boldsymbol{\Sigma}_{(j+1)}^{1/2} \right\|_F \right)$$
$$\sup_{0 \le t \le 1} \sup_{0 \le s \le 1} \frac{1}{\sqrt{2}} \left\| \boldsymbol{W}(t+s) - \boldsymbol{W}(t) \right\|$$
$$= O_P(1),$$

where the last line follows from the almost sure continuity of paths of a Wiener process and the compactness of the considered interval $[0, 1]$. Hence, since the number of change points q is finite we receive

$$\max_{k \in B_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{R}_{G,n}(k) \right\| = \max_{1 \le l \le q} \max_{k_{l,n}-G<k<k_{l,n}+G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{R}_{G,n}(k) \right\| = O_P(1).$$

Furthermore, similar arguments can be used to show that

$$\begin{aligned}
&\max_{k_{j,n}-G<k<k_{j,n}+G} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\| \\
&\leq \sup_{k_{j,n}-2G\leq t\leq k_{j,n}+G} \sup_{0\leq s\leq G} \frac{\sqrt{2}}{\sqrt{G}} \|\boldsymbol{W}(t+s) - \boldsymbol{W}(t)\| \\
&\overset{D}{=} \sup_{\frac{k_{j,n}}{G}-2\leq t\leq \frac{k_{j,n}}{G}+1} \sup_{0\leq s\leq 1} \sqrt{2} \|\boldsymbol{W}(t+s) - \boldsymbol{W}(t)\| \\
&\overset{D}{=} \sup_{0\leq t\leq 3} \sup_{0\leq s\leq 1} \sqrt{2} \|\boldsymbol{W}(t+s) - \boldsymbol{W}(t)\| = O_P(1)
\end{aligned}$$

and thus

$$\begin{aligned}
&\max_{k\in B_{n,G}} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\| \\
&= \max_{1\leq l\leq q} \max_{k_{l,n}-G<k<k_{l,n}+G} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\| \\
&= O_P(1).
\end{aligned}$$

On noting that $\lim_{n\to\infty} \frac{a(n/G)}{b(n/G)} = 0$, we obtain

$$\max_{k\in B_{n,G}} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\| = o_P\left(\frac{b(n/G)}{a(n/G)}\right) \tag{2.5}$$

and

$$\begin{aligned}
&\max_{k\in B_{n,G}} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) + \boldsymbol{R}_{G,n}(k)\| \qquad (2.6)\\
&\leq \max_{k\in B_{n,G}} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\| + \max_{k\in B_{n,G}} \frac{1}{\sqrt{2G}} \|\boldsymbol{R}_{G,n}(k)\| \\
&= o_P\left(\frac{b(n/G)}{a(n/G)}\right).
\end{aligned}$$

With $\{G,\dots,n-G\} = A_{n,G} + B_{n,G}$ the maximum over all time points can be regarded as the maximum over the two maxima of the subsets, i.e. $\max_{k\in\{G,\dots,n-G\}} a_k = \max(\max_{k\in A_{n,G}} a_k, \max_{k\in B_{n,G}} a_k)$ holds for each sequence $\{a_k\}_{k\in\mathbb{N}}$. Hence, applying Lemma E.2.3 and (2.6) yields that

$$a(n/G) \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G) + \boldsymbol{R}_{G,n}(k)\| - b(n/G)$$

has the same limit distribution as

$$a(n/G) \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\| - b(n/G) \tag{2.7}$$

since $\boldsymbol{R}_{G,n}(k) = \boldsymbol{0}$ for all $k \in A_{n,G}$. By using Lemma E.2.3 together with (2.5) we obtain that the limit distribution of (2.7) coincides with the limit distribution of

$$a(n/G) \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \|\boldsymbol{W}(k+G) - 2\boldsymbol{W}(k) + \boldsymbol{W}(k-G)\| - b(n/G)$$

which is asymptotically Gumbel distributed as shown in the proof of Theorem 2.1.1. □

Lemma 2.1.4. *Let Assumptions A.1.1, A.2.1, A.2.2 and A.2.3 hold for some* $\widetilde{\boldsymbol{\theta}}$.

(a) Then

$$\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|=O_P\left(\sqrt{\log(n/G)}\right).$$

(b) Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ *be a sequence of estimators fulfilling Assumption A.2.4 for* $\widetilde{\boldsymbol{\theta}}$*. Then,*

$$\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|=O_P\left(\sqrt{\log(n/G)}\right).$$

(c) Furthermore, the covariance matrix $\boldsymbol{\Sigma}_k$ *can be replaced by an estimator* $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *satisfying Assumption A.2.5 (a) without changing the results of part (a) and (b).*

Proof. (a) Applying Lemma 2.1.3 yields

$$a(n/G)\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|-b(n/G)=O_P(1),$$

which implies that

$$\begin{aligned}\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|&=\frac{O_P(1)+b(n/G)}{a(n/G)}\\&=O_P\left(\sqrt{\log(n/G)}\right).\end{aligned}$$

(b) On noting that

$$\max_{G\le k\le n-G}\left\|\boldsymbol{\Sigma}_k^{-1/2}\right\|_F=\max_{l\in\{1,\ldots,q+1\}}\left\|\boldsymbol{\Sigma}_{(l)}^{-1/2}\right\|_F=O(1),$$

Assumption A.2.4 in connection with Lemma E.1.5 lead to

$$\begin{aligned}&\left|\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|\right.\\&\quad\left.-\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|\right|\\&\le\max_{G\le k\le n-G}\left\|\boldsymbol{\Sigma}_k^{-1/2}\right\|_F\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\|=O_P\left(\sqrt{\log(n/G)}\right).\end{aligned}$$

Thus, together with part (a) we obtain

$$\begin{aligned}&\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|\\&=\max_{G\le k\le n-G}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|+O_P\left(\sqrt{\log(n/G)}\right)\\&=O_P\left(\sqrt{\log(n/G)}\right).\end{aligned}$$

(c) With the result of part (a) and since

$$\max_{G \le k \le n-G} \left\| \boldsymbol{\Sigma}_k^{1/2} \right\|_F = \max_{l \in \{1,\dots,q+1\}} \left\| \boldsymbol{\Sigma}_{(l)}^{1/2} \right\|_F = O(1),$$

we receive

$$\begin{aligned}
&\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) \right\| \\
&\le \max_{G \le k \le n-G} \left\| \boldsymbol{\Sigma}_k^{1/2} \right\|_F \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right) \right\| \\
&= O_P\left(\sqrt{\log(n/G)}\right).
\end{aligned}$$

Hence, by combining Assumption A.2.5 (a) and Lemma E.1.5, we can conclude that

$$\begin{aligned}
&\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right) \right\| \\
&\le \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) \right\| \max_{G \le k \le n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \right\|_F \\
&= O_P\left(\sqrt{\log(n/G)}\right) O_P(1) = O_P\left(\sqrt{\log(n/G)}\right).
\end{aligned}$$

Furthermore, in an analogous manner we get

$$\max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right) \right\| = O_P\left(\sqrt{\log(n/G)}\right).$$

□

Now we are able to show that the constructed test has asymptotic power one, i.e. the probability that the test rejects the null hypothesis under the alternative converges to one as n goes to infinity.

Theorem 2.1.5. *Let Assumptions A.1.1, A.2.1, A.2.2, A.2.3 and A.2.6 hold for some* $\widetilde{\boldsymbol{\theta}}$.

(a) Then, under H_1*, we obtain for any* $z \in \mathbb{R}$

$$\lim_{n\to\infty} P(a(n/G)T_n(G, \widetilde{\boldsymbol{\theta}}) - b(n/G) \ge z) = 1.$$

(b) Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ *be a sequence of estimators fulfilling Assumption A.2.4 for* $\widetilde{\boldsymbol{\theta}}$*. Then under* H_1,

$$\lim_{n\to\infty} P(a(n/G)T_n(G, \widehat{\boldsymbol{\theta}}_n) - b(n/G) \ge z) = 1$$

for any $z \in \mathbb{R}$.

(c) Furthermore, the long-run covariance matrix $\boldsymbol{\Sigma}_k$ *can be replaced by an estimator* $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *satisfying Assumption A.2.5 (a) and (c) without changing the results of part (a) and (b).*

Proof. (a) As the inequality $a(n/G)T_n(G,\widetilde{\boldsymbol{\theta}}) - b(n/G) \geq z$ is equivalent to

$$T_n(G,\widetilde{\boldsymbol{\theta}}) - \frac{z + b(n/G)}{a(n/G)} \geq 0,$$

it is sufficient to show that $T_n(G,\widetilde{\boldsymbol{\theta}}) - \frac{z+b(n/G)}{a(n/G)} \xrightarrow{P} \infty$.

Note that we use $\boldsymbol{d_j} = E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right) - E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right)$ for all $j \in \{1, \ldots, q\}$. By Assumption A.2.6 there exists $j^* \in \{1, \ldots, q\}$ fulfilling $\boldsymbol{d_{j^*}} \neq \boldsymbol{0}$. Furthermore, applying the following inequality

$$\max_k \|\boldsymbol{a}_k + \boldsymbol{b}_k\| \geq \max_k |\|\boldsymbol{a}_k\| - \|\boldsymbol{b}_k\|| \geq \max_k \|\boldsymbol{a}_k\| - \max_k \|\boldsymbol{b}_k\|,$$

for any vector-valued sequences $\boldsymbol{a}_k, \boldsymbol{b}_k$, yields

$$\begin{aligned} T_n(G,\widetilde{\boldsymbol{\theta}}) &= \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\| \\ &\geq \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\| - \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\| \\ &\geq \frac{\sqrt{G}}{\sqrt{2}} \left\|\boldsymbol{\Sigma}_{(j^*)}^{-1/2} \boldsymbol{d_{j^*}}\right\| + O_P\left(\sqrt{\log(n/G)}\right), \end{aligned}$$

where the last line follows from Lemma 2.1.4 (a) and Lemma 2.1.2 for $k_{j^*,n}$. Since $\boldsymbol{d}_{j^*} \neq \boldsymbol{0}$ and $\boldsymbol{\Sigma}_{(j^*)}^{-1}$ is a positive definite matrix, we obtain

$$\left\|\boldsymbol{\Sigma}_{(j^*)}^{-1/2} \boldsymbol{d}_{j^*}\right\| = \left\|\boldsymbol{\Sigma}_{(j^*)}^{-1/2} \boldsymbol{d}_{j^*}\right\|_F = \sqrt{\boldsymbol{d}_{j^*}^T \boldsymbol{\Sigma}_{(j^*)}^{-1} \boldsymbol{d}_{j^*}} > c$$

for some $c > 0$. Thus, we receive $T_n(G,\widetilde{\boldsymbol{\theta}}) \geq \frac{\sqrt{G}}{\sqrt{2}} c + O_P\left(\sqrt{\log(n/G)}\right)$. Furthermore, on noting that $\frac{z+b(n/G)}{a(n/G)} = O(\sqrt{\log(n/G)})$ and $\sqrt{\log(n/G)} = o(\sqrt{G})$ by Assumption A.1.1, we can conclude

$$T_n(G,\widetilde{\boldsymbol{\theta}}) - \frac{z + b(n/G)}{a(n/G)} \geq \frac{\sqrt{G}}{\sqrt{2}} c + O_P\left(\sqrt{\log(n/G)}\right) = \sqrt{G}\left(\frac{c}{\sqrt{2}} + o_P(1)\right) \xrightarrow{P} \infty,$$

which implies the assertion.

(b) The assertion of part (b) can be shown in an analogous manner to part (a) by using Lemma 2.1.4 (b).

(c) Similar to (a), applying Lemma 2.1.4 (c), which requires Assumption A.2.5 (a), and Lemma 2.1.2 yields

$$\widehat{T}_n(G,\widetilde{\boldsymbol{\theta}}) = \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\| \geq \frac{\sqrt{G}}{\sqrt{2}} \left\|\widehat{\boldsymbol{\Sigma}}_{k_{j^*,n},n}^{-1/2} \boldsymbol{d}_{j^*}\right\| + O_P\left(\sqrt{\log(n/G)}\right).$$

In order to replace the estimator sequence of the covariance matrix by some deterministic matrix sequence, with Assumption A.2.5 (c) and Lemma E.1.5 we get

$$\left|\left\|\widehat{\boldsymbol{\Sigma}}_{k_{j^*,n},n}^{-1/2} \boldsymbol{d}_{j^*}\right\| - \left\|\boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1/2} \boldsymbol{d}_{j^*}\right\|\right| \leq \left\|\left(\widehat{\boldsymbol{\Sigma}}_{k_{j^*,n},n}^{-1/2} - \boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1/2}\right) \boldsymbol{d}_{j^*}\right\|$$

$$\leq \left\| \widehat{\boldsymbol{\Sigma}}_{k_{j^*,n},n}^{-1/2} - \boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1/2} \right\|_F \|\boldsymbol{d}_{j^*}\| \leq \max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F \|\boldsymbol{d}_{j^*}\| = o_P(1),$$

which implies $\left\| \widehat{\boldsymbol{\Sigma}}_{k_{j^*,n},n}^{-1/2} \boldsymbol{d}_{j^*} \right\| = \left\| \boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1/2} \boldsymbol{d}_{j^*} \right\| + o_P(1)$.
Furthermore, Assumption A.2.5 (c) shows

$$\sup_n \left\| \boldsymbol{\Sigma}_{A,k_{j^*,n}} \right\|_F < \sup_n \sup_{k\in B_{n,G}} \|\boldsymbol{\Sigma}_{A,k}\|_F < \infty.$$

Hence, Lemma E.1.10 ensures that for the sequence $\{\boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1}\}$ there exists $c > 0$ such that $\min_{l\in\{1,\dots,p\}} (\lambda_{l,n}) \geq c$ holds for all n, where $\lambda_{1,n}, \dots, \lambda_{p,n}$ are the eigenvalues of the matrix $\boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1}$. Thus, in combination with Lemma E.1.12 we obtain

$$\begin{aligned} \left\| \boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1/2} \boldsymbol{d}_{j^*} \right\| &= \left\| \boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1/2} \boldsymbol{d}_{j^*} \right\|_F = \sqrt{\operatorname{tr}\left(\boldsymbol{d}_{j^*}^T \boldsymbol{\Sigma}_{A,k_{j^*,n}}^{-1} \boldsymbol{d}_{j^*} \right)} \\ &\geq \sqrt{\min_{l\in\{1,\dots,p\}} (\lambda_{l,n})}\, \|\boldsymbol{d}_{j^*}\| \geq \sqrt{c}\, \|\boldsymbol{d}_{j^*}\| = \widetilde{c} > 0, \end{aligned} \tag{2.8}$$

since $\boldsymbol{d}_{j^*} \neq \mathbf{0}$. Finally, we can conclude

$$\widehat{T}_n(G, \widetilde{\boldsymbol{\theta}}) - \frac{z + b(n/G)}{a(n/G)} \geq \sqrt{G} \left(\frac{\widetilde{c}}{\sqrt{2}} + o_P(1) \right) \xrightarrow{P} \infty.$$

Similar arguments can be used to show that the test has still asymptotic power one if, in addition, $\widetilde{\boldsymbol{\theta}}$ is replaced by an estimator $\widehat{\boldsymbol{\theta}}_n$ satisfying Assumption A.2.4 in the statistic.

□

Remark 2.1.6. *If Assumption A.2.5 (c) is replaced by the following statement, the assertion of Theorem 2.1.5 remains true.*
Let $\widehat{\boldsymbol{\Sigma}}_{\lfloor \lambda_j n \rfloor,n} \xrightarrow{P} \boldsymbol{\Sigma}_{A,\lambda_j}$*, for* $j = 1, \dots, q$*, where* $\boldsymbol{\Sigma}_{A,\lambda_j}$ *is a positive definite matrix.*

Applying Lemma E.1.8 yields $\widehat{\boldsymbol{\Sigma}}_{k_{j^*,n},n}^{-1/2} = \widehat{\boldsymbol{\Sigma}}_{\lfloor \lambda_{j^*} n \rfloor,n}^{-1/2} \xrightarrow{P} \boldsymbol{\Sigma}_{A,\lambda_{j^*}}^{-1/2}$. Hence, we obtain

$$\left\| \widehat{\boldsymbol{\Sigma}}_{k_{j^*,n},n}^{-1/2} \boldsymbol{d}_{j^*} \right\| = \left\| \boldsymbol{\Sigma}_{A,\lambda_{j^*}}^{-1/2} \boldsymbol{d}_{j^*} \right\| + o_P(1).$$

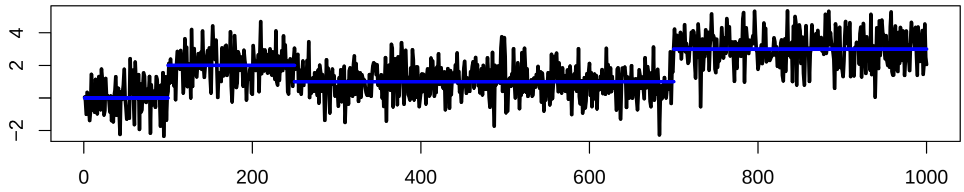
Thus, $\left\| \boldsymbol{\Sigma}_{A,\lambda_{j^*}}^{-1/2} \boldsymbol{d}_{j^*} \right\|$ can be used in the approximation above.

2.1.3. MOSUM-based Estimators

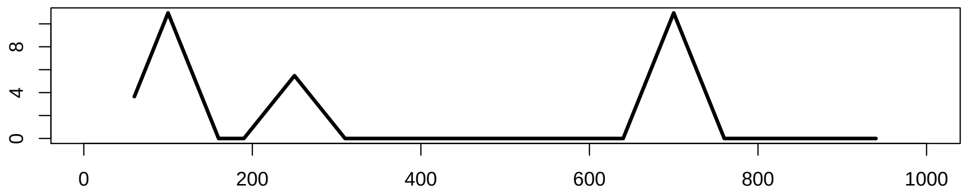
In this subsection, we focus on the estimation of the number and the locations of the changes. After introducing a MOSUM procedure based on MOSUM Score-type statistics, we show consistency of the corresponding estimators.

2.1.3.1. The MOSUM Procedure

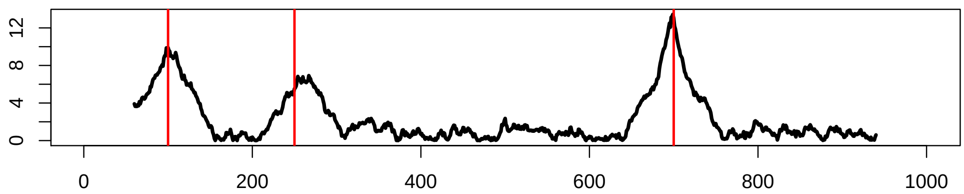
Let us at first look at an example in the classical mean change model where the MOSUM statistic investigated by Eichinger & Kirch (2018) is applied. This statistic is a special case of the MOSUM statistic based on estimating functions, as already mentioned in the introductory part, and it is used to illustrate the idea of the MOSUM procedure. However, please keep in mind that our MOSUM score-type approach is not limited to the mean change problem and can be applied to general parameter change problems. The following plot shows a time series with changes in the mean at the time points $100, 250, 700$.



In the second plot, the signal of the classical MOSUM statistic ($G = 60$) is illustrated which achieves its local maxima at the true change points and is equal to zero in time points lying far away from any change.



Unfortunately, we cannot observe or compute the signal of the statistic in practice. We only get a noisy version which is the actual statistic and shown in the graph below.



Nevertheless, the statistic performs quite well. It gets large in intervals around the true locations of the changes with local maximal points close to the true change points. However, there are also some smaller peaks, e.g. between 300 and 400, which could be interpreted as changes as well if the number and the locations of the changes were unknown. Hence, we need some kind of threshold which helps us to decide what peaks or intervals of time points should be considered in the estimation process. Theorem 2.1.1, where we have determined the limit distribution of the maximum of the statistic under the null, provides a reasonable threshold, $D_n(\alpha_n, G) := (c_{\alpha_n} + b(n/G))/a(n/G)$, as the statistic behaves like under the null hypothesis in time points which are far away from any change.
Consequently, similar to Eichinger & Kirch (2018), we propose the following **MOSUM**

procedure to determine estimators for the number and the locations of the changes:

We consider all pairs of time points $(v_{j,n}, w_{j,n})$ *with*

$$T_{k,n}(G, \widetilde{\boldsymbol{\theta}}) \geq D_n(\alpha_n, G) \ \ \textit{for} \ \ v_{j,n} \leq k \leq w_{j,n}, \tag{2.9}$$

$$T_{k,n}(G, \widetilde{\boldsymbol{\theta}}) < D_n(\alpha_n, G) \ \ \textit{for} \ \ k = v_{j,n} - 1, w_{j,n} + 1, \tag{2.10}$$

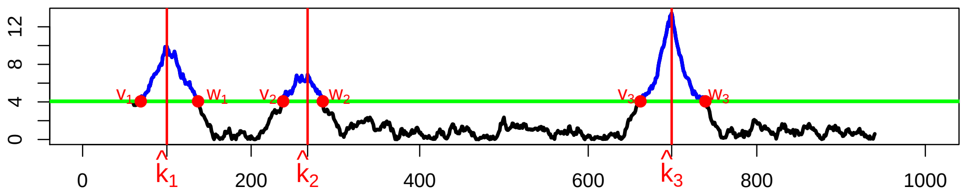
$$w_{j,n} - v_{j,n} \geq \varepsilon G \quad \textit{with} \ \ 0 < \varepsilon < 1/2 \ \ \textit{arbitrary but fixed.} \tag{2.11}$$

We take the number of these pairs as an estimator for the number of changes:

$$\widehat{q}_n = \widehat{q}_n(\widetilde{\boldsymbol{\theta}}) \mathrel{\widehat{=}} \ \textit{number of pairs} \ (v_{j,n}, w_{j,n}).$$

Furthermore, we determine the local maxima between $v_{j,n}$ *and* $w_{j,n}$*,* $j = 1, \ldots, \widehat{q}_n$*, and use them as estimators for the locations of the change points:*

$$\widehat{k}_{j,n} = \widehat{k}_{j,n}(\widetilde{\boldsymbol{\theta}}) := \underset{v_{j,n} \leq k \leq w_{j,n}}{\arg\max} \ T_{k,n}(G, \widetilde{\boldsymbol{\theta}})$$



The pairs $(v_{j,n}, w_{j,n})$, $j = 1, \ldots, \widehat{q}_n$, give start and end points of intervals on which the statistic exceeds the threshold. For this reason we call $[v_{j,n}, w_{j,n}]$, $j = 1, \ldots, \widehat{q}_n$, intervals of exceedings or exceeding intervals.
Condition (2.11) in the procedure restricts our attention to intervals of length greater than εG and prevents us from considering exceeding intervals produced by spurious local maxima in the estimating process. This is very important because the statistic is noisy and, for instance, in an environment of a change point it can actually happen that the statistic goes beyond the threshold but only for a short period and falls below it before shortly exceeding the threshold again. Hence, in this sense, Condition (2.11) avoids overestimation.
Furthermore, note that Eichinger & Kirch (2018) have also proposed an alternative version of the MOSUM procedure. Instead of Condition (2.11) they consider all intervals of exceedings and determine their local maximal points before checking whether these points are local maxima in their cG-environments as well. Consistency for the corresponding estimators can be derived by using similar arguments as for the original MOSUM procedure in the classical mean change model according to Eichinger & Kirch (2018). This will probably hold for our general setting too, but we concentrate on the above case here.

2.1.3.2. Consistency of the Estimators

In this subsection we derive consistency for the estimators of the number and the locations of the changes obtained by the MOSUM procedure which has been described

in the previous subsection. In doing so, we have to introduce further notation and assumptions.
We have already learned that a change in the parameter vector $\boldsymbol{\theta}$ can only be detected or localized by the MOSUM score-type statistic if it causes a change in the expectation of the transformed series $\boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})$. This is formalized in the following way.

A.2.7:
Let $\widetilde{Q} = \widetilde{Q}\left(\widetilde{\boldsymbol{\theta}}\right)$ be the set of indices of all rescaled change points causing a change in the expected value of the transformed series (detectable changes), i.e.

$$E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right) \neq E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right)$$

holds for all $j \in \widetilde{Q}$ and

$$E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right) = E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right)$$

for all $j \in \{1, \ldots, q\} \setminus \widetilde{Q}$.
Furthermore, let $\widetilde{q} = \widetilde{q}\left(\widetilde{\boldsymbol{\theta}}\right)$ be the number of elements of $\widetilde{Q}$ which is the number of detectable changes.

Note that the number of detectable changes does not need to coincide with the total number of changes in general. If the score-type statistic based on the Z-estimator computed on the whole sample is used in the MOSUM procedure at least one change is detectable which will be shown in Lemma 2.3.10 under some regularity conditions. The problem of detectability will be discussed in detail in Section 2.4.2.
Thus, in general we have to distinguish between detectable and non-detectable changes and therefore we define the following sets:

$$\widetilde{A}_{n,G} := \left\{k \in \{G, \ldots, n-G\} : |k - k_{j,n}| \geq G \;\forall\, j \in \widetilde{Q}\right\}, \tag{2.12}$$

$$\bar{B}_{n,G} := \left\{k \in \{G, \ldots, n-G\} : \exists\, j \in \widetilde{Q} : |k - k_{j,n}| < (1-\varepsilon)G\right\} \tag{2.13}$$

with ε as in (2.11).

The following condition is of more technical nature. Instead of using a fixed significance level, we will need a sequence of significance levels in Theorem 2.1.8 satisfying:

A.2.8 *Let the sequence of significance levels α_n fulfill*

$$\alpha_n \to 0 \quad \text{and} \quad \frac{c_{\alpha_n}}{a(n/G)\sqrt{G}} = o(1).$$

Moreover, if we want to use an estimator sequence $\widehat{\boldsymbol{\theta}}_n$ instead of $\widetilde{\boldsymbol{\theta}}$ in the statistic and show consistency for the corresponding estimators, we have to introduce additional conditions.

A.2.9:
Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ be a sequence of estimators fulfilling

(I) $\max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = o_P\left((\log(n/G))^{-1/2}\right)$,
with $A_{n,G}$ *as in* (2.2).

(II) $\max_{k\in \widetilde{A}_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = o_P\left(\sqrt{\log(n/G)}\right)$,
with $\widetilde{A}_{n,G}$ *as in* (2.12).

In the following lemma we consider the maximum of the statistic over all time points $k \in \widetilde{A}_{n,G}$, which do not lie in a G-environment of any detectable change, and derive its limit distribution under the alternative. This result will be crucial for proving consistency in Theorem 2.1.8.

Lemma 2.1.7. *Let Assumptions A.1.1, A.2.1, A.2.2,A.2.3 and A.2.7 hold for some* $\widetilde{\boldsymbol{\theta}}$.

(a) Then,

$$a(n/G) \max_{k\in \widetilde{A}_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G) \xrightarrow{D} E$$

with E *as Gumbel distributed random variable as in Theorem 2.1.1.*

(b) Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ *be a sequence of estimators fulfilling Assumption A.2.9 for* $\widetilde{\boldsymbol{\theta}}$. *Then, if* $\widetilde{\boldsymbol{\theta}}$ *is replaced by* $\widehat{\boldsymbol{\theta}}_n$ *in the statistic the result of (a) remain true.*

(c) Furthermore, the long-run covariance matrix $\boldsymbol{\Sigma}_k$ *can be replaced by an estimator* $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *satisfying the Assumptions A.2.5 (a) and (b) without changing the results of part (a) and (b).*

Proof. (a) At first, note that $A_{n,G} \subset \widetilde{A}_{n,G}$, with $A_{n,G}$ as in (2.2) denoting the set of all points not lying in a G-environment of any change, and $\widetilde{A}_{n,G} \setminus A_{n,G} \subset B_{n,G}$, with $B_{n,G}$ as in (2.3) containing the time points of the G-environments of all changes. By combining (2.4) and (2.6), which have been shown in the proof of Lemma 2.1.3, we obtain

$$\max_{k\in B_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right) \right\| = o_P\left(\frac{b(n/G)}{a(n/G)}\right).$$

Hence, since $E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right) = \boldsymbol{0}$ for all $k \in \widetilde{A}_{n,G}$ and $A_{n,G} \cup B_{n,G} = \{G, \ldots, n-G\}$, which implies that

$$\max_{G\le k\le n-G} \|\cdot\| = \max\left(\max_{A_{n,G}} \|\cdot\|, \max_{B_{n,G}} \|\cdot\|\right),$$

applying Lemma E.2.3 together with Lemma 2.1.3 yields that

$$a(n/G) \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G) \tag{2.14}$$

is asymptotically Gumbel distributed. Furthermore, with $\widetilde{A}_{n,G} \setminus A_{n,G} \subset B_{n,G}$ we get

$$\begin{aligned} &\max_{k\in\widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \\ &\leq \max_{k\in B_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\| = o_P\left(\frac{b(n/G)}{a(n/G)}\right). \end{aligned} \tag{2.15}$$

Thus, Lemma E.2.3 can be used again to conclude that

$$a(n/G) \max_{k\in\widetilde{A}_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G)$$

has the same limit distribution as (2.14) completing the proof of part (a).

(b) On noting that

$$\max_{k\in\widetilde{A}_{n,G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \right\|_F \leq \max_{l\in\{1,\ldots,q+1\}} \left\| \boldsymbol{\Sigma}_{(l)}^{-1/2} \right\|_F = O(1),$$

by using Lemma E.1.5 and Assumption A.2.9 (I) we obtain

$$\begin{aligned} &\left| \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}},k} \right\| - \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \right| \\ &\leq \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}},k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \max_{k\in A_{n,G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \right\|_F = o_P\left(a(n/G)^{-1}\right), \end{aligned}$$

implying that

$$a(n/G) \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} \right\| - b(n/G)$$

and

$$a(n/G) \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G)$$

have the same limit distribution, which is a Gumbel distribution as shown in part (a). Furthermore, Assumption A.2.9 (II) in combination with (2.15) yields

$$\begin{aligned} &\max_{k\in\widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} \right\| = \max_{k\in\widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} + \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right) \right\| \\ &\leq \max_{k\in\widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \max_{l\in\{1,\ldots,q+1\}} \left\| \boldsymbol{\Sigma}_{(l)}^{-1/2} \right\|_F \\ &\quad + \max_{k\in\widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \\ &= o_P\left(\frac{b(n/G)}{a(n/G)}\right). \end{aligned}$$

Finally, by Lemma E.2.3 we get the assertion.

(c) By using (2.14) and Lemma E.1.4 we receive

$$\begin{aligned}
&\max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{1/2} \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \\
&\leq \max_{1\leq l\leq q+1} \left\| \boldsymbol{\Sigma}_{(l)}^{1/2} \right\|_F \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = O_P\left(\sqrt{\log(n/G)}\right).
\end{aligned}$$

Thus, applying Assumption A.2.5 (b) and Lemma E.1.4 yields

$$\begin{aligned}
&\left| \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \right| \\
&\leq \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \left(\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_k^{-1/2} \right) \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \\
&\leq \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \max_{k\in A_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_k^{-1/2} \right\|_F = o_P\left(a(n/G)^{-1} \right),
\end{aligned}$$

implying that

$$a(n/G) \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G)$$

and

$$a(n/G) \max_{k\in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| - b(n/G)$$

have the same limit distribution, which is a Gumbel distribution as shown in part (a). Furthermore, by (2.15) we get

$$\begin{aligned}
&\max_{k\in \widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = \max_{k\in \widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{1/2} \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \\
&\leq \max_{1\leq l\leq q+1} \left\| \boldsymbol{\Sigma}_{(l)}^{1/2} \right\|_F \max_{k\in \widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = o_P\left(\frac{b(n/G)}{a(n/G)} \right).
\end{aligned}$$

Hence, in connection with Assumption A.2.5 (a) and Lemma E.1.5 we receive

$$\begin{aligned}
&\max_{k\in \widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \leq \max_{k\in \widetilde{A}_{n,G}\setminus A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \max_{k\in \widetilde{A}_{n,G}\setminus A_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \right\|_F \\
&= o_P\left(\frac{b(n/G)}{a(n/G)} \right) O_P(1) = o_P\left(\frac{b(n/G)}{a(n/G)} \right).
\end{aligned}$$

Finally, the assertion follows from Lemma E.2.3.

□

Now, we are ready to show that the estimator for the number of changes $\widehat{q}_n$ is consistent for the number of detectable changes $\widetilde{q}$.

Theorem 2.1.8. *Let the Assumptions A.1.1, A.2.1, A.2.2, A.2.3 and A.2.7 hold for some* $\widetilde{\boldsymbol{\theta}}$. *Furthermore, assume that the sequence* $\{\alpha_n\}_{n\in\mathbb{N}}$ *fulfills Assumption A.2.8.*

(a) Then, for any $\widetilde{\boldsymbol{\theta}}$ *with corresponding* $\widetilde{q} = \widetilde{q}(\widetilde{\boldsymbol{\theta}})$,

$$P(\widehat{q}_n(\widetilde{\boldsymbol{\theta}}) = \widetilde{q}) \to 1 \quad \textit{as } n \to \infty.$$

(b) Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ *be a sequence of estimators for* $\widetilde{\boldsymbol{\theta}}$ *fulfilling Assumption A.2.4 and A.2.9. Then, the statistic* $T_n(G,\widehat{\boldsymbol{\theta}}_n)$ *can be used in the MOSUM procedure without changing the result of part (a), i.e.*

$$P(\widehat{q}_n(\widehat{\boldsymbol{\theta}}_n)=\widetilde{q})\to 1 \quad \text{as } n\to\infty.$$

(c) Furthermore, the consistency statements of (a) and (b) remain true if the long-run covariance matrix $\boldsymbol{\Sigma}_k$ *is replaced by an estimator* $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *satisfying Assumptions A.2.5 (a), (b) and (c).*

Proof. (a) The basic idea of this proof, going back to Muhsal (2013) (Theorem 6.1.) or Eichinger & Kirch (2018) (Theorem 3.1. a), is to show that the statistic is below the threshold on $\widetilde{A}_{n,G}$ as in (2.12) while exceeding the threshold on $\bar{B}_{n,G}$ as in (2.13) with probability tending to one. Note that it is sufficient to analyse the asymptotic behavior of the statistic on these two sets since the omitted intervals are too small to be taken into account for estimation due to Condition (2.11) of the MOSUM procedure.
Using the simple inequality

$$P(A\cap B)=P(A)+P(B)-P(A\cup B)\geq P(A)+P(B)-1, \tag{2.16}$$

we obtain

$$\begin{aligned}
&P(\widehat{q}_n=\widetilde{q})\\
&\geq P\left(\max_{k\in\widetilde{A}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}})<D_n(\alpha_n,G),\ \min_{k\in\bar{B}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}})\geq D_n(\alpha_n,G)\right)\\
&\geq P\left(\max_{k\in\widetilde{A}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}})<D_n(\alpha_n,G)\right)+P\left(\min_{k\in\bar{B}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}})\geq D_n(\alpha_n,G)\right)-1.
\end{aligned}$$

Hence, it is sufficient to show that

$$\begin{aligned}
&(1)\ P\left(\max_{k\in\widetilde{A}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}})<D_n(\alpha_n,G)\right)\to 1 \ \text{ and}\\
&(2)\ P\left(\min_{k\in\bar{B}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}})\geq D_n(\alpha_n,G)\right)\to 1
\end{aligned}$$

as n goes to infinity.

Part (1):
On noting that c_{α_n} is the $(1-\alpha_n)$-quantile of the Gumbel distribution given in Theorem 2.1.1, applying Lemma 2.1.7 (a) yields

$$\begin{aligned}
&P\left(\max_{k\in\widetilde{A}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}})\geq D_n(\alpha_n,G)\right)=P\left(\max_{k\in\widetilde{A}_{n,G}}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\|\geq D_n(\alpha_n,G)\right)\\
&=P\left(a(n/G)\max_{k\in\widetilde{A}_{n,G}}\frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_k^{-1/2}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\|-b(n/G)\geq c_{\alpha_n}\right)
\end{aligned}$$

$$= \alpha_n + o(1) \to 0 \quad \text{as } n \to \infty, \tag{2.17}$$

since α_n converges to 0 by Assumption A.2.8 and as the Gumbel distribution is continuous.

Part (2):
Let $j = j(k)$ where $j(k)$ is the the index of the closest change point to k as in Lemma 2.1.2. By using Lemma 2.1.2 for $k \in \bar{B}_{n,G}$ we get

$$\left\|\boldsymbol{\Sigma}_k^{-1/2} E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\| = (G - |k - k_{j,n}|) \left\|\boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{d}_j\right\| \geq \varepsilon G \min_{k \in \bar{B}_{n,G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{d}_j\right\|,$$

which shows that the signal part of the statistic grows with a rate of G on the set $\bar{B}_{n,G}$. Furthermore, as Assumption A.2.7 ensures that $\boldsymbol{d}_l \neq \mathbf{0}$ for all $l \in \widetilde{Q}$ and as $\boldsymbol{\Sigma}_{(1)}^{-1}, \ldots, \boldsymbol{\Sigma}_{(q+1)}^{-1}$ are positive definite due to the positive definiteness of $\boldsymbol{\Sigma}_{(1)}, \ldots, \boldsymbol{\Sigma}_{(q+1)}$ we obtain

$$\min_{k \in \bar{B}_{n,G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{d}_j\right\| = \min_{k \in \bar{B}_{n,G}} \sqrt{\boldsymbol{d}_j^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{d}_j} > c \ \text{ for some } c > 0.$$

The results above in connection with Lemma 2.1.4 and the following inequality

$$\begin{aligned} &\min_{k \in \bar{B}_{n,G}} T_{k,n}(G, \widetilde{\boldsymbol{\theta}}) \\ &\geq \min_{k \in \bar{B}_{n,G}} \frac{1}{\sqrt{2G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\| - \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\|\boldsymbol{\Sigma}_k^{-1/2}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)\right\|, \end{aligned}$$

imply that

$$\begin{aligned} &P\left(\min_{k \in \bar{B}_{n,G}} T_{k,n}(G, \widetilde{\boldsymbol{\theta}}) \geq D_n(\alpha_n, G)\right) \qquad (2.18)\\ &\geq P\left(\min_{k \in \bar{B}_{n,G}} \frac{1}{\sqrt{2G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\| + O_P\left(\sqrt{\log(n/G)}\right) \geq D_n(\alpha_n, G)\right) \\ &\geq P\left(\varepsilon \sqrt{\frac{G}{2}} \min_{k \in \bar{B}_{n,G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{d}_j\right\| + O_P\left(\sqrt{\log(n/G)}\right) \geq D_n(\alpha_n, G)\right) \\ &= P\left(\varepsilon \min_{k \in \bar{B}_{n,G}} \left\|\boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{d}_j\right\| + o_P(1) \geq \frac{\sqrt{2} D_n(\alpha_n, G)}{\sqrt{G}}\right) \to 1, \end{aligned}$$

since $\frac{D_n(\alpha_n,G)}{\sqrt{G}} = \frac{b(n/G)}{a(n/G)\sqrt{G}} + \frac{c_{\alpha_n}}{a(n/G)\sqrt{G}} = o(1)$ follows from Assumptions A.1.1 and A.2.8.

(b) Part (1):
The result can be shown in an analogous manner to (a) by using Lemma 2.1.7 (b).

Part (2):
Result (b) of Lemma 2.1.4 can be used to prove part (2) similarly to (a).

(c) Part (1):
With Lemma 2.1.7 (c) the result can be shown analogously to (a).

Part (2):
With $j = j(k)$ as in Lemma 2.1.2 we receive by that lemma

$$\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\| = (G - |k - k_{j,n}|)\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\boldsymbol{d}_j\right\| \geq \varepsilon G \min_{k\in\bar{B}_{n,G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\boldsymbol{d}_j\right\|,$$

which holds for all $k \in \bar{B}_{n,G}$. Furthermore, on noting that

$$\max_{k\in\bar{B}_{n,G}} \|\boldsymbol{d}_j\| \leq \max_{l\in\{1,\ldots,q\}} \|\boldsymbol{d}_l\| = O(1),$$

by Assumption A.2.5 (c) combined with Lemma E.1.3 and Lemma E.1.5 we get

$$\begin{aligned}&\left|\min_{k\in\bar{B}_{n,G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\boldsymbol{d}_j\right\| - \min_{k\in\bar{B}_{n,G}}\left\|\boldsymbol{\Sigma}_{A,k}^{-1/2}\boldsymbol{d}_j\right\|\right| \leq \max_{k\in\bar{B}_{n,G}}\left\|\left(\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2}\right)\boldsymbol{d}_j\right\|\\&\leq \max_{k\in\bar{B}_{n,G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2}\right\|_F \max_{k\in\bar{B}_{n,G}}\|\boldsymbol{d}_j\| = o_P(1),\end{aligned}$$

implying that

$$\min_{k\in\bar{B}_{n,G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\boldsymbol{d}_j\right\| = \min_{k\in\bar{B}_{n,G}}\left\|\boldsymbol{\Sigma}_{A,k}^{-1/2}\boldsymbol{d}_j\right\| + o_P(1).$$

Applying Assumption A.2.5 (c) in combination with Lemma E.1.10 and Lemma E.1.12 as in (2.8) yields that there exists $c > 0$ such that

$$\min_{k\in\bar{B}_{n,G}}\left\|\boldsymbol{\Sigma}_{A,k}^{-1/2}\boldsymbol{d}_j\right\| \geq c + o(1).$$

Thus, in an analogous manner to (a) we receive

$$\begin{aligned}&P\left(\min_{k\in\bar{B}_{n,G}} \widehat{T}_{k,n}(G,\widetilde{\boldsymbol{\theta}}) \geq D_n(\alpha_n, G)\right)\\&= P\left(\min_{k\in\bar{B}_{n,G}} \frac{1}{\sqrt{2G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\| \geq D_n(\alpha_n, G)\right)\\&\geq P\left(\min_{k\in\bar{B}_{n,G}} \frac{1}{\sqrt{2G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\| + O_P\left(\sqrt{\log(n/G)}\right) \geq D_n(\alpha_n, G)\right)\\&\geq P\left(\varepsilon \min_{k\in\bar{B}_{n,G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\boldsymbol{d}_j\right\| + o_P(1) \geq \frac{\sqrt{2}D_n(\alpha_n, G)}{\sqrt{G}}\right)\\&= P\left(\varepsilon \min_{k\in\bar{B}_{n,G}}\left\|\boldsymbol{\Sigma}_{A,k}^{-1/2}\boldsymbol{d}_j\right\| + o_P(1) \geq \frac{\sqrt{2}D_n(\alpha_n, G)}{\sqrt{G}}\right)\\&\geq P\left(\varepsilon c + o_P(1) \geq \frac{\sqrt{2}D_n(\alpha_n, G)}{\sqrt{G}}\right) \to 1,\end{aligned}$$

since $\frac{D_n(\alpha_n,G)}{\sqrt{G}} = o(1)$ by Assumptions A.1.1 and A.2.8.

With similar arguments we obtain

$$P\left(\min_{k\in\bar{B}_{n,G}} \widehat{T}_{k,n}(G,\widehat{\boldsymbol{\theta}}_n) \geq D_n(\alpha_n,G)\right) \to 1,$$

completing the proof of this theorem.

□

Remark 2.1.9. *If $\widetilde{Q} = \{1,\ldots,q\}$, i.e. each change in the parameter vector $\boldsymbol{\theta}$ causes a mean change in the transformed sequence of the estimating function, Theorem 2.1.8 yields*

$$\lim_{n\to\infty} P\left(\widehat{q}_n = q\right) = 1.$$

Moreover, the following corollary in combination with Remark 2.1.11 proves a weak consistency statement for the change point estimators $\widehat{k}_{j,n}$.

Corollary 2.1.10. *Let the assumptions of Theorem 2.1.8 hold. Then,*

$$P\left(\max_{j\in\widetilde{Q}} \min_{1\leq l\leq\hat{q}_n} \left|\widehat{k}_{l,n} - k_{j,n}\right| < G\right) \to 1,$$

i.e. with probability tending to one every detectable change point has at least one estimator in its G-environment.

Remark 2.1.11. *By Theorem 2.1.8 there are exactly $\widetilde{q}$ change point estimators with asymptotic probability one. Since the distance between two adjacent change points is asymptotically greater than 2G an estimator can only lie in the G-environment of one change point. Thus, combining Theorem 2.1.8 and Corollary 2.1.10 yields that every detectable change point has exactly one estimator in its G-environment with probability tending to one.*

Proof of Corollary 2.1.10. On noting that

$$\left\{\min_{k\in\bar{B}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}}) \geq D_n(\alpha_n,G)\right\} \subset \left\{\max_{j\in\widetilde{Q}} \min_{1\leq l\leq\hat{q}_n} \left|\widehat{k}_{l,n} - k_{j,n}\right| < G\right\},$$

applying (2.18) yields

$$P\left(\max_{j\in\widetilde{Q}} \min_{1\leq l\leq\hat{q}_n} \left|\widehat{k}_{l,n} - k_{j,n}\right| < G\right) \geq P\left(\min_{k\in\bar{B}_{n,G}} T_{k,n}(G,\widetilde{\boldsymbol{\theta}}) \geq D_n(\alpha_n,G)\right) \to 1,$$

which shows the assertion. □

The results above do not show consistency in the classical sense since we only get a weak convergence rate depending on the bandwidth G which tends to infinity. This rate can be improved under stronger assumptions on the series which will be shown in the following subsection. Nevertheless, the corollary above enables us to construct estimators of the rescaled change points $\widehat{\lambda}_{j,n} := \frac{\widehat{k}_{j,n}}{n}$, $j\in\widetilde{Q}$, which are actually consistent for the true rescaled changes λ_j, $j\in\widetilde{Q}$, in the classical sense as shown in the following corollary.

Corollary 2.1.12. *Let the assumptions of Theorem 2.1.8 hold. Then,*

$$\max_{j\in\widetilde{Q}}\min_{1\le l\le\hat{q}_n}\left|\widehat{\lambda}_{l,n}-\lambda_j\right| = O_P\left(\frac{G}{n}\right) = o_P(1).$$

Proof. By applying Corollary 2.1.10 and on noting that

$$|k_{j,n}-\lambda_j n| = |\lfloor\lambda_j n\rfloor-\lambda_j n| \le 1,$$

we obtain

$$\max_{j\in\widetilde{Q}}\min_{1\le l\le\hat{q}_n}\left|\frac{\widehat{k}_{l,n}}{n}-\lambda_j\right| = \frac{1}{n}\max_{j\in\widetilde{Q}}\min_{1\le l\le\hat{q}_n}\left|\widehat{k}_{l,n}-\lambda_j n\right| \le \frac{1}{n}\max_{j\in\widetilde{Q}}\min_{1\le l\le\hat{q}_n}\left|\widehat{k}_{l,n}-k_{j,n}\right| + \frac{1}{n}$$
$$= O_P\left(\frac{G}{n}\right).$$

Furthermore, by Assumption A.1.1 we know that $\frac{G}{n}\to 0$. □

2.2. Convergence Rates

In the previous section we have derived consistency for the estimators of the number and the locations of the changes. These results can be improved in terms of getting better convergence rates under some stronger assumptions which are described in the following.

A.2.10 *Let the following forward and backward Hájek-Rényi-type inequalities hold for some $\gamma > 2$:*

(a) For all $j\in\{1,\dots,q+1\}$ and for any positive and non-increasing sequence $b_1 \ge b_2 \ge \dots \ge b_n > 0$ there exists a constant $B(\gamma)$ such that

$$E\left(\max_{1\le k\le n} b_k\left\|\sum_{i=1}^{k}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma} \le B(\gamma)\sum_{k=1}^{n} b_k^{\gamma}k^{\gamma/2-1}.$$

(b) For all $j\in\{1,\dots,q+1\}$ and for any positive and non-decreasing sequence $0 < a_1 \le a_2 \le \dots \le a_n$ there exists a constant $A(\gamma)$ such that

$$E\left(\max_{1\le k\le n} a_k\left\|\sum_{i=k+1}^{n}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma} \le A(\gamma)\sum_{k=1}^{n} a_k^{\gamma}(n-k)^{\gamma/2-1}.$$

Furthermore, we need an additional assumption on the estimator sequence $\widehat{\boldsymbol{\theta}}_n$ which will allow us to replace $\widetilde{\boldsymbol{\theta}}$ in the statistic without changing the theoretical results.

A.2.11 *Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ be a sequence of estimators fulfilling, for any $m\in\mathbb{N}$ and for each $j\in\{1,\dots,q+1\}$,*

(i) $\max_{1\le k\le n} \frac{1}{k} \left\| \sum_{i=m-k+1}^{m} \left(\boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widehat{\boldsymbol{\theta}}_n\right) - \boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}\right)\right)\right\| = o_P(1)$
and

(ii) $\max_{1\le k\le n} \frac{1}{k} \left\| \sum_{i=m+1}^{m+k} \left(\boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widehat{\boldsymbol{\theta}}_n\right) - \boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}\right)\right)\right\| = o_P(1)$

for some $\widetilde{\boldsymbol{\theta}}$.

In order to prove the main result in Theorem 2.2.5 we need the following auxiliary lemmata.

Lemma 2.2.1. *Let Assumptions A.2.2 and A.2.10 hold for some* $\widetilde{\boldsymbol{\theta}}$ *and let* $\{b_k\}_{k\ge 1}$ *be a positive and non-increasing sequence with* $b_1 \ge \ldots \ge b_k$. *Then, it holds for any* $1 \le l \le u$, *any* $m \in \mathbb{N}_0$, *any* $\delta > 0$ *and for each* $j \in \{1, \ldots, q+1\}$

(a)

$$\delta^\gamma P\left(\max_{l\le k\le u} b_k \left\| \sum_{i=m+1}^{m+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| \ge \delta\right) \le \widetilde{C}\left(b_l^\gamma l^{\gamma/2} + \sum_{k=l+1}^{u} b_k^\gamma k^{\gamma/2-1}\right),$$

(b)

$$\delta^\gamma P\left(\max_{l\le k\le u} b_k \left\| \sum_{i=m-k+1}^{m} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| \ge \delta\right) \le \widetilde{C}\left(b_l^\gamma l^{\gamma/2} + \sum_{k=l+1}^{u} b_k^\gamma k^{\gamma/2-1}\right),$$

where $\widetilde{C}$ *only depends on* γ *of Assumption A.2.10.*

Proof. (a) The result of this lemma can be shown similarly to Lemma 3.1 in Eichinger & Kirch (2018). On noting that for $k > l$ holds

$$\sum_{i=m+1}^{m+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) = \sum_{i=m+1}^{m+l} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) + \sum_{i=m+l+1}^{m+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}),$$

the triangle inequality and the monotonicity of the sequence yield

$$\begin{aligned} &\max_{l\le k\le u} b_k \left\| \sum_{i=m+1}^{m+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| \\ &\le b_l \left\| \sum_{i=m+1}^{m+l} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| + \max_{l< k\le u} b_k \left\| \sum_{i=m+l+1}^{m+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\|. \end{aligned}$$

Hence, by Chebychev inequality we receive

$$P\left(\max_{l\le k\le u} b_k \left\| \sum_{i=m+1}^{m+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| \ge \delta\right)$$

$$\leq P\left(b_l\left\|\sum_{i=m+1}^{m+l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{2}\right)+P\left(\max_{l<k\leq u}b_k\left\|\sum_{i=m+l+1}^{m+k}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{2}\right)$$
$$\leq E\left(b_l\left\|\sum_{i=m+1}^{m+l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}\left(\frac{\delta}{2}\right)^{-\gamma}$$
$$+E\left(\max_{l<k\leq u}b_k\left\|\sum_{i=m+l+1}^{m+k}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}\left(\frac{\delta}{2}\right)^{-\gamma}.$$

Since the series $\{\boldsymbol{H}(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\}_{i\geq 1}$ is stationary and the constant sequence $\widetilde{b}_k\equiv l$ fulfills the conditions of Assumption A.2.10 (a) we obtain

$$E\left(b_l\left\|\sum_{i=m+1}^{m+l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}=E\left(b_l\left\|\sum_{i=1}^{l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}$$
$$\leq E\left(\max_{1\leq k\leq l}b_l\left\|\sum_{i=1}^{k}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}\leq\widetilde{C}_1\sum_{k=1}^{l}b_l^{\gamma}k^{\gamma/2-1}\leq\widetilde{C}_1b_l^{\gamma}l^{\gamma/2},$$

where $\widetilde{C}_1$ is a constant only depending on γ. Furthermore, with $\widetilde{b}_k=b_{k+l}$ and an index shift to $h=k-l$ the stationarity of the series and Assumption A.2.10 (a) can be used again to get

$$E\left(\max_{l<k\leq u}b_k\left\|\sum_{i=m+l+1}^{m+k}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}$$
$$=E\left(\max_{1\leq h\leq u-l}\widetilde{b}_h\left\|\sum_{i=m+l+1}^{m+l+h}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}=E\left(\max_{1\leq h\leq u-l}\widetilde{b}_h\left\|\sum_{i=1}^{h}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}$$
$$\leq\widetilde{C}_2\sum_{h=1}^{u-l}\widetilde{b}_h^{\gamma}h^{\gamma/2-1}=\widetilde{C}_2\sum_{k=l+1}^{u}b_k^{\gamma}(k-l)^{\gamma/2-1}\leq\widetilde{C}_2\sum_{k=l+1}^{u}b_k^{\gamma}k^{\gamma/2-1}.$$

Thus, with $\widetilde{C}=2^{\gamma}\left(\max\left(\widetilde{C}_1,\widetilde{C}_2\right)\right)$ we can complete the proof of (a).

(b) On noting that by the triangle inequality and the monotonicity of the sequence $\{b_k\}$

$$\max_{l\leq k\leq u}b_k\left\|\sum_{i=m-k+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|$$
$$\leq b_u\left\|\sum_{i=m-u+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|+\max_{l\leq k<u}b_k\left\|\sum_{i=m-k+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|$$
$$\leq b_u\left\|\sum_{i=m-u+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|+\max_{l<k<u}b_k\left\|\sum_{i=m-k+1}^{m-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|+b_l\left\|\sum_{i=m-l+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|,$$

we receive

$$P\left(\max_{l\leq k\leq u}b_k\left\|\sum_{i=m-k+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\delta\right)$$

$$\leq P\left(b_u\left\|\sum_{i=m-u+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{3}\right)+P\left(\max_{l<k<u}b_k\left\|\sum_{i=m-k+1}^{m-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{3}\right)$$
$$+P\left(b_l\left\|\sum_{i=m-l+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{3}\right).$$

Furthermore, with the stationarity of the sequence $\{\boldsymbol{H}(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\}_{i\geq 1}$ we get

$$b_u\left\|\sum_{i=m-u+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\overset{D}{=}b_u\left\|\sum_{i=1}^{u}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\leq\max_{l\leq k\leq u}b_k\left\|\sum_{i=1}^{k}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|.$$

Thus, applying part (a) yields

$$\left(\frac{\delta}{3}\right)^{\gamma}P\left(b_u\left\|\sum_{i=m-u+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{3}\right)\leq\widetilde{C}_1\left(b_l^{\gamma}l^{\gamma/2}+\sum_{k=l+1}^{u}b_k^{\gamma}k^{\gamma/2-1}\right).$$

Besides, as $b_l\left\|\sum_{i=m-l+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\overset{D}{=}b_l\left\|\sum_{i=1}^{l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|$ we obtain by part (a)

$$\left(\frac{\delta}{3}\right)^{\gamma}P\left(b_l\left\|\sum_{i=m-l+1}^{m}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{3}\right)\leq\widetilde{C}_2\sum_{k=1+l}^{u}b_k^{\gamma}k^{\gamma/2-1}.$$

Moreover, with an index shift to $h=u-k$ and $\widetilde{b}_h=b_{u-h}$ the stationarity of the series can be used again to get

$$\max_{l<k<u}b_k\left\|\sum_{i=m-k+1}^{m-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|=\max_{1\leq h<u-l}\tilde{b}_h\left\|\sum_{i=m-u+h+1}^{m-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|$$
$$=\max_{1\leq h<u-l}\tilde{b}_h\left\|\sum_{i=m-u+h+1}^{m-u+u-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\overset{D}{=}\max_{1\leq h<u-l}\tilde{b}_h\left\|\sum_{i=h+1}^{u-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|.$$

Hence, since the sequence $\{\widetilde{b}_h\}$ is positive and non-decreasing Assumption A.2.10 (b) in connection with the Chebychev inequality yields

$$\left(\frac{\delta}{3}\right)^{\gamma}P\left(\max_{l\leq k<u}b_k\left\|\sum_{i=m-k+1}^{m-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{3}\right)$$
$$\leq\left(\frac{\delta}{3}\right)^{\gamma}P\left(\max_{1\leq h\leq u-l}\tilde{b}_h\left\|\sum_{i=h+1}^{u-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\delta}{3}\right)$$
$$\leq E\left(\max_{1\leq h\leq u-l}\tilde{b}_h\left\|\sum_{i=h+1}^{u-l}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\right)^{\gamma}\leq\tilde{C}_3\sum_{h=1}^{u-l}\tilde{b}_h^{\gamma}(u-l-h)^{\gamma/2-1}$$
$$=\tilde{C}_3\sum_{k=l}^{u-1}b_k^{\gamma}(k-l)^{\gamma/2-1}=\tilde{C}_3\sum_{k=l+1}^{u-1}b_k^{\gamma}(k-l)^{\gamma/2-1}\leq\tilde{C}_3\sum_{k=l+1}^{u}b_k^{\gamma}(k-l)^{\gamma/2-1}$$

$$\leq \tilde{C}_3 \sum_{k=l+1}^{u} b_k^{\gamma} k^{\gamma/2-1},$$

completing the proof of part (b).

□

In the following lemma we use the notation

$$\boldsymbol{E}_1(k,G,\boldsymbol{\theta}) = \frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\boldsymbol{\theta},k_{j,n}} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right) - \boldsymbol{A}_{\boldsymbol{\theta},k} + E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right) \quad \text{and} \tag{2.19}$$
$$\boldsymbol{E}_2(k,G,\boldsymbol{\theta}) = \frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\boldsymbol{\theta},k_{j,n}} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right) + \boldsymbol{A}_{\boldsymbol{\theta},k} - E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right),$$

for some fixed $\widetilde{\boldsymbol{\theta}}$ which appears in the expectations.

Lemma 2.2.2. *Let Assumptions A.2.1, A.2.2 and A.2.10 hold for some $\widetilde{\boldsymbol{\theta}}$. Then, for any $\beta > 0$, $0 < u < k_{j,n}$ and $0 < C \leq G$*

(a) $P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C} \frac{\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\|}{(k_{j,n}-k)} \geq \beta\right) = O\left((\beta^2 GC)^{-\gamma/2}\right)$

(b) $P\left(\max_{k_{j,n}-u\leq k\leq k_{j,n}} \left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\| \geq \beta\right) = O\left(\beta^{-\gamma}\left(\frac{u}{G}\right)^{\gamma/2}\right)$,

(c) $P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C} \left\|\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\| \geq \beta\right) = O\left(\beta^{-\gamma}\right)$,

with $\boldsymbol{E}_1(k,G,\boldsymbol{\theta})$ and $\boldsymbol{E}_2(k,G,\boldsymbol{\theta})$ as in (2.19).

Proof. Similar arguments as in the proof of Lemma 5.2 in Eichinger & Kirch (2018) can be used here.

(a) By Assumptions A.2.1 and A.2.2 we obtain, for all $k_{j,n} - G \leq k \leq k_{j,n} - C$,

$$\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})$$
$$= \frac{1}{\sqrt{2G}}\left(\sum_{i=k+G+1}^{k_{j,n}+G} \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}}) + \sum_{i=k-G+1}^{k_{j,n}-G} \boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}}) - 2\sum_{i=k+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right).$$

This implies in connection with an index shift of $l = k_{j,n} - k$

$$P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C} \frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)} \geq \beta\right)$$
$$\leq P\left(\max_{C\leq l\leq G} \frac{1}{l}\left\|\sum_{i=k_{j,n}-l+G+1}^{k_{j,n}+G} \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right\| \geq \frac{\beta\sqrt{2G}}{3}\right)$$
$$+ P\left(\max_{C\leq l\leq G} \frac{1}{l}\left\|\sum_{i=k_{j,n}-l-G+1}^{k_{j,n}-G} \boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\| \geq \frac{\beta\sqrt{2G}}{3}\right)$$
$$+ P\left(\max_{C\leq l\leq G} \frac{1}{l}\left\|\sum_{i=k_{j,n}-l+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\| \geq \frac{\beta\sqrt{2G}}{6}\right).$$

Applying part (b) of Lemma 2.2.1 on the first summand of the inequality above together with the monotony of integrals yields

$$
\begin{aligned}
&P\left(\max_{C\le l\le G}\frac{1}{l}\left\|\sum_{i=k_{j,n}-l+G+1}^{k_{j,n}+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta\sqrt{2G}}{3}\right)\\
&\le\widetilde{C}\left(C^{-\gamma/2}+\sum_{l=C+1}^{G}l^{-\gamma/2-1}\right)\left(\frac{\beta\sqrt{2G}}{3}\right)^{-\gamma}\\
&\le\widetilde{C}\left(C^{-\gamma/2}+\int_C^G x^{-\gamma/2-1}\mathrm{d}x\right)\left(\frac{\beta\sqrt{2G}}{3}\right)^{-\gamma}\\
&\le\widetilde{C}\left(C^{-\gamma/2}+\int_C^\infty x^{-\gamma/2-1}\mathrm{d}x\right)\left(\frac{\beta\sqrt{2G}}{3}\right)^{-\gamma}\\
&=\left(\beta^2GC\right)^{-\gamma/2}\widetilde{C}\left(\frac{\sqrt{2}}{3}\right)^{-\gamma}\left(1+\frac{2}{\gamma}\right)=O\left(\left(\beta^2CG\right)^{-\gamma/2}\right).
\end{aligned}
$$

By stationarity, this also implies corresponding assertions for the other two summands completing the proof of part (a).

(b) Since $\left\|\boldsymbol{E}_1(k_{j,n},G,\widetilde{\boldsymbol{\theta}})\right\|=0$, it is sufficient to consider the maximum over $k_{j,n}-u\le k<k_{j,n}$. Similar to (a), by the triangle inequality and an index shift to $l=k_{j,n}-k$ we obtain

$$
\begin{aligned}
&P\left(\max_{k_{j,n}-u\le k<k_{j,n}}\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|\ge\beta\right)\\
&\le P\left(\max_{1\le l\le u}\left\|\sum_{i=k_{j,n}-l+G+1}^{k_{j,n}+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta\sqrt{2G}}{3}\right)\\
&\quad+P\left(\max_{1\le l\le u}\left\|\sum_{i=k_{j,n}-l-G+1}^{k_{j,n}-G}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta\sqrt{2G}}{3}\right)\\
&\quad+P\left(\max_{1\le l\le u}\left\|\sum_{i=k_{j,n}-l+1}^{k_{j,n}}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta\sqrt{2G}}{6}\right).
\end{aligned}
$$

Considering the first summand, with $b_k\equiv 1$ we apply Lemma 2.2.1 (b) to get

$$
\begin{aligned}
&P\left(\max_{1\le l\le u}\left\|\sum_{i=k_{j,n}-l+G+1}^{k_{j,n}+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta\sqrt{2G}}{3}\right)\\
&\le\widetilde{C}\left(u^{\gamma/2}+\sum_{l=1}^{u}l^{\gamma/2-1}\right)\left(\frac{\beta\sqrt{2G}}{3}\right)^{-\gamma}=O\left(\beta^{-\gamma}\left(\frac{u}{G}\right)^{\gamma/2}\right),
\end{aligned}
$$

since $\gamma/2-1>0$. By stationarity, this also implies corresponding assertions for the other two summands.

(c) On noting that

$$\begin{aligned}&\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\\&=-\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})+\frac{\sqrt{2}}{\sqrt{G}}\left(\sum_{i=k_{j,n}+1}^{k_{j,n}+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})-\sum_{i=k_{j,n}-G+1}^{k_{j,n}}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right)\end{aligned}$$

holds for all $k_{j,n}-G\le k\le k_{j,n}-C$ we receive

$$\begin{aligned}&P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\left\|\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\|\ge\beta\right)\\&\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta}{3}\right)\\&\quad+P\left(\frac{1}{\sqrt{G}}\left\|\sum_{i=k_{j,n}+1}^{k_{j,n}+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta}{3\sqrt{2}}\right)\\&\quad+P\left(\frac{1}{\sqrt{G}}\left\|\sum_{i=k_{j,n}-G+1}^{k_{j,n}}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta}{3\sqrt{2}}\right).\end{aligned}\tag{2.20}$$

For the first summand, applying the result of part (b) yields

$$\begin{aligned}&P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta}{3}\right)\\&\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}}\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta}{3}\right)=O\left(\beta^{-\gamma}\right).\end{aligned}$$

Furthermore, by Lemma 2.2.1 (a) we obtain, for some constant $\widetilde{C}>0$,

$$P\left(\frac{1}{\sqrt{G}}\left\|\sum_{i=k_{j,n}+1}^{k_{j,n}+G}\boldsymbol{H}_0(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta}{3\sqrt{2}}\right)\le\left(\frac{\beta}{3\sqrt{2}}\right)^{-\gamma}\widetilde{C}=O\left(\beta^{-\gamma}\right),$$

and similarly with Lemma 2.2.1 (b)

$$P\left(\frac{1}{\sqrt{G}}\left\|\sum_{i=k_{j,n}-G+1}^{k_{j,n}}\boldsymbol{H}_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\beta}{3\sqrt{2}}\right)=O\left(\beta^{-\gamma}\right),$$

completing the proof of part (c).

□

Lemma 2.2.3. *Let the Assumption A.1.1, A.2.1 and A.2.2 hold for some $\widetilde{\boldsymbol{\theta}}$. Furthermore, assume that $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ is a sequence of estimators fulfilling Assumption A.2.11 for $\widetilde{\boldsymbol{\theta}}$. Then,*

(a)

$$\max_{k_{j,n}-G\le k\le k_{j,n}-C}\frac{\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}}-\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\|}{k_{j,n}-k}=o_P(1)$$

(b)

$$\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}+\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}}-\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}+\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)\right\|=o_P\left(G\right)$$

Proof. (a) Assumptions A.2.1 and A.2.2 and the triangle inequality in connection with an index shift to $l=k_{j,n}-k$ yield

$$\begin{aligned}
&\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}}-\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\|}{k_{j,n}-k}\\
&\leq\max_{C\leq l\leq G}\frac{\left\|\sum_{i=k_{j,n}+G-l+1}^{k_{j,n}+G}\left(\boldsymbol{H}(\mathbb{X}_i^{(j+1)},\widehat{\boldsymbol{\theta}}_n)-\boldsymbol{H}(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right)\right\|}{l}\\
&\quad+\max_{C\leq l\leq G}\frac{\left\|\sum_{i=k_{j,n}-l+1}^{k_{j,n}}\left(\boldsymbol{H}(\mathbb{X}_i^{(j)},\widehat{\boldsymbol{\theta}}_n)-\boldsymbol{H}(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right)\right\|}{l}\\
&\quad+\max_{C\leq l\leq G}\frac{\left\|\sum_{i=k_{j,n}-G-l+1}^{k_{j,n}-G}\left(\boldsymbol{H}(\mathbb{X}_i^{(j)},\widehat{\boldsymbol{\theta}}_n)-\boldsymbol{H}(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right)\right\|}{l}=o_P(1),
\end{aligned}$$

where the last line follows directly from Assumption A.2.11.

(b) We get

$$\begin{aligned}
&\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}+\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}}-\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}+\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)\right\|}{G}\\
&\leq\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}}-\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\|}{k_{j,n}-k}+\frac{2}{G}\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right\|\\
&\leq\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}}-\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\|}{k_{j,n}-k}\\
&\quad+\frac{2}{G}\left\|\sum_{i=k_{j,n}+1}^{k_{j,n}+G}\left(\boldsymbol{H}(\mathbb{X}_i^{(j+1)},\widehat{\boldsymbol{\theta}}_n)-\boldsymbol{H}(\mathbb{X}_i^{(j+1)},\widetilde{\boldsymbol{\theta}})\right)\right\|\\
&\quad+\frac{2}{G}\left\|\sum_{i=k_{j,n}-G+1}^{k_{j,n}}\left(\boldsymbol{H}(\mathbb{X}_i^{(j)},\widehat{\boldsymbol{\theta}}_n)-\boldsymbol{H}(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right)\right\|\\
&=o_P(1),
\end{aligned}$$

where the last line follows from Assumption A.2.11 and the result of part (a).

□

The following lemma gives more information about the intervals of exceedings obtained by the MOSUM score-type procedure. It shows that the start and end points v_j and w_j, $j\in\widetilde{Q}$, lie in the G-environment of the corresponding change point $k_{j,n}$ with probability tending to one.

Lemma 2.2.4. *Let the assumptions of Theorem 2.1.8 hold. Furthermore, let* $[v_{j,n}, w_{j,n}]$ *for* $j \in \widetilde{Q}$ *be the intervals of exceedings of the statistic* $T_{k,n}\left(G, \widetilde{\boldsymbol{\theta}}\right)$, $T_{k,n}\left(G, \widehat{\boldsymbol{\theta}}_n\right)$ *or* $\widehat{T}_{k,n}\left(G, \widehat{\boldsymbol{\theta}}_n\right)$. *Then,*

$$\lim_{n\to\infty} P\left(k_{j,n} - G < v_{j,n} < k_{j,n} < w_{j,n} < k_{j,n} + G\right) = 1.$$

Proof. The statement follows directly from the results shown in the proof of Theorem 2.1.8 since

$$\begin{aligned} &P\left(k_{j,n} - G < v_{j,n} < k_{j,n} < w_{j,n} < k_{j,n} + G\right) \\ &\geq P\left(\left\{\max_{k\in\widetilde{A}_{n,G}} T_{k,n}\left(G, \widetilde{\boldsymbol{\theta}}\right) < D_n\left(\alpha_n, G\right)\right\} \cap \left\{\min_{k\in\bar{B}_{n,G}} T_{k,n}\left(G, \widetilde{\boldsymbol{\theta}}\right) \geq D_n\left(\alpha_n, G\right)\right\}\right) \\ &\geq P\left(\max_{k\in\widetilde{A}_{n,G}} T_{k,n}\left(G, \widetilde{\boldsymbol{\theta}}\right) < D_n\left(\alpha_n, G\right)\right) + P\left(\min_{k\in\bar{B}_{n,G}} T_{k,n}\left(G, \widetilde{\boldsymbol{\theta}}\right) \geq D_n\left(\alpha_n, G\right)\right) - 1, \end{aligned}$$

with $\widetilde{A}_{n,G}$ and $\bar{B}_{n,G}$ as in (2.12) and (2.13). □

Now, we are almost ready to state the main result of this section. However, note that the long-run covariance matrix of $\boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})$ is usually unknown in applications so that an estimator is used in the statistic. In order to get a better convergence rate of the change point estimators for these situations as well we need to modify the MOSUM procedure in the following way.

Let $\widehat{\boldsymbol{\Sigma}}_{k,n}$ be an estimator for the long-run covariance matrix $\boldsymbol{\Sigma}_k$ fulfilling Assumption A.2.5. This covariance matrix estimator depends on k, which means that we can get different estimates of the covariance matrix for every time point k. For this reason we call it a local estimator of the long-run covariance matrix. The estimators and the corresponding intervals of exceedings obtained by the MOSUM procedure, which uses this local estimator in the statistic, are denoted by $\widehat{q}_n, \widehat{k}_{j,n}$ and $[v_{j,n}, w_{j,n}]$, for $j \in \widetilde{Q}$. Furthermore, let $\widehat{\boldsymbol{\Sigma}}_{j,n}$, $j = 1, \ldots, q+1$, be an estimator of the long-run covariance matrix computed on the whole sample or a fixed subsample which allows to apply different estimators $\widehat{\boldsymbol{\Sigma}}_{j,n}$ for different regimes. We call $\widehat{\boldsymbol{\Sigma}}_{j,n}$ a global estimator and assume that this estimator sequence is consistent for the true long-run covariance matrix $\boldsymbol{\Sigma}$ under the null and converges in probability to some positive definite matrix $\boldsymbol{\Sigma}_{A,j}$ under alternative.

At first, we determine the intervals of exceedings $[v_{j,n}, w_{j,n}]$ by the MOSUM statistic which uses the local estimator $\widehat{\boldsymbol{\Sigma}}_{k,n}$. Then, the change point estimators are computed by finding the maxima of the MOSUM statistic, which bases on the global estimator $\widehat{\boldsymbol{\Sigma}}_{j,n}$, on $[v_{j,n}, w_{j,n}]$, $j \in \widetilde{Q}$. We define

$$\bar{k}_{j,n} := \underset{v_{j,n}\leq k\leq w_{j,n}}{\arg\max} \frac{1}{\sqrt{2G}} \sqrt{\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}^T \widehat{\boldsymbol{\Sigma}}_{j,n}^{-1} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}} \quad \text{for} \quad j \in \widetilde{Q}. \tag{2.21}$$

Note that not all of them need to exist in finite samples.

In the following theorem we derive a better convergence rate for these modified change point estimators in comparison to Corollary 2.1.10.

Theorem 2.2.5. *Let Assumption A.1.1 on the bandwidth and Assumptions A.2.1, A.2.2,A.2.3, A.2.7 and A.2.10 hold for some* $\widetilde{\boldsymbol{\theta}}$. *Assume that a sequence* $\{\alpha_n\}_{n\in\mathbb{N}}$ *fulfills Assumption A.2.8. Furthermore, let* $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *be a local estimator and* $\widehat{\boldsymbol{\Sigma}}_{j,n}$ *be a global estimator for the long-run covariance matrix fulfilling Assumption A.2.12.*

(a) Then,

$$\max_{j\in\widetilde{Q}} \min_{1\le l\le \hat{q}_n} \left|\bar{k}_{l,n} - k_{j,n}\right| = O_P(1).$$

(b) Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ *be a sequence of estimators fulfilling Assumption A.2.4, A.2.9 and A.2.11 for* $\widetilde{\boldsymbol{\theta}}$. *Then,* $\widetilde{\boldsymbol{\theta}}$ *can be replaced by* $\widehat{\boldsymbol{\theta}}_n$ *in the statistic without changing the result of part (a).*

Proof. (a) The basic idea of this proof goes back to Eichinger & Kirch (2018) (Theorem 3.2.).

Since the number of change points is finite it is sufficient to prove that $\min_{1\le l\le \hat{q}_n}\left|\bar{k}_{l,n} - k_{j,n}\right| = O_P(1)$ holds for all $j \in \widetilde{Q}$. Hence, we want to show that for each $\epsilon > 0$ there exists a constant $C > 0$ such that

$$\begin{aligned}
&P\left(\min_{1\le l\le \hat{q}_n}\left|\bar{k}_{l,n} - k_{j,n}\right| > C\right) \le P\left(\left|\bar{k}_{j,n} - k_{j,n}\right| > C\right)\\
&= P\left(\{\bar{k}_{j,n} > k_{j,n} + C\} \cup \{\bar{k}_{j,n} < k_{j,n} - C\}\right)\\
&= P\left(\bar{k}_{j,n} > k_{j,n} + C\right) + P\left(\bar{k}_{j,n} < k_{j,n} - C\right) \le \epsilon.
\end{aligned}$$

We define $I_{n,G} := \{k_{j,n} - G < v_{j,n} < k_{j,n} < w_{j,n} < k_{j,n} + G\}$ and $M_{n,G} := \{\widehat{q}_n = \widetilde{q}\} \cap I_{n,G}$, with $v_{j,n}, w_{j,n}$ and $\widehat{q}_n$ obtained by the MOSUM procedure using the local covariance matrix estimator. Furthermore, we get

$$\begin{aligned}
\bar{k}_{j,n} &= \underset{v_{j,n}\le k\le w_{j,n}}{\arg\max} \frac{1}{\sqrt{2G}}\sqrt{\boldsymbol{A}^T_{\widetilde{\boldsymbol{\theta}},k}\widehat{\boldsymbol{\Sigma}}^{-1}_{j,n}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}} = \underset{v_{j,n}\le k\le w_{j,n}}{\arg\max} \frac{1}{\sqrt{2G}}\left\|\widehat{\boldsymbol{\Sigma}}^{-1/2}_{j,n}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\|\\
&= \underset{v_{j,n}\le k\le w_{j,n}}{\arg\max} V^{(j)}_{k,n}(G,\widetilde{\boldsymbol{\theta}}),
\end{aligned}$$

for all $j \in \widetilde{Q}$, where $V^{(j)}_{k,n}(G,\widetilde{\boldsymbol{\theta}}) = \frac{1}{2G}\left\|\widehat{\boldsymbol{\Sigma}}^{-1/2}_{j,n}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\|^2 - \frac{1}{2G}\left\|\widehat{\boldsymbol{\Sigma}}^{-1/2}_{j,n}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right\|^2$. On noting that $V^{(j)}_{k_{j,n},n}(G,\widetilde{\boldsymbol{\theta}}) = 0$ and $k_{j,n} \in (v_{j,n}, w_{j,n})$ on $M_{n,G}$, we receive, for some $0 < C < G$,

$$\begin{aligned}
&P\left(\bar{k}_{j,n} < k_{j,n} - C\right)\\
&= P\left(\bar{k}_{j,n} < k_{j,n} - C, M_{n,G}\right) + P\left(\bar{k}_{j,n} < k_{j,n} - C, M^C_{n,G}\right)\\
&\le P\left(\underset{v_{j,n}\le k\le w_{j,n}}{\arg\max} V^{(j)}_{k,n}(G,\widetilde{\boldsymbol{\theta}}) < k_{j,n} - C, M_{n,G}\right) + P\left(M^C_{n,G}\right)\\
&= P\left(\max_{v_{j,n}\le k< k_{j,n}-C} V^{(j)}_{k,n}(G,\widetilde{\boldsymbol{\theta}}) \ge \max_{k_{j,n}-C\le k\le w_{j,n}} V^{(j)}_{k,n}(G,\widetilde{\boldsymbol{\theta}}), M_{n,G}\right)\\
&\quad + P\left(M^C_{n,G}\right)\\
&\le P\left(\max_{v_{j,n}\le k< k_{j,n}-C} V^{(j)}_{k,n}(G,\widetilde{\boldsymbol{\theta}}) \ge 0, M_{n,G}\right) + P\left(M^C_{n,G}\right)
\end{aligned}$$

$$= P\left(\max_{v_{j,n}\le k<k_{j,n}-C} V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}})\ge 0, M_{n,G}\right)+o(1),$$

where the last line follows from Theorem 2.1.8 and Lemma 2.2.4 since

$$P(M_{n,G}^C)\le P\left(\widehat{q}_n\neq\widetilde{q}\right)+P\left(I_{n,G}\right).$$

Furthermore, we obtain

$$P\left(\max_{v_{j,n}\le k<k_{j,n}-C} V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}})\ge 0, M_{n,G}\right)\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C} V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}})\ge 0, M_{n,G}\right).$$

Hence, it suffices to investigate the maximum of $V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}})$ over all time points $k\in\{k_{j,n}-G,\ldots,k_{j,n}-C\}$. On noting that $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{y}=\boldsymbol{y}^T\boldsymbol{B}\boldsymbol{x}$ holds in general for a symmetric $p\times p$ matrix $\boldsymbol{B}$ and vectors $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^p$, we obtain

$$\begin{aligned}
&2G\,V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}})=\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\|^2-\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right\|^2\\
&=\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}=\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}+\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)\\
&=-\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}+\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right).
\end{aligned}$$

Furthermore, with $\boldsymbol{d}_j=E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\widetilde{\boldsymbol{\theta}})\right)-E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\widetilde{\boldsymbol{\theta}})\right)$, we receive by Lemma 2.1.2, for $k\in\{k_{j,n}-G,\ldots,k_{j,n}-C\}$,

$$\begin{aligned}
&\frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\\
&=\frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}+E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)+(2G)^{-1/2}\left(k_{j,n}-k\right)\boldsymbol{d}_j\\
&=:\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})+(2G)^{-1/2}\left(k_{j,n}-k\right)\boldsymbol{d}_j
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}+\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)\\
&=\frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)+\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}-E\left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)+(2G)^{-1/2}\left(2G+k-k_{j,n}\right)\boldsymbol{d}_j\\
&=:\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})+(2G)^{-1/2}\left(2G+k-k_{j,n}\right)\boldsymbol{d}_j.
\end{aligned}$$

Hence, we get

$$\begin{aligned}
&V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}})\\
&=-\left(\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})+(2G)^{-1/2}\left(k_{j,n}-k\right)\boldsymbol{d}_j\right)^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\\
&\qquad\left(\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})+(2G)^{-1/2}\left(2G+k-k_{j,n}\right)\boldsymbol{d}_j\right)
\end{aligned}$$

$$\begin{aligned}
&= -\Big(\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}}) \\
&\quad +(2G)^{-1/2}\left(2G+k-k_{j,n}\right)\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j \\
&\quad +(2G)^{-1/2}\left(k_{j,n}-k\right)\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}}) \\
&\quad +(2G)^{-1}(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j\Big).
\end{aligned} \tag{2.22}$$

Considering the last summand in (2.22) first, we want to replace the estimator $\widehat{\boldsymbol{\Sigma}}_{j,n}$ by the positive definite matrix $\boldsymbol{\Sigma}_A$ given by Assumption A.2.12. Applying Assumption A.2.12 and Lemma E.1.8 yields $\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}-\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F = o_P(1)$. Thus, on noting that $\|\boldsymbol{d}_j\| = O(1)$, Lemma E.1.5 can be used to receive

$$\begin{aligned}
&\left|\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{d}_j\right\|^2-\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2\right| \\
&= \left|\left(\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{d}_j\right\|-\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|\right)\left(\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{d}_j\right\|+\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|\right)\right| \\
&\leq \left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}-\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F\|\boldsymbol{d}_j\|\left(\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}-\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F\|\boldsymbol{d}_j\|+2\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|\right) = o_P(1),
\end{aligned}$$

implying that $\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j = \boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j + o_P(1)$. Furthermore, since the matrix $\boldsymbol{\Sigma}_{A,j}^{-1}$ is positive definite and $\boldsymbol{d}_j \neq \boldsymbol{0}$ holds for all $j \in \widetilde{Q}$ we obtain

$$(2G)^{-1}(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j \geq \frac{C}{2}\boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j = \frac{C}{2}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2 > 0,$$

for all $k_{j,n}-G \leq k \leq k_{j,n}-C$. Hence, in connection with (2.22) we get

$$\begin{aligned}
&V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}}) \\
&= -\Big(\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}}) \\
&\quad +(2G)^{-1/2}\left(2G+k-k_{j,n}\right)\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j \\
&\quad +(2G)^{-1/2}\left(k_{j,n}-k\right)\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}}) \\
&\quad +(2G)^{-1}(2G+k-k_{j,n})(k_{j,n}-k)\left(\boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j+o_P(1)\right)\Big) \\
&= -(2G)^{-1}(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j \\
&\quad\Bigg(\frac{2G\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j}+\frac{\sqrt{2G}\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j}{(k_{j,n}-k)\boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} \\
&\quad +\frac{\sqrt{2G}\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})\boldsymbol{d}_j^T\boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j)}+1+o_P(1)\Bigg).
\end{aligned}$$

This can be used to obtain

$$P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C} V_{k,n}^{(j)}(G,\widetilde{\boldsymbol{\theta}}) \geq 0, M_{n,G}\right)$$

$$
\begin{aligned}
&= P\Bigg(\max_{k_{j,n}-G \le k \le k_{j,n}-C} -(2G)^{-1}(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j \\
&\quad \Bigg(\frac{2G\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} + \frac{\sqrt{2G}\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j}{(k_{j,n}-k)\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} \\
&\quad + \frac{\sqrt{2G}\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j)} + 1 + o_P(1)\Bigg) \ge 0, M_{n,G}\Bigg) \\
&= P\Bigg(\max_{k_{j,n}-G \le k \le k_{j,n}-C} \Bigg(\frac{2G\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} \\
&\quad + \frac{\sqrt{2G}\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j}{(k_{j,n}-k)\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} + \frac{\sqrt{2G}\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j)} + o_P(1)\Bigg) \le -1, M_{n,G}\Bigg) \\
&\le P\Bigg(\max_{k_{j,n}-G \le k \le k_{j,n}-C} \Bigg| \frac{2G\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})(k_{j,n}-k)\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} \\
&\quad + \frac{\sqrt{2G}\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j}{(k_{j,n}-k)\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} + \frac{\sqrt{2G}\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})}{(2G+k-k_{j,n})\boldsymbol{d}_j^T \boldsymbol{\Sigma}_{A,j}^{-1}\boldsymbol{d}_j} + o_P(1)\Bigg| \ge 1, M_{n,G}\Bigg) \\
&\le P\Bigg(\max_{k_{j,n}-G \le k \le k_{j,n}-C} \frac{2G\left|\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right|}{(2G+k-k_{j,n})(k_{j,n}-k)\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2} \\
&\quad + \max_{k_{j,n}-G \le k \le k_{j,n}-C} \frac{\sqrt{2G}\left|\boldsymbol{E}_1^T(k,G,\widetilde{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{d}_j\right|}{(k_{j,n}-k)\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2} \\
&\quad + \max_{k_{j,n}-G \le k \le k_{j,n}-C} \frac{\sqrt{2G}\left|\boldsymbol{d}_j^T\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right|}{(2G+k-k_{j,n})\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2} + o_P(1) \ge 1, M_{n,G}\Bigg) \\
&\le P\Bigg(\max_{k_{j,n}-G \le k \le k_{j,n}-C} \frac{\sqrt{2G}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|} \\
&\quad \max_{k_{j,n}-G \le k \le k_{j,n}-C} \frac{\sqrt{2}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|} \\
&\quad + \max_{k_{j,n}-G \le k \le k_{j,n}-C} \frac{\sqrt{2G}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\| \left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{d}_j\right\|}{(k_{j,n}-k)\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2} \\
&\quad + \max_{k_{j,n}-G \le k \le k_{j,n}-C} \frac{\sqrt{2}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\| \left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\boldsymbol{d}_j\right\|}{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2} + o_P(1) \ge 1, M_{n,G}\Bigg),
\end{aligned}
$$

where the last line follows from Cauchy-Schwarz inequality and $(2G+k-k_{j,n}) \ge G$.

After splitting the probability, applying Lemma E.1.5 yields

$$
\begin{aligned}
&\leq P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)}\geq\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2G}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F}\right)\\
&\quad+P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\left\|\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F}\right)\\
&\quad+P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)}\geq\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2}{4\sqrt{2G}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F^2\|\boldsymbol{d}_j\|}\right)\\
&\quad+P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\left\|\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\|\geq\frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2}{4\sqrt{2}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F^2\|\boldsymbol{d}_j\|}\right)+o(1).
\end{aligned}
$$

In order to use Lemma 2.2.2 we need to get rid of the covariance matrix estimator in the probability statements above. Therefore, we define

$$
F_{n,\varepsilon}:=\left\{\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F\leq\left\|\boldsymbol{\Sigma}_{A}^{-1/2}\right\|_F+\varepsilon\right\}
$$

for some $\varepsilon>0$. On noting that

$$
\begin{aligned}
P\left(F_{n,\varepsilon}^C\right)&=P\left(\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F-\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F>\varepsilon\right)\\
&\leq P\left(\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}-\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F>\varepsilon\right)\to 0\quad\text{as}\quad n\to\infty,
\end{aligned}
$$

applying Lemma 2.2.2 (a) and (c) yields

$$
\begin{aligned}
&P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)}\geq\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2G}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F}\right)\\
&\leq P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)}\geq\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2G}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F},F_{n,\varepsilon}\right)+P\left(F_{n,\varepsilon}^C\right)\\
&\leq P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)}\geq\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2G}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F+\varepsilon\right)},F_{n,\varepsilon}\right)\\
&\quad+P\left(F_{n,\varepsilon}^C\right)\\
&\leq P\left(\max_{k_{j,n}-G\leq k\leq k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)}\geq\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2G}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F+\varepsilon\right)}\right)+P\left(F_{n,\varepsilon}^C\right)\\
&=O\left(C^{-\gamma/2}\right)+o(1)=o(1)
\end{aligned}
$$

and

$$P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\left\|\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2}\left\|\widehat{\boldsymbol{\Sigma}}_{j,n}^{-1/2}\right\|_F}\right)$$
$$\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\left\|\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\|\ge\frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F+\varepsilon\right)}\right)+P\left(F_{n,\varepsilon}^C\right)$$
$$=O\left(G^{-\gamma/2}\right)+o(1)=o(1).$$

The remaining probabilities can be approximated in an analogous manner. Consequently, we receive

$$P\left(\bar{k}_{j,n}<k_{j,n}-C\right)=O\left(C^{-\gamma/2}\right)+o(1).$$

The second part of the assertion

$$P\left(\bar{k}_{j,n}>k_{j,n}+C\right)=O\left(C^{-\gamma/2}\right)+o(1)$$

can be shown analogously with a modified version of Lemma 2.2.2. Hence, we can conclude that

$$P\left(\left|\bar{k}_{j,n}-k_{j,n}\right|>C\right)=O\left(C^{-\gamma/2}\right)+o(1),$$

which proves the assertion.

(b) Similar to (a), we obtain

$$P\left(\bar{k}_{j,n}<k_{j,n}-C\right)\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}V_{k,n}^{(j)}(G,\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}})\ge 0,M_{n,G}\right)+o(1)$$

since the results of Theorem 2.1.8 and Lemma 2.2.4 remain true if $\widetilde{\boldsymbol{\theta}}$ is replaced by an estimator satisfying Assumptions A.2.4 and A.2.9. Moreover, we receive

$$P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}V_{k,n}^{(j)}(G,\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}})\ge 0,M_{n,G}\right)$$
$$\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widehat{\boldsymbol{\theta}}_n)\right\|}{(k_{j,n}-k)}\ge\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2G}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F+\varepsilon\right)}\right)$$
$$+P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\left\|\boldsymbol{E}_2(k,G,\widehat{\boldsymbol{\theta}}_n)\right\|\ge\frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F+\varepsilon\right)}\right)$$
$$+P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\frac{\left\|\boldsymbol{E}_1(k,G,\widehat{\boldsymbol{\theta}}_n)\right\|}{(k_{j,n}-k)}\ge\frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2}{4\sqrt{2G}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F+\varepsilon\right)^2\|\boldsymbol{d}_j\|}\right)$$
$$+P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C}\left\|\boldsymbol{E}_2(k,G,\widehat{\boldsymbol{\theta}}_n)\right\|\ge\frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|^2}{4\sqrt{2}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F+\varepsilon\right)^2\|\boldsymbol{d}_j\|}\right)+o(1)$$
$$=o(1).$$

On noting that

$$\boldsymbol{E}_1(k,G,\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}}) + \frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}} - \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right)$$

and

$$\boldsymbol{E}_2(k,G,\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}}) + \frac{1}{\sqrt{2G}}\left(\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} + \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}} - \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} + \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)\right),$$

combining Lemma 2.2.3 and the results of part (a) yields

$$\begin{aligned}
&P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C} \frac{\left\|\boldsymbol{E}_1(k,G,\widehat{\boldsymbol{\theta}}_n)\right\|}{(k_{j,n}-k)} \ge \frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2G}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F + \varepsilon\right)}\right)\\
&\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C} \frac{\left\|\boldsymbol{E}_1(k,G,\widetilde{\boldsymbol{\theta}})\right\|}{(k_{j,n}-k)} \ge \frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{4\sqrt{2G}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F + \varepsilon\right)}\right)\\
&\quad + P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C} \frac{\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}} - \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right)\right\|}{(k_{j,n}-k)} \ge \frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{4\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F + \varepsilon\right)}\right)\\
&= O\left(C^{-\gamma/2}\right) + o(1) = o(1)
\end{aligned}$$

and

$$\begin{aligned}
&P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C} \left\|\boldsymbol{E}_2(k,G,\widehat{\boldsymbol{\theta}}_n)\right\| \ge \frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{2\sqrt{2}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F + \varepsilon\right)}\right)\\
&\le P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C} \left\|\boldsymbol{E}_2(k,G,\widetilde{\boldsymbol{\theta}})\right\| \ge \frac{\sqrt{G}\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{4\sqrt{2}\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F + \varepsilon\right)}\right)\\
&\quad + P\left(\max_{k_{j,n}-G\le k\le k_{j,n}-C} \frac{\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} + \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k_{j,n}} - \left(\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} + \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k_{j,n}}\right)\right\|}{G} \ge \frac{\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\boldsymbol{d}_j\right\|}{4\left(\left\|\boldsymbol{\Sigma}_{A,j}^{-1/2}\right\|_F + \varepsilon\right)}\right)\\
&= O\left(G^{-\gamma/2}\right) + o(1) = o(1),
\end{aligned}$$

which proves the assertion. □

2.3. Some Considerations on the Assumptions

The assumptions of Theorem 2.1.1 and Theorem 2.1.8 are stated in quite a general way. Hence, before applying the MOSUM procedure to a specific model one needs to check whether these general assumptions are satisfied or not. Here, we consider two examples, an i.i.d. sequence and a stationary and strongly mixing sequence, and show that they satisfy the main assumptions under some moment conditions which are summarized in

Section B. In doing so, we first assume that there exists an estimator sequence which is $\sqrt{n}$-consistent. Later on, we will examine a specific class of estimator sequences and derive their $\sqrt{n}$-consistency under the null hypothesis and the alternative.
In this chapter, we assume that the estimating function $\boldsymbol{H}$ is twice continuously differentiable on $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta} \subset \mathbb{R}^p$ is a compact parameter space, and that $\boldsymbol{H}$ and its derivatives are measurable with respect to $\mathbb{X}_i$. For convenience, let the co-domain of $\boldsymbol{H}$ be a subset of $\mathbb{R}^p$, i.e. $\boldsymbol{H}$ is a vector valued function such that $\boldsymbol{H}(\boldsymbol{x},\boldsymbol{\theta}) = (H_1(\boldsymbol{x},\boldsymbol{\theta}),\dots,H_p(\boldsymbol{x},\boldsymbol{\theta}))^T$. Furthermore, note that the following notation is used:

- $\boldsymbol{H}_0(\mathbb{X}_i,\boldsymbol{\theta}) = \boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}) - E\left(\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\right)$,
- $\nabla\boldsymbol{H}_0(\mathbb{X}_i,\boldsymbol{\theta}) = \nabla\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}) - E\left(\nabla\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\right)$ with
  $\nabla\boldsymbol{H}_0(\mathbb{X}_i,\boldsymbol{\theta}) = (\nabla H_{1,0},\dots\nabla H_{p,0})$, where $\nabla H_{j,0}$ denotes the centered gradient vector of H_j,
- $\nabla^2 H_{j,0}(\mathbb{X}_i,\boldsymbol{\theta}) = \nabla^2 H_j(\mathbb{X}_i,\boldsymbol{\theta}) - E\left(\nabla^2 H_j(\mathbb{X}_1,\boldsymbol{\theta})\right)$ representing the centered Hessian matrix of H_j,
- under the null hypothesis: $\boldsymbol{V}(\boldsymbol{\theta}) = E\left(\nabla\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\right)^T$ and
- under alternative: $\boldsymbol{V}_j(\boldsymbol{\theta}) = E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right)^T$.

2.3.1. Under the Null Hypothesis

We concentrate on the following two examples satisfying Assumption A.1.2:

(E1) $\mathbb{X}_1,\dots,\mathbb{X}_n$ are an i.i.d. sequence of random vectors or

(E2) $\mathbb{X}_1,\dots,\mathbb{X}_n$ are a stationary and strongly mixing sequence of random vectors with a mixing rate $\alpha(n)$ satisfying $\alpha(n) = O(n^{-\beta})$ for some $\beta > 1 + 2/\nu$, where ν is as in Assumption A.1.3.

The strong mixing condition introduced by Rosenblatt (1956) describes a specific type of dependence. The following definition can be found in Bradley (2007) on page 28.

Definition 2.3.1. *Let $\{X_i\}_{i\geq 1}$ be a sequence of random variables and let $\mathcal{F}_l^u := \sigma(X_i, l \leq i \leq u)$ denote the σ-field generated by $(X_i, l \leq i \leq u)$. Furthermore, let*

$$\alpha(n) := \sup_{j\in\mathbb{N}} \sup_{A\in\mathcal{F}_1^j, B\in\mathcal{F}_{j+n}^\infty} \left|P(A\cap B) - P(A)P(B)\right|.$$

Then, the sequence $\{X_i\}$ is called strongly mixing if $\alpha(n) \to 0$ as $n \to \infty$.

Furthermore, we assume that the estimator sequence $\widehat{\boldsymbol{\theta}}_n$ is $\sqrt{n}$-consistent for some $\widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ under the null hypothesis and that the series fulfill the moment conditions below which are listed in Section B.1 as well.

B.1.1 *Let $E\left(\|\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\|\right) < \infty$ hold for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.*

B.1.2 *Let* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1,\widetilde{\boldsymbol{\theta}})\right\|^2\right)<\infty$.

B.1.3 *Let* $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\nabla\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\|_F\right)<\infty$.

B.1.4 $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\nabla^2 H_l(\mathbb{X}_1,\boldsymbol{\theta})\|_F\right)<\infty$ *hold for all* $l=1,\ldots,p$.

B.1.5 *There exists a* $\nu>0$ *such that* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1,\widetilde{\boldsymbol{\theta}})\right\|^{2+\nu}\right)<\infty$.

B.1.6 *There exists a* $\nu>0$ *such that* $E\left(\|\nabla\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\|_F^{2+\nu}\right)<\infty$ *holds for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$.

2.3.1.1. Assumptions A.1.4 and A.1.3

In this paragraph we prove that the main assumptions of Theorem 2.1.1 are satisfied by these specific time series.

Lemma 2.3.2. *Let* $\{\mathbb{X}_i : i\geq 1\}$ *be a series of type (E1) or type (E2). Furthermore, let Assumption A.1.1 hold on the bandwidth and let* $\widehat{\boldsymbol{\theta}}_n$ *be a global estimator sequence which is* $\sqrt{n}$*-consistent for some* $\widetilde{\boldsymbol{\theta}}$ *under the null. Then,*

- *if* $\{\mathbb{X}_i : i\geq 1\}$ *fulfills Condition B.1.5 , Assumption A.1.3 is satisfied.*
- *if* $\{\mathbb{X}_i : i\geq 1\}$ *fulfills the Conditions B.1.4 and B.1.6, Assumptions A.1.4 is satisfied.*

Proof.

- Assumption A.1.4:
  By a Taylor expansion of each component ($l=1,\ldots,p$) there exists a $\boldsymbol{\xi}_{l,n,k}$ such that $\left\|\boldsymbol{\xi}_{l,n,k}-\widetilde{\boldsymbol{\theta}}\right\|\leq\left\|\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}\right\|$ with

$$\sum_{i=k+1}^{k+G} H_l(\mathbb{X}_i,\widehat{\boldsymbol{\theta}}_n)-\sum_{i=k-G+1}^{k} H_l(\mathbb{X}_i,\widehat{\boldsymbol{\theta}}_n)-\left(\sum_{i=k+1}^{k+G} H_l(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})-\sum_{i=k-G+1}^{k} H_l(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})\right) \tag{2.23}$$

$$=\left(\sum_{i=k+1}^{k+G}\nabla H_l(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})-\sum_{i=k-G+1}^{k}\nabla H_l(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})\right)^T\left(\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}\right)$$
$$+\frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}\right)^T\left(\sum_{i=k+1}^{k+G}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k})-\sum_{i=k-G+1}^{k}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k})\right)\left(\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}\right),$$

  where $\nabla H_l(\mathbb{X}_i,\boldsymbol{\theta})$ denotes the gradient with respect to $\boldsymbol{\theta}$ and $\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\theta})$ is the Hessian matrix.
  We start with approximating the first summand and use the following notation:

$$\nabla H_l(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})=\widetilde{\boldsymbol{X}}_i=(\widetilde{X}_{i,1},\ldots,\widetilde{X}_{i,p})^T.$$

Since ∇H_l and the projection map are measurable with respect to $\mathbb{X}_i$ we get that the sequences $\{\widetilde{X}_{i,1}\}_{i\geq 1}, \ldots, \{\widetilde{X}_{i,p}\}_{i\geq 1}$ are i.i.d. (E1) or stationary and strongly mixing with at least the same rate as the original series (E2). Without loss of generality we can assume that the long-run variance of these random variables is equal to 1. Furthermore, note that Lemma E.1.6 (a) and Assumption B.1.6 imply $E\left(\left|\widetilde{X}_{i,m}\right|^{2+\nu}\right) < \infty$ for $m = 1, \ldots, p$. Hence, for a sequence of type (E1) the invariance principle proved by Komlós *et al.* (1975), Komlós *et al.* (1976) and Major (1976) can be applied to obtain

$$\left|\sum_{i=1}^{k} \widetilde{X}_{i,m} - kE\left(\widetilde{X}_{1,m}\right) - W(k)\right| = O\left(k^{1/(2+\nu)}\right) \quad a.s., \text{ for } k \to \infty,$$

where $W(t)$ is a standard Wiener process. We get a similar result for sequences of type (E2) by using Theorem 4 of Kuelbs & Philipp (1980). Consequently, applying Theorem 2.1. of Eichinger & Kirch (2018), which is a result on the null asymptotics of the classical MOSUM statistic, to each sequence $\{\widetilde{X}_{i,m}\}$ and Lemma E.1.6 (b) yields

$$\begin{aligned}
&\max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \left\|\sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\right\| \\
&\leq \sum_{m=1}^{p} \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \left|\sum_{i=k+1}^{k+G} \widetilde{X}_{i,m} - \sum_{i=k-G+1}^{k} \widetilde{X}_{i,m}\right| = O_P\left(\sqrt{\log(n/G)}\right).
\end{aligned}$$

Thus, together with the Cauchy-Schwarz inequality and the $\sqrt{n}$-consistency of the estimator sequence we receive

$$\begin{aligned}
&\max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \left|\left(\sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\right)^T \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right)\right| \\
&\leq \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \left\|\sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\right\| \left\|\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right\| \\
&= O_P\left(\sqrt{\frac{\log(n/G)}{n}}\right) = o_P\left((\log(n/G))^{-1/2}\right),
\end{aligned}$$

since $\frac{\log(n/G)}{\sqrt{n}} \leq \frac{\log(n)}{\sqrt{n}} \to 0$. Now, the next step is to approximate the remainder term of the Taylor expansion in (2.23). With the measurability of the second derivatives of $\boldsymbol{H}$ we get that the random variables $\nabla^2 H_l(\mathbb{X}_1, \boldsymbol{\theta}), \ldots, \nabla^2 H_l(\mathbb{X}_n, \boldsymbol{\theta})$ are i.i.d. (E1) or stationary and strongly mixing (E2) and, thus, stationary and ergodic. Hence, as

$$E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_1, \boldsymbol{\theta})\right\|_F\right) < \infty$$

by Assumption B.1.4 the Uniform Law of Large Numbers of Theorem E.2.6 shows

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n} \left\|\sum_{i=1}^{n} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\theta})\right\|_F \tag{2.24}$$

$$\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{i=1}^{n} \left\|\nabla^2 H_{l,0}(\mathbb{X}_i,\boldsymbol{\theta})\right\|_F + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|E\left(\nabla^2 H_l(\mathbb{X}_1,\boldsymbol{\theta})\right)\right\|_F = O(1) \ \ a.s.,$$

since $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|E(\nabla^2 H_l(\mathbb{X}_1,\boldsymbol{\theta}))\|_F \leq E(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla^2 H_l(\mathbb{X}_1,\boldsymbol{\theta})\|_F) = O(1)$. This implies

$$\begin{aligned}
&\left\|\sum_{i=k+1}^{k+G} \nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k}) - \sum_{i=k-G+1}^{k} \nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k})\right\|_F \\
&\leq (k+G)\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{k+G}\left\|\sum_{i=1}^{k+G}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\theta})\right\|_F + 2k\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\frac{1}{k}\left\|\sum_{i=1}^{k}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\theta})\right\|_F \\
&\quad + (k-G)\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\frac{1}{k-G}\left\|\sum_{i=1}^{k-G}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\theta})\right\|_F \\
&= O(k+G) = O(n) \ \ a.s. \text{ uniformly in } k.
\end{aligned}$$

Thus, combining Lemma E.1.5 with the $\sqrt{n}$-consistency of the estimator sequence yields

$$\begin{aligned}
&\max_{G\leq k\leq n-G}\frac{1}{\sqrt{G}}\left\|\left(\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}\right)^T\left(\sum_{i=k+1}^{k+G}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k})\right.\right. \\
&\qquad\qquad\qquad \left.\left. - \sum_{i=k-G+1}^{k}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k})\right)\left(\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}\right)\right\| \\
&\leq \frac{1}{\sqrt{G}}\left\|\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}\right\|^2 \max_{G\leq k\leq n-G}\left\|\sum_{i=k+1}^{k+G}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k}) - \sum_{i=k-G+1}^{k}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,n,k})\right\|_F \\
&= O_P\left(\frac{1}{\sqrt{G}}\right) = o_P\left((\log(n/G))^{-1/2}\right).
\end{aligned}$$

Finally, with Lemma E.1.6 (b) we can conclude that

$$\begin{aligned}
&\max_{G\leq k\leq n-G}\frac{1}{\sqrt{G}}\left\|\boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k}-\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}\right\| \\
&\leq \sum_{l=1}^{p}\max_{G\leq k\leq n-G}\frac{1}{\sqrt{G}}\left|\sum_{i=k+1}^{k+G}\left(H_l(\mathbb{X}_i,\widehat{\boldsymbol{\theta}}_n)-H_l(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})\right)\right. \\
&\qquad\qquad\qquad \left. - \sum_{i=k-G+1}^{k}\left(H_l(\mathbb{X}_i,\widehat{\boldsymbol{\theta}}_n)-H_l(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})\right)\right| \\
&= o_P\left((\log(n/G))^{-1/2}\right).
\end{aligned}$$

- Assumption A.1.3:
  Under Condition (E1), we know that $\boldsymbol{H}(\mathbb{X}_1,\widetilde{\boldsymbol{\theta}}),\ldots,\boldsymbol{H}(\mathbb{X}_n,\widetilde{\boldsymbol{\theta}})$ are i.i.d. With Assumption B.1.5 Theorem 2 of Einmahl (1989) can be applied to receive

$$\left\|\boldsymbol{\Sigma}^{-1/2}\left(\sum_{i=1}^{k}\boldsymbol{H}(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}})-kE\left(\boldsymbol{H}(\mathbb{X}_1,\widetilde{\boldsymbol{\theta}})\right)\right)-\boldsymbol{W}(k)\right\| = O(k^{1/(2+\nu)}) \ \ a.s.$$

For further explanation we refer to chapter 1 of Aue (2003).
A similiar result can be derived for sequences of type (E2) by using Theorem 4 of Kuelbs & Philipp (1980) since $\boldsymbol{H}(\mathbb{X}_1, \widetilde{\boldsymbol{\theta}}), \ldots, \boldsymbol{H}(\mathbb{X}_n, \widetilde{\boldsymbol{\theta}})$ are strongly mixing with at least the same rate as the original sequence by the measurability of $\boldsymbol{H}$ with respect to $\mathbb{X}_i$.

□

2.3.1.2. General Z-Estimators

In the introductory chapter, we have already considered classical Z-estimators or M-estimators which are determined by solving the estimating equation system $\sum_{i=1}^{n} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) \stackrel{!}{=} \mathbf{0}$. Here, we investigate the asymptotic behavior of a broader class of Z-estimators based only on a part of the sample. Therefore, we define the general Z-estimators $\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}$ as the solution of

$$\frac{1}{n} \sum_{i=\lfloor \gamma_1 n \rfloor + 1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) \stackrel{!}{=} \mathbf{0}, \quad \text{for every } n, \tag{2.25}$$

where $\gamma_1, \gamma_2 \in [0, 1]$ and $\gamma_1 < \gamma_2$. Furthermore, let $\boldsymbol{\theta}_0$ be the unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right)$, i.e. $\boldsymbol{\theta}_0$ is the true parameter vector under the null hypothesis in a correctly specified model and the best approximating parameter under misspecification as in Assumption A.1.2.
Now we want to prove that the estimator sequence $\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}$ is $\sqrt{n}$-consistent for $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$. Firstly, we have to show that $\left\| \widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0 \right\| = o_P(1)$, i.e. it is consistent for $\boldsymbol{\theta}_0$.

Lemma 2.3.3. *Let $\{\mathbb{X}_i : i \geq 1\}$ be a series of type (E1) or type (E2) fulfilling Assumptions B.1.1 and B.1.3. Then,*

$$\left\| \boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} \right\| = o_P(1).$$

Proof. Consistency of the estimator sequence can be proved by applying Lemma E.2.11. However, first of all we have to show that the assumptions of this lemma are fulfilled here.
The Uniform Law of Large Numbers in Theorem E.2.8 can be used to derive the uniform convergence condition. Therefore, we check if the assumptions of this theorem are satisfied as well. First, note that $\boldsymbol{H}$ is measurable with respect to $\mathbb{X}_i$ and that $\{\mathbb{X}_i\}_{i\geq 1}$ is i.i.d. or stationary and strongly mixing and therefore stationary and ergodic. Moreover, condition (i) of Theorem E.2.8 holds by Assumption B.1.1. By a first order Taylor expansion and Lemma E.1.5 we get

$$\left\| \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\theta}_1) - \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\theta}_2) \right\| \leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\theta}) \right\|_F \left\| \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \right\|, \tag{2.26}$$

which is well defined at least almost surely with respect to $P_{\mathbb{X}_1}$ since $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta}) \right\|_F\right) < \infty$ (condition (iii)) holds by Assumption B.1.3. The continuity of the supremum and the Frobenius norm in combination with the measurability of the first derivatives of the estimating function imply that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\|_F$ is a measurable function on $\mathbb{X}_i$ which completes condition (ii) of Theorem E.2.8. Consequently, applying Theorem E.2.8 yields

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - (\gamma_2 - \gamma_1)\, E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| \\
&\stackrel{D}{=} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=1}^{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - (\gamma_2 - \gamma_1)\, E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| \\
&\leq \frac{\lfloor\gamma_2 n\rfloor - \lfloor\gamma_1 n\rfloor}{n} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{\lfloor\gamma_2 n\rfloor - \lfloor\gamma_1 n\rfloor} \sum_{i=1}^{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta}) \right\| \\
&\quad + \left(\frac{\lfloor\gamma_2 n\rfloor - \lfloor\gamma_1 n\rfloor}{n} - (\gamma_2 - \gamma_1) \right) \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| \\
&= o_P(1) + o(1) \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\|,
\end{aligned}$$

since $\frac{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor}{n} \to \gamma_2 - \gamma_1$ as n goes to infinity. Furthermore, note that $E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right)$ is (Lipschitz) continuous in $\boldsymbol{\theta}$ since due to (2.26) we have

$$\begin{aligned}
&\left\| E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta_2})\right) - E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta_1})\right) \right\| \leq E\left(\left\| \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta_2}) - \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta_1}) \right\| \right) \\
&\leq E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\|_F \right) \|\boldsymbol{\theta_2} - \boldsymbol{\theta_1}\|,
\end{aligned}$$

and $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\|_F\right) < \infty$ by Assumption B.1.3. Hence, together with the compactness of the parameter space we obtain

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - (\gamma_2 - \gamma_1)\, E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o_P(1). \tag{2.27}$$

Moreover, Lemma E.2.10 shows that $\boldsymbol{\theta}_0$ is the unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right)$ in the strict sense. Finally, $\left\| \widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0 \right\| = o_P(1)$ follows from Lemma E.2.11. □

Asymptotic properties of the global estimator sequence $\widehat{\boldsymbol{\theta}}_{0,1}$ have already been investigated in other papers and books. For instance, asymptotic normality was shown by Van der Vaart (2007) (Theorem 5.41 on page 68) for the i.i.d. case. Similar arguments can be used to derive asymptotic normality of general Z-estimators as well which is demonstrated in the following theorems.

Theorem 2.3.4. *Let $\{\mathbb{X}_i : i \geq 1\}$ be a series of type (E1) fulfilling the Assumptions B.1.1 to B.1.4. Furthermore, let $\boldsymbol{V}(\boldsymbol{\theta}_0) = E\left(\nabla \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta}_0)\right)^T$ be a non-singular matrix, where $\nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) = (\nabla H_1(\mathbb{X}_i, \boldsymbol{\theta}), \ldots, \nabla H_p(\mathbb{X}_i, \boldsymbol{\theta}))$ denotes the matrix of gradients with respect to $\boldsymbol{\theta}$. Then,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0 \right) = -(\gamma_2 - \gamma_1)^{-1} \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) + o_P(1),$$

as well as

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right)\xrightarrow{D} N_p\left(\mathbf{0},(\gamma_2-\gamma_1)^{-1}\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1})^T\right).$$

Proof. This proof is based on the proof of Theorem 5.41 in Van der Vaart (2007) on page 68 for the univariate setting.
By componentwise Taylor series expansions around $\boldsymbol{\theta}_0$ there exist $\boldsymbol{\xi}_{1,k,n},\ldots,\boldsymbol{\xi}_{p,k,n}$ such that $\left\|\boldsymbol{\xi}_{l,k,n}-\boldsymbol{\theta}_0\right\|\leq\left\|\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right\|$, for $l=1,\ldots,p$, with

$$\begin{aligned}
&-\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}H_l(\mathbb{X}_i,\boldsymbol{\theta}_0)\\
&=\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla H_l(\mathbb{X}_i,\boldsymbol{\theta}_0)^T\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right)\\
&\quad+\frac{1}{2}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right)^T\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,k,n})\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right)\\
&=\left(\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla H_l(\mathbb{X}_i,\boldsymbol{\theta}_0)^T+\frac{1}{2n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right)^T\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,k,n})\right)\\
&\qquad\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right).
\end{aligned}$$

Since $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\nabla^2 H_l(\mathbb{X}_1,\boldsymbol{\theta})\|_F\right)<\infty$ by Assumption B.1.4 and as the measurability of the second derivatives on $\mathbb{X}_i$ yields that $\nabla^2 H_l(\mathbb{X}_1,\boldsymbol{\theta}),\ldots,\nabla^2 H_l(\mathbb{X}_n,\boldsymbol{\theta})$ are i.i.d. and therefore stationary and ergodic, for all $l=1,\ldots,p$, the Uniform Law of Large Numbers in Corollary E.2.7 can be applied to obtain

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\theta})\right\|_F \qquad (2.28)\\
&\overset{D}{=}\frac{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor}{n}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor}\sum_{i=1}^{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\theta})\right\|_F\\
&\leq\frac{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor}{n}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor}\sum_{i=1}^{\lfloor\gamma_2 n\rfloor-\lfloor\gamma_1 n\rfloor}\nabla^2 H_{l,0}(\mathbb{X}_i,\boldsymbol{\theta})\right\|_F\right.\\
&\qquad\left.+E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^2 H_l(\mathbb{X}_1,\boldsymbol{\theta})\right\|_F\right)\right)\\
&=O_P(1).
\end{aligned}$$

Hence, together with Lemma 2.3.3 and Lemma E.1.5 we get

$$\left\|\frac{1}{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_0\right)^T\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,k,n})\right\|$$

$$\leq \left\| \widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0 \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\theta}) \right\|_F = o_P(1).$$

Furthermore, note that $E\left(\|\nabla H_l(\mathbb{X}_1, \boldsymbol{\theta}_0)\|\right) < \infty$ by Assumption B.1.3 and Lemma E.1.6 (c) and that the sequence $\{\nabla H_l(\mathbb{X}_i, \boldsymbol{\theta}_0)\}_{i\geq 1}$ is i.i.d. due to the measurability of the first derivatives with respect to $\mathbb{X}_i$. Thus, we obtain

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \nabla H_l(\mathbb{X}_i, \boldsymbol{\theta}_0) - (\gamma_2 - \gamma_1)\, E\left(\nabla H_l(\mathbb{X}_1, \boldsymbol{\theta}_0)\right) \right\| \\
&\overset{D}{=} \left\| \frac{1}{n} \sum_{i=1}^{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \nabla H_l(\mathbb{X}_i, \boldsymbol{\theta}_0) - (\gamma_2 - \gamma_1)\, E\left(\nabla H_l(\mathbb{X}_1, \boldsymbol{\theta}_0)\right) \right\| \\
&\leq \left(\frac{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor}{n} - (\gamma_2 - \gamma_1) \right) \left\| E\left(\nabla H_l(\mathbb{X}_1, \boldsymbol{\theta}_0)\right) \right\| \\
&\quad + \frac{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor}{n} \left\| \frac{1}{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \sum_{i=1}^{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \nabla H_{l,0}(\mathbb{X}_i, \boldsymbol{\theta}_0) \right\| \\
&= o_P(1),
\end{aligned}$$

where the last line follows from

$$\left\| \frac{1}{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \sum_{i=1}^{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \nabla H_{l,0}(\mathbb{X}_i, \boldsymbol{\theta}_0) \right\| = o_P(1) \tag{2.29}$$

as given by the Law of Large Numbers and since $\lim_{n\to\infty} \frac{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor}{n} = \gamma_2 - \gamma_1$. This yields

$$\frac{1}{n} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \nabla H_l(\mathbb{X}_i, \boldsymbol{\theta}_0) = (\gamma_2 - \gamma_1) E\left(\nabla H_l(\mathbb{X}_1, \boldsymbol{\theta}_0)\right) + o_P(1).$$

By combining the Taylor expansions of all components, we receive

$$-\frac{1}{n} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) = \left((\gamma_2 - \gamma_1)\boldsymbol{V}(\boldsymbol{\theta}_0) + o_P(1)\right) \left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0 \right).$$

Since $f(\boldsymbol{V}) = \boldsymbol{V}^{-1}$ is a continuous function on the elements of $\boldsymbol{V}$ by Theorem 5.19 in Schott (1997) on page 188 the Continuous Mapping Theorem gives

$$\left((\gamma_2 - \gamma_1)\boldsymbol{V}(\boldsymbol{\theta}_0) + o_P(1)\right)^{-1} = (\gamma_2 - \gamma_1)^{-1} \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1} + o_P(1).$$

Thus, we get

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0 \right) = -\left((\gamma_2 - \gamma_1)^{-1} \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1} + o_P(1) \right) \frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0). \tag{2.30}$$

With Assumption B.1.2 we apply a multivariate version of the Central Limit Theorem, which can be easily derived by using the Cramér-Wold Theorem and the univariate CLT, on the sequence $\{\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0)\}_{i\geq 1}$. Hence, we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) \stackrel{D}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) \tag{2.31}$$

$$= \sqrt{\frac{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor}{n}} \frac{1}{\sqrt{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor}} \sum_{i=1}^{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) \stackrel{D}{\to} N_p\left(\mathbf{0}, (\gamma_2 - \gamma_1)\boldsymbol{\Sigma}\right),$$

which also implies that $\frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) = O_P(1)$. Consequently, we can conclude that

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0\right) = -(\gamma_2 - \gamma_1)^{-1} \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) + o_P(1)$$

$$\stackrel{D}{\to} N_p\left(\mathbf{0}, (\gamma_2 - \gamma_1)^{-1} \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{\Sigma} (\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1})^T\right).$$

□

Theorem 2.3.5. *Let $\{\mathbb{X}_i : i \geq 1\}$ be a series of type (E2) fulfilling the Assumptions B.1.1, B.1.3, B.1.4 and B.1.5. Then, the result of Theorem 2.3.4 remains true.*

Proof. The proof is analogous to that of Theorem 2.3.4 with the exception that the statements in (2.29) and (2.31) are derived by using different arguments as explained in the following.

On noting that the pattern of the original sequence described by type (E2) is inherited by the sequence $\nabla H_l(\mathbb{X}_1, \boldsymbol{\theta}_0), \ldots, \nabla H_l(\mathbb{X}_n, \boldsymbol{\theta}_0)$ due to the measurability of the first derivatives, together with Assumption B.1.3 we get that the sequence is stationary and ergodic with existing first moment. Hence, the Ergodic Theorem shows (2.29). Furthermore, Assumption B.1.5 enables us to apply a strong invariance principle by Kuelbs & Philipp (1980) (Theorem 4) to the sequence $\{\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0)\}_{i\geq 1}$. We receive

$$\left\| \sum_{i=1}^{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) - \widetilde{\boldsymbol{W}}(\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor) \right\| = O\left(n^{1/(2+\nu)}\right) a.s.,$$

where $\widetilde{\boldsymbol{W}}(t)$ denotes a p-dimensional Wiener process with covariance matrix $\boldsymbol{\Sigma}$. Thus, in connection with the stationarity of the sequence we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) \stackrel{D}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \widetilde{\boldsymbol{W}}(\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor) + o_P(1).$$

Furthermore, the self-similarity of the Wiener process and the almost sure continuity of its paths lead to

$$\frac{1}{\sqrt{n}} \widetilde{\boldsymbol{W}}(\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor) \stackrel{D}{=} \widetilde{\boldsymbol{W}}\left(\frac{\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor}{n}\right) \stackrel{P}{\to} \widetilde{\boldsymbol{W}}(\gamma_2 - \gamma_1),$$

which proves (2.31) since $\widetilde{\boldsymbol{W}}(\gamma_2 - \gamma_1) \sim N_p\left(\mathbf{0}, (\gamma_2 - \gamma_1)\boldsymbol{\Sigma}\right)$. □

Corollary 2.3.6 ($\sqrt{n}$-Consistency)**.** *Under the assumptions of Theorem 2.3.4 and Theorem 2.3.5, respectively, the estimator sequence* $\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}$ *is* $\sqrt{n}$*-consistent for* $\boldsymbol{\theta}_0$*.*

Proof. Applying Theorem 2.3.4 or Theorem 2.3.5 yields $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_0\right) = O_P(1)$ which shows the assertion. □

2.3.2. Under the Alternative

Under the alternative we have to take into consideration that the sequence $\{\mathbb{X}_i\}_{i\geq 1}$ is only piecewise stationary, i.e.

$$\mathbb{X}_i = \begin{cases} \mathbb{X}_i^{(1)}, & \text{if} \quad i \leq k_{1,n} \\ \mathbb{X}_i^{(2)}, & \text{if} \quad k_{1,n} < i \leq k_{2,n} \\ \vdots & \\ \mathbb{X}_i^{(q+1)}, & \text{if} \quad i > k_{q,n} \end{cases} .$$

Furthermore, we assume that the estimator sequence $\widehat{\boldsymbol{\theta}}_n$ is $\sqrt{n}$-consistent for some $\widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ under the alternative and that $\{\mathbb{X}_i^{(j)}\}_{i\geq 1}$, $j = 1, \ldots, q+1$, are sequences of type (E1) or (E2) satisfying Assumption A.2.2 and the following moment conditions which are summarized in Section B.2 again.

B.2.1 *Let* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|\right) < \infty$ *hold for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*,* $j = 1, \ldots, q+1$*.*

B.2.2 *Let* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right\|^2\right) < \infty$*,* $j = 1, \ldots, q+1$*.*

B.2.3 *Let* $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$*,* $j = 1, \ldots, q+1$*.*

B.2.4 $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$ *hold, for all* $l = 1, \ldots, p$*,* $j = 1, \ldots, q+1$*.*

B.2.5 *There exists a* $\nu > 0$ *such that* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right\|^{2+\nu}\right) < \infty$*,* $j = 1, \ldots, q+1$*.*

B.2.6 *There exists a* $\nu > 0$ *such that* $E\left(\left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F^{2+\nu}\right) < \infty$ *holds for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*,* $j = 1, \ldots, q+1$*.*

2.3.2.1. Assumptions A.2.3, A.2.4 and A.2.9

In this part we show that the main assumptions of Theorem 2.1.8 are fulfilled by time series belonging to type (E1) or (E2).

Lemma 2.3.7. *Let* $\{\mathbb{X}_i^{(j)} : i \geq 1\}$*,* $j = 1, \ldots, q+1$*, be series of type (E1) or type (E2) fulfilling the Assumptions B.2.4, B.2.5 and B.2.6. Furthermore, let Assumption A.1.1 hold on the bandwidth and let* $\widehat{\boldsymbol{\theta}}_n$ *be an estimator sequence which is* $\sqrt{n}$*-consistent for some* $\widetilde{\boldsymbol{\theta}}$ *under the alternative. Then, the Assumptions A.2.3, A.2.4 and A.2.9 are satisfied.*

Proof. • Assumption A.2.3:
We want to derive a strong invariance principle as in Assumption A.1.3. Thus, the assertion follows from Lemma 2.3.2 as the sequence $\{\mathbb{X}_i^{(j)}\}$, $j = 1, \ldots, q+1$, satisfies all the assumptions of this lemma.

• Assumptions A.2.4 and A.2.9:
At first we consider the set $B_{n,G}$ containing all time points lying in a G-environment of a change point. By a Taylor expansion of each component ($l = 1, \ldots, p$) there exists a $\boldsymbol{\xi}_{l,n,k}$ such that $\left\|\boldsymbol{\xi}_{l,n,k} - \widetilde{\boldsymbol{\theta}}\right\| \leq \left\|\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right\|$ with

$$\begin{aligned}
&\sum_{i=k+1}^{k+G} H_l(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_n) - \sum_{i=k-G+1}^{k} H_l(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_n) - \left(\sum_{i=k+1}^{k+G} H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\right) \\
&= \left(\sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\right)^T \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right) \qquad (2.32) \\
&\quad + \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right)^T \left(\sum_{i=k+1}^{k+G} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k}) - \sum_{i=k-G+1}^{k} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k})\right) \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right).
\end{aligned}$$

Furthermore, we assume that $k_{j,n} < k \leq k_{j,n} + G$ while noting that the following statements can be derived similarly for $k_{j,n} - G < k \leq k_{j,n}$. Thus, we receive

$$\begin{aligned}
&\sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \\
&= \sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k_{j,n}} \nabla H_l(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k_{j,n}+1}^{k} \nabla H_l(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}).
\end{aligned}$$

Since the sequence $\{\mathbb{X}_i^{(j)}\}_{i\geq 1}$, $j = 1, \ldots, q+1$, belongs to type (E1) or (E2) and the first derivatives of the estimating function are measurable with respect to $\mathbb{X}_i$ we get that $\{\nabla H_1(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\}_{i\geq 1}, \ldots, \{\nabla H_p(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\}_{i\geq 1}$, $j = 1, \ldots, q+1$, are i.i.d. or stationary and strongly mixing as well. Furthermore, with $\nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) = (\nabla H_1(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}), \ldots, \nabla H_p(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}))$, Assumption B.2.6 and Lemma E.1.6 we obtain

$$E\left(\left\|\nabla H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|^{2+\nu}\right) < \infty,$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $l = 1, \ldots, p$ and $j = 1, \ldots, q+1$. Hence, applying Lemma E.2.14 together with Assumption B.2.6 yields

$$\frac{1}{\sqrt{G}}\left\|\sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})\right\|$$

$$
\begin{aligned}
&\leq \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k+G} \nabla H_{l,0}(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}) \right\| + \frac{1}{\sqrt{G}} \left\| \sum_{i=k_{j,n}+1}^{k} \nabla H_{l,0}(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}) \right\| \\
&\quad + \frac{1}{\sqrt{G}} \left\| \sum_{i=k-G+1}^{k_{j,n}} \nabla H_{l,0}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\| \\
&\quad + \frac{k_{j,n} - k + G}{\sqrt{G}} \left\| E\left(\nabla H_l(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right) - E\left(\nabla H_l(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right) \right\| \\
&= O_P\left(\sqrt{G}\right) \text{ uniformly in } k \in \{k_{j,n}+1, \ldots, k_{j,n}+G\},
\end{aligned}
$$

with $\nabla H_{l,0}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) := \nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) - E\left(\nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right)$. In connection with the submultiplicativity of the Euclidean norm and the $\sqrt{n}$-consistency of the estimator sequence we are able to approximate the first summand of the Taylor expansion in (2.32)

$$
\begin{aligned}
&\frac{1}{\sqrt{G}} \left| \left(\sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \right)^T \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}} \right) \right| \\
&\leq \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k+G} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \nabla H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \right\| \left\| \widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}} \right\| = O_P\left(\sqrt{\frac{G}{n}} \right),
\end{aligned}
$$

which holds uniformly in $k \in \{k_{j,n}+1, \ldots, k_{j,n}+G\}$.
Moreover, the Assumption B.2.4 can be combined with the Uniform Law of Large Numbers in Theorem E.2.6 as in (2.24) which leads to

$$
\begin{aligned}
&\left\| \sum_{i=k+1}^{k+G} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k}) - \sum_{i=k-G+1}^{k} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k}) \right\|_F \\
&= \left\| \sum_{i=k+1}^{k+G} \nabla^2 H_l(\mathbb{X}_i^{(j+1)}, \boldsymbol{\xi}_{l,n,k}) - \sum_{i=k-G+1}^{k_{j,n}} \nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\xi}_{l,n,k}) \right. \\
&\qquad \left. - \sum_{i=k_{j,n}+1}^{k} \nabla^2 H_l(\mathbb{X}_i^{(j+1)}, \boldsymbol{\xi}_{l,n,k}) \right\|_F \\
&\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \sum_{i=1}^{k+G} \nabla^2 H_l(\mathbb{X}_i^{(j+1)}, \boldsymbol{\theta}) \right\|_F + 2 \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \sum_{i=1}^{k} \nabla^2 H_l(\mathbb{X}_i^{(j+1)}, \boldsymbol{\theta}) \right\|_F \\
&\quad + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \sum_{i=1}^{k_{j,n}} \nabla^2 H_l(\mathbb{X}_i^{(j+1)}, \boldsymbol{\theta}) \right\|_F + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \sum_{i=1}^{k_{j,n}} \nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \\
&\quad + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \sum_{i=1}^{k-G} \nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \\
&= O(k+G) = O(n) \;\; a.s. \text{ uniformly in } k \in \{k_{j,n}+1, \ldots, k_{j,n}+G\}.
\end{aligned}
$$

Hence, with Lemma E.1.5 and the $\sqrt{n}$-consistency of the estimator sequence we receive, uniformly in $k \in \{k_{j,n}+1, \dots, k_{j,n}+G\}$,

$$\begin{aligned}
&\left\| \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right)^T \left(\sum_{i=k+1}^{k+G} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k}) - \sum_{i=k-G+1}^{k} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k}) \right) \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}\right) \right\| \\
&\leq \left\| \widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}} \right\|^2 \left\| \sum_{i=k+1}^{k+G} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k}) - \sum_{i=k-G+1}^{k} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,n,k}) \right\|_F \\
&= O_P(1) .
\end{aligned}$$

Thus, by considering (2.32) we can conclude

$$\begin{aligned}
&\max_{k \in B_{n,G}} \frac{1}{\sqrt{G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| \qquad (2.33)\\
&\leq \sum_{l=1}^{p} \max_{k \in B_{n,G}} \frac{1}{\sqrt{G}} \left| \sum_{i=k+1}^{k+G} \left(H_l(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_n) - H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \right) \right. \\
&\qquad \left. - \sum_{i=k-G+1}^{k} \left(H_l(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_n) - H_l(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \right) \right| \\
&= O_P\left(\sqrt{\frac{G}{n}} \right) + O_P\left(\frac{1}{\sqrt{G}} \right) = o_P\left((\log(n/G))^{-1/2} \right),
\end{aligned}$$

where the last line follows from Assumption A.1.1 since $\frac{n}{G} \to \infty$ implies $\frac{\log(n/G)}{n/G} \to 0$. Furthermore, on noting that the subsequences $\mathbb{X}_{k-G+1}, \dots, \mathbb{X}_{k+G}$ are stationary for all $k \in A_{n,G}$, similar arguments as in Lemma 2.3.2 can be applied here to get

$$\max_{k \in A_{n,G}} \frac{1}{\sqrt{G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = o_P\left((\log(n/G))^{-1/2} \right).$$

Hence, we receive $\max_{G \leq k \leq n-G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = o_P\left((\log(n/G))^{-1/2} \right)$ showing Assumptions A.2.4 and A.2.9.

□

2.3.2.2. General Z-Estimators

Similar to our considerations under the null hypothesis, we examine the behavior of general Z-estimators $\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}$ defined by (2.25) and prove $\sqrt{n}$-consistency for these estimator sequences under the alternative.
Therefore, let s be the number of change points lying in the interval (γ_1, γ_2): $s := |(\gamma_1, \gamma_2) \cap \mathcal{C}_n|$, where $\mathcal{C}_n := \{k_{1,n}, \dots, k_{q,n}\}$ denotes the set of all change points. Furthermore, we define

$$j^* := \min\{j \in \{1, \dots, q+1\} : (\lambda_j - \gamma_1) > 0\}$$

and we consider the sequence

$$\widetilde{\lambda}_{j^*-1}, \widetilde{\lambda}_{j^*}, \dots, \widetilde{\lambda}_{j^*+s} \qquad (2.34)$$

with $\widetilde{\lambda}_{j^*-1} = \gamma_1, \widetilde{\lambda}_{j^*+s} = \gamma_2$ and $\widetilde{\lambda}_j = \lambda_j$ for $j \in \{j^*, \ldots, j^* + s - 1\}$

Moreover, let $\boldsymbol{\theta}_{\gamma_1,\gamma_2}$ be the unique zero of

$$\boldsymbol{F}_{\gamma_1,\gamma_2}(\boldsymbol{\theta}) = \sum_{j=j^*}^{j^*+s} \left(\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}\right) E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right). \tag{2.35}$$

Now we want to show that the estimator sequence $\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}$ is $\sqrt{n}$-consistent for $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_{\gamma_1,\gamma_2}$ under the alternative. At first, we derive classical consistency in the following lemma.

Lemma 2.3.8. *Let $\{\mathbb{X}_i : i \geq 1\}$ be a series of type (E1) or type (E2) fulfilling the Assumptions B.2.1 and B.2.3. Then,*

$$\left\|\boldsymbol{\theta}_{\gamma_1,\gamma_2} - \widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}\right\| = o_P(1).$$

Proof. The proof of this result is similar to that of Lemma 2.3.3. By (2.27) it holds, for $j \in \{j^*, \ldots, j^* + s - 1\}$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) - \left(\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}\right) E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right) \right\| = o_P(1),$$

as by Assumptions B.2.1 and B.2.3 $\{\mathbb{X}_i^{(j)}\}$ fulfills all the assumptions of Lemma 2.3.3. This implies together with (2.35)

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}_{\gamma_1,\gamma_2}(\boldsymbol{\theta}) \right\| \\
&= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \sum_{j=j^*}^{j^*+s} \frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) - \left(\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}\right) E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right) \right\| \\
&\leq \sum_{j=j^*}^{j^*+s} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) - \left(\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}\right) E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right) \right\| \\
&= o_P(1),
\end{aligned}$$

which shows that the uniform convergence condition of Lemma E.2.11 holds. Moreover, as the expectation of the estimating function is continuous on $\boldsymbol{\Theta}$ we get the continuity of $\boldsymbol{F}_{\gamma_1,\gamma_2}(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$. Thus, by Lemma E.2.10 we receive that $\boldsymbol{\theta}_{\gamma_1,\gamma_2}$ is the unique zero of $\boldsymbol{F}_{\gamma_1,\gamma_2}(\boldsymbol{\theta})$ in the strict sense. Finally, $\left\|\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)} - \boldsymbol{\theta}_{\gamma_1,\gamma_2}\right\| = o_P(1)$ follows from Lemma E.2.11. □

With the result of the lemma above we are able to derive $\sqrt{n}$-consistency for general Z-estimators. This is shown in the theorem below where we additionally get asymptotic normality of the estimator sequence in the i.i.d. case.

Theorem 2.3.9. *Let the matrix* $\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})$ *be invertible with* $\widetilde{\lambda}_j$ *as in (2.34). Furthermore, let*

(a) $\{\mathbb{X}_i^{(1)}:i\geq 1\},\ldots,\{\mathbb{X}_i^{(q+1)}:i\geq 1\}$ *be sequences of type (E1) fulfilling the Assumptions B.2.1, B.2.3, B.2.4 and B.2.2 with* $\widetilde{\boldsymbol{\theta}}=\boldsymbol{\theta}_{\gamma_1,\gamma_2}$. *Then,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)\xrightarrow{D}N_p\left(\mathbf{0},\boldsymbol{V}^{-1}\left(\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{\Sigma}_{(j)}\right)\left(\boldsymbol{V}^T\right)^{-1}\right),$$

where $\boldsymbol{\Sigma}_{(j)}=\boldsymbol{\Sigma}_{(j)}(\boldsymbol{\theta}_{\gamma_1,\gamma_2})$ *denoting the long-run covariance matrix of* $\boldsymbol{H}(\mathbb{X}_i^{(j)},\boldsymbol{\theta}_{\gamma_1,\gamma_2})$ *and* $\boldsymbol{V}=\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})$. *In particular, this implies that the estimator sequence* $\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}$ *is* $\sqrt{n}$*-consistent for* $\boldsymbol{\theta}_{\gamma_1,\gamma_2}$.

(b) $\{\mathbb{X}_i^{(1)}:i\geq 1\},\ldots,\{\mathbb{X}_i^{(q+1)}:i\geq 1\}$ *be sequences of type (E2) fulfilling the Assumptions B.2.1, B.2.3, B.2.4 and B.2.5 with* $\widetilde{\boldsymbol{\theta}}=\boldsymbol{\theta}_{\gamma_1,\gamma_2}$. *Then, the estimator sequence* $\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}$ *is* $\sqrt{n}$*-consistent for* $\boldsymbol{\theta}_{\gamma_1,\gamma_2}$.

Proof. The proof is similar to that of Theorem 2.3.4 and Theorem 2.3.5.
By componentwise second order Taylor series expansions around $\boldsymbol{\theta}_{\gamma_1,\gamma_2}$ there exist $\boldsymbol{\xi}_{1,\gamma_1,\gamma_2,n},\ldots,\boldsymbol{\xi}_{p,\gamma_1,\gamma_2,n}$ such that $\left\|\boldsymbol{\xi}_{l,\gamma_1,\gamma_2,n}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right\|\leq\left\|\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right\|$ for $l=1,\ldots,p$ with

$$\begin{aligned}
&-\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}H_l(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})\\
&=\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla H_l(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})^T\left(\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)\\
&\quad+\frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)^T\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,\gamma_1,\gamma_2,n})\left(\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)\\
&=\left(\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla H_l(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})^T+\frac{1}{2n}\left(\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)^T\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla^2 H_l(\mathbb{X}_i,\boldsymbol{\xi}_{l,\gamma_1,\gamma_2,n})\right)\\
&\qquad\left(\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right).
\end{aligned}$$

On noting that the second derivatives of the estimating function are measurable with respect to $\mathbb{X}_i$, the i.i.d. or stationary, strong mixing structure of the original sequence $\{\mathbb{X}_i^{(j)}\}_{i\geq 1}$, $j=1,\ldots,q+1$, carries over to the transformed sequence $\{\nabla^2 H_l(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\}_{i\geq 1}$, $l=1,\ldots,p$, i.e. $\{\nabla^2 H_1(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\}_{i\geq 1},\ldots,\{\nabla^2 H_p(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\}_{i\geq 1}$, $j=1,\ldots,q+1$, are random sequences of type (E1) or type (E2) and, consequently, stationary and ergodic. Hence, by (2.28) (on noting that $\{\mathbb{X}_i^{(j)}\}$ fufills the assumptions of Theorem 2.3.4) we get, for $j \in \{j^*, \dots, j^*+s-1\}$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = O_P(1).$$

This implies

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\theta}) \right\|_F = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \sum_{j=j^*}^{j^*+s} \frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\theta}) \right\|_F \\
&\leq \sum_{j=j^*}^{j^*+s} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = O_P(1).
\end{aligned}$$

Hence, applying Lemma 2.3.8 and Lemma E.1.5 yields

$$\begin{aligned}
&\left\| \frac{1}{n} \left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)^T \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\xi}_{l,\gamma_1,\gamma_2,n}) \right\| \\
&\leq \left\| \widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_{\gamma_1,\gamma_2} \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \nabla^2 H_l(\mathbb{X}_i, \boldsymbol{\theta}) \right\|_F = o_P(1).
\end{aligned}$$

Furthermore, note that the pattern of the original sequence $\{\mathbb{X}_i^{(j)}\}_{i\geq 1}$, $j = 1, \dots, q+1$, described by type (E1) or (E2) is inherited by the sequence $\{\nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2})\}$, $l = 1, \dots, p$, due to the measurability of the first derivatives. Together with Assumption B.2.3 we get that the transformed sequence is stationary and ergodic with existing first moment. Hence, the Ergodic Theorem yields

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) - \left(\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}\right) E\left(\nabla H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2})\right) \right\| \\
&\stackrel{D}{=} \left\| \frac{1}{n} \sum_{i=1}^{\lfloor\widetilde{\lambda}_j n\rfloor - \lfloor\widetilde{\lambda}_{j-1}n\rfloor} \nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) - \left(\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}\right) E\left(\nabla H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2})\right) \right\| \\
&\leq \left(\frac{\lfloor\widetilde{\lambda}_j n\rfloor - \lfloor\widetilde{\lambda}_{j-1}n\rfloor}{n} - (\gamma_2 - \gamma_1) \right) \left\| E\left(\nabla H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2})\right) \right\| \\
&\quad + \frac{\lfloor\widetilde{\lambda}_j n\rfloor - \lfloor\widetilde{\lambda}_{j-1}n\rfloor}{n} \left\| \frac{1}{\lfloor\widetilde{\lambda}_j n\rfloor - \lfloor\widetilde{\lambda}_{j-1}n\rfloor} \sum_{i=1}^{\lfloor\widetilde{\lambda}_j n\rfloor - \lfloor\widetilde{\lambda}_{j-1}n\rfloor} \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_0) \right\| \\
&= o_P(1), \quad j \in \{j^*, \dots, j^*+s-1\}
\end{aligned}$$

which shows that

$$\frac{1}{n} \sum_{i=\lfloor\widetilde{\lambda}_{j-1}n\rfloor+1}^{\lfloor\widetilde{\lambda}_j n\rfloor} \nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) = (\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}) E\left(\nabla H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2})\right) + o_P(1).$$

Thus, we obtain

$$\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\nabla H_l(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})=\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)E\left(\nabla H_l(\mathbb{X}_1^{(j)},\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)+o_P(1).$$

By merging the Taylor expansions of all components, we get

$$\begin{aligned}&-\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})\\&=\left(\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})+o_P(1)\right)\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right).\end{aligned}$$

Since the matrix $\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})$ is invertible by assumption and as $f(\boldsymbol{V})=\boldsymbol{V}^{-1}$ is a continuous function on the elements of $\boldsymbol{V}$ (see Theorem 5.19 in Schott (1997) page 188) the Continuous Mapping Theorem leads to

$$\left(\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})+o_P(1)\right)^{-1}=\left(\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)^{-1}+o_P(1).$$

Thus, we receive

$$\begin{aligned}&\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)\\&=-\left(\left(\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)^{-1}+o_P(1)\right)\frac{1}{\sqrt{n}}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2}).\end{aligned}$$

Furthermore, on noting that (with $\boldsymbol{F}_{\gamma_1,\gamma_2}$ as in (2.35))

$$\begin{aligned}&\left\|\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}E\left(\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)\right\|=\left\|\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}E\left(\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)-n\boldsymbol{F}_{\gamma_1,\gamma_2}(\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right\|\\&\leq\sum_{j=j*}^{j^*+s}\left|(\widetilde{\lambda}_{j-1}n-\lfloor\widetilde{\lambda}_{j-1}n\rfloor)-(\widetilde{\lambda}_j n-\lfloor\widetilde{\lambda}_j n\rfloor)\right|\left\|E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)\right\|\\&\leq\sum_{j=j^*}^{j^*+s}\left\|E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)\right\|\leq\sum_{j=j^*}^{j^*+s}E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right\|\right)=O(1),\end{aligned}$$

we obtain

$$\begin{aligned}&\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}-\boldsymbol{\theta}_{\gamma_1,\gamma_2}\right)\qquad(2.36)\\&=-\left(\left(\sum_{j=j^*}^{j^*+s}\left(\widetilde{\lambda}_j-\widetilde{\lambda}_{j-1}\right)\boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)^{-1}+o_P(1)\right)\frac{1}{\sqrt{n}}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor}\boldsymbol{H}_0(\mathbb{X}_i,\boldsymbol{\theta}_{\gamma_1,\gamma_2})+O\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

For proving part (a), with Assumption B.2.2 we apply a multivariate version of the Central Limit Theorem on the sequence $\{\boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2})\}_{i\geq 1}$, $j = 1, \ldots, q+1$. and get

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=\lfloor \widetilde{\lambda}_{j-1} n\rfloor+1}^{\lfloor \widetilde{\lambda}_j n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) \overset{D}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor \widetilde{\lambda}_j n\rfloor - \lfloor \widetilde{\lambda}_{j-1} n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) \\
&= \sqrt{\frac{\lfloor \widetilde{\lambda}_j n\rfloor - \lfloor \widetilde{\lambda}_{j-1} n\rfloor}{n}} \frac{1}{\sqrt{\lfloor \widetilde{\lambda}_j n\rfloor - \lfloor \widetilde{\lambda}_{j-1} n\rfloor}} \sum_{i=1}^{\lfloor \widetilde{\lambda}_j n\rfloor - \lfloor \widetilde{\lambda}_{j-1} n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) \\
&\overset{D}{\to} N_p\left(\boldsymbol{0}, (\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1})\boldsymbol{\Sigma}_{(j)}\right),
\end{aligned}$$

implying $\frac{1}{\sqrt{n}} \sum_{i=\lfloor \widetilde{\lambda}_{j-1} n\rfloor+1}^{\lfloor \widetilde{\lambda}_j n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) = O_P(1)$ as well. Furthermore, by the latter one and equation (2.36) we receive

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\gamma_1,\gamma_2}^{(n)} - \boldsymbol{\theta}_{\gamma_1,\gamma_2}\right) \\
&= -\left(\sum_{j=j^*}^{j^*+s} \left(\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1}\right) \boldsymbol{V}_j(\boldsymbol{\theta}_{\gamma_1,\gamma_2})\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n\rfloor+1}^{\lfloor \gamma_2 n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) + o_P(1).
\end{aligned}$$

Moreover, since the sequences $\mathbb{X}_{\lfloor \widetilde{\lambda}_{j-1} n\rfloor+1}^{(j)}, \ldots, \mathbb{X}_{\lfloor \widetilde{\lambda}_j n\rfloor}^{(j)}$, $j = 1, \ldots, q+1$, are independent we can conclude that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n\rfloor+1}^{\lfloor \gamma_2 n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) &= \sum_{j=j^*}^{j^*+s} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \widetilde{\lambda}_{j-1} n\rfloor+1}^{\lfloor \widetilde{\lambda}_j n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) \\
&\overset{D}{\to} N_p\left(\boldsymbol{0}, \sum_{j=j^*}^{j^*+s} (\widetilde{\lambda}_j - \widetilde{\lambda}_{j-1})\boldsymbol{\Sigma}_{(j)}\right),
\end{aligned}$$

which shows the assertion of part (a).
In order to derive the assertion in (b), note that Assumption A.2.3 holds by Condition B.2.5 by Lemma 2.3.7. Hence, for every $j \in \{1, \ldots, q+1\}$ there exists a p-dimensional Wiener process $\{\widetilde{\boldsymbol{W}}^{(j)}(t)\}_{t\geq 0}$ with covariance matrix $\boldsymbol{\Sigma}_{(j)}$ such that

$$\left\| \sum_{i=1}^{n} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) - \widetilde{\boldsymbol{W}}^{(j)}(n) \right\| = O(n^{1/(2+\nu)}) \ \ a.s.$$

This yields together with the self-similarity of the Wiener process and Lemma E.1.5

$$\begin{aligned}
&\left\| \frac{1}{\sqrt{n}} \sum_{i=\lfloor \widetilde{\lambda}_{j-1} n\rfloor+1}^{\lfloor \widetilde{\lambda}_j n\rfloor} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) \right\| \\
&\leq \frac{1}{\sqrt{n}} \left\| \widetilde{\boldsymbol{W}}^{(j)}(\lfloor \widetilde{\lambda}_j n\rfloor) - \widetilde{\boldsymbol{W}}^{(j)}(\lfloor \widetilde{\lambda}_{j-1} n\rfloor) \right\| + O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{n}}\right)
\end{aligned}$$

$$\stackrel{D}{=} \left\| \widetilde{\boldsymbol{W}}^{(j)}\left(\frac{\lfloor \widetilde{\lambda}_j n \rfloor}{n}\right) - \widetilde{\boldsymbol{W}}^{(j)}\left(\frac{\lfloor \widetilde{\lambda}_{j-1} n \rfloor}{n}\right) \right\| + O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{n}}\right)$$
$$\leq 2 \left\| \boldsymbol{\Sigma}^{1/2} \right\|_F \sup_{0\leq t\leq 1} \| \boldsymbol{W}(t) \| + O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{n}}\right)$$
$$= O_P(1),$$

where the last line follows from the almost sure continuity of the paths of a Wiener process and the compactness of $[0,1]$. Thus, we get

$$\frac{1}{\sqrt{n}} \sum_{i=\lfloor \gamma_1 n \rfloor+1}^{\lfloor \gamma_2 n \rfloor} \boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) = \sum_{j=j^*}^{j^*+s} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \widetilde{\lambda}_{j-1} n \rfloor+1}^{\lfloor \widetilde{\lambda}_j n \rfloor} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_{\gamma_1,\gamma_2}) = O_P(1).$$

which implies $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} - \boldsymbol{\theta}_{\gamma_1,\gamma_2}\right) = O_P(1)$ by (2.36). □

Let us finish this subsection with a result linking to the discussion on the problem in detectability which somehow depends on the choice of the global estimator. The following lemma tells us that, under some regularity conditions, by taking the classical Z-estimator in the MOSUM statistic at least one change is detectable or even all changes are detectable if there are only two possible regimes, i.e. the parameter vector only switches between two values.

Lemma 2.3.10. *Let the assumptions of Theorem 2.3.9 be satisfied. Furthermore, let $\boldsymbol{\theta}_j$ denote the unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)$, for $j \in \{1,\ldots,q+1\}$ and let $\boldsymbol{\theta}_l \neq \boldsymbol{\theta}_{l+1}$ hold for all $l = 1,\ldots,q$. Moreover, suppose that the classical Z-estimator $\widehat{\boldsymbol{\theta}}^{(n)}_{0,1}$ is used in the MOSUM score-type statistic, $T_{k,n}(G, \widehat{\boldsymbol{\theta}}^{(n)}_{0,1})$.*

(a) Then, at least one change is detectable by the MOSUM statistic, i.e. $\widetilde{q} \geq 1$ with $\widetilde{q}$ as in Assumption A.2.7.

(b) If there are only two distinct regimes then all changes are detectable, i.e. $\widetilde{q} = q$.

Proof. By Theorem 2.3.9 we know that $\widehat{\boldsymbol{\theta}}^{(n)}_{0,1}$ is $\sqrt{n}$-consistent for $\boldsymbol{\theta}_{0,1}$ which is the unique zero of

$$\boldsymbol{F}_{0,1}(\boldsymbol{\theta}) = \sum_{j=1}^{q+1} (\lambda_j - \lambda_{j-1})\, E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)$$

according to (2.35). Hence, for proving the result of part (a) we have to show that $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{0,1})\right) \neq E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_{0,1})\right)$ holds for at least one $j \in \{1,\ldots,q\}$. Let us now assume that this not true, which means

$$E\left(\boldsymbol{H}(\mathbb{X}_1^{(1)}, \boldsymbol{\theta}_{0,1})\right) = \ldots = E\left(\boldsymbol{H}(\mathbb{X}_1^{(q+1)}, \boldsymbol{\theta}_{0,1})\right).$$

Then, by definition of $\boldsymbol{\theta}_{0,1}$ we would get

$$\boldsymbol{0} = \boldsymbol{F}_{0,1}(\boldsymbol{\theta}_{0,1}) = \sum_{j=1}^{q+1} (\lambda_j - \lambda_{j-1})\, E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{0,1})\right) = E\left(\boldsymbol{H}(\mathbb{X}_1^{(1)}, \boldsymbol{\theta}_{0,1})\right),$$

implying that $\boldsymbol{\theta}_j = \boldsymbol{\theta}_{0,1}$, for all $j = 1, \ldots, q+1$, which contradicts the assumption. Consequently, we can conclude that there exists at least one $j \in \{1, \ldots, q\}$ with $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{0,1})\right) \neq E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_{0,1})\right)$.
This immediately implies the assertion in (b) as in such a case

$$\left| E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_{0,1})\right) - E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_{0,1})\right) \right|$$

are equal for all $j \in \{1, \ldots, q\}$. □

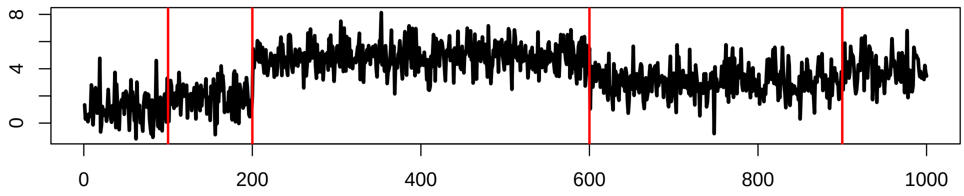
2.4. Possible Problems of the Procedure

2.4.1. The Choice of the Bandwidth

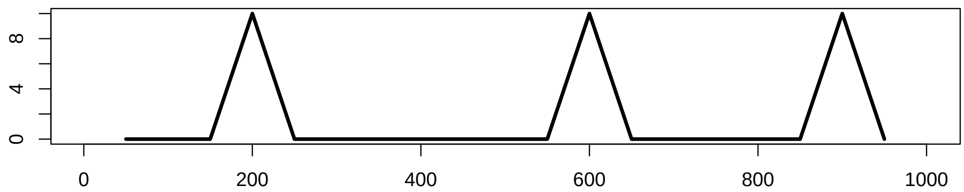
In the Section 2.1.3 we could derive consistency for the estimator sequences which is some kind of quality criterion theoretically justifying the usage of the MOSUM procedure. Nevertheless, this property only holds asymptotically but in practice we want to apply the procedure on data sets of finite sample size. Thus, we need to assess how well or bad the procedure performs on finite samples and yet some questions come up: How do we choose the length of the moving window for finite samples? Is there an optimal bandwidth?
Eichinger & Kirch (2018) already recognized that the choice of the bandwidth has a decisive influence on the performance of the MOSUM procedure in the classical mean change model. This holds for the general model as well as we will see in the simulations of Chapter 4.
In theory it would make sense to choose the bandwidth G as large as possible so that the minimal distance between two adjacent structural breaks is still greater than $2G$. This would ensure that in each time point the statistic is contaminated by at most one change implying that the signal of the statistic takes its local maxima only at the locations of the detectable changes as described in Section 2.1.3.1. However, for real data sets the change points and its distances to each other are unknown so that there is no possibility to check whether a bandwidth G satisfies the condition above or not. Moreover, for localizing small changes lying far from any other change, we need a large bandwidth whereas detecting large changes being close to other changes requires a small window length. Hence, problems arise if we have a combination of both scenarios and only one bandwidth is used. One possible solution is to run the procedure with different bandwidths and merge the results in an appropriate way. Cho & Kirch (2018) have introduced such a multiscale method for the classical mean change model which can be adapted to our general setting. Further explanations on that and first results are given in Chapter 5.

2.4.2. The Problem in Detectability

In Section 2.1.3.2, we have already discussed when a change can be detected by MOSUM score-type statistics, at least asymptotically. We have learned that the signal of the statistic needs to be strictly positive in an interval around the change which is the case if a change in the parameter vector of the underlying distribution causes a change in the expectation of $\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})$ as well. Hence, the number of detectable changes somehow depends on the choice of the estimating function. There are examples in which all changes are actually detectable by this MOSUM procedure, e.g. the classical MOSUM statistic in the mean change model or the MOSUM score-type statistic based on the least squares method in a simple linear regression model. However, this is not true in general and there exist some examples, even for the classical mean change setting, where the number of detectable changes is not equal to the total number of changes. Let us consider the following example.

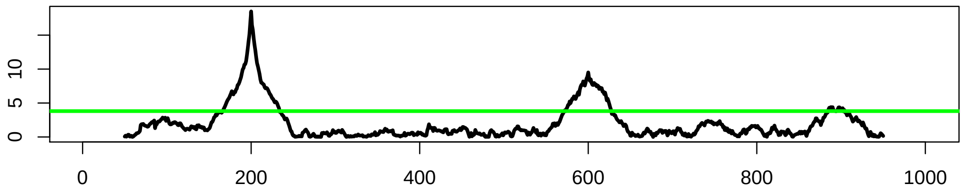


The plot shows a time series which randomly deviates from a mean value changing at the time points $100, 200, 600, 900$. By applying the MOSUM statistic based on the estimating function of the classical median, which is $H(x, \mu) = \operatorname{sgn}(x - \mu)$, we would not be able to localize the first change point as the signal of the statistic is equal to zero at this point as illustrated in the graph below.

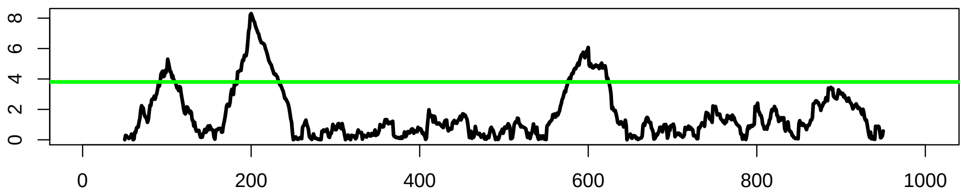


Nevertheless, as shown at the end of Section 2.3, under some regularity conditions in the general model at least one change point is detectable if the global estimator computed on the whole sample is used in the statistic. Thus, we could theoretically find all changes by doing some kind of binary segmentation as explained in the following. At first we determine the estimated change points obtained by the global Z-estimator computed on the whole sample. After splitting the data sequence in these points into different segments, an Z-estimator can be computed on each segment and employed in the score-type statistic to detect further changes. By recursively repeating this step we would be able to detect all changes for large n.

However, even if we choose the estimating function such that the signal of the MOSUM score-type statistic is strictly positive in every change point, which ensures detectability in the asymptotics, we could still fail in localizing some of the changes in finite samples. For instance, if we use the median-like estimator and its estimating function, $H(x, \mu) = \frac{2}{\pi} \arctan(\mu - x)$, for the example above and compute the corresponding MOSUM score-type statistic ($G = 50$) we are again not able to detect the first change point in 100 as plotted below.



In this statistic we compute the median-like estimator on the whole sample, $\widehat{\mu}_{1,n}$, and employ it in the estimating function to get the transformed series, $H(X_i, \widehat{\mu}_{1,n}) = \frac{2}{\pi}\arctan(\widehat{\mu}_{1,n} - X_i)$, on which the MOSUM statistic is based. However, our theoretical results are not restricted to this specific global estimator and we can think of choosing a different global estimator for the mean μ in the statistic. Therefore, let $\widehat{\mu}_{1,100}$ denote the median-like estimator calculated on the subsample $X_1, \ldots, X_{100}$. Note that this broader class of Z-estimators computed on a part of the sample has been examined in Section 2.3. The behavior of the MOSUM score-type statistic, where the alternative estimator is used, is shown in the following graph.



Here, we succeed in finding the first change but we slightly fail in detecting the last one. Hence, applying several global estimators in the procedure and merging the results reasonably could solve the problem in detectability. Consequently, similar to the bandwidth problem, described in the previous subsection, adapting the multiscale method introduced by Cho & Kirch (2018) will be essential for improving the performance of the procedure.

3. MOSUM Wald-Type Statistics

As already discussed in Section 1.2, the classical MOSUM statistic investigated by Eichinger & Kirch (2018) is constructed by comparing the sample means of subsamples of size G around each time point. Hence, a quite natural way for generalizing this procedure to several parameter change problems would be to use MOSUM Wald-type statistics based on differences of local estimators as described in the following definition:

Definition 3.0.1 (MOSUM Wald-Type Statistic)**.** *A MOSUM Wald-type statistic based on Z-estimators is given by*

$$W_n(G) = \max_{G \le k \le n-G} W_{k,n}(G)$$
$$\text{with } W_{k,n}(G) = \frac{\sqrt{G}}{\sqrt{2}} \sqrt{\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)^T \boldsymbol{\Gamma}_k^{-1} \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)},$$

where $\widehat{\boldsymbol{\theta}}_{l,u}$ is a local Z-estimator satisfying $\sum_{i=l}^{u} \boldsymbol{H}(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_{l,u}) \overset{!}{=} \mathbf{0}$ and $\boldsymbol{\Gamma}_k$ is the asymptotic covariance matrix of $\sqrt{G}\widehat{\boldsymbol{\theta}}_{k-G+1,k}$ which is assumed to be positive definite.

Note that the matrix $\boldsymbol{\Gamma}_k$ in specified in (3.1) below.
This chapter is organized as follows. In the first section, we consider the MOSUM Wald-type statistic in a general setting under i.i.d. and strong mixing assumptions on the observations. In Section 3.1.1 the asymptotic behavior of the statistic under the null hypothesis is examined whereas in Section 3.1.2 we show consistency for the corresponding test and estimators. In Section 3.2 we focus on the linear regression model and derive similar results under the null hypothesis and the alternative.

3.1. General Setting

Throughout this section, we assume that the estimating function $\boldsymbol{H}$ is twice continuously differentiable on a compact parameter space $\boldsymbol{\Theta}$, where $\boldsymbol{H}$ and its derivatives are measurable with respect to $\mathbb{X}_i$. Furthermore, note that the same notation as in Section 2.3 is used here.

3.1.1. Asymptotics Under the Null Hypothesis

Similar to Section 2.3.1, we assume that $\mathbb{X}_1, \ldots, \mathbb{X}_n$ is a series of type (E1) or type (E2). Let $\boldsymbol{\theta}_0$ be the unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right)$ in the strict sense according to Definition E.2.9 such that $\boldsymbol{\theta}_0$ is the true parameter vector under the null hypothesis in a correctly specified model and the best approximating parameter under misspecification. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ be the long-run covariance matrix of $\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta}_0)$, which is assumed to be positive definite. Moreover, we consider the assumptions of Section B.1 with $\underline{\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0}$

and additionally introduce the following assumptions:

B.1.7 *Let* $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla^2 H_j(\mathbb{X}_1,\boldsymbol{\theta})\|_F^{2+\nu}\right) < \infty$ *hold for all* $j = 1,\ldots,p$ *and for some* $\nu > 0$.

B.1.8 *Let* $\boldsymbol{V}(\boldsymbol{\theta}) = E\left(\nabla \boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\right)^T$ *be a regular matrix for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and let*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{V}(\boldsymbol{\theta})^{-1}\right\|_F < \infty.$$

For investigating the asymptotic behavior of the Wald-type statistics we need to understand how the local Z-estimators behave if n goes to infinity. Similar to the previous section, we can derive consistency results for these estimator sequences holding pointwise in k.

Lemma 3.1.1. *Let* $\{\mathbb{X}_i : i \geq 1\}$ *be a series of type (E1) or type (E2) fulfilling the Assumptions B.1.1 and B.1.3. Then, it holds pointwise for any* $k = G,\ldots,n-G$

$$\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right\| = o_P(1) \quad \textit{and} \quad \left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \boldsymbol{\theta}_0\right\| = o_P(1)$$

i.e. the local Z-estimators are consistent for the true parameter vector.

Proof. Similar arguments as in the proof of Lemma 2.3.3 can be used here. However, note that the condition of uniform convergence in Lemma E.2.11 (statement (2.27) in Lemma 2.3.3) follows directly from the stationarity of the sequence and the Uniform Law of Large Numbers in Theorem E.2.8

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{G}\sum_{i=k+1}^{k+G} \boldsymbol{H}_0(\mathbb{X}_i,\boldsymbol{\theta})\right\| \overset{D}{=} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{G}\sum_{i=1}^{G} \boldsymbol{H}_0(\mathbb{X}_i,\boldsymbol{\theta})\right\| = o_P(1).$$

□

Moreover, the local estimator sequences asymptotically follow a normal distribution which is shown in the following two theorems.

Theorem 3.1.2. *Let* $\{\mathbb{X}_i : i \geq 1\}$ *be a series of type (E1) fulfilling the Assumptions B.1.1 to B.1.4 and let Assumption A.1.1 hold on the bandwidth. Then, it holds pointwise for any* $k = G,\ldots,n-G$

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_{\mathbf{0}}\right) = -\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}}) + o_P\left(1\right).$$

and

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \boldsymbol{\theta}_{\mathbf{0}}\right) = -\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\frac{1}{\sqrt{G}}\sum_{i=k-G+1}^{k} \boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}}) + o_P\left(1\right),$$

as well as

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_{\mathbf{0}}\right) \overset{D}{\to} N_p\left(\mathbf{0}, \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1})^T\right)$$

and

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k-G+1,k}-\boldsymbol{\theta}_{\mathbf{0}}\right)\xrightarrow{D}N_p\left(\mathbf{0},\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1})^T\right),$$

where $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ denotes the long-run covariance matrix of $\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta}_0)$, which is assumed to be positive definite.

Proof. Similar arguments as in (2.30) in the proof of Theorem 2.3.4 can be used to get

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\boldsymbol{\theta}_{\mathbf{0}}\right)=-\left(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}+o_P(1)\right)\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}}).$$

Applying a multivariate version of the Central Limit Theorem and the stationarity of the sequence yields

$$\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}})\overset{D}{=}\frac{1}{\sqrt{G}}\sum_{i=1}^{G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}})\xrightarrow{D}N_p\left(\mathbf{0},\boldsymbol{\Sigma}\right),$$

which implies

$$\begin{aligned}&\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\boldsymbol{\theta}_{\mathbf{0}}\right)\\&=-\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}})+o_P(1)\xrightarrow{D}N_p\left(\mathbf{0},\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1})^T\right).\end{aligned}$$

The second statement can be derived similarly. □

Theorem 3.1.3. *Let $\{\mathbb{X}_i : i\geq 1\}$ be a series of type (E2) fulfilling the Assumptions B.1.1, B.1.3, B.1.4 and B.1.5 and let Assumption A.1.1 hold on the bandwidth. Then, the result of Theorem 3.1.2 remains true.*

Proof. Similar arguments as for (2.30) in the proof of Theorem 2.3.4 can be used here where the statement in (2.29) is given by the Ergodic Theorem which can be applied by Assumption B.1.3 and as $\nabla H_l(\mathbb{X}_1,\boldsymbol{\theta}_0),\ldots,\nabla H_l(\mathbb{X}_n,\boldsymbol{\theta}_0)$ is of type (E2) due to the measurability of the first derivatives. Thus, we get

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\boldsymbol{\theta}_{\mathbf{0}}\right)=-\left(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}+o_P(1)\right)\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}}).$$

Furthermore, Assumption B.1.5 enables us to apply a strong invariance principle by Kuelbs & Philipp (1980) (Theorem 4) to the sequence $\{\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_0)\}_{i\geq 1}$ which is of type (E2) since the estimating function is measurable with respect to $\mathbb{X}_i$. We receive

$$\left\|\sum_{i=1}^{G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_{\mathbf{0}})-\widetilde{\boldsymbol{W}}(G)\right\|=O\left(G^{1/(2+\nu)}\right)\ a.s.,$$

where $\widetilde{\boldsymbol{W}}(t)$ denotes a p-dimensional Wiener process with covariance matrix $\boldsymbol{\Sigma}$. Hence, with the stationarity of the sequence in connection with the self-similarity of the Wiener process we receive

$$\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta_0}) \stackrel{D}{=} \frac{1}{\sqrt{G}}\sum_{i=1}^{G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta_0}) = \frac{1}{\sqrt{G}}\widetilde{\boldsymbol{W}}(G) + o_P(1) \stackrel{D}{=} \widetilde{\boldsymbol{W}}(1) + o_P(1),$$

which proves the assertion since $\widetilde{\boldsymbol{W}}(1) \sim N_p\left(\boldsymbol{0},\boldsymbol{\Sigma}\right)$.
The second statement can be shown similarly. □

The results of Theorem 3.1.2 and Theorem 3.1.3 also show that the local estimator sequences are $\sqrt{G}$-consistent for $\boldsymbol{\theta}_0$ which is given in the following corollary.

Corollary 3.1.4 ($\sqrt{G}$-Consistency). *Under the assumptions of Theorem 3.1.2 or Theorem 3.1.3, the estimator sequences $\widehat{\boldsymbol{\theta}}_{k+1,k+G}$ and $\widehat{\boldsymbol{\theta}}_{k-G+1,k}$ are (pointwise for any $k = G,\dots,n-G$) $\sqrt{G}$-consistent for $\boldsymbol{\theta}_0$.*

Proof. By Theorem 3.1.2 or Theorem 3.1.3 we get $\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\boldsymbol{\theta}_0\right) = O_P(1)$ and $\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k-G+1,k}-\boldsymbol{\theta}_0\right) = O_P(1)$ showing the assertion. □

Moreover, applying these theorems enables us to specify the asymptotic covariance matrix

$$\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma} = \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1})^T, \quad \text{for all } k\in\{G,\dots,n-G\}. \tag{3.1}$$

In addition, Theorem 3.1.2 gives the asymptotic covariance matrix of the difference of the local estimators in the i.i.d. case. Since the local estimators are computed on disjoint and therefore independent subsamples we obtain

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\widehat{\boldsymbol{\theta}}_{k-G+1,k}\right) \stackrel{D}{\to} N_p\left(\boldsymbol{0}, 2\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}(\boldsymbol{\theta}_0)^{-1})^T\right).$$

This justifies the use of the factor $\frac{1}{\sqrt{2}}$ in the statistic.
Nevertheless, we have to be aware of that all these results, including Lemma 3.1.1, only hold pointwise and they do not need to hold uniformly in k. However, in order to derive a limit distribution for the MOSUM Wald-type statistic we need the following uniform statement:

$$\begin{aligned}&\max_{0\le k\le n-G}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_0) - \frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{V}(\boldsymbol{\theta}_0)\left(\boldsymbol{\theta}_0-\widehat{\boldsymbol{\theta}}_{k+1,k+G}\right)\right)\right\| \\ &= o_P\left((\log(n/G))^{-1/2}\right).\end{aligned} \tag{3.2}$$

The following lemma provides a uniform convergence rate for the local Z-estimators. This will help us to show that Conditon (3.2) is satisfied under some moment conditions.

Lemma 3.1.5. *Let $\{\mathbb{X}_i : i\ge 1\}$ be a series of type (E1) or type (E2) satisfying the Assumptions B.1.5, B.1.6, B.1.7 and B.1.8. Then*

$$\max_{G\le k\le n-G}\sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\boldsymbol{\theta}_0\right\| = O_P\left(\sqrt{\log(n/G)}\right)$$

Proof. A Taylor expansion in $\widehat{\boldsymbol{\theta}}_{k+1,k+G}$ around $\boldsymbol{\theta}_0$ yields that there exists a $\boldsymbol{\xi}_{k,n}$ with $\left\|\boldsymbol{\xi}_{k,n} - \boldsymbol{\theta}_0\right\| \leq \left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right\|$ such that

$$-\frac{1}{\sqrt{G}} \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta_0}) = \left(\frac{1}{G} \sum_{i=k+1}^{k+G} \nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\xi}_{k,n})\right)^T \sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right). \tag{3.3}$$

At first, we want to show that the assumptions of Theorem E.2.16 are fulfilled. Since we assume that $\{\mathbb{X}_i\}_{i\geq 1}$ is of type (E1) or type (E2) and that the first derivatives of the estimating function are measurable with respect to $\mathbb{X}_i$ only Conditions (i) to (iii) of the theorem have to be checked.
On noting that $\nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) = (\nabla H_1(\mathbb{X}_i, \boldsymbol{\theta}), \ldots, \nabla H_p(\mathbb{X}_i, \boldsymbol{\theta}))$, Assumption B.1.6 in connection with Lemma E.1.6 (c) yields $E\left(\|\nabla H_j(\mathbb{X}_1, \boldsymbol{\theta})\|^{2+\nu}\right) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, hence (i). Moreover, by a first order Taylor expansion we get

$$\|\nabla H_j(\boldsymbol{x}, \boldsymbol{\theta}) - \nabla H_j(\boldsymbol{x}, \boldsymbol{\xi})\| \leq \left\|\nabla^2 H_j(\boldsymbol{x}, \boldsymbol{\eta})\right\|_F \|\boldsymbol{\theta} - \boldsymbol{\xi}\| \leq \sup_{\boldsymbol{\eta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_j(\boldsymbol{x}, \boldsymbol{\eta})\right\|_F \|\boldsymbol{\theta} - \boldsymbol{\xi}\|,$$

which is well defined at least almost surely with respect to $P_{\mathbb{X}_1}$ since $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla^2 H_j(\mathbb{X}_1, \boldsymbol{\theta})\|_F\right) < \infty$ by Assumption B.1.7. The continuity of the supremum and the Frobenius norm in combination with the measurability of the second derivatives of the estimating function imply that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla^2 H_j(\cdot, \boldsymbol{\theta})\|_F$ is a measurable function with respect to $\mathbb{X}_i$. Hence, Condition (ii) of Theorem E.2.16 is satisfied. Furthermore, we have $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla^2 H_j(\mathbb{X}_1, \boldsymbol{\theta})\|_F^{2+\nu}\right) < \infty$ by Assumption B.1.7 which shows Condition (iii) in the theorem. Consequently, applying Theorem E.2.16 in connection with Lemma E.1.6 (d) yields

$$\begin{aligned}\max_{1\leq k\leq n-G} \left\|\frac{1}{G} \sum_{i=k+1}^{k+G} \nabla \boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\xi}_{k,n})\right\|_F &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{1\leq k\leq n-G} \left\|\frac{1}{G} \sum_{i=k+1}^{k+G} \nabla \boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta})\right\|_F \\ &\leq \sum_{j=1}^{p} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{1\leq k\leq n-G} \left\|\frac{1}{G} \sum_{i=k+1}^{k+G} \nabla H_{j,0}(\mathbb{X}_i, \boldsymbol{\theta})\right\| = o_P(1).\end{aligned} \tag{3.4}$$

Hence, by considering the Taylor expansion in (3.3) again, we obtain

$$-\frac{1}{\sqrt{G}} \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) = \left(o_P(1) + \boldsymbol{V}(\boldsymbol{\xi}_{k,n})\right) \sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right) \text{ uniformly in } k.$$

which shows together with Lemma E.2.21 and Assumption B.1.8

$$-\boldsymbol{V}(\boldsymbol{\xi}_{k,n})^{-1} \frac{1}{\sqrt{G}} \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) = (o_P(1) + \boldsymbol{I}_p) \sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right) \text{ uniformly in } k.$$

On noting that $E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta}_0)\right) = \boldsymbol{0}$ and $E\left(\|\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta}_0)\|^{2+\nu}\right) < \infty$ by Assumption B.1.5, Corollary E.2.13 can be used since the pattern of the original sequence described by type

(E1) or type (E2) is inherited by the sequence $\{\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_0)\}_{i\geq 1}$ due to the measurability of $\boldsymbol{H}$. We receive

$$\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_0)=O_P\left(\sqrt{\log(n/G)}\right)\quad\text{uniformly in } k. \tag{3.5}$$

Thus, applying Lemma E.2.21 in combination with Assumption B.1.8 yields

$$O_P\left(\sqrt{\log(n/G)}\right)=(o_P(1)+\boldsymbol{I}_p)\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\boldsymbol{\theta}_0\right)\ \text{uniformly in } k.$$

Hence, the assertion follows from Lemma E.2.22. □

Remark 3.1.6. *The condition in Assumption B.1.8 that* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{V}(\boldsymbol{\theta})^{-1}\|_F<\infty$ *is satisfied if*

$$\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\lambda_{min}\left(\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{V}(\boldsymbol{\theta})^T\right)>c,\ \textit{for some } c>0,$$

where $\lambda_{min}(\cdot)$ *denotes the smallest eigenvalue of a matrix.*

Proof. Note that the spectral norm of a matrix $\boldsymbol{A}$ is defined by (see e.g. Horn & Johnson (1990) on page 295)

$$\|\boldsymbol{A}\|_2=\sqrt{\lambda_{max}(\boldsymbol{A}^T\boldsymbol{A})}$$

with $\lambda_{max}(\cdot)$ denoting the largest eigenvalue of a matrix. Now, the result can be shown by using the following inequality of the Frobenius norm and the spectral norm:

$$\|\boldsymbol{A}\|_F=\sqrt{\operatorname{tr}(\boldsymbol{A}^T\boldsymbol{A})}=\sqrt{\sum_{i=1}^{p}\lambda_i\left(\boldsymbol{A}^T\boldsymbol{A}\right)}\leq\sqrt{p\lambda_{max}\left(\boldsymbol{A}^T\boldsymbol{A}\right)}=\sqrt{p}\,\|\boldsymbol{A}\|_2\,,$$

where $\lambda_i(\cdot)$ represents an eigenvalue of a matrix.
Hence, we obtain

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{V}(\boldsymbol{\theta})^{-1}\right\|_F\leq\sqrt{p}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{V}(\boldsymbol{\theta})^{-1}\right\|_2=\sqrt{p}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sqrt{\lambda_{max}\left((\boldsymbol{V}(\boldsymbol{\theta})^{-1})^T\boldsymbol{V}(\boldsymbol{\theta})^{-1}\right)}\\&=\sqrt{p}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\frac{1}{\sqrt{\lambda_{min}\left(\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{V}(\boldsymbol{\theta})^T\right)}}=\sqrt{p}\frac{1}{\sqrt{\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\lambda_{min}\left(\boldsymbol{V}(\boldsymbol{\theta}))\boldsymbol{V}(\boldsymbol{\theta})^T\right)}}\leq\sqrt{\frac{p}{c}}<\infty.\end{aligned}$$

□

The following lemma shows that Assumption (3.2) can be derived under some moment conditions.

Lemma 3.1.7. *Let* $\{\mathbb{X}_i : i\geq 1\}$ *be a series of type (E1) or type (E2) satisfying the Assumptions B.1.5, B.1.6, B.1.7 and B.1.8. Then, Condition* (3.2)

$$\begin{aligned}&\max_{0\leq k\leq n-G}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta}_0)-\frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{V}(\boldsymbol{\theta}_0)\left(\boldsymbol{\theta}_0-\widehat{\boldsymbol{\theta}}_{k+1,k+G}\right)\right)\right\|\\&=o_P\left((\log(n/G))^{-1/2}\right).\end{aligned}$$

is satisfied.

Proof. By a componentwise Taylor expansion, for each component ($j = 1, \dots, p$) there exists a $\boldsymbol{\xi}_{j,n,k}$ such that $\left\|\boldsymbol{\xi}_{j,n,k} - \boldsymbol{\theta}_0\right\| \leq \left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right\|$ with

$$
\begin{aligned}
-\sum_{i=k+1}^{k+G} H_j(\mathbb{X}_i, \boldsymbol{\theta}_0) = &\left(\sum_{i=k+1}^{k+G} (\nabla H_j(\mathbb{X}_i, \boldsymbol{\theta}_0))\right)^T \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right) \\
&+ \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right)^T \left(\sum_{i=k+1}^{k+G} \nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\xi}_{j,n,k})\right)\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right).
\end{aligned}
$$

Hence, we obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{2G}} \sum_{i=k+1}^{k+G} H_j(\mathbb{X}_i, \boldsymbol{\theta}_0) \qquad (3.6)\\
&= \left(E\left(\nabla H_j(\mathbb{X}_1, \boldsymbol{\theta}_0)\right) + \frac{1}{G}\sum_{i=k+1}^{k+G} \nabla H_{j,0}(\mathbb{X}_i, \boldsymbol{\theta}_0)\right)^T \frac{\sqrt{G}}{\sqrt{2}}\left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right) \\
&\quad + \frac{\sqrt{G}}{\sqrt{8}}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right)^T \left(\frac{1}{G}\sum_{i=k+1}^{k+G} \nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\xi}_{j,n,k})\right)\left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right).
\end{aligned}
$$

Furthermore, note that the function $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_j(\cdot, \boldsymbol{\theta})\right\|_F$ is measurable with respect to $\mathbb{X}_i$ due to the continuity of the supremum and the Frobenius norm and since the second derivatives of the estimating function are measurable with respect to $\mathbb{X}_i$. Thus, the i.i.d. or strong mixing structure of the original sequence carries over to the transformed sequence, i.e. $\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\theta})\|\}_{i\geq 1}$ is a random sequence of type (E1) or type (E2) as well. With Assumption B.1.7 and Assumption A.1.1, Corollary E.2.13 can be applied to receive

$$
\begin{aligned}
&\frac{1}{G}\left\|\sum_{i=k+1}^{k+G} \nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\xi}_{j,n,k})\right\|_F \leq \frac{1}{G}\sum_{i=k+1}^{k+G} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\theta})\right\|_F \\
&\leq \frac{1}{G}\left|\sum_{i=k+1}^{k+G}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\theta})\right\|_F - E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_j(\mathbb{X}_1, \boldsymbol{\theta})\right\|_F\right)\right)\right| \\
&\quad + E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_j(\mathbb{X}_1, \boldsymbol{\theta})\right\|_F\right) \\
&= O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right) + O(1) = O_P(1) \quad \text{uniformly in } k,
\end{aligned}
$$

where the last line follows directly from Assumption B.1.7 and Assumption A.1.1. This yields in connection with Lemma E.1.5 and Lemma 3.1.5

$$
\begin{aligned}
&\left|\frac{\sqrt{G}}{\sqrt{8}}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right)^T \left(\frac{1}{G}\sum_{i=k+1}^{k+G} \nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\xi}_{j,n,k})\right)\left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right)\right| \\
&\leq \frac{\sqrt{G}}{\sqrt{8}}\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_0\right\|^2 \left\|\frac{1}{G}\sum_{i=k+1}^{k+G} \nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\xi}_{j,n,k})\right\|_F
\end{aligned}
$$

$$= O_P\left(\frac{\log(n/G)}{\sqrt{G}}\right) = o_P\left((\log(n/G))^{-1/2}\right) \quad \text{uniformly in } k,$$

since, for large n,

$$\frac{\log(n/G)\sqrt{\log(n/G)}}{\sqrt{G}} \leq \frac{n^{1/(2+\nu)}\sqrt{\log(n)}}{\sqrt{G}} \to 0 \quad \text{as } n \to \infty$$

by Assumption A.1.1. Furthermore, with Assumption B.1.6 and as the sequence $\{\nabla H_j(\mathbb{X}_i, \boldsymbol{\theta}_0)\}_{i\geq 1}$ is of type (E1) or type (E2) due to the measurability of the first derivative Corollary E.2.13 can be used to get

$$\frac{1}{G}\sum_{i=k+1}^{k+G} \nabla H_{j,0}(\mathbb{X}_i, \boldsymbol{\theta}_0) = O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right).$$

Combining this with the result of Lemma 3.1.5, the Cauchy-Schwarz inequality and Assumption A.1.1 leads to

$$\begin{aligned}
&\left|\frac{1}{G}\sum_{i=k+1}^{k+G} \nabla H_{j,0}(\mathbb{X}_i, \boldsymbol{\theta}_0)^T \frac{\sqrt{G}}{\sqrt{2}}\left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right)\right| \\
&\leq \left\|\frac{1}{G}\sum_{i=k+1}^{k+G} \nabla H_{j,0}(\mathbb{X}_i, \boldsymbol{\theta}_0)\right\| \frac{\sqrt{G}}{\sqrt{2}} \left\|\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right\| \\
&= O_P\left(\frac{\log(n/G)}{\sqrt{G}}\right) = o_P\left((\log(n/G))^{-1/2}\right) \quad \text{uniformly in } k.
\end{aligned}$$

Thus, with (3.6) we can conclude that

$$\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G} H_j(\mathbb{X}_i, \boldsymbol{\theta}_0) = E\left(\nabla H_j(\mathbb{X}_1, \boldsymbol{\theta}_0)\right)^T \frac{\sqrt{G}}{\sqrt{2}}\left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right) + o_P\left((\log(n/G))^{-1/2}\right)$$

holds uniformly in k. Finally, we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0) \\
&= \boldsymbol{V}(\boldsymbol{\theta}_0)\frac{\sqrt{G}}{\sqrt{2}}\left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right) + o_P\left((\log(n/G))^{-1/2}\right) \quad \text{uniformly in } k,
\end{aligned}$$

which shows (3.2). □

If Assumption (3.2) is satisfied we are able to determine the limit distribution of the Wald-type statistic under the null as stated in Theorem 3.1.8. The basic idea there is to show that the Wald-type statistic is asymptotically equivalent to the score-type statistic considered in the previous chapter so that both statistics behave similarly in the limit and $a(n/G)W_n(G) - b(n/G)$ has the same limit distribution as its score-type counterpart described in Theorem 2.1.1.

Theorem 3.1.8. *Let $\{\mathbb{X}_i : i \geq 1\}$ be a series of type (E1) or type (E2) satisfying Assumption B.1.5 and let Assumption A.1.1 on the bandwidth hold. Furthermore, assume that* (3.2) *is fulfilled.*

(a) Then, under H_0,

$$a(n/G)W_n(G) - b(n/G) \overset{D}{\to} E$$

with E as Gumbel distributed random variable as in Theorem 2.1.1 and with $a(x)$ and $b(x)$ as in (2.1).

(b) The long-run covariance matrix $\boldsymbol{\Sigma}$ can be replaced by a positive definite estimator sequence $\widehat{\boldsymbol{\Sigma}}_{k,n}$ and the expectation of the gradient matrix $\boldsymbol{V}(\boldsymbol{\theta}_0)$ can be replaced by a regular estimator sequence $\widehat{\boldsymbol{V}}_{k,n}$ satisfying the following assumption

$$\max_{G\leq k\leq n-G} \left\| \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}(\boldsymbol{\theta}_0) - \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\widehat{\boldsymbol{V}}_{k,n} \right\|_F = o_P\left(\left(\log(n/G)\right)^{-1} \right)$$

without changing the result of part (a).

Proof. (a) By (3.1) we know that $\boldsymbol{\Gamma}_k^{-1/2} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}(\boldsymbol{\theta}_0)$ holds for all $k \in \{G, \ldots, n-G\}$ under the null hypothesis. Similar to the MOSUM score-type statistic (see Remark 2.0.2), we receive

$$W_n(G) = \max_{G\leq k\leq n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}(\boldsymbol{\theta}_0) \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\|.$$

On noting that

$$\begin{aligned}
&\max_{0\leq k\leq n-G} \left\| \boldsymbol{\Sigma}^{-1/2} \left(\frac{1}{\sqrt{2G}} \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta_0}) - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G} \right) \right) \right\| \\
&= \max_{G\leq l\leq n} \left\| \boldsymbol{\Sigma}^{-1/2} \left(\frac{1}{\sqrt{2G}} \sum_{i=l-G+1}^{l} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta_0}) - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{l-G+1,l} \right) \right) \right\|
\end{aligned}$$

by shifting the index to $l = k + G$, Assumption (3.2) yields

$$\begin{aligned}
&|T_n(G, \boldsymbol{\theta}_0) - W_n(G)| \\
&= \Bigg| \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{A}_{\boldsymbol{\theta}_0,k} \right\| \\
&\qquad - \max_{G\leq k\leq n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}(\boldsymbol{\theta}_0) \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\| \Bigg| \\
&\leq \max_{G\leq k\leq n-G} \left\| \boldsymbol{\Sigma}^{-1/2} \left(\frac{1}{\sqrt{2G}} \boldsymbol{A}_{\boldsymbol{\theta}_0,k} - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \widehat{\boldsymbol{\theta}}_{k+1,k+G} \right) \right) \right\| \\
&\leq \max_{0\leq k\leq n-G} \left\| \boldsymbol{\Sigma}^{-1/2} \left(\frac{1}{\sqrt{2G}} \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta_0}) - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G} \right) \right) \right\| \\
&\quad + \max_{G\leq k\leq n} \left\| \boldsymbol{\Sigma}^{-1/2} \left(\frac{1}{\sqrt{2G}} \sum_{i=k-G+1}^{k} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta_0}) - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right) \right\|
\end{aligned}$$

$$= 2 \max_{0 \le k \le n-G} \left\| \boldsymbol{\Sigma}^{-1/2} \left(\frac{1}{\sqrt{2G}} \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta_0}) - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{k+1,k+G} \right) \right) \right\|$$
$$= o_P\left((\log(n/G))^{-1/2} \right),$$

which implies that $W_n(G) = T_n(G, \boldsymbol{\theta}_0) + o_P\left((\log(n/G))^{-1/2} \right)$. Furthermore, Assumption A.1.3 of Theorem 2.1.1 is satisfied by Lemma 2.3.2 and Theorem 2.1.1 can be applied to finish the proof.

(b) The result of part (a) yields

$$W_n(G) = \max_{G \le k \le n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\|$$
$$= O_P\left(\frac{b(n/G)}{a(n/G)} \right) = O_P\left(\sqrt{\log(n/G)} \right).$$

Furthermore, by Lemma E.1.5 we receive

$$\max_{G \le k \le n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right\|$$
$$= \max_{G \le k \le n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\|$$
$$\le \left\| \boldsymbol{V}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{\Sigma}^{1/2} \right\|_F \max_{G \le k \le n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\|$$
$$= O_P\left(\sqrt{\log(n/G)} \right).$$

Hence, we obtain

$$\Bigg| \max_{G \le k \le n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{V}(\boldsymbol{\theta}_0) \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\|$$
$$- \max_{G \le k \le n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \widehat{\boldsymbol{V}}_{k,n} \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\| \Bigg|$$
$$\le \max_{G \le k \le n-G} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{V}(\boldsymbol{\theta}_0) - \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \widehat{\boldsymbol{V}}_{k,n} \right\|_F \max_{G \le k \le n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right\|$$
$$= o_P\left((\log(n/G))^{-1/2} \right),$$

which implies the assertion.

□

Remark 3.1.9. *The assumption on the estimator sequences in part (b) is fulfilled if*

$$\max_{G \le k \le n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}^{-1/2} \right\|_F = o_P\left((\log(n/G))^{-1} \right) \text{ and}$$
$$\max_{G \le k \le n-G} \left\| \widehat{\boldsymbol{V}}_{k,n} - \boldsymbol{V}(\boldsymbol{\theta}_0) \right\|_F = o_P\left((\log(n/G))^{-1} \right).$$

Proof. By using the triangle inequality and the submultiplicativity of the Frobenius norm, we receive

$$\begin{aligned}
&\max_{G\le k\le n-G}\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}(\boldsymbol{\theta}_0)-\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\widehat{\boldsymbol{V}}_{k,n}\right\|_F\\
&=\max_{G\le k\le n-G}\Big\|\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{V}(\boldsymbol{\theta}_0)-\widehat{\boldsymbol{V}}_{k,n}\right)+\left(\boldsymbol{\Sigma}^{-1/2}-\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\right)\left(\widehat{\boldsymbol{V}}_{k,n}-\boldsymbol{V}(\boldsymbol{\theta}_0)\right)\\
&\qquad\qquad+\left(\boldsymbol{\Sigma}^{-1/2}-\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\right)\boldsymbol{V}(\boldsymbol{\theta}_0)\Big\|_F\\
&\le\left\|\boldsymbol{\Sigma}^{-1/2}\right\|_F\max_{G\le k\le n-G}\left\|\boldsymbol{V}(\boldsymbol{\theta}_0)-\widehat{\boldsymbol{V}}_{k,n}\right\|_F\\
&\quad+\max_{G\le k\le n-G}\left\|\boldsymbol{V}(\boldsymbol{\theta}_0)-\widehat{\boldsymbol{V}}_{k,n}\right\|_F\max_{G\le k\le n-G}\left\|\boldsymbol{\Sigma}^{-1/2}-\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\right\|_F\\
&\quad+\|\boldsymbol{V}(\boldsymbol{\theta}_0)\|_F\max_{G\le k\le n-G}\left\|\boldsymbol{\Sigma}^{-1/2}-\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\right\|_F\\
&=o_P\left((\log(n/G))^{-1}\right).
\end{aligned}$$

□

3.1.2. Asymptotics Under the Alternative

As already described in Section 2.1.2, we consider alternatives with q structural breaks where q denotes the unknown number of change points. Under the assumption of piecewise stationarity, there exist stationary sequences $\{\mathbb{X}_i^{(j)} : i \ge 1\}$, $j = 1, \dots, q+1$, such that

$$\mathbb{X}_i = \mathbb{X}_i^{(j)}, \quad \text{for} \quad k_{j-1,n} < i \le k_{j,n}.$$

Furthermore, we assume that $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}$ is the unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)$ in the strict sense according to Definition E.2.9, for $j = 1, \dots, q+1$, and $\boldsymbol{\theta}_j \ne \boldsymbol{\theta}_{j+1}$ for all $j = 1, \dots, q$. Hence, $\boldsymbol{\theta}_j$ is the true parameter vector of the underlying distribution of $\{\mathbb{X}_i^{(j)} : i \ge 1\}$ in a correctly specified model and the best approximating parameter under misspecification. Besides, let $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_j(\boldsymbol{\theta}_j)$ denote the long-run covariance matrix of $\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_j)$ which is assumed to be positive definite. Moreover, we consider the same assumptions as in Section 2.3.2 with $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_j$ for $\{\mathbb{X}_i^{(j)}\}$. In addition to that, the following conditions are needed in this section:

B.2.7 *Let* $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^2 H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F^{2+\nu}\right) < \infty$, *for all* $l = 1, \dots, p$ *and for some* $\nu > 0$, $j = 1, \dots, q+1$.

B.2.8 *There exists a* $\nu > 0$ *such that* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|^{2+\nu}\right) < \infty$, *for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $j = 1, \dots, q+1$.

B.2.9 *Let* $\boldsymbol{V}_j(\boldsymbol{\theta}) = E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)^T$, $j = 1, \dots, q+1$, *be a regular matrix for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and let*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{V}_j(\boldsymbol{\theta})^{-1}\right\|_F < \infty.$$

B.2.10 *Let $\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta})$ be a regular matrix for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\delta \in [0,1]$ and let*

$$\sup_{\delta \in [0,1]} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \left(\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta) \boldsymbol{V}_{j+1}(\boldsymbol{\theta}) \right)^{-1} \right\|_F < \infty, \quad j = 1, \ldots, q.$$

B.2.11 *There exists a $\nu > 0$ such that $E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}) \right\|_F^{2+\nu} \right) < \infty$, for $j = 1, \ldots, q+1$.*

Note that all conditions are summarized in Section B.2.
In order to show consistency for the MOSUM-based test and estimators, we need to examine the behavior of the MOSUM Wald-type statistic on several subsets of $\{G, \ldots, n-G\}$. Remember that

$$A_{n,G} := \{k \in \{G, \ldots, n-G\} : |k - k_{j,n}| \geq G \ \ \forall j \in \{1, \ldots, q\}\}$$

is the set of all time points being far away from any change point as in (2.2). Furthermore, let

$$A_{j,n,G} := \{k \in A_{n,G} : k_{j-1,n} < k \leq k_{j,n}\} \quad (\text{for } j = 1, \ldots, q+1) \tag{3.7}$$

such that $A_{1,n,G}, \ldots, A_{q+1,n,G}$ is a partition of $A_{n,G}$ since for each k there exists exactly one $j \in \{1, \ldots, q+1\}$ with $k_{j-1,n} < k \leq k_{j,n}$. Moreover, we define

$$B_{n,G}^{(1)} := \bigcup_{j=1}^{q+1} B_{j,n,G}^{(1)} \quad \text{with } B_{j,n,G}^{(1)} := \{k \in \{G, \ldots, n-G\} : k < k_{j,n} < k + G\} \tag{3.8}$$

and

$$B_{n,G}^{(2)} := \bigcup_{j=1}^{q+1} B_{j,n,G}^{(2)} \quad \text{with } B_{j,n,G}^{(2)} := \{k \in \{G, \ldots, n-G\} : k - G < k_{j,n} \leq k\}. \tag{3.9}$$

Note that the sets $A_{1,n,G}, \ldots, A_{q+1,n,G}, B_{1,n,G}^{(1)}, \ldots, B_{q+1,n,G}^{(1)}, B_{1,n,G}^{(2)}, \ldots, B_{q+1,n,G}^{(2)}$ are pairwise disjoint for n sufficiently large by Assumption A.1.1 and A.2.1 so that they together built a partition of $\{G, \ldots, n-G\}$.
Under the null hypothesis we used the asymptotic equivalence to the score-type statistic to determine the limit distribution whereas under the alternative this might be only useful on segments far away from any change. The main advantage of the MOSUM Wald-type statistic is that a change in the parameter vector directly affects the statistic which implies that the signal of the statistic is strictly positive on the G-environment of every change. In order to use this strength in detectability, we have to examine the behavior of Wald-type statistics on intervals around the change points. In doing so

we need to understand how the local estimator sequences $\widehat{\boldsymbol{\theta}}_{k+1,k+G}$ behave under the alternative. Therefore, we introduce two functions:

$$F_1(k,n,G,\boldsymbol{\theta}) = \begin{cases} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right), & \text{if} \quad k \in A_{j,n,G} \\ E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\boldsymbol{\theta})\right), & \text{if} \quad k \in B_{j,n,G}^{(2)} \\ \frac{k_{j,n}-k}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right) + \frac{k+G-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\boldsymbol{\theta})\right), & \\ \qquad \text{if} \quad k \in B_{j,n,G}^{(1)} \end{cases} \tag{3.10}$$

and

$$F_2(k,n,G,\boldsymbol{\theta}) = \begin{cases} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right), & \text{if} \quad k \in A_{j,n,G} \\ E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right), & \text{if} \quad k \in B_{j,n,G}^{(1)} \\ \frac{k_{j,n}-k+G}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right) + \frac{k-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\boldsymbol{\theta})\right), & \\ \qquad \text{if} \quad k \in B_{j,n,G}^{(2)} \end{cases} . \tag{3.11}$$

Furthermore, let $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)}$ be the unique zero of $F_1(k,n,G,\cdot)$ and $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)}$ denote the unique zero of $F_2(k,n,G,\cdot)$. By definition of F_1 and F_2 we get $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)} = \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)} = \boldsymbol{\theta}_j$ for $k \in A_{j,n,G}$ ($j = 1,\ldots,q+1$) and $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)} = \boldsymbol{\theta}_{j+1}, \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)} = \boldsymbol{\theta}_j$ for $k = k_{j,n}$, $j = 1,\ldots,q$. In the following lemma we derive $\sqrt{G}$-consistency pointwise for all change points and all time points being far away from any change.

Lemma 3.1.10. *Let $\{\mathbb{X}_i^{(1)} : i \geq 1\},\ldots,\{\mathbb{X}_i^{(q+1)} : i \geq 1\}$ be sequences of type (E1) satisfying the Assumptions B.2.1 to B.2.4 or sequences of type (E2) fulfilling the Assumptions B.2.1, B.2.3, B.2.4 and B.2.5, with $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_j$ for $\{\mathbb{X}_i^{(j)}\}$. Furthermore, let Assumption A.1.1 on the bandwidth and Assumption A.2.1 hold. Then,*

(a) $\sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k_{j,n}+1,k_{j,n}+G} - \boldsymbol{\theta}_{j+1}\right\| = O_P(1)$ *and* $\sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k_{j,n}-G+1,k_{j,n}} - \boldsymbol{\theta}_j\right\| = O_P(1)$ *for all change points* $k_{j,n}$, $j = 1,\ldots,q$, *and*

(b) $\sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_j\right\| = O_P(1)$ *and* $\sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \boldsymbol{\theta}_j\right\| = O_P(1)$ *pointwise for all* $k \in A_{j,n,G}$.

Proof. First note that, for all change points $k_{j,n}, j = 1,\ldots,q$, $(\mathbb{X}_{k_{j,n}+1},\ldots,\mathbb{X}_{k_{j,n}+G}) = (\mathbb{X}_{k_{j,n}+1}^{(j+1)},\ldots,\mathbb{X}_{k_{j,n}+G}^{(j+1)})$ and $(\mathbb{X}_{k_{j,n}-G+1},\ldots,\mathbb{X}_{k_{j,n}}) = (\mathbb{X}_{k_{j,n}-G+1}^{(j)},\ldots,\mathbb{X}_{k_{j,n}}^{(j)})$ hold for all n. Hence, $\widehat{\boldsymbol{\theta}}_{k_{j,n}-G+1,k_{j,n}}$ and $\widehat{\boldsymbol{\theta}}_{k_{j,n}+1,k_{j,n}+G}$, respectively, is determined by solving the estimating equation system $\sum_{i=k_{j,n}-G+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)},\boldsymbol{\theta}) = \mathbf{0}$ or $\sum_{i=k_{j,n}+1}^{k_{j,n}+G} \boldsymbol{H}(\mathbb{X}_i^{(j+1)},\boldsymbol{\theta}) = \mathbf{0}$. Since the sequences $\{\mathbb{X}_i^{(1)}\},\ldots,\{\mathbb{X}_i^{(q+1)}\}$ are stationary and fulfill the assumptions of Theorem 3.1.2 or Theorem 3.1.3 the assertion in (a) follows from Corollary 3.1.4. Similar arguments can be applied to show (b). □

The lemma above is very helpful in the consistency proof for the test. However, for showing consistency of the change point estimators we need results which enable us to split the statistic into noise and signal and to approximate the noise part in a uniform way. Hence, the next lemma will be essential for the proof of Theorem 3.1.15.

Lemma 3.1.11. *Let* $\{\mathbb{X}_i^{(1)} : i \geq 1\}, \ldots, \{\mathbb{X}_i^{(q+1)} : i \geq 1\}$ *be sequences of type (E1) or (E2) satisfying the Assumptions B.2.7, B.2.8, B.2.10 and B.2.11. Furthermore, let Assumption A.1.1 on the bandwidth and Assumption A.2.1 hold. Then,*

(a) $\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)}\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right)$ *and* $\left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)}\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right)$ *uniformly on* $A_{n,G}$*,*

(b) $\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)}\right\| = o_P(1)$ *and* $\left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)}\right\| = O_P\left(\frac{1}{\sqrt{G}}\right)$ *uniformly on* $B_{n,G}^{(1)}$*,*

(c) $\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)}\right\| = O_P\left(\frac{1}{\sqrt{G}}\right)$ *and* $\left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)}\right\| = o_P(1)$ *uniformly on* $B_{n,G}^{(2)}$*.*

Proof. (a) First note that $(\mathbb{X}_{k+1}, \ldots, \mathbb{X}_{k+G}) = (\mathbb{X}_{k+1}^{(j)}, \ldots, \mathbb{X}_{k+G}^{(j)})$, $(\mathbb{X}_{k-G+1}, \ldots, \mathbb{X}_k) = (\mathbb{X}_{k-G+1}^{(j)}, \ldots, \mathbb{X}_k^{(j)})$ and $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)} = \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)} = \boldsymbol{\theta}_j$ hold for all time points $k \in A_{j,n,G}$. Thus, $\widehat{\boldsymbol{\theta}}_{k-G+1,k}$ and $\widehat{\boldsymbol{\theta}}_{k+1,k+G}$ are computed on a stationary sequence $\{\mathbb{X}_i^{(j)}\}_{i\geq 1}$ which fulfills the assumptions of Lemma 3.1.5 since Assumption B.2.6 follows from Assumption B.2.11 and as Assumption B.2.10 implies (with $\delta = 1$ or $\delta = 0$) that $\boldsymbol{V}_j(\boldsymbol{\theta})$ is regular for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{V}_j(\boldsymbol{\theta})^{-1}\|_F < \infty$, $j = 1, \ldots, q+1$. Thus, Lemma 3.1.5 can be applied here to receive

$$\max_{k\in A_{n,G}} \sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)}\right\| \leq \sum_{j=1}^{q+1} \max_{k\in A_{j,n,G}} \sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_j\right\|$$
$$= O_P\left(\sqrt{\log(n/G)}\right)$$

and

$$\max_{k\in A_{n,G}} \sqrt{G}\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)}\right\| = O_P\left(\sqrt{\log(n/G)}\right).$$

(b) We start with the proof of the second statement. Note that $(\mathbb{X}_{k-G+1}, \ldots, \mathbb{X}_k) = (\mathbb{X}_{k-G+1}^{(j)}, \ldots, \mathbb{X}_k^{(j)})$ and $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)} = \boldsymbol{\theta}_j$ hold for all time points $k \in B_{j,n,G}^{(1)}$ and that the assumptions of Lemma 3.1.5 are satisfied by $\{\mathbb{X}_i^{(j)}\}$ as explained in (a). Consequently, the same arguments as in Lemma 3.1.5 can be used here. However, the rate in (3.5) can be improved by Lemma E.2.15 as follows

$$\frac{1}{\sqrt{G}} \sum_{i=k-G+1}^{k} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_0) = O_P(1) \quad \text{uniformly in } k \in B_{n,G}^{(1)}.$$

Finally, the proof can be finished in an analogous manner to Lemma 3.1.5 with the new rate.

For the first statement we have to take into consideration that $k_{j,n} \in \{k+1, \ldots k+$

$G-1\}$ which makes it a bit more complicated. A Taylor expansion in $\widehat{\boldsymbol{\theta}}_{k+1,k+G}$ around $\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}$ yields that there exists a $\boldsymbol{\xi}_{k,n,G}$ with

$$\left\|\boldsymbol{\xi}_{k,n}-\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right\|\leq\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right\|$$

such that

$$-\frac{1}{\sqrt{G}}\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G})=\left(\frac{1}{G}\sum_{i=k+1}^{k+G}\nabla\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\xi}_{k,n,G})\right)^T\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right), \tag{3.12}$$

which is by Assumption A.2.1 equivalent to

$$\begin{aligned}&-\frac{1}{\sqrt{G}}\left(\sum_{i=k+1}^{k_{j,n}}\boldsymbol{H}(\mathbb{X}^{(j)}_i,\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G})+\sum_{i=k_{j,n}+1}^{k+G}\boldsymbol{H}(\mathbb{X}^{(j+1)}_i,\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G})\right)\\&=\left(\frac{1}{G}\sum_{i=k+1}^{k_{j,n}}\nabla\boldsymbol{H}(\mathbb{X}^{(j)}_i,\boldsymbol{\xi}_{k,n,G})+\frac{1}{G}\sum_{i=k_{j,n}+1}^{k+G}\nabla\boldsymbol{H}(\mathbb{X}^{(j+1)}_i,\boldsymbol{\xi}_{k,n,G})\right)^T\\&\qquad\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}-\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right).\end{aligned}$$

Similar to Lemma 3.1.5, one can show that the sequences $\{\mathbb{X}^{(j)}_i\}_{i\geq 1}$ or $\{\mathbb{X}^{(j+1)}_i\}_{i\geq 1}$ and the function ∇H_l satisfy the assumptions of Theorem E.2.16 coinciding with the assumptions of Lemma E.2.17. For further explanation we refer to (3.4) in the proof of Lemma 3.1.5. Hence, applying Lemma E.2.17 on $\nabla H_l(\mathbb{X}^{(j)}_i,\boldsymbol{\theta})$ in combination with Lemma E.1.6 (d) and Assumption A.1.1 yields

$$\max_{k\in B^{(1)}_{j,n,G}}\left\|\frac{1}{G}\sum_{i=k+1}^{k_{j,n}}\nabla\boldsymbol{H}_0(\mathbb{X}^{(j)}_i,\boldsymbol{\xi}_{k,n,G})\right\|_F\leq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\max_{k\in B^{(1)}_{j,n,G}}\left\|\frac{1}{G}\sum_{i=k+1}^{k_{j,n}}\nabla\boldsymbol{H}_0(\mathbb{X}^{(j)}_i,\boldsymbol{\theta})\right\|_F \tag{3.13}$$

$$\leq\sum_{l=1}^{p}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\max_{k\in B^{(1)}_{j,n,G}}\left\|\frac{1}{G}\sum_{i=k+1}^{k_{j,n}}\nabla H_{l,0}(\mathbb{X}^{(j)}_i,\boldsymbol{\theta})\right\|=o_P(1)$$

and

$$\begin{aligned}&\max_{k\in B^{(1)}_{j,n,G}}\left\|\frac{1}{G}\sum_{i=k_{j,n}+1}^{k+G}\nabla\boldsymbol{H}_0(\mathbb{X}^{(j+1)}_i,\boldsymbol{\xi}_{k,n,G})\right\|_F\\&\leq\sum_{l=1}^{p}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\max_{k\in B^{(1)}_{j,n,G}}\left\|\frac{1}{G}\sum_{i=k_{j,n}+1}^{k+G}\nabla H_{l,0}(\mathbb{X}^{(j+1)}_i,\boldsymbol{\theta})\right\|=o_P(1).\end{aligned} \tag{3.14}$$

Consequently, by considering the Taylor expansion in (3.12) again we obtain

$$-\frac{1}{\sqrt{G}}\left(\sum_{i=k+1}^{k_{j,n}}\boldsymbol{H}(\mathbb{X}^{(j)}_i,\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G})+\sum_{i=k_{j,n}+1}^{k+G}\boldsymbol{H}(\mathbb{X}^{(j+1)}_i,\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G})\right)$$

$$= \left(o_P(1) + \frac{k_{j,n} - k}{G} \boldsymbol{V}_j(\boldsymbol{\xi}_{k,n,G}) + \frac{k + G - k_{j,n}}{G} \boldsymbol{V}_{j+1}(\boldsymbol{\xi}_{k,n,G}) \right)$$
$$\sqrt{G} \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} \right) \quad \text{uniformly on } B^{(1)}_{j,n,G}.$$

In addition, Lemma E.2.17 can be used to approximate the left hand side of the equation above. Therefore, we have to verify that the assumptions of this lemma, which are explicitly described in Theorem E.2.16, are satisfied. On noting that $\{\mathbb{X}_i^{(j)}\}_{i\geq 1}$ is of type (E1) or type (E2) and that the estimating function $\boldsymbol{H}$ is measurable with respect to $\mathbb{X}_i^{(j)}$ by assumption, only Conditions (i) to (iii) of Theorem E.2.16 have to be derived.
By Assumption B.2.8 we have $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|^{2+\nu}\right) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Moreover, a first order Taylor expansion yields

$$\|\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\xi})\| \leq \|\nabla \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\eta})\|_F \|\boldsymbol{\theta} - \boldsymbol{\xi}\| \leq \sup_{\boldsymbol{\eta}\in\boldsymbol{\Theta}} \|\nabla \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\eta})\|_F \|\boldsymbol{\theta} - \boldsymbol{\xi}\|, \tag{3.15}$$

which is well defined, at least almost surely with respect to $P_{\mathbb{X}_1^{(j)}}$, since $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$ by Assumption B.2.11. By the continuity of the supremum and the Frobenius norm in combination with the measurability of the first derivatives of the estimating function we can conclude that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\nabla \boldsymbol{H}(\cdot, \boldsymbol{\theta})\|_F$ is a measurable function with respect to $\mathbb{X}_i^{(j)}$ for $j = 1, \ldots, q+1$. Hence, Condition (ii) is satisfied and Assumption B.2.11 shows Condition (iii). Consequently, since $\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}$ is the unique zero of $\frac{k_{j,n}-k}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right) + \frac{k+G-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta})\right)$ applying Lemma E.2.17 yields

$$\max_{k\in B^{(1)}_{j,n,G}} \left\| \frac{1}{\sqrt{G}} \left(\sum_{i=k+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}) + \sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}) \right) \right\| \tag{3.16}$$
$$= \max_{k\in B^{(1)}_{j,n,G}} \sqrt{G} \left\| \frac{1}{G} \left(\sum_{i=k+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}) + \sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}) \right) \right\|$$
$$\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{k\in B^{(1)}_{j,n,G}} \sqrt{G} \left\| \frac{1}{G} \left(\sum_{i=k+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right) \right\|$$
$$+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{k\in B^{(1)}_{j,n,G}} \sqrt{G} \left\| \frac{1}{G} \left(\sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \boldsymbol{\theta}) \right) \right\|$$
$$= o_P\left(\sqrt{G}\right).$$

Hence, we receive

$$\left(o_P(1) + \frac{k_{j,n} - k}{G} \boldsymbol{V}_j(\boldsymbol{\xi}_{k,n,G}) + \frac{k + G - k_{j,n}}{G} \boldsymbol{V}_{j+1}(\boldsymbol{\xi}_{k,n,G}) \right) \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} \right)$$

$= o_P(1)$ uniformly on $B^{(1)}_{j,n,G}$.

Furthermore, by combining Assumption B.2.10 and Lemma E.2.21 we can multiply both sides of the equation above with the inverse of the convex combination $\frac{k_{j,n}-k}{G}\boldsymbol{V}_j(\boldsymbol{\xi}_{k,n,G}) + \frac{k+G-k_{j,n}}{G}\boldsymbol{V}_{j+1}(\boldsymbol{\xi}_{k,n,G})$ and get

$$o_P(1) = (o_P(1) + \boldsymbol{I}_p)\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right) \quad \text{uniformly on } B^{(1)}_{j,n,G}.$$

Thus, Lemma E.2.22 shows

$$\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right\| = o_P(1) \quad \text{uniformly on } B^{(1)}_{j,n,G}$$

and we obtain

$$\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right\| = o_P(1) \quad \text{uniformly on } B^{(1)}_{n,G},$$

since the number of changes q is finite.

(c) These assertions can be shown analogously to (b).

□

Note that Assumption B.2.10 is very restrictive and might not be verifiable in some models. For further discussion on that we refer to Remark 3.1.19 below.
Moreover, the matrix $\boldsymbol{\Gamma}_k$ has to be specified under the alternative as it constitutes an important part of our Wald-type statistic. According to Assumption A.2.2, for $k_{j-1,n} < k \leq k_{j,n}$, $j = 1, \ldots, q+1$, we set

$$\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_j = \boldsymbol{V}_j(\boldsymbol{\theta}_j)^{-1}\boldsymbol{\Sigma}_j(\boldsymbol{V}_j(\boldsymbol{\theta}_j)^{-1})^T. \tag{3.17}$$

3.1.2.1. Asymptotic Power of the MOSUM-Based Test

By applying Theorem 3.1.8 an **asymptotic level $\boldsymbol{\alpha}$ test** for testing the null hypothesis of $q = 0$, i.e. no change or structural break occurs in the considered time period, has been constructed:

$$\text{Reject } H_0 \text{ if } W_n(G) > D_n(G,\alpha),$$
$$\text{with } D_n(G,\alpha) = \frac{b(n/G) + c_\alpha}{a(n/G)},$$

where $c_\alpha := -\log\log\frac{1}{\sqrt{1-\alpha}}$ denotes the $(1-\alpha)$-quantile of the Gumbel distribution.
The following theorem shows that this test correctly rejects the null hypothesis under the alternative with probability tending to 1. The main result is stated in part (a) followed by part (b) which is of particular interest for applications where the matrix $\boldsymbol{\Gamma}_k$ is usually unknown and estimators are used.

Theorem 3.1.12. *Let $\{\mathbb{X}^{(1)}_i : i \geq 1\}, \ldots, \{\mathbb{X}^{(q+1)}_i : i \geq 1\}$ be sequences of type (E1) or (E2) satisfying the Assumptions B.2.3, B.2.7, B.2.8 and B.2.9. with $\widetilde{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_j$ for $\{\mathbb{X}^{(j)}_i : i \geq 1\}$. Furthermore, let Assumption A.1.1 on the bandwidth and Assumption A.2.1 hold.*

(a) Then, under H_1, we obtain for any $z \in \mathbb{R}$

$$\lim_{n\to\infty} P(a(n/G)W_n(G) - b(n/G) \geq z) = 1,$$

i.e. the test has asymptotic power one.

(b) The matrix $\mathbf{\Gamma}_k$ as in (3.17) can be replaced by a positive definite estimator sequence $\widehat{\mathbf{\Gamma}}_{k,n}$ satisfying the following assumption:

(I) $\max_{k\in B_{n,G}} \left\| \widehat{\mathbf{\Gamma}}_{k,n}^{-1/2} - \mathbf{\Gamma}_{A,k}^{-1/2} \right\|_F = o_P(1)$, where $\{\mathbf{\Gamma}_{A,k}\}_{k\geq 1}$ is a sequence of positive definite matrices fulfilling $\sup_k \|\mathbf{\Gamma}_{A,k}\|_F < \infty$ and $\sup_k \left\|\mathbf{\Gamma}_{A,k}^{-1/2}\right\|_F < \infty$.

Proof. (a) Similar to the proof of Theorem 2.1.5, it is sufficient to show that $W_n(G) - \frac{z+b(n/G)}{a(n/G)} \xrightarrow{P} \infty$ since the inequality $a(n/G)W_n(G) - b(n/G) \geq z$ is equivalent to

$$W_n(G) - \frac{z + b(n/G)}{a(n/G)} \geq 0.$$

First we use $W_n(G) \geq W_{k_{j,n},n}(G)$ before we split the statistic $W_{k_{j,n},n}(G)$ into noise and signal. Then, since the Assumptions B.2.7 and B.2.8 imply B.2.1, B.2.2 and B.2.4 Lemma 3.1.10 (a) can be applied which shows together with Lemma E.1.5

$$\begin{aligned}
W_n(G) &= \max_{G\leq k\leq n-G} W_{k,n}(G) \\
&\geq W_{k_{j,n},n}(G) = \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}+1,k_{j,n}+G} - \widehat{\boldsymbol{\theta}}_{k_{j,n}-G+1,k_{j,n}} \right) \right\| \\
&\geq \frac{\sqrt{G}}{\sqrt{2}} \left(\left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j) \right\| \right. \\
&\quad \left. - \left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}+1,k_{j,n}+G} - \boldsymbol{\theta}_{j+1} - \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-G+1,k_{j,n}} - \boldsymbol{\theta}_j \right) \right) \right\| \right) \\
&\geq \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j) \right\| \\
&\quad - \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} \right\|_F \left(\left\| \widehat{\boldsymbol{\theta}}_{k_{j,n}+1,k_{j,n}+G} - \boldsymbol{\theta}_{j+1} \right\| + \left\| \widehat{\boldsymbol{\theta}}_{k_{j,n}-G+1,k_{j,n}} - \boldsymbol{\theta}_j \right\| \right) \\
&= \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j) \right\| + O_P(1),
\end{aligned}$$

as $\left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} \right\|_F = O(1)$. Furthermore, Assumption B.2.9 in combination with Lemma E.1.7 yields that $\mathbf{\Gamma}_{k_{j,n}} = \boldsymbol{V}_j(\boldsymbol{\theta}_j)^{-1} \mathbf{\Sigma}_j (\boldsymbol{V}_j(\boldsymbol{\theta}_j)^{-1})^T$ is positive definite which implies that $\mathbf{\Gamma}_{k_{j,n}}^{-1}$ is positive definite as well. Hence, on noting that $\boldsymbol{\theta}_{j+1} \neq \boldsymbol{\theta}_j$ and that the Euclidean norm coincides with the Frobenius norm for vectors in $\mathbb{R}^d$ as shown in Lemma E.1.4, we receive

$$\left\| \mathbf{\Gamma}_{k_{j,n}}^{-1/2} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j) \right\| = \sqrt{(\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j)^T \mathbf{\Gamma}_{k_{j,n}}^{-1} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j)}$$

$$= \sqrt{(\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j)^T \left(\boldsymbol{V}_j(\boldsymbol{\theta}_j)^{-1}\boldsymbol{\Sigma}_j(\boldsymbol{V}_j(\boldsymbol{\theta}_j)^{-1})^T\right)^{-1} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j)} \geq c$$

holds for some $c > 0$. Since $\frac{z+b(n/G)}{a(n/G)} = o(\sqrt{G})$ by Assumption A.1.1 we can conclude that

$$W_n(G) - \frac{z + b(n/G)}{a(n/G)} \geq \frac{\sqrt{G}c}{\sqrt{2}} + O_P(1) - \frac{z + b(n/G)}{a(n/G)} = \sqrt{G}\left(\frac{c}{\sqrt{2}} + o_P(1)\right) \xrightarrow{P} \infty,$$

which implies the assertion in (a).

(b) If the matrix $\boldsymbol{\Gamma}_k$ is replaced by an estimator $\widehat{\boldsymbol{\Gamma}}_{k,n}$ in the statistic we use the following notation

$$\widehat{W}_n(G) = \max_{G \leq k \leq n-G} \widehat{W}_{k,n}(G) = \max_{G \leq k \leq n-G} \frac{\sqrt{G}}{\sqrt{2}} \left\| \widehat{\boldsymbol{\Gamma}}_{k,n}^{-1/2} \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right\|.$$

On noting that by Assumption (I) on the estimator sequence

$$\left\| \widehat{\boldsymbol{\Gamma}}_{k_{j,n},n}^{-1/2} \right\|_F \leq \max_{k \in B_{n,G}} \left\| \widehat{\boldsymbol{\Gamma}}_{k,n}^{-1/2} - \boldsymbol{\Gamma}_{A,k}^{-1/2} \right\|_F + \max_{k \in B_{n,G}} \left\| \boldsymbol{\Gamma}_{A,k}^{-1/2} \right\|_F = O_P(1),$$

the same arguments as in (a) can be applied here to obtain

$$\widehat{W}_n(G) \geq \frac{\sqrt{G}}{\sqrt{2}} \left\| \widehat{\boldsymbol{\Gamma}}_{k_{j,n},n}^{-1/2} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j) \right\| + O_P(1).$$

In analogous manner to the proof of Theorem 2.1.5 (c) we can show that

$$\left\| \widehat{\boldsymbol{\Gamma}}_{k_{j,n},n}^{-1/2} (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j) \right\| \geq c$$

hold for some $c > 0$. For a detailed explanation we refer to that proof. Finally, similar to (a) we can conclude that

$$\widehat{W}_n(G) - \frac{z + b(n/G)}{a(n/G)} \geq \sqrt{G}(c + o_P(1)) \to \infty,$$

which yields the statement of (b).

□

Remark 3.1.13. *The assumption on the estimator sequence of the long-run covariance matrix in part (b) is fulfilled if* $\widehat{\boldsymbol{\Gamma}}_{k,n} = \widehat{\boldsymbol{V}}_{k,n}^{-1}\widehat{\boldsymbol{\Sigma}}_{k,n}(\widehat{\boldsymbol{V}}_{k,n}^{-1})^T$, *where*

- $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *is a positive definite estimator sequence satisfying:* $\max_{k \in B_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F = o_P(1)$, *with* $\{\boldsymbol{\Sigma}_{A,k}\}_{k \geq 1}$ *denoting a sequence of positive definite matrices fulfilling* $\sup_k \|\boldsymbol{\Sigma}_{A,k}\|_F < \infty$ *and* $\sup_k \left\| \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F < \infty$ *and*

- $\widehat{\boldsymbol{V}}_{k,n}$ *is a regular estimator sequence satisfying:* $\max_{k \in B_{n,G}} \left\| \widehat{\boldsymbol{V}}_{k,n} - \boldsymbol{V}_{A,k} \right\|_F = o_P(1)$, *with* $\{\boldsymbol{V}_{A,k}\}_{k \geq 1}$ *denoting a sequence of regular matrices fulfilling* $\sup_k \left\| \boldsymbol{V}_{A,k}^{-1} \right\|_F < \infty$ *and* $\sup_k \|\boldsymbol{V}_{A,k}\|_F < \infty$.

Proof. We receive

$$\max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{\Gamma}}_{k,n}^{-1/2} - \boldsymbol{\Gamma}_{A,k}^{-1/2} \right\|_F = \max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{V}}_{k,n} \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{V}_{A,k}\boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F$$
$$= \max_{k\in B_{n,G}} \Big\| \left(\widehat{\boldsymbol{V}}_{k,n} - \boldsymbol{V}_{A,k} \right) \boldsymbol{\Sigma}_{A,k}^{-1/2}$$
$$+ \left(\widehat{\boldsymbol{V}}_{k,n} - \boldsymbol{V}_{A,k} \right) \left(\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right) + \boldsymbol{V}_{A,k} \left(\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right) \Big\|_F$$
$$\le \max_{k\in B_{n,G}} \left\| \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F \max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{V}}_{k,n} - \boldsymbol{V}_{A,k} \right\|_F$$
$$+ \max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{V}}_{k,n} - \boldsymbol{V}_{A,k} \right\|_F \max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F$$
$$+ \max_{k\in B_{n,G}} \left\| \boldsymbol{V}_{A,k} \right\|_F \max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F$$
$$= o_P(1),$$

which shows the first part of Assumption (I). Furthermore, note that $\boldsymbol{\Gamma}_{A,k} = \boldsymbol{V}_{A,k}^{-1}\boldsymbol{\Sigma}_{A,k}\left(\boldsymbol{V}_{A,k}^{-1}\right)^T$ is a sequence of positive definite matrices by Lemma E.1.7 with

$$\sup_k \left\| \boldsymbol{\Gamma}_{A,k} \right\|_F = \sup_k \left\| \boldsymbol{V}_{A,k}^{-1}\boldsymbol{\Sigma}_{A,k}\left(\boldsymbol{V}_{A,k}^{-1}\right)^T \right\|_F \le \sup_k \left\| \boldsymbol{V}_{A,k}^{-1} \right\|_F^2 \sup_k \left\| \boldsymbol{\Sigma}_{A,k} \right\|_F < \infty$$

and

$$\sup_k \left\| \boldsymbol{\Gamma}_{A,k}^{-1/2} \right\|_F = \sup_k \left\| \boldsymbol{V}_{A,k}\boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F \le \sup_k \left\| \boldsymbol{V}_{A,k} \right\|_F \sup_k \left\| \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F < \infty$$

completing the proof of Assumption (I). □

3.1.2.2. MOSUM Wald-Type Estimators

The estimators for the number of changes and the locations are determined in an analogous manner to Section 2.1.3.1. We consider all pairs of time points $(v_{j,n}, w_{j,n})$ with

$$\begin{aligned} &W_{k,n}(G) \ge D_n(\alpha_n, G) \ \text{ for } \ v_{j,n} \le k \le w_{j,n}, \\ &W_{k,n}(G) < D_n(\alpha_n, G) \ \text{ for } \ k = v_{j,n}-1, w_{j,n}+1, \\ &w_{j,n} - v_{j,n} \ge \varepsilon G \quad \text{with } \ 0 < \varepsilon < 1/2 \quad \text{fixed.} \end{aligned} \tag{3.18}$$

The estimator for the number of changes $\widehat{q}_n$ is given by the number of pairs and we take the maximal points of these exceeding intervals $[v_{j,n}, w_{j,n}]$ as estimators for the location of the change points:

$$\widehat{k}_{j,n} := \underset{v_{j,n}\le k\le w_{j,n}}{\arg\max}\, W_{k,n}(G).$$

Before we prove that these estimators are consistent for the true values in some sense, we need to examine the distance between $\boldsymbol{\theta}_j$ and the unique zeros $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(1)}$ and $\widetilde{\boldsymbol{\theta}}_{k,n,G}^{(2)}$ on specific intervals around the change points. Therefore, we define

$$\begin{aligned} &\bar{B}_{j,n,G}^{(1)} := B_{j,n,G}^{(1)} \cap \{k : |k - k_{j,n}| < (1-\varepsilon)G\} \text{ and} \\ &\bar{B}_{j,n,G}^{(2)} := B_{j,n,G}^{(2)} \cap \{k : |k - k_{j,n}| < (1-\varepsilon)G\} \end{aligned} \tag{3.19}$$

with $B_{j,n,G}^{(1)}$ and $B_{j,n,G}^{(2)}$ as in (3.8) and (3.9) and ε as in (3.18).

Lemma 3.1.14. *Let Assumption A.1.1 on the bandwidth and Assumption A.2.1 be fulfilled. Moreover let* $\{\mathbb{X}_i^{(1)} : i \geq 1\}, \ldots, \{\mathbb{X}_i^{(q+1)} : i \geq 1\}$ *be sequences of type (E1) or (E2) satisfying Assumption B.2.11. Then, for all* $j = 1, \ldots, q+1$*,*

(a) $\max_{k \in \bar{B}^{(1)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| = O(1)$ *and* $\max_{k \in \bar{B}^{(2)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\theta}}^{(2)}_{k,n,G} - \boldsymbol{\theta}_{j+1} \right\| = O(1)$*,*

(b) $\min_{k \in \bar{B}^{(1)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| > c$ *and* $\min_{k \in \bar{B}^{(2)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\theta}}^{(2)}_{k,n,G} - \boldsymbol{\theta}_{j+1} \right\| > c$ *for some* $c > 0$.

Proof. (a) The assertions follow directly from the compactness of the parameter space.

(b) First note that by definition of $\boldsymbol{\theta}_j$ and $\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}$

$$\begin{aligned}
&\frac{k+G-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_j)\right) \\
&= \frac{k_{j,n}-k}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_j)\right) + \frac{k+G-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_j)\right) \\
&\quad - \frac{k_{j,n}-k}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G})\right) - \frac{k+G-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G})\right).
\end{aligned}$$

Furthermore, with the approximation in (3.15) and by using the properties of expected values we receive

$$\begin{aligned}
&\frac{k+G-k_{j,n}}{G} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_j)\right) - E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right) \right\| \\
&\leq \frac{k+G-k_{j,n}}{G} E\left(\left\| \boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_j) - \boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} \right\| \right) \\
&\leq \frac{k+G-k_{j,n}}{G} E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}) \right\|_F \right) \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| \\
&\leq E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}) \right\|_F \right) \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| < \infty, \text{ for all } k \in \bar{B}^{(1)}_{j,n,G},
\end{aligned}$$

where the last line follows from Assumption B.2.11. Similarly, we obtain

$$\begin{aligned}
&\frac{k_{j,n}-k}{G} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_j)\right) - E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right) \right\| \\
&\leq \frac{k_{j,n}-k}{G} E\left(\left\| \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_j) - \boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} \right\| \right) \\
&\leq \frac{k_{j,n}-k}{G} E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}) \right\|_F \right) \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| \\
&\leq (1-\varepsilon) E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}) \right\|_F \right) \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| < \infty, \text{ for all } k \in \bar{B}^{(1)}_{j,n,G},
\end{aligned}$$

Hence, this can be combined to

$$\begin{aligned}
&\left\| \frac{k+G-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_j)\right) \right\| \\
&\leq \left(E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}) \right\|_F \right) + (1-\varepsilon) E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}) \right\|_F \right) \right) \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\|.
\end{aligned}$$

Moreover, since $\boldsymbol{\theta}_{j+1}$ is the unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\cdot)\right)$ and $\boldsymbol{\theta}_{j+1} \neq \boldsymbol{\theta}_j$ we know that there exists a $\widetilde{c} > 0$ such that

$$\left\| \frac{k+G-k_{j,n}}{G} E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\boldsymbol{\theta}_j)\right)\right\| \geq \varepsilon \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\boldsymbol{\theta}_j)\right)\right\| > \widetilde{c}.$$

Finally, we can conclude

$$\min_{k\in\bar{B}^{(1)}_{j,n,G}} \left\|\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right\| > c \text{ with}$$

$$c := \frac{\widetilde{c}}{E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j+1)},\boldsymbol{\theta})\right\|_F\right) + (1-\varepsilon)E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right\|_F\right)} > 0.$$

The second statement can be derived in an analogous manner.

□

Now, we have all the ingredients together to show that the estimator for the number of changes is equal to the true number with probability tending to one.

Theorem 3.1.15. *Let Assumption A.1.1 on the bandwidth and Assumption A.2.1 be fulfilled. Moreover, let* $\{\mathbb{X}_i^{(1)} : i \geq 1\}, \ldots, \{\mathbb{X}_i^{(q+1)} : i \geq 1\}$ *be sequences of type (E1) or (E2) satisfying the Assumptions (B.2.7), (B.2.8), (B.2.10) and (B.2.11). Furthermore, assume that the sequence* $\{\alpha_n\}_{n\in\mathbb{N}}$ *fulfills Assumption A.2.8.*

(a) Then, it holds

$$P(\widehat{q}_n = q) \to 1, \quad \text{as } n \to \infty.$$

(b) The consistency statement in (a) remains true if the matrix $\boldsymbol{\Gamma}_k$ *is replaced by an estimator* $\widehat{\boldsymbol{\Gamma}}_{k,n}$ *satisfying the following assumptions.*

(I) $\max_{k\in B_{n,G}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n} - \boldsymbol{\Gamma}^{-1/2}_{A,k}\right\|_F = o_P(1)$, *where* $\{\boldsymbol{\Gamma}_{A,k}\}_{k\geq 1}$ *is a sequence of positive definite matrices fulfilling* $\sup_k \|\boldsymbol{\Gamma}_{A,k}\|_F < \infty$ *and* $\sup_k \left\|\boldsymbol{\Gamma}^{-1/2}_{A,k}\right\|_F < \infty$.

(II) $\max_{k\in A_{n,G}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n} - \boldsymbol{\Gamma}^{-1/2}_{k}\right\|_F = o_P\left(\log(n/G)^{-1}\right).$

Proof. (a) The basic idea of this proof is similar to that of Theorem 2.1.8. Due to Condition 3 in the MOSUM procedure, we need to consider only the set of time points being far away from any change, $A_{n,G}$, and the sets of time points lying in an $((1-\varepsilon)G)$- environment of a change as in (3.19) such that

$$\bar{B}_{n,G} := \bigcup_{j=1}^{q+1} \bar{B}_{j,n,G} \quad \text{with} \quad \bar{B}_{j,n,G} := \bar{B}^{(1)}_{j,n,G} \cup \bar{B}^{(2)}_{j,n,G}.$$

We obtain

$$P(\widehat{q}_n = q)$$

$$\geq P\left(\max_{k\in A_{n,G}} W_{k,n}(G) < D_n(\alpha_n, G), \min_{k\in \bar{B}_{n,G}} W_{k,n}(G) \geq D_n(\alpha_n, G)\right)$$
$$\geq P\left(\max_{k\in A_{n,G}} W_{k,n}(G) < D_n(\alpha_n, G)\right) + P\left(\min_{k\in \bar{B}_{n,G}} W_{k,n}(G) \geq D_n(\alpha_n, G)\right) - 1.$$

Thus, it is sufficient to show that

$$(1)\ P\left(\max_{k\in A_{n,G}} W_{k,n}(G) < D_n(\alpha_n, G)\right) \to 1 \ \text{ and}$$
$$(2)\ P\left(\min_{k\in \bar{B}_{n,G}} W_{k,n}(G) \geq D_n(\alpha_n, G)\right) \to 1$$

as n goes to infinity.
Part (1):
We start with proving statement (1). Let $T^{(j)}_{k,n}(G, \boldsymbol{\theta})$ be the MOSUM statistic based on estimating functions according to Definition 2.0.1 computed on the sequence $\{\mathbb{X}^{(j)}_i\}_{i\geq 1}$. Furthermore, note that applying Lemma 3.1.11 yields

$$\max_{k\in A_{j,n,G}} \left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \boldsymbol{\theta}_j\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right) \quad \text{and}$$
$$\max_{k\in A_{j,n,G}} \left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \boldsymbol{\theta}_j\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right).$$

Thus, since $(\mathbb{X}_{k-G+1}, \ldots, \mathbb{X}_{k+G}) = (\mathbb{X}^{(j)}_{k-G+1}, \ldots, \mathbb{X}^{(j)}_{k+G})$ holds for all $k \in A_{j,n,G}$ with $\{\mathbb{X}^{(j)}_i\}_{i\geq 1}$ fulfilling the Assumptions B.2.5, B.2.6, B.2.7 and B.2.9 (following from Assumption B.2.10 with $\delta = 1$) Lemma 3.1.7 can be used here to receive

$$\max_{k\in A_{j,n,G}} \left\|\boldsymbol{\Sigma}_j^{-1/2}\left(\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}^{(j)}_i, \boldsymbol{\theta}_j) - \frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{V}_j(\boldsymbol{\theta}_j)\left(\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{k+1,k+G}\right)\right)\right\|$$
$$= o_P\left((\log(n/G))^{-1/2}\right)$$

and

$$\max_{k\in A_{j,n,G}} \left\|\boldsymbol{\Sigma}_j^{-1/2}\left(\frac{1}{\sqrt{2G}}\sum_{i=k-G+1}^{k} \boldsymbol{H}(\mathbb{X}^{(j)}_i, \boldsymbol{\theta}_j) - \frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{V}_j(\boldsymbol{\theta}_j)\left(\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)\right)\right\|$$
$$= o_P\left((\log(n/G))^{-1/2}\right).$$

This implies

$$\left|\max_{k\in A_{j,n,G}} T^{(j)}_{k,n}(G, \boldsymbol{\theta}_j) - \max_{k\in A_{j,n,G}} W_{k,n}(G)\right|$$
$$= \left|\max_{k\in A_{j,n,G}} \frac{1}{\sqrt{2G}}\left\|\boldsymbol{\Sigma}_j^{-1/2}\left(\sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}^{(j)}_i, \boldsymbol{\theta}_j) - \sum_{i=k-G+1}^{k+G} \boldsymbol{H}(\mathbb{X}^{(j)}_i, \boldsymbol{\theta}_j)\right)\right\|\right.$$
$$\left. - \max_{k\in A_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}}\left\|\boldsymbol{\Sigma}_j^{-1/2}\boldsymbol{V}_j(\boldsymbol{\theta}_j)\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)\right\|\right|$$

$$\leq \max_{k\in A_{j,n,G}} \left\| \boldsymbol{\Sigma}_j^{-1/2} \left(\frac{1}{\sqrt{2G}} \left(\sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) - \sum_{i=k-G+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right) \right.\right.$$
$$\left.\left. - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}_j(\boldsymbol{\theta}_j) \left(\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \widehat{\boldsymbol{\theta}}_{k+1,k+G} \right) \right) \right\|$$
$$\leq \max_{k\in A_{j,n,G}} \left\| \boldsymbol{\Sigma}_j^{-1/2} \left(\frac{1}{\sqrt{2G}} \sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}_j(\boldsymbol{\theta}_j) \left(\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{k+1,k+G} \right) \right) \right\|$$
$$+ \max_{k\in A_{j,n,G}} \left\| \boldsymbol{\Sigma}_j^{-1/2} \left(\frac{1}{\sqrt{2G}} \sum_{i=k-G+1}^{k} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) - \frac{\sqrt{G}}{\sqrt{2}} \boldsymbol{V}_j(\boldsymbol{\theta}_j) \left(\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{k-G+1,k} \right) \right) \right\|$$
$$= o_P\left((\log(n/G))^{-1/2} \right),$$

which shows that $W_{k,n}(G) = T_{k,n}^{(j)}(G, \boldsymbol{\theta}_j) + o_P\left((\log(n/G))^{-1/2}\right)$ uniformly on $A_{j,n,G}$. Moreover, this yields

$$\begin{aligned} &P\left(a(n/G) \max_{k\in A_{j,n,G}} W_{k,n}(G) - b(n/G) < z \right) \\ &= P\left(a(n/G) \max_{k\in A_{j,n,G}} T_{k,n}^{(j)}(G, \boldsymbol{\theta}_j) - b(n/G) + o_P(1) < z \right) \\ &\geq P\left(a(n/G) \max_{G\leq k\leq n-G} T_{k,n}^{(j)}(G, \boldsymbol{\theta}_j) - b(n/G) + o_P(1) < z \right) \\ &= \exp(-2\exp(-z)) + o(1) \quad \text{for some } z \in \mathbb{R}, \end{aligned} \tag{3.20}$$

where the last line follows from Theorem 2.1.1 since Assumption A.1.3 is satisfied by $\{\boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\}_{i\geq 1}$ under the Conditions B.2.8 or B.2.5 as shown in Lemma 2.3.2. Hence, on noting that c_{α_n} is the $1-\alpha_n$-quantile of the Gumbel distribution, applying Lemma E.2.18 yields

$$\begin{aligned} &P\left(\max_{k\in A_{n,G}} W_{k,n}(G) < D_n(\alpha_n, G) \right) \\ &= P\left(\max_{1\leq j\leq q+1} \max_{k\in A_{j,n,G}} W_{k,n}(G) < D_n(\alpha_n, G) \right) \\ &= P\left(\bigcap_{j=1}^{q+1} \left\{ \max_{k\in A_{j,n,G}} W_{k,n}(G) < D_n(\alpha_n, G) \right\} \right) \\ &\geq \sum_{j=1}^{q+1} P\left(\max_{k\in A_{j,n,G}} W_{k,n}(G) < D_n(\alpha_n, G) \right) - (q+1) + 1 \\ &= \sum_{j=1}^{q+1} P\left(a(n/G) \max_{k\in A_{j,n,G}} W_{k,n}(G) - b(n/G) < c_{\alpha_n} \right) - q \\ &\geq (q+1)(1-\alpha_n) + o(1) - q = 1 - (q+1)\alpha_n + o(1). \end{aligned} \tag{3.21}$$

Thus, we can conclude that $P\left(\max_{k\in A_{n,G}} W_{k,n}(G) < D_n(\alpha_n, G)\right) \to 1$ by Assumption A.2.8.

Part (2):
For proving the second statement, it is necessary to approximate
$P\left(\min_{k\in\bar{B}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n, G)\right)$.
Note that $\mathbf{\Gamma}_k = \mathbf{\Gamma}_j$ and $\widetilde{\boldsymbol{\theta}}^{(2)}_{k,n,G} = \boldsymbol{\theta}_j$ hold for all $k \in \bar{B}^{(1)}_{j,n,G}$. After splitting the statistic into noise and signal, applying Lemma E.1.5 and Lemma 3.1.11 yields

$$\begin{aligned}
\min_{k\in\bar{B}^{(1)}_{j,n,G}} W_{k,n}(G) &= \min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_j^{-1/2}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)\right\| \\
&\geq \min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_j^{-1/2}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\| \\
&\quad - \max_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_j^{-1/2}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} - \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right)\right\| \\
&\geq \min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_j^{-1/2}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\| \\
&\quad - \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_j^{-1/2}\right\|_F \left(\max_{k\in B^{(1)}_{j,n,G}} \left\| \widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right\| + \max_{k\in B^{(1)}_{j,n,G}} \left\| \widehat{\boldsymbol{\theta}}_{k-G+1,k} - \boldsymbol{\theta}_j\right\|\right) \\
&= \min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_j^{-1/2}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\| + o_P\left(\sqrt{G}\right).
\end{aligned}$$

By using the result above and on noting that

$$\left\|\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right\| = \left\| \mathbf{\Gamma}_j^{1/2}\mathbf{\Gamma}_j^{-1/2}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\| \leq \left\|\mathbf{\Gamma}_j^{1/2}\right\|_F \left\| \mathbf{\Gamma}_j^{-1/2}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\|$$

according to Lemma E.1.5 and

$$\min_{k\in\bar{B}^{(1)}_{j,n,G}} \left\|\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right\| > c$$

by Lemma 3.1.14, we receive

$$\begin{aligned}
&P\left(\min_{k\in\bar{B}^{(1)}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n, G)\right) \\
&\geq P\left(\min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_j^{-1/2}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\| + o_P\left(\sqrt{G}\right) \geq D_n(\alpha_n, G)\right) \\
&\geq P\left(\min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\|\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right\| \left(\left\|\mathbf{\Gamma}_j^{1/2}\right\|_F\right)^{-1} + o_P\left(\sqrt{G}\right) \geq D_n(\alpha_n, G)\right) \\
&\geq P\left(\sqrt{G}\left(c + o_P(1)\right) \geq D_n(\alpha_n, G)\right) = P\left(c + o_P(1) \geq \frac{D_n(\alpha_n, G)}{\sqrt{G}}\right) \to 1,
\end{aligned}$$

where the last line follows from $\left(\left\|\mathbf{\Gamma}_j^{1/2}\right\|_F\right)^{-1} = O(1)$ due to the positive definiteness of the matrix $\mathbf{\Gamma}_j$ and from Assumptions A.1.1 and A.2.8 ensuring that

$\frac{D_n(\alpha_n,G)}{\sqrt{G}} = o(1)$. Similar arguments can be used to obtain

$$P\left(\min_{k\in\bar{B}^{(2)}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right) \to 1 \quad \text{as } n\to\infty.$$

Hence, with Lemma E.2.18 we get

$$\begin{aligned}
&P\left(\min_{k\in\bar{B}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right)\\
&= P\left(\left\{\min_{k\in\bar{B}^{(1)}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right\} \cap \left\{\min_{k\in\bar{B}^{(2)}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right\}\right)\\
&\geq P\left(\min_{k\in\bar{B}^{(1)}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right)\\
&\quad + P\left(\min_{k\in\bar{B}^{(2)}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right) - 1 \to 1.
\end{aligned}$$

Finally, we can conclude

$$\begin{aligned}
&P\left(\min_{k\in\bar{B}_{n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right)\\
&= P\left(\min_{1\leq j\leq q+1}\min_{k\in\bar{B}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right)\\
&= P\left(\bigcap_{j=1}^{q+1}\left\{\min_{k\in\bar{B}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right\}\right)\\
&\geq \sum_{j=1}^{q+1} P\left(\min_{k\in\bar{B}_{j,n,G}} W_{k,n}(G) \geq D_n(\alpha_n,G)\right) - (q+1) + 1 \to 1,
\end{aligned}$$

where the approximation in the last line follows from Lemma E.2.18.

(b) Part (1):
Combining Lemma 3.1.11 with Lemma E.1.5 and Assumption (II) on the estimator sequence yields

$$\begin{aligned}
&\left|\max_{k\in A_{n,G}} \widehat{W}_{k,n}(G) - \max_{k\in A_{n,G}} W_{k,n}(G)\right|\\
&= \Bigg|\max_{k\in A_{n,G}} \frac{\sqrt{G}}{\sqrt{2}}\left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)\right\|\\
&\qquad - \max_{k\in A_{n,G}} \frac{\sqrt{G}}{\sqrt{2}}\left\|\boldsymbol{\Gamma}^{-1/2}_{k}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)\right\|\Bigg|\\
&\leq \max_{k\in A_{n,G}} \frac{\sqrt{G}}{\sqrt{2}}\left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right\| \max_{k\in A_{n,G}}\left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n} - \boldsymbol{\Gamma}^{-1/2}_{k}\right\|_F
\end{aligned}$$

$$= O_P\left(\sqrt{\log(n/G)}\right) o_P\left(\log(n/G)^{-1}\right) = o_P\left(\log(n/G)^{-1/2}\right),$$

which shows that $\widehat{W}_{k,n}(G) = W_{k,n}(G) + o_P\left((\log(n/G))^{-1/2}\right)$ uniformly on $A_{n,G}$. Thus, with (3.20) we get

$$\begin{aligned}
&P\left(a(n/G) \max_{k\in A_{j,n,G}} \widehat{W}_{k,n}(G) - b(n/G) < z\right)\\
&= P\left(a(n/G) \max_{k\in A_{j,n,G}} W_{k,n}(G) - b(n/G) + o_P(1) < z\right)\\
&\geq \exp(-2\exp(-z)) + o(1) \quad \text{for some } z \in \mathbb{R}.
\end{aligned}$$

Hence, in an analogous manner to (3.21) we can conclude

$$P\left(\max_{k\in A_{n,G}} \widehat{W}_{k,n}(G) < D_n(\alpha_n, G)\right) \to 1.$$

Part (2):
Similar to (a), we get with Lemma E.1.5

$$\begin{aligned}
&\min_{k\in \bar{B}^{(1)}_{j,n,G}} \widehat{W}_{k,n}(G) = \min_{k\in \bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}\right)\right\|\\
&\geq \min_{k\in \bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\|\\
&\quad - \max_{k\in \bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widehat{\boldsymbol{\theta}}_{k-G+1,k} - \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right)\right\|\\
&\geq \min_{k\in \bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\|\\
&\quad - \frac{\sqrt{G}}{\sqrt{2}} \max_{k\in \bar{B}^{(1)}_{j,n,G}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\right\|_F \left(\max_{k\in \bar{B}^{(1)}_{j,n,G}} \left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right\| + \max_{k\in \bar{B}^{(1)}_{j,n,G}} \left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \boldsymbol{\theta}_j\right\|\right)\\
&\geq \min_{k\in \bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\|\\
&\quad - \frac{\sqrt{G}}{\sqrt{2}} \left(\max_{k\in \bar{B}^{(1)}_{j,n,G}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n} - \boldsymbol{\Gamma}^{-1/2}_{A,k}\right\|_F + \max_{k\in \bar{B}^{(1)}_{j,n,G}} \left\|\boldsymbol{\Gamma}^{-1/2}_{A,k}\right\|_F\right)\\
&\qquad \left(\max_{k\in B^{(1)}_{j,n,G}} \left\|\widehat{\boldsymbol{\theta}}_{k+1,k+G} - \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G}\right\| + \max_{k\in B^{(1)}_{j,n,G}} \left\|\widehat{\boldsymbol{\theta}}_{k-G+1,k} - \boldsymbol{\theta}_j\right\|\right)\\
&= \min_{k\in \bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\|\widehat{\boldsymbol{\Gamma}}^{-1/2}_{k,n}\left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j\right)\right\| + o_P\left(\sqrt{G}\right),
\end{aligned}$$

where the last line follows from Lemma 3.1.11 and Assumption (I). Furthermore, applying Lemma E.1.3 and Lemma E.1.5 in connection with Lemma 3.1.14 (a) and

Assumption (I) yields

$$\begin{aligned}
&\left| \min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \widehat{\mathbf{\Gamma}}_{k,n}^{-1/2} \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right) \right\| - \min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_{A,k}^{-1/2} \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right) \right\| \right| \\
&\leq \max_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \widehat{\mathbf{\Gamma}}_{k,n}^{-1/2} \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right) - \mathbf{\Gamma}_{A,k}^{-1/2} \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right) \right\| \\
&\leq \max_{k\in\bar{B}^{(1)}_{j,n,G}} \left\| \widehat{\mathbf{\Gamma}}_{k,n}^{-1/2} - \mathbf{\Gamma}_{A,k}^{-1/2} \right\|_F \max_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| \\
&= o_P \left((\log(n/G)^{-1} \right) O_P \left(\sqrt{G} \right) = o_P \left(\sqrt{G} \right).
\end{aligned}$$

Thus, by combining these results we obtain

$$\begin{aligned}
&P \left(\min_{k\in\bar{B}^{(1)}_{j,n,G}} \widehat{W}_{k,n}(G) \geq D_n(\alpha_n, G) \right) \\
&\geq P \left(\min_{k\in\bar{B}^{(1)}_{j,n,G}} \frac{\sqrt{G}}{\sqrt{2}} \left\| \mathbf{\Gamma}_{A,k}^{-1/2} \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right) \right\| + o_P \left(\sqrt{G} \right) \geq D_n(\alpha_n, G) \right).
\end{aligned}$$

Similar arguments as in (2.8) in the proof of Theorem 2.1.5 (c) can be applied here to show that $\min_{k\in\bar{B}^{(1)}_{j,n,G}} \left\| \mathbf{\Gamma}_{A,k}^{-1/2} \boldsymbol{x} \right\| \geq \sqrt{c_1} \|\boldsymbol{x}\|$ for every $\boldsymbol{x} \in \mathbb{R}^p$ with $\boldsymbol{x} \neq \mathbf{0}$, where $c_1 > 0$ denotes a lower bound for the minimal eigenvalues of $\mathbf{\Gamma}_{A,k}^{-1}$, for every $k \in \bar{B}^{(1)}_{j,n,G}$, obtained by Lemma E.1.10. For further explanation we refer to the proof of Theorem 2.1.5 (c). Moreover, from Lemma 3.1.14 we know that $\min_{k\in\bar{B}^{(1)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| > c_2$ for some $c_2 > 0$. Thus, we receive

$$\left\| \mathbf{\Gamma}_{A,k}^{-1/2} \left(\widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right) \right\| \geq \sqrt{c_1} \left\| \widetilde{\boldsymbol{\theta}}^{(1)}_{k,n,G} - \boldsymbol{\theta}_j \right\| > c \text{ with } c := \sqrt{c_1} c_2.$$

Finally, in analogous manner to (a) we can conclude that

$$P \left(\min_{k\in\bar{B}^{(1)}_{j,n,G}} \widehat{W}_{k,n}(G) \geq D_n(\alpha_n, G) \right) \geq P \left(c + o_P(1) \geq \frac{D_n(\alpha_n, G)}{\sqrt{G}} \right) \to 1,$$

since $\frac{D_n(\alpha_n,G)}{\sqrt{G}} = o(1)$ by Assumptions A.1.1 and A.2.8.

□

Moreover, the following corollary in combination with Remark 3.1.17 proves a weak consistency statement for the change point estimators $\widehat{k}_{j,n}$.

Corollary 3.1.16. *Let the assumptions of Theorem 3.1.15 hold. Then,*

$$P \left(\max_{1\leq j\leq q} \min_{1\leq l\leq \hat{q}_n} \left| \widehat{k}_{l,n} - k_{j,n} \right| < G \right) \to 1,$$

i.e. with probability tending to one every change point has at least one estimator in its G-environment.

Proof. Similar to the proof of Corollary 2.1.10, the assertion follows from part (2) in the proof of Theorem 3.1.15. □

Remark 3.1.17. *By Theorem 3.1.15 there are exactly q change point estimators with asymptotic probability one. Since the distance between two adjacent change points is asymptotically greater than $2G$ an estimator can only lie in the G-environment of one change point. Thus, combining Theorem 3.1.15 and Corollary 3.1.16 yields that every change point has exactly one estimator in its G-environment with probability tending to one.*

With the help of the results above we get that the estimators of the rescaled change points $\widehat{\lambda}_{j,n} := \frac{\widehat{k}_{j,n}}{n}$, $j \in \{1, \ldots, q\}$, are consistent for the true rescaled changes λ_j, $j \in \{1, \ldots, q\}$, in the classical sense shown by the corollary below.

Corollary 3.1.18. *Let the assumptions of Theorem 3.1.15 hold. Then,*

$$\max_{1\le j\le q} \min_{1\le l\le \hat{q}_n} \left|\widehat{\lambda}_{l,n} - \lambda_j\right| = O_P\left(\frac{G}{n}\right) = o_P(1).$$

Proof. The assertion follows immediately from Corollary 3.1.16. □

Remark 3.1.19. *In order to prove the consistency results above we have to assume that Assumption B.2.10 holds which is very restrictive and can not be verified in some models. Thus, we have been thinking about relaxing this assumtpion to Assumption B.2.9. From theory we know that under Assumption B.2.9, for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, there are at most finitely many δ such that $\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta})$ is not regular. Unfortunately, this statement is not strong enough and it does not rule out that there exists a $k \in \bar{B}^{(1)}_{j,n,G}$ so that*

$$\frac{k_{j,n} - k}{G} \boldsymbol{V}_j(\boldsymbol{\xi}_{k,n,G}) + \frac{k + G - k_{j,n}}{G} \boldsymbol{V}_{j+1}(\boldsymbol{\xi}_{k,n,G})$$

is not invertible. This would suspend a main argument in the proof of Lemma 3.1.11 which would implicate that we are not able to assess the value of the statistic in this point. Hence, it would be theoretically possible that the statistic takes extremely small or large values on $\bar{B}^{(1)}_{j,n,G}$. Whereas large outliers are not problematic, as long as we are not interested in improving the convergence rates, small outliers, which fall below the critical value, can have severe consequences. In this scenario we could have two intervals of exceedings of length greater than εG in the G-environment of a change point such that the number of changes would be overestimated. A solution could be to modify the ε-criterion in the MOSUM procedure. One can think of choosing the estimators of the locations and the number of the changes in the following way:

Let $U_{n,G} := \{k \in \{G, \ldots, n-G\} : W_{k,n}(G) \ge D_n(\alpha, G)\}$. Then, take the local maximum $\tilde{k}_{j,n}$ as a change point estimator if it satisfies

$$\left|[\tilde{k}_{j,n} - G, \tilde{k}_{j,n} + G] \cap U_{n,G}\right| > (1 - \tilde{\varepsilon})G, \quad \text{with } 0 < \tilde{\varepsilon} < 1/2,$$

i.e. the statistic exceeds the critical value in more than $(1 - \tilde{\varepsilon})G$ time points of the G-environment of the local maximum $\tilde{k}_{j,n}$. Furthermore, use the number of these local maxima as an estimator for the number of changes.

Nevertheless, investigating the test and the estimators obtained by this modified version of the MOSUM procedure would go beyond the scope of this work but should be examined in the future.

3.2. The Linear Regression Model

The linear regression model has been investigated and applied in many different fields of statistics. Due its structure the model is easier to analyse in comparison to non-linear models. Even if one is interested in a more general setting it can be helpful to consider the linear regression model first in order to gain insight how to generalize a statistical method. This is exactly what we did in this work. We actually started with investigating MOSUM Wald-type statistics in the linear regression model before we were able to develop the general theory of Section 3.1. Nevertheless, linear regression has to be considered separately since the parameter space is not compact, which is an important assumption in the section before.
It is not surprising that we are not the first ones who want to detect changes in this model. At-most-one-change situations have been discussed extensively under various assumptions in the change point literature, we refer to Csörgö & Horváth (1997) Chapter 3 and references therein or Zeileis *et al.* (2002). For instance, Hawkins (1989) and Horváth & Shao (1995) considered tests based on Wald-type statistics which for each time point k compare the least squares estimator computed on the subsample of the first k observations with the estimator calculated on the last $n-k$ observations. Some papers also focus on the change point detection in linear autoregressive time series. Whereas Horváth (1993) considered test statistics of unweighted partial sums of residuals, Hušková *et al.* (2007) used statistics based on partial sums of weighted residuals belonging to the group of score-type statistics.
In comparison, alternatives of multiple changes have received less attention. Just to name a few contributions, Liu *et al.* (1997) applied a modified Schwarz criterion for identifying different segments in a multivariate regression model. Bai & Perron (1998) and Bai & Perron (2003) estimated the changes by minimizing the sum of squared residuals and determined the number of changes by conducting a consistent test. More recently, Perron & Qu (2006) extended their work to models with linear restrictions on the regression coefficients.
We consider a linear regression model under minimal restrictive assumptions such that simple linear regression models as well as autoregressive structures or models with exogenous and endogenous regressors can be analysed.

3.2.1. Asymptotics Under the Null Hypothesis

Let $\boldsymbol{\beta}_0$ be the true parameter of the model if no change occurs, i.e.

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta}_0 + \varepsilon_i$$

holds for all $i = 1, \ldots, n$ under the null. In a random design model we assume that the vector of the regressors $\boldsymbol{X}_i$ is random whereas a fixed design requires $\boldsymbol{X}_i$ being deterministic. Here we focus on the first one, the random design model.
Furthermore, the following assumptions are used in this section:

(R1) The sequence $\{\boldsymbol{X}_i\}_{i \geq 1}$ is stationary and ergodic with $E\left(\|\boldsymbol{X}_1\|\right) < \infty$.

(R2) Let $\mathcal{F}_t = \sigma\left(\boldsymbol{X}_j, \varepsilon_{j-1}, j \leq t\right)$. We assume that ε_t and $\mathcal{F}_t$ are independent.

(R3) $\boldsymbol{C} := E\left(\boldsymbol{X}_1 \boldsymbol{X}_1^T\right)$ is a positive definite matrix.

(R4) The sequence $\{\varepsilon_i\}_{i\geq 1}$ is i.i.d. with $E(\varepsilon_1)=0$, $0<E(\varepsilon_1^2):=\sigma^2<\infty$.

(R5) Let the components of $\{\boldsymbol{X}_i\boldsymbol{X}_i^T-\boldsymbol{C}\}_{i\geq 1}$ satisfy a strong invariance principle similar to that in Assumption A.1.3.

(R6) Let $\{\boldsymbol{X}_i\varepsilon_i\}_{i\geq 1}$ be a series with positive definite long-run covariance matrix $\boldsymbol{\Sigma}$ satisfying a strong invariance principle similar to that in Assumption A.1.3.

The (local) least-squares estimator $\widehat{\boldsymbol{\beta}}_{l,u}$ based on the subsample $\left(Y_l,\boldsymbol{X}_l^T\right)^T,\ldots,\left(Y_u,\boldsymbol{X}_u^T\right)^T$ solves the normal equation

$$\sum_{i=l}^{u}\boldsymbol{X}_i\boldsymbol{X}_i^T\boldsymbol{\beta}=\sum_{i=l}^{u}\boldsymbol{X}_iY_i \text{ and is given by } \widehat{\boldsymbol{\beta}}_{l,u}=\left(\sum_{i=l}^{u}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\sum_{i=l}^{u}\boldsymbol{X}_iY_i. \tag{3.22}$$

This estimator minimises the sum of the squared residuals $\sum_{i=l}^{u}\left(Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}\right)^2$ and is therefore the solution of the following estimating equation system

$$\sum_{i=l}^{u}\boldsymbol{H}(Y_i,\boldsymbol{X}_i,\boldsymbol{\beta})=\boldsymbol{0} \quad \text{with} \quad \boldsymbol{H}(Y_i,\boldsymbol{X}_i,\boldsymbol{\beta}):=-\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}\right), \tag{3.23}$$

where the vector $\mathbb{X}_i$ from Section 3.1 is given by $\left(Y_i,\boldsymbol{X}_i^T\right)^T$. On noting that the parameter space $\boldsymbol{\Theta}=\mathbb{R}^p$ is not compact, Lemma 3.1.1 can not be applied here. However, by using the formula for the least-squares estimator in (3.22) the consistency of the estimator sequences $\widehat{\boldsymbol{\beta}}_{k+1,k+G}$ and $\widehat{\boldsymbol{\beta}}_{k-G+1,k}$ can be derived directly.

Lemma 3.2.1. *Let the Assumptions (R1) to (R4) and (R6) hold for the random design model and let Assumption A.1.1 hold on the bandwidth. Then, under H_0*

$$\left\|\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right\|=O_P\left(\frac{1}{\sqrt{G}}\right) \quad \textit{and} \quad \left\|\widehat{\boldsymbol{\beta}}_{k-G+1,k}-\boldsymbol{\beta}_0\right\|=O_P\left(\frac{1}{\sqrt{G}}\right)$$

pointwise for all $k=G,\ldots,n-G$.

Proof. On noting that by Conditions (R1) and (R3) $\{\boldsymbol{X}_i\boldsymbol{X}_i^T\}_{i\geq 1}$ is a stationary and ergodic sequence with existing first moment, applying the Ergodic Theorem yields

$$\frac{1}{G}\sum_{i=1}^{G}\boldsymbol{X}_i\boldsymbol{X}_i^T \quad \stackrel{a.s.}{\to} \quad \boldsymbol{C}.$$

Since the matrix $\boldsymbol{C}$ is invertible by Assumption (R3) we get that $\frac{1}{G}\sum_{i=1}^{G}\boldsymbol{X}_i\boldsymbol{X}_i^T$ is invertible as well for large G. Thus, with the continuity of the matrix inverse function the Continuous Mapping Theorem can be used to get

$$\left(\frac{1}{G}\sum_{i=1}^{G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1} \quad \stackrel{a.s.}{\to} \quad \boldsymbol{C}^{-1}. \tag{3.24}$$

Furthermore, by the properties of conditional expectations, the measurability of $\boldsymbol{X}_i$ with respect to the filtration $\mathcal{F}_i$ defined in Assumption (R2) and as ε_i is independent of $\mathcal{F}_i$ we get

$$E\left(\boldsymbol{X}_i\varepsilon_i\right) = E\left(E\left(\boldsymbol{X}_i\varepsilon_i|\mathcal{F}_i\right)\right) = E\left(\boldsymbol{X}_i E\left(\varepsilon_i|\mathcal{F}_i\right)\right) = E\left(\boldsymbol{X}_i E\left(\varepsilon_i\right)\right) \\ = E\left(\boldsymbol{X}_i\right) E\left(\varepsilon_i\right) = \mathbf{0}. \tag{3.25}$$

Moreover, thanks to the invariance principle of Assumption (R6) there exists a Wiener process $\widetilde{\boldsymbol{W}}(t)$ with covariance matrix being equal to the long-run covariance matrix of $\boldsymbol{X}_i\varepsilon_i$ such that

$$\left\|\frac{1}{\sqrt{G}}\sum_{i=1}^{G}\boldsymbol{X}_i\varepsilon_i\right\| \leq \left\|\frac{1}{\sqrt{G}}\left(\sum_{i=1}^{G}\boldsymbol{X}_i\varepsilon_i - \widetilde{\boldsymbol{W}}(G)\right)\right\| + \left\|\frac{1}{\sqrt{G}}\widetilde{\boldsymbol{W}}(G)\right\| \tag{3.26}$$
$$= O_P\left(\frac{G^{1/(2+\nu)}}{\sqrt{G}}\right) + \left\|\frac{1}{\sqrt{G}}\widetilde{\boldsymbol{W}}(G)\right\| \overset{D}{=} O_P\left(G^{-\nu/(4+2\nu)}\right) + \left\|\widetilde{\boldsymbol{W}}(1)\right\| = O_P(1),$$

where the last line follows from the self-similarity of the Wiener process. Furthermore, note that

$$\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \boldsymbol{\beta}_0 \tag{3.27}$$
$$= \left(\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i Y_i - \left(\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\boldsymbol{\beta}_0$$
$$= \left(\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\left(Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}_0\right) = \left(\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\varepsilon_i.$$

Finally, with (3.24), (3.26) and the stationarity of the sequences we can conclude that

$$\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \boldsymbol{\beta}_0$$
$$= \left(\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\varepsilon_i \overset{D}{=} \left(\frac{1}{G}\sum_{i=1}^{G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\frac{1}{G}\sum_{i=1}^{G}\boldsymbol{X}_i\varepsilon_i$$
$$= \left(\boldsymbol{C}^{-1} + o_P(1)\right) O_P\left(\frac{1}{\sqrt{G}}\right) = O_P\left(\frac{1}{\sqrt{G}}\right).$$

□

The lemma above shows that the local estimator sequences $\widehat{\boldsymbol{\beta}}_{k+1,k+G}$ are $\sqrt{G}$-consistent for the true parameter vector $\boldsymbol{\beta}_0$ holding pointwise for each k. However, investigating the MOSUM Wald-type statistic requires a result holding uniformly in k which is given in the follwing lemma.

Lemma 3.2.2. *Let the Assumptions (R1) to (R6) hold for the random design model and let Assumption A.1.1 hold on the bandwidth. Then, under* H_0

$$\left\|\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \boldsymbol{\beta}_0\right\| = O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right) \quad \textit{uniformly in } k = 0, \ldots, n-G.$$

Proof. Firstly, note that by (3.27) we get

$$\left(\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right)=\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\varepsilon_i. \tag{3.28}$$

Furthermore, by Condition (R5) Theorem E.2.12 can be applied to each component of the matrix-valued sequence $\{\boldsymbol{X}_i\boldsymbol{X}_i^T-\boldsymbol{C}\}_{i\geq 1}$. Thus, with $\boldsymbol{X}_i\boldsymbol{X}_i^T-\boldsymbol{C}=(\widetilde{X}_{rs}(i))$ we receive

$$\frac{1}{G}\left|\sum_{i=k+1}^{k+G}\widetilde{X}_{rs}(i)\right|=O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right)=o_P(1)\quad\text{uniformly in } k,$$

where the last line follows from Assumption A.1.1. Hence, in connection with Lemma E.1.6 (b) we get

$$\frac{1}{G}\left\|\sum_{i=k+1}^{k+G}\left(\boldsymbol{X}_i\boldsymbol{X}_i^T-\boldsymbol{C}\right)\right\|_F=O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right)=o_P(1)\quad\text{uniformly in } k. \tag{3.29}$$

Moreover, the uniform statement

$$\max_{0\leq k\leq n-G}\frac{1}{G}\left\|\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\varepsilon_i\right\|=O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right)$$

follows from Assumption (R6) and Theorem E.2.12. Hence, by considering equation (3.28) again we can conclude

$$\left(\boldsymbol{C}+o_P(1)\right)\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right)=O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right)\quad\text{uniformly in } k.$$

Finally, the assertion follows from Lemma E.2.22 in connection with Assumption (R3). □

Now, we want to show that the Condition (3.2) is satisfied in order to derive a limit distribution for the Wald-type statistic with the help of Theorem 3.1.8. Under the null hypothesis, we obtain

$$\boldsymbol{H}(Y_i,\boldsymbol{X}_i,\boldsymbol{\beta}_0)=-\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}_0\right)=-\boldsymbol{X}_i\varepsilon_i$$

with expectation $E\left(\boldsymbol{H}(Y_i,\boldsymbol{X}_i,\boldsymbol{\beta}_0)\right)=E\left(-\boldsymbol{X}_i\varepsilon_i\right)=\boldsymbol{0}$ as shown in (3.25) and long-run covariance matrix $\boldsymbol{\Sigma}$. Furthermore, we receive

$$\nabla\boldsymbol{H}\left(Y_i,\boldsymbol{X}_i,\boldsymbol{\beta}\right)=\frac{\partial\boldsymbol{H}\left(Y_i,\boldsymbol{X}_i,\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}}=\boldsymbol{X}_i\boldsymbol{X}_i^T,$$

implying that $\boldsymbol{V}(\boldsymbol{\beta}_0)=\boldsymbol{C}$. Thus, we get that the asymptotic covariance matrix of $\sqrt{G}\widehat{\boldsymbol{\beta}}_{k-G+1,k}$ is $\boldsymbol{\Gamma}_k=\boldsymbol{C}^{-1}\boldsymbol{\Sigma}\boldsymbol{C}^{-1}$. Consequently, under the null hypothesis the MOSUM Wald-type statistic is given by:

$$\begin{aligned}
&W_n^{linear}(G)\\
&=\max_{G\leq k\leq n-G}\sqrt{G}\sqrt{\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\widehat{\boldsymbol{\beta}}_{k-G+1,k}\right)^T\boldsymbol{C}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\widehat{\boldsymbol{\beta}}_{k-G+1,k}\right)}\\
&=\max_{G\leq k\leq n-G}\sqrt{G}\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{C}\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\widehat{\boldsymbol{\beta}}_{k-G+1,k}\right)\right\|.
\end{aligned}$$

Lemma 3.2.3. *Let Assumption A.1.1 hold on the bandwidth. Furthermore, assume that Conditions (R1) to (R6) are fulfilled. Then, under H_0, the Assumption* (3.2) *is satisfied for the linear regression model, i.e.*

$$\max_{0\le k\le n-G}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}_0\right)-\frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{C}\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right)\right)\right\| \quad (3.30)$$
$$= o_P\left((\log(n/G))^{-1/2}\right).$$

Proof. By (3.22) we know that

$$\sum_{i=k+1}^{k+G}\boldsymbol{X}_iY_i=\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{k+1,k+G},$$

which is equivalent to

$$\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}_0\right)=\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right).$$

By multiplying $\frac{1}{\sqrt{2G}}$ and subtracting $\frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{C}\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right)$ from both sides of the equation above, we obtain

$$\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}_0\right)-\frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{C}\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right)$$
$$=\frac{1}{G}\sum_{i=k+1}^{k+G}\left(\boldsymbol{X}_i\boldsymbol{X}_i^T-\boldsymbol{C}\right)\frac{\sqrt{G}}{\sqrt{2}}\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right).$$

Hence, by Lemma E.1.5, Lemma 3.2.2 and since

$$\frac{1}{G}\left\|\sum_{i=k+1}^{k+G}\left(\boldsymbol{X}_i\boldsymbol{X}_i^T-\boldsymbol{C}\right)\right\|_F=O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}}\right)$$

holds uniformly in k by (3.29), we receive

$$\max_{0\le k\le n-G}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\frac{1}{\sqrt{2G}}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}_0\right)-\frac{\sqrt{G}}{\sqrt{2}}\boldsymbol{C}\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right)\right)\right\|$$
$$\le\left\|\boldsymbol{\Sigma}^{-1/2}\right\|_F\max_{0\le k\le n-G}\frac{1}{G}\left\|\sum_{i=k+1}^{k+G}\left(\boldsymbol{X}_i\boldsymbol{X}_i^T-\boldsymbol{C}\right)\right\|_F\max_{0\le k\le n-G}\frac{\sqrt{G}}{\sqrt{2}}\left\|\widehat{\boldsymbol{\beta}}_{k+1,k+G}-\boldsymbol{\beta}_0\right\|$$
$$=O_P\left(\frac{\log(n/G)}{\sqrt{G}}\right)=o_P\left((\log(n/G))^{-1/2}\right),$$

where the last step follows from Assumption A.1.1 with

$$\frac{\log(n/G)\sqrt{\log(n/G)}}{\sqrt{G}}\le\frac{n^{1/(2+\nu)}\sqrt{\log n}}{\sqrt{G}}\to 0\quad\text{as } n\to\infty.$$

This shows (3.30). □

Now, the limit distribution of the statistic can be derived under the null hypothesis.

Theorem 3.2.4. *Let Assumption A.1.1 hold on the bandwidth. Furthermore, assume that Conditions (R1) to (R6) are fulfilled.*

(a) Then, under H_0,

$$a(n/G)W_n^{linear}(G) - b(n/G) \xrightarrow{D} E$$

with E as Gumbel distributed random variable as in Theorem 2.1.1 and with $a(x)$ and $b(x)$ as in (2.1).

(b) The matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{C}$ can be replaced by estimator sequences $\widehat{\boldsymbol{\Sigma}}_{k,n}$ and $\widehat{\boldsymbol{C}}_{k,n}$ fulfilling

$$\max_{G \le k \le n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} \widehat{\boldsymbol{C}}_{k,n} - \boldsymbol{\Sigma}^{-1/2} \boldsymbol{C} \right\|_F = o_P\left((\log(n/G))^{-1} \right)$$

without changing the results of part (a).

Proof. (a) By Lemma 3.2.3 we know that Condition (3.2) from Section 3.1.1, which has been specified for the linear regression model in (3.30), holds. Moreover, Assumption A.1.3 is directly given by Condition (R6) as

$$\boldsymbol{H}(Y_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0) = -\boldsymbol{X}_i \varepsilon_i$$

holds for all $i = 1, \ldots, n$ under the null hypothesis. Thus, applying the same arguments as in the proof of Theorem 3.1.8 yields the assertion.

(b) The result can be shown similarly to part (b) of Theorem 3.1.8.

□

For models with strictly exogenous regressors we propose to use a global estimator, computed on the whole sample, for the expectation matrix $\boldsymbol{C}$ since the estimation of this matrix would not be influenced by changes under the alternative and therefore all the information available should be incorporated. In contrast, the estimation of the matrix $\boldsymbol{\Sigma}$ integrates the estimation of the error variance σ^2 in some way which is based on estimated residuals and would be contaminated by changes. For the mean change model, Muhsal (2013) and Eichinger & Kirch (2018) pointed out that the classical variance estimator, computed on the whole sample, overestimates the error variance under the alternative as it is contaminated by the changes. They further proposed to use a MOSUM-type estimator which gives a single estimate for each time point k computed on the G-environment of k. This estimator has the nice property that it consistently estimates the error variance at the change points and at time points being far from any change while overestimating the variance on intervals around the changes which leads to tighter peaks of the statistic close to true changes possibly improving the performance of the procedure. Hence, it might be interesting to check whether a MOSUM-type estimator for the error variance in the linear regression shows a similiar behavior. For a further discussion on that we refer to Section 4.1.4.

3.2.2. Asymptotics Under the Alternative

Under the alternative we allow for multiple changes in the regression coefficients:

$$Y_i = \begin{cases} \boldsymbol{X}_i^T \boldsymbol{\beta}_1 + \varepsilon_i, & \text{if} \quad 1 \le i \le k_{1,n} \\ \boldsymbol{X}_i^T \boldsymbol{\beta}_2 + \varepsilon_i, & \text{if} \quad k_{1,n} < i \le k_{2,n} \\ \quad \vdots & \\ \boldsymbol{X}_i^T \boldsymbol{\beta}_{q+1} + \varepsilon_i, & \text{if} \quad k_{q,n} < i \le n \end{cases},$$

with $\boldsymbol{X}_i = (1, X_{2,i}, \ldots, X_{p,i})^T$, $\boldsymbol{\beta}_j = (\beta_{1,j}, \ldots, \beta_{p,j})^T$ and q being the number of changes. Note that a change in $\boldsymbol{\beta}$ results in the non-stationarity of the sequence $\{Y_i\}_{i \ge 1}$, whereas the sequence $\{\varepsilon_i\}_{i \ge 1}$ is still stationary. In order to incorporate autoregressive structures, we merely assume that the regressor sequence $\{\boldsymbol{X}_i\}_{i \ge 1}$ is piecewise stationary so that

$$\boldsymbol{X}_i = \begin{cases} \boldsymbol{X}_i^{(1)}, & \text{if} \quad 1 \le i \le k_{1,n} \\ \boldsymbol{X}_i^{(2)}, & \text{if} \quad k_{1,n} < i \le k_{2,n} \\ \quad \vdots & \\ \boldsymbol{X}_i^{(q+1)}, & \text{if} \quad k_{q,n} < i \le n \end{cases},$$

where $\{\boldsymbol{X}_i^{(j)} : i \ge 1\}$, $j = 1, \ldots, q+1$, is stationary and satisfies the Assumptions (R1), (R3), (R5) and (R6). This can probably be relaxed to allow for starting values from the other regime under some additional technical effort.
We get that the response sequence $\{Y_i\}_{i \ge 1}$ is piecewise stationary as well with

$$Y_i = Y_i^{(j)} = \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j + \varepsilon_i,$$

for $k_{j-1,n} < i \le k_{j,n}$ and $j = 1, \ldots, q+1$.
Due to the structure of the regressor sequence we slightly change the Assumptions (R1) to (R6) as follows:

(R1*) The sequence $\{\boldsymbol{X}_i^{(j)}\}_{i \ge 1}$ is stationary and ergodic with $E\left(\left\|\boldsymbol{X}_1^{(j)}\right\|\right) < \infty$, for $j = 1, \ldots, q+1$.

(R2*) Let $\mathcal{F}_t = \sigma\left(\boldsymbol{X}_j, \varepsilon_{j-1}, j \le t\right)$. We assume that ε_t and $\mathcal{F}_t$ are independent.

(R3*) $\boldsymbol{C}_{(j)} := E\left(\boldsymbol{X}_1^{(j)} \boldsymbol{X}_1^{(j)T}\right)$ is a positive definite matrix, for $j = 1, \ldots, q+1$.

(R4*) The sequence $\{\varepsilon_i\}_{i \ge 1}$ is i.i.d. with $E(\varepsilon_1) = 0$, $0 < E(\varepsilon_1^2) := \sigma^2 < \infty$.

(R5*) Let the components of $\{\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\}_{i \ge 1}$ satisfy a strong invariance principle similar to that in Assumption A.1.3, for $j = 1, \ldots, q+1$.

(R6*) Let $\{\boldsymbol{X}_i^{(j)} \varepsilon_i\}_{i \ge 1}$ be a series with positive definite long-run covariance matrix $\boldsymbol{\Sigma}_{(j)}$ satisfying a strong invariance principle similar to that in Assumption A.1.3, for $j = 1, \ldots, q+1$.

Furthermore, we need an additional assumption on the expectation matrices of $\boldsymbol{X}_i \boldsymbol{X}_i^T$:

(R7*) Let the matrix $\delta \boldsymbol{C}_{(j)} + (1-\delta)\boldsymbol{C}_{(j+1)}$ be positive definite for all $\delta \in [0,1]$ and assume that $\sup_{\delta\in[0,1]} \left\| \left(\delta \boldsymbol{C}_{(j)} + (1-\delta)\boldsymbol{C}_{(j+1)}\right)^{-1} \right\|_F < \infty$, for all $j = 1, \ldots, q$.

Note that this assumption coincides with Assumption (R3*) if the regressors are strictly exogenous and not effected by the changes so that $\{\boldsymbol{X}_i\}$ is stationary.
In order to prove consistency for the test and the change point estimators the following lemmata are needed.

Lemma 3.2.5. *Let the sequences $\{\boldsymbol{X}_i\}_{i\geq 1}$ and $\{\varepsilon_i\}_{i\geq 1}$ satisfy the Assumptions (R1*), (R2*), (R3*), (R4*) and (R6*). Furthermore, let Assumption A.1.1 on the bandwidth and Assumption A.2.1 hold. Then,*

$$\sqrt{G}\left\|\widehat{\boldsymbol{\beta}}_{k_{j,n}+1,k_{j,n}+G} - \boldsymbol{\beta}_{j+1}\right\| = O_P(1) \text{ and } \sqrt{G}\left\|\widehat{\boldsymbol{\beta}}_{k_{j,n}-G+1,k_{j,n}} - \boldsymbol{\beta}_{j}\right\| = O_P(1)$$

for all change points $k_{j,n}$, $j = 1, \ldots, q$.

Proof. By Assumption A.2.1 we know that the estimator sequences $\widehat{\boldsymbol{\beta}}_{k_{j,n}+1,k_{j,n}+G}$ and $\widehat{\boldsymbol{\beta}}_{k_{j,n}-G+1,k_{j,n}}$ are computed on stationary subsamples $(Y^{(j+1)}_{k_{j,n}+1}, \ldots, Y^{(j+1)}_{k_{j,n}+G})$ and $(Y^{(j)}_{k_{j,n}-G+1}, \ldots, Y^{(j)}_{k_{j,n}})$, respectively. Hence, on noting that $\{\boldsymbol{X}_i\}_{i\geq 1}$ and $\{\varepsilon_i\}_{i\geq 1}$ satisfy the Assumptions (R1*), (R2*), (R3*), (R4*) and (R6*) and since

$$\boldsymbol{H}(Y_i^{(j)}, \boldsymbol{X}_i^{(j)}, \boldsymbol{\beta}_j) = -\boldsymbol{X}_i^{(j)}\left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T}\boldsymbol{\beta}_j\right) = -\boldsymbol{X}_i^{(j)}\varepsilon_i, \quad j = 1, \ldots, q+1,$$

Lemma 3.2.1 shows the assertion. □

This lemma shows $\sqrt{G}$-consistency of the estimator sequences holding pointwise in $k_{j,n}$ which will be used for proving asymptotic power one of the test. However, this is not sufficient if we want to derive consistency for the corresponding estimators. Hence, we need to derive some uniform results as in the previous section. Therefore, we define $\widetilde{\boldsymbol{\beta}}^{(l)}_{k,n,G}$ similar to $\widetilde{\boldsymbol{\theta}}^{(l)}_{k,n,G}$ as unique zero of $F_l(k, n, G, \boldsymbol{\theta})$ in (3.10) and (3.11), $l = 1, 2$. Since $E(\boldsymbol{X}_i\varepsilon_i) = \boldsymbol{0}$, as shown in (3.25), holds under the alternative as well we obtain

$$\begin{aligned}
&E\left(\boldsymbol{H}(Y_i^{(j)}, \boldsymbol{X}_i^{(j)}, \boldsymbol{\beta})\right) = E\left(-\boldsymbol{X}_i^{(j)}\left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T}\boldsymbol{\beta}\right)\right)\\
&= E\left(-\boldsymbol{X}_i^{(j)}\left(\boldsymbol{X}_i^{(j)T}\boldsymbol{\beta}_j + \varepsilon_i - \boldsymbol{X}_i^{(j)T}\boldsymbol{\beta}\right)\right) = E\left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T}\right)\left(\boldsymbol{\beta} - \boldsymbol{\beta}_j\right) - E\left(\boldsymbol{X}_i^{(j)}\varepsilon_i\right)\\
&= \boldsymbol{C}_{(j)}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_j\right),
\end{aligned}$$

implying that

$$\begin{aligned}
&\frac{k_{j,n}-k}{G}E\left(\boldsymbol{H}(Y_i^{(j)}, \boldsymbol{X}_i^{(j)}, \boldsymbol{\beta})\right) + \frac{k+G-k_{j,n}}{G}E\left(\boldsymbol{H}(Y_i^{(j+1)}, \boldsymbol{X}_i^{(j+1)}, \boldsymbol{\beta})\right)\\
&= \frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_j\right) + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_{j+1}\right)\\
&= \left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)} + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)\boldsymbol{\beta}\\
&\quad - \left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right)
\end{aligned}$$

and similarly

$$\begin{aligned}&\frac{k_{j,n}-k+G}{G}E\left(\boldsymbol{H}(Y_i^{(j)},\boldsymbol{X}_i^{(j)},\boldsymbol{\beta})\right)+\frac{k-k_{j,n}}{G}E\left(\boldsymbol{H}(Y_i^{(j+1)},\boldsymbol{X}_i^{(j)},\boldsymbol{\beta})\right)\\&=\left(\frac{k_{j,n}-k+G}{G}\boldsymbol{C}_{(j)}+\frac{k-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)\boldsymbol{\beta}\\&\quad-\left(\frac{k_{j,n}-k+G}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j+\frac{k-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right).\end{aligned}$$

Thus, we get

$$F_1(k,n,G,\boldsymbol{\beta})=\begin{cases}\boldsymbol{C}_{(j)}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_j\right), & \text{if } k\in A_{j,n,G}\\ \boldsymbol{C}_{(j+1)}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_{j+1}\right), & \text{if } k\in B^{(2)}_{j,n,G}\\ \left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}+\frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)\boldsymbol{\beta} & \\ \quad-\left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j+\frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right), & \text{if } k\in B^{(1)}_{j,n,G}\end{cases}$$

and

$$F_2(k,n,G,\boldsymbol{\beta})=\begin{cases}\boldsymbol{C}_{(j)}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_j\right), & \text{if } k\in A_{j,n,G}\\ \boldsymbol{C}_{(j)}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_{j}\right), & \text{if } k\in B^{(1)}_{j,n,G}\\ \left(\frac{k_{j,n}-k+G}{G}\boldsymbol{C}_{(j)}+\frac{k-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)\boldsymbol{\beta} & \\ \quad-\left(\frac{k_{j,n}-k+G}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j+\frac{k-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right), & \text{if } k\in B^{(2)}_{j,n,G}\end{cases},$$

such that the unique zeros $\widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}$ and $\widetilde{\boldsymbol{\beta}}^{(2)}_{k,n,G}$ of the functions above can be specified as follows

$$\widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}=\begin{cases}\boldsymbol{\beta}_j, & \text{if } k\in A_{j,n,G}\\ \boldsymbol{\beta}_{j+1}, & \text{if } k\in B^{(2)}_{j,n,G}\\ \left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}+\frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)^{-1} & \\ \quad\left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j+\frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right), & \text{if } k\in B^{(1)}_{j,n,G}\end{cases}\tag{3.31}$$

and

$$\widetilde{\boldsymbol{\beta}}^{(2)}_{k,n,G}=\begin{cases}\boldsymbol{\beta}_j, & \text{if } k\in A_{j,n,G}\\ \boldsymbol{\beta}_{j}, & \text{if } k\in B^{(1)}_{j,n,G}\\ \left(\frac{k_{j,n}-k+G}{G}\boldsymbol{C}_{(j)}+\frac{k-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)^{-1} & \\ \quad\left(\frac{k_{j,n}-k+G}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j+\frac{k-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right), & \text{if } k\in B^{(2)}_{j,n,G}\end{cases}\tag{3.32}$$

with $A_{j,n,G}$, $B^{(1)}_{j,n,G}$ and $B^{(2)}_{j,n,G}$ as defined in (3.7), (3.8) and (3.9).

Lemma 3.2.6. *Let the sequences $\{\boldsymbol{X}_i\}_{i\geq 1}$ and $\{\varepsilon_i\}_{i\geq 1}$ satisfy the Assumptions (R1*) to (R7*). Furthermore, let Assumption A.1.1 on the bandwidth and Assumption A.2.1 hold. Then,*

(a) $\left\|\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right)$ *and* $\left\|\widehat{\boldsymbol{\beta}}_{k-G+1,k} - \widetilde{\boldsymbol{\beta}}^{(2)}_{k,n,G}\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right)$ *uniformly on* $A_{n,G}$,

(b) $\left\|\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right\| = O_P\left(\frac{1}{\sqrt{G}}\right)$ *and* $\left\|\widehat{\boldsymbol{\beta}}_{k-G+1,k} - \widetilde{\boldsymbol{\beta}}^{(2)}_{k,n,G}\right\| = O_P\left(\frac{1}{\sqrt{G}}\right)$ *uniformly on* $B_{n,G}$.

Proof. (a) On noting that

$$Y_i = Y_i^{(j)} = \boldsymbol{X}_i^{(j)}\boldsymbol{\beta}_j + \varepsilon_i, \quad i = k-G+1, \ldots, k+G,$$

holds for all $k \in A_{j,n,G}$, and that the Assumptions (R1*) to (R6*) are satisfied, Lemma 3.2.2 can be applied to receive

$$\max_{k \in A_{j,n,G}} \left\|\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \boldsymbol{\beta}_j\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right)$$

and

$$\max_{k \in A_{j,n,G}} \left\|\widehat{\boldsymbol{\beta}}_{k-G+1,k} - \boldsymbol{\beta}_j\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right).$$

Since q is finite and $A_{n,G} = \bigcup_{j=1}^{q+1} A_{j,n,G}$ we can conclude that

$$\max_{k \in A_{n,G}} \left\|\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \boldsymbol{\beta}_j\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right)$$

and

$$\max_{k \in A_{n,G}} \left\|\widehat{\boldsymbol{\beta}}_{k-G+1,k} - \boldsymbol{\beta}_j\right\| = O_P\left(\sqrt{\frac{\log(n/G)}{G}}\right).$$

(b) Since $B_{n,G} = B^{(1)}_{n,G} \cup B^{(2)}_{n,G}$, it is sufficient to show that the assertions hold uniformly on these two subsets. We only consider the set $B^{(1)}_{n,G}$ as the results on $B^{(2)}_{n,G}$ can be derived in an analogous manner.
For the first statement we have to mind that $k_{j,n} \in \{k+1, \ldots, k+G-1\}$. By (3.22) we know that

$$\sum_{i=k+1}^{k+G} \boldsymbol{X}_i Y_i = \sum_{i=k+1}^{k+G} \boldsymbol{X}_i \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{k+1,k+G},$$

which is equivalent to

$$\sum_{i=k+1}^{k+G} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right) = \sum_{i=k+1}^{k+G} \boldsymbol{X}_i \boldsymbol{X}_i^T \left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right). \tag{3.33}$$

At first, we approximate the left hand side of the equation above. As $(Y_{k+1},\ldots,Y_{k_{j,n}}) = (Y^{(j)}_{k+1},\ldots,Y^{(j)}_{k_{j,n}})$ and $(Y_{k_{j,n}+1},\ldots,Y_{k+G}) = (Y^{(j+1)}_{k_{j,n}+1},\ldots,Y^{(j+1)}_{k+G})$ hold for all $k \in B^{(1)}_{j,n,G}$ we obtain

$$
\begin{aligned}
&\sum_{i=k+1}^{k+G} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right) \qquad (3.34)\\
&= \sum_{i=k+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T} \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right) + \sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{X}_i^{(j+1)} \left(Y_i^{(j+1)} - \boldsymbol{X}_i^{(j+1)T} \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right)\\
&= \sum_{i=k+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right) + \sum_{i=k+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i\\
&\quad + \sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} \left(\boldsymbol{\beta}_{j+1} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right) + \sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{X}_i^{(j+1)} \varepsilon_i\\
&= \left(\sum_{i=k+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right) + (k_{j,n} - k)\boldsymbol{C}_{(j)}\right) \left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right)\\
&\quad + \left(\sum_{i=k_{j,n}+1}^{k+G} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)}\right) + (k+G-k_{j,n})\boldsymbol{C}_{(j+1)}\right) \left(\boldsymbol{\beta}_{j+1} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right)\\
&\quad + \sum_{i=k+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i + \sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{X}_i^{(j+1)} \varepsilon_i.
\end{aligned}
$$

Furthermore, note that by (3.31) and Assumption (R7*) in connection with Lemma E.1.5

$$
\begin{aligned}
&\max_{k\in B^{(1)}_{j,n,G}} \left\|\widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right\| \qquad (3.35)\\
&= \max_{k\in B^{(1)}_{j,n,G}} \Bigg\| \left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)} + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)^{-1}\\
&\qquad\qquad \left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right)\Bigg\|\\
&\le \sup_{\delta\in[0,1]} \left\|\left(\delta\boldsymbol{C}_{(j)} + (1-\delta)\boldsymbol{C}_{(j+1)}\right)^{-1}\right\|_F \left(\left\|\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j\right\| + \left\|\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}\right\|\right) = O(1).
\end{aligned}
$$

Moreover, by Lemma E.2.14 applied to each component of $\{\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\}$, $j = 1,\ldots,q+1$, and Lemma E.1.6 (a) we get

$$
\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{G}\left\|\sum_{i=k+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right)\right\|_F = O_P\left(\frac{1}{\sqrt{G}}\right) \quad \text{and} \qquad (3.36)
$$

$$
\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{G}\left\|\sum_{i=k_{j,n}+1}^{k+G} \left(\boldsymbol{X}_i^{(j+1)}\boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)}\right)\right\|_F = O_P\left(\frac{1}{\sqrt{G}}\right).
$$

Thus, in combination with (3.35) and Lemma E.1.5 we receive

$$\max_{k\in B^{(1)}_{j,n,G}} \left\| \frac{1}{G}\sum_{i=k+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right)\left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right)\right\|$$
$$\leq \max_{k\in B^{(1)}_{j,n,G}} \left\| \frac{1}{G}\sum_{i=k+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right)\right\|_F \max_{k\in B^{(1)}_{j,n,G}} \left\|\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right\| = O_P\left(\frac{1}{\sqrt{G}}\right)$$

and similarly

$$\max_{k\in B^{(1)}_{j,n,G}} \left\| \frac{1}{G}\sum_{i=k_{j,n}+1}^{k+G} \left(\boldsymbol{X}_i^{(j+1)}\boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)}\right)\left(\boldsymbol{\beta}_{j+1} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right)\right\| = O_P\left(\frac{1}{\sqrt{G}}\right).$$

Furthermore, as by (3.25) $E\left(\boldsymbol{X}_i^{(j)}\varepsilon_i\right) = \mathbf{0}$, $j = 1,\ldots,q+1$, Lemma E.2.14 together with Assumption (R6*) shows

$$\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{G}\left\|\sum_{i=k+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)}\varepsilon_i\right\| = O_P\left(\frac{1}{\sqrt{G}}\right)$$

and

$$\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{G}\left\|\sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{X}_i^{(j+1)}\varepsilon_i\right\| = O_P\left(\frac{1}{\sqrt{G}}\right).$$

Hence, by considering (3.34) again we can conclude that

$$\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\left(Y_i - \boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right)$$
$$= \frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}\left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right) + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\left(\boldsymbol{\beta}_{j+1} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right) + O_P\left(\frac{1}{\sqrt{G}}\right)$$
$$= \frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\boldsymbol{\beta}_{j+1}$$
$$- \left(\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)} + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}\right)\widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G} + O_P\left(\frac{1}{\sqrt{G}}\right)$$
$$= O_P\left(\frac{1}{\sqrt{G}}\right) \text{ uniformly in } k \in B^{(1)}_{j,n,G},$$

where the last line follows directly from the definition of $\widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}$ in (3.31). Moreover, applying (3.36) yields

$$\frac{1}{G}\sum_{i=k+1}^{k+G}\boldsymbol{X}_i\boldsymbol{X}_i^T\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G}\right)$$

$$= \left(o_P(1) + \frac{k_{j,n} - k}{G} \boldsymbol{C}_{(j)} + \frac{k + G - k_{j,n}}{G} \boldsymbol{C}_{(j+1)} \right) \left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G} \right),$$
uniformly in $k \in B^{(1)}_{j,n,G}$.

Finally, with (3.33) we obtain

$$O_P\left(\frac{1}{\sqrt{G}}\right) = \left(o_P(1) + \frac{k_{j,n} - k}{G} \boldsymbol{C}_{(j)} + \frac{k + G - k_{j,n}}{G} \boldsymbol{C}_{(j+1)} \right) \left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G} \right),$$
uniformly in $k \in B^{(1)}_{j,n,G}$.

By Assumption (R7*) in combination with Lemma E.2.21, multiplying the inverse of $\frac{k_{j,n}-k}{G}\boldsymbol{C}_{(j)} + \frac{k+G-k_{j,n}}{G}\boldsymbol{C}_{(j+1)}$ to both sides of the equation above leads to

$$O_P\left(\frac{1}{\sqrt{G}}\right) = \left(o_P(1) + \boldsymbol{I}_p\right) \left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G} \right), \text{ uniformly in } k \in B^{(1)}_{j,n,G},$$

which shows the assertion on $B^{(1)}_{j,n,G}$ by Lemma E.2.22. Since q is finite and $B^{(1)}_{n,G} = \bigcup_{j=1}^{q+1} B^{(1)}_{j,n,G}$ we get

$$\max_{k \in B^{(1)}_{n,G}} \left\| \widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G} \right\| = O_P\left(\frac{1}{\sqrt{G}}\right).$$

For proving the second statement in (b) note that $Y_i = Y_i^{(j)}$, $i = k-G+1, \ldots, k$, and $\widetilde{\boldsymbol{\beta}}^{(2)}_{k,n,G} = \boldsymbol{\beta}_j$ holds for all time points $k \in B^{(1)}_{j,n,G}$. Thus, by (3.22) we get

$$\sum_{i=k-G+1}^{k} \boldsymbol{X}_i^{(j)} \left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j \right) = \sum_{i=k-G+1}^{k} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \boldsymbol{\beta}_j \right). \quad (3.37)$$

Furthermore, as by (3.25) $E\left(\boldsymbol{X}_i^{(j)} \varepsilon_i\right) = \boldsymbol{0}$, $j = 1, \ldots, q+1$, Lemma E.2.15 with Assumption (R6*) yields

$$\max_{k \in B^{(1)}_{j,n,G}} \frac{1}{G} \left\| \sum_{i=k-G+1}^{k} \boldsymbol{X}_i^{(j)} \left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j \right) \right\| = \max_{k \in B^{(1)}_{j,n,G}} \frac{1}{G} \left\| \sum_{i=k-G+1}^{k} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{G}}\right).$$

Moreover, by Lemma E.2.15 applied to each component of $\{\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\}$, $j = 1, \ldots, q+1$, and Lemma E.1.6 (a) we get

$$\max_{k \in B^{(1)}_{j,n,G}} \left\| \frac{1}{G} \sum_{i=k-G+1}^{k} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\| = O_P\left(\frac{1}{\sqrt{G}}\right)$$

implying that

$$\frac{1}{G}\sum_{i=k-G+1}^{k} \boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} = \boldsymbol{C}_{(j)} + o_P(1) \text{ uniformly in } k \in B_{j,n,G}^{(1)}.$$

Thus, by considering (3.37) again, we obtain

$$O_P\left(\frac{1}{\sqrt{G}}\right) = \left(o_P(1) + \boldsymbol{C}_{(j)}\right)\left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \boldsymbol{\beta}_j\right) \text{ uniformly in } k \in B_{j,n,G}^{(1)}.$$

Finally, Lemma E.2.22 combined with Assumption (R3*) completes the proof.

□

Moreover, we have to take into consideration that the asymptotic covariance matrix $\boldsymbol{\Gamma}_k$ of $\sqrt{G}\widehat{\boldsymbol{\beta}}_{k-G+1,k}$ actually changes with k under the alternative. It can be specified as follows:

$$\boldsymbol{\Gamma}_k = \boldsymbol{C}_k^{-1}\boldsymbol{\Sigma}_k\boldsymbol{C}_k^{-1} \tag{3.38}$$

with $\boldsymbol{C}_k = \boldsymbol{C}_{(j)}$ and $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{(j)}$ for $k_{j-1,n} < k \le k_{j,n},\ j = 1, \ldots, q+1.$

The following theorem shows that the test for the linear regression model, which is similar to that in Section 3.1.2.1, correctly rejects the null hypothesis under the alternative with probability tending to one.

Theorem 3.2.7. *Let the sequences $\{\boldsymbol{X}_i\}_{i\ge1}$ and $\{\varepsilon_i\}_{i\ge1}$ satisfy the Assumptions (R1*) to (R4*) and (R6*). Furthermore, let Assumption A.1.1 on the bandwidth and Assumption A.2.1 hold.*

(a) Then, under H_1, we obtain for any $z \in \mathbb{R}$

$$\lim_{n\to\infty} P(a(n/G)W_n^{linear}(G) - b(n/G) \ge z) = 1,$$

i.e. the test has asymptotic power one.

(b) The matrices $\boldsymbol{\Sigma}_k$ and $\boldsymbol{C}_k$ can be replaced by estimator sequences $\widehat{\boldsymbol{\Sigma}}_{k,n}$ and $\widehat{\boldsymbol{C}}_{k,n}$ satisfying the assumption:

(I) $\max_{k\in B_{n,G}} \left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\widehat{\boldsymbol{C}}_{k,n} - \boldsymbol{\Gamma}_{A,k}^{-1/2}\right\|_F = o_P(1)$, with $B_{n,G} = B_{n,G}^{(1)} \cup B_{n,G}^{(2)}$ and where $\{\boldsymbol{\Gamma}_{A,k}\}_{k\ge1}$ is a sequence of positive definite matrices fulfilling $\sup_k \|\boldsymbol{\Gamma}_{A,k}\|_F < \infty$ and $\sup_k \left\|\boldsymbol{\Gamma}_{A,k}^{-1/2}\right\|_F < \infty$.

Proof. The result of part (a) can be shown in an analoguous manner to part (a) of Theorem 3.1.12 by using Lemma 3.2.5 and on noting that $\boldsymbol{\Gamma}_{k_{j,n}} = \boldsymbol{C}_{(j)}^{-1}\boldsymbol{\Sigma}_{(j)}\boldsymbol{C}_{(j)}^{-1}$ as in (3.38) is positive definite due to Lemma E.1.7 in combination with the positive definiteness of $\boldsymbol{\Sigma}_{(j)}$ and the regularity of $\boldsymbol{C}_{(j)}$.
Moreover, we can derive the assertion in (b) by using similar arguments as in Theorem 3.1.12 (b).

□

The estimators for the number and the locations of the changes denoted by $\widehat{q}_n$ and $\widehat{k}_{j,n}$, $j = 1, \ldots, \widehat{q}_n$, are determined in an analogous manner to Section 3.1.2.2. Similar to the general setting, for proving consistency of these estimators we need the following results.

Lemma 3.2.8. *Let Assumption A.1.1 on the bandwidth and Assumption A.2.1 be fulfilled. Furthermore, assume that Assumption (R7*) is satisfied. Then, for all* $j = 1, \ldots, q$,

(a) $\max_{k \in \bar{B}^{(1)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G} - \boldsymbol{\beta}_j \right\| = O(1)$ *and* $\max_{k \in \bar{B}^{(2)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\beta}}^{(2)}_{k,n,G} - \boldsymbol{\beta}_j \right\| = O(1)$,

(b) $\min_{k \in \bar{B}^{(1)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\beta}}^{(1)}_{k,n,G} - \boldsymbol{\beta}_j \right\| > c$ *and* $\min_{k \in \bar{B}^{(2)}_{j,n,G}} \left\| \widetilde{\boldsymbol{\beta}}^{(2)}_{k,n,G} - \boldsymbol{\beta}_j \right\| > c$ *for some* $c > 0$,

with $\bar{B}^{(l)}_{j,n,G}$, $l = 1, 2$, *as in (3.19).*

Proof. (a) The first statement follows directly from (3.35). The second assertion can be derived in an analogous manner by using (3.32) and Assumption (R7*).

(b) The same arguments as in the proof of Lemma 3.1.14 (b) can be used here $\left(\text{with } E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}^{(j)}_1, \boldsymbol{\theta}) \right\|_F\right) = \boldsymbol{C}_{(j)}\right)$.

□

The following theorem states the main result of this section and shows that the estimator for the number of changes is consistent for the true number q.

Theorem 3.2.9. *Let Assumption A.1.1 on the bandwidth and Assumption A.2.1 be fulfilled. Moreover, let the sequences* $\{\boldsymbol{X}_i\}_{i\geq 1}$ *and* $\{\varepsilon_i\}_{i\geq 1}$ *satisfy the Assumptions (R1*) to (R7*). Furthermore, assume that the sequence* $\{\alpha_n\}_{n\in\mathbb{N}}$ *fulfills Assumption A.2.8.*

(a) Then, it holds

$$P(\widehat{q}_n = q) \to 1, \quad \text{as } n \to \infty.$$

(b) The result remains true if the matrices $\boldsymbol{\Sigma}_k$ *and* $\boldsymbol{C}_k$ *are replaced by estimator sequences* $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *and* $\widehat{\boldsymbol{C}}_{k,n}$ *fulfilling:*

(I) $\max_{k\in B_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}^{-1/2}_{k,n} \widehat{\boldsymbol{C}}_{k,n} - \boldsymbol{\Gamma}^{-1/2}_{A,k} \right\|_F = o_P(1)$, *where* $\{\boldsymbol{\Gamma}_{A,k}\}_{k\geq 1}$ *is a sequence of positive definite matrices fulfilling* $\sup_k \|\boldsymbol{\Gamma}_{A,k}\|_F < \infty$ *and* $\sup_k \left\| \boldsymbol{\Gamma}^{-1/2}_{A,k} \right\|_F < \infty$.

(II) $\max_{k\in A_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}^{-1/2}_{k,n} \widehat{\boldsymbol{C}}_{k,n} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{C} \right\|_F = o_P\left((\log(n/G))^{-1}\right)$.

Proof. The results can be derived in an analogous manner to Theorem 3.1.15 by using Lemma 3.2.8 and Lemma 3.2.6. Furthermore, note that $\boldsymbol{H}(Y^{(j)}_i, \boldsymbol{X}^{(j)}_i, \boldsymbol{\beta}_j) = \boldsymbol{X}^{(j)}_i \varepsilon_i$, $j = 1, \ldots, q+1$, and that the sequence $\{\boldsymbol{X}^{(j)}_i \varepsilon_i\}_{i\geq 1}$ satisfies Assumption A.1.3 by Condition (R6*). Hence, Theorem 2.1.1 can be used here as well. □

Similar to the general setting, the following assertions can be proved as well.

Corollary 3.2.10. *Let the assumptions of Theorem 3.2.9 hold. Then,*

$$P\left(\max_{1\leq j\leq q}\min_{1\leq l\leq \hat{q}_n}\left|\widehat{k}_{l,n}-k_{j,n}\right|<G\right)\to 1.$$

Corollary 3.2.11. *Let the assumptions of Theorem 3.2.9 hold. Then,*

$$\max_{1\leq j\leq q}\min_{1\leq l\leq \hat{q}_n}\left|\widehat{\lambda}_{l,n}-\lambda_j\right|=O_P\left(\frac{G}{n}\right)=o_P(1).$$

3.3. Possible Problems of the Procedure

The choice of the bandwidth is not only an issue of the MOSUM score-type procedure, the performance of the MOSUM Wald-type procedure on finite samples also depends on the selection of the bandwidth. For a detailed discussion on that we refer to Section 2.4.1. As already described there, a possible solution for this problem is to run the procedure with several window lengths and merge the results appropriately by using a multiscale method as described in Chapter 5.

Moreover, note that the validity of the results for the MOSUM Wald-type statistic in some sense base on the condition that the estimator sequences $\widehat{\boldsymbol{\theta}}_{k+1,k+G}$ are identifiably unique which implies that the distance between all possible solutions of the estimating equation system $\sum_{i=k+1}^{k+G}\boldsymbol{H}(\mathbb{X}_i,\boldsymbol{\theta})=\mathbf{0}$ goes to zero. See, for example, Pötscher & Prucha (1997) Chapter 3 and Section 4.6. This condition has not been directly mentioned but it is implicated by the assumptions on the general setting and the linear regression model. This is an important condition as it guarantees that the signal of the Wald-type statistic can only be strictly positive on intervals around the true changes in the asymptotics. A violation of this assumption leads to an estimation error which possibly causes overestimation of the changes under alternative or a size problem under the null hypothesis.

4. Simulation Studies

4.1. Example: Linear Regression

We consider a linear regression model where the regressors are strictly exogenous and the regressors as well as the error are modelled by i.i.d. normally distributed random variables. This is a simple model in the sense that it does not incorporate any dependence structure and that there exists an explicit solution of the estimating equation system such that the least squares estimators can be computed easily without applying numerical methods. By conducting a simulation study on this linear regression model we want to get an impression about the pure difference in performance of the MOSUM Wald-type and score-type procedure in the absence of numerical errors and disturbances caused by the variance estimation in dependence settings. Intuitively, we would expect that the MOSUM Wald-type statistic performs much better as it directly focuses on the difference in the parameter vectors without making a detour over the estimating function like in the score-type approach.
However, before we start to analyse the simulation results, we need to check whether the assumptions of the MOSUM procedures are satisfied in this specific linear regression model.

4.1.1. Proving the Assumptions of the Wald-Type and Score-Type Approach

In the simulations, we consider an example of the follwing simple linear regression model:

- Under the null hypothesis:

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta}_0 + \varepsilon_i \quad \text{for } i = 1, \ldots, n,$$

- Under the alternative:

$$Y_i = Y_i^{(j)} = \boldsymbol{X}_i^T \boldsymbol{\beta}_j + \varepsilon_i \quad \text{for } k_{j-1,n} < i \leq k_{j,n}, j = 1, \ldots, q+1,$$

where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is an i.i.d. series with $\boldsymbol{X}_i = (1, X_{i,1}, X_{i,2})^T$, $X_{i,1} \sim N(\mu_1, \sigma_1^2)$ and $X_{i,2} \sim N(\mu_2, \sigma_2^2)$ and $\varepsilon_1, \ldots, \varepsilon_n$ is i.i.d. with $\varepsilon_i \sim N(0, \sigma^2)$.
At first, we consider the Assumptions of the MOSUM Wald-type statistics given in Section 3.2. Note that due to the stationarity of the sequence $\{\boldsymbol{X}_i\}$ under the null hypothesis and the alternative the Assumptions (R1) to (R6) coincide with the Assumptions (R1*) to (R6*). Moreover, Condition (R7*) simplifies to Assumption (R3). The Conditions (R1), (R2) and (R4) are obviously satisfied. Furthermore, we get that

the sequence $\{\boldsymbol{X}_i\varepsilon_i\}$ is i.i.d. with existing second and higher moments so that the invariance principle in Theorem 2 of Einmahl (1989) shows Assumption (R6). Besides, as the components of the sequence $\{\boldsymbol{X}_i\boldsymbol{X}_i^T - \boldsymbol{C}\}$ are univariate i.i.d. sequences with zero mean and existing second and higher moments Assumption (R5) can be obtained by the invariance principle proved by Komlós *et al.* (1975), Komlós *et al.* (1976) and Major (1976). Furthermore, we get that the positive semi-definite matrix

$$\boldsymbol{C} = E\left(\boldsymbol{X}_1\boldsymbol{X}_1^T\right) = \begin{pmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \sigma_1^2+\mu_1^2 & \mu_1\mu_2 \\ \mu_2 & \mu_1\mu_2 & \sigma_2^2+\mu_2^2 \end{pmatrix}$$

is positive definite as $\det(\boldsymbol{C}) = \sigma_1^2\sigma_2^2 > 0$.

For the score-type procedure, we show that the general assumptions of Theorem 2.1.1 and Theorem 2.1.8, stating the main results for MOSUM score-type statistics, are fulfilled.

Assumptions A.1.4 and A.1.3 under the Null Hypothesis:
We use the global least squares estimator $\widehat{\boldsymbol{\beta}}_{1,n}$ which is computed on the whole sample for calculating the MOSUM score-type statistic. This classical estimator is $\sqrt{n}$-consistent for the true parameter vector $\boldsymbol{\beta}_0$ under the null hypothesis so that $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$ and

$$\boldsymbol{H}(Y_i, \boldsymbol{X}_i, \widetilde{\boldsymbol{\beta}}) = \boldsymbol{H}(Y_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0) = -\boldsymbol{X}_i\varepsilon_i.$$

Thus, Condition (R6) directly yields Assumption A.1.3. Furthermore, with (3.23) we receive

$$\begin{aligned}
&\max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\beta}}_{1,n},k} - \boldsymbol{A}_{\boldsymbol{\beta}_0,k} \right\| \\
&= \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \left(\sum_{i=k+1}^{k+G} \boldsymbol{X}_i\boldsymbol{X}_i^T - \sum_{i=k-G+1}^{k} \boldsymbol{X}_i\boldsymbol{X}_i^T \right) \left(\widehat{\boldsymbol{\beta}}_{1,n} - \boldsymbol{\beta}_0 \right) \right\| \\
&\le \left(\max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \sum_{i=k+1}^{k+G} \boldsymbol{X}_i\boldsymbol{X}_i^T \right\|_F + \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \sum_{i=k-G+1}^{k} \boldsymbol{X}_i\boldsymbol{X}_i^T \right\|_F \right) \left\| \widehat{\boldsymbol{\beta}}_{1,n} - \boldsymbol{\beta}_0 \right\|,
\end{aligned}$$

where the last line follows from Lemma E.1.5. Since the arguments used here are very similar to that of previous proofs we only give a brief explanation. Applying Theorem E.2.12 and Condition (R5) in combination with Lemma E.1.6 (b) and Assumption A.1.1 together with the $\sqrt{n}$-consistency of the least-squares estimator leads to Assumption A.1.4.

Assumptions A.2.3, A.2.4 and A.2.9 under the Alternative:
Let us start with Assumptions A.2.4 and A.2.9. In Lemma 5.4.24 we will see that the least-squares estimator $\widehat{\boldsymbol{\beta}}_{1,n}$ is $\sqrt{n}$-consistent for $\widetilde{\boldsymbol{\beta}} = \sum_{j=1}^{q+1}(\lambda_j - \lambda_{j-1})\boldsymbol{\beta}_j$ under the alternative. Furthermore, we obtain

$$\max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\beta}}_{1,n},k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\beta}},k} \right\|$$

$$
\begin{aligned}
&= \max_{G \le k \le n-G} \frac{1}{\sqrt{2G}} \left\| \left(\sum_{i=k+1}^{k+G} \boldsymbol{X}_i \boldsymbol{X}_i^T - \sum_{i=k-G+1}^{k} \boldsymbol{X}_i \boldsymbol{X}_i^T \right) \left(\widehat{\boldsymbol{\beta}}_{1,n} - \widetilde{\boldsymbol{\beta}} \right) \right\| \\
&= o_P\left(a(n/G)^{-1} \right),
\end{aligned}
$$

where the last line follows from the same arguments which we used for proving Assumption A.1.3. This implies the statements of Assumptions A.2.4 and A.2.9. For deriving Assumption A.2.3, note that the transformed sequence $\{\boldsymbol{H}(Y_i^{(j)}, \boldsymbol{X}_i, \widetilde{\boldsymbol{\beta}})\}_{i \ge 1}$ with

$$
-\boldsymbol{H}(Y_i^{(j)}, \boldsymbol{X}_i, \widetilde{\boldsymbol{\beta}}) = \boldsymbol{X}_i \boldsymbol{X}_i^T \left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}} \right) + \boldsymbol{X}_i \varepsilon_i, \quad j = 1, \ldots, q+1,
$$

is i.i.d. with existing second and higher moments. Hence, the assumption follows from Theorem 2 in Einmahl (1989). Moreover, we can show that all changes are detectable by that statistic since

$$
E\left(\boldsymbol{H}(Y_i^{(j)}, \boldsymbol{X}_i, \widetilde{\boldsymbol{\beta}}) \right) = \boldsymbol{C}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_j \right) \ne \boldsymbol{C}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_{j+1} \right) = E\left(\boldsymbol{H}(Y_i^{(j+1)}, \boldsymbol{X}_i, \widetilde{\boldsymbol{\beta}} \right)
$$

holds for all $j = 1, \ldots, q$.

4.1.2. Simulating the Data

Let $\boldsymbol{X}_i = (1, X_{i,1}, X_{i,2})^T$ with $X_{i,1} \sim N(1,1)$ and $X_{i,2} \sim N(2,1)$ and $\varepsilon_i \sim N(1,1)$. We simulate a data sample of length $n = 1000$ and use 1000 replications in the study. Under the null hypothesis, let $\boldsymbol{\beta}_0 = (1,2,2)^T$ so that

$$
Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta}_0 + \varepsilon_i \quad \text{for } i = 1, \ldots, 1000.
$$

For evaluating the performance of the procedures under alternative we include three change points, $q = 3$, at $k_{1,1000} = 200, k_{2,1000} = 500$ and $k_{3,1000} = 800$. Furthermore, with $\boldsymbol{\beta}_1 = (1,2,2)^T, \boldsymbol{\beta}_2 = (1,1,2)^T, \boldsymbol{\beta}_3 = (2,1,2)^T$ and $\boldsymbol{\beta}_4 = (2,1,1)^T$ we get the following model

$$
Y_i = \begin{cases} \boldsymbol{X}_i^T \boldsymbol{\beta}_1 + \varepsilon_i, & \text{if } \; i \le 200 \\ \boldsymbol{X}_i^T \boldsymbol{\beta}_2 + \varepsilon_i, & \text{if } \; 200 < i \le 500 \\ \boldsymbol{X}_i^T \boldsymbol{\beta}_3 + \varepsilon_i, & \text{if } \; 500 < i \le 800 \\ \boldsymbol{X}_i^T \boldsymbol{\beta}_4 + \varepsilon_i, & \text{if } \; i > 800 \end{cases}.
$$

4.1.3. Estimating the Covariance Matrices

First note that in this specific setting, where the regressors and the error are independent i.i.d. sequences, the long-run covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\widetilde{\boldsymbol{\beta}})$ coincides with the covariance matrix of the estimating function. As the errors and the regressors are independent and the components of the regressor vector $\boldsymbol{X}_i$ are indenpendent as well a straightforward calculation of the covariance matrix shows that:

$$
\operatorname{Cov}\left(\boldsymbol{H}(\mathbb{X}_1, \widetilde{\boldsymbol{\beta}}) \right) = \sigma^2 \boldsymbol{C}
$$

holds under the null hypothesis and the alternative. Hence, estimating the covariance matrices for the Wald-type and score-type statistic comes down to estimating the error variance and the expectation matrix $\boldsymbol{C}$. We consider the MOSUM score-type statistic

$$\widehat{T}_{n,k}(G,\widehat{\boldsymbol{\beta}}_{1,n}) = \frac{1}{\sqrt{2G}\widehat{\sigma}_{n,k}} \left\| \widehat{\boldsymbol{C}}_n^{-1/2} \left(\sum_{i=k+1}^{k+G} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{1,n} \right) - \sum_{i=k-G+1}^{k} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{1,n} \right) \right) \right\|$$

and the MOSUM Wald-type statistic

$$\widehat{W}_{n,k}(G) = \frac{\sqrt{G}}{\widehat{\sigma}_{n,k}\sqrt{2}} \left\| \widehat{\boldsymbol{C}}_n^{1/2} \left(\widehat{\boldsymbol{\beta}}_{k+1,k+G} - \widehat{\boldsymbol{\beta}}_{k-G+1,k} \right) \right\|,$$

where $\widehat{\boldsymbol{C}}_n$ denotes the estimator of $\boldsymbol{C}$ and $\widehat{\sigma}^2_{n,k}$ represents the estimator of the error variance σ^2. As already explained in Section 3.2, in this setting it is reasonable to use a global estimator, calculated on the whole sample, for the expectation matrix $\boldsymbol{C}$. Thus, we take the sample mean $\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T$ as an estimator for $\boldsymbol{C}$ which is not contaminated by changes in the parameter vector $\boldsymbol{\beta}$ under the alternative since the distribution of the $\boldsymbol{X}_i$ does not change. In comparison to that estimators of the error variance are based on estimated residuals which are highly contaminated by changes in the regression parameter. Therefore, we prefer to apply a MOSUM-type estimator for σ^2 which is time dependent and calculates a variance estimate for each time point k on its G-environment. Nevertheless, in order to investigate how the performance of the procedures depends on the choice of estimator for the error variance, we use three different estimators in the simulations:

- GLOBAL:
$$\widehat{\sigma}^2_{n,k} = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{1,n})^2$$
We fit a linear regression model on the whole data sample and compute the sample variance of the corresponding residuals.

- LOCAL1:
$$\widehat{\sigma}^2_{n,k} = \frac{1}{2G} \left(\sum_{i=k-G+1}^{k} (Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{k-G+1,k})^2 + \sum_{i=k+1}^{k+G} (Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{k+1,k+G})^2 \right)$$
For each time point k we fit a linear regression model on the subsample from $k-G+1$ to k and on the subsample from $k+1$ to $k+G$ and take the sum of the corresponding squared residuals divided by $2G$.

- LOCAL2:
$$\widehat{\sigma}^2_{n,k} = \frac{1}{2G} \left(\sum_{i=k-G+1}^{k} (\hat{\epsilon}_{i,n} - \bar{\epsilon}_{k-G+1,k})^2 + \sum_{i=k+1}^{k+G} (\hat{\epsilon}_{i,n} - \bar{\epsilon}_{k+1,k+G})^2 \right)$$
with $\hat{\epsilon}_{i,n} := Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{1,n}$ and $\bar{\epsilon}_{l,u} := \frac{1}{u-l+1}\sum_{i=l}^{u} \hat{\epsilon}_{i,n}$.
We fit a linear regression model on the complete data sequence and compute the corresponding residuals denoted by $\hat{\epsilon}_{i,n}$. For each time point k we compute the sample variance of $\hat{\epsilon}_{k-G+1,n}, \ldots, \hat{\epsilon}_{k,n}$ and $\hat{\epsilon}_{k+1,n}, \ldots, \hat{\epsilon}_{k+G,n}$ and take the sum of both values multiplied by $\frac{G-1}{2G}$ as the LOCAL2 estimator.

4.1.4. Results of the Simulations

Table 4.1 shows the performance of the MOSUM Wald-type test with different bandwidths under the null hypothesis and the alternative at significance level $\alpha = 0.05$ for the different variance estimators. Under the null hypothesis, the empirical size of the tests gets relatively small if we choose a large bandwidth in the procedure. The smallest bandwidth of 20 is not appropriate, in particular if the LOCAL1 estimator is employed in the statistic, since the empirical size is too high whereas the tests with a bandwidth greater than 50 are conservative as the empirical value is less than the theoretical value of 0.05. Note that the tests based on the GLOBAL and LOCAL2 estimator reveal a more conservative behavior in comparison to the LOCAL1 estimator. Under the alternative, the size adjusted power is equal to one or at least close to one for all considered window lengths and variance estimators that shows the consistency of the MOSUM Wald-type tests empirically. Furthermore, note that the MOSUM score-type tests show similar results for the size adjusted power under the alternative, which are given in Table 4.2, but are even more conservative under the null hypothesis. Thus, we can conclude that the tests perform quite well.

Table 4.1.: Simulation results for the test based on the MOSUM Wald-type statistic

	$\boldsymbol{H_0}$ Empirical size	$\boldsymbol{H_1}$ Size adjusted power
GLOBAL		
$G = 20$	0.198	0.748
$G = 50$	0.030	1
$G = 80$	0.013	1
$G = 100$	0.014	1
$G = 120$	0.012	1
$G = 150$	0.013	1
LOCAL1		
$G = 20$	0.652	0.975
$G = 50$	0.087	1
$G = 80$	0.029	1
$G = 100$	0.029	1
$G = 120$	0.025	1
$G = 150$	0.022	1
LOCAL2		
$G = 20$	0.196	0.957
$G = 50$	0.040	1
$G = 80$	0.015	1
$G = 100$	0.018	1
$G = 120$	0.014	1
$G = 150$	0.014	1

Table 4.2.: Simulation results for the test based on the MOSUM score-type statistic

	H_0 Empirical size	H_1 Size adjusted power
GLOBAL		
$G = 20$	0.026	0.867
$G = 50$	0.010	1
$G = 80$	0.009	1
$G = 100$	0.008	1
$G = 120$	0.013	1
$G = 150$	0.007	1
LOCAL1		
$G = 20$	0.312	0.998
$G = 50$	0.047	1
$G = 80$	0.022	1
$G = 100$	0.022	1
$G = 120$	0.020	1
$G = 150$	0.012	1
LOCAL2		
$G = 20$	0.064	0.991
$G = 50$	0.014	1
$G = 80$	0.008	1
$G = 100$	0.012	1
$G = 120$	0.009	1
$G = 150$	0.007	1

Nevertheless, we are more interested in estimation as in testing and for this reason we analyse the simualtion results for the change point estimators in more detail. The Tables 4.3 and 4.4 show the performance of the estimator for the number and the locations of the changes based on the MOSUM Wald-type and MOSUM score-type statistic. The simulation results for the estimated number are summarized in the columns two to six. The entries represent the proportions of repetitions in which the estimated number of changes was less than or equal to one or equal to two and so on. For example, the MOSUM Wald-type procedure with LOCAL1 estimator for the error variance and bandwidth $G = 100$ correctly estimates the number of change points in 94.5% of the simulated samples. In the last three columns, the performance of the estimators for the change point locations are recorded as follows. The entries are the proportions of repetitions in which we got a change point estimate lying in an interval of $[k_{j,n} - 20, k_{j,n} + 20]$ around the true change point, $j = 1, 2, 3$. For instance, the MOSUM Wald-type procedure with LOCAL1 estimator and bandwidth $G = 100$ produced a change point estimate in the interval $[480, 520]$ in 93.8% of the cases. Note that we used $\alpha = 0.05$ and $\epsilon = 0.2$ in the estimation process.

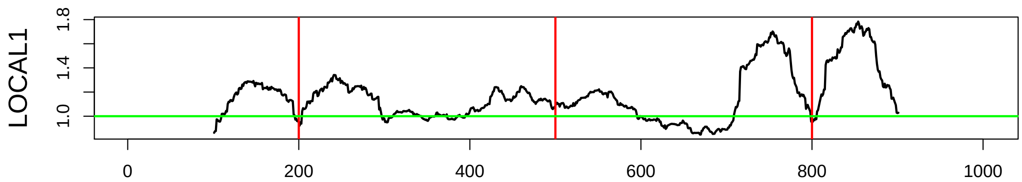
Table 4.3.: Simulation results for the estimator of the number and the locations of the changes based on the MOSUM Wald-type statistic

	Estimated number					**Estimated change point in**		
	≤ 1	2	$q = 3$	4	≥ 5	$[180, 220]$	$[480, 520]$	$[730, 770]$
GLOBAL								
$G = 20$	0.990	0.010	0.000	0.000	0.000	0.010	0.000	0.442
$G = 50$	0.523	0.469	0.008	0.000	0.000	0.467	0.016	1.000
$G = 80$	0.023	0.745	0.230	0.002	0.000	0.965	0.231	1.000
$G = 100$	0.000	0.480	0.516	0.004	0.000	0.991	0.510	1.000
$G = 120$	0.000	0.222	0.772	0.006	0.000	0.990	0.764	1.000
$G = 150$	0.000	0.045	0.951	0.004	0.000	0.990	0.926	1.000
LOCAL1								
$G = 20$	0.404	0.406	0.154	0.033	0.003	0.461	0.140	0.957
$G = 50$	0.018	0.445	0.501	0.035	0.001	0.963	0.539	1.000
$G = 80$	0.000	0.094	0.866	0.040	0.000	0.998	0.885	1.000
$G = 100$	0.000	0.030	0.945	0.025	0.000	0.998	0.938	1.000
$G = 120$	0.000	0.008	0.981	0.011	0.000	0.992	0.957	1.000
$G = 150$	0.000	0.000	0.991	0.009	0.000	0.995	0.970	1.000
LOCAL2								
$G = 20$	0.866	0.127	0.007	0.000	0.000	0.090	0.062	0.790
$G = 50$	0.092	0.544	0.354	0.010	0.000	0.852	0.409	1.000
$G = 80$	0.001	0.155	0.822	0.021	0.001	0.992	0.825	1.000
$G = 100$	0.000	0.044	0.942	0.014	0.000	0.996	0.925	1.000
$G = 120$	0.000	0.012	0.980	0.008	0.000	0.993	0.954	1.000
$G = 150$	0.000	0.000	0.993	0.007	0.000	0.992	0.970	1.000

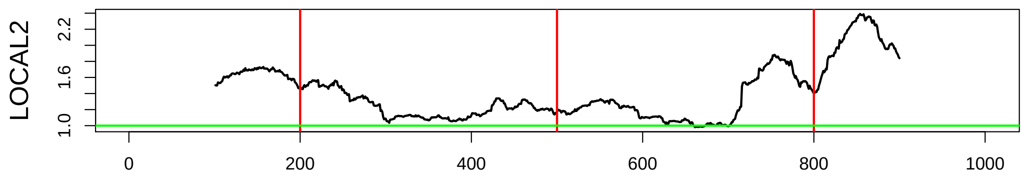
Table 4.4.: Simulation results for the estimator of the number and the locations of the of changes based on the MOSUM score-type statistic

	Estimated number					**Estimated change point in**		
	≤ 1	2	$q = 3$	4	≥ 5	$[180, 220]$	$[480, 520]$	$[730, 770]$
GLOBAL								
$G = 20$	0.950	0.045	0.004	0.001	0.000	0.043	0.000	0.452
$G = 50$	0.505	0.470	0.024	0.001	0.000	0.471	0.017	0.997
$G = 80$	0.047	0.714	0.238	0.001	0.000	0.904	0.241	0.999
$G = 100$	0.006	0.451	0.529	0.013	0.001	0.969	0.516	0.997
$G = 120$	0.112	0.462	0.421	0.005	0.000	0.960	0.743	0.536
$G = 150$	0.051	0.942	0.007	0.000	0.000	0.969	0.900	0.000
LOCAL1								
$G = 20$	0.080	0.163	0.231	0.211	0.315	0.499	0.127	0.913
$G = 50$	0.023	0.309	0.453	0.158	0.057	0.932	0.542	1.000
$G = 80$	0.000	0.086	0.795	0.111	0.008	0.981	0.875	1.000
$G = 100$	0.001	0.036	0.919	0.043	0.001	0.979	0.936	0.986
$G = 120$	0.006	0.592	0.396	0.005	0.001	0.975	0.947	0.391
$G = 150$	0.002	0.988	0.010	0.000	0.000	0.987	0.953	0.000
LOCAL2								
$G = 20$	0.741	0.206	0.045	0.003	0.005	0.160	0.046	0.746
$G = 50$	0.107	0.528	0.329	0.034	0.002	0.807	0.415	1.000
$G = 80$	0.001	0.146	0.807	0.044	0.002	0.976	0.825	1.000
$G = 100$	0.001	0.043	0.925	0.031	0.000	0.983	0.916	0.999
$G = 120$	0.003	0.431	0.558	0.008	0.000	0.981	0.943	0.565
$G = 150$	0.003	0.990	0.007	0.000	0.000	0.989	0.952	0.000

As expected, the estimator based on score-type statistics performs worse than the Wald-type estimator. Its performance highly depends on the selection of the bandwidth and can only compete with the Wald-type procedure for $G = 100$. For bandwidths smaller than 100, exept 20, the score-type statistic performs quite well in detecting the first and third change point but has problems in finding the second one. In comparison to that, for bandwidth 120 and 150 the score-type procedure localises the first and the second change in more than 90% of the cases but does poorly in dectecting the third one. Hence $G = 100$ seems to be the optimal choice of the bandwidth here. This is exactly half of the minimal distance between two adjacent structural breaks $\min_{0\leq j\leq q} |k_{j+1,n} - k_{j,n}|$ with $k_{0,n} = 1$ and $k_{q+1,n} = n$ which is a theoretically reasonable window length as explained in Section 2.4.1. However, this value is unknown and cannot be determined in practice so that the application of a multiscale procedure as described in Chapter 5, which merges the results obtained by different bandwidths in an appropriate way, is essential to make the MOSUM score-type approach competitive. In contrast, the MOSUM Wald-type procedure seems to be less sensitive to the choice of the bandwidth and shows very good results for window lengths of $100, 120, 150$. Moreover, the performance of the procedures also depends on which estimator of the error variance has been used in the statistic. The MOSUM score-type statistic as well as the MOSUM Wald-type statistic perform worse if the GLOBAL error variance estimator is used in comparison to the results for the LOCAL1 and LOCAL2 variance estimator which perform both very well. In the MOSUM Wald-type procedure the LOCAL1 estimator shows slightly better results than LOCAL2 whereas in the MOSUM score-type procedure the LOCAL2 estimator performs best.
Now, we consider the estimators of the error variance more closely in order to explain the differences in performance. The GLOBAL variance estimator relies on the residuals of the global estimator $\widehat{\boldsymbol{\beta}}_{1,n}$ which are obviously affected by changes in the parameter vector. This leads to overestimation of the error variance and, thus, deterioration of the performance of the procedure as the values of the statistic get smaller. In contrast, the LOCAL1 estimator bases on residuals obtained from the local estimators $\widehat{\boldsymbol{\beta}}_{k-G+1,k}$ and $\widehat{\boldsymbol{\beta}}_{k+1,k+G}$ and is therefore able to react to changes in $\boldsymbol{\beta}$ in some way. It reveals a similar behavior as the MOSUM-type estimator of the error variance in the classical mean change model used by Eichinger & Kirch (2018). The plot below shows the estimates by using the LOCAL1 estimator for one of the simulated data samples where the red vertical lines give the change points and the green horizontal line illustrates the true error variance $\sigma^2 = 1$.



We can see that the LOCAL1 estimates get quite large in intervals around the true changes whereas the estimates at the changes and at time points which are far away from any change are relatively small. This improves the performance of the procedure by making the peaks at the changes tighter. Consequently, we would recommend to apply the LOCAL1 estimator in general. However, from a pratical point of view it would not make much sense to use this estimator in the score-type statistic as it requires the computation of the local parameter estimators $\widehat{\boldsymbol{\beta}}_{k-G+1,k}$ and $\widehat{\boldsymbol{\beta}}_{k+1,k+G}$ so that the computational advantages of the score-type procedure would vanish. For this reason we have introduced the LOCAL2 estimator for the error variance which is a MOSUM-type estimator based on the global residuals obtained by fitting a linear regression model on the whole sample. The plot below shows the LOCAL2 estimates of the simulation example.



As a compromis between less complexity in computation and high accuracy the LOCAL2 estimator shows a similar behavior as the LOCAL1 estimator but less pronounced. Furthermore, note that the global residuals are contaminated by the changes which carries over to the variance estimator so that it tends to overestimate the error variance. For this reason we would have expected that the MOSUM procedures using the LOCAL2 estimator does not perform as quite as well as the LOCAL1 alternative. However, this does not hold for the MOSUM score-type procedure where we are slightly more successful by applying the LOCAL2 variance estimator. One possible explanation for this is that the MOSUM score-type statistic is based on the weighted global residuals so that the LOCAL2 estimator, which rests upon the global residuals as well, is more suitable to mimic the variation in the statistic.

4.1.5. Comparison of the Run Time

Although both statistics, the MOSUM Wald-type and score-type, can be computed in linear time, the difference of the run time of the two procedures could be quite large. In order to investigate this computational aspect we calculated the two statistics for several sample sizes $n = 250, 500, 1000, 2000, 4000, 8000, 16000, 32000$ with bandwidth $G = G(n) = n^{2/3}$ so that $G(1000) = 100$. For the MOSUM Wald-type statistic the LOCAL1 variance estimator has been used whereas the LOCAL2 variance estimator has been employed in the score-procedure. The results are summarized in Figure 4.1 and Figure 4.2 shows the performance of the score-type statistic (in seconds) separately. We can see that the score-type procedure performs much better in terms of computation times which is not surprising as the procedure only requires the calculation of one global estimate. For instance, for a sample size of $n = 32000$ the run time of the Wald-type statistic is about 38.78 seconds which is still acceptable. However, in comparison to that the calculation of the MOSUM score-type statistic only takes 1.18 seconds on average for $n = 32000$.

Figure 4.1.: The graph shows the average computation time of the MOSUM Wald-type statistic and the score-type statistic for 100 replications.

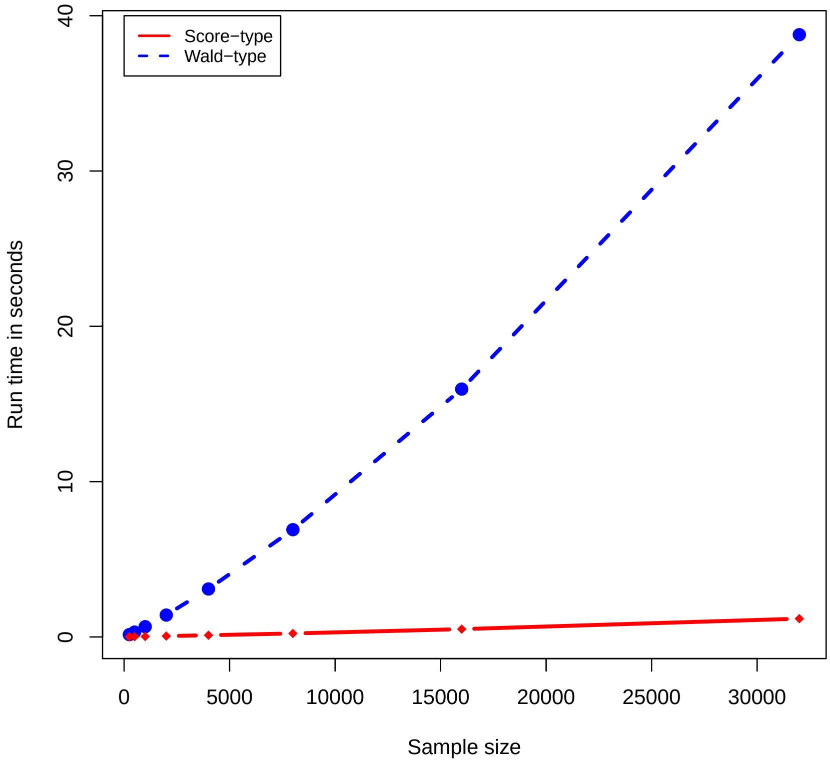
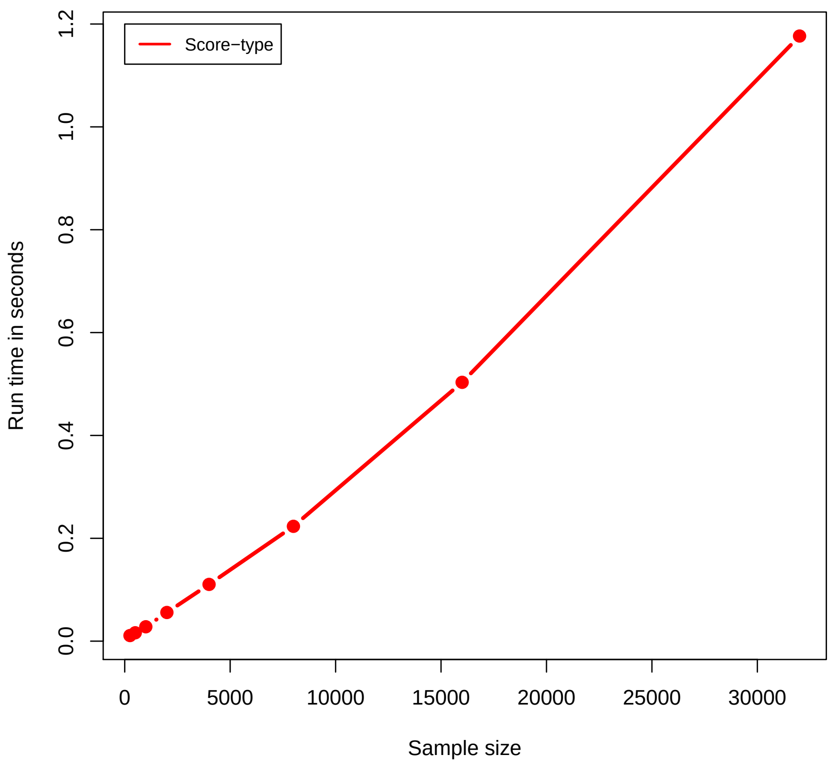


Figure 4.2.: The graph shows the average computation time of the MOSUM score-type statistic for 100 replications.



4.2. Example: Poisson Autoregressive Model

In this section, we consider a Poisson autoregressive model of order one, the INARCH(1) model. This is a very popular example of integer-valued time series and can be applied to different data sets in practise as it is able to describe overdispersion. For instance, Weiß (2010) applied the INARCH(1) model to monthly strike data of the U.S. labor market and Zhu & Wang (2010) used it to model daily download counts of a program. In general, INARCH models are a specific class of INGARCH time series which were introduced by Heinen (2003) as an analog of continuous GARCH models for count data. Since then INGARCH models have been examined by several authors, e.g. Ferland *et al.* (2006), Fokianos *et al.* (2009) and Weiß (2009).
More recently, these time series of counts have received more attention in the context of change point detection. Franke *et al.* (2012) used CUSUM-type statistics based on conditional least squares residuals for detecting structural breaks in a general class of Poisson autoregressive models of order one. Related to that Kirch & Tadjuidje Kamgaing (2016) considered a CUSUM score-type test statistic based on the least squares approach and derived consistency for the corresponding change point test and estimator in a quite general setting. Furthermore, Doukhan & Kengne (2013) proposed several Wald-type test statistics and investigated the behavior of the corresponding tests under the null hypothesis and alternatives of multiple changes.

4.2.1. The Model and the Statistics

The time series $\{Y_i\}_{i\geq 0}$ follows an INARCH(1) model if the observation Y_i conditioned on the past is Poisson distributed with parameter $\lambda_i = \theta_1 + \theta_2 Y_{i-1}$:

$$Y_i|\mathcal{F}_{i-1} \sim P(\lambda_i), \quad \text{with } \lambda_i = \theta_1 + \theta_2 Y_{i-1}. \tag{4.1}$$

According to Fokianos *et al.* (2009) we assume that Y_0 is fixed and we set it to zero in the simulations. In addition, similar to Franke *et al.* (2012) we constrain the parameters to $\delta \leq \theta_1 \leq \Delta$ and $0 \leq \theta_2 \leq 1-\delta$, for some small $0 < \delta < 1$ and some large $\Delta < \infty$, in order to get a compact parameter space $\boldsymbol{\Theta}$. According to Ferland *et al.* (2006) all moments of the Poisson autoregressive series exist. Furthermore, by Neumann (2011) we know that there exists a stationary ergodic solution of (4.1) which is β-mixing with exponential rate if the autocorrelation coefficient $\theta_2 < 1$. This implies that $\{Y_i\}$ is stationary and strongly mixing (α-mixing) with exponential rate (see e.g. Bradley (2007)). Since the mixing property of a sequence is preserved by measurable transformations we can conclude that the series $\{\mathbb{Y}_i\}$ with $\mathbb{Y}_i = (Y_{i-1}, Y_i)^T$ is of type (E2).
Under the alternative with q structural breaks we assume that there are $q+1$ INARCH(1) time series $\{Y_i^{(j)}\}$, $j = 1, \ldots, q+1$, such that

$$Y_i = \begin{cases} Y_i^{(1)}, & \text{if } \; i \leq k_{1,n} \\ Y_i^{(2)}, & \text{if } \; k_{1,n} < i \leq k_{2,n} \\ \quad\vdots & \\ Y_i^{(q+1)}, & \text{if } \; i > k_{q,n} \end{cases}, \tag{4.2}$$

with corresponding parameter vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{q+1} \in \boldsymbol{\Theta}$ where $\boldsymbol{\theta}_j \neq \boldsymbol{\theta}_{j+1}$ for all $j = 1, \ldots, q$. With the same arguments as before we get that the sequences $\{\mathbb{Y}_i^{(j)}\}$, $j = 1, \ldots, q+1$, are of type (E2). Furthermore, we use the notation $\boldsymbol{Y}_{i-1} = (1, Y_{i-1})^T$.
In the following, we consider the MOSUM Wald-type and score-type statistics based on the conditional least squares approach and the conditional maximum likelihood approach. The least squares estimator $\widehat{\boldsymbol{\theta}}_{1,n}^{LS}$ is the solution of the estimating equation system $\sum_{i=1}^{n} \boldsymbol{Y}_{i-1}\left(Y_i - \boldsymbol{Y}_{i-1}^T \boldsymbol{\theta}\right) = \mathbf{0}$ and has the explicit expression (see Weiß (2010) on page 1278)

$$\widehat{\boldsymbol{\theta}}_{1,n}^{LS} = \begin{pmatrix} (1-\hat{\rho}_{1,n})\bar{X}_{1,n} \\ \hat{\rho}_{1,n} \end{pmatrix}, \tag{4.3}$$

where $\hat{\rho}_{1,n}$ denotes the lag 1 sample autocorrelation and $\bar{X}_{1,n}$ represents the sample mean. In contrast, the (conditional) maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{1,n}^{Ml}$ has to be determined numerically by solving the estimating equation system $\sum_{i=1}^{n} \boldsymbol{Y}_{i-1}\left(\frac{Y_i - \boldsymbol{Y}_{i-1}^T\boldsymbol{\theta}}{\boldsymbol{Y}_{i-1}^T\boldsymbol{\theta}}\right) = \mathbf{0}$. The estimating functions of the two approaches and their derivatives are shown in the following table.

	Least squares approach	Likelihood approach
Estimating function $\boldsymbol{H}(\mathbb{Y}_i, \boldsymbol{\theta})$	$-2\boldsymbol{Y}_{i-1}\left(Y_i - \boldsymbol{Y}_{i-1}^T\boldsymbol{\theta}\right)$	$-2\boldsymbol{Y}_{i-1}\left(\frac{Y_i - \boldsymbol{Y}_{i-1}^T\boldsymbol{\theta}}{\boldsymbol{Y}_{i-1}^T\boldsymbol{\theta}}\right)$
First derivatives $\nabla\boldsymbol{H}(\mathbb{Y}_i, \boldsymbol{\theta})$	$2\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}^T$	$2\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}^T\frac{Y_i}{(\boldsymbol{Y}_{i-1}^T\boldsymbol{\theta})^2}$
Second derivatives		
$\nabla^2 H_1(\mathbb{Y}_i, \boldsymbol{\theta})$	$\mathbf{0}$	$-4\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}^T\frac{Y_i}{(\boldsymbol{Y}_{i-1}^T\boldsymbol{\theta})^3}$
$\nabla^2 H_2(\mathbb{Y}_i, \boldsymbol{\theta})$	$\mathbf{0}$	$-4\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}^T\frac{Y_iY_{i-1}}{(\boldsymbol{Y}_{i-1}^T\boldsymbol{\theta})^3}$

The MOSUM score-type statistic and the MOSUM Wald-type statistic based on the least squares approach are given by

$$\begin{aligned}&\widehat{T}_{k,n}^{LS}\left(G, \widehat{\boldsymbol{\theta}}_{1,n}^{LS}\right) \qquad (4.4)\\ &= \frac{\sqrt{2}}{\sqrt{G}}\left\|\widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\left(\sum_{i=k+1}^{k+G}\boldsymbol{Y}_{i-1}\left(Y_i - \boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}_{1,n}^{LS}\right) - \sum_{i=k-G+1}^{k}\boldsymbol{Y}_{i-1}\left(Y_i - \boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}_{1,n}^{LS}\right)\right)\right\|\end{aligned}$$

and

$$\widehat{W}_{k,n}^{LS}(G) = \frac{\sqrt{G}}{\sqrt{2}}\left\|\widehat{\boldsymbol{\Gamma}}_{k,n}^{-1/2}\left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}^{LS} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}^{LS}\right)\right\|. \tag{4.5}$$

Furthermore, the likelihood based Wald-type and score-type statistic are defined as follows

$$\begin{aligned}&\widehat{T}_{k,n}^{ML}\left(G, \widehat{\boldsymbol{\theta}}_{1,n}^{ML}\right) \qquad (4.6)\\ &= \frac{\sqrt{2}}{\sqrt{G}}\left\|\widetilde{\boldsymbol{\Sigma}}_{k,n}^{-1/2}\left(\sum_{i=k+1}^{k+G}\boldsymbol{Y}_{i-1}\left(\frac{Y_i - \boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}_{1,n}^{ML}}{\boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}_{1,n}^{ML}}\right) - \sum_{i=k-G+1}^{k}\boldsymbol{Y}_{i-1}\left(\frac{Y_i - \boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}_{1,n}^{ML}}{\boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}_{1,n}^{ML}}\right)\right)\right\|\end{aligned}$$

and

$$\widehat{W}_{k,n}^{ML}(G) = \frac{\sqrt{G}}{\sqrt{2}} \left\| \widetilde{\mathbf{\Gamma}}_{k,n}^{-1/2} \left(\widehat{\boldsymbol{\theta}}_{k+1,k+G}^{ML} - \widehat{\boldsymbol{\theta}}_{k-G+1,k}^{ML} \right) \right\|. \tag{4.7}$$

4.2.2. Proving the Assumptions of the MOSUM Procedures

As the series $\{\mathbb{Y}_i\}$ is of type (E2) as mentioned in the previous subsection we only have to show that the moment conditions in Section 2.3 and 3.1, which are summarized in Appendix B, are satisfied. We focus on the assumptions under the alternative as the assumptions under the null hypothesis are similar to that with stationary sequence $\{Y_i\}$ instead of $\{Y_i^{(j)}\}$.

Least Squares Approach:
At first, note that the matrix $E\left(\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)$ is regular since

$$\det\left(E\left(\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)\right) = \det\left(E\begin{pmatrix} 1 & Y_{i-1}^{(j)} \\ Y_{i-1}^{(j)} & Y_{i-1}^{(j)2} \end{pmatrix}\right) \tag{4.8}$$
$$= \left(E\left(Y_{i-1}^{(j)2}\right) - E\left(Y_{i-1}^{(j)}\right)^2\right) = \operatorname{Var}(Y_{i-1}^{(j)}) > 0.$$

Moreover, applying the computation rules for conditional expectations in combination with (4.1) and (4.2) yields

$$E\left(-\frac{1}{2}\boldsymbol{H}(\mathbb{Y}_i^{(j)},\boldsymbol{\theta})\right) = E\left(\boldsymbol{Y}_{i-1}^{(j)}\left(Y_i^{(j)} - \boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}\right)\right) \tag{4.9}$$
$$= E\left(E\left(\boldsymbol{Y}_{i-1}^{(j)}\left(Y_i^{(j)} - \boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}\right)|\mathcal{F}_{i-1}\right)\right) = E\left(\boldsymbol{Y}_{i-1}^{(j)}\left(E\left(Y_i^{(j)}|\mathcal{F}_{i-1}\right) - \boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}\right)\right)$$
$$= E\left(\boldsymbol{Y}_{i-1}^{(j)}\left(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j - \boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}\right)\right) = E\left(\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)(\boldsymbol{\theta}_j - \boldsymbol{\theta}),$$

which directly shows that $\boldsymbol{\theta}_j$ is the unique zero of $E\left(\boldsymbol{H}(\mathbb{Y}_i^{(j)},\boldsymbol{\theta})\right)$ as the matrix $E\left(\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)$ is regular by (4.8). Furthermore, we get

$$E\left(-\frac{1}{2}\boldsymbol{H}(\mathbb{Y}_i^{(j)},\boldsymbol{\theta})\right) = E\left(\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{\theta}_j - \boldsymbol{\theta})\right) = E\left(\begin{pmatrix} 1 & Y_{i-1}^{(j)} \\ Y_{i-1}^{(j)} & Y_{i-1}^{(j)2} \end{pmatrix}(\boldsymbol{\theta}_j - \boldsymbol{\theta})\right).$$

Hence, since the parameter space $\boldsymbol{\Theta}$ is compact and as all moments of $Y_{i-1}^{(j)}$ exist by Proposition 6 of Ferland *et al.* (2006), we know that all moments of the components of $\boldsymbol{H}(\mathbb{Y}_i^{(j)},\boldsymbol{\theta})$ exist. Thus, with the help of Lemma E.1.6 (b) we can conclude that $E\left(\left\|\boldsymbol{H}(\mathbb{Y}_1^{(j)},\boldsymbol{\theta})\right\|\right) < \infty$ holds for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ showing Assumption B.2.1 and $E\left(\left\|\boldsymbol{H}(\mathbb{Y}_1^{(j)},\boldsymbol{\theta})\right\|^{2+\nu}\right) < \infty$ holds for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, for some $\nu > 0$, which gives Assumption B.2.8.
Moreover, as

$$\nabla\boldsymbol{H}\left(\mathbb{Y}_i^{(j)},\boldsymbol{\theta}\right) = 2\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T} = 2\begin{pmatrix} 1 & Y_{i-1}^{(j)} \\ Y_{i-1}^{(j)} & Y_{i-1}^{(j)2} \end{pmatrix}$$

holds for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the Assumptions B.2.3 and B.2.6 (B.2.11) follow directly from Proposition 6 of Ferland *et al.* (2006) and Lemma E.1.6 (b) again. To prove Assumption B.2.10, we consider the convex combination

$$\gamma E\left(2\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right) + (1-\gamma)E\left(2\boldsymbol{Y}_{i-1}^{(j+1)}\boldsymbol{Y}_{i-1}^{(j+1)T}\right),$$

which is regular, for all $\gamma \in [0,1]$, as a straightforward calculation shows

$$\begin{aligned} &\det\left(\gamma E\left(2\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right) + (1-\gamma)E\left(2\boldsymbol{Y}_{i-1}^{(j+1)}\boldsymbol{Y}_{i-1}^{(j+1)T}\right)\right) \qquad (4.10)\\ &= 4\gamma \mathrm{Var}\left(Y_{i-1}^{(j)}\right) + 4(1-\gamma)\mathrm{Var}\left(Y_{i-1}^{(j+1)}\right) + 4\gamma(1-\gamma)\left(E\left(Y_{i-1}^{(j)}\right) - E\left(Y_{i-1}^{(j+1)}\right)\right)^2 \\ &\geq 4\gamma \mathrm{Var}\left(Y_{i-1}^{(j)}\right) + 4(1-\gamma)\mathrm{Var}\left(Y_{i-1}^{(j+1)}\right) \geq 4\min\left\{\mathrm{Var}\left(Y_{i-1}^{(j)}\right), \mathrm{Var}\left(Y_{i-1}^{(j+1)}\right)\right\} > 0, \end{aligned}$$

uniformly in $\gamma \in [0,1]$. Moreover, since all moments of $Y_{i-1}^{(j)}$ and $Y_{i-1}^{(j+1)}$ exist we obtain

$$\begin{aligned} &\sup_{\gamma\in[0,1]} \left\|\gamma E\left(2\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right) + (1-\gamma)E\left(2\boldsymbol{Y}_{i-1}^{(j+1)}\boldsymbol{Y}_{i-1}^{(j+1)T}\right)\right\|_F \qquad (4.11)\\ &\leq \left\|E\left(2\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)\right\|_F + \left\|E\left(2\boldsymbol{Y}_{i-1}^{(j+1)}\boldsymbol{Y}_{i-1}^{(j+1)T}\right)\right\|_F < \infty. \end{aligned}$$

Hence, by Lemma E.1.11 combined with (4.10) and (4.11) we can conclude that

$$\sup_{\gamma\in[0,1]} \left\|\left(\gamma E\left(2\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right) + (1-\gamma)E\left(2\boldsymbol{Y}_{i-1}^{(j+1)}\boldsymbol{Y}_{i-1}^{(j+1)T}\right)\right)^{-1}\right\|_F < \infty,$$

which shows Assumption B.2.10. Furthermore, note that the Assumptions B.2.4 and B.2.7 are satisfied because the second derivatives of the estimating function are equal to zero.

Likelihood Approach:

At first, we want to show that the matrix $E\left(\frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)$ is regular. By computing the determinant we obtain

$$\begin{aligned} &\det\left(E\left(\frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)\right) = \det\left(E\begin{pmatrix} \frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}} & \frac{Y_{i-1}^{(j)}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}} \\ \frac{Y_{i-1}^{(j)}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}} & \frac{Y_{i-1}^{(j)2}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}} \end{pmatrix}\right) \qquad (4.12)\\ &= E\left(\frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\right)E\left(\frac{Y_{i-1}^{(j)2}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\right) - E\left(\frac{Y_{i-1}^{(j)}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\right)^2 \\ &= \mathrm{Var}\left(\frac{Y_{i-1}^{(j)}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\right) - \mathrm{Cov}\left(\frac{Y_{i-1}^{(j)2}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}, \frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\right) \geq \mathrm{Var}\left(\frac{Y_{i-1}^{(j)}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\right) > 0, \end{aligned}$$

where the last line follows from $\mathrm{Cov}\left(\frac{Y_{i-1}^{(j)2}}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}, \frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\right) \leq 0$. Furthermore, by using similar arguments as in (4.9) we obtain

$$E\left(-\frac{1}{2}\boldsymbol{H}(\mathbb{Y}_i^{(j)}, \boldsymbol{\theta})\right) = E\left(\frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\right)(\boldsymbol{\theta}_j - \boldsymbol{\theta}),$$

showing that $\boldsymbol{\theta}_j$ is the unique zero of $E\left(\boldsymbol{H}(\mathbb{Y}_i^{(j)}, \boldsymbol{\theta})\right)$ by (4.12). Besides, as the parameter space $\boldsymbol{\Theta} = [\delta, \Delta] \times [0, 1-\delta]$ we get $\frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}} \leq \frac{1}{\delta}$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Hence, the results from the least squares section can be used as follows

$$E\left(\left\|\frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{\theta}_j - \boldsymbol{\theta})\right\|\right) \leq \frac{1}{\delta}E\left(\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{\theta}_j - \boldsymbol{\theta})\right\|\right) < \infty$$

and

$$E\left(\left\|\frac{1}{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}}\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{\theta}_j - \boldsymbol{\theta})\right\|^{2+\nu}\right) \leq \frac{1}{\delta^{2+\nu}}E\left(\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{\theta}_j - \boldsymbol{\theta})\right\|^{2+\nu}\right) < \infty,$$

proving the Assumptions B.2.1 and B.2.8. Furthermore, by applying the computation rules for the conditional expectation similar to (4.9), we receive

$$E\left(\nabla\boldsymbol{H}\left(\mathbb{Y}_i^{(j)}, \boldsymbol{\theta}\right)\right) = E\left(2\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\frac{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j}{(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta})^2}\right).$$

On noting that $E\left(\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j)\right\|_F\right) < \infty$ and $E\left(\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j)\right\|_F^{2+\nu}\right) < \infty$ holds since all moments of $Y_{i-1}^{(j)}$ exist by Proposition 6 of Ferland *et al.* (2006), we can conclude that

$$E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\frac{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j}{(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta})^2}\right\|_F\right) \leq \frac{1}{\delta^2}E\left(\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j)\right\|_F\right) < \infty$$

and

$$E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}\frac{\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j}{(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta})^2}\right\|_F^{2+\nu}\right) \leq \frac{1}{\delta^{4+2\nu}}E\left(\left\|\boldsymbol{Y}_{i-1}^{(j)}\boldsymbol{Y}_{i-1}^{(j)T}(\boldsymbol{Y}_{i-1}^{(j)T}\boldsymbol{\theta}_j)\right\|_F^{2+\nu}\right) < \infty,$$

which show the Assumptions B.2.3 and B.2.6 or B.2.11. The moment conditions B.2.4 and B.2.7 can be derived by using similar arguments. In contrast, proving Assumption B.2.10 is more complicated and needs to be examined in detail. However, this would go beyond the scope of this thesis and will be part of future work.

4.2.3. Estimating the Covariance Matrices

In order to apply the MOSUM score-type procedure we have to find an appropriate estimator of the long-run covariance matrix of $\boldsymbol{H}(\mathbb{Y}_i, \widetilde{\boldsymbol{\theta}})$. A first idea is to ignore the dependency and to use a MOSUM-version of the empirical covariance matrix estimator of $\boldsymbol{H}(\mathbb{Y}_i, \widehat{\boldsymbol{\theta}}_{1,n})$ as follows:

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_{k,n} = \frac{1}{2G}\Bigg(&\sum_{i=k-G+1}^{k}\left(\boldsymbol{H}(\mathbb{Y}_i, \widehat{\boldsymbol{\theta}}_{1,n}) - \overline{\boldsymbol{H}}_{k-G+1,k}\right)\left(\boldsymbol{H}(\mathbb{Y}_i, \widehat{\boldsymbol{\theta}}_{1,n}) - \overline{\boldsymbol{H}}_{k-G+1,k}\right)^T \\ &+ \sum_{i=k+1}^{k+G}\left(\boldsymbol{H}(\mathbb{Y}_i, \widehat{\boldsymbol{\theta}}_{1,n}) - \overline{\boldsymbol{H}}_{k+1,k+G}\right)\left(\boldsymbol{H}(\mathbb{Y}_i, \widehat{\boldsymbol{\theta}}_{1,n}) - \overline{\boldsymbol{H}}_{k+1,k+G}\right)^T\Bigg),\end{aligned} \tag{4.13}$$

where $\overline{\boldsymbol{H}}_{l,u}$ denotes the sample mean of $\boldsymbol{H}(\mathbb{Y}_l, \widehat{\boldsymbol{\theta}}_{1,n}), \ldots, \boldsymbol{H}(\mathbb{Y}_u, \widehat{\boldsymbol{\theta}}_{1,n})$. We applied this type of estimators in the least squares and the likelihood based procedures. However, for future simulation studies it might be reasonable to consider more complex estimators created for estimating long-run covariance matrices as well.
For the Wald-type procedures, we need to find suitable estimators for the log-run covariance matrices of $\widehat{\boldsymbol{\theta}}^{LS}_{k-G+1,k}$ and $\widehat{\boldsymbol{\theta}}^{ML}_{k-G+1,k}$.
The asymptotic covariance matrix of the least squares estimator $\widehat{\boldsymbol{\theta}}^{LS}_{1,n}$ is under the null hypothesis given ($\boldsymbol{\theta} = (\theta_1, \theta_2)^T$) by:

$$\begin{pmatrix} \frac{\theta_1}{1-\theta_2}\left(\theta_1 + \theta_1\theta_2 + \frac{1+2\theta_2^4}{1+\theta_2+\theta_2^2}\right) & -\theta_1 - \theta_1\theta_2 - \frac{\theta_2^3+2\theta_2^4}{1+\theta_2+\theta_2^2} \\ -\theta_1 - \theta_1\theta_2 - \frac{\theta_2^3+2\theta_2^4}{1+\theta_2+\theta_2^2} & (1-\theta_2^2)\left(1 + \frac{\theta_2+2\theta_2^3}{\theta_1(1+\theta_2+\theta_2^2)}\right) \end{pmatrix},$$

which can be found in Weiß (2010) on page 1278. Hence, by replacing the parameter vector $\boldsymbol{\theta}$ by the local estimators $\widehat{\boldsymbol{\theta}}^{LS}_{k-G+1,k} = (\widehat{\theta}_{1,k}, \widehat{\theta}_{2,k})^T$ we get the following MOSUM-type estimator:

$$\widehat{\boldsymbol{\Gamma}}_{k,n} = \begin{pmatrix} \frac{\widehat{\theta}_{1,k}}{1-\widehat{\theta}_{2,k}}\left(\widehat{\theta}_{1,k} + \widehat{\theta}_{1,k}\widehat{\theta}_{2,k} + \frac{1+2\widehat{\theta}_{2,k}^4}{1+\widehat{\theta}_{2,k}+\widehat{\theta}_{2,k}^2}\right) & -\widehat{\theta}_{1,k} - \widehat{\theta}_{1,k}\widehat{\theta}_{2,k} - \frac{\widehat{\theta}_{2,k}^3+2\widehat{\theta}_{2,k}^4}{1+\widehat{\theta}_{2,k}+\widehat{\theta}_{2,k}^2} \\ -\widehat{\theta}_{1,k} - \widehat{\theta}_{1,k}\widehat{\theta}_{2,k} - \frac{\widehat{\theta}_{2,k}^3+2\widehat{\theta}_{2,k}^4}{1+\widehat{\theta}_{2,k}+\widehat{\theta}_{2,k}^2} & (1-\widehat{\theta}_{2,k}^2)\left(1 + \frac{\widehat{\theta}_{2,k}+2\widehat{\theta}_{2,k}^3}{\widehat{\theta}_{1,k}(1+\widehat{\theta}_{2,k}+\widehat{\theta}_{2,k}^2)}\right) \end{pmatrix}, \tag{4.14}$$

which has been used in the simulations.
As described in Weiß (2010) on page 1277, under the null hypothesis, the asymptotic covariance matrix of the likelihood estimator $\widehat{\boldsymbol{\theta}}^{ML}_{1,n}$ coincides with the inverse of the expected Fisher information matrix which does not have an explicit expression. However, it can be approximated by the observed Fisher information matrix given by:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{(\boldsymbol{Y}_{i-1}^T\boldsymbol{\theta})^2} \begin{pmatrix} Y_i & Y_iY_{i-1} \\ Y_iY_{i-1} & Y_iY_{i-1}^2 \end{pmatrix}.$$

Thus, we propose to estimate the Fisher information matrix by applying the following MOSUM-type estimator which combines the mean of the estimated observed Fisher Information on $k-G+1, ..., k$ and on $k+1, ...k+G$:

$$\widetilde{\boldsymbol{\Gamma}}^{-1}_{k,n} = \frac{1}{2G}\left(\sum_{i=k-G+1}^{k} \frac{1}{(\boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}^{ML}_{k-G+1,k})^2} \begin{pmatrix} Y_i & Y_iY_{i-1} \\ Y_iY_{i-1} & Y_iY_{i-1}^2 \end{pmatrix} + \sum_{i=k+1}^{k+G} \frac{1}{(\boldsymbol{Y}_{i-1}^T\widehat{\boldsymbol{\theta}}^{ML}_{k+1,k+G})^2} \begin{pmatrix} Y_i & Y_iY_{i-1} \\ Y_iY_{i-1} & Y_iY_{i-1}^2 \end{pmatrix}\right).$$

This estimator has been applied in the simulations.

4.2.4. Simulating the Data

Under the null hypothesis we simulate 1000 samples of size $n = 1000$ from the following model:

$$Y_i|\mathcal{F}_{i-1} \sim P(\lambda_i), \quad \text{with } \lambda_i = 1 + 0.5Y_{i-1}, i = 1, \ldots, n.$$

Under the alternative, we let three changes occur. At time point 250 the parameter vector changes from $\boldsymbol{\theta}_1 = (1, 0.5)^T$ to $\boldsymbol{\theta}_2 = (2.5, 0.5)^T$, at 500 to $\boldsymbol{\theta}_3 = (2.5, 0.2)^T$ and at change point 750 the paramter vector goes to $\boldsymbol{\theta}_4 = (1, 0.5)^T$.

4.2.5. Results of the Simulations

On the simulated data, we test the null hypothesis of having no change at significance level $\alpha = 0.05$ by using MOSUM score-type and Wald-type statistics based on the least squares approach. The empirical size under the null hypothesis and the size adjusted power under the alternative are given in Table 4.5. For both tests, we observe that the empirical size gets smaller and the size adjusted power grows with increasing bandwidth. Whereas the score-type test is very conservative with empirical size less than $\alpha = 0.05$ for all considered bandwidths, the MOSUM Wald-type test holds the level only for the largest bandwidth of 150. Under the alternative, the score-type test performs quite well with a size adjusted power close to one for bandwidths greater than 80. In comparison to that, the size adjusted power of the Wald-type test is smaller and lies between 0.8 and 0.9. Furthermore, the results for score-type and Wald-type test based on the likelihood approach are given in Table 4.6. Note that the likelihood version of the score-type test shows similar results as the score-type test considered above. On the contrary, the likelihood based Wald-type test performs much better than its least squares counterpart. But this is not surprising as the maximum likelihood estimator is used in the statistic which usually performs better than the least squares estimator. We can see that the empirical size is around 0.05 for bandwidth between 80 and 120 and that the test is a bit conservative for the largest bandwidth of 150. Furthermore, the size adjusted power is for all bandwidth close to one. Thus, we can conclude that the MOSUM Wald-type test based on the likelihood approach performs best among the considered tests.

Table 4.5.: Simulation results for the tests based on the least squares approach

	$\boldsymbol{H_0}$ Empirical size	$\boldsymbol{H_1}$ Size adjusted power
Score-type (least squares)		
$G = 50$	0.017	0.554
$G = 80$	0.010	0.879
$G = 100$	0.010	0.949
$G = 120$	0.002	0.981
$G = 150$	0.006	1
Wald-type (least squares)		
$G = 50$	0.292	0.771
$G = 80$	0.125	0.835
$G = 100$	0.090	0.843
$G = 120$	0.060	0.854
$G = 150$	0.034	0.859

Table 4.6.: Simulation results for the tests based on the likelihood approach

	$\boldsymbol{H_0}$ Empirical size	$\boldsymbol{H_1}$ Size adjusted power
Score-type (least squares)		
$G = 50$	0.015	0.542
$G = 80$	0.010	0.876
$G = 100$	0.009	0.953
$G = 120$	0.003	0.983
$G = 150$	0.008	1
Wald-type (likelihood)		
$G = 50$	0.170	0.918
$G = 80$	0.067	0.992
$G = 100$	0.060	0.998
$G = 120$	0.032	1
$G = 150$	0.025	1

Now, we want to compare the performance of the different procedures in terms of estimation. Therefore, for each procedure, we recorded the estimated number and the estimated locations of the changes as in the simulations for the linear regression. Note that we use $\alpha = 0.2$ and $\epsilon = 0.1$ in the estimation process. The results of the MOSUM procedures based on the least squares approach are summarized in Table 4.7. In the MOSUM score-type and Wald-type procedure the performance of the estimator for the number of changes depends on the choice of the bandwidth. We can observe that the estimator performs better if the bandwidth gets larger which even holds for a bandwidth greater than the proposed length of half of the minimal distance between two adjacent structural breaks. For $G = 150$, the MOSUM score-type procedure correctly estimates the number of changes for 35.8% of the simulated samples whereas the Wald-type approach estimates the number by 3 in 58.2% of the cases. Thus, the estimator based on the score-type statistic performs worse than the Wald-type estimator. This coincides with our expectations as the Wald-type statistics are directly based on the local least squares estimators which can computed by an explicit formula without applying any numerical method. Furthermore, by considering the results for the estimated change point locations we can see that the score-type procedure performs quite well in detecting the first change and can compete with the Wald-type procedure for bandwidths greater than 80. It even shows better results than the Wald-type approach for localising the third change point. However, the score-type statistics are not really able to detect the second change since it estimates a change between 480 and 520 for only 33.2% of the simulated samples ($G = 150$). In contrast, the MOSUM Wald-type procedure provides an appropriate change point estimate for the second change in 80.7% of the cases. Nevertheless, the question arises if the detection of the second change by the score-type procedure can be improved by using another global estimator in the statistic. By replacing the estimator $\widehat{\boldsymbol{\theta}}_{1,1000}^{LS}$ by the least squares estimator $\widehat{\boldsymbol{\theta}}_{300,700}^{LS}$

calculated on the sample $Y_{300}, \ldots, Y_{700}$ we get the following results which are given in Table 4.8. Thus, we would prefer to use this score-type procedure as it performs better in detecting the second change point and shows similar results for the first and third change in comparison to the original one. Unfortunately, in practice we do not know what global estimators are more suitable to detect several changes. Hence, it would be reasonable to use a set of global estimators in the score-type procedure and to merge the results by a multiscale method as described in Chapter 5.

Table 4.7.: Simulation results for the MOSUM score-type and Wald-type statistic based on the least squares approach

	Estimated number					**Estimated change point in**		
	≤ 1	2	$q = 3$	4	≥ 5	$[230, 270]$	$[480, 520]$	$[730, 770]$
Score-type (least squares)								
$G = 50$	0.901	0.080	0.018	0.001	0.000	0.352	0.010	0.125
$G = 80$	0.627	0.228	0.073	0.012	0.000	0.734	0.036	0.326
$G = 100$	0.407	0.435	0.131	0.023	0.004	0.871	0.077	0.460
$G = 120$	0.223	0.486	0.237	0.043	0.011	0.931	0.161	0.588
$G = 150$	0.081	0.428	0.358	0.111	0.022	0.920	0.332	0.724
Wald-type (least squares)								
$G = 50$	0.390	0.374	0.175	0.051	0.010	0.891	0.429	0.044
$G = 80$	0.181	0.472	0.253	0.071	0.023	0.916	0.664	0.156
$G = 100$	0.089	0.438	0.338	0.098	0.037	0.916	0.731	0.292
$G = 120$	0.036	0.346	0.435	0.153	0.03	0.914	0.781	0.431
$G = 150$	0.008	0.184	0.582	0.188	0.038	0.894	0.807	0.635

Table 4.8.: Simulation results for the MOSUM score-type based on the least squares estimator $\widehat{\boldsymbol{\theta}}^{LS}_{300,700}$

	Estimated number					**Estimated change point in**		
	≤ 1	2	$q = 3$	4	≥ 5	$[230, 270]$	$[480, 520]$	$[730, 770]$
Score-type with estimator $\widehat{\boldsymbol{\theta}}^{LS}_{300,700}$								
$G = 50$	0.264	0.339	0.230	0.102	0.065	0.798	0.037	0.484
$G = 80$	0.100	0.397	0.300	0.143	0.060	0.932	0.121	0.684
$G = 100$	0.038	0.341	0.413	0.155	0.053	0.923	0.259	0.759
$G = 120$	0.011	0.249	0.490	0.204	0.046	0.932	0.408	0.772
$G = 150$	0.018	0.162	0.596	0.194	0.030	0.923	0.660	0.749

Furthermore, the results of the MOSUM procedures based on the likelihood approach are given in Table 4.9. The procedures perform better than their least squares based counterparts as the maximum likelihood estimator is used in the statistics. The MOSUM Wald-type statistic of likelihood approach shows the best performance in terms of estimating the number and the location of the changes. Besides, the score-type procedure with the global estimator $\widehat{\boldsymbol{\theta}}^{ML}_{300,700}$ shows for $G = 150$ similar results as the Wald-type approach but is more effected by the choice of the bandwidth and performs worse for smaller window length. However, this problem can be solved by implementing a multiscale method in order to appropriately combine results produced by MOSUM score-type procedures with different bandwidths and different global estimators. This could considerably improve the performance of the score-type approach so that it could compete with the Wald-type procedures.

Table 4.9.: Simulation results for the MOSUM score-type and Wald-type statistic based on the likelihood approach

	Estimated number					**Estimated change point in**		
	≤ 1	2	$q = 3$	4	≥ 5	$[230, 270]$	$[480, 520]$	$[730, 770]$
Score-type (likelihood) with estimator $\widehat{\boldsymbol{\theta}}^{ML}_{1,1000}$								
$G = 50$	0.903	0.077	0.019	0.001	0.000	0.312	0.028	0.097
$G = 80$	0.619	0.288	0.063	0.028	0.002	0.713	0.135	0.242
$G = 100$	0.385	0.379	0.183	0.045	0.008	0.857	0.267	0.367
$G = 120$	0.190	0.418	0.276	0.096	0.020	0.925	0.406	0.494
$G = 150$	0.056	0.321	0.449	0.137	0.037	0.921	0.583	0.623
Score-type (likelihood) with estimator $\widehat{\boldsymbol{\theta}}^{ML}_{300,700}$								
$G = 50$	0.264	0.339	0.230	0.102	0.065	0.839	0.061	0.545
$G = 80$	0.100	0.397	0.300	0.143	0.060	0.936	0.199	0.734
$G = 100$	0.038	0.341	0.413	0.155	0.053	0.929	0.389	0.780
$G = 120$	0.011	0.249	0.490	0.204	0.046	0.929	0.561	0.782
$G = 150$	0.018	0.162	0.596	0.194	0.030	0.919	0.724	0.742
Wald-type (likelihood)								
$G = 50$	0.330	0.320	0.225	0.090	0.035	0.789	0.352	0.429
$G = 80$	0.069	0.295	0.373	0.199	0.064	0.890	0.603	0.645
$G = 100$	0.017	0.205	0.415	0.259	0.104	0.908	0.702	0.714
$G = 120$	0.005	0.115	0.546	0.248	0.086	0.908	0.766	0.764
$G = 150$	0.001	0.040	0.629	0.261	0.069	0.896	0.809	0.803

4.2.6. Comparison of the Run Time

Similar to the considerations in the linear regression we want to compare the Wald-type and score-type procedures in terms of their actual computation times. Thus, we calculated the two statistics based on the least squares and the likelihood approach for several sample sizes $n = 250, 500, 1000, 2000, 4000, 8000, 16000, 32000$ with bandwidth

$G = G(n) = n^{2/3}$ and recorded the run times of 100 repetitions. Figure 4.3 shows the average run time of the MOSUM score-type and Wald-type statistic based on the least squares approach. We can see that the score-type statistic performs slightly better in terms of computation times. Since the procedure only requires the calculation of one global estimate we would have expected a larger difference in run time as in the linear regression example in Section 4.1.5. An explanation for this could be that we applied a MOSUM-version of the empirical covariance matrix $\widehat{\boldsymbol{\Sigma}}_{k,n}$ (given in (4.13)) as an estimator for the long-run covariance matrix of the estimating function which seems to slow down the computation in total. Nevertheless, both procedures perform quite well. For a sample size of $n = 32000$ the computation of the score-type statistic only takes 9.48 seconds on average followed by the Wald-type statistic with 10.94 seconds. Figure 4.4 gives the average run time of the two statistics based on the likelihood approach (in minutes) and Figure 4.5 shows the performance of the score-type statistic (in seconds) separately. The plots show that the score-type statistic clearly outperforms the Wald-type statistic in terms of computation time. For sample size $n = 1000$ ($n = 32000$) it only takes 0.71 (35.24) seconds to calculate the score-type statistic whereas the Wald-type statistic requires on average 3.23 (221.34) minutes of computation time. This illustrates a main disadvantage of the Wald-type procedure in contrast to the score-type approach. As numerical methods are needed to calculate the local estimates in the Wald-type statistics computation time increases dramatically in large data sets.

Figure 4.3.: The graph shows the average computation time of the MOSUM Wald-type statistic and the score-type statistic based on the least squares approach for 100 replications.

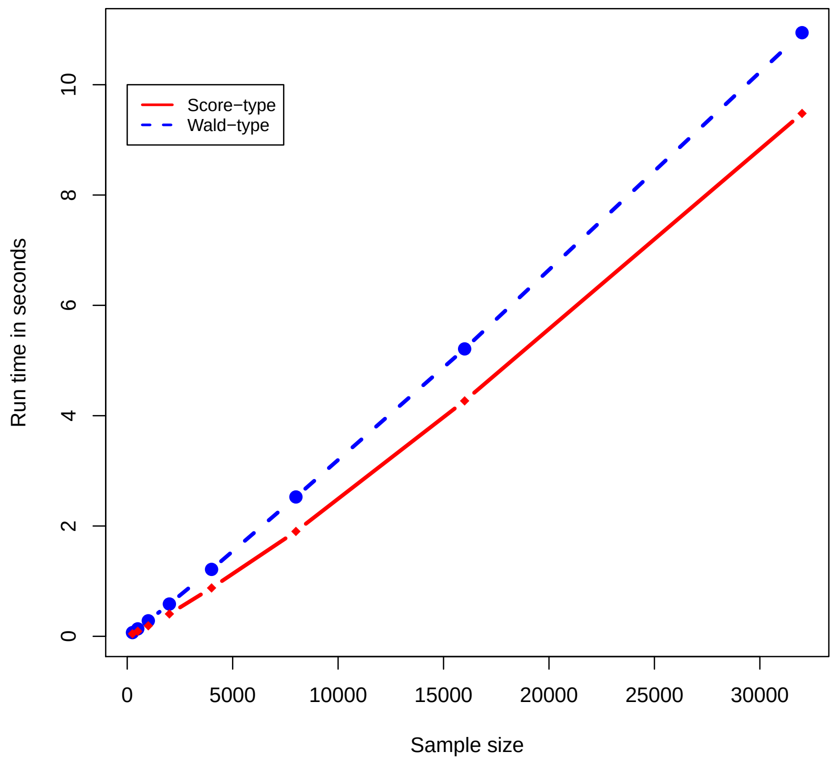


Figure 4.4.: The graph shows the average computation time of the MOSUM Wald-type statistic and the score-type statistic based on the likelihood approach (in minutes) for 100 replications.

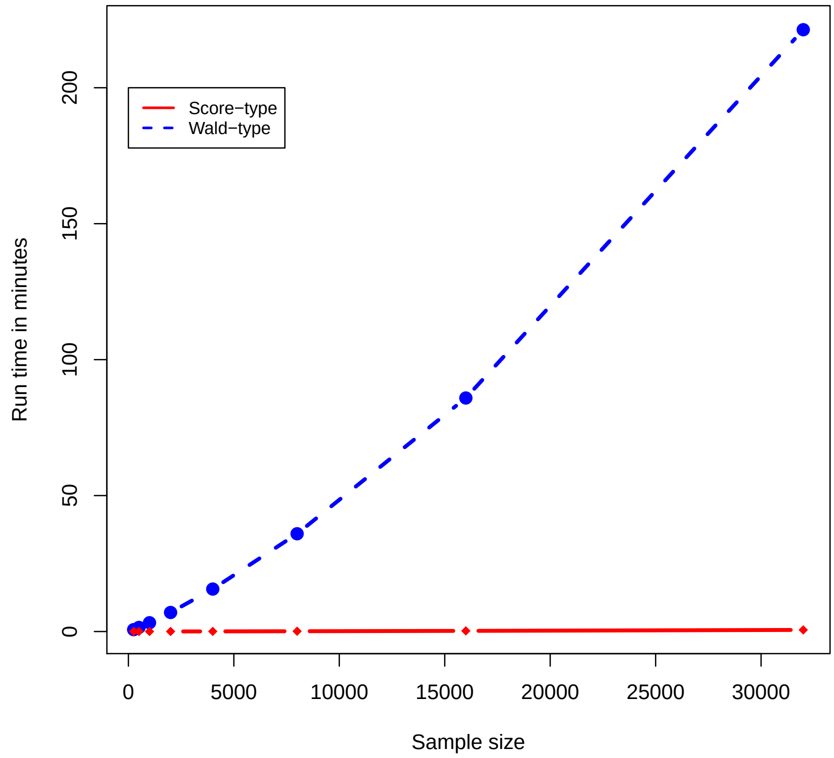
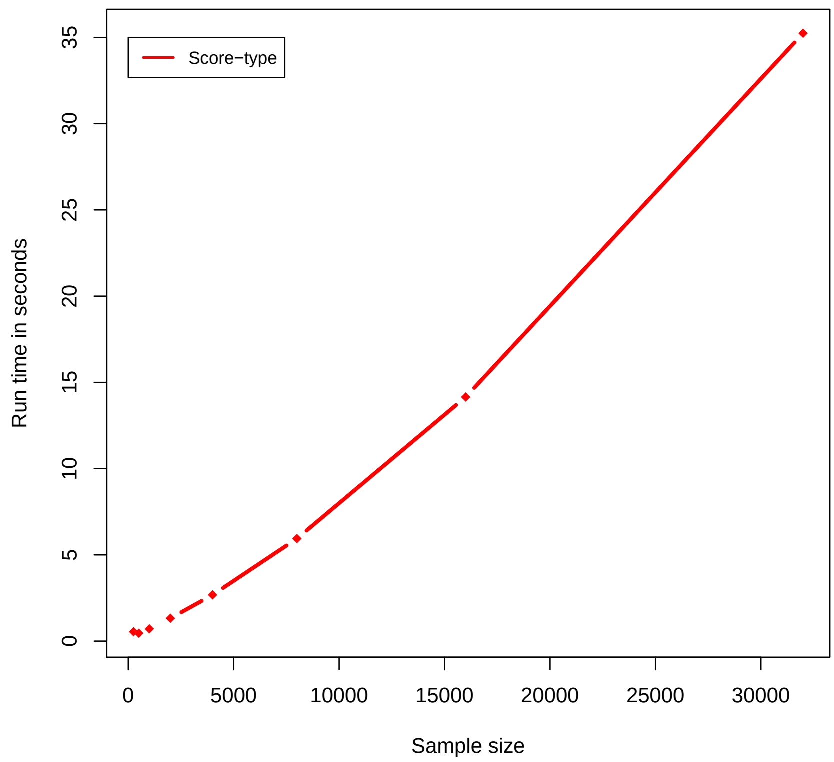


Figure 4.5.: The graph shows the average computation time of the MOSUM score-type statistic based on the likelihood approach (in seconds) for 100 replications.



5. The Multiscale MOSUM Procedure

5.1. Introduction

In this chapter we consider the multiscale MOSUM procedure with localised pruning introduced by Cho & Kirch (2018), which is decsribed in the following subsection, and we adapt it to the MOSUM score-type and MOSUM Wald-type procedure in order to solve the problem of choosing the bandwidth and the problem in detectability of the score-MOSUM. In Section 5.3 we explain how the method can be extended to the linear regression model as given in Section 3.2 and a general model which is similar to that in Section 3.1 but we do not restrict our considerations to i.i.d. (type (E1)) or stationary and strongly mixing (type (E2)) sequences. Before stating a first result on the output of the multiscale method, we derive some auxiliary lemmata which are crucial for that proof and for showing consistency of the estimators as future work.

5.2. The Multiscale MOSUM Procedure with Localized Pruning by Cho and Kirch

The choice of the bandwidth has a crucial impact on the performance of the MOSUM procedures, even in the classical mean change model. One possibility to solve this problem is to run the MOSUM procedure on a set of different bandwidths and to merge the results in an appropriate way. For instance, Messer *et al.* (2014), using MOSUM statistics for detecting changes in point processes, proposed a multiscale method which takes all the estimates obtained from the smallest bandwidth and adds recursively estimates produced by the next largest bandwidth but only if their G-environments do not cover any of the previous estimates. However, one main drawback of this procedure is that it might fail to eliminate some estimates produced by spurious local maxima. Furthermore, as already mentioned by Eichinger & Kirch (2018), it is questionable how to extend this method to asymmetric bandwidths where the length of the right window does not coincide with length of the left window. Such bandwidth constellations have been considered for the classical mean change model in order to locate small changes with one close and one distant neighboring change and it would be reasonable to incorporate this in our MOSUM procedures as well. Moreover, the multiscale procedure mentioned above cannot be used to merge results produced by MOSUM score-type statistics of different global estimators which is essential for solving the problem in detectability.

In comparison to that, Cho & Kirch (2018) adopted the idea of Fryzlewicz (2014) to use an information criterion to combine the estimates produced by the MOSUM procedure

with different window lengths. Starting with Yao (1988), who used Schwarz's criterion (also known as BIC – Bayesian information criterion) to estimate the number of changes in the mean under normality assumption, information criteria have been widely used in multiple change point detection. Just to mention a few contributions, Liu *et al.* (1997) applied a modified Schwarz criterion for localizing changes or different segments in a multivariate regression model. Pan & Chen (2006) considered a general parameter change problem of independent observations and used a modified information criterion. Moreover, Kühn (2001) extended the result of Yao (1988) to a non-parametric setting by using a weak invariance principle. Nevertheless, in all of these examples the change point estimates are obtained by an optimization of the considered criterion over all time points which can be computationally expensive. In contrast, Cho & Kirch (2018) apply a modified Schwarz criterion ($sBIC$) to choose among a set of candidates produced by different bandwidths and propose a localized pruning approach which is shortly described in the following.
We consider a set of bandwidths and for each of these window lengths we run the MOSUM procedure to get estimates for the change points. These estimates are stored together with the information about the bandwidth, the values of the corresponding MOSUM statistic and the p-values, derived from the limit distribution of the test statistic under the null, in order to create an initial candidate set for the multiscale procedure. In the next step, the candidates are ranked according to their p-values or jump sizes, which are computed by dividing the MOSUM statistics by the square root of the corresponding bandwidth. We start with the candidate $\widehat{k}^*$, which has the smallest p-value or the largest jump size, and determine its conflicting candidates in the set which are estimates possibly generated by the same structural break. Two estimates are in conflict if they lie in the computation interval of the other, where the comptutation interval of a candidate is the window on which the associated MOSUM statistic has been calculated. Then, we determine all the subsets of $\widehat{k}^*$ and its conflicting candidates satisfying the following conditions and add them to the set of the final candidate sets.

(C1) Adding further candidates to the set monotonically increases $sBIC$.

(C2) Removing any single candidate from the set increases $sBIC$.

After removing the estimates, which have already been considered, from the candidate set, we repeat the same steps until the initial candidate set is empty. Finally, the output consists of sets of candidates $\widehat{\mathcal{A}}$ satifying the conditions (C1) and (C2). Furthermore, the cardinality of the output is defined as the minimal cardinality among its candidate sets and is used as an estimator for the number of changes.

5.3. Adapting the Procedure

Cho & Kirch (2018) apply the following modified Bayesian information criterion

$$sBIC(\mathcal{A}_n) = \frac{n}{2}\log\left(\frac{RSS(\mathcal{A}_n)}{n}\right) + |\mathcal{A}_n|\xi_n,$$

where $\mathcal{A}_n = \{c_{1,n}, \ldots, c_{m,n}\}$ is a generic set of possible change points with cardinality $|\mathcal{A}_n| = m$. The choice of ξ_n, which penalizes the complexity of the model, depends in a sense on the assumptions of the model. Moreover, the sum of squared residuals of the set $\mathcal{A}_n$ is given by

$$RSS(\mathcal{A}_n) = \sum_{j=0}^{m} \sum_{i=c_{j,n}+1}^{c_{j+1,n}} (X_i - \bar{X}_{c_{j,n}+1,c_{j+1,n}})^2,$$

where $c_{0,n} = 0$ and $c_{m+1,n} = n$. Under normality assumptions on the random variables the RSS represents the log-likelihood function whereas in a non-parametric world we would call it pseudo-log-likelihood which has the following relation to the estimator. The sample mean computed on the subsample $X_l, \ldots, X_u$, denoted by $\bar{X}_{l,u}$, minimizes the function $\sum_{i=l}^{u}(X_i - \mu)^2$ and thus belongs to the class of M-estimators. This connection enables us to generalize the information criterion as follows.

At first, let us think about the classical M-estimator theory in the absence of change points. With $Q(\mathbb{X}_i, \boldsymbol{\theta})$ denoting the criterion function, we call the estimator sequence $\widehat{\boldsymbol{\theta}}_{1,n}$ an M-estimator for $\boldsymbol{\theta}_0 := \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} E\left(Q(\mathbb{X}_1, \boldsymbol{\theta})\right)$ if $\widehat{\boldsymbol{\theta}}_{1,n}$ minimizes $\frac{1}{n}\sum_{i=1}^{n} Q(\mathbb{X}_i, \boldsymbol{\theta})$ for every n. Moreover, we assume that the criterion function Q is positive and continuously differentiable on the compact parameter space $\boldsymbol{\Theta}$ such that the estimating function is the gradient vector of the criterion function with respect to $\boldsymbol{\theta}$: $\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) = \nabla Q(\mathbb{X}_i, \boldsymbol{\theta})$. Hence, minimizing $\frac{1}{n}\sum_{i=1}^{n} Q(\mathbb{X}_i, \boldsymbol{\theta})$ comes down to solving the estimating equation system $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) = \mathbf{0}$. Under the moment conditions

(I) $E\left(Q(\mathbb{X}_i, \boldsymbol{\theta})\right) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and

(II) $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\|\right) < \infty$,

the dominated convergence theorem yields $\nabla E(Q(\mathbb{X}_1, \boldsymbol{\theta})) = E\left(\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right)$ (see Bauer (2001) page 102 ff). This shows that these M-estimators belong to the class of Z-estimators considered in the previous sections where the unique zero of the expectation of the estimating function coincides with the unique minimizer $\boldsymbol{\theta}_0$ of the criterion function. Thus, the estimator $\widehat{\boldsymbol{\theta}}_{l,u}$ minimizing the function $\sum_{i=l}^{u} Q(\mathbb{X}_i, \boldsymbol{\theta})$ can be called M-estimator or Z-estimator on the subsample $\mathbb{X}_l, \ldots, \mathbb{X}_u$.

Under the alternative, we assume piecewise stationarity as in Assumption A.2.2 where the sequences $\{\mathbb{X}_i^{(j)} : j \geq 1\}$, $j = 1, \ldots, q+1$, additionally satisfy Conditions (I) and (II) such that the unique minimizer of $E\left(Q(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}\right)$, denoted by $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}$, is equal to the unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)$.

Consequently, we propose the following information criterion for the MOSUM score-type and the MOSUM Wald-type procedure:

$$sBIC(\mathcal{A}_n) = \frac{n}{2} \log\left(gRSS(\mathcal{A}_n)\right) + |\mathcal{A}_n|\xi_n, \tag{5.1}$$

where

$$gRSS(\mathcal{A}_n) = \frac{1}{n} \sum_{j=0}^{m} \sum_{i=c_{j,n}+1}^{c_{j+1,n}} Q(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_{c_{j,n}+1,c_{j+1,n}}) \tag{5.2}$$

and the penalty fufills

$$\xi_n = o(n). \tag{5.3}$$

Throughout this chapter, we consider the general parameter change model and the linear regression model which satisfy Assumptions A.2.1 and A.2.2 under the alternative. In the general model we assume that the estimating function $\boldsymbol{H}$ is twice continuously differentiable on $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta} \subset \mathbb{R}^p$ is a compact parameter space, and that $\boldsymbol{H}$ and its derivatives are measurable with respect to $\mathbb{X}_i$. Further conditions on that model will be introduced in the following subsection. In the linear regression model, where the parameter space is not compact, we consider Assumptions (R1*) to (R7*) given in Section 3.2.2 and we will introduce additional conditions the Section 5.4.2. According to the least-squares approach we get the following criterion function:

$$Q(Y_i, \boldsymbol{X}_i, \boldsymbol{\beta}) = \left(Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}\right)^2,$$

such that $gRSS$ of an arbitrary set is given by:

$$gRSS(\mathcal{A}_n) = \frac{1}{n} \sum_{j=0}^{m} \sum_{i=c_{j,n}+1}^{c_{j+1,n}} \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{c_{j,n}+1, c_{j+1,n}}\right)^2. \tag{5.4}$$

5.4. Theoretical Foundation

In this section, we derive some auxiliary results helping us to prove a first result for the multiscale method. We first consider the general model before we show similar results for the linear regression model.
To simplify the notation in the proofs we shorten the expressions of uniformity if it is clear what scenarios are considered. For instance, in Lemma 5.4.11 we want to show that $\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\theta}_j \right\| = O_P\left(\max\left\{\frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n}\right\}\right)$ holds. Hence, within the proof of that lemma we only write that a derived statement holds 'uniformly in g and h' instead of 'uniformly for all $v_n \le h < \delta_{j,n}$ and $1 \le g \le \tilde{v}_n$'.

5.4.1. The General Model

In the general parameter change model we assume that, for all $j = 1, \ldots, q+1$, the following conditions are satisfied:

(M1) Let $\{\mathbb{X}_i^{(j)}\}_{i \ge 1}$ be a stationary and ergodic sequence in $\mathbb{R}^p$.

(M2) Let $\boldsymbol{S}(j, k, \boldsymbol{\theta}_j) = \sum_{i=1}^k \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)$ fulfill a strong invariance principle such that (possibly after changing the probability space) there exists a p-dimensional standard Wiener process $\{\boldsymbol{W}(k) : k \ge 0\}$ with identity matrix $\boldsymbol{I}_p$ as covariance matrix and $\nu > 0$ such that

$$\left\| \boldsymbol{\Sigma}_{(j)}^{-1/2} \left(\boldsymbol{S}(j, k, \boldsymbol{\theta}_j) - E(\boldsymbol{S}(j, k, \boldsymbol{\theta}_j))\right) - \boldsymbol{W}(k) \right\| = O(k^{1/(2+\nu)}) \;\; a.s.$$

as k goes to infinity.

(M3) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for all $l = 1, \ldots, p$, let the sequence $\left\{\nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\}$ fulfill a strong invariance principle as described in (M2).

(M4) For all $l = 1, \ldots, p$, let $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$ and let the sequence $\left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\|_F\right\}$ satisfy a strong invariance principle as in (M2).

(M5) Let $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$ hold.

(M6) Let the following forward and backward Hájek-Rényi-type inequalities hold for $\widetilde{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{1+q}\}$, for any $m_n \in \mathbb{N}_0$ and a positive deterministic sequence $\{v_n\}$ with $v_n \to \infty$ (which will be specified later):

$$\max_{v_n \le k \le n-m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n+1}^{m_n+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\| = O_P(1)$$

and

$$\max_{v_n \le k \le m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n-k+1}^{m_n} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\| = O_P(1)$$

(M7) Let $\boldsymbol{V}_j(\boldsymbol{\theta}) = E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)^T$ be a regular matrix for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and let

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{V}_j(\boldsymbol{\theta})^{-1}\right\|_F < \infty, \text{ for all } j = 1, \ldots, q+1.$$

(M8) Let $\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta})$ be a regular matrix for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\delta \in [0,1]$ and let

$$\sup_{\delta\in[0,1]} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|(\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta}))^{-1}\right\|_F < \infty, \text{ for all } j = 1, \ldots, q.$$

(M9) For $s \ge 1$ we assume that $\sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) \boldsymbol{V}_l(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left(\sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) \boldsymbol{V}_l(\boldsymbol{\theta}) \right)^{-1} \right\|_F < \infty.$$

(M10) Let $E\left(Q(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right) < \infty$ where $\widetilde{\boldsymbol{\theta}}$ denotes the unique zero of

$$\sum_{j=1}^{q+1} (\lambda_j - \lambda_{j-1})\, E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right).$$

Remark 5.4.1. *Note that Assumption (M6) holds for any positive deterministic sequence $\{v_n\}$ with $v_n \to \infty$ if the series $\{\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\}$ is time-reversible and satisfies Assumption (M3). For instance, i.i.d sequences of type (E1) and stationary and strongly mixing sequences of type (E2) as considered in Section 2.3 and Section 3.1 fulfill these conditions. If the series satisfies Assumption (M3) but is not time-reversible, we only get that Assumption (M6) holds for sequences $\{v_n\}$ with $\frac{n^{1/(2+\nu)}}{\sqrt{v_n}} = O(1)$.*

The first results up to Lemma 5.4.8 give uniform statements on the approximation of sums of transformed sequences which will enable us to investigate the asymptotic behavior of the local estimator sequences in a uniform manner.
Remember that $k_{1,n}, \ldots, k_{q,n}$ denote the change points and that we set $k_{0,n} = 1$ and $k_{q+1,n} = n$ for all n. Furthermore, we use

$$\delta_{j,n} = k_{j,n} - k_{j-1,n}, \; j = 1, \ldots, q+1. \tag{5.5}$$

Lemma 5.4.2. *Let the Assumptions (M1) and (M2) be fufilled. Then, for any positive deterministic sequence $\{v_n\}$, it holds*

$$\max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\| \frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right\| = O_P\left(\max\left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right).$$

Proof. At first, we assume that $\boldsymbol{\Sigma}_{(j)} = \boldsymbol{I}_p$. On noting that $E\left(\boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right) = \boldsymbol{0}$, the strong invariance principle in (M2) and Lemma E.1.6 (b) can be applied to receive

$$\begin{aligned}
&\max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\| \frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right\| \leq \frac{1}{v_n} O_P\left(n^{1/(2+\nu)}\right) + \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \frac{1}{h} \left\| \boldsymbol{W}(l+h) - \boldsymbol{W}(l) \right\| \\
&\leq O_P\left(\frac{n^{1/(2+\nu)}}{v_n}\right) + \frac{1}{\sqrt{v_n}} \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \frac{1}{\sqrt{h}} \left\| \boldsymbol{W}(l+h) - \boldsymbol{W}(l) \right\| \\
&\leq O_P\left(\frac{n^{1/(2+\nu)}}{v_n}\right) + \frac{1}{\sqrt{v_n}} \sum_{m=1}^{p} \max_{0 \leq l_1 < l_2 \leq n} \frac{1}{\sqrt{l_2 - l_1}} \left| W_m(l_2) - W_m(l_1) \right| \\
&= O_P\left(\frac{n^{1/(2+\nu)}}{v_n}\right) + O_P\left(\frac{\sqrt{\log(n)}}{\sqrt{v_n}}\right),
\end{aligned}$$

where the last line follows from Lemma 1 in Yao (1988) since the increments of a Wiener process can be represented as sums of independent standard normal distributed random variables.
If $\boldsymbol{\Sigma}_{(j)} \neq \boldsymbol{I}_p$, Lemma E.1.5 yields

$$\begin{aligned}
&\max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\| \frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right\| = \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\| \boldsymbol{\Sigma}_{(j)}^{1/2} \boldsymbol{\Sigma}_{(j)}^{-1/2} \frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right\| \\
&\leq \left\| \boldsymbol{\Sigma}_{(j)}^{1/2} \right\|_F \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\| \frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{\Sigma}_{(j)}^{-1/2} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right\| = O_P\left(\max\left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right),
\end{aligned}$$

where the last line follows from the first the part and since $\left\| \boldsymbol{\Sigma}_{(j)}^{1/2} \right\|_F = O(1)$. □

Remark 5.4.3. *The value ν appearing in the rate in the lemma above comes from the invariance principle of the considered transformed sequence. In the following lemmata we need invariance principles for different sequences (but only finitely many), for instance $\{\boldsymbol{H}(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\}_{i\geq 1}$, $\{\nabla H_l(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\}_{i\geq 1}$. Hence, throughout this chapter we choose ν as the minimum over the values of all these invariance principles.*

Lemma 5.4.4. *(a) Let the Assumptions (M1) and (M3) be fufilled. Then, for all $l = 1,\dots,p$ and for any positive deterministic sequence $\{v_n\}$, it holds*

$$\max_{\substack{h\geq v_n\\ 0\leq l\leq n-h}} \left\| \frac{1}{h}\sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\right\| = O_P\left(\max\left\{\frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}}\right\}\right).$$

(b) Let the Assumptions (M1) and (M4) be satisfied. Then, for all $l = 1,\dots,p$ and for any positive deterministic sequence $\{v_n\}$, it holds

$$\max_{\substack{h\geq v_n\\ 0\leq l\leq n-h}} \frac{1}{h}\left|\sum_{i=l+1}^{l+h} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_{0,l}(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\right\|_F - E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^2 H_{0,l}(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\right\|_F\right)\right|$$
$$= O_P\left(\max\left\{\frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}}\right\}\right).$$

Proof. The results can be derived in an analogous manner to Lemma 5.4.2. □

Lemma 5.4.5. *Let the Assumptions (M1), (M3) and (M4) be satisfied. Then, for any positive deterministic sequence $\{v_n\}$ with $\frac{n^{1/2+\nu)}}{v_n}\to 0$,*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{\substack{h\geq v_n\\ 0\leq l\leq n-h}} \left\| \frac{1}{h}\sum_{i=l+1}^{l+h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)},\boldsymbol{\theta})\right\|_F = o_P(1).$$

Proof. The basic idea of this proof is well known in non-parametric statistics and it can be used in general to derive uniform results on a compact space. We have only adapted the arguments to our specific setting.
There are three main arguments:

(1) The compactness of the parameter space $\boldsymbol{\Theta}$ implies that for each $\delta > 0$ there exist a finite number $M = M(\delta) \geq 1$ and $\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M \in \boldsymbol{\Theta}$ such that for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ there is an $m \leq M$ with $\|\boldsymbol{\theta}-\boldsymbol{\theta}_m\| < \delta$.

(2) For fixed M and $\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M$, applying Lemma 5.4.4 (a) in connection with Assumption (M3) yields

$$\sup_{m\leq M} \max_{\substack{h\geq v_n\\ 0\leq l\leq n-h}} \left\| \frac{1}{h}\sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)},\boldsymbol{\theta}_m)\right\| \leq \sum_{m=1}^{M} \max_{\substack{h\geq v_n\\ 0\leq l\leq n-h}} \left\| \frac{1}{h}\sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)},\boldsymbol{\theta}_m)\right\|$$
$$= O_P\left(\max\left\{\frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}}\right\}\right) = o_P(1),$$

as $\frac{\sqrt{\log(n)}}{\sqrt{v_n}} = O\left(\frac{\sqrt{n^{1/(2+\nu)}}}{\sqrt{v_n}}\right)$ implies $\max\left\{\frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}}\right\} \to 0$.

(3) Since the estimating function is twice continuously differentiable on $\boldsymbol{\Theta}$ a first order Taylor expansion on ∇H_l shows, for any $\boldsymbol{\theta}, \boldsymbol{\xi} \in \boldsymbol{\Theta}$,

$$\|\nabla H_l(\boldsymbol{x}, \boldsymbol{\theta}) - \nabla H_l(\boldsymbol{x}, \boldsymbol{\xi})\| \leq \sup_{\boldsymbol{\eta} \in \boldsymbol{\Theta}} \left\|\nabla^2 H_l(\boldsymbol{x}, \boldsymbol{\eta})\right\|_F \|\boldsymbol{\theta} - \boldsymbol{\xi}\|,$$

which is well defined at least almost surely with respect to $P_{\mathbb{X}_1}$ since $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$ holds by (M4).
Let $L(\mathbb{X}_i^{(j)}) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\|_F$. Hence, for any $\boldsymbol{\theta}, \boldsymbol{\xi} \in \boldsymbol{\Theta}$ with $\|\boldsymbol{\theta} - \boldsymbol{\xi}\| < \delta$, we obtain

$$\begin{aligned} &\frac{1}{h} \sum_{i=l+1}^{l+h} \left\|\nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) - \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\xi})\right\| \\ &\leq \delta \left(\frac{1}{h} \sum_{i=l+1}^{l+h} \left(L(\mathbb{X}_i^{(j)}) - E(L(\mathbb{X}_1^{(j)}))\right) + 2E(L(\mathbb{X}_1^{(j)}))\right). \end{aligned}$$

Furthermore, for each $\varepsilon > 0$ we can choose a $\delta > 0$ such that $\frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1^{(j)})) > 0$ as $E(L(\mathbb{X}_1^{(j)})) < \infty$ by Assumption (M4). Thus, combining (2) and (3) shows

$$\begin{aligned} &P\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\|\frac{1}{h} \sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\| > \varepsilon\right) \\ &= P\left(\sup_{m \leq M} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_m\|} \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\|\frac{1}{h} \sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\| > \varepsilon\right) \\ &= P\left(\sup_{m \leq M} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_m\|} \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\|\frac{1}{h} \sum_{i=l+1}^{l+h} \left(\nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) - \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_m)\right)\right.\right. \\ &\qquad\qquad \left.\left. + \frac{1}{h} \sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_m)\right\| > \varepsilon\right) \\ &\leq P\left(\sup_{m \leq M} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_m\|} \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\|\frac{1}{h} \sum_{i=l+1}^{l+h} \left(\nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) - \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_m)\right)\right\| > \frac{\varepsilon}{2}\right) \\ &\quad + P\left(\sup_{m \leq M} \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\|\frac{1}{h} \sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_m)\right\| > \frac{\varepsilon}{2}\right) \\ &\leq P\left(\max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \frac{1}{h} \sum_{i=l+1}^{l+h} \left(L(\mathbb{X}_i^{(j)}) - E(L(\mathbb{X}_1^{(j)}))\right) > \frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1^{(j)}))\right) + o(1) = o(1), \end{aligned}$$

where the last step follows from Lemma 5.4.4 (b) and since $\frac{\varepsilon}{2\delta} - 2E\left(L\left(\mathbb{X}_1^{(j)}\right)\right) > 0$ by the choice of δ. Finally, applying Lemma E.1.6 (d) completes the proof as

$$\max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\|\frac{1}{h} \sum_{i=l+1}^{l+h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\|_F \leq \sum_{l=1}^{p} \max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\|\frac{1}{h} \sum_{i=l+1}^{l+h} \nabla H_{0,l}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\| = o_P(1).$$

□

Lemma 5.4.6. *Let the Assumptions (M1) and (M5) hold. Then, for any positive deterministic sequence $\{v_n\}$ with $v_n \to \infty$,*

$$\sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=k_{j,n}+1}^{k_{j,n}+h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = o_P(1)$$

and

$$\sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = o_P(1).$$

Proof. Note that the first derivatives of the estimating function are measurable with respect to $\mathbb{X}_i^{(j)}$ so that the transformed sequence $\{\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\}$ is stationary and ergodic by Assumption (M1). Furthermore, with the stationarity of the sequence we obtain

$$\sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=k_{j,n}+1}^{k_{j,n}+h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \stackrel{D}{=} \sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=1}^{h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F .$$

Moreover, applying Assumption (M5) together with the Uniform Ergodic Theorem given in Corollary E.2.7 yields

$$\sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=1}^{h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = o(1) \quad a.s.$$

The second assertion can be derived in an analogous manner with

$$\sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \stackrel{D}{=} \sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=n-h+1}^{n} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F$$

and since the series in reversed time is ergodic as well which implies that

$$\sup_{h \geq v_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=n-h+1}^{n} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = o(1) \quad a.s.$$

by the Uniform Ergodic Theorem in Corollary E.2.7. □

Lemma 5.4.7. *Let the Assumptions (M1) and (M5) hold. Furthermore, let $\{v_n\}$ and $\{\tilde{v}_n\}$ be positive deterministic sequences with $v_n \to \infty$ as $n \to \infty$. Then,*

$$\max_{g \leq \tilde{v}_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{g + v_n} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = o_P(1)$$

and

$$\max_{g \leq \tilde{v}_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{g + v_n} \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = o_P(1).$$

Proof. With the stationarity of the sequence, we receive

$$\begin{aligned}
&\max_{g\le\tilde v_n}\sup_{\boldsymbol\theta\in\boldsymbol\Theta}\left\|\frac{1}{v_n+g}\sum_{i=k_{j,n}+1}^{k_{j,n}+g}\nabla\boldsymbol H_0(\mathbb X_i^{(j)},\boldsymbol\theta)\right\|_F \overset{D}{=} \max_{g\le\tilde v_n}\sup_{\boldsymbol\theta\in\boldsymbol\Theta}\left\|\frac{1}{v_n+g}\sum_{i=1}^{g}\nabla\boldsymbol H_0(\mathbb X_i^{(j)},\boldsymbol\theta)\right\|_F\\
&\le \max_{g<\sqrt{v_n}}\sup_{\boldsymbol\theta\in\boldsymbol\Theta}\frac{g}{v_n}\left\|\frac1g\sum_{i=1}^{g}\nabla\boldsymbol H_0(\mathbb X_i^{(j)},\boldsymbol\theta)\right\|_F+\max_{g\ge\sqrt{v_n}}\sup_{\boldsymbol\theta\in\boldsymbol\Theta}\left\|\frac1g\sum_{i=1}^{g}\nabla\boldsymbol H_0(\mathbb X_i^{(j)},\boldsymbol\theta)\right\|_F\\
&\le \max_{g<\sqrt{v_n}}\sup_{\boldsymbol\theta\in\boldsymbol\Theta}\frac{1}{\sqrt{v_n}}\left\|\frac1g\sum_{i=1}^{g}\nabla\boldsymbol H_0(\mathbb X_i^{(j)},\boldsymbol\theta)\right\|_F+\max_{g\ge\sqrt{v_n}}\sup_{\boldsymbol\theta\in\boldsymbol\Theta}\left\|\frac1g\sum_{i=1}^{g}\nabla\boldsymbol H_0(\mathbb X_i^{(j)},\boldsymbol\theta)\right\|_F\\
&= o_P(1).
\end{aligned}$$

The first summand in the inequality above converges to zero as $\frac{1}{\sqrt{v_n}}\to 0$ and since

$$\max_{1\le g\le n}\sup_{\boldsymbol\theta\in\boldsymbol\Theta}\left\|\frac1g\sum_{i=1}^{g}\nabla\boldsymbol H_0(\mathbb X_i^{(j)},\boldsymbol\theta)\right\| = O(1)\quad a.s.$$

by the Uniform Ergodic Theorem in Corollary E.2.7. The convergence of the second summand follows from Corollary E.2.7 as well.
The second assertion can be shown similarly by using the ergodicity of the sequence in reversed time. □

Lemma 5.4.8. *Let the Assumption (M1) hold. Furthermore, let Assumption (M6) be satisfied by $\{v_n\}$, for $\widetilde{\boldsymbol\theta}\in\{\boldsymbol\theta_1,\dots,\boldsymbol\theta_{q+1}\}$. Then, for any positive deterministic sequence $\{\tilde v_n\}$,*

$$\max_{g\le\tilde v_n}\left\|\frac{1}{v_n+g}\sum_{i=k_{j,n}+1}^{k_{j,n}+g}\boldsymbol H_0(\mathbb X_i^{(j)},\widetilde{\boldsymbol\theta})\right\| = O_P\left(\frac{1}{\sqrt{v_n}}\right)$$

and

$$\max_{g\le\tilde v_n}\left\|\frac{1}{v_n+g}\sum_{i=k_{j,n}-g+1}^{k_{j,n}}\boldsymbol H_0(\mathbb X_i^{(j)},\widetilde{\boldsymbol\theta})\right\| = O_P\left(\frac{1}{\sqrt{v_n}}\right)$$

Proof. By the stationarity of the sequence we get

$$\begin{aligned}
&\max_{g\le\tilde v_n}\left\|\frac{1}{v_n+g}\sum_{i=k_{j,n}+1}^{k_{j,n}+g}\boldsymbol H_0(\mathbb X_i^{(j)},\widetilde{\boldsymbol\theta})\right\| \overset{D}{=} \max_{g\le\tilde v_n}\left\|\frac{1}{v_n+g}\sum_{i=1}^{g}\boldsymbol H_0(\mathbb X_i^{(j)},\widetilde{\boldsymbol\theta})\right\| \qquad (5.6)\\
&\le \max_{g<\sqrt{v_n}}\frac{1}{\sqrt{v_n}}\left\|\frac1g\sum_{i=1}^{g}\boldsymbol H_0(\mathbb X_i^{(j)},\widetilde{\boldsymbol\theta})\right\|+\max_{g\ge\sqrt{v_n}}\left\|\frac1g\sum_{i=1}^{g}\boldsymbol H_0(\mathbb X_i^{(j)},\widetilde{\boldsymbol\theta})\right\| = O_P\left(\frac{1}{\sqrt{v_n}}\right),
\end{aligned}$$

where the approximation of the second summand follows directly from the Hájek-Rényi-type inequality of Assumption (M6). The first summand is $O_P\left(\frac{1}{\sqrt{v_n}}\right)$ since the Ergodic Theorem shows

$$\max_{1\le g\le n}\left\|\frac1g\sum_{i=1}^{g}\boldsymbol H_0(\mathbb X_i^{(j)},\widetilde{\boldsymbol\theta})\right\| = O(1)\quad a.s.$$

The second statement can be derived in an analogous manner by using the backward Hájek-Rényi-type inequality and on noting that the series in reversed time is ergodic as well. □

Now we are able to investigate the behavior of the estimator sequences. In the following two lemmata we concentrate on situations where no change occurs between the considered time points, i.e. the estimators are computed on stationary subsequences of the time series.

Lemma 5.4.9. *Let the Assumptions (M1), (M2), (M3), (M4) and (M7) be satisfied. Then, for any positive deterministic sequence $\{v_n\}$ with $\frac{n^{1/2+\nu)}}{v_n} \to 0$, it holds*

$$\max_{\substack{v_n \le h < \delta_{j,n} \\ k_{j-1,n} < l \le k_{j,n}-h}} \left\| \widehat{\boldsymbol{\theta}}_{l+1,l+h} - \boldsymbol{\theta}_j \right\| = O_P\left(\max\left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right).$$

Proof. A first order Taylor expansion yields, that there exists a $\boldsymbol{\eta}_{l,h,n}$ with $\left\| \boldsymbol{\eta}_{l,h,n} - \widehat{\boldsymbol{\theta}}_{l+1,l+h} \right\| \le \left\| \boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{l+1,l+h} \right\|$ such that

$$\begin{aligned}
\frac{1}{h}\sum_{i=l+1}^{l+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) &= \frac{1}{h}\sum_{i=l+1}^{l+h} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{l,h,n})^T \left(\widehat{\boldsymbol{\theta}}_{l+1,l+h} - \boldsymbol{\theta}_j \right) \\
&= \left(\frac{1}{h}\sum_{i=l+1}^{l+h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{l,h,n}) + E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{l,h,n}) \right) \right)^T \left(\widehat{\boldsymbol{\theta}}_{l+1,l+h} - \boldsymbol{\theta}_j \right) \\
&= \left(o_P(1) + E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{l,h,n}) \right) \right)^T \left(\widehat{\boldsymbol{\theta}}_{l+1,l+h} - \boldsymbol{\theta}_j \right) \quad \text{uniformly in } l \text{ and } h,
\end{aligned}$$

where the last line follows from Lemma 5.4.5 since

$$\begin{aligned}
&\max_{\substack{v_n \le h < \delta_{j,n} \\ k_{j-1,n} < l \le k_{j,n}-h}} \left\| \frac{1}{h}\sum_{i=l+1}^{l+h} \left(\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{l,h,n}) \right) \right\|_F \\
&\le \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \max_{\substack{v_n \le h < \delta_{j,n} \\ k_{j-1,n} < l \le k_{j,n}-h}} \left\| \frac{1}{h}\sum_{i=l+1}^{l+h} \left(\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right) \right\|_F = o_P(1).
\end{aligned}$$

Moreover, Lemma 5.4.2 gives

$$\max_{\substack{v_n \le h < \delta_{j,n} \\ k_{j-1,n} < l \le k_{j,n}-h}} \left\| \frac{1}{h}\sum_{i=l+1}^{l+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right\| = O_P\left(\max\left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right).$$

Hence, we can conclude that

$$\begin{aligned}
&O_P\left(\max\left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right) \\
&= \left(o_P(1) + E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{l,h,n}) \right) \right)^T \left(\widehat{\boldsymbol{\theta}}_{l+1,l+h} - \boldsymbol{\theta}_j \right) \quad \text{uniformly in } l \text{ and } h.
\end{aligned}$$

Furthermore, after multiplying both sides of the equation above with the inverse of the expectation matrix, Lemma E.2.21 in connection with Assumption (M7) can be applied to get

$$O_P\left(\max\left\{\frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}}\right\}\right) = (o_P(1) + \boldsymbol{I}_p)^T \left(\widehat{\boldsymbol{\theta}}_{l+1,l+h} - \boldsymbol{\theta}_j\right) \text{ uniformly in } l \text{ and } h.$$

Finally, Lemma E.2.22 shows the assertion. □

This rate of convergence can be improved if we have further information about the start and end points of the subsample on which we compute the estimator sequence.

Lemma 5.4.10. *Let the Assumptions (M1), (M5) and (M7) hold. Furthermore, let Assumption (M6) be satisfied by* $\delta_{j,n}$ *as in* (5.5)*. Then, for any positive deterministic sequences* $\{\bar{v}_n\}$ *and* $\{\tilde{v}_n\}$ *fulfilling that* $\delta_{j,n} - \bar{v}_n - \tilde{v}_n > 0$ *and* $\frac{n}{\delta_{j,n}-\bar{v}_n-\tilde{v}_n} = O(1)$,

$$\max_{\substack{0\le h\le \bar{v}_n \\ 0\le g\le \tilde{v}_n}} \left\|\widehat{\boldsymbol{\theta}}_{k_{j-1,n}+h+1,k_{j,n}-g} - \boldsymbol{\theta}_j\right\| = O_P\left(\frac{1}{\sqrt{n}}\right),$$

Proof. A first order Taylor expansion shows

$$\begin{aligned}
&\frac{1}{n}\sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \qquad (5.7)\\
&= \frac{1}{n}\sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{g,h,n})^T \left(\widehat{\boldsymbol{\theta}}_{k_{j-1,n}+h+1,k_{j,n}-g} - \boldsymbol{\theta}_j\right)
\end{aligned}$$

for some $\boldsymbol{\eta}_{g,h,n}$ with

$$\left\|\boldsymbol{\eta}_{g,h,n} - \widehat{\boldsymbol{\theta}}_{k_{j-1,n}+h+1,k_{j,n}-g}\right\| \le \left\|\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{k_{j-1,n}+h+1,k_{j,n}-g}\right\|.$$

On noting that $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}_j)\right) = \boldsymbol{0}$, applying Lemma 5.4.8 together with Assumption (M6) yields

$$\begin{aligned}
&\max_{\substack{0\le h\le \bar{v}_n \\ 0\le g\le \tilde{v}_n}} \left\|\frac{1}{n}\sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right\| \qquad (5.8)\\
&= \max_{\substack{0\le h\le \bar{v}_n \\ 0\le g\le \tilde{v}_n}} \left\|\frac{1}{n}\left(\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) - \sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) - \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right)\right\|\\
&\le \frac{\delta_{j,n}}{n}\left\|\frac{1}{\delta_{j,n}}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right\| + \frac{\delta_{j,n}+\bar{v}_n}{n}\max_{0\le h\le \bar{v}_n}\left\|\frac{1}{\delta_{j,n}+h}\sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right\|\\
&\quad + \frac{\delta_{j,n}+\tilde{v}_n}{n}\max_{0\le g\le \tilde{v}_n}\left\|\frac{1}{\delta_{j,n}+\tilde{v}_n}\sum_{i=k_{j,n}-g+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right\|\\
&= O_P\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Moreover, by Lemma 5.4.7 we obtain

$$\begin{aligned}
&\max_{\substack{0\le h\le \bar{v}_n\\ 0\le g\le \tilde{v}_n}} \left\| \frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{g,h,n}) \right\|_F \\
&= \max_{\substack{0\le h\le \bar{v}_n\\ 0\le g\le \tilde{v}_n}} \left\| \frac{1}{n} \left(\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{g,h,n}) - \sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{g,h,n}) \right.\right. \\
&\qquad\qquad \left.\left. - \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{g,h,n}) \right) \right\|_F \\
&\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=k_{j-1,n}+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F + \max_{0\le h\le \bar{v}_n} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \\
&\quad + \max_{0\le g\le \tilde{v}_n} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \\
&= o_P(1),
\end{aligned}$$

implying that

$$\frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{g,h,n}) = \frac{\delta_{j,n} - h - g}{n} E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{g,h,n}) \right) + o_P(1),$$

holds uniformly in g and h. Thus, together with (5.7) and (5.8) we can conclude that

$$O_P\left(\frac{1}{\sqrt{n}}\right) = \left(\frac{\delta_{j,n} - h - g}{n} E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{g,h,n}) \right) + o_P(1) \right)^T \left(\widehat{\boldsymbol{\theta}}_{k_{j-1,n}+h+1,k_{j,n}-g} - \boldsymbol{\theta}_j \right)$$

uniformly in g and h.

On noting that

$$\frac{n}{\delta_{j,n} - h - g} \left\| E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{g,h,n}) \right)^{-1} \right\|_F \le \frac{n}{\delta_{j,n} - \bar{v}_n - \tilde{v}_n} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}) \right)^{-1} \right\|_F = O(1)$$

holds uniformly in g and h by Assumption (M7), with Lemma E.2.21 both sides of the equation above can be multiplied with the inverse of the expectation matrix in order to get

$$O_P\left(\frac{1}{\sqrt{n}}\right) = (\boldsymbol{I}_p + o_P(1))^T \left(\widehat{\boldsymbol{\theta}}_{k_{j-1,n}+h+1,k_{j,n}-g} - \boldsymbol{\theta}_j \right) \text{ uniformly in } g \text{ and } h,$$

which completes the proof by Lemma E.2.22. □

In the next lemma we focus on situations in which exactly one change arises between the two considered time points so that the sample, on which an estimator is computed, can be divided into two subsamples: before and after the change. Lemma 5.4.11 in particular shows that we still get uniform convergence if one subsample increases with n whereas the second subsample is dominated by the other one.

Lemma 5.4.11. *Let the Assumptions (M1) and (M5) hold. Furthermore, let (M6) be satisfied by $\{v_n\}$ and let $\delta_{j,n}$ as in* (5.5).

(a) If Assumption (M7) is fulfilled, then, for any positive deterministic sequence $\{\tilde{v}_n\}$ with $\tilde{v}_n = o(v_n)$,

$$\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j,n}+g} - \boldsymbol{\theta}_j \right\| = O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right)$$

and

$$\max_{\substack{v_n \le h < \delta_{j+1,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \widehat{\boldsymbol{\theta}}_{k_{j,n}-g+1,k_{j,n}+h} - \boldsymbol{\theta}_{j+1} \right\| = O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right).$$

(b) If Assumption (M8) holds, we get the same result as in (a) for any positive deterministic sequence $\{\tilde{v}_n\}$ with $\tilde{v}_n \le \delta_{j+1,n}$ and $\tilde{v}_n \le \delta_{j,n}$ respectively.

Proof. We only derive the first assertion since the second statement in (a) can be shown in an analogous manner.
First note that h, which is greater than v_n, denotes the length of the subsample before the change and g represents the length of the subsample after the change point bounded by $\tilde{v}_n$. A first order Taylor expansion yields, that there exists an $\boldsymbol{\eta}_{h,g,n}$ with $\left\| \boldsymbol{\eta}_{h,g,n} - \widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j,n}+g} \right\| \le \left\| \boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j,n}+g} \right\|$ such that

$$\begin{aligned}
&\frac{1}{h+g} \sum_{i=k_{j,n}-h+1}^{k_{j,n}+g} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_j) \qquad (5.9)\\
&= \frac{1}{h+g} \sum_{i=k_{j,n}-h+1}^{k_{j,n}+g} \nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\eta}_{h,g,n})^T \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-g+1,k_{j,n}+h} - \boldsymbol{\theta}_j \right)\\
&= \Bigg(\frac{1}{h+g} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{h,g,n}) \right)\\
&\quad + \frac{1}{h+g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \boldsymbol{\eta}_{h,g,n}) \right) + \frac{h}{h+g} E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{h,g,n}) \right)\\
&\quad + \frac{g}{h+g} E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\eta}_{h,g,n}) \right) \Bigg)^T \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j,n}+g} - \boldsymbol{\theta}_j \right).
\end{aligned}$$

Furthermore, we can apply Lemma 5.4.6 in order to receive

$$\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \frac{1}{h+g} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{h,g,n}) \right\|_F$$

$$\leq \max_{v_n \leq h < \delta_{j,n}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F = o_P(1).$$

Moreover, by Lemma 5.4.7 we obtain

$$\max_{\substack{v_n \leq h < \delta_{j,n} \\ 1 \leq g \leq \tilde{v}_n}} \left\| \frac{1}{h+g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{h,g,n}) \right) \right\|_F$$
$$\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \max_{1 \leq g \leq \tilde{v}_n} \left\| \frac{1}{v_n+g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right) \right\|_F = o_P(1).$$

Thus, by considering (5.9) again, with $\frac{g}{h+g} \leq \frac{\tilde{v}_n}{v_n} \to 0$ and Assumption (M5) we get

$$\frac{1}{g+h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}+g} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)$$
$$= \left(o_P(1) + \frac{h}{h+g} E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{h,g,n}) \right) \right)^T \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\theta}_j \right)$$

uniformly in h and g.

Besides, the left hand side of the equation above can be approximated in the following way. Applying the Hájek-Rényi-type inequality in Assumption (M6) and Lemma 5.4.8 yields

$$\max_{\substack{v_n \leq h < \delta_{j,n} \\ 1 \leq g \leq \tilde{v}_n}} \left\| \frac{1}{h+g} \sum_{i=k_{j,n}-h+1}^{k_{j,n}+g} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_j) \right\|$$
$$= \max_{\substack{v_n \leq h < \delta_{j,n} \\ 1 \leq g \leq \tilde{v}_n}} \left\| \frac{1}{h+g} \left(\sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) + \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{H}(\mathbb{X}_i^{(j+1)}, \boldsymbol{\theta}_j) \right) \right\|$$
$$\leq \max_{v_n \leq h < \delta_{j,n}} \left\| \frac{1}{h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j) \right\| + \frac{\tilde{v}_n}{v_n} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \boldsymbol{\theta}_j) \right) \right\|$$
$$+ \max_{1 \leq g \leq \tilde{v}_n} \left\| \frac{1}{v_n+g} \left(\sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{H}_0(\mathbb{X}_i^{(j+1)}, \boldsymbol{\theta}_j) \right) \right\|$$
$$= O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right).$$

Hence, we receive, uniformly in h and g,

$$O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right)$$
$$= \left(o_P(1) + \frac{h}{h+g} E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\eta}_{h,g,n}) \right) \right)^T \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\theta}_j \right).$$

Furthermore, with Assumption (M7) and since $\frac{\tilde{v}_n}{v_n} + o(1)$ we get

$$\frac{h+g}{h}\left\|E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\eta}_{h,g,n})\right)^{-1}\right\|_F \leq \left(1+\frac{\tilde{v}_n}{v_n}\right)\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right)^{-1}\right\|_F = O(1).$$

Thus, by Lemma E.2.21 we can multiply the inverse of $\frac{h}{h+g}E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\eta}_{h,g,n})\right)$ to both sides of the equation above to get

$$O_P\left(\max\left\{\frac{1}{\sqrt{v_n}},\frac{\tilde{v}_n}{v_n}\right\}\right) = (o_P(1)+\boldsymbol{I}_p)^T\left(\widehat{\boldsymbol{\theta}}_{l+1,l+h}-\boldsymbol{\theta}_j\right) \text{ uniformly in } h \text{ and } g.$$

Finally, Lemma E.2.22 shows the assertion of (a).
For proving part (b), similar arguments as in (a) can be used to obtain

$$\begin{aligned}&\frac{1}{g+h}\sum_{i=k_{j,n}-h+1}^{k_{j,n}+g}\boldsymbol{H}(\mathbb{X}_i^{(j)},\boldsymbol{\theta}_j)\\&=\left(o_P(1)+\frac{h}{h+g}E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\eta}_{h,g,n})\right)+\frac{g}{h+g}E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(j+1)},\boldsymbol{\eta}_{h,g,n})\right)\right)^T\\&\qquad\qquad\left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j,n}+g}-\boldsymbol{\theta}_j\right)\quad\text{uniformly in } h \text{ and } g.\end{aligned}$$

By Assumption (M8) and Lemma E.2.21 we receive

$$O_P\left(\max\left\{\frac{1}{\sqrt{v_n}},\frac{\tilde{v}_n}{v_n}\right\}\right) = (o_P(1)+\boldsymbol{I}_p)^T\left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j,n}+g}-\boldsymbol{\theta}_j\right) \text{ uniformly in } h \text{ and } g,$$

so that Lemma E.2.22 completes the proof. □

Until now we have focused on scenarios with at most one change occuring in the considered subsample. However, it might be of interest as well to examine the behavior of estimator sequences computed on subsamples in which more changes arise. The following lemma incorporates such scenarios.

Lemma 5.4.12. *Let the Assumptions (M1), (M5) and (M9), for some $s \geq 1$, hold. Furthermore, let (M6) be satisfied by a sequence $\{v_n\}$ of order n. Then, for any positive deterministic sequences $\{\bar{v}_n\}$ and $\{\tilde{v}_n\}$ with $\bar{v}_n = o(n)$ and $\tilde{v}_n = o(n)$,*

$$\max_{\substack{1\leq h\leq\bar{v}_n\\1\leq g\leq\tilde{v}_n}}\left\|\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j+s,n}+g}-\widetilde{\boldsymbol{\theta}}_{j+1,j+s}\right\| = O_P\left(\max\left\{\frac{1}{\sqrt{n}},\frac{\bar{v}_n}{n},\frac{\tilde{v}_n}{n}\right\}\right),$$

where $\widetilde{\boldsymbol{\theta}}_{j+1,j+s}$ is the unique zero of $\sum_{l=j+1}^{j+s}(\lambda_l-\lambda_{l-1})E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)},\boldsymbol{\theta})\right)$.

Proof. In the proof we use the short version $\widetilde{\boldsymbol{\theta}}$ for $\widetilde{\boldsymbol{\theta}}_{j+1,j+s}$ and δ_n for $\delta_{j,j+s,n} := k_{j+s,n} - k_{j,n}$. Furthermore, remember that $k_{j,n} = \lfloor\lambda_j n\rfloor$. By a first order Taylor expansion we get

$$\frac{1}{\delta_n+h+g}\sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g}\boldsymbol{H}(\mathbb{X}_i,\widetilde{\boldsymbol{\theta}}) \tag{5.10}$$

$$= \frac{1}{\delta_n + h + g} \sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g} \nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\eta}_{h,g,n})^T \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j+s,n}+g} - \widetilde{\boldsymbol{\theta}} \right).$$

Besides, with $\sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) = \mathbf{0}$ and

$$\begin{aligned}
&\left\| n \sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) - \sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| \qquad (5.11)\\
&= \left\| \sum_{l=j+1}^{j+s} (n\lambda_l - \lfloor n\lambda_l \rfloor - (n\lambda_{l-1} - \lfloor n\lambda_{l-1} \rfloor)) E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| \\
&\leq \sum_{l=j+1}^{j+s} (n\lambda_l - \lfloor n\lambda_l \rfloor) \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| + \sum_{l=j+1}^{j+s} (n\lambda_{l-1} - \lfloor n\lambda_{l-1} \rfloor) \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| \\
&\leq 2 \sum_{l=j+1}^{j+s} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| = O(1)
\end{aligned}$$

we receive $\sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) = O(1)$. Hence, applying Assumption (M6) and Lemma 5.4.8 yields

$$\begin{aligned}
&\max_{\substack{1 \leq h \leq \bar{v}_n \\ 1 \leq g \leq \tilde{v}_n}} \left\| \frac{1}{\delta_n + h + g} \sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \right\| \qquad (5.12)\\
&= \max_{\substack{1 \leq h \leq \bar{v}_n \\ 1 \leq g \leq \tilde{v}_n}} \left\| \frac{1}{\delta_n + h + g} \left(\sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) + \sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \boldsymbol{H}(\mathbb{X}_i^{(l)}, \widetilde{\boldsymbol{\theta}}) \right.\right. \\
&\qquad \left.\left. + \sum_{i=k_{j+s,n}+1}^{k_{j+s,n}+g} \boldsymbol{H}(\mathbb{X}_i^{(j+s+1)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| \\
&\leq \max_{1 \leq h \leq \bar{v}_n} \frac{1}{\delta_n + h} \left\| \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\| + \frac{\bar{v}_n}{\delta_n} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| \\
&\quad + \sum_{l=j+1}^{j+s} \frac{1}{k_{l,n} - k_{l-1,n}} \left\| \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(l)}, \widetilde{\boldsymbol{\theta}}) \right\| \\
&\quad + \frac{1}{\delta_n} \left\| \sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) E\left(\boldsymbol{H}(\mathbb{X}_1^{(l)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| \\
&\quad + \max_{1 \leq g \leq \tilde{v}_n} \frac{1}{\delta_n + g} \left\| \sum_{i=k_{j+s,n}+1}^{k_{j+s}+g} \boldsymbol{H}_0(\mathbb{X}_i^{(j+s+1)}, \widetilde{\boldsymbol{\theta}}) \right\| + \frac{\tilde{v}_n}{\delta_n} \left\| E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+s+1)}, \widetilde{\boldsymbol{\theta}}) \right) \right\| \\
&= O_P\left(\max\left\{ \frac{1}{\sqrt{n}}, \frac{\bar{v}_n}{n}, \frac{\tilde{v}_n}{n} \right\} \right).
\end{aligned}$$

Moreover, by Assumptions (M1) and (M5) the Uniform Ergodic Theorem in Corollary E.2.7 and Lemma 5.4.7 can be used to obtain

$$\begin{aligned}
&\Bigg\| \frac{1}{\delta_n + h + g} \sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g} \nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\eta}_{h,g,n}) \\
&\qquad - \frac{1}{\delta_n + h + g} \sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(l)}, \boldsymbol{\eta}_{h,g,n})\right) \Bigg\|_F \qquad (5.13)\\
&= \frac{1}{\delta_n + h + g} \Bigg\| \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_{h,g,n}) + \sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(l)}, \boldsymbol{\eta}_{h,g,n}) \\
&\qquad\qquad + \sum_{i=k_{j+s,n}+1}^{k_{j+s,n}+g} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j+s+1)}, \boldsymbol{\eta}_{h,g,n}) \Bigg\|_F \\
&\le \max_{1 \le h \le \bar{v}_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{\delta_n + h} \Bigg\| \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \Bigg\|_F + \frac{\bar{v}_n}{\delta_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right) \right\|_F \\
&\quad + \sum_{l=j+1}^{j+s} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{k_{l,n} - k_{l-1,n}} \Bigg\| \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(l)}, \boldsymbol{\theta}) \Bigg\|_F \\
&\quad + \max_{1 \le g \le \tilde{v}_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{\delta_n + g} \Bigg\| \sum_{i=k_{j+s,n}+1}^{k_{j+s}+g} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j+s+1)}, \boldsymbol{\theta}) \Bigg\|_F + \frac{\tilde{v}_n}{\delta_n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j+s+1)}, \boldsymbol{\theta})\right) \right\|_F \\
&= o_P(1),
\end{aligned}$$

where the last line follows from Assumption (M5) and since $\bar{v}_n = o(n)$ and $\tilde{v}_n = o(n)$. Furthermore, by considering the Taylor expansion in (5.10) the statements in (5.12) and (5.13) together with

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Bigg\| \sum_{l=j+1}^{j+s} \left((k_{l,n} - k_{l-1,n}) - n(\lambda_l - \lambda_{l-1})\right) E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(l)}, \boldsymbol{\theta})\right) \Bigg\|_F \\
&\le 2 \sum_{l=j+1}^{j+s} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(l)}, \boldsymbol{\theta})\right) \right\|_F = O(1)
\end{aligned}$$

can be combined to

$$\begin{aligned}
&O_P\left(\max\left\{\frac{1}{\sqrt{n}}, \frac{\bar{v}_n}{n}, \frac{\tilde{v}_n}{n}\right\}\right) \qquad (5.14)\\
&= \left(o_P(1) + \frac{n}{\delta_n + h + g} \sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(l)}, \boldsymbol{\eta}_{h,g,n})\right)\right)^T \qquad (5.15)\\
&\qquad\qquad \left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1, k_{j+s,n}+g} - \widetilde{\boldsymbol{\theta}}\right) \quad \text{uniformly in } h \text{ and } g.
\end{aligned}$$

With $\frac{\delta_n+\bar{v}_n+\tilde{v}_n}{n}=O(1)$ and Assumption (M9) we get

$$\frac{\delta_n+h+g}{n}\left\|\left(\sum_{l=j+1}^{j+s}(\lambda_l-\lambda_{l-1})E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(l)},\boldsymbol{\eta}_{h,g,n})\right)\right)^{-1}\right\|_F$$
$$\leq\frac{\delta_n+\bar{v}_n+\tilde{v}_n}{n}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\left(\sum_{l=j+1}^{j+s}(\lambda_l-\lambda_{l-1})E\left(\nabla\boldsymbol{H}(\mathbb{X}_1^{(l)},\boldsymbol{\theta})\right)\right)^{-1}\right\|_F=O(1).$$

Hence, by Lemma E.2.21 multiplying the inverse matrix to both sides of the equation in(5.14) yields

$$O_P\left(\max\left\{\frac{1}{\sqrt{n}},\frac{\bar{v}_n}{n},\frac{\tilde{v}_n}{n}\right\}\right)=(o_P(1)+\boldsymbol{I}_p)^T\left(\widehat{\boldsymbol{\theta}}_{k_{j,n}-h+1,k_{j+s,n}+g}-\widetilde{\boldsymbol{\theta}}\right)\text{ uniformly in } h \text{ and } g,$$

which completes the proof by Lemma E.2.22. □

In order to get an upper bound for $gRSS$ of an arbitrary candidate set, we need to examine the behavior of the global estimator $\widehat{\boldsymbol{\theta}}_{1,n}$ under alternative. Therefore, let $\widetilde{\boldsymbol{\theta}}$ be the unique zero of

$$\sum_{j=1}^{q+1}(\lambda_j-\lambda_{j-1})\,E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right).\tag{5.16}$$

Lemma 5.4.13. *Let the Assumptions (M1), (M5) and (M9) for $j=0$ and $s=q+1$ hold. Furthermore, let (M6) be satisfied by a sequence $\{v_n\}$ of order n. Then,*

$$\left\|\widehat{\boldsymbol{\theta}}_{1,n}-\widetilde{\boldsymbol{\theta}}\right\|=O_P\left(\frac{1}{\sqrt{n}}\right).$$

Proof. The result follows directly from Lemma 5.4.12 with $\bar{v}_n\equiv 0$ and $\tilde{v}_n\equiv 0$. □

Lemma 5.4.14. *Let the assumptions of Lemma 5.4.13 be satified. Furthermore, assume that Assumption (M10) holds. Then, there exists a constant C_1 such that*

$$gRSS(\mathcal{A}_n)\leq C_1+o_P(1)$$

holds for all sets $\mathcal{A}_n\subset\{2,\dots,n-1\}$.

Proof. By definition we know that the $gRSS$ of an arbitrary set $\mathcal{A}_n$ is less than or equal to the $gRSS$ of the empty set, which is

$$gRSS(\emptyset)=\frac{1}{n}\sum_{i=1}^{n}Q(\mathbb{X}_i,\widehat{\boldsymbol{\theta}}_{1,n})=\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}}Q(\mathbb{X}_i^{(j)},\widehat{\boldsymbol{\theta}}_{1,n}).$$

Lemma 5.4.13 shows that the estimator sequence $\{\widehat{\boldsymbol{\theta}}_{1,n}\}_{i\geq 1}$ is $\sqrt{n}$-consistent for $\widetilde{\boldsymbol{\theta}}$ as defined in (5.16). Furthermore, a second order Taylor expansion shows, that there exists an $\boldsymbol{\eta}_n$ with $\left\|\boldsymbol{\eta}_n-\widehat{\boldsymbol{\theta}}_{1,n}\right\|\leq\left\|\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_{1,n}\right\|$ such that

$$gRSS(\emptyset)=\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}}Q(\mathbb{X}_i^{(j)},\widehat{\boldsymbol{\theta}}_{1,n})\tag{5.17}$$

$$= \frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} Q(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) + \frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})^T \left(\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}\right)$$
$$+ \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}\right)^T \left(\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_n)\right)\left(\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}\right).$$

On noting that

$$\left\|\sum_{j=1}^{q+1}(k_{j,n} - k_{j-1,n})E\left(\boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right)\right\| = O(1),$$

which has been derived in Lemma 5.4.12 in (5.11) (with $s = q+1$ and $j = 0$), by Assumption (M6) we get

$$\left\|\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| \leq \left\|\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| + O_P\left(\frac{1}{n}\right)$$
$$\leq \sum_{j=1}^{q+1}\left\|\frac{1}{n}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| + O_P\left(\frac{1}{n}\right) = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Thus, in connection with the submultiplicativity of the Euclidean norm and Lemma 5.4.13 we obtain

$$\left|\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})^T \left(\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}\right)\right| \tag{5.18}$$
$$\leq \left\|\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right\| \left\|\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}\right\| = O_P\left(\frac{1}{n}\right).$$

Furthermore, with the Assumptions (M1) and (M5) the Uniform Ergodic Theorem in Corollary E.2.7 can be applied to receive

$$\left\|\frac{1}{n}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_n)\right\|_F \overset{D}{=} \left\|\frac{1}{n}\sum_{i=1}^{k_{j,n}-k_{j-1,n}} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_n)\right\|_F$$
$$\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{n}\sum_{i=1}^{k_{j,n}-k_{j-1,n}} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\|_F$$
$$\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{n}\sum_{i=1}^{k_{j,n}-k_{j-1,n}} \nabla \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \boldsymbol{\theta})\right\|_F + O_P(1) = O_P(1),$$

where the last line follows from the triangle inequality and Assumption (M5). Thus, we get

$$\left\|\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \nabla \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\eta}_n)\right\|_F = O_P(1).$$

Moreover, the $\sqrt{n}$-consistency of the estimator sequence and Lemma E.1.4 in combination with Lemma E.1.5 show

$$\begin{aligned}&\left|\left(\widehat{\boldsymbol{\theta}}_{1,n}-\widetilde{\boldsymbol{\theta}}\right)^T\left(\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}}\nabla\boldsymbol{H}(\mathbb{X}_i^{(j)},\boldsymbol{\eta}_n)\right)\left(\widehat{\boldsymbol{\theta}}_{1,n}-\widetilde{\boldsymbol{\theta}}\right)\right| \\ &\leq\left\|\widehat{\boldsymbol{\theta}}_{1,n}-\widetilde{\boldsymbol{\theta}}\right\|^2\left\|\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}}\nabla\boldsymbol{H}(\mathbb{X}_i^{(j)},\boldsymbol{\eta}_n)\right\|_F=O_P\left(\frac{1}{n}\right).\end{aligned} \tag{5.19}$$

Besides, with Assumption (M1) and as the function Q is measurable with respect to $\mathbb{X}_i^{(j)}$ the sequence $Q(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})$ is stationary and ergodic. Hence, by Assumption (M10) the Ergodic Theorem can be used to get

$$\left|\frac{1}{n}\sum_{i=k_{j-1,n}+1}^{k_{j,n}}Q_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right|\overset{D}{=}\left|\frac{1}{n}\sum_{i=1}^{k_{j,n}-k_{j-1,n}}Q_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right|=o_P(1),$$

implying that

$$\left|\frac{1}{n}\sum_{j=1}^{q+1}\sum_{i=k_{j-1,n}+1}^{k_{j,n}}Q_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right|\leq\sum_{j=1}^{q+1}\left|\frac{1}{n}\sum_{i=k_{j-1,n}+1}^{k_{j,n}}Q_0(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\right|=o_P(1).$$

Finally, by considering the Taylor expansion in (5.17) again, the statement above combined with (5.18) and (5.19) yields

$$\begin{aligned}gRSS(\emptyset)&=\sum_{j=1}^{q+1}\frac{k_{j,n}-k_{j-1,n}}{n}E\left(Q(\mathbb{X}_1^{(j)},\widetilde{\boldsymbol{\theta}})\right)+o_P(1)\\&=\sum_{j=1}^{q+1}(\lambda_j-\lambda_{j-1})E\left(Q(\mathbb{X}_1^{(j)},\widetilde{\boldsymbol{\theta}})\right)+o_P(1),\end{aligned}$$

which completes the proof as $0<\sum_{j=1}^{q+1}(\lambda_j-\lambda_{j-1})E\left(Q(\mathbb{X}_1^{(j)},\widetilde{\boldsymbol{\theta}})\right):=C_1<\infty$. □

Remark 5.4.15. *Note that we could get a rate of convergence in Lemma 5.4.14 by imposing additional assumptions on the transformed sequence* $\{Q(\mathbb{X}_i^{(j)},\widetilde{\boldsymbol{\theta}})\}$.

5.4.2. The Linear Regression Model

We consider the linear regression model introduced in Section 3.2.2 under the Assumptions (R1*) to (R7*). Furthermore, we need forward and backward Hájek-Rényi-type inequalities as given in the following:

(R8*) For all $j=1,\dots,q+1$, let the series $\{\boldsymbol{X}_i^{(j)}\varepsilon_i\}_{i\geq1}$ satisfy the following forward and backward Hájek-Rényi-type inequalities, for any $m_n\in\mathbb{N}_0$ and a positive deterministic sequence $\{v_n\}$ with $v_n\to\infty$ (which will be specified later):

$$\max_{v_n\leq k\leq n-m_n}\left\|\frac{\sqrt{v_n}}{k}\sum_{i=m_n+1}^{m_n+k}\boldsymbol{X}_i^{(j)}\varepsilon_i\right\|=O_P(1)$$

and

$$\max_{v_n \leq k \leq m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n-k+1}^{m_n} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| = O_P(1).$$

(R9*) For all $j = 1, \ldots, q+1$, let the series $\{\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\}_{i \geq 1}$ fulfill the following forward and backward Hájek-Rényi-type inequalities, for any $m_n \in \mathbb{N}_0$ and a positive deterministic sequence $\{v_n\}$ with $v_n \to \infty$ (which will be specified later):

$$\max_{v_n \leq k \leq n-m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n+1}^{m_n+k} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F = O_P(1)$$

and

$$\max_{v_n \leq k \leq m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n-k+1}^{m_n} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F = O_P(1).$$

Moreover, we introduce the following additional assumption:

(R10*) Let the matrix $\sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) \boldsymbol{C}_{(l)}$ be positive definite for some $s \geq 1$.

Note that Assumption (R10*) follows directly from Assumption (R3*) if the regressors are strictly exogenous and not effected by changes.
Now, we derive similar statements as in the previous section. The results up to Lemma 5.4.19 are mainly needed to investigate the asymptotic behavior of the estimator sequences in the subsequent lemmata.

Lemma 5.4.16. *Let the Assumptions (R1*), (R2*), (R4*) and (R6*) be satisfied. Then,*

$$\max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\| \frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| = O_P \left(\max \left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right).$$

Proof. First, note that by Assumptions (R1*), (R2*) and (R4*) we get

$$E\left(\boldsymbol{X}_i^{(j)} \varepsilon_i\right) = E\left(E\left(\boldsymbol{X}_i^{(j)} \varepsilon_i | \mathcal{F}_i\right)\right) = E\left(\boldsymbol{X}_i^{(j)} E\left(\varepsilon_i | \mathcal{F}_i\right)\right) = E\left(\boldsymbol{X}_i^{(j)} E\left(\varepsilon_i\right)\right) \quad (5.20)$$
$$= E\left(\boldsymbol{X}_i^{(j)}\right) E\left(\varepsilon_i\right) = \boldsymbol{0}.$$

Hence, with Assumption (R6*) we can apply Lemma 5.4.2, which shows the assertion. □

Lemma 5.4.17. *Let the sequence $\{\boldsymbol{X}_i^{(j)}\}_{i \geq 1}$ satisfy the Assumptions (R1*), (R3*) and (R5*). Then,*

$$\max_{\substack{h \geq v_n \\ 0 \leq l \leq n-h}} \left\| \frac{1}{h} \sum_{i=l+1}^{l+h} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F = O_P \left(\max \left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right).$$

Proof. By Assumption (R3*) we know that $\boldsymbol{C}_{(j)}$ is the expectation of $\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T}$ so that the component sequences of $\{\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\}_{i\geq 1}$ have zero expectation. Furthermore, they satisfy a strong invariance principle by Assumption (R5*). Thus, applying Lemma 5.4.2 to each component in connection with Lemma E.1.6 (b) yields the assertion. □

Lemma 5.4.18. *Let the Assumptions (R1*), (R2*) and (R4*) hold. Furthermore, let (R8*) be satisfied by $\{v_n\}$ and let $\delta_{j,n}$ as in* (5.5). *Then, for any positive deterministic sequence $\{\tilde{v}_n\}$ with $\tilde{v}_n \leq \delta_{j+1,n}$ or $\tilde{v}_n \leq \delta_{j,n}$,*

$$\max_{g\leq \tilde{v}_n} \left\| \frac{1}{v_n+g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j)}\varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{v_n}}\right)$$

and

$$\max_{g\leq \tilde{v}_n} \left\| \frac{1}{v_n+g} \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)}\varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{v_n}}\right)$$

Proof. Note that by Assumptions (R1*), (R2*) and (R4*) $E\left(\boldsymbol{X}_i^{(j)}\varepsilon_i\right) = \boldsymbol{0}$ which has been shown in (5.20). Furthermore, we know that the sequence $\{\boldsymbol{X}_i^{(j)}\varepsilon_i\}$ is stationary and ergodic so that the same arguments as in the proof of Lemma 5.4.8 (see (5.6)) can be applied here to show the assertion. □

Lemma 5.4.19. *Let the Assumptions (R1*) and (R3*) hold. Furthermore, let (R9*) be satisfied by $\{v_n\}$. Then, for any positive deterministic sequence $\{\tilde{v}_n\}$ with $\tilde{v}_n \leq \delta_{j+1,n}$ or $\tilde{v}_n \leq \delta_{j,n}$,*

$$\max_{g\leq \tilde{v}_n} \left\| \frac{1}{v_n+g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right) \right\|_F = O_P\left(\frac{1}{\sqrt{v_n}}\right)$$

and

$$\max_{g\leq \tilde{v}_n} \left\| \frac{1}{v_n+g} \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right) \right\|_F = O_P\left(\frac{1}{\sqrt{v_n}}\right)$$

Proof. As the series $\{\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\}$ is stationary and ergodic by Assumption (R1*) the same arguments as in the proof of Lemma 5.4.8 (see (5.6)) can be used again. □

In the following lemma we examine the behavior of local least-squares estimators which are computed on stationary subsamples of appropriate length.

Lemma 5.4.20. *Let the Assumptions (R1*) to (R6*) be satisfied. Then, for any positive deterministic sequence $\{v_n\}$ with $\frac{n^{1/(2+\nu)}}{v_n} \to 0$ and $v_n < \delta_{j,n}$,*

$$\max_{\substack{v_n\leq h<\delta_{j,n} \\ k_{j-1,n}\leq l\leq k_{j,n}-h}} \left\| \widehat{\boldsymbol{\beta}}_{l+1,l+h} - \boldsymbol{\beta}_j \right\| = O_P\left(\max\left\{\frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}}\right\}\right).$$

Proof. By the normal equation in (3.22) and on noting that there is no change point between $l+1$ and $l+h$ we get

$$\sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} Y_i^{(j)} = \sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \widehat{\boldsymbol{\beta}}_{l+1,l+h},$$

implying that

$$\sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} \left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j \right) = \sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \left(\widehat{\boldsymbol{\beta}}_{l+1,l+h} - \boldsymbol{\beta}_j \right),$$

which is equivalent to

$$\frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} \varepsilon_i = \frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \left(\widehat{\boldsymbol{\beta}}_{l+1,l+h} - \boldsymbol{\beta}_j \right)$$

for $k_{j-1,n} \leq l \leq l+h \leq k_{j,n}$. Hence, since $\frac{n^{1/(2+\nu)}}{v_n} \to 0$ implies that $\max\left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} = o_P(1)$ Lemma 5.4.17 yields

$$\frac{1}{h} \sum_{i=l+1}^{l+h} \boldsymbol{X}_i^{(j)} \varepsilon_i = \left(o_P(1) + \boldsymbol{C}_{(j)} \right) \left(\widehat{\boldsymbol{\beta}}_{l+1,l+h} - \boldsymbol{\beta}_j \right), \quad \text{uniformly in } l \text{ and } h \geq v_n.$$

Furthermore, by Lemma 5.4.16 we receive, uniformly in l and $h \geq v_n$,

$$O_P\left(\max\left\{ \frac{n^{1/(2+\nu)}}{v_n}, \frac{\sqrt{\log(n)}}{\sqrt{v_n}} \right\} \right) = \left(o_P(1) + \boldsymbol{C}_{(j)} \right) \left(\widehat{\boldsymbol{\beta}}_{l+1,l+h} - \boldsymbol{\beta}_j \right).$$

Thus, Lemma E.2.22 together with Condition (R3*) shows the assertion. □

The convergence rate of the result above can be improved by imposing additional restrictions as in the following lemma.

Lemma 5.4.21. *Let the Assumptions (R1*), (R2*), (R3*) and (R4*) hold. Furthermore, let (R8*) and (R9*) be satisfied by $\delta_{j,n}$ as in* (5.5)*. Then, for any positive deterministic sequences $\{\bar{v}_n\}$ and $\{\tilde{v}_n\}$ fulfilling that $\delta_{j,n} - \bar{v}_n - \tilde{v}_n > 0$ and $\frac{n}{\delta_{j,n} - \bar{v}_n - \tilde{v}_n} = O(1)$,*

$$\max_{\substack{0 \leq h \leq \bar{v}_n \\ 0 \leq g \leq \tilde{v}_n}} \left\| \widehat{\boldsymbol{\beta}}_{k_{j-1,n}+h+1, k_{j,n}-g} - \boldsymbol{\beta}_j \right\| = O_P\left(\frac{1}{\sqrt{n}} \right),$$

Proof. By the normal equation in (3.22) we know that, for $0 \leq h \leq \bar{v}_n$ and $0 \leq g \leq \tilde{v}_n$,

$$\frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \boldsymbol{X}_i^{(j)} \left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j \right) \tag{5.21}$$

$$= \frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \left(\widehat{\boldsymbol{\beta}}_{k_{j-1,n}+h+1, k_{j,n}-g} - \boldsymbol{\beta}_j \right).$$

We start with approximating the left hand side of the equation above. By Lemma 5.4.18 and Assumption (R8*) we receive

$$\begin{aligned}
&\max_{\substack{0\le h\le \bar{v}_n\\ 0\le g\le \tilde{v}_n}} \left\| \frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \boldsymbol{X}_i^{(j)} \left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j \right) \right\| = \max_{\substack{0\le h\le \bar{v}_n\\ 0\le g\le \tilde{v}_n}} \left\| \frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| \\
&= \max_{\substack{0\le h\le \bar{v}_n\\ 0\le g\le \tilde{v}_n}} \left\| \frac{1}{n} \left(\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i - \sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \boldsymbol{X}_i^{(j)} \varepsilon_i - \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i \right) \right\| \\
&\le \frac{\delta_{j,n}}{n} \left\| \frac{1}{\delta_{j,n}} \sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| + \frac{\delta_{j,n} + \bar{v}_n}{n} \max_{0\le h\le \bar{v}_n} \left\| \frac{1}{\delta_{j,n}+h} \sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| \\
&\quad + \frac{\delta_{j,n} + \tilde{v}_n}{n} \max_{0\le g\le \tilde{v}_n} \left\| \frac{1}{\delta_{j,n}+g} \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| \\
&= O_P\left(\frac{1}{\sqrt{n}} \right). \qquad (5.22)
\end{aligned}$$

Furthermore, applying Lemma 5.4.19 and Assumption (R9*) and yields

$$\begin{aligned}
&\max_{\substack{0\le h\le \bar{v}_n\\ 0\le g\le \tilde{v}_n}} \left\| \frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F \qquad (5.23) \\
&= \max_{\substack{0\le h\le \bar{v}_n\\ 0\le g\le \tilde{v}_n}} \left\| \frac{1}{n} \left(\sum_{i=k_{j-1,n}+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) - \sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right.\right. \\
&\qquad \left.\left. - \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right) \right\|_F \\
&\le \frac{\delta_{j,n}}{n} \left\| \frac{1}{\delta_{j,n}} \sum_{i=k_{j-1,n}+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F \\
&\quad + \frac{\delta_{j,n} + \bar{v}_n}{n} \max_{0\le h\le \bar{v}_n} \left\| \frac{1}{\delta_{j,n}+h} \sum_{i=k_{j-1,n}+1}^{k_{j-1,n}+h} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F \\
&\quad + \frac{\delta_{j,n} + \tilde{v}_n}{n} \max_{0\le g\le \tilde{v}_n} \left\| \frac{1}{\delta_{j,n}+g} \sum_{i=k_{j,n}-g+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F \\
&= O_P\left(\frac{1}{\sqrt{n}} \right) = o_P(1),
\end{aligned}$$

which shows that

$$\frac{1}{n} \sum_{i=k_{j-1,n}+h+1}^{k_{j,n}-g} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} = \frac{\delta_{j,n} - h - g}{n} \boldsymbol{C}_{(j)} + o_P(1) \text{ uniformly in } g \text{ and } h.$$

Hence, together with (5.21) and (5.22) we can conclude that

$$O_P\left(\frac{1}{\sqrt{n}}\right) = \left(\frac{\delta_{j,n}-h-g}{n}\boldsymbol{C}_{(j)} + o_P(1)\right)^T \left(\widehat{\boldsymbol{\beta}}_{k_{j-1,n}+h+1,k_{j,n}-g} - \boldsymbol{\beta}_j\right)$$

holds uniformly in g and h. On noting that $\frac{n}{\delta_{j,n}-h-g} \leq \frac{n}{\delta_{j,n}-\bar{v}_n-\tilde{v}_n} = O(1)$, by Lemma E.2.21 multiplying $\frac{n}{\delta_{j,n}-h-g}$ to both sides of the equation above leads to

$$O_P\left(\frac{1}{\sqrt{n}}\right) = \left(\boldsymbol{C}_{(j)} + o_P(1)\right)^T \left(\widehat{\boldsymbol{\beta}}_{k_{j-1,n}+h+1,k_{j,n}-g} - \boldsymbol{\beta}_j\right), \text{ uniformly in } g \text{ and } h.$$

Finally, Assumption (R3*) in combination with Lemma E.2.22 completes the proof. □

In the next lemma we concentrate on estimator sequences calculated on subsamples with exactly one change.

Lemma 5.4.22. *Let the Assumptions (R1*), (R2*), (R3*) and (R4*) hold. Furthermore, let (R8*) and (R9*) be satisfied by the sequence $\{v_n\}$.*

(a) Then, for any positive deterministic sequence $\{\tilde{v}_n\}$ with $\tilde{v}_n = o(v_n)$,

$$\max_{\substack{v_n \leq h < \delta_{j,n} \\ 1 \leq g \leq \tilde{v}_n}} \left\| \widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1,k_{j,n}+g} - \boldsymbol{\beta}_j \right\| = O_P\left(\max\left\{\frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n}\right\}\right)$$

and

$$\max_{\substack{v_n \leq h < \delta_{j+1,n} \\ 1 \leq g \leq \tilde{v}_n}} \left\| \widehat{\boldsymbol{\beta}}_{k_{j,n}-g+1,k_{j,n}+h} - \boldsymbol{\beta}_{j+1} \right\| = O_P\left(\max\left\{\frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n}\right\}\right).$$

(b) If Assumption (R7) holds in addition, we get the same result as in (a) for any positive deterministic sequence $\{\tilde{v}_n\}$ with $\tilde{v}_n \leq \delta_{j+1,n}$ and $\tilde{v}_n \leq \delta_{j,n}$ respectively.*

Proof. We only derive the first statement since the second statement in (a) can be proved in an analogous manner.
With the normal equation in (3.22) we get

$$\frac{1}{h+g}\sum_{i=k_{j,n}-h+1}^{k_{j,n}+g} \boldsymbol{X}_i\left(Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}_j\right) = \frac{1}{h+g}\sum_{i=k_{j,n}-h+1}^{k_{j,n}+g} \boldsymbol{X}_i\boldsymbol{X}_i^T\left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1,k_{j,n}+g} - \boldsymbol{\beta}_j\right),$$

which is equivalent to

$$\frac{1}{h+g}\left(\sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)}\varepsilon_i + \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j+1)}\varepsilon_i\right) + \frac{1}{h+g}\sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j+1)}\boldsymbol{X}_i^{(j+1)T}\left(\boldsymbol{\beta}_{j+1} - \boldsymbol{\beta}_j\right) \tag{5.24}$$

$$= \frac{1}{h+g}\left(\sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} + \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j+1)}\boldsymbol{X}_i^{(j+1)T}\right)\left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1,k_{j,n}+g} - \boldsymbol{\beta}_j\right),$$

as $Y_i = \boldsymbol{X}_i^{(j)T}\boldsymbol{\beta}_j + \varepsilon_i$ for $i = k_{j,n} - h + 1, \ldots, k_{j,n}$ and $Y_i = \boldsymbol{X}_i^{(j+1)T}\boldsymbol{\beta}_{j+1} + \varepsilon_i$ for $i = k_{j,n} + 1, \ldots, k_{j,n} + g$. Furthermore, by Lemma 5.4.19 we receive

$$\begin{aligned}
&\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \frac{1}{h+g} \left\| \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F \\
&\le \max_{g \le \tilde{v}_n} \left\| \frac{1}{v_n + g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F = O_P\left(\frac{1}{\sqrt{v_n}} \right).
\end{aligned} \tag{5.25}$$

This implies in connection with Lemma E.1.5 and the triangle inequality

$$\begin{aligned}
&\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \frac{1}{h+g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} \left(\boldsymbol{\beta}_{j+1} - \boldsymbol{\beta}_j \right) \right\| \\
&\le \left\| \boldsymbol{\beta}_{j+1} - \boldsymbol{\beta}_j \right\| \left(\max_{1 \le g \le \tilde{v}_n} \left\| \frac{1}{v_n + g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F + \frac{\tilde{v}_n}{v_n} \left\| \boldsymbol{C}_{(j+1)} \right\|_F \right) \\
&= O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right).
\end{aligned}$$

Moreover, by Lemma 5.4.18 we receive

$$\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \frac{1}{h+g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j+1)} \varepsilon_i \right\| \le \max_{g \le \tilde{v}_n} \left\| \frac{1}{v_n + g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j+1)} \varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{v_n}} \right)$$

and with the backward Hájek-Rényi-type inequality of Assumption (R8*) we get

$$\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \frac{1}{h+g} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| \le \max_{v_n \le h < \delta_{j,n}} \left\| \frac{1}{h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{v_n}} \right).$$

Thus, by considering equation (5.24) again we obtain

$$\begin{aligned}
&O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right) \\
&= \frac{1}{h+g} \left(\sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} + \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} \right) \left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\beta}_j \right),
\end{aligned} \tag{5.26}$$

uniformly in h and g.

Furthermore, applying the backward Hájek-Rényi-type inequality of Assumption (R9*) yields

$$\max_{\substack{v_n \le h < \delta_{j,n} \\ 1 \le g \le \tilde{v}_n}} \left\| \frac{1}{h+g} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|$$

$$\leq \max_{v_n \leq h < \delta_{j,n}} \left\| \frac{1}{h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F = O_P\left(\frac{1}{\sqrt{v_n}} \right) = o_P(1),$$

Hence, together with (5.25) and (5.26) we obtain

$$O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right) \tag{5.27}$$
$$= \left(o_P(1) + \frac{h}{h+g} \boldsymbol{C}_{(j)} + \frac{g}{h+g} \boldsymbol{C}_{(j+1)} \right) \left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\beta}_j \right), \text{ uniformly in } h \text{ and } g.$$

In part (a) we assume that $\tilde{v}_n = o(v_n)$ so that

$$\left\| \frac{g}{h+g} \boldsymbol{C}_{(j+1)} \right\|_F \leq \left\| \frac{\tilde{v}_n}{v_n} \boldsymbol{C}_{(j+1)} \right\|_F = o(1), \text{ uniformly in } h \text{ and } g,$$

and (5.27) simplifies to

$$O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right)$$
$$= \left(o_P(1) + \frac{h}{h+g} \boldsymbol{C}_{(j)} \right) \left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\beta}_j \right), \text{ uniformly in } h \text{ and } g.$$

Furthermore, as $\frac{h+g}{h} \leq \frac{v_n + \tilde{v}_n}{v_n} = o(1)$ holds uniformly in h and g as well and since $\boldsymbol{C}_{(j)}$, which does not depend on h and g, is invertible we can multiply both sides of the equation above by $\frac{h+g}{h} \boldsymbol{C}_{(j)}^{-1}$ while preserving the uniformity by Lemma E.2.21. Thus, we get

$$O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right) = \left(o_P(1) + \boldsymbol{I}_p \right) \left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\beta}_j \right), \text{ uniformly in } h \text{ and } g.$$

Hence, the assertion follows from Lemma E.2.22.
For proving part (b), we consider the matrix of the convex combination $\frac{h}{h+g} \boldsymbol{C}_{(j)} + \frac{g}{h+g} \boldsymbol{C}_{(j+1)}$. By Assumption (R7*) we know that the inverse of this matrix exists and is uniformly bounded from above. Hence, we can multiply both sides of equation (5.27) with the inverse and get by Lemma E.2.21

$$O_P\left(\max\left\{ \frac{1}{\sqrt{v_n}}, \frac{\tilde{v}_n}{v_n} \right\} \right) = \left(o_P(1) + \boldsymbol{I}_p \right) \left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1, k_{j,n}+g} - \boldsymbol{\beta}_j \right), \text{ uniformly in } h \text{ and } g,$$

completing the proof with Lemma E.2.22. □

The following lemma considers local least-squares estimators calculated on subsamples with more than one structural break and shows that they behave quite nicely if the start and end point of the subsample lie in some sense close to a change point.

Lemma 5.4.23. *Let the Assumptions (R1*), (R2*), (R3*), (R4*) and (R10*), for some* $s \geq 1$, *hold. Furthermore, let (R8*) and (R9*) be satisfied by the sequence* $\{v_n\}$ *of order* *n. Then, for any positive deterministic sequences* $\{\bar{v}_n\}$ *and* $\{\tilde{v}_n\}$ *with* $\bar{v}_n = o(n)$ *and* $\tilde{v}_n = o(n)$,

$$\max_{\substack{1\le h\le \bar{v}_n\\ 1\le g\le \tilde{v}_n}} \left\| \widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1,k_{j+s,n}+g} - \widetilde{\boldsymbol{\beta}}_{j+1,j+s} \right\| = O_P\left(\max\left\{ \frac{1}{\sqrt{n}}, \frac{\bar{v}_n}{n}, \frac{\tilde{v}_n}{n} \right\} \right),$$

with $\widetilde{\boldsymbol{\beta}}_{j+1,j+s} := \left(\sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) \boldsymbol{C}_{(l)} \right)^{-1} \sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) \boldsymbol{C}_{(l)} \boldsymbol{\beta}_l$.

Proof. In the proof we use the short version $\widetilde{\boldsymbol{\beta}}$ for $\widetilde{\boldsymbol{\beta}}_{j+1,j+s}$ and δ_n for $\delta_{j,j+s,n} := k_{j+s,n} - k_{j,n}$. By the normal equation in (3.22) we know that

$$\begin{aligned} &\frac{1}{h+g+\delta_n} \sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}} \right) \\ &= \frac{1}{h+g+\delta_n} \sum_{i=k_{j,n}-h+1}^{k_{j,n}+g} \boldsymbol{X}_i \boldsymbol{X}_i^T \left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1,k_{j,n}+g} - \widetilde{\boldsymbol{\beta}} \right). \end{aligned} \tag{5.28}$$

Considering the left hand side of the equation above, with $Y_i = Y_i^{(j)} = \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j + \varepsilon_i$, for $k_{j-1,n} < i \le k_{j,n}$ we receive

$$\begin{aligned} &\sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}} \right) \\ &= \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \left(Y_i^{(j)} - \boldsymbol{X}_i^{(j)T} \widetilde{\boldsymbol{\beta}} \right) + \sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \boldsymbol{X}_i^{(l)} \left(Y_i^{(l)} - \boldsymbol{X}_i^{(l)T} \widetilde{\boldsymbol{\beta}} \right) \\ &\quad + \sum_{i=k_{j+s,n}+1}^{k_{j+s,n}+g} \boldsymbol{X}_i^{(j+s+1)} \left(Y_i^{(j+s+1)} - \boldsymbol{X}_i^{(j+s+1)T} \widetilde{\boldsymbol{\beta}} \right) \\ &= \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}} \right) + \boldsymbol{X}_i^{(j)} \varepsilon_i \right) \\ &\quad + \sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \left(\boldsymbol{X}_i^{(l)} \boldsymbol{X}_i^{(l)T} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) + \boldsymbol{X}_i^{(l)} \varepsilon_i \right) \\ &\quad + \sum_{i=k_{j+s,n}+1}^{k_{j+s,n}+g} \left(\boldsymbol{X}_i^{(j+s+1)} \boldsymbol{X}_i^{(j+s+1)T} \left(\boldsymbol{\beta}_{j+s+1} - \widetilde{\boldsymbol{\beta}} \right) + \boldsymbol{X}_i^{(j+s+1)} \varepsilon_i \right). \end{aligned} \tag{5.29}$$

Moreover, since

$$\sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) \boldsymbol{C}_{(l)} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) = \boldsymbol{0}$$

holds by the definition of $\widetilde{\boldsymbol{\beta}}$ we obtain

$$\left\| E \left(\sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \boldsymbol{X}_i^{(l)} \boldsymbol{X}_i^{(l)T} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right) \right\|$$

$$
\begin{aligned}
&= \left\| \sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) \boldsymbol{C}_{(l)} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right\| \\
&= \left\| \sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) \boldsymbol{C}_{(l)} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) - n \sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1}) \boldsymbol{C}_{(l)} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right\| \\
&= \left\| \sum_{l=j+1}^{j+s} (n\lambda_l - \lfloor \lambda_l n \rfloor - (n\lambda_{l-1} - \lfloor \lambda_{l-1} n \rfloor)) \boldsymbol{C}_{(l)} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right\| \\
&\leq \sum_{l=j+1}^{j+s} \left\| \boldsymbol{C}_{(l)} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right\| \leq \max_{j<l\leq j+s} \left\| \boldsymbol{C}_{(l)} \right\|_F \sum_{l=j+1}^{j+s} \left\| \boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right\| = O(1),
\end{aligned}
$$

where the last line follows from Lemma E.1.5. Thus, on noting that $k_{l,n} - k_{l-1,n} = \lfloor \lambda_l n \rfloor - \lfloor \lambda_{l-1} n \rfloor$ is of order n, applying the Hájek-Rényi-type inequality of Assumption (R9*) together with Lemma E.1.5 yields

$$
\begin{aligned}
&\max_{\substack{1\leq h\leq \bar{v}_n \\ 1\leq g\leq \tilde{v}_n}} \left\| \frac{1}{h+g+\delta_n} \sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \boldsymbol{X}_i^{(l)} \boldsymbol{X}_i^{(l)T} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right\| \\
&\leq \frac{1}{\delta_n} \left\| \sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \left(\boldsymbol{X}_i^{(l)} \boldsymbol{X}_i^{(l)T} - \boldsymbol{C}_{(l)} \right) \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right. \\
&\qquad\qquad \left. + \sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) \boldsymbol{C}_{(l)} \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right\| \\
&\leq \sum_{l=j+1}^{j+s} \left\| \frac{1}{k_{l,n} - k_{l-1,n}} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \left(\boldsymbol{X}_i^{(l)} \boldsymbol{X}_i^{(l)T} - \boldsymbol{C}_{(l)} \right) \left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right) \right\| + O\left(\frac{1}{n} \right) \\
&\leq \sum_{l=j+1}^{j+s} \left\| \frac{1}{k_{l,n} - k_{l-1,n}} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \left(\boldsymbol{X}_i^{(l)} \boldsymbol{X}_i^{(l)T} - \boldsymbol{C}_{(l)} \right) \right\|_F \left\| \boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}} \right\| + O\left(\frac{1}{n} \right) \\
&= O_P\left(\frac{1}{\sqrt{n}} \right) + O_P\left(\frac{1}{n} \right) = O_P\left(\frac{1}{\sqrt{n}} \right).
\end{aligned}
$$

Furthermore, with the Hájek-Rényi-type inequality of Assumption (R8*) we obtain

$$
\begin{aligned}
&\max_{\substack{1\leq h\leq \bar{v}_n \\ 1\leq g\leq \tilde{v}_n}} \left\| \frac{1}{h+g+\delta_n} \sum_{l=j+1}^{j+s} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \boldsymbol{X}_i^{(l)} \varepsilon_i \right\| \\
&\leq \sum_{l=j+1}^{j+s} \left\| \frac{1}{k_{l,n} - k_{l-1,n}} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \boldsymbol{X}_i^{(l)} \varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{n}} \right).
\end{aligned}
$$

Besides, by Lemma 5.4.19 and Lemma E.1.5 we get

$$
\max_{\substack{1\leq h\leq \bar{v}_n \\ 1\leq g\leq \tilde{v}_n}} \left\| \frac{1}{h+g+\delta_n} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}} \right) \right\|
$$

$$\leq \max_{1\leq h\leq \bar{v}_n} \left\| \frac{1}{\delta_n + h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} + \boldsymbol{C}_{(j)} \right) \left(\boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}} \right) \right\|$$
$$\leq \left(\max_{1\leq h\leq \bar{v}_n} \left\| \frac{1}{\delta_n + h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F + \frac{\bar{v}_n}{\delta_n} \left\| \boldsymbol{C}_{(j)} \right\|_F \right) \left\| \boldsymbol{\beta}_j - \widetilde{\boldsymbol{\beta}} \right\|$$
$$= O_P\left(\max\left\{ \frac{1}{\sqrt{n}}, \frac{\bar{v}_n}{n} \right\} \right).$$

Similarly, we obtain

$$\max_{\substack{1\leq h\leq \bar{v}_n \\ 1\leq g\leq \tilde{v}_n}} \left\| \frac{1}{h+g+\delta_n} \sum_{i=k_{j+s,n}+1}^{k_{j+s,n}+g} \boldsymbol{X}_i^{(j+s+1)} \boldsymbol{X}_i^{(j+s+1)T} \left(\boldsymbol{\beta}_{j+s+1} - \widetilde{\boldsymbol{\beta}} \right) \right\| = O_P\left(\max\left\{ \frac{1}{\sqrt{n}}, \frac{\tilde{v}_n}{n} \right\} \right).$$

Furthermore, Lemma 5.4.18 shows

$$\max_{\substack{1\leq h\leq \bar{v}_n \\ 1\leq g\leq \tilde{v}_n}} \left\| \frac{1}{h+g+\delta_n} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| \leq \max_{1\leq h\leq \bar{v}_n} \left\| \frac{1}{\delta_n + h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{n}} \right)$$

and similarly

$$\max_{\substack{1\leq h\leq \bar{v}_n \\ 1\leq g\leq \tilde{v}_n}} \left\| \frac{1}{h+g+\delta_n} \sum_{i=k_{j+s,n}+1}^{k_{j+s,n}+g} \boldsymbol{X}_i^{(j+s+1)} \varepsilon_i \right\| = O_P\left(\frac{1}{\sqrt{n}} \right).$$

Hence, by considering the decomposition in (5.29) again we can conclude that

$$\frac{1}{h+g+\delta_n} \sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}} \right) \tag{5.30}$$
$$= O_P\left(\max\left\{ \frac{1}{\sqrt{n}}, \frac{\bar{v}_n}{n}, \frac{\tilde{v}_n}{n} \right\} \right) \text{ uniformly in } h \text{ and } g.$$

Now, we focus on the right hand side of equation (5.28). Applying Lemma 5.4.19 and the Hájek-Rényi-type inequalities of Assumption (R9*) yields

$$\max_{\substack{1\leq h\leq \bar{v}_n \\ 1\leq g\leq \tilde{v}_n}} \left\| \frac{1}{h+g+\delta_n} \left(\sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g} \boldsymbol{X}_i \boldsymbol{X}_i^T - \sum_{l=j+1}^{j+s} (k_{l,n} - k_{l-1,n}) \boldsymbol{C}_{(l)} \right) \right\|_F$$
$$\leq \sum_{l=j+1}^{j+s} \left\| \frac{1}{k_{l,n} - k_{l-1,n}} \sum_{i=k_{l-1,n}+1}^{k_{l,n}} \left(\boldsymbol{X}_i^{(l)} \boldsymbol{X}_i^{(l)T} - \boldsymbol{C}_{(l)} \right) \right\|_F + \frac{\bar{v}_n}{\delta_n} \left\| \boldsymbol{C}_{(j)} \right\|_F$$
$$+ \max_{1\leq h\leq \bar{v}_n} \left\| \frac{1}{\delta_n + h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F + \frac{\tilde{v}_n}{\delta_n} \left\| \boldsymbol{C}_{(j+s+1)} \right\|_F$$
$$+ \max_{1\leq g\leq \tilde{v}_n} \left\| \frac{1}{\delta_n + g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j+s+1)} \boldsymbol{X}_i^{(j+s+1)T} - \boldsymbol{C}_{(j+s+1)} \right) \right\|_F$$
$$= o_P(1),$$

implying that

$$\frac{1}{h+g+\delta_n}\sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g}\boldsymbol{X}_i\boldsymbol{X}_i^T$$
$$=\sum_{l=j+1}^{j+s}\frac{k_{l,n}-k_{l-1,n}}{h+g+\delta_n}\boldsymbol{C}_{(l)}+o_P(1)\ \text{ uniformly in } h \text{ and } g.$$

Furthermore, on noting that

$$\sum_{l=j+1}^{j+s}\frac{k_{l,n}-k_{l-1,n}}{h+g+\delta_n}\boldsymbol{C}_{(l)}=\frac{\delta_n}{h+g+\delta_n}\sum_{l=j+1}^{j+s}\frac{k_{l,n}-k_{l-1,n}}{\delta_n}\boldsymbol{C}_{(l)}$$
$$=\frac{\delta_n}{h+g+\delta_n}\sum_{l=j+1}^{j+s}\frac{\lfloor\lambda_l n\rfloor-\lfloor\lambda_{l-1}n\rfloor}{\lfloor\lambda_{j+s}n\rfloor-\lfloor\lambda_j n\rfloor}\boldsymbol{C}_{(l)}=(1+o(1))\left(\sum_{l=j+1}^{j+s}\frac{\lambda_l-\lambda_{l-1}}{\lambda_{j+s}-\lambda_j}\boldsymbol{C}_{(l)}+o(1)\right)$$
$$=\sum_{l=j+1}^{j+s}\frac{\lambda_l-\lambda_{l-1}}{\lambda_{j+s}-\lambda_j}\boldsymbol{C}_{(l)}+o(1)\ \text{ uniformly in } h \text{ and } g,$$

we can conclude that

$$\frac{1}{h+g+\delta_n}\sum_{i=k_{j,n}-h+1}^{k_{j+s,n}+g}\boldsymbol{X}_i\boldsymbol{X}_i^T=\sum_{l=j+1}^{j+s}\frac{\lambda_l-\lambda_{l-1}}{\lambda_{j+s}-\lambda_j}\boldsymbol{C}_{(l)}+o_P(1)\ \text{ uniformly in } h \text{ and } g. \tag{5.31}$$

Hence, in combination with (5.28) and (5.30) we receive

$$O_P\left(\max\left\{\frac{1}{\sqrt{n}},\frac{\bar{v}_n}{n},\frac{\tilde{v}_n}{n}\right\}\right)$$
$$=\left(o_P(1)+\sum_{l=j+1}^{j+s}\frac{\lambda_l-\lambda_{l-1}}{\lambda_{j+s}-\lambda_j}\boldsymbol{C}_{(l)}\right)\left(\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1,k_{j,n}+g}-\widetilde{\boldsymbol{\beta}}\right)\ \text{ uniformly in } h \text{ and } g.$$

Finally, Lemma E.2.22 together with Assumption (R10*) completes the proof. □

Similar to the general parameter change model, we show in Lemma 5.4.25 that the *gRSS* of an arbitrary candidate set is bounded from above. In order to prove this result, we have to investigate the behavior of the global estimator $\widehat{\boldsymbol{\beta}}_{1,n}$ under the alternative which is done in the following lemma. Therefore, let $\widetilde{\boldsymbol{\beta}}$ be the unique zero of $\sum_{j=1}^{q+1}(\lambda_j-\lambda_{j-1})\,E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)},\boldsymbol{\theta})\right)$, which is given by

$$\widetilde{\boldsymbol{\beta}}=\left(\sum_{j=1}^{q+1}(\lambda_j-\lambda_{j-1})\,\boldsymbol{C}_{(j)}\right)^{-1}\sum_{j=1}^{q+1}(\lambda_j-\lambda_{j-1})\,\boldsymbol{C}_{(j)}\boldsymbol{\beta}_j. \tag{5.32}$$

Lemma 5.4.24. *Let the Assumptions (R1*), (R2*), (R3*), (R4*) and (R10*) for $j=0$ and $s=q+1$ hold. Furthermore, let (R8*) and (R9*) be satisfied by a sequence $\{v_n\}$ of order n. Then,*

$$\left\|\widehat{\boldsymbol{\beta}}_{1,n}-\widetilde{\boldsymbol{\beta}}\right\|=O_P\left(\frac{1}{\sqrt{n}}\right).$$

Proof. The result follows directly from Lemma 5.4.23 with $\bar{v}_n \equiv 0$ and $\tilde{v}_n \equiv 0$ and with $s = q + 1$ and $j = 0$. □

Lemma 5.4.25. *Let the assumptions of Lemma 5.4.24 hold. Then, there exists a constant C_1 such that*

$$gRSS(\mathcal{A}_n) \leq C_1 + o_P(1)$$

holds for all sets $\mathcal{A}_n \subset \{2, \dots, n-1\}$.

Proof. The $gRSS$ of the empty set gives an upper bound for the $gRSS$ of any candidate set and is given by

$$gRSS(\emptyset) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{1,n} \right)^2 .$$

A second order Taylor expansion about $\widetilde{\boldsymbol{\beta}}$, as defined in (5.32), yields

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{1,n} \right)^2 \qquad (5.33) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}} \right)^2 - 2 \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}} \right)^T \left(\widehat{\boldsymbol{\beta}}_{1,n} - \widetilde{\boldsymbol{\beta}} \right) \\ &+ \left(\widehat{\boldsymbol{\beta}}_{1,n} - \widetilde{\boldsymbol{\beta}} \right)^T \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T \right) \left(\widehat{\boldsymbol{\beta}}_{1,n} - \widetilde{\boldsymbol{\beta}} \right) . \end{aligned}$$

Furthermore, the following auxiliary results have already been derived in the proof of Lemma 5.4.23 (with $j = 0$, $s = q + 1$, $\bar{v}_n \equiv 0$ and $\tilde{v}_n \equiv 0$). By (5.30) we know

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i^T \widetilde{\boldsymbol{\beta}} \right) = O_P \left(\frac{1}{\sqrt{n}} \right)$$

and with (5.31) we have

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T = \sum_{l=1}^{q+1} (\lambda_l - \lambda_{l-1}) \boldsymbol{C}_{(l)} + o_P(1) .$$

Thus, together with the $\sqrt{n}$-consistency of the estimator sequence, shown in Lemma 5.4.24, and Lemma E.1.5 we obtain

$$\begin{aligned} &\left| \left(\widehat{\boldsymbol{\beta}}_{1,n} - \widetilde{\boldsymbol{\beta}} \right)^T \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T \right) \left(\widehat{\boldsymbol{\beta}}_{1,n} - \widetilde{\boldsymbol{\beta}} \right) \right| \leq \left\| \widehat{\boldsymbol{\beta}}_{1,n} - \widetilde{\boldsymbol{\beta}} \right\|^2 \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T \right\|_F \\ &= O_P \left(\frac{1}{n} \right) \left(\sum_{l=1}^{q+1} (\lambda_l - \lambda_{l-1}) \boldsymbol{C}_{(l)} + o_P(1) \right) = O_P \left(\frac{1}{n} \right) . \end{aligned}$$

Moreover, Lemma 5.4.24 in combination with the submultiplicativity of the Euclidean norm leads to

$$\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}\right)^T\left(\widehat{\boldsymbol{\beta}}_{1,n}-\widetilde{\boldsymbol{\beta}}\right)\right|\leq\left\|\widehat{\boldsymbol{\beta}}_{1,n}-\widetilde{\boldsymbol{\beta}}\right\|\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}\right)\right\|$$
$$=O_P\left(\frac{1}{n}\right).$$

By considering (5.33) we can conclude that

$$gRSS(\emptyset)=\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{1,n}\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}\right)^2+O_P\left(\frac{1}{n}\right). \tag{5.34}$$

Furthermore, we receive

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}\right)^2&=\frac{1}{n}\sum_{l=1}^{q+1}\sum_{i=k_{l-1,n}+1}^{k_{l,n}}\left(\boldsymbol{X}_i^{(l)T}\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)+\varepsilon_i\right)^2\\&=\sum_{l=1}^{q+1}\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)^T\left(\frac{1}{n}\sum_{i=k_{l-1,n}+1}^{k_{l,n}}\boldsymbol{X}_i^{(l)}\boldsymbol{X}_i^{(l)T}\right)\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)\\&\quad+2\sum_{l=1}^{q+1}\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)^T\left(\frac{1}{n}\sum_{i=k_{l-1,n}+1}^{k_{l,n}}\boldsymbol{X}_i^{(l)}\varepsilon_i\right)+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2.\end{aligned}$$

On noting that by Assumption (R9*)

$$\left\|\frac{1}{k_{l,n}-k_{l-1,n}}\sum_{i=k_{l-1,n}+1}^{k_{l,n}}\left(\boldsymbol{X}_i^{(l)}\boldsymbol{X}_i^{(l)T}-\boldsymbol{C}_{(l)}\right)\right\|_F=O_P\left(\frac{1}{\sqrt{n}}\right).$$

and $\frac{k_{l,n}-k_{l-1,n}}{n}\to(\lambda_l-\lambda_{l-1})$, the first summand in equation above can be approximated in the following way

$$\begin{aligned}&\sum_{l=1}^{q+1}\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)^T\left(\frac{1}{n}\sum_{i=k_{l-1,n}+1}^{k_{l,n}}\boldsymbol{X}_i^{(l)}\boldsymbol{X}_i^{(l)T}\right)\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)\\&=\sum_{l=1}^{q+1}\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)^T\left((\lambda_l-\lambda_{l-1})\boldsymbol{C}_{(l)}\right)\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)+O_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Moreover, by using Assumption (R8*) together with the triangle inequality and the submultiplicativity of the Euclidean norm, we receive

$$\sum_{l=1}^{q+1}\left(\boldsymbol{\beta}_l-\widetilde{\boldsymbol{\beta}}\right)^T\left(\frac{1}{n}\sum_{i=k_{l-1,n}+1}^{k_{l,n}}\boldsymbol{X}_i^{(l)}\varepsilon_i\right)=O_P\left(\frac{1}{\sqrt{n}}\right).$$

Besides, by Assumption (R4*) we know that the sequence $\{\varepsilon_i^2\}$ is i.i.d. with existing first moment $E(\varepsilon_i^2) = \sigma^2$ so that the Law of Large Numbers shows

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = \sigma^2 + o_P(1).$$

Finally, together with (5.34) the results above can be summarized to

$$gRSS(\emptyset) = \sum_{l=1}^{q+1}\left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}}\right)^T\left((\lambda_l - \lambda_{l-1})\boldsymbol{C}_{(l)}\right)\left(\boldsymbol{\beta}_l - \widetilde{\boldsymbol{\beta}}\right) + \sigma^2 + o_P(1).$$

Consequently, we get that there exists a constant C_1 such that

$$gRSS(\mathcal{A}_n) \le gRSS(\emptyset) \le C_1 + o_P(1).$$

□

Remark 5.4.26. *Note that we could get a rate of convergence in Lemma 5.4.25 if we use stronger moment conditions for the error sequence.*

5.5. Estimating the Number of Change Points - A First Result

Throughout this section, we assume that for the linear regression model the Assumptions (R1*) to (R7*) and (R10*) are fufilled and that the Assumptions (R8*) and (R9*) are satisfied by any sequence $\{v_n\}$ with $v_n \le n$ and $\frac{n}{v_n} = O(1)$. For the general parameter change model, let the Assumptions (M1) to (M10) hold and let (M6) be satisfied by any sequence $\{v_n\}$ with $v_n \le n$ and $\frac{n}{v_n} = O(1)$.
Similar to Cho & Kirch (2018) we use an algorithm which returns a set of final candidate sets satifying the following conditions with $sBIC$ as defined in (5.1):

(C1) Adding further candidates to the set monotonically increases $sBIC$.

(C2) Removing any single candidate from the set increases $sBIC$.

However, in contrast to Cho & Kirch (2018) we omit the pruning step and perform an exhaustive search on the whole set of change point candidates obtained from the MOSUM Wald-type procedure with different bandwidths or the MOSUM score-type procedure with several bandwidths and/or different global estimators. Furthermore, we take the cardinality of the output, which is defined as the minimal cardinality among its final candidate sets, as an estimator for the number of changes and we state a first result which will be the basis for proving consistency in the future.
In doing so, we basically consider specific candidate sets $\mathcal{A}_n$ and we want to know how the $gRSS$, defined in (5.2) and (5.4), changes if we add a candidate to such a set: $gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})$, where $\tilde{l}_n$ represents the added candidate. Furthermore, let $l_{0,n}$ and $l_{1,n}$ denote the elements of the set $\mathcal{A}_n$ lying closest to the left and to the right of $\tilde{l}_n$ such that $l_{0,n} < \tilde{l}_n < l_{1,n}$. We use the following definitions and notation:

- Let $\mathcal{G}$ be the set of bandwidths, which are used in the MOSUM procedure to produce the initial candidates for the algorithm.
- Let $G \in \mathcal{G}$ denote a bandwidth satisfying Assumption A.1.1.
- Let $\mathcal{L} = \mathcal{L}(\mathcal{G})$ be the set of initial candidates obtained from the bandwidths of $\mathcal{G}$.
- A candidate (estimate) l_n is called valid for a change point $k_{j,n}$ if $|k_{j,n} - l_n| < u_n$ with

$$u_n := \min_{1 \le j \le q+1} |k_{j,n} - k_{j-1,n}|/2, \tag{5.35}$$

  which is half of the minimal distance between two adjacent structural breaks. The set of valid candidates of a change point $k_{j,n}$ is denoted by:

$$\mathcal{V}_{j,n} = \{l_n \in \mathcal{L} : |k_{j,n} - l_n| < u_n\}. \tag{5.36}$$

- A candidate (estimate) l_n is called strictly valid for a change point $k_{j,n}$ if $|k_{j,n} - l_n| < G$. The set of strictly valid candidates of a change point $k_{j,n}$ is denoted by:

$$\mathcal{V}^*_{j,n} = \{l_n \in \mathcal{L} : |k_{j,n} - l_n| < G\}. \tag{5.37}$$

- A candidate l_n is called invalid if $l_n \notin \mathcal{V}_{j,n}$ holds for all $j = 1, \ldots, q$.

Furthermore, the relationship between likelihood-ratio and Wald-type statistic plays an important role and needs to be examined in detail. Note that in the linear regression model these statistics are equivalent to each other, as we will see later, whereas in the general model this only holds asymptotically.
The difference in $gRRS$ of two sets $\mathcal{A}_n = \{l_{1,n}, \ldots, l_{m,n}\}$ and $\mathcal{A}_n \cup \{\tilde{l}_n\}$

$$\begin{aligned} &n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right) \\ &= \sum_{i=l_{0,n}+1}^{l_{1,n}} Q(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_{l_{0,n}+1,l_{1,n}}) - \sum_{i=l_{0,n}+1}^{\tilde{l}_n} Q(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_{l_{0,n}+1,\tilde{l}_n}) - \sum_{i=\tilde{l}_n+1}^{l_{1,n}} Q(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_{\tilde{l}_n+1,l_{1,n}}) \end{aligned}$$

can be regarded as a generalization or a non-parametric version of the likelihood-ratio statistic where a general criterion function Q is used instead of the logarithm of a specific probability density function. For further information on the classical likelihood approach in change point analysis we refer to Csörgö & Horváth (1997), Chapter 1.
For the linear regression model we get the following likelihood-ratio statistic

$$\begin{aligned} &n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right) \\ &= \sum_{i=l_{0,n}+1}^{l_{1,n}} \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{l_{0,n}+1,l_{1,n}})\right)^2 - \sum_{i=l_{0,n}+1}^{\tilde{l}_n} \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n}\right)^2 \\ &\quad - \sum_{i=\tilde{l}_n+1}^{l_{1,n}} \left(Y_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right)^2, \end{aligned}$$

being equal to a Wald-type statistic given by

$$\left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n}-\widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right)^T \boldsymbol{C}_{l_{0,n}+1,\tilde{l}_n}\boldsymbol{C}^{-1}_{l_{0,n}+1,l_{1,n}}\boldsymbol{C}_{\tilde{l}_n+1,l_{1,n}}\left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n}-\widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right) \quad (5.38)$$

$$\text{with} \quad \boldsymbol{C}_{l,u} := \sum_{i=l}^{u} \boldsymbol{X}_i\boldsymbol{X}_i^T,$$

which was shown by Csörgö & Horváth (1997) on page 226f. After some further long calculations we also obtain that this equivalent to

$$\left(\sum_{i=l_{0,n}+1}^{\tilde{l}_n}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,l_{1,n}}\right)\right)^T\boldsymbol{C}^{-1}_{l_{0,n}+1,\tilde{l}_n}\boldsymbol{C}_{l_{0,n}+1,l_{1,n}}$$
$$\boldsymbol{C}^{-1}_{\tilde{l}_n+1,l_{1,n}}\left(\sum_{i=l_{0,n}+1}^{\tilde{l}_n}\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,l_{1,n}}\right)\right),$$

which represents a score-type statistic. Understanding the connection between the score-type statistic and the Wald-type or the likelihood-ratio statistic will be important as well for proving consistency later.
The relationship between the Wald-type and the likelihood-ratio statistic in the linear regression model can be used to derive the following result which gives a modified asymptotic Wald-type representation of the difference in $gRSS$ for specific settings.

Lemma 5.5.1. *Let the bandwidth G satisfy Assumption A.1.1. Then,*

$$\left(n\left(gRSS(\mathcal{A}_n)-gRSS(\mathcal{A}_n\cup\{\tilde{l}_n\})\right)\right)^{1/2}$$
$$=\sqrt{\frac{(\tilde{l}_n-l_{0,n})(l_{1,n}-\tilde{l}_n)}{l_{1,n}-l_{0,n}}}$$
$$\left\|\left(\left(\frac{k_{j,n}-l_{0,n}}{l_{1,n}-l_{0,n}}\boldsymbol{C}^{-1}_{(j+1)}+\frac{l_{1,n}-k_{j,n}}{l_{1,n}-l_{0,n}}\boldsymbol{C}^{-1}_{(j)}\right)^{-1}+o_P(1)\right)^{1/2}\left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n}-\widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right)\right\|$$

holds uniformly in $l_{0,n}\in(k_{j-1,n}-G,k_{j,n}-u_n]$, $\tilde{l}_n\in(k_{j,n}-G,k_{j,n}+G)$ and $l_{1,n}\in(k_{j,n}+u_n,k_{j+1,n}+G)$.

Proof. We only prove the assertion for $\tilde{l}_n>k_{j,n}$ and note that the proof is similar for $\tilde{l}_n\le k_{j,n}$ as all the results of Section 5.4.2, which are applied here, are stated in a forward and backward way. Furthermore, we distinguish between four cases:

(i) $k_{j-1,n}\le l_{0,n}<k_{j,n}<\tilde{l}_n<l_{1,n}\le k_{j+1,n}$,

(ii) $l_{0,n}<k_{j-1,n}<k_{j,n}<\tilde{l}_n<k_{j+1,n}<l_{1,n}$,

(iii) $l_{0,n}<k_{j-1,n}<k_{j,n}<\tilde{l}_n<l_{1,n}\le k_{j+1,n}$ and

(iv) $k_{j-1,n}\le l_{0,n}<k_{j,n}<\tilde{l}_n<k_{j+1,n}<l_{1,n}$.

First note that in each scenario the distances between the candidates, i.e. $\tilde{l}_n - l_{0,n}$ and $l_{1,n} - \tilde{l}_n$, are greater than $u_n - G$ which is of order n by Assumption A.1.1.
We start with scenario (i). With the forward and backward Hájek-Rényi-type inequalities of Assumption (R9*) (with $v_n = u_n - G$ so that $n = O(v_n)$ by Assumption A.1.1) we obtain with $\delta_{j,n}$ as in (5.5)

$$\begin{aligned}
&\left\| \frac{1}{l_{1,n} - l_{0,n}} \left(\sum_{i=l_{0,n}+1}^{l_{1,n}} \boldsymbol{X}_i \boldsymbol{X}_i^T - (k_{j,n} - l_{0,n}) \boldsymbol{C}_{(j)} - (l_{1,n} - k_{j,n}) \boldsymbol{C}_{(j+1)} \right) \right\|_F \\
&= \left\| \frac{1}{l_{1,n} - l_{0,n}} \left(\sum_{i=l_{0,n}+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) + \sum_{i=k_{j,n}+1}^{l_{0,n}} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right) \right\|_F \\
&\leq \max_{u_n - G \leq h \leq \delta_{j,n}} \left\| \frac{1}{h} \sum_{i=k_{j,n}-h}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F \\
&\quad + \max_{u_n - G \leq g \leq \delta_{j+1,n}} \left\| \frac{1}{g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F \\
&= O_P\left(\frac{1}{\sqrt{n}} \right), \text{ uniformly in } l_{0,n} \text{ and } l_{1,n} \text{ of case (i).}
\end{aligned}$$

Moreover, by Assumption (R7*) we know that the matrix $\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)}$ is positive definite and that its inverse is uniformly bounded from above. Thus, Lemma E.2.21 in connection with Corollary E.2.20 shows

$$\begin{aligned}
&\left(\frac{1}{l_{1,n} - l_{0,n}} \sum_{i=l_{0,n}+1}^{l_{1,n}} \boldsymbol{X}_i \boldsymbol{X}_i^T \right)^{-1} \qquad (5.39) \\
&= \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)} + o_P(1) \right)^{-1} \\
&= \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)} \right)^{-1} (\boldsymbol{I}_p + o_P(1))^{-1} \\
&= \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)} \right)^{-1} (\boldsymbol{I}_p + o_P(1)) \\
&= \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)} \right)^{-1} + o_P(1),
\end{aligned}$$

uniformly in $l_{0,n}$ and $l_{1,n}$ of case (i). Furthermore, by (5.23) with $\bar{v}_n = G$ and $\tilde{v}_n = \delta_{j+1,n} - u_n$ and on noting that $\delta_{j+1,n} - G - (\delta_{j+1,n} - u_n) = u_n - G > 0$ is of order n, we obtain

$$\left\| \frac{1}{l_{1,n} - \tilde{l}_n} \sum_{i=\tilde{l}_n+1}^{l_{1,n}} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F = O_P\left(\frac{1}{\sqrt{n}} \right) = o_P(1)$$

holding uniformly in $\tilde{l}_n$ and $l_{1,n}$ of case (i). Besides, applying Lemma 5.4.19 and the Hájek-Rényi-type inequalities of Assumption (R9*) (with $v_n = u_n$) yields

$$\begin{aligned}
&\left\|\frac{1}{\tilde{l}_n - l_{0,n}} \sum_{i=l_{0,n}+1}^{\tilde{l}_n} \left(\boldsymbol{X}_i \boldsymbol{X}_i^T - \boldsymbol{C}_{(j)}\right)\right\|_F \\
&= \left\|\frac{1}{\tilde{l}_n - l_{0,n}} \left(\sum_{i=l_{0,n}+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} + \sum_{i=k_{j,n}+1}^{\tilde{l}_n} \boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T}\right) - \boldsymbol{C}_{(j)}\right\|_F \\
&\leq \frac{\delta_{j,n}}{u_n - G} \max_{u_n \leq h \leq \delta_{j,n}} \left\|\frac{1}{h} \sum_{i=k_{j,n}-h+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right)\right\|_F \\
&\quad + \max_{1 \leq g \leq G} \left\|\frac{1}{u_n + g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)}\right)\right\|_F + \frac{G}{u_n - G} \left\|\boldsymbol{C}_{(j+1)} - \boldsymbol{C}_{(j)}\right\|_F \\
&= o_P(1), \text{ uniformly in } l_{0,n} \text{ and } \tilde{l}_n \text{ of case (i),}
\end{aligned}$$

as $\frac{\delta_{j,n}}{u_n - G} = O(1)$ and $\frac{G}{u_n - G} = o(1)$ by Assumption A.1.1. Hence, with (5.39), Assumption (R7*) and Lemma E.2.21 we can conclude

$$\begin{aligned}
&\frac{l_{1,n} - l_{0,n}}{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)} \boldsymbol{C}_{l_{0,n}+1,\tilde{l}_n} \boldsymbol{C}^{-1}_{l_{0,n}+1,l_{1,n}} \boldsymbol{C}_{\tilde{l}_n+1,l_{1,n}} \qquad (5.40)\\
&= \left(\frac{1}{\tilde{l}_n - l_{0,n}} \sum_{i=l_{0,n}+1}^{\tilde{l}_n} \boldsymbol{X}_i \boldsymbol{X}_i^T\right) \left(\frac{1}{l_{1,n} - l_{0,n}} \sum_{i=l_{0,n}+1}^{l_{1,n}} \boldsymbol{X}_i \boldsymbol{X}_i^T\right)^{-1} \left(\frac{1}{l_{1,n} - \tilde{l}_n} \sum_{i=\tilde{l}_n+1}^{l_{1,n}} \boldsymbol{X}_i \boldsymbol{X}_i^T\right) \\
&= \left(\boldsymbol{C}_{(j)} + o_P(1)\right) \left(\left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)}\right)^{-1} + o_P(1)\right) \left(\boldsymbol{C}_{(j+1)} + o_P(1)\right) \\
&= \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}^{-1}_{(j+1)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}^{-1}_{(j)}\right)^{-1} + o_P(1),
\end{aligned}$$

uniformly in $\tilde{l}_n, l_{0,n}$ and $l_{1,n}$ of case (i).

Thus, with (5.38) we obtain

$$\begin{aligned}
&n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right) \\
&= \left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right)^T \boldsymbol{C}_{l_{0,n}+1,\tilde{l}_n} \boldsymbol{C}^{-1}_{l_{0,n}+1,l_{1,n}} \boldsymbol{C}_{\tilde{l}_n+1,l_{1,n}} \left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right) \\
&= \frac{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)}{l_{1,n} - l_{0,n}} \\
&\quad \left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right)^T \left(\left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}^{-1}_{(j+1)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}^{-1}_{(j)}\right)^{-1} + o_P(1)\right) \\
&\qquad \left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right)
\end{aligned}$$

$$= \frac{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)}{l_{1,n} - l_{0,n}}$$
$$\left\| \left(\left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)}^{-1} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)}^{-1} \right)^{-1} + o_P(1) \right)^{1/2} \left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}} \right) \right\|_F^2,$$

holding uniformly in $\tilde{l}_n, l_{0,n}, l_{1,n}$ of case (i). This shows the assertion for case (i) and we can continue with case (ii).
By considering case (ii), with the Hájek-Rényi-type inequalities of Assumption (R9*) (with $v_n = \delta_{j,n}$) and Lemma 5.4.19 we receive

$$\left\| \frac{1}{l_{1,n} - l_{0,n}} \left(\sum_{i=l_{0,n}+1}^{l_{1,n}} \boldsymbol{X}_i \boldsymbol{X}_i^T - (k_{j,n} - l_{0,n}) \boldsymbol{C}_{(j)} - (l_{1,n} - k_{j,n}) \boldsymbol{C}_{(j+1)} \right) \right\|_F$$
$$= \left\| \frac{1}{l_{1,n} - l_{0,n}} \left(\sum_{i=l_{0,n}+1}^{k_{j-1,n}} \left(\boldsymbol{X}_i^{(j-1)} \boldsymbol{X}_i^{(j-1)^T} - \boldsymbol{C}_{(j-1)} \right) + \sum_{i=k_{j-1,n}+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)^T} - \boldsymbol{C}_{(j)} \right) \right. \right.$$
$$+ \sum_{i=k_{j,n}+1}^{k_{j+1,n}} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)^T} - \boldsymbol{C}_{(j+1)} \right) + \sum_{i=k_{j+1,n}+1}^{l_{0,n}} \left(\boldsymbol{X}_i^{(j+2)} \boldsymbol{X}_i^{(j+2)^T} - \boldsymbol{C}_{(j+2)} \right)$$
$$\left. \left. + (k_{j-1,n} - l_{0,n}) \left(\boldsymbol{C}_{(j-1)} - \boldsymbol{C}_{(j)} \right) + (l_{0,n} - k_{j+1,n}) \left(\boldsymbol{C}_{(j+2)} - \boldsymbol{C}_{(j+1)} \right) \right) \right\|_F$$
$$\leq \max_{1 \leq h \leq G} \left\| \frac{1}{\delta_{j,n} + h} \sum_{i=k_{j-1,n}-h}^{k_{j-1,n}} \left(\boldsymbol{X}_i^{(j-1)} \boldsymbol{X}_i^{(j-1)T} - \boldsymbol{C}_{(j-1)} \right) \right\|_F + \frac{G}{\delta_{j,n}} \left\| \boldsymbol{C}_{(j-1)} - \boldsymbol{C}_{(j)} \right\|_F$$
$$+ \max_{1 \leq g \leq G} \left\| \frac{1}{\delta_{j,n} + g} \sum_{i=k_{j+1,n}+1}^{k_{j+1,n}+g} \left(\boldsymbol{X}_i^{(j+2)} \boldsymbol{X}_i^{(j+2)T} - \boldsymbol{C}_{(j+2)} \right) \right\|_F + \frac{G}{\delta_{j,n}} \left\| \boldsymbol{C}_{(j+2)} - \boldsymbol{C}_{(j+1)} \right\|_F$$
$$+ \left\| \frac{1}{\delta_{j,n}} \sum_{i=k_{j-1,n}+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F$$
$$+ \left\| \frac{1}{\delta_{j+1,n}} \sum_{i=k_{j,n}+1}^{k_{j+1,n}} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F$$
$$= O_P\left(\frac{1}{\sqrt{n}} \right) + O_P\left(\frac{G}{n} \right) = o_P(1), \text{ uniformly in } l_{0,n} \text{ and } l_{1,n} \text{ of case (ii)},$$

since $\frac{G}{n} = o(1)$ by Assumption A.1.1. Thus, similar to (5.39) we get

$$\left(\frac{1}{l_{1,n} - l_{0,n}} \sum_{i=l_{0,n}+1}^{l_{1,n}} \boldsymbol{X}_i \boldsymbol{X}_i^T \right)^{-1} = \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)} \right)^{-1} + o_P(1),$$

uniformly in $l_{0,n}$ and $l_{1,n}$ of case (ii). Furthermore, using Assumption (R9*) (with $v_n = \delta_{j,n}$ for the first part and $v_n = \delta_{j+1,n} - G$ for the second), Lemma 5.4.19 and

Assumption A.1.1 again shows

$$\begin{aligned}
&\left\| \frac{1}{\tilde{l}_n - l_{0,n}} \sum_{i=l_{0,n}+1}^{\tilde{l}_n} \boldsymbol{X}_i \boldsymbol{X}_i^T - \boldsymbol{C}_{(j)} \right\|_F \\
&= \left\| \frac{1}{\tilde{l}_n - l_{0,n}} \left(\sum_{i=l_{0,n}+1}^{k_{j-1,n}} \boldsymbol{X}_i^{(j-1)} \boldsymbol{X}_i^{(j-1)T} + \sum_{i=k_{j-1,n}+1}^{k_{j,n}} \boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} \right.\right. \\
&\qquad\qquad \left.\left. + \sum_{i=k_{j,n}+1}^{\tilde{l}_n} \boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} \right) - \boldsymbol{C}_{(j)} \right\|_F \\
&\leq \max_{1 \leq h \leq G} \left\| \frac{1}{\delta_{j,n} + h} \sum_{i=k_{j-1,n}-h}^{k_{j-1,n}} \left(\boldsymbol{X}_i^{(j-1)} \boldsymbol{X}_i^{(j-1)T} - \boldsymbol{C}_{(j-1)} \right) \right\|_F + \frac{G}{\delta_{j,n}} \left\| \boldsymbol{C}_{(j-1)} - \boldsymbol{C}_{(j)} \right\|_F \\
&\quad + \max_{1 \leq g \leq G} \left\| \frac{1}{\delta_{j,n} + g} \sum_{i=k_{j,n}+1}^{k_{j,n}+g} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F + \frac{G}{\delta_{j,n}} \left\| \boldsymbol{C}_{(j+1)} - \boldsymbol{C}_{(j)} \right\|_F \\
&\quad + \left\| \frac{1}{\delta_{j,n}} \sum_{i=k_{j-1,n}+1}^{k_{j,n}} \left(\boldsymbol{X}_i^{(j)} \boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)} \right) \right\|_F \\
&= O_P\left(\frac{1}{\sqrt{n}} \right) + O_P\left(\frac{G}{n} \right) = o_P(1), \text{ uniformly in } l_{0,n} \text{ and } \tilde{l}_n \text{ of case (ii)}
\end{aligned}$$

and

$$\begin{aligned}
&\left\| \frac{1}{l_{1,n} - \tilde{l}_n} \sum_{i=\tilde{l}_n+1}^{l_{1,n}} \boldsymbol{X}_i \boldsymbol{X}_i^T - \boldsymbol{C}_{(j+1)} \right\|_F \\
&= \left\| \frac{1}{l_{1,n} - \tilde{l}_n} \left(\sum_{i=\tilde{l}_n+1}^{k_{j+1,n}} \boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} + \sum_{i=k_{j+1,n}+1}^{l_{1,n}} \boldsymbol{X}_i^{(j+2)} \boldsymbol{X}_i^{(j+2)T} \right) - \boldsymbol{C}_{(j+1)} \right\|_F \\
&\leq \max_{1 \leq g \leq G} \left\| \frac{1}{\delta_{j+1,n} - G + g} \sum_{i=k_{j+1,n}+1}^{k_{j+1,n}+g} \left(\boldsymbol{X}_i^{(j+2)} \boldsymbol{X}_i^{(j+2)T} - \boldsymbol{C}_{(j+2)} \right) \right\|_F \\
&\quad + \frac{G}{\delta_{j+1,n} - G} \left\| \boldsymbol{C}_{(j+2)} - \boldsymbol{C}_{(j+1)} \right\|_F \\
&\quad + \max_{\delta_{j+1,n} - G \leq h \leq \delta_{j+1,n}} \left\| \frac{1}{h} \sum_{i=k_{j+1,n}-h}^{k_{j+1,n}} \left(\boldsymbol{X}_i^{(j+1)} \boldsymbol{X}_i^{(j+1)T} - \boldsymbol{C}_{(j+1)} \right) \right\|_F \\
&= O_P\left(\frac{1}{\sqrt{n}} \right) + O_P\left(\frac{G}{n} \right) = o_P(1), \text{ uniformly in } \tilde{l}_n \text{ and } l_{1,n} \text{ of case (ii),}
\end{aligned}$$

since $\delta_{j+1,n} - G$ is of order n. Hence, similar to (5.40) the results above can be summarized to

$$\frac{l_{1,n} - l_{0,n}}{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)} \boldsymbol{C}_{l_{0,n}+1,\tilde{l}_n} \boldsymbol{C}_{l_{0,n}+1,l_{1,n}}^{-1} \boldsymbol{C}_{\tilde{l}_n+1,l_{1,n}}$$

$$= \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)}^{-1} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)}^{-1} \right)^{-1} + o_P(1),$$

uniformly in $\tilde{l}_n, l_{0,n}$ and $l_{1,n}$ of case (ii)

and the proof can be completed in an analogous manner to case (i).
In the scenarios (iii) and (iv,) the arguments used for (i) and (ii) can be combined to derive the assertion. □

Remark 5.5.2. *Note that the matrix* $\left(\frac{k_{j,n}-l_{0,n}}{l_{1,n}-l_{0,n}} \boldsymbol{C}_{(j+1)}^{-1} + \frac{l_{1,n}-k_{j,n}}{l_{1,n}-l_{0,n}} \boldsymbol{C}_{(j)}^{-1} \right)^{-1}$ *is positive definite. By Assumption (R7*) it can be rewritten as a product of invertible matrices*

$$\boldsymbol{C}_{(j)} \left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)} \right)^{-1} \boldsymbol{C}_{(j+1)},$$

which has already been used in (5.40)*, and is therefore invertible. Moreover, the positive definiteness of the matrix follows from the positive definiteness of its inverse which is a convex combination of positive definite matrices by Assumption (R3*).*

For proving the main result of this chapter for the general parameter change model we need a similar statement or at least an asymptotic Wald-type lower bound of the difference in $gRSS$ for scenarios described in Lemma 5.5.1. Since the Likelihood-ratio, the Wald-type and the Score-type statistic are in the general setting only asymptotically equivalent it is more complicated to get a uniform statement as in the lemma above. Investigating this asymptotic relationship would go beyond the scope of this thesis and will be a part of future work. Hence, at the moment we can only solve this problem by imposing an additional assumption on the general model.

Assumption 5.5.3. *Let $C > 0$ be a constant such that*

$$\left(n \left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\}) \right) \right)^{1/2}$$
$$\geq \sqrt{\frac{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)}{l_{1,n} - l_{0,n}}} \left(C + o_P(1) \right) \left\| \widehat{\boldsymbol{\theta}}_{l_{0,n}+1, \tilde{l}_n} - \widehat{\boldsymbol{\theta}}_{\tilde{l}_n+1, l_{1,n}} \right\|$$

holds uniformly in $l_{0,n} \in (k_{j-1,n} - G, k_{j,n} - u_n]$, $\tilde{l}_n \in (k_{j,n} - G, k_{j,n} + G)$ and $l_{1,n} \in (k_{j,n} + u_n, k_{j+1,n} + G)$.

The following lemma is essential for proving Theorem 5.5.6. It rules out that a candidate set obtained by the algorithm does not include a valid candidate for a change point while containing strictly valid candidates for the neighboring changes.

Lemma 5.5.4. *Let $\mathcal{A}_n$ be a candidate set which does not contain any valid estimate for a change point $k_{j,n}$. Furthermore, let $\mathcal{A}_n \cap \mathcal{V}_{j-1,n}^* \neq \emptyset$ and $\mathcal{A}_n \cap \mathcal{V}_{j+1,n}^* \neq \emptyset$, i.e. the set contains at least one strictly valid candidate for both of the neighboring changes. Then, adding a strictly valid candidate $\tilde{l}_n$ for $k_{j,n}$ to the set decreases the information criterion with probability tending to one:*

$$sBIC\left(\mathcal{A}_n\right) > sBIC\left(\mathcal{A}_n \cup \{\tilde{l}_n\}\right).$$

Proof. For the asymptotics we have to take into consideration that $l_{0,n}$ and $l_{1,n}$ are returned by the algorithm. If n grows we always get new results from the algorithm and we only have some information on the difference between $l_{0,n}$, $l_{1,n}$ and $\tilde{l}_n$. Hence, we treat these candidates as arbitrary while imposing assumptions on the distances. We assume that $\tilde{l}_n > k_{j,n}$ and note that the proof is similar for $\tilde{l}_n \leq k_{j,n}$ as the results from Section 5.4, which will be used here, are stated in a forward and backward manner. Furthermore, we have to distinguish between the following four scenarios which have all in common that the distances between the considered time points are of order n, in particular $\tilde{l}_n - l_{0,n} > u_n$ and $l_{1,n} - \tilde{l}_n > u_n - G$ with u_n as in (5.35).

(i) $k_{j-1,n} \leq l_{0,n} < k_{j,n} < \tilde{l}_n < l_{1,n} \leq k_{j+1,n}$: The candidates $l_{0,n}$ and $l_{1,n}$ can be (strictly) valid for $k_{j-1,n}$ and $k_{j+1,n}$, respectively, or invalid.

(ii) $l_{0,n} < k_{j-1,n} < k_{j,n} < \tilde{l}_n < l_{1,n} \leq k_{j+1,n}$: The candidate $l_{1,n}$ can be (strictly) valid for $k_{j+1,n}$ or invalid whereas $l_{0,n}$ must be strictly valid for $k_{j-1,n}$.

(iii) $l_{0,n} < k_{j-1,n} < k_{j,n} < \tilde{l}_n < k_{j+1,n} < l_{1,n}$: The candidates $l_{0,n}$ and $l_{1,n}$ are strictly valid for $k_{j-1,n}$ and $k_{j+1,n}$, respectively.

(iv) $k_{j-1,n} \leq l_{0,n} < k_{j,n} < \tilde{l}_n < k_{j+1,n} < l_{1,n}$: The candidate $l_{0,n}$ can be (strictly) valid for $k_{j-1,n}$ or invalid whereas $l_{1,n}$ must be strictly valid for $k_{j+1,n}$.

<u>Linear regression model:</u>

At first, we consider the difference of the local estimators $\left(\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\right)$ and decompose it into noise and signal. Therefore, the uniform results on the convergence of the estimators derived in Section 5.4 are needed. By considering the estimator sequence $\{\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n}\}$, for the cases (i) and (iv) Lemma 5.4.22 can be used to receive

$$\left\|\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \boldsymbol{\beta}_j\right\| \leq \max_{\substack{u_n \leq h < \delta_{j,n} \\ 1 \leq g \leq G}} \left\|\widehat{\boldsymbol{\beta}}_{k_{j,n}-h+1,k_{j,n}+g} - \boldsymbol{\beta}_j\right\| = O_P\left(\max\left\{\frac{1}{\sqrt{n}}, \frac{G}{n}\right\}\right) = o_P(1),$$

uniformly in $l_{0,n}$ and $\tilde{l}_n$ of cases (i) and (iv), where the last line follows from Assumption A.1.1. In the scenarios (ii) and (iii) applying Lemma 5.4.23 (with $\bar{v}_n = \tilde{v}_n = G$, $s = 1$ and $\widetilde{\boldsymbol{\beta}}_{j,j} = \boldsymbol{\beta}_j$) yields

$$\left\|\widehat{\boldsymbol{\beta}}_{l_{0,n}+1,\tilde{l}_n} - \boldsymbol{\beta}_j\right\| \leq \max_{\substack{1 \leq h \leq G \\ 1 \leq g \leq G}} \left\|\widehat{\boldsymbol{\beta}}_{k_{j-1,n}-h+1,k_{j,n}+g} - \boldsymbol{\beta}_j\right\| = O_P\left(\max\left\{\frac{1}{\sqrt{n}}, \frac{G}{n}\right\}\right) = o_P(1),$$

uniformly in $l_{0,n}$ and $\tilde{l}_n$ of cases (ii) and (iii). Investigating the behavior of the second estimator sequence $\{\widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}}\}$, in the cases (i) and (ii) Lemma 5.4.21 shows

$$\left\|\widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}} - \boldsymbol{\beta}_{j+1}\right\| \leq \max_{\substack{0 \leq h \leq G \\ 0 \leq g \leq \delta_{j+1,n}-u_n}} \left\|\widehat{\boldsymbol{\beta}}_{k_{j,n}+h+1,k_{j+1,n}-g} - \boldsymbol{\beta}_{j+1}\right\| = O_P\left(\frac{1}{\sqrt{n}}\right) = o_P(1),$$

uniformly in $\tilde{l}_n$ and $l_{1,n}$ of cases (i) and (ii), since $\delta_{j+1,n} - G - (\delta_{j+1,n} - u_n) = u_n - G > 0$ is of order n. Furthermore, for the scenarios (iii) and (iv), by Lemma 5.4.22 we obtain

$$\left\|\widehat{\boldsymbol{\beta}}_{\tilde{l}_n+1,l_{1,n}} - \boldsymbol{\beta}_{j+1}\right\| \leq \max_{\substack{\delta_{j+1,n}-G \leq h \leq \delta_{j+1,n} \\ 1 \leq g \leq G}} \left\|\widehat{\boldsymbol{\beta}}_{k_{j+1,n}-h+1,k_{j+1,n}+g} - \boldsymbol{\beta}_{j+1}\right\|$$

$$= O_P\left(\max\left\{\frac{1}{\sqrt{n}}, \frac{G}{n}\right\}\right) = o_P(1),$$

uniformly in $\tilde{l}_n$ and $l_{1,n}$ of cases (iii) and (iv). Thus, by applying Lemma 5.5.1, the results above can be combined to

$$n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right) \tag{5.41}$$
$$= \frac{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)}{l_{1,n} - l_{0,n}}$$
$$\left\| \left(\left(\frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j+1)}^{-1} + \frac{l_{1,n} - k_{j,n}}{l_{1,n} - l_{0,n}} \boldsymbol{C}_{(j)}^{-1} \right)^{-1} + o_P(1) \right)^{1/2} \left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) + o_P(1) \right) \right\|_F^2,$$

uniformly in $l_{0,n} \in (k_{j-1,n} - G, k_{j,n} - u_n]$, $\tilde{l}_n \in (k_{j,n} - G, k_{j,n} + G)$ and $l_{1,n} \in (k_{j,n} + u_n, k_{j+1,n} + G)$. We use the following notation to simply the expression above for further calculations. Let $\gamma_n = \gamma(l_{0,n}, l_{r+,n}, n) := \frac{k_{j,n} - l_{0,n}}{l_{1,n} - l_{0,n}}$ so that

$$\boldsymbol{M}_{\gamma_n} := \left(\gamma_n \boldsymbol{C}_{(j+1)}^{-1} + (1 - \gamma_n)\boldsymbol{C}_{(j)}^{-1}\right)^{-1}$$

represents the matrix above which is positive definite as described in Remark 5.5.2. By Proposition E.6 and Example E.7.c in Marshall *et al.* (2011) (on page 676ff) we know that the difference

$$\left(\gamma_n \boldsymbol{C}_{(j+1)} + (1 - \gamma_n)\boldsymbol{C}_{(j)}\right) - \boldsymbol{M}_{\gamma_n}$$

is a positive semi-definite matrix. On noting that the matrices $\boldsymbol{C}_{(j)}$ and $\boldsymbol{C}_{(j+1)}$ are positive definite by Assumption (R3*), this implies in combination with Lemma E.1.4 and Lemma E.1.5, for any vector $\boldsymbol{x} \in \mathbb{R}^p$,

$$\begin{aligned} 0 < \boldsymbol{x}^T \boldsymbol{M}_{\gamma_n} \boldsymbol{x} &\leq \boldsymbol{x}^T \left(\gamma_n \boldsymbol{C}_{(j+1)} + (1 - \gamma_n)\boldsymbol{C}_{(j)}\right) \boldsymbol{x} = \gamma_n \boldsymbol{x}^T \boldsymbol{C}_{(j+1)} \boldsymbol{x} + (1 - \gamma_n) \boldsymbol{x}^T \boldsymbol{C}_{(j)} \boldsymbol{x} \\ &\leq \boldsymbol{x}^T \boldsymbol{C}_{(j+1)} \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{C}_{(j)} \boldsymbol{x} = \boldsymbol{x}^T \left(\boldsymbol{C}_{(j+1)} + \boldsymbol{C}_{(j)}\right) \boldsymbol{x}, \\ &\text{uniformly in } \gamma_n \in [0, 1]. \end{aligned}$$

Thus, by Min-Max Theorem the eigenvalues of $\boldsymbol{M}_{\gamma_n}$, which are denoted by $\lambda_{\gamma_n,i}, i = 1, \ldots, p$, are bounded by the eigenvalues of $\boldsymbol{C}_{(j+1)} + \boldsymbol{C}_{(j)}$ uniformly in γ_n. Hence, on noting that $\|\boldsymbol{M}_{\gamma_n}\|_F = \sqrt{\sum_{i=1}^p \lambda_{\gamma_n,i}^2}$, we can conclude that $\|\boldsymbol{M}_{\gamma_n}\|_F$ is bounded uniformly in γ_n. Together with Lemma E.2.21 we obtain

$$\begin{aligned} &\left\| (\boldsymbol{M}_{\gamma_n} + o_P(1))^{1/2} \left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) + o_P(1) \right) \right\|_F^2 \\ &= \left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) + o_P(1) \right)^T (\boldsymbol{M}_{\gamma_n} + o_P(1)) \left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) + o_P(1) \right) \\ &= (\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1})^T \boldsymbol{M}_{\gamma_n} (\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) + o_P(1), \end{aligned}$$

holds uniformly in $l_{0,n} \in (k_{j-1,n} - G, k_{j,n} - u_n]$, $\tilde{l}_n \in (k_{j,n} - G, k_{j,n} + G)$ and $l_{1,n} \in (k_{j,n} + u_n, k_{j+1,n} + G)$. Moreover, note that the inverse of $\boldsymbol{M}_{\gamma_n}$ is given by $\gamma_n \boldsymbol{C}_{(j+1)}^{-1} + (1 - \gamma_n)\boldsymbol{C}_{(j)}^{-1}$ satisfying

$$\sup_{\gamma_n \in [0,1]} \left\| \boldsymbol{M}_{\gamma_n}^{-1} \right\|_F \leq \left\| \boldsymbol{C}_{(j+1)}^{-1} \right\|_F + \left\| \boldsymbol{C}_{(j)}^{-1} \right\|_F < \infty.$$

Thus, by Lemma E.1.10 in connection with Lemma E.1.12 we can conclude that there exists a constant $C_1 > 0$ such that

$$\begin{aligned}&\left\|(\boldsymbol{M}_{\gamma_n} + o_P(1))^{1/2}\left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) + o_P(1)\right)\right\|_F^2 \\ &= \left(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}\right)^T \boldsymbol{M}_{\gamma_n}\left(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}\right) + o_P(1) \geq C_1 + o_P(1),\end{aligned}$$

which holds uniformly in $l_{0,n} \in (k_{j-1,n} - G, k_{j,n} - u_n]$, $\tilde{l}_n \in (k_{j,n} - G, k_{j,n} + G)$. Furthermore, as the fraction $\frac{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)}{l_{1,n} - l_{0,n}}$ is of order n there exists a constant $C_2 > 0$ such that $\frac{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)}{l_{1,n} - l_{0,n}} > C_2 n$ holds for all $l_{0,n}, l_{1,n}, \tilde{l}_n$ of this setting. Hence, together with (5.41) we receive

$$n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right) \geq nC_2\left(C_1 + o_P(1)\right) = n\left(\tilde{C} + o_P(1)\right), \quad (5.42)$$

holds uniformly in $l_{0,n} \in (k_{j-1,n} - G, k_{j,n} - u_n]$, $\tilde{l}_n \in (k_{j,n} - G, k_{j,n} + G)$. Thus, we get

$$\begin{aligned}&sBIC(\mathcal{A}_n) - sBIC(\mathcal{A}_n \cup \{\tilde{l}_n\}) \\ &= \frac{n}{2}\left(\log(gRSS(\mathcal{A}_n)) - \log(gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\}))\right) - \xi_n \\ &= \frac{n}{2}\log\left(\frac{gRSS(\mathcal{A}_n)}{gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})}\right) - \xi_n \\ &\geq \frac{n}{2}\left(1 - \frac{gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})}{gRSS(\mathcal{A}_n)}\right) - \xi_n = \frac{n}{2}\frac{\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right)}{gRSS(\mathcal{A}_n)} - \xi_n \\ &\geq \frac{n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right)}{C_3 + o_P(1)} - \xi_n \geq n\left(\widetilde{C} + o_P(1)\right) - \xi_n,\end{aligned}$$

where the second last line follows from the property of the natural logarithm that $\log(x) \geq 1 - \frac{1}{x}$ for all $x > 0$ and the last line is obtained by applying Lemma 5.4.25 and (5.42). Finally, since $\xi_n = o(n)$ by (5.3), this implies

$$sBIC(\mathcal{A}_n) - sBIC(\mathcal{A}_n \cup \{\tilde{l}_n\}) \geq n\left(\widetilde{C} + o_P(1)\right),$$

completing the proof.
Note that there are two special cases. The first one is that there is no valid candidate for $k_{1,n}$ and no other candidate between 1 and $k_{j,n}$ in $\mathcal{A}_n$. Then, we set $l_{0,n} = 1$ and proceed as before. In the second case, there is no valid candidate for $k_{q,n}$ and no candidate between $k_{q,n}$ and n in the set $\mathcal{A}_n$. By setting $l_{1,n} = n$ we can use the same arguments as in usual case again.

General parameter change model:
By Assumption 5.5.3 we know that

$$\begin{aligned}&n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right) \\ &\geq \frac{(\tilde{l}_n - l_{0,n})(l_{1,n} - \tilde{l}_n)}{l_{1,n} - l_{0,n}}\left(C + o_P(1)\right)\left\|\widehat{\boldsymbol{\theta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\theta}}_{\tilde{l}_n+1,l_{1,n}}\right\|^2\end{aligned}$$

holds uniformly in $l_{0,n} \in (k_{j-1,n} - G, k_{j,n} - u_n]$, $\tilde{l}_n \in (k_{j,n} - G, k_{j,n} + G)$ and $l_{1,n} \in (k_{j,n} + u_n, k_{j+1,n} + G)$. Analogously to the linear regression, by applying Lemma 5.4.10, Lemma 5.4.11 and Lemma 5.4.12, we obtain

$$\left\| \widehat{\boldsymbol{\theta}}_{l_{0,n}+1,\tilde{l}_n} - \widehat{\boldsymbol{\theta}}_{\tilde{l}_n+1,l_{1,n}} \right\| = \|\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j+1}\| + o_P(1), \text{ uniformly in } l_{0,n}, \tilde{l}_n, l_{1,n}.$$

Thus, similar to the first part we get

$$n\left(gRSS(\mathcal{A}_n) - gRSS(\mathcal{A}_n \cup \{\tilde{l}_n\})\right) \geq n\left(\widetilde{C} + o_P(1)\right),$$

for some constant $\widetilde{C} > 0$, and the proof can be finished in an analogous manner to the linear regression. □

Furthermore, we have to verify that there is at least one valid candidate for each change point in the initial candidate set.

Lemma 5.5.5. *(a) Let $G \in \mathcal{G}$ satisfy Assumption A.1.1 and let $\mathcal{L}$ denote the set of initial candidates obtained by the MOSUM Wald-type procedure. Then, for every change point $k_{j,n}$, there exists a candidate $\widehat{k}_{j,n} \in \mathcal{L}$ such that*

$$P\left(\left|\widehat{k}_{j,n} - k_{j,n}\right| > G\right) \to 0.$$

(b) The result remains true for the MOSUM score-type procedure if we additionally ensure that all changes are detectable by choosing an appropriate set of global estimators.

Proof. The assertion in (a) follows directly from Corollary 3.1.16 and Corollary 3.2.10, respectively, whereas the statement in (b) can be derived from Corollary 2.1.10. □

The main result of this chapter is stated in the following theorem showing that a final candidate set of the algorithm's output contains at least q candidates with probability tending to one.

Theorem 5.5.6. *Let $\mathcal{A}_n$ be a final candidate set of the algorithm's output satisfying (C1) and (C2) and let the assumptions of Lemma 5.5.5 be fulfilled. Then,*

$$P\left(|\mathcal{A}_n| \geq q\right) \to 1 \quad \textit{as } n \to \infty.$$

Proof. The assertion can be proved by contradiction. Therefore, assume that $|\mathcal{A}_n| < q$. Hence, there would be at least one change point without valid candidate in $\mathcal{A}_n$. By Lemma 5.5.5 there is a strictly valid candidate for each change point in the initial candidate set. Thus, by adding strictly valid candidates to the set $\mathcal{A}_n$ we can create a new set $\widetilde{\mathcal{A}}_n$ such that there is exactly one change point $k_{j^*,n}$ without valid estimate in $\widetilde{\mathcal{A}}_n$ while the set contains strictly valid candidates for its neighboring change points. By Lemma 5.5.4 we know that adding a strictly valid candidate for $k_{j^*,n}$ to the set $\widetilde{\mathcal{A}}_n$ decreases the information criterion. Hence, the candidate set $\widetilde{\mathcal{A}}_n$ does not fulfill Condition (C1) implying that $\mathcal{A}_n$ cannot satisfy Condition (C1) as $\mathcal{A}_n \subset \widetilde{\mathcal{A}}_n$. This would contradict the assumption that $\mathcal{A}_n$ is in the output of the algorithm completing the proof.

□

5.6. Outlook

For justifying the usage of the procedure theoretically we need to show consistency for the estimators of the number and the locations of the changes obtained by the multiscale method. As a part of future work, this will be based on the theoretical results of Section 5.4 and Theorem 5.5.6. Moreover, it will be necessary to implement the localised pruning approach like Cho & Kirch (2018) to make the procedure competitive in terms of computation time compared to other detection algorithms. Furthermore, simulation studies for different change point problems will be conducted in order to assess the performance of the method empirically. Simulations in the classical mean change model done by Cho & Kirch (2018) have shown that the multiscale MOSUM procedure with localised pruning, which is already implemented in the mosum R-package, perfoms quite well in comparative studies. We would expect to get similar results for the linear regression model or examples of the general parameter change model.

A. Assumptions of Chapter 2

A.1. Assumptions Under the Null Hypothesis

Assumption A.1.1. *Let the bandwidth G depend on n, i.e. $G = G(n)$. Furthermore, for $\nu > 0$ assume that*

$$\frac{n}{G} \to \infty \quad and \quad \frac{n^{\frac{2}{2+\nu}} \log(n)}{G} \to 0 \quad for \quad n \to \infty.$$

Assumption A.1.2. *Let $\{\mathbb{X}_i : i \geq 1\}$ be a stationary series following a distribution determined by $\boldsymbol{\theta}_0$ in a correctly specified model. Under misspecification let $\boldsymbol{\theta}_0$ be the best approximating parameter for $\{\mathbb{X}_i : i \geq 1\}$ in the sense of $E\left(\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}_0)\right) = \mathbf{0}$. Furthermore, we assume that the stationary sequence $\{\boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) : i \geq 1\}$ has a positive definite long-run covariance matrix $\boldsymbol{\Sigma}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}$.*

Assumption A.1.3. *Let $\boldsymbol{S}(k, \widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{k} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})$ fulfill a strong invariance principle. So possibly after changing the probability space there exists a p-dimensional standard Wiener process $\{\boldsymbol{W}(k) : k \geq 0\}$ with identity matrix $\boldsymbol{I}_p$ as covariance matrix and $\nu > 0$ such that*

$$\left\| \boldsymbol{\Sigma}^{-1/2} \left(\boldsymbol{S}(k, \widetilde{\boldsymbol{\theta}}) - E(\boldsymbol{S}(k, \widetilde{\boldsymbol{\theta}})) \right) - \boldsymbol{W}(k) \right\| = O(k^{1/(2+\nu)}) \quad a.s.$$

as k goes to infinity.

Assumption A.1.4. *Let*

$$\begin{aligned} &\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n, k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}}, k} \right\| \\ &= \max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\| \sum_{i=k+1}^{k+G} \left(\boldsymbol{H}(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}) - \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \right) - \sum_{i=k-G+1}^{k} \left(\boldsymbol{H}(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}) - \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) \right) \right\| \\ &= o_P\left((\log(n/G))^{-1/2} \right) \end{aligned}$$

hold for some $\widetilde{\boldsymbol{\theta}}$.

Assumption A.1.5. *The estimator $\widehat{\boldsymbol{\Sigma}}_{k,n}$ of the long-run covariance matrix $\boldsymbol{\Sigma}$ can depend on k and satisfies*

$$\max_{G \leq k \leq n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}^{-1/2} \right\|_F = o_P\left(\left(\log(n/G) \right)^{-1} \right)$$

under the null hypothesis.

A.2. Assumptions Under the Alternative

Assumption A.2.1. *Let q be the number of change points, occuring in the time period, which is unknown but fixed. Furthermore, let $k_{1,n} < \ldots < k_{q,n}$ be the change points depending on the sample size n in the following way: $k_{j,n} = \lfloor \lambda_j n \rfloor$ with λ_j as rescaled change point being a constant but unknown value in $(0,1)$, for $j = 1, \ldots, q$*

Assumption A.2.2. *Let $\{\mathbb{X}_i : i \geq 1\}$ be a piecewise stationary series such that*

$$\mathbb{X}_i = \begin{cases} \mathbb{X}_i^{(1)}, & \text{if } 1 \leq i \leq k_{1,n} \\ \mathbb{X}_i^{(2)}, & \text{if } k_{1,n} < i \leq k_{2,n} \\ \vdots & \\ \mathbb{X}_i^{(q+1)}, & \text{if } k_{q,n} < i \leq n \end{cases},$$

where $\{\mathbb{X}_i^{(j)} : i \geq 1\}$ is stationary following a distribution determined by $\boldsymbol{\theta}_j$, for $j = 1, \ldots, q+1$, in a correctly specified model. Under misspecification let $\boldsymbol{\theta}_j$ be the best approximating parameter for $\{\mathbb{X}_i^{(j)} : i \geq 1\}$ in the sense of $E\left(\boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)\right) = \mathbf{0}$. Furthermore, we assume that the stationary sequence $\{\boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) : i \geq 1\}$ has a positive definite long-run covariance matrix $\boldsymbol{\Sigma}_{(j)}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}_{(j)}$, for all $j = 1, \ldots, q+1$.

Assumption A.2.3. *Let $\boldsymbol{S}(j,k,\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{k} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})$ fulfill a strong invariance principle for all $j = 1, \ldots, q+1$. So possibly after changing the probability space there exists a p-dimensional standard Wiener process $\{\boldsymbol{W}(k) : k \geq 0\}$ with identity matrix $\boldsymbol{I}_p$ as covariance matrix and $\nu > 0$ such that*

$$\left\| \boldsymbol{\Sigma}_{(j)}^{-1/2} \left(\boldsymbol{S}(j,k,\widetilde{\boldsymbol{\theta}}) - E(\boldsymbol{S}(j,k,\widetilde{\boldsymbol{\theta}})) \right) - \boldsymbol{W}(k) \right\| = O(k^{1/(2+\nu)}) \quad a.s., \quad k \to \infty.$$

Assumption A.2.4. *Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n \in \mathbb{N}}$ be a sequence of estimators fulfilling*

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = O_P\left(\sqrt{\log(n/G)} \right),$$

for some $\widetilde{\boldsymbol{\theta}}$.

Assumption A.2.5. *The estimator $\widehat{\boldsymbol{\Sigma}}_{k,n}$ of the long-run covariance matrix $\boldsymbol{\Sigma}_k$ is positive definite and satisfies*

(a) $\max_{G \leq k \leq n-G} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_k^{-1/2} \right\|_F = O_P(1)$

(b) $\max_{k \in A_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_k^{-1/2} \right\|_F = o_P\left(\log(n/G)^{-1}\right)$
with $A_{n,G} := \{k \in \{G, \ldots, n-G\} : |k - k_{j,n}| > G \; \forall \, j = 1, \ldots, q\}$,

(c) $\max_{k \in B_{n,G}} \left\| \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1/2} - \boldsymbol{\Sigma}_{A,k}^{-1/2} \right\|_F = o_P(1)$,
where $B_{n,G} := \{k \in \{G, \ldots, n-G\} : \exists \, j \in \{1, \ldots, q\}$ with $|k - k_{j,n}| \leq G\}$ and $\{\boldsymbol{\Sigma}_{A,k}\}$ is a sequence of positive definite matrices fulfilling $\sup_k \|\boldsymbol{\Sigma}_{A,k}\|_F < \infty$.

Assumption A.2.6. *For at least one $j \in \{1, \ldots, q\}$ it holds that*

$$E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right) \neq E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right).$$

Assumption A.2.7. *Let $\widetilde{Q} = \widetilde{Q}\left(\widetilde{\boldsymbol{\theta}}\right)$ be the set of indices of all rescaled change points causing a change in the expected value of the transformed series (detectable changes), i.e.*

$$E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right) \neq E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right)$$

holds for all $j \in \widetilde{Q}$ and

$$E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right) = E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right)$$

for all $j \in \{1, \ldots, q\} \setminus \widetilde{Q}$.
Furthermore, let $\widetilde{q} = \widetilde{q}\left(\widetilde{\boldsymbol{\theta}}\right)$ be the number of elements of $\widetilde{Q}$ which is the number of detectable changes.

Assumption A.2.8. *Let the sequence of significance levels α_n fulfill*

$$\alpha_n \to 0 \quad \text{and} \quad \frac{c_{\alpha_n}}{a(n/G)\sqrt{G}} = o(1),$$

where $a(x) = \sqrt{\log(x)}$ and c_{α_n} is the $(1-\alpha_n)$-quantile of the Gumbel distribution.

Assumption A.2.9. *Let $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ be a sequence of estimators fulfilling*

(I) $\max_{k \in A_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = o_P\left((\log(n/G))^{-1/2}\right)$,
where $A_{n,G} = \{k \in \{G, \ldots, n-G\} : |k - \lfloor \lambda_j n \rfloor| \geq G \ \forall\, j \in \{1, \ldots, q\}\}$.

(II) $\max_{k \in \widetilde{A}_{n,G}} \frac{1}{\sqrt{2G}} \left\| \boldsymbol{A}_{\widehat{\boldsymbol{\theta}}_n,k} - \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k} \right\| = o_P\left(\sqrt{\log(n/G)}\right)$,
where $\widetilde{A}_{n,G} = \left\{k \in \{G, \ldots, n-G\} : |k - \lfloor \lambda_j n \rfloor| \geq G \ \forall\, j \in \widetilde{Q}\right\}$ with $\widetilde{Q}$ denoting the set of indices of detectable rescaled change points defined by Assumption A.2.7.

Assumption A.2.10. *Let the following forward and backward Hájek-Rényi-type inequalities hold for some $\gamma > 2$:*

(a) For all $j \in \{1, \ldots, q+1\}$ and for any positive and non-increasing sequence $b_1 \geq b_2 \geq \ldots \geq b_n > 0$ there exists a constant $B(\gamma)$ such that

$$E\left(\max_{1\leq k\leq n} b_k \left\| \sum_{i=1}^{k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\|\right)^{\gamma} \leq B(\gamma) \sum_{k=1}^{n} b_k^{\gamma} k^{\gamma/2-1}.$$

(b) For all $j \in \{1, \ldots, q+1\}$ and for any positive and non-decreasing sequence $0 < a_1 \leq a_2 \leq \ldots \leq a_n$ there exists a constant $A(\gamma)$ such that

$$E\left(\max_{1\leq k\leq n} a_k \left\| \sum_{i=k+1}^{n} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\|\right)^{\gamma} \leq A(\gamma) \sum_{k=1}^{n} a_k^{\gamma} (n-k)^{\gamma/2-1}.$$

Assumption A.2.11. *Let* $\{\widehat{\boldsymbol{\theta}}_n\}_{n\in\mathbb{N}}$ *be a sequence of estimators fulfilling, for any* $m \in \mathbb{N}$ *and for each* $j \in \{1, \ldots, q+1\}$,

(i) $\max_{1\le k\le n} \frac{1}{k} \left\| \sum_{i=m-k+1}^{m} \left(\boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widehat{\boldsymbol{\theta}}_n\right) - \boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}\right)\right)\right\| = o_P(1)$
and

(ii) $\max_{1\le k\le n} \frac{1}{k} \left\| \sum_{i=m+1}^{m+k} \left(\boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widehat{\boldsymbol{\theta}}_n\right) - \boldsymbol{H}\left(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}\right)\right)\right\| = o_P(1)$

for some $\widetilde{\boldsymbol{\theta}}$.

Assumption A.2.12. *Let* $\widehat{\boldsymbol{\Sigma}}_{k,n}$ *be a local estimator for the long-run covariance matrix* $\boldsymbol{\Sigma}_k$ *which is positive definite and fufills Assumption A.2.5. Furthermore, let* $\widehat{\boldsymbol{\Sigma}}_{j,n}$, $j = 1, \ldots, q+1$, *be a positive definite global estimator which is consistent for the true long-run covariance matrix* $\boldsymbol{\Sigma}$ *under the null and which converges in probability to some positive definite matrix* $\boldsymbol{\Sigma}_{A,j}$ *under alternative.*

B. Assumptions of Section 2.3 and Section 3.1

We consider the following type of series:

(E1) $\mathbb{X}_1, \ldots, \mathbb{X}_n$ are an i.i.d. sequence of random vectors or

(E2) $\mathbb{X}_1, \ldots, \mathbb{X}_n$ are a stationary and strongly mixing sequence of random vectors with a mixing rate $\alpha(n)$ satisfying $\alpha(n) = O(n^{-\beta})$ for some $\beta > 1 + 2/\nu$, where ν is as in Assumption A.1.3.

B.1. Under the Null Hypothesis

Assumption B.1.1.
Let $E\left(\|\boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\|\right) < \infty$ *hold for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Assumption B.1.2.
Let $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1, \widetilde{\boldsymbol{\theta}})\right\|^2\right) < \infty$.

Assumption B.1.3.
Let $E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\nabla \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\|_F\right) < \infty$.

Assumption B.1.4.
$E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\nabla^2 H_j(\mathbb{X}_1, \boldsymbol{\theta})\|_F\right) < \infty$ *hold for all* $j = 1, \ldots, p$.

Assumption B.1.5.
There exists a $\nu > 0$ *such that* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1, \widetilde{\boldsymbol{\theta}})\right\|^{2+\nu}\right) < \infty$.

Assumption B.1.6.
There exists a $\nu > 0$ *such that* $E\left(\|\nabla \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\|_F^{2+\nu}\right) < \infty$ *holds for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Assumption B.1.7.
Let $E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\nabla^2 H_j(\mathbb{X}_1, \boldsymbol{\theta})\|_F^{2+\nu}\right) < \infty$ *hold for all* $j = 1, \ldots, p$ *and for some* $\nu > 0$.

Assumption B.1.8.
Let $\boldsymbol{V}(\boldsymbol{\theta})$ *be a regular matrix for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and let*

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\|\boldsymbol{V}(\boldsymbol{\theta})^{-1}\right\|_F < \infty,$$

with $\boldsymbol{V}(\boldsymbol{\theta}) = E\left(\nabla \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right)^T$.

B.2. Under the Alternative

Assumption B.2.1.
Let $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|\right) < \infty$ *hold for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $j = 1, \ldots, q+1$.

Assumption B.2.2.
Let $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right\|^2\right) < \infty$, $j = 1, \ldots, q+1$.

Assumption B.2.3.
Let $E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$, $j = 1, \ldots, q+1$.

Assumption B.2.4.
$E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F\right) < \infty$ *hold for all* $l = 1, \ldots, p$, $j = 1, \ldots, q+1$.

Assumption B.2.5.
There exists a $\nu > 0$ *such that* $E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right\|^{2+\nu}\right) < \infty$, $j = 1, \ldots, q+1$.

Assumption B.2.6.
There exists a $\nu > 0$ *such that* $E\left(\left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F^{2+\nu}\right) < \infty$ *holds for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $j = 1, \ldots, q+1$.

Assumption B.2.7.
Let $E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\|\nabla^2 H_l(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F^{2+\nu}\right) < \infty$ *hold for all* $l = 1, \ldots, p$ *and for some* $\nu > 0$, $j = 1, \ldots, q+1$.

Assumption B.2.8.
There exists a $\nu > 0$ *such that*
$E\left(\left\|\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|^{2+\nu}\right) < \infty$ *holds for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $j = 1, \ldots, q+1$.

Assumption B.2.9.
Let $\boldsymbol{V}_j(\boldsymbol{\theta})$ *be a regular matrix for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and let*

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\|\boldsymbol{V}_j(\boldsymbol{\theta})^{-1}\right\|_F < \infty,$$

with $\boldsymbol{V}_j(\boldsymbol{\theta}) = E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)^T$, $j = 1, \ldots, q+1$.

Assumption B.2.10. *Let*

$$\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta})$$

be a regular matrix for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and all* $\delta \in [0, 1]$ *and let*

$$\sup_{\delta \in [0,1]} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\|(\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta}))^{-1}\right\|_F < \infty,$$

$j = 1, \ldots, q$.

Assumption B.2.11. *There exists a* $\nu > 0$ *such that*
$E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\|\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right\|_F^{2+\nu}\right) < \infty$ *holds*, $j = 1, \ldots, q+1$.

C. Assumptions of Section 3.2

C.1. Assumptions Under the Null Hypothesis

Let $\boldsymbol{\beta}_0$ be the true parameter of the model if no change occurs, i.e.

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta}_0 + \varepsilon_i$$

holds for all $i = 1, \ldots, n$ under the null. Furthermore, the following assumptions are used in that section:

(R1) The sequence $\{\boldsymbol{X}_i\}_{i \geq 1}$ is stationary and ergodic with $E(\|\boldsymbol{X}_1\|) < \infty$.

(R2) Let $\mathcal{F}_t = \sigma(\boldsymbol{X}_j, \varepsilon_{j-1}, j \leq t)$. We assume that ε_t and $\mathcal{F}_t$ are independent.

(R3) $\boldsymbol{C} := E\left(\boldsymbol{X}_1 \boldsymbol{X}_1^T\right)$ is a positive definite matrix.

(R4) The sequence $\{\varepsilon_i\}_{i \geq 1}$ is i.i.d. with $E(\varepsilon_1) = 0$, $0 < E(\varepsilon_1^2) := \sigma^2 < \infty$.

(R5) Let the components of $\{\boldsymbol{X}_i \boldsymbol{X}_i^T - \boldsymbol{C}\}_{i \geq 1}$ satisfy a strong invariance principle similar to that in Assumption A.1.3.

(R6) Let $\{\boldsymbol{X}_i \varepsilon_i\}_{i \geq 1}$ be a series with positive definite long-run covariance matrix $\boldsymbol{\Sigma}$ satisfying a strong invariance principle similar to that in Assumption A.1.3.

C.2. Assumptions Under the Alternative

Under the alternative we allow for multiple changes in the regression coefficients and get a piecewise stationary response sequence $\{Y_i\}_{i \geq 1}$ with

$$Y_i = Y_i^{(j)} = \boldsymbol{X}_i^{(j)T} \boldsymbol{\beta}_j + \varepsilon_i,$$

for $k_{j-1,n} < i \leq k_{j,n}$ and $j = 1, \ldots, q+1$. The following assumptions are used in that section.

(R1*) The sequence $\{\boldsymbol{X}_i^{(j)}\}_{i \geq 1}$ is stationary and ergodic with $E\left(\left\|\boldsymbol{X}_1^{(j)}\right\|\right) < \infty$, for $j = 1, \ldots, q+1$.

(R2*) Let $\mathcal{F}_t = \sigma(\boldsymbol{X}_j, \varepsilon_{j-1}, j \leq t)$. We assume that ε_t and $\mathcal{F}_t$ are independent.

(R3*) $\boldsymbol{C}_{(j)} := E\left(\boldsymbol{X}_1^{(j)} \boldsymbol{X}_1^{(j)T}\right)$ is a positive definite matrix, for $j = 1, \ldots, q+1$.

(R4*) The sequence $\{\varepsilon_i\}_{i \geq 1}$ is i.i.d. with $E(\varepsilon_1) = 0$, $0 < E(\varepsilon_1^2) := \sigma^2 < \infty$.

(R5*) Let the components of $\{\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T}-\boldsymbol{C}_{(j)}\}_{i\geq 1}$ satisfy a strong invariance principle similar to that in Assumption A.1.3, for $j=1,\ldots,q+1$.

(R6*) Let $\{\boldsymbol{X}_i^{(j)}\varepsilon_i\}_{i\geq 1}$ be a series with positive definite long-run covariance matrix $\boldsymbol{\Sigma}_{(j)}$ satisfying a strong invariance principle similar to that in Assumption A.1.3, for $j=1,\ldots,q+1$.

(R7*) Let the matrix $\delta\boldsymbol{C}_{(j)}+(1-\delta)\boldsymbol{C}_{(j+1)}$ be positive definite for all $\delta\in[0,1]$ and assume that $\sup_{\delta\in[0,1]}\left\|\left(\delta\boldsymbol{C}_{(j)}+(1-\delta)\boldsymbol{C}_{(j+1)}\right)^{-1}\right\|_F<\infty$, for all $j=1,\ldots,q$.

D. Assumptions of Chapter 5

D.1. The General Model

In the general parameter change model we assume that, for all $j = 1, \ldots, q+1$, the following conditions are satisfied:

(M1) Let $\{\mathbb{X}_i^{(j)}\}_{i\geq 1}$ be a stationary and ergodic sequence in $\mathbb{R}^p$.

(M2) Let $\boldsymbol{S}(j,k,\boldsymbol{\theta}_j) = \sum_{i=1}^{k} \boldsymbol{H}(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}_j)$ fulfill a strong invariance principle such that (possibly after changing the probability space) there exists a p-dimensional standard Wiener process $\{\boldsymbol{W}(k) : k \geq 0\}$ with identity matrix $\boldsymbol{I}_p$ as covariance matrix and $\nu > 0$ such that

$$\left\| \boldsymbol{\Sigma}_{(j)}^{-1/2} \left(\boldsymbol{S}(j,k,\boldsymbol{\theta}_j) - E(\boldsymbol{S}(j,k,\boldsymbol{\theta}_j)) \right) - \boldsymbol{W}(k) \right\| = O(k^{1/(2+\nu)}) \;\; a.s.$$

as k goes to infinity.

(M3) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for all $l = 1, \ldots, p$, let the sequence $\left\{ \nabla H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\}$ fulfill a strong invariance principle as described in (M2).

(M4) For all $l = 1, \ldots, p$, let $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \right) < \infty$ and let the sequence $\left\{ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^2 H_l(\mathbb{X}_i^{(j)}, \boldsymbol{\theta}) \right\|_F \right\}$ satisfy a strong invariance principle as in (M2).

(M5) Let $E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}) \right\|_F \right) < \infty$ hold.

(M6) Let the following forward and backward Hájek-Rényi-type inequalities hold for $\widetilde{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{1+q}\}$, for any $m_n \in \mathbb{N}_0$ and a positive deterministic sequence $\{v_n\}$ with $v_n \to \infty$ (which will be specified later):

$$\max_{v_n \leq k \leq n-m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n+1}^{m_n+k} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\| = O_P(1)$$

and

$$\max_{v_n \leq k \leq m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n-k+1}^{m_n} \boldsymbol{H}_0(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}}) \right\| = O_P(1)$$

(M7) Let $\boldsymbol{V}_j(\boldsymbol{\theta}) = E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta}) \right)^T$ be a regular matrix for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and let

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \boldsymbol{V}_j(\boldsymbol{\theta})^{-1} \right\|_F < \infty, \;\; \text{for all } j = 1, \ldots, q+1.$$

(M8) Let $\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta})$ be a regular matrix for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\delta \in [0,1]$ and let

$$\sup_{\delta\in[0,1]} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left(\delta \boldsymbol{V}_j(\boldsymbol{\theta}) + (1-\delta)\boldsymbol{V}_{j+1}(\boldsymbol{\theta})\right)^{-1} \right\|_F < \infty, \text{ for all } j = 1, \dots, q.$$

(M9) For $s \geq 1$ we assume that $\sum_{l=j+1}^{j+s}(\lambda_l - \lambda_{l-1})\boldsymbol{V}_l(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left(\sum_{l=j+1}^{j+s} (\lambda_l - \lambda_{l-1})\boldsymbol{V}_l(\boldsymbol{\theta}) \right)^{-1} \right\|_F < \infty.$$

(M10) Let $E\left(Q(\mathbb{X}_i^{(j)}, \widetilde{\boldsymbol{\theta}})\right) < \infty$ where $\widetilde{\boldsymbol{\theta}}$ denotes the unique zero of

$$\sum_{j=1}^{q+1} (\lambda_j - \lambda_{j-1})\, E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right).$$

D.2. The Linear Regression Model

We consider the linear regression model introduced in Section 3.2.2 under the Assumptions (R1*) to (R7*) which can be found in Section C.2 as well. Furthermore, the following conditions are used.

(R8*) For all $j = 1, \dots, q+1$, let the series $\{\boldsymbol{X}_i^{(j)}\varepsilon_i\}_{i\geq 1}$ satisfy the following forward and backward Hájek-Rényi-type inequalities, for any $m_n \in \mathbb{N}_0$ and a positive deterministic sequence $\{v_n\}$ with $v_n \to \infty$ (which will be specified later):

$$\max_{v_n \leq k \leq n-m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n+1}^{m_n+k} \boldsymbol{X}_i^{(j)}\varepsilon_i \right\| = O_P(1)$$

and

$$\max_{v_n \leq k \leq m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n-k+1}^{m_n} \boldsymbol{X}_i^{(j)}\varepsilon_i \right\| = O_P(1).$$

(R9*) For all $j = 1, \dots, q+1$, let the series $\{\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\}_{i\geq 1}$ fulfill the following forward and backward Hájek-Rényi-type inequalities, for any $m_n \in \mathbb{N}_0$ and a positive deterministic sequence $\{v_n\}$ with $v_n \to \infty$ (which will be specified later):

$$\max_{v_n \leq k \leq n-m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n+1}^{m_n+k} \left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right) \right\|_F = O_P(1)$$

and

$$\max_{v_n \leq k \leq m_n} \left\| \frac{\sqrt{v_n}}{k} \sum_{i=m_n-k+1}^{m_n} \left(\boldsymbol{X}_i^{(j)}\boldsymbol{X}_i^{(j)T} - \boldsymbol{C}_{(j)}\right) \right\|_F = O_P(1).$$

(R10*) Let the matrix $\sum_{l=j+1}^{j+s}(\lambda_l - \lambda_{l-1})\boldsymbol{C}_{(l)}$ be positive definite for some $s \geq 1$.

E. Theoretical Results

E.1. Norms and Matrices

Remark E.1.1. *According to Roy & Banerjee (2014) (page 492f), the Frobenius norm is a matrix norm which is defined as follows*

$$\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{A}^T\boldsymbol{A})} = \sqrt{\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^T)} \;\; \textit{for} \;\; \boldsymbol{A} \in \mathbb{R}^{p\times q}.$$

An important property of this norm is the submultiplicativity,

$$\|\boldsymbol{A}\boldsymbol{B}\|_F \leq \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F .$$

Lemma E.1.2. *Let $\{X_n\}$ and $\{Y_n\}$ be deterministic or stochastic sequences of vectors in $\mathbb{R}^p$ and let $A \subseteq \mathbb{N}$. Then,*

$$\left| \max_{n\in A} \|X_n\| - \max_{n\in A} \|Y_n\| \right| \leq \max_{n\in A} \|X_n - Y_n\| .$$

Proof.

$$\begin{aligned}\max_{n\in A} \|X_n\| &= \max_{n\in A} \|X_n - Y_n + Y_n\| \leq \max_{n\in A} \left(\|X_n - Y_n\| + \|Y_n\|\right) \\ &\leq \max_{n\in A} \|X_n - Y_n\| + \max_{n\in A} \|Y_n\|\end{aligned}$$

yields

$$\max_{n\in A} \|X_n\| - \max_{n\in A} \|Y_n\| \leq \max_{n\in A} \|X_n - Y_n\| .$$

Similarly we obtain

$$\max_{n\in A} \|Y_n\| - \max_{n\in A} \|X_n\| \leq \max_{n\in A} \|X_n - Y_n\| ,$$

showing the assertion. □

Lemma E.1.3. *Let $\{X_n\}$ and $\{Y_n\}$ be deterministic or stochastic sequences of vectors in $\mathbb{R}^p$ and let $A \subseteq \mathbb{N}$. Then,*

$$\left| \min_{n\in A} \|X_n\| - \min_{n\in A} \|Y_n\| \right| \leq \max_{n\in A} \|X_n - Y_n\| .$$

Proof. Without loss of generality we assume that $\min_{n\in A} \|X_n\| \geq \min_{n\in A} \|Y_n\|$. We receive

$$\|X_n - Y_n\| \geq \|X_n\| - \|Y_n\| \geq \min_{n\in A} \|X_n\| - \|Y_n\| ,$$

which holds for all $n \in A$. Hence, maximizing the left and right site does not change the inequality sign and we obtain

$$\max_{n\in A} \|X_n - Y_n\|$$
$$\geq \max_{n\in A} \left(\min_{n\in A} \|X_n\| - \|Y_n\| \right) = \min_{n\in A} \|X_n\| - \min_{n\in A} \|Y_n\| = \left| \min_{n\in A} \|X_n\| - \min_{n\in A} \|Y_n\| \right| .$$

□

Lemma E.1.4. *Let $\boldsymbol{a} = (a_1, \ldots, a_p)^T$ be a vector in $\mathbb{R}^p$. Then,*

$$\|\boldsymbol{a}\|_F = \|\boldsymbol{a}\| ,$$

where $\|\cdot\|$ denotes the Euclidean norm and $\|\cdot\|_F$ the Frobenius norm.

Proof. By definition of the Frobenius and the Euclidean norm it holds that

$$\|\boldsymbol{a}\|_F = \sqrt{\operatorname{tr}\left(\boldsymbol{a}^T\boldsymbol{a}\right)} = \sqrt{a_1^2 + \ldots + a_p^2} = \|\boldsymbol{a}\| .$$

□

Lemma E.1.5. *Let $\boldsymbol{A}$ be a $p \times p$-matrix and $\boldsymbol{x} \in \mathbb{R}^p$. Then,*

$$\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\|_F \|\boldsymbol{x}\| ,$$

where $\|\cdot\|$ denotes the Euclidean norm and $\|\cdot\|_F$ the Frobenius norm of matrices.

Proof. On noting that $\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^p$, applying lemma E.1.4 and the submultiplicativity of the Frobenius norm yield

$$\|\boldsymbol{A}\boldsymbol{x}\| = \|\boldsymbol{A}\boldsymbol{x}\|_F \leq \|\boldsymbol{A}\|_F \|\boldsymbol{x}\|_F = \|\boldsymbol{A}\|_F \|\boldsymbol{x}\| .$$

□

Lemma E.1.6. *Let $\boldsymbol{A} = (a_{ij})$ be an $r \times s$-matrix. Furthermore, let $a_{i\bullet}$ be the i-th row and $a_{\bullet j}$ be the j-th column of the matrix. Then,*

(a) $\max_{i,j} |a_{i,j}| \leq \|\boldsymbol{A}\|_F$,

(b) $\|\boldsymbol{A}\|_F \leq \sum_{i,j} |a_{ij}|$,

(c) $\max_{1\leq i\leq r} \|a_{i\bullet}\| \leq \|\boldsymbol{A}\|_F$ *and* $\max_{1\leq j\leq s} \|a_{\bullet j}\| \leq \|\boldsymbol{A}\|_F$ *and*

(d) $\|\boldsymbol{A}\|_F \leq \sum_{i=1}^{r} \|a_{i\bullet}\|$ *and* $\|\boldsymbol{A}\|_F \leq \sum_{j=1}^{s} \|a_{\bullet j}\|$.

Proof. By using the definition of the Frobenius norm, we get

$$\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^T)} = \sqrt{\sum_{i,j} |a_{ij}|^2} \geq \sqrt{|a_{kl}|^2} = |a_{kl}| \quad \text{for all } k \text{ and } l$$

and

$$\|\boldsymbol{A}\|_F^2 = \sum_{i,j} |a_{ij}|^2 \leq \sum_{k,l} \sum_{i,j} |a_{ij}||a_{kl}| = \left(\sum_{i,j} |a_{ij}| \right)^2 .$$

The first inequality proves part (a) and the second implies result (b).
Moreover, we obtain

$$\|\boldsymbol{A}\|_F^2 = \sum_{i,j} |a_{ij}|^2 = \sum_{i=1}^{r} \|a_{i.}\|^2 \leq \sum_{k=1}^{r} \sum_{i=1}^{r} \|a_{i.}\| \, \|a_{k.}\| = \left(\sum_{i=1}^{r} \|a_{i.}\| \right)^2 \quad \text{and}$$

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\sum_{i=1}^{r} \|a_{i.}\|^2} \geq \sqrt{\|a_{i.}\|^2} = \|a_{i.}\| \quad \text{for all } i,$$

showing the results (c) and (d) of the lemma. Similar inequalities can be derived for the columns of the matrix in order to complete the proof. □

Lemma E.1.7. *Let $\boldsymbol{A} \in \mathbb{R}^{p \times p}$ be a symmetric positive definite matrix and $\boldsymbol{B} \in \mathbb{R}^{p \times p}$ a regular matrix. Then, $\boldsymbol{C} := \boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^T$ is a symmetric positive definite matrix.*

Proof. By determining the transpose $(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^T)^T = \boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^T$ we get that $\boldsymbol{C}$ is symmetric. Furthermore, let $\boldsymbol{x} \in \mathbb{R}^p$ be an arbitrary non-zero vector ($\boldsymbol{x} \neq \boldsymbol{0}$). Since $\boldsymbol{B}$ is a regular Matrix $\boldsymbol{y} := \boldsymbol{B}^T\boldsymbol{x}$ is a non-zero vector in $\mathbb{R}^p$ as well. Hence, on noting that $\boldsymbol{A}$ is positive definite, we receive

$$\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y} > 0,$$

which shows the assertion. □

Lemma E.1.8. *Let $\{\boldsymbol{B}_n\}$ be a sequence of symmetric positive definite $p \times p$ matrices. If $\boldsymbol{B}_n \overset{P}{\to} \boldsymbol{B}$, where $\boldsymbol{B}$ is a positive definite matrix,*
then it holds

$$\boldsymbol{B}_n^{-1/2} \overset{P}{\to} \boldsymbol{B}^{-1/2}.$$

Proof. First note that the square root of the inverse matrix $\boldsymbol{f}(\boldsymbol{A}) = \boldsymbol{A}^{-1/2}$ is a primary matrix function on the set of positive definite matrices with scalar-valued stem function $f(t) = t^{-1/2}$ according to Definition 6.2.4 in Horn & Johnson (1991) on page 410. Since the stem function $f(t) = t^{-1/2}$ is continuous on $(0, \infty)$, applying Theorem 6.2.37 of Horn & Johnson (1991) on page 433 yields that the matrix function $\boldsymbol{f}$ is continuous on the open cone of positive definite matrices. Thus, the assertion follows from the continuous mapping theorem.

□

Lemma E.1.9. *Let $\boldsymbol{A}$ be a symmetric positive definite $p \times p$-matrix. Furthermore, let λ_{max} denote the largest eigenvalue of matrix $\boldsymbol{A}$ and δ_{min} the smallest eigenvalue of the inverse matrix $\boldsymbol{A}^{-1}$. Then, the following inequalities hold:*

(a) $\lambda_{max} \leq \|\boldsymbol{A}\|_F$

(b) $\delta_{min} \geq \|\boldsymbol{A}\|_F^{-1}$.

Proof. (a) Let $\lambda_1, \ldots, \lambda_p$ denote the eigenvalues of $\boldsymbol{A}$. On noting that the eigenvalues of the squared matrix $\boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{A}$ are given by $\lambda_1^2, \ldots, \lambda_p^2$, we receive

$$\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{A}^T\right)} = \sqrt{\operatorname{tr}\left(\boldsymbol{A}^2\right)} = \sqrt{\sum_{i=1}^{p} \lambda_i^2} \geq \sqrt{\lambda_{max}^2} = \lambda_{max}.$$

(b) Let $\delta_1, \ldots, \delta_p$ denote the eigenvalues of $\boldsymbol{A}^{-1}$. Since $\delta_i = \frac{1}{\lambda_i}$ for $i = 1, \ldots, p$, the result of part (a) can be used to obtain

$$\delta_{min} = \frac{1}{\lambda_{max}} \geq \|\boldsymbol{A}\|_F^{-1}.$$

□

Lemma E.1.10. *Let $\{\boldsymbol{A}_k\}$ be a sequence of symmetric positive definite $p \times p$- matrices. If $\sup_k \|\boldsymbol{A}_k\|_F < \infty$, then there exists $c > 0$ such that $\lambda_{min}(\boldsymbol{A}_k^{-1}) \geq c$ holds for all k, where $\lambda_{min}\left(\boldsymbol{A}_k^{-1}\right)$ denotes the smallest eigenvalue of matrix $\boldsymbol{A}_k^{-1}$.*

Proof. Applying Lemma E.1.9 (b) yields

$$\lambda_{min}\left(\boldsymbol{A}_k^{-1}\right) \geq \|\boldsymbol{A}_k\|_F^{-1} \quad \text{for all } k.$$

Hence, by assumption we obtain

$$\inf_k \left(\lambda_{min}\left(\boldsymbol{A}_k^{-1}\right)\right) \geq \inf_k \left(\|\boldsymbol{A}_k\|_F^{-1}\right) = \left(\sup_k \|\boldsymbol{A}_k\|_F\right)^{-1} > 0,$$

which shows the assertion. □

Lemma E.1.11. *Let $\boldsymbol{A}$ be a regular 2×2-matrix. Then,*

$$\left\|\boldsymbol{A}^{-1}\right\|_F = \frac{1}{|\det(\boldsymbol{A})|} \|\boldsymbol{A}\|_F.$$

Proof. With $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ the inverse of the matrix is given by

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Thus, by the definition of the Frobenius norm we receive

$$\left\|\boldsymbol{A}^{-1}\right\|_F = \sqrt{\sum_{i,j} \left(\frac{a_{ij}}{\det(\boldsymbol{A})}\right)^2} = \frac{1}{|\det(\boldsymbol{A})|} \sqrt{\sum_{i,j} a_{ij}^2} = \frac{1}{|\det(\boldsymbol{A})|} \|\boldsymbol{A}\|_F.$$

□

Lemma E.1.12. *Let $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ be a symmetric positive definite matrix with minimal eigenvalue $\lambda_{min} := \min_{1\leq j\leq p} \lambda_j$. Then, for any vector $\boldsymbol{x} \in \mathbb{R}^p$ with $\boldsymbol{x} \neq \boldsymbol{0}$, it holds*

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq \lambda_{min} \left\|\boldsymbol{x}\right\|^2 .$$

Proof. An eigendecomposition of the positive definite matrix $\boldsymbol{A}$ yields that there exists an orthogonal matrix $\boldsymbol{Q}$ such that $\boldsymbol{A} = \boldsymbol{Q}\mathrm{diag}(\lambda_i)\boldsymbol{Q}^T$. Hence, with $\boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{x}$, the computation rules for the trace of a matrix can be used to receive

$$\begin{aligned}
&\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \mathrm{tr}\left(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\right) = \mathrm{tr}\left(\boldsymbol{x}^T\boldsymbol{Q}\mathrm{diag}(\lambda_i)\boldsymbol{Q}^T\boldsymbol{x}\right) = \mathrm{tr}\left(\mathrm{diag}(\lambda_i)\boldsymbol{Q}^T\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{Q}\right)\\
&= \mathrm{tr}\left(\mathrm{diag}(\lambda_i)\boldsymbol{y}\boldsymbol{y}^T\right) = \sum_{i=1}^{p}\lambda_{i,n}y_i^2 \geq \sum_{i=1}^{p}\lambda_{min}y_i^2 = \lambda_{min}\left\|\boldsymbol{y}\right\|^2 = \lambda_{min}\left\|\boldsymbol{Q^T}\boldsymbol{x}\right\|^2\\
&= \lambda_{min}\left\|\boldsymbol{x}\right\|^2,
\end{aligned}$$

as the matrix $\boldsymbol{Q}$ is orthogonal. □

E.2. Probability Theory and Statistics

Lemma E.2.1. *Let $\{W(t) : t \in [0,\infty)\}$ be a standard Wiener process, then it holds*

$$\sup_{0\leq t\leq T-1}\sup_{0\leq s\leq 1}|W(t+s)-W(t)| = O(\sqrt{\log T}) \;\; a.s.$$

Proof. By Theorem 1.2.1 (1.2.6) of Csörgo & Révész (1981) on page 30 we obtain

$$\lim_{T\to\infty}\sup_{0\leq t\leq T-1}\sup_{0\leq s\leq 1}\frac{|W(t+s)-W(t)|}{\sqrt{2(\log T+\log\log T)}} = 1 \;\; a.s.$$

This implies

$$\begin{aligned}
\sup_{0\leq t\leq T-1}\sup_{0\leq s\leq 1}|W(t+s)-W(t)| &= O(\sqrt{2(\log T+\log\log T)}) \;\; a.s.\\
&= O(\sqrt{\log T}) \;\; a.s.
\end{aligned}$$

□

Lemma E.2.2. *Let $\{\boldsymbol{W}(t) : t \in [0,\infty)\}$ be a p-dimensional standard Wiener process with identity matrix $\boldsymbol{I}_p$ as covariance matrix, then it holds*

$$\sup_{0\leq t\leq T-1}\sup_{0\leq s\leq 1}\|\boldsymbol{W}(t+s)-\boldsymbol{W}(t)\| = O(\sqrt{\log T}) \;\; a.s.$$

Proof. The component processes $W_1(t),\ldots,W_p(t)$ of the p-dimensional standard Wiener process $\boldsymbol{W}(t)$ are independent standard Wiener processes. Hence, applying Lemma E.2.1 yields

$$\begin{aligned}
&\sup_{0\leq t\leq T-1}\sup_{0\leq s\leq 1}\|\boldsymbol{W}(t+s)-\boldsymbol{W}(t)\| \leq \sup_{0\leq t\leq T-1}\sup_{0\leq s\leq 1}\sum_{i=1}^{p}|W_i(t+s)-W_i(t)|\\
&\leq \sum_{i=1}^{p}\sup_{0\leq t\leq T-1}\sup_{0\leq s\leq 1}|W_i(t+s)-W_i(t)| = O(\sqrt{\log T}) \;\; a.s.
\end{aligned}$$

□

The following lemma and the corresponding proof goes back to Kirch (2008) page 14.

Lemma E.2.3. *Let* $\{a_n\}$ *and* $\{b_n\}$ *be sequences of positive real numbers with* $\lim_{n\to\infty} b_n = \infty$. *Furthermore, consider two sequences of random variables* $\{A_n\}$ *and* $\{B_n\}$ *with* $A_n = o_P(b_n/a_n)$. *Then the random sequences* $a_n \max(A_n, B_n) - b_n$ *and* $a_n B_n - b_n$ *have the same limit distribution, i.e. it holds*

$$\lim_{n\to\infty} |P(a_n \max(A_n, B_n) - b_n \le y) - P(a_n B_n - b_n \le y)| = 0, \quad \forall\, y \in \mathbb{R}.$$

Proof. The proof can also be found in Kirch (2008) on page 14.
Since the maximum can be rewritten as an intersection of two sets, the following inequality holds

$$\begin{aligned}
&P(a_n \max(A_n, B_n) - b_n \le y) = P(\{a_n A_n - b_n \le y\} \cap \{a_n B_n - b_n \le y\}) \\
&= P(a_n A_n - b_n \le y) + P(a_n B_n - b_n \le y) \\
&\qquad\qquad - P(\{a_n A_n - b_n \le y\} \cup \{a_n B_n - b_n \le y\}) \\
&\ge P(a_n A_n - b_n \le y) + P(a_n B_n - b_n \le y) - 1 \\
&= P(a_n B_n - b_n \le y) + P\left(\frac{a_n}{b_n} A_n \le \frac{y}{b_n} + 1\right) - 1.
\end{aligned}$$

Since b_n goes to infinity for n going to infinity, for all $y \in \mathbb{R}$, there exists an $m(y) \in \mathbb{N}$ such that $y/b_n + 1 \ge 1/2$ holds for all $n \ge m(y)$. With $\frac{a_n}{b_n} A_n = o_P(1)$ we obtain $\lim_{n\to\infty} P\big(\frac{a_n}{b_n} A_n \le 1/2\big) = 1$. This implies

$$\begin{aligned}
P(a_n \max(A_n, B_n) - b_n \le y) &\ge P(a_n B_n - b_n \le y) + P\left(\frac{a_n}{b_n} A_n \le 1/2\right) - 1 \\
&= P(a_n B_n - b_n \le y) + o(1)
\end{aligned}$$

for large n.
Furthermore, we obtain

$$P(a_n \max(A_n, B_n) - b_n \le y) \le P(a_n B_n - b_n \le y)$$

since $\{a_n A_n - b_n \le y\} \cap \{a_n B_n - b_n \le y\} \subseteq \{a_n B_n - b_n \le y\}$ holds for all $n \in \mathbb{N}$. □

The following lemma can be derived by using some results of Steinebach & Eastwood (1996).

Lemma E.2.4. *Let* $\{\boldsymbol{Z}(t) : t \ge 0\}$ *be a separable stationary Gaussian process with values in* $\mathbb{R}^p$ *and independent standardized component processes. Let the covariance functions of these components fulfill*

$$\begin{aligned}
&r_i(h) = 1 - C|h| + o(|h|) \quad \text{as } h \to 0 \\
&r_i(h) = o(1/\log h) \quad \text{as } h \to \infty \text{ for all } i = 1, \ldots, p,
\end{aligned}$$

for some constant $C > 0$. *Then,*

$$a(m) \sup_{0\le t\le m} \|\boldsymbol{Z}(t)\| - b(m) \xrightarrow{D} E \quad \text{as } m \to \infty,$$

where

$$a(m) = \sqrt{2\log m},$$
$$b(m) = 2\log m + \frac{p}{2}\log\log m - \log\left(C^{-1}\Gamma\left(\frac{p}{2}\right)\right)$$

and E follows a Gumbel distribution with

$$P(E \le x) = \exp\big(-2\exp(-x)\big) \ \text{for all} \ x \in \mathbb{R}.$$

Proof. The assertion follows from Lemma 3.1. in combination with Remark 3.1. of Steinebach & Eastwood (1996) on page 289 with $\alpha = 1$ and $C_1 = \ldots = C_p = C$. □

Lemma E.2.5. *Let $\{\boldsymbol{Z}(t) : t \ge 0\}$ be a stochastic process with $\boldsymbol{Z}(t) = \frac{1}{\sqrt{2}}(2\boldsymbol{W}(t+1) - \boldsymbol{W}(t) - \boldsymbol{W}(t+2))$ where $\{\boldsymbol{W}(t) : t \ge 0\}$ is a p-dimensional Wiener process with identity matrix $\boldsymbol{I}_p$ as covariance matrix. Then, the covariance function of the component processes is given by*

$$r(h) = \begin{cases} 1 - \frac{3}{2}|h|, & \text{for } |h| \le 1 \\ \frac{1}{2}|h| - 1, & \text{for } 1 < |h| \le 2 \\ 0, & \text{else} \end{cases} .$$

Proof. With $\boldsymbol{W}(t) = (W_1(t), \ldots, W_p(t))^T$ and $\boldsymbol{Z}(t) = (Z_1(t), \ldots, Z_p(t))^T$ the covariance function of a component process $\{Z_i(t)\}$ can be determined as follows.
At first, we consider the case $h \ge 0$.

- Let $h > 2$.

$$\begin{aligned} r_i(h) &= r(h) = \operatorname{Cov}(Z_i(t), Z_i(t+h)) \\ &= \operatorname{Cov}\Bigg(\frac{1}{\sqrt{2}}\big((W_i(t+1) - W_i(t)) - (W_i(t+2) - W_i(t+1)\big), \\ &\qquad \frac{1}{\sqrt{2}}\big((W_i(t+1+h) - W_i(t+h)) - (W_i(t+2+h) - W_i(t+h+1))\big)\Bigg) \\ &= 0. \end{aligned}$$

- Let $1 < h \le 2$.

$$\begin{aligned} &\operatorname{Cov}(Z_i(t), Z_i(t+h)) \\ &= \operatorname{Cov}\Bigg(\frac{1}{\sqrt{2}}\big((W_i(t+1) - W_i(t)) \\ &\qquad - (W_i(t+2) - W_i(t+h)) - (W_i(t+h) - W_i(t+1))\big), \\ &\qquad \frac{1}{\sqrt{2}}\big((W_i(t+1+h) - W_i(t+2)) + (W_i(t+2) - W_i(t+h)) \\ &\qquad - (W_i(t+2+h) - W_i(t+h+1))\big)\Bigg) \\ &= -\frac{1}{2}\operatorname{Var}(W_i(t+2) - W_i(t+h)) = -\frac{2-h}{2} = \frac{1}{2}h - 1. \end{aligned}$$

- Let $0 \le h \le 1$.

$$
\begin{aligned}
&\operatorname{Cov}(Z_i(t), Z_i(t+h)) \\
&= \operatorname{Cov}\Bigg(\frac{1}{\sqrt{2}}\big(2W_i(t+1) - W_i(t) - W_i(t+2)\big), \\
&\qquad\qquad \frac{1}{\sqrt{2}}\big(2W_i(t+1+h) - W_i(t+h) - W_i(t+2+h)\big)\Bigg) \\
&= \operatorname{Cov}\Bigg(\frac{1}{\sqrt{2}}\big((W_i(t+1) - W_i(t+h)) + (W_i(t+h) - W_i(t)) \\
&\qquad - (W_i(t+2) - W_i(t+1+h)) - (W_i(t+1+h) - W_i(t+1))\big), \\
&\qquad\qquad \frac{1}{\sqrt{2}}\big((W_i(t+1+h) - W_i(t+1)) + (W_i(t+1) - W_i(t+h)) \\
&\qquad - (W_i(t+2+h) - W_i(t+2)) - (W_i(t+2) - W_i(t+1+h))\big)\Bigg) \\
&= -\frac{1}{2}\operatorname{Var}(W_i(t+1+h) - W_i(t+1)) + \frac{1}{2}\operatorname{Var}(W_i(t+1) - W_i(t+h)) \\
&\quad + \frac{1}{2}\operatorname{Var}(W_i(t+2) - W_i(t+1+h)) \\
&= -\frac{h}{2} + \frac{1-h}{2} + \frac{1-h}{2} = 1 - \frac{3}{2}h.
\end{aligned}
$$

Due to the symmetry we obtain for $h < 0$

$$
\begin{aligned}
&\operatorname{Cov}(Z_i(t+h), Z_i(t)) = \operatorname{Cov}(Z_i(s), Z_i(s+|h|)) \ \ (s = t+h) \\
&= \begin{cases} 1 - \frac{3}{2}|h|, & \text{for } \ -1 \le h \le 0 \\ \frac{1}{2}|h| - 1, & \text{for } \ -2 \le h < -1 \\ 0, & \text{for } \qquad\quad h < -2 \end{cases},
\end{aligned}
$$

which completes the proof. □

Theorem E.2.6 (Uniform Law of Large Numbers, Theorem 6.5 in Rao (1962))**.** *Let $\|\cdot\|$ be any norm on $\mathbb{R}^d$ and let $\boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta})$ be a stationary and ergodic random sequence with values in $\mathbb{R}^d$ satisfying*

$$
E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\|\right) < \infty,
$$

then

$$
\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| \to 0 \ \ a.s. \quad as \ \ n \to \infty.
$$

Corollary E.2.7. *Let $\|\cdot\|_F$ be the Frobenius norm for matrices and let $\boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta})$ be a stationary and ergodic random sequence with values in $\mathbb{R}^{r \times s}$ satisfying*

$$
E\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\|_F\right) < \infty,
$$

then

$$
\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\|_F \to 0 \ \ a.s. \quad as \ \ n \to \infty.
$$

Proof. Let $F_1(\mathbb{X}_i, \boldsymbol{\theta}), \ldots, F_s(\mathbb{X}_i, \boldsymbol{\theta})$ denote the columns of $\boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta})$. Then, the moment assumption and Lemma E.1.6 (c) yield

$$E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|F_j(\mathbb{X}_1, \boldsymbol{\theta})\|\right) \leq E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\|_F\right) < \infty \quad \text{for all } j = 1, \ldots, s.$$

Hence, by Lemma E.1.6 (d) and Theorem E.2.6 applied to each column vector we receive

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\|_F \\
&\leq \sum_{j=1}^{s} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=1}^{n} F_j(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(F_j(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o(1) \; a.s.
\end{aligned}$$

□

The same statement can be obtained under different assumptions as well which is shown in the following theorem.

Theorem E.2.8 (Uniform Law of Large Numbers II).
Let the parameter space $\boldsymbol{\Theta}$ be compact, let $\{\mathbb{X}_i\}_{i\geq 1}$ be a stationary and ergodic sequence of p-dimensional random vectors and let $\boldsymbol{F} : (\mathbb{R}^p, \boldsymbol{\Theta}) \to \mathbb{R}^d$ be a measurable function with respect to $\mathbb{X}_i$ such that the following assumptions are fulfilled:

(i) Let $E\left(\|\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\|\right) < \infty$ hold for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(ii) Let $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\theta})$ be Lipschitz continuous in $\boldsymbol{\theta}$, i.e. there exists a function $L(\boldsymbol{x}) > 0$, which is measurable with respect to $\mathbb{X}_i$, such that

$$\|\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\xi})\| \leq L(\boldsymbol{x}) \|\boldsymbol{\theta} - \boldsymbol{\xi}\|$$

for all $\boldsymbol{\theta}, \boldsymbol{\xi} \in \boldsymbol{\Theta}$ and

(iii) $E\left(L(\mathbb{X}_1)\right) < \infty$.

Then,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o_P(1).$$

Proof. The proof of this theorem is well known in non-parametric statistics.
There are three main arguments:

(1) The parameter space $\boldsymbol{\Theta}$ is compact. This implies that for each $\delta > 0$ there exist a finite number $M = M(\delta) \geq 1$ and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M \in \boldsymbol{\Theta}$ such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ there is an $m \leq M$ with $\|\boldsymbol{\theta} - \boldsymbol{\theta}_m\| < \delta$.

(2) Since $\boldsymbol{F}$ is measurable with respect to $\mathbb{X}_i$ the ergodicity of the sequence $\{\mathbb{X}_i\}_{i\geq 1}$ carries over to the transformed sequence $\{\boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta})\}_{i\geq 1}$. Hence, for fixed M and

$\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$, the Law of Large Numbers and the Ergodic Theorem, respectively, and assumption (i) yield

$$\begin{aligned}
&\sup_{m \leq M} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}_m) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta}_m)\right) \right\| \\
&\leq \sum_{m=1}^{M} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}_m) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta}_m)\right) \right\| = o_P(1).
\end{aligned}$$

(3) Let $\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) = \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right)$. For any $\boldsymbol{\theta}, \boldsymbol{\xi} \in \boldsymbol{\Theta}$ with $\|\boldsymbol{\theta} - \boldsymbol{\xi}\| < \delta$ we obtain

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\xi})\| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\xi})\| + \|E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\xi})\right)\| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} L(\mathbb{X}_i)\delta + E(L(\mathbb{X}_1))\delta = \delta \left(\frac{1}{n} \sum_{i=1}^{n} (L(\mathbb{X}_i) - E(L(\mathbb{X}_1))) + 2E(L(\mathbb{X}_1)) \right),
\end{aligned}$$

where the last line follows from assumption (ii).

Now, for each $\varepsilon > 0$ we can choose a $\delta > 0$ such that $\frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1)) > 0$ by assumption (iii). Note that the Law of Large Numbers and the Ergodic Theorem, respectively, can be applied to the sequence $\{L(\mathbb{X}_i)\}_{i \geq 1}$ by assumption (iii) and since the sequence $\{L(\mathbb{X}_i)\}_{i \geq 1}$ is stationary and ergodic by the measurability of L and the ergodicity of $\{\mathbb{X}_i\}_{i \geq 1}$. Hence, with M and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$ according to (1) and by applying (2) and (3) we get

$$\begin{aligned}
&P\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) \right\| > \varepsilon \right) = P\left(\sup_{m \leq M} \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_m\|} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) \right\| > \varepsilon \right) \\
&= P\left(\sup_{m \leq M} \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_m\|} \left\| \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m)) + \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m) \right\| > \varepsilon \right) \\
&\leq P\left(\sup_{m \leq M} \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_m\|} \left\| \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m)) \right\| > \frac{\varepsilon}{2} \right) \\
&\quad + P\left(\sup_{m \leq M} \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m) \right\| > \frac{\varepsilon}{2} \right) \\
&\leq P\left(\frac{1}{n} \sum_{i=1}^{n} (L(\mathbb{X}_i) - E(L(\mathbb{X}_1))) > \frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1)) \right) + o(1) = o(1),
\end{aligned}$$

since $\frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1) > 0$ by the choice of δ. □

Definition E.2.9. *Let F be a function on $\boldsymbol{\Theta}$ and let $\widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ be the unique zero of $F(\boldsymbol{\theta})$. Then, $\widetilde{\boldsymbol{\theta}}$ is called the unique zero in the strict sense if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\|F(\boldsymbol{\theta})\| > \delta$ whenever $\left\|\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}\right\| > \epsilon$.*

Lemma E.2.10. *Let $\boldsymbol{\Theta}$ be a compact set and let $F : \boldsymbol{\Theta} \to \mathbb{R}^p$ be a continuous function on $\boldsymbol{\Theta}$. Furthermore, let $\widetilde{\boldsymbol{\theta}}$ be the unique zero of $F(\boldsymbol{\theta})$.*
Then, $\widetilde{\boldsymbol{\theta}}$ is the unique zero in the strict sense according to Definition E.2.9.

Proof. We want to prove this by contradiction. Hence we assume that $\widetilde{\boldsymbol{\theta}}$ is not the unique zero in the strict sense, i.e.:

$$\exists\, \varepsilon > 0 \;\; \forall\, \delta > 0 \; \exists\, \boldsymbol{\theta}_\delta \text{ with } \left\|\boldsymbol{\theta}_\delta - \widetilde{\boldsymbol{\theta}}\right\| > \epsilon : \; \|F(\boldsymbol{\theta}_\delta)\| \leq \delta.$$

Considering the sequence $\{\delta_n\}$ with $\delta_n = \frac{1}{n}$ for all $n \in \mathbb{N}$, we obtain

$$\|F(\boldsymbol{\theta}_{\delta_n})\| \leq \frac{1}{n} \to 0 \;\text{ as }\; n \to \infty. \tag{E.1}$$

Since $\boldsymbol{\Theta}$ is compact the sequence $\{\boldsymbol{\theta}_{\delta_n}\}$ has a convergent subsequence $\{\boldsymbol{\theta}_{\delta_{\alpha(n)}}\}$ with a limit $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. The continuity of F implies $F(\boldsymbol{\theta}_{\delta_{\alpha(n)}}) \to F(\boldsymbol{\theta}^*)$. Furthermore, $F(\boldsymbol{\theta}^*) = \mathbf{0}$ follows from equation (E.1). But $\boldsymbol{\theta}^* \neq \widetilde{\boldsymbol{\theta}}$, which contradicts the assumption that $\widetilde{\boldsymbol{\theta}}$ is the unique zero. □

Lemma E.2.11. *Let $\widetilde{\boldsymbol{\theta}}$ be the unique zero of some function $\boldsymbol{F}(\boldsymbol{\theta})$ in the strict sense as defined in E.2.9. Furthermore, it holds that*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{n} \sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}(\boldsymbol{\theta}) \right\| = o_P(1).$$

Then, the estimator sequence $\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}$ with $\frac{1}{n}\sum_{i=\lfloor\gamma_1 n\rfloor+1}^{\lfloor\gamma_2 n\rfloor} \boldsymbol{H}(\mathbb{X}_i, \widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}) = \mathbf{0}$ for every n satisfies $\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2} \xrightarrow{\mathcal{P}} \widetilde{\boldsymbol{\theta}}$.

Proof. We refer to the proof of proposition 10.1 in Kirch & Tadjuidje Kamgaing (2016) on page 240. □

Theorem E.2.12. *Let $\{\mathbb{X}_i\}_{i\geq 1}$ be a stationary sequence in $\mathbb{R}^p$ with $E(\mathbb{X}_1) = \mathbf{0}$ and positive definite long-run covariance matrix $\boldsymbol{\Sigma}$. Furthermore, let a strong invariance principle be fufilled so that there exists a p-dimensional Wiener process $\boldsymbol{W}(t)$ with*

$$\left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=1}^{k} \mathbb{X}_i - \boldsymbol{W}(k) \right\| = O\left(k^{1/(2+\nu)}\right) \;\; a.s., \quad k \to \infty.$$

Moreover, let Assumption A.1.1 hold on the bandwidth G. Then,

$$\max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k+G} \mathbb{X}_i \right\| = O_P\left(\sqrt{\log(n/G)}\right).$$

Proof. Applying the invariance principle and the self-similarity of the Wiener process yields

$$\max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=k+1}^{k+G} \mathbb{X}_i \right\|$$

$$\begin{aligned}
&\leq \max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=k+1}^{k+G} \mathbb{X}_i - (\boldsymbol{W}(k+G) - \boldsymbol{W}(k)) \right\| \\
&\quad + \max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \|\boldsymbol{W}(k+G) - \boldsymbol{W}(k)\| \\
&\leq \frac{1}{\sqrt{G}} \left(\max_{0\leq k\leq n-G} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=1}^{k+G} \mathbb{X}_i - \boldsymbol{W}(k+G) \right\| + \max_{0\leq k\leq n-G} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=1}^{k} \mathbb{X}_i - \boldsymbol{W}(k) \right\| \right) \\
&\quad + \max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \|\boldsymbol{W}(k+G) - \boldsymbol{W}(k)\| \\
&= O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{G}} \right) + \max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \|\boldsymbol{W}(k+G) - \boldsymbol{W}(k)\| \\
&\overset{D}{=} O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{G}} \right) + \max_{0\leq k\leq n-G} \left\| \boldsymbol{W}\left(\frac{k}{G} + 1 \right) - \boldsymbol{W}\left(\frac{k}{G} \right) \right\| \\
&\leq O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{G}} \right) + \sup_{0\leq t\leq \frac{n}{G}-1} \sup_{0\leq s\leq 1} \|\boldsymbol{W}(t+s) - \boldsymbol{W}(t)\| \\
&= O_P\left(\frac{n^{1/(2+\nu)}}{\sqrt{G}} \right) + O_P\left(\sqrt{\log(n/G)} \right) = O_P\left(\sqrt{\log(n/G)} \right),
\end{aligned}$$

where the last line follows from Lemma E.2.2 and Assumption A.1.1. Finally, by Lemma E.1.5 we can conclude

$$\begin{aligned}
&\max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k+G} \mathbb{X}_i \right\| = \max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{1/2} \mathbf{\Sigma}^{-1/2} \sum_{i=k+1}^{k+G} \mathbb{X}_i \right\| \\
&\leq \left\| \mathbf{\Sigma}^{1/2} \right\|_F \max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=k+1}^{k+G} \mathbb{X}_i \right\| = O_P\left(\sqrt{\log(n/G)} \right).
\end{aligned}$$

□

Corollary E.2.13. *Let $\{\mathbb{X}_i\}_{i\geq 1}$ be a vector-valued sequence of type (E1) or type (E2) with $\mathbb{X}_i = (X_{1i}, \ldots, X_{pi})^T$ and $E(\mathbb{X}_1) = \mathbf{0}$. Furthermore, let Assumption A.1.1 hold on the bandwidth G. Then, if $E\left(\|\mathbb{X}_1\|^{2+\nu}\right) < \infty$ for some $\nu > 0$, we have*

$$\max_{0\leq k\leq n-G} \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k+G} \mathbb{X}_i \right\| = O_P\left(\sqrt{\log(n/G)} \right).$$

Proof. For a sequence of type (E1) the invariance principle proved by Einmahl (1989) in Theorem 2 can be applied to obtain

$$\left\| \mathbf{\Sigma}^{-1/2} \sum_{i=1}^{k} \mathbb{X}_i - \boldsymbol{W}(k) \right\| = O\left(k^{1/(2+\nu)} \right) \quad a.s., \quad k \to \infty,$$

with $\boldsymbol{W}(t)$ as a standard p-dimensional Wiener process. We get a similar result for sequences of type (E2) by using Theorem 4 of Kuelbs & Philipp (1980). Thus, Theorem E.2.12 completes the proof. □

Lemma E.2.14. *Let the assumptions of Theorem E.2.12 or Corollary E.2.13 hold. Then,*

(a) with $B_{j,n,G}^{(1)} := \{k \in \{1,\dots,n\} : k < k_{j,n} < k+G\}$,

$$\max_{k \in B_{j,n,G}^{(1)}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k_{j,n}} \mathbb{X}_i \right\| = O_P(1) \quad \text{and} \quad \max_{k \in B_{j,n,G}^{(1)}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k_{j,n}+1}^{k+G} \mathbb{X}_i \right\| = O_P(1),$$

(b) with $B_{j,n,G}^{(2)} := \{k \in \{1,\dots,n\} : k - G < k_{j,n} \le k\}$,

$$\max_{k \in B_{j,n,G}^{(2)}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k-G+1}^{k_{j,n}} \mathbb{X}_i \right\| = O_P(1) \quad \text{and} \quad \max_{k \in B_{j,n,G}^{(2)}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k_{j,n}+1}^{k} \mathbb{X}_i \right\| = O_P(1).$$

Proof. We only prove the first result of (a) since the second assertion in (a) and the assertions in (b) can be shown in an analogous manner.
With the stationarity of the sequence and by changing the index to $l = k - k_{j,n} + G$ we get

$$\begin{aligned}
\max_{k \in B_{j,n,G}^{(1)}} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=k+1}^{k_{j,n}} \mathbb{X}_i \right\| &= \max_{k_{j,n}-G<k<k_{j,n}} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=k+1}^{k_{j,n}} \mathbb{X}_i \right\| \\
= \max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=k_{j,n}-G+1+l}^{k_{j,n}} \mathbb{X}_i \right\| &\overset{D}{=} \max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=l+1}^{G} \mathbb{X}_i \right\|.
\end{aligned}$$

Furthermore, applying the invariance principle, which is directly given or can be derived under the assumptions of Corollary E.2.13, in combination with the self-similarity of the Wiener process results in

$$\begin{aligned}
&\max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=l+1}^{G} \mathbb{X}_i \right\| \\
&\le \max_{0<l<G} \frac{1}{\sqrt{G}} \left(\left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=1}^{G} \mathbb{X}_i - \boldsymbol{W}(G) \right\| + \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=1}^{l} \mathbb{X}_i - \boldsymbol{W}(l) \right\| \right) \\
&\quad + \max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{W}(G) - \boldsymbol{W}(l) \right\| \\
&= O_P\left(\frac{G^{1/(2+\nu)}}{\sqrt{G}} \right) + \max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{W}(G) - \boldsymbol{W}(l) \right\| \\
&\overset{D}{=} O_P\left(G^{-\nu/(4+2\nu)} \right) + \max_{0<l<G} \left\| \boldsymbol{W}(1) - \boldsymbol{W}\left(\frac{l}{G} \right) \right\| \\
&\le O_P\left(G^{-\nu/(4+2\nu)} \right) + \sup_{0 \le t \le 1} \left\| \boldsymbol{W}(1) - \boldsymbol{W}(t) \right\| = O_P(1),
\end{aligned}$$

where the last step follows from the almost sure continuity of paths of a Wiener process and the compactness of the considered interval $[0, 1]$. Finally, Lemma E.1.5 completes the proof since

$$\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k_{j,n}} \mathbb{X}_i \right\| = \max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{-1/2} \sum_{i=k+1}^{k_{j,n}} \mathbb{X}_i \right\|$$
$$\leq \left\| \mathbf{\Sigma}^{1/2} \right\|_F \max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=k+1}^{k_{j,n}} \mathbb{X}_i \right\| = O_P(1).$$

□

Lemma E.2.15. *Let the assumptions of Theorem E.2.12 or Corollary E.2.13 hold. Then,*

$$\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k-G+1}^{k} \mathbb{X}_i \right\| = O_P(1) \;\; \textit{and} \;\; \max_{k\in B^{(2)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k+1}^{k+G} \mathbb{X}_i \right\| = O_P(1).$$

Proof. We only prove the first result as the second assertion can be derived by using similar arguments.
With the stationarity of the sequence we obtain

$$\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=k-G+1}^{k} \mathbb{X}_i \right\| = \max_{k_{j,n}-G<k<k_{j,n}} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=k-G+1}^{k} \mathbb{X}_i \right\|$$
$$= \max_{k_{j,n}-2G<l<k_{j,n}-G} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=l+1}^{l+G} \mathbb{X}_i \right\| \overset{D}{=} \max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=l+1}^{l+G} \mathbb{X}_i \right\|.$$

Moreover, the invariance principle, which is directly given or can be derived under the assumptions of Corollary E.2.13, in connection with the self-similarity of the Wiener process shows

$$\max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=l+1}^{l+G} \mathbb{X}_i \right\|$$
$$\leq \max_{0<l<G} \frac{1}{\sqrt{G}} \left(\left\| \mathbf{\Sigma}^{-1/2} \sum_{i=1}^{l+G} \mathbb{X}_i - \boldsymbol{W}(l+G) \right\| + \left\| \mathbf{\Sigma}^{-1/2} \sum_{i=1}^{l} \mathbb{X}_i - \boldsymbol{W}(l) \right\| \right)$$
$$+ \max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{W}(l+G) - \boldsymbol{W}(l) \right\|$$
$$= O_P\left(\frac{G^{1/(2+\nu)}}{\sqrt{G}} \right) + \max_{0<l<G} \frac{1}{\sqrt{G}} \left\| \boldsymbol{W}(l+G) - \boldsymbol{W}(l) \right\|$$
$$\overset{D}{=} O_P\left(G^{-\nu/(4+2\nu)} \right) + \max_{0<l<G} \left\| \boldsymbol{W}\left(\frac{l}{G} + 1 \right) - \boldsymbol{W}\left(\frac{l}{G} \right) \right\|$$
$$\leq O_P\left(G^{-\nu/(4+2\nu)} \right) + \sup_{0\leq t\leq 1} \sup_{0\leq s\leq 1} \left\| \boldsymbol{W}(t+s) - \boldsymbol{W}(t) \right\| = O_P(1),$$

where the last step follows from the almost sure continuity of paths of a Wiener process and the compactness of the considered interval $[0,1]$. Finally, applying Lemma E.1.5

finishes the proof since

$$\max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \sum_{i=k-G+1}^{k} \mathbb{X}_i \right\| = \max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2} \sum_{i=k-G+1}^{k} \mathbb{X}_i \right\|$$
$$\leq \left\| \boldsymbol{\Sigma}^{1/2} \right\|_F \max_{k\in B^{(1)}_{j,n,G}} \frac{1}{\sqrt{G}} \left\| \boldsymbol{\Sigma}^{-1/2} \sum_{i=k-G+1}^{k} \mathbb{X}_i \right\| = O_P(1).$$

□

Theorem E.2.16.
Let the parameter space $\boldsymbol{\Theta}$ be compact and let Assumption A.1.1 hold on the bandwidth G. Furthermore, assume that $\{\mathbb{X}_i\}_{i\geq 1}$ is a sequence of type (E1) or type (E2) and let $\boldsymbol{F} : (\mathbb{R}^p, \boldsymbol{\Theta}) \to \mathbb{R}^d$ be a measurable function with respect to $\mathbb{X}_i$ such that the following assumptions are fulfilled:

(i) Let $E\left(\|\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\|^{2+\nu}\right) < \infty$ hold for some $\nu > 0$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(ii) Let $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\theta})$ be Lipschitz continuous in $\boldsymbol{\theta}$, i.e. there exists a function $L(\boldsymbol{x}) > 0$, which is measurable with respect to $\mathbb{X}_i$, such that

$$\|\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\xi})\| \leq L(\boldsymbol{x}) \|\boldsymbol{\theta} - \boldsymbol{\xi}\|$$

for all $\boldsymbol{\theta}, \boldsymbol{\xi} \in \boldsymbol{\Theta}$ and

(iii) $E\left(|L(\mathbb{X}_1)|^{2+\nu}\right) < \infty$ for some $\nu > 0$.

Then,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{0\leq k\leq n-G} \frac{1}{G} \left\| \sum_{i=k+1}^{k+G} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o_P(1).$$

Proof. Similar arguments as in the proof of Theorem E.2.8 can be used here. Argument (1) is the same as in the proof of Thereom E.2.8. We need to modify argument (2) and (3) slightly to:

(2') Let $\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) = \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right)$. Since $\boldsymbol{F}$ is measurable with respect to $\mathbb{X}_i$, the function $\boldsymbol{F}_0$ is measurable as well. Thus, the pattern of the original sequence $\{\mathbb{X}_i\}_{i\geq 1}$ described by type (E1) or type (E2) is inherited by the transformed sequence $\{\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta})\}_{i\geq 1}$. Hence, for fixed M and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$, Assumption (i) and Corollary E.2.13 yield

$$\sup_{m\leq M} \max_{0\leq k\leq n-G} \frac{1}{G} \left\| \sum_{i=k+1}^{k+G} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m) \right\| \leq \sum_{m=1}^{M} \max_{0\leq k\leq n-G} \frac{1}{G} \left\| \sum_{i=k+1}^{k+G} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m) \right\|$$
$$= O_P\left(\frac{\sqrt{\log(n/G)}}{\sqrt{G}} \right) = o_P(1),$$

where the last line follows from Assumption A.1.1.

(3') For any $\boldsymbol{\theta}, \boldsymbol{\xi} \in \boldsymbol{\Theta}$ with $\|\boldsymbol{\theta} - \boldsymbol{\xi}\| < \delta$ we obtain

$$\frac{1}{G} \sum_{i=k+1}^{k+G} \|\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\xi})\| \leq \delta \left(\frac{1}{G} \sum_{i=k+1}^{k+G} (L(\mathbb{X}_i) - E(L(\mathbb{X}_1))) + 2E(L(\mathbb{X}_1)) \right),$$

which is given by Assumption (ii).

Now, for each $\varepsilon > 0$ we can choose a $\delta > 0$ such that $\frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1)) > 0$ by Assumption (iii). Note that Corollary E.2.13 can be applied to the sequence $\{L(\mathbb{X}_i)\}_{i\geq 1}$ with Assumption (iii) and since the measurability of L with respect to $\mathbb{X}_i$ ensures that the pattern of the original sequence carries over such that the sequence $\{L(\mathbb{X}_i)\}_{i\geq 1}$ is of type (E1) or type (E2) as well. Hence, in connection with (2') and (3') we get

$$\begin{aligned}
&P\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{1\leq k\leq n-G} \frac{1}{G} \left\| \sum_{i=k+1}^{k+G} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) \right\| > \varepsilon \right) \\
&= P\left(\sup_{m\leq M} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_m\|} \max_{1\leq k\leq n-G} \frac{1}{G} \left\| \sum_{i=k+1}^{k+G} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) \right\| > \varepsilon \right) \\
&= P\left(\sup_{m\leq M} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_m\|} \max_{1\leq k\leq n-G} \left\| \frac{1}{G} \sum_{i=k+1}^{k+G} (\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m)) + \frac{1}{G} \sum_{i=k+1}^{k+G} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m) \right\| > \varepsilon \right) \\
&\leq P\left(\sup_{m\leq M} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_m\|} \max_{1\leq k\leq n-G} \left\| \frac{1}{G} \sum_{i=k+1}^{k+G} (\boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}) - \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m)) \right\| > \frac{\varepsilon}{2} \right) \\
&\quad + P\left(\sup_{m\leq M} \max_{1\leq k\leq n-G} \left\| \frac{1}{G} \sum_{i=k+1}^{k+G} \boldsymbol{F}_0(\mathbb{X}_i, \boldsymbol{\theta}_m) \right\| > \frac{\varepsilon}{2} \right) \\
&\leq P\left(\max_{1\leq k\leq n-G} \frac{1}{G} \sum_{i=k+1}^{k+G} (L(\mathbb{X}_i) - E(L(\mathbb{X}_1))) > \frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1)) \right) + o(1) = o(1),
\end{aligned}$$

since $\frac{\varepsilon}{2\delta} - 2E(L(\mathbb{X}_1) > 0$ by the choice of δ.

□

Lemma E.2.17. *Let the assumptions of Theorem E.2.16 hold. Then, if* $E\left(\|\boldsymbol{X}_1\|^{2+\nu}\right) < \infty$ *for some* $\nu > 0$, *it holds*

(a)

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{k\in B^{(1)}_{j,n,G}} \frac{1}{G} \left\| \sum_{i=k+1}^{k_{j,n}} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o_P(1) \quad \textit{and}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{k\in B^{(1)}_{j,n,G}} \frac{1}{G} \left\| \sum_{i=k_{j,n}+1}^{k+G} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o_P(1),$$

(b)

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{k\in B^{(2)}_{j,n,G}} \frac{1}{G} \left\| \sum_{i=k-G+1}^{k_{j,n}} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o_P(1) \quad \textit{and}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{k\in B^{(2)}_{j,n,G}} \frac{1}{G} \left\| \sum_{i=k_{j,n}+1}^{k} \boldsymbol{F}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{F}(\mathbb{X}_1, \boldsymbol{\theta})\right) \right\| = o_P\left(1\right).$$

Proof. These results can be derived in analogous manner to Theorem E.2.16 by applying Lemma E.2.14 instead of Corollary E.2.13. □

Lemma E.2.18. *Let $(\Omega, \mathfrak{A}, P)$ be a probability space and $A_i \in \mathfrak{A}$ be events.Then, for $n \in \mathbb{N}$*

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - n + 1.$$

Proof. The assertion can be proved by induction. We start with $n = 2$:

$$P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) \geq P(A_1) + P(A_2) - 1,$$

which shows the base clause. The induction hypothesis is that

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - n + 1.$$

holds for $n \in \mathbb{N}$. Hence, by using the base clause and the hypothesis we obtain

$$\begin{aligned} P\left(\bigcap_{i=1}^{n+1} A_i\right) &\geq P\left(\bigcap_{i=1}^{n} A_i\right) + P(A_{n+1}) - 1 \\ &\geq \sum_{i=1}^{n} P(A_i) - n + 1 + P(A_{n+1}) - 1 = \sum_{i=1}^{n+1} P(A_i) - (n+1) + 1, \end{aligned}$$

which completes the proof. □

Theorem E.2.19. *Let $\{a_{k,n}\}$, $1 \leq k \leq n$, be a sequence of random variables on some probabilty space $(\Omega, \mathcal{B}, P)$ with values in $A \subset \mathbb{R}$ which converges to some deterministic value $a \in A$ uniformly in k as $n \to \infty$, i.e. $\sup_k |a_{k,n} - a| = o_P(1)$. Furthermore, let f denote a continuous function on A. Then,*

$$\sup_k |f(a_{k,n}) - f(a)| = o_P(1).$$

Proof. The result can be shown by applying the subsequence principle. For each subsequence $\sup_k \left|a_{k,\alpha(n)} - a\right|$ there exists a further subsequence $\sup_k \left|a_{k,\beta(\alpha(n))} - a\right|$ such that

$$\sup_k \left|a_{k,\beta(\alpha(n))} - a\right| \overset{a.s.}{\to} 0,$$

which means that $\sup_k \left|a_{k,\beta(\alpha(n))}(\omega) - a\right| \to 0$ holds for all $\omega \in \Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$. This implies that

$$\forall \delta > 0 \quad \exists n_0 \in \mathbb{N} \quad \forall n \geq n_0: \quad \sup_k \left|a_{k,\beta(\alpha(n))}(\omega) - a\right| < \delta. \tag{E.2}$$

Furthermore, by the continuity of f in a we get

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x \text{ with } |x - a| < \delta : \quad |f(x) - f(a)| < \frac{\varepsilon}{2}. \tag{E.3}$$

The statements in (E.2) and (E.3) can be combined as follows. For every $\varepsilon > 0$ with corresponding δ as above there exists an $n_0 \in \mathbb{N}$ such that $|a_{k,\beta(\alpha(n))}(\omega) - a| < \delta$ for all $n \geq n_0$ and all $1 \leq k \leq n$ implying that

$$\sup_k |f(a_{k,\beta(\alpha(n))}(\omega)) - f(a)| < \varepsilon, \quad \forall n \geq n_0,$$

which shows the convergence $\sup_k |f(a_{k,\beta(\alpha(n))}(\omega)) - f(a)| \to 0$. Hence, we obtain $\sup_k |f(a_{k,\beta(\alpha(n))}) - f(a)| \overset{a.s.}{\to} 0$ and the subsequence principle can be applied again to receive $\sup_k |f(a_{k,n}) - f(a)| \overset{P}{\to} 0$ ☐

Corollary E.2.20. *Let $\{\boldsymbol{B}_{k,n}\}$ be a random sequence of $p \times p$ matrices. If $\boldsymbol{B}_{k,n} \overset{P}{\to} \boldsymbol{B}$ holds uniformly in k, where $\boldsymbol{B}$ is a regular deterministic matrix, then*

$$\boldsymbol{B}_{k,n}^{-1} \overset{P}{\to} \boldsymbol{B}^{-1} \quad \textit{uniformly in } k.$$

Proof. First note that $\boldsymbol{B}_{k,n}$ is invertible for large n as by the continuity of the determinant and Theorem E.2.19 $|\det(\boldsymbol{B}_{k,n}) - \det(\boldsymbol{B})| = o_P(1)$ implying that $|\det(\boldsymbol{B}_{k,n})| = |\det(\boldsymbol{B})| + o_P(1)$ holds uniformly in k. Furthermore, since the matrix inverse $f(\boldsymbol{A}) = \boldsymbol{A}^{-1}$ is a continuous function on the elements of $\boldsymbol{A}$ by Theorem 5.19 in Schott (1997) on page 188 applying Theorem E.2.19 shows the assertion. ☐

Lemma E.2.21. *Let $\{\boldsymbol{c}_{k,n}\}$ be a sequence of random vectors in $\mathbb{R}^p$. Furthermore, let $\{a_n\}$ be a deterministic positive sequence and let $\{\boldsymbol{B}_{k,n}\}$ denote a deterministic sequence of $\mathbb{R}^{p\times p}$ matrices satisfying $\sup_k \|\boldsymbol{B}_{k,n}\|_F < \infty$. Then, if*

- $\sup_k \|\boldsymbol{c}_{k,n}\| = O_P(a_n)$, *it holds that*

$$\boldsymbol{B}_{k,n}\boldsymbol{c}_{k,n} = O_P(a_n) \textit{ holds uniformly in } k.$$

- $\sup_k \|\boldsymbol{c}_{k,n}\| = o_P(a_n)$, *it holds that*

$$\boldsymbol{B}_{k,n}\boldsymbol{c}_{k,n} = o_P(a_n) \textit{ holds uniformly in } k.$$

Proof. By Lemma E.1.5 we obtain

$$\sup_k \|\boldsymbol{B}_{k,n}\boldsymbol{c}_{k,n}\| \leq \sup_k \|\boldsymbol{B}_{k,n}\|_F \sup_k \|\boldsymbol{c}_{k,n}\| = O\left(\sup_k \|\boldsymbol{c}_{k,n}\|\right),$$

hence the assertions follow. ☐

Lemma E.2.22. *Let $\{\boldsymbol{a}_{k,n}\}$ be a sequence of random vectors. Furthermore, let a_n be a deterministic sequence and $\boldsymbol{B}$ a regular deterministic matrix.*

- *Then,*

$$O_P(a_n) = (o_P(1) + \boldsymbol{B})\, \boldsymbol{a}_{k,n}, \quad \textit{uniformly in } k,$$

implies that

$$\boldsymbol{a}_{k,n} = O_P(a_n) \textit{ holds uniformly in } k.$$

- *Then,*

$$o_P(a_n) = (o_P(1) + \boldsymbol{B})\,\boldsymbol{a}_{k,n}, \quad \textit{uniformly in } k,$$

implies that

$$\boldsymbol{a}_{k,n} = o_P(a_n) \textit{ holds uniformly in } k.$$

Proof. By Corollary E.2.20 we can multiply both sides of the equation above with the inverse of $(o_P(1) + \boldsymbol{B})$ and get

$$(o_P(1) + \boldsymbol{B}^{-1})O_P(a_n) = \boldsymbol{A}_{k,n}, \quad \text{uniformly in k},$$

which shows the assertion as $\boldsymbol{B}^{-1} = O(1)$. □

Bibliography

Aggarwal, Reena, Inclan, Carla, & Leal, Ricardo. 1999. Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis*, **34**(1), 33–55.

Andrews, Donald WK. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**(4), 821–856.

Aue, Alexander. 2003. *Sequential change-point analysis based on invariance principles.* Ph.D. thesis, Universität zu Köln.

Bai, Jushan, & Perron, Pierre. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**(1), 47–78.

Bai, Jushan, & Perron, Pierre. 2003. Computation and analysis of multiple structural change models. *Journal of applied econometrics*, **18**, 1–22.

Bauer, Heinz. 2001. *Wahrscheinlichkeitstheorie.* Walter de Gruyter.

Bauer, Peter, & Hackl, Peter. 1980. An extension of the MOSUM technique for quality control. *Technometrics*, **22**(1), 1–7.

Bradley, Richard C. 2007. *Introduction to strong mixing conditions.* Vol. 1. Kendrick Press.

Braun, J. V., Braun, R. K., & Müller, H.-G. 2000. Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika*, **87**(2), 301–314.

Cho, Haeran, & Kirch, Claudia. 2018. *Multiscale MOSUM procedure with localised pruning.* Unpublished Manuscript.

Chu, Chia-Shang J., Hornik, Kurt, & Kaun, Chung-Ming. 1995. MOSUM tests for parameter constancy. *Biometrika*, **82**(3), 603–617.

Csörgö, Miklós, & Horváth, Lajos. 1997. *Limit theorems in change-point analysis.* John Wiley & Sons Inc.

Csörgo, Miklos, & Révész, Pál. 1981. *Strong approximations in probability and statistics.* Akadémiai Kiadó and Academic Press.

Davis, Richard A., Lee, Thomas C. M., & Rodriguez-Yam, Gabriel A. 2006. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, **101**(473), 223–239.

Doukhan, Paul, & Kengne, William. 2013. Inference and testing for structural change in time series of counts model. *arXiv preprint arXiv:1305.1751*.

Eichinger, Birte, & Kirch, Claudia. 2018. A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, **24**(1), 526–564.

Einmahl, Uwe. 1989. Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of multivariate analysis*, **28**(1), 20–68.

Ferland, René, Latour, Alain, & Oraichi, Driss. 2006. Integer-valued GARCH process. *Journal of Time Series Analysis*, **27**(6), 923–942.

Fokianos, Konstantinos, Rahbek, Anders, & Tjøstheim, Dag. 2009. Poisson autoregression. *Journal of the American Statistical Association*, **104**(488), 1430–1439.

Franke, Jürgen, Kirch, Claudia, & Kamgaing, Joseph Tadjuidje. 2012. Changepoints in times series of counts. *Journal of Time Series Analysis*, **33**(5), 757–770.

Fryzlewicz, Piotr. 2014. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, **42**(6), 2243–2281.

Hawkins, D.L. 1987. A test for a change point in a parametric model based on a maximal Wald-type statistic. *Sankhyā: The Indian Journal of Statistics, Series A*, **49**(3), 368–376.

Hawkins, D.L. 1989. A UI approach to retrospective testing for shifting parameters in a linear model. *Communications in Statistics-Theory and Methods*, **18**(8), 3117–3134.

Heinen, Andréas. 2003. *Modelling time series count data: an autoregressive conditional Poisson model.* Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1117187`.

Horn, Roger A., & Johnson, Charles R. 1990. *Matrix analysis.* Cambridge University Press.

Horn, Roger A., & Johnson, Charles R. 1991. *Topics in matrix analysis.* Cambridge University Press.

Horváth, Lajos. 1993. Change in autoregressive processes. *Stochastic processes and their Applications*, **44**(2), 221–242.

Horváth, Lajos, & Shao, Qi-Man. 1995. Limit theorems for the union-intersection test. *Journal of statistical planning and inference*, **44**(2), 133–148.

Hušková, Marie. 1990. Asymptotics for robust MOSUM. *Commentationes Mathematicae Universitatis Carolinae*, **31**(2), 345–356.

Hušková, Marie. 1996. Tests and estimators for the change point problem based on M-statistics. *Statistics & Decisions*, **14**, 115–136.

Hušková, Marie, & Slabý, Aleš. 2001. Permutation tests for multiple changes. *Kybernetika*, **37**(5), 605–622.

Hušková, Marie, Prášková, Zuzana, & Steinebach, Josef. 2007. On the detection of changes in autoregressive time series I. Asymptotics. *Journal of Statistical Planning and Inference*, **137**(4), 1243–1259.

Killick, Rebecca, & Eckley, Idris A. 2014. changepoint: An R package for changepoint analysis. *Journal of statistical software*, **58**(3), 1–19.

Killick, Rebecca, Eckley, Idris A., Ewans, Kevin, & Jonathan, Philip. 2010. Detection of changes in variance of oceanographic time-series using changepoint analysis. *Ocean Engineering*, **37**, 1120–1126.

Killick, Rebecca, Fearnhead, Paul, & Eckley, Idris A. 2012. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, **107**(500), 1590–1598.

Kirch, Claudia. 2008. *Lecture notes – Introduction to Change-Point Analysis.*

Kirch, Claudia, & Tadjuidje Kamgaing, Josef. 2016. Detection of change points in discrete valued time series. *Pages 219–244 of:* Davis, Richard A., Holan, Scott H., Lund, Robert, & Ravishanker, Nalini (eds), *Handbook of discrete valued time series.* CRC Press.

Komlós, János, Major, Péter, & Tusnády, Gábor. 1975. An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **32**, 111–131.

Komlós, János, Major, Péter, & Tusnády, Gábor. 1976. An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **34**, 33–58.

Kuelbs, James, & Philipp, Walter. 1980. Almost sure invariance principles for partial sums of mixing B-valued random variables. *The Annals of Probability*, **8**(6), 1003–1036.

Kühn, Christoph. 2001. An estimator of the number of change points based on a weak invariance principle. *Statistics & probability letters*, **51**, 189–196.

Liu, Jian, Wu, Shiying, & Zidek, James V. 1997. On segmented multivariate regression. *Statistica Sinica*, **7**, 497–525.

Major, Péter. 1976. The approximation of partial sums of independent rv's. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **35**, 213–220.

Marshall, Albert W., Olkin, Ingram, & Arnold, Barry C. 2011. *Inequalities: theory of majorization and its applications.* Springer.

Messer, Michael, Kirchner, Marietta, Schiemann, Julia, Roeper, Jochen, Neininger, Ralph, & Schneider, Gaby. 2014. A multiple filter test for the detection of rate changes in renewal processes with varying variance. *The Annals of Applied Statistics*, **8**(4), 2027–2067.

Muhsal, Birte Chantal Simone. 2013. *Change-point methods for multivariate autoregressive models and multiple structural breaks in the mean.* Ph.D. thesis, Karlsruher Institut für Technologie (KIT).

Neumann, Michael H. 2011. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli*, **17**(4), 1268–1284.

Pan, Jianmin, & Chen, Jiahua. 2006. Application of modified information criterion to multiple change point problems. *Journal of multivariate analysis*, **97**, 2221–2241.

Perron, Pierre, & Qu, Zhongjun. 2006. Estimating restricted structural change models. *Journal of Econometrics*, **134**, 373–399.

Pötscher, Benedikt M., & Prucha, Ingmar R. 1997. *Dynamic nonlinear econometric models: Asymptotic theory.* Springer-Verlag.

Rao, R. Ranga. 1962. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, **33**(2), 659–680.

Rosenblatt, Murray. 1956. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences*, **42**(1), 43–47.

Roy, Anindya, & Banerjee, Sudipto. 2014. *Linear algebra and matrix analysis for statistics.* Chapman and Hall/CRC.

Schott, James R. 1997. *Matrix analysis for statistics.* John Wiley & Sons.

Steinebach, Josef, & Eastwood, Vera R. 1996. Extreme value asymptotics for multivariate renewal processes. *Journal of multivariate analysis*, **56**(2), 284–302.

Van der Vaart, Aad W. 2007. *Asymptotic statistics.* Cambridge University Press.

Vostrikova, Ludmilla Ju. 1981. Detecting disorder in multidimensional random processes. *Soviet Mathematics Doklady*, **24**, 55–59.

Weiß, Christian H. 2009. Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, **18**(4), 507–519.

Weiß, Christian H. 2010. The INARCH (1) model for overdispersed time series of counts. *Communications in Statistics-Simulation and Computation*, **39**(6), 1269–1291.

Yao, Yi-Ching. 1988. Estimating the number of change-points via Schwarz'criterion. *Statistics & Probability Letters*, **6**, 181–189.

Yao, Yi-Ching, & Au, Siu-Tong. 1989. Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics, Series A*, **51**(3), 370–381.

Zeileis, Achim, Leisch, Friedrich, Hornik, Kurt, & Kleiber, Christian. 2002. strucchange: An R package for testing for structural change in linear regression models. *Journal of statistical software*, **7**(2), 1–38.

Zhu, Fukang, & Wang, Dehui. 2010. Diagnostic checking integer-valued ARCH (p) models using conditional residual autocorrelations. *Computational Statistics & Data Analysis*, **54**(2), 496–508.

Notation

$X_1, \ldots, X_n$	Observations
$\{\mathbb{X}_i : i \geq 1\}$	Vector-valued sequence of observations
$\boldsymbol{\theta}$	Parameter of interest
$\boldsymbol{\Theta}$	Parameter space
p	Dimension of the parameter space
$\boldsymbol{H}$	Estimating function
$T_{k,n}(G, \widetilde{\boldsymbol{\theta}})$	MOSUM score-type statistic according to Defintion 2.0.1
$T_n(G, \widetilde{\boldsymbol{\theta}})$	Maximum of the MOSUM score-type statistic: $\max_{G \leq k \leq n-G} T_{k,n}(G, \widetilde{\boldsymbol{\theta}})$
G	Bandwidth
$\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}$	$\sum_{i=k+1}^{k+G} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}}) - \sum_{i=k-G+1}^{k} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})$ as in Defintion 2.0.1
$\boldsymbol{\Sigma}_k$	Long-run covariance matrix of $\boldsymbol{H}(\mathbb{X}_k, \widetilde{\boldsymbol{\theta}})$ as in Defintion 2.0.1
$\boldsymbol{\Sigma}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}$	Long-run covariance matrix of $\boldsymbol{H}(\mathbb{X}_1, \widetilde{\boldsymbol{\theta}})$
$\widehat{\boldsymbol{\Sigma}}_{k,n}$	Estimator of $\boldsymbol{\Sigma}_k$
$\widehat{T}_{k,n}(G, \widetilde{\boldsymbol{\theta}})$	MOSUM score-type statistic with estimated long-run covariance matrix: $\frac{1}{\sqrt{2G}}\sqrt{\boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}^T \widehat{\boldsymbol{\Sigma}}_{k,n}^{-1} \boldsymbol{A}_{\widetilde{\boldsymbol{\theta}},k}}$
$\widehat{T}_n(G, \widetilde{\boldsymbol{\theta}})$	Corresponding maximum
$\Vert\cdot\Vert_F$	Frobenius norm
$\Vert\cdot\Vert$	Euclidean norm
$\boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta})$	$\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\right)$
$\nabla \boldsymbol{H}_0(\mathbb{X}_i, \boldsymbol{\theta})$	Matrix of centered gradient vectors $\nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\nabla \boldsymbol{H}(\mathbb{X}_i, \boldsymbol{\theta})\right)$
$\nabla^2 H_{j,0}(\mathbb{X}_i, \boldsymbol{\theta})$	Centered Hessian matrix $\nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\theta}) - E\left(\nabla^2 H_j(\mathbb{X}_i, \boldsymbol{\theta})\right)$
$\boldsymbol{V}(\boldsymbol{\theta})$	Expectation matrix of the gradient vectors under the null hypothesis $E\left(\nabla \boldsymbol{H}(\mathbb{X}_1, \boldsymbol{\theta})\right)^T$
$\boldsymbol{V}_j(\boldsymbol{\theta})$	Expectation matrix of the gradient vectors under the alternative $E\left(\nabla \boldsymbol{H}(\mathbb{X}_1^{(j)}, \boldsymbol{\theta})\right)^T$
$\widetilde{\boldsymbol{\theta}}_n$, $\widehat{\boldsymbol{\theta}}_n$	Estimator of $\boldsymbol{\theta}$
$\widehat{\boldsymbol{\theta}}_{l,u}$	Z-estimator (M-estimator) computed on the subsample $X_l, \ldots, X_u$
$\boldsymbol{S}(k, \widetilde{\boldsymbol{\theta}})$	Partial sum process $\sum_{i=1}^{k} \boldsymbol{H}(\mathbb{X}_i, \widetilde{\boldsymbol{\theta}})$
$\{\boldsymbol{W}(t) : t \geq 0\}$	p-dimensional standard Wiener process with identity matrix $\boldsymbol{I}_p$ as covariance matrix
$a(x), b(x)$	Normalizing functions defined in (2.1)
$\Gamma(x)$	Gamma function
$A_{n,G}$	Set of time points defined in (2.2) $\{k \in \{G, \ldots, n-G\} : \lvert k - \lfloor \lambda_j n \rfloor \rvert \geq G \; \forall \; j \in \{1, \ldots, q\}\}$
$B_{n,G}$	Set of time points defined in (2.3) $\{k \in \{G, \ldots, n-G\} : \exists \; j \in \{1, \ldots, q\} : \lvert k - k_{j,n} \rvert \leq G\}$
$\{\boldsymbol{\Sigma}_{A,k}\}$	Sequence of positive definite matrices
$j(k)$	Index of the closest change point to k
$\boldsymbol{d}_j$	Difference in exected values $E\left(\boldsymbol{H}(\mathbb{X}_1^{(j+1)}, \widetilde{\boldsymbol{\theta}})\right) - E\left(\boldsymbol{H}(\mathbb{X}_1^{(j)}, \widetilde{\boldsymbol{\theta}})\right)$
$(v_{j,n}, w_{j,n})$	Pair of time points considered in the MOSUM procedure

	fulfilling (2.9) to (2.11)
$\widehat{q}_n$	Estimator of the number of change points
$\widehat{q}_n(\boldsymbol{\theta})$	Estimator of the number of change points based on $T_{k,n}(G,\boldsymbol{\theta})$
$\widehat{k}_{j,n}$	Estimator for the location of a change point
$\widehat{k}_{j,n}(\boldsymbol{\theta})$	Estimator for the location based on $T_{k,n}(G,\boldsymbol{\theta})$
$\widetilde{Q}$	Set of indices of detectable changes
$\widetilde{q}$	Number of detectable changes
$\widetilde{A}_{n,G}$	Set of time points defined in (2.12) $\left\{k \in \{G,\ldots,n-G\} : \lvert k-k_{j,n}\rvert \geq G \;\forall\, j \in \widetilde{Q}\right\}$
$\bar{B}_{n,G}$	Set of time points defined in (2.13) $\left\{k \in \{G,\ldots,n-G\} : \exists\, j \in \widetilde{Q} : \lvert k-k_{j,n}\rvert < (1-\varepsilon)G\right\}$
$\widehat{\boldsymbol{\Sigma}}_{j,n}$	Global estimator of the long-run covariance matrix $\boldsymbol{\Sigma}$ satisfying Assumption A.2.12
$\boldsymbol{\Sigma}_{A,j}$	Positive definite matrix as in Assumption A.2.12
$\bar{k}_{j,n}$	Estimator for the location of a change point defined in (2.21)
$\boldsymbol{E}_1(k,G,\boldsymbol{\theta})$ and $\boldsymbol{E}_2(k,G,\boldsymbol{\theta})$	Defined in (2.19)
Type (E1)	i.i.d. sequence of random vectors
Type (E2)	Stationary and strongly mixing sequence of random vectors with a mixing rate $\alpha(n)$ satisfying $\alpha(n) = O(n^{-\beta})$ for some $\beta > 1 + 2/\nu$, where ν is as in Assumption A.1.3
$\widehat{\boldsymbol{\theta}}^{(n)}_{\gamma_1,\gamma_2}$	General Z-estimator defined as the solution of (2.25)
$\boldsymbol{\theta}_0$	Unique zero of $E\left(\boldsymbol{H}(\mathbb{X}_1,\boldsymbol{\theta})\right)$ (in the strict sense according to Definition E.2.9) under the null hypothesis
$\boldsymbol{\theta}_{\gamma_1,\gamma_2}$	Unique zero of (2.35) under the alternative
$W_{k,n}(G)$	MOSUM Wald-type statistic according to Defintion 3.0.1
$W_n(G)$	Maximum of the MOSUM score-type statistic: $\max_{G \leq k \leq n-G} W_{k,n}(G)$
$\boldsymbol{\Gamma}_k$	Asymptotic covariance matrix of $\sqrt{G}\widehat{\boldsymbol{\theta}}_{k-G+1,k}$ specified in (3.1)
$\boldsymbol{\theta}_j$	Unique zero of $E\left(\boldsymbol{H}\left(\mathbb{X}_1^{(j)},\boldsymbol{\theta}\right)\right)$ (in the strict sense according to Definition E.2.9) under the alternative
$\delta_{j,n}$	Distance between two adjacent change points $k_{j,n} - k_{j-1,n}$
$sBIC$	Information criterion given in (5.1)
$gRSS$	Generalized sum of squared residuals defined in (5.2) and (5.4)
ξ_n	Penalty specified in (5.3)
$\mathcal{G}$	Set of bandwidths
$\mathcal{L} = \mathcal{L}(\mathcal{G})$	Set of initial candidates obtained from the bandwidths of $\mathcal{G}$
u_n	Half of the minimal distance between two adjacent structural breaks defined in (5.35)
$\mathcal{V}_{j,n}$	Set of valid candidates for a change point $k_{j,n}$ as in (5.36)
$\mathcal{V}^*_{j,n}$	Set of strictly valid candidates for a change point $k_{j,n}$ as in (5.37)