Asymptotics for change-point tests and change-point estimators

Vom Fachbereich Mathematik

der Technischen Universität Kaiserslautern

zur Verleihung des akademischen Grades

Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)

genehmigte Dissertation

Stefanie Schwaar

16. Dezember 2016

Gutachter: Prof. Dr. Jürgen Franke
Technische Universität Kaiserslautern
Prof. Dr. Claudia Kirch
Otto-von-Guericke-Universität Magdeburg

D 386

Danksagung

An dieser Stelle möchte ich allen Arbeitskollegen, Bekannten und Freunden sowie meiner Familie, die mich in der Zeit meiner Promotion unterstützt haben einen großen Dank aussprechen. Ein besonderer Dank geht an meinen Betreuer Prof. Jürgen Franke, meine Gutachterin Prof. Claudia Kirch und Dr. Jean-Pierre Stockis.

Für die große Unterstützung nicht nur im wissenschaftlichen Bereich danke ich von Herzen Dr. Martin Anders. Auch meinen Eltern und meiner Schwester danke ich für Ihre Hilfe und Stütze auf meinem bisherigen Weg.

Abstract

In change-point analysis the point of interest is to decide if the observations follow one model or if there is at least one time-point, where the model has changed. This results in two subfields, the testing of a change and the estimation of the time of change. This thesis considers both parts but with the restriction of testing and estimating for at most one change-point.

A well known example is based on independent observations having one change in the mean. Based on the likelihood ratio test a test statistic with an asymptotic Gumbel distribution was derived for this model. As it is a well-known fact that the corresponding convergence rate is very slow, modifications of the test using a weight function were considered. Those tests have a better performance. We focus on this class of test statistics.

The first part gives a detailed introduction to the techniques for analysing test statistics and estimators. Therefore we consider the multivariate mean change model and focus on the effects of the weight function. In the case of change-point estimators we can distinguish between the assumption of a fixed size of change (fixed alternative) and the assumption that the size of the change is converging to 0 (local alternative). Especially, the fixed case in rarely analysed in the literature. We show how to come from the proof for the fixed alternative to the proof of the local alternative. Finally, we give a simulation study for heavy tailed multivariate observations.

The main part of this thesis focuses on two points. First, analysing test statistics and, secondly, analysing the corresponding change-point estimators. In both cases, we first consider a change in the mean for independent observations but relaxing the moment condition. Based on a robust estimator for the mean, we derive a new type of change-point test having a randomized weight function. Secondly, we analyse non-linear autoregressive models with unknown regression function. Based on neural networks, test statistics and estimators are derived for correctly specified as well as for misspecified situations. This part extends the literature as we analyse test statistics and estimators not only based on the sample residuals. In both sections, the section on tests and the one on the change-point estimator, we end with giving regularity conditions on the model as well as the parameter estimator.

Finally, a simulation study for the case of the neural network based test and estimator is given. We discuss the behaviour under correct and mis-specification and apply the neural network based test and estimator on two data sets.

Abstract

Die Change-point Analyse beschäftigt sich mit der Analyse von Beobachtungen hinsichtlich Veränderungen. Im Fokus steht die Fragestellung ob die Beobachtungen einem Modell folgt oder ob es Zeitabschnitte gibt in denen ein anderes Modell zugrunde liegt. Hieraus ergeben sich zwei Teilgebiete, eines welches sich mit dem Testen auf Modellwechsel beschäftigt und ein anderes, welches das Schätzen des Zeitpunktes zum Ziel hat. Diese Arbeit beschäftigt sich mit beiden Gebieten, jedoch mit der Einschränkung das maximal ein Zeitpunkt der Änderung erwartet wird.

Ein bekanntes Beispiel basiert auf unabhängigen Beobachtungen und betrachtet die Änderungen im Mittelwert. Hier erhält man auf Basis des Likelihood- ratio Tests eine Statistik deren asymptotische Verteilung Gumbel ist. Aus der Extremwerttheorie ist bekannt, dass die Konvergenz gegen diese Verteilung sehr langsam erfolgt. Daher wurde die Teststatistik mit Gewichtsfunktionen modifiziert. Diese Tests haben ein besseres Konvergenzverhalten. Wir untersuchen hier diese Klasse von Teststatistiken.

Der erste Abschnitt dieser Arbeit gibt eine Einführung in die verwendeten Techniken zur Analyse von Change-point Tests und Schätzern. Wir nutzen hierfür das mehrdimensionale Modell einer Mittelwertänderung bei unabhängigen Beobachtungen. Ein Schwerpunkt dieses Abschnittes liegt auf dem Verständnis der Gewichtsfunktion. Für den Schätzer des Zeitpunktes der Änderung wird unterschieden zwischen der Annahme einer festen Änderungsgröße (feste Alternative) und der Annahme das diese Änderung asymptotisch verschwindet (lokale Alternative). Für den Fall der festen Alternative gibt es kaum Literatur. Wir zeigen hier, wie der Beweis für die feste Alternative die Grundlage zum Beweis der lokalen Alternative bildet. Abschließend geben wir die Resultate einer Simulationsstudie für mehrdimensionale heavy-tailed Beobachtungen.

Im Hauptteil der Arbeit werden zuerst Teststatistiken und anschließend die zugehörigen Schätzer untersucht. Hierbei wird jeweils zuerst eine Mittelwertsänderung für unabhängige Beobachtungen untersucht, wobei Momentenannahmen abgeschwächt wurden. Basierend auf robusten Schätzern für Mittelwerte, erhalten wir eine neue Art von Change-point Tests deren Gewichtsfunktion zufällig ist. Das zweite Modell untersucht nichtlineare (auto-)regressive Zeitreihen, bei denen die Regressionsfunktion unbekannt ist. Basierend auf Neuronalen Netzen werden Tesstatistiken sowie Schätzer für korrekt spezifizierte und misspezifizierte Modelle analysiert. Dieser Teil erweitert die Literatur in dem Sinne, dass die Statistiken nicht nur auf den geschätzten Residuen basieren. Beide Abschnitte, Test und Schätzer, werden abgeschlossen mit der Ana- lyse von Regularisierungsbedingungen für das Modell sowie die Parameterschätzer, so dass das asymptotische Verhalten der Tests und Change-point Schätzer gezeigt werden kann.

Abschließend wird eine Simulationsstudie vorgestellt welche das Verhalten des Tests und des Schätzers basierend auf Neuronalen Netzen im Fall der Misspezifikation und auch der Korrektspezifikation analysiert. Diese Tests und Schätzer werden auf zwei Datensätze angewendet.

Contents

Notation

LLN, sLLN	law of large number, strong
CLT	central limit theorem
FCLT	functional central limit theorem
LIL	law of iterated logarithm
UCLT	uniform law of large number
w.r.t.	with respect to
$\langle x, y \rangle_A$	$x^{\mathsf{T}} A y$
$\|x\|_A$	$\sqrt{\langle x, x \rangle_A}$
$\tilde{\varepsilon}$	for random vector ε, the standardised vector
$\xrightarrow{a.s}$, $\xrightarrow{p}$, $\xrightarrow{d}$	stochastic convergence: almost surely, in probability, in distribution
$O_{a.s.}$, $o_{a.s.}$, O_P, o_P	stochastic Landau symbols

1. Introduction

The first step toward change is awareness.
The second step is acceptance.

Nathaniel Branden

1.1. From the beginning of change-point theory to our set-up

Changes happen all the time. So it is natural to become interested in detecting and estimating changes. The statistical literature on change-point analysis goes back to the 1950's when Darling and Erdös [1956] published a first study of a change-point test. Many publications analysing different models or adopting different points of view followed.

Besides testing for a change, the estimation of the time of the change is also of interest. A first publication on the analysis of the change-point estimator was published in 1995, see Antoch et al. [1995]. They considered the estimation of the change-point for the mean change model based on independent and identical distributed (i.i.d.) random variables. For an overview of the different models and concepts on change-point test and estimator we refer to Csörgő and Horváth [1997].

As changes are not a problem of a specific scientific field, the theory is applied in many areas. Change-point tests are of interest in industrial quality management (production monitoring), finance (stock prices), climate studies (global warming) as well as in medicine (online-monitoring of intensive-care patients) or geoscience (annual water volume of rivers) to name a few.

In practice we usually do not know the correct model, but statistical analysis often starts with one. This problem is approached by analysing a change-point test and estimator allowing for model misspecification. In this thesis we contribute to the discussion of misspecification effects on change-point analysis.

As we have measurements of the same object over time, we consider it as a time-series. This time-series is allowed to depend on its past values or on some other variable. The structure of the dependence is unknown. We present a test statistic and estimator using the general approximation property of neural networks to adapt the unknown relation.

Besides the unknown model, a second problem in many data sets is the possible existence of a few measurements far away from the rest. To handle such "outliers" approaches like

M-statistics (compare, e.g. Hušková [1996]) are given even in the context of change-point theory. Motivated by Zhou and Liu [2008], we derive a new modification of the change-point test and estimator in case of heavy-tailed distributions producing outliers.

We conclude the analysis of the change-point test and estimator using the concept of estimation functions which provides a general framework for finding estimators and discussing their properties. The definition of estimation functions was given by Godambe [1960]. A general discussion om estimation functions in parametric models can be found in Sørensen [1999].

1.2. Structure of this thesis

This thesis is organised as follows. In subsection 1.3 the general assumptions on the model we are going to analyse are introduced as well as the notations we use further on. As an introductory example we consider in section 2 the mean change model based on multivariate i.i.d. observations. We give detailed proofs and discuss the main ideas. The proofs are sourced out in seperate sections. We first analyse the change-point test and secondly consider the change-point estimator. As the ideas of the proofs are easier to understand under the assumption of a fixed alternative, meaning the size of the change is constant, we first show the assertions under this fixed alternatives. In section 2.4 we relax this assumption and consider the local alternative. The chapter is concluded with a simulation study for multivariate heavy-tailed observations.

We consider separately the change-point test,chapter 3, and the change-point estimator,chapter 4, where we analyse them for the same models. In both chapters, we first consider the a mean change model with residuals possibly not having a second moment. Afterwards, inspired by Kirch and Tadjuidje Kamgaing [2012], we show how to handle misspecification for a time-series model using neural networks (section 3.2 and 4.2) and close the discussion with the general framework of determining change-point test and change-point estimator using estimation functions, section 3.3 and 4.3.

In chapter 5 we give a simulation study for the neural network based test and estimator derived in sections 3.2 and 4.2. We discuss the behaviour under correct specification and misspecification in view of the expanded version.

Finally, we give two real-data examples in chapter 6, one based on DAX data, section 6.1, and one based on calcium concentrations derived from an ice block drilled in the Antarctica (section 6.2).

1.3. At most one change (AMOC)

In Darling and Erdös [1956], they assumed at most one change, so called AMOC-model. We consider the AMOC-model, too. Let the observations $\{X_t\}$, $t = 1, \ldots, n$, be a time-series in $\mathbb{R}^d$ fulfilling

$$X_t = \begin{cases} X_t^{(1)} & t \leq m\,, \\ X_t^{(2)} & t > m\,, \end{cases} \tag{1.1}$$

where the two sub-series $\{X_t^{(1)},\ t = 1, \ldots, m\}$, and $\{X_t^{(2)},\ t = m+1, \ldots, n\}$, follow distinguishable probability measures. The change is assumed to be in a characteristic or a parameter of the probability measure. We consider a parametric statistic model, in the sense of a probability space with a parametrized family of probability measures $(\Omega, \mathcal{F}, \{P_\theta : \theta \in \Theta\})$, $\Theta \subset \mathbb{R}^q$. The parameter $\theta \in \Theta \subset \mathbb{R}^q$ is assumed to change at $t = m$.

We assume that $\{X_t^{(2)}\}$ is an independent process w.r.t. $X_t^{(1)}$ which results in a kind of discontinuity in the observations at m. If the observations depend on their past the change of a parameter would result observations of a time-series having starting values not out of their stationary distribution. We assume the observations after the change are based on a stationary time-series.

If $m < n$, the distribution of the process has changed and m is called change-point. This change-point is assumed to grow linear with the number of observation to insure asymptotic results. Therefore, the following assumption is made:

G.1 There exists a $\lambda \in (0,1]$ such that the unknown time-point m of the change fulfils

$$m = \lfloor \lambda n \rfloor\,,$$

where n is the number of observations.

There is at-most one unknown time-point where the change occurs.

We first consider the testing problem. Unless otherwise stated, we test the following hypotheses:

$$H_0:\ m = n \text{ (no change occurs)} \qquad \text{vs.} \qquad H_1:\ m < n \text{ (there is a change)}. \tag{1.2}$$

In the following we will use the terms hypothesis and alternative for $\lambda = 1$ and $\lambda \in (0,1)$, respectively.

Weighted CUSUM statistic

To derive a change-point test, suppose the time of change is known. Based on the log-likelihood ratio test statistic the weighted CUSUM is given as

$$T_n = \max_{1 \leq k < n} \sqrt{\frac{n}{k(n-k)}} \left| \sum_{t=1}^{k} (X_t - \overline{X}_n) \right| .$$

Details on the construction are given in section 2.2.1, see page 13.

This statistic has the disadvantage of converging to ∞ for $n \to \infty$. Therefore, an asymptotic distribution only exists after transforming the statistic. The critical value then depends on n and the convergence to the asymptotic distribution is quite slow. As an alternative we consider modifications of the weighted CUSUM statistic.

CUSUM statistic

The reason for the disadvantage is the weight function $\sqrt{\frac{n}{k(n-k)}}$ considering the statistic

$$T_n = \max_{1 \le k < n} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} (X_t - \overline{X}_n) \right|$$

shows the existence of a limit distribution. But for this test statistic the power depends on the position of the change-point. Besides the CUSUM and the weighted CUSUM modifications of the weight function inbetween are considered.

q-weighted CUSUM statistic

Let the weight function q be of the class

$$Q_{0,1} = \{q : \ q \text{ is non-decreasing in a neighborhood of zero, non-increasing in a neighbourhood of one and } \inf_{\eta \le t \le 1-\eta} q(t) > 0 \text{ for all } 0 < \eta < \tfrac{1}{2}\}\,.$$

The q-weighted CUSUM statistic is given as

$$T_n = \max_{1 \le k < n} \frac{1}{\sqrt{n} q(\frac{k}{n})} \left| \sum_{t=1}^{k} (X_t - \overline{X}_n) \right| .$$

To derive asymptotic results Csörgő and Horváth [1997] introduced the integral

$$I(q,c) = \int_0^1 \frac{1}{s(1-s)} \exp\left\{ -c \frac{q^2(s)}{s(1-s)} \right\} ds\,.$$

In Csörgő and Horváth [1993] they showed that for functions $q \in Q_{0,1}$ with $I(q,c) < \infty$ for all $c > 0$ the asymptotics hold true for the q-weighted CUSUM statistics.

Modified weighted CUSUM statistic

These statistics are all of the form

$$T_n(\eta,\gamma;A) = \max_{0<k<n} \frac{w(k/n)}{\sqrt{n}} \left(\sum_{t=1}^{n} H(\mathbb{X}_t;\hat{\theta}_n) \right)^{\mathsf{T}} A \sum_{t=1}^{n} H(\mathbb{X}_t;\hat{\theta}_n)\,, \tag{1.3}$$

where $\mathbb{X}_t$ is some vector of observations up to time t, $\hat{\theta}_n = \hat{\theta}_n(X_1,\dots,X_n)$ is an estimator of the parameter $\tilde{\theta}$, H some function and A a suitable matrix, mostly an estimator of the covariance matrix of the residuals. Thereby, the weight function w is assumed to fulfil the following assumption.

Let $w : [0,1] \to \mathbb{R}_+$ be a non-negative continuous weight function fulfilling

$$\lim_{s\to 0} s^\beta w(s) < \infty\,, \qquad \lim_{s\to 1} (1-s)^\beta w(s) < \infty \qquad \text{for some } \beta \in \left[0, \tfrac{1}{2}\right) \tag{1.4}$$

and for all $\alpha \in (0, \frac{1}{2})$

$$\sup_{\alpha < s < 1-\alpha} w(s) < \infty\,. \tag{1.5}$$

A common choice for the weight function is $w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta < s < (1-\eta)\}} \left(s(1-s)\right)^{-\gamma}$ with γ and η such that $w_{\eta,\gamma}$ fulfils the assumptions above. In this case the assumption on the weight function reduces to the following.

G.2 Let $w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta < s < (1-\eta)\}} \left(s(1-s)\right)^{-\gamma}$, with γ and η be either $\eta \in (0, \frac{1}{2})$ and $\gamma \in [0, \frac{1}{2}]$ or $\eta = 0$ and $\gamma \in [0, \frac{1}{2})$.

Observe, that the case $\eta = 0$ and $\gamma = 0.5$ is not covered.

change-point estimator

Besides testing for a change, we are also interested in estimating the time of the change. As the maximum often occurs in the area of the true change-point, a good choice for the estimator is, where the test statistic has a global maximum. Thus the estimator for the change-point is given by

$$\hat{m}(\eta, \gamma; A) = \arg\max_{1 \le k < n} \frac{w_{\eta,\gamma}(k/n)}{\sqrt{n}} \left(\sum_{t=1}^{n} H(\mathbb{X}_t; \hat{\theta}_n)\right)^{\mathsf{T}} A \sum_{t=1}^{n} H(\mathbb{X}_t; \hat{\theta}_n)\,. \tag{1.6}$$

Under some additional regularity conditions we are able to show that (1.6) is an appropriate estimator for the change-point λ. Moreover, it is an asymptotically consistent estimator and we determine the asymptotic distribution.

Notations

The considered test statistics consist of partial sum processes. To simplify the notation, we introduce the following notation.

N.1 If not stated otherwise we define for some function G, parameter θ and $k = 1, 2, \ldots, N$ the partial sum

$$S_G(k; \theta) = \sum_{t=1}^{k} G(\mathbb{X}_t; \theta)\,.$$

If $G(\mathbb{X}_t; \theta) = (X_t - \theta)$ we also write $S_G(k; \theta) = S_k(\theta)$.

N.2 For a symmetric and positive semi-definite matrix A we write $\langle x, y\rangle_A := x^{\mathsf{T}} A y = \langle A^{\frac{1}{2}} x, A^{\frac{1}{2}} y\rangle$, with $\langle \cdot, \cdot\rangle$ denoting the scalar-product on $Im(A)$ and $A^{\frac{1}{2}}$ a Cholesky-Decomposition of A.

N.3 The norm $\|\cdot\|_A$ is induced by $\langle \cdot, \cdot\rangle_A$, i.e. $\|x\|_A = \sqrt{x^{\mathsf{T}} A x}$.

The modified weighted CUSUM statistic, we are interested in, is then of the form

$$T_n(\eta, \gamma; A) = \max_{0 < k < n} \frac{w_{\eta,\gamma}(k/n)}{\sqrt{n}} \left\| S_G(k; \hat{\theta}_n) \right\|_A\,, \tag{1.7}$$

with A fulfilling the following conditions.

G.3 Let A be a symmetric and positive semidefinite matrix.

This class covers many statistics. In most of the cases we refer to the parts of the magnitude δ which is of interest. As we can control the parts of interest with A we introduce the following notation.

N.4 Let δ_A, δ_K be so that $\delta_n = \delta_A + \delta_K$ and $\delta_A \in Im(A)$ and $\delta_K \in Kern(A)$. Then δ_A is called the detectable part of δ w.r.t. A and δ_K the non-detectable part of δ w.r.t. A.

Based on the decision matrix A we introduce the A-fixed and A-local alternatives. The magnitude of the change will be represented with the vector $\delta_n \underset{n\to\infty}{\longrightarrow} \delta$. Depending on the limit δ, different assumptions on the alternatives are considered.

G.5.a) (A-fixed alternative) Let δ_n be an r-dimensional vector with $\|\delta_n\|_A \equiv \mathcal{D} > 0$.

G.5.b) (A-local alternative) Let δ_n be an r-dimensional vector with $\|\delta_n\|_A = \mathcal{D}_n > 0$ and $\mathcal{D}_n \underset{n\to\infty}{\longrightarrow} 0$ as well as $\sqrt{n}\mathcal{D}_n \underset{n\to\infty}{\longrightarrow} \infty$.

In the case of A-fixed alternative, the interesting part of δ w.r.t. A is non zero. Otherwise for A-local alternatives δ_A is assumed to be the zero-vector.

Especially for multi-dimensional test statistics it becomes of interest which kind of alternatives are detectable. So we add a matrix A' and define $A := \Sigma^{-\frac{1}{2}} A' \Sigma^{-\frac{1}{2}}$, where Σ is a particular symmetric, positive definite matrix we will have later on.

N.5 Let $A' := \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ denote the decision matrix.

It is clear, that A' is a symmetric and positive semi-definite matrix.

Observe, that we will use the term decision matrix for both A and A' depending on the situation.

2. Multidimensional mean change model

One of the first results in change-point analysis was given by Page [1957] using the mean change model, assuming the time of the possible change is known. He assumed the mean under H_0 to be known and the errors to be independent identically normally distributed. His results were explored, e.g. Sen and Srivaastava [1975], James et al. [1987] and up to the non-parametric set-up given in Csörgő and Horváth [1997]. We introduce the technique of the proofs we use in the rest of the thesis with the non-parametric multivariate mean change model. In [Csörgő and Horváth, 1997, section 2.1.] the observed process $X_t \in \mathbb{R}$ consists of i.i.d. (independent identically distributed) random variables, where the (unknown) mean is changing at some unknown time m. Besides the fact that we are allowing the observations to be multidimensional, i.e. $X_t \in \mathbb{R}^d$, we use the same model. So, the observed process follows the model (1.1), where the process before and after the unknown time-point m, which fulfils assumption **G.1**, is given by

$$X_t = \begin{cases} X_t^{(1)} = \theta + \varepsilon_t & 1 \leq t \leq m\,, \\ X_t^{(2)} = \theta + \delta_n + \varepsilon_t & m < t \leq n\,, \end{cases} \tag{2.1}$$

where $\theta \in \Theta(\subseteq \mathbb{R}^d)$, $\delta_n \in \mathbb{R}^d$ and the errors $\{\varepsilon_t,\ 1 \leq t \leq n\}$, are unknown. We make the following assumption.

L.1 The errors ε_i are i.i.d. with zero mean, finite second moment and unknown covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$.

We do not make any assumptions on the error distribution, i.e. we are in the so-called "non-parametric" set-up. One way to derive estimators is to assume normality for the error distribution and calculate the corresponding estimator based on likelihood considerations. In a second step one shows that this estimator is still valid for error distributions fulfilling **L.1** due to the law of large numbers (LLN) and the central limit theorem (CLT). Under the normality assumption on the distribution of the errors it is well known, that under H_0 (1.2) the MLE (maximum-likelihood estimator) and the LSE (least squares estimator) for θ are equal. Motivated by this result, we are going to use the LSE's for the unknown parameter θ for the model with unknown error distribution.

To derive the test statistic, we first analyse the behaviour of the parameter estimator. We use the least squares estimator, i.e. $\hat{\theta}_n$ solves

$$\sum_{t=1}^{n} (X_t - \theta) = 0\,. \tag{2.2}$$

The corresponding estimator is

$$\hat{\theta}_n = \overline{X}_n\,,$$

which is an unbiased and $\sqrt{n}$-consistent estimator due to the law of large numbers (LLN), Theorem C.2.1, and the central limit theorem (CLT), Proposition C.2.1, respectively. The results are shown in section 2.1. In section 2.2 we are constructing the statistic of the change-point test. If no change occurs the estimated mean over all the observed data should be close to the mean of a shorter subsample. So, the test statistic $T_n(\eta, \gamma; A)$ (1.7) is given by

$$\begin{aligned} T_n(\eta, \gamma; A) &= \max_{1 \le k < n} \frac{1}{\sqrt{n}} \left\| S_G(k; \hat{\theta}_n) \right\|_A \\ &= \max_{1 \le k < n} \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_A \\ &= \max_{1 \le k < n} \left(\frac{k}{\sqrt{n}} \right) \left\| \overline{X}_k - \hat{\theta}_n \right\|_A . \end{aligned}$$

and should have no significant value. It is also possible to weight the statistic to improve the power of the test for specific values of the change-point (i.e. of λ). To achieve this, statistics of the form

$$T_n(\eta, \gamma) = \max_{1 \le k < n} w_{\eta,\gamma}(k/n) \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_A \tag{2.3}$$

are of interest, where the weight function $w_{\eta,\gamma}(s)$ is usually given as

$$w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta < s < (1-\eta)\}} \left(s(1-s) \right)^{-\gamma} \tag{2.4}$$

with either $\eta \in (0, \frac{1}{2})$, $\gamma \in [0, \frac{1}{2}]$ or $\eta = 0$ and $\gamma \in [0, \frac{1}{2})$, see **G.2**.

For a significance test, we determine the distribution of the test statistic for a finite sample size n. Due to the fact that we have not made any assumptions on the distribution of the innovations (non-parameteric set-up), determining the distribution of the test statistic $T_n(\eta, \gamma)$ is not possible. However, we determine the asymptotic distributions of these statistics and further show the consistency of the corresponding tests.

For one-dimensional data the asymptotic distribution of the test statistic $T_n(\eta, \gamma)$ (2.3) with the estimator $\hat{\theta}_n = \overline{X}_n$ is given in [Csörgő and Horváth, 1997, section 2.1]. The asymptotic behaviour for the corresponding change-point estimator, i.e.

$$\hat{m}(\eta, \gamma) := \hat{m}(\eta, \gamma; \Sigma^{-1}) = \arg\max_{1 \le k < n} w_{\eta,\gamma}(k/n) \left\| \sum_{i=1}^{k} \left(X_i - \overline{X}_n \right) \right\|_{\Sigma^{-1}}$$

was analysed and proved in [Antoch et al., 1995]. The proofs are well known for the one-dimensional case. We show the results for

$$\hat{m}(\eta, \gamma; A) = \arg\max_{1 \le k < n} w_{\eta,\gamma}(k/n) \left\| \sum_{i=1}^{k} \left(X_i - \overline{X}_n \right) \right\|_A \tag{2.5}$$

multidimensional data using techniques we are going to improve in the rest of the thesis.

For the change-point test and change-point estimator we are going to analyse the different types of alternatives separately. In the literature so-called fixed alternative, i.e. $\delta_n \equiv \delta$ constant and local alternatives ($\|\delta_n\| \underset{n \to \infty}{\longrightarrow} 0$) are analysed. In the multi-dimensional case we introduce the A-fixed alternative, i.e. $\|\delta_n\|_A = \mathcal{D} > 0$ and A-local alternative, i.e. $\|\delta_n\|_A = \mathcal{D}_n \underset{n\to\infty}{\longrightarrow} 0$ (compare **G.5.a)** and **G.5.b)**).

We start with analysing the parameter estimator under H_0 and H_1 in section 2.1. The fixed and the local alternative are considered in this section. In sections 2.2 and 2.3 we show the asymptotic results under the A-fixed alternative. We present the corresponding results for the A-local alternative in section 2.4. In section 2.5 we present first results on a simulation study for the mean change model with heavy-tailed innovations.

2.1. Parameter estimator

In this section we analyse the behaviour of the parameter estimator for the change-point model (2.1). We have assumed that the change-point is of the form $m = \lfloor \lambda n \rfloor$, with $\lambda \in (0, 1]$. Then the hypothesis H_0 is equivalent to $\lambda = 1$ and the alternative H_1 means $\lambda \in (0, 1)$.

Under the hypothesis H_0 (1.2) it is clear by the LLN (law of large numbers) for i.i.d. random variables that we get a consistent estimator. Moreover, by the CLT (central limit theorem) for i.i.d. random variables we get the asymptotic normality of the estimator.

The question arises how the estimator behaves under the alternative H_1, i.e. if there is a change at observation time $m = \lfloor \lambda n \rfloor$, $\lambda \in (0, 1)$. In the following we show that under H_0 as well as under H_1 the estimator is $\sqrt{n}$-consistent.

2.1.1. Asymptotic behaviour

Under H_0 and H_1 we have a consistent estimator for the unknown parameter of our model (but with different limits). Thus, under H_0 (no change) the parameter of the observations will be correctly determined with growing sample size. If we have observations with a change, H_1, the estimator still converges to some limit. This limit is a convex combination of the true parameters. It is just the mean of the parameters, if the change is in the middle. And for example, if the change-point is less than 0.5 we will have more observations after the change, thus the limit will be weighted in this direction. To be more precise, the result is the following:

Theorem 2.1.1 *Let $\tilde{\theta} \in \Theta$ be defined as*

$$\tilde{\theta}_n = \begin{cases} \tilde{\theta}_0 = \theta & \text{if } H_0 \text{ is true,} \\ \tilde{\theta}_1 = \theta + (1-\lambda)\delta_n & \text{if } H_1 \text{ is true} \end{cases}$$

with δ_n bounded, i.e. there exists $c > 0$ such that $\|\delta_n\| < c$ for all n. Then, under H_0 and under H_1, it holds

$$\|\hat{\theta}_n - \tilde{\theta}_n\| \underset{n\to\infty}{\longrightarrow} 0 \qquad a.s.$$

Differently to the one-dimensional case, δ_n is allowed to contain both constant entries and vanishing ones (for each $i = 1, \ldots, q$ either $(\delta_n)_i \equiv \delta_i \neq 0$ or $(\delta_n)_i \underset{n\to\infty}{\longrightarrow} 0$). The results for the parameter estimator still hold true.

To prove this, we use that the independence of the ε_t results in the independence of the X_t due to the chosen model. Under H_0 we can directly apply the LLN. For the alternative we split the statistic into independent sums. One before the change-point and one for observations after the change-point. Due to the model both sums are sums of i.i.d. random vectors with constant mean, which allows to apply the LLN.

The specific form of the asymptotic limit can be seen in the following way. Assume, $\delta_n \equiv \delta$. Then the sample mean can be split into two sums, one over the observations before and one over the observations after the change-point. The first sum is weighted by $\frac{m}{n}$ (converging to λ, which is the percentage of the observations before the change) and the second sum is weighted by $1 - \frac{m}{n}$ (converging to $1 - \lambda$, the percentage of the observations after the change). We can think of the limit as a weighted mean of the parameter before the change (θ) and after the change ($\theta + \delta$).

Under the local alternative ($\delta_n \underset{n\to\infty}{\longrightarrow} 0$) the corresponding entries converge always to the unchanged parameter entry. Why this kind of alternative is of interest will be discussed in the analysis of the change-point test and the change-point estimator under the local alternative in section 2.4.

For the parameter estimator we are able to prove $\sqrt{n}$-convergence. We show that this is the best rate, by determining the asymptotic distribution.

Theorem 2.1.2 *Let $\tilde{\theta}$ and δ_n be as in Theorem 2.1.1. Under H_0 as well as under H_1 it holds*

$$\mathcal{L}\left(\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n)\right) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \Sigma) ,$$

where $\Sigma = \mathbb{C}\mathrm{ov}(\varepsilon_1)$.

The asymptotic distribution can be derived by the CLT (see e.g., Theorem C.2.1) and by using the same partition for the alternative as used in the proof of consistency.

Notice, the covariance matrix of the limit distribution is not changing for the alternative. This covariance matrix is just the covariance matrix of the errors, i.e. the estimator depends on the scale of the process.

2.1.2. Proofs

Theorem 2.1.1

Let $\tilde{\theta} \in \Theta$ be defined as

$$\tilde{\theta}_n = \begin{cases} \tilde{\theta}_0 = \theta & \text{if } H_0 \text{ is true,} \\ \tilde{\theta}_1 = \theta + (1-\lambda)\delta_n & \text{if } H_1 \text{ is true} \end{cases}$$

with δ_n is bounded, i.e. there exists $c > 0$ such that $\|d_n\| < c$ for all n. Then, under H_0 and under H_1, it holds

$$\|\hat{\theta}_n - \tilde{\theta}_n\| \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{a.s.}$$

For this proof we use the following notation.

N.6 Let Y_i, $i = 1, \dots, n$, be a sequence of observations and $a, b \in \mathbb{N}$ with $0 < a < b \leq n$, then

$$\overline{Y}_{a,b} = \frac{1}{b-a+1} \sum_{i=a}^{b} Y_i \,. \tag{2.6}$$

We are going to use this notation in the rest of this thesis.

Proof of Theorem 2.1.1:
First we analyse the behaviour under H_0. By the LLN (Theorem C.2.1) we will directly get the result if we observe that

$$\hat{\theta}_n = \theta + \overline{\varepsilon}_n \tag{2.7}$$

due to the given model 2.1. The expectation of the ε_i, $1 \leq i \leq n$, is assumed to be zero, which proves the claim.

Let us now take a look at what happens under H_1. Observe that

$$\begin{aligned} \hat{\theta}_n &= \frac{m}{n}\overline{X}_m + \frac{n-m}{n}\overline{X}_{m+1,n} \\ &= \frac{m}{n}\theta + \frac{m}{n}\overline{\varepsilon}_m + \frac{(n-m)}{n}\theta + (1-\frac{m}{n})\delta_n + \frac{n-m}{n}\overline{\varepsilon}_{m+1,n} \\ &= \overline{\varepsilon}_n + \theta + (1-\frac{m}{n})\delta_n = \overline{\varepsilon}_n + \tilde{\theta}_n + (\lambda - \frac{m}{n})\delta_n \,. \end{aligned} \tag{2.8}$$

The LLN C.2.1 for the sample mean of the ε_i and the fact that $\frac{m}{n} = \frac{\lfloor \lambda n \rfloor}{n} \underset{n\to\infty}{\longrightarrow} \lambda$ completes the proof. ■

Theorem 2.1.2

Let $\tilde{\theta}_n$ and δ_n be as in Theorem 2.1.1. Under H_0 as well as under H_1 it holds

$$\mathcal{L}\left(\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n)\right) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \Sigma) \,,$$

where $\Sigma = \mathbb{C}\mathrm{ov}\,(\varepsilon_1)$.

Proof:
Under H_0:
With (2.7) and the CLT for $\bar{\varepsilon}_n$ we get the result under H_0.

Under H_1:
Using the representation of the estimator as in the proof of Theorem 2.1.1 yields

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) = \sqrt{n}\bar{\varepsilon}_n + \sqrt{n}(\lambda - \frac{m}{n})\delta_n \,.$$

Observe that $\frac{m}{n} - \lambda = \frac{\lfloor \lambda n \rfloor}{n} - \lambda = O(\frac{1}{n})$ and so $\sqrt{n}(\frac{m}{n} - \lambda) = O(\frac{1}{\sqrt{n}}) = o(1)$. This fact together with the CLT C.2.1 for $\bar{\varepsilon}_n$ and Slutzky C.1.4 finishes the proof. ■

2.2. Change-point test

In this and the following section, we only consider the A-fixed alternative (A as in **G.3**), i.e. the model we are going to analyse is

$$X_t = \begin{cases} \theta + \varepsilon_t & t \leq m, \\ \theta + \delta_n + \varepsilon_t & t > m, \end{cases} \tag{2.9}$$

where the errors $\{\varepsilon_t\}$ fulfil **L.1**, the unknown time-point m fulfils assumption **G.1** and we have an A-fixed alternative (**G.5.a)**).

First we take a look on how we can derive a change-point test (section 2.2.1). There we introduce the matrix A. We are going to discuss different versions of change-point tests, which we are using in this thesis further on. This versions differ only in the weightfunction. To understand this functions we take a closer look on them in section 2.2.2. Then, we show and discuss the asymptotic results for the change-point test in section 2.2.3. Section 2.2.4 provides the detailed proofs for the asymptotic results.

2.2.1. Test-statistic

In our model 2.9, we do not make any assumptions on the distribution of the errors, the so-called non-parametric set-up. Nevertheless, one way to derive test statistics is to use the likelihood ratio, where we pretend that the errors are i.i.d. standard normal. In a second step we prove that this statistic is still reasonable for other distributions fulfilling assumption **L.1**. This makes sense in view of the CLT.

We also have to solve the problem that the change-point is unknown. The idea to overcome this is to first assume the change-point to be known.

To derive the so-called pseudo-likelihood ratio statistic we first need the maximum likelihood estimators. It is well known, that this estimators are equal to the least squares estimators under standard normal assumption on the errors. The parameter estimator under H_0 is given as

$$\hat{\theta}_n = \overline{X}_n \tag{2.10}$$

and the estimators under H_1 are given as

$$\hat{\theta}_n^0 = \overline{X}_m \text{ (before change-point)} \qquad \text{and} \qquad \hat{\theta}_n^1 = \overline{X}_{m+1,n} \text{ (after change-point)}^* . \tag{2.11}$$

*For the notation see **N.6**.

Instead of the likelihood-ratio statistic, we consider the log-likelihood ratio $l(X_1,\ldots,X_n)$. Defining $f_\theta(x)$ as the density of a normal distributed random variable with mean θ and some fixed covariance matrix Σ, we get

$$
\begin{aligned}
l(X_1,\ldots,X_n) =& \log \frac{\prod_{t=1}^{n} f_{\hat\theta_n}(x_t)}{\prod_{t=1}^{m} f_{\hat\theta_n^0}(x_t)\ \prod_{i=m+1}^{n} f_{\hat\theta_n^1}(x_i)} \\
=& \frac{1}{2}\left(\sum_{i=1}^{n}(X_i-\hat\theta_n)^t\Sigma^{-1}(X_i-\hat\theta_n)\right. \\
& \left. -\left(\sum_{i=1}^{m}(X_i-\hat\theta_n^0)^t\Sigma^{-1}(X_i-\hat\theta_n^0)+\sum_{i=m+1}^{n}(X_i-\hat\theta_n^1)^t\Sigma^{-1}(X_i-\hat\theta_n^1)\right)\right)
\end{aligned}
$$

Observe, that it holds

$$
\begin{aligned}
\sum_{i=1}^{n}(X_i-\hat\theta_n)^t\Sigma^{-1}(X_i-\hat\theta_n) =& \sum_{i=1}^{n} X_i^t\Sigma^{-1}X_i - \sum_{i=1}^{n}\hat\theta_n^t\Sigma^{-1}X_i - \sum_{i=1}^{n} X_i^t\Sigma^{-1}\hat\theta_n \\
& + n\hat\theta_n^t\Sigma^{-1}\hat\theta_n \\
=& \sum_{i=1}^{n} X_i^t\Sigma^{-1}X_i - n\,\hat\theta_n^t\Sigma^{-1}\hat\theta_n
\end{aligned}
$$

The same result follows for the sum up to m by replacing $\hat\theta_n$ with $\hat\theta_n^0$ and for the sum from $m+1$ to n by replacing with $\hat\theta_n^1$. We get for the log-likelihood ratio

$$
\begin{aligned}
l(X_1,\ldots,X_n) =& \frac{1}{2}\left(m\,(\hat\theta_n^0)^t\Sigma^{-1/2}\hat\theta_n^0+(n-m)\,(\hat\theta_n^1)^t\Sigma^{-1}\hat\theta_n^1-n\,(\hat\theta_n)^t\Sigma^{-1}\hat\theta_n\right) \\
=& \frac{1}{2}\left(m\,(\hat\theta_n^0)^t\Sigma^{-1}\hat\theta_n^0+\frac{1}{n-m}\left(n\hat\theta_n-m\hat\theta_n^0\right)^t\Sigma^{-1}\left(n\hat\theta_n-m\hat\theta_n^0\right)-n\,(\hat\theta_n)^t\Sigma^{-1}\hat\theta_n\right) \\
=& \frac{1}{2}\frac{nm}{n-m}\left\|\hat\theta_n^0-\hat\theta_n\right\|_{\Sigma^{-1}}^2 \\
=& \frac{1}{2}\frac{n}{m(n-m)}\left\|\sum_{t=1}^{m}\left(X_t-\hat\theta_n\right)\right\|_{\Sigma^{-1}}^2
\end{aligned}
$$

Usually, we do not know the point of change. So we maximize over all possible values of the change-point m, looking for the time-point m which minimizes the variance under the alternative, i.e. most plausible time-point for the change for the given observations.

The result does not change if we use the square root of the log-likelihood function as test statistic. This test statistic is the so-called weighted CUSUM (cumulated sum) given by

$$
\max_{1\le k<n}\sqrt{\frac{n}{k(n-k)}}\left\|\sum_{t=1}^{k}\left(X_t-\hat\theta_n\right)\right\|_{\Sigma^{-1}}.
$$

It is known that this statistic has no limit distribution under H_0 unless it is transformed. Then, it becomes asymptotically Gumbel-distributed Darling and Erdös [1956]. In several simulations it was observable that this convergence is rather slow, that is why modifications of the test statistic became of interest. The modified test statistics are of the form

$$
T_n(\eta,\gamma) = \max_{\eta n<k<(1-\eta)n}\left(\frac{n^2}{k(n-k)}\right)^{\gamma}\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k}\left(X_t-\hat\theta_n\right)\right\|_{\Sigma^{-1}}
$$

with $\eta \in (0, \frac{1}{2})$ and $\gamma \in [0, \frac{1}{2}]$ or $\eta = 0$ and $\gamma \in [0, \frac{1}{2})$.

These test statistics can be seen as one type of statistic, but with different weight functions $\mathbf{1}_{\{\eta n < k < (1-\eta)n\}} \left(\frac{n^2}{k(n-k)}\right)^{\gamma}$.

This is a possibility to test for changes in the mean vector. If one is only interested in testing for specific dimensions of the mean vector or if one dimension is much more important than the others, we replace Σ^{-1} with a matrix A. The test statistic is then defined as

$$T_n(\eta, \gamma; A) = \max_{\eta n < k < (1-\eta)n} \left(\frac{n^2}{k(n-k)}\right)^{\gamma} \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^{k} \left(X_t - \hat{\theta}_n\right) \right\|_A \tag{2.12}$$

With the matrix A we can decide which alternatives we are interested in. An easy example is, if we are only interested in changes in the l-th dimension (projection), then

$$A_{ij} = 0 \quad (i,j) \neq (l,l) \qquad \text{and} \qquad A_{ll} = 1\,.$$

In the same way one can weight the dimensions to detect for smaller changes in a more important one. Defining the decision matrix A' as given in **N.5** the asymptotics for the change-point test and later the change-point estimator become easier.

A different matrix other than the covariance matrix of the residuals, for an other model was given e.g. in Hušková et al. [2007]. A first introduction in the offline set-up of such a matrix is given in Kirch et al. [2015]. They discuss the role of the matrix A (in their publication called H) in the context of estimating the covariance matrix for large dimensions. If the dimension is quite large in relation to the observation number, the estimators for covariance structures have high fluctuations. This motivates to use some kind of projection realised with the matrix A (in Kirch et al. [2015] called H), which is then singular. In the simulation study in section 2.5 we will discuss this effect for some specific examples.

The weight function on the other hand gives the sensitivity of the test statistics against different alternatives, in the sense of the position of the change. For a better understanding of the weight function we take a closer look at this class of function in the next section.

2.2.2. Understanding the weight functions

We are going to take a closer look at the weight function $\mathbf{1}_{\{\eta<s<(1-\eta)\}}\,(s(1-s))^{-\gamma}$, with η and γ fulfilling assumption **G.2** (i.e. $\eta \in (0, \frac{1}{2})$, $\gamma \in [0, \frac{1}{2}]$ or $\eta = 0$ and $\gamma \in [0, \frac{1}{2})$) and its effect on the change-point test. For simplicity of notation we introduce the following function.

N.7 Let $w_{\eta,\gamma} : (0,1) \to \mathbb{R}_+$, $w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta<s<(1-\eta)\}}\,(s(1-s))^{-\gamma}$, with $\eta \in (0, \frac{1}{2})$ and $\gamma \in [0, \frac{1}{2}]$ or with $\eta = 0$ and $\gamma \in [0, \frac{1}{2})$. In the case of $\eta = 0$ we write $w_\gamma(s) \equiv w_{0,\gamma}(s)$.

The effect of η is clear, it controls the region of detectable times for a change. Sometimes, we do not make any reduction on the region of the possible change-point, i.e. we would choose $\eta = 0$. Therefore, the effect of the parameter γ on the change-point test becomes of interest. A study of this effect is rarely done in the literature.

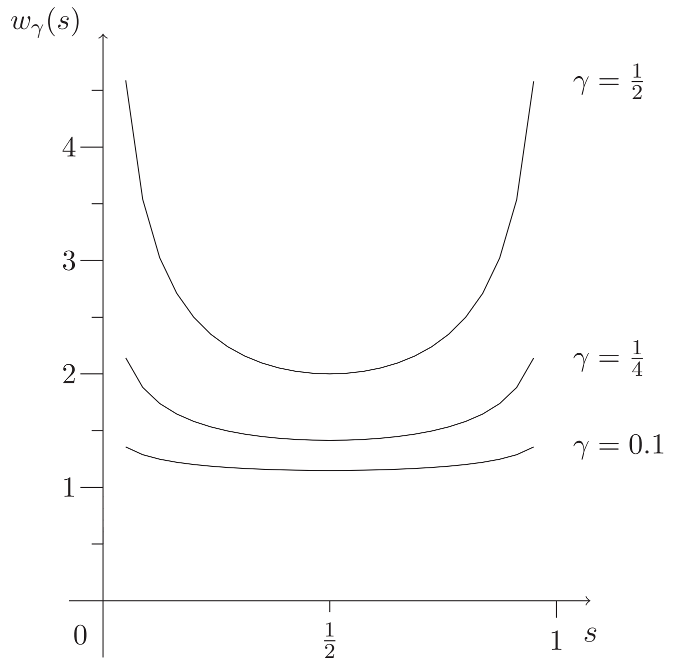


Figure 2.1.: weight function $w_\gamma(s)$ with $\gamma = 0.1, 0.25, 0.5$

In Figure 2.1 the weight function $w_\gamma(s)$, $s \in (0,1)$, is illustrated for $\gamma = 0.1, 0.25$ and $\gamma = 0.5$. The weight function which we use to modify the statistic increases values at the beginning and at the end. If the change is at the beginning, the partial sum $|S_k(\hat{\theta}_n)|$ only increases for the first few observations up to the change. The contribution of the data before the change to the overall mean will be small, so it is more likely that the change goes undetected. The parameter γ allows for stronger weighting of these terms at the maximum which corresponds to potential change-points close to the boundaries. This enables detection of smaller changes with a constant variability.

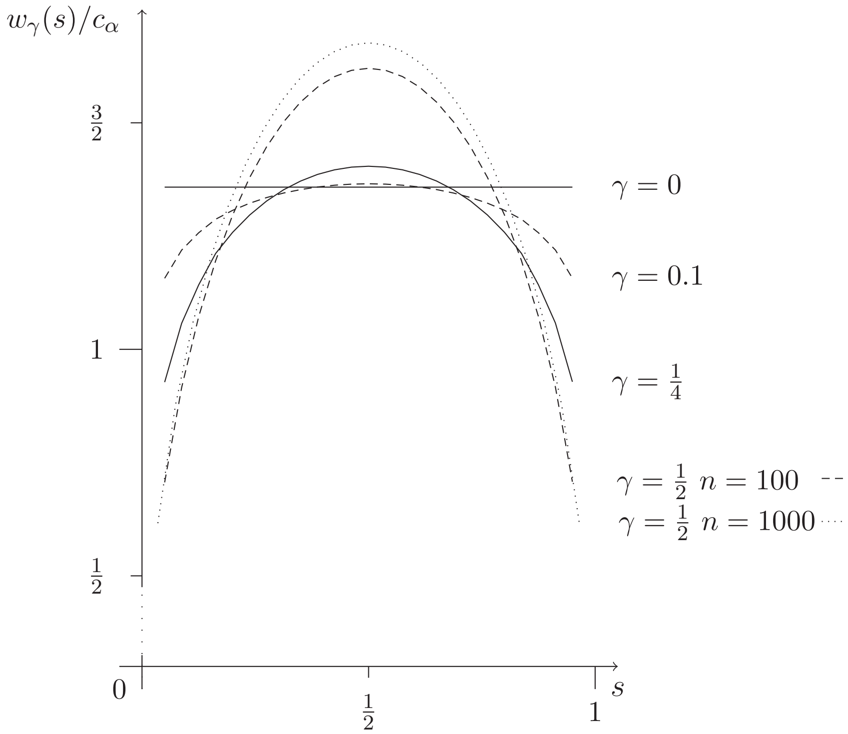


Figure 2.2.: weight function $w_\gamma(s)/c_\alpha$ with $\gamma = 0, 0.1, 0.25, 0.5$ and $\alpha = 0.05$

But for each choice of the parameter γ, we have different critical values for the test statistic (see Theorem 2.2.1). For a better comparison we divide the critical value by the corresponding value of the weight function for $s \in (0, 1)$, see Figure 2.2. Assuming the maximum of the partial sum process is at some fixed s. A lower value of the quotient at this point corresponds to detection of smaller change-points by the corresponding test. This substantiate once more the relation between the parameter γ and the position of the change.

Let us take a closer look at the effect. Assume one dimensional normal distributed observations. For the illustration we restrict the discussion to the weight functions with $\eta = \gamma = 0$ and $\eta = 0$, $\gamma = \frac{1}{2}$. The test statistic can be written as

$$T_n(0, \gamma) = \max_{1 \leq k < n} \left(\frac{n^2}{k(n-k)} \right)^\gamma \frac{1}{\sqrt{n}} \left| S_k(\hat{\theta}_n) \right| = \max_{1 \leq k < n} w_\gamma(k/n) \left| V_k(\hat{\theta}_n) \right|$$

with $V_k = \frac{1}{\sqrt{n}} S_k(\hat{\theta}_n)$ and $S_k(\hat{\theta}_n)$ as in **N.1**. It is clear, that the V_k are all normally distributed but with different variances and expectations. The expectation, variance and covariance for $V_k(\hat{\theta}_n)$ are given as

$$\mathbb{E}[V_k(\hat{\theta}_n)] = \begin{cases} -\frac{k(n-m)}{n}\frac{\delta}{\sqrt{n}} & ,k \le m \\ -\frac{m(n-k)}{n}\frac{\delta}{\sqrt{n}} & ,k > m \end{cases} \tag{2.13}$$

$$\mathbb{V}\mathrm{ar}[V_k(\hat{\theta}_n)] = \frac{k(n-k)}{n^2}\sigma^2 \equiv \sigma_k^2\,, \tag{2.14}$$

$$\mathbb{C}\mathrm{ov}[V_k(\hat{\theta}_n), V_l(\hat{\theta}_n)] = \frac{1}{n}\mathbb{C}\mathrm{ov}[S_k(\hat{\theta}_n), S_l(\hat{\theta}_n)] = \frac{(n-\max(k,l))\ \min(k,l)}{n^2}\sigma^2\,. \tag{2.15}$$

Observe, that multiplying with the weight function $w_{\frac{1}{2}}(k/n)$ ($s = \frac{k}{n}$) would lead to partial sums with constant variance, i.e.

$$\mathbb{E}[w_{\frac{1}{2}}(k/n)V_k(\hat{\theta}_n)] = \begin{cases} -\frac{\sqrt{k}(n-m)}{\sqrt{(n-k)}}\frac{\delta}{\sqrt{n}} & ,k \le m \\ -\frac{m\sqrt{n-k}}{\sqrt{k}}\frac{\delta}{\sqrt{n}} & ,k > m \end{cases} \tag{2.16}$$

$$\mathbb{V}\mathrm{ar}\left[w_{\frac{1}{2}}(k/n)V_k(\hat{\theta}_n)\right] = \sigma^2\,, \tag{2.17}$$

$$\mathbb{C}\mathrm{ov}\left[w_{\frac{1}{2}}(k/n)V_k(\hat{\theta}_n), w_{\frac{1}{2}}(l/n)\,V_l(\hat{\theta}_n)\right] = \frac{n\,\mathbb{C}\mathrm{ov}\left[S_k(\hat{\theta}_n), S_l(\hat{\theta}_n)\right]}{\sqrt{k\,l\,(n-k)(n-l)}}\,. \tag{2.18}$$

The test statistic is based on the absolute values of the unweighted V_k and the weighted V_k. Due to Jensen's inequality we have $\mathbb{E}[|V_k(\hat{\theta}_n)|] \ge |\mathbb{E}[V_k(\hat{\theta}_n)]|$. Observe that the expectations of the V_k and the weighted V_k have both their maximum for $k = m$. If the maximum expectation is larger compared to the standard deviation, the test statistic will exhibit the value V_m with high probability.

Let us take a closer look at the unweighted and the weighted partial sums, i.e. $|V_{\lfloor sn \rfloor}|$ and $w_\gamma(\lfloor sn \rfloor/n)|V_{\lfloor sn \rfloor}|$, respectively. In Figure 2.3 the expectations and standard deviations of the unweighted partial sums $|V_{\lfloor sn \rfloor}|$ are shown. We have used $\delta = -0.2$ and $n = 50$. We compare the behaviour for different standard deviations, $\sigma = 0.5$ and $\sigma = 0.2$ in the first and second row, respectively. In the left column the change is in the middle ($\lambda = 0.5$) and in the right column after the first quarter ($\lambda = 0.25$, so called early change). The corresponding results for the weighted partial sums $w_\gamma(\lfloor sn \rfloor/n)|V_{\lfloor sn \rfloor}|$ are given in Figure 2.4 in the same order.

In Figure 2.3 the expectations (thick black lines) and the expectations plus and minus the standard deviations (black lines) are shown for different parameters. The blue line is the critical value of the corresponding test statistic. The maximum mean of the unweighted partial sum is smaller for the early change ((b) and (d)) than for the change in the middle (2.3a and 2.3c). Moreover the value is always below the critical value, so the test has problems in detecting early changes correctly. In case the maximum mean is below the critical value, the power can not be increased with a lower variance of the residuals. As we can see, it may happen that all three lines are below the critical value (probability of rejection is quite small, i.e. low power). Furthermore we see that for a change in the middle the power of the test increases for smaller variances.

In Figure 2.4 again the expectations (thick black lines) and the expectations plus and minus the standard deviations (black lines) are shown but now for the weighted partial sum $w_\gamma(\lfloor sn \rfloor/n)|V_{\lfloor sn \rfloor}|$. The blue line is the corresponding critical value of the test statistic. The maximum expectation is always higher than the critical value, while we use the same parameters as in Figure 2.3. Here we observe that the test enables the detection of the early and middle changes. For fixed variance of the residuals, the maximum expectation and the maximum expectation minus the average spread are higher than the critical value (see 2.4a and 2.4b). This means that the power of the test is not so significantly dependent on the time of change (compare 2.4b and 2.4d). For some other choice of δ this could be different.

Nevertheless, there exist values of δ for which the unweighted test has smaller power than the weighted, depending on the time of change. The illustration supports the interpretation we have made, i.e. for early changes the unweighted test statistic will give better results. After all, we get the impression that the weighted test statistic should be preferred. We had stated earlier (section 2.2.1, page 14) that the weighted test statistic with parameter $\gamma = \frac{1}{2}$ converges slowly to the asymptotic distribution, that is why this distribution is not preferred. A closer look on the choice of the parameter γ and the properties of the corresponding test statistic for different distributions (also heavy tailed) can be found in section 2.5.

We conclude that for higher values of γ, $\gamma \in [0, \frac{1}{2})$ the corresponding modified change-point test detects early changes better. If we have no difficulties in reducing the region of the detectable times of change, i.e. $\eta \in (0, \frac{1}{2})$, we can use $\gamma = \frac{1}{2}$. Otherwise, we would choose a γ quite close to $\frac{1}{2}$. Several choices are possible, e.g. assume we would have a truncation of η_0, then we could set $\gamma = \frac{1}{2} - \eta_0(1 - \eta_0)$.

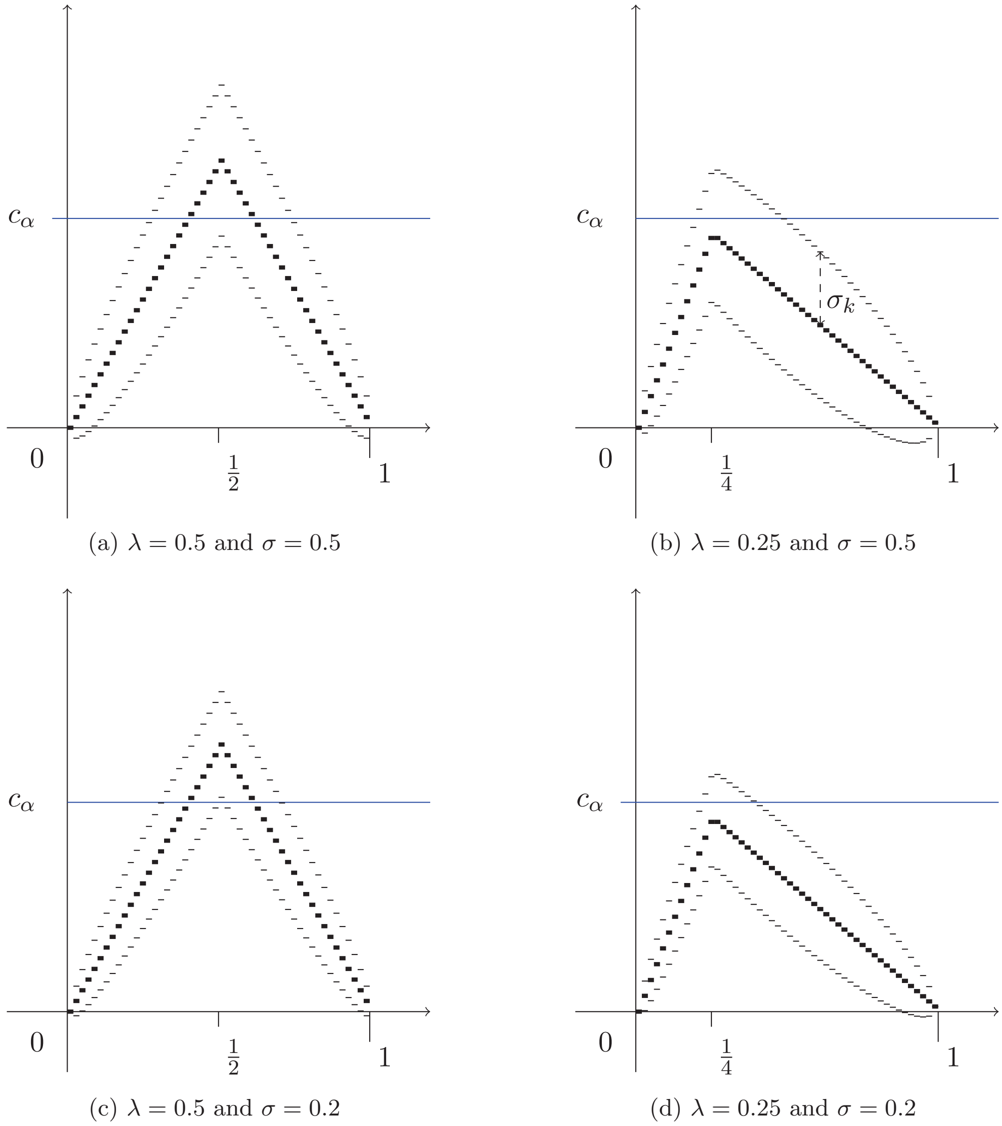
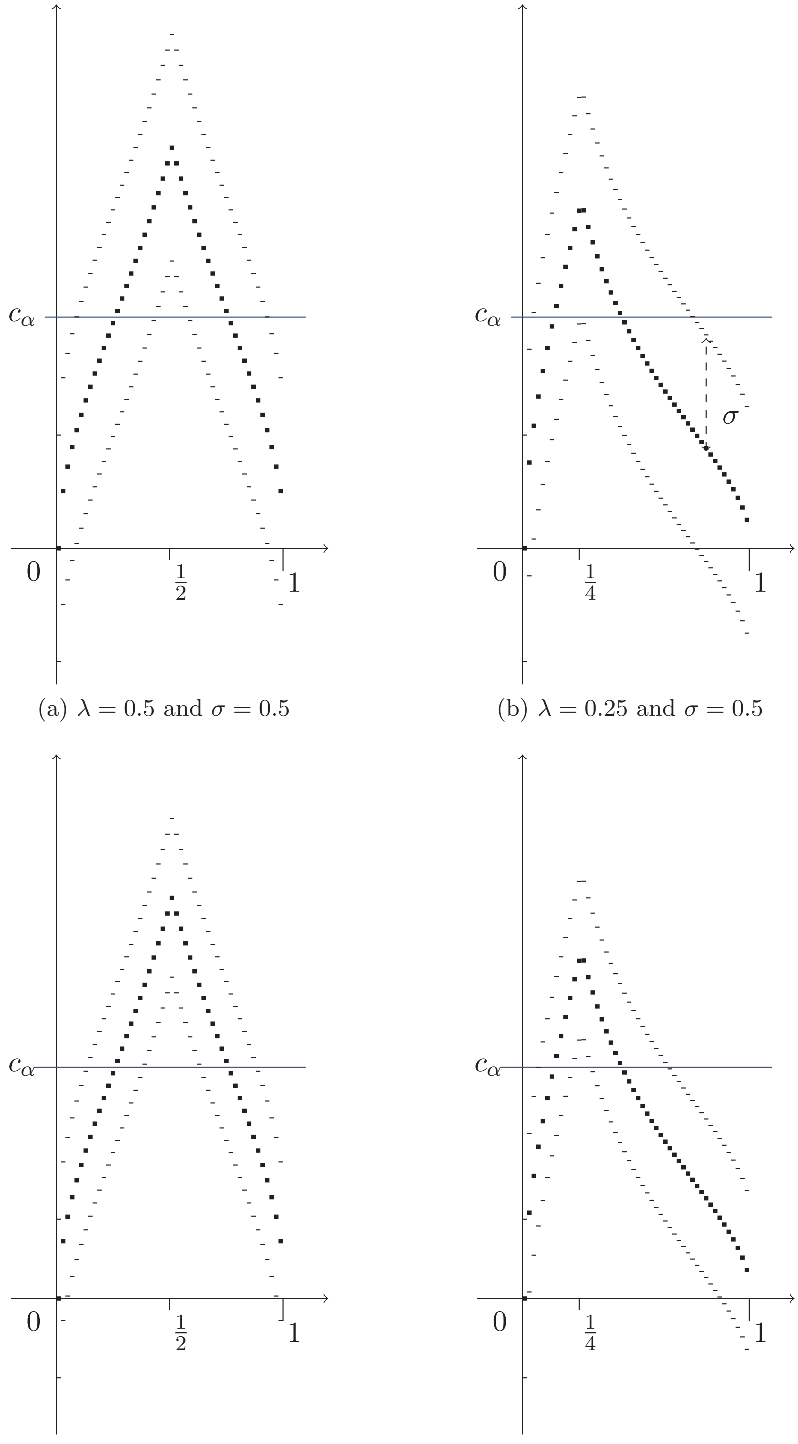


Figure 2.3.: Expectation and standard deviation of partial sums $V_{\lfloor sn \rfloor}(\hat{\theta}_n)$, $s \in (0,1)$ for $n = 50$, $\delta = 0.5$ with the corresponding critical value c_α (blue line) based on the asymptotic distribution



(a) $\lambda = 0.5$ and $\sigma = 0.5$

(b) $\lambda = 0.25$ and $\sigma = 0.5$

(c) $\lambda = 0.5$ and $\sigma = 0.2$

(d) $\lambda = 0.25$ and $\sigma = 0.2$

Figure 2.4.: Expectation and standard deviation of weighted partial sums $w_\gamma(\lfloor sn \rfloor / n)|V_{\lfloor sn \rfloor}|$, $s \in (0,1)$ for $n = 50$, $\delta = 0.5$ with the coresponding critical value c_α (blue line) based on the asymptotic distribution with $\gamma = \frac{1}{2}$

2.2.3. Asymptotic results

Now, we can analyse the asymptotic distribution of the test statistics. Due to the non-parametric set-up we can not determine the finite dimensional distribution of the test statistic. We prove the asymptotic distribution for the generalised mean change test statistic.

Theorem 2.2.1 *Under the model* (2.9)*, for the weight function $w_{\eta,\gamma}(k/n)$ (as in **N.7**) and under H_0* (1.2) *we have*

$$T_n(\eta, \gamma; A) = \max_{1 \le k < n} w_{\eta,\gamma}(k/n) \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_A \xrightarrow{d} \sup_{\eta < s < 1-\eta} \frac{\|W(s) - sW(1)\|_A}{(s(1-s))^\gamma},$$

where $\{W(s)\}$ is a Wiener process having covariance matrix Σ.

The proof of this theorem is analogous to the one with $A = \Sigma^{-1}$. For the one-dimensional case, this can be found in Csörgő and Horváth [1997].

Corollary 2.2.1 *Under the assumptions of Theorem 2.2.1 we have with $A' = \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$*

$$T_n(\eta, \gamma) = \max_{1 \le k < n} w_{\eta,\gamma}(k/n) \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_A \xrightarrow{d} \sup_{\eta < s < 1-\eta} \frac{\|B(s)\|_{A'}}{(s(1-s))^\gamma},$$

where $\{B(s)\}$ is a standard Brownian Bridge.

For the proofs we are going to use the following notation for standardized random vectors.

N.8 The notation $\tilde{\ }$ is used to indicate the multiplication from the left by $\Sigma^{-\frac{1}{2}}$.

Then the residuals $\{\varepsilon_t\}$ become standardized i.i.d. random vectors $\{\tilde{\varepsilon}_t\}$. It becomes clear how to prove that the test statistic gives an asymptotic level α-test.

The proof is done in three steps. First we show that the result holds true for $T_n(0, 0)$. To this end, we make use of the assumed model and then apply the functional central limit theorem (FCLT) for multidimensional i.i.d. random variables (Theorem C.2.2 gives an invariance principle for strong mixing, which covers the i.i.d. case and the FCLT follows from this), such that

$$\left\{ \Sigma^{-\frac{1}{2}} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \varepsilon_i \,,\ s \in [0,1] \right\} \stackrel{d}{=} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \tilde{\varepsilon}_i \,,\ s \in [0,1] \right\} \xrightarrow{d} \{W(s)\,,\ s \in [0,1]\}\,.$$

The conclusion of the theorem follows then from the continuous mapping theorem.

In the situation of $\eta \in (0, \frac{1}{2})$ and $\gamma \in [0, \frac{1}{2}]$ the result follows directly. Because the weight function is well defined, the FCLT holds true for $w_{\eta,\gamma}(k/n) \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_{\Sigma^{-1}}$ and since the weight function is bounded, the conclusion follows by the continuous mapping theorem.

If we do not restrict to the smaller inner interval (i.e. $\eta = 0$), we can use a modification of the weight function with $\gamma \in [0, \frac{1}{2})$. Due to the unbounded behaviour of the weight function at the boundaries, we have to use a truncation argument. Obviously in the inner interval we gain the result directly. At the boundaries we are going to show that uniformly in n the values are negligible if the cutting value tends to the boundary.

We have derived an asymptotic level α-test. The choice of the critical value for finite samples will be discussed in section 2.5.

For the different tests, we have identified the asymptotic distribution but it is also important to know that the test is consistent, i.e. we have an asymptotic power one test. Here, the assumption on a A-fixed alternative is important.

Theorem 2.2.2 *Let model* (2.9)*, assumption* ***G.1*** *and* H_1 (1.2) *be true. If* $\eta < \lambda < (1-\eta)$*, then*

$$T_n(\eta, \gamma; A) \xrightarrow{p} \infty\,.$$

To prove this, we do not have to make any use of a truncation argument, because the statistic is greater or equal to the weighted norm of the partial sum up to the change-point m. Making use of the CLT the asymptotic behaviour follows.

It is crucial to recognize that the constant assumption on δ_n is made w.r.t. the decision matrix A ($\|\delta_n\|_A \equiv \mathcal{D}$). So we allow the non-interesting entries of δ_n ($\|\delta_n - \delta'\|_A = 0$ with δ' equals δ_n) to behave arbitrary, as long as the estimator is still a $\sqrt{n}$-consistent estimator, i.e. δ_n is non-increasing (see Theorem 2.1.1 and 2.1.2).

Theorem 2.2.3 *Let* A *be a positive semi-definite matrix and*

$$T_n(A) := \max_{1 \le k < n} w(n,k) \|S(k;\theta)\|_A$$

a test statistic, with $w(n,k)$ *a possibly random weight function and* $S(k;\theta)$ *such that Theorem 2.2.1 and Theorem 2.2.2 hold true for this* A *as well as for* $A = I_d$*. Let* $\hat{A}_n$ *be an a.s. positive definite consistent estimator for* A*. For* $T_n(\hat{A}_n)$ *the result of Theorem 2.2.1 holds true. If* $\hat{A}_n$ *is additionally* $\sqrt{n}$*-consistent estimator Theorem 2.2.2 is also true for* $T_n(\hat{A}_n)$*.*

The key of the proof is to show that the difference between the test statistic with the estimator and the one with the true matrix is vanishing asymptotically.

2.2.4. Proofs

Lemma 2.2.1

Under the model (2.9) *and* H_0 (1.2) *we have*

$$T_n(\eta, \gamma; A) \xrightarrow{d} \sup_{\eta < s < (1-\eta)} w_{\eta,\gamma}(s) \|W(s) - sW(1)\|_A\,, \qquad (2.19)$$

where $\{W(s)\}$ *is Wiener process with covariance matrix* Σ, $\eta \in (0, \frac{1}{2})$ *and* $\gamma \in [0, \frac{1}{2}]$ *or* $\eta = 0 = \gamma$.

Proof:
We show the result in 2 steps. First we assume $\eta = \gamma = 0$ and secondly we show the result for $\eta \in (0, \frac{1}{2})$ and $\gamma \in [0, \frac{1}{2}]$.

Let us start with $\eta = \gamma = 0$. The main idea of the proof is, instead of analysing $\sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right)$ we can analyse the behaviour of $\sum_{i=1}^{k} (\varepsilon_i - \overline{\varepsilon}_n)$. The functional central limit theorem for multivariate i.i.d. random variables gives us

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{sn} \varepsilon_i \,,\ s \in [0,1] \right\} \xrightarrow{d} \{W(s)\,,\ s \in [0,1]\} \ ,$$

where $\{W(s)\}$ denotes a d-dimensional Wiener process with covariance matrix Σ. As projections are continuous a first application of the continous mapping theorem (see Theorem C.1.6) we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \varepsilon_i - \frac{\lfloor sn \rfloor}{n} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \xrightarrow{d} W(l) - lW(1) \qquad \text{for all } s \in [0,1] \,.$$

With the tightness of $\{\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor \cdot n \rfloor} \varepsilon_i - \frac{\lfloor \cdot n \rfloor}{n} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i\}$ and an additional application of the continuous mapping theorem, the proof is completed.

For $\eta \in (0, \frac{1}{2})$ and $\gamma \in [0, \frac{1}{2}]$ we need to analyse the weight function. We know

$$1_{\eta < s < (1-\eta)} \left(\frac{n^2}{sn(n - sn)} \right)^{\frac{1}{2}} = 1_{\eta < s < (1-\eta)} \frac{1}{\sqrt{s(1-s)}} \ ,$$

is a well defined function for $0 < \eta < \frac{1}{2}$. It follows by Slutsky C.1.4 that a FCLT is fulfilled. Therefore, by applying the monotone mapping theorem we get the result.

■

Lemma 2.2.2

Under the same assumptions as in Theorem 2.2.1 we derive for $\eta = 0$ and $\gamma \in (0, 1/2)$

$$T_n\big(\eta, \gamma; A\big) \xrightarrow{d} \sup_{0<s<1} \frac{\|W(s) - sW(1)\|_A}{(s(1-s))^{\gamma}} \,.$$

Proof:
We truncate the range of the maximum such that the weight function is bounded. At the boundaries, we are going to show that uniformly in n the effect is negligible if the truncation value τ tends to 0. To this end we use

$$\|Mx\|_A \leq \|M\|_A \, \|x\| \tag{2.20}$$

for $x \in \mathbb{R}^d$, $M \in \mathbb{R}^{d \times d}$ and $\|M\|_A = \sup_{\|v\|=1} \|Mv\|_A$. Which gives, analysing w.r.t. $\|\cdot\|$ is enough.

Since the weight function w_γ fulfils the assumptions of Theorem C.1.7, we have for $\tau \to 0$

$$\sup_{0<s<\tau,\,(1-\tau)<s<1} w_\gamma^2(s)\|B(s)\|^2 = o_P(1)\,,$$

where $\{B(s)\}$ is a d-dimensional standardised Brownian Bridge. Then we have

$$\begin{aligned}\sup_{0<s<\tau,\,(1-\tau)<s<1} w_\gamma^2(s)\|W(s)-sW(1)\|_A^2 &\le \left\|\Sigma^{\frac{1}{2}}\right\|_A^2 \sup_{0<s<\tau,\,(1-\tau)<s<1} w_\gamma^2(s)\|B(s)\|^2 \\ &= o_P(1)\,.\end{aligned}$$

On the other hand, we have for $\tau \to 0$

$$\begin{aligned}\max_{1\le k<\tau n} w_\gamma(k/n)\frac{k}{n}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n \varepsilon_t\right\| &\le \max_{1/n\le s<\tau} w_\gamma(s)s\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n \varepsilon_t\right\| \\ &= O_P(\tau^{1-\gamma})\,. \end{aligned} \tag{2.21}$$

So, for analysing $\max_{1\le k<\tau n,\,(1-\tau)n<k<n} w_\gamma(k/n)\frac{1}{\sqrt{n}}\left\|S_k(\hat\theta_n)\right\|$ it is left to analyse $\max_{1\le k<\tau n,\,(1-\tau)n<k<n} w_\gamma(k/n)\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^k \varepsilon_t\right\|$. From the Hájek-Rényi inequality we get for an i.i.d. sequence $\{\varepsilon_t\}$ that

$$\max_{a_n\le k\le b_n}\left\|c_k\sum_{t=1}^k \varepsilon_t\right\| = O_P\left((a_nc_{a_n}^2+\sum_{k=a_n+1}^{b_n} c_k^2)^{\frac{1}{2}}\right)\,.$$

Here, $a_n = 1$, $b_n = \tau n$ and $c_k = w_\gamma(k/n)$, such that for all fixed n we have for $\tau \to 0$

$$\begin{aligned}a_nc_{a_n}^2+\sum_{k=a_n+1}^{b_n} c_k^2 &\le w_{2\gamma}(1/n) + n\int_{1/n}^{\tau} w_{2\gamma}(s)ds \\ &\le 2^{2\gamma}n^{2\gamma} + (-2\gamma+1)^{-1}(1-\tau)^{-2\gamma}n\left(\tau^{-2\gamma+1}-n^{2\gamma-1}\right) \\ &\le 2^{2\gamma}n^{2\gamma} + \frac{2^{2\gamma}}{1-2\gamma}n\tau^{-2\gamma+1} - \frac{2^{2\gamma}}{1-2\gamma}n^{2\gamma} \le \frac{2^{2\gamma}}{1-2\gamma}n\tau^{1-2\gamma}\,.\end{aligned}$$

Thus it holds

$$\max_{1\le k\le \tau n}\left\|w(k/n)\sum_{t=1}^k \varepsilon_t\right\| = O_P\left(\left(n\tau^{1-2\gamma}\right)^{\frac{1}{2}}\right) \tag{2.22}$$

Combining (2.21) and (2.22) we get uniformly in n

$$\begin{aligned}&\max_{1\le k<\tau n} w_\gamma(k/n)\frac{1}{\sqrt{n}}\left\|S_k(\hat\theta_n)\right\| \\ &\quad\le \frac{1}{\sqrt{n}}\max_{1\le k<\tau n} w_\gamma(k/n)\left\|\sum_{t=1}^k \varepsilon_t\right\| \\ &\qquad + \max_{1\le k<\tau n} w_\gamma(k/n)\frac{k}{n}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n \varepsilon_t\right\| \\ &\quad= O_P(\tau^{\frac{1}{2}-\gamma}) + O_P(\tau^{1-\gamma}) \\ &\quad= O_P(\tau^{\frac{1}{2}-\gamma})\,. \end{aligned} \tag{2.23}$$

For $(1-\tau)n < k < n$ we observe

$$\max_{(1-\tau)n<k<n} w_\gamma(k/n)\frac{1}{\sqrt{n}}\left\|S_k(\hat\theta_n)\right\| = \max_{(1-\tau)n<k<n} w_\gamma(k/n)\frac{1}{\sqrt{n}}\left\|\sum_{t=k+1}^{n}\varepsilon_t - \frac{n-k}{n}\sum_{t=1}^{n}\varepsilon_t\right\| .$$

Using the symmetry of $w_\gamma(k/n)$ and $l = n-k$, we derive equivalent sums as for $1 \le k \le \tau n$. Analogue arguments yield

$$\max_{(1-\tau)n<k<n} w_\gamma(k/n)\frac{1}{\sqrt{n}}\left\|S_k(\hat\theta_n)\right\| = O_P(\tau^{\frac{1}{2}-\gamma}) . \tag{2.24}$$

From Lemma 2.2.1 we have for each fixed but arbitrary $\tau > 0$ and $x \in \mathbb{R}$

$$P(T_n(\tau,\gamma;A) < x) \underset{n\to\infty}{\longrightarrow} P(\sup_{\tau<s<(1-\tau)} w_\gamma(s)\|W(s) - sW(1)\|_A < x) .$$

We can conclude that for all $x \in \mathbb{R}$ it holds

$$\begin{aligned}\lim_{n\to\infty} P(T_n(0,\gamma;A) < x) &= \lim_{\tau\to 0}\lim_{n\to\infty} P(T_n(\tau,\gamma;A) < x)\\ &= \lim_{\tau\to 0} P(\sup_{\tau<s<(1-\tau)} w_\gamma(s)\|W(s) - sW(1)\|_A < x)\\ &= P(\sup_{0<s<1} w_\gamma(s)\|W(s) - sW(1)\|_A < x) .\end{aligned}$$

■

Theorem 2.2.1

Under the model (2.9), for the weight function $w_{\eta,\gamma}(n,k)$ (as in **N.7**) and H_0 (1.2) we have

$$T_n(\eta,\gamma;A) = \max_{1\le k<n} w_{\eta,\gamma}(k/n)\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k}\left(X_t - \hat\theta_n\right)\right\|_A \xrightarrow{d} \sup_{\eta<s<1-\eta}\frac{\|W(s) - sW(1)\|_A}{(s(1-s))^\gamma} ,$$

where $\{W(s)\}$ is a d-dimensional Wiener process with covariance matrix Σ.

Proof:
The result follows from Lemma 2.2.1 (for $\eta \in (0,\frac{1}{2})$ and $\gamma \in [0,\frac{1}{2}]$ or $\eta = 0 = \gamma$) and Lemma 2.2.2 ($\eta = 0$ and $\gamma \in [0,\frac{1}{2})$). ■

Corollary 2.2.2 *Under the assumptions of Theorem 2.2.1 and with $A' = \Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}$, we get*

$$T_n(\eta,\gamma;A) \xrightarrow{d} \sup_{\eta<s<1-\eta}\frac{\|B(s)\|_{A'}}{(s(1-s))^\gamma} ,$$

where $\{B(s)\}$ is a standard Brownian Bridge.

Proof:
Follows directly from $\|x\|_A = \left\|\Sigma^{-\frac{1}{2}}x\right\|_{A'}$. We have

$$\left\{\Sigma^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{sn}\varepsilon_t\,,\ s\in[0,1]\right\} \overset{d}{=} \left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{sn}\tilde\varepsilon_t\,,\ s\in[0,1]\right\} \xrightarrow{d} \{W(s)\,,\ s\in[0,1]\} ,$$

with $\{W(s)\}$ being a standard Wiener process. Thus, from Theorem 2.2.1 we get

$$T_n(\eta,\gamma) \xrightarrow{d} \sup_{\eta<s<1-\eta} \frac{\|W(s)-sW(1)\|_{A'}}{(s(1-s))^{\gamma}} = \sup_{\eta<s<1-\eta} \frac{\|B(s)\|_{A'}}{(s(1-s))^{\gamma}}\,.$$

■

Theorem 2.2.2

Let model (2.9), assumption **G.1** and H_1 (1.2) be true, then

$$T_n\big(\eta,\gamma;A\big) \xrightarrow{p} \infty\,.$$

Proof:

Let $k_0 = \lfloor \kappa n \rfloor$, with $\kappa = \lambda$ if $\eta < \lambda < (1-\eta)$ and $\kappa = \eta$ otherwise. The main idea is to show that the sum up to m converges to ∞. Observe that

$$\begin{aligned} T_n(\eta,\gamma;A) &\geq w_{\gamma}(k_0/n)\frac{1}{\sqrt{n}}\left\|S_{k_0}(\hat{\theta}_n)\right\|_A \\ &\geq \frac{1}{\sqrt{n}}\left\|S_{k_0}(\hat{\theta}_n)\right\|_A \\ &\geq \frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k_0}(X_t-\hat{\theta}_n)\right\|_A \\ &= \frac{1}{\sqrt{n}}\left\|\left(\sum_{t=1}^{k_0}(\varepsilon_t-\overline{\varepsilon}_n)-\delta\frac{k_0(n-m)}{n}\right)\right\|_A. \end{aligned}$$

Because it holds

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{k_0}(\varepsilon_t-\overline{\varepsilon}_n)-\frac{k_0(n-m)}{n\sqrt{n}}\delta\right\|_A \geq \left|\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k_0}(\varepsilon_t-\overline{\varepsilon}_n)\right\|_A - \sqrt{n}\frac{k_0(n-m)}{n^2}\|\delta\|_A\right|,$$

with $\kappa(1-\lambda)-\frac{k_0}{n}(1-\frac{m}{n}) = \kappa-\frac{k_0}{n}-\kappa(\lambda-\frac{m}{n})-\lambda(\kappa-\frac{k_0}{n}) = O\left(\frac{1}{n}\right)$ we have

$$\begin{aligned} T_n(\eta,\gamma;A) &\geq \frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k_0}(\varepsilon_t-\overline{\varepsilon}_n)-\frac{k_0(n-m)}{n}\delta\right\|_A \\ &= \left|O_P(1)-\sqrt{n}\,(\kappa(1-\lambda)\|\delta\|_A)+o(1)\right| \\ &\underset{n\to\infty}{\longrightarrow} \infty\,. \end{aligned}$$

■

Theorem 2.2.3

Let A be a positive semi-definite matrix and

$$T_n(A) := \max_{1\leq k<n} w(n,k)\|S(k;\theta)\|_A$$

a test statistic, with $w(n,k)$ a possibly random weight function and $S(k;\theta)$ such that Theorem 2.2.1 and Theorem 2.2.2 hold true for this A as well as for $A = I_d$. Let $\hat{A}_n$ be an a.s. positive definite consistent estimator for A. For $T_n(\hat{A}_n)$ the result of Theorem 2.2.1 holds true. If $\hat{A}_n$ is additionally $\sqrt{n}$-consistent estimator Theorem 2.2.2 is also true for $T_n(\hat{A}_n)$.

Proof:
For the test statistic we have

$$\begin{aligned}\left|T_n(\eta,\gamma;\hat{A}_n)-T_n(\eta,\gamma,A)\right| &\leq \max_{1\leq k<n} w(k/n)\left\|\hat{A}_n^{\frac{1}{2}}S_k(\hat{\theta}_n)-A^{\frac{1}{2}}S_k(\hat{\theta}_n)\right\| \\ &\leq \left\|\hat{A}_n^{\frac{1}{2}}-A^{\frac{1}{2}}\right\| \max_{1\leq k<n} w(k/n)\left\|S_k(\hat{\theta}_n)\right\| \\ &= \left\|\hat{A}_n^{\frac{1}{2}}-A^{\frac{1}{2}}\right\| T_n(I_d)\,.\end{aligned}$$

That $\left\|\hat{A}_n^{\frac{1}{2}}-A^{\frac{1}{2}}\right\| \xrightarrow{p} 0$ follows directly from the consistency of the estimator. Hence, the convergence under H_0 is not changed as $T_n(I_d)=O_P(1)$.

For the consistency, we observe, with $k_0=\lfloor \max(\eta,\lambda)n\rfloor$,

$$\begin{aligned}T_n(\eta,\gamma;\hat{A}) &\geq \frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k_0}(\varepsilon_t-\overline{\varepsilon}_n)-\frac{k_0(n-m)}{n}\delta\right\|_{\hat{A}_n} \\ &= \left|O_P(1)-\sqrt{n}(\kappa(1-\lambda))\|\delta\|_A+O_P(1)+o(1)\right| \\ &\underset{n\to\infty}{\longrightarrow} \infty\,,\end{aligned}$$

where we used analogous arguments as above for $\left\|\sum_{t=1}^{k_0}(\varepsilon_t-\overline{\varepsilon}_n)\right\|_{\hat{A}_n}$ and the $\sqrt{n}$-consistency for $\|\delta\|_{\hat{A}_n}$.

■

2.3. Change-point estimator

After deriving a consistent and asymptotic level-α test, we are interested in the estimator of the change-point. We want to determine the rate of convergence and verify it by determining the asymptotic distribution.

We still assume a mean change model (2.1) with a fixed but arbitrary size of the change ($\delta_n \equiv \delta$ constant). The test statistic is given by

$$T_n(\eta, \gamma; A) = \max_{\eta n < k < (1-\eta)n} \left(\frac{n^2}{k(n-k)} \right)^{\gamma} \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_A$$

as we derived in section 2.2 (compare (2.12)). With the notation **N.7** of the weight function $w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta < s < (1-\eta)\}} \left(s(1-s) \right)^{-\gamma}$ the test statistic is defined as

$$T_n(\eta, \gamma; A) = \max_{1 \le k < n} w_{\eta,\gamma}(k/n) \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_A .$$

We are going to see in subsection 2.3.1, that the argument of the maximum is a good estimator for the change-point. The effect of the weight function for the change-point estimator will be discussed in subsection 2.3.2. In this section we also state useful properties of this function. The main results for the asymptotics of the estimator are given in section 2.3.3. We show n-consistency of the estimator and verify it by determining the asymptotic distribution. Detailed proofs are given in section 2.3.4.

2.3.1. The estimator

We have derived the test statistic T_n and discussed the effect of weight function. To this end, we calculated the expectation for the unweighted and weighted partial sums (see (2.13) and (2.16) on page 18). We observed, that the maximum of the expectations is approached at the time of change. Thus the argumentum maximum is a good estimator for the change-point. So there is a close relation between change-point estimator and test statistic. With the test statistic $T_n(\eta, \gamma; A)$ the change-point estimator is of the form

$$\hat{m}(\eta, \gamma; A) = \operatorname*{arg\,max}_{1 \le k < n} w_{\eta,\gamma}(k/n) \left\| \sum_{i=1}^{k} \left(X_i - \hat{\theta}_n \right) \right\|_A , \tag{2.25}$$

with $w_{\eta,\gamma}(n, k)$ as in **N.7** on page 16.

In the one-dimensional case asymptotic results for the change-point estimator are given in Antoch et al. [1995] and Csörgő and Horváth [1997] (section 2.8.1). Multivariate analyses of change-point estimators are rare. For some models multivariate analyses of change-point estimators are published, e.g. in the case of linear multiple regression Bai [1997] a analyse of fixed and shrinking magnitude are done for the non-weighted test statistic. We use the opportunity to state the results for the multivariate case. The proofs follow the ideas of Csörgő and Horváth [1997].

As in the proofs for the asymptotic behaviour of the change-point test, we also make use of properties of the weight function here. We take a closer look at the weight function to obtain a better understanding of its effect on the change-point estimator and state the properties needed later on.

2.3.2. Effect and properties of the weight function

In this section we take a closer look at the influence of the weight function on the change-point estimator as we did in section 2.2.2 for the test statistic. We therefore assume i.i. normally distributed residuals. We compare the expectation and standard deviation for the unweighted and weighted partial sums in case of an early change and in a situation where the change occurs in the middle.

In Figure 2.5 the expectations (thick lines) and expectations plus/minus the standard deviations (thin lines) for the weighted (first row) and the unweighted partial sums (second row) are shown. The region for each partial sum where it is likely observed (around 70%) is between the thin vertical lines. The horizontal solid line corresponds to the maximal value of the partial sum minus the standard deviation. The event that the maximum of the partial sums is less than this value is unlikely. All values of k for which the partial sum has mean plus the standard deviation larger than the solid black line are also likely values of the change-point estimator. So, we compare this regions, drawn in blue lines, for the weighted and unweighted estimators.

We observe, that the intervals (indicated by the blue line) for a change in the middle are smaller for the unweighted estimator. For early changes the weighted estimator (second row) has a smaller interval. In conclusion, for changes in the middle the unweighted estimator is preferable and for early changes the weighted one.

For the proofs of the asymptotics for the weighted change-point estimator we are going to make use of properties of the weight function. A key property is the Lipschitz continuity of the weight function $w_\gamma(k/n)$ defined on a compact set $k \in [\alpha n, \beta n]$.

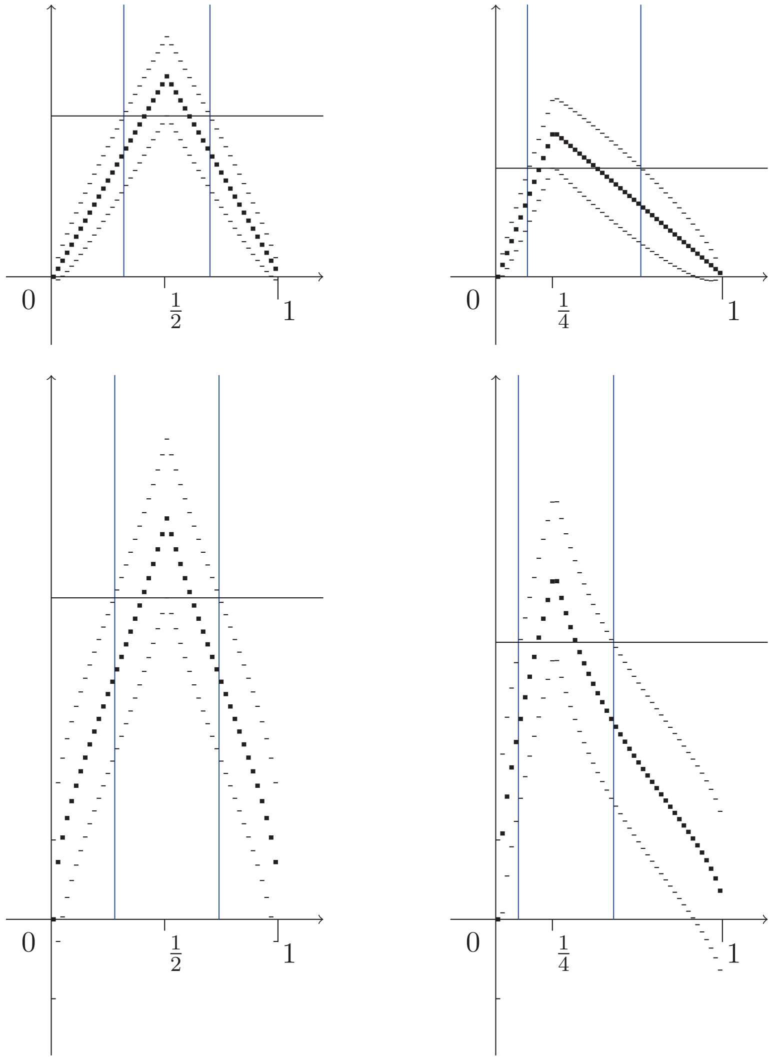


Figure 2.5.: Comparison of unweighted (first row) and weighted (second row) partial sums $w_\gamma(\lfloor sn \rfloor/n)V_{\lfloor sn \rfloor}$ ($\gamma = 0$ and $\gamma = \frac{1}{2}$, respectively) in view of estimating the change-point based on $n = 50$ observations

Lemma 2.3.1

For the function $w_\gamma(k/n) = \left(\frac{k}{n}(1 - \frac{k}{n})\right)^{-\gamma}$, $\gamma > 0$, and $k, l \in [\alpha n, \beta n]$, $0 < \alpha < \beta < 1$, there exists a constant $c > 0$ with

$$|w_\gamma(k/n) - w_\gamma(l/n)| \leq |k - l| c\, n^{-1}\,.$$

Proof:
By the mean value theorem applied to $w_\gamma(s) = s^{-\gamma}(1 - s)^{-\gamma}$ we have

$$|w_\gamma(k/n) - w_\gamma(m/n)| = |k - m||w'_\gamma(\tilde{s})\frac{1}{n}|\,, \tag{2.26}$$

with $\tilde{s}$ between k/n and m/n and derivative

$$\begin{aligned} w'_\gamma(s) &= -\gamma s^{-\gamma-1}(1-s)^{-\gamma} + \gamma s^{-\gamma}(1-s)^{-\gamma-1} \\ &= \gamma w_{\gamma+1}(s)(2s-1)\,. \end{aligned}$$

This is a bounded function on $[\alpha,\beta] \subset [0,1]$ such that

$$|w'_\gamma(\tilde{s})| \leq \max_{\alpha<s<\beta} |w'(s)| =: c\,. \tag{2.27}$$

■

A very useful result, not only in this special case but for all bounded and Lipschitz continuous functions, is, that this family is closed under multiplication. This results in the following important properties of the weight function.

Corollary 2.3.1 *Let $w_\gamma(k/n)$, k, l, α and β be as in Lemma 2.3.1*

$$|w_\gamma^2(k/n) - w_\gamma^2(l/n)| \leq |k-l| c_1 n^{-1} \tag{2.28}$$

$$|w_\gamma^2(k/n)k - w_\gamma^2(l/n)l| \leq |k-l| c_2 \tag{2.29}$$

$$|k-l| c_3 n \leq |w_\gamma^2(k/n)k^2 - w_\gamma^2(l/n)l^2| \leq |k-l| c'_3 n \tag{2.30}$$

with c_1, c'_1, c_2 and c_3 constants depending on α, β and γ. Moreover, there exists a constant c_4 such that

$$w_\gamma^2(l/n)l^2 - w_\gamma^2(k/n)k^2 > c_4 > 0 \tag{2.31}$$

with $1 \leq k < l < n$.

Proof:
The inequality (2.28) is clear since $w_\gamma^2(s) = w_{2\gamma}(s)$. With the Lipschitz property for linear functions on a bounded set, the inequalities (2.29) and (2.30) (right hand side) follow directly.

The left inequality of (2.30) follows with the mean value theorem equivalent to the proof of Lemma 2.3.1, just using $|w'_\gamma(\tilde{s})| \geq \min_{\alpha<s<\beta} |w'_\gamma(s)|$ for $\tilde{s} \in [k/n, l/n]$.

Due to the specific choice of the weight function, we can prove (2.31) by calculation. From the mean value theorem we get for $\tilde{s} \in [k/n, l/n]$

$$\begin{aligned} w_\gamma^2(l/n)l^2 - w_\gamma^2(k/n)k^2 &= (l-k)n\,2\,w_\gamma(\tilde{s})\tilde{s}\big(\gamma w_{\gamma+1}(\tilde{s})(2\tilde{s}-1)\tilde{s} + w_\gamma(\tilde{s})\big) \\ &= (l-k)w_{2\gamma+1}(\tilde{s})\tilde{s}^2 n 2\left(\gamma\,(2\tilde{s}-1) + 1 - \tilde{s}\right) \\ &= \underbrace{(l-k)\,2\,w_{2\gamma+1}(\tilde{s})\tilde{s}^2 n}_{>c>0}\Bigg(\underbrace{(1-\gamma)\,\tilde{s} + \gamma\tilde{s}}_{=f(\tilde{s})}\Bigg) \end{aligned}$$

with $f(s) > 0$ for all $\alpha < s < \beta$ and c some constant depending on k, l and n, where we use $\gamma \leq \frac{1}{2}$. ■

Knowing how the weight function behaves in a truncated area, we also need to know how it behaves at the boundaries.

Lemma 2.3.2

The weight function $w_\gamma(\cdot)$ fulfils

$$\lim_{\alpha\to 0}\sup_{s\in(0,\alpha)} w_\gamma(s)s = 0 \qquad \textit{and} \qquad \lim_{\alpha\to 0}\sup_{s\in(1-\alpha,1)} w_\gamma(s)(1-s) = 0\,.$$

Proof:
We show $\lim_{\alpha\to 0}\sup_{s\in(0,\alpha)} w_\gamma(s)s < \infty$; the other inequality follows analogously.
It holds

$$\begin{aligned}\lim_{\alpha\to 0}\sup_{s\in(0,\alpha)} w_\gamma(s)s &= \lim_{\alpha\to 0}\sup_{s\in(0,\alpha)} s^{1-\gamma}(1-s)^{-\gamma}\\ &\le \lim_{\alpha\to 0}\sup_{s\in(0,\alpha)} s^{1-\gamma}(1-\alpha)^{-\gamma}\\ &= 0\,.\end{aligned}$$

■

2.3.3. Asymptotic results

First of all it is important to know whether, for increasing n, the estimator is consistent. This is not only of interest for the practical application of the estimator, it also helps in further proofs. In [Csörgő and Horváth, 1997] they proved weak consistency. For other examples, almost sure consistency is also proven (see, e.g. Antoch et al. [1995] or Kirch and Tadjuidje Kamgaing [2012]). Later on we only need weak consistency for the change-point estimator, so we prove convergence in probability for the non-rescaled version of the change-point estimator.

Theorem 2.3.1 *Assume assumption **G.1**, the model* (2.1) *and let the alternative H_1* (1.2) *hold. If the change-point estimator* (2.25) *fulfils $\eta < \lambda < (1-\eta)$, then*

$$\hat{m}(\eta,\gamma;A) - m = o_P(n)\,.$$

In particular $\frac{\hat{m}(\eta,\gamma;A)}{n}$ is a consistent estimator for λ.

The key of the proof is to show uniform convergence of the test statistic. Then with the Lemma 3.1 of [Pötscher and Prucha, 1997] (see C.1.2 on page 209) we get the claim. For the proof of the uniform convergence, we use the LIL (see (C.3)). The detailed proof is given in section 2.3.4.

As we mentioned, Theorem 2.3.1 shows consistency. In applications, we are also interested in confidence regions for the time of change. To this end, we first determine the rate of convergence. Secondly, we show that this rate is the best by determining the asymptotic distribution.

Csörgő and Horváth [1997] (proof of Theorem 2.8.1) stated that with the weak consistence it is enough to prove the rate of convergence only for the truncated change-point estimator. The proof makes use of the definition of the change-point estimator only, and not of its specific form depending on the model.

Corollary 2.3.2 *Let $\hat{m}(\eta, \gamma; A)$ and $\hat{m}(\alpha, \gamma; A)$ denote two change-point estimators of the form*

$$\hat{m}(\beta, \gamma; A) = \arg\max_{1 \le k < n} w_{\beta,\gamma}(k/n) \left\| \sum_{t=1}^{k} H(X_t, \hat{\theta}_n) \right\|_A$$

with η, α, γ, A and function H such that

$$\hat{m}(\beta, \gamma; A) - m = o_P(n)$$

for $\beta = \eta$ or $\beta = \alpha$ and $m = \lfloor \lambda n \rfloor$. Then it holds

$$P(\hat{m}(\eta, \gamma; A) = \hat{m}(\alpha, \gamma; A)) \underset{n \to \infty}{\longrightarrow} 1 \,. \tag{2.32}$$

As we see, this Corollary is for a general change-point estimator. The proof follows essentially from the $o_P(n)$ convergence as well as from the specific definition as the argument of the maximum. It is clear, that in our specific situation we have the assumptions fulfilled as shown in Theorem 2.3.1.

The main idea of the proof is that for $\eta < \alpha < \frac{1}{2}$

$$\begin{aligned} &P(|\hat{m}(\eta, \gamma; A) - \hat{m}(\alpha, \gamma; A)| > 0) \\ =&P(\max_{\substack{\eta n \le k < \alpha n \\ (1-\alpha)n \le k < (1-\eta)n}} w_{0,\gamma}(k/n) \left\| S_k(\overline{X}_n) \right\|_A > \max_{\alpha n \le k < (1-\alpha)n} w_{0,\gamma}(k/n) \left\| S_k(\overline{X}_n) \right\|_A) \end{aligned}$$

results in the event that $|\frac{\hat{m}(\eta,\gamma;A)}{n} - \lambda| > \max(\lambda - \alpha, 1 - \alpha - \lambda)$. Using the definition of the stochastic Landau symbols (see Appendix A.2) and Theorem 2.3.1, which gives the consistency of $\hat{m}/n$, completes the proof.

In a next step as done in [Antoch et al., 1995] and [Csörgő and Horváth, 1997], we prove the rate of convergence, i.e. $\hat{m}(\eta, \gamma; A) - m = O_P(1)$. This is not the best rate in the case of the local alternative, but the proof is analogue. In the literature the proof for the fixed alternative is rarely shown. We show the proof of the multivariate case in detail. For the rest of the thesis we will refer to the main ideas and discuss the difficulties in the special cases.

Theorem 2.3.2 *Under the assumptions of Theorem 2.3.1 the change-point estimator $\hat{m}(\eta, \gamma; A)$* (2.25) *fulfils*

$$\hat{m}(\eta, \gamma; A) - m = O_P(1) \,.$$

For the proof, we use that the estimator is n-consistent and therefore focus on the estimators with truncated maximum range, because asymptotically they are equivalent. We then replace the weighted norm of the partial sum by another but equivalent value, i.e. the argument of the maximum of $w_{0,\gamma}(k/n)\left\|S_k(\hat{\theta}_n)\right\|_A$ is the same as that of V_k, with

$$\begin{aligned}V_k = w_{0,\gamma}(k/n)&\left\|S_k(\hat{\theta}_n)\right\|_A \\ &+ w_{0,\gamma}(k/n)w_{0,\gamma}(m/n)S_k^{\mathrm{T}}(\hat{\theta}_n)AS_m(\hat{\theta}_n) \\ &- w_{0,\gamma}(k/n)w_{0,\gamma}(m/n)S_m^{\mathrm{T}}(\hat{\theta}_n)AS_k(\hat{\theta}_n) \\ &- w_{0,\gamma}(m/n)\left\|S_m(\hat{\theta}_n)\right\|_A .\end{aligned}$$

By the properties of the weight function in addition to the Hájek-Rényi inequality and the CLT, we determine by the right decomposition the dominating parts. Details can be found in section 2.3.4.

We have seen, that the rescaled change-point estimator $(\frac{\hat{m}}{n})$ converges with rate $\frac{1}{n}$. From the proof of Theorem 2.3.2 we get the decomposition as well as the dominating parts. This is useful for the proof of the asymptotic distribution. To justify that this rate cannot be improved, we determine the asymptotic distribution.

Theorem 2.3.3 *Let the assumptions of Theorem 2.3.1 hold true. Then*

$$\hat{m}(\eta,\gamma;A) - m \xrightarrow{d} \arg\max\{W(s) - |s|g(s)\mathcal{D}^2,\ s \in \mathbb{Z}\}$$

with $\mathcal{D} = \|\delta\|_A$,

$$W(s) = \begin{cases} 0 & , \ s = 0\,, \\ \delta^{\mathrm{T}}A\Sigma^{\frac{1}{2}}\sum_{i=s}^{-1}\xi_i^{(2)} & , \ s < 0\,, \\ \delta^{\mathrm{T}}A\Sigma^{\frac{1}{2}}\sum_{i=1}^{s}\xi_i^{(1)} & , \ s > 0\,, \end{cases}$$

$\xi_i^{(z)}$ *are i.i.d. with* $\mathcal{L}\left(\xi_1^{(z)}\right) = \mathcal{L}\left(\tilde{\varepsilon}_1\right)$ *for* $z = 1, 2$ *and*

$$g(s) = \begin{cases} (1-\gamma)(1-\lambda) + \gamma\lambda & , \ s < 0\,, \\ 0 & , \ s = 0\,, \\ \gamma(1-\lambda) + (1-\gamma)\lambda & , \ s > 0\,. \end{cases}$$

With the same decomposition as we used in proof of the theorem about the convergence rate, we are able to determine the asymptotic distribution. From before, we know the asymptotic dominating terms, so the asymptotic distribution is given by these parts. For the proof, we observe that the difference can also be analysed on a sufficiently large subset. Due to the consistency the probability of observing values outside this subset will tend to zero. As the maximum is now taken over a finite set, it follows that all the other parts are negligible, which can be shown using again the triangle inequality, Cauchy-Schwarz inequality as well as the CLT and the Hájek-Rényi inequality. For the deterministic part we make use of properties of the weight function to determine the limit behaviour. On the other hand we show the limit for the stochastic part, by rewriting this part as a function on $k - m$.

In the one-dimensional case, see Antoch et al. [1995], the distribution is given as

$$\hat{m}(\eta,\gamma) - m \xrightarrow{d} \arg\max\{\delta\, W(s) - |s| g(s)\delta^2,\ s \in \mathbb{Z}\}$$

with $\{W(s)\}$ denoting a two-sided standardised random walk which is equivalent to our result.

We have discussed the influence of the weight function $w_{\eta,\gamma}$ in section 2.3.2. We now observe the influence of the parameters η and γ. First of all, for large observations the influence of η vanishes. But the effect of γ still exists. We can see, for a change in the middle, the decision of γ has no effect. But on the other hand, if $\lambda \neq \frac{1}{2}$, i.e. early or late changes, we would prefer $\gamma = \frac{1}{2}$ as then the distribution is symmetric. This coincides with our observations in section 2.3.2.

2.3.4. Proofs

Theorem 2.3.1

Assume the model (2.1), assumption **G.1** and H_1 (1.2). If the change-point estimator (2.25) fulfils $\eta < \lambda < (1-\eta)$, then

$$\hat{m}(\eta,\gamma;A) - m = o_P(n)\,.$$

In particular $\frac{\hat{m}(\eta,\gamma;A)}{n}$ is a consistent estimator for λ.

For the proof of Theorem 2.3.1 as for the rest of the thesis we are going to use the following notation:

$$g_n(s) = \begin{cases} -\frac{\lfloor sn \rfloor}{n}\left(1 - \frac{\lfloor \lambda n \rfloor}{n}\right) & , s < \lambda \\ -\frac{\lfloor \lambda n \rfloor}{n}\left(1 - \frac{\lfloor sn \rfloor}{n}\right) & , s \geq \lambda\,. \end{cases} \tag{2.33}$$

Proof of Theorem 2.3.1:
We first give the proof for $\eta = 0$ and $\gamma = 0$. In the case $\eta \in (0, \frac{1}{2})$ we can directly get the result from the first part. But for $\eta = 0$ and $\gamma \in (0, \frac{1}{2})$ we have to use a truncation argument as for the proof of Lemma 2.2.2.

First observe that by the model (2.1) we have

$$\begin{aligned} \sum_{t=1}^{\lfloor sn \rfloor}\left(X_t - \hat{\theta}_n\right) &= \sum_{t=1}^{\lfloor sn \rfloor}(\varepsilon_t - \overline{\varepsilon}_n) + \left(\mathbf{1}_{\{s>\lambda\}}(\lfloor sn \rfloor - \lfloor \lambda n \rfloor) - \lfloor sn \rfloor\left(1 - \frac{\lfloor \lambda n \rfloor}{n}\right)\right)\delta\,. \\ &= \sum_{t=1}^{\lfloor sn \rfloor}(\varepsilon_t - \overline{\varepsilon}_n) + n g_n(s)\delta \end{aligned} \tag{2.34}$$

Let

$$L_\gamma(s) := \begin{cases} (s(1-s))^{-\gamma}\left\|s(1-\lambda)\delta\right\|_A & , s < \lambda \\ (s(1-s))^{-\gamma}\left\|\lambda(1-s)\delta\right\|_A & , s \geq \lambda\,. \end{cases} \tag{2.35}$$

If we can prove

$$\sup_{s\in(\eta,1-\eta)}\left| w_\gamma\left(\tfrac{\lfloor sn \rfloor}{n}\right)\frac{1}{n}\left\|\sum_{t=1}^{\lfloor sn \rfloor}\left(X_t - \hat{\theta}_n\right)\right\|_A - L_\gamma(s)\right| = o_{a.s.}(1)\,, \tag{2.36}$$

where $w_\gamma(s) = (s(1-s))^{-\gamma}$, Lemma 3.1 of Pötscher and Prucha [1997] (Theorem C.1.2) yields

$$\frac{\hat{m}(\eta,\gamma)}{n} - \frac{m}{n} = o_{a.s.}(1)\,.$$

Let us prove equation (2.36).

We first assume $\eta = 0$ and $\gamma = 0$.
Observe, that the law of iterated logarithm holds true such that

$$\begin{aligned}
\left\| \sum_{t=1}^{\lfloor sn \rfloor} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A &= O(1) \left\| \sum_{t=1}^{\lfloor sn \rfloor} (\tilde{\varepsilon}_t - \overline{\tilde{\varepsilon}}_n) \right\| \\
&= O(1) \left(\left\| \sum_{t=1}^{\lfloor sn \rfloor} \varepsilon_t \right\| + \frac{\lfloor sn \rfloor}{n} \left\| \sum_{t=1}^{n} \varepsilon_t \right\| \right) \\
&= O_{a.s.}\left(\sqrt{\lfloor sn \rfloor \log\log(\lfloor sn \rfloor)} \right) + \frac{\lfloor sn \rfloor}{n} O_{a.s.}\left(\sqrt{n \log\log(n)} \right) \\
&= O_{a.s.}\left(\sqrt{n \log\log(n)} \right) \qquad (2.37)
\end{aligned}$$

uniformly in s.

First using equation (2.34) and secondly the triangle inequality gives

$$\begin{aligned}
\left\| \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \hat{\theta}_n \right) \right\|_A &= \left\| \sum_{t=1}^{\lfloor sn \rfloor} (\varepsilon_t - \overline{\varepsilon}_n) + n g_n(s)\delta \right\|_A \\
&\leq \left\| \sum_{t=1}^{\lfloor sn \rfloor} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A + n \| g_n(s)\delta \|_A\,,
\end{aligned}$$

with $g_n(s)$ as in (2.33). By the convergence of $g_n(s)$ and (2.37) it holds

$$\left\| \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \hat{\theta}_n \right) \right\|_A = O_{a.s.}\left(\sqrt{n \log\log(n)} \right) + n(L_0(s) + o_{a.s.}(1))$$

with $L_0(s)$ as in (2.35). The result (2.36) follows directly by weighting with $\frac{1}{n}$. Because the convergence of the difference does not depend on s, the supremum converges with the same rate.

We derive the proof for $\eta \in (0, \frac{1}{2})$ and $\gamma \in (0, \frac{1}{2}]$ as follows.
The weight function is bounded on $[\eta, 1-\eta]$. Thus, with

$$w_\gamma\left(\frac{\lfloor sn \rfloor}{n} \right) L_0(s) - L_\gamma(s) = o(1)$$

uniformly in $s \in [\eta, (1-\eta)]$ the result follows from the case above.

It is left to prove the claim for $\eta = 0$ and $\gamma \in (0, \frac{1}{2})$.
Here, we have to be careful because the weight function is not well behaved on $(0,1)$. Let α be such that $0 < \alpha < \lambda < 1 - \alpha < 1$. In the truncated area $[\alpha, 1-\alpha]$ we know from the case above, that the claim holds true. It is left to prove the (uniformly in n) negligibility of the values outside this area, i.e.

$$\lim_{\alpha \to 0} \max_{\substack{0<s<\alpha, \\ (1-\alpha)<s<1}} \left| w_\gamma \left(\frac{\lfloor sn \rfloor}{n} \right) \left\| \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \hat{\theta}_n \right) \right\|_A - L_\gamma(s) \right| = o_P(1) \, .$$

Due to the special choice of the weight function (see Lemma 2.3.2), it follows

$$\lim_{\alpha \to 0} \sup_{s \in (0,\alpha)} w_\gamma(s) s \|(1-\lambda)\delta\|_A = 0$$

and

$$\lim_{\alpha \to 0} \sup_{s \in (1-\alpha,1)} w_\gamma(s)(1-s)\|\lambda\delta\|_A = 0 \, .$$

As also the quotient $w_\gamma(\lfloor sn \rfloor / n) \frac{\lfloor sn \rfloor}{n}$ is bounded, we get the uniformly convergence

$$\lim_{\alpha \to 0} \sup_{s \in (0,\alpha)} w_\gamma(\lfloor sn \rfloor / n) |g_n(s)| \|\delta\|_A = 0$$

as above. On the other hand we can argue as in Lemma 2.3.2 and derive

$$\lim_{\alpha \to 0} \sup_{s \in (1-\alpha,1)} w_\gamma(\lfloor sn \rfloor / n) |g_n(s)| \|\delta\|_A = 0 \, .$$

By (2.23) and (2.24) we have uniformly in n

$$\lim_{\alpha \to 0} \sup_{\substack{0<s<\alpha, \\ (1-\alpha)<s<1}} w_\gamma(\lfloor sn \rfloor / n) \left\| \sum_{t=1}^{\lfloor sn \rfloor} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A = o_P(1) \, .$$

Thus, it is allowed to ex-change limits for $\alpha \to 0$ and $n \to \infty$. This finishes the proof.

■

Corollary 2.3.2

Let $\hat{m}(\eta, \gamma; A)$ and $\hat{m}(\alpha, \gamma; A)$ denote two change-point estimators of the form

$$\hat{m}(\beta, \gamma; A) = \arg\max_{1 \le k < n} w_{\beta,\gamma}(k/n) \left\| \sum_{t=1}^{k} H(X_t, \hat{\theta}_n) \right\|_A$$

with η, α, γ, A and function H such that

$$\hat{m}(\beta, \gamma; A) - m = o_P(n)$$

for $\beta = \eta$ or $\beta = \alpha$ and $m = \lfloor \lambda n \rfloor$. Then it holds

$$P(\hat{m}(\eta, \gamma; A) = \hat{m}(\alpha, \gamma; A)) \underset{n \to \infty}{\longrightarrow} 1 \, .$$

Proof:
Without loss of generality let us assume $\eta < \alpha < \frac{1}{2}$. For $\eta = \alpha$ the result is clear. Denote $S_k(\hat{\theta}_n) = \sum_{t=1}^{k} H(X_t, \hat{\theta}_n)$. Then we get

$$\begin{aligned}
&P(|\hat{m}(\eta,\gamma;A) - \hat{m}(\alpha,\gamma;A)| > 0)\\
&\quad = P\Big(\max_{\substack{\eta n \le k < \alpha n \\ (1-\alpha)n \le k < (1-\eta)n}} w_\gamma(k/n) \Big\| S_k(\hat{\theta}_n) \Big\|_A > \max_{\alpha n \le k < (1-\alpha)n} w_\gamma(k/n) \Big\| S_k(\hat{\theta}_n) \Big\|_A \Big)\\
&\quad = P(\{\tfrac{\hat{m}(\eta,\gamma;A)}{n} < \alpha\} \cup \{\tfrac{\hat{m}(\eta,\gamma;A)}{n} > 1-\alpha\})\\
&\quad = P(\tfrac{\hat{m}(\eta,\gamma;A)}{n} - \lambda < \alpha - \lambda) + P(\tfrac{\hat{m}(\eta,\gamma;A)}{n} - \lambda > 1 - \alpha - \lambda))\,. \qquad (2.38)
\end{aligned}$$

From Theorem 2.3.1 we have $\frac{\hat{m}(\eta,\gamma;A)}{n} - \lambda = o_P(1)$, so (2.38) converges to 0 as $\alpha < \lambda < 1-\alpha$.
■

Before we prove the optimal rate of convergence, we show some useful properties of the partial sums.

Lemma 2.3.3

Let Y_t, $1 \le t \le n$, be i.i.d. random vectors with finite second moments and A symmetric, positive semi-definite matrix. It follows

$$\max_{1\le k\le n} \left\| \sum_{t=1}^{k} (Y_t - \overline{Y}_n) \right\|_A \le 2 \max_{1 \le k \le n} \left\| \sum_{t=1}^{k} Y_t \right\|_A = O_P\left(n^{\frac{1}{2}}\right) \qquad (2.39)$$

and for $\kappa_n > 0$

$$\begin{aligned}
\max_{1\le k\le m-\kappa_n} \left\| \frac{1}{m-k} \sum_{t=k+1}^{m} (Y_t - \overline{Y}_n) \right\|_A &\le \max_{1\le k\le m-\kappa_n} \left\| \frac{\sum_{t=k+1}^{m} Y_t}{m-k} \right\|_A + o_P(1)\\
&= O_P\left(\kappa_n^{-\frac{1}{2}}\right) + o_P(1)\,, \qquad (2.40)\\
\max_{m+\kappa_n\le k< n} \left\| \frac{1}{k-m} \sum_{t=m+1}^{k} (Y_t - \overline{Y}_n) \right\|_A &\le \max_{m+\kappa_n\le k<n} \left\| \frac{\sum_{t=m+1}^{k} Y_t}{k-m} \right\|_A + o_P(1)\\
&= O_P\left(\kappa_n^{-\frac{1}{2}}\right) + o_P(1)\,. \qquad (2.41)
\end{aligned}$$

Proof:
Observe, that w.l.o.g. we can assume $\mathbb{E}[Y_1] = 0$, otherwise define $Y_i^\star = Y_i - \mathbb{E}[Y_i]$ and show the behaviour for $Y_i^\star$.

From the Hájek-Rényi inequality we get for an i.i.d. sequence Y_t and for sequences a_n, b_n and non-increasing sequence c_k that the partial sums $\sum_{t=1}^k Y_t$ follow

$$\max_{a_n \le k \le b_n} \left\| c_k \sum_{t=1}^{k} Y_t \right\|_A = O_P\left((a_n c_{a_n}^2 + \sum_{k=a_n+1}^{b_n} c_k^2)^{\frac{1}{2}} \right) .$$

The equation in (2.39) can be proved using the Hájek-Rényi inequality with $c_k = 1$. For the inequality in (2.39) we use the triangle inequality.

For the inequalities in (2.40) and (2.41), we use the triangle inequality. The CLT (see Theorem C.2.1) gives

$$\left\| \overline{Y}_n \right\|_A = O_P\left(n^{-\frac{1}{2}} \right) .$$

To prove the other relations we apply again the Hájek-Rényi inequality. The sequence $\{Y_{m-k+1}\}$ is a sequence of i.i.d. random vectors, so the Hájek-Rényi inequality can be applied to partial sums of them. For (2.40) we get

$$\begin{aligned}
\max_{1 \le k \le m-\kappa_n} \left\| \frac{\sum_{t=k+1}^m Y_t}{m-k} \right\|_A = \max_{\kappa_n \le l \le m-1} \left\| \frac{1}{l} \sum_{k=1}^{l} Y_{m-k+1} \right\|_A
&= O_p\left(\left(\frac{1}{\kappa_n} + \sum_{k=\kappa_n+1}^{m-1} \frac{1}{k^2} \right)^{\frac{1}{2}} \right) \\
&= O_p\left(\left(\frac{1}{\kappa_n} + \int_{\kappa_n}^{m} \frac{1}{x^2} dx \right)^{\frac{1}{2}} \right) \\
&= O_p\left(\left(\frac{1}{\kappa_n} + \int_{\kappa_n}^{\infty} \frac{1}{x^2} dx \right)^{\frac{1}{2}} \right) \\
&= O_p\left(\kappa_n^{-\frac{1}{2}} \right) .
\end{aligned}$$

The equation in (2.41) follows in the same way

$$\begin{aligned}
\max_{m+\kappa_n \le k < n} \left\| \frac{\sum_{t=m+1}^k Y_t}{k-m} \right\|_A &= \max_{\kappa_n \le l \le n-m-1} \left\| \frac{1}{l} \sum_{t=1}^{l} Y_{t+m} \right\|_A = O_p\left(\left(\frac{1}{\kappa_n} + \int_{\kappa_n}^{\infty} \frac{1}{x^2} dx \right)^{\frac{1}{2}} \right) \\
&= O_p\left(\kappa_n^{-\frac{1}{2}} \right) .
\end{aligned}$$

■

Theorem 2.3.2

Under the assumptions of Theorem 2.3.1 the change-point estimator $\hat{m}(\eta,\gamma)$ (2.25) fulfils

$$\hat{m}(\eta,\gamma;A) - m = O_P(1)\,.$$

Proof:
From Corollary 2.3.2 we know, that it is enough to analyse the behaviour in a truncated area. For estimators $\hat{m}(\eta,\gamma;A) = \hat{m}(0,\gamma;A)$, it is enough to analyse $\hat{m}(\alpha,\gamma;A)$ with $\alpha \in \left(0,\frac{1}{2}\right)$ fixed but with $\alpha < \lambda < 1-\alpha$. In the other cases, i.e. $\eta \in (0,\lambda)$, let α be equal to η. So we have to prove

$$\hat{m}(\alpha,\gamma;A) - m = O_P(1)$$

for $\alpha \in (0,\frac{1}{2})$. This means we have to show that for every $\epsilon > 0$ exists a $\kappa > 0$ such that for all n

$$P(|\hat{m}(\alpha,\gamma;A) - m| > \kappa) = P(\hat{m}(\alpha,\gamma;A) < m-\kappa) + P(\hat{m}(\alpha,\gamma;A) > m+\kappa) \le \epsilon\,. \quad (2.42)$$

We will consider the two parts $\hat{m}(\alpha,\gamma;A) < m-\kappa$ and $\hat{m}(\alpha,\gamma;A) > m+\kappa$ separately.

Before we come to the proofs for each part, we observe

$$\hat{m}(\alpha,\gamma;A) = \underset{\alpha n \le k \le (1-\alpha)n}{\arg\max}\; w_{0,\gamma}(k/n)\left\|\sum_{t=1}^{k}(X_t - \hat{\theta}_n)\right\|_A = \underset{\alpha n \le k \le (1-\alpha)n}{\arg\max}\; V_k\,, \quad (2.43)$$

with

$$\begin{aligned}
V_k &= \left\|w_\gamma(k/n)S_k(\hat{\theta}_n) - w_\gamma(\hat{m}/n)S_k(\hat{\theta}_n)\right\|_A \\
&= \left(w_\gamma(k/n)S_k(\hat{\theta}_n)\right)^{\mathsf{T}} A\left(w_\gamma(k/n)S_k(\hat{\theta}_n)\right) - \left(w_\gamma(m/n)S_m(\hat{\theta}_n)\right)^{\mathsf{T}} A\left(w_\gamma(k/n)S_k(\hat{\theta}_n)\right) \\
&\quad + \left(w_\gamma(k/n)S_k(\hat{\theta}_n)\right)^{\mathsf{T}} A\left(w_\gamma(m/n)S_m(\hat{\theta}_n)\right) - \left(w_\gamma(m/n)S_m(\hat{\theta}_n)\right)^{\mathsf{T}} A\left(w_\gamma(m/n)S_m(\hat{\theta}_n)\right) \\
&= -\left(w_\gamma(m/n)S_m(\hat{\theta}_n) - w_\gamma(k/n)S_k(\hat{\theta}_n)\right)^{\mathsf{T}} A\left(w_\gamma(m/n)S_m(\hat{\theta}_n) + w_\gamma(k/n)S_k(\hat{\theta}_n)\right) \\
&= -\left\langle w_\gamma(m/n)S_m(\hat{\theta}_n) - w_\gamma(k/n)S_k(\hat{\theta}_n),\, w_\gamma(m/n)S_m(\hat{\theta}_n) + w_\gamma(k/n)S_k(\hat{\theta}_n)\right\rangle_A
\end{aligned}$$

and $S_k(\theta) = \sum_{t=1}^{k}(X_t - \theta)$, see **N.1**. Recall, $w_{0,\gamma}(k/n) = w_\gamma(k/n)$ by notation **N.7** and $\langle\cdot,\cdot\rangle_A$ as in notation **N.2**.

First, let us consider $0 < \hat{m} < m - \kappa$.
With $S_k(\hat{\theta}_n) = \sum_{t=1}^{k} \varepsilon_t - k\overline{\varepsilon}_n - k(1 - \frac{m}{n})\delta$ for $k < m$ we get

$$\begin{aligned}
V_k = &- \Big\langle (w_\gamma(m/n) - w_\gamma(k/n)) \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) + w_\gamma(m/n) \sum_{t=k+1}^{m} \varepsilon_t \\
&\qquad - w_\gamma(m/n)(m-k)\overline{\varepsilon}_n - (w_\gamma(m/n)m - w_\gamma(k/n)k)(1 - \frac{m}{n})\delta\,, \\
&\quad w_\gamma(m/n) \sum_{t=k+1}^{m} (\varepsilon_t - \overline{\varepsilon}_n) + (w_\gamma(m/n) + w_\gamma(k/n)) \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \\
&\qquad - (w_\gamma(m/n)m + w_\gamma(k/n)k)(1 - \frac{m}{n})\delta \Big\rangle_A . \\
= &- \Bigg\langle (w_\gamma(m/n) - w_\gamma(k/n)) \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n), (w_\gamma(k/n) + w_\gamma(m/n)) \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \\
&\qquad + w_\gamma(m/n) \sum_{t=k+1}^{m} (\varepsilon_t - \overline{\varepsilon}_n) - (w_\gamma(k/n)k + w_\gamma(m/n)m)(1 - \frac{m}{n})\delta \Bigg\rangle_A \\
&- \Bigg\langle w_\gamma(m/n) \sum_{t=k+1}^{m} (\varepsilon_t - \overline{\varepsilon}_n), (w_\gamma(k/n) + w_\gamma(m/n)) \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \\
&\qquad + w_\gamma(m/n) \sum_{t=k+1}^{m} (\varepsilon_t - \overline{\varepsilon}_n) \Bigg\rangle_A \\
&+ w_\gamma(m/n)(w_\gamma(k/n)k + w_\gamma(m/n)m) \Bigg\langle \sum_{t=k+1}^{m} \varepsilon_t, (1 - \frac{m}{n})\delta \Bigg\rangle_A \\
&- w_\gamma(m/n)(w_\gamma(k/n)k + w_\gamma(m/n)m) \Bigg\langle \overline{\varepsilon}_n, (1 - \frac{m}{n})\delta \Bigg\rangle_A \\
&+ (w_\gamma(m/n)m - w_\gamma(k/n)k) \Bigg\langle (1 - \frac{m}{n})\delta, (w_\gamma(k/n) + w_\gamma(m/n)) \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \\
&\qquad + w_\gamma(m/n) \sum_{t=k+1}^{m} \varepsilon_t - w_\gamma(m/n)(m-k)\overline{\varepsilon}_n - (w_\gamma(k/n)k + w_\gamma(m/n)m)(1 - \frac{m}{n})\delta \Bigg\rangle_A .
\end{aligned}$$

Straight forward calculation gives

$$\begin{aligned}
V_k = &(w_\gamma(k/n) - w_\gamma(m/n))(w_\gamma(k/n) + w_\gamma(m/n)) \left\| \sum_{t=1}^{k} \varepsilon_t - k\overline{\varepsilon}_n \right\|_A^2 \\
&- w_\gamma^2(m/n) \Bigg\langle \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) + \sum_{t=1}^{m} (\varepsilon_t - \overline{\varepsilon}_n), \sum_{t=k+1}^{m} (\varepsilon_t - \overline{\varepsilon}_n) \Bigg\rangle_A \\
&- 2(w_\gamma^2(m/n)m - w_\gamma^2(k/n)k) \Bigg\langle \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n), (1 - \frac{m}{n})\delta \Bigg\rangle_A
\end{aligned}$$

$$
\begin{aligned}
&- 2w_\gamma^2(m/n)m\left\langle (m-k)\overline{\varepsilon}_n, (1-\frac{m}{n})\delta \right\rangle_A \\
&+ 2w_\gamma^2(m/n)m\left\langle \sum_{t=k+1}^{m} \varepsilon_t, (1-\frac{m}{n})\delta \right\rangle_A \\
&- (w_\gamma^2(m/n)m^2 - w_\gamma^2(k/n)k^2)\left\|(1-\frac{m}{n})\delta\right\|_A^2 \\
=&B_1 + B_2 + B_3 + B_4 + B_5 + B_6\,. \qquad (2.44)
\end{aligned}
$$

We determine the parts of V_k which can not be dominated by the deterministic part B_6. First observe that from the properties of the weight function (equation (2.30) in Corollary 2.3.1), we have

$$
\begin{aligned}
\max_{\alpha n<k\leq m-\kappa} \frac{m-k}{|B_6|} &= \max_{\alpha n<k\leq m-\kappa} \frac{m-k}{|(w_\gamma^2(m/n)m^2 - w_\gamma^2(k/n)k^2)\left\|(1-\frac{m}{n})\delta\right\|_A^2|} \\
&= O\left(n^{-1}\left\|(1-\frac{m}{n})\delta\right\|_A^{-2}\right)\,. \qquad (2.45)
\end{aligned}
$$

Now, we analyse each part of V_k. For the first part observe

$$
\begin{aligned}
\max_{\alpha n<k\leq m-\kappa} \left|\frac{B_1}{m-k}\right| &\leq \max_{\alpha n<k\leq m-\kappa} \frac{|w_\gamma^2(k/n) - w_\gamma^2(m/n)|}{m-k} \max_{\alpha n<k\leq m-\kappa} \left\|\sum_{t=1}^{k} \varepsilon_t - k\overline{\varepsilon}_n\right\|_A^2 \\
&\leq O(n^{-1})\left(\max_{\alpha n<k\leq m-\kappa}\left\|\sum_{t=1}^{k}\varepsilon_t\right\|_A + \max_{\alpha n<k\leq m-\kappa} k^2\|\overline{\varepsilon}_n\|_A\right)^2,
\end{aligned}
$$

with the triangle inequality and equation (2.28). In the equation (2.28) the constant is $O(1)$ because $\kappa/n \underset{n\to\infty}{\longrightarrow} 0$. Using $(a+b)^2 \leq a^2+b^2$ for $a,b\in\mathbb{R}$ and Lemma 2.3.3 (with $\kappa_n \equiv \kappa > 0$ fixed but arbitrary), we get

$$
\begin{aligned}
\max_{\alpha n<k\leq m-\kappa} \left|\frac{B_1}{m-k}\right| &\leq O(n^{-1})\left(\max_{\alpha n<k\leq m-\kappa}\left\|\sum_{t=1}^{k}\varepsilon_t\right\|_A^2 + \max_{\alpha n<k\leq m-\kappa} k^2\|\overline{\varepsilon}_n\|_A^2\right) \\
&= O(n^{-1})O_P(n) = O_P(1)\,. \qquad (2.46)
\end{aligned}
$$

For the analysis of B_2 we use the Cauchy-Schwarz inequality to conclude

$$
\begin{aligned}
\max_{\alpha n<k\leq m-\kappa} \left|\frac{B_2}{m-k}\right| \leq& w_\gamma^2(m/n) \max_{\alpha n<k\leq m-\kappa}\left\|\sum_{t=1}^{k}(\varepsilon_t - \overline{\varepsilon}_n) + \sum_{t=1}^{m}(\varepsilon_t - \overline{\varepsilon}_n)\right\|_A \\
&\cdot \max_{\alpha n<k\leq m-\kappa}\left\|\frac{1}{m-k}\sum_{t=k+1}^{m}(\varepsilon_t - \overline{\varepsilon}_n)\right\|_A\,.
\end{aligned}
$$

Recall that $m = \lfloor \lambda n \rfloor$ (**G.1**), so the weight function at m/n is bounded. Applying the triangle inequality to $\sum_{t=1}^{k} \varepsilon_t - k\bar{\varepsilon}_n + \sum_{t=1}^{m} \varepsilon_t - m\bar{\varepsilon}_n$ two-times and using Lemma 2.3.3, we get

$$\begin{aligned}
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_2}{m-k} \right| &\le O(1) \left(2 \max_{\alpha n < k \le m} \left\| \sum_{t=1}^{k} \varepsilon_t \right\|_A + 2m \|\bar{\varepsilon}_n\|_A \right) \\
&\quad \cdot \left(\max_{\alpha n < k \le m-\kappa} \left\| \frac{1}{m-k} \sum_{t=k+1}^{m} \varepsilon_t \right\|_A + \|\bar{\varepsilon}_n\|_A \right) \\
&= \left(O_P(n^{\frac{1}{2}}) + O(n) O_P(n^{-\frac{1}{2}}) \right) \left(O_P(\kappa^{-\frac{1}{2}}) + O_P(n^{-\frac{1}{2}}) \right) \\
&= O_P\left(n^{\frac{1}{2}} \kappa^{-\frac{1}{2}} \right) + O_P(1) . \qquad (2.47)
\end{aligned}$$

For B_3 and B_4 we use the Cauchy-Schwarz inequality together with Lemma 2.3.3 (again $\kappa_n \equiv \kappa > 0$) and (2.29). Thus, it follows

$$\begin{aligned}
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_3}{m-k} \right| &\le O(1) \max_{\alpha n < k \le m-\kappa} \left\| \sum_{t=1}^{k} \varepsilon_t - k\bar{\varepsilon}_n \right\|_A \left\| (1 - \frac{m}{n}\delta) \right\|_A \\
&= O_P\left(n^{\frac{1}{2}} \left\| (1 - \frac{m}{n}\delta) \right\|_A \right) , \qquad (2.48)
\end{aligned}$$

$$\begin{aligned}
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_4}{m-k} \right| &\le O(n) \|\bar{\varepsilon}_n\|_A \left\| (1 - \frac{m}{n}\delta) \right\|_A \\
&= O_P\left(n^{\frac{1}{2}} \left\| (1 - \frac{m}{n}\delta) \right\|_A \right) . \qquad (2.49)
\end{aligned}$$

To approximate the remaining term we again use Lemma 2.3.3 equation (2.40) and get

$$\begin{aligned}
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_5}{m-k} \right| &\le O(n) \max_{\alpha n \le k \le m-\kappa} \left\| \frac{1}{m-\kappa} \sum_{t=k+1}^{m} \varepsilon_t \right\|_A \left\| (1 - \frac{m}{n})\delta \right\|_A \\
&= O_P\left(n\kappa^{-\frac{1}{2}} \left\| (1 - \frac{m}{n}\delta) \right\|_A \right) . \qquad (2.50)
\end{aligned}$$

Thus, we can conclude

$$\begin{aligned}
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_1}{B_6} \right| &= O\left(n^{-1} \left\| (1 - \frac{m}{n})\delta \right\|_A^{-2} \right) O_P(1) = o_P(1) , \\
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_2}{B_6} \right| &= O\left(n^{-1} \left\| (1 - \frac{m}{n})\delta \right\|_A^{-2} \right) \left(O_P(n^{\frac{1}{2}} \kappa^{-\frac{1}{2}}) + O_P(1) \right) = o_P(1) , \\
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_3}{B_6} \right| &= O\left(n^{-1} \left\| (1 - \frac{m}{n})\delta \right\|_A^{-2} \right) O_P\left(n^{\frac{1}{2}} \left\| (1 - \frac{m}{n}\delta) \right\|_A \right) = o_P(1) , \\
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_4}{B_6} \right| &= O\left(n^{-1} \left\| (1 - \frac{m}{n})\delta \right\|_A^{-2} \right) O_P\left(n^{\frac{1}{2}} \left\| (1 - \frac{m}{n})\delta \right\|_A \right) = o_P(1) , \\
\max_{\alpha n < k \le m-\kappa} \left| \frac{B_5}{B_6} \right| &= O\left(n^{-1} \left\| (1 - \frac{m}{n})\delta \right\|_A^{-2} \right) O_P\left(n\kappa^{-\frac{1}{2}} \left\| (1 - \frac{m}{n}\delta) \right\|_A \right) = O_P(\kappa^{-\frac{1}{2}}) .
\end{aligned}$$

This leads to

$$\begin{aligned}
P(\hat{m}(\alpha,\gamma;A) < m-\kappa) &= P\left(\max_{\alpha n<k<m-\kappa} V_k > \max_{m-\kappa\le k\le(1-\alpha)n} V_k\right)\\
&\le P\left(\max_{\alpha n<k<m-\kappa} V_k \ge V_m\right)\\
&= P\left(\max_{\alpha n<k<m-\kappa} V_k \ge 0\right)\\
&= P\left(\max_{\alpha n<k<m-\kappa} \left\{B_6\left(1+o_P(1)+\frac{B_5}{B_6}\right)\right\} \ge 0\right). \qquad (2.51)
\end{aligned}$$

From Corollary 2.3.1 we have $B_6 < 0$ for $k < m$, with $\max_{1\le k\le m-\kappa} B_6 \le c < 0$ for some c. Thus, we conclude

$$\begin{aligned}
P(\hat{m}(\alpha,\gamma;A) < m-\kappa) &\le P\left(\max_{\alpha n<k<m-\kappa}\left|\frac{B_5}{B_6}\right| \ge 1+o_P(1)\right)\\
&\le P\left(\max_{\alpha n<k<m-\kappa}\left|\frac{B_5}{B_6}\right| \ge 1-\tau\right) + P\left(1+o_P(1) \le \max_{\alpha n<k<m-\kappa}\left|\frac{B_5}{B_6}\right| \le 1-\tau\right)\\
&\le P\Big(O_P(1) \ge (1-\tau)\kappa^{\frac{1}{2}}\Big) + o(1)
\end{aligned}$$

with $0 < \tau < 1$ arbitrary. This term becomes arbitrarily small for a sufficiently large $\kappa > 0$.

The proof for other case, $m+\kappa < \hat{m} < n$, is similar to the case before. However, in the literature this case is hardly dealt with. We take the opportunity to look at it closely.
As in (2.44) we determine a decomposition of V_k. By using $S_m(\hat{\theta}_n) = S_k(\hat{\theta}_n) - S_{m+1,k}(\hat{\theta}_n)$ we get

$$\begin{aligned}
V_k &= -\Big\langle w_\gamma(m/n)S_m(\hat{\theta}_n) - w_\gamma(k/n)S_k(\hat{\theta}_n),\; w_\gamma(m/n)S_m(\hat{\theta}_n) + w_\gamma(k/n)S_k(\hat{\theta}_n)\Big\rangle_A\\
&= -\Big\langle (w_\gamma(m/n) - w_\gamma(k/n))S_k(\hat{\theta}_n) - w_\gamma(m/n)S_{m+1,k}(\hat{\theta}_n),\; (w_\gamma(m/n) + w_\gamma(k/n))S_k(\hat{\theta}_n)\\
&\qquad - w_\gamma(m/n)S_{m+1,k}(\hat{\theta}_n)\Big\rangle_A .
\end{aligned}$$

For the specific model we have for $k > m$

$$\begin{aligned}
S_m(\hat{\theta}_n) &= \sum_{t=1}^{m}\varepsilon_t - m\overline{\varepsilon}_n - (n-m)\frac{m}{n}\delta\,,\\
S_{m+1,k}(\hat{\theta}_n) &= \sum_{t=m+1}^{k}\varepsilon_t + (k-m)\delta - (k-m)\overline{\varepsilon}_n - (k-m)(1-\frac{m}{n})\delta\\
&= \sum_{t=m+1}^{k}\varepsilon_t - (k-m)\overline{\varepsilon}_n + (k-m)\frac{m}{n}\delta\\
&= \sum_{t=m+1}^{k}\varepsilon_t - (k-m)\overline{\varepsilon}_n + (n-m-(n-k))\frac{m}{n}\delta
\end{aligned}$$

and for $k \geq m$

$$\begin{aligned} S_k(\hat{\theta}_n) &= S_m(\hat{\theta}_n) + S_{m+1,k}(\hat{\theta}_n) \\ &= \sum_{t=1}^{k} \varepsilon_t - k\overline{\varepsilon}_n - (n-m)\frac{m}{n}\delta + (k-m)\frac{m}{n}\delta \\ &= \sum_{t=1}^{k} \varepsilon_t - k\overline{\varepsilon}_n - (n-k)\frac{m}{n}\delta\,. \end{aligned}$$

Then we get

$$\begin{aligned} V_k = &- \Big\langle (w_\gamma(m/n) - w_\gamma(k/n)) \sum_{t=1}^{k}(\varepsilon_t - \overline{\varepsilon}_n) - w_\gamma(m/n) \sum_{t=m+1}^{k} \varepsilon_t + w_\gamma(m/n)(k-m)\overline{\varepsilon}_n \\ &\qquad - (w_\gamma(m/n)(n-m) - w_\gamma(k/n)(n-k))\frac{m}{n}\delta\,, \\ &\quad (w_\gamma(m/n) + w_\gamma(k/n)) \sum_{t=1}^{k}(\varepsilon_t - \overline{\varepsilon}_n) - w_\gamma(m/n) \sum_{t=m+1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \\ &\qquad - (w_\gamma(m/n)(n-m) + w_\gamma(k/n)(n-k))\frac{m}{n}\delta \Big\rangle_A \\ =&(w_\gamma(k/n) - w_\gamma(m/n))\Big\langle \sum_{t=1}^{k}(\varepsilon_t - \overline{\varepsilon}_n)\,,\, (w_\gamma(k/n) + w_\gamma(m/n)) \sum_{t=1}^{k}(\varepsilon_t - \overline{\varepsilon}_n) \\ &\qquad - w_\gamma(m/n) \sum_{t=m+1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) - (w_\gamma(k/n)(n-k) + w_\gamma(m/n)(n-m))\frac{m}{n}\delta \Big\rangle_A \\ &+ w_\gamma(m/n)\Big\langle \sum_{t=m+1}^{k} (\varepsilon_t - \overline{\varepsilon}_n)\,,\, (w_\gamma(k/n) + w_\gamma(m/n)) \sum_{t=1}^{k}(\varepsilon_t - \overline{\varepsilon}_n) \\ &\qquad - w_\gamma(m/n) \sum_{t=m+1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \Big\rangle_A \\ &- w_\gamma(m/n)(w_\gamma(k/n)(n-k) + w_\gamma(m/n)(n-m))\Big\langle \sum_{t=m+1}^{k} \varepsilon_t, \frac{m}{n}\delta \Big\rangle_A \\ &+ w_\gamma(m/n)(w_\gamma(k/n)(n-k) + w_\gamma(m/n)(n-m))\Big\langle (k-m)\overline{\varepsilon}_n, \frac{m}{n}\delta \Big\rangle_A \\ &+ (w_\gamma(m/n)(n-m) - w_\gamma(k/n)(n-k))\Big\langle \frac{m}{n}\delta, (w_\gamma(k/n) + w_\gamma(m/n)) \sum_{t=1}^{k}(\varepsilon_t - \overline{\varepsilon}_n) \\ &\qquad - w_\gamma(m/n) \sum_{t=m+1}^{k} \varepsilon_t + w_\gamma(m/n)(k-m)\overline{\varepsilon}_n - (w_\gamma(m/n)(n-m) + w_\gamma(k/n)(n-k))\frac{m}{n}\delta \Big\rangle_A\,. \end{aligned}$$

Straight forward calculation leads to

$$\begin{aligned}
V_k =&(w_\gamma^2(k/n) - w_\gamma^2(m/n))\left\|\sum_{t=1}^{k}\varepsilon_t - k\bar{\varepsilon}_n\right\|_A^2\\
&- w_\gamma^2(m/n)\Big\langle\sum_{t=1}^{k}(\varepsilon_t-\bar{\varepsilon}_n)+\sum_{t=1}^{m}(\varepsilon_t-\bar{\varepsilon}_n), \sum_{t=m+1}^{k}(\varepsilon_t-\bar{\varepsilon}_n)\Big\rangle_A\\
&- 2(w_\gamma^2(k/n)(n-k) - w_\gamma^2(m/n)(n-m))\Big\langle\sum_{t=1}^{k}(\varepsilon_t-\bar{\varepsilon}_n), \frac{m}{n}\delta\Big\rangle_A\\
&+ 2w_\gamma^2(m/n)m\Big\langle(k-m)\bar{\varepsilon}_n, (1-\frac{m}{n})\delta\Big\rangle_A\\
&- 2w_\gamma^2(m/n)m\Big\langle\sum_{t=m+1}^{k}\varepsilon_t, (1-\frac{m}{n})\delta\Big\rangle_A\\
&+ (w_\gamma^2(k/n)(n-k)^2 - w_\gamma^2(m/n)(n-m)^2)\left\|\frac{m}{n}\delta\right\|_A^2\\
=&\tilde{B}_1+\tilde{B}_2+\tilde{B}_3+\tilde{B}_4+\tilde{B}_5+\tilde{B}_6\,. \qquad (2.52)
\end{aligned}$$

We use the properties of the weight function (see Corollary 2.3.1) to prove the asymptotic behaviour. For the deterministic part (B_6) we use the left inequality of (2.30). With the symmetry of $w_\gamma(s)$ and $l = n-k$ we have

$$\begin{aligned}
\max_{m+\kappa\le k\le(1-\alpha)n}\frac{k-m}{|\tilde{B}_6|} &= \max_{\alpha n\le l\le n-m-\kappa}\frac{(n-m)-l}{w_\gamma^2(1-\frac{l}{n})l^2 - w_\gamma^2(m/n)(n-m)^2}\\
&= \max_{\alpha n\le l\le n-m-\kappa}\frac{(n-m)-l}{w_\gamma^2(\frac{l}{n})l^2 - w_\gamma^2(\frac{n-m}{n})(n-m)^2} = O\left(n^{-1}\left\|(1-\frac{m}{n})\delta\right\|_A^{-2}\right)\,.
\end{aligned}$$

Again we analyse the asymptotic behaviour of the other parts. With analogue arguments for $\tilde{B}_1$, $\tilde{B}_2$, $\tilde{B}_3$ and $\tilde{B}_4$ we get

$$\begin{aligned}
\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_1}{m-k}\right| &= \max_{m+\kappa\le k\le(1-\alpha)n}\frac{|w_\gamma^2(k/n)-w_\gamma^2(m/n)|}{k-m}\left\|\sum_{t=1}^{k}\varepsilon_t + k\bar{\varepsilon}_n\right\|_A^2\\
&= O(n^{-1})O_P(n) = O_P(1)\,.
\end{aligned}$$

Equivalently, we get with the Cauchy-Schwarz inequality and the triangle inequality

$$\begin{aligned}
\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_2}{m-k}\right| &\le O(1)\left(2\max_{\alpha n\le k\le m}\left\|\sum_{t=1}^{k}\varepsilon_t\right\|_A + 2m\|\bar{\varepsilon}_n\|_A\right)\\
&\quad\cdot\left(\max_{\alpha n\le k\le m-\kappa}\left\|\frac{1}{k-m}\sum_{t=m+1}^{k}\varepsilon_t\right\|_A + \|\bar{\varepsilon}_n\|_A\right)\\
&= \left(O_P(n^{\frac{1}{2}}) + O(n)O_P(n^{-\frac{1}{2}})\right)\left(O_P(\kappa^{-\frac{1}{2}}) + O_P(n^{-\frac{1}{2}})\right)\\
&= O_P\left(n^{\frac{1}{2}}\kappa^{-\frac{1}{2}}\right) + O_P(1)\,.
\end{aligned}$$

For the analysis of $\tilde{B}_3$ we use the Cauchy-Schwarz inequality and the symmetry of the weight function to conclude

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_3}{k-m}\right| \le \max_{\alpha n\le n-k\le n-m-\kappa}\left|\frac{w_\gamma^2(1-k/n)(n-k)-w_\gamma^2(1-m/n)(n-m)}{n-m-(n-k)}\right|$$
$$\max_{m+\kappa\le k\le(1-\alpha)n}\left(\left\|\sum_{t=1}^{k}(\varepsilon_t-\overline{\varepsilon}_n)\right\|_A\right)\left(\left\|\frac{m}{n}\delta\right\|_A\right).$$

Using property (2.29) of the weight function and Lemma 2.3.3 we get

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_3}{m-k}\right| \le O(1)O_P\left(n^{\frac{1}{2}}\left\|\frac{m}{n}\delta\right\|_A\right) = O_P\left(n^{\frac{1}{2}}\left\|(1-\frac{m}{n}\delta)\right\|_A\right).$$

For $\tilde{B}_4$ and $\tilde{B}_5$ we also use the Cauchy-Schwarz inequality and Lemma 2.3.3. Thus, it follows

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_4}{m-k}\right| \le O(n)\|\overline{\varepsilon}_n\|_A\left\|(1-\frac{m}{n}\delta)\right\|_A = O_P\left(n^{\frac{1}{2}}\left\|(1-\frac{m}{n}\delta)\right\|_A\right),$$

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_5}{k-m}\right| \le O(n)\max_{m+\kappa\le k\le(1-\alpha)n}\left\|\frac{1}{k-m}\sum_{t=m+1}^{k}\varepsilon_t\right\|_A\left\|(1-\frac{m}{n}\delta)\right\|_A$$
$$= O_P\left(n\kappa^{-\frac{1}{2}}\left\|(1-\frac{m}{n}\delta)\right\|_A\right).$$

Thus, we can conclude

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_1}{\tilde{B}_6}\right| = O\left(n^{-1}\left\|(1-\frac{m}{n})\delta\right\|_A^{-2}\right)O_P(1) = o_P(1),$$

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_2}{B_6}\right| = O\left(n^{-1}\left\|(1-\frac{m}{n})\delta\right\|_A^{-2}\right)(O_P(n^{\frac{1}{2}}\kappa^{-\frac{1}{2}})+O_P(1)) = o_P(1),$$

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_3}{\tilde{B}_6}\right| = O\left(n^{-1}\left\|(1-\frac{m}{n})\delta\right\|_A^{-2}\right)O_P\left(n^{-\frac{1}{2}}\left\|(1-\frac{m}{n}\delta)\right\|_A\right) = o_P(1),$$

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_4}{\tilde{B}_6}\right| = O\left(n^{-1}\left\|(1-\frac{m}{n})\delta\right\|_A^{-2}\right)O_P\left(n^{\frac{1}{2}}\left\|(1-\frac{m}{n}\delta)\right\|_A\right) = o_P(1),$$

$$\max_{m+\kappa\le k\le(1-\alpha)n}\left|\frac{\tilde{B}_5}{\tilde{B}_6}\right| = O\left(n^{-1}\left\|(1-\frac{m}{n})\delta\right\|_A^{-2}\right)O_P\left(n\kappa^{-\frac{1}{2}}\left\|(1-\frac{m}{n}\delta)\right\|_A\right)$$
$$= O_P\left(\kappa^{-\frac{1}{2}}\left\|(1-\frac{m}{n}\delta)\right\|_A^{-1}\right).$$

As in equation (2.51) we have

$$P(\hat{m}(\alpha,\gamma)>m+\kappa) = P\left(\max_{m+\kappa<k\le(1-\alpha)n}V_k > \max_{\alpha n<k\le m+\kappa}V_k\right)$$
$$\le P\left(\max_{m+\kappa<k\le(1-\alpha)n}V_k\ge 0\right)$$
$$= P\left(\max_{m+\kappa<k\le(1-\alpha)n}\left\{\tilde{B}_6\left(1+o_P(1)+\frac{\tilde{B}_5}{\tilde{B}_6}\right)\right\}\ge 0\right)$$

By Corollary 2.3.1 together with the symmetry of $w_\gamma(k/n)$, i.e. $w_\gamma(k/n) = w_\gamma((n-k)/n)$, we have $\tilde{B}_6 < 0$ for $m < k$, with $\max_{m+\kappa\leq k\leq(1-\alpha)n} \tilde{B}_6 \leq c < 0$ for c fixed but arbitrary. Equivalent to the case $k < m$ we have for $k > m$

$$\begin{aligned}
&P(\hat{m}(\alpha,\gamma;A) > m+\kappa) \\
&\qquad \leq P\left(\max_{m+\kappa<k\leq(1-\alpha)n} \left|\frac{\tilde{B}_5}{\tilde{B}_6}\right| \geq 1-\tau\right) + P\left(1+o_P(1) \leq \max_{m+\kappa<k\leq(1-\alpha)n} \left|\frac{\tilde{B}_5}{\tilde{B}_6}\right| \leq 1-\tau\right) \\
&\qquad \leq P\Big(O_P(1) \geq (1-\tau)\kappa^{\frac{1}{2}}\Big) + o(1)\,,
\end{aligned}$$

with $0 < \tau < 1$ arbitrary. This term becomes arbitrarily small for a sufficiently large $\kappa > 0$.

■

Lemma 2.3.4

For Y_t, $t = 1, \ldots, n$, being i.i.d. with finite second moment and $\kappa_n > 0$ a deterministic sequence. If $\frac{\kappa_n}{n} \underset{n\to\infty}{\longrightarrow} 0$ then we have

$$\max_{m-\kappa_n<k<m+\kappa_n} \left\|\sum_{t=1}^{k}(Y_t - \overline{Y}_n)\right\|_A = O_P(\sqrt{n})\,, \tag{2.53}$$

$$\max_{m-\kappa_n<k<m} \left\|\sum_{t=k+1}^{m}(Y_t - \overline{Y}_n)\right\|_A = O_P(\sqrt{\kappa_n})\,, \tag{2.54}$$

$$\max_{m<k<m+\kappa_n} \left\|\sum_{t=m}^{k+1}(Y_t - \overline{Y}_n)\right\|_A = O_P(\sqrt{\kappa_n})\,. \tag{2.55}$$

Proof:
First observe that w.l.o.g. we can set $A = I_d$ and $\mathbb{E}[Y_1] = 0$.

For the first equation (2.53) we observe that by the triangle inequality we have

$$\begin{aligned}
\max_{m-\kappa_n<k<m+\kappa_n} \left\|\sum_{t=1}^{k}(Y_t - \overline{Y}_n)\right\| &\leq \max_{m-\kappa_n<k<m+\kappa_n} \left\|\sum_{t=1}^{k} Y_t\right\| + \max_{m-\kappa_n<k<m+\kappa_n} k\left\|\overline{Y}_n\right\| \\
&\leq \max_{1<k<m+\kappa_n} \sum_{t=1}^{k} \|Y_t\| + O(n+\kappa_n)\left\|\overline{Y}_n\right\| \\
&= O_P\left(\sqrt{n}\left(1+\sqrt{\frac{\kappa_n}{n}}\right)\right) + O_P\left(\sqrt{n}\left(1+\frac{\kappa_n}{n}\right)\right)\,.
\end{aligned}$$

The claim follows by the CLT (see e.g. Theorem C.2.1).

With the same arguments we get

$$\begin{aligned}\max_{m-\kappa_n<k<m}\left\|\sum_{t=k+1}^{m}(Y_t-\overline{Y}_n)\right\| &= \max_{1\le l<\kappa_n}\left\|\sum_{t=1}^{l}Y_{m-t+1}-l\overline{Y}_n\right\| \\ &\le \max_{1\le l<\kappa_n}\left\|\sum_{t=1}^{l}Y_{m-t+1}\right\|+\kappa_n\left\|\overline{Y}_n\right\| \\ &= O_P(\sqrt{\kappa_n})+O_p\left(\frac{\kappa_n}{\sqrt{n}}\right).\end{aligned}$$

For the equation (2.55) equivalent argumentation finishes the proof. ■

Theorem 2.3.3

Let the assumptions of Theorem 2.3.1 hold true. Then

$$\hat{m}(\eta,\gamma;A)-m \xrightarrow{d} \arg\max\{W(s)-|s|g(s)\mathcal{D}^2,\ s\in\mathbb{Z}\}$$

with $\mathcal{D}=\|\delta\|_A$,

$$W(s)=\begin{cases}0 & ,\ s=0\,,\\ \delta^{\mathrm{T}}A\Sigma^{\frac{1}{2}}\ \sum_{t=s}^{-1}\xi_t^{(2)} & ,\ s<0\,,\\ \delta^{\mathrm{T}}A\Sigma^{\frac{1}{2}}\ \sum_{t=1}^{s}\xi_t^{(1)} & ,\ s>0\,,\end{cases}$$

$\xi_i^{(z)}$ are i.i.d. with $\mathcal{L}\left(\xi_1^{(z)}\right)=\mathcal{L}\left(\tilde{\varepsilon}_1\right)$ for $z=1,2$ and

$$g(s)=\begin{cases}(1-\gamma)(1-\lambda)+\gamma\lambda & ,\ s<0\,,\\ 0 & ,\ s=0\,,\\ \gamma(1-\lambda)+(1-\gamma)\lambda & ,\ s>0\,.\end{cases}$$

Proof:

To simplify the notation we use $\hat{m}:=\hat{m}(\eta,\gamma;A)$.

We show the claim by analysing the behaviour of the parts from the decomposition of V_k. Let (w.l.o.g.) $x>0$ and $\kappa>x$ be both fixed but arbitrary. We get

$$P(\hat{m}-m\le x)=P(m-\kappa\le\hat{m}\le m+x,|\hat{m}-m|<\kappa)+P(\hat{m}-m\le x,|\hat{m}-m|>\kappa)\,.$$

The second term on the right hand side is bounded by $P(|\hat{m}-m|>\kappa)$. From Theorem 2.3.2 we can conclude that for all n the second term of the right hand side becomes arbitrary small for large enough κ.

Hence, for fixed n and $\kappa\to\infty$ we have uniformly convergence and know from Theorem 2.3.2 that the limit exists. This implies that we can interchange the limits. For determining the asymptotic distribution we consider the convergence in n for fixed $\kappa>0$. Then we derive the asymptotic distribution by letting $\kappa\to\infty$. To determine the limit we use

$$\begin{aligned}&\lim_{n\to\infty}P(m-\kappa\le\hat{m}\le m+x,|\hat{m}-m|<\kappa)\\ &\quad=\lim_{n\to\infty}P(\max_{(k-m)\in(-\kappa,x]}V_k\ge\max_{(k-m)\in(x,\kappa)}V_k)\,,\end{aligned} \tag{2.56}$$

where V_k is as in the proof of Theorem 2.3.2.

Again, we consider the two different cases for $k \in (m-\kappa, m)$ and $k \in (m, m+\kappa)$ separately. We start with analysing the asymptotic behaviour for $k \in (m-\kappa, m)$. For this side, we have by equation (2.44) the following decomposition

$$V_k = B_1 + B_2 + B_3 + B_4 + B_5 + B_6$$

with

$$\begin{aligned}
B_1 &= (w_\gamma^2(k/n) - w_\gamma^2(m/n)) \left\| \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A^2 , \\
B_2 &= -w_\gamma^2(m/n) \Big\langle \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) + \sum_{t=1}^{m} (\varepsilon_t - \overline{\varepsilon}_n), \sum_{t=k+1}^{m} (\varepsilon_t - \overline{\varepsilon}_n) \Big\rangle_A , \\
B_3 &= -2(w_\gamma^2(m/n)m - w_\gamma^2(k/n)k) \Big\langle \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n), (1 - \frac{m}{n})\delta \Big\rangle_A , \\
B_4 &= 2w_\gamma^2(m/n)m \Big\langle (m-k)\overline{\varepsilon}_n, (1 - \frac{m}{n})\delta \Big\rangle_A , \\
B_5 &= -2w_\gamma^2(m/n)m \Big\langle \sum_{t=k+1}^{m} \varepsilon_t, (1 - \frac{m}{n})\delta \Big\rangle_A , \\
B_6 &= -(w_\gamma^2(m/n)m^2 - w_\gamma^2(k/n)k^2) \left\| (1 - \frac{m}{n})\delta \right\|_A^2 \\
&= -2(w_\gamma(m/n)m - w_\gamma(k/n)k) w_\gamma(m/n)m \|(1-\lambda)\delta\|_A^2 \\
&\quad + (w_\gamma(m/n)m - w_\gamma(k/n)k)^2 \|(1-\lambda)\delta\|_A^2 .
\end{aligned}$$

Recall, from (2.31) we have $B_6 < 0$.

We state the results in relation to a fixed but arbitrary constant κ. This will be useful later (see Theorem2.4.4). With Lemma 2.3.4 and the properties of the weight function (Corollary 2.3.1) we have

$$\begin{aligned}
\max_{k \in (m-\kappa, m)} |B_1| &\le \max_{m-\kappa \le k \le m} \left| w_\gamma^2(m/n) - w_\gamma^2(k/n) \right| \left(\max_{m-\kappa \le k \le m} \left\| \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A^2 \right) \\
&= O_P(c_2 \kappa n^{-1}) O_P(n) \\
&= O_P(\kappa) .
\end{aligned} \tag{2.57}$$

Again using the triangle inequality in addition to the Hájek-Rényi inequality and with Lemma 2.3.4 we get, recalling that $m = \lfloor \lambda n \rfloor$,

$$\begin{aligned}
\max_{k \in (m-\kappa, m)} |B_2| &\le w_\gamma^2(m/n) \left(\max_{k \in (m-\kappa, m)} \left\| \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A + \left\| \sum_{t=1}^{m} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A \right) \\
&\qquad \max_{k \in (m-\kappa, m)} \left\| \sum_{t=k+1}^{m} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A \\
&= O_P(\sqrt{n}\sqrt{\kappa}) .
\end{aligned}$$

Analogously we get

$$\begin{aligned}\max_{k\in(m-\kappa,m)}|B_3| &\leq 2\max_{k\in(m-\kappa,m)}|w_\gamma^2(m/n)m - w_\gamma^2(k/n)k| \\ &\qquad \max_{k\in(m-\kappa,m)}\left\|\sum_{t=1}^{k}(\varepsilon_t-\overline{\varepsilon}_n)\right\|_A \left\|(1-\frac{m}{n})\delta\right\|_A \\ &= O_P\left(\kappa\sqrt{n}\left\|(1-\frac{m}{n})\delta\right\|_A\right)\end{aligned}$$

and

$$\begin{aligned}\max_{k\in(m-\kappa,m)}|B_4| &\leq 2w_\gamma^2(m/n)m\max_{k\in(m-\kappa,m)}(m-k)\|\overline{\varepsilon}_n\|_A\left\|(1-\frac{m}{n})\delta\right\|_A \\ &= O_P\left(\kappa\sqrt{n}\left\|(1-\frac{m}{n})\delta\right\|_A\right).\end{aligned}$$

For the dominating parts we observe the following

$$\begin{aligned}\max_{k\in(m-\kappa,m)}|B_5| &= \max_{k\in(m-\kappa,m)}\left|-2w_\gamma^2(m/n)m(1-\frac{m}{n})\delta^{\mathrm{T}}A\sum_{t=k+1}^{m}\tilde{\varepsilon}_t\right| \\ &=O_P(n\left\|(1-\frac{m}{n})\delta\right\|_A)\end{aligned}$$

and

$$\begin{aligned}\max_{k\in(m-\kappa,m)}|B_6| &= \max_{k\in(m-\kappa,m)}\left|-(w_\gamma(m/n)m-w_\gamma(k/n)k)(w_\gamma(m/n)m+w_\gamma(k/n)k)\left\|(1-\frac{m}{n})\delta\right\|_A^2\right| \\ &= \max_{k\in(m-\kappa,m)}\Big|-2(w_\gamma(m/n)m-w_\gamma(k/n)k)w_\gamma(m/n)m \\ &\qquad + (w_\gamma(m/n)m-w_\gamma(k/n)k)^2\Big|\left\|(1-\frac{m}{n})\delta\right\|_A^2 \\ &= \max_{k\in(m-\kappa,m)}\Big|-2(w_\gamma(m/n)m-w_\gamma(k/n)k)w_\gamma(m/n)m+o_P(1)\Big|\left\|(1-\frac{m}{n})\delta\right\|_A^2 \\ &=O_P(n\left\|(1-\frac{m}{n})\delta\right\|_A^2). \end{aligned} \tag{2.58}$$

Moreover, we have due to the property (2.26) of the weight function, that

$$\begin{aligned}(w_\gamma(m/n)m&-w_\gamma((m+l)/n)(m+l)) \\ &= -l\left(w'_\gamma(\tilde{l}/n)\tilde{l}+w_\gamma(\tilde{l}/n)\right) \\ &= -lw_{\gamma+1}(\tilde{l}/n)\left(\frac{\tilde{l}}{n}\gamma(2\frac{\tilde{l}}{n}-1)+(\tilde{l}/n(1-\tilde{l}/n))\right),\end{aligned} \tag{2.59}$$

where $\tilde{l}\in[m+l,m]$ (because $l\in(-\kappa,0)$). It follows from $\frac{m+l}{n}\underset{n\to\infty}{\longrightarrow}\lambda$ for all $l\in(-\kappa,0)$ that

$$w_\gamma(m/n)m-w_\gamma((m+l)/n)(m+l)\underset{n\to\infty}{\longrightarrow}|l|w_{\gamma+1}(\lambda)\lambda(\gamma(2\lambda-1)+1-\lambda).$$

Let $\tilde{\varepsilon}_{m+i+1} = \Sigma^{-\frac{1}{2}}\varepsilon_{m+i+1}$ (see Notation **N.8**), then we have

$$\begin{aligned}
&\max_{l\in(-\kappa,0)} \frac{w_{-2\gamma}(m/n)}{m} V_{m+l} \\
&\quad = \max_{l\in(-\kappa,0)} \left(2(1-\lambda)\delta^{\mathrm{T}} A \Sigma^{\frac{1}{2}} \sum_{t=l}^{-1} \tilde{\varepsilon}_{m+t+1} - 2|l| \frac{\gamma\lambda + (1-\gamma)(1-\lambda)}{(1-\lambda)} \|(1-\lambda)\delta\|_A^2 + o_P(1) \right) .
\end{aligned}$$

Define $\xi_i^{(1)} \stackrel{d}{=} \tilde{\varepsilon}_1$ and

$$V_l^{(1)} = 2\delta^{\mathrm{T}} A \Sigma^{\frac{1}{2}} \sum_{t=l}^{-1} \xi_t^{(1)} - 2|l| \frac{(1-\gamma)(1-\lambda)+\gamma\lambda}{(1-\lambda)^2} \|(1-\lambda)\delta\|_A^2 .$$

Then we have asymptomatic convergence and get

$$\max_{l\in(-\kappa,0)} \frac{w_{-2\gamma}(m/n)}{m(1-\frac{m}{n})} V_{m+l} \stackrel{d}{\longrightarrow} \max_{l\in(-\kappa,0)} V_l^{(1)} .$$

Now we analyse the other side.

As before the case $m < k < m+\kappa$ is hardly considered in the literature, because it is analogue. For the completeness, we give the ideas. Analogously to the first case, we analyse this one using the decomposition we used in the proof of Theorem 2.3.2 (see equation (2.52)), i.e.

$$V_k = \tilde{B}_1 + \tilde{B}_2 + \tilde{B}_3 + \tilde{B}_4 + \tilde{B}_5 + \tilde{B}_6 .$$

With the same argumentations, i.e. triangle inequality and Cauchy-Schwarz inequality as well as the CLT together with the Hájek-Rényi inequality, we get

$$\begin{aligned}
\max_{k\in(m,m+\kappa)} \left|\tilde{B}_1\right| &= \max_{k\in(m,m+\kappa)} -\left(w_\gamma^2(k/n) - w_\gamma^2(m/n)\right) \left\| \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \right\|_A^2 \\
&= O_P(\kappa) , \\
\max_{k\in(m,m+\kappa)} \left|\tilde{B}_2\right| &= \max_{k\in(m,m+\kappa)} w_\gamma^2(m/n) \left\langle \sum_{t=1}^{k} (e_t - \overline{\varepsilon}_n) + \sum_{t=1}^{m} (\varepsilon_t - \overline{\varepsilon}_n), \sum_{t=m+1}^{k} (\varepsilon_t - \overline{\varepsilon}_n) \right\rangle_A \\
&= O_P(\sqrt{n})
\end{aligned}$$

and

$$\begin{aligned}
\max_{k\in(m,m+\kappa)} \left|\tilde{B}_3\right| &= \max_{k\in(m,m+\kappa)} 2\left(w_\gamma^2(k/n)(n-k) - w_\gamma^2(m/n)(n-m)\right) \left\langle \sum_{t=1}^{k} (\varepsilon_t - \overline{\varepsilon}_n), \frac{m}{n}\delta \right\rangle_A \\
&= O_P\left(\kappa\sqrt{n} \left\| \frac{m}{n}\delta \right\|_A\right) , \\
\max_{k\in(m,m+\kappa)} \left|\tilde{B}_4\right| &= \max_{k\in(m,m+\kappa)} 2w_\gamma^2(m/n)(n-m) \left\langle (k-m)\overline{\varepsilon}_n, \frac{m}{n}\delta \right\rangle_A \\
&= O_P\left(\kappa\sqrt{n} \left\| \frac{m}{n}\delta \right\|_A\right) .
\end{aligned}$$

For the dominating parts we observe, that

$$
\begin{aligned}
\max_{k\in(m,m+\kappa)} |\tilde{B}_5| &= 2w_\gamma^2(m/n)m \max_{k\in(m,m+\kappa)} \left\langle \sum_{t=m+1}^{k} \varepsilon_t, (1-\frac{m}{n})\delta \right\rangle_A \\
&= O_P\left(n\left\|(1-\frac{m}{n})\delta\right\|_A\right), \\
\max_{k\in(m,m+\kappa)} |\tilde{B}_6| &= \max_{k\in(m,m+\kappa)} \Big| 2(w_\gamma(m/n)(n-m) - w_\gamma(k/n)(n-k))w_\gamma(m/n)(n-m) \\
&\qquad - (w_\gamma(m/n)(n-m) - w_\gamma(k/n)(n-k))^2 \Big| \left\|\frac{m}{n}\delta\right\|_A^2 \\
&= \max_{k\in(m,m+\kappa)} \Big| 2(w_\gamma(m/n)(n-m) - w_\gamma(k/n)(n-k))w_\gamma(m/n)(n-m) \\
&\qquad - O(1) \Big| \left\|\frac{m}{n}\delta\right\|_A^2 \\
&= O_P\left(\kappa n \frac{n}{n-m}\frac{m^2}{n^2}\left\|(1-\frac{m}{n})\delta\right\|_A^2\right).
\end{aligned}
$$

Observe, that we have by the property (2.26) of the weight function

$$
\begin{aligned}
n(w_\gamma(m/n) - w_\gamma((m+l)/n)) &= -l\gamma w_{\gamma+1}(\tilde{l}/n)(2\frac{\tilde{l}}{n}-1) \qquad , \tilde{l}\in(m,m+l) \\
&\underset{n\to\infty}{\longrightarrow} -lw_{\gamma+1}(\lambda)\gamma(2\lambda-1).
\end{aligned}
$$

Together with (2.59), we have

$$
\begin{aligned}
(w_\gamma(m/n)(n-m) &- w_\gamma((m+l)/n)(n-(m+l))) \\
&= n(w_\gamma(m/n) - w_\gamma((m+l)/n)) - (w_\gamma(m/n)m - w_\gamma((m+l)/n)(m+l)) \\
&\underset{n\to\infty}{\longrightarrow} -lw_{\gamma+1}(\lambda)\left(\gamma(2\lambda-1) - \lambda\gamma(2\lambda-1) - \lambda(1-\lambda)\right).
\end{aligned}
$$

Observe, that $(\gamma(2\lambda-1) - \lambda\gamma(2\lambda-1) - \lambda(1-\lambda)) = (1-\lambda)\left(\gamma(\lambda-(1-\lambda)) - \lambda\right)$. So, we can conclude

$$
\begin{aligned}
\max_{l\in(0,\kappa)} \Big| & 2(w_\gamma(m/n)(n-m) - w_\gamma((m+l)/n)(n-(m+l)))\frac{m}{n-m}\left\|(1-\frac{m}{n})d\right\|_A^2 \\
&- 2lw_{2\gamma+1}(\lambda)\lambda(\gamma(1-\lambda) - \gamma\lambda + \lambda)\|(1-\lambda)d\|_A^2 \Big| \\
= o_P(1).
\end{aligned}
$$

With $\tilde{\varepsilon}_{i+m} = \Sigma^{-\frac{1}{2}}\varepsilon_{i+m}$ we have

$$
\begin{aligned}
\max_{l\in(0,\kappa)} \frac{w_{-2\gamma}(m/n)}{m} V_{m+l} \overset{d}{=} \max_{l\in(0,\kappa)} \Bigg(& 2(1-\lambda)d^T\Sigma^{\frac{1}{2}}\sum_{t=1}^{l}\xi_t^{(2)} \\
&- 2l\frac{\gamma(1-\lambda)+(1-\gamma)\lambda}{1-\lambda}\|(1-\lambda)d\|_A^2 + o_P(1)\Bigg).
\end{aligned}
$$

Let $V_l^{(2)}$ be defined as

$$
V_l^{(2)} = 2d^TA\Sigma^{\frac{1}{2}}\sum_{t=1}^{l}\xi_t^{(2)} - 2l\frac{\gamma(1-\lambda)+(1-\gamma)\lambda}{(1-\lambda)^2}\|(1-\lambda)d\|_A^2, \tag{2.60}
$$

with $\xi_t^{(2)}$ are i.i.d. with $\mathcal{L}\left(\xi_1^{(2)}\right) = \mathcal{L}(\tilde{\varepsilon}_1)$. Then it holds

$$\max_{l\in(0,\kappa)} \frac{w_{-2\gamma}(m/n)}{m(1-\frac{m}{n})} V_{m+l} \xrightarrow{d} \max_{l\in(0,\kappa)} V_l^{(2)} .$$

Define

$$W_l = \begin{cases} \frac{1}{2}V_l^{(1)} & ,l<0 \\ 0 & ,l=0 \\ \frac{1}{2}V_l^{(2)} & ,l>0 \end{cases} ,$$

then for large enough but fixed κ we have

$$\begin{aligned}
\lim_{n\to\infty} & P(\hat{m}-m \leq x, |\hat{m}-m| \leq \kappa) \\
&= \lim_{n\to\infty} P(m-\kappa \leq \hat{m} \leq m+x) \\
&= \lim_{n\to\infty} P\left(\max_{(k-m)\in(-\kappa,x]} V_k \geq \max_{(k-m)\in(x,\kappa)} V_k\right) \\
&= \lim_{n\to\infty} P\left(\frac{1}{w_{2\gamma}(m/n)m} \max_{(k-m)\in(-\kappa,x]} V_k \geq \frac{1}{w_{2\gamma}(m/n)m} \max_{(k-m)\in(x,\kappa)} V_k\right) \\
&= P\left(\max_{l\in(-\kappa,x]} W_l \geq \max_{l\in(x,\kappa)} W_l\right) .
\end{aligned}$$

So, letting $\kappa \to \infty$, we get the claim. ■

2.4. A-local alternative

In this section we are going to analyse, what happens if we allow the size of the change to vanish in the crucial entries, i.e. we have a A-local alternative. The corresponding model is given as

$$X_t = \begin{cases} \theta + \varepsilon_t & t \leq m, \\ \theta + \delta_n + \varepsilon_t & t > m, \end{cases} \tag{2.61}$$

where the errors $\{\varepsilon_t\}$ fulfil the **L.1**, the unknown time-point m fulfils assumption **G.1**. The size of the change, i.e. δ_n, is assumed to fulfil $\|\delta_n\|_A = \mathcal{D}_n \underset{n\to\infty}{\longrightarrow} 0$.

First we take a closer look on the test statistic and derive conditions on $\mathcal{D}_n$ such that we still have the asymptotic behaviour we derived before, i.e. a consistent level α test. In section 2.4.2 we are going to analyse the change-point estimator. Here, we are going to see that for the local alternative we derive a distribution free asymptotic distribution of the change-point estimator.

2.4.1. Change-point test

The difference between the local and non-local case occurs only under the alternative, i.e. the asymptotic behaviour of the test statistic under H_0 is still valid. We have to analyse the behaviour under H_1. To this end, we have to check under which additional conditions on δ_n

the test is still an asymptotic consistent level α-test.

If the difference between the two states is too small for a given finite sample size the test has a very low power. Increasing the sample size n will increase the power for all fixed values of the difference. Now, we allow the entries we are interested in (defined by A) of the difference δ to depend on the sample size n and decrease with n, i.e. $\|\delta_n\|_A \underset{n\to\infty}{\longrightarrow} 0$. It is clear, that this converges too fast, the increase of the sample size will not lead to a power one test. Therefore, we have to determine, how fast the change can become smaller, depending on n. We assume the A-local alternative, see assumption **G.5.b)**.

Theorem 2.4.1 *Assume assumptions **L.1,G.1, G.3** and **G.5.b)**. For model* (2.61) *and under* H_1 (1.2)

$$T_n(\eta,\gamma;A) \xrightarrow{p} \infty$$

if $\eta < \lambda < (1-\eta)$.

Proof:
To prove that the test has asymptotic power one, we recall from the proof of Theorem 2.2.2 that

$$T_n \geq \left| \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^{m} (\tilde{\varepsilon}_t - \bar{\tilde{\varepsilon}}_n) \right\| - O(\sqrt{n}\mathcal{D}_n) + o(1) \right| .$$

With $\mathcal{D}_n$ being such that $\mathcal{D}_n \underset{n\to\infty}{\longrightarrow} 0$ but $\sqrt{n}\mathcal{D}_n \underset{n\to\infty}{\longrightarrow} \infty$, we still have a power one test.

■

2.4.2. Change-point estimator

For the change-point estimator under a local alternative we have to do a little bit more. First we check if it is still a consistent estimator.

Theorem 2.4.2 *Under the assumptions of Theorem 2.4.1, we have*

$$\hat{m}(\eta,\gamma;A) - m = o_P\left(\mathcal{D}_n^{-2}\right) ,$$

i.e. $\frac{\hat{m}(\eta,\gamma;A)}{n}$ *is a consistent estimator for* λ.

Proof:
To prove the consistency in Theorem 2.3.1 we have used Lemma 3.1 of Pötscher and Prucha [1997] (Theorem C.1.2), which gives us that from

$$\sup_{s\in(\eta,1-\eta)} \left| w_n(s)^\gamma \frac{1}{n} \left\| \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \hat{\theta}_n \right) \right\|_A - L_\gamma(s) \right| = o_{a.s.}(1) \, ,$$

follows that the estimator $\hat{m}$ maximizing

$$w_n(s)^\gamma \frac{1}{n} \left\| \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \hat{\theta}_n \right) \right\|_A$$

converges to λ (which maximizes $\L_\gamma(s)$). Because the Lemma 3.1 of Pötscher and Prucha [1997] holds true if we use

$$L_\gamma^n(s) := \begin{cases} (s(1-s))^{-\gamma} \left\|-s(1-\lambda)\delta_n\right\|_A & , s < \lambda \\ (s(1-s))^{-\gamma} \left\|-\lambda(1-s)\delta_n\right\|_A & , s \geq \lambda \end{cases}$$

instead of $L_\gamma(s)$ which is also maximised by λ, the consistency follows directly.

■

In determining the convergence rate we have to go a little more into details. In the proofs we replace κ by κ_n and check under which assumptions on κ_n and δ_n the result holds true. It follows that all remains true if we assume $\sqrt{n}\mathcal{D}_n \underset{n\to\infty}{\longrightarrow} \infty$ (assumption **G.5.b)**)and choose $\kappa = \kappa_n = K/\mathcal{D}_n^2$ with $K > 0$.

Theorem 2.4.3 *Under assumptions of Theorem 2.4.1 the change-point estimator* $\hat{m}(\eta, \gamma; A)$ (2.25) *fulfils*

$$\hat{m}(\eta, \gamma; A) - m = O_P(\mathcal{D}_n^{-2})$$

with $\mathcal{D}_n = \|\delta_n\|_A$.

The main steps of the proof follows the same idea as for the fixed alternative, but we have to be careful at some parts.

Proof:
First we have to determine the dominating parts of

$$V_k = -\left\langle w_\gamma(m/n)S_m(\hat{\theta}_n) - w_\gamma(k/n)S_k(\hat{\theta}_n), w_\gamma(m/n)S_m(\hat{\theta}_n) + w_\gamma(k/n)S_k(\hat{\theta}_n)\right\rangle_A .$$

To this end, we use the same decomposition as in the proof of Theorem 2.3.2 (compare equation (2.44))

$$\begin{aligned}
V_k =&(w_\gamma(k/n) - w_\gamma(m/n))(w_\gamma(k/n) + w_\gamma(m/n))\left\|\sum_{t=1}^{k} \varepsilon_t - k\bar{\varepsilon}_n\right\|_A^2 \\
&- w_\gamma^2(m/n)\left\langle \sum_{t=1}^{k}(\varepsilon_t - \bar{\varepsilon}_n) + \sum_{t=1}^{m}(\varepsilon_t - \bar{\varepsilon}_n), \sum_{t=k+1}^{m}(\varepsilon_t - \bar{\varepsilon}_n)\right\rangle_A \\
&+ 2w_\gamma^2(m/n)m\left\langle (m-k)\bar{\varepsilon}_n, (1-\frac{m}{n})\delta_n\right\rangle_A \\
&- 2w_\gamma^2(m/n)m\left\langle \sum_{t=k+1}^{m} \varepsilon_t, (1-\frac{m}{n})\delta\right\rangle_A \\
&- 2(w_\gamma^2(m/n)m - w_\gamma^2(k/n)k)\left\langle \sum_{t=1}^{k}(\varepsilon_t - \bar{\varepsilon}_n), (1-\frac{m}{n})\delta_n\right\rangle_A \\
&+ (w_\gamma^2(m/n)m - w_\gamma^2(k/n)k)\left\|(1-\frac{m}{n})\delta_n\right\|_A^2 \\
=&B_1 + B_2 + B_3 + B_4 + B_5 + B_6 .
\end{aligned}$$

With the CLT and Hájek-Rényi inequality (see Theorem C.2.1 and C.2.2 for the strong mixing equivalents) we get the same estimations of the parts B_1 till B_6 but for the fractions we must be careful because δ now dependence on n. That is, why we replace κ by κ_n. Nevertheless, we still have the same bounding of the maxima of the weight function, as long as $\kappa_n/n \underset{n\to\infty}{\longrightarrow} 0$. With the assumption $\kappa_n = K/\mathcal{D}_n^2$ and $\sqrt{n}\mathcal{D}_n \underset{n\to\infty}{\longrightarrow} \infty$ this is fulfilled and we get

$$\begin{aligned}
\max_{\alpha n\le k\le m-\kappa_n}\left|\frac{B_1}{B_6}\right| =& O\left(n^{-1}\left\|(1-\frac{m}{n})\delta_n\right\|_A^{-2}\right) O_P(1) = o_P(1)\\
\max_{\alpha n\le k\le m-\kappa_n}\left|\frac{B_2}{B_6}\right| =& O\left(n^{-1}\left\|(1-\frac{m}{n})\delta_n\right\|_A^{-2}\right) O_P(n^{\frac{1}{2}}\kappa_n^{-\frac{1}{2}})\\
=& O_P\left((\sqrt{n\kappa_n}\left\|(1-\frac{m}{n})\delta_n\right\|_A^{2})^{-1}\right)\\
\max_{\alpha n< k\le m-\kappa_n}\left|\frac{B_3}{B_6}\right| =& O\left(n^{-1}\left\|(1-\frac{m}{n})\delta_n\right\|_A^{-2}\right) O_P\left(n^{\frac{1}{2}}\left\|(1-\frac{m}{n}\delta_n)\right\|_A\right) = o_P(1)\\
\max_{\alpha n< k\le m-\kappa_n}\left|\frac{B_4}{B_6}\right| =& O\left(n^{-1}\left\|(1-\frac{m}{n})\delta_n\right\|_A^{-2}\right) O_P\left(n^{\frac{1}{2}}\left\|(1-\frac{m}{n}\delta_n)\right\|_A\right) = o_P(1)\\
\max_{\alpha n< k\le m-\kappa_n}\left|\frac{B_5}{B_6}\right| =& O\left(n^{-1}\left\|(1-\frac{m}{n})\delta_n\right\|_A^{-2}\right) O_P(n\kappa_n^{-\frac{1}{2}}\left\|(1-\frac{m}{n})\delta_n\right\|_A)\\
=& O_P\left(\left(\sqrt{\kappa_n}\left\|(1-\frac{m}{n}\delta_n)\right\|_A\right)^{-1}\right)
\end{aligned}$$

As we observe, the fractions of B_2/B_6 and $B_5/B6$ depend on the choice of κ_n. With $\kappa_n = K/\mathcal{D}_n^2$ we get the same asymptotic behaviour for this fractions as in the fixed alternative part.

It follows that

$$\begin{aligned}
P(\tilde{m} < m-\kappa_n) =& P\left(\max_{\alpha n<k\le m-\kappa_n} V_k \ge 0\right)\\
\le& P\left(\max_{\alpha n<k\le m-\kappa_n}\left|\frac{B_5}{B_6}\right| \ge 1+o_P(1)\right)\\
\le& P\left(\max_{\alpha n<k\le m-\kappa_n}\left|\frac{B_5}{B_6}\right| \ge 1-\eta\right) + P\left(1+o_P(1)\le \max_{\alpha n<k\le m-\kappa_n}\left|\frac{B_5}{B_6}\right| \le 1-\eta\right)\\
\le& P\Big(O_P(1)\ge (1-\eta)K\Big) + o(1)\,.
\end{aligned}$$

This finishes the proof for the case $\alpha n \le k \le m-\kappa_n$. The same analysise can be done for $m+\kappa_n \le k \le (1-\alpha)n$, which finishes the proof.

■

From **G.5.b)**, we have that δ_n converges to 0 slower than $\frac{1}{\sqrt{n}}$. Hence, $\mathcal{D}_n^2$ is of smaller order than n. As $\hat{m}$, m both increase like n, $\hat{m}(\eta,\gamma;A)-m$ becomes relatively smaller (to n) with increasing sample size.

Theorem 2.4.4 *Let the assumptions of Theorem 2.4.1 be fulfilled. For the mean change model* (2.61) *it holds* $\frac{\|\delta_n\|_{A\Sigma A}}{\mathcal{D}_n} \underset{n\to\infty}{\longrightarrow} c$ *(some constant) and*

$$\mathcal{D}_n^2\Big(\hat{m}(\eta,\gamma;A)-m\Big) \xrightarrow{d} \arg\max\{cW(s)-|s|g(s),\ s\in\mathbb{R}\}$$

with $W(s)$ being a two sided standard Wiener process, $\mathcal{D}_n = \|\delta_n\|_A$ and

$$g(s) = \begin{cases} (1-\gamma)(1-\lambda)+\gamma\lambda & s<0 \\ 0 & s=0 \\ \gamma(1-\lambda)+(1-\gamma)\lambda & s>0 \end{cases}.$$

Proof:
We use the main ideas as in the proof of Theorem 2.4.3. So we start with the decomposition as in Theorem 2.4.3 and determine the limit of V_k for $k \in (m - C\mathcal{D}_n^{-2}, m)$ and for $k \in (m, m + C\mathcal{D}_n^{-2})$. Observe that the bounds for the non-dominating parts are still true if we replace C by $C\mathcal{D}_n^{-2}$.

Since $\mathcal{D}_n^2 n \underset{n\to\infty}{\longrightarrow} \infty$ the maxima of the weight function fulfil the same properties. For example, the constant c_2 from 2.57 is bounded since it is the maximum of the derivative calculated within $\frac{m-\kappa_n}{n}$ and m (from Corollary 2.3.1). The range the maximum is taken decreases in n and so the derivative. For the case $k \in (m - C\mathcal{D}_n^{-2}, m)$ we then have

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_1| = O_P\left(C\mathcal{D}_n^{-2}\right),$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_2| = O_P\left(\sqrt{n}\right),$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_3| = O_P\left(\sqrt{n}C\mathcal{D}_n^{-1}\right),$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_4| = O_P\left(C\sqrt{n}\mathcal{D}_n^{-1}\right) = o_P(1).$$

Equivalently we can handle $k \in (m, m + C\mathcal{D}_n^{-2})$. Thus, with the same weight of V_k we get

$$\frac{w_{-2\gamma}(m/n)}{2m(1-\frac{m}{n})} V_k$$

$$= \begin{cases} 2\frac{\delta_n^{\mathsf{T}} A \Sigma^{\frac{1}{2}}}{\mathcal{D}_n} \mathcal{D}_n \sum\limits_{t=k-m}^{-1} \xi_{nt}^{(1)} - 2|k-m| \frac{\gamma(2\lambda-1)+1-\lambda}{(1-\lambda)} \mathcal{D}_n^2 + o_P(1) & -C\mathcal{D}_n^{-2} \le k-m \le -1 \\ 0 & k-m=0 \\ 2\frac{\delta_n^{\mathsf{T}} A \Sigma^{\frac{1}{2}}}{\mathcal{D}_n} \mathcal{D}_n \sum\limits_{t=1}^{k-m} \xi_{nt}^{(2)} - 2(k-m)\lambda(\gamma(1-2\lambda)+\lambda)\mathcal{D}_n^2 + o_P(1) & 1 \le k-m \le C{\mathcal{D}_n}^{-2} \end{cases},$$

with $\mathcal{L}\left(\xi_t^{(1)}\right) = \mathcal{L}\left(\tilde{\varepsilon}_{-t}\right)$ and $\mathcal{L}\left(\xi_t^{(2)}\right) = \mathcal{L}\left(\tilde{\varepsilon}_t\right)$. By the functional central limit theorem (FCLT) for i.i.d. random vectors (the strong mixing equivalent is discussed in section C.2), we get

$$\left\{\mathcal{D}_n \sum_{t=s\|\mathcal{D}_n\|^{-2}}^{-1} \xi_{nt}^{(1)},\ s<0\right\} \xrightarrow{d} \left\{W'^{(1)}(s),\ s<0\right\}, \tag{2.62}$$

where $\{W'^{(1)}(s)\}$ is a d-dimensional standard Wiener process. This yields for every $x \in \mathbb{R}^d$

$$\left\{x^{\mathsf{T}} A \Sigma^{\frac{1}{2}}\ \mathcal{D}_n \sum_{t=s\mathcal{D}_n^{-2}}^{-1} \xi_{nt}^{(1)},\ s>0\right\} \xrightarrow{d} \left\{\|x\|_{A\Sigma A}\, W^{(1)}(s),\ s>0\right\}, \tag{2.63}$$

with $\{W^{(1)}(s)\}$ a standard Wiener process. As $\frac{\|\delta_n\|_{A\Sigma A}}{\mathcal{D}_n} \underset{n\to\infty}{\longrightarrow} c$, we define the processes

$$V_s^{(1)} = cW^{(1)}(s) - |s|g(s) \qquad s < 0\,,$$
$$V_s^{(2)} = cW^{(2)}(s) - |s|g(s) \qquad s > 0$$

and combine them to

$$W_s = \begin{cases} V_s^{(1)} & s < 0\,, \\ 0 & s = 0\,, \\ V_s^{(2)} & s > 0\,. \end{cases} \tag{2.64}$$

Thus, we can conclude[†]

$$\begin{aligned}\lim_{n\to\infty} \underset{-C<(k-m)\mathcal{D}_n^2<C}{\arg\max} \frac{w_{-2\gamma}(m/n)}{2m(1-\frac{m}{n})} V_k &= \lim_{n\to\infty} \underset{-C<s<C}{\arg\max} \frac{w_{-2\gamma}(m/n)}{2m(1-\frac{m}{n})} V_{\lfloor ns \rfloor} \\ &= \underset{-C<s<C}{\arg\max}\, W_s\,.\end{aligned}$$

■

Observe, that the asymptotic distribution now depends on the lowest decreasing index. But we the local alternative is usually analysed to derive asymptotic distribution not depending on the chosen model. In the multivariate case we have to make an additional assumption on A.

Lemma 2.4.1

Let the assumption of Theorem 2.4.4 hold. If $A' = \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ is idempotent, then

$$\mathcal{D}_n^2\Big(\hat{m}(\eta,\gamma;A) - m\Big) \xrightarrow{d} \arg\max\{W(s) - |s|g(s),\ s\in\mathbb{R}\}$$

with $W(s)$ being a two sided standard Wiener process and

$$g(s) = \begin{cases} (1-\gamma)(1-\lambda) + \gamma\lambda & s < 0 \\ 0 & s = 0 \\ \gamma(1-\lambda) + (1-\gamma)\lambda & s > 0 \end{cases}.$$

Proof:
Follows directly by Theorem 2.4.4 and with $\|\delta_n\|_{A\Sigma A} = \|\delta_n\|_{\Sigma^{-\frac{1}{2}}A'A'\Sigma^{-\frac{1}{2}}} = \|\delta_n\|_{\Sigma^{-\frac{1}{2}}A'\Sigma^{-\frac{1}{2}}} = \mathcal{D}_n$.

■

[†] This holds, because for values $k+\epsilon$, $\epsilon \in (0,1)$, from the dominating parts only the deterministic one is affected. The dominating deterministic parts of $V_{k+\epsilon}$ and of $V_{-(k+\epsilon))}$ $(k > m)$ decrease in ϵ.

2.5. Simulation

Usually we are going to use a distribution free test and estimator. To this end, we have $A = \Sigma^{-\frac{1}{2}} A' \Sigma^{-\frac{1}{2}}$ where A' is the decision matrix and Σ is the true covariance matrix of the errors. Σ is usually unknown, hence, before we come to the simulation results, we discuss the estimation of the unknown covariance matrix. In the next section we are going to introduce the estimator we used in the simulations.

2.5.1. Covariance estimator

In practice, the covariance matrix Σ is usually unknown. So, the test statistics are not directly applicable. Replacing the covariance matrix with a consistent estimator gives the same asymptotic results as it is shown in Theorem 2.2.3. For the one-dimensional case, i.e. for the test statistic

$$T_n(\eta, \gamma) = \frac{1}{\sigma} \max_{1 \le k < n} w_{\eta,\gamma}(k/n) \frac{1}{\sqrt{n}} \left| S_k(\hat{\theta}_n) \right| ,$$

different estimators for the variance were discussed, see Csörgő and Horváth [1997]. It is clear, that the sample variance could be used, but also the splitted sample variance, i.e. estimating the minimal variance under the assumption of a change. If we split the time-series at the change-point we have two stationary series. Calculating for each separately the sample variance returns under H_0 the same variance estimation after appropriate weighting. But under H_1 the estimated value will be smaller. This results in a larger value of the test statistic. As we can see, this estimator gains sensitivity also against the alternative of a change in the variance. This estimator is therefore a good and reasonable one.

In practice we cannot split at the change-point as it is unknown. To this end, we first estimate a representant of the change-point $\tilde{m}$ using $A = A'$. The resulting covariance estimator is then given as

$$\hat{\Sigma}_n = \frac{\tilde{m}}{n} \hat{\Sigma}_1 + \frac{n - \tilde{m}}{n} \hat{\Sigma}_2 \tag{2.65}$$

where $\hat{\theta}_1$, $\hat{\theta}_2$ are the parameter estimators based on the observations before and after the change-point estimator $\hat{m}$, respectively.

In the univariate case estimating the change-point with the scaled or non-scaled estimator does not matter. In the multivariate case the estimator for the change-point based on $A = A'$ and $A = \Sigma^{\frac{1}{2}} A' \Sigma^{-\frac{1}{2}}$ may differ.

2.5.2. Weight function

As we have discussed in section 2.2.2 and 2.3.2 the choice of the weight function, i.e. the choice of the parameters γ and η in

$$w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta<s<(1-\eta)\}}\,(s(1-s))^{-\gamma}$$

fulfilling **N.7**, can improve the power of the test as well as the estimation of the change-point. For the estimation of the change-point depending on the (unknown) position of the change-point we choose the power of the weight function (γ) to improve the estimation. As the position is unknown the parameter could be derived using a Plug-In technique, i.e. depending on the behaviour of the non-weighted test statistic.

Instead of a truncated weight function, with truncation parameter $\eta = \eta_0$ and power $\frac{1}{2}$, one could use a non-truncated one (i.e. $\eta = 0$) with a modified power like $\gamma = \frac{1}{2} - \eta_0(1-\eta_0)$. We are going to analyse the behaviour of both a truncated with power $\frac{1}{2}$ and a non-truncated with a power modified as proposed. The following parameter constellations are chosen:

η	0			0.01	0.05	0.1
γ	0.4901	0.4525	0.41	0.5		

Table 2.1.: Parameter choice for the weight function

The simulated corresponding critical values of the asymptotic distribution for dimensions $d = 1$, $d = 2$ and $d = 3$ are given in the appendix D.

2.5.3. Models

We have derived a class of test statistics and discussed their behaviour. In this section we are going to validate the theoretical results and take a closer look at the heavy tailed case.

Csörgő and Horváth [1997] have discussed the goodness of the asymptotic critical value for the likelihood ratio test in the univariate case. They analysed exponential, Poisson and normal observations. We focus on the multivariate case and especially on heavy-tailed observations. For the simulations we used the concept of copula. An introduction and the relation to multivariate heavy-tailed random variables is given in section B.

As we are in the mean change model (compare (2.61)) we only consider here the simulation of the residuals ε_i. For simulating a multivariate random variable based on copula, we use the algorithm 3.16 from Kroese et al. [2011].

Algorithm 1 Simulation of n random vectors based on copula

1: Generate U having distribution $C(u_1, \ldots, u_d)$.
2: Output $X = (X_1, \ldots, X_d)^{\mathrm{T}} = (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d))^{\mathrm{T}}$.
3: Repeat n times steps 1 & 2.

The following three distributions for the residuals are analysed.

Example 2.5.1 *Multivariate normal distributed, i.e* $\mathcal{L}(\varepsilon_i) = \mathcal{N}(0, \Sigma)$. *For the simulation we use that a multivariate standard normal distributed random variable has i.i.d. entries. Thus, let* ϵ_i *be independent multivariate normal distributed random variables with expectation* 0 *and covariance matrix* I_d, *then*

$$\varepsilon_t = \Sigma^{\frac{1}{2}} \epsilon_t \,,$$

with $\Sigma^{\frac{1}{2}}$ *being the Cholesky-decomposition of the positive definite covariance matrix.*

Example 2.5.2 *Two-sided multivariate log-normal distribution, i.e.* $\varepsilon_t = \varrho_t \epsilon_t$, *where each entry of* ϱ_t *is independent of the others with* 1 *or* -1 *having probability* $\frac{1}{2}$ *and* ϵ_t *are independent multivariate log-normal distributed with* 0 *and* Σ, *i.e.* $\mathcal{L}(\log \epsilon_t) = \mathcal{N}(0, \Sigma)$. *The corresponding copula (as mentioned in Example B.1.2) is given by*

$$\mathbf{C}(u_1, \ldots, u_d) = \phi_\Sigma(\phi^{-1}_{\sigma^2_{11}}(u_1), \ldots, \phi^{-1}_{\sigma^2_{dd}}(u_d)) \,.$$

The simulation of the U *in step 1 of Algorithm 1 results in simulating first a random variable* Y *which is multivariate normal distributed with expectation* 0 *and covariance matrix* Σ *and secondly calculate* U *using* $U = (\phi_{\sigma^2_{11}}(Y_1), \ldots, \phi_{\sigma^2_{dd}}(Y_d))^{\mathsf{T}}$.

The marginal distributions of the residuals are normally distributed with variances which may differ from the corresponding variances of the marginals in the copula. We know that the Gaussian copula fulfils the condition for long-tail dependence from example B.2.1. Thus the marginal distribution decides if the joint distribution of the residuals is long-tail dependent.

Example 2.5.3 *Let the observations be t-distributed. Thus, the residuals are assumed to be multivariate t-distributed with expectation* 0. *For the marginal distributions we know it is long-tail dependent for every degree of freedom* ν *greater than 2. In the multivariate case we choose the t-copula (see Example B.1.3 on page 206)*

$$\mathbf{C}(u_1, \ldots, u_d) = T_{\nu,\Sigma}(T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d)) \,.$$

As we know from Example B.2.2 the joint distribution of the random vector having t-distributed marginals and a t-copula is long-tail dependent. We observe, that if the marginals are all of the same degree of freedom t-distributed as chosen for the copula, the random vector is just multivariate t-distributed with expectation 0 *and covariance matrix* Σ. *Such a random vector can be simulated in the following way (compare Kroese et al. [2011]).*

Algorithm 2 Simulation t-distributed random vector with expectation μ and covariance matrix Σ

1: Generate $Z = (Z_1, \ldots, Z_d)^{\mathsf{T}}$ with Y_i being t_ν distributed.
2: Output $Y = (Y_1, \ldots, Y_d)^{\mathsf{T}} = \mu + \Sigma^{\frac{1}{2}} Z$ with $(\Sigma^{\frac{1}{2}})^{\mathsf{T}} \Sigma^{\frac{1}{2}} = \Sigma$.

Observe, that this is analogue as for normal distributed random variables, which can also be seen as for $\nu \to \infty$ *the t-distribution becomes the normal distribution.*

Now, to simulate the U *in step 1 of Algorithm 1, we first simulate* Y *having distribution* $T_{\nu,\Sigma}$ *as in Algorithm 2. Then* $U := (T_\nu(Y_1), \ldots, T_\nu(Y_d))^{\mathsf{T}}$.

Observe, that the marginal distributions do not have to be of the same degree of freedoms and especially their degree of freedom can be different to the one of the copula.

2.5.4. Results

Using copulas we analyse the change-point test and change-point estimator based on simulated multivariate normal and multivariate heavy tailed random variables. As introduced in Section 2.5.1, we estimate the covariance matrix of the residuals based on the splitted covariance estimator.

The algorithm implemented for the analysis of the change-point test and change-point estimator is given as follows.

Algorithm 3 Change-point test and estimator for mean change

1: Generate $X_t = (X_{t1}, \dots, X_{td})^{\mathsf{T}}$, $t = 1, \dots, n$ observations (including change or without change).
2: For $A = A'$ determine the possible change-point $\tilde{m}$.
3: Estimate the covariance matrix Σ using (2.65) with $\tilde{m}$ from the step before.
4: Calculate the test statistic with $A = \hat{\Sigma}^{\frac{1}{2}} A' \hat{\Sigma}^{\frac{1}{2}}$ and the corresponding change-point estimator $\hat{m}$.
5: For a given level α compare the test statistic with the critical value of the asymptotic distribution (possibly simulated).
6: Output test decision and $\hat{m}$.

We use the following parameter constellations. Let the expectation μ change from 0 to

$$a)\ \mu_2 = \begin{pmatrix} 0.5 \\ 0 \\ 0 \end{pmatrix}, \qquad b)\ \mu_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad c)\ \mu_2 = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}, \qquad d)\ \mu_2 = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}$$

while we use the following covariance matrices

$$i)\ \Sigma = 0.25 I_d \qquad ii)\ \Sigma = I_d \qquad iii)\ \Sigma = 10 I_d$$

$$iv)\ \Sigma = \begin{pmatrix} 1 & 0.9 & 0 \\ 0.9 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad v)\ \Sigma = \begin{pmatrix} 1 & 0 & 0.9 \\ 0 & 1 & 0 \\ 0.9 & 0 & 1 \end{pmatrix} \qquad vi)\ \Sigma = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix}.$$

For the multivariate normal case (Example 2.5.2) this is the covariance matrix. In the other cases (Example 2.5.2 and 2.5.3) we use Σ for the joint distribution in the copula. Here, we have to define additional parameters.

In the log-normal case the marginal distributions are with parameters $\mu = 0$ and $\sigma^2 = 1$, i.e. $\mathcal{L}(\log(\epsilon_{ti})) = \mathcal{N}(0, 1)$ for every $t = 1, \dots, n$ and $i = 1, \dots, d$.

The t-distribution has finite second moment only for a degree of freedom greater than 2. We choose for the marginal distribution a degree of freedom of 3 and for the copula used 2.

Size

First of all, we compare the size of the different statistics. In Table 2.2, 2.3 and 2.4 the size for the multivariate normal, log-normal and t-distributions are given for the 6 different covariance matrices i) to vi) from page 64. Thereby, we distinguish between diagonal covariance matrices (case i), ii) and iii)) and non-diagonal covariance matrices (case iv), v) and vi)). We are interested in the behaviour for small sample sizes. Hence, we run the test for $N = 20, 50, 100$ observations with $M = 10^6$ repetitions. In the case of t-distributed observations we run into calculation problems, concerning the root of the covariance matrix, for the sample size $N = 100$. So, they are missing.

In Table 2.2 we observe that the size of the test converges to the level for normally distributed observations having a diagonal covariance matrix (case i) to iii)). Otherwise, the size is quite high (between 0.15 and 0.38). A decrease is observable in the cases iv) and v), but in the case vi) it seems to be constant.

The estimations of the size under heavy tailed observations (Table 2.3 and Table 2.4) give comparable results. For non-diagonal covariance matrices (cases iv),v) and vi)) the size is always higher than 0.56, even 1 for t-distributed observations.

In the case of a diagonal covariance matrix, the estimated size is between 0.04 and 0.52. Let us take a closer look at the results for log-normal distributed observations (Table 2.3). In the case i) and ii) the estimated size is between 0.04 and 0.27. A tendency is hardly observable. But for iii) we can see that the size seems to fall. For example in the case of the weighted CUSUM we have estimated values between 0.5 and 0.54 for $N = 20$ and between 0.17 and 0.37 for $N = 100$. Nevertheless, the values are quite high. This is not the case for t-distributed observations (with a diagonal covariance matrix). Although we have only estimations for $N = 20$ and $N = 50$, we observe that the estimated size for i) and ii) is comparable to that in the case of log-normal distribution. But there is no significant difference observable in between the cases i), ii) and iii) for t-distributed observations.

In conclusion, a dependence of the size of the test on the underlying covariance matrix of the observations is discernable. A closer look at the estimation of the covariance matrix has to be done. The use of robust change-point tests even for heavy tailed observations with second moment should be analysed in comparison to the weighted CUSUM. Also dimension reduction should be considered in view of the size.

N	η	γ	i)	ii)	iii)	iv)	v)	vi)
20	0	0	0.174	0.175	0.175	0.307	0.233	0.325
	0.01	0.5	0.148	0.147	0.148	0.301	0.222	0.341
	0	0.4901	0.144	0.143	0.144	0.293	0.215	0.332
	0.05	0.5	0.172	0.171	0.172	0.335	0.256	0.379
	0	0.4525	0.174	0.173	0.174	0.327	0.247	0.365
	0.1	0.5	0.202	0.201	0.202	0.352	0.271	0.378
	0	0.41	0.188	0.187	0.188	0.339	0.259	0.374
50	0	0	0.085	0.084	0.085	0.196	0.165	0.275
	0.01	0.5	0.073	0.073	0.074	0.217	0.181	0.337
	0	0.4901	0.068	0.068	0.069	0.205	0.17	0.322
	0.05	0.5	0.088	0.088	0.088	0.228	0.191	0.334
	0	0.4525	0.085	0.085	0.086	0.229	0.193	0.346
	0.1	0.5	0.091	0.091	0.092	0.225	0.19	0.323
	0	0.41	0.091	0.091	0.092	0.233	0.196	0.344
100	0	0	0.063	0.063	0.063	0.168	0.152	0.269
	0.01	0.5	0.056	0.056	0.056	0.206	0.186	0.366
	0	0.4901	0.05	0.05	0.05	0.191	0.172	0.346
	0.05	0.5	0.064	0.064	0.064	0.204	0.185	0.344
	0	0.4525	0.062	0.062	0.062	0.207	0.188	0.361
	0.1	0.5	0.065	0.065	0.065	0.194	0.176	0.319
	0	0.41	0.065	0.065	0.066	0.206	0.186	0.351

Table 2.2.: Size for multivariate normal distributed random vectors with $\alpha = 0.05$

N	η	γ	i)	ii)	iii)	iv)	v)	vi)
20	0	0	0.171	0.178	0.377	0.673	0.665	0.868
	0.01	0.5	0.183	0.224	0.519	0.702	0.695	0.884
	0	0.4901	0.175	0.214	0.511	0.697	0.689	0.881
	0.05	0.5	0.21	0.254	0.539	0.726	0.72	0.896
	0	0.4525	0.196	0.23	0.519	0.719	0.713	0.894
	0.1	0.5	0.235	0.262	0.5	0.719	0.713	0.892
	0	0.41	0.201	0.227	0.506	0.726	0.72	0.898
50	0	0	0.074	0.069	0.184	0.595	0.591	0.849
	0.01	0.5	0.167	0.212	0.403	0.678	0.674	0.892
	0	0.4901	0.156	0.201	0.394	0.67	0.666	0.888
	0.05	0.5	0.133	0.161	0.361	0.663	0.659	0.888
	0	0.4525	0.159	0.201	0.397	0.691	0.688	0.901
	0.1	0.5	0.107	0.119	0.26	0.645	0.641	0.88
	0	0.41	0.148	0.184	0.384	0.693	0.69	0.903
100	0	0	0.107	0.047	0.096	0.563	0.563	0.843
	0.01	0.5	0.211	0.213	0.361	0.682	0.68	0.907
	0	0.4901	0.199	0.199	0.351	0.672	0.67	0.902
	0.05	0.5	0.151	0.129	0.304	0.651	0.65	0.898
	0	0.4525	0.193	0.191	0.347	0.69	0.689	0.913
	0.1	0.5	0.125	0.083	0.173	0.62	0.619	0.879
	0	0.41	0.175	0.167	0.327	0.688	0.687	0.913

Table 2.3.: Size for multivariate log-normal distributed random vectors with $\alpha = 0.05$

N	η	γ	i)	ii)	iii)	iv)	v)	vi)
20	0	0	0.196	0.195	0.192	1	1	1
	0.01	0.5	0.212	0.208	0.2	1	1	1
	0	0.4901	0.205	0.201	0.194	1	1	1
	0.05	0.5	0.24	0.237	0.228	1	1	1
	0	0.4525	0.229	0.226	0.219	1	1	1
	0.1	0.5	0.261	0.258	0.252	1	1	1
	0	0.41	0.237	0.234	0.228	1	1	1
50	0	0	0.088	0.088	0.087	1	1	1
	0.01	0.5	0.171	0.168	0.159	1	1	1
	0	0.4901	0.16	0.157	0.149	1	1	1
	0.05	0.5	0.143	0.141	0.136	1	1	1
	0	0.4525	0.166	0.164	0.156	1	1	1
	0.1	0.5	0.119	0.118	0.115	1	1	1
	0	0.41	0.156	0.154	0.148	1	1	1

Table 2.4.: Size for multivariate t distributed random vectors with $\alpha = 0.05$

Power

For the analysis of the power, we restrict to the case ii) for the covariance matrix, i.e. we use the unit matrix. We analyse the 4 different cases a), b), c) and d) of the mean change introduced on page 64.

First, we consider a change in the middle. The results are given in table 2.5, 2.6 and 2.7. Observe, that even for these small sample sizes, the power for multivariate normal distributed observations is astonishingly good, especially in the case b). We observe that the power in c) is constantly less than in b).

N	η	γ	a)	b)	c)	d)
20	0	0	0.257	0.495	0.351	0.45
	0.01	0.5	0.201	0.376	0.267	0.345
	0	0.4901	0.197	0.372	0.263	0.341
	0.05	0.5	0.231	0.416	0.301	0.382
	0	0.4525	0.236	0.429	0.309	0.394
	0.1	0.5	0.264	0.454	0.337	0.42
	0	0.41	0.256	0.459	0.334	0.421
50	0	0	0.281	0.79	0.496	0.679
	0.01	0.5	0.188	0.638	0.349	0.519
	0	0.4901	0.182	0.633	0.342	0.513
	0.05	0.5	0.221	0.682	0.393	0.566
	0	0.4525	0.227	0.698	0.405	0.581
	0.1	0.5	0.238	0.709	0.419	0.595
	0	0.41	0.249	0.731	0.439	0.616
100	0	0	0.463	0.981	0.788	0.935
	0.01	0.5	0.307	0.941	0.628	0.845
	0	0.4901	0.299	0.94	0.622	0.842
	0.05	0.5	0.35	0.954	0.673	0.873
	0	0.4525	0.363	0.959	0.69	0.883
	0.1	0.5	0.376	0.962	0.701	0.889
	0	0.41	0.397	0.968	0.724	0.902

Table 2.5.: Power for multivariate normal distributed random vectors with $\lambda = 0.5$ and $\alpha = 0.05$

For the heavy tailed distributions the weighted CUSUM is significantly better than the CUSUM statistic. For example the power of the CUSUM statistic is 0.047 in the multivariate log-normal case for $N = 50$ and mean change a). But the weighted CUSUM has power between 0.139 and 0.25.

Now, we consider an early change. In table 2.8, 2.9 and 2.10 the power for an early change at $\lambda = 0.1$ is given. Let us compare the truncated weight function with the proposed untruncated

N	η	γ	a)	b)	c)	d)
20	0	0	0.113	0.205	0.157	0.239
	0.01	0.5	0.222	0.269	0.231	0.261
	0	0.4901	0.209	0.258	0.218	0.249
	0.05	0.5	0.253	0.305	0.264	0.299
	0	0.4525	0.219	0.278	0.233	0.272
	0.1	0.5	0.261	0.327	0.278	0.317
	0	0.41	0.206	0.274	0.224	0.27
50	0	0	0.047	0.155	0.101	0.228
	0.01	0.5	0.25	0.314	0.26	0.297
	0	0.4901	0.236	0.3	0.247	0.284
	0.05	0.5	0.191	0.276	0.207	0.255
	0	0.4525	0.233	0.31	0.25	0.302
	0.1	0.5	0.139	0.242	0.161	0.217
	0	0.41	0.209	0.296	0.232	0.297
100	0	0	0.04	0.211	0.139	0.364
	0.01	0.5	0.255	0.363	0.28	0.361
	0	0.4901	0.239	0.347	0.264	0.346
	0.05	0.5	0.167	0.31	0.205	0.308
	0	0.4525	0.229	0.359	0.267	0.378
	0.1	0.5	0.116	0.287	0.163	0.285
	0	0.41	0.2	0.346	0.249	0.383

Table 2.6.: Power for multivariate log-normal distributed random vectors with $\lambda = 0.5$ and $\alpha = 0.05$

N	η	γ	a)	b)	c)	d)
20	0	0	0.099	0.136	0.118	0.147
	0.01	0.5	0.176	0.198	0.181	0.189
	0	0.4901	0.169	0.191	0.174	0.182
	0.05	0.5	0.201	0.225	0.207	0.216
	0	0.4525	0.185	0.211	0.192	0.204
	0.1	0.5	0.229	0.256	0.236	0.247
	0	0.41	0.185	0.214	0.194	0.208
50	0	0	0.04	0.085	0.066	0.113
	0.01	0.5	0.171	0.203	0.178	0.189
	0	0.4901	0.159	0.191	0.165	0.177
	0.05	0.5	0.145	0.187	0.154	0.169
	0	0.4525	0.159	0.198	0.169	0.186
	0.1	0.5	0.117	0.166	0.129	0.147
	0	0.41	0.143	0.186	0.155	0.177

Table 2.7.: Power for multivariate t-distributed random vectors with $\lambda = 0.5$ and $\alpha = 0.05$

version. In the multivariate normal case, the untruncated version has slightly lower power. For the log-normal case the power is better if $\eta = 0.05, 0.1$ and $N = 50, 100$. In the case of the t-distribution the behaviour is comparable. We conclude that for heavy tailed distributions the untruncated version has good power (compared to the truncated weight function) for sample sizes not less than 50. Clearly, this effect could be caused by the different sizes. A further study, especially in view of size correction and also on the change-point estimator should be considered.

N	η	γ	a)	b)	c)	d)
20	0	0	0.18	0.192	0.185	0.193
	0.01	0.5	0.157	0.191	0.17	0.187
	0	0.4901	0.153	0.186	0.166	0.182
	0.05	0.5	0.183	0.22	0.197	0.215
	0	0.4525	0.184	0.218	0.197	0.214
	0.1	0.5	0.215	0.256	0.231	0.251
	0	0.41	0.198	0.231	0.211	0.227
50	0	0	0.096	0.135	0.109	0.125
	0.01	0.5	0.099	0.22	0.136	0.182
	0	0.4901	0.093	0.207	0.127	0.171
	0.05	0.5	0.118	0.252	0.16	0.211
	0	0.4525	0.112	0.231	0.149	0.194
	0.1	0.5	0.12	0.252	0.161	0.211
	0	0.41	0.117	0.23	0.152	0.195
100	0	0	0.087	0.183	0.115	0.15
	0.01	0.5	0.114	0.417	0.203	0.313
	0	0.4901	0.104	0.396	0.188	0.294
	0.05	0.5	0.133	0.46	0.232	0.351
	0	0.4525	0.12	0.42	0.208	0.317
	0.1	0.5	0.128	0.449	0.223	0.34
	0	0.41	0.121	0.407	0.204	0.308

Table 2.8.: Power for multivariate normal distributed random vectors with $\lambda = 0.1$ and $\alpha = 0.05$

N	η	γ	a)	b)	c)	d)
20	0	0	0.192	0.196	0.197	0.213
	0.01	0.5	0.384	0.403	0.392	0.415
	0	0.4901	0.371	0.391	0.38	0.402
	0.05	0.5	0.41	0.43	0.419	0.443
	0	0.4525	0.375	0.396	0.384	0.408
	0.1	0.5	0.384	0.409	0.396	0.423
	0	0.41	0.357	0.377	0.366	0.389
50	0	0	0.03	0.039	0.031	0.035
	0.01	0.5	0.231	0.262	0.228	0.231
	0	0.4901	0.218	0.248	0.214	0.217
	0.05	0.5	0.179	0.223	0.177	0.186
	0	0.4525	0.21	0.243	0.208	0.212
	0.1	0.5	0.128	0.181	0.131	0.143
	0	0.41	0.187	0.219	0.185	0.19
100	0	0	0.009	0.022	0.011	0.014
	0.01	0.5	0.226	0.283	0.225	0.228
	0	0.4901	0.209	0.263	0.208	0.21
	0.05	0.5	0.128	0.207	0.131	0.14
	0	0.4525	0.189	0.248	0.19	0.194
	0.1	0.5	0.078	0.165	0.087	0.102
	0	0.41	0.153	0.211	0.154	0.158

Table 2.9.: Power for multivariate log-normal distributed random vectors with $\lambda = 0.1$ and $\alpha = 0.05$

N	η	γ	a)	b)	c)	d)
20	0	0	0.132	0.135	0.134	0.138
	0.01	0.5	0.298	0.307	0.302	0.309
	0	0.4901	0.288	0.296	0.291	0.298
	0.05	0.5	0.324	0.333	0.328	0.335
	0	0.4525	0.294	0.303	0.298	0.306
	0.1	0.5	0.313	0.325	0.318	0.327
	0	0.41	0.283	0.292	0.287	0.294
50	0	0	0.016	0.019	0.018	0.021
	0.01	0.5	0.153	0.17	0.159	0.168
	0	0.4901	0.142	0.158	0.148	0.157
	0.05	0.5	0.122	0.143	0.13	0.141
	0	0.4525	0.137	0.154	0.144	0.153
	0.1	0.5	0.091	0.113	0.099	0.11
	0	0.41	0.119	0.136	0.125	0.135

Table 2.10.: Power for multivariate t-distributed random vectors with $\lambda = 0.1$ and $\alpha = 0.05$

3. Change-point tests

Since the 1950's many different models have been analysed in deriving off-line test statistics and estimators in the field of change-point analysis. In this chapter we analyse change-point tests for some examples. The first example considered in section 3.1 considers a mean change model with i.i.d. random variables only assuming the existence of the first moment. We construct the change-point test having a randomized weight function. We are able to prove the asymptotic behaviour for this test statistic.

As a second example we decided to analyse a NLAR(p)-process or non-linear regression for a change in the unknown regression function (section 3.2). To overcome the non-parametric problem, we used neural-networks to approximate the regression function. This approach is close to practical situations as we allow for possible misspecification.

We complete this chapter giving regularity conditions for change-point tests based on the concept of estimation functions. The key assumptions are given in this section, 3.3. We also show that for smooth enough function and some moment conditions these key assumptions can be derived.

3.1. Randomized weight functions

Up to now, we considered the weighted CUSUM statistic with the deterministic weight function

$$w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta<s<1-\eta\}}\,(s(1-s))^{-\gamma} \qquad s \in (0,1)\,.$$

This is a typical form of the weight function considered in the literature, but a generalised class of test statistics so-called q-weighted CUSUM statistics are also considered (see Kirch and Tadjuidje Kamgaing [2012] or Csörgő and Horváth [1997] section 4.1). The weight function w is then defined as $1/q$ where q fulfils some integral conditions. As we mentioned before, this conditions are equivalent to (1.4) and (1.5).

In all cases the weight function is deterministic and free to choose. A randomized weight function in the set-up of autroregressive residuals is considered in Zhou and Liu [2008]. It turns out that the constructed test has a different asymptotic distribution as usual, depending on the expectation of the weighted autoregressive process. But the consistency of the estimator is still given. We use the idea for the case of i.i.d. observations with mean change under possible infinite second moment. We derive a modified change-point test which has the usual asymptotic properties. The weight function of this change-point test is randomized.

In section 3.1.1 we construct the change-point test motivated by the robust statistic. We then show the asymptotics for the test statistic.

3.1.1. Construction

As explained in section 2, we usually assume errors with a finite second moment. Here, we are going to introduce a method for i.i.d. observations under the weaker assumption that only the first moment exists. Robust techniques are known like the change-point test based on an M-estimator, analysed for the case of i.i.d. observations with change in mean in Hušková [1996] and for the mean change under strong mixing residuals in Hušková and Marušiaková [2012]. In the case of regressions we refer to Prášková and Chochola [2014]. In Hušková [1996] the unweighted version of our test statistic is analysed along with the MOSUM and the pseudo maximum likelihood statistic while in Hušková and Marušiaková [2012] the weighted test statistic based on M-estimators is studied. The proofs are missing in this publication. We do not only analyse the weighted test statistic based on a subclass of M-estimator but also discuss a weight function which is data driven. The motivation comes from the paper Zhou and Liu [2008], where autoregressive processes with infinite variances are analysed. At the end of this section a new version of weight functions is derived. The idea of the proofs is used to define a class of weight functions, such that the weighted CUSUM statistics still have the known asymptotics (see section 2.2).

Let us assume that $X_1, \ldots, X_n \in \mathbb{R}$ follow the AMOC-model

$$X_t = \begin{cases} \theta + \varepsilon_t & , 1 \le t \le m \\ \theta + \delta_n + \varepsilon_t & , m < t \le n \end{cases}, \tag{3.1}$$

where m fulfils **G.1** and $\delta_n \underset{n\to\infty}{\longrightarrow} 0$. Recall the definition of the AMOC-model (1.1), then $X_t^{(1)} = \theta + \varepsilon_t$ and $X_t^{(2)} = \theta + \delta_n + \varepsilon_t$. Differently to section 2 we only assume the errors to have at least first moment.

L.1 Let $\{\varepsilon_t\}$ be i.i.d. with $\mathbb{E}[\varepsilon_1] = 0$ and $\mathbb{E}[|\varepsilon_1|^\nu] < \infty$ for some $\nu > 1$.

To handle the difficulties with the possibly infinite variance, in Hušková [1996] M-estimators are considered. So, the parameter estimator solves the equation

$$\sum_{t=1}^{n} \phi(X_t, \theta) = 0\,, \tag{3.2}$$

where $\phi(x, \theta)$ is assumed to be non-decreasing in θ and $\mathbb{E}[\phi^2(X_1, \theta)] < \infty$. We are going to take a closer look at estimators as solutions of (3.2) where ϕ can be written as

$$\phi(x, \theta) = g(x)(x - \theta)\,.$$

For simplicity of notation we introduce the following sum.

N.9

$$S_g(n, \theta) := \sum_{t=1}^{n} g(X_t)(X_t - \theta)\,.$$

The weight function $g(\cdot)$ should be chosen in the way that the weighted process has finite second moment. An intuitive example is a truncation function such as $g(x) = \mathbf{1}_{\{|x-\theta|\leq K\}} + \mathbf{1}_{\{\{x-\theta\}>K\}} \frac{K}{|x-\theta|}$, with $K >> 0$. Further examples are given on page 78.

L.2 Let $g : \mathbb{R}^d \to \mathbb{R}$ be a measurable function such that $g(X_1^{(z)})$ is a.s. positive, bounded and $\mathbb{E}\left[g^2(X_1^{(z)})\left(1 + \left\|X_1^{(z)}\right\|^2\right)\right] < \infty$, $z = 1, 2$.

Besides the condition on the existence of the second moment for the weighted process, we will need the existence of the finite second moment of the random variable $g(X_1^{(z)})$. This is not a strong assumption as the function g itself is chosen in the way that it reduces the variability of the process. But, the assumption is not only under H_0 but also under H_1, i.e. for the time series after the change, which is needed for the change-point estimator in section 4.1.

L.3 Furthermore, let g fulfil $\mathbb{E}[g(X_1^{(1)})\varepsilon_1] = 0$.

Notize, the assumption **L.3**, the zero expectation of the modified errors, is only true for the unchanged part of the time-series. So we have to be careful especially in the proofs under the alternative. The parameter estimator is given as

$$\hat{\theta}_n(g) = \frac{\sum_{t=1}^{n} g(X_t)X_t}{\sum_{t=1}^{n} g(X_t)}\,. \tag{3.3}$$

We can prove that the estimator is also consistent under both hypotheses, H_0 and H_1.

Theorem 3.1.1 *Assume **G.1**, **L.1** – **L.3**. Let $\hat{\theta}_n$ be the solution of $S_g(n,\theta) = 0$ (see **N.9**) under H_0 as well as under H_1. If δ_n fulfils either **G.5.a)** or **G.5.b)** and $\delta = \lim_{n\to\infty} \delta_n$, then we have:*

a) $\hat{\theta}_n(g)$ is a consistent estimator, i.e

$$\hat{\theta}_n(g) \xrightarrow{a.s} \tilde{\theta}\,,$$

where $\tilde{\theta} = \theta$ under H_0 and under H_1 $\tilde{\theta} = \theta + (1-\lambda)(c_1\delta + c_2)$ with c_1 and c_2 some constants depending on g and ε.

b)

$$\sqrt{n}(\hat{\theta}_n(g) - \tilde{\theta}) \xrightarrow{d} \mathcal{N}(0, V)\,,$$

with $\tilde{\theta}$ as in a) and V the covariance matrix depending on $g(X_1^{(z)})\varepsilon_1$ as well as $g(X_1^{(z)})$, $z = 1, 2$.

This theorem gives us $|\hat{\theta}_n - \theta| = O_P\left(n^{-\frac{1}{2}}\right)$. For the detailed proof see page 81.

Observe, that even under the local alternative, we do not necessary have $\tilde{\theta} = \theta$. Later on, we will need this, so we make an additional assumption which guarantees $\tilde{\theta} = \theta$ under the local alterantive.

L.4 For $\delta_n \underset{n\to\infty}{\longrightarrow} 0$ it follows $\mathbb{E}[g(X_1^{(2)})\varepsilon_1] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[g(X_1^{(1)})\varepsilon_1]$.

Proposition 3.1.1

*Let $\hat{\theta}_n(g)$, g and $\{X_t\}$ be as in Theorem 3.1.1 and δ_n fulfil **G.5.b)**. If additionally assumption **L.4** holds true, the results of Theorem 3.1.1 are true for $\tilde{\theta} = \theta$.*

For the construction of the test statistic let $d = 1$. Recall that the least squares estimator under H_0 is given as

$$\sum_{t=1}^{n} g(X_t)(X_t - \hat{\theta}_n)^2 = \min_{\theta}\left(\sum_{t=1}^{n} g(X_t)(X_t - \theta)^2\right)$$

then

$$\hat{\theta}_n(g) = \frac{\sum_{t=1}^{n} g(X_t)X_t}{\sum_{t=1}^{n} g(X_t)}\,.$$

Equivalently, we determine the estimators under the alternative. They are given as

$$\sum_{t=1}^{k} g(X_t)(X_t - \hat{\theta}_1)^2 = \min_{\theta_1} \sum_{t=1}^{k} g(X_t)(X_t - \theta_1)^2$$
$$\sum_{t=k+1}^{n} g(X_t)(X_t - \hat{\theta}_2)^2 = \min_{\theta_2} \sum_{t=k+1}^{n} g(X_t)(X_t - \theta_2)^2$$

with

$$\hat{\theta}_1(g) = \hat{\theta}_k \qquad \text{and} \qquad \hat{\theta}_2(g) = \frac{\sum_{t=k+1}^{n} g(X_t)X_t}{\sum_{t=k+1}^{n} g(X_t)} =: \hat{\theta}_{k+1,n} \,. \tag{3.4}$$

As in the case of the mean change model (section 2.2.1), we consider the difference of the sums based on the estimation under H_0 or under H_1. We get

$$\begin{aligned}
&\left(\sum_{t=1}^{n} g(X_t)(X_t - \hat{\theta}_n(g))^2\right) - \left(\sum_{t=1}^{k} g(X_t)(X_t - \hat{\theta}_1(g))^2 + \sum_{t=k+1}^{n} g(X_t)(X_t - \hat{\theta}_2(g))^2\right) \\
&\quad = \sum_{t=1}^{n} g(X_t)(X_t - \hat{\theta}_n(g))^2 - \sum_{t=1}^{k} g(X_t)(X_t - \hat{\theta}_k(g))^2 - \sum_{t=k+1}^{n} g(X_t)(X_t - \hat{\theta}_{k+1,n}(g))^2 \\
&\quad = \sum_{t=1}^{k} g(X_t)\left(X_t - \hat{\theta}_k(g) - (\hat{\theta}_n(g) - \hat{\theta}_k(g))\right)^2 \\
&\qquad + \sum_{t=k+1}^{n} g(X_t)\left(X_t - \hat{\theta}_{k+1,n}(g) - (\hat{\theta}_n(g) - \hat{\theta}_{k+1,n}(g))\right)^2 \\
&\qquad - \sum_{t=1}^{k} g(X_t)(X_t - \hat{\theta}_k(g))^2 - \sum_{t=k+1}^{n} g(X_t)(X_t - \hat{\theta}_{k+1,n}(g))^2 \\
&\quad = (\hat{\theta}_n(g) - \hat{\theta}_k(g))^2 \left(\sum_{t=1}^{k} g(X_t)\right) + (\hat{\theta}_n - \hat{\theta}_{k+1,n}(g))^2 \left(\sum_{t=k+1}^{n} g(X_t)\right)
\end{aligned}$$

with $\sum_{t=1}^{k} g(X_t)(X_t - \hat{\theta}_k(g)) = 0$ and $\sum_{t=k+1}^{n} g(X_t)(X_t - \hat{\theta}_{k+1,n}(g)) = 0$. Using the definition of the estimators (see (3.3) and (3.4)) we have

$$\begin{aligned}
&(\hat{\theta}_n(g) - \hat{\theta}_k(g))^2 \left(\sum_{t=1}^{k} g(X_t)\right) + (\hat{\theta}_n(g) - \hat{\theta}_{k+1,n}(g))^2 \left(\sum_{t=k+1}^{n} g(X_t)\right) \\
&\quad = \left(\sum_{t=1}^{k} g(X_t)X_t - \left(\sum_{t=1}^{k} g(X_t)\right)\hat{\theta}_n(g)\right)^2 \frac{1}{\sum_{t=1}^{k} g(X_t)} \\
&\qquad + \left(\left(\sum_{t=1}^{n} g(X_t) - \sum_{t=1}^{k} g(X_t)\right)\hat{\theta}_n(g) - \left(\sum_{t=1}^{n} g(X_t)X_t - \sum_{t=1}^{k} g(X_t)X_t\right)\right)^2 \frac{1}{\sum_{t=k+1}^{n} g(X_t)} \\
&\quad = \left(\sum_{t=1}^{k} g(X_t)X_t - \left(\sum_{t=1}^{k} g(X_t)\right)\hat{\theta}_n(g)\right)^2 \left(\frac{1}{\sum_{t=1}^{k} g(X_t)} + \frac{1}{\sum_{t=k+1}^{n} g(X_t)}\right) \\
&\quad = \left(\sum_{t=1}^{k} g(X_t)X_t - \left(\sum_{t=1}^{k} g(X_t)\right)\hat{\theta}_n(g)\right)^2 \left(\frac{\sum_{t=1}^{n} g(X_t)}{\sum_{t=1}^{k} g(X_t)\left(\sum_{t=1}^{n} g(X_t) - \sum_{t=1}^{k} g(X_t)\right)}\right) .
\end{aligned}$$

Due to the fact, that the change-point is unknown, we take the maximum over all possible values k. The change-point test statistic is given as

$$T_n = \max_{1 \le k < n} \left(\frac{\sum_{t=1}^{n} g(X_t)}{\sum_{t=1}^{k} g(X_t)\left(\sum_{t=1}^{n} g(X_t) - \sum_{t=1}^{k} g(X_t)\right)}\right)^{\frac{1}{2}} \left|S_g(k, \hat{\theta}_n(g))\right| . \tag{3.5}$$

In Zhou and Liu [2008] they analysed the truncated version of this kind of statistics. If $g \equiv 1$ we get the weighted CUSUM statistic which converges to a Gumble distribution. To derive the asymptotics, we modified the weight function, see section 2.2. Equivalently, we multiply the statistic in (3.5) with $(\overline{g(X)}_n)^{\frac{1}{2}}$. Since $\mathbb{E}[g(X_t)] \neq 0$ the asymptotic statistic of the truncated version of (3.5) can be derived using Slutsky C.1.4. We are going to analyse the asymptotic behaviour of the modified multivariate version.

N.10 Denote with $w_{\eta,\gamma;g}$ the weight function

$$w_{\eta,\gamma;g}(k,n) = \mathbf{1}_{\{\eta n < k < (1-\eta)n\}} \left(\frac{\left(\sum_{t=1}^n g(X_t)\right)^2}{\sum_{t=1}^k g(X_t) \left(\sum_{t=1}^n g(X_t) - \sum_{t=1}^k g(X_t) \right)} \right)^\gamma ,$$

with $\eta \in (0, \frac{1}{2})$.

With this notation we analyse test statistics of the form

$$T_n(\eta, \gamma; g) = \max_{1 \le k < n} w_{\eta,\gamma;g}(k,n) \frac{1}{\sqrt{n}} \left\| S_g(k, \hat{\theta}_n(g)) \right\|_{\Sigma_g^{-1}}, \tag{3.6}$$

with $\Sigma_g = \mathbb{V}\mathrm{ar}[g(X_t)\varepsilon_t]$. Observe, that if we have $g(x) \equiv 1$, $\Sigma_g = \Sigma$ and the errors ε_t have finite second moment. The test statistic is then the weighted CUSUM as mentioned before, i.e.

$$T_n(\eta, \gamma) = T_n(\eta, \gamma; \Sigma^{-1}) = \max_{1 \le k < n} w_{\eta,\gamma}(k/n) \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^k X_t - \frac{k}{n} \sum_{t=1}^n X_t \right\|_{\Sigma^{-1}} .$$

So, the derived test coincides with the known theory.

Let us for the moment allow g to depend on θ, too, and consider some examples. Clearly, these functions are not applicable, because they depend on the unknown parameter θ. We are going to see how we can modify the weight function.

Example 3.1.1 *Let θ be the true parameter under H_0 and ε_t be a sequence of continuous random variables. An intuitive idea is to truncate all values having a large error value. Then, the weight function can be chosen as $\tilde{g}(x) = \mathbf{1}_{\{|x-\theta| \le K\}} + \mathbf{1}_{\{\{x-\theta\} > K\}} \frac{K}{|x-\theta|}$, with $K >> 0$. This function fulfils assumption **L.2**, if the distribution of ε_1 is symmetric around zero. At some points later on we will need a g which is two times differentiable. Of course $\tilde{g}$ is not. Hence, we use a function g which is C^2 and equals $\tilde{g}$ outside $(-K-\epsilon, -K+\epsilon) \cup (K-\epsilon, K+\epsilon)$ instead. Such a function g can easily be found by replacing $\tilde{g}$ at $(-K-\epsilon, -K+\epsilon) \cup (K-\epsilon, K+\epsilon)$ by suitable polynomials of order 5 or higher.*

Example 3.1.2 *We can also construct examples given by $g(x) = \phi(x-\theta)/(x-\theta)$ and ϕ being some generating function for an M-estimator, like the Tukey function or Welsh function. As long as ε_1 is symmetric around zero, the assumption **L.2** is fulfilled.*

Example 3.1.3 *Let $g(\cdot)$ be as in Example 3.1.1. The corresponding test is a consistent level-α test. In practice this test is not applicable due to the dependence of $g(\cdot)$ on the unknown parameter θ, i.e. $g(x) = g(x;\theta)$. Replacing the unknown parameter with a robust estimator will solve this problem. Let $\hat{\theta}_n$ be a robust estimator for θ, then $g_n(x) = g(x;\hat{\theta}_n)$ is the function of interest.*

From now on, we define $g(x)$ as in the above example.

N.11 Let $g(x) = g(x,\theta)$, where θ is the true parameter under H_0 or up to time m. If θ is replaced by an estimator $\hat{\theta}_n$ we denote $g_n(X_t) = g(X_t;\hat{\theta}_n)$.

Observe, that the function $g_n(x)$ given in example 3.1.3 is evaluated at $\{X_t,\ t = 1,\ldots,n\}$. So we did not have a function $g_n(X)$ with X independent of $X_1,\ldots,X_n$ as the estimator $\dot{\theta}$ depends on $X_1,\ldots,X_n$. Therefore, we have to make additional assumptions on the weight function.

First, we need to make an assumption on the asymptotic behaviour of the parameter estimator $\hat{\theta}_n$ used for $g_n(x)$.

L.5 The estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent estimator w.r.t. θ under H_0 and under H_1 (local alternative, i.e. **G.5.b)**).

We only consider the local alternative from now on, as under the fixed alternative the estimator does not always converge against θ. But this is necessary, since we are going to use that the function $g_n(x)$ is converges against $g(x)$.

L.6 Let $g_n(X_t)$ be an a.s. positive function with $\mathbb{E}[g_n^2(X_1^{(z)})(1 + \left\|X_1^{(z)}\right\|^2)] < c < \infty$, for all n, and

$$P\left(\lim_{n\to\infty} g_n\left(X_t^{(z)}\right) = g\left(X_t^{(z)}\right)\right) = 1 \qquad \forall t\,, \tag{3.7}$$

where $z = 1,2$ and $g(\cdot)$ fulfils **L.2** and **L.3**. *

L.7 Let $g(x) = g(x;\theta)$ be 2-times differentiable w.r.t. θ with $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla g(X_1^{(z)})\varepsilon_1\right\|\right] < \infty$ and $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^2 g(X_1^{(z)})\varepsilon_1\right\|\right] < \infty$.

Observe, the weight function in Example 3.1.3 fulfils this assumption due to the convergence of the median.

Notice, the assumption **L.6** implies that we only consider the local alternative. It is important to recognize, that under this assumption the strong law of large numbers holds true for $g_n(X_t)$.

Lemma 3.1.1

*Let g_n and g fulfil the assumption **L.6**. Then for the model* (3.1) *it holds under H_0 and H_1 (i.e. **G.5.b)**)*

$$\frac{1}{n}\sum_{t=1}^{n}(g_n(X_t) - g(X_t))(1+\varepsilon_t) = o_{a.s.}(1)\,.$$

This implies that under H_0 and under H_1 the limit of the arithmetic sum w.r.t. $g_n(X_t)$ can be identified as the limit of the arithmetic sum of $g(X_t)$. But for determining the rate, we have had to make the additional assumptions on the derivatives. Then we derive the following Theorem.

*E.g. for $g(x,\theta)$ continuous in $\theta = \tilde{\theta}$ for $P_{X_t^{(z)}}$, $z = 1,2$, a.s. all x this assumption is fulfilled.

Theorem 3.1.2 *Assume **G.1** and **L.1** – **L.7**. Let $\hat{\theta}_n(g_n)$ be the solution of $S_{g_n}(n,\theta) = 0$ (see **N.9**) under H_0 as well as H_1 (i.e. **G.5.b)**). Then $\hat{\theta}_n(g_n)$ is a $\sqrt{n}$-consistent estimator, i.e*

$$\hat{\theta}_n(g_n) - \theta = O_P\left(\frac{1}{\sqrt{n}}\right).$$

The corresponding test statistic is given as

$$T_n(\eta,\gamma;g_n) = \max_{1\le k<n} w_{\eta,\gamma;g_n}(k,n)\frac{1}{\sqrt{n}}\left\|S_{g_n}(k,\hat{\theta}_n(g_n))\right\|_{\Sigma_{g_n}^{-1}}. \tag{3.8}$$

At first we are going to analyse the behaviour of the test statistic $T_n(\eta,\gamma;g)$. Then we show that also for $T_n(\eta,\gamma;g_n)$ the results hold true.

3.1.2. Asymptotics for randomly weighted change-point tests

We constructed test statistics with weight functions depending on the observations. For these weight functions we are going to show that the test statistics still fulfil the asymptotics as in the case of i.i.d. errors with finite second moment.

To handle the influence of the estimator $\hat{\theta}_n$ used for $g_n(x)$ we make the following assumptions.

L.8 Let $\hat{\theta}_n$ be the solution of

$$\sum_{t=1}^{n}\psi(X_t;\theta) = 0 \tag{3.9}$$

with ψ being dominated integrable and 2-times differentiable with dominated integrable derivatives and $\mathbb{E}[\nabla\psi(X_1^{(z)},\theta)] \neq 0$.

Theorem 3.1.3 *Assume **L.1** and let η, γ fulfil **G.2**. Under H_0, we have*

*a) if **L.2** and **L.3** hold, then*

$$T_n(\eta,\gamma;g) = \max_{1\le k<n} w_{\eta,\gamma;g}(k,n)\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k} g(X_t)X_t - \left(\sum_{t=1}^{k} g(X_t)\right)\hat{\theta}_n(g)\right\|_{\Sigma_g^{-1}}$$
$$\xrightarrow{d} \sup_{s\in(\eta,1-\eta)} \frac{\|B(s)\|}{(s(1-s))^{\gamma}},$$

with $w_{\eta,\gamma;g}(k,n) = \mathbf{1}_{\{\eta n<k<(1-\eta)n\}}\left(\frac{\left(\sum_{t=1}^{n} g(X_t)\right)^2}{\sum_{t=1}^{k} g(X_t)\left(\sum_{t=1}^{n} g(X_t)-\sum_{t=1}^{k} g(X_t)\right)}\right)^{\gamma}.$

*b) if **L.2** – **L.8** hold, then*

$$T_n(\eta,\gamma;g_n) = \max_{1\le k<n} w_{\eta,\gamma;g_n}(k,n)\frac{1}{\sqrt{n}}\left\|S_{g_n}(k,\hat{\theta}_n(g_n))\right\|_{\Sigma_{g_n}^{-1}}$$
$$\xrightarrow{d} \sup_{s\in(\eta,1-\eta)} \frac{\|B(s)\|}{(s(1-s))^{\gamma}},$$

with $\Sigma_{g_n} = \mathbb{V}\mathrm{ar}[g_n(X_1)\varepsilon_1]$ *and* $S_{g_n}(k,\cdot)$, $w_{\eta,\gamma;g_n}(k,n)$ *as in **N.9**, **N.10**.*

For a) the important part is that we have the *a.s.* convergence against the deterministic weight function $w(\eta, \gamma)$ from before. Then the rest follows straight forward as in the proof of Theorem 2.2.1 . To show b) we use the result a). Therefore, we show that asymptotically there is no difference between these two test statistics.

As before we do not only guarantee a level-α test but also an asymptotic power one test.

Theorem 3.1.4 *Let $\{B(s)\}$ be a standard Brownian bridge. Assume **L.1**, **G.1** and let η, γ fulfil **G.2** . Under H_1, we have:*

*a) if **L.2** and **L.3** hold true and δ_n fulfilling either **G.5.a)** or **G.5.b)**, then (with $A = \Sigma_g^{-1}$)*

$$T_n(\eta, \gamma; g) \xrightarrow{p} \infty\,.$$

*b) if **L.2** – **L.7** hold true and δ_n fulfilling **G.5.b)**, then (with $A = \Sigma_{g_n}^{-1}$)*

$$T_n(\eta, \gamma; g_n) \xrightarrow{p} \infty\,.$$

In section 2.2.3 we have used some properties of the considered deterministic weight function to prove the asymptotics. Specifically, we used results based on (1.4) and (1.5). As we are only interested in determining sufficient conditions, we combine those with the results here and derive regularity conditions for the weight functions. This will allow us to choose weight functions independently of the considered test statistic. Details are given in section 3.3.

3.1.3. Proofs

Theorem 3.1.1

Assume **G.1**, **L.1** – **L.3**. Let $\hat{\theta}_n$ be the solution of $S_g(n, \theta) = 0$ (see **N.9**) under H_0 as well as under H_1. If δ_n fulfils either **G.5.a)** or **G.5.b)** and $\delta = \lim_{n\to\infty} \delta_n$, then we have:

a) $\hat{\theta}_n(g)$ is a consistent estimator, i.e

$$\hat{\theta}_n(g) \xrightarrow{a.s} \tilde{\theta}\,,$$

where $\tilde{\theta} = \theta$ under H_0 and under H_1 $\tilde{\theta} = \theta + (1-\lambda)(c_1\delta + c_2)$ with c_1 and c_2 some constants depending on g and ε.

b)

$$\sqrt{n}(\hat{\theta}_n(g) - \tilde{\theta}) \xrightarrow{d} \mathcal{N}(0, V)\,,$$

with $\tilde{\theta}$ as in a) and V the covariance matrix depending on $g(X_1^{(z)})\varepsilon_1$ as well as $g(X_1^{(z)})$, $z = 1, 2$.

Proof:
First we show a).
From assumption **L.1** it follows that $g(X_i)$ and $g(X_t)X_t$, $t = 1, \ldots, n$, are i.i.d. random variables under H_0. So, the sums of these random variables fulfil a strong LLN (for strong

mixing Theorem C.2.1). From assumption **L.2** and **L.3** together with Slutsky C.1.4 we have under H_0

$$\frac{\sum_{t=1}^n g(X_t)X_t}{\sum_{t=1}^n g(X_t)} \xrightarrow{a.s} \frac{\mathbb{E}[g(X_1)X_1]}{\mathbb{E}[g(X_1)]} = \theta \,.$$

Analogue to the proof of Theorem 2.1.1 we can prove consistency under H_1 by splitting the sums at m. Then we have

$$\frac{\sum_{t=1}^n g(X_t)X_t}{\sum_{t=1}^n g(X_t)} = \theta + \frac{\sum_{t=m+1}^n g(X_t)\delta_n}{\sum_{t=1}^m g(X_t) + \sum_{t=m+1}^n g(X_t)} + \frac{\sum_{t=1}^m g(X_t)\varepsilon_t + \sum_{t=m+1}^n g(X_t)\varepsilon_t}{\sum_{t=1}^m g(X_t) + \sum_{t=m+1}^n g(X_t)} \quad (3.10)$$

Applying the LLN for each part of the sums finishes the proof.

Now, we prove b) in two steps.
First we assume H_0. The difference $\hat{\theta}_n(g) - \tilde{\theta}$ can be written (using the model (3.1)) as a fraction of sums of i.i.d. positive random variables with finite second moment, i.e.

$$\frac{\sum_{t=1}^n g(X_t)X_t}{\sum_{t=1}^n g(X_t)} - \theta = \frac{\sum_{t=1}^n g(X_t)\varepsilon_t}{\sum_{t=1}^n g(X_t)} \,.$$

Applying the CLT for a sum of i.i.d. random variables in the numerator and the LLN in the denominator yields the claim together with Slutsky, see Theorem C.1.4.

The proof under H_1 uses that for X_t, $t = 1, \ldots, n$, i.i.d. random variables it holds

$$\frac{1}{\frac{1}{n}\sum_{t=1}^n X_t} - \frac{1}{\mathbb{E}[X_1]} = O_P\left(\frac{1}{\sqrt{n}}\right) .$$

The decompotion (3.10) and analogous arguments as in the proof of Theorem 2.1.2 finish the proof.

■

Lemma 3.1.1

Let g_n and g fulfil the assumption **L.6**. Then for the model (3.1) it holds under H_0 and H_1

$$\frac{1}{n}\sum_{t=1}^n (g_n(X_t) - g(X_t))(1 + \varepsilon_t) = o_{a.s.}(1) \,.$$

Proof:
We proof this in 2 steps.

First observe that it is enough to show $\frac{1}{n}\sum_{t=1}^n |g_n(X_t) - g(X_t)| = o_{a.s.}(1)$. Let

$$A := \bigcap_{t=1}^{\infty} \{\omega | \lim_{n\to\infty} g_n(X_t) = g(X_t)\} \,.$$

From the assumption **L.6** we have $P(A) = 1$. Then for all $\omega \in A$ and all ϵ there exists $n_0(\epsilon, \omega)$ such that for all t and all $n \geq n_0(\epsilon, \omega)$ it holds

$$|g_n(X_t(\omega)) - g(X_t(\omega))| \leq \epsilon \,.$$

Then we have

$$\frac{1}{n}\sum_{t=1}^{n}|g_n(X_t(\omega)) - g(X_t(\omega))| \leq \frac{1}{n}\sum_{t=1}^{n_0(\epsilon,\omega)}|g_n(X_t(\omega)) - g(X_t(\omega))| + \frac{n-n_0(\epsilon,\omega)}{n}\epsilon\,,$$

where the weighted sum of the differences on the left hand side converges to 0 as it is a fixed sum and $n \to \infty$ and $\frac{n-n_0(\epsilon,\omega)}{n} \underset{n\to\infty}{\longrightarrow} 1$. Then we can define the sets

$$B_\epsilon := \left\{\omega \middle| \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}|g_n(X_t(\omega)) - g(X_t(\omega))| < \epsilon\right\}$$

and

$$B_\epsilon \underset{\epsilon\to 0}{\longrightarrow} B = \left\{\omega \middle| \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}|g_n(X_t(\omega)) - g(X_t(\omega))| = 0\right\}.$$

As for all $\omega \in A$ it holds $\omega \in B_\epsilon$, we have $A \subseteq B$. This implies $P(B) = 1$.

Now define $A\prime = \{\omega|\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t \underset{n\to\infty}{\longrightarrow} 0\}$, which has probability 1 due to the sLLN. Because $P(A\cap A\prime) = P(A) + P(A\prime) - P(A\cup A\prime) = 1$, we can consider an equivalent argumentation as above and gain the result.

■

Theorem 3.1.2

Assume **G.1** and **L.1** – **L.7**. Let $\hat{\theta}_n(g_n)$ be the solution of $S_{g_n}(n,\theta) = 0$ (see **N.9**) under H_0 as well as H_1 (i.e. **G.5.b)**). If $\hat{\theta}_n(g_n)$ is a $\sqrt{n}$-consistent estimator, i.e

$$\hat{\theta}_n(g_n) - \theta = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Proof:
With the assumption **L.7**, we have the rate of convergence and can conclude

$$\begin{aligned}
\hat{\theta}_n(g_n) - \theta &= \hat{\theta}_n(g_n) - \hat{\theta}_n(g) + \hat{\theta}_n(g) - \theta \\
&= \frac{\sum_{t=1}^{n}(g_n(X_t) - g(X_t))X_t}{\sum_{t=1}^{n}g_n(X_t)} + \frac{\sum_{t=1}^{n}g(X_t)X_t}{\sum_{t=1}^{n}g(X_t)\ \sum_{t=1}^{n}g_n(X_t)}\left(\sum_{t=1}^{n}g_n(X_t) - \sum_{t=1}^{n}g(X_t)\right) \\
&\quad + O_P(n^{-1/2}) \\
&= O_P(n^{-1/2}) + o_P(n^{-1/2}) + O_P(n^{-1/2})\,.
\end{aligned}$$

We used

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}(g_n(X_t) - g(X_t))X_t \\
&\qquad = \left(\frac{1}{n}\sum_{t=1}^{n}(\nabla g(X_t))X_t\right)(\hat{\theta}_n - \theta) + (\hat{\theta}_n - \theta)^{\mathtt{T}}\left(\frac{1}{n}\sum_{t=1}^{n}(\nabla g(X_t))X_t\right)(\hat{\theta}_n - \theta) \\
&\qquad = O_P(1)O_P(n^{-1/2}) + O_P(n^{-1})\,. \qquad (3.11)
\end{aligned}$$

■

Lemma 3.1.2

Let $\{Z_t\}$ be a sequence of i.i.d. positive random variables with $E[Z_1] = \mu$. For every ϵ exists Δ with $\mu - \Delta > 0$ and $n_0 = n_0(\epsilon, \Delta)$ such that such that $B_{n_0} := \bigcap_{n>n_0} \{|\overline{Z}_n - \mu| < \Delta\}$ fulfils

$$P(B_{n_0}) > 1 - \epsilon\,. \tag{3.12}$$

Proof:
By the strong LLN we have for all $\Delta > 0$

$$P\left(\limsup_{k\geq n} |\overline{Z}_k - \mu| \geq \Delta\right) = 0$$

and it holds

$$P\left(\limsup_{k\geq n} |\overline{Z}_k - \mu| \geq \Delta\right) = \lim_{n\to\infty} P\left(\sup_{k\geq n} |\overline{Z}_k - \mu| \geq \Delta\right)\,.$$

Choose $\Delta < \mu$ and $\epsilon > 0$, then there exists an n_0 such that

$$P\left(\sup_{k\geq n_0} |\overline{Z}_k - \mu| \geq \Delta\right) \leq \epsilon$$

Now, we have

$$\begin{aligned} P(\neg B_{n_0}) &= P\left(\bigcup_{n>n_0} \{|\overline{Z}_n - \mu| \geq \Delta\}\right) \\ &\leq P\left(\{\exists n > n_0 :\ |\overline{Z}_n - \mu| \geq \Delta\}\right) \\ &\leq P\left(\sup_{n\geq n_0} |\overline{Z}_n - \mu| \geq \Delta\right) \\ &\leq \epsilon\,. \end{aligned}$$

■

Lemma 3.1.3

Let $\{Z_t\}$ be a sequence of i.i.d. positive random variables with $\mathbb{E}[Z_1] = \mu < \infty$. For arbitrary but fixed $\tau \leq 1$ let $M_{\tau,n} = \min_{1\leq k\leq \tau n} \overline{Z}_k$. Then for every $\epsilon > 0$ exists $\Delta > 0$ such that for all n and τ we have

$$P\left(\frac{M_{\tau,n}}{\overline{Z}_n} > \Delta\right) > 1 - \epsilon\,.$$

Proof:
Observe that if we have

$$P\left(\frac{M_{1,n}}{\overline{Z}_n} > \Delta\right) > 1 - \epsilon\,, \tag{3.13}$$

then for all $\tau < 1$ we have [†]

$$P\left(\frac{M_{\tau,n}}{\overline{Z}_n} > \delta\right) \geq P\left(\frac{M_{1,n}}{\overline{Z}_n} > \Delta\right) > 1 - \epsilon\,.$$

[†]Since $M_{\tau,n} \geq M_{1,n}$.

So without loss of generality we can assume $\tau = 1$. For easiness of notation we set $M_n := M_{1,n}$.

Let ϵ be arbitrary but fixed. From Lemma 3.1.2 we know there exists $n_0 = n_0(\epsilon_1, \Delta_1)$ such that for every ϵ_1 exists Δ_1 with $\mu - \Delta_1 > 0$ such that $B_{n_0} := \bigcap_{n>n_0}\{|\overline{Z}_n - \mu| < \Delta_1\}$

$$P(B_{n_0}) > 1 - \epsilon_1 \,. \tag{3.14}$$

Let $n \leq n_0$, then the minimum M_n consists of only finitely many a.s. positive random variables. So, for every ϵ_2 there exists Δ_n such that

$$P\left(\frac{M_n}{\overline{Z}_n} > \Delta_n\right) = 1 - \epsilon_2 \,.$$

Define $\Delta_2 = \min_{n \leq n_0}(\Delta_n)$. Then also

$$P\left(\frac{M_n}{\overline{Z}_n} > \Delta_2\right) > 1 - \epsilon_2 \,. \tag{3.15}$$

The question occurs, how it behaves for large n.

Consider $n \geq n_0$, then we have

$$\begin{aligned} P\left(\frac{M_n}{\overline{Z}_n} > \Delta\right) &= P\left(\frac{M_n}{\overline{Z}_n} > \Delta,\, B_{n_0}\right) + P\left(\frac{M_n}{\overline{Z}_n} > \Delta,\, \neg B_{n_0}\right) \\ &\geq P\left(\frac{M_n}{\overline{Z}_n} > \Delta,\, B_{n_0}\right) \,. \end{aligned}$$

From B_{n_0} we have $\overline{Z}_n < \mu + \Delta_1$ a.s. and $M_n \geq \min(M_{n_0}, \mu - \Delta_1)$, so we get

$$\begin{aligned} P\left(\frac{M_n}{\overline{Z}_n} > \Delta\right) &\geq P\left(\frac{M_n}{\mu + \delta_1} > \Delta,\, B_{n_0}\right) \\ &\geq P\left(\frac{\min(M_{n_0}, \mu - \Delta_1)}{\mu + \Delta_1} > \delta,\, B_{n_0}\right) \\ &\geq 1 - P\left(\min(M_{n_0}, \mu - \Delta_1) < \delta(\mu + \Delta_1)\right) - P(\neg B_{n_0}) \,. \end{aligned}$$

We can choose $\Delta < \frac{\mu - \Delta_1}{\mu + \Delta_1}$, then we have

$$\begin{aligned} P\left(\frac{M_n}{\mu + \Delta_1} > \Delta\right) &\geq 1 - \left(P\left(M_{n_0} < \Delta(\mu + \Delta_1)\right) + P(\neg B_{n_0})\right) \\ &\geq 1 - \epsilon \,. \end{aligned}$$

The last line follows from (3.15) and (3.14) with $\epsilon = \epsilon_1 + \epsilon_2$ and $\Delta \leq \Delta_2/(\mu + \Delta_1)$.

In conclusion for every $\epsilon = \epsilon_1 + \epsilon_2$ we find $0 < \Delta < \min(\Delta_2/(\mu + \Delta_1), \Delta_2, \frac{\mu - \Delta_1}{\mu + \Delta_1})$ such that for all τ and n the claim holds true.

■

Theorem 3.1.3

Let $\{B(s)\}$ be a standard Brownian bridge. Assume **L.1** and let η, γ fulfil **G.2**. Under H_0, we have

a) if **L.2** and **L.3** hold, then

$$T_n(\eta,\gamma;g) = \max_{1\le k<n} w_{\eta,\gamma;g}(k,n)\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k} g(X_t)X_t - \left(\sum_{t=1}^{k} g(X_t)\right)\hat{\theta}_n(g)\right\|_{\Sigma_g^{-1}}$$
$$\xrightarrow{d} \sup_{s\in(\eta,1-\eta)} \frac{\|B(s)\|}{(s(1-s))^{\gamma}},$$

with $w_{\eta,\gamma;g}(k,n) = \mathbf{1}_{\{\eta n<k<(1-\eta)n\}}\left(\frac{\left(\sum_{t=1}^{n} g(X_t)\right)^2}{\sum_{t=1}^{k} g(X_t)\left(\sum_{t=1}^{n} g(X_t)-\sum_{t=1}^{k} g(X_t)\right)}\right)^{\gamma}$.

b) if **L.2** – **L.8** holds, then

$$T_n(\eta,\gamma;g_n) = \max_{1\le k<n} w_{\eta,\gamma;g_n}(k,n)\frac{1}{\sqrt{n}}\left\|S_{g_n}(k,\hat{\theta}_n(g_n))\right\|_{\Sigma_{g_n}^{-1}}$$
$$\xrightarrow{d} \sup_{s\in(\eta,1-\eta)} \frac{\|B(s)\|}{(s(1-s))^{\gamma}},$$

with $\Sigma_{g_n} = \mathbb{V}\mathrm{ar}[g_n(X_1)\varepsilon_1]$ and $S_{g_n}(k,\cdot)$, $w_{\eta,\gamma;g_n}(k,n)$ as in **N.9**, **N.10**.

Proof:
At first we prove a).
Notice, that $\{g(X_t)\varepsilon_t\}$ is a sequence of i.i.d. random variables with expectation zero and finite second moment. With the FCLT we have

$$\left\{s\in[0,1] : \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor} g(X_t)\varepsilon_t\right\} \xrightarrow{d} \{s\in[0,1] : W(s)\}, \tag{3.16}$$

where $\{W(s)\}$ is a Wiener process with covariance matrix Σ_g. The LLN gives us $\frac{1}{n}\sum_{t=1}^{\lfloor ns\rfloor} g(X_t) \xrightarrow{p} s\mathbb{E}[g(X_1)]$ for all $s\in[0,1]$ since $g(X_1)$ are i.i.d. with finite variance. By the Cramer-World device C.1.5 and the tightness we get for the random vector

$$\left\{s\in[0,1] : \begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor} g(X_t)\varepsilon_t\\ \frac{1}{n}\sum_{t=1}^{\lfloor ns\rfloor} g(X_t)\end{pmatrix}\right\} \xrightarrow{d} \left\{s\in[0,1] : \begin{pmatrix}W(s)\\ s\mathbb{E}[g(X_1)]\end{pmatrix}\right\}. \tag{3.17}$$

Observe, that we have

$$\sum_{t=1}^{k} g(X_t)(X_t-\hat{\theta}_n) = \sum_{t=1}^{k} g(\theta+\varepsilon_t)\varepsilon_t - \frac{\sum_{t=1}^{k} g(\theta+\varepsilon_t)}{\sum_{t=1}^{n} g(\theta+\varepsilon_t)}\sum_{t=1}^{n} g(\theta+\varepsilon_t)\varepsilon_t.$$

In the proof of Lemma 2.2.1 we used for $\eta\in(0,\frac{1}{2})$ and $\gamma\in[0,\frac{1}{2}]$ or $\eta=0=\gamma$ that the statistic could be written as a sum over i.i.d. random variables with mean zero. Then the functional central limit theorem together with the continuous mapping theorem finished the proof.

For $\eta\in(0,\frac{1}{2})$ and $\gamma\in[0,\frac{1}{2}]$ or $\eta=0=\gamma$ we use the analogue idea. The equation 3.17

together with the continuous mapping theorem yield the result.

It is left to show the asymptotic distribution for $\eta = 0$ and $\gamma \in (0, \frac{1}{2})$.

As in the proof of Lemma 2.2.2 we show that uniformly in n the values outside a truncated range are asymptotically negligible. To this end we analyse the behaviour for $k \in [1, \tau n]$ and $k \in [(1-\tau)n, n)$.

Let us consider the case $k \in [1, \tau n]$.

Observe that we therefore have to analyse the fraction

$$\begin{aligned}\frac{w_{0,\gamma;g}(k,n)}{w_\gamma(k/n)} &= \left(\overline{g(X)}_n^2 \frac{k(n-k)}{\sum_{t=1}^k g(X_t)\left(\sum_{t=1}^n g(X_t) - \sum_{t=1}^k g(X_t)\right)}\right)^\gamma \\ &= \left(\frac{\overline{g(X)}_n}{\overline{g(X)}_k}\right)^\gamma \left(\frac{\overline{g(X)}_n}{\frac{1}{n-k}\sum_{t=k+1}^n g(X_t)}\right)^\gamma .\end{aligned}$$

First observe that the second fraction on the right hand-side is of the same form as the first one. Let us consider the first fraction.

The random sequence $\{g(X_t)\}$ is by **L.2** a.s. positive sequence of i.i.d random variables. Hence it fulfils the assumptions of Lemma 3.1.3 and we gain

$$\max_{1 \le k \le \tau n} \frac{\overline{g(X)}_n}{\overline{g(X)}_k} = O_P(1)\,.$$

Then for every τ, we have that

$$\begin{aligned}\max_{1 \le k \le \tau n} \frac{w_{0,\gamma;g}(k,n)}{w_\gamma(k/n)} &\le \max_{1 \le k \le \tau n} \left(\frac{\overline{g(X)}_n}{\overline{g(X)}_k}\right)^\gamma \max_{1 \le k \le \tau n} \left(\frac{\overline{g(X)}_n}{\frac{1}{n-k}\sum_{t=k+1}^n g(X_t)}\right)^\gamma \\ &= O_P(1)\end{aligned}$$

Now we have

$$\begin{aligned}&\max_{1 \le k < \tau n} w_{0,\gamma;g}(k,n) \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^k g(\theta+\varepsilon_t)\varepsilon_t - \frac{\sum_{t=1}^k g(\theta+\varepsilon_t)}{\sum_{t=1}^n g(\theta+\varepsilon_t)} \sum_{t=1}^n g(\theta+\varepsilon_t)\varepsilon_t \right\|_{\Sigma_g^{-1}} \\ &\qquad \le \max_{1 \le k < \tau n} \frac{w_{0,\gamma;g}(k,n)}{w_\gamma(k/n)} \left(\max_{1 \le k < \tau n} w_\gamma(k/n) \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^k g(\theta+\varepsilon_t)\varepsilon_t \right\|_{\Sigma_g^{-1}} \right. \\ &\qquad\quad \left. + \max_{1 \le k < \tau n} \frac{\overline{g(X)}_n}{\overline{g(X)}_{n-k,n}} \max_{1 \le k < \tau n} w_\gamma(k/n) \frac{k}{n} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n g(\theta+\varepsilon_t)\varepsilon_t \right\|_{\Sigma_g^{-1}} \right) .\end{aligned}$$

From above we have

$$\max_{1 \le k < \tau n} \frac{w_{0,\gamma;g}(k,n)}{w_\gamma(k/n)} = O_P(1) \qquad \text{and} \qquad \max_{1 \le k < \tau n} \frac{\overline{g(X)}_n}{\overline{g(X)}_{n-k,n}} = O_P(1)\,.$$

As in (2.21) and (2.22), we can conclude

$$\max_{1\le k<\tau n} w_\gamma(k/n)\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k} g(\theta+\varepsilon_t)\varepsilon_t\right\|_{\Sigma_g^{-1}} = O_P(\tau^{\frac{1}{2}-\gamma})$$

$$\max_{1\le k<\tau n} w_\gamma(k/n)\frac{k}{n}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(\theta+\varepsilon_t)\varepsilon_t\right\|_{\Sigma_g^{-1}} = O_P(\tau^{1-\gamma})$$

since this are sums of i.i.d. random variables. Finally, we have uniformly in n for $\tau \to 0$

$$\max_{1\le k<\tau n} w_{0,\gamma;g}(k,n)\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k} g(\theta+\varepsilon_t)\varepsilon_t - \frac{\sum_{t=1}^{k} g(\theta+\varepsilon_t)}{\sum_{t=1}^{n} g(\theta+\varepsilon_t)}\sum_{t=1}^{n} g(\theta+\varepsilon_t)\varepsilon_t\right\|_{\Sigma_g^{-1}} = o_P(1)\,.$$

For $\tau \to 0$ this means this part is (uniformly in n) negligible.

In the other case, i.e. $k \in [(1-\tau)n, n)$, the result follows with analogue arguments. Exchanging limits is therefore allowed, which finishes for a).

Now, we show b).
Observe, from assumption **L.6** we have from Lemma 3.1.1 that

$$\frac{1}{n}\sum_{t=1}^{\lfloor ns\rfloor} g_n(X_t) \xrightarrow{a.s} s\mathbb{E}[g(X_1)]\,. \tag{3.18}$$

With

$$C(\theta) = \lambda\mathbb{E}[\nabla g(X_1^{(1)})\varepsilon_1]\mathbb{E}^{-1}[\nabla\psi(X_1^{(1)},\theta)] + (1-\lambda)\mathbb{E}[\nabla g(X_1^{(2)})\varepsilon_1]\mathbb{E}^{-1}[\nabla\psi(X_1^{(2)},\theta)]$$

on the other hand, we can determine

$$\begin{aligned}\left(\sum_{t=1}^{\lfloor ns\rfloor} g_n(X_t)\varepsilon_t - \sum_{t=1}^{\lfloor ns\rfloor} g(X_t)\varepsilon_t + \frac{\lfloor ns\rfloor}{n}C(\theta)\sum_{t=1}^{n}\psi(X_t,\theta)\right)&\\ = \Bigg(\sum_{t=1}^{\lfloor ns\rfloor}\nabla g(X_t,\theta)\varepsilon_t(\hat\theta_n-\theta) + \sum_{t=1}^{\lfloor ns\rfloor}(\hat\theta_n-\theta)^{\mathsf T}\nabla^2 g(X_t,\theta)(\hat\theta_n-\theta)\varepsilon_t&\\ + \frac{\lfloor ns\rfloor}{n}C(\theta)\sum_{t=1}^{n}(\psi(X_t,\theta)-\psi(X_t,\hat\theta_n))\Bigg)\,.&\end{aligned}$$

Centring each sum leads to

$$\begin{aligned}\left(\sum_{t=1}^{\lfloor ns\rfloor} g_n(X_t)\varepsilon_t - \sum_{t=1}^{\lfloor ns\rfloor} g(X_t)\varepsilon_t + \frac{\lfloor ns\rfloor}{n} C(\theta)\sum_{t=1}^{n}\psi(X_t,\theta)\right)&\\ = \sum_{t=1}^{\lfloor ns\rfloor}\left(\nabla g(X_t,\theta)\varepsilon_t - \mathbb{E}[\nabla g(X_1,\theta)\varepsilon_1]\right)(\hat\theta_n-\theta)&\\ + \frac{\lfloor ns\rfloor}{n} C(\theta)\sum_{t=1}^{n}\left(\nabla\psi(X_t,\theta) - \mathbb{E}[\nabla\psi(X_1,\theta)]\right)(\hat\theta_n-\theta)&\\ - \lfloor ns\rfloor\left(\mathbb{E}[\nabla g(X_1,\theta)\varepsilon_1] + C(\theta)\mathbb{E}[\nabla\psi(X_1,\theta)]\right)(\hat\theta_n-\theta)&\\ - (\hat\theta_n-\theta)^{\mathsf T}\sum_{t=1}^{\lfloor ns\rfloor}\nabla^2 g(X_t,\xi)\varepsilon_t(\hat\theta_n-\theta)&\\ + \frac{\lfloor ns\rfloor}{n} C(\theta)(\hat\theta_n-\theta)^{\mathsf T}\sum_{t=1}^{n}\nabla^2\psi(X_t,\xi)(\hat\theta_n-\theta)&\\ = A_1(\hat\theta_n-\theta) + A_2(\hat\theta_n-\theta) + A_3(\hat\theta_n-\theta)&\\ + (\hat\theta_n-\theta)^{\mathsf T}A_4(\hat\theta_n-\theta) + (\hat\theta_n-\theta)^{\mathsf T}A_5(\hat\theta_n-\theta)\,.&\end{aligned}$$

Observe that A_1, A_2 fulfil the assumptions of the LIL. As we have assumed that $\hat\theta_n$ is a $\sqrt{n}$-consistent estimator, ξ is is for large enough n in a compact area around θ. Then we get

$$\mathbb{E}[\sup_{\xi\in U_\theta}\|\psi(X_t,\xi)\|] < \infty\,, \tag{3.19}$$

and so we have a ULLN (uniform law of large numbers, see Theorem C.1.1). Hence, we get

$$\begin{aligned}A_1(\hat\theta_n-\theta) + A_2(\hat\theta_n-\theta) + (\hat\theta_n-\theta)^{\mathsf T}A_4(\hat\theta_n-\theta) + (\hat\theta_n-\theta)^{\mathsf T}A_5(\hat\theta_n-\theta)&\\ = O_P(\sqrt{\log\log n}) + O_P(\sqrt{\log\log n}) + O_P(1) + O_P(1)\,.&\end{aligned}$$

It is left to analyse the term with A_3. Choosing $C(\theta) = -\mathbb{E}[\nabla g(X_1,\theta)\varepsilon_1]\mathbb{E}^{-1}[\nabla\psi(X_1,\theta)]$ (assumption **L.8**) gives us $o_P(1)$ for $A_3(\hat\theta_n-\theta)$. In conclusion we have

$$\sup_{s\in(0,1)}\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor ns\rfloor} g_n(X_t)\varepsilon_t - \sum_{t=1}^{\lfloor ns\rfloor} g(X_t)\varepsilon_t + \frac{\lfloor ns\rfloor}{n} C(\theta)\sum_{t=1}^{n}\psi(X_t,\theta)\right) = o_P(1). \tag{3.20}$$

Hence, we only have to analyse the asymptotic distribution of

$$\sum_{t=1}^{\lfloor ns\rfloor} g(X_t)\varepsilon_t + \frac{\lfloor ns\rfloor}{n} C(\theta)\sum_{t=1}^{n}\psi(X_t,\theta)\,.$$

With (3.11) and the FCLT we can conclude

$$\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor} g_n(X_t),\ s\in(0,1)\right\} \xrightarrow{d} \{W(s) - sW_\psi(1),\ s\in(0,1)\}\,,$$

where $\{W_\psi(s)\}$ is the Wiener process $\frac{1}{\sqrt{n}}\{C(\theta)\sum_{t=1}^{\lfloor sn\rfloor}\psi(X_t,\theta)$ is converging to and $\{W(s)\}$ the Wiener process from (3.16). So, we get

$$\max_{1\le k<n} w_{0,\gamma;g_n}(k,n)\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k} g_n(\theta+\varepsilon_t)\varepsilon_t - \frac{\sum_{t=1}^{k} g_n(\theta+\varepsilon_t)}{\sum_{t=1}^{n} g_n(\theta+\varepsilon_t)}\sum_{t=1}^{n} g_n(\theta+\varepsilon_t)\varepsilon_t\right\|_{\Sigma_{g_n}^{-1}}$$
$$\xrightarrow{d} \sup_{s\in(\eta,1-\eta)} w_\gamma(s)\left\|W(s) - sW_\psi(1) - s(W(1)-W_\psi(1))\right\| ,$$

which gives the claim.

For $\eta = 0$ and $\gamma \in (0,\frac{1}{2})$ we first observe that Lemma 3.1.3 holds true for sums over g_n. Moreover, the Hájek-Rényi inequality is also fulfilled, after replacing. Therefore, analogue arguments as in the proof of a) yield the results. ■

Theorem 3.1.4

Assume **L.1**, **G.1** and let η, γ fulfil **G.2** . Under H_1, we have:

a) if **L.2** and **L.3** hold true and δ_n fulfilling either **G.5.a)** or **G.5.b)**, then (with $A = \Sigma_g^{-1}$)
$$T_n(\eta,\gamma;g) \xrightarrow{p} \infty .$$

b) if **L.2** – **L.7** hold true and δ_n fulfilling **G.5.b)**, then (with $A = \Sigma_{g_n}^{-1}$)
$$T_n(\eta,\gamma;g_n) \xrightarrow{p} \infty .$$

Proof:
Consider the claim a).
First observe that by assumption **L.2** we have
$$w_{\eta,\gamma;g}(k,n) \ge 1 .$$
Equivalently to the proof of Theorem 2.2.2, we see analysing $\left\|S_m(\hat\theta_n;g)\right\|_{\Sigma_g^{-1}}$ is enough. From model (3.1), we get

$$\frac{1}{\sqrt{n}}\left\|S_m(\hat\theta_n;g)\right\|_{\Sigma_g^{-1}} \ge \left|\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{m}(g(X_t)\varepsilon_t - \overline{g(X_t)\varepsilon}_n)\right\|_{\Sigma_g^{-1}} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(X_t)w_{0,1;g}(m,n)\|\delta_n\|_{\Sigma_g^{-1}}\right| .$$

The key in the proof of Theorem 2.2.2 was the CLT applied to the random sum. The $\{g(X_t)\varepsilon_t\}$ are i.i.d. with finite variance (assumptions **L.1** and **L.2**), so the CLT is also fulfilled. Moreover, we have the a.s. convergence of $w_{0,1;g}(m,n)$ and $\overline{g(X)}_n$. Thus, we conclude
$$\left\|S_m(\hat\theta_n;g)\right\| = \left|O_P(1) - O_P(\sqrt{n}\|\delta_n\|_{\Sigma_g^{-1}})\right| \xrightarrow{p} \infty \qquad \text{for } n\to\infty .$$
Due to the strong LLN we have with the equivalent inequalities as in the proof of Theorem 2.2.2.

The proof of b) goes equivalently.
Assumptions **L.1** and **L.6** yield that $\sum_{t=1}^{k} g_n(X_t)\varepsilon_t$, $k = m, n$, fulfil a CLT and $w_{0,1;g_n}(m/n) = O_P(1)$, $\overline{g_n(X)}_n = O_P(1)$. Then the proof follows the ideas of the proof of a). ■

3.2. Non-linear (auto-)regressive processes and neural network functions

Let $X_1, \ldots, X_N$ be the observations of a non-linear autoregressive process of order $p > 0$. It is only possible to detect changes in the observations $X_{p+1}, \ldots, X_N$. We denote with n the number of observations in which the change is detectable, i.e. $X_{-p+1}, \ldots, X_0, X_1, \ldots, X_n$ with $n = N - p$.

The considered time series model with a change after an unknown time point $1 \leq m = m(n) \leq n$ is given as

$$X_t = \begin{cases} g_1(\mathbb{X}_t) + \varepsilon_t & 1 \leq t \leq m\,, \\ g_2(\mathbb{X}_t) + \varepsilon_t & m < t \leq n\,, \end{cases} \tag{3.21}$$

where g_1 and g_2 are some functions, such that $\{X_t,\ t \leq m\}$ and $\{X_t,\ t > m\}$ differ in distribution, $\mathbb{X}_t = (X_{t-1}, \ldots, X_{t-p})$ is the regression vector and ε_t is a sequence of i.i.d. zero mean random variables with $2 + \phi$ moments ($\phi > 0$). The unknown parameter $m = \lfloor \lambda n \rfloor$, $\lambda \in (0, 1)$, is called the change-point if $m < n$. For $m = n$ no change occurs.

After the change-point the new time-series X_t with autoregression function g_2 has starting values not from the stationary distribution. Therefore, one needs some more assumptions for this time series. We assume X_t to be α-mixing with polynomial rate for (auto-)regression functions g_1 and g_2.

Neural networks have a universal approximation property, i.e. a large class of functions can be approximated by a neural network to any degree of accuracy (Hornik et al. [1989]). This motivates to overcome the problem of the unknown regression function by approximating with neural networks. Under H_0, the unknown regression function is approximated by the neural network

$$f(x, \theta) = \nu_0 + \sum_{i=1}^{H} \nu_i \psi(< a_i, x > + b_i)$$

with $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{ip})^{\intercal}$, $\theta = (\nu_0, \ldots, \nu_H, a_1, \ldots, a_H, b_1, \ldots, b_H) \in \Theta$ compact and $\Theta \subset \mathbb{R}^{(2+p)H+1}$. The ψ is assumed to be a sigmoid function with

$$\psi(x) = 1 - \psi(-x) \qquad \lim_{x \to \infty} \psi(x) = 1 \qquad \lim_{x \to -\infty} \psi(x) = 0\,.$$

Observe, that θ is identifiable up to permutations, if the network is not redundant (see section 3.2.2). It is necessary to assume that the neural network parameters, approximating each function, are distinguishable, i.e.

$$g_1(x) \approx f(x, \theta_1) \qquad g_2(x) \approx f(x, \theta_2)$$

with $\theta_1 \neq \theta_2$ (not of the same equivalence class). Therefore, we have the following model

$$X_t = \begin{cases} f(\mathbb{X}_t, \theta_1) + e_t & t \leq m\,, \\ f(\mathbb{X}_t, \theta_2) + e_t & t > m\,. \end{cases}$$

If we assume X_t to be a sequence of α-mixing random variables (with polynomial rate), then invariance principles hold true and we are able to derive asymptotic results. We want to mention, that in the correctly specified case the residuals e_t form a sequence of i.i.d. random variables with e_t independent of $\{X_s,\ s < t\}$.

A technical problem occurs if the time-series after the change-point does not start in the stationary distribution. To overcome this problem we use the quasi model, i.e.

$$X_t = \begin{cases} X_t^{(1)} = g_1(\mathbb{X}_t^{(1)}) + \varepsilon_t & t \leq m\,, \\ X_t^{(2)} = g_2(\mathbb{X}_t^{(2)}) + \varepsilon_t & t > m\,, \end{cases}$$

where $\{X_t^{(1)}\}$ and $\{X_t^{(2)}\}$ are 2 independent time series, which differ in distribution. Then the quasi neural network model is given by

$$X_t = \begin{cases} f(\mathbb{X}_t^{(1)}, \theta_1) + e_t^{(1)} & t \leq m\,, \\ f(\mathbb{X}_t^{(2)}, \theta_2) + e_t^{(2)} & t > m\,, \end{cases}$$

where $\{e_t^{(1)}\}$, $\{e_t^{(2)}\}$ are sequences of zero mean, stationary and α-mixing random variables with polynomial rate, but the time series before ($\{X_t^{(1)}\}$) and after ($\{X_t^{(2)}\}$) the change point do not coincide.

Observe that under correct specification it the $\{e_t^{(1)}\}$ and $\{e_t^{(2)}\}$ become $\{\varepsilon_t\}$ from the original model. In this case some of the assumptions can be relaxed. We are going to analyse the behaviour for the misspecified situation.

In Kirch and Tadjuidje Kamgaing [2012] tests for change-points for these models are introduced. Based on the sample residuals they introduced, besides other statistics, the test statistic

$$T_{n2} = \max_{1 \leq k < n} \frac{1}{\sqrt{n}\, q(\frac{k}{n})} \left| \sum_{i=1}^{k} (X_i - f(\mathbb{X}_i, \hat{\theta}_n)) \right|\,,$$

where q is a weight function defined on $(0, 1)$. They assumed that the weight function q belongs to the class

$$\begin{aligned} Q_{0,1} = \{q:\ & q \text{ is non-decreasing in a neighborhood of 0, non-increasing in a} \\ & \text{neighbourhood of 1 and } \inf_{\eta \leq t \leq 1-\eta} q(t) > 0 \text{ for all } 0 < \eta < \frac{1}{2}\}\,. \end{aligned}$$

The convergence of the test statistic is then based on results of Csörgő and Horváth [1993]. To this end, an additional integral condition has to be fulfilled. We discuss the assumptions and the relation to our weight function in section 3.3. Then, this kind of test statistic is covered by the tests we are considering. But we are going to generalise this kind of test statistic to a more general model and to be sensitive against different alternatives.

Besides the test statistic Kirch and Tadjuidje Kamgaing [2012] also proved the consistency of the change-point estimator using

$$\hat{m} = \arg\max \left\{ \left| S_H(k; \hat{\theta}_n) \right| : 1 \leq k < n \right\}\,,$$

where $S_H(k;\hat{\theta}_n)$ is the partial sum of estimated residuals using the least-squares estimator $\hat{\theta}_n$, i.e.

$$S_H(k;\hat{\theta}_n) = \sum_{t=1}^{k} \hat{\varepsilon}_t = \sum_{t=1}^{k} \left(X_t - f(\mathbb{X}_t, \hat{\theta}_n) \right) ,$$

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} \left(X_t - f(\mathbb{X}_t, \theta) \right)^2 \tag{3.22}$$

for the sample $X_{-p}, \ldots, X_n$. We are going to analyse the change-point estimator based on the test statistic. For more informations on the change-point estimator we refer to section 4.2.

In the following we use a more general set-up where we combine regression and autoregression model. We show that with a some more assumptions on the observations as well as on the generating function of the neural network we derive equivalent results.

3.2.1. Model

Recall that we specified the situation to observing times series

$$X_t = \begin{cases} g_1(\mathbb{X}_t, Z_t) + \varepsilon_t & t \leq m \\ g_2(\mathbb{X}_t, Z_t) + \varepsilon_t & t > m \end{cases} \tag{3.23}$$

where $Z_t \in \mathbb{R}^d$, possibly random but independent of ε_t, and $1 \leq m = m(n) = \lfloor \lambda n \rfloor \leq n$ ($\lambda \in (0,1)$) is called the change-point. To simplify notation we introduce the following random vector.

N.12 Let $\mathbb{X}_t \in \mathbb{R}^p$ be the autoregression and $Z_t \in \mathbb{R}^d$ the regression vector. Define $Y_t = (\mathbb{X}_t, Z_t) \in \mathbb{R}^{p+d}$ the vector containing in the first p coordinates the autoregression and in the last d coordinates the regression vector.

In the given model 3.23 the regression functions g_1 and g_2 are assumed to be unknown, which motivates us to use approximation via neuronal networks instead. This leads to the model

$$X_t = \begin{cases} X_t^{(1)} = f((Y_t^{(1)}), \theta_1) + \varepsilon_t^{(1)} & t \leq m \\ X_t^{(2)} = f((Y_t^{(2)}), \theta_2) + \varepsilon_t^{(2)} & t > m \end{cases} , \tag{3.24}$$

where $\{\varepsilon_t^{(1)}\}$, $\{\varepsilon_t^{(2)}\}$ are sequences of stationary and α-mixing random variables of polynomial order and θ_1, θ_2 define different neural networks. $\{X_t^{(1)}\}$ and $\{X_t^{(2)}\}$ are 2 independent (strictly) stationary time series, which differ distributionally.

Observe that like in Kirch and Tadjuidje Kamgaing [2012] we restrict our model to be able to make use of results like the invariance principle (Theorem C.2.2) and the Hajek-Renyi inequality (Lemma C.2.4).

L.1 Let $\{X_t^{(1)} : t \in \mathbb{Z}\}$ and $\{X_t^{(2)} : t \in \mathbb{Z}\}$ be stationary time series and $\mathbb{E}\left[\left|X_1^{(1)}\right|^{\upsilon}\right] < \infty$ and $\mathbb{E}\left[\left|X_1^{(2)}\right|^{\upsilon}\right] < \infty$ for some $\upsilon \geq 3$.

L.2 Let $\{X_t^{(1)} : t \in \mathbb{Z}\}$ and $\{X_t^{(2)} : t \in \mathbb{Z}\}$ be independent and each α-mixing with rate $\alpha(j) = o(j^{-c})$, $c > \upsilon/(\upsilon - 2)$ where $\upsilon > 3$.

Observe, that we have to make stronger assumptions on the observed time series than in Kirch and Tadjuidje Kamgaing [2012]. This is caused by the multidimensional test statistic. For the regression parameter $\{Z_t\}$ we make the following assumptions.

L.3 Let $\{Z_t : t \in \mathbb{Z}\}$, $Z_t \in \mathbb{R}^d$, be either a deterministic sequence or a stationary time series with finite third moment and independent of $\{\varepsilon_t : t \in \mathbb{Z}\}$.

For the regression part, we also need to have a third moment condition. If the test statistic contains only the first h derivatives or the last h ones of the neural network function this assumption can be relaxed to the existence of the second moment.

3.2.2. Properties of the neural network estimator

As mentioned we approximate the regression function using non-linear (auto-)regressive models with a one-layer neural network as (auto-)regressive function. The one-layer neural network $f : \mathbb{R}^{p+d} \to \mathbb{R}$ is given by

$$f(y, \theta) = \nu_0 + \sum_{i=1}^{h} \nu_i \psi(< a_i, y > +b_i) \tag{3.25}$$

with $\theta = (\nu_0, \ldots, \nu_h, a_{11}, \ldots, a_{1(p+d)}, a_{21}, \ldots, a_{h(p+d)}, b_1, \ldots, b_h)^T \in \Theta$ compact and $\Theta \subset \mathbb{R}^{(2+p+d)h+1}$. The generating function ψ is assumed to be a sigmoid function, i.e. a continuous function with

$$\psi(y) = 1 - \psi(-y) \quad \lim_{y\to\infty} \psi(y) = 1 \quad \lim_{y\to-\infty} \psi(y) = 0\,.$$

Moreover, we need the following assumption.

L.4 The sigmoid function ψ of the neural network

$$f(y, \theta) = \nu_0 + \sum_{i=1}^{h} \nu_i \psi(< a_i, y > +b_i)$$

has to be 3-times differentiable w.r.t. $\theta = (\nu_0, \ldots, \nu_h, a_{11}, \ldots, a_{1(p+d)}, a_{21}, \ldots, a_{h(p+d)}, b_1, \ldots, b_h)^T$ with bounded derivatives.

Observe, that we need 3-times differentiable but in Kirch and Tadjuidje Kamgaing [2012] 2-times is enough.

For the definition of the limit of the parameter estimator under H_0 as well as under H_1, we assume **G.1**. Then the loss function is defined as

$$\mathbb{E}_\theta = \lambda \mathbb{E}\left(X_1^{(1)} - f(Y_1^{(1)}, \theta)\right)^2 + (1 - \lambda)\mathbb{E}\left(X_1^{(2)} - f(Y_1^{(2)}, \theta)\right)^2 \tag{3.26}$$

for $0 < \lambda \leq 1$. The non-linear least squares estimator of the neural network parameter is derived by minimizing

$$Q_n(\theta) := \sum_{t=1}^{n} (X_t - f(Y_t, \theta))^2 \tag{3.27}$$

with respect to θ, which leads to

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\theta) \tag{3.28}$$

for a suitable compact set Θ.

L.5 Let

$$\mathbb{E}_\theta = \lambda \mathbb{E}\left(X_1^{(1)} - f(Y_1^{(1)}, \theta)\right)^2 + (1-\lambda)\mathbb{E}\left(X_1^{(2)} - f(Y_1^{(2)}, \theta)\right)^2 .$$

There exists $\tilde{\theta} = \arg\min_\theta \mathbb{E}_\theta$, which is the unique minimizer of $\mathbb{E}_\theta$ and lies in the interior of the compact parameter set $\Theta \subset \mathbb{R}^q$, with $q = (p + d + 2)h + 1$.

L.6 Let

$$M := \nabla^2 E_{\tilde{\theta}}$$

be positive definite.

The existence and identifiability of $\tilde{\theta}$ is necessary to derive the asymptotic distribution for the test statistic and later for the change-point estimator. It is known that each neural network is not identifiable in the common sense. So, we have to clarify the term identifiably in the context of neural networks. Following Hwang and Ding [1997] we get for a non-redundant and irreducible network the identifiability of the parameter up to a symmetry transformation and transposition (see Lemma A.1.1). So we can define a equivalence class and define identifiability as identifying the equivalence class.

The equivalence class is defined as follows. Let $\theta \in \Theta$, $\theta = (\nu_0, \mu_1, \ldots, \mu_h)$, with $\mu_i = (\nu_i, \alpha_i, \beta_i)$ for $i = 1, \ldots, h$, then a parameter θ_2 is of the equivalence class of θ if there exists a finite number of transpositions $(\pi_{i,k})$ and symmetry transformations (π_k), i.e. a function π a composite function of $\{\pi_l, \pi_{i,k} | i, k, l = 1, \ldots, h \quad i \neq k\}$ such that $\theta_2 = \pi(\theta)$. For the notation and definitions see section A.1.

For the test statistic this definition of identifiability is enough, because we are interested in determining a possible difference in the regression function. So it is enough to check if $f(x, \theta_1)$ is close to $f(x, \theta_2)$ where θ_1 is the estimator based on the observations before the possible change and θ_2 the one after.

We want to analyse the parameter estimator for the non-linear (auto-)regressive model (3.24). As in the proof of a consistent change-point estimator in the mean change model (proof of Theorem 2.3.1 given in section 2.3.4) we are going to make use of Lemma 3.1 of Pötscher and Prucha [1997] (see Theorem C.1.2). So we first prove the uniform convergence of $\frac{1}{n}Q_n(\cdot)$ against the loss function and then conclude that the estimator is consistent.

Kirch and Tadjuidje Kamgaing [2012] stated in Proposition 1

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} (X_t - f(\mathbb{X}_t, \theta))^2 - E_\theta \right| = o(1) \ a.s. \quad \forall \theta \in \Theta. \tag{3.29}$$

We show the result also holds true in the (auto-)regressive model.

Proposition 3.2.1

Assume assumptions ***L.1***, ***L.3*** *and* ***L.4***. *Let $Q_n(\cdot)$ be as in* (3.27) *and E. as in* (3.26). *Then under the model* (3.24) *and for $\Theta \subset \mathbb{R}^q$ $(q = (p+d+2)h+1)$ we have for H_0 as well as H_1 that for $n \to \infty$*

a)

$$\sup_{\theta\in\Theta} \left| \frac{1}{n} Q_n(\theta) - E_\theta \right| \xrightarrow{a.s} 0$$

b) and

$$\sup_{\theta\in\Theta} \left\| \frac{1}{n} \nabla^2 Q_n(\theta) - \nabla^2 E_\theta \right\| \xrightarrow{a.s} 0\,,$$

where ∇^2 denotes the Hesse matrix with respect to θ.

This is essentially a ULLN for $(X_i - f(Y_i,\theta)^2)$ (a) and $\nabla^2(X_i - f(Y_i,\theta))^2$. To prove this, we have to assume 3-times differentiability of the generating function. In Kirch and Tadjuidje Kamgaing [2012] they assumed 2-times, which for the change-point test is enough if the model is correctly specified. Otherwise, the 3rd moment is necessary.

With this result, the $\sqrt{n}$-consistency follows directly. In Kirch and Tadjuidje Kamgaing [2012] Theorem 2.1 and Theorem 2.2 state the consistency

$$\hat{\theta}_n - \tilde{\theta} = o(1)\ a.s.\,, \tag{3.30}$$

where $\tilde{\theta}$ is given by **L.5** and the rate of convergence

$$\|\hat{\theta}_n - \tilde{\theta}\| = O_p\left(\frac{1}{\sqrt{n}}\right)\,, \tag{3.31}$$

respectively. We derive equivalent results for our model (3.24).

Theorem 3.2.1 *Assume* ***L.1***, ***L.3*** *and* ***L.4***, ***L.5***. *Then*

a) the parameter estimator is strong consistent, i.e.

$$\hat{\theta}_n \xrightarrow{a.s} \tilde{\theta} \qquad \text{as } n \to \infty\,,$$

where $\tilde{\theta}$ as in ***L.5***.

b) If additionally assumptions ***L.2*** *and* ***L.6*** *hold true, then*

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}) \xrightarrow{d} \mathcal{N}(0, M^{-1}VM^{-1})\,,$$

where

$$V = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\nabla Q_n(\tilde{\theta})(\nabla Q_n(\tilde{\theta}))^{\mathtt{T}}\right]$$

and M as in ***L.6***.

Notice, the limit V exists but can be singular.

The proof of the consistency follows directly from Proposition 3.2.1 using Lemma 3.1 of Pötscher and Prucha [1997] (Theorem C.1.2). For the asymptotic normality analogue techniques are used as for the sample mean (see Theorem 2.1.2). With the assumption of the existence for the Hesse matrix the result then follows due to the mixing assumptions and because of the sum structure of the estimation function.

We are going to see, that the normality of the estimator is not important but the rate of convergence is. In the section 3.3.2 this fact is presented clearly.

3.2.3. Asymptotics of the test-statistic

Differently to Kirch and Tadjuidje Kamgaing [2012] we do not observe the sum over the sample residuals, corresponding to the derivative of the least squares function w.r.t. the constant parameter of the neural network (ν_0). As can be seen in the simulations in Kirch and Tadjuidje Kamgaing [2012] for a correctly specified model, the sensitivity against a change in the non-linear parameter can happen to be not as good as for a misspecified model. One idea to overcome this problem is to not necessarily consider the first derivative but to allow for more flexibility by, e.g. considering gradients w.r.t. only a part of the parameter vector. The test statistic is then given as

$$T_n(\eta, \gamma; A) = \max_{1 \le k < n} w_{\eta,\gamma}(k/n) \left\| S_k(\hat{\theta}_n) \right\|_A \tag{3.32}$$

$$S_k(\theta) = \sum_{i=1}^{k} \nabla f(Y_i, \theta)(X_i - f(Y_i, \theta))$$

with Y_t as given in **N.12** and A as in **G.3**. The result follows with analogue argumentation as in Kirch and Tadjuidje Kamgaing [2012]. They first observed that the sample residuals in the partial sums can be replaced with $\zeta_t - \overline{\zeta_n}$, where $\zeta_t = X_t - f(\mathbb{X}_t, \tilde{\theta})$. Then the main assertion followed by standard argumentation.

Analogous, we first show that the replacement in the modified model and for the modified test statistic still holds true. To simplify the notation we introduce the following function.

N.13 Let $q(t, \theta) := \nabla f(Y_t, \theta)(X_t - f(Y_t, \theta))$ for $\theta \in \Theta$.

Theorem 3.2.2 *Assume* ***L.1- L.6*** *and define $q(t, \theta)$ as in* ***N.13***, *$\zeta(t) = q(t, \tilde{\theta})$. Then under H_0 it holds*

$$\max_{1 \le k < n} \frac{n}{k(n-k)} \left\| \sum_{t=1}^{k} (q(t, \hat{\theta}_n) - (\zeta(t) - \bar{\zeta}_n)) \right\|^2 = O_p\left(\frac{\log \log n}{n}\right).$$

Observe that the centering with the sample mean is necessary. This controls the variability due to the estimator. For $k = n$ the sum of $q(t, \hat{\theta}_n)$ is 0 but the sum of ζ_t is not. We are going to use this result to formulate regularity conditions for change-point tests in section 3.3.

With the replacement, determining the asymptotic distribution of T_n follows with equivalent arguments as in Theorem 2.2.1.

Theorem 3.2.3 *Assume **L.1**- **L.6**. Under H_0 the two series*

$$\Gamma_{ij} = \mathbb{E}[Z_{1i}Z_{1j}] + 2\sum_{l\geq 2}^{\infty} \mathbb{E}[Z_{1i}Z_{lj}] + \sum_{l\geq 2}^{\infty} \mathbb{E}[Z_{li}Z_{1j}]\,,$$

with $Z_{1i} = q_i(t, \tilde{\theta}) = \nabla_i f(Y_1, \tilde{\theta})(X_1 - f(Y_1, \tilde{\theta}))$, converge absolutely and we have

$$T_n(\eta, \gamma; A) \xrightarrow{d} \sup_{\eta < s < (1-\eta)} \frac{\|W(s) - sW(1)\|_A}{(s(1-s))^\gamma}\,,$$

where $\{W(t)\}$ is a Wiener process having a covariance matrix $\Gamma = (\Gamma_{ij})_{1\leq i,j\leq q}$.

Observe, that Γ denotes the long-run variance of the residuals. In the multivariate case with $A_{11} = 1$ and $A_{ij} = 0$ for $i + j \neq 2$ we have for the correctly specified model that Γ is equal to the variance of the residuals. In this situation estimation of Γ becomes easier, but in the rest of the cases, as already mentioned in the case of the variance estimator (section 2.5.1), the estimation becomes its own problem.

With the special choice of the matrix A we were able to derive the limit distribution w.r.t. a Brownian bridge. Therefore, we used the square root of the covariance matrix of the residuals which corresponds, due to the model, with the covariance matrix of the Wiener process. Here, the covariance matrix of the Wiener process is the long run variance of the observed process. So we define our decision matrix as follows.

L.7 There exists a matrix A' such that $A = \Gamma^{-\frac{1}{2}} A' \Gamma^{-\frac{1}{2}}$, with Γ as in Theorem 3.2.3.

Corollary 3.2.1 *Under the assumptions of Theorem 3.2.3 and **L.7**, then*

$$T_n(\eta, \gamma; A) \xrightarrow{d} \sup_{0<t<1} \|B(t)\|_{A'}\,,$$

with $\{B(t)\}$ being a standard Brownian bridge in $\mathbb{R}^q$.

In practice the long-run variance of the observed process is not known. From Theorem 2.2.3 we know that replacing it with a consistent estimator will not change the asymptotic results.

Estimating the long-run variance in finite samples leads to estimation problems. In the section 5.1 we are going to discuss this problem and propose a possible solution.

Now we have determined the asymptotic behaviour of T_n under H_0. In the next step we show that the change-point test is a consistent one.

L.8 There exists $c > 0$ such that

$$\left\| \mathbb{E}\left[\nabla f(Y_1^{(1)}, \tilde{\theta})(X_1^{(1)} - f(Y_1^{(1)}, \tilde{\theta}) \right] \right\|_A > c\,.$$

Before we consider the asymptotic behaviour of the change-point test under H_1, we state a useful property of the neural network estimator.

Proposition 3.2.2

Assume the model (3.24) *and the assumptions **L.1**, **L.3**, **L.4**, **L.5** and **L.8**. Under H_1 we have*

$$\tilde{\theta} \neq \theta_z \qquad z = 1, 2\,. \tag{3.33}$$

To prove this proposition, we are going to use, that $\tilde{\theta}$ is also the root of the existing derivative of E_θ. We will use this result also for the change-point test in proving the consistency and especially for the proofs of the asymptotic behaviour of the change-point estimator.

Lemma 3.2.1

*Let assumption **L.1-L.8** hold true. Under H_1 the parameter $\tilde{\theta}$ solves $\nabla E_\theta = 0$, with*

$$\nabla E_\theta = \lambda \mathbb{E}[\nabla f(Y_1^{(1)}, \theta)(X_1^{(1)} - f(Y_1^{(1)}, \theta))] + (1-\lambda)\mathbb{E}[\nabla f(Y_1^{(2)}, \theta)(X_1^{(2)} - f(Y_1^{(2)}, \theta))]\,.$$

This Lemma is going to be proven using the properties of the derivative of neural networks as well as the Dominated Convergence Theorem C.1.3.

With the Proposition 3.2.2 and Lemma 3.2.1 we are able to prove the consistency of the test. The proof follows essentially the same idea as the proof of Theorem 2.2.2 but requires higher technically effort due to the neural network.

Theorem 3.2.4 *Let assumptions **L.1**, **L.3**, **L.4**, **L.5** and **L.8** hold. If $\eta < \lambda < (1-\eta)$, then under H_1*

$$T_n(\eta, \gamma; A) \xrightarrow{p} \infty\,.$$

The main idea is to take a look at the sum up to m and analyse this behaviour. Then T_n is bounded from below by this value.

For the boundedness from below one can use the ULLN (Theorem C.1.1)but also the UCLT (uniform central limit theorem) as shown in Kirch and Tadjuidje Kamgaing [2012] and Kirch and Kamgaing [2014], respectively. Here we have the derivative of the neural network function within the test statistic, thus we follow the argumentation with the ULLN to show the consistency. Now, Lemma 3.2.1 helps to show the consistency equivalently to the proof of Theorem 2.2.2.

3.2.4. Proofs

For all the proofs we use a following of the neural network properties which guarantees the existence of the uniform LLN.

Lemma 3.2.2

*Assume assumptions **L.1**, **L.3** and **L.4**. Then we have*

$$\begin{aligned}
\mathbb{E}[\sup_{\theta\in\Theta} \|\nabla f(Y_t,\theta)\|] &< \infty \\
\mathbb{E}[\sup_{\theta\in\Theta} \left\|\nabla^2 f(Y_t,\theta)\right\|] &< \infty \\
\mathbb{E}[\sup_{\theta\in\Theta} \left\|\nabla^2 (\nabla_i f(Y_t,\theta))\right\|] &< \infty \qquad i=1,\ldots,p+d\,.
\end{aligned}$$

Proof:
In Lemma A.1.2 the derivatives of the neural network functions are calculated and approximated under the assumption Θ is compact. From this we gain there exists constants c_1, c_2 and c_3 such that

$$\begin{aligned}
\mathbb{E}[\sup_{\theta\in\Theta} \|\nabla f(Y_t,\theta)\|] &\le c_1 \mathbb{E}[\max_{j=1,\ldots,p+d} |Y_{t-j}|] \\
\mathbb{E}[\sup_{\theta\in\Theta} \left\|\nabla^2 f(Y_t,\theta)\right\|] &\le c_2 \mathbb{E}[\max_{j=1,\ldots,p+d} Y_{t-j}^2] \\
\mathbb{E}[\sum_{i=1}^{p+d} \sup_{\theta\in\Theta} \left\|\nabla^2 (\nabla_i f(Y_t,\theta))\right\|] &\le c_3 \mathbb{E}[\max_{j=1,\ldots,p+d} |Y_{t-j}^3|]\,.
\end{aligned}$$

With the definition of Y_t **N.12** we get the existence of the moments due to the assumptions **L.2** and **L.3**. ∎

Corollary 3.2.2 *Under the assumptions of Lemma 3.2.2 we have*

$$\begin{aligned}
\mathbb{E}[\sup_{\theta\in\Theta} \|\nabla f(Y_t,\theta)(X_t - f(Y_t,\theta))\|] &< \infty \\
\mathbb{E}[\sup_{\theta\in\Theta} \left\|\nabla^2 (X_t - f(Y_t,\theta))^2\right\|] &< \infty \\
\mathbb{E}[\sup_{\theta\in\Theta} \left\|\nabla^2 (\nabla_i f(Y_t,\theta)(X_t - f(Y_t,\theta))\right\|] &< \infty
\end{aligned}$$

Proof:
Using the triangle inequality and Lemma 3.2.2 we get

$$\begin{aligned}
\mathbb{E}[\sup_{\theta\in\Theta}\|\nabla f(Y_t,\theta)(X_t - f(Y_t,\theta))\|] &\leq c_1\mathbb{E}[|X_t| \max_{j=1,\dots,p+d}|Y_{t-j}|] + c_2\mathbb{E}[\max_{j=1,\dots,p+d}|Y_{t-j}|] \\
&< \infty\,, \\
\mathbb{E}[\sup_{\theta\in\Theta}\|\nabla(\nabla f(Y_t,\theta)(X_t - f(Y_t,\theta)))\|] &\leq c_3\mathbb{E}[|X_t| \max_{j=1,\dots,p+d}Y_{t-j}^2] + c_4\mathbb{E}[\max_{j=1,\dots,p+d}Y_{t-j}^2] \\
&< \infty\,, \\
\mathbb{E}[\sup_{\theta\in\Theta}\left\|\nabla^2(\nabla_i f(Y_t,\theta)(X_t - f(Y_t,\theta)))\right\|] &\leq c_6\mathbb{E}[|X_t| \max_{j=1,\dots,p+d}|Y_{t-j}^3|] + c_4\mathbb{E}[\max_{j=1,\dots,p+d}|Y_{t-j}^3|] \\
&< \infty\,.
\end{aligned}$$

■

Proposition 3.2.1

Assume assumptions **L.1**, **L.3** and **L.4**. Let $Q_n(\cdot)$ be as in (3.27) and $E.$ as in (3.26). Then under the model (3.24) and for $\Theta \subset \mathbb{R}^q$ ($q = (p+d+2)h+1$) we have for H_0 as well as H_1 that for $n \to \infty$

a)
$$\sup_{\theta\in\Theta}\left|\frac{1}{n}Q_n(\theta) - E_\theta\right| \xrightarrow{a.s} 0$$

b) and
$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\nabla^2 Q_n(\theta) - \nabla^2 E_\theta\right\| \xrightarrow{a.s} 0\,,$$

where ∇^2 denotes the Hesse matrix with respect to θ.

Proof:
The proof is analogous to the one of Proposition 1 in Kirch and Tadjuidje Kamgaing [2012]. Therefore, we only discuss the necessary modifications.

With the additional assumptions on the independent regression vector Z_t we still have for H_0 that $Q_n(\theta)$ and $\nabla^2 Q_n(\theta)$ are a sums of stationary and ergodic processes defined on $C(\Theta,\mathbb{R})$ and $C(\Theta,\mathbb{R}^{(q\times q)})$, respectively. In both cases the functions are of the form $\sum_{i=1}^n v_i(\theta)$ (see (3.27)). Then we use under H_1, that we have

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n} v_i(\theta) - E(\theta)\right\| \\
&\quad\leq \sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor\lambda n\rfloor} v_i(\theta) - \lambda E_1(\theta)\right\| + \sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=\lfloor\lambda n\rfloor+1}^{n} v_i(\theta) - (1-\lambda)E_2(\theta)\right\|
\end{aligned}$$

for every $E(\theta) = \lambda E_1(\theta) + (1-\lambda)E_2(\theta)$. Notice the sums of the mixed parts, i.e. where the function $v_i(\theta)$ depends on the observations before the change-point, are negligible by **L.1** and **L.3**. So it is enough to show the result separately for each part. The separate parts are again sums of stationary ergodic processes. Thus one assumption of the uniform LLN (see Theorem

C.1.1) is fulfilled for the least squares function $Q_n(\theta)$ and the second derivative of the least squares function $\nabla^2 Q_n(\theta)$.

For a) we have from **L.4** that $\sup_{\theta\in\Theta}|f(y,\theta)| \leq D_1$ for some $D_1 > 0$ and $y \in \mathbb{R}^{p+d}$. With the assumptions **L.1** and **L.3** the uniform LLN is applicable under H_0 as well as H_1.

To show b) we use Ranga Rao C.1.1. The stationarity and ergodicity follows from assumptions on X and Z. It is left to show

$$\mathbb{E}[\sup_{\theta\in\Theta}\left\|\nabla^2(X_t - f(Y_t,\theta))^2\right\|] < \infty\,. \tag{3.34}$$

We observe that $\mathbb{E}[\sup_{\theta\in\Theta}\left\|\nabla^2(X_t - f(Y_t,\theta))^2\right\|] = 2\mathbb{E}[\sup_{\theta\in\Theta}\|\nabla(\nabla f(Y_t,\theta)(X_t - f(Y_t,\theta)))\|]$. From Corollary 3.2.2 we get directly 3.34 and from the dominated convergence theorem C.1.3 we can exchange the derivatives and the expectations. This shows $\nabla^2 E_\theta$ is the expectation.

■

Theorem 3.2.1

Assume **L.1**, **L.3** and **L.4**, **L.5**. Then

a) the parameter estimator is strongly consistent, i.e.

$$\hat{\theta}_n \xrightarrow{a.s} \tilde{\theta} \qquad \text{as } n \to \infty\,,$$

where $\tilde{\theta}$ as in **L.5**.

b) If additionally assumptions **L.2** and **L.6** hold true, then

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}) \xrightarrow{d} \mathcal{N}(0, M^{-1}VM^{-1})\,,$$

where

$$V = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\nabla Q_n(\tilde{\theta})(\nabla Q_n(\tilde{\theta}))^{\mathsf{T}}\right]$$

and M as in **L.6**.

Notice, the limit V exists but can be singular.

Proof:

With Proposition 3.2.1 and assumption **L.5** the assumptions of Lemma 3.1 in Pötscher and Prucha [1997] (see Theorem C.1.2) hold true. The assertion a) follows then directly.

The proof of assertion b) is given in two steps. First we prove it for H_0 and secondly for H_1.

Under H_0 we have from Lemma A.1.2 $\sup_\theta\|\nabla f(y,\theta)\| \leq c\max(|x_1|,\ldots,|x_p|,|z_1|,\ldots,|z_d|)$, where $y = (x_1,\ldots,x_p,z_1,\ldots,z_d)$. By **L.1** and **L.3** it follows from the dominated convergence Theorem C.1.1 that we can exchange derivative and expectation. With the definition of $\tilde{\theta}$ (see assumption **L.5**) we know, $\tilde{\theta}$ solves

$$0 = \nabla E_\theta = \mathbb{E}[\nabla(X_1 - f((\mathbb{X}_1, Z_1),\theta))^2]\,.$$

Because $\nabla Q_n(\tilde{\theta})$ is a sum over α-mixing stationary time series, a CLT holds true (see Proposition C.2.1) and we have

$$\frac{1}{\sqrt{n}}\nabla Q_n(\tilde{\theta}) \xrightarrow{d} \mathcal{N}(0, V)\,.$$

From the Taylor expansion there exists a $\theta_n^\star$, with $\left\|\theta_n^\star - \tilde{\theta}\right\| \leq \left\|\hat{\theta}_n - \tilde{\theta}\right\|$ and

$$\begin{aligned} 0 &= \nabla Q_n(\hat{\theta}_n) \\ &= \nabla Q_n(\tilde{\theta}) + (\hat{\theta}_n - \tilde{\theta})\nabla^2 Q_n(\theta_n^\star)\,. \end{aligned}$$

The almost sure consistency, shown in a), provides $\theta_n^\star \xrightarrow{a.s} \tilde{\theta}$. Proposition 3.2.1 assertion b) yields

$$\frac{1}{n}\nabla^2 Q_n(\theta_n^\star) \xrightarrow{a.s} \nabla^2 E_{\tilde{\theta}}\,.$$

With assumption **L.6** the left hand side is an invertible matrix. Putting everything together we have first

$$\nabla Q_n(\tilde{\theta}) = -(\hat{\theta}_n - \tilde{\theta})\nabla^2 Q_n(\theta_n^\star) = -n(\hat{\theta}_n - \tilde{\theta})M + o_p(\sqrt{n})$$

and secondly

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}) = -\frac{1}{\sqrt{n}}\nabla Q_n(\tilde{\theta})M^{-1} \xrightarrow{d} \mathcal{N}(0, M^{-1}VM^{-1})\,.$$

Under H_1 we use the independence of the sums up to m and from $m+1$ till n. Each part is separately a sum over α-mixing stationary time series, so the result follows with the same argumentation. With the linearity of the derivative we can directly get the assertion.

■

Theorem 3.2.2

Assume **L.1- L.6** and define $q(t,\theta)$ as in **N.13**, $\zeta(t) = q(t,\tilde{\theta})$. Then under H_0 it holds

$$\max_{1\leq k<n} \frac{n}{k(n-k)} \left\| \sum_{t=1}^{k} (q(t,\hat{\theta}_n) - (\zeta(t) - \overline{\zeta}_n)) \right\|^2 = O_p\left(\frac{\log\log n}{n}\right)\,.$$

Proof:
From

$$\left\| \sum_{t=1}^{k} (q(t,\hat{\theta}_n) - (\zeta_t - \overline{\zeta}_n)) \right\|^2 = \sum_{i=1}^{q} \left(\sum_{t=1}^{k} (q_i(t,\hat{\theta}_n) - (\zeta_i(t) - \overline{\zeta}_{i,n})) \right)^2$$

we get that it is enough to show

$$\max_{1\leq k<n} \sqrt{\frac{n}{k(n-k)}} \left| \sum_{t=1}^{k} (q_i(t,\hat{\theta}_n) - (\zeta_i(t) - \overline{\zeta}_{in})) \right| = O_p\left(\sqrt{\frac{\log\log n}{n}}\right)$$

for $i = 1, \ldots, q$. The proof follows analogously to the proof of Lemma 3 in Kirch and Tadjuidje Kamgaing [2012], except we have the derivative w.r.t. θ of the neural network. That is the reason for the higher moment condition. We use the possibility to show the proof in detailed steps.

Assume n large enough such that $\hat{\theta}_n \in \Theta$. Because $\sum_{t=1}^{n} q(t,\hat{\theta}_n) = 0$ we have

$$\begin{aligned}
\sum_{t=1}^{k}&(q_i(t,\hat{\theta}_n) - (\zeta_i(t) - \overline{\zeta_{in}})) \\
&= \sum_{t=1}^{k}(q_i(t,\hat{\theta}_n) - \zeta_i(t)) - \frac{k}{n}\sum_{t=1}^{n}(q_i(t,\hat{\theta}_n) - \zeta_i(t)) \\
&= \sum_{t=1}^{k}\left(q_i(t,\hat{\theta}_n) - q_i(t,\tilde{\theta})\right) - \frac{k}{n}\sum_{t=1}^{n}\left(q_i(t,\hat{\theta}_n) - q_i(t,\tilde{\theta})\right) \\
&= \sum_{t=1}^{k}\left(\nabla q_i(t,\tilde{\theta})^T(\hat{\theta}_n - \tilde{\theta})\right) + \frac{1}{2}\sum_{t=1}^{k}\left((\hat{\theta}_n - \tilde{\theta})^T\nabla^2 q_i(t,\xi)(\hat{\theta}_n - \tilde{\theta})\right) \\
&\quad - \frac{k}{n}\sum_{t=1}^{n}\left(\nabla q_i(t,\tilde{\theta})^T(\hat{\theta}_n - \tilde{\theta})\right) - \frac{k}{2n}\sum_{t=1}^{n}\left((\hat{\theta}_n - \tilde{\theta})^T\nabla^2 q_i(t,\xi)(\hat{\theta}_n - \tilde{\theta})\right),
\end{aligned}$$

with ξ elementwise between $\hat{\theta}_n$ and $\tilde{\theta}$. From consistency of $\hat{\theta}_n$ we have $\xi \in \dot{\Theta}$. Using that $\nabla q_i(t,\tilde{\theta})$, for each $i = 1,\dots,q$, is again a stationary α-mixing time series with finite $2+\delta$ moments, the assumptions of the LIL (C.3) are fulfilled. From this and $\|\hat{\theta}_n - \tilde{\theta}\| = O_p(\frac{1}{\sqrt{n}})$ we derive

$$\begin{aligned}
\max_{1\le k<\frac{n}{2}}\frac{1}{\sqrt{k}}&\left|\sum_{t=1}^{k}\left(\nabla q_i(t,\tilde{\theta})^T(\hat{\theta}_n - \tilde{\theta})\right) - \frac{k}{n}\sum_{t=1}^{n}\left(\nabla q_i(t,\tilde{\theta})^T(\hat{\theta}_n - \tilde{\theta})\right)\right| \\
&\le \max_{1\le k<\frac{n}{2}}\left\|\frac{1}{\sqrt{k}}\sum_{t=1}^{k}\left(\nabla q_i(t,\tilde{\theta}) - \mathbb{E}(\nabla q_i(1,\tilde{\theta})\right)\right\|\left\|\hat{\theta}_n - \tilde{\theta}\right\| \\
&\quad + \left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\nabla q_i(1,\tilde{\theta}) - \mathbb{E}(\nabla q_i(1,\tilde{\theta}))\right)\right\|\left\|\hat{\theta}_n - \tilde{\theta}\right\| \\
&= O_p(\sqrt{\log\log n})O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

It is left to show that the parts with the second derivative vanishes faster. We have from Lemma

$$\max_{1\le k\le\frac{n}{2}}\frac{1}{k}\sum_{t=1}^{k}\|\nabla^2 q_i(t,\xi)\|_\infty = \max_{1\le k\le\frac{n}{2}}\frac{1}{k}\sum_{t=1}^{k}\sup_{\theta\in\Theta}\left\|\nabla^2(\nabla_i f(Y_t,\theta))\right\| = O_p(1) \qquad i = 1,\dots,q.$$

Then we get

$$\begin{aligned}
\max_{1\le k\le\frac{n}{2}}\frac{1}{\sqrt{k}}&\left|\sum_{t=1}^{k}\left((\hat{\theta}_n - \tilde{\theta})^{\mathrm{T}}\nabla^2 q_i(t,\xi)(\hat{\theta}_n - \tilde{\theta})\right) - \frac{k}{n}\sum_{t=1}^{n}\left((\hat{\theta}_n - \tilde{\theta})^{\mathrm{T}}\nabla^2 q_i(t,\xi)(\hat{\theta}_n - \tilde{\theta})\right)\right| \\
&\le \sqrt{n}\left\|\hat{\theta}_n - \tilde{\theta}\right\|^2\max_{1\le k\le\frac{n}{2}}\frac{1}{k}\sum_{t=1}^{k}\|\nabla^2 q_i(t,\xi)\|_\infty + \sqrt{n}\left\|\hat{\theta}_n - \tilde{\theta}\right\|^2\frac{1}{n}\sum_{t=1}^{n}\|\nabla^2 q_i(t,\xi)\|_\infty \\
&= O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

From the properties of the stochastic Landau symbols (A.2) we get the claim for the maximum up to $\frac{n}{2}$. For the other part, we observe

$$\max_{\frac{n}{2}\le k<n} \frac{1}{\sqrt{n-k}}\left|\sum_{t=1}^{k}(q_i(t,\hat{\theta}_n)-(\zeta_i(t)-\overline{\zeta_{in}}))\right| = \max_{\frac{n}{2}\le k<n} \frac{1}{\sqrt{n-k}}\left|\sum_{t=k+1}^{n}(q_i(t,\hat{\theta}_n)-(\zeta_i(t)-\overline{\zeta_{in}}))\right|.$$

With the Tailor expansion we get an equivalent splitting except the sums are from $k+1$ up to n. Using the stationarity we get $\{\sum_{t=k+1}^{n}\nabla q_i(t,\tilde{\theta}),\ \frac{n}{2}\le k\le n-1\}$ is distributionally equal to $\{\sum_{t=1}^{l}\nabla q_i(-t,\tilde{\theta}),\ 1\le l\le n-\frac{n}{2}\}$. The mixing property does not change, therefore we can argue analogue as before. Doing the same for the second derivative, we get

$$\begin{aligned}\max_{1\le k<n}\frac{n}{k(n-k)}&\left|\sum_{t=1}^{k}(q_i(t,\hat{\theta}_n)-(\zeta_i(t)-\overline{\zeta}_{i,n}))\right|\\ &= O(1)\max\left\{\max_{1\le k\le\frac{n}{2}}\frac{1}{\sqrt{k}}\left|\sum_{t=1}^{k}(q_i(t,\hat{\theta}_n)-(\zeta_i(t)-\overline{\zeta}_{i,n}))\right|,\right.\\ &\qquad\left.\max_{\frac{n}{2}\le k<n}\frac{1}{\sqrt{n-k}}\left|\sum_{t=1}^{k}(q_i(t,\hat{\theta}_n)-(\zeta_i(t)-\overline{\zeta}_{i,n}))\right|\right\}.\end{aligned}$$

Then the claim follows. ■

Theorem 3.2.3

Assume **L.1- L.6**. Under H_0 the two series

$$\Gamma_{ij}=\mathbb{E}[Z_{1i}Z_{1j}]+2\sum_{l\ge2}^{\infty}\mathbb{E}[Z_{1i}Z_{lj}]+\sum_{l\ge2}\mathbb{E}[Z_{li}Z_{1j}]$$

with $Z_{1i}=q_i(t,\tilde{\theta})=\nabla_i f(Y_1,\tilde{\theta})(X_1-f(Y_1,\tilde{\theta}))$, converges absolutely and we have

$$T_n(\eta,\gamma;A)\xrightarrow{d}\sup_{\eta<s<(1-\eta)}\frac{\|W(s)-sW(1)\|_A}{(s(1-s))^\gamma},$$

where $\{W(t)\}$ is a Wiener process having a covariance matrix $\Gamma=(\Gamma_{ij})_{1\le i,j\le q}$.

Proof:
By Theorem 3.2.2 we have

$$\begin{aligned}\left|\max_{1\le k<n}w(\eta,\gamma)\left\|\sum_{t=1}^{k}q(t,\hat{\theta}_n)\right\|_A-\max_{1\le k<n}w(\eta,\gamma)\left\|\sum_{t=1}^{k}(\zeta(t)-\overline{\zeta}_n)\right\|_A\right|&\\ \le\max_{1\le k<n}w(\eta,\gamma)\left\|\sum_{t=1}^{k}(q(t,\hat{\theta}_n)-(\zeta(t)-\overline{\zeta}_n))\right\|_A&\\ =o_P(1)&\end{aligned}$$

with $q(t,\theta)=\nabla f(Y_t,\theta)(X_t-f(Y_t,\theta))$ and $\zeta(t)=q(t,\tilde{\theta})$. Observing that

$$\max_{1\le k<n}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{k}\zeta(t)-\frac{k}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\zeta(t)\right\|_A=\max_{\frac{1}{n}\le s<1}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor sn\rfloor}\zeta(t)-s\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\zeta(t)\right\|_A$$

and $\zeta(t)$ fulfils a invariance principle for α-mixing processes (see C.2.2). With the invariance principle and the continuous mapping theorem we get

$$\max_{1\le k<n}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{k}\zeta_t-\frac{k}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\zeta_t\right\|_A \xrightarrow{d} \sup_{0<s<1}\|W(s)-s\,W(1)\|_A$$

■

Lemma 3.2.1

Let assumption **L.1-L.8** hold true. Under H_1 the parameter $\tilde{\theta}$ solves $\nabla E_\theta = 0$, with

$$\nabla E_\theta = \lambda\mathbb{E}[\nabla f(Y_1^{(1)},\theta)(X_1^{(1)}-f(Y_1^{(1)},\theta))]+(1-\lambda)\mathbb{E}[\nabla f(Y_1^{(2)},\theta)(X_1^{(2)}-f(Y_1^{(2)},\theta))]\,.$$

Proof:
From the Corollary of the Dominated Convergence Theorem (see Corollary C.1.1) we know the derivative of E_θ exists if we can show that $\|\nabla f(Y_1^{(1)},\theta)(X_1^{(1)}-f(Y_1^{(1)},\theta))\|$ as well as $\|\nabla f(Y_1^{(2)},\theta)(X_1^{(2)}-f(Y_1^{(2)},\theta))\|$ are dominated integrable. This is given in Corollary 3.2.2, which gives the claim.

■

Proposition 3.2.2

Assume the model (3.24) and the assumption **L.1- L.8**. Under H_1 we have

$$\tilde{\theta}\neq\theta_z \qquad z=1,2\,. \tag{3.35}$$

Proof:
Assume $\tilde{\theta}=\theta_1$, i.e. $\mathbb{E}\left[\nabla f(Y_1^{(1)},\tilde{\theta})\left(X_1^{(1)}-f(Y_1^{(1)},\tilde{\theta})\right)\right]=0$ which is a contradiction to **L.8**.

Let $\tilde{\theta}=\theta_2$, i.e. $\mathbb{E}\left[\nabla f(Y_1^{(2)},\tilde{\theta})\left(X_1^{(2)}-f(Y_1^{(2)},\tilde{\theta})\right)\right]=0$.
Using **L.8** we get the existence of d such that

$$\left\|\mathbb{E}[\nabla f(Y_1^{(1)},\tilde{\theta})(X_1^{(1)}-f(Y_1^{(1)},\tilde{\theta})]\right\|>d>0\,.$$

From Lemma 3.2.1 we get that

$$\lambda\mathbb{E}\left[\nabla f(Y_1^{(1)},\tilde{\theta})\left(X_1^{(1)}-f(Y_1^{(1)},\tilde{\theta})\right)\right]+(1-\lambda)\mathbb{E}\left[\nabla f(Y_1^{(2)},\tilde{\theta})\left(X_1^{(2)}-f(Y_1^{(2)},\tilde{\theta})\right)\right]=0\,.$$

Therefore, it follows

$$\begin{aligned}\left\|\mathbb{E}\left[\nabla f(Y_1^{(2)},\tilde{\theta})\left(X_1^{(2)}-f(Y_1^{(2)},\tilde{\theta})\right)\right]\right\| &= \frac{\lambda}{1-\lambda}\left\|\mathbb{E}[\nabla f(Y_1^{(1)},\tilde{\theta})(X_1^{(1)}-f(Y_1^{(1)},\tilde{\theta})]\right\|\\ &> \frac{\lambda}{1-\lambda}d>0\,.\end{aligned}$$

This is a contradiction to the definition of θ_1, but we assumed $\tilde{\theta}=\theta_1$. So it follows $\tilde{\theta}\neq\theta_z$ for $z=1,2$.

■

Theorem 3.2.4

Let assumptions **L.1-L.8** hold. Under H_1

$$T_n(\eta,\gamma;A) \xrightarrow{p} \infty\,.$$

Proof:
As in the proof of Theorem 2.2.2 (see page 27) we can conclude

$$T_n(\eta,\gamma;A) \geq \frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{m} \nabla f(Y_t,\hat{\theta}_n)\left(X_t - f(Y_t,\hat{\theta}_n)\right)\right\|_A .$$

In Kirch and Tadjuidje Kamgaing [2012] they rewrite the sum up to m as a sum over centered stationary and α-mixing sequence plus some expectation as well as a part which could be handled using the ULLN for neural networks. In our situation we get for $i = 1,\ldots,q$

$$\begin{aligned}\sum_{t=1}^{m}&\left((\nabla f)_i(Y_t,\hat{\theta}_n)\left(X_t - f(Y_t,\hat{\theta}_n)\right)\right)\\ &= m\left(\mathbb{E}[q_i(Y_1,\tilde{\theta})]\right)\\ &\quad+\sum_{t=1}^{m}\left(q_i(Y_t,\hat{\theta}_n) - q_i(Y_t,\tilde{\theta})\right)\\ &\quad+O\left(\sup_{\theta\in K}\sum_{t=1}^{m}(q_i(Y_t,\theta) - \mathbb{E}[q_i(Y_1,\theta)])\right).\end{aligned}$$

For the last part on the right hand side, we know from Corollary 3.2.2 that the assumptions of Ranga Rao C.1.1 are fulfilled. It converges to 0, which gives $o_P(n)$. It is left to analyse

$$\sum_{t=1}^{m}\left(q_i(Y_t,\hat{\theta}_n) - q_i(Y_t,\tilde{\theta})\right) = \sum_{t=1}^{m}\left(\nabla q_i(Y_t,\tilde{\theta})\left(\hat{\theta}_n - \tilde{\theta}\right)\right) + \sum_{t=1}^{m}\left(\hat{\theta}_n - \tilde{\theta}\right)^{\mathsf{T}}\nabla^2 q_i(Y_t,\xi)\left(\hat{\theta}_n - \tilde{\theta}\right),$$

where $\left\|\xi - \tilde{\theta}\right\| \leq \left\|\hat{\theta}_n - \tilde{\theta}\right\|$. First observe that $\hat{\theta}_n$ is $\sqrt{n}$-consistent, i.e. $\left\|\hat{\theta}_n - \tilde{\theta}\right\| = O_P(\frac{1}{\sqrt{n}})$. Using the ULLN (Theorem C.1.1) which holds by Corollary 3.2.2, we get

$$\begin{aligned}\left|\sum_{t=1}^{m}\left(\left(\hat{\theta}_n - \tilde{\theta}\right)^{\mathsf{T}}\nabla q_i(Y_t,\tilde{\theta})\right)\right| &\leq \left\|\hat{\theta}_n - \tilde{\theta}\right\|\left\|\sum_{t=1}^{m}\nabla q_i(Y_t,\tilde{\theta})\right\|\\ &= o_P(n)\,,\end{aligned}$$

$$\begin{aligned}\left|\sum_{t=1}^{m}\left(\hat{\theta}_n - \tilde{\theta}\right)^{\mathsf{T}}\nabla^2 q_i(Y_t,\xi)\left(\hat{\theta}_n - \tilde{\theta}\right)\right| &\leq \left\|\hat{\theta}_n - \tilde{\theta}\right\|^2\sum_{t=1}^{m}\left\|\nabla^2 q_i(Y_t,\xi)\right\|_\infty\\ &= O_P(1)\,.\end{aligned}$$

Then we gain

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{m} q_i(Y_t,\hat{\theta}_n) - \sum_{t=1}^{m} q_i(Y_t,\tilde{\theta}) = o_P(\sqrt{n})\,.$$

For the test statistic this results in

$$\begin{aligned} T_n(\eta,\gamma;A) &\geq \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^{m} q(t,\hat{\theta}_n) \right\|_A \\ &= \sqrt{n}\lambda \left\| \mathbb{E}[q(Y_1,\tilde{\theta})] \right\|_A + o_P(\sqrt{n}) + o_P(n)\,. \end{aligned}$$

■

3.3. Generalized class of change-point test

In the section 2.2 as well as in section 3.2 we have analysed change-point tests for different model assumptions. Nevertheless, the proofs are quite analogue. This does not only hold true for this two examples. The non-parameteric change-point test is analysed not only for the mean change model, but has been subsequently been extended to many different models. For instance, Gombay [2010] analysed a change-point test for linear regression model with time-series errors, Gombay et al. [1996] considered change-point tests for variance changes and robust techniques are also studied, e.g. for M-tests see Hušková [1996], see 3.1. The test statistics seems to be different, but they are all based on the same idea and are proven in a similar way. Hence, it seems more plausible to develop a generalized class of change-point tests.

For the sequential set-up regularity conditions on the change-point test were derived in Kirch and Tadjuidje Kamgaing [2014]. In the sequential set-up the estimator is based on an independent historical dataset. This assumption is not made in the offline case; on the contrary the estimator depends on the same observations as the test statistic.

We considered the weighted CUSUM statistic with the deterministic weight function

$$w_{\eta,\gamma}(s) = \mathbf{1}_{\{\eta<s<1-\eta\}}\,(s(1-s))^{-\gamma} \qquad s \in (0,1)\,.$$

This is a typical form of weight function considered in the literature. But also a generalised class of test statistics so-called q-weighted CUSUM stastistics are considered (see Kirch and Tadjuidje Kamgaing [2012] or Csörgő and Horváth [1997] section 4.1). The weight function w is then defined as $1/q$ where q is out of the class

$$Q_{0,1} = \{q:\ q \text{ is non-decreasing in a neighborhood of zero, non-increasing in a neighbourhood of one and } \inf_{\eta\le t\le 1-\eta} q(t) > 0 \text{ for all } 0<\eta<\tfrac{1}{2}\}\,.$$

To derive asymptotic results they introduced the integral

$$I(q,c) = \int_0^1 \frac{1}{s(1-s)} \exp\left\{-c\frac{q^2(s)}{s(1-s)}\right\} ds\,.$$

In Csörgő and Horváth [1993] they showed that for functions $q \in Q_{0,1}$ the integral $I(q,c)$ is finite for all $c>0$ if and only if we have

$$\limsup_{t\searrow 0} |B(t)|/q(t) = 0 \quad a.s. \qquad \text{and} \qquad \limsup_{t\nearrow 1} |B(t)|/q(t) = 0 \quad a.s.$$

One can check that the weight function $w_{\eta,\gamma}(s)$ fulfils these conditions.

As we are interested in sufficient conditions on the weight function to derive the asymptotics, we define the following class of weight functions (as done in Kirch and Tadjuidje Kamgaing [2014])

$$\begin{aligned}\mathcal{L}(\eta,\gamma) := \{\rho:\ &\text{non-negative functions with } \lim_{\alpha<s<1-\alpha} \rho(s) < \infty \\ &\text{for all } \alpha \in (\eta,\tfrac{1}{2}) \text{ and for } \gamma \in (0,\tfrac{1}{2}) \lim_{s\to 0} \rho(s)s^{\gamma} < \infty \\ &\text{and } \lim_{s\to 1} \rho(s)(1-s)^{\gamma} < \infty\}\,.\end{aligned} \tag{3.36}$$

Notize, the asymptotic behaviour given in (3.36) is used to prove the sufficiency of the integral condition. For informations on the proof we refer to Shao et al. [1991]. So, functions satisfying the integral conditions also fulfil the asymptotic conditions in (3.36), in consequence, they are equivalent. In section 3.1, we analysed a special form of the randomized weight function. We use the main idea to show that for a general class of possible randomized weight functions the asymptotic results still hold true under some regularity conditions.In the following sections we study the general principle of the test statistic and identify regularity conditions under which the asymptotic results are derived. In section 3.3.1, we describe the statistical model and the general concept of the test statistic of interest. The regularity conditions for deriving the asymptotic distribution as well as having a consistent test for a general set-up will be discussed in section 3.3.2. In the case of smooth functions, we give the conditions on the function and the model such that we still have the same asymptotics (section 3.3.3). This topic is completed with technical proofs in section 3.3.4.

3.3.1. Change-point model and construction of change-point test

Let $\{X_t\}$ be the observed process. Under $\lambda = 1$, the process $\{X_t\}$ has probability measure P_0. We consider the following class of probability measures $\mathcal{P}_\theta := \{P_\theta : \theta \in \Theta\}$. In most of the cases one assumes that P_0 is out of $\mathcal{P}_\theta$. As we have seen in section 3.2, it is enough to assume that there exists an identifiable parameter $\tilde{\theta}$ such that $P_{\tilde{\theta}}$ fits P_0 best. We assume for $\lambda < 1$ that there exists also a parameter out of Θ such that the measure to this parameter fits best, but is neither equal to the measure before the change-point nor to the measure after the change-point. In both cases we denote the best fitting parameter with $\tilde{\theta}$. It is important that under $\lambda = 1$ the parameter $\tilde{\theta}$ is not the same as under $\lambda < 1$.

For the analysis of estimators, Godambe [1960] introduced the class of estimation functions. A discussion of estimation functions in parametric statistic model is given in Sørensen [1999]. The most interesting property of this class of functions is the structure. They are assumed to be sums and the estimator is defined as a root. In the following those functions are denoted by estimating functions.

A common method in change-point theory to derive test statistics is to use the estimating function of the parameter as the test function. Let G denote the estimating function for the change-point, i.e.

$$S_G(n, \theta) = \sum_{t=1}^{n} G(\mathbb{X}_t; \theta)\,,$$

and the corresponding parameter estimator $\hat{\theta}_n$ solves

$$S_G(n, \theta) = 0\,.$$

The estimation function is assumed to be an unbiased estimating function, i.e.

$$\mathbb{E}[G(\mathbb{X}_1, \tilde{\theta})] = 0\,. \tag{3.37}$$

In our situation this is only true under H_0. That is why we assume, the parameter $\tilde{\theta}$ under H_0 to be the unique solution of the equation $\mathbb{E}[G(\mathbb{X}_1, \theta)] = 0$. The corresponding test statistic is then given as

$$T_n(\hat{\theta}_n; A) = \max_{0<k<n} \frac{w(k/n)}{\sqrt{n}} \left\| S_G(k; \hat{\theta}_n) \right\|_A\,,$$

with $w(k/n) \in \mathcal{L}(\eta, \gamma)$ (see (3.36)). A is again given as before (see chapter 2). A simple example is the projection on the lth dimension, i.e.

$$A_{ij} = 0 \quad (i,j) \neq (l,l) \qquad \text{and} \qquad A_{ll} = 1\,. \tag{3.38}$$

In the same way, one can weight the dimensions due to the importance. For example, there is one dimension where even small changes have to be detectable, then this dimension is weighted higher. In consequence, we can control the power and increase it for some alternatives at the cost of loosing power in others. The matrix A is a positive semi-definite symmetric matrix. From the previous examples, we can see that besides weighting the importance of the alternatives, the matrix is also used to handle the asymptotic covariance Σ of the Wiener process. To derive the asymptotic distribution based on a standard Brownian bridge we replace the matrix A by $\Sigma^{-\frac{1}{2}} A' \Sigma^{-\frac{1}{2}}$. For the applications we usually have to replace $\Sigma^{-\frac{1}{2}}$ with a consistent estimator, but this does not change the asymptotic results (see Theorem 2.2.3). Moreover, the choice of A defines the detectable alternatives as well as the sensitivity to different alternatives.

In Ciupera [2013] different functions are considered for testing and estimation. Motivated by this publication, we allow the general set-up, where one might use different estimation functions for testing and for deriving the parameter. Additionally, we allow the weight function to depend on the data. The corresponding test statistic is then given by

$$T_n(\hat{\theta}_n; A) = \max_{0<k<n} \frac{w_n(n,k)}{\sqrt{n}} \left\| S_H(k; \hat{\theta}_n) \right\|_A , \tag{3.39}$$

where $S_H(k;\theta) = \sum_{t=1}^{k} H(\mathbb{X}_t, \theta)$ is some unbiased estimating function, but does not be an estimating function such that it has unique solution of $\mathbb{E}[H(X_1, \theta)] = 0$. We consider the following example.

Example 3.3.1 *Let $X_1, \ldots, X_n$ be log-normal distributed with μ_0, σ_0^2 then we have to solve*

$$\sum_{t=1}^{n} (\log(X_t) - \mu) = 0$$
$$\sum_{t=1}^{n} \left((\log(X_t) - \mu)^2 - \sigma^2 \right) = 0$$

This means an estimating function is given by

$$G(x, \theta) = \begin{pmatrix} \log(x) - \mu \\ (\log(x) - \mu)^2 - \sigma^2 \end{pmatrix} . \tag{3.40}$$

The change, we are interested in, occurs in the mean of the log-normal distributed observations, i.e. an intuitively choice of the test function would be

$$S_H(k;\theta) = \sum_{t=1}^{k} \left(X_t - \exp(\mu + \tfrac{1}{2}\sigma^2) \right) ,$$

with $\theta = (\mu, \sigma^2)$.

3.3.2. Regularity conditions for asymptotics of change-point tests

Before we come to the parameter estimation and change-point estimation function we derive general assumptions, further on called regularity conditions. We derive an asymptotic test-statistic with asymptotic power 1. We consider the following change-point test-statistic

$$T_n(\hat{\theta}_n) := \max_{1\le k<n} w(n,k)\sqrt{S_H^{\mathsf{T}}(k;\hat{\theta}_n)A\,S_H(k;\hat{\theta}_n)} = \max_{1\le k<n} w(n,k)\left\|S_H(k;\hat{\theta}_n)\right\|_A, \tag{3.41}$$

where $\|\cdot\|_A$ is as defined in **N.3** and

$$S_H(k;\theta) = \sum_{i=1}^{k} H(\mathbb{X}_i,\theta)\,. \tag{3.42}$$

Further on we will call $S_H(k;\cdot)$ the statistic function. Our test statistic is not only based on the statistic function, we also have the so called weight function $w(n,k)$. This function has to fulfil the following assumption.

G.7 The weight-function $w_n(n,k) \equiv w(n,k;X_1,\ldots,X_n)$ is a measurable non-negative random function and there exists a continuous function $\rho(s) \in \mathcal{L}(\eta,\gamma)$ (see (3.36)), such that

$$\sup_{0<s<1} |w_n(n,\lfloor sn\rfloor) - \rho(s)| = o_P(1)\,.$$

Note that on every inner interval of $(0,1)$ the function $\rho(s)$ is bounded. In most of the cases this assumption is fulfilled due to the fact that the weight-function is a non-random function of the form $w_n(n,k) \equiv w(n,k) = \frac{1}{\sqrt{n}}\rho(k/n)$. But, as we have seen in section 3.1 other weight functions are also possible.

It should be mentioned that the form of the statistic function does not need to be a sum as described here, as long as the asymptotic behaviour ,given as assumptions in this section, still hold true. The sum assumption makes it easier to verify these assumptions.

Null hypothesis

Let us first consider under which regularity conditions we derive an asymptotic level-α test. Thus, in the following H_0 is true and so $\lambda = 1$ or $m = n$ (see (1.2) and **G.1**). For determining the asymptotic distribution of the test statistic, we saw in the section 2.2 that we rewrite the statistic in terms of i.i.d. random variables. In section 3.2, we needed to replace the test function such that we were able to determine the asymptotic distribution. The key is to replace the estimator with the true parameter under H_0, see Prášková and Chochola [2014], Gombay [2010], among others. For an overview we refer to Csörgő and Horváth [1997]. So the regularity condition for the asymptotic distribution of the test statistic is the following.

G.8 There exists a matrix $C(\theta)$ such that

$$\max_{1\le k<n}\left\|S_H(k;\hat{\theta}_n) - \left(S_H(k;\tilde{\theta}) - \frac{k}{n}C(\tilde{\theta})S_G(n;\tilde{\theta})\right)\right\|_A = O_P(1)\,. \tag{3.43}$$

Thereby, the matrix $C(\tilde{\theta})$ has the suitable dimension, such that the operations are well defined.

Observe that the weight function here is necessary, as we have stated in the section 3.2. There, the replacement with the sample mean was only necessary due to the weight function to control the convergence in the neighbourhood of 1 and to guarantee the use of the LIL. As an example for the dimension reduction, let us consider the Example from before.

Example 3.3.2 *On going with example 3.3.1. Define the matrix*

$$C(\tilde{\theta}) = -\exp(\mu + \tfrac{1}{2}\sigma^2)\left(1, \frac{1}{2}\right)\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \exp(\mu + \sigma^2)\left(1, \frac{1}{2}\right).$$

Then, we have

$$\begin{aligned}
&S_H(k;\hat{\theta}_n) - S_H(k;\tilde{\theta}) + \frac{k}{n}C(\tilde{\theta})S_G(n;\tilde{\theta}) \\
&= \sum_{t=1}^{k} H(X_t,\hat{\theta}_n) - H(X_t,\tilde{\theta}) - C(\tilde{\theta})\frac{k}{n}\sum_{t=1}^{n} G(X_t,\hat{\theta}_n) - G(X_t,\tilde{\theta}) \\
&= \sum_{t=1}^{k} \nabla H(X_t,\tilde{\theta})(\hat{\theta}_n - \tilde{\theta}) + \sum_{t=1}^{k}(\hat{\theta}_n - \tilde{\theta})^{\mathsf{T}}\nabla^2 H(X_t,\xi)(\hat{\theta}_n - \tilde{\theta}) \\
&\quad - C(\tilde{\theta})\frac{k}{n}\left(\sum_{t=1}^{n} \nabla G(X_t,\tilde{\theta})(\hat{\theta}_n - \tilde{\theta}) + \sum_{t=1}^{n}\sum_{i=1}^{2} \mathbf{e}_i(\hat{\theta}_n - \tilde{\theta})^{\mathsf{T}}\nabla^2 G_i(X_t,\xi_i)(\hat{\theta}_n - \tilde{\theta})\right),
\end{aligned}$$

with $\tilde{\theta} = (\mu, \sigma^2)^{\mathsf{T}}$, $\mathbf{e}_i$ the ith unit vector and $\xi_i \in \Theta$, where it is in each component between $\hat{\theta}_n$ and $\tilde{\theta}$. With the choice of C we have

$$\sum_{t=1}^{k} \nabla H(X_t,\tilde{\theta})(\hat{\theta}_n - \tilde{\theta}) - C(\tilde{\theta})\frac{k}{n}\left(\sum_{t=1}^{n} \nabla G(X_t,\tilde{\theta})(\hat{\theta}_n - \tilde{\theta})\right) = 0$$

Then we have to show that the rest converges to 0. This holds as we have that the second derivatives are uniformly bounded. For the test function we have

$$\nabla^2 H(X_t,\xi) = -\exp(\mu_\xi + \frac{1}{2}\sigma_\xi^2)\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

which is bounded for a parameter $\xi = (\mu_\xi, \sigma_\xi^2) \in \Theta$. The second derivative of $G(X_t,\xi)$ is either -2 or 0 not depending on ξ, in particular constant. Then we have

$$\sum_{t=1}^{n}\sum_{i=1}^{2} \mathbf{e}_i(\hat{\theta}_n - \tilde{\theta})^{\mathsf{T}}\nabla^2 G_i(X_t,\xi)(\hat{\theta}_n - \tilde{\theta}) = n\left\|\hat{\theta}_n - \tilde{\theta}\right\|^2\begin{pmatrix} 0 \\ -2 \end{pmatrix}.$$

This leads us with

$$\begin{aligned}
&\max_{1\le k\le \frac{n}{2}} \rho\left(\frac{k}{n}\right)\left\|S_H(k;\hat{\theta}_n) - \left(S_H(k;\tilde{\theta}) - \frac{k}{n}C(\tilde{\theta})S_G(n;\tilde{\theta})\right)\right\|_A \\
&= \left\|\hat{\theta}_n - \tilde{\theta}\right\|^2 \max_{1\le k\le \frac{n}{2}} \rho\left(\frac{k}{n}\right) kO_P(1) \\
&= O_P\left(\frac{1}{n}\right),
\end{aligned}$$

which shows the result for the first half. It is left to proof it for the second half, which essentially goes an equivalent argumentation for

$$\max_{\frac{n}{2}<k<n} \rho\left(\frac{k}{n}\right) \Big\| S_H(n-k,\hat{\theta}_n) - S_H(n-k,\tilde{\theta}) + \frac{n-k}{n} C(\tilde{\theta}) S_G(n,\tilde{\theta}) - S_H(n,\hat{\theta}_n) + S_H(n,\tilde{\theta}) - C(\tilde{\theta}) S_G(n,\tilde{\theta}) \Big\|_A .$$

To determine the asymptotic distribution, we used the functional central limit theorem. This is then applied on the replacement.

G.9 Let the fCLT hold true, i.e.

$$\left\{ \frac{1}{\sqrt{n}} \begin{pmatrix} S_H(\lfloor sn \rfloor; \tilde{\theta}) \\ C(\tilde{\theta}) S_G(\lfloor sn \rfloor; \tilde{\theta}) \end{pmatrix} : \ s \in [0,1] \right\}$$

converges towards a Wiener process $\{W(s) : \ 0 \le s \le 1\}$, $W(s) = (W_H(s), W_G(s))^T$ with covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_H & \Sigma_{H,G} \\ \Sigma_{H,G}^{\mathrm{T}} & \Sigma_G \end{pmatrix} .$$

If the statistic function $S_H(k;\cdot)$ is given by (3.42), this assumption implies $\mathbb{E}[H(\mathbb{X}_1^{(1)}, \tilde{\theta})] = 0$. In the context of estimation function, such functions are called unbiased as we have stated in the introduction.

To handle the use of some weight function we have to assume forward and backward Hájek-Rényi inequalities.

G.10 For $0 < \gamma < \frac{1}{2}$ it holds

$$\max_{1 \le k < \frac{n}{2}} \frac{1}{m^{\frac{1}{2}-\gamma} \, k^{\gamma}} \Big\| S_H(k;\tilde{\theta}) \Big\|_A = O_P(1) ,$$

$$\max_{\frac{n}{2} \le k < n} \frac{1}{m^{\frac{1}{2}-\gamma} \, (n-k)^{\gamma}} \Big\| S_H(n;\tilde{\theta}) - S_H(k;\tilde{\theta}) \Big\|_A = O_P(1) .$$

Theorem 3.3.1 *Assume assumptions **G.7** and **G.8** – **G.10**. Under H_0 we have*

$$T_n(\hat{\theta}_n; A) \xrightarrow{d} \sup_{0<s<1} \rho(s) \| W_H(s) - s W_G(1) \|_A , \tag{3.44}$$

with $\{W_H(s)\}$ and $\{W_G(s)\}$ are Wiener processes with covariance matrix Σ_H and Σ_G, respectively.
The covariance matrix of the Wiener processes is given by $\Sigma_{H,G}$.

Example 3.3.3 *For the example 3.3.1 we show that there exists such an Wiener process. We have*

$$\mathbb{E}[H(\lfloor sn \rfloor, \tilde{\theta}) C(\tilde{\theta}) G(\lfloor sn \rfloor, \tilde{\theta})] = c(\tilde{\theta}) \left(\mathbb{E}[H(\lfloor sn \rfloor, \tilde{\theta}) G_1(\lfloor sn \rfloor, \tilde{\theta})] + \frac{1}{2} \mathbb{E}[H(\lfloor sn \rfloor, \tilde{\theta}) G_2(\lfloor sn \rfloor, \tilde{\theta})] \right) .$$

As we can write this as moments of Y_i and $\exp(Y_i)$, with Y_i normally i.i.d. having expectation 0, we get the existence and in the same way, we can show that it converges.

As we already know from the examples, we just want to mention that the matrix A can be replaced by a consistent estimator, not changing the asymptotic distribution.

Corollary 3.3.1 *Let the assumptions of Theorem 3.3.1 hold true and let $\hat{A}_n$ be a consistent estimator for A. Then it holds*

$$T_n(\hat{\theta}_n; \hat{A}_n) \xrightarrow{d} \sup_{0<s<1} \rho(s) \|W_H(s) - sW_G(1)\|_A$$

Alternative

Now, we want to focus on the power of the test. In Kirch and Tadjuidje Kamgaing [2012], and here, we make the following assumption.

G.11 The statistic function $S_H(k; \cdot)$ has to fulfils a CLT, i.e. there exists a vector δ_n such that

$$\left\| \frac{1}{m} S_H(m; \tilde{\theta}) - \delta_n \right\|_A = O_p\left(\frac{1}{\sqrt{m}}\right) .$$

We are going to see, in view of the change-point estimator this assumption is not strong enough to proof the asymptotics of these estimator. But for the moment that is all we need.

For sufficiently smooth functions and $\sqrt{n}$-consistent parameter estimator these conditions are fulfilled, see section 3.3.3.

Theorem 3.3.2 *Let H_1 hold true as well as the assumptions* ***G.1****,* ***G.7****,* ***G.8*** *and* ***G.11****. If $\rho(\lambda) > 0$ and δ_n from* ***G.11*** *fulfils either assumption* ***G.5.a)*** *or* ***G.5.b)****, then*

$$T_n(\hat{\theta}_n; A) \xrightarrow{p} \infty . \tag{3.45}$$

Let us show that this condition is fulfilled for the example of the log-normal distributed observations introduced in Example 3.3.1.

Example 3.3.4 *For the test function given in Example 3.3.1, we have*

$$\left\| \frac{1}{m} S_H(m; \tilde{\theta}) - \mathbb{E}[H(X_1, \tilde{\theta})] \right\|_A = \left\| \frac{1}{m} \sum_{t=1}^{m} (H(X_t, \tilde{\theta}) - \mathbb{E}[H(X_1, \tilde{\theta})]) \right\|_A .$$

The sum is taken over i.i.d. centered random variables with finite second moment. From the CLT we have this converges with the rate of $\sqrt{n}$.

Observe that under H_0 the test function $H(X_1, \tilde{\theta})$ has expectation 0 but as we are under the alternative and allow fixed alternative, we have to center with this expectation. In the local alternative, the expectation depends on n. Then we have to make assumptions on the rate of convergence (compare section 3.2).

For smooth enough function, criteria given below, we can replace the assumptions **G.8** and **G.11** with some moment conditions.

3.3.3. Smooth functions

In the case of the mean change for i.i.d. random variables as in the case of NLAR(p)-processes we consider smooth function. These functions fulfil some moment conditions such that we are able to proof the regularization conditions. Here, we are going to state such moment conditions for the testing and the estimating function to guarantee the replacement assumption **G.8**. Afterwards, we also consider under which additional assumptions the consistency of the test follows, i.e. the assumption **G.11**.

We follow the same idea as used in Kirch and Tadjuidje Kamgaing [2012]. The key ideas are still the same, nevertheless we have to be careful since the estimator $\hat{\theta}_n$ depends on the same observations as the testing function is evaluated.

$\sqrt{\mathbf{n}}$-consistent parameter estimator

We show first, under which conditions we derive a $\sqrt{n}$-consistent parameter estimator.

L.1 Assume $\{X_t\}$ stationary and ergodic under H_0.

The following assumptions are made for the estimating function.

L.2 $\mathbb{E}[\sup_{\theta\in\Theta}\|G(\mathbb{X}_1,\theta)\|]<\infty$

L.3 $\tilde{\theta}$ unique root of $\mathbb{E}[G(\mathbb{X}_1,\theta)]$

L.4 G continuously differentiable w.r.t. θ in a convex environment $U_{\tilde{\theta}}$ of $\tilde{\theta}$ such that $\mathbb{E}[\nabla G(\mathbb{X}_1,\tilde{\theta})]$ is positive definite and $\mathbb{E}[\sup_{\theta\in U_{\tilde{\theta}}}\|\nabla G(\mathbb{X}_1,\theta)\|]<\infty$

L.5 $\sum_{t=1}^{n} G(\mathbb{X}_t,\tilde{\theta}) = O_P(\sqrt{n})$.

As in Sørensen [1999] and Kirch and Tadjuidje Kamgaing [2012] stated the last assumption follows from a central limit theorem for $G(X_t,\tilde{\theta})$ under moment conditions in addition to weak dependence assumptions. Then the $\sqrt{n}$-consistency of the parameter estimator follows.

Moreover, we know from the examples that the stationary assumption and the uniformly boundedness of the estimating function w.r.t. the expectation we have from Theorem C.1.1 that there exists $F(\theta)$ such that

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}S_G(n;\theta)-F(\theta)\right\| = o_P(1)\,.$$

Then we can use the Theorem C.1.2 and gain the consistency. Equivalent argumentation together with the $\sqrt{n}$ assumption on the estimating function gives us then that the parameter estimator is $\sqrt{n}$-consistent.

Proposition 3.3.1 (Kirch and Tadjuidje Kamgaing [2012] Proposition 5.1)

1. *Under assumptions **L.1**–**L.3** the estimator $\hat{\theta}_n$ is consistent for $\tilde{\theta}$.*
2. *Under the assumptions **L.1**–**L.5** the estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent for $\tilde{\theta}$.*

Replacement assumption

For $\sqrt{n}$-consistent estimator which are derived under the above assumptions, with the following additional assumptions the replacement assumption **G.8**.

L.6 Let $\mathbb{E}\left[\left\|\nabla H(X_1,\tilde{\theta})\right\|\right] < \infty$ and

L.7 for $j = 1, \dots, r$

$$\mathbb{E}\left[\sup_{\theta\in U_{\tilde{\theta}}} \left\|\nabla^2 H_j(X_1,\theta)\right\|_\infty\right] < \infty\,, \qquad \mathbb{E}\left[\sup_{\theta\in U_{\tilde{\theta}}} \left\|\nabla^2 (C(\tilde{\theta})G)_j\ (X_1,\theta)\right\|_\infty\right] < \infty\,.$$

This gives us a ULLN.

Proposition 3.3.2 (Kirch and Tadjuidje Kamgaing [2012], Proposition 5.2)

*Under the assumptions **L.1**–**L.7** and with*

$$C(\tilde{\theta}) = \mathbb{E}[\nabla H(X_1,\tilde{\theta})](\mathbb{E}[\nabla G(X_1,\tilde{\theta})])^{-1} \tag{3.46}$$

*the assumption **G.8** follows with ρ fulfilling **G.7**.*

The proof follows the equivalent ideas of the proof of Proposition 5.2 in Kirch and Tadjuidje Kamgaing [2012]. Due to the Taylor-expansion and the triangle inequality we can follow their proof.

Observe that this assumptions hold true for the example 3.3.1. So for deriving the matrix $C(\tilde{\theta})$ we just calculated the derivatives.

Alternative assumption

To derive the asymptotic conditions for the change-point test, we need some moment conditions on the testing function evaluated for the time-series after the change-point.

L.8 Let $\{X_t^{(2)}\}$ be a stationary and ergodic time-series such that for a convex environment $U_{\tilde{\theta}}$ of $\tilde{\theta}$ we have for $j = 1, \dots, r$

$$\mathbb{E}\left[\left\|\nabla H(X_1^{(2)},\tilde{\theta})\right\|\right] < \infty\,, \qquad \mathbb{E}\left[\sup_{\theta\in U_{\tilde{\theta}}} \left\|\nabla^2 H_j(X_1^{(2)},\theta)\right\|_\infty\right] < \infty$$

and there exists $\tilde{\theta}_n$ being the root of $\lambda\mathbb{E}[G(X_1^{(1)},\theta)] + (1-\lambda)\mathbb{E}[G(X_1^{(2)},\theta)]]$ for each n.

This guarantees the interchange of integration and limit. So we are able to identify the limit with

$$\lim_{n\to\infty} \delta_n = \mathbb{E}\left[H(\mathbb{X}_1^{(1)},\tilde{\theta})\right] = \delta$$

for stationary observations up to m. In the case of the A-fixed alternative we have $\delta_n \equiv \delta$ and $\mathbb{E}\left[H(\mathbb{X}_1^{(1)},\tilde{\theta})\right] \neq \mathbb{E}\left[H(\mathbb{X}_1^{(2)},\tilde{\theta})\right]$. For the A-local alternative the equality of the expectations hold true.

Proposition 3.3.3

*Assume **L.8** then under H_1 we have **G.11**.*

Until now, we always considered abrupt changes. However, even in the case of an AR(p)-process with an abrupt change of the autoregressive function the observations after the change have starting values not out of the stationary distribution. Therefore, Kirch and Tadjuidje Kamgaing [2014] considered under which conditions on the time-series after the change the proofs still hold true.

3.3.4. Proofs

Theorem 3.3.1

Assume assumptions **G.7** and **G.8** – **G.10**. Under H_0 we have

$$T_n(\hat{\theta}_n; A) \xrightarrow{d} \sup_{0<s<1} \rho(s) \|W_H(s) - sW_G(1)\|_A \,, \tag{3.47}$$

with $\{W_H(s)\}$ and $\{W_G(s)\}$ are Wiener processes with covariance matrix Σ_H and Σ_G, respectively.
The covariance matrix of the Wiener processes is given by $\Sigma_{H,G}$.

Proof of Theorem 3.3.1:
First we consider the behaviour of

$$\left| \left\| S_H(k; \hat{\theta}_n) \right\|_A - \left\| S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta}) \right\|_A \right| \leq \left\| S_H(k; \hat{\theta}_n) - (S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta})) \right\|_A$$

Note that it holds $\forall a_k, \; b_k \geq 0 \; k = 1, \ldots, d$

$$|\max_{1 \leq k \leq d} a_k - \max_{1 \leq k \leq d} b_k| \leq \max_{1 \leq k \leq d} |a_k - b_k| \,.$$

This inequality in addition with assumption **G.8** gives us, that

$$\begin{aligned}
&\left| \max_{1 \leq k < n} \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \hat{\theta}_n) \right\|_A - \max_{1 \leq k < n} \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta}) \right\|_A \right| \\
&\quad \leq \max_{1 \leq k < n} \left| \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \hat{\theta}_n) \right\|_A - \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta}) \right\|_A \right| \\
&\quad \leq \max_{1 \leq k < n} \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \hat{\theta}_n) - (S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta})) \right\|_A \\
&\quad \to 0 \,.
\end{aligned}$$

So it follows

$$\max_{1 \leq k < n} \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \hat{\theta}_n) \right\|_A = \max_{1 \leq k < n} \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta}) \right\|_A + o_p(1) \,.$$

With the matrixes D_1 and D_2 given by

$$D_1 = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad D2 = \begin{pmatrix} 0 & 0 \\ I_p & 0 \end{pmatrix}$$

and we get from the monotone mapping theorem the following result. Then for every $\eta \in (0, \frac{1}{2})$ we have

$$\max_{\eta n \le k < (1-\eta)n} \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k;\tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n;\tilde{\theta}) \right\|_A$$

$$= \sup_{\eta < s < 1-\eta} \rho\left(\frac{\lfloor sn \rfloor}{n}\right) \left\| \frac{1}{\sqrt{n}} D_1 \begin{pmatrix} S_H(\lfloor sn \rfloor;\tilde{\theta}) \\ C(\tilde{\theta}) S_G(\lfloor sn \rfloor;\tilde{\theta}) \end{pmatrix} - \frac{\lfloor sn \rfloor}{n} D_2 \begin{pmatrix} S_H(n;\tilde{\theta}) \\ C(\tilde{\theta}) S_G(n;\tilde{\theta}) \end{pmatrix} \right\|_A$$

$$\xrightarrow{d} \sup_{\eta < s < 1-\eta} \rho(s) \| W_H(s) - s W_G(1) \|_A .$$

It is left to proof that for all $\eta \to 0$ the limit also exists and is given as defined.

$$\max_{1 \le k < \eta n} \rho^2\left(\frac{k}{n}\right) \frac{1}{n} \left\| S_H(k;\tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n;\tilde{\theta}) \right\|_A^2$$

$$\le \sup_{0 < k < \eta n} \rho^2\left(\frac{k}{n}\right) k^{2\alpha} \frac{1}{m^{2\alpha}} \cdot \left(\sup_{1 \le k < \eta n} \frac{1}{n^{1-2\alpha} k^{2\alpha}} \left\| S_H(k;\tilde{\theta}) \right\|_A^2 + \eta^{1-2\alpha} \left\| \frac{1}{\sqrt{n}} C(\tilde{\theta}) S_G(n,\tilde{\theta}) \right\|_A^2 \right)$$

With the forward Hájek-Rényi inequality (see assumption **G.10**) we gain this is $o_P(1)$. Analogously, we handle the supremum over $((1-\eta)n, n)$. So for deterministic weight functions fulfilling of the class $\mathcal{L}(\eta, \gamma)$ we are finished.

It is left to show that for the randomized weight function $w_n(n,k)$ the same limit is reached. From the a.s. convergence and because ρ is a.s. positive, we have

$$\left| \max_{1 \le k < n} w_n(n,k) \frac{1}{\sqrt{n}} \left\| S_H(k;\hat{\theta}_n) \right\|_A - \max_{1 \le k < n} \rho(\tfrac{k}{n}) \frac{1}{\sqrt{n}} \left\| S_H(k;\hat{\theta}_n) \right\|_A \right|$$

$$\le \left| \max_{1 \le k < n} \left(w_n(n,k) - \rho(\tfrac{k}{n}) \right) \frac{1}{\sqrt{n}} \left\| S_H(k;\hat{\theta}_n) \right\|_A \right|$$

$$\le \max_{1 \le k < n} \left| w_n(n,k) - \rho(\tfrac{k}{n}) \right| \left| \max_{1 \le k < n} \frac{1}{\sqrt{n}} \left\| S_H(k;\hat{\theta}_n) \right\|_A \right|$$

Observe, from the assumption **G.8** and **G.9** the second maximum on the right hand side is $O_P(1)$ as shown above. From the assumption on the weight function we get

$$\max_{1 \le k < n} \left| w_n(n,k) - \rho(\tfrac{k}{n}) \right| = o_P(1) .$$

This gives us the claim.

■

Corollary 3.3.1

Let the assumptions of Theorem 3.3.1 hold true and let $\hat{A}_n$ be a consistent estimator for A. Then it holds

$$T_n(\hat{\theta}_n; \hat{A}_n) \xrightarrow{d} \sup_{0 < s < 1} \rho(s) \| W_H(s) - s W_G(1) \|_A$$

Proof:
The proof follows essentially the idea of the proof of Theorem 2.2.3. ■

Theorem 3.3.2

Let H_1 hold true as well as the assumptions **G.1**, **G.7**, **G.8** and **G.11**. If $\rho(\lambda) > 0$ and δ_n from **G.11** fulfils either assumption **G.5.a)** or **G.5.b)**, then

$$T_n(\hat{\theta}_n; A) \xrightarrow{p} \infty \,. \tag{3.48}$$

Proof:
The main idea is to show that $w(n,m)\left\|S_H(m;\hat{\theta}_n)\right\|_A$ is converging to ∞.

$$\begin{aligned}\max_{1\le k\le n} w_n(n,k)\left\|S_H(k;\hat{\theta}_n)\right\|_A &\ge m\, w_n(n,m)\left\|\frac{1}{m}S_H(m;\hat{\theta}_n)\right\|_A \\ &= O_P\left(\rho\left(\frac{m}{n}\right)\right)\sqrt{n}\left\|\frac{1}{m}S_H(m;\hat{\theta}_n)\right\|_A\end{aligned}$$

Due to the assumptions **G.8** and **G.11** we have

$$\begin{aligned}\left|\left\|\frac{1}{m}S_H(m;\hat{\theta}_n)\right\|_A - \left\|\frac{1}{m}S_H(m;\tilde{\theta})\right\|_A\right| &\le \left\|\frac{1}{m}S_H(m;\hat{\theta}_n) - \frac{1}{m}S_H(m;\tilde{\theta})\right\|_A \\ &= O_P\left(\frac{1}{\sqrt{m}}\right) \\ \left|\left\|\frac{1}{m}S_H(m;\tilde{\theta})\right\|_A - \|\delta_n\|_A\right| &\le \left\|\frac{1}{m}S_H(m;\tilde{\theta}) - \delta_n\right\|_A \\ &= O_P\left(\frac{1}{\sqrt{m}}\right)\end{aligned}$$

Therefore, we get with assumption **G.1** for the A-local (**G.5.b)**) and the A-fixed (**G.5.a)**) alternative

$$\begin{aligned}T_n(\hat{\theta}_n) &\ge w_n(n,m)\left\|S_H(m;\hat{\theta}_n)\right\|_A \\ &\ge \left|\sqrt{m}\,\rho(\frac{m}{n})\|\delta_n\|_A + O_P(1) + \left|\rho(\frac{m}{n}) - \sqrt{n}w_n(n,m)\right|(\|\delta_n\|_A + O_p(1))\right| \to \infty \,.\end{aligned}$$

■

Proposition 3.3.1 (Kirch and Tadjuidje Kamgaing [2012] Proposition 5.1)

1. Under assumptions **L.1–L.3** the estimator $\hat{\theta}_n$ is consistent for $\tilde{\theta}$.
2. Under the assumptions **L.1–L.5** the estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent for $\tilde{\theta}$.

Proof:
From the assumption **L.1** and **L.2** we get with Theorem C.1.1 a ULLN. With the assumption **L.3** it follows from Theorem C.1.2 that $\hat{\theta}_n$ is consistent w.r.t. $\tilde{\theta}$.

With the assumption **L.4** the differentiation is given and from **L.5** we have a CLT for the estimating function evaluated at $\tilde{\theta}$. With the mean value theorem get the claim. ■

Proposition 3.3.2 (Kirch and Tadjuidje Kamgaing [2012], Proposition 5.2)

Under the assumptions **L.1**–**L.7** and with

$$C(\tilde{\theta}) = \mathbb{E}[\nabla H(X_1, \tilde{\theta})](\mathbb{E}[\nabla G(X_1, \tilde{\theta})])^{-1} \tag{3.49}$$

the assumption **G.8** follows with ρ fulfilling **G.7**.

Proof:
First of all we observe, the assumption **L.4** guarantees the existence of $(\mathbb{E}[\nabla G(X_1, \tilde{\theta})])^{-1}$ and that it is unique. From **L.6** we have that $\mathbb{E}[\nabla H(X_1, \tilde{\theta})]$ does also exist. Then we have to show that for this $C(\tilde{\theta})$ we have

$$\max_{1 \le k < n} \rho\left(\frac{k}{n}\right) \frac{1}{\sqrt{n}} \left\| S_H(k; \hat{\theta}_n) - \left(S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta}) \right) \right\|_A = o_P(1)\,.$$

As $\hat{\theta}_n$ solves $S_G(n; \hat{\theta}_n) = 0$ and we have $\{X_t\}$ is a sequence of stationary ergodic random variables **L.1** we get

$$\begin{aligned}
S_H(k; \hat{\theta}_n) - &\left(S_H(k; \tilde{\theta}) - \frac{k}{n} C(\tilde{\theta}) S_G(n; \tilde{\theta}) \right) \\
&= S_H(k; \hat{\theta}_n) - S_H(k; \tilde{\theta}) + \frac{k}{n} C(\tilde{\theta}) \left(S_G(n; \tilde{\theta}) - S_G(n; \tilde{\theta}) \right) \\
&= \nabla \left(S_H(k; \tilde{\theta}) - k\mathbb{E}\left[\nabla H(X_1, \tilde{\theta})\right] \right) (\hat{\theta}_n - \tilde{\theta}) + k\mathbb{E}\left[\nabla H(X_1, \tilde{\theta})\right] (\hat{\theta}_n - \tilde{\theta}) \\
&\quad - \frac{k}{n} C(\tilde{\theta}) \left(\left(S_G(n; \tilde{\theta}) - n\mathbb{E}\left[\nabla G(X_1, \tilde{\theta})\right] \right) (\hat{\theta}_n - \tilde{\theta}) + n\mathbb{E}\left[\nabla G(X_1, \tilde{\theta})\right] (\hat{\theta}_n - \tilde{\theta}) \right)
\end{aligned}$$

■

4. Change-point estimator

The literature contains many publications analysing change-point tests for different models. The number of publications on the asymptotic analysis of change-point estimator, i.e. the rate and asymptotic distribution, is even smaller.

In section 4.1 we present the analysis of the change-point estimator for i.i.d. observations with possible infinite variance. The corresponding change-point test is constructed in section 3.1. The change-point test and estimator are randomly weighted, as the weight function depends on the observations. We analyse the estimator and derive analogous results as in the case of i.i.d. observations with finite variance.

In section 4.2 we analyse for a non-linear (auto)regressive model a change-point estimator based on neural networks. Also under misspecification we can derive the asymptotic distribution of the change-point estimator.

With the concept of estimation functions we are able to determine regularity conditions for the change-point estimator. In section 4.3 we give the regularity conditions and prove the asymptotic behaviour.

4.1. Randomized weight function

As in section 3.1.1, we are going to consider a mean change model with i.i.d. observations with finite first moment. The model (3.1) is defined as

$$X_t = \begin{cases} \mu + \varepsilon_t & ,1 \le t \le m \\ \mu + \delta_n + \varepsilon_t & ,m < t \le n \end{cases}$$

with $\{\varepsilon_t\}$ i.i.d sequence and $\mathbb{E}[|\varepsilon_1|] < \infty$. In contrast to the model in section 2, the residuals do not need to have a second moment. Based on the idea of M-estimator we constructed a change-point test having a randomized weight function (section 3.1.1). We now focus on the corresponding change-point estimator and analyse its asymptotic behaviour.

In section 4.1.1 we first introduce the change-point estimator of interest and then give the main results. The proofs are given in section 4.1.2.

4.1.1. Asymptotics of random weighted change-point estimator

Again we make use of a self weighted estimator for the unknown parameter μ which results from a weighted least squares, where the weight function is allowed to depend on the ith observation. As discussed in section 3.1.1 on the one hand we have the test statistic

$$T_n(\eta, \gamma; g) = \max_{1 \le k < n} w_{\eta,\gamma;g}(k, n) \frac{1}{\sqrt{n}} \left\| S_g(k, \hat{\theta}_n(g)) \right\|_{\Sigma_g^{-1}},$$

with

$$w_{\eta,\gamma;g}(k, n) = \mathbf{1}_{\{\eta n < k < (1-\eta)n\}} \left(\frac{\left(\sum_{t=1}^n g(X_t)\right)^2}{\sum_{t=1}^k g(X_t) \left(\sum_{t=1}^n g(X_t) - \sum_{t=1}^k g(X_t) \right)} \right)^{\gamma}$$

and γ, η fulfilling assumption **G.2**, and on the other hand $T_n(\eta, \beta; g_n)$. In view of applications, we consider the estimator w.r.t. the weight function g_n.

As in the introductory example (section 2.3) we are going to use the argument of the maximum as estimator for the change-point, i.e. the change-point estimator is given as

$$\hat{m}(\eta, \gamma; g_n) = \arg\max_{1 \le k < n} w_{\eta,\gamma;g_n}(k, n) \left\| \sum_{t=1}^k g_n(X_t) X_t - \frac{\sum_{t=1}^k g_n(X_t)}{\sum_{t=1}^n g_n(X_t)} \sum_{t=1}^n g_n(X_t) X_t \right\| . \tag{4.1}$$

The consistency of $\hat{m}(\eta, \gamma; g_n)$ follows equivalently as in the mean change model, but is technically more efficient. We have to make an assumption on the behaviour of the expectation of g w.r.t. the time-series after the change. If g is dominated integrable, we can change limit and expectation. If additionally the function is continuous, we do not need any more assumptions. As we assumed g to be bounded, this is fulfilled.

Theorem 4.1.1 *Under H_1 and the assumptions **L.1** – **L.7*** we have for weight functions g_n fulfilling **L.6***

$$\frac{\hat{m}(\eta, \gamma; g_n)}{n} - \lambda = o_P(1) .$$

*See section 3.1.

One can ask, how fast does the estimator converge against the true parameter. An answer gives the next Theorem.

Theorem 4.1.2 *Under H_1, assumptions **L.1** – **L.8** and if $\delta_n \underset{n\to\infty}{\longrightarrow} 0$ the change-point estimator $\hat{m}(\eta, \gamma; g_n)$ given in 4.1 fulfils*

$$\hat{m}(\eta, \gamma; g_n) - m = O_P(\mathcal{D}_n^{-2}),$$

as long as $\mathcal{D}_n \sqrt{n} \underset{n\to\infty}{\longrightarrow} \infty$, with $\mathcal{D}_n = \left\| \mathbb{E}[g(X_1^{(1)} + \delta_n)(\varepsilon_1 + \delta_n)] \right\| \underset{n\to\infty}{\longrightarrow} 0$.

The rate derived in Theorem 4.1.2 is the best rate as can be seen by the next theorem.

Theorem 4.1.3 *Let H_1 be true and the change-point fulfil assumption **G.1**. Under the assumptions of Theorem 4.1.2, the asymptotic distribution of the change-point estimator is given as*

$$\mathcal{D}_n^2(\hat{m}(\eta, \gamma; g_n) - m) \overset{d}{\longrightarrow} \arg\max \{W(s) - |s| g_{\gamma,\lambda}(s);\ s \in \mathbb{R}\}$$

with D_n as in Theorem 4.1.2 , $\{W(s)\}$ a two-sided standard Wiener process and

$$g_{\gamma,\lambda}(s) = \begin{cases} (1-\gamma)(1-\lambda) + \gamma\lambda & , s < 0, \\ 0 & , s = 0, \\ \gamma(1-\lambda) + (1-\gamma)\lambda & , s > 0. \end{cases}$$

So the results are exactly the same as in the case of finite variance.

4.1.2. Proofs

Theorem 4.1.1

Under H_1 and the assumptions **L.1** – **L.7** we have for weight functions g_n fulfilling **L.6**

$$\frac{\hat{m}(\eta, \gamma; g_n)}{n} - \lambda = o_P(1).$$

Proof:
First we observe that the random weight function converges to the non-random weight function even under the alternative as the a.s. convergence holds true ($\delta_n \to 0$).

It is left to analyse

$$\left\| S_{g_n}(k, \hat{\theta}_n(g_n)) \right\| = \left\| \sum_{t=1}^{k} g_n(X_t)\ X_t - \frac{\sum_{t=1}^{k} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)} \sum_{t=1}^{n} g_n(X_t) X_t \right\|.$$

Notice, from assumption **L.6** we have by Lemma 3.1.1 that (3.18) holds. So, it follows

$$\frac{\sum_{t=1}^{\lfloor ns \rfloor} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)} \xrightarrow{p} s\,. \tag{4.2}$$

Let $k < m$, then we have

$$\begin{aligned}&\left\|\sum_{t=1}^{k} g_n(X_t)\ X_t - \frac{\sum_{t=1}^{k} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)} \sum_{t=1}^{n} g_n(X_t) X_t\right\| \\ &\qquad = \left\|\sum_{t=1}^{k} g_n(X_t)\varepsilon_t - \frac{\sum_{t=1}^{k} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)} \sum_{t=1}^{n} g_n(X_t)(\varepsilon_t + \delta_n)\right\|\,.\end{aligned}$$

Using (4.2) and Lemma 3.1.1 we first show that $\left\|S_{g_n}(k, \hat{\theta}_n(g_n))\right\|$ can be replaced by

$$\|S_g(k,\theta)\| = \left\|\sum_{t=1}^{k} g(X_t)\varepsilon_t - \frac{k}{n}\sum_{t=1}^{n} g(X_t)(\varepsilon_t + \delta_n)\right\|\,,$$

Have in mind that $g_n(X_t) = g(X_t;\hat{\theta}_n)$ and $g(X_t) = g(X_t;\theta)$. Using $C(\theta)$ as in (3.20) and (4.2), we conclude

$$\begin{aligned}&\left\|\sum_{t=1}^{\lfloor ns \rfloor}(g_n(X_t) - g(X_t))\varepsilon_t - \frac{\sum_{t=1}^{\lfloor ns \rfloor} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)}\left(\sum_{t=1}^{n}(g_n(X_t) - g(X_t))(\varepsilon_t + \delta_n)\right)\right. \\ &\left.\quad + \left(\frac{\sum_{t=1}^{\lfloor ns \rfloor} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)} - \frac{\lfloor ns \rfloor}{n}\right)\sum_{t=1}^{n} g(X_t)(\varepsilon_t + \delta_n)\right\| = o_P(n)\,.\end{aligned}$$

An equivalent replacement follows for $k > m$. Hence, it is left to analyse the replacement.

Now, using the CLT, which holds true for $\{g(X_t)\varepsilon_t\}$, we get uniformly in s

$$\left\|\sum_{t=1}^{\lfloor ns \rfloor} g(X_t)\varepsilon_t - E_{g\varepsilon,n}(\lfloor ns \rfloor) - s\sum_{t=1}^{n} g(X_t)\varepsilon_t - sE_{g\varepsilon,n}(n)\right\| = O_P(\sqrt{n}) + o_P\left(\frac{1}{n}\right)\,,$$

with

$$E_{g\varepsilon,n}(k) = \begin{cases} 0 & ,1 \le k \le m, \\ (k-m)\mathbb{E}[g(X_1^{(1)} + \delta_n)\varepsilon_1] & ,m < k \le n\,. \end{cases}$$

Then we can conclude that $\left\|\frac{1}{n}(E_{g\varepsilon,n}(k) - \frac{k}{n}E_{g\varepsilon,n}(1)\right\| =: \|E(k,n)\|$ has a maximum at $k = m$, since

$$E(k,n) = \begin{cases} -\frac{k}{n}(n-m)\mathbb{E}[g(X_1^{(1)} + \delta_n)(\varepsilon_1 + \delta_n)] & ,\ k \le m\,, \\ -\frac{m}{n}(n-k)\mathbb{E}[g(X_1^{(1)} + \delta_n)(\varepsilon_1 + \delta_n)] & ,\ k > m\,. \end{cases}$$

So, in conclusion we have

$$\max_{1 \le k < n} \frac{1}{n}\left\|S_{g_n}(k, \hat{\theta}_n(g_n)) - E(k/n)\right\| = O_P\left(\frac{1}{\sqrt{n}}\right)\,.$$

Thus, with the Theorem C.1.2 we get the result.

For a deterministic weight function

$$\rho(s) = \mathbf{1}_{\{\eta < s < (1-\eta)\}} \left(\frac{1}{s(1-s)} \right)^{\gamma}$$

fulfilling assumption **G.2**, the result follows analogously as in the proof of Theorem 2.3.1.

In the case of the randomised weight function $w_{\eta,\gamma;g_n}(k,n)$, we have from (3.18), that

$$\sup_{s\in(0,1)} |w_{\eta,\gamma;g_n}(s) - \rho(s)| = o_P(1)\,,$$

where

$$\rho(s) = \mathbf{1}_{\{\eta < s < (1-\eta)\}} \left(\frac{(\mathbb{E}[g(X_1)])^2}{s\mathbb{E}[g(X_1)](\mathbb{E}[g(X_1)] - s\mathbb{E}[g(X_1)])} \right)^{\gamma} = \frac{1}{(s(1-s))^{\gamma}}\,.$$

With

$$\begin{aligned}
&\max_{1\le k<n} \left\| w_{\eta,\gamma;g_n}(k/n) \frac{1}{n} S_{g_n}(k, \hat\theta_n(g_n)) - \rho(k/n) E(k/n) \right\| \\
&\quad \le \sup_{s\in(0,1)} |w_{\eta,\gamma;g_n}(s) - \rho(s)| \max_{1\le k<n} \left\| \frac{1}{n} S_{g_n}(k, \hat\theta_n(g_n)) \right\| \\
&\quad\quad + \max_{1\le k<n} \left\| \rho(k/n) \frac{1}{n} S_{g_n}(k, \hat\theta_n(g_n)) - \rho(k/n) E(k/n) \right\| = o_P(1)
\end{aligned}$$

the proof is finished. ■

Theorem 4.1.2

Under H_1, assumptions **L.1** – **L.8** and if $\delta_n \underset{n\to\infty}{\longrightarrow} 0$ the change-point estimator $\hat m(\eta,\gamma;g_n)$ given in 4.1 fulfils

$$\hat m(\eta,\gamma;g_n) - m = O_P(\mathcal{D}_n^{-2})\,,$$

as long as $\mathcal{D}_n\sqrt{n} \underset{n\to\infty}{\longrightarrow} \infty$, with $\mathcal{D}_n = \left\| \mathbb{E}[g(X_1^{(1)} + \delta_n)(\varepsilon_1 + \delta_n)] \right\| \underset{n\to\infty}{\longrightarrow} 0$.

Proof:
From Corollary 2.3.2we know, it is enough to analyse a truncated version of the change-point estimator. To simplify the notation, we define

$$S_{g_n}(k;\hat\theta_n(g_n)) = \sum_{t=1}^{k} g_n(X_t)\varepsilon_t - \frac{\sum_{t=1}^{k} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)} \sum_{t=1}^{n} g_n(X_t)\varepsilon_t - \left(\sum_{t=1}^{k} g_n(X_t) \right) \left(1 - \frac{\sum_{t=1}^{k} g_n(X_t)}{\sum_{t=1}^{n} g_n(X_t)} \right) \delta_n\,,$$

$$S_g(k;\theta) = \sum_{t=1}^{k} g(X_t)\varepsilon_t - \frac{k}{n} \sum_{t=1}^{n} g(X_t)\varepsilon_t - \frac{k}{n} \sum_{t=1}^{n} g(X_t)\delta_n\,.$$

The change-point estimator

$$\hat m = \arg\max_{1\le k<n} \left\| w_{\eta,\gamma;g_n}(k/n) S_{g_n}(k;\hat\theta_n(g_n)) \right\| =: \arg\max_{1\le k<n} \{V_k\}$$

can be represented as in the proof of Theorem 2.3.2 by

$$
\begin{aligned}
V_k = &- \Big\langle w_{\eta,\gamma;g_n}(m/n)S_{g_n}(m;\hat{\theta}_n(g_n)) - w_{\eta,\gamma;g_n}(k/n)S_{g_n}(k;\hat{\theta}_n(g_n)),\\
&\qquad w_{\eta,\gamma;g_n}(k/n)S_{g_n}(k;\hat{\theta}_n(g_n)) + w_{\eta,\gamma;g_n}(m/n)S_{g_n}(m;\hat{\theta}_n(g_n))\Big\rangle\\
= &- \Big\langle (w_{\eta,\gamma;g_n}(m/n) - w_{\eta,\gamma;g_n}(k/n))S_{g_n}(k,\hat{\theta}_n(g_n))\\
&\qquad + w_{\eta,\gamma;g_n}(m/n)\Big(S_{g_n}(m,\hat{\theta}_n(g_n)) - S_{g_n}(k,\hat{\theta}_n(g_n))\Big),\\
&\quad w_{\eta,\gamma;g_n}(m/n)\Big(S_{g_n}(m,\hat{\theta}_n(g_n)) - S_{g_n}(k,\hat{\theta}_n(g_n))\Big)\\
&\qquad + (w_{\eta,\gamma;g_n}(m/n) + w_{\eta,\gamma;g_n}(k/n))S_{g_n}(k,\hat{\theta}_n(g_n))\Big\rangle\,.
\end{aligned}
$$

Observe, from (3.18) we have

$$
\sup_{s\in(0,1)} |w_{\eta,\gamma;g_n}(s) - \rho(s)| = o_P(1)\,, \tag{4.3}
$$

where

$$
\rho(s) = \mathbf{1}_{\{\eta<s<(1-\eta)\}}\left(\frac{(\mathbb{E}[g(X_1)])^2}{s\mathbb{E}[g(X_1)](\mathbb{E}[g(X_1)] - s\mathbb{E}[g(X_1)])}\right)^{\gamma} = \mathbf{1}_{\{\eta<s<(1-\eta)\}}\frac{1}{(s(1-s))^{\gamma}}\,.
$$

Later on we are interested in the asymptotic behaviour of V_k. From (4.3) we can replace $w_{\eta,\gamma;g}(k/n)$ by the Lipschitz continuous function $\rho(k/n)$ in V_k. To simplify the notation we define

$$
E(k,n) = \begin{cases} -\frac{k}{n}(n-m)\mathbb{E}[g(X_1+\delta_n)(\varepsilon_1+\delta_n)] & k \le m\,,\\ -\frac{m}{n}(n-k)\mathbb{E}[g(X_1+\delta_n)(\varepsilon_1+\delta_n)] & k > m\,. \end{cases} \tag{4.4}
$$

Replacing $S_{g_n}(k,\hat{\theta}_n(g_n))$ by $S_{g_n}(k,\tilde{\theta})$, with $\tilde{\theta} = \theta$ since $\delta_n \underset{n\to\infty}{\longrightarrow}$, and centring $S_{g_n}(k,\theta)$ add expectation, we derive the following representation for V_k analogously as in Theorem 2.3.2

$$
\begin{aligned}
V_k = &- \Bigg\langle (\rho(m/n) - \rho(k/n))\Big(S_{g_n}(k,\hat{\theta}_n(g_n)) - E(k,n)\Big)\\
&\qquad + \rho(m/n)\Big(S_{g_n}(m,\hat{\theta}_n(g_n)) - S_g(m,\tilde{\theta}) - S_{g_n}(k,\hat{\theta}_n) + S_g(k,\tilde{\theta})\Big)\\
&\qquad + \rho(m/n)\Big(S_g(m,\tilde{\theta}) - E(m,n) - S_g(k,\tilde{\theta}) + E(k,n)\Big)\\
&\qquad + (E(m,n)\rho(m/n) - E(k,n)\rho(k/n))\,,\\
&\quad \rho(m/n)\Big(S_{g_n}(m,\hat{\theta}_n(g_n)) - E(m,n) - S_{g_n}(k,\hat{\theta}_n(g_n)) + E(k,n)\Big)\\
&\qquad + (\rho(m/n) + \rho(k/n))\Big(S_{g_n}(k,\hat{\theta}_n(g_n)) - E(k,n)\Big)\\
&\qquad + (E(m,n)\rho(m/n) + E(k,n)\rho(k/n))\Bigg\rangle\\
= &B_1 + B_2 + B_3 + B_4 + B_5 + B_6\,.
\end{aligned}
$$

We can calculate in an analogous way as in Theorem 2.3.2, using

$$\sum_{t=1}^{k} \varepsilon_t - k\bar{\varepsilon}_n = S_{g_n}(k, \hat{\theta}_n) - E(k,n)\,, \qquad \sum_{t=1}^{k} \varepsilon_t = S_g(k, \tilde{\theta}) - E(k,n)\,,$$
$$\bar{\varepsilon}_n = -\left(S_{g_n}(k, \hat{\theta}_n) - S_g(k, \tilde{\theta})\right)\,.$$

Hence, we derive

$$\begin{aligned}
B_1 &= (\rho^2(k/n) - \rho^2(m/n)) \left\| S_{g_n}(k, \hat{\theta}_n) - E(k,n) \right\|^2 \\
B_2 &= -\rho^2(m/n)\Big\langle S_{g_n}(k, \hat{\theta}_n) - E(k,n) + S_{g_n}(m, \hat{\theta}_n) - E(m,n), \\
&\qquad\qquad S_{g_n}(m, \hat{\theta}_n) - E(m,n) - S_{g_n}(k, \hat{\theta}_n) - E(k,n)\Big\rangle \\
B_3 &= -(\rho(m/n) - \rho(k/n)) \Big\langle S_{g_n}(k, \hat{\theta}_n) - E(k,n),\ (E(m,n)\rho(m/n) + E(k,n)\rho(k/n))\Big\rangle \\
&\quad + (\rho(m/n) + \rho(k/n)) \Big\langle S_{g_n}(k, \hat{\theta}_n) - E(k,n),\ (E(m,n)\rho(m/n) - E(k,n)\rho(k/n))\Big\rangle \\
&= -2(\rho^2(m/n)m - \rho^2(k/n)k) \Big\langle S_{g_n}(k, \hat{\theta}_n) - E(k,n),\ \left(1 - \frac{m}{n}\right) \mathbb{E}[g(X_1 + \delta_n)(\varepsilon_1 + \delta_n)]\Big\rangle \\
B_4 &= -\rho(m/n)\Big\langle S_{g_n}(m, \hat{\theta}_n) - S_g(m,\theta) - S_{g_n}(k, \hat{\theta}_n) + S_g(k,\theta), \\
&\qquad\qquad \rho(k/n)E(k,n) + \rho(m/n)E(m,n)\Big\rangle \\
&\quad - \rho(m/n)\Big\langle \rho(m/n)E(m,n) - \rho(k/n)E(k,n), \\
&\qquad\qquad S_{g_n}(m, \hat{\theta}_n) - S_g(m,\theta) - S_{g_n}(k, \hat{\theta}_n) + S_g(k,\theta)\Big\rangle \\
&= -2\rho^2(m/n)m\Big\langle S_{g_n}(m, \hat{\theta}_n) - S_g(m,\theta) - S_{g_n}(k, \hat{\theta}_n) + S_g(k,\theta), \\
&\qquad\qquad \left(1 - \frac{m}{n}\right) \mathbb{E}[g(X_1 + \delta_n)(\varepsilon_1 + \delta_n)]\Big\rangle \\
B_5 &= 2\rho^2(m/n)m \Big\langle S_g(m,\theta) - E(m,n) - S_g(k,\theta) + E(k,n),\ \left(1 - \frac{m}{n}\right) \mathbb{E}[g(X_1 + \delta_n)(\varepsilon_1 + \delta_n)]\Big\rangle \\
B_6 &= -\left(\|E(m,n)\|^2 \rho^2(m/n) - \|E(k,n)\|^2 \rho^2(k/n)\right) \\
&= -\left(\rho^2(m/n)m^2 - \rho^2(k/n)k^2\right) \left\| (1 - \frac{m}{n})\mathbb{E}[g(X_1 + \delta_n)(\varepsilon_1 + \delta_n)] \right\|^2 .
\end{aligned}$$

As $\rho(s)$ equals the weight function $w_{\eta,\gamma}(k/n)$ from section 2. So, we can conclude the following approximations.

We have

$$\max_{\alpha n < k \leq m - \kappa_n} \frac{m-k}{|B_6|} = O\left(\frac{1}{n\mathcal{D}_n^2}\right)\,,$$

with $\mathcal{D}_n = \left\| \mathbb{E}[g(X_1^{(1)} + \delta_n)(\varepsilon_1 + \delta_n)] \right\|$. Hence, it is enough to examine the order of

$$\max_{\alpha n < k < m} \left| \frac{B_i}{m-k} \right|$$

for $i = 1, \ldots, 5$. From the proof of Theorem 4.1.1 we get

$$\max_{\alpha n < k < m - \kappa_n} \frac{1}{m-k} \left\| S_{g_n}(m, \hat{\theta}_n) - S_g(m, \theta) - S_{g_n}(k, \hat{\theta}_n) + S_g(k, \theta) \right\| = o_P(1)$$

and

$$\max_{\alpha n < k < m - \kappa_n} \frac{1}{m-k} \left\| S_{g_n}(m, \hat{\theta}_n) - E(m,n) - S_{g_n}(k, \hat{\theta}_n) + E(k,n) \right\| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Additionally, we are going to use the Hájek-Rényi inequality for i.i.d. random variables, which gives us

$$\max_{\alpha n < k < m - \kappa_n} \frac{1}{m-k} \left\| S_g(m, \theta) - S_g(k, \theta) - (E(k,n) - E(m,n)) \right\| = O_P(\kappa_n^{-1/2}).$$

For B_1 we then conclude

$$\begin{aligned}
&\max_{\alpha n < k < m - \kappa_n} \left| \frac{B_1}{(m-k)} \right| \\
&\quad \leq \max_{\alpha n < k < m - \kappa_n} \left| \frac{\rho^2(k/n) - \rho^2(m/n)}{(m-k)} \right| \max_{\alpha n < k < m - \kappa_n} \left\| S_{g_n}(k, \hat{\theta}_n) - E(k,n) \right\|^2 \\
&\quad = O(n^{-1}) O_P(n).
\end{aligned}$$

From the Cauchy-Schwarz inequality we get

$$\begin{aligned}
&\max_{\alpha n < k < m - \kappa_n} \left| \frac{B_2}{(m-k)} \right| \\
&\quad \leq \rho^2(m/n) \max_{\alpha n < k < m - \kappa_n} \left\| S_{g_n}(k, \hat{\theta}_n) - E(k,n) + S_{g_n}(m, \hat{\theta}_n) - E(m,n) \right\| \\
&\qquad \cdot \max_{\alpha n < k < m - \kappa_n} \frac{1}{m-k} \left\| S_{g_n}(m, \hat{\theta}_n) - E(m,n) - S_{g_n}(k, \hat{\theta}_n) + E(k,n) \right\| \\
&\quad = O(1) O_P(\sqrt{n}) \left(O_P(\kappa_n^{-\frac{1}{2}}) + O_P(n^{-\frac{1}{2}}) \right) = O_P(n^{\frac{1}{2}} \kappa_n^{-\frac{1}{2}}) + O_P(1).
\end{aligned}$$

The terms B_3 and B_4 follow with the triangel inequality and arguments as above, i.e. we get

$$\begin{aligned}
\max_{\alpha n < k < m - \kappa_n} \left| \frac{B_3}{(m-k)} \right| &\leq \max_{\alpha n < k < m - \kappa_n} \left| \frac{\rho(m/n)m - \rho(k/n)k}{(m-k)} \right| \\
&\qquad \cdot \max_{\alpha n < k < m - \kappa_n} \left\| S_{g_n}(k, \hat{\theta}_n) - E(k,n) \right\| \left(1 - \frac{m}{n}\right) \mathcal{D}_n \\
&= O(1) O_P(\sqrt{n}) O(\mathcal{D}_n),
\end{aligned}$$

$$\begin{aligned}
\max_{\alpha n < k < m - \kappa_n} \left| \frac{B_4}{(m-k)} \right| &\leq 2\rho^2(m/n) m \left(1 - \frac{1}{2}\right) \mathcal{D}_n \\
&\qquad \cdot \max_{\alpha n < k < m - \kappa_n} \frac{1}{m-k} \left\| S_{g_n}(m, \hat{\theta}_n) - S_{g_n}(m, \theta) - S_{g_n}(k, \hat{\theta}_n) + S_{g_n}(k, \theta) \right\| \\
&= O(1) O_P(n \mathcal{D}_n) O_P(n^{-\frac{1}{2}}).
\end{aligned}$$

The last 2 terms are the dominating ones and give the rate of convergence. We get

$$\begin{aligned}
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_5}{(m-k)}\right| &\leq 2\rho^2(m/n)m\left(1-\frac{m}{n}\right)\mathcal{D}_n \\
&\quad\cdot \max_{\alpha n<k<m-\kappa_n}\frac{1}{(m-k)}\left\|S_{g_n}(m,\tilde{\theta})-E(m,n)-S_{g_n}(k,\tilde{\theta})+E(k,n)\right\| \\
&= O(n\mathcal{D}_n)O_P(\kappa_n^{-1/2})\,.
\end{aligned}$$

For $\kappa_n = K/\mathcal{D}_n^2$ this leads to

$$\begin{aligned}
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_1}{B_6}\right| &= o_P\,(1)\ , \\
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_2}{B_6}\right| &= o_P\,(1)\ , \\
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_3}{B_6}\right| &= o_P\,(1)\ , \\
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_4}{B_6}\right| &= o_P\,(1)\ , \\
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_5}{B_6}\right| &= K^{-1/2}O_P\,(1)\ .
\end{aligned}$$

Combining this, gives us for $k < m$

$$\begin{aligned}
P(\hat{m} < m-\kappa_n) &= P\left(\max_{\alpha n<k<m-\kappa} V_k > \max_{m-\kappa\leq k<(1-\alpha)n} V_k\right) \\
&\leq P\left(\max_{\alpha n<k<m-\kappa_n} V_k \geq V_m\right) \\
&= P\left(\max_{\alpha n<k<m-\kappa_n} V_k \geq 0\right) \\
&= P\left(\max_{\alpha n<k<m-\kappa_n}\left\{B_6\left(1+o_P(1)+\frac{B_5}{B_6}\right)+o_P(1)\right\}\geq 0\right) \\
&\leq P\left(\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_5}{B_6}\right|\geq 1+o_P(1)\right) \\
&\leq P\left(\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_5}{B_6}\right|\geq 1-\tau\right)+P\left(1+o_p(1)\leq \max_{\alpha n<k<m-\kappa_n}\left|\frac{B_5}{B_6}\right|\leq 1-\tau\right) \\
&\leq P\left(O_P(1)\geq (1-\tau)K^{1/2}\right)+o(1)\,,
\end{aligned}$$

with $0 < \tau < 1$ arbitrary. This term becomes arbitrarily small for a sufficiently large $K > 0$. Analogously we can show the other direction ($k > m$).

■

Theorem 4.1.3

Let H_1 be true and the change-point fulfil assumption **G.1**. Under the assumptions of Theorem 4.1.2, the asymptotic distribution of the change-point estimator is given as

$$\mathcal{D}_n^2(\hat{m}-m) \xrightarrow{d} \arg\max\left\{W(s)-|s|g_{\gamma,\lambda}(s);\ s\in\mathbb{R}\right\}$$

with D_n as in Theorem 4.1.2 , $\{W(s)\}$ a two-sided standard Wiener process and

$$g_{\gamma,\lambda}(s) = \begin{cases} (1-\gamma)(1-\lambda)+\gamma\lambda & , s<0\,, \\ 0 & , s=0\,, \\ \gamma(1-\lambda)+(1-\gamma)\lambda & , s>0\,. \end{cases}$$

Proof:

We show the claim by analysing the behaviour of the parts from the decomposition of V_k. Let (w.l.o.g.) $x > 0$ and $C > x$ be both fixed but arbitrary. We get

$$P(\hat{m}-m \le x) = P(m - C\mathcal{D}_n^{-2} \le \hat{m} \le m+x) + P(\hat{m}-m \le x, |\hat{m}-m| > C\mathcal{D}_n^{-2})$$

The second term on the right side becomes arbitrary small for large enough C because of the consistency of $\hat{m}$. Therefore, we consider

$$P(m - C\mathcal{D}_n^{-2} \le \hat{m} \le m+x) \le P(\max_{(k-m)\in(-C\mathcal{D}_n^{-2},x]} V_k \ge \max_{(k-m)\in(x,C\mathcal{D}_n^{-2})} V_k) + o_P(1)\,.$$

The last approximation follows with the same argumentation. Consider the decomposition of V_k similar to the one we used in the proof of Theorem 4.1.2

$$V_k = B_1 + B_2 + B_3 + B_4 + B_5 + B_6\,.$$

Recall the representation was given as

$$B_1 = (\rho^2(k/n) - \rho^2(m/n)) \left\| S_{g_n}(k,\hat{\theta}_n) - E(k,n) \right\|^2$$

$$B_2 = -\rho^2(m/n) \Big\langle S_{g_n}(k,\hat{\theta}_n) - E(k,n) + S_{g_n}(m,\hat{\theta}_n) - E(m,n), \quad S_{g_n}(m,\hat{\theta}_n) - E(m,n) - S_{g_n}(k,\hat{\theta}_n) - E(k,n) \Big\rangle$$

$$B_3 = -2(\rho^2(m/n)m - \rho^2(k/n)k) \left\langle S_{g_n}(k,\hat{\theta}_n) - E(k,n),\, \left(1-\frac{m}{n}\right) \mathbb{E}[g(X_1+\delta_n)(\varepsilon_1+\delta_n)] \right\rangle$$

$$B_4 = -2\rho^2(m/n)m \Big\langle S_{g_n}(m,\hat{\theta}_n) - S_g(m,\theta) - S_{g_n}(k,\hat{\theta}_n) + S_g(k,\theta), \quad \left(1-\frac{m}{n}\right) \mathbb{E}[g(X_1+\delta_n)(\varepsilon_1+\delta_n)] \Big\rangle$$

$$B_5 = 2\rho^2(m/n)m \left\langle S_g(m,\theta) - E(m,n) - S_g(k,\theta) + E(k,n),\, \left(1-\frac{m}{n}\right) \mathbb{E}[g(X_1+\delta_n)(\varepsilon_1+\delta_n)] \right\rangle$$

$$B_6 = -\left(\rho^2(m/n)m^2 - \rho^2(k/n)k^2\right) \left\| (1-\frac{m}{n}) \mathbb{E}[g(X_1+\delta_n)(\varepsilon_1+\delta_n)] \right\|^2\,.$$

Thus, analogously as in Theorem 2.4.4 we get

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_1| = O_P(C\mathcal{D}_n^{-1/2})\,, \qquad \max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_2| = O_P(\sqrt{n})\,,$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_3| = O_P(\sqrt{n})\,, \qquad \max_{k\in(m-C\mathcal{D}_n^{-2},m)} |B_4| = o_P(n)\,,$$

and for $k \in (m - C\mathcal{D}_n^{-2}, m)$

$$\frac{\rho^{-2}(m/n)}{m(1-\frac{m}{n})} B_5 = \left((\mathbb{E}[g(X_1+\delta_n)(\varepsilon_1+\delta_n)])^{\mathsf{T}} \sum_{t=k+1}^{m} (g(X_t)\varepsilon_t) + O_P\left(\frac{C}{\mathcal{D}_n\sqrt{n}}\right) \right)\,.$$

With the FCLT for i.i.d. random vectors we have

$$\left\{ \mathcal{D}_n \left(\sum_{t=s\mathcal{D}_n^{-2}}^{-1} (g(X_t)\varepsilon_t) \right),\ s \in (0,1) \right\} \xrightarrow{d} \left\{ W_s', ßs \in (0,1) \right\} ,$$

with W_s' being a Wiener process with covariance matrix $\Sigma_{g\varepsilon}$. Then, we can conclude

$$\left\{ \frac{d_n^{\mathrm{T}}}{\mathcal{D}_n} \Sigma_{g\varepsilon}^{\frac{1}{2}} \mathcal{D}_n \left(\Sigma_{g\varepsilon}^{-\frac{1}{2}} \sum_{t=s\mathcal{D}_n^{-2}}^{-1} (g(X_t)\varepsilon_t) \right),\ s \in (0,1) \right\} \xrightarrow{d} \{\mathcal{D}W(s),\ s \in (0,1)\} ,$$

with $\{W(s)\}$ being a standard Wiener process, $d_n := \mathbb{E}[g(X_1 + \delta_n)(\varepsilon_1 + \delta_n)]$ and $D = \lim_{n\to\infty} \frac{\|d_n\|_{\Sigma_{g\varepsilon}}}{\|d_n\|}$, which clearly exists.

As $\rho(s)$ is equal to $w_{\eta,\gamma}(s)$ from (2.58), we have

$$\begin{aligned} \max_{k\in(m-C\mathcal{D}_n^{-2},m)} \frac{\rho^{-2}(m/n)}{m(1-\frac{m}{n})} V_k &= \max_{k\in(m-C\mathcal{D}_n^{-2},m)} V_k^{(1)} \\ &= \max_{k\in(m-C\mathcal{D}_n^{-2},m)} \left(2\frac{d_n^T A}{\mathcal{D}_n} \mathcal{D}_n \xi_{m,\lfloor sn \rfloor}^{(1)} - 2(m-k) g_\lambda(k/n) \mathcal{D}_n^2 + o_P(1) \right) . \end{aligned}$$

Doing the same for $m < k < m+C$, we can argue analogously as in the proofs of the Theorem 2.3.3. ■

4.2. Non-linear (auto-)regressive processes and neural network functions

We considered in section 3.2 the change-point test for a non-linear (auto-)regressive model with unknown regression function. For the change-point test we used the general approximation property of neural networks and derived the test statistic for possible misspecified models. For the non-linear autoregressive model Kirch and Tadjuidje Kamgaing [2012] analysed a univariate test statistic. We used results from Kirch and Tadjuidje Kamgaing [2012] and showed the asymptotics for the multivariate test for a non-linear (auto-)regressive model.

As the estimator is usually based on the test statistic we are going to analyse the corresponding estimator to the test statistic for non-linear (auto)-regressive model derived in section 3.2. Before we do so, let us notice that Kirch and Tadjuidje Kamgaing [2012] also introduced a change-point estimator given as

$$\hat{m} = \arg\max \left\{ \left| S_H(k;\hat{\theta}_n) \right| : 1 \leq k < n \right\} ,$$

where $S_H(k;\hat{\theta}_n)$ is the partial sum of estimated residuals using the least-squares estimator $\hat{\theta}_n$, i.e.

$$\begin{aligned} S_H(k;\hat{\theta}_n) &= \sum_{t=1}^{k} \hat{\varepsilon}_t = \sum_{t=1}^{k} \Big(X_t - f(\mathbb{X}_t, \hat{\theta}_n) \Big) , \\ \hat{\theta}_n &= \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} (X_t - f(\mathbb{X}_t, \theta))^2 \end{aligned} \tag{4.5}$$

for the sample $X_{-p}, \ldots, X_n$. In Kirch and Tadjuidje Kamgaing [2012] Corollary 3.1 they showed $\hat{m} - m = o_p(n)$, with $m = \lfloor \lambda n \rfloor$, i.e it defines a consistent estimator for the change-point λ.

As we have discussed in section 2.3 the weighted version of the change-point estimator allows better estimation for different positions of the change-point. We consider the non-linear (auto-)regressive model introduced in section 3.2.1 and analyse the weighted multivariate change-point estimator given as

$$\begin{aligned} \hat{m}(\eta,\gamma;A) &= \arg\max_{1\leq k<n} w_{\eta,\gamma}(k/n) \left\| S(k,\hat{\theta}_n) \right\|_A , & (4.6) \\ S(k,\theta) &= \sum_{t=1}^{k} \nabla f(Y_t,\theta)(X_t - f(Y_t,\theta)) \\ &= \sum_{t=1}^{k} q(t,\theta) , \end{aligned}$$

where $w_{\eta,\gamma}$ as in **N.7** and A fulfils **G.3**. We show that the weighted multivariate change-point estimator is consistent and determine the best convergence rate. To verify that it is the best rate, we determine the asymptotic distribution.

4.2.1. Preliminary results

Before we state the main results we first determine some properties of sums and maxima of neural networks and its derivatives. To derive asymptotic results for the change-point estimator we first state a result following from (3.29) and (3.30).

Lemma 4.2.1

For the model (3.24) *we assume* ***L.1**- **L.5** and **L.8**. We get for all $\theta \in \Theta$ and $z = 1, 2$*

$$\frac{1}{l}\sum_{t=1}^{l}\Big(\nabla f(Y_t^{(z)},\theta)f(Y_t^{(z)},\theta) - \nabla f(Y_t^{(z)},\hat{\theta}_n)f(Y_t^{(z)},\hat{\theta}_n)\Big) = \mathbb{E}\Big(\nabla f(Y_1^{(z)},\theta)f(Y_1^{(z)},\theta) - \nabla f(Y_1^{(z)},\tilde{\theta})f(Y_1^{(z)},\tilde{\theta})\Big) + o_{a.s.}(1) \tag{4.7}$$

for $l \to \infty$.

Remark, that because of the ergodicity of $\{f(Y_t^{(z)},\theta)\}$ and because of the properties of $o_{a.s.}$ and $O_{a.s.}$ it holds for $m \to \infty$

$$\max_{1\le l\le m}\left\|\frac{1}{l}\sum_{t=1}^{l}\nabla f(Y_t^{(z)},\theta)\, f(Y_t^{(z)},\theta) - \mathbb{E}\Big(\nabla f(Y_1^{(z)},\theta)\, f(Y_1^{(z)},\theta)\Big)\right\| = O_{a.s.}(1)\,.$$

Together with Lemma 4.2.1 we get for $m \to \infty$

$$\max_{1\le l\le m}\left\|\frac{1}{l}\sum_{t=1}^{l}\Big(\nabla f(Y_t^{(z)},\hat{\theta}_n)f(\mathbb{X}_t^{(z)},\hat{\theta}_n) - \mathbb{E}\Big[\nabla f(Y_1^{(z)},\tilde{\theta})f(Y_1^{(z)},\tilde{\theta})\Big]\Big)\right\| = O_{a.s.}(1)\,. \tag{4.8}$$

But this result is not strong enough to prove the consistency of the change-point estimator. Nevertheless its proof contains ideas which are needed in the proof of the consistency. Therefore, we state the following Lemma which needs the Hajek-Renyi-Typ inequality (Lemma C.2.4 b)) from section C.2.
Recall the notation **N.13**, i.e. we define $q(t,\theta) := \nabla f(Y_t,\theta)(X_t - f(Y_t,\theta))$ for $\theta \in \Theta$.

Lemma 4.2.2

*Let $q(t,\theta)$ be as in **N.13** and $d := \mathbb{E}\Big[q(1,\tilde{\theta})\Big]$. Assume that the model* (3.24) *fulfils assumption **L.1**- **L.5** and **L.8**. Then there exists $\varsigma \in (0,1)$ such that for fixed $\kappa > 0$ we have*

$$\max_{1\le k\le m-\kappa}\left\|\frac{1}{m-k}\sum_{t=k+1}^{m}\Big(q(t,\tilde{\theta}) - d\Big)\right\| = \kappa^{-\frac{\varsigma}{4+2\varsigma}}O_P(1)\,,$$

$$\max_{1\le k\le m-\kappa}\left\|\frac{1}{m-k}\sum_{t=k+1}^{m}\Big(q(t,\hat{\theta}_n) - q(t,\tilde{\theta})\Big)\right\| = o_P(1)\,,$$

$$\max_{1\le k\le m-\kappa}\frac{1}{n}\left\|\sum_{t=1}^{k}\Big(q(t,\hat{\theta}_n) - d\Big) + \sum_{t=1}^{m}\Big(q(t,\hat{\theta}_n) - d\Big)\right\| = O\left(\max_{1\le k\le m-\kappa}\frac{1}{n}\left\|\sum_{t=1}^{k}\Big(q(t,\hat{\theta}_n) - d\Big)\right\|\right) = o_P(1)\,.$$

Observe, that we used X_t, Y_t instead of $X_t^{(1)}$, $Y_t^{(1)}$ this is caused by $k \leq m$ which implies that the underlying process is then $X_t^{(1)}$ and $Y_t^{(1)}$.

To determine the asymptotic distribution we use analogous arguments as in Theorem 2.3.3. We know that for $n \to \infty$ the limit exists. To determine the limit we use a truncation argument and determine first the limit for a fixed area around m. As we only consider the A-fixed alternative, this distance is fixed. For the local alternative we have to consider the behaviour in a growing distance and would derive conditions for the rate of growing. Since we use the equivalent decomposition as for the proof of the rate. Hence we determine the following convergence.

Lemma 4.2.3

With the notation and under the assumptions of Lemma 4.2.2 and with the definitions from there, we get

$$\max_{0<m-k<C} \left\| \sum_{t=k+1}^{m} (q(t,\tilde{\theta}) - d) \right\| = O_P(1)\,,$$

$$\max_{0<m-k<C} \left\| \sum_{t=k+1}^{m} (q(t,\hat{\theta}_n) - q(t,\tilde{\theta})) \right\| = o_P(1)\,,$$

$$\max_{0<m-k<C} \left\| \sum_{t=1}^{k} (q(t,\hat{\theta}_n) - d) + \sum_{t=1}^{m} (q(t,\hat{\theta}_n) - d) \right\|$$
$$= O_P\left(\max_{0<m-k<C} \left\| \sum_{t=1}^{k} (q(t,\hat{\theta}_n) - d) \right\| \right) = O_P\left(\sqrt{m}\right)\,,$$

where C is a constant.

4.2.2. Asymptotics

The first result is that the weighted change-point estimator is consistent.

Theorem 4.2.1 *Under the model* (3.24) *and assumptions **G.1**, **L.1**- **L.5** and **L.8**, the change-point estimator $\hat{m}(\eta,\gamma)$* (4.6) *is a consistent estimator for λ if $\eta < \lambda < 1-\eta$, i.e.*

$$\frac{\hat{m}(\eta,\gamma)}{n} \xrightarrow{P} \lambda\,. \tag{4.9}$$

In Kirch and Tadjuidje Kamgaing [2012] they proved the consistency of the change-point estimator under a NLAR(p)-process for $\eta = 0$, $\gamma = 0$ and $A = (a_{ij})_{1\leq i,j\leq r}$, $r = (2+p+d)h+1$, where $a_{11} = 1$ and $a_{ij} = 0$ for $i \cdot j \neq 1$. Now we also show that the multivariate weighted change-point estimator is a consistent estimator. From the discussion in section 2.3.2 we know that for different positions of the change-point different weight functions are preferable. In the next steps we determine the rate and the asymptotic distribution.

From Corollary 2.3.2 we have seen, that due to the construction of the estimator and the $o_P(n)$ convergence for the proofs of the asymptotics observing the truncated version of the weighted change-point estimator is sufficient.

To determine the consistency rate we use Lemma 4.2.2.

Theorem 4.2.2 *Under assumptions of Theorem 4.2.1 we have*

$$\hat{m}(\eta,\gamma) - m = O_p(1)\,.$$

This Theorem gives an upper bound for the convergence rate. To verify that it is the best rate of convergence and to be able to derive a confidence interval for m, we are interested in determining the asymptotic distribution.

Theorem 4.2.3 *Let the assumptions of Theorem 4.2.1 hold true. Then*

$$\hat{m}(\eta,\gamma) - m \xrightarrow{d} \arg\max\{W(s) - |s| g(s)\mathcal{D}^2,\ s \in \mathbb{Z}\}$$

with $\mathcal{D} = \|\delta\|_A = \left\|\mathbb{E}[(\nabla f(Y_1^{(1)},\tilde{\theta})(X_1^{(1)} - f(Y_1^{(1)},\tilde{\theta}_0))]\right\|_A$,

$$W(s) = \begin{cases} 0 & , \ s = 0\,, \\ \delta^{\mathsf{T}} A \Gamma^{\frac{1}{2}} \sum_{i=s}^{-1} \xi_i^{(2)} & , \ s < 0\,, \\ \delta^{\mathsf{T}} A \Gamma^{\frac{1}{2}} \sum_{i=1}^{s} \xi_i^{(1)} & , \ s > 0\,, \end{cases}$$

$\xi_i^{(z)} = \nabla f(Y_i^{(z)},\tilde{\theta})(X_i^{(z)} - f(Y_i^{(z)},\tilde{\theta}) - \mathbb{E}[(\nabla f(Y_1^{(z)},\tilde{\theta})(X_1^{(z)} - f(Y_1^{(z)},\tilde{\theta}_0))]$ *for* $z = 1, 2$ *and*

$$g(s) = \begin{cases} (1-\gamma)(1-\lambda) + \gamma\lambda & , \ s < 0\,, \\ 0 & , \ s = 0\,, \\ \gamma(1-\lambda) + (1-\gamma)\lambda & , \ s > 0\,. \end{cases}$$

Note that remembering the original model (3.23), $\xi_i^{(z)} = e_i + g_z(Y_i^{(z)}) - f(Y_i^{(z)},\tilde{\theta}_0) - \mathbb{E}(g_z(Y_i^{(z)}) - f(Y_i^{(z)},\tilde{\theta}_0))$ is a stationary, centered at 0, α-mixing time series with $\alpha(j) = o(j^{-c})$. But in the correct specified case or if we have a non-linear regression model with iid regressors, the errors are i.i.d and we get an equivalent result as in section 2.3.

Recalling that under correct specification (i.e. g is a neural network) a neural network with $h = 0$ hidden layer is then a mean change model. In this case, we have $\xi_i^{(z)} = \varepsilon_i$ and get the same result as in Theorem 2.3.3 .

4.2.3. Proofs

Lemma 4.2.1

For the model (3.24) we assume assumption **L.1**- **L.5** and **L.8**. We get for all $\theta \in \Theta$ and $z = 1, 2$

$$\begin{aligned} \frac{1}{l}\sum_{t=1}^{l} \Big(\nabla f(Y_t^{(z)},\theta) f(Y_t^{(z)},\theta) - \nabla f(Y_t^{(z)},\hat{\theta}_n) f(Y_t^{(z)},\hat{\theta}_n)\Big) \\ = \mathbb{E}\Big(\nabla f(Y_1^{(z)},\theta) f(Y_1^{(z)},\theta) - \nabla f(Y_1^{(z)},\tilde{\theta}) f(Y_1^{(z)},\tilde{\theta})\Big) + o(1) \qquad a.s. \end{aligned} \tag{4.10}$$

for $l \to \infty$.

Proof:
First notice that it is enough to analyse each entry of the vectors the sum is taken over. Consider the following decomposition of them. Let $j = 1, \dots, (2+p+d)H+1$, then

$$\begin{aligned}
&\frac{1}{l}\sum_{t=1}^{l}\Big((\nabla f)_j(Y_t^{(z)},\theta)f(Y_t^{(z)},\theta) - (\nabla f)_j(Y_t^{(z)},\hat{\theta}_n)f(Y_t^{(z)},\hat{\theta}_n)\Big)\\
&\qquad= \frac{1}{l}\sum_{t=1}^{l}(\nabla f)_j(Y_t^{(z)},\theta)f(Y_t^{(z)},\theta) - \mathbb{E}\left[(\nabla f)_j(Y_1^{(z)},\theta)f(Y_1^{(z)},\theta)\right]\\
&\qquad\quad+\mathbb{E}\left[(\nabla f)_j(Y_1^{(z)},\theta)f(Y_1^{(z)},\theta)\right] - \mathbb{E}\left[(\nabla f)_j(Y_1^{(z)},\theta)f(Y_1^{(z)},\tilde{\theta})\right]\\
&\qquad\quad+\mathbb{E}\left[(\nabla f)_j(Y_t^{(z)},\tilde{\theta})f(Y_1^{(z)},\tilde{\theta})\right] - \frac{1}{l}\sum_{t=1}^{l}(\nabla f)_j(Y_t^{(z)},\hat{\theta}_n)f(Y_t^{(z)},\hat{\theta}_n)\\
&\qquad=: A_1 + A_2 + A_3
\end{aligned}$$

We know $\{Y_t^{(z)}\}$ is a sequence of α-mixing random vectors and f as well as $(\nabla f)_j$ are measurable functions, such that $(\nabla f)_j(Y_t^{(z)},\theta)f(Y_t^{(z)},\theta)$ is α-mixing. From (C.1) we get that $A_1 = o(1)$ *a.s.* The constant on the right hand of (4.10) if A_2. Thus, we need to show $A_3 = o(1)$ *a.s.* It holds

$$\begin{aligned}
&\left|\mathbb{E}\big[(\nabla f)_j(Y_t^{(z)},\tilde{\theta})f(Y_1^{(z)},\tilde{\theta})\big] - \frac{1}{l}\sum_{t=1}^{l}(\nabla f)_j(Y_t^{(z)},\hat{\theta}_n)f(Y_t^{(z)},\hat{\theta}_n)\right|\\
&\qquad\le \left(\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}(\nabla f)_j(Y_t^{(z)},\theta)f(Y_t^{(z)},\theta) - \mathbb{E}[(\nabla f)_j(Y_1^{(z)},\theta)f(Y_1^{(z)},\theta)]\right|\right)\\
&\qquad\quad+\mathbb{E}\left|(\nabla f)_j(Y_1^{(z)},\tilde{\theta})f(Y_1^{(z)},\tilde{\theta}) - (\nabla f)_j(Y_t^{(z)},\theta)f(Y_1^{(z)},\theta)\right|\Bigg|_{\theta=\hat{\theta}_n}.
\end{aligned}$$

Using from Corollary 3.2.2 and Theorem C.1.1 the first term on the right side is $o_{a.s.}(1)$. As we can see in the proof of this corollary, there exists an integrable dominating variable. Theorem C.1.3 gives that we can interchange limit and integration in the second term. With the consistency of the parameter estimator $\hat{\theta}_n$ given in a) in Theorem 3.2.1 and the continuity of the generating function, the desired result follows.

■

Lemma 4.2.2

Let $q(t,\theta)$ be as in **N.13** and $d := \mathbb{E}\left[q(1,\tilde{\theta})\right]$. Assume that the model (3.24) fulfils assumption **L.1**- **L.5** and **L.8**. Then there exists $\varsigma \in (0,1)$ such that for fixed $\kappa > 0$ we have

$$\max_{1\le k\le m-\kappa}\left\|\frac{1}{m-k}\sum_{t=k+1}^{m}\left(q(t,\tilde{\theta})-d\right)\right\| = \kappa^{-\frac{\varsigma}{4+2\varsigma}}O_P(1)\,, \tag{4.11}$$

$$\max_{1\le k\le m-\kappa}\left\|\frac{1}{m-k}\sum_{t=k+1}^{m}\left(q(t,\hat{\theta}_n)-q(t,\tilde{\theta})\right)\right\| = o_p(1)\,, \tag{4.12}$$

$$\begin{aligned}\max_{1\le k\le m-\kappa}\frac{1}{n}&\left\|\sum_{t=1}^{k}\left(q(t,\hat{\theta}_n)-d\right)+\sum_{t=1}^{m}\left(q(t,\hat{\theta}_n)-d\right)\right\|\\ &= O\left(\max_{1\le k\le m}\frac{1}{n}\left\|\sum_{t=1}^{k}\left(q(t,\hat{\theta}_n)-d\right)\right\|\right) = o_p(1)\,.\end{aligned} \tag{4.13}$$

Proof:
Observe, that we used X_t, Y_t is stationary since $k < m$. Again it is enough to analyse the behaviour of the maxima for each entry of the vectors. Define for every $i = 1,\ldots,(2+p+d)h+1$ the sequence $\{\zeta_i(t)\}$ with $\zeta_i(t) := q_i(-t,\tilde{\theta}) - d_i$. Then this is a zero mean, stationary and α-mixing time series. From Lemma C.2.1 we have that $\zeta_i(t)$ is again polynomial mixing with $\alpha_\zeta(j) = \alpha_X(j-p+1)$. Furthermore, we have for $\upsilon = 2+\varsigma+\Delta$, $\varsigma\in(0,1]$, that

$$\begin{aligned}\mathbb{E}[|\zeta_i(t)|^{2+\varsigma+\Delta}] =&\mathbb{E}[|\zeta_i(1)|^{2+\varsigma+\Delta}]\\ \le& c_1\max_{j=1,\ldots,p+d}\mathbb{E}[|Y_{1,j}X_1|^{2+\varsigma+\Delta}]\\ &+c_2\max_{j=1,\ldots,p+d}\mathbb{E}[|Y_{1,j}f(Y_1,\tilde{\theta})|^{2+\varsigma+\Delta}]+d^{2+\varsigma+\Delta}\\ \le& c_3\end{aligned} \tag{4.14}$$

for some constants $0 < c_1$, $c_2 < c_3 < \infty$, by $Y_{1,j} \ne X_1$ and assumption **L.1**, **L.2**, **L.3** as well as the existence of $0 < K < \infty$ such that $|f(Y_1,\tilde{\theta})| < K < \infty$ (**L.4**).

For showing (4.11) we use $\left\{\sum_{t=k+1}^{m}\left(q(t,\tilde{\theta})-d\right),\,1\le t< m-\kappa\right\}$ is distributional equal to $\left\{\sum_{t=1}^{l}\left(q(-l,\tilde{\theta})-d\right),\,\kappa< t< m\right\}$ for 4.11. Then $\zeta_i(t) := q_i(-l,\tilde{\theta})-d_i$ is again alpha mixing of polynomial rate. From Lemma C.2.4 b) with ς as in (4.14) we have

$$\begin{aligned}\max_{1\le k\le m-\kappa}\frac{1}{m-k}\left|\sum_{t=k+1}^{m}(q_i(t,\tilde{\theta})-d_i)\right| &= \max_{\kappa\le l\le m-1}\frac{1}{l}\left|\sum_{t=1}^{l}\zeta_i(t)\right|\\ &= O_P\left(\left(\sum_{l=\kappa}^{m-1}l^{-\frac{2+\varsigma}{2}}\right)^{\frac{1}{2+\varsigma}}\right)\\ &= O_P\left(\left(\int_{\kappa}^{\infty}x^{-\frac{2+\varsigma}{2}}\,dx\right)^{\frac{1}{2+\varsigma}}\right)\\ &= \kappa^{-\frac{\varsigma}{4+2\varsigma}}O_P(1)\,.\end{aligned}$$

We will use a property of the neural network to prove (4.12). Let $\theta = (\theta_1, \dots, \theta_{(2+p+d)h+1})$, with $(\hat{\theta}_n)_i \le \theta_i \le (\tilde{\theta})_i$, $i = 1, \dots, (2+p+d)h+1$. Then it holds

$$
\begin{aligned}
|q_i(t, \hat{\theta}_n) - q_i(t, \tilde{\theta})| &= |(\nabla q_i(t, \theta))^t (\hat{\theta}_n - \tilde{\theta})| \\
&\le D \left(\max_{j=1,\dots,p+d} |Y_{t,j}^2| + \max_{j=1,\dots,p+d} |Y_{t,j}^2 X_t| \right) \|\hat{\theta}_n - \tilde{\theta}\| , \qquad (4.15)
\end{aligned}
$$

from mean value theorem and from (**L.4**)

$$
\sup_{\theta \in \Theta} \|\nabla q_i(Y_t, \theta)\| \le D \left(\max_{j=1,\dots,p+d} |Y_{t,j}^2| + \max_{j=1,\dots,p+d} |Y_{t,j}^2 X_t| \right) .
$$

Observe, by Lemma C.2.1 the time series $\max_{j=1,\dots,p+d} |Y_{t,j}^2|$ is α-mixing. Hence $\{\frac{1}{m-k} \sum_{t=k+1}^m \max_{j=1,\dots,p} |Y_{t,j}^2|, 1 \le k \le m - \kappa\}$ is distributional equal to $\{\frac{1}{l} \sum_{t=1}^l \max_{j=1,\dots,p} |Y_{-l,j}^2|, \kappa \le l < m\}$. Using Theorem C.2.1 gives

$$
\begin{aligned}
\max_{1 \le k \le m-\kappa} \left| \frac{1}{m-k} \sum_{t=k+1}^m \left(q(t, \hat{\theta}_n) - q(t, \tilde{\theta}) \right) \right| &= O(1) \left\| \hat{\theta}_n - \tilde{\theta} \right\| \max_{1 \le k \le m-\kappa} \frac{1}{m-k} \sum_{t=k+1}^m \max_{j=1,\dots,p} |Y_{t,j}^2| \\
&\le O(1) o_P(1) \max_{\kappa \le l < m} \frac{1}{l} \sum_{t=1}^l \max_{j=1,\dots,p} |Y_{-l,j}^2| \\
&= o_p(1) O(1) \qquad a.s.
\end{aligned}
$$

For the norm of the difference of the parameter esitmator and its limit we used a) of Theorem 3.2.1. The result (4.12) follows from properties of the stochastic Landau symbols.

For (4.13) we get with the triangle inequality that we have

$$
\begin{aligned}
&\max_{1 \le k \le m-\kappa} \frac{1}{n} \left\| \sum_{t=1}^k \left(q(t, \hat{\theta}_n) - d \right) + \sum_{t=1}^m \left(q(t, \hat{\theta}_n) - d \right) \right\| \\
&\qquad \le \max_{1 \le k \le m-\kappa} \frac{1}{n} \left\| \sum_{t=1}^k \left(q(t, \hat{\theta}_n) - d \right) \right\| + \frac{1}{n} \left\| \sum_{t=1}^m \left(q(t, \hat{\theta}_n) - d \right) \right\| \\
&\qquad \le 2 \max_{1 \le k \le m} \frac{1}{n} \left\| \sum_{t=1}^k \left(q(t, \hat{\theta}_n) - d \right) \right\|
\end{aligned}
$$

Hence, it is enough to consider

$$
\max_{1 \le k \le m} \frac{1}{n} \left\| \sum_{t=1}^k \left(q(t, \hat{\theta}_n) - d \right) \right\| .
$$

From the triangle inequality it follows that analysing the two maxima

$$
\max_{1 \le k \le m} \left| \sum_{t=1}^k \left(q_i(t, \hat{\theta}_n) - q_i(t, \tilde{\theta}) \right) \right| \qquad \text{and} \qquad \max_{1 \le k \le m} \left| \sum_{t=1}^k \zeta_i(t) \right| , \qquad (4.16)
$$

where $\zeta_i(t) := q_i(t, \tilde{\theta}) - d_i$, is enough. The second maximum in (4.16) is $O_P(\sqrt{n})$ using $\{\zeta_i(t)\}$ is stationary α-mixing time-series and Remark C.2.1. An analogous argument as for (4.12) gives that the first maximum in (4.16) is $O_P(n)o_P(1)$. These give us (4.13).

■

Lemma 4.2.3

Under the assumptions of Lemma 4.2.2 and with the definitions from there, we get

$$\max_{0<m-k<C}\left\|\sum_{t=k+1}^{m}(q(t,\tilde{\theta})-d)\right\| = O_P(1)\,, \tag{4.17}$$

$$\max_{0<m-k<C}\left\|\sum_{t=k+1}^{m}(q(t,\hat{\theta}_n)-q(t,\tilde{\theta}))\right\| = o_P(1)\,, \tag{4.18}$$

$$\begin{aligned}\max_{0<m-k<C}&\left\|\sum_{t=1}^{k}(q(t,\hat{\theta}_n)-d)+\sum_{t=1}^{m}(q(t,\hat{\theta}_n)-d)\right\|\\ &= O_P\left(\max_{0<m-k<C}\left\|\sum_{t=1}^{k}(q(t,\hat{\theta}_n)-d)\right\|\right) = O_P\left(\sqrt{m}\right)\,,\end{aligned} \tag{4.19}$$

where C is a constant.

Proof:
As before in Lemma 4.2.2 we consider the entries of the vectors and show their asymptotic behaviour.
To show (4.17), note that by **L.1**, **L.2** and **L.3** we can use the invariance principle and get

$$\begin{aligned}\max_{0<m-k<C}\left|\sum_{t=k+1}^{m}\zeta_i(t)\right| &\leq \max_{0<m-k<C}(m-k)^{\frac{1}{2}}\max_{0<m-k<C}\frac{1}{(m-k)^{\frac{1}{2}}}\left|\sum_{t=k+1}^{m}\zeta_i(t)\right|\\ &= O_P(1)\end{aligned}$$

with $\zeta_i(t) = q_i(t,\tilde{\theta}) - d_i$ fulfilling the assumptions of Theorem C.2.2 (analogously to the proof of (4.11)).

Equation (4.18) can be shown, using (4.15) and knowing that $\max_{0<m-k<C}|\max_{j=1,\dots,p+d} Y^2_{t+k,j}| = O_P(1)$, by stationarity, as well as (a)) in Theorem 3.2.1 hold true.

For (4.19), we first observe that the left hand side is equal to

$$\max_{0<m-k<C}\left\|2\sum_{t=1}^{k}(q(t,\hat{\theta}_n)-d)+\sum_{t=k+1}^{m}(q(t,\hat{\theta}_n)-d)\right\|\,. \tag{4.20}$$

Now, it follows

$$\begin{aligned}\max_{0<m-k<C}\sum_{t=k+1}^{m}(q_i(t,\hat{\theta}_n)-d_i) \leq &\max_{0<m-k<C}\left|\sum_{t=k+1}^{m}(q_i(t,\tilde{\theta})-d_i)\right|\\ &+\max_{0<m-k<C}\sum_{t=k+1}^{m}\left|q_i(t,\hat{\theta}_n)-q_i(t,\tilde{\theta})\right|\\ \leq &O_P(1)+o_P(1)\,,\end{aligned} \tag{4.21}$$

where the behaviour of the first term follows by stationarity. For the second sum we used the Lipschitz-inequality as in (4.15), which leads to

$$\max_{0<m-k<C} \sum_{t=k+1}^{m} \left| q_i(t, \hat{\theta}_n) - q_i(t, \tilde{\theta}) \right| \leq c_1 \left\| \hat{\theta}_n - \tilde{\theta} \right\| \max_{0<m-k<C} \sum_{t=k+1}^{m} \max_{j=1,\ldots,p+d} |Y_{t,j}^2| \,, \tag{4.22}$$

where c_1 is some constant. From Theorem 3.2.1 we have $\|\hat{\theta}_n - \tilde{\theta}_0\| = o_P(1)$ using stationarity of $\{Y_t^2\}$ we get

$$\max_{0<m-k<C} \sum_{i=k+1}^{m} \max_{j=1,\ldots,p+d} |Y_{t,j}^2| = O_P(1) \,. \tag{4.23}$$

Combine (4.23) with (4.22) and (4.21) is shown.

It is left to observe the behaviour of the first term in (4.20). We use the same decomposition as for the second term and apply the Lipschitz-inequality, too. Thus, we gain

$$\max_{0<m-k<C} \left| \sum_{t=1}^{k} (q_i(t, \hat{\theta}_n) - d_i) \right| \leq \max_{k \in (m-C, m)} \left| \sum_{t=1}^{k} (q_i(t, \tilde{\theta}) - d_i) \right|$$
$$+ \|\hat{\theta}_n - \tilde{\theta}\| \max_{k \in (m-C, m)} \sum_{t=1}^{k} \max_{j=1,\ldots,p+d} |Y_{t,j}^2| \,.$$

Using $\{\max_{j=1,\ldots,p+d} |Y_{t,j}^2|\}$ is stationary, α-mixing time series, Theorem C.2.1 gives us that the second term is $O_P(\sqrt{m})$. For the first term we apply the Hajek-Renyi Inequality (Lemma C.2.4), which gives the claim.

■

Lemma 4.2.4

Under the model (3.24) *and assumptions* ***G.1***, ***L.1***- ***L.5*** *and* ***L.8*** *we have that*

$$\sup_{s \in [0,1]} \left| \frac{1}{n} \left\| S(\lfloor sn \rfloor, \hat{\theta}_n) \right\|_A - \left\| E_s(\tilde{\theta}) \right\|_A \right| = o_P(1) \,,$$

with $E_s(\theta) = g(s)\mathbb{E}[\nabla f(Y_1, \theta)(X_1 - f(Y_1, \theta))]$ *and*

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

Proof:
From triangle inequality it follows

$$\sup_{s \in [0,1]} \left| \frac{1}{n} \left\| S(\lfloor sn \rfloor, \hat{\theta}_n) \right\|_A - \left\| E_s(\tilde{\theta}) \right\|_A \right|$$
$$\leq \sup_{s \in [0,1]} \left\| \frac{1}{n} S(\lfloor sn \rfloor, \hat{\theta}_n) - E_s(\tilde{\theta}) \right\|_A \,.$$

Furthermore, we have

$$\begin{aligned}\sup_{s\in(0,1)}&\left\|\frac{1}{n}S(\lfloor sn\rfloor,\hat{\theta}_n)-E_s(\tilde{\theta})\right\|_A\\ &\le \sup_{s\in(0,1)}\left\|\frac{1}{n}S(\lfloor sn\rfloor,\hat{\theta}_n)-\frac{1}{n}S(\lfloor sn\rfloor,\tilde{\theta})\right\|_A+\sup_{s\in(0,1)}\left\|\frac{1}{n}S(\lfloor sn\rfloor,\tilde{\theta})-E_s(\tilde{\theta})\right\|_A \\ &= o_P(1)\,.\end{aligned} \tag{4.24}$$

The second term on the right hand side can be proven equivalently to b)of Proposition 3.2.1. It is left to prove the convergence of the first supremum in 4.24.

Assume $s \le \lambda$ the other side follows from splitting the sum at λ. We use the mean value theorem and the $\sqrt{n}$-consistency of $\hat{\theta}_n$. By the properties of the neural network we get $\sup_{\theta\in\Theta}\|\nabla q_i(t,\theta)\|_\infty \le c(\max_{j=1,\dots,p+d}|X_tY_{t,j}^2| + \max_{j=1,\dots,p+d}|Y_{t,j}^2|)$ for some constant $c>0$. Since by the assumptions **L.1** and **L.3** both maxima are stationary α-mixing time-series, we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{\lfloor sn\rfloor} q_i(t,\hat{\theta}_n)-q_i(t,\tilde{\theta}) &\le \frac{1}{n}\sum_{t=1}^{\lfloor sn\rfloor}\sup_{\theta\in\Theta}\|\nabla q_i(t,\theta)\|_\infty\left\|\hat{\theta}_n-\tilde{\theta}\right\|\\ &\le \left\|\hat{\theta}_n-\tilde{\theta}\right\|\left(\frac{1}{n}\sum^{\lfloor sn\rfloor}\max_{j=1,\dots,p+d}|X_tY_{t,j}^2|+\frac{1}{n}\sum_{t=1}^{\lfloor sn\rfloor}\max_{j=1,\dots,p+d}|Y_{t,j}^2|\right)\\ &= O_P(n^{-\frac{1}{2}})O_P(1)\,.\end{aligned}$$

Since the convergence holds for every entry of the difference uniformly, it holds for the norm of the difference vector.

■

Theorem 4.2.1

Under the model (3.24) and assumptions **G.1**, **L.1**- **L.5** and **L.8**, we get the change-point estimator $\hat{m}(\eta,\gamma)$ (4.6) defines an consistent estimator for λ if $\eta<\lambda<1-\eta$, i.e.

$$\frac{\hat{m}(\eta,\gamma)}{n}\xrightarrow{p}\lambda\,. \tag{4.25}$$

Proof:
First observe, that in Kirch and Tadjuidje Kamgaing [2012] they essentially used

$$\sup_{s\in[0,1]}\left|\frac{1}{n}\left|\sum_{t=1}^{\lfloor sn\rfloor}(X_t-f(\mathbb{X},\hat{\theta}_n))\right|-L_n(s)\right|=o_P(1)\,,$$

where

$$L_n(s)=\begin{cases} s\left|\mathbb{E}[X_1^{(1)}-f(\mathbb{X}_1^{(1)},\theta)]|_{\theta=\hat{\theta}_n}\right|, & s<\lambda\\ \max\left(\lambda\left|\mathbb{E}[X_1^{(1)}-f(\mathbb{X}_1^{(1)},\theta)]|_{\theta=\hat{\theta}_n}\right|,(1-s)\left|\mathbb{E}[X_1^{(2)}-f(\mathbb{X}_1^{(2)},\theta)]|_{\theta=\hat{\theta}_n}\right|\right), & s=\lambda\\ (1-s)\left|\mathbb{E}[X_1^{(2)}-f(\mathbb{X}_1^{(2)},\theta)]|_{\theta=\hat{\theta}_n}\right|. & s>\lambda\end{cases}$$

We instead use the result of Lemma 4.2.4 and combine it with Theorem C.1.2.

■

Theorem 4.2.2

Under the model (3.24) and assumptions **L.1- L.5** and **L.8**, we get for $\eta < \lambda < 1-\eta$

$$\hat{m}(\eta,\gamma) - m = O_P(1)\,.$$

Proof:
Observe that by Corollary 2.3.2 the change-point estimator $\hat{m}(\eta,\gamma)$ is asymptotically equal to $\hat{m}(\alpha,\gamma)$ with $0<\alpha<\min(\lambda,1-\lambda)$. For the proof we therefore consider $\hat{m}=\hat{m}(\alpha,\gamma)$.

With the notation **N.7**, we define

$$\hat{m} = \arg\max\{w(\eta,\gamma)\left\|S_k(\hat{\theta}_n)\right\|_A : 1 \le k < n\} = \arg\max\{V_k, \alpha_n < k < (1-\alpha)n\}\,,$$

where

$$V_k = -\langle w_\gamma(m/n)S_m(\hat{\theta}_n) - w_\gamma(k/n)S_k(\hat{\theta}_n), w_\gamma(m/n)S_m(\hat{\theta}_n) + w_\gamma(k/n)S_k(\hat{\theta}_n)\rangle_A\,.$$

We have to check (2.42) for this model. As we done there, we check each side separately.

Lets consider the first event $0<\hat{m}<m-\kappa$. Thus, we have to determine

$$P(\hat{m}<m-\kappa) = P(0<\hat{m}<m-\kappa) = P\left(\max_{\alpha n<k<m-\kappa} V_k \ge \max_{m-\kappa\le k<(1-\alpha)n} V_k\right)\,.$$

We observe that V_k can be written in the following way

$$\begin{aligned}
V_k = &-\Big\langle (w_\gamma(m/n)-w_\gamma(k/n))S_m(\hat{\theta}_n) + w_\gamma(k/n)S_{k+1,n}(\hat{\theta}_n),\\
&\qquad w_\gamma(m/n)S_{k+1,m}(\hat{\theta}_n) + (w_\gamma(m/n)+w_\gamma(k/n))S_m(\hat{\theta}_n)\Big\rangle_A\\
= &-\Bigg\langle (w_\gamma(m/n)-w_\gamma(k/n))\sum_{t=1}^{m}\Big(q(t,\hat{\theta}_n)-\mathbb{E}[q(1,\tilde{\theta})]\Big) + w_\gamma(k/n)\sum_{t=k+1}^{m}\Big(q(t,\hat{\theta}_n)-q(t,\tilde{\theta})\Big)\\
&\quad + w_\gamma(k/n)\sum_{t=k+1}^{m}\Big(q(t,\tilde{\theta})-\mathbb{E}[q(1,\tilde{\theta})]\Big) + (mw_\gamma(m/n)-kw_\gamma(k/n))\mathbb{E}[q(1,\tilde{\theta})],\\
&\quad w_\gamma(m/n)\sum_{t=k+1}^{m}\Big(q(t,\hat{\theta}_n)-\mathbb{E}[q(1,\tilde{\theta})]\Big) + (w_\gamma(m/n)+w_\gamma(k/n))\sum_{t=1}^{k}\Big(q(t,\hat{\theta}_n)-\mathbb{E}[q(1,\tilde{\theta})]\Big)\\
&\quad + (mw_\gamma(m/n)+kw_\gamma(k/n))\mathbb{E}[q(1,\tilde{\theta})]\Bigg\rangle_A\\
= &\; B_1+B_2+B_3+B_4+B_5+B_6\,.
\end{aligned}$$

Where we use the equivalent calculation as for (2.44) with replacing

$$\varepsilon_t-\bar{\varepsilon}_n \hat{=} q(t,\hat{\theta}_n)-\mathbb{E}[q(1,\tilde{\theta})]\,,\qquad \bar{\varepsilon}_n \hat{=} -\left(q(t,\hat{\theta}_n)-q(t,\tilde{\theta})\right)\qquad \varepsilon_t \hat{=} q(t,\tilde{\theta})-\mathbb{E}[q(1,\tilde{\theta})]\,.$$

Thus, we have B_i, $i = 1, \ldots, 6$, given as

$$B_1 = (w_{2\gamma}(k/n) - w_{2\gamma}(m/n)) \left\| \sum_{t=1}^{k} (q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta}))\right\|_A^2$$

$$B_2 = -w_{2\gamma}(m/n) \left\langle \sum_{t=1}^{k} \left(q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta})]\right) + \sum_{t=1}^{m} \left(q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta})]\right), \sum_{t=k+1}^{m} \left(q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta})]\right) \right\rangle_A$$

$$B_3 = -2(w_{2\gamma}(m/n)m - w_{2\gamma}(k/n)k) \left\langle \sum_{t=1}^{k} \left(q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta})]\right), \mathbb{E}[q(1, \tilde{\theta})] \right\rangle_A$$

$$B_4 = 2w_{2\gamma}(m/n)m \left\langle \sum_{t=k+1}^{m} \left(q(t, \hat{\theta}_n) - q(t, \tilde{\theta})\right), \mathbb{E}[q(1, \tilde{\theta})] \right\rangle_A$$

$$B_5 = 2w_{2\gamma}m \left\langle \sum_{t=k+1}^{m} \left(q(t, \tilde{\theta}) - \mathbb{E}[q(1, \tilde{\theta})]\right), \mathbb{E}[q(1, \tilde{\theta})] \right\rangle_A$$

$$B_6 = -(m^2 w_\gamma^2(m/n) - k^2 w_\gamma^2(k/n)) \left\| \mathbb{E}[q(1, \tilde{\theta})] \right\|_A^2$$

Let $\delta_n \equiv \delta := \mathbb{E}[q(1, \tilde{\theta})]$, then $B_6 = -(m^2 w_{\eta,\gamma}^2(m/n) - k^2 w_{\eta,\gamma}^2(k/n))\|\delta\|_A^2$. This is equivalent to B_6 from Theorem 2.3.2 (see page 43). Thus, we have the same approximation as in (2.45). Hence, it is enough to examine the order of

$$\max_{1 \le k < m} \left| \frac{B_i}{m-k} \right|$$

for $i = 1, \ldots, 5$. By Lemma 4.2.2 there exists ς such that for chosen large enough κ, we get

$$\begin{aligned}
\max_{\alpha n < k < m - \kappa} \left| \frac{B_1}{(m-k)} \right| &\le \max_{\alpha n < k < m - \kappa} \left| \frac{w_{2\gamma}(k/n) - w_{2\gamma}(m/n)}{(m-k)} \right| \max_{\alpha n < k < m - \kappa} \left\| \sum_{t=1}^{k} (q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta}))\right\|_A^2 \\
&= O(n^{-1}) \, (o_P(n)) \ .
\end{aligned}$$

We used the Lipschitz property of the weight function (2.28), equivalently as for (2.46). For the following approximations we will also make use of the properties of the weight function given in Corollary 2.3.1.

$$\begin{aligned}
&\max_{\alpha n < k < m - \kappa} \left| \frac{B_2}{(m-k)} \right| \\
&\le w_{2\gamma}(m/n) \max_{\alpha n < k < m - \kappa} \left\| \sum_{t=1}^{k} \left(q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta})]\right) + \sum_{t=1}^{m} \left(q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta})]\right) \right\|_A \\
&\quad \cdot \max_{\alpha n < k < m - \kappa} \left\| \sum_{t=k+1}^{m} \left(q(t, \hat{\theta}_n) - \mathbb{E}[q(1, \tilde{\theta})]\right) \right\|_A \\
&= O(1) \, (o_p(1))
\end{aligned}$$

Therefore, we used that from the triangle inequality we get

$$\begin{aligned}
&\max_{1\le k\le m-\kappa} \frac{1}{n}\left|\left(\sum_{t=1}^{m}\left(q_i(t,\hat\theta_n)\ \ -d_i\right)+\sum_{t=1}^{k}\left(q_i(t,\hat\theta_n)-d_i)\right)\right)\right| \\
&\le \frac{1}{n}\left(\max_{1\le k\le m-\kappa}\left|\sum_{t=1}^{m}\left(q_i(t,\hat\theta_n)-d_i\right)\right|+\max_{1\le k\le m-\kappa}\left|\sum_{t=1}^{k}\left(q_i(t,\hat\theta_n)-d_i)\right)\right|\right) \\
&\le \frac{2}{n}\max_{1\le k\le m-\kappa}\left|\sum_{t=1}^{k}\left(q_i(t,\hat\theta_n)-d_i)\right)\right| .
\end{aligned}$$

For B_3 and B_4 we directly can apply Corollary 2.3.1. and get

$$\begin{aligned}
&\max_{\alpha n<k<m-\kappa}\left|\frac{B_3}{(m-k)}\right| \\
&\quad\le \max_{\alpha n<k<m-\kappa}\left|\frac{w_{2\gamma}(m/n)m-w_{2\gamma}(k/n)k}{(m-k)}\right|\max_{\alpha n<k<m-\kappa}\left\|\sum_{t=1}^{k}\left(q(t,\hat\theta_n)-\mathbb{E}[q(1,\tilde\theta)]\right)\right\|_A\|\delta\|_A \\
&\quad= O(1)O_P(\sqrt{n}\|\delta\|_A)\,, \\
&\max_{\alpha n<k<m-\kappa}\left|\frac{B_4}{(m-k)}\right| \\
&\quad\le 2w_{2\gamma}(m/n)m\max_{\alpha n<k<m-\kappa}\left\|\frac{1}{m-k}\sum_{t=k+1}^{m}\left(q(t,\hat\theta_n)-q(t,\tilde\theta)\right)\right\|_A\|\delta\|_A \\
&\quad= O(n)o_P(1)\|\delta\|_A\,.
\end{aligned}$$

Observe, that for B_4 we used $\sqrt{n}$-consistency of $\hat\theta_n$ and that $\{Y_t^2\}$ is a α-mixing stationary time series.

$$\begin{aligned}
&\max_{\alpha n<k<m-\kappa}\left|\frac{B_5}{(m-k)}\right| \\
&\qquad\le 2w_{2\gamma}(m/n)m\max_{\alpha n<k<m-\kappa}\left\|\frac{1}{m-k}\sum_{t=k+1}^{m}\left(q(t,\tilde\theta)-\mathbb{E}[q(1,\tilde\theta)]\right)\right\|_A\|\delta\|_A \\
&\qquad= O(n)O_P(\kappa^{-\frac{\varsigma}{6}})\|\delta\|_A\,.
\end{aligned}$$

It leads to

$$\begin{aligned}
\max_{\alpha n\le k<m-\kappa}\left|\frac{B_1}{B_6}\right| &= o_P\left(1\right)\,, \\
\max_{\alpha n\le k<m-\kappa}\left|\frac{B_2}{B_6}\right| &= \kappa^{-\frac{\varsigma}{6}}O_P\left(1\right)\,, \\
\max_{\alpha n\le k<m-\kappa}\left|\frac{B_3}{B_6}\right| &= o_P\left(1\right)\,, \\
\max_{\alpha n\le k<m-\kappa}\left|\frac{B_4}{B_6}\right| &= o_P\left(1\right)\,, \\
\max_{\alpha n\le k<m-\kappa}\left|\frac{B_5}{B_6}\right| &= o_P\left(1\right)\,.
\end{aligned}$$

Using this results and $B_6 < 0$ for $k < m$, with $\max_{1\leq k<m-\kappa} B_6 \leq c < 0$ for c fixed but arbitrary, it yields

$$\begin{aligned}
P(\hat{m} < m-\kappa) &= P\left(\max_{1\leq k<m-\kappa} V_k > \max_{m-\kappa\leq k\leq n} V_k\right)\\
&\leq P\left(\max_{1\leq k<m-\kappa} V_k \geq V_m\right)\\
&= P\left(\max_{1\leq k<m-\kappa} V_k \geq 0\right)\\
&= P\left(\max_{1\leq k<m-\kappa} \left\{B_6\left(1+o_P(1)+\frac{B_2}{B_6}\right)\right\} \geq 0\right)\\
&\leq P\left(\max_{1\leq k<m-\kappa} \left|\frac{B_2}{B_6}\right| \geq 1+o_P(1)\right)\\
&\leq P\left(\max_{1\leq k<m-\kappa} \left|\frac{B_2}{B_6}\right| \geq 1-\eta\right) + P\left(1+o_p(1) \leq \max_{1\leq k<m-\kappa} \left|\frac{B_2}{B_6}\right| \leq 1-\eta\right)\\
&\leq P\left(O_P(1) \geq (1-\eta)\kappa^{\frac{5}{6}}\right) + o(1)\,,
\end{aligned}$$

with $0 < \eta < 1$ arbitrary. This term becomes arbitrarily small for a sufficiently large $\kappa > 0$. The other way around works similar, just using that $\sum_{i=1}^{k}(X_i - f(Y_i,\hat{\theta}_N)) = -\sum_{i=k+1}^{m}(X_i - f(Y_i,\hat{\theta}_N))$.

■

Theorem 4.2.3

Let the assumptions of Theorem 4.2.2 hold true. Then

$$\hat{m}(\eta,\gamma;A) - m \xrightarrow{d} \arg\max\{W(s) - |s|g(s)\mathcal{D}^2,\ s\in\mathbb{Z}\}$$

with $\mathcal{D} = \|\delta\|_A = \left\|\mathbb{E}[(\nabla f(Y_1^{(1)},\tilde{\theta})(X_1^{(1)} - f(Y_1^{(1)},\tilde{\theta}))]\right\|_A$,

$$W(s) = \begin{cases} 0 & ,\ s=0\,,\\ \delta^{\mathsf{T}} A\Gamma^{\frac{1}{2}} \sum_{i=s}^{-1}\xi_i^{(2)} & ,\ s<0\,,\\ \delta^{\mathsf{T}} A\Gamma^{\frac{1}{2}} \sum_{i=1}^{s}\xi_i^{(1)} & ,\ s>0\,,\end{cases}$$

$\xi_i^{(z)} = \nabla f(Y_i^{(z)},\tilde{\theta})(X_i^{(z)} - f(Y_i^{(z)},\tilde{\theta}) - \mathbb{E}[(\nabla f(Y_1^{(z)},\tilde{\theta})(X_1^{(z)} - f(Y_1^{(z)},\tilde{\theta}_0))]$ for $z = 1,2$ and

$$g(s) = \begin{cases} (1-\gamma)(1-\lambda)+\gamma\lambda & ,\ s<0\,,\\ 0 & ,\ s=0\,,\\ \gamma(1-\lambda)+(1-\gamma)\lambda & ,\ s>0\,.\end{cases}$$

Proof:
To simplify the notation we use $\hat{m} := \hat{m}(\eta,\gamma;A)$, as η, γ and A are fixed.

As in the proof of Theorem 2.3.3 (see page 50) it is only of interest to determine the asymptotic distribution of

$$P(m-\kappa \leq \hat{m} \leq m+x, |\hat{m}-m| < \kappa) = P\left(\max_{(k-m)\in(-\kappa,x]} V_k \geq \max_{(k-m)\in(x,\kappa)} V_k\right)\,,$$

(compare (2.56)). V_k is as in Theorem 4.2.2. For the proof we have to analyse V_k for $k \in (m-\kappa, m)$ and for $k \in (m, m+\kappa)$. Here, we show that with Lemma 4.2.3 and analoguous argumentation as in the proof of Theorem 2.3.3 we can determine the asymptotic distribution for $k \in (m-\kappa, m)$. The other side follows equivalently.

Let (w.l.o.g.) $x > 0$ and $\kappa > x$ be both fixed but arbitrary. Then we get from Corollary 2.3.1 and Lemma 4.2.3

$$\begin{aligned}
\max_{k\in(m-C,m)} |B_1| &= \max_{k\in(m-C,m)} |w_{2\gamma}(k/n) - w_{2\gamma}(m/n)| \left\| \sum_{t=1}^{k} (q(t,\hat{\theta}_n) - \mathbb{E}[q(1,\tilde{\theta}))\right\|_A^2 \\
&= O_P\left(n^{-1}\right) O_P(n)\,, \\
\max_{k\in(m-C,m)} |B_2| &\leq w_{2\gamma}(m/n) \max_{k\in(m-C,m)} \left\| \sum_{t=1}^{k} \left(q(t,\hat{\theta}_n) - \mathbb{E}[q(1,\tilde{\theta})]\right) + \sum_{t=1}^{m} \left(q(t,\hat{\theta}_n) - \mathbb{E}[q(1,\tilde{\theta})]\right)\right\|_A \\
&\qquad \cdot \max_{k\in(m-C,m)} \left\| \sum_{t=k+1}^{m} \left(q(t,\hat{\theta}_n) - \mathbb{E}[q(1,\tilde{\theta})]\right)\right\|_A \\
&= O_P(1) o_P\left(\sqrt{n}\right)\,, \\
\max_{k\in(m-C,m)} |B_3| &\leq 2 \max_{k\in(m-C,m)} |w_{2\gamma}(m/n)m - w_{2\gamma}(k/n)k| \max_{k\in(m-C,m)} \left\| \sum_{t=1}^{k} \left(q(t,\hat{\theta}_n) - \mathbb{E}[q(1,\tilde{\theta})]\right)\right\|_A \\
&\qquad \cdot \max_{k\in(m-C,m)} \left\| \mathbb{E}[q(1,\tilde{\theta})]\right\|_A \\
&= O_P(1) O_P(\sqrt{n})\,, \\
\max_{k\in(m-C,m)} |B_4| &\leq 2 w_{2\gamma}(m/n) m \max_{k\in(m-C,m)} \left\| \sum_{t=k+1}^{m} \left(q(t,\hat{\theta}_n) - q(t,\tilde{\theta})\right)\right\|_A \left\| \mathbb{E}[q(1,\tilde{\theta})]\right\|_A \\
&= O_P(n) o_P(1)\,, \\
\max_{k\in(m-C,m)} |B_5| &= 2 w_{2\gamma} m \max_{k\in(m-C,m)} \left\| \sum_{t=k+1}^{m} \left(q(t,\tilde{\theta}) - \mathbb{E}[q(1,\tilde{\theta})]\right)\right\|_A \left\| \mathbb{E}[q(1,\tilde{\theta})]\right\|_A \\
&= O_P(n) O_P(1)\,, \\
\max_{k\in(m-C,m)} |B_6| &= \max_{k\in(m-C,m)} |m^2 w_{\gamma}^2(m/n) - k^2 w_{\gamma}^2(k/n)| \left\| \mathbb{E}[q(1,\tilde{\theta})]\right\|_A^2 = O(n)\,.
\end{aligned}$$

This coincides with the results in the proof of Theorem 2.3.3 (page 50). Observe, that the result (2.58) only depends on the weight function. Thus we get the same asymptotic for B_6 (with $(1-\lambda)\delta = \mathbb{E}[q(1,\tilde{\theta})]$ in (2.58)). Analogously, we conclude

$$\begin{aligned}
&\max_{l\in(-\kappa,0)} \frac{w_{-2\gamma}(m/n)}{m} V_{m+l} \\
&\quad = \max_{l\in(-\kappa,0)} \left(2\delta^{\mathrm{T}} A \Gamma^{\frac{1}{2}} \sum_{t=l}^{-1} \tilde{\zeta}(m+t+1) - 2|l| \frac{\gamma\lambda + (1-\gamma)(1-\lambda)}{(1-\lambda)} \mathcal{D} + o_P(1)\right)\,,
\end{aligned}$$

with $\zeta(t) = q(t, \tilde{\theta}) - \mathbb{E}[q(1, \tilde{\theta})] = \nabla f(Y_t, \tilde{\theta})(X_t - f(Y_t, \tilde{\theta}) - \mathbb{E}[\nabla f(Y_t, \tilde{\theta})(X_t - f(Y_t, \tilde{\theta})]$ and $\mathcal{D} = \|\delta\|_A^2$. Define $\xi_t^{(1)} \overset{d}{=} \tilde{\zeta}(-t)$, again stationary α-mixing time-series, and

$$V_l^{(1)} = 2\delta^{\mathsf{T}} A \Sigma^{\frac{1}{2}} \sum_{t=l}^{-1} \xi_t^{(1)} - 2|l| \frac{(1-\gamma)(1-\lambda) + \gamma\lambda}{(1-\lambda)^2} \mathcal{D} .$$

Then we have asymptomatic convergence and get

$$\max_{l \in (-\kappa, 0)} \frac{w_{-2\gamma}(m/n)}{m(1 - \frac{m}{n})} V_{m+l} \xrightarrow{d} \max_{l \in (-\kappa, 0)} V_l^{(1)} .$$

Doing the same for $m < k < m + \kappa$, we get

$$\max_{l \in (0, \kappa)} \frac{w_{-2\gamma}(m/n)}{m(1 - \frac{m}{n})} V_{m+l} \xrightarrow{d} \max_{l \in (0, \kappa)} V_l^{(2)} .$$

with

$$V_l^{(2)} = 2\delta^{\mathsf{T}} A \Sigma^{\frac{1}{2}} \sum_{t=1}^{l} \xi_t^{(2)} - 2|l| \frac{\gamma(1-\lambda) + (1-\gamma)\lambda}{(1-\lambda)^2} \mathcal{D} .$$

Define

$$W_l = \begin{cases} \frac{1}{2} V_l^{(1)}, & l < 0 \\ 0, & l = 0 \\ \frac{1}{2} V_l^{(2)}, & l > 0 \end{cases} ,$$

we can conclude equivalently as in the proof of Theorem 2.3.3 (see page 50).

∎

4.3. Generalized class of change-point estimator

In section 3.3, we analysed a consistent level -α test. Now, we are interested in determining the best convergence rate of the corresponding change-point estimator, i.e.

$$\hat{m} := \arg\max_{1\le k<n} w_n(n,k)\left\|S_H(k;\hat{\theta}_n)\right\|_A , \tag{4.26}$$

with $S_H(k;\hat{\theta}_n)$ as given in (3.42) and $w_n(n,k)$ fulfils assumption **G.7**. We verify the result by determining the asymptotic distribution of the estimator.

The main idea is to decompose the statistic such that we can analyse the behaviour of the parts. To this end, we will make use of the representation of $S_H(k;\theta)$ as a sum.

4.3.1. Asymptotic results for the change-point estimator

We first consider the key regularity conditions and secondly prove them for the example 3.3.1 from section 3.3.

To handle the randomized weight function we recall that this weightfunction converges uniformly to a deterministic weight function, such that the influence vanishes asymptotically. In the proofs we use this to change directly to the deterministic function.

Lemma 4.3.1

*Let $w_n(n,k)$ denote a random weight function fulfilling assumption **G.7**. Then it holds for $\alpha \in (0,\lambda)$ and κ_n non-decreasing, that*

$$\max_{\alpha n<k<m-\kappa_n}\left|\frac{w_n(k,n)-\rho(k/n)}{m-k}\right| = o_P(\kappa_n^{-1}) = o_P(1)\,,$$
$$\max_{\alpha n<k<m-\kappa_n}\left|\frac{w_n^2(k,n)-\rho^2(k/n)}{m-k}\right| = o_P(\kappa_n^{-1}) = o_P(1)\,,$$

Consistency

First we want to show, that the rescaled change-point estimator $\frac{\hat{m}}{n}$ is consistent to λ, i.e.

$$\left|\frac{\hat{m}}{n}-\lambda\right| = o_P(1)\,.$$

Due to the fact, that we are interested on assumptions besides properties of the process, we have the following properties for the estimation function as well as for the test function. It is ensured that for the asymptotic behaviour we can replace the statistic function evaluated at $\hat{\theta}_n$ by the statistic function evaluated at $\tilde{\theta}$. Especially, we have to assume the following.

G.12 There exists $\tilde{\theta} \in \Theta$ being unique root of $\mathbb{E}[G(\mathbb{X}_1,\theta)]$ under H_1.

This is some basic assumption, as we need to define the asymptotic behaviour under H_1 of the change-point estimator. Let us state the asymptotic value under H_1 for the example of the log-normal observations.

Example 4.3.1 *We assume that for the observations introduced in Example 3.3.1 a change in the expectation of the log-normal observations occurs. This can either follow from a change in the expectation ($\delta_{n,\mu}$) or in the variance (δ_{n,σ^2}) or in both of the log-transformed observations. We denote with δ_n the change given by*

$$\exp(\mu_0 + \delta_{n,\mu} + \frac{1}{2}(\sigma_0^2 + \delta_{n,\sigma^2})) = \delta_n \exp(\mu_0 + \frac{1}{2}\sigma_0^2) \,.$$

The parameter estimator converges under H_1 against

$$\tilde{\theta} = \begin{pmatrix} \mu_0 + (1-\lambda)\delta_{n,\mu} \\ \sigma_0^2 + (1-\lambda)\delta_{n,\sigma^2} \end{pmatrix} .$$

Now, as we have the convergence of the parameter estimator, we are interested in conditions for deriving a consistent change-point estimator. To this end we need some replacement assumption as well as an assumption on the convergence rate.

G.13 There exists a function $\psi(s)$, $s \in (0,1)$ and a matrix $C(\tilde{\theta})$ such that

$$\sup_{0<s<1} \frac{\psi(\lfloor sn \rfloor/n)}{n} \left\| S_H(\lfloor sn \rfloor; \hat{\theta}_n) - S_H(\lfloor sn \rfloor; \tilde{\theta}) + \frac{\lfloor sn \rfloor}{n} C(\tilde{\theta}) S_G(n, \tilde{\theta}) \right\|_A = O_P\left(\frac{1}{\sqrt{n}}\right) ,$$

with G being an unbiased estimation function.

G.14 There exists a function $\psi(s)$, $s \in (0,1)$ and a function $E_{sn,n}(\tilde{\theta})$ such that

$$\sup_{0<s<1} \frac{\psi(\lfloor sn \rfloor/n)}{n} \left\| S_H(\lfloor ns \rfloor; \tilde{\theta}) - \frac{\lfloor sn \rfloor}{n} C(\tilde{\theta}) S_G(n, \tilde{\theta}) - E_{\lfloor sn \rfloor,n}(\tilde{\theta}) \right\|_A = O_P\left(\frac{1}{\sqrt{n}}\right) .$$

Moreover, $\left\| E_{k,n}(\tilde{\theta}) \right\|_A$ has a unique maximum at $k = m$. We denote by $\mathcal{D}_n = \|\delta_n\|_A = \max_{1 \le k < n} \left\| \frac{1}{n} E_{k,n}(\tilde{\theta}) \right\|_A$.

As we assumed that G is an unbiased estimation function and we have defined a limit to exist for the parameter estimator, the part $\frac{\lfloor sn \rfloor}{n} C(\tilde{\theta}) S_G(n, \tilde{\theta})$ usually converges to 0. Hence, $\left\| E_{\lfloor sn \rfloor,n}(\tilde{\theta}) \right\|_A$ does not depend on the estimating function G.

Thus in the following we are going to prove the results for the deterministic weight function. We will see that the results then also hold true for the randomized weight function.

Theorem 4.3.1 *Let $\hat{m}$ be the change-point estimator given by* (4.26) *with a possibly random weight function fulfilling assumption **G.7**. If the assumptions **G.1**, **G.13** and **G.14**, with $\psi(s) = \rho(s)$ and $\psi(s) = 1$, are fulfilled, then $\hat{m}/n$ is a consistent estimator for λ.*

Let us consider the example of the log-normal distributed observations. We show that the conditions are fulfilled and determine the function $E_{\lfloor sn \rfloor,n}(\tilde{\theta})$.

Example 4.3.2 *Recall the example of independent identical log-normal distributed observations, introduced in Example 3.3.1. We were interested if the expectation of the log-normal observations changes. We considered the test statistic*

$$T_n(\hat{\theta}_n; A) = \max_{0<k<n} w_n(n,k) \left\| \sum_{t=1}^{k} \left(X_t - \exp(\hat{\mu}_n + \tfrac{1}{2}\hat{\sigma}_n^2) \right) \right\|_A ,$$

with $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)^{\mathsf{T}}$ *are the least-squares estimators of the log-observations. Let us proof that the assumption* ***G.14*** *is fulfilled. To this end we consider*

$$\begin{aligned}
&\frac{1}{n} S_H(\lfloor ns \rfloor; \hat{\theta}_n) \\
&\quad = \frac{1}{n} \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \exp(\hat{\mu}_n + \tfrac{1}{2}\hat{\sigma}_n^2) - X_t + \exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) \right) + \frac{1}{n} \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) \right) \\
&\quad = \frac{\lfloor sn \rfloor}{n} \left(\exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) - \exp(\hat{\mu}_n + \tfrac{1}{2}\hat{\sigma}_n^2) \right) + \frac{1}{n} \sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) \right) ,
\end{aligned}$$

with $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma}^2)^{\mathsf{T}}$.

The last sum on the right hand-side is a sum is a sum over i.i.d. random variables, if $s \le \lambda$. *For* $s > \lambda$ *we split the sum at* $s = \lambda$, *then we get the convergence.*

For the assumption ***G.14*** *we need to determine the rate of convergence. This essentially comes from the CLT applied to each stationary sum of* $\frac{1}{n}\sum_{t=1}^{\lfloor sn \rfloor} \left(X_t - \exp(\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2) \right)$ *after splitting at* $\lfloor \lambda n \rfloor$. *But we have to be careful since* $X_t - \exp(\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2)$ *has not expectation* 0. *We use the same* $C(\tilde{\theta})$ *as in example 3.3.2 evaluated for the parameter* $\tilde{\theta}$ *under the alternative. Then we have*

$$\begin{aligned}
\frac{1}{n} S_H(\lfloor ns \rfloor; \hat{\theta}_n) - \frac{1}{n} S_H(\lfloor ns \rfloor; \tilde{\theta}) + \frac{\lfloor ns \rfloor}{n} C(\tilde{\theta}) S_G(n, \tilde{\theta})
&= \frac{\lfloor sn \rfloor}{n} \left(\exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) - \exp(\hat{\mu}_n + \tfrac{1}{2}\hat{\sigma}_n^2) \right) + \frac{\lfloor ns \rfloor}{n} C(\tilde{\theta}) S_G(n, \tilde{\theta}) \\
&= \frac{\lfloor sn \rfloor}{n} \left(\exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) - \exp(\hat{\mu}_n + \tfrac{1}{2}\hat{\sigma}_n^2) \right) \\
&\quad + \lfloor ns \rfloor \exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) \left(\tilde{\mu} - \hat{\mu}_n + \frac{1}{2}(\hat{\sigma}_n^2 - \tilde{\sigma}^2) \right) \\
&\quad + \frac{\lfloor ns \rfloor}{n} \exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) \left(\sum_{t=1}^{n} (\frac{1}{2}(\log(X_t) - \tilde{\mu})^2 - \frac{1}{2}(\log(X_t) - \hat{\mu}_n)^2) \right)
\end{aligned}$$

Now calculating the Taylor expansion up to degree 2, *gives with analogue argumentation as in Example 3.3.2 the rate* $O_P(1/\sqrt{n})$.

The question occurs of the limit $E_{\lfloor sn\rfloor,n}(\tilde\theta)$. Assuming the change is additive in one of the parameters leads to

$$\frac{1}{n}\sum_{t=m+1}^{\lfloor sn\rfloor}\left(X_t-\exp(\mu_0+\tfrac{1}{2}{\sigma_0}^2+d_n)\right)$$
$$=\frac{1}{n}\sum_{t=m+1}^{\lfloor sn\rfloor}\left(X_t-\exp(\mu_0+\tfrac{1}{2}{\sigma_0}^2)\right)+\frac{\lfloor sn\rfloor-m}{n}(1-\delta_n)\exp(\mu_0+\tfrac{1}{2}{\sigma_0}^2)\,.$$

Observe that G is an unbiased estimation function even under H_1. So, $E_{\lfloor sn\rfloor,n}(\tilde\theta)$ is given as

$$E_{\lfloor sn\rfloor,n}(\tilde\theta)=\begin{cases}\lfloor sn\rfloor(\exp(\mu_0+\frac{1}{2}\sigma_0^2)-\exp(\tilde\mu+\frac{1}{2}\tilde\sigma^2)) & ,s\le\lambda\\ m(1-\delta_n)(\exp(\mu_0+\frac{1}{2}\sigma_0^2))+\lfloor sn\rfloor(\delta_n\exp(\mu_0+\frac{1}{2}\sigma_0^2)-\exp(\tilde\mu+\frac{1}{2}\tilde\sigma^2)) & ,s>\lambda\end{cases},\tag{4.27}$$

with $\theta_0=(\mu_0,\sigma_0^2)^{\mathrm T}$ being the true value from the time-series before the change and δ denote the size of the change in the expectation of the log-normal distributed observations (see Example 4.3.1). This function has its the maximum or its minimum at $k=m$. For $\delta_n>0$ is greater 1 *we have $\mu_0+\frac{1}{2}\sigma_0^2<\tilde\mu+\frac{1}{2}\tilde\sigma^2<\mu_0+\frac{1}{2}\sigma_0^2+\log(\delta_n)$, which includes that $E_{k,n}(\tilde\theta)$ is decreasing for $k<m$ and increasing for $k>m$ always with negative values. Then we see, that $\left\|E_{k,n}(\tilde\theta)\right\|_A$ has its maximum at $k=m$.*

Rate of convergence

In this section, we are going to determine the rate of convergence and identify the part of the statistic which gives us the asymptotic behaviour. Before we state the main result, we give the proofs of the behaviour for the decomposition parts.

To derive our result, we have to make some assumptions.

G.15 Let the following assumptions hold true for the process before the change-point.

a) The parameter estimator can be replaced by its consistent asymptotic value without changing the asymptotic behaviour, i.e.

$$\max_{1\le k<m}\frac{1}{m-k}\left\|\sum_{t=k+1}^{m}\left(H(\mathbb{X}_t,\hat\theta_n)-H(\mathbb{X}_t,\tilde\theta)\right)\right\|_A=O_P\left(\frac{1}{\sqrt{n}}\right).$$

b) Let a Hájek-Rényi-type condition hold true, i.e. there exists $\kappa_n>0$ non-decreasing with $\kappa_n/n\underset{n\to\infty}{\longrightarrow}0$, such that

$$\max_{1\le k\le m-\kappa_n}\frac{1}{m-k}\left\|S_H(m,\tilde\theta)-E_{m,n}(\tilde\theta)-S_H(k,\tilde\theta)+E_{k,n}(\tilde\theta)\right\|_A=\kappa_n^{-1/2}O_P(1)\,.$$

G.16 Let the following assumptions hold true for the process after the change-point.

a) The parameter estimator can be replaced by its consistent asymptotic value without changing the asymptotic behaviour, i.e.

$$\max_{m\le k<n}\frac{1}{k-m}\left\|\sum_{t=m+1}^{k}\left(H(\mathbb{X}_t^{(2)},\hat\theta_n)-H(\mathbb{X}_t^{(2)},\tilde\theta)\right)\right\|_A=O_P\left(\frac{1}{\sqrt{n}}\right).$$

b) Let a Hájek-Rényi-type condition hold true, i.e. there exists $\kappa_n > 0$ non-decreasing with $\kappa_n/n \underset{n\to\infty}{\longrightarrow} 0$, such that

$$\max_{m+\kappa_n \le k < n} \frac{1}{k-m} \left\| S_H(k,\tilde{\theta}) - E_{k,n}(\tilde{\theta}) - S_H(m,\tilde{\theta}) + E_{m,n}(\tilde{\theta}) \right\|_A = \kappa_n^{-1/2} O_P(1) .$$

Example 4.3.3 *Ongoing example of log-normal distributed observations (see first mentioned in Example 3.3.1). We proof that the assumptions* ***G.15****a and* ***G.15****b hold true. The other site follows equivalently.*

Let us consider the replacement condition. We have

$$\left.\begin{array}{l} \displaystyle\max_{1\le k\le m} \frac{1}{m-k} \left\| \sum_{t=k+1}^{m} \Big(H(\mathbb{X}_t,\hat{\theta}_n) - H(\mathbb{X}_t,\tilde{\theta}) \Big) \right\|_A \\ \displaystyle\max_{m<k<n} \frac{1}{k-m} \left\| \sum_{t=m+1}^{k} \Big(H(\mathbb{X}_t,\hat{\theta}_n) - H(\mathbb{X}_t,\tilde{\theta}) \Big) \right\|_A \end{array}\right\} = \left\| \left(\exp(\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2) - \exp(\hat{\mu}_n + \tfrac{1}{2}\hat{\sigma}_n^2) \right) \right\|_A .$$

So, it vanishes as the parameter estimator is consistent and exp *is continuous.*

For the Hájek-Rényi-type condition we use $E_{\lfloor sn\rfloor,n}(\tilde{\theta})$ *as given in* (4.27)*. Then it follows that we have*

$$S_H(m,\tilde{\theta}) - E_{m,n}(\tilde{\theta}) - S_H(k,\tilde{\theta}) + E_{k,n}(\tilde{\theta}) = \sum_{t=k+1}^{m} (X_t - \mathbb{E}[X_t]) .$$

This is the same situation as given in Lemma 2.3.3.Analogously we calculate for the difference for $k > m$

$$\begin{aligned} &\max_{m+\kappa_n \le k < n} \frac{1}{m-k} \left\| S_H(k,\tilde{\theta}) - E_{k,n}(\tilde{\theta}) - S_H(m,\tilde{\theta}) + E_{m,n}(\tilde{\theta}) \right\|_A \\ &\qquad = \max_{m+\kappa_n \le k < n} \frac{1}{m-k} \left\| \sum_{t=m+1}^{k} (X_t - \mathbb{E}[X_t]) \right\|_A \\ &\qquad = O_P(\kappa_n^{-\frac{1}{2}}) . \end{aligned}$$

As the sum is taken over stationary ergodic random variables with expectation 0*.*

Lemma 4.3.2

Let the change-point be given by ***G.1*** *and let the assumption* ***G.14****, with* $\psi \equiv 1$*, be fulfilled.*
If $\{X_t^{(1)}\}$ *fulfils the assumption* ***G.15****, then it holds*

$$\max_{1 \le k < m} \frac{1}{n} \left\| S_H(k; \hat{\theta}_n) + S_H(m; \hat{\theta}_n) - E_{k,n}(\tilde{\theta}) - E_{m,n}(\tilde{\theta}) \right\|_A = o_P(1) \,. \tag{4.28}$$

If $\{X_t^{(2)}\}$ *fulfils the assumptions* ***G.16****, then we have*

$$\max_{m < k < n} \frac{1}{n} \left\| S_H(k; \hat{\theta}_n) + S_H(m; \hat{\theta}_n) - E_{k,n}(\tilde{\theta}) - E_{m,n}(\tilde{\theta}) \right\|_A = o_P(1) \,. \tag{4.29}$$

For the rate we have to make some additional assumptions on the weight function, as it has an influence on the asymptotic distribution.

G.17 Let the weight function $\rho : (0,1) \to \mathbb{R}_+$ be local Lipschitz continuous and

$$c_1 |k - l| n^{-1} \ge |\rho^2(k/n) - \rho^2(l/n)|$$

with $c_1 > 0$ and $k, l \in [\alpha n, \beta n]$ for $0 < \alpha < \beta < 1$.

G.18 The function $E_{\lfloor sn \rfloor, n}(\tilde{\theta})$ is Lipschitz continuous with

$$|s - r| n \mathcal{D}_n c_2 \le \| E_{\lfloor sn \rfloor, n}(\tilde{\theta}) - E_{\lfloor rn \rfloor, n}(\tilde{\theta}) \| \le |s - r| n \mathcal{D}_n c_2' \qquad s,\, r \in (\alpha, \beta) \,,$$

where $0 < \alpha < \beta < 1$, c_2, c_2' some constants, $\mathcal{D}_n$ (from assumption **G.14**) is non-increasing and there exists a constant $\mathcal{D}$ such that $\lim_{n \to \infty} \mathcal{D}_n = \mathcal{D}$.

G.19 For $E_{\lfloor sn \rfloor, n}(\tilde{\theta})$ and $\rho(s)$, $s \in (0,1)$ we have

$$\max_{\alpha n \le k < m} \frac{m - k}{\rho^2(k/n) \| E_{k,n}(\tilde{\theta}) \|^2 - \rho^2(m/n) \| E_{m,n}(\tilde{\theta}) \|^2} = O(n^{-1} \mathcal{D}_n^{-2})$$

and

$$\max_{m < k \le (1-\alpha) n} \frac{k - m}{\rho^2(k/n) \| E_{k,n}(\tilde{\theta}) \|^2 - \rho^2(m) \| E_{m,n}(\tilde{\theta}) \|^2} = O(n^{-1} \mathcal{D}_n^{-2}) \,.$$

Lemma 4.3.3

For a function $\rho \in \mathcal{L}(\eta, \gamma)$ *and* ***G.17*** *and a function* $E_{s,n}$ *fulfilling* ***G.18*** *there exists constants* c_1'*,* c_2' *and* c_3' *constants depending on* α*,* β *such that it holds for* $k, l \in [\alpha n, \beta n]$

$$\begin{aligned} |\rho^2(k/n) - \rho^2(l/n)| &\le |k - l| c_1' n^{-1} \\ \left| \rho(k/n) \left\| E_{k,n}(\tilde{\theta}) \right\| - \rho(l/n) \left\| E_{l,n}(\tilde{\theta}) \right\| \right| &\le |k - l| c_2' n^{-1} \mathcal{D}_n \\ |\rho^2(k/n) \| E_{k,n}(\tilde{\theta}) \|^2 - \rho^2(l) \| E_{l,n}(\tilde{\theta}) \|^2 | &\le |k - l| c_3' n \mathcal{D}_n^2 \,. \end{aligned}$$

This is an equivalent result to the properties of the usual used weight function we needed in section 2.3 (see Corollary 2.3.1).

Finally, we are able to determine the rate of convergence, at least an upper bound. That it is the best rate is proven in Theorem 4.3.3.

Theorem 4.3.2 *Let assumptions* ***G.1****,* ***G.5.a)*** *or* ***G.5.b)*** *and* ***G.12*** *–* ***G.16*** *hold true. Then the change-point estimator* $\hat{m}$*,* (4.26)*, with* $\lambda \in (\eta, 1-\eta)$ *fulfils*

$$\hat{m} - m = O_P(\mathcal{D}_n^{-2}) \,. \tag{4.30}$$

Asymptotic distribution

We will make use of the same decomposition of V_k as in the proof for the rate. Therefore, we first observe the behaviour of the parts from the decomposition.

Lemma 4.3.4

Let H_1 *hold true and the change-point fulfil assumption* ***G.1****. If the assumptions* ***G.5.a)*** *(or* ***G.5.b)****) and* ***G.14*** *hold true and the process* $\{X_t^{(1)}\}$ *fulfils the assumption* ***G.15****, then for fixed but arbitrary* C *we have*

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} \left\| S_H(m,\hat{\theta}_n) - S_H(m,\tilde{\theta}) - S_H(k,\hat{\theta}_n) + S_H(k,\tilde{\theta}) \right\|_A = o_P(1) \,,$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} \left\| S_H(m;\tilde{\theta}) - E_{m,n}(\tilde{\theta}) - S_H(k;\tilde{\theta}) + E_{k,n}(\tilde{\theta}) \right\|_A = O_P(1) \,,$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} \left\| S_H(m;\hat{\theta}_n) - E_{m,n}(\tilde{\theta}) + S_H(k;\hat{\theta}_n) - E_{k,n}(\tilde{\theta}) \right\|_A = O_P(\sqrt{n}) \,.$$

Secondly, we have to make some more assumptions on the test function.

G.20 a) Define

$$\xi_n^{(1)}(s) := \delta^{\mathrm{T}} A \left(S_H(m,\tilde{\theta}) - S_H(m - \mathcal{D}_n^{-2}s,\tilde{\theta}) - E_{(m,n)}(\tilde{\theta}) + E_{(m-\mathcal{D}_n^{-2}s,n)}(\tilde{\theta}) \right) \quad 0 < s \,,$$

$$\xi_n^{(2)}(s) := \delta^{\mathrm{T}} A \left(S_H(m - \mathcal{D}_n^{-2}s,\tilde{\theta}) - S_H(m,\tilde{\theta}) - E_{(m-\mathcal{D}_n^{-2}s,n)}(\tilde{\theta}) + E_{(m,n)}(\tilde{\theta}) \right) \quad s < 0$$

and $\xi_n(s) := \mathbf{1}_{\{s>0\}}\xi_n^{(1)}(s) + \mathbf{1}_{\{s<0\}}\xi_n^{(2)}(s)$. Let there exists a process $\{W(s),\ s \in \mathbb{Z}\}$ such that for the A-fixed case we have

$$\{\xi_n(s),\ s \in \mathbb{Z}\} \xrightarrow{d} \{W(s),\ s \in \mathbb{Z}\} \,.$$

b) Let

$$\xi_n^{(1)}(s) := \delta_n^{\mathrm{T}} A(S_H(m,\tilde{\theta}) - E_{m,n}(\tilde{\theta}) - S_H(m - \mathcal{D}_n^{-2}s,\tilde{\theta}) + E_{(m-\mathcal{D}_n^{-2}s),n}(\tilde{\theta})) \quad 0 < s\,,$$
$$\xi_n^{(2)}(s) := \delta_n^{\mathrm{T}} A(S_H(m - \mathcal{D}_n^{-2}s,\tilde{\theta}) - S_H(m,\tilde{\theta}) - E_{(m-\mathcal{D}_n^{-2}s),n}(\tilde{\theta}) + E_{m,n}(\tilde{\theta})) \quad s < 0\,.$$

Under the A-local alternative there exists a process $W(s)$ such that $\xi_n(s) = \mathbf{1}_{\{s>0\}}\xi_n^{(1)}(s) + \mathbf{1}_{\{s<0\}}\xi_n^{(2)}(s)$ converge in distribution, i.e.

$$\{\xi_n(s),\ s \in \mathbb{R}\} \xrightarrow{d} \{W(s),\ s \in \mathbb{R}\}\,.$$

G.21 a) For

$$h(s,n) := \frac{\left(\rho^2(m/n)\left\|E_{(m,n)}(\tilde{\theta})\right\|_A^2 - \rho^2((m-s)/n)\left\|E_{((m-s),n)}(\tilde{\theta})\right\|_A^2\right)}{\rho^2(m/n)\,n}$$

we have under the A-fixed alternative, there exists a function $g_\lambda(s)$ such that uniformly in C it holds

$$\sup_{s\in(-C,C)} \left|h(s,n) - |s|g_\lambda(s)\mathcal{D}^2\right| = o_P(1)\,,$$

with $\mathcal{D} = \lim_{n\to\infty}\mathcal{D}_n$.

b) For

$$h(s,n) := \frac{\left(\rho^2(m/n)\left\|E_{(m,n)}(\tilde{\theta})\right\|_A^2 - \rho^2((m-\mathcal{D}_n^{-2}s)/n)\left\|E_{((m-\mathcal{D}_n^{-2}s),n)}(\tilde{\theta})\right\|_A^2\right)}{\rho^2(m/n)\,n}$$

we have under the A-local alternative, there exists a function $g_\lambda(s)$ with

$$\sup_{s\in(-C,C)} |h(s,n) - |s|g_\lambda(s)| = o_P(1)\,.$$

Theorem 4.3.3 *Let the assumptions of Theorem 4.3.2 hold true.*

1. *Under the A-fixed alternative, if additionally the assumptions **G.20**a and **G.21**a are true, then*
$$\hat{m} - m \xrightarrow{d} \arg\max\{W(s) - |s|g_\lambda(s)D^2,\ s \in \mathbb{Z}\}\,. \tag{4.31}$$
2. *Under the A-local alternative, if additionally the assumptions **G.20**b and **G.21**b hold true, then*
$$\mathcal{D}_n^2(\hat{m}(\eta,\gamma) - m) \xrightarrow{d} \arg\max\{W(s) - |s|g_\lambda(s),\ s \in \mathbb{R}\}\,.$$

4.3.2. Smooth functions

We have seen in section 3.3.3 that we derive moment conditions for the estimating and testing functions if they are smooth enough. The smoothness is meant in the sense of the existence and boundedness of the derivatives and the functions itself. Let us consider under which assumptions we derive the regularity conditions.

Replacement assumption

As in the case of H_0, we have to assume differentiability and stationarity of the observations before and after the change-point as well as some moment conditions.

L.9 The observations before ($\{X_t^{(1)}\}$) and after ($\{X_t^{(2)}\}$) the change-point are independent and each is stationary and ergodic.

L.10 For each stationary part the estimating function G fulfills the assumptions **L.2**, **L.4** and **L.5** hold true.

L.11 The testing function H fulfils for each stationary part the assumptions **L.6** and **L.7**.

Proposition 4.3.1

Under the assumptions ***L.9*** *–* ***L.11*** *we have* ***G.13*** *with*

$$C(\tilde{\theta}) = \lambda\mathbb{E}[\nabla H(X_1^{(1)}, \tilde{\theta})](\mathbb{E}[\nabla G(X_1^{(1)}, \tilde{\theta})])^{-1} + (1-\lambda)\mathbb{E}[\nabla H(X_1^{(2)}, \tilde{\theta})](\mathbb{E}[\nabla G(X_1^{(2)}, \tilde{\theta})])^{-1}.$$

Deterministic function $\mathbf{E}_{\lfloor \mathbf{sn} \rfloor,\mathbf{n}}(\tilde{\theta})$

Now, let us analyse the deterministic function $E_{\lfloor sn \rfloor,n}(\tilde{\theta})$. Therefore, we assume H to be an unbiased estimation function. Then, we use that $S_H(k,\theta)$ is given as a sum. This allows us to determine the following representation of $E_{\lfloor sn \rfloor,n}(\tilde{\theta})$.

Proposition 4.3.2

Under the assumptions ***G.12*** *and* ***L.8*** *–* ***L.11*** *we have for H being an unbiased estimation function for θ under H_1, i.e. $\lambda\mathbb{E}[H(X_1^{(1)}, \tilde{\theta})] + (1-\lambda)\mathbb{E}[H(X_1^{(2)}, \tilde{\theta})] = 0$, then*

$$E_{\lfloor sn \rfloor,n}(\tilde{\theta}) = g(\lfloor sn \rfloor)\mathbb{E}[H(X_1^{(1)}, \tilde{\theta}_n)]$$

with $\tilde{\theta}_n$ as in ***L.8*** *and*

$$g(\lfloor sn \rfloor) = \begin{cases} \lfloor sn \rfloor & , s \leq \lambda \\ \frac{1}{1-\lambda}(m - \lfloor sn \rfloor \lambda) & , s > \lambda \end{cases}.$$

We have this form in section 2.3, 4.1 and 4.2. But also in the mentioned publications this form was used.

L.12 The weight function ρ is symmetric and differentiable with $0 < s\rho'(s) + \rho(s) < c$ for all $s \in (0,1)$.

Proposition 4.3.3

Let the assumptions of Proposition 4.3.2 hold true. If additionally the assumption ***L.12*** *holds, then* ***G.18*** *and* ***G.21****a,* ***G.21****b hold true with each having a limit function*

$$g(s) = \begin{cases} 2\lambda\rho(\lambda)(\rho(\lambda) + \lambda\rho'(\lambda)) & , s < 0 \\ 2\rho'(\lambda) - \lambda\rho(\lambda)(\rho(\lambda) + \lambda\rho'(\lambda)) & , s > 0 \end{cases}$$

either with $s \in \mathbb{Z}$ or $s \in \mathbb{R}$.

4.3.3. Proofs

Lemma 4.3.1

Let $w_n(n,k)$ denote a random weight function fulfilling assumption **G.7**.
Then it holds for $\alpha \in (0,\lambda)$ and κ_n non-decreasing, that

$$\max_{\alpha n<k<m-\kappa_n} \left| \frac{w_n(n,k)-\rho(k/n)}{m-k} \right| = o_P(\kappa_n^{-1}) = o_P(1)\,,$$
$$\max_{\alpha n<k<m-\kappa_n} \left| \frac{w_n^2(n,k)-\rho^2(k/n)}{m-k} \right| = o_P(\kappa_n^{-1}) = o_P(1)\,,$$

Proof:
First notice, $\max(a_n/b_n) \leq \max(a_n)/\min(b_n)$. Applying this inequation and using the uniform convergence of $w_n(n,k)$, finishes the proof.

■

Theorem 4.3.1

Let $\hat{m}$ be the change-point estimator given by (4.26) with a possibly random weight function fulfilling assumption **G.7**. If the assumptions **G.1**, **G.13** and **G.14**, with $\psi(s)=\rho(s)$ and $\psi(s)=1$, are fulfilled, then $\hat{m}/n$ is a consistent estimator for λ.

Proof:
First notice, the random weight function can be replaced by the deterministic function, since **G.13** and **G.14** hold true for $\psi(s)=1$.

Secondly, from assumption **G.13** and **G.14**, we get

$$\sup_{s\in(0,1)} \frac{\rho(\frac{\lfloor sn\rfloor}{n})}{n} \left\| S_H(\lfloor ns\rfloor;\hat{\theta}_n) - E_{\lfloor ns\rfloor,n}(\tilde{\theta}) \right\|_A = O_P\left(\frac{1}{\sqrt{n}}\right) \tag{4.32}$$

Under the assumption **G.14**, the maximum of $E_{\lfloor ns\rfloor,n}(\tilde{\theta})$ is taken at m. So by Lemma 3.1 of Pötscher and Prucha [1997] (Theorem C.1.2), we have $\frac{\hat{m}}{n}$ is consistent to λ ($\frac{m}{n} \xrightarrow{p} \lambda$).

■

Lemma 4.3.2

Let the change-point be given by **G.1** and let the assumption **G.14**, with $\psi \equiv 1$, be fulfilled.
If $\{X_t^{(1)}\}$ fulfils the assumption **G.15**, then it holds

$$\max_{1\le k<m} \frac{1}{m+k} \left\| S_H(k;\hat\theta_n) + S_H(m;\hat\theta_n) - E_{k,n}(\tilde\theta_1) - E_{m,n}(\tilde\theta) \right\|_A = o_p(1)\,. \tag{4.33}$$

If $\{X_t^{(2)}\}$ fulfils the assumptions **G.16**, then we have

$$\max_{1\le k<m} \frac{1}{m+k} \left\| S_H(k;\hat\theta_n) + S_H(m;\hat\theta_n) - E_{k,n}(\tilde\theta) - E_{m,n}(\tilde\theta) \right\|_A = o_p(1)\,. \tag{4.34}$$

Proof:
We are able to show (4.33) using the assumption **G.14** in the following way.

$$\begin{aligned}
&\max_{1\le k<m} \frac{1}{m+k} \left\| S_H(k;\hat\theta_n) + S_H(m;\hat\theta_n) - E_{k,n}(\tilde\theta) - E_{m,n}(\tilde\theta) \right\|_A \\
&\quad \le \left(\max_{1\le k<m} \frac{n}{m+k} \right) \left(\left\| \frac{1}{n} (S_H(m;\hat\theta_n) - E_{m,n}(\tilde\theta)) \right\|_A \right. \\
&\qquad \left. + \max_{1\le k<m} \frac{1}{n} \left\| S_H(k;\hat\theta_n) - E_{k,n}(\tilde\theta) \right\|_A \right) \\
&\quad = O_P(1)\,(o_P(1) + o_P(1))
\end{aligned}$$

■

Lemma 4.3.3

For a function $\rho \in \mathcal{L}(\eta,\gamma)$ and **G.17** and a function $E_{s,n}$ fulfilling **G.18** there exists constants c_1', c_2' and c_3' constants depending on α, β such that it holds

$$|\rho^2(k/n) - \rho^2(l/n)| \le |k-l| c_1' n^{-1} \tag{4.35}$$

$$\left| \rho(k/n) \left\| E_{k,n}(\tilde\theta) \right\| - \rho(l/n) \left\| E_{l,n}(\tilde\theta) \right\| \right| \le |k-l| c_2' n^{-1} \mathcal{D}_n \tag{4.36}$$

$$|\rho^2(k/n) \|E_{k,n}(\tilde\theta)\|^2 - \rho^2(l) \|E_{l,n}(\tilde\theta)\|^2| \le |k-l| c_3' n \mathcal{D}_n^2\,. \tag{4.37}$$

Proof:
The properties (4.35), (4.36) and the right hand side of (4.37) follow directly from the Lipschitz property and the boundedness.

■

Theorem 4.3.2

Let assumptions **G.1**, **G.5.a)** or **G.5.b)** and **G.12** – **G.16** hold true. Then the change-point estimator $\hat m$, (4.26), with $\lambda \in (\eta, 1-\eta)$ fulfils

$$\hat m - m = O_P(\mathcal{D}_n^{-2})\,.$$

Proof:
First observe, we can reduce to a truncated version of the change-point estimator by Corollary 2.3.2. Secondly, we analyse the change-point estimator with the deterministic weight function $\rho(s)$ instead of $w_n(s)$. Thirdly, we finish the proof by showing that the randomized weight function can be replaced by the deterministic one asymptotically.
The change-point estimator

$$\hat{m} = \underset{1\le k<n}{\arg\max}\{\rho(k/n)S_H^T(k;\hat{\theta}_n)\, A\, \rho(k/n)S_H(k;\hat{\theta}_n)\} =: \underset{1\le k<n}{\arg\max}\{V_k\}$$

can be represented as in the proof of Theorem 2.3.2, Theorem 2.4.3 and Theorem 4.2.2 with

$$\begin{aligned} V_k &= (\rho(k/n)S_H(k;\hat{\theta}_n) - \rho(m/n)S_H(m;\hat{\theta}_n))^T A(\rho(m/n)S_H(m;\hat{\theta}_n) + \rho(k/n)S_H(k;\hat{\theta}_n)) \\ &= -\left\langle \rho(m/n)S_H(m;\hat{\theta}_n) - \rho(k/n)S_H(k;\hat{\theta}_n), \rho(k/n)S_H(k;\hat{\theta}_n) + \rho(m/n)S_H(m;\hat{\theta}_n)\right\rangle_A \\ &= -\Big\langle (\rho(m/n) - \rho(k/n))S_H(k,\hat{\theta}_n) + \rho(m/n)\left(S_H(m,\hat{\theta}_n) - S_H(k,\hat{\theta}_n)\right), \\ &\qquad \rho(m/n)\left(S_H(m,\hat{\theta}_n) - S_H(k,\hat{\theta}_n)\right) + (\rho(m/n) + \rho(k/n))S_H(k,\hat{\theta}_n)\Big\rangle_A . \end{aligned}$$

In the examples we always considered a decomposition of V_k and prove then that for every ϵ

$$P(\hat{m} < m - \kappa_n) = P\left(\max_{1\le k<m-\kappa_n} V_k \ge \max_{m-\kappa_n\le k<n} V_k\right) < \epsilon .$$

Observe that we usually should use, for every ϵ exists a constant δ such that

$$P(\hat{m} < m - \delta\kappa_n) < \epsilon .$$

Without loss of generality, we can assume $\delta = 1$.

The decomposition we analyse is given as

$$\begin{aligned} V_k = -\Bigg\langle & (\rho(m/n) - \rho(k/n))\left(S_H(k,\hat{\theta}_n) - E_{k,n}(\tilde{\theta})\right) \\ & + \rho(m/n)\left(S_H(m,\hat{\theta}_n) - S_H(m,\tilde{\theta}) - S_H(k,\hat{\theta}_n) + S_H(k,\tilde{\theta})\right) \\ & + \rho(m/n)\left(S_H(m,\tilde{\theta}) - E_{m,n}(\tilde{\theta}) - S_H(k,\tilde{\theta}) + E_{k,n}(\tilde{\theta})\right) \\ & + \left(E_{m,n}(\tilde{\theta})\rho(m/n) - E_{k,n}(\tilde{\theta})\rho(k/n)\right), \\ & \rho(m/n)\left(S_H(m,\hat{\theta}_n) - E_{m,n}(\tilde{\theta}) - S_H(k,\hat{\theta}_n) + E_{k,n}(\tilde{\theta})\right) \\ & + (\rho(m/n) + \rho(k/n))\left(S_H(k,\hat{\theta}_n) - E_{k,n}(\tilde{\theta})\right) \\ & + \left(E_{m,n}(\tilde{\theta})\rho(m/n) + E_{k,n}(\tilde{\theta})\rho(k/n)\right) \Bigg\rangle_A \\ =& B_1 + B_2 + B_3 + B_4 + B_5 + B_6 . \end{aligned}$$

Observe, we are in an equivalent situation as in the case of the neural-network based change-point estimator. Careful consideration and analogously calculation leads to the following representation

$$
\begin{aligned}
B_1 &= (\rho^2(k/n) - \rho^2(m/n))\left\|S_H(k,\hat{\theta}_n) - E_{k,n}(\tilde{\theta})\right\|_A^2 \\
B_2 &= -\rho^2(m/n)\Big\langle S_H(k,\hat{\theta}_n) - E_{k,n}(\tilde{\theta}) + S_H(m,\hat{\theta}_n) - E_{m,n}(\tilde{\theta}), \\
&\qquad S_H(m,\hat{\theta}_n) - E_{m,n}(\tilde{\theta}) - S_H(k,\hat{\theta}_n) - E_{k,n}(\tilde{\theta})\Big\rangle_A \\
B_3 &= 2\Big\langle S_H(k,\hat{\theta}_n) - E_{k,n}(\tilde{\theta}), E_{m,n}(\tilde{\theta})\rho^2(m/n) + E_{k,n}(\tilde{\theta})\rho^2(k/n)\Big\rangle_A \\
B_4 &= -2\rho^2(m/n)\left\langle S_H(m,\hat{\theta}_n) - S_H(m,\tilde{\theta}) - S_H(k,\hat{\theta}_n) + S_H(k,\tilde{\theta}),\ E_{m,n}(\tilde{\theta})\right\rangle_A \\
B_5 &= 2\rho^2(m/n)\left\langle S_H(m,\tilde{\theta}) - E_{m,n}(\tilde{\theta}) - S_H(k,\tilde{\theta}) + E_{k,n}(\tilde{\theta}),\ E_{m,n}(\tilde{\theta})\right\rangle_A \\
B_6 &= -\left(\left\|E_{m,n}(\tilde{\theta})\right\|_A^2\rho^2(m/n) - \left\|E_{k,n}(\tilde{\theta})\right\|_A^2\rho^2(k/n)\right).
\end{aligned}
$$

With the assumption **G.17** and **G.18**

$$
\max_{\alpha n\le k<m-\kappa_n}\frac{m-k}{|B_6|} = O\left(\frac{1}{n\mathcal{D}_n^2}\right)
$$

This is equivalent to the property of B_6 in Theorem 2.3.2 (see page 43) or Theorem 2.4.3. In the last one, we see the relation between the size of the change δ_n and κ_n from the assumptions. Nevertheless, we have the same approximation as in (2.45). Hence, it is enough to examine the order of

$$
\max_{1\le k<m}\left|\frac{B_i}{m-k}\right|
$$

for $i = 1,\dots,5$. From the assumptions **G.15** and **G.15**a as well as the Lipschitz continuoity of ρ, we get

$$
\begin{aligned}
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_1}{(m-k)}\right| &\le \max_{\alpha n<k<m-\kappa_n}\left|\frac{\rho^2(k/n)-\rho^2(m/n)}{(m-k)}\right| \max_{\alpha n<k<m-\kappa_n}\left\|S_H(k,\hat{\theta}_n) - E_{k,n}(\tilde{\theta})\right\|_A^2 \\
&= O(n^{-1})O_P(n).
\end{aligned}
$$

We used the Lipschitz property of the weight function (2.28), equivalently as for (2.46). For the following approximations we use the properties of the weight function in combination with $E_{s,n}(\tilde{\theta})$ given in Lemma 4.3.3.

Combining the assumptions **G.13**, **G.15**b and the Lemma 4.3.2, the Cauchy-Schwarz inequality gives

$$\begin{aligned}
&\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_2}{(m-k)}\right| \\
&\quad\le \rho^2(m/n)\max_{\alpha n<k<m-\kappa_n}\left\|S_H(k,\tilde{\theta})-E_{k,n}(\tilde{\theta})+S_H(m,\hat{\theta}_n)-E_{m,n}(\tilde{\theta})\right\|_A \\
&\qquad\cdot\max_{\alpha n<k<m-\kappa_n}\frac{1}{m-k}\left\|S_H(m,\hat{\theta}_n)-E_{m,n}(\tilde{\theta})-S_H(k,\hat{\theta}_n)-E_{k,n}(\tilde{\theta})\right\|_A \\
&\quad= O(1)O_P(\sqrt{n})O_P(\kappa_n^{-1/2})\,.
\end{aligned}$$

For B_3 and B_4 using again the triangle inequality, Cauchy-Schwarz and Lemma 4.3.2, we conclude

$$\begin{aligned}
&\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_3}{(m-k)}\right| \\
&\quad\le \max_{\alpha n<k<m-\kappa_n}\left|\frac{\rho(m/n)-\rho(k/n)}{(m-k)}\right|\max_{\alpha n<k<m-\kappa_n}\left\|S_H(k,\hat{\theta}_n)-E_{k,n}(\tilde{\theta})\right\|_A \\
&\qquad\left(\max_{\alpha n<k<m-\kappa_n}\left\|\left(E_{m,n}(\tilde{\theta})\rho(m/n)-E_{k,n}(\tilde{\theta})\rho(k/n)\right)\right\|_A+\left\|-2\left(E_{m,n}(\tilde{\theta})\right)\right\|_A\right) \\
&\quad= O(n^{-1})O_P(\sqrt{n})\left(O(n\mathcal{D}_n)\right)\,,
\end{aligned}$$

$$\begin{aligned}
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_4}{(m-k)}\right| &\le 2\rho^2(m/n)\left\|E_{m,n}(\tilde{\theta})\right\|_A \\
&\quad\cdot\max_{\alpha n<k<m-\kappa_n}\frac{1}{m-k}\left\|S_H(m,\hat{\theta}_n)-S_H(m,\tilde{\theta})-S_H(k,\hat{\theta}_n)+S_H(k,\tilde{\theta})\right\|_A \\
&= O(1)O_P(n\mathcal{D}_n)O_P(n^{-\frac{1}{2}})\,.
\end{aligned}$$

Using the triangle inequality we get

$$\begin{aligned}
\max_{\alpha n<k<m-\kappa_n}\left|\frac{B_5}{(m-k)}\right| &\le 2\rho^2(m/n)\left\|E_{m,n}(\tilde{\theta})\right\|_A \\
&\quad\cdot\max_{\alpha n<k<m-\kappa_n}\frac{1}{(m-k)}\left\|S_H(m,\tilde{\theta})-E_{m,n}(\tilde{\theta})-S_H(k,\tilde{\theta})+E_{k,n}(\tilde{\theta})\right\|_A \\
&= O(n\mathcal{D}_n)O_P(\kappa_n^{-1/2})\,.
\end{aligned}$$

For $\kappa_n = K/\mathcal{D}_n^2$ this leads to

$$\max_{\alpha n < k < m-\kappa_n} \left| \frac{B_1}{B_6} \right| = o_P(1) \,,$$
$$\max_{\alpha n < k < m-\kappa_n} \left| \frac{B_2}{B_6} \right| = o_P(1) \,,$$
$$\max_{\alpha n < k < m-\kappa_n} \left| \frac{B_3}{B_6} \right| = o_P(1) \,,$$
$$\max_{\alpha n < k < m-\kappa_n} \left| \frac{B_4}{B_6} \right| = o_P(1) \,,$$
$$\max_{\alpha n < k < m-\kappa_n} \left| \frac{B_5}{B_6} \right| = K^{-1/2} O_P(1) \,.$$

Using this results and $B_6 < 0$ for $k < m$, with $\max_{1 \leq k < m-\kappa_n} B_6 \leq c < 0$ for c fixed but arbitrary, it yields

$$\begin{aligned}
P(\hat{m} < m - \kappa_n) &= P\left(\max_{\alpha n < k < m-\kappa_n} V_k > \max_{m-\kappa_n \leq k \leq (1-\alpha)n} V_k \right) \\
&\leq P\left(\max_{\alpha n < k < m-\kappa_n} V_k \geq V_m \right) \\
&= P\left(\max_{\alpha n < k < m-\kappa_n} V_k \geq 0 \right) \\
&= P\left(\max_{\alpha n < k < m-\kappa_n} \left\{ B_6 \left(1 + o_P(1) + \frac{B_2}{B_6} \right) \right\} \geq 0 \right) \\
&\leq P\left(\max_{\alpha n < k < m-\kappa_n} \left| \frac{B_2}{B_6} \right| \geq 1 + o_P(1) \right) \\
&\leq P\left(\max_{\alpha n < k < m-\kappa_n} \left| \frac{B_2}{B_6} \right| \geq 1 - \tau \right) + P\left(1 + o_p(1) \leq \max_{\alpha n < k < m-\kappa_n} \left| \frac{B_2}{B_6} \right| \leq 1 - \tau \right) \\
&\leq P\left(O_P(1) \geq (1-\tau) K^{1/2} \right) + o(1) \,,
\end{aligned}$$

with $0 < \tau < 1$ arbitrary. This term becomes arbitrarily small for a sufficiently large $K > 0$. The other way around works similar.

It is left to analyse the behaviour for a change-point estimator having a randomized weight function. We use the equivalent representation of V_k, i.e. $V_k = \tilde{B}_1 + \tilde{B}_2 + \tilde{B}_3 \tilde{B}_4 + \tilde{B}_5 + \tilde{B}_6$. For analysing

$$\max \left| \frac{\tilde{B}_i}{B_6} \right| \qquad i = 1, \ldots, 6 \,,$$

with B_6 as above, we use Lemma 4.3.1 to handle the randomized weight function. This is finishes the proof.

■

Lemma 4.3.4

Let H_1 hold true and the change-point fulfil assumption **G.1**. If the assumptions **G.5.a)** (or **G.5.b)**) and **G.14** hold true and the process $\{X_t^{(1)}\}$ fulfils the assumption **G.15**, then for fixed but arbitrary C we have

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} \left\| S_H(m,\hat{\theta}_n) - S_H(m,\tilde{\theta}) - S_H(k,\hat{\theta}_n) + S_H(k,\tilde{\theta}) \right\|_A = o_P(1)\,, \tag{4.38}$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} \left\| S_H(m;\tilde{\theta}) - E_{m,n}(\tilde{\theta}) - S_H(k;\tilde{\theta}) + E_{k,n}(\tilde{\theta}) \right\|_A = O_P(1)\,, \tag{4.39}$$

$$\max_{k\in(m-C\mathcal{D}_n^{-2},m)} \left\| S_H(m;\hat{\theta}_n) - E_{m,n}(\tilde{\theta}) + S_H(k;\hat{\theta}_n) - E_{k,n}(\tilde{\theta}) \right\|_A = O_P(\sqrt{n})\,. \tag{4.40}$$

Proof:
The part (4.38) can be shown by the assumption **G.15**a using that the maximum is taken over a set of maximal constant size as $\mathcal{D}_n^{-\beta}$ is non-increasing.
For the equation (4.39) we use assumption **G.15**b with $\kappa_n \equiv 1$ and get the claim using analogously to (4.38).
To show (4.40) we make use of the assumption **G.14** and get analogously to the other parts the result.

■

Theorem 4.3.3

Let the assumptions of Theorem 4.3.2 hold true.

1. Under the A-fixed alternative, if additionally the assumptions **G.20**a and **G.21**a are true, then

$$\hat{m} - m \xrightarrow{d} \arg\max\{\delta^{\mathsf{T}} A^{\frac{1}{2}} W(s) - |s| g_\lambda(s)\mathcal{D}^2,\ s \in \mathbb{Z}\}\,. \tag{4.41}$$

2. Under the A-local alternative, if additionally the assumptions **G.20**b and **G.21**b hold true, then

$$\mathcal{D}_n^2(\hat{m}(\eta,\gamma) - m) \xrightarrow{d} \arg\max\{W(s) - |s| g_\lambda(s),\ s \in \mathbb{R}\}\,.$$

Proof:
The proof is shown for the local alternative, the fixed alternative follows equivalently to the proof of Theorem 2.3.3.

We show the claim by analysing the behaviour of the parts from the decomposition of V_k. Let (w.l.o.g.) $x > 0$ and $C > x$ be both fixed but arbitrary. We get

$$P(\mathcal{D}_n^2(\hat{m}-m) \le x) = P(m - C\mathcal{D}_n^{-2} \le \hat{m} \le m + \mathcal{D}_n^{-2}x) + P(\hat{m} - m \le \mathcal{D}_n^{-2}x, |\hat{m}-m| > C\mathcal{D}_n^{-2})$$

The second term on the right side becomes arbitrary small for large enough C because of the consistency of $\hat{m}$. Therefore, we consider

$$P(m - C\mathcal{D}_n^{-2} \le \hat{m} \le m + \mathcal{D}_n^{-2}x) \le P(\max_{\mathcal{D}_n^2(k-m)\in(-C,x]} V_k \ge \max_{\mathcal{D}_n^2(k-m)\in(x,C)} V_k) + o_P(1)\,,$$

by the same argumentation. Consider the decomposition of V_k similar to the one we used in the Proof of Theorem 4.3.2. Again we first analyse the decomposition with

$$\begin{aligned}V_k = &- \Big\langle \rho(m,n)S_H(m;\hat\theta_n) + \rho(k,n)S_H(k;\hat\theta_n) - \rho(m,n)E_{m,n}(\tilde\theta) - \rho(k,n)E_{k,n}(\tilde\theta) \\ &\qquad + \rho(m,n)E_{m,n}(\tilde\theta) + \rho(\tfrac{k}{n})E_{k,n}(\tilde\theta)\,, \\ &\quad \rho(m,n)S_H(m,\hat\theta_n) - \rho(m,n)S_H(m,\tilde\theta) - \rho(k,n)S_H(k,\hat\theta_n) + \rho(k,n)S_H(k,\tilde\theta) \\ &\qquad + \rho(m,n)S_H(m,\tilde\theta) - \rho(k,n)S_H(k,\tilde\theta) - \rho(m,n)E_{m,n}(\tilde\theta) + \rho(k,n)E_{k,n}(\tilde\theta) \\ &\qquad + \rho(m,n)E_{m,n}(\tilde\theta) - \rho(k,n)E_{k,n}(\tilde\theta)\Big\rangle_A \\ =&B_1 + B_2 + B_3 + B_4 + B_5 + B_6\,.\end{aligned}$$

Recall that we then have

$$\begin{aligned}B_1 &= (\rho^2(k/n) - \rho^2(m/n))\Big\|S_H(k,\hat\theta_n) - E_{k,n}(\tilde\theta)\Big\|_A^2 \\ B_2 &= -\rho^2(m/n)\Big\langle S_H(k,\hat\theta_n) - E_{k,n}(\tilde\theta) + S_H(m,\hat\theta_n) - E_{m,n}(\tilde\theta), \\ &\qquad\qquad S_H(m,\hat\theta_n) - E_{m,n}(\tilde\theta) - S_H(k,\hat\theta_n) - E_{k,n}(\tilde\theta)\Big\rangle_A \\ B_3 &= 2\Big\langle S_H(k,\hat\theta_n) - E_{k,n}(\tilde\theta), E_{m,n}(\tilde\theta)\rho^2(m/n) + E_{k,n}(\tilde\theta)\rho^2(k/n)\Big\rangle_A \\ B_4 &= -2\rho^2(m/n)\left\langle S_H(m,\hat\theta_n) - S_H(m,\tilde\theta) - S_H(k,\hat\theta_n) + S_H(k,\tilde\theta),\, E_{m,n}(\tilde\theta)\right\rangle_A \\ B_5 &= 2\rho^2(m/n)\left\langle S_H(m,\tilde\theta) - E_{m,n}(\tilde\theta) - S_H(k,\tilde\theta) + E_{k,n}(\tilde\theta),\, E_{m,n}(\tilde\theta)\right\rangle_A \\ B_6 &= -\left(\Big\|E_{m,n}(\tilde\theta)\Big\|_A^2\rho^2(m/n) - \Big\|E_{k,n}(\tilde\theta)\Big\|_A^2\rho^2(k/n)\right) \\ &= -2\left(\rho(m/n)\Big\|E_{m,n}(\tilde\theta)\Big\|_A - \rho(k/n)\Big\|E_{k,n}(\tilde\theta)\Big\|_A\right)\rho(m/n)\Big\|E_{m,n}(\tilde\theta)\Big\|_A \\ &\quad + \left(\rho(m/n)\Big\|E_{m,n}(\tilde\theta)\Big\|_A - \rho(k/n)\Big\|E_{k,n}(\tilde\theta)\Big\|_A\right)^2\,.\end{aligned}$$

From Lemma 4.3.4 we get

$$\max_{k\in(m-C\mathcal{D}_n^{-\beta},m)}|B_1| = O_P(n)\,,\qquad \max_{k\in(m-C\mathcal{D}_n^{-\beta},m)}|B_2| = O_P(\sqrt{n})\,,$$

$$\max_{k\in(m-C\mathcal{D}_n^{-\beta},m)}|B_3| = O_P(\tfrac{\mathcal{D}_n}{\sqrt{n}})\,,\qquad \max_{k\in(m-C\mathcal{D}_n^{-\beta},m)}|B_4| = o_P(n)\,.$$

Assumption **G.20**b gives us for

$$\xi_n(s) := \delta_n^{\mathtt{T}}A(S_H(m,\tilde\theta) - E_{m,n}(\tilde\theta) - S_H(m-\mathcal{D}_n^{-2}s,\tilde\theta) + nE_{m-\mathcal{D}_n^{-2}s,n}(\tilde\theta))$$

that

$$\Big\{\xi_n(s),\ s>0\Big\} \xrightarrow{d} \{W_s,\ s>0\}$$

and for

$$h(s,n) := \frac{\rho(m/n)\Big\|E_{(m,n)}(\tilde\theta)\Big\|_A - \rho((m-\mathcal{D}_n^{-1/\beta}s)/n)\Big\|E_{((m-\mathcal{D}_n^{-1/\beta}s),n)}(\tilde\theta)\Big\|_A}{\rho^2(m/n)n}$$

we have from **G.21**b

$$\max_{s\in(-C,0)} |h(s,n) - |s|g_\lambda(s)| = o_P(1)\,.$$

Let $\xi^{(1)}_{m,k} = (S_H(m,\tilde\theta) - E_{m,n}(\tilde\theta) - S_H(k,\tilde\theta) + E_{k,n}(\tilde\theta))$, then we have

$$\begin{aligned}
&\max_{\mathcal{D}_n^2(m-k)\in(0,C)} \frac{1}{\rho^2(m/n)\,n} V_k \\
&\quad= \max_{\mathcal{D}_n^2(m-k)\in(0,C)} V_k^{(1)} \\
&\quad= \max_{\mathcal{D}_n^2(m-k)\in(0,C)} \left(d_n^T A\xi^{(1)}_{m,k} - 2(m-k)g_\lambda((m-k)\mathcal{D}_n^2)\mathcal{D}_n^2 + o_P(1)\right)\,.
\end{aligned}$$

Doing the same for $m < k < m + C\mathcal{D}_n^{-2}$, we can now argue analogously as in the proofs of the Theorem 2.3.3 leads to the claim. ■

Smooth functions

Proposition 4.3.1

Under the assumptions **L.9** – **L.11** we have **G.13** with

$$C(\tilde\theta) = \lambda\mathbb{E}[\nabla H(X_1^{(1)},\tilde\theta)](\mathbb{E}[\nabla G(X_1^{(1)},\tilde\theta)])^{-1} + (1-\lambda)\mathbb{E}[\nabla H(X_1^{(2)},\tilde\theta)](\mathbb{E}[\nabla G(X_1^{(2)},\tilde\theta)])^{-1}\,.$$

Proof:
The proof follows by splitting the sums at m. Each sum is now a sum of stationary and ergodic observations. Analogous argumenation as in the proof of Proposition 3.3.2 the claim follows. ■

Deterministic function $\mathbf{E}_{\lfloor \mathbf{sn}\rfloor,\mathbf{n}}(\tilde\theta)$

Proposition 4.3.2

Under the assumptions **G.12** and **L.8** – **L.11** we have for H being an unbiased estimation function for θ under H_1, i.e. $\lambda\mathbb{E}[H(X_1^{(1)},\tilde\theta)] + (1-\lambda)\mathbb{E}[H(X_1^{(2)},\tilde\theta)] = 0$, then

$$E_{\lfloor sn\rfloor,n}(\tilde\theta) = g(\lfloor sn\rfloor)\mathbb{E}[H(X_1^{(1)},\tilde\theta_n)]$$

with $\tilde\theta_n$ as in **L.8** and

$$g(\lfloor sn\rfloor) = \begin{cases} \lfloor sn\rfloor & , s\le\lambda \\ \frac{1}{1-\lambda}(m - \lfloor sn\rfloor\lambda) & , s>\lambda \end{cases}\,.$$

Proof:
Observe, with the assumptions the Proposition 4.3.1 holds true and $\frac{1}{n}S_G(n,\tilde\theta) \xrightarrow{p} 0$. The only relevant part is then

$$\frac{1}{n}S_H(\lfloor ns\rfloor;\tilde\theta)\,.$$

As we have stationary and ergodic parts which are independent, we get from the ergodic theorem the convergence against the expectation in each. Now using the unbiasedness under H_1, gives the claim. ■

Proposition 4.3.3

Let the assumptions of Proposition 4.3.2 hold true. If additionally the assumption **L.12** holds, then **G.18** and **G.21**a, **G.21**b hold true with each having a limit function

$$g_\lambda(s) = \begin{cases} \lambda\rho(\lambda)(\rho(\lambda) + \lambda\rho'(\lambda)) & , s < 0 \\ \rho'(\lambda) - \lambda\rho(\lambda)(\rho(\lambda) + \lambda\rho'(\lambda)) & , s > 0 \end{cases}$$

either with $s \in \mathbb{Z}$ or $s \in \mathbb{R}$.

Proof:

For **G.18** we use the special form of $E_{\lfloor sn\rfloor,n}(\tilde{\theta})$ and calculate

$$\begin{aligned} \left| \left\| E_{\lfloor sn\rfloor,n}(\tilde{\theta}) \right\|_A - \left\| E_{\lfloor rn\rfloor,n}(\tilde{\theta}) \right\|_A \right| &= |\lfloor sn\rfloor - \lfloor rn\rfloor| \mathcal{D}_n \\ &\le c_2 |s - r| n \mathcal{D}_n \, , \end{aligned}$$

with

$$\mathcal{D}_n = \left\| E[H(X_1^{(1)}, \tilde{\theta}_n)] \right\|_A^2 .$$

To show the results for the two alternatives we notice first

$$\begin{aligned} \frac{1}{n^2} \left(\left\| E_{\lfloor sn\rfloor,n}(\tilde{\theta}) \right\|_A^2 \rho^2(s) - \left\| E_{m,n}(\tilde{\theta}) \right\|_A^2 \rho^2(\lambda) \right) \\ &= \left\| E[H(X_1^{(1)}, \tilde{\theta}_n)] \right\|_A^2 \left(\left(\frac{\lfloor sn\rfloor}{n} \right)^2 \rho^2(s) - \left(\frac{m}{n} \right)^2 \rho^2(\lambda) \right) \\ &= \left(2\tilde{s}\rho(\tilde{s}) \left(\rho(\tilde{s}) + \tilde{s}\rho\prime(\tilde{s}) \right) (s - \lambda) + o(1) \right) \mathcal{D}_n^2 \\ &\ge c > 0 \, . \end{aligned}$$

for $s > \lambda$. On the other hand $\tilde{s} \in (\lambda, 1 - \alpha)$, we have the boundedness.

Additionally, we can conclude for the maximum

$$\begin{aligned} \max_{k \in (m - K\mathcal{D}_n^{-2}, m)} \frac{1}{n} \Big(- \left\| E_{m,n}(\tilde{\theta}) \right\|_A^2 \rho^2(\lambda) - \left\| E_{k,n}(\tilde{\theta}) \right\|_A^2 \rho^2(k/n) \Big) \\ = 2 \max_{\frac{k}{n} \in (\lambda - \frac{K}{\mathcal{D}_n^2 n}, \lambda);\ \tilde{l} \in (m,k)} \left(\frac{\tilde{l}}{n} \rho\left(\frac{\tilde{l}}{n}\right) \left(\rho\left(\frac{\tilde{l}}{n}\right) + \frac{\tilde{l}}{n}\rho'\left(\frac{\tilde{l}}{n}\right) \right) (m - k) \right) \mathcal{D}_n^2 \, . \end{aligned}$$

Since $\mathcal{D}_n^2 n \underset{n\to\infty}{\longrightarrow} \infty$, we get

$$\begin{aligned} \sup_{\mathcal{D}_n^{-2}(m-k) \in (0,K)} \Bigg| n^{-1}\rho^{-2}(m/n) \left(\left\| E_{k,n}(\tilde{\theta}) \right\|_A^2 \rho^2(k/n) - \left\| E_{m,n}(\tilde{\theta}) \right\|_A^2 \rho^2(\lambda) \right) \\ -2(k - m)\mathcal{D}_n^2 \lambda\rho^{-1}(\lambda)(\rho(\lambda) + \lambda\rho'(\lambda)) \Bigg| = o_P(1) \, . \end{aligned}$$

■

5. Simulations - Neural Network based change-point test and estimator for NLAR(p)-processes

In this section we are interested in simulation studies of the asymptotic behaviour of the CP-estimator. For analysing the estimated distribution and the theoretical limit distribution, we also have to approximate the second one. The algorithms are implemented in **R**.

The change-point test and the change-point estimator depend on the long-run variance. As we have shown, we can replace the long-run variance with an estimator without changing the asymptotic distribution of the test. In section 5.1 we discuss some difficulties in the common long-run variance estimator and state the used one.

Besides the problems with the long-run variance estimator. We get some problems for the analysis of the asymptotic distribution of the change-point estimator. The main problem is the unknown limit of the neural network estimator $\tilde{\theta}_0$. In section 5.2 we discuss how to find a suitable representative of $\tilde{\theta}$.

A simulation study on the power of the test and the distribution of the change-point estimator is done. Thereby, we first consider the case of correct specification, section 5.3. In section 5.4 we analyse the test and estimator under misspecification.

5.1. Long-run variance estimator

As we usually do not know the true long-run variance, we replace it with a consistent estimator. Even in the univariate case the estimation of the long-run variance leads to statistical problems. The estimations mostly contain higher errors than in the case of the variance estimation. In the multivariate case this leads also to the problem of the increase of incorrect estimations. Moreover, we would like the estimator to converge even under the alternative to some reasonable matrix.

In Kirch and Tadjuidje Kamgaing [2012], the authors used the splitted variance estimator instead of the long-run variance. The idea is that the error made with the long-run variance estimator is higher as the error made with the splitted variance estimator for a good approximation. Have in mind, even in the case of multivariate i.i.d. observations with a high dimension of unknown parameters, we have difficulties in deriving a good estimation. So, in the case of long-run variance this problem increases, as we have a higher error probability. This will lead to size distortion. A discussion on this can be found in Kirch et al. [2015].

We are going to use the proposed splitted variance estimator for the simulation study and later on in the applications.

5.2. Representative of $\tilde{\theta}$

We are interested in the simulation of $\tilde{\theta}$ as we want to verify if the simulated distribution based on a small finite number of observations is close to the asymptotic distribution or not. Normally, we would try different sample sizes, estimate the unknown parameter and observe, if the parameter differs from one number of observations to the next more than some threshold. But this algorithm is not possible, as we have to deal with the non identifiability of $\tilde{\theta}$. So we have to transform the problem. From the asymptotic distribution, Theorem 4.2.3 we know that we are only interested in the distance of the neural network functions on the support of the regression vector. So we can transform the problem in finding a representant $\hat{\theta}_N$ such that

$$|f(\hat{\theta}_N, \cdot) - f(\tilde{\theta}, \cdot)|$$

becomes necessary small.

In section 3.2 we stated in Theorem 3.2.1 that the neural network estimator is consistent. So our problem reduces to finding an N such that

$$P(|f(\hat{\theta}_N, \cdot) - f(\tilde{\theta}, \cdot)| < \epsilon) \geq 1 - \Delta$$

for some $\Delta, \ \epsilon > 0$.

Because we do not know $\tilde{\theta}$, we cannot calculate the lower bound for ϵ directly. To solve this problem of the unknown $\tilde{\theta}_0$ we analyse the distances of several independently estimated $\hat{\theta}_N^i$, $i = 1, \ldots, M$. Then we have the following theoretical result for the lower bound of the ϵ distance.

Lemma 5.2.1

Let $\hat{\theta}_N^i$, $i = 1, 2$, be independent estimates for the neural network parameter given by (4.6)*, with*

$$\mathbb{E}\Big(f\left(\hat{\theta}_N^1, \mathbb{X}_1\right) - f\left(\hat{\theta}_N^2, \mathbb{X}_1\right)\Big)^2 < c$$

for some $c > 0$, where X_1 is independent of $\hat{\theta}_N^i$ for $i = 1, 2$. Then for every neural network estimator $\hat{\theta}_N$ given by (4.6) *it follows*

$$P(|f(\hat{\theta}_N, \mathbb{X}_1) - f(\tilde{\theta}, \mathbb{X}_1)| < \epsilon) \geq 1 - \frac{c}{2\epsilon^2}$$

Proof:
Observe that $\mathcal{L}\left(\hat{\theta}_N^i\right) = \mathcal{L}\left(\hat{\theta}_N\right)$, $i = 1, 2$, such that it holds

$$\begin{aligned}
c &> \mathbb{E}\left[\left(f\left(\hat{\theta}_N^1, \mathbb{X}_1\right) - f\left(\hat{\theta}_N^2, \mathbb{X}_1\right)\right)^2\right] \\
&= \mathbb{E}\left[\left(f\left(\hat{\theta}_N^1, \mathbb{X}_1\right) - f\left(\tilde{\theta}, \mathbb{X}_1\right) + f\left(\tilde{\theta}, \mathbb{X}_1\right) - f\left(\hat{\theta}_N^2, \mathbb{X}_1\right)\right)^2\right] \\
&= 2\mathbb{E}\left[\left(f\left(\hat{\theta}_N^1, \mathbb{X}_1\right) - f\left(\tilde{\theta}, \mathbb{X}_1\right)\right)^2\right] \\
&\quad +2\mathbb{E}\left[\mathbb{E}\left[\left(f\left(\tilde{\theta}, \mathbb{X}_1\right) - f\left(\hat{\theta}_N^1, \mathbb{X}_1\right)\right)\left(f\left(\tilde{\theta}, \mathbb{X}_1\right) - f\left(\hat{\theta}_N^2, \mathbb{X}_1\right)\right)\Big|\mathbb{X}_1\right]\right] \\
&=^{\ddagger} 2\mathbb{E}\left[\left(f\left(\hat{\theta}_N^1, \mathbb{X}_1\right) - f\left(\tilde{\theta}, \mathbb{X}_1\right)\right)^2\right] + 2\mathbb{E}\left[\mathbb{E}\left[\left(f\left(\tilde{\theta}, \mathbb{X}_1\right) - f\left(\hat{\theta}_N, \mathbb{X}_1\right)\right)\Big|\mathbb{X}_1\right]^2\right] \\
&\geq 2\mathbb{E}\left[\left(f\left(\hat{\theta}_N, \mathbb{X}_1\right) - f\left(\tilde{\theta}, \mathbb{X}_1\right)\right)^2\right]
\end{aligned}$$

Then it follows

$$\begin{aligned}
P(|f(\hat{\theta}_N, \mathbb{X}_1) - f(\tilde{\theta}, \mathbb{X}_1)| < \epsilon) &\geq 1 - \frac{1}{\epsilon^2}\mathbb{E}\left[\left(f\left(\hat{\theta}_N, \mathbb{X}_1\right) - f\left(\tilde{\theta}, \mathbb{X}_1\right)\right)^2\right] \\
&\geq 1 - \frac{c}{2\epsilon^2}
\end{aligned}$$

■

To estimate the expected distance we used the statistic

$$\frac{1}{l}\sum_{r=1}^{l}\frac{2}{(M-1)M}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\left(f\left(\hat{\theta}_N^{i,l}, \mathbb{X}_l\right) - f\left(\hat{\theta}_N^{j,l}, \mathbb{X}_l\right)\right)^2 .$$

Thereby, $\hat{\theta}_n^{j,l}$ indicates for the lth repetition the jth estimator.

We determine a representative N based on 2 examples, which we use later on for the analysis of the change-point test and estimator. Thereby, we decided for one example of a correct specified model, i.e. regression function is a neural network, and one of a misspecified model. As we are under the alternative, we have the following models.

The correct specified model, called GAR 1, is given as

$$X_t = \begin{cases} 0.5 + (1 + \exp(0.5 * (1+\mathbf{0.7}X_{t-1})))^{-1} + \varepsilon_t & t \leq m \\ 0.5 + (1 + \exp(0.5 * (1-\mathbf{0.7}X_{t-1})))^{-1} + \varepsilon_t & t > m \end{cases} .$$

In the misspecified case we have the AR 1 case, given by

$$X_t = \begin{cases} 0.3X_{t-1} + \varepsilon_t & t \leq m \\ 0.5 + 0.1X_{t-1} + \varepsilon_t & t > m \end{cases} .$$

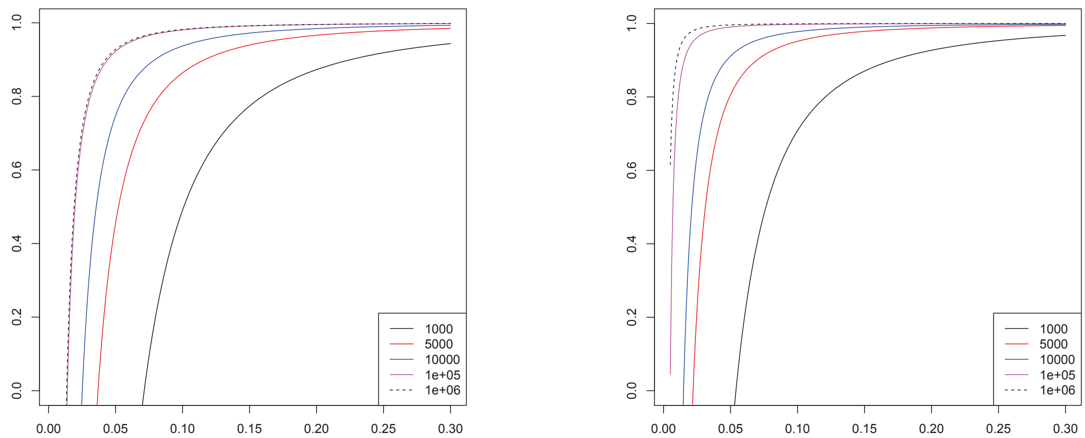
‡Using first the independence of θ^1 and θ^2, and secondly that $\mathcal{L}\left(\mathbb{E}\left[\left(f\left(\tilde{\theta}, \mathbb{X}_1\right) - f\left(\hat{\theta}_N^1, \mathbb{X}_1\right)\right)\Big|\mathbb{X}_1\right]\right) = \mathcal{L}\left(\mathbb{E}\left[\left(f\left(\tilde{\theta}, \mathbb{X}_1\right) - f\left(\hat{\theta}_N^2, \mathbb{X}_1\right)\right)\Big|\mathbb{X}_1\right]\right)$.

In Table 5.1 we present the estimation of the expected distance for the AR 1 and the GAR 1 model. Depending on N we can see that there is a strong increase for $N = 1\,000, 5\,000$ and $10\,000$. After this the increase is not very significant. Even the difference between $5\,000$ and $10\,000$ is not so big, so one may say this is a good size for the estimation of the expected distance. But as we have to divide by the maximal distance we allow on the neural network, we should choose the expected distance as small as possible but with acceptable calculation time.

N	1 000	5 000	10 000	100 000	1 000 000
AR 1	1.02 e-02	2.72 e-03	1.26 e-03	3.89 e-04	3.62 e-04
GAR 1	5.83 e-03	9.69 e-04	4.44 e-04	4.78 e-05	1.92 e-05

Table 5.1.: estimated expected distance of the neural network based on $M = 100$ replications of $\hat{\theta}_n$ for $\lambda = 0.5$

Therefore, we calculate based on the expected distance the lower bound for the probability that the neural network differs at most ϵ for two independent estimators based on N observations. Figure 5.1 shows the estimated lower bound for this probability. The y-axes contains the estimated lower bound and on the x-axes the maximal distance ϵ. The figure confirms that the increase of the sample size from $N = 10^5$ to $N = 10^6$ does not result in a strong increase of the probability of the distance.



(a) AR 1

(b) GAR 1

Figure 5.1.: estimated lower relative frequency against ϵ for different number of observations

We observe that the estimation for the parameter in the AR 1 model seems to be worse than in the GAR 1 case. Having in mind that the approximation for a linear function, as the regression line is in the AR 1 case, the neural network can not be approximate for finite h properly. This results in a slower convergence for an acceptable representative of the best approximating neural network.

As a conclusion we propose a sample size of $N = 10^5$ for the simulation of the asymptotic distribution in the analysed models.

5.3. Correct specified Model

Let's assume that the regression functions before and after the change are given by a normal network. The first example is the classical mean change model which is approximated by a neural network with 0 hidden layers.

Observe change in mean model introduced by Antoch et al. [1995]. Let $m = \lfloor \lambda N \rfloor$ and

$$X_t = \begin{cases} \mu + \varepsilon_t, & t \leq m \\ \mu + \delta + \varepsilon_t, & t > m \,. \end{cases}$$

Denote by $\theta_1 = \mu$ and $\theta_2 = \mu + \delta$ the expectation before the change and after the change, respectively. We want to approximate the regressive function by an empty neuron one layer neural network, i.e. $f(\mathbb{X}_t, \theta) = \theta$ and $\theta \subseteq \Theta \subset \mathbb{R}$. Then, by definition of $\hat{\theta}_N$ and of the consistency we get

$$\begin{aligned}\tilde{\theta}_0 =& \lambda E(\mu + \varepsilon_1) + (1-\lambda)E(\mu + \delta + \varepsilon_1) \\ =& \lambda\theta_1 + (1-\lambda)\theta_2 \,.\end{aligned}$$

Then we get for $z = 1, 2$

$$\begin{aligned}e_i^{(z)} =& X_i^{(z)} - f(\mathbb{X}_i^{(z)}, \tilde{\theta}_0) - \mathbb{E}(X_i^{(z)} - f(\mathbb{X}_i^{(z)}, \tilde{\theta}_0)) \\ =& \varepsilon_i + \theta_z - \tilde{\theta}_0 - \theta_z + \tilde{\theta}_0 \\ =& \varepsilon_i \,,\end{aligned}$$

where ε_i are the errors of the true time series. So the limit distributions becomes

$$\arg\max\{D\, W_e(s) - |s|\, D^2\, g_\lambda(s),\ s \in \mathbb{Z}\}\,,$$

where $D = (\theta_2 - \theta_1)$ and

$$g_\lambda(s) = \begin{cases} \lambda, & s \leq 0 \\ 1-\lambda, & s > 0\,. \end{cases}$$

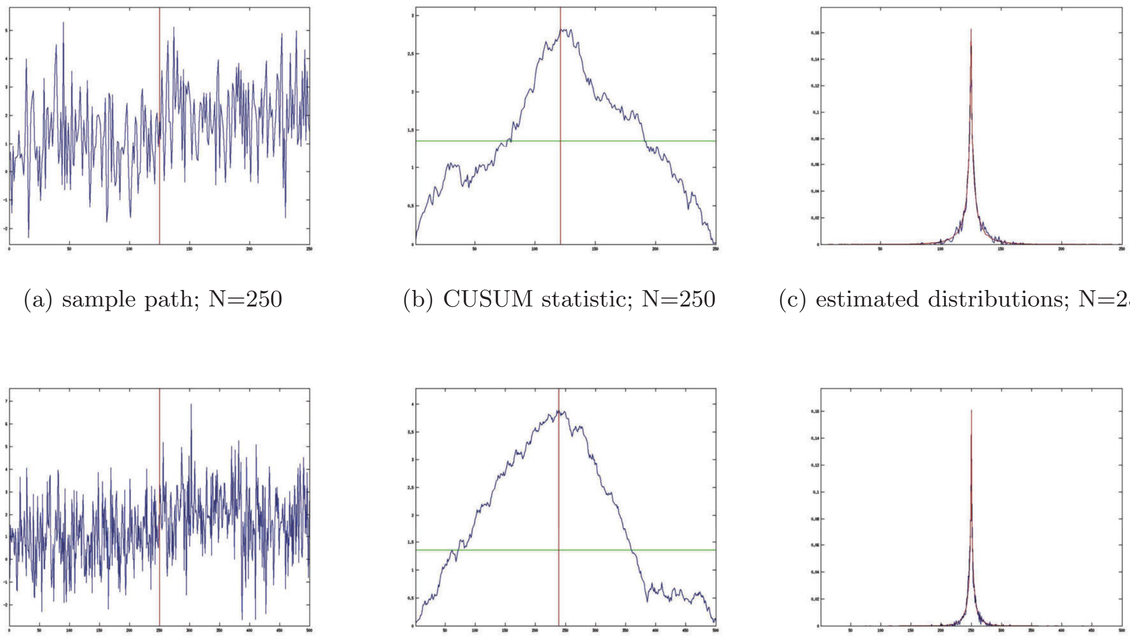
This has the same distribution as the result in Antoch et al. [1995].

In Figure 5.2 a sample path of a mean change model with $\mu = 1$ and $\delta = 1$ is given, where $\lambda = 0.5$ and $N = 250,\ 500$. The corresponding test statistic and the empirical as well as the estimated asymptotic distributions using 200 simulation steps are given in Figure 5.2, too.

Let us observe one more model which has regression function given by a neural network. In Kirch and Tadjuidje Kamgaing [2012] they analysed the power of the following example

$$X_t = \begin{cases} 0.5 + (1 + \exp(0.5 * (1 + 0.7X_{t-1})))^{-1} + \varepsilon_t & t \leq m \\ \mu + \alpha(1 + \exp(0.5 * (1 + \beta X_{t-1})))^{-1} + \varepsilon_t & t > m \end{cases},$$

with



(a) sample path; N=250 (b) CUSUM statistic; N=250 (c) estimated distributions; N=250

(d) sample path; N=500 (e) CUSUM statistic; N=500 (f) estimated distributions; N=500

Figure 5.2.: Mean Change Model with $\mu = 1$, $\delta = 1$ and standard normally distributed errors

- GAR 1: $\mu = 0.1$, $\alpha = 1$, $\beta = 0.7$
- GAR 2: $\mu = 0.5$, $\alpha = -1$, $\beta = 0.7$
- GAR 3: $\mu = 0.5$, $\alpha = 1$, $\beta = -0.7$
- GAR 4: $\mu = 0.5$, $\alpha = -1$, $\beta = -0.7$

As they discussed the power of the test for the different parameter combinations (see table 5.2. As in the cases of GAR 1, GAR 2 and GAR 4 the power is quite good, we expect the asymptotic distribution to be quite good approximated by the empirical distribution of the estimator using $M = 1000$ replications of estimates of m. The empirical and asymptotic distribution as well as a sample path and its CUSUM statistic can be found in the Figures 5.3, 5.4, 5.5 and 5.6.

N	GAR 1	GAR 2	GAR 3	GAR 4	GAR 3 (derivative w.r.t. β)
250	0.42	0.94	0.09	0.96	0.77
500	0.74	1	0.16	1	0.92

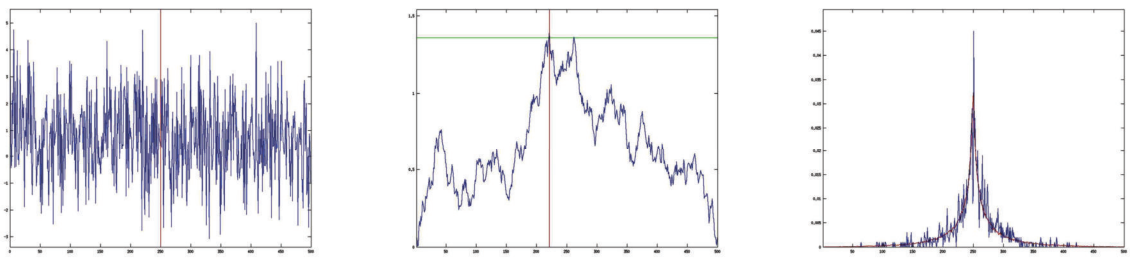
Table 5.2.: Power for the correct specified model based on 10^4 repetitions

Notice, that the structure of the asymptotic distribution is related to the power of the test. For example the GAR 3 has power 0.16 and the corresponding distribution of the estimator has larger standard deviation. So even if the model is correct specified this test has difficulties in detecting the change and estimating the time of change.

This follows as we have discussed that the used test statistic is constructed to detect changes in the mean. In the case of GAR 1, the mean changes linearly. For GAR 2 and GAR 4 we see that the positive dependence changed to a negative one. All this three cases are clearly change the expectation. But in GAR 3 the intensity of the regressor changed.

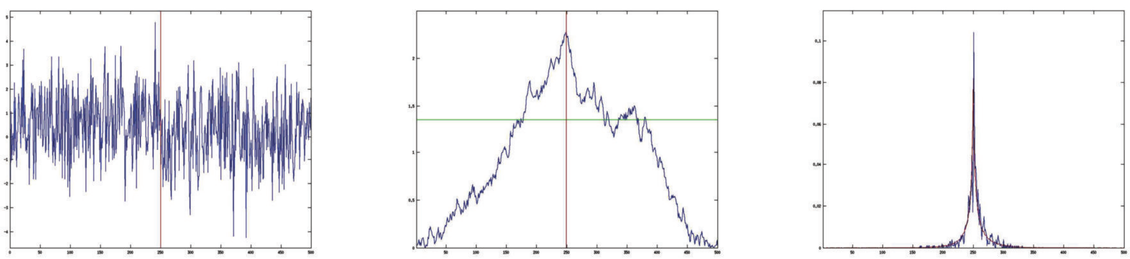
To improve the power and the detection of the change-point, we can use the derivative w.r.t. β as the test statistic. This is equivalent to using a different A in (3.32) and in (4.6). Observe that this has an advantage in the power.

We also observe that the asymptotic distribution is still wider and the estimator has problems in correct detection.



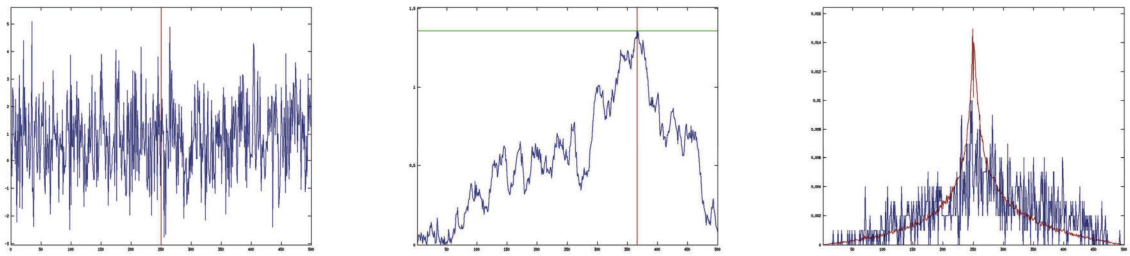
(a) GAR 1: sample path (b) GAR 1: CUSUM statistic (c) GAR 1: estimated distributions

Figure 5.3.: GAR 1 with 1 hidden layer neural network as regression function and standard normally distributed errors, N=500



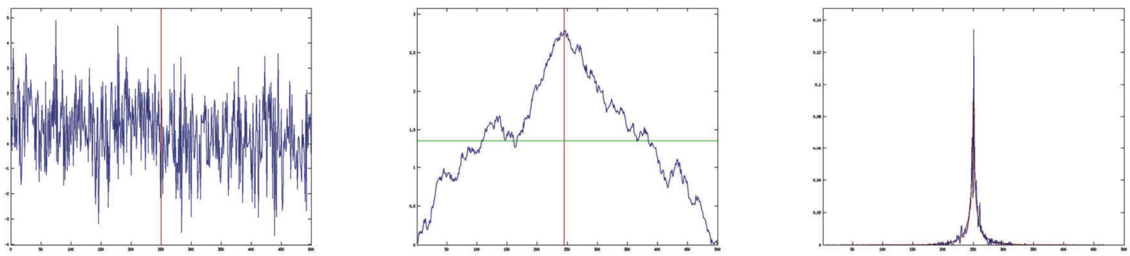
(a) GAR 2: sample path (b) GAR 2: CUSUM statistic (c) GAR 2: estimated distributions

Figure 5.4.: GAR 2 with 1 hidden layer neural network as regression function and standard normally distributed errors, N=500



(a) GAR 3: sample path (b) GAR 3: CUSUM statistic (c) GAR 3: estimated distributions

Figure 5.5.: GAR 3 with 1 hidden layer neural network as regression function and standard normally distributed errors, N=500



(a) GAR 4: sample path (b) GAR 4: CUSUM statistic (c) GAR 4: estimated distributions

Figure 5.6.: GAR 4 with 1 hidden layer neural network as regression function and standard normally distributed errors, N=500

5.4. Misspecified Model

Again we use the results of Kirch and Tadjuidje Kamgaing [2012] which discussed the behaviour of the test statistic as well as the the number of the Hidden layers for AR(1) and TAR(1)-processes. Based on this results we choose the following Models

$$X_t = \begin{cases} g_1(X_{t-1}) + \varepsilon_t & t \leq m \\ g_2(X_{t-1}) + \varepsilon_t & t > m \end{cases}$$

with

- AR 1: $g_1(x) = 0.3x$, $g_2(x) = 0.5 + 0.1x$
- AR 2: $g_1(x) = 0.3x$, $g_2(x) = 1 - 0.1x$
- TAR 1: $g_1(x) = 0.3x1_{\{x \geq 0\}} - 0.1x1_{\{x<0\}}$, $g_2(x) = (0.5 + 0.5x)1_{\{x \geq 0\}} - 0.3x1_{\{x<0\}}$
- TAR 2: $g_1(x) = 0.3x1_{\{x \geq 0\}} - 0.1x1_{\{x<0\}}$, $g_2(x) = (1 - 0.1x)1_{\{x \geq 0\}} + (0.5 + 0.1x)1_{\{x<0\}}$

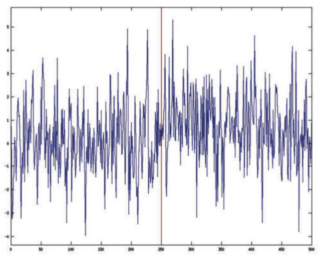
For the simulation we choose $N = 500$ observations and $M = 1000$ replications. In the AR models we approximated the regression function with a 2 layer hidden neural network using the logistic function. Where as we used in the TAR models 3 hidden layer neural network but again with the logistic function.

In Kirch and Tadjuidje Kamgaing [2012] they analysed the power of these models. Our calculation coincides with the results give there. Table 5.3 shows the results.

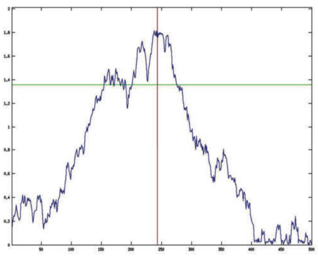
N	AR 1	AR 2	TAR 1	TAR 2
250	0.627	0.963	0.762	0.477
500	0.911	0.999	0.984	0.795

Table 5.3.: Power for the misspecified specified model based on 10^4 repetitions

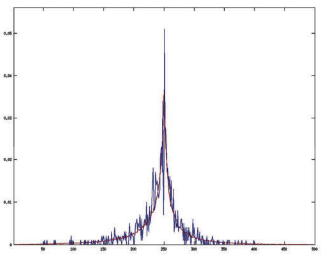
From the power we assume that the asymptotic and the empirical distributions should be close to each other, which is confirmed by the graphics in Figure 5.7 (AR 1 and AR 2) and Figure 5.8 (TAR 1 and TAR 2). Although, we have miss-specification and we know the neural network has difficulties in approximating linear functions, we still have good results. In view of the correct specification even petter, as we the GAR 3 has for different test statistics a power lower than 50%.



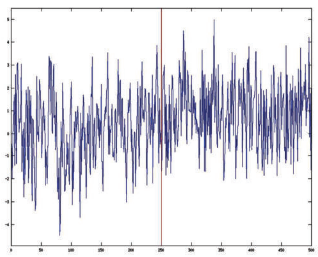
(a) AR 1: sample path



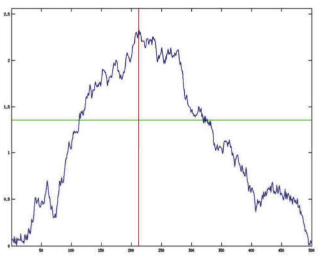
(b) AR 1: CUSUM statistic



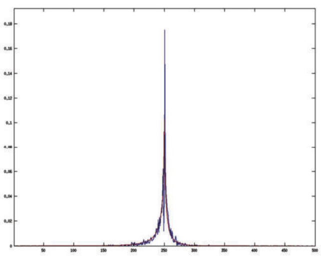
(c) AR 1: estimated distributions



(d) AR 2: sample path

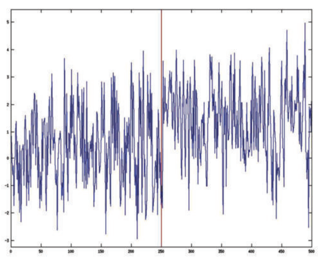


(e) AR 2: CUSUM statistic

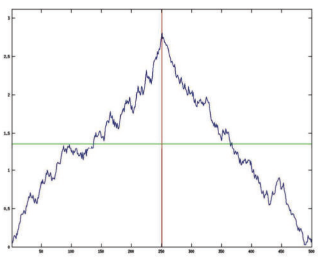


(f) AR 2: estimated distributions

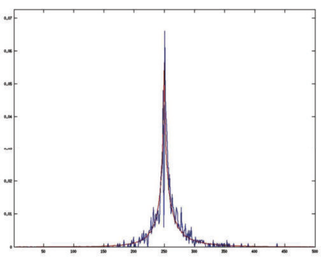
Figure 5.7.: AR(1)-Models with standard normally distributed errors, N=500



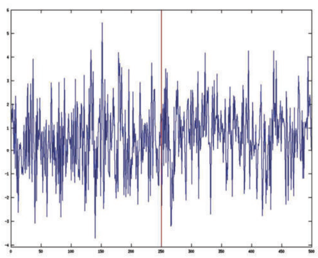
(a) TAR 1: sample path



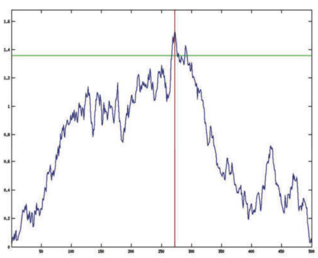
(b) TAR 1: CUSUM statistic



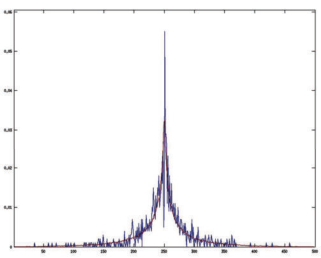
(c) TAR 1: estimated distributions



(d) TAR 2: sample path



(e) TAR 2: CUSUM statistic



(f) TAR 2: estimated distributions

Figure 5.8.: TAR(1)-Models with standard normally distributed errors, N=500

6. Applications

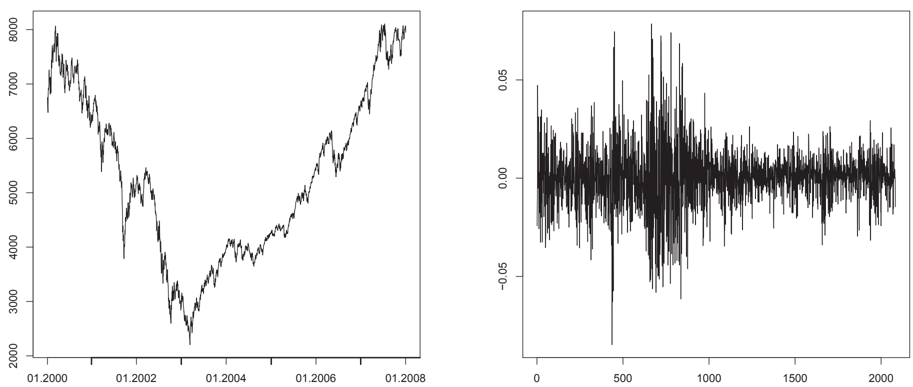
6.1. Dax - Dot-com bubble

As an example for application of change-point detection, we are going to analyse the DAX-data from 01.2000-01.2008 in which the Dot-com bubble occurred. With the Dot-com bubble massive speculations on internet-based companies in the end of the 1990's are meant. After 1995 many new companies are founded with an almost immediate emission of publicly traded stocks. High expectations in those companies having a domain name ".com" increase their stock prices. But the speculations could not be fulfilled by the companies, such that during 1999-2001 some of them failed completely and the crisis took place. Until 2004 the stock prices were falling and even afterwards most companies were underrated, compared to their starting prices.

We are going to check if we are able to detect and estimate the time of stabilization of the stock market using the NLAR set-up with a change-point test and estimator based on neural networks.

Fit time series

First we have to fit a NLAR-process to the data. Let $\{Y_t,\ t = 1, \ldots, T\}$ denote the DAX-data from 01.2000-01.2008, i.e. 2084 measurements. Usually, one is interested in the returns, i.e. $r_t = (Y_t - Y_{t-1})/Y_{t-1}$. The DAX-values and the corresponding returns are given in Figure 6.1.



(a) DAX (b) Returns

Figure 6.1.: DAX-data from 01.2000-01.2008 and the corresponding returns

A possible model for the returns is given by $r_t = \sigma_t Z_t$. In Stărică and Granger [2005] they proposed to choose σ_t piecewise constant and Z_t some autoregressive process. Then the log-returns would result in $\log(r_t^2) = 2\log(\sigma_t) + \log(Z_t^2)$. In Figure 6.1 we observe that the change occurs in the variance. As the test and estimator are constructed for change in the mean, we can use them for detecting a change in the structure r_t in applying them to $\log(r_t^2)$. But the squared returns can become also 0 due to rounding effects. So, instead of the usual analysis of the log-returns we are going to work with the Fuller log-transformed returns of the DAX-data. In Fuller [1996] this transformation is given as

$$X_t = \log(r_t^2 + \rho\hat{\sigma}_r^2) - \frac{\rho\hat{\sigma}_r^2}{r_t^2 + \rho\hat{\sigma}_r^2}, \tag{6.1}$$

where $\hat{\sigma}_r^2$ is the sample variance of the returns and $\rho > 0$ some weight. We will call it the Fuller transformation. With the Fuller transformation we come close to our assumption of a NLAR-process. In the following we use $\rho = 0.02$.

We need to find the order of the stationary parts. Therefore, we analyse the first 600 and the last 600 observations. In Figure 6.2 we used the graphical methods of acf and pacf to get a first idea about the orders. Interpreting the graphics leads to the idea of a maximal AR(6). But to justify the model we used the `auto.arima`* function in **R** based on the AIC to identify the model and get the following:

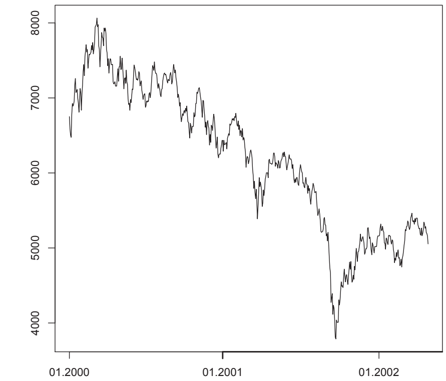
```
ARIMA(1,0,1) with non-zero mean

Coefficients:
         ar1      ma1  intercept
      0.9157  -0.8269    -9.5192
s.e.  0.0374   0.0508     0.1529

sigma^2 estimated as 3.408:  log likelihood=-1217.76
AIC=2443.52   AICc=2443.58   BIC=2461.1
```

This leads to an ARMA(1,1) with an intercept. Avoiding a moving average part we gain an AR(6):

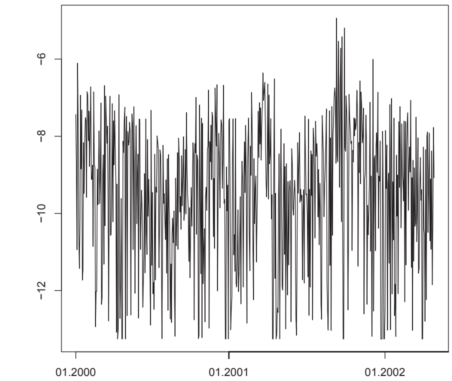
```
ARIMA(6,0,0) with non-zero mean

Coefficients:
         ar1     ar2     ar3     ar4     ar5     ar6  intercept
      0.0735  0.0575  0.0414  0.0871  0.0757  0.1224    -9.5201
s.e.  0.0405  0.0405  0.0404  0.0406  0.0407  0.0407     0.1371

sigma^2 estimated as 3.394:  log likelihood=-1214.53
AIC=2445.05   AICc=2445.29   BIC=2480.23
```

For the change-point test and estimator we have to use one order for the whole time-series. But with a too large chosen order we run into the problem of overfitting the data. Therefore, we analyse the assumed second stationary part and check which order we would assume there. In Figure 6.3 we have again a graphical analysis using the acf and pacf. As for the first part it is hard to tell which kind of model to fit, but at least the order seems to be smaller than

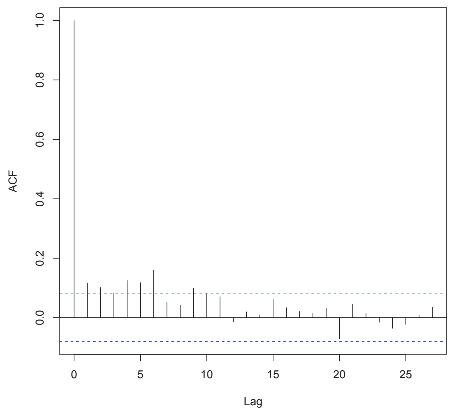
*Contained in the forecast package.



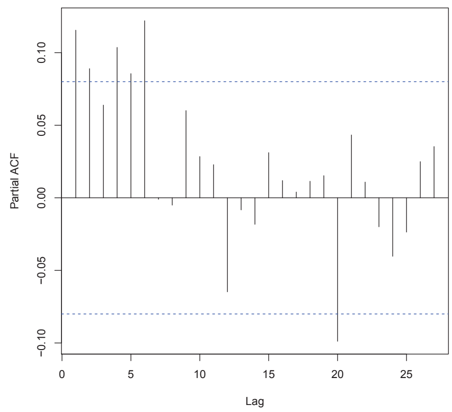
(a) DAX first part



(b) Fullertransformed returns of the first part



(c) acf

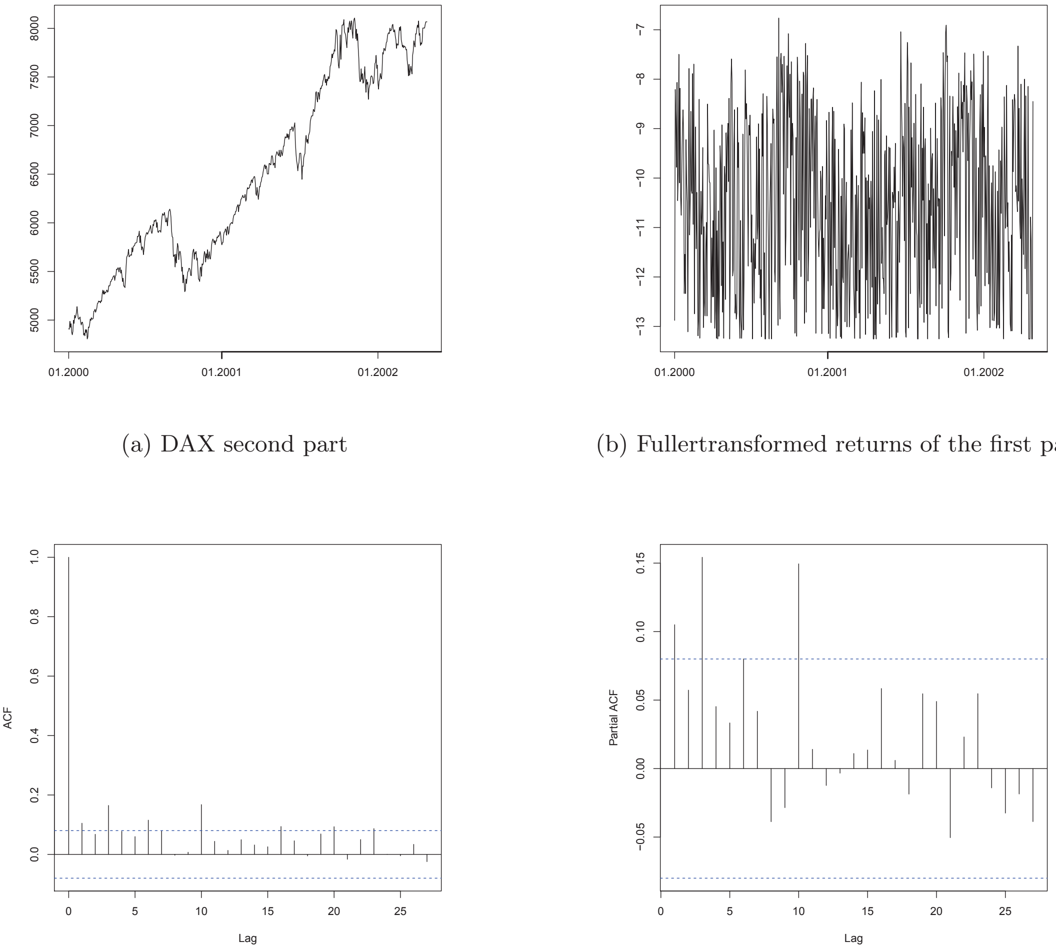


(d) pacf

Figure 6.2.: graphical analysis of the first 600 observations

10. The analogue estimation using the `auto.arima` function from the **R**-package `forecast` leads to either an ARMA(2,1) with an intercept or to an AR(3).

```
ARIMA(2,0,1) with non-zero mean

Coefficients:
         ar1     ar2      ma1  intercept
      0.9426  0.0025  -0.8760   -10.5273
s.e.  0.0722  0.0476   0.0593     0.1554

sigma^2 estimated as 2.954:  log likelihood=-1174.36
AIC=2358.73   AICc=2358.83   BIC=2380.71
```



(a) DAX second part

(b) Fullertransformed returns of the first part

(c) acf

(d) pacf

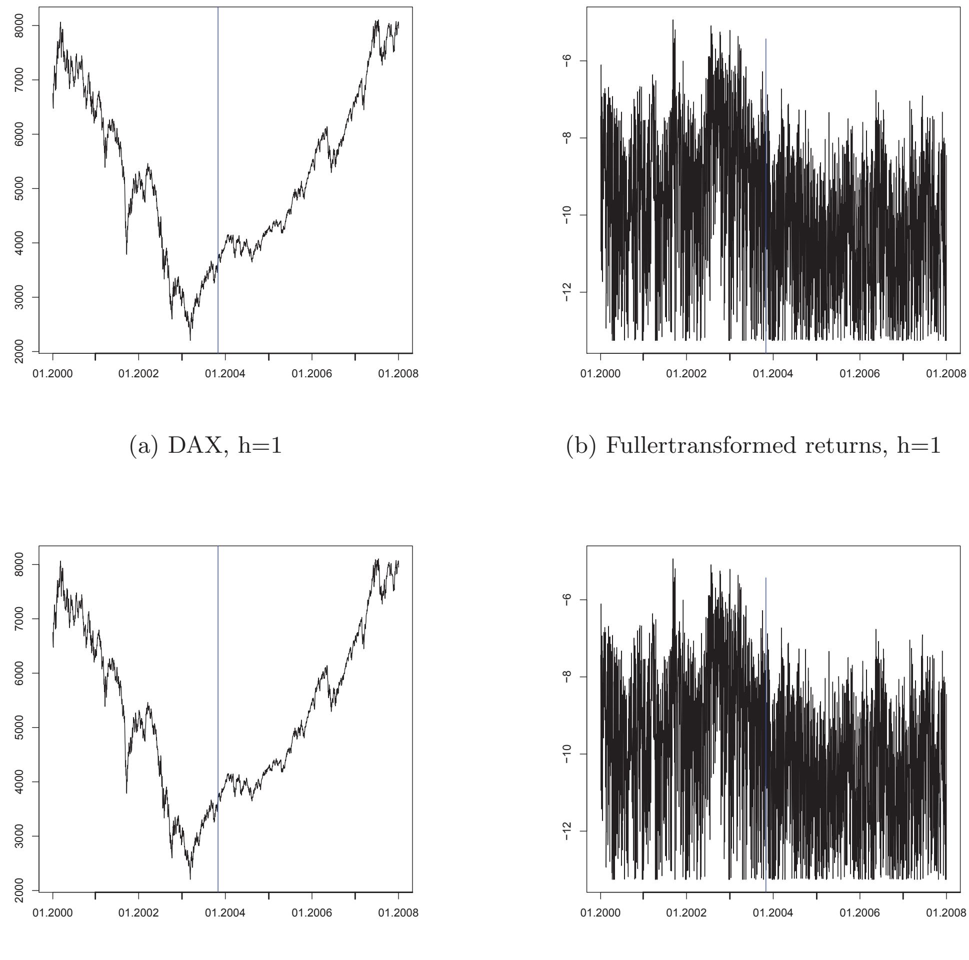
Figure 6.3.: graphical analysis of the last 600 observations

```
ARIMA(3,0,0) with non-zero mean

Coefficients:
         ar1     ar2     ar3  intercept
      0.0893  0.0417  0.1559   -10.5270
s.e.  0.0404  0.0407  0.0406     0.0981

sigma^2 estimated as 2.967:  log likelihood=-1175.64
AIC=2361.28   AICc=2361.38   BIC=2383.27
```

Finally, we decide for the smaller order 3 to avoid overfitting. For neural networks there is still the decision of the number of hidden neurons to make. There are different test procedures even in combination to the choice of the order. A good overview about this part is given in Anders [1997] but we will leave this out and just try different numbers of hidden neurons. Note that the application of order selection methods based on linear autoregressions does not necessarily provide good orders for non-linear autoregressions, e.g. neural networks. We rather expect the latter to be smaller as the non-linear models offer more freedom in characterizing the dependence. However, for the change-point tests we do not need a very good model but only a reasonable one as we use a change-point approach designed for model misspecification anyhow.

6.1.1. Change-point detection and estimation

If we want to use the neural network based change-point test and change-point estimator for a NLAR(p)-process, we have to find p and h (number of hidden neurons). As discussed in the section before, we use a NLAR(3)-model for the test with different numbers of hidden neurons. We choose $h = 1, \ldots, 5$. The results for $h = 1, \ldots, 5$ to the significance level $\alpha = 0.05$ do not differ much, we give exemplarily the results for $h = 1, 2$ in the Figure 6.4.



(a) DAX, h=1

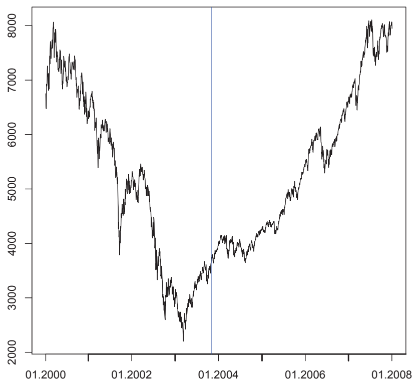
(b) Fullertransformed returns, h=1

(c) DAX, h=2

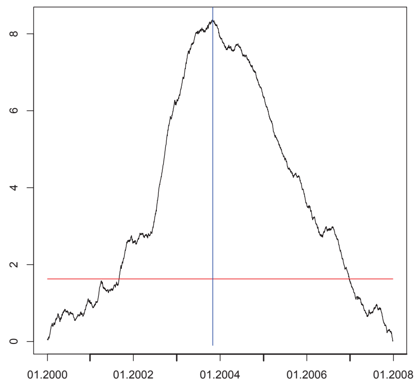
(d) Fullertransformed returns, h=2

Figure 6.4.: DAX and Fullertransformed returns with the detected and estimated change-point (blue line), $\alpha = 0.05$

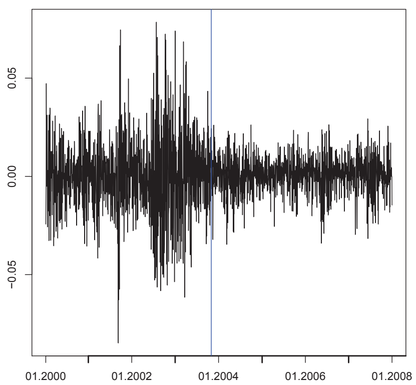
We mentioned at the beginning, that the stock market for ".com" companies stabilised after 2004. This can be detected at the change in the variance of the returns. To show the significance, we finish this section with the Figure 6.5 giving the DAX, the cumulated sum of the sample residuals (Cumsum, not to be confused with CUSUM, the test statistic), the returns and the Fuller transformed returns for the change-point test to the significance level $\alpha = 0.01$. Even for this level the change-point is quite significant, so the dot-com bubble had a significant influence on the stock prices of the DAX.



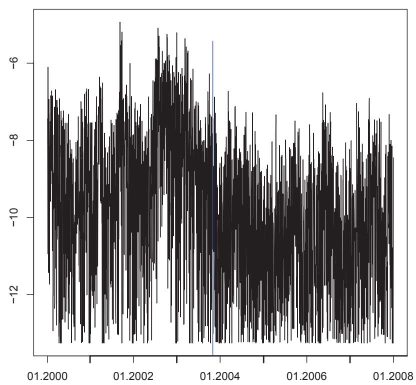
(a) DAX, $\alpha = 0.01$



(b) Cumsum, $\alpha = 0.01$



(c) returns, $\alpha = 0.01$

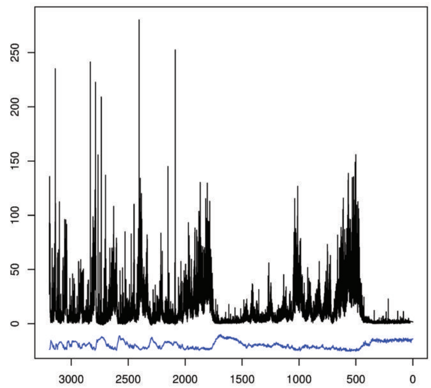


(d) Fuller transformed returns, $\alpha = 0.01$

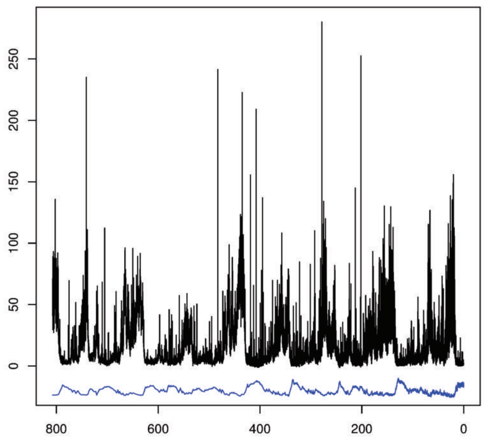
Figure 6.5.: estimated change-point using significance level $\alpha = 0.01$ with constant number of hidden neurons $h = 1$

6.2. Ice Data

The Antarctic ice sheet is a huge persistent ice body existing over millions of years. It is a valuable archive for reconstructing climate time series. We consider a set of measurements done on an ice core drilled in the framework of EPICA (European partnership of ice coring in Antarctica) at Dome C. For further informations see Bigler et al. [2010]. The measurements were performed on a continuously melted core section (Continuous flow analysis) giving depth profiles of different trace elements and other species. Here, we focus on the depth profiles of calcium ion concentration (a tracer of dust in the atmosphere) and temperature estimated from measured water isotopes. The data are equidistant with depth with some missing data points on breaks and other core imperfections. The age of the ice was estimated by a complex flow model that allows for ice flow under continuously increasing compression. The resulting age-depth relation was used to interpolate the measured data points to a time-equidistant data point series. The dataset contains the depth in m, the calcium concentrations in ppb as well as the estimated temperature at the Antarctic (based on isotopic data) and the estimated time in ka (based on other measurements - here 1 ka=1000 years), each has a length of 281 000 values.



(a) against depth



(b) against age

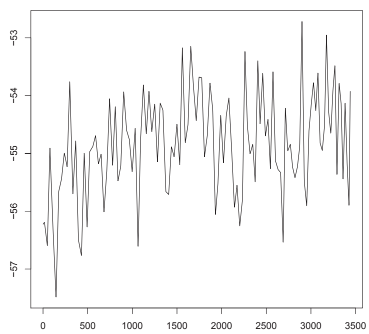
Figure 6.6.: black line is the measured calcium values, blue line is the estimated temperature

From Figure 6.6 we can see that if temp is low, we have a higher mean as well as a bigger variance of the calcium concentration. That confirms that there might be a correlation between the temperature and the calcium value. The data also contains many outliers, probably errors based on measurements,that is why normality is not assumed. Additionally, the measurements are not equidistant w.r.t. the depth. But the interpolation was done such that we can assume equidistant measurements w.r.t. the time (age).

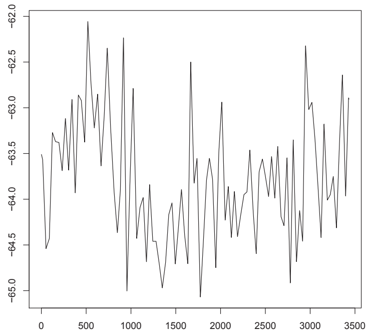
Dependent on the cold and warm periods at the Antarctic, we expect changes in the calcium concentration. We are interested in detecting and estimating these change-points. As we want to be flexible in the model we use the change-point test and change-point estimator for NLAR-processes based on neural network functions. For the theoretical results we have assumed p and h fixed, compare section 3.2 and 4.2. To determine p we first analyse which order the NLAR with a neural network as regression function can be assumed for both cold and warm periods.

6.2.1. Fit non-linear autoregressive model

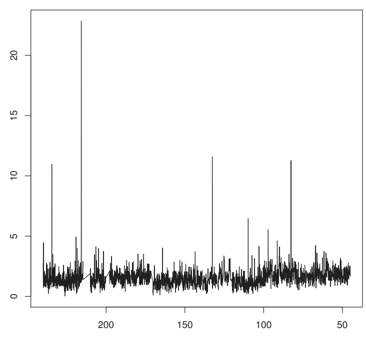
We fit a non-linear autoregressive model (NLAR) with neural networks as regression functions for the cold and the warm periods, where we first split the data. The first part (warm period) is chosen in depth between 45 and 240 and the second part (cold period) is in depth between 470 and 631 (see Figure 6.7). We first fit to each subsample a NLAR.



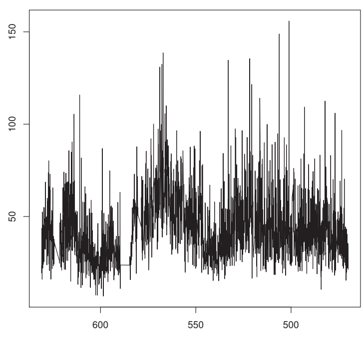
(a) temperature warm period



(b) temperature cold period



(c) calcium warm period



(d) calcium cold period

Figure 6.7.: temperature of the desired stationary representation parts and the corresponding calcium levels

As we can see in Figure 6.8 the first observations between depth 45 and 240, i.e. between 0.368 and 12.183 thousand years ago, seem to be stationary up to some linear part caused by the interpolation of missing data. Thus, we use this part to fit an underlying model. To model the value of calcium as a time-series $\{X_t\}$ with t denotes the number of the measurement, where the first measurement X_1 corresponds to the measured value in depth $d_1 = 3192.391$

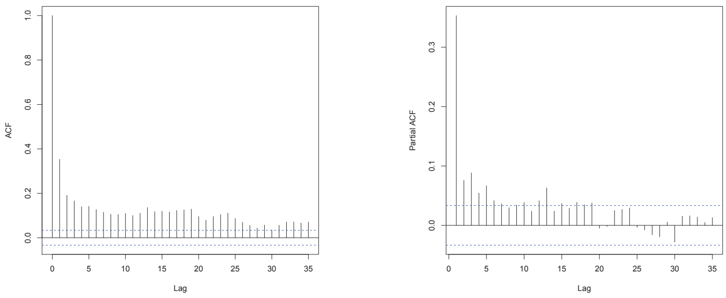


Figure 6.8.: acf and pacf of the observations in depth between 45 and 240

and the last measured value X_{281000} in depth $d_{281000} = 0$. So we are assuming a causal time-series, i.e.

$$X_t = m(X_{t-1}, \ldots, X_{t-p}) + \eta_t \tag{6.2}$$

with m an unknown function and η_t some noise.

We can see, that both acf and pacf indicate no pure AR or MA process. The slow decrease of the acf indicates an ARIMA process. Identifying the order of the underlying time-series is not possible by graphical methods. We use the `auto.arima` function in **R** to identify the model on one hand and force this function to fit an AR(p)-process.

```
Series: Ca.first
ARIMA(1,1,1)

Coefficients:
         ar1      ma1
      0.2689  -0.9687
s.e.  0.0182   0.0064

sigma^2 estimated as 0.4751:  log likelihood=-3600.1
AIC=7206.2   AICc=7206.21   BIC=7224.63
```

```
Series: Ca.first
ARIMA(10,0,0) with non-zero mean

Coefficients:
         ar1     ar2     ar3     ar4     ar5     ar6     ar7     ar8     ar9    ar10  intercept
      0.3037  0.0338  0.0604  0.0256  0.0472  0.0269  0.0235  0.0180  0.0224  0.0388     1.5423
s.e.  0.0170  0.0178  0.0178  0.0178  0.0178  0.0179  0.0179  0.0179  0.0179  0.0171     0.0294

sigma^2 estimated as 0.479:  log likelihood=-3609.76
AIC=7243.53   AICc=7243.62   BIC=7317.25
```

The result coincides with the further result. The ARIMA(1,1,1) seems to fit the calcium levels best. But we would like to get a NLAR(p)-process, therefore we prefer an order of 10.

We do have mentioned that we only analysed the part of the process which seems to be stationary. Let us know find a model for a part which obviously is different to the first observations.

The part in depth between 470 and 631 is one candidate. We proceed as before. Again we

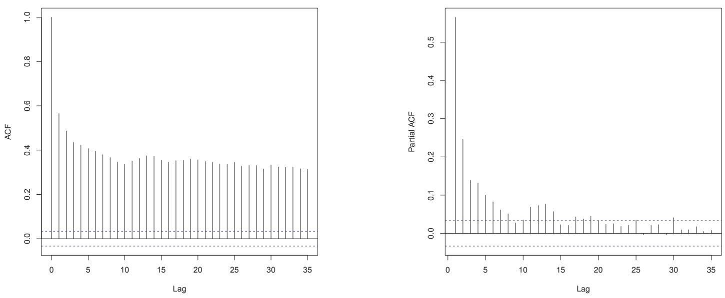


Figure 6.9.: acf and pacf of the observations in depth between 470 and 631

see, the acf is slowly falling which indicates a integrated part. To verify our assumption we determine the most plausible orders for this time-series based on AICc.

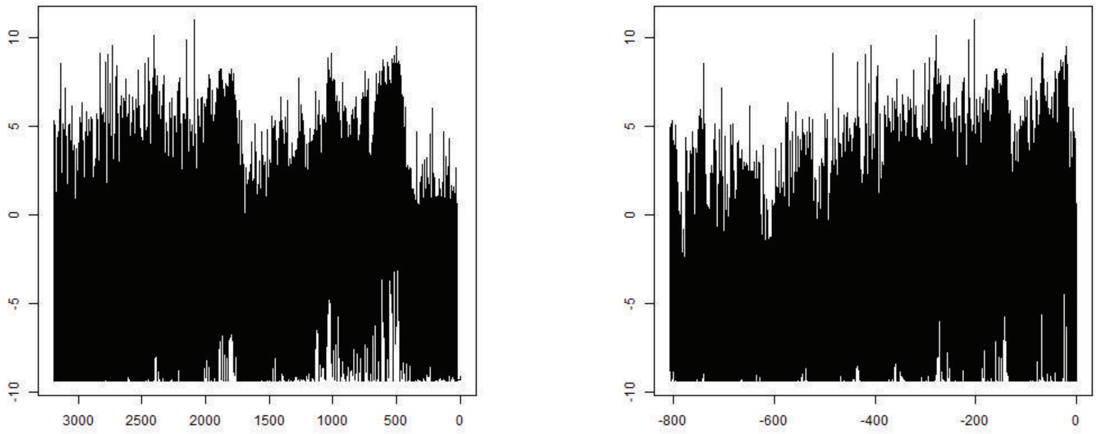
```
Series: Ca.first
ARIMA(2,1,2)

Coefficients:
         ar1      ar2      ma1     ma2
      0.7469  -0.0479  -1.4307  0.4525
s.e.  0.1503   0.0526   0.1488  0.1404

sigma^2 estimated as 161.9:  log likelihood=-13597.8
AIC=27205.6   AICc=27205.62   BIC=27236.31
```

```
Series: Ca.second
ARIMA(8,0,0) with non-zero mean

Coefficients:
         ar1     ar2     ar3     ar4     ar5     ar6     ar7     ar8  intercept
      0.3441  0.1381  0.0579  0.0733  0.0576  0.0539  0.0437  0.0515    42.7447
s.e.  0.0170  0.0180  0.0181  0.0181  0.0181  0.0181  0.0180  0.0170     1.2177

sigma^2 estimated as 166.4:  log likelihood=-13646.11
AIC=27312.22   AICc=27312.28   BIC=27373.63
```

We derive that the best fit would be given with an ARIMA(2,1,2) but as we use a non-linear autoregressive process we would prefer the forced autoregressive order for the neural network function as the number of explanatory variables, i.e. $p = 8$.

In combination we choose the smaller one, as we want to avoid over-fitting. Thus for the calcium levels itself we choose a NLAR(11)-process as model. Then it is left to determine the number of hidden neurons.

As our data are concentrations with high variability, we apply a normalizing and variance stabilizing transformation. Again we choose the Fuller-transformation of section 6.1 to the first difference of the calcium values. The latter are considered to remove the effect of integration.



(a) Depth (b) Age

Figure 6.10.: Fuller transformed calcium innovations

As in the previous subsection we first select the order of a linear auto-regression based on the data- here often differencing to remove the effect of integration. The results based on the `auto.arima` function from **R**-package `forecast` are given below. In order to verify the result we determine the difference and check again the acf and pacf. Based on the graphics in figure 6.11 we may fit a MA(2). From the result of the order selection based on AICc for the first part of the calcium levels, we know that a better model might be a ARMA(p,q) process. We again determine the order of the process based on the AICc for the innovations of the calcium levels.

```
Series: Ca.first.diff
ARIMA(1,0,1) with zero mean

Coefficients:
         ar1      ma1
      0.2689  -0.9687
s.e.  0.0182   0.0064

sigma^2 estimated as 0.4751:  log likelihood=-3600.1
AIC=7206.2   AICc=7206.21   BIC=7224.63
```

```
Series: Ca.first.diff
ARIMA(5,0,0) with zero mean

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -0.5992  -0.4915  -0.3577  -0.2575  -0.1349
s.e.   0.0169   0.0193   0.0201   0.0193   0.0169
```

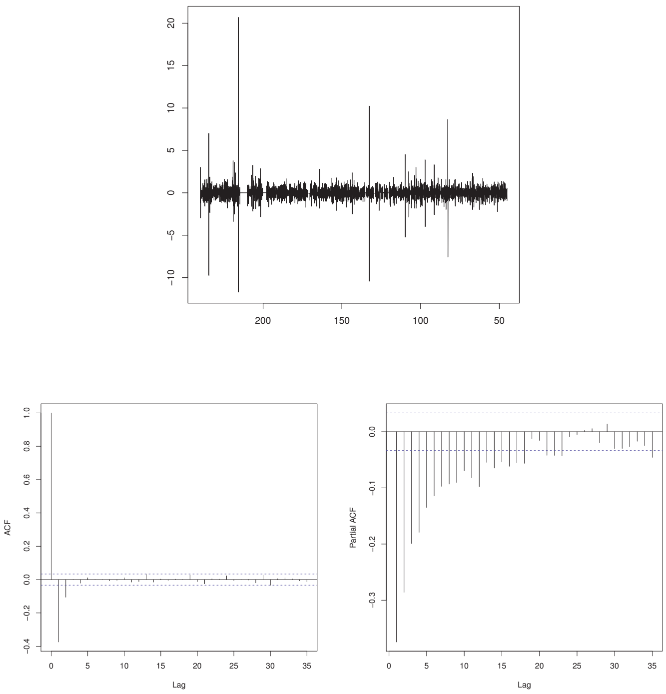


Figure 6.11.: first line: values of the calcium concentration differences in depth between 45 and 240; second line: acf and pacf of these observations

```
sigma^2 estimated as 0.5229:  log likelihood=-3762.79
AIC=7537.59   AICc=7537.61   BIC=7574.44
```

Let us do the same for the second part. Let us verify this by determining the difference and check the assumption that this might be a ARMA(2,2) process.

```
Series: Ca.second.diff
ARIMA(2,0,2) with zero mean

Coefficients:
         ar1      ar2      ma1     ma2
      0.7469  -0.0479  -1.4307  0.4525
s.e.  0.1503   0.0526   0.1488  0.1404

sigma^2 estimated as 161.9:  log likelihood=-13597.8
AIC=27205.6   AICc=27205.62   BIC=27236.31
```

```
Series: Ca.second.diff
ARIMA(5,0,0) with zero mean

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -0.6059  -0.4294  -0.3349  -0.2250  -0.1245
s.e.   0.0169   0.0195   0.0200   0.0195   0.0169
```

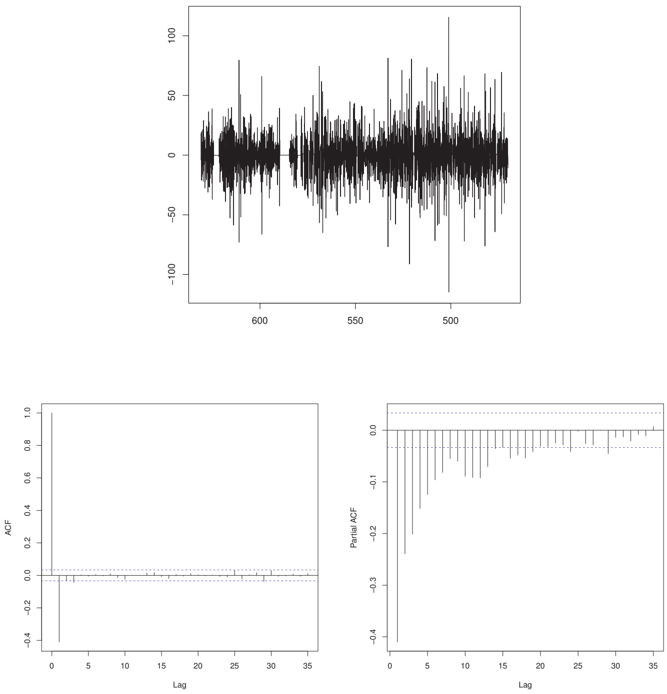


Figure 6.12.: first line: calcium innovations in depth between 470 and 631; second line: acf and pacf of the reversed observations

```
sigma^2 estimated as 173.7:  log likelihood=-13717.95
AIC=27447.9   AICc=27447.93   BIC=27484.75
```

Here we observe, that a fit of a ARMA(2,2) would fit best. If we enforce an AR(p)-process we end up with the order 5.

As we allow a non-linear AR(p)-process and due to the general approximation property of neural networks, we have to choose p and h in appropriate way. The order and the estimated coefficients for the AR(p) fitted processes are in the first and second part nearly the same. Therefore, we choose for the Fuller transformed innovations of the calcium level an NLAR(p)-process of order $p = 2$ based on the general fittings. For the calcium level itself, we have decided to choose $p = 8$ as this is the smallest order and we want to avoid over-fitting.

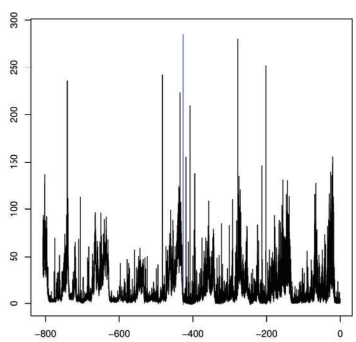
6.2.2. Neuronal network based change-point test and estimator

In this example we are not interested in detecting only one change but we expect many to exist. The presented test statistic is an AMOC, i.e. at most one change, statistic. To overcome this problem the binary segmentation (BS) algorithm was introduced. This algorithm recursively checks for change-points. The idea of the BS algorithm is given in Algorithm 4.

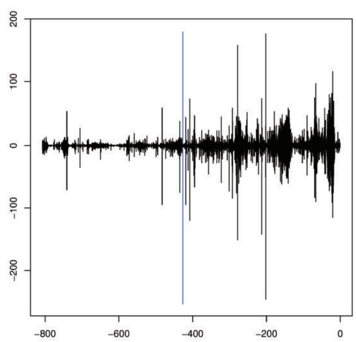
Algorithm 4 BS (binary segmentation)

1: Start with only one subample equal to the whole sample $\{X_t\}$, $t = 1, \ldots, n$. For each current subsample, say $\{X'_t\}$, $t = 1, \ldots, M$, test for a change-point using the critical value ζ_n.
2: If a change-point is detected,, estimate it by $\hat{m}_0$ and add the latter to the set of already detected change-points. Split the subsample $\{X'_t\}$, $t = 1, \ldots, M$ into two new subsamples $\{X'_t\}$, $t = 1, \ldots, \hat{m}_0$ and $\{X'_t\}$, $t = \hat{m} + 1, \ldots, M$. If no change-point is detected, then fix the subsample $\{X'_t\}$, $t = 1, \ldots, M$ and do not consider it further on.
3: Repeat steps 1-2 for any not yet fixed subsample until all subsamples are fixed, i.e. in none of them a $change - point$ is detected.

The algorithm given here is from Fryzlewicz [2014]. He proposed to choose $\zeta_n = C\sqrt{2 \log n}$. Then it remains to determine C , which he has done using a simulation study. As the time was not enough for a wide simulation study, we choose $C = 1$ to run the BS. It is left to determine the number of explanatory variables as well as the number of hidden neurons. Due to the discussion above, we use a dependence of order 8 and 1. For the number of hidden neurons we used 1 till 5. Additionally, we run the test for the Fuller transformed differences of the calcium levels. Therefore, we used 2 and 5 as the number of explanatory variables but with the same number of hidden neurons, i.e. $h = 1, \ldots, 5$. The figures show only the values where the procedure detected at least one change. The other combinations fail to detect a change point.



(a) Observations: h= 1 and p=2



(b) Innovations: h=1 and p=2

Figure 6.13.: Time-series and innovations with the detected change-points based on NeuNet-statistic applied on time-series modelled as NLAR(1)-process with the $C = 1$

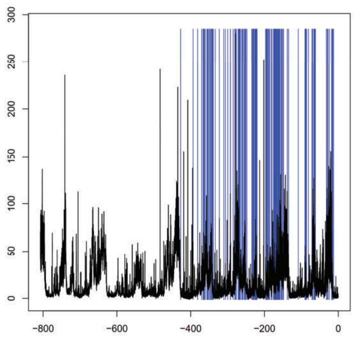
In the graphics we present the results using the binary segmentation algorithm. We observe that the NLAR(1) model in combination with a boundary parameter $C = 1$ detects only one change.

Using an NLAR(8) model we get a better result, see Figure 6.14. But also observe that with an increase of the number of hidden neurons the detection get worse.

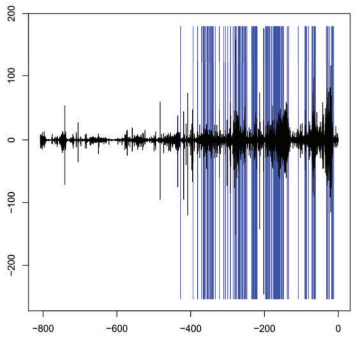
Differently to this result we observe that in the case of the Fuller transformations we gain better results. In Figure 6.15 and Figure 6.16 we have the results for the Fuller transformed calcium innovations modelled as an NLAR(2)-process. The results are promising, especially for $H = 1$ and $H = 2$. Again we observe that increasing the number of hidden neurons leads to underestimation of the number of changes, as the neural network fits can adapt the change. In the case of $H = 5$ we get worse results as expected, see Figure 6.17 and Figure 6.18. In the fit we have seen that for a linear AR(5)-process the parameter for the cold and for the warm period did not differ much. So it is not surprising that the algorithm has trouble in detecting changes.

We observe, that the result based on the Fuller transformed calcium innovations with $p = 5$ and $h = 2$ yields a result we would expect. Although this one fails in detecting between 200ka and 400ka the results are the best. A simulation study for a more reasonable choice of the constant may give an even better result.

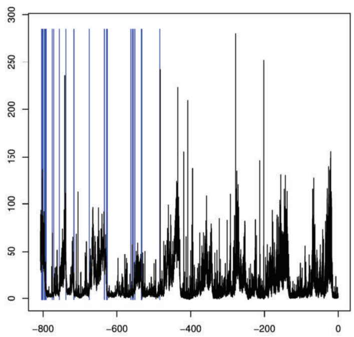
We can see, the result clearly depends on the choice of the number of hidden neurons. There exists some algorithms for automatically choosing the number of hidden neurons and the order of the process. An overview is given in Anders [1997]. We applied one model selection strategy based on hypotheses tests (Teräsvirta/Lin/Granger 1993) which often produces good results (Anders [1997]). But due to the interpolations, the data contain linear dependent parts. These parts are reasons why the algorithm fails and always prefers one hidden layer (the starting value). Alternatively, one may try some other strategies or even other test statistics as robust ones for the change-point detection.



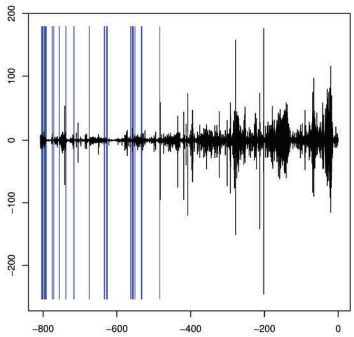
(a) Observations: h= 1 and p=8



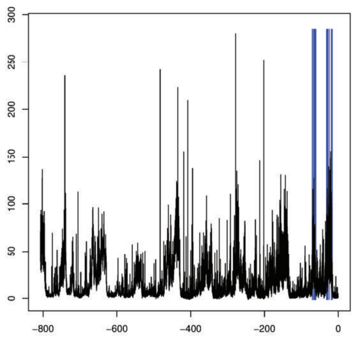
(b) Innovations: h=1 and p=8



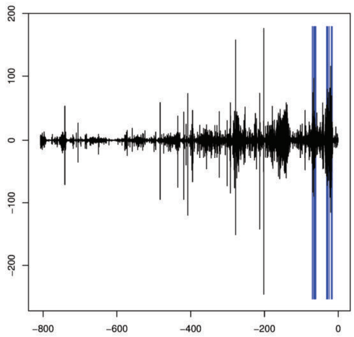
(c) Observations: h= 2 and p=8



(d) Innovations: h=2 and p=8

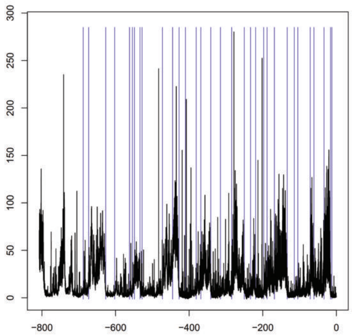


(e) Observations: h= 3 and p=8

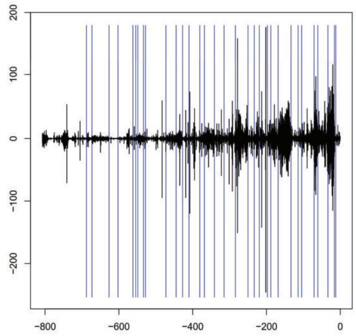


(f) Innovations: h=3 and p=8

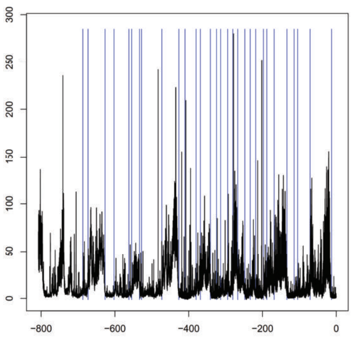
Figure 6.14.: Time-series and innovations with the detected change-points based on NeuNet-statistic applied on time-series modelled as NLAR(8)-process with the $C = 1$



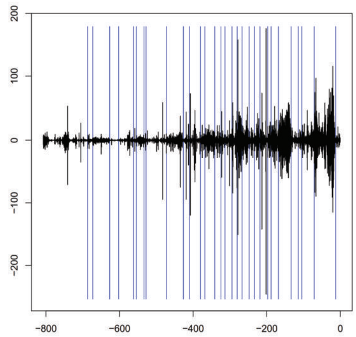
(a) Observations: h=1 and p=2



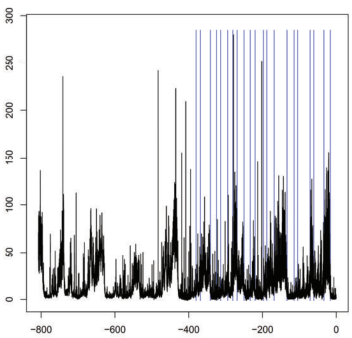
(b) Innovations: h=1 and p=2



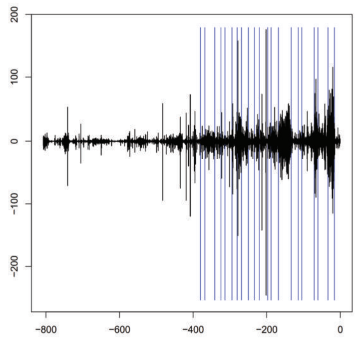
(c) Observations: h=2 and p=2



(d) Innovations: h=2 and p=2

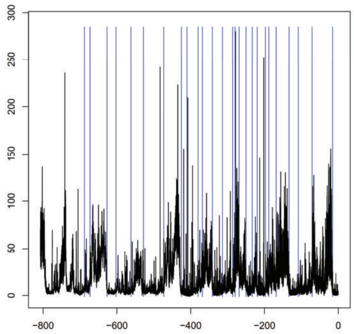


(e) Observations: h=3 and p=2

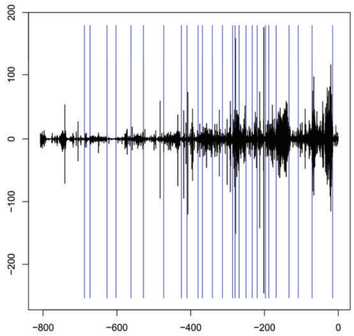


(f) Innovations: h=3 and p=2

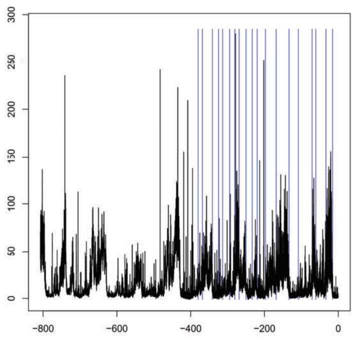
Figure 6.15.: Time-series and innovations with the detected change-points based on NeuNet-statistic applied on the Fuller transformed innovations of the time-series to modelled as NLAR(2)-process, $C = 1$, for $h = 1, 2, 3$



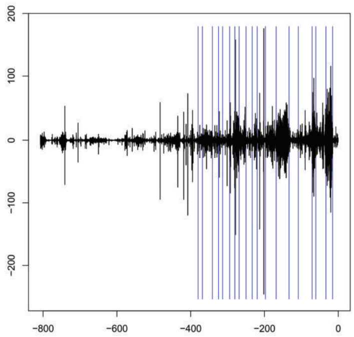
(a) Observations: h=4 and p=2



(b) Innovations: h=4 and p=2

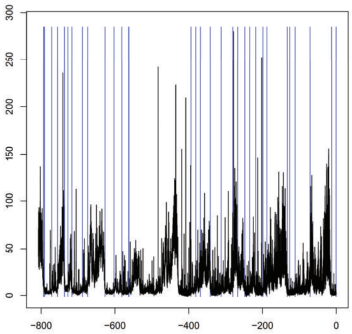


(c) Observations: h=5 and p=2

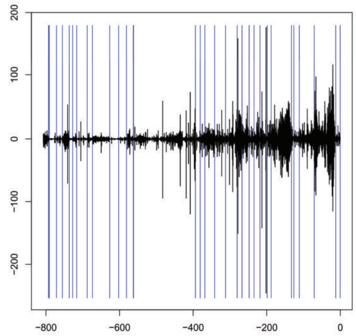


(d) Innovations: h=5 and p=2

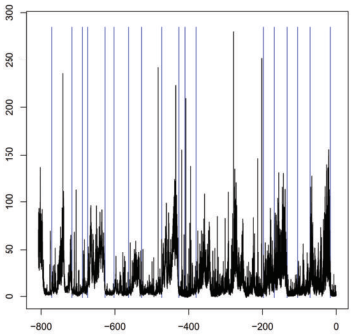
Figure 6.16.: Time-series and innovations with the detected change-points based on NeuNet-statistic applied on the Fuller transformed innovations of the time-series to modelled as NLAR(2)-process, $C = 1$,for $h = 4, 5$



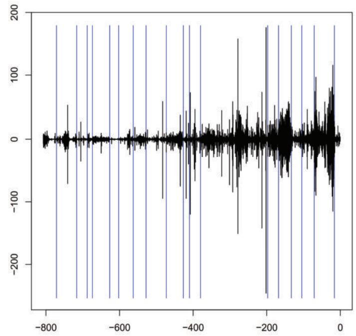
(a) Observations: h=1 and p=5



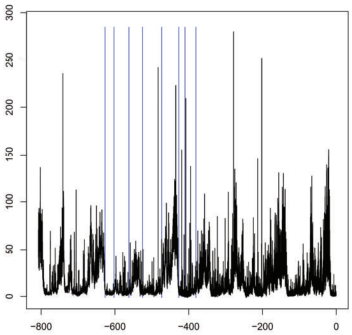
(b) Innovations: H=1 and p=5



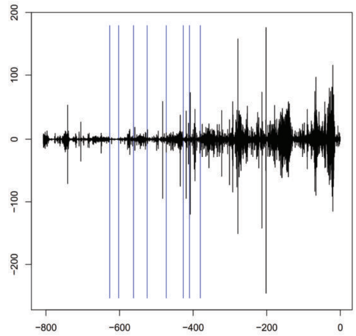
(c) Observations: h=2 and p=5



(d) Innovations: h=2 and p=5

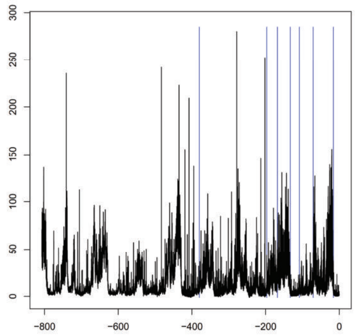


(e) Observations: h=3 and p=5

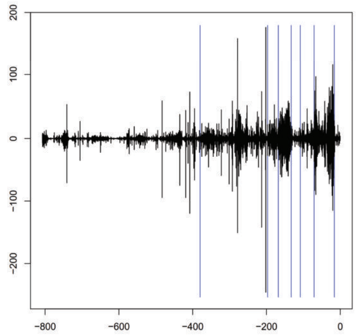


(f) Innovations: h=3 and p=5

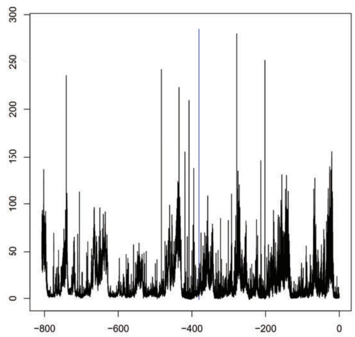
Figure 6.17.: Time-series and innovations with the detected change-points based on NeuNet-statistic applied on the Fuller transformed innovations of the time-series to modelled as NLAR(5)-process, $C = 1$, for $h = 1, 2, 3$



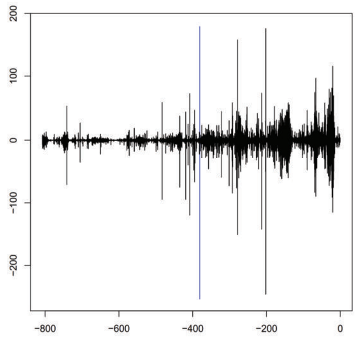
(a) Observations: h=4 and p=5



(b) Innovations: h=4 and p=5



(c) Observations: h=5 and p=5



(d) Innovations: h=5 and p=5

Figure 6.18.: Time-series and innovations with the detected change-points based on NeuNet-statistic applied on the Fuller transformed innovations of the time-series to modelled as NLAR(5)-process, $C = 1$, for $h = 4, 5$

7. Further research

In this thesis, we derived a general framework for change-point tests and estimators based on estimation functions. This allows to show the asymptotic results for different change-point models by verifying the regularity conditions.

As we have asymptotic results and data consisting of a finite number of observations, a common technique to derive a critical value is via bootstrap. Bootstrap can also be used to derive confidence intervals for the change-point estimator. In both cases, we would first show that the bootstrap parameter estimator is $\sqrt{n}$-consistent even under H_1. Having done this, it is left to prove that the regularity conditions are fulfilled.

For i.i.d. data, this likely can be done based on results for the bootstrap in misspecified regression models using neural network approximations. The $\sqrt{n}$-consistency of the bootstrap estimator for neural networks under H_0 was proven in Franke and Neumann [1998]. With the splitting technique it should be possible to derive the asymptotics of this estimator even under H_1. Proving then that the sample residuals still fulfil the regularity conditions would give us the result.

To handle the multiple change-points in the ice data, we used the technique of binary segmentation. Another approach would be the wild binary segmentation. In both cases the algorithms depend on parameter choices. To determine a suitable choice, a simulation study has to be done. We mentioned that the automatic algorithms for determining the order of the non-linear autoregressive model are not applicable due to the interpolations. Another possible way of detecting the change could be based on the original data. They have the disadvantage that they are not equidistantly measured. But the relation between time (years) and depth (meter) of the measurement is known. For multiple changes the MOSUM statistic (moving sum) could also be used. This statistic usually has a fixed size window moving over the data. In each step the test statistic is calculated. If the value of the test statistic reaches some critical value a change is detected. Direct application for the interpolated data would be possible, but also a modified version changing the window size in relation to the change in time would be possible.

We also want to mention that even the set-up with the change-point test and estimator having randomized weight function could be advantageous for our applications. We could think about plugin-weights replacing the power in the usually used weight function with an estimator based on the observations. But of course, we have to ensure that the regularity conditions are still fulfilled.

A. Neural network and Landau symbolds

A.1. Identifiability of neural network parameter

Definition A.1.1

Let $\theta = (\nu_0, \mu_1, \ldots, \mu_h)$. Then

a) *a transposition, i.e. a permutation changing only two elements, π of $(\nu_0, \mu_1, \ldots, \mu_h)$, is given for $i < k$ by*

$$\pi_{ik}(\nu_0, \mu_1, \ldots, \mu_h) = (\nu_0, \mu_1, \ldots, \mu_{i-1}, \mu_k, \mu_{i+1}, \ldots, \mu_{k-1}, \mu_i, \mu_{k+1}, \ldots, \mu_h)$$

b) *a symmetry transformation π_k of $(\nu_0, \mu_1, \ldots, \mu_h)$ is defined as $\pi_k(\nu_0, \mu_1, \ldots, \mu_h) = (\nu_0 + \mu_k, \mu_1, \ldots, \mu_{k-1}, -\mu_k, \mu_{k+1}, \ldots, \mu_h)$.*

Lemma A.1.1

Assume

a) *$f(y, \theta)$ is not redundant (i.e. there exists no other networks with fewer hidden neurons ($h' < h$) that represent exactly the same relationship function,*

b) *$f(y, \theta)$ is irreducible, i.e. for all $i \neq 0$, $j \neq 0$*

   a) $\nu_i \neq 0$

   b) $\alpha_i \neq 0$

   c) $(\alpha_i, \beta_i) \neq (\alpha_j, \beta_j)$ for all $i \neq j$.

Moreover, in Hwang and Ding [1997] Theorem 2.3 a) a sufficient condition on identifiability is given.

Theorem A.1.1 *Hwang and Ding [1997] Theorem 2.3 a) Assume*

a) *$f(y, \theta)$ is not redundant (i.e. there exists no other networks with fewer hidden neurons ($h' < h$) that represent exactly the same relationship function,*

b) *$f(y, \theta)$ is irreducible, i.e. for all $i \neq 0$, $j \neq 0$*

   a) $\nu_i \neq 0$

   b) $\alpha_i \neq 0$

   c) $(\alpha_i, \beta_i) \neq (\alpha_j, \beta_j)$ for all $i \neq j$.

then θ is identifiable up to a family of symmetry transformations and transpositions.

We are going to observe the behaviour of the derivatives.

Lemma A.1.2

Assume the generating function ϕ be bounded from below and above and 2-times differentiable. Let $Y = (Y_1, \dots, Y_q)^\mathsf{T}$ be a random vector with finite second moment. For a compact set Θ we have

$$\sup_{\theta\in\Theta} \|\nabla f(Y,\theta)\|_\infty \leq c \max_{i=1,\dots,q} |Y_i|$$

and for all $i = 1, \dots, q$

$$\sup_{\theta\in\Theta} \|\nabla(\nabla f)_i(Y,\theta)\|_\infty \leq c \max_{j,i=1,\dots,q} |Y_iY_j|\,.$$

Proof:
First we calculate the derivatives of $f(y,\theta) = \nu_0 + \sum_{i=1}^h \nu_i\phi(\langle\alpha_i, y\rangle + \beta_i)$.

The first derivative is given as

$$\frac{\partial}{\partial\nu_0} f(y,\theta) = 1\,, \qquad \frac{\partial}{\partial\alpha_{ij}} f(y,\theta) = \nu_i\phi'(\langle\alpha_i, y\rangle + \beta_i)y_j\,,$$

$$\frac{\partial}{\partial\nu_i} f(y,\theta) = \phi(\langle\alpha_i, y\rangle + \beta_i)\,, \qquad \frac{\partial}{\partial\beta_i} f(y,\theta) = \nu_i\phi'(\langle\alpha_i, y\rangle + \beta_i)\,.$$

The second derivative is given as

$$\begin{aligned}
\frac{\partial}{\partial\nu_k}\left(\frac{\partial}{\partial\nu_i} f(y,\theta)\right) &= 0 \\
\frac{\partial}{\partial\alpha_{kj}}\left(\frac{\partial}{\partial\nu_i} f(y,\theta)\right) &= \begin{cases}\phi'(\langle\alpha_i, y\rangle + \beta_i)y_j & k = i\\ 0 & \text{else}\end{cases} \\
\frac{\partial}{\partial\beta_k}\left(\frac{\partial}{\partial\nu_i} f(y,\theta)\right) &= \begin{cases}\phi'(\langle\alpha_i, y\rangle + \beta_i) & k = i\\ 0 & \text{else}\end{cases} \\
\frac{\partial}{\partial\nu_k}\left(\frac{\partial}{\partial\alpha_{ij}} f(y,\theta)\right) &= \begin{cases}\phi'(\langle\alpha_i, y\rangle + \beta_i)y_j & k = i\\ 0 & \text{else}\end{cases} \\
\frac{\partial}{\partial\alpha_{kl}}\left(\frac{\partial}{\partial\alpha_{ij}} f(y,\theta)\right) &= \begin{cases}\nu_i\phi'(\langle\alpha_i, y\rangle + \beta_i)y_jy_l & k = i\\ 0 & \text{else}\end{cases} \\
\frac{\partial}{\partial\beta_k}\left(\frac{\partial}{\partial\alpha_{ij}} f(y,\theta)\right) &= \begin{cases}\nu_i\phi'(\langle\alpha_i, y\rangle + \beta_i)y_j & k = i\\ 0 & \text{else}\end{cases} \\
\frac{\partial}{\partial\nu_k}\left(\frac{\partial}{\partial\beta_i} f(y,\theta)\right) &= \begin{cases}\phi'(\langle\alpha_i, y\rangle + \beta_i) & k = i\\ 0 & \text{else}\end{cases} \\
\frac{\partial}{\partial\alpha_{kl}}\left(\frac{\partial}{\partial\beta_i} f(y,\theta)\right) &= \begin{cases}\nu_i\phi'(\langle\alpha_i, y\rangle + \beta_i)y_l & k = i\\ 0 & \text{else}\end{cases} \\
\frac{\partial}{\partial\beta_k}\left(\frac{\partial}{\partial\beta_i} f(y,\theta)\right) &= \begin{cases}\nu_i\phi'(\langle\alpha_i, y\rangle + \beta_i) & k = i\\ 0 & \text{else}\end{cases}
\end{aligned}$$

The statement follows from the boundedness of the generating function ϕ.

■

A.2. Analytic and stochastic Landau symbols

Landau's symbols are named after the german number theoretician Edmund Landau who invented the notation. He used O because the rate of growth of a function is also called its order. We first define the analytic landau symbols and some properties. The proofs of the rules are left out. Secondly we show the corresponding stochastic versions. Informations to the proofs can be found in van der Vaart [1998], section 2.2.

Definition A.2.1

Let x_n and u_n be deterministic sequences with $u_n > 0$, then

$$x_n = o(u_n) \quad \Leftrightarrow \quad \frac{x_n}{u_n} \underset{n\to\infty}{\longrightarrow} 0\,,$$

$$x_n = O(u_n) \quad \Leftrightarrow \quad \exists\, n_0, C: \quad |x_n| \le C u_n \qquad \forall n \ge n_0\,.$$

We are going to state some useful rules, the analogues versions we are going to use later.

Lemma A.2.1

Let x_n, y_n, u_n and v_n be deterministic sequences with u_n, v_n positive.

i) $x_n = o(u_n) \quad \Rightarrow \quad x_n = O(u_n)$.

ii) $x_n = O(u_n)$, $y_n = O(v_n) \quad \Rightarrow \quad x_n \pm y_n = O(\max(u_n, v_n))$, $x_n y_n = O(u_n v_n)$.

iii) $x_n = o(u_n)$, $y_n = o(v_n) \quad \Rightarrow \quad x_n \pm y_n = o(\max(u_n, v_n))$.

iv) $x_n = O(u_n)$, $y_n = o(v_n) \quad \Rightarrow \quad x_n y_n = o(u_n v_n)$.

v) $x_n = O(u_n)$, $u_n = O(v_n) \quad \Rightarrow \quad x_n = O(v_n)$.

vi) If one of the Landau symbols on the left hand side in v) is o instead of O we get $x_n = o(v_n)$.

vii) If $x_n \underset{n\to\infty}{\longrightarrow} x$, *then* $x_n = O(1)$.

viii) $x_n = O(1) \quad \Rightarrow \quad \max_{1\le i\le n} x_i = O(1)$.

The proof is omitted.

Now, we define the stochastic versions of the Landau symbols. Here we have for each, asymptotically dominated or bounded, we define the almost sure and the convergence in probability.

Definition A.2.2

Let X_n, U_n be stochastic sequences with $U_n > 0$ almost surely, then

$$
\begin{aligned}
X_n = o_P(U_n) \quad &\Leftrightarrow \quad \frac{X_n}{U_n} \xrightarrow{p} 0 \quad \text{for} \quad n \to \infty\,,\\
X_n = o(U_n)\ a.s. \quad &\Leftrightarrow \quad \frac{X_n}{U_n} \xrightarrow{a.s} 0 \quad \text{for} \quad n \to \infty\,,\\
X_n = O_P(U_n) \quad &\Leftrightarrow \quad \forall\,\epsilon\,\exists\, C : P(|X_n| > CU_n) \le \epsilon \quad \forall\, n\,,\\
X_n = O(U_n)\ a.s. \quad &\Leftrightarrow \quad \exists\, C : |X_n| \le CU_n \qquad a.s. \quad \forall n\,,
\end{aligned}
$$

where C in last line might be random.

Observe, that $O_P(1)$ does not necessarily mean the existence of a limit distribution.

Example A.2.1 *Let $\{X_t\}$ be i.i.d. random variables with finite variance. From the CLT we know the sample mean converges with rate $\sqrt{n}$. Define the divergent rate $r(n) = \sqrt{n} * b(n)$ with $b(n) = 1$ if n odd and $b(n) = 2$ if n even. The sample mean multiplied with $r(n)$ is $O_P(1)$ but does not have a limit distribution.*

Lemma A.2.2

Let X_n, Y_n, U_n and V_n stochastic sequences and U_n as well as V_n are almost surely positive.

i) Assertion i)-viii) in Lemma A.2.1 also hold true replacing the Landau symbols with the a.s. ones.

ii) Replacing the Landau symbols in assertion i)-vii), and in vii) the convergence, in Lemma A.2.1 with the corresponding P ones.

iii) $X_n = O(U_n) \quad a.s. \qquad \Rightarrow \qquad X_n = O_p(1)$
and $X_n = o(U_n) \quad a.s. \qquad \Rightarrow \qquad X_n = o_P(1)$

iv) $X_n = O(1) \quad a.s. \qquad \Rightarrow \qquad \max_{1\le k\le n} X_k = O(1) \qquad a.s.$

v) The assertion viii) is in general not true for replacing O with O_P.

Some of the proofs can be found in van der Vaart [1998] section 2.2.

B. Multivariate heavy-tailed random variables and copulas

We are going to analyse the change-point test and estimator for multivariate heavy-tailed random variables. To this end, we define what we understand as multivariate heavy-tailed random variables and introduce how copulas can be used for the simulation's.

B.1. Copula

In the context of multidimensional random variables copulas are useful. An introduction to copulas can be found in Nelsen [1999] and in the context with applications we refer to Embrechts et al. [1997]. We briefly summarise the necessary results here based on [Kroese et al., 2011, section 3.2.1].

Definition B.1.1

> *A function* $\mathbf{C} : [0,1]^d \to [0,1]$ *is called a copula function if there are dependent uniform random variables* $U_1, \ldots, U_d$ *taking values in* $[0,1]$ *such that* $\mathbf{C}$ *is their joint distribution function,i.e.*
> $$\mathbf{C}(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d)\,.$$

We are interested in simulating a random vector X with a cumulative distribution function F. Sklar's Theorem describes the relation between a copula $\mathbf{C}$ and a cumulative distribution function F. For a given copula $\mathbf{C}$ and marginal distributions F_i, $i = 1, \ldots, d$, there exists a distribution function F such that

$$F(x_1, \ldots, x_d) = \mathbf{C}(F_1(x_1), \ldots, F_d(x_d))\,.$$

On the other hand for a given distribution function F and marginal distributions F_i, $i = 1, \ldots, d$, a copula $\mathbf{C}$ exists if it holds

$$\mathbf{C}(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))\,.$$

Observe, that for continuous $F_1, \ldots, F_d$ the copula $\mathbf{C}$ is unique. This also shows, that with copulas we can change the dependence structure without changing the marginal distributions.

Example B.1.1 *Let* $\mathbf{C}$ *be given as*

$$\mathbf{C}(u_1, \ldots, u_d) = u_1 \cdot \ldots \cdot u_d\,.$$

Then $\mathbf{C}$ *is called independence copula as the entries of* X *will be independent.*

Example B.1.2 *For multivariate normal distributed random variables with covariance matrix $\Sigma = (\sigma_{ij}^2)_{1\le i,j\le d}$, the copula is given as*

$$\mathbf{C}(u_1, \ldots, u_d) = \Phi_\Sigma(\Phi^{-1}_{\sigma^2_{11}}(u_1), \ldots, \Phi^{-1}_{\sigma^2_{dd}}(u_d))\,.$$

This is also the copula for log-normal distributed random vectors. Let X be normal distributed with expectation μ and covariance matrix Σ, then $Y = \exp(X)$ is multivariate log-normally distributed with the parameters μ and Σ. This relation holds also for each entry, that is why we have

$$F_Y(y_1, \ldots, y_d) = F_X(\log(y_1), \ldots, \log(y_d))\,. \tag{B.1}$$

For $z = F_{Y_i}(y_i) = F_{X_i}(\log(y_i)) = F_{X_i}(x_i)$ we have

$$\begin{aligned} F^{-1}_{Y_i}(z) &= F^{-1}_{Y_i}(F_{Y_i}(y_i)) \\ &= y_i \\ &= \exp(F^{-1}_{X_i}(z))\,. \end{aligned} \tag{B.2}$$

Combining the results (B.1) *and* (B.2) *shows that the copula is the same.*

Example B.1.3 *A commonly used copula is the t-copula given by*

$$\mathbf{C}(u_1, \ldots, u_d) = T_{\nu,\Sigma}(T^{-1}_\nu(u_1), \ldots, T^{-1}_\nu(u_d))\,,$$

where $T_{\nu,\Sigma}$ is the distribution function of a $t_\nu(0, \Sigma)$ distributed random vector with ν degree of freedom, mean vector 0 and covariance matrix Σ.

B.2. Heavy-tailed random vectors

Now we are familiar with simulating dependent random vectors. In the univariate case, heavy tailed distributions with finite second moments exist and especially in finance they are of interest. An introduction to the topic of heavy-tailed random vectors in the context of insurance and finance is given in Embrechts et al. [1997]. We focus only on the relevant definitions and relations. Based on Weng and Zhang [2012] we introduce the following definitions.

Definition B.2.1

Let $\overline{F} := 1 - F$ denote the survival function of an univariate random variable. It belongs to the class $\mathcal{L}(\alpha)$ (called long-tailed class) for some $\alpha \ge 0$ if and only if

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = e^{\alpha y} \qquad \text{for } y > 0\,. \tag{B.3}$$

For the multivariate case we need the following notations.

N.14 The Borel σ-field $\mathcal{B}_d$ on $\mathbb{E}_d := [0,\infty)^d \setminus 0$ (here 0 is the d-dimensional zero vector)

N.15 The space $M_+(\mathbb{E}_d)$ of non-negative Radon measures on $\mathbb{E}_d$ and $\xrightarrow{v}$ stands for vague convergence on this space.

Definition B.2.2

A random vector X in $[0,\infty)^d$ belongs to the multivariate long-tailed class $\mathcal{L}(\nu, b)$ if and only if there exist a non-null Radon measure ν on $\mathcal{B}_d$ and a function vector $b(u) := (b_1(u), \ldots, b_d(u))$ satisfying $b_i(u) \to \infty$ as $u \to \infty$ for $i = 1, \ldots, d$ such that

$$uP(X - b(u) \in \cdot) \xrightarrow{v} \nu(\cdot)$$

in $M_+(\mathbb{E}_d)$.

It is quite technical to check this condition. With copula however we are able to give conditions on the copula and the marginal distributions such that the random vector is long-tailed.

Definition B.2.3

A function $\hat{\mathbf{C}} : [0,1]^d \to [0,1]$ is called the survival copula induced by the copula $\mathbf{C}$ if it is given as

$$\hat{\mathbf{C}}(u_1, \ldots, u_d) = u_1 + \cdots + u_d - 1 + \mathbf{C}(1 - u_1, \ldots, 1 - u_d)\,.$$

It follows that for the joint distribution F and the marginals F_i, $i = 1, \ldots, d$, the survival copula is given as

$$\hat{\mathbf{C}}(u_1, \ldots, u_d) = \overline{F}(\overline{F}_1^{-1}(u_1), \ldots, \overline{F}_d^{-1}(u_1))\,,$$

where $\overline{F}$ is the survival function of F and $\overline{F}_i^{-1}$ is the inverse of the survival functions of F_i, $i = 1, \ldots, d$. The following assumptions have to be made on the marginal distributions of the random vector X.

L.1 The marginal distributions of X belong to the class of long-tailed distributions.

L.2 The marginal distributions of X have equivalent tails, i.e.

$$\overline{F}_i = O(\overline{F}_1) \qquad \forall i = 2, \ldots, d\,.$$

Then we get the following result bringing together the long-tail property of the marginal distributions and properties of the survival copula to ensure that the random vector X is long-tail dependent.

Theorem B.2.1 (Weng and Zhang [2012] Theorem 4.1)

*Let X be a random vector with survival Copula $\hat{\mathbf{C}}$ and univariate marginal distributions F_i, $i = 1, \ldots, d$ satisfying assumptions **L.1** and **L.2**. If additionally the joint distribution of X is continuous and the lower tail dependence functions of $\hat{\mathbf{C}}$ exist, i.e. for $k = 1, \ldots, d$ there exists $\lambda_k(u_1, \ldots, u_k)$ with*

$$\lambda_k(u_1, \ldots, u_k) := \lim_{s \to 0^+} \frac{\hat{\mathbf{C}}(su_1, \ldots, su_d)}{s}\,, \tag{B.4}$$

then X belongs to the multivariate long-tailed class.

For the given copulas in section B.1 we show, that the survival Copula fulfils the conditions.

Example B.2.1 *Let us consider the Gaussian copula from example B.1.2. Observe, that it holds*

$$\lim_{s\to 0^+} \frac{\mathbf{C}(1-su_1,\ldots,1-su_d)}{s} = \lim_{s\to 0^+} \frac{\Phi(\Phi^{-1}_{\sigma^2_{11}}(1-su_1)),\ldots,\Phi^{-1}_{\sigma^2_{dd}}(1-su_d)}{s}\,.$$

From example 3.4 in Embrechts et al. [2003] we know

$$\begin{aligned}\lambda_k(u_1,\ldots,u_k) :&= \lim_{s\to 0^+} \frac{\hat{\mathbf{C}}(su_1,\ldots,su_d)}{s}\\ &= 0\,.\end{aligned}$$

Example B.2.2 *The lower tail dependence function for the* t*-copula was analysed and determined in Embrechts et al. [2003] (section 5.3). Thus this copula also fulfils the assumptions for the long-tail dependence.*

For further informations we advise the reader to study Embrechts et al. [2003].

C. Limittheory

C.1. Used basic limit theorems

Theorem C.1.1 (Ranga Rao [1962] Theorem 6.5 (cited from Kirch and Tadjuidje Kamgaing [2012] Theorem 6))

> *Let $\|\cdot\|$ be any norm on $\mathbb{R}^d$ and $\{v_t(\theta)\}$ be a stationary ergodic random sequence with values in $C(\Theta, \mathbb{R}^d)$ satisfying*
>
> $$\mathbb{E}[\sup_{\theta \in \Theta} \|v_1(\theta)\|] < \infty ,$$
>
> *then*
>
> $$\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{t=1}^{n} v_t(\theta) - \mathbb{E}[v_1(\theta)] \right\| \underset{n \to \infty}{\longrightarrow} 0 \qquad a.s.$$

Proof:
The proof follows from Theorem 6.5 in Ranga Rao [1962]. Observe that the Theorem 6.5 is based on Theorem 6.4. The g within Theorem 6.4 is here $\sup_{\theta \in \Theta} \|v_1(\theta)\|$. ■

Theorem C.1.2 (Pötscher and Prucha [1997] Lemma 3.1)

> *Let $f_n : \Omega \times \Theta \to \mathbb{R}$ and $\overline{f}_n : \Theta \to \mathbb{R}$ ($\Theta \subset \mathbb{R}^p$) be two sequences of functions such that a.s. (or in probability)*
>
> $$\sup_{\theta \in \Theta} |f_n(\omega, \theta) - \overline{f}_n(\theta)| \underset{n \to \infty}{\longrightarrow} 0 .$$
>
> *Let $\overline{\theta}_n$ be an identifiably unique sequence of minimizers of $\overline{f}_n(\theta)$, then for any sequence $\hat{\theta}_n$ such that eventually*
>
> $$f_n(\omega, \hat{\theta}_n) = \inf_{\theta \in \Theta} \overline{f}_n(\theta)$$
>
> *holds, we have $\left\| \hat{\theta}_n - \overline{\theta}_n \right\|$ a.s. (or in probability).*

Proof:
Just the Lemma 3.1 with a specific $\rho_B(\cdot) = \|\cdot\|$. ■

The next Theorems are stated without proofs as the references are given and no difference was made (except notation).

Theorem C.1.3 (Dominated Convergence Theorem, Theorem 16.4 in Billingsley [1995], page 209)

> *Let $\{X_n\}$ be a sequence of integrable random variables and let the limit $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ exist for all $\omega \in \Omega$. If there is a non-negative integrable random variable Y such that $|X_n(\omega)| \leq Y(\omega)$ for all $\omega \in \Omega$ and all n, then X is integrable and $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.*

Corollary C.1.1 *Let $f(X,\theta)$ be an integrable random variable for $\theta \in \Theta \subset \mathbb{R}^p$ and let $f(x,\theta)$ be differentiable in $\theta \in \Theta$ and for all x. If there exists an integrable random variable $M(X)$ such that for all x*

$$\left|\frac{\partial}{\partial\theta} f(x,\theta)\right| \leq M(x)\,,$$

then $\mathbb{E}[f(X,\theta)]$ has derivative given by

$$\frac{\partial}{\partial\theta}\mathbb{E}[f(X,\theta)] = \mathbb{E}\left[\frac{\partial}{\partial\theta} f(X,\theta)\right]\,.$$

on Θ.

Proof:
Follows directly from the Dominated Convergence Theorem. A detailed proof can be found in Billingsley [1995] proof of Theorem 16.8 on page 212. ∎

Theorem C.1.4 (Slutsky from Lemma 2.8 in van der Vaart [1998] page 11)

> *$\{X_n\}$ and $\{Y_n\}$ sequences of random variables such that $X_n \xrightarrow{d} X$ (random variable) and $Y_n \xrightarrow{p} y$ (constant) then*
>
> 1. *$X_n * Y_n \xrightarrow{d} y * X$*
> 2. *$X_n + Y_n \xrightarrow{d} X + y$*
> 3. *$Y_n^{-1} X_n \xrightarrow{d} y^{-1} X$ if $y \neq 0$.*

Theorem C.1.5 (Cramer-Wold device from Theorem 29.4 in Billingsley [1995], page 383)

> *For random vectors $X_d = (X_{n1}, \ldots, X_{nd})^\intercal$ and $X = (X_1, \ldots, X_d)^\intercal$, a necessary and sufficient condition for $X_n \xrightarrow{d} X$ is that $a^\intercal X_n \xrightarrow{d} a^\intercal X$ for each vector $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$.*

Theorem C.1.6 (Continuous mapping theorem from Theorem 2.3 in van der Vaart [1998], page 7)

> *Let $\{X_n\}$, X be random vectors in $\mathbb{R}^d$ and $X_n \xrightarrow{d} X$. For every continuous function $f : \mathbb{R}^d \to \mathbb{R}^r$, it holds*
>
> $$f(X_n) \xrightarrow{d} f(X)\,.$$

Let $B(t)$ denote a standard Brownian bridge. We give useful results for asymptotic behaviour of this process. A useful result by Shao et al. [1991] gives necessary and sufficient conditions for the behaviour of functions such that the supreme of the weighted Brownian bridge still exists or is zero if we tend near to the boundary.

Theorem C.1.7 (Csörgő and Horváth [1997], page 373 Theorem A.5.1)

Let $q(t)$ be a positive function on $(0,1)$ and non-decreasing in a neighbourhood of zero and non-increasing in a neighbourhood of one. Denote by $I_{0,1}(q,c)$ the integral

$$\int_0^1 \frac{1}{t(1-t)} \exp\left(-c\frac{q^2(t)}{t(1-t)}\right) \partial t \,.$$

Then,

1. *$I_{0,1}(q,c) < \infty$ for some $c > 0$ if and only if*

$$\limsup_{t\searrow 0} |B(t)|/q(t) < \infty \quad a.s. \qquad and \qquad \limsup_{t\nearrow 1} |B(t)|/q(t) < \infty \quad a.s.$$

2. *$I_{0,1}(q,c) < \infty$ for all $c > 0$ if and only if*

$$\limsup_{t\searrow 0} |B(t)|/q(t) = 0 \quad a.s. \qquad and \qquad \limsup_{t\nearrow 1} |B(t)|/q(t) = 0 \quad a.s.$$

C.2. Strong mixing time series

In the Change-point theory we are going to analyse functions over discrete observations X_i. We want to get asymptotic results. As we have seen in section 2 we are going to use some general results for the underlying process. For the rest of the thesis the observed process is assumed to be strong mixing. We are going to repeat needed results and show a Hájek-Rényi-type inequality.

Let $\{Z_t\}$ be a strictly stationary (in the following we only say stationary), α-mixing time series with rate of polynomial order. First we give the definition of α-mixing of polynomial order.

Definition C.2.1

A stationary stochastic process $\{Z_t\}$ is called α- or strong mixing with rate $\alpha(\cdot)$, if

$$\alpha(j) = sup_{A\in\mathcal{F}^0_{-\infty}(Z),\ B\in\mathcal{F}^\infty_j(Z)} |P(A\cap B) - P(A)P(B)| \to 0 \ as\ j\to\infty\,,$$

where $\mathcal{F}^0_{-\infty}(Z)$ is the σ-algebra generated by $Z_0,\ Z_{-1},\ \ldots$ and $\mathcal{F}^\infty_j(Z)$ is the σ-algebra generated by $Z_j,\ Z_{j+1},\ \ldots$. It is called α-mixing with rate of polynomial order if $a(j) = O(j^{-c})$ for some $1 < c < \infty$.

For the next steps we need some basic properties of α-mixing time series, which we state here.

Lemma C.2.1

For the time series $\{Z_t\}$ the following hold.

1. *$\{Z_t\}$ is ergodic.*
2. *For every measurable function $f : \mathbb{R}^M \to \mathbb{R}$ ($M > 0$) the time series $f(Z_t, \ldots, Z_{t+M-1})$ is α-mixing with $\alpha_f(j) = \alpha_Z(j - M + 1)$ for $j \geq M$.*
3. *$\{Z_{-t}\}$ is α-mixing time series with the same mixing rate.*

These are basic results. For more informations we refer to [Bradley, 2007, volume 1, Remark 1.8, Theorem 2.14].

The first result needed is the strong law of large numbers.

Theorem C.2.1 *Assuming $g : \mathbb{R}^p \to \mathbb{R}$ is a measurable function. Then for Z_t, we get*

$$\frac{1}{l}\sum_{t=1}^{l} g(\mathbb{Z}_t) - E\left(g(\mathbb{Z}_1)\right) = o(1) \ a.s., \qquad (l \to \infty) \tag{C.1}$$

where $\mathbb{Z}_t = (Z_{t-1}, \ldots, Z_{t-p})$.

Proof:
Theorem 6.28 and Proposition 6.31 in Breiman [1992], together, give that (C.1) follows from stationarity and ergodicity of the processes and from g being measurable.

■

Observe, the strong law of large numbers holds true for multivariate strong mixing processes.

Theorem C.2.2 *Let $\{Z_t\}$ be a stationary sequence of random vectors in $\mathbb{R}^d$, centered and having $(2 + \phi)$th moments with $\phi > 0$. Suppose that $\{Z_t\}$ is α-mixing of polynomial order $c > \frac{2+\phi}{\phi}$. Then the two series in*

$$\gamma_{ij} = \mathbb{E}[Z_{1i}Z_{1j}] + 2\sum_{k\geq 2}^{\infty} \mathbb{E}[Z_{1i}Z_{kj}] + \sum_{k\geq 2} \mathbb{E}[Z_{ki}Z_{1j}]$$

converges absolutely. Let Γ denote the matrix of γ_{ij}, $1 \leq i,j \leq d$. Then we can redefine the sequence $\{Z_t\}$ on a new probability space together with a Wiener process $\{W_s\}$ having covariance matrix Γ such that

$$\sum_{t=1}^{k} Z_t - W_k = O\left(k^{\frac{1}{2}-\lambda}\right) \quad a.s.. \qquad (k \to \infty)$$

with $0 < \lambda < \frac{1}{2}$ depending on ϕ, c and d.

Proof:
The proof can be found in Berkes and Philipp [1979] remark 4.4.4 .

■

Proposition C.2.1

Let $\{Z_t\}$ fulfill the conditions of the invariance principle (Theorem C.2.2). Then

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(Z_t - \mathbb{E}Z_t) \xrightarrow{d} \mathcal{N}(0,\tau)\,.$$

Proof:
See Kuelbs and Philipp [1980] Proposition 2.1. or Pötscher and Prucha [1997] S. 102. ■

Remark C.2.1

Under the conditions of Theorem C.2.2 one gets

$$\max_{1\le k\le n}|\sum_{t=1}^{k} Z_t| = O_p(\sqrt{n})\,.$$

Proof:
From Theorem C.2.2 we know that also $\sum_{t=1}^{s}\epsilon_t$, where ϵ_t are i.i.d standard normal random variables, converges to a Wiener process $\{W_s\}$ and $\tau = 1$. Then it follows

$$\begin{aligned}\max_{1\le k\le n}|\sum_{t=1}^{k} Z_t| &\le \max_{1\le k\le n} k^{\frac{1}{2}-\nu}\max_{1\le k\le n}\frac{1}{k^{\frac{1}{2}-\nu}}|\sum_{t=1}^{k} Z_t - \tau\sum_{t=1}^{k}\epsilon_t| + \tau\max_{1\le k\le n}|\sum_{t=1}^{k}\epsilon_t| \\ &= o(n^{\frac{1}{2}})O_{a.s.}(1) + O_p(n^{\frac{1}{2}})\,,\end{aligned}$$

where the last line holds because of the Hajek-Renyi inequality for martingal differences, Lemma C.2.2, applied to i.i.d. random variables and properties of the stochastic Landau-Symbols*.

■

Lemma C.2.2

Let $\{\epsilon_t\}$ be sequence of martingal differences with $E|\epsilon_t^2| < \infty$ and $Y_s := \sum_{t=1}^{s}\epsilon_t$. Then for all $c > 0$ and non increasing positive sequence b_s

$$P(\max_{1\le s\le n} b_s|Y_s| \ge c) \le c^{-2}\sum_{s=1}^{n} b_s^2\,\mathbb{V}\mathrm{ar}(\epsilon_k)$$

Proof:
See Gänssler and Stute [1977], p.230. ■

As we can see the Hajek-Renyi inequality is quite useful. By Theorem 1 from Yokoyama [1980] we get such a result for stationary α-mixing sequences with polynomial mixing order. For geometric ergodic time series this follows by the exponential rate. Under the weaker assumption of mixing rate with polynomial order we have to restrict the order w.r.t. the existing moments.

*From $o(\,)$ $a.s.$ follows $O(\,)$ $a.s.$. Moreover if $X_n = O(1)$ $a.s.$ this implies $\max(X_1,\ldots,X_n) = O(1)$ $a.s.$.

Lemma C.2.3

Let $\{Z_t\}$ be α-mixing with $\alpha(j) = o(j^{-c})$ for some $c > 1$. If there exists $\phi > 0$ and $D > 0$ such that

$$\mathbb{E}|Z_t|^{2+\phi} \leq D \quad \forall t \in \mathbb{Z}\,,$$

and

$$\alpha(j) = o(j^{-c}) \text{ and } \phi > \frac{2}{c-1}, \tag{C.2}$$

then there exists $\delta \in (0, 1]$ and $\Delta > 0$ with $\phi = \delta + \Delta$ and

$$\mathbb{E}\left|\sum_{t=1}^{n} Z_t\right|^{2+\delta} \leq \Gamma(D, \alpha, \delta, \Delta) n^{\frac{(2+\delta)}{2}}\,,$$

where Γ is a function depending on D, α, δ and Δ.

Proof:
The proof follows by Theorem 1 of Yokoyama [1980] showing that under the conditions above (3.1) in that Theorem is fulfilled. Let $\phi = \delta + \Delta$, then (3.1) in Theorem 1 of Yokoyama [1980] becomes

$$\sum_{k=0}^{\infty} (k+1)^{\frac{\delta}{2}} \left((k)^{-c}\right)^{\frac{\Delta}{2+\delta+\Delta}} = \sum_{k=0}^{\infty} (k+1)^{\frac{\delta}{2}} k^{-c\frac{\Delta}{2+\delta+\Delta}} < K < \infty\,.$$

It is enough to show[†]

$$\frac{\delta}{2} - c\frac{\Delta}{2+\delta+\Delta} < -1\,,$$

which is equivalent to

$$\frac{(2+\delta)^2}{2c-(2+\delta)} < \Delta = \phi - \delta\,.$$

The existence of such a δ follows because $\frac{(2+\delta)^2}{2c-(2+\delta)} + \delta$ is continuous increasing in δ and for $\delta \to 0$ we derive $\frac{2}{c-1} < \phi$ (as required). Therefore, we can find for all $\phi = \epsilon + \frac{2}{c-1}$ a δ such that $\phi - \left(\frac{(2+\delta)^2}{2c-(2+\delta)} + \delta\right) \geq \frac{\epsilon}{2}$. ■

Lemma C.2.4

Let $\{Z_t\}$ be stationary, α-mixing process of polynomial order $c > 1$ and assume the $(2 + \phi)$-moments exist, with $\phi > \frac{2}{c-1}$. Then exist $\delta \in (0, 1]$ and $\Delta > 0$, such that $\phi = \delta + \Delta$. Moreover, for any $n \geq 1$, $1 \leq m \leq n$ and any positive decreasing sequence b_k, we have

a) $\mathbb{E}\left(\max_{k=1,\ldots,n} b_k|S_k|\right)^{2+\delta} \leq CA(\delta) \sum_{k=1}^{n} b_k^{2+\delta} k^{\frac{2+\delta}{2}}$

b) $\max_{m\leq k\leq n} b_k|S_k| = O_p\left(\left[\sum_{k=m}^{n} b_k^{2+\delta} k^{\frac{2+\delta}{2}}\right]^{\frac{1}{2+\delta}}\right)$,

[†]Observe that for $k > 1$ it holds $\frac{(\frac{1}{2}k)^a}{(2k)^b} \leq \frac{(k+1)^a}{k^b} \leq \frac{(2k)^a}{(\frac{1}{2}k)^b}$. This implies that $\sum_{k=0}^{\infty} \frac{(k+1)^a}{k^b}$ is converging iff $a - b < -1$.

where $S_k = \sum_{t=1}^{k} Z_t$, *C is a constant and* $A(\delta)$ *is some constant depending on* δ, *which is as in Lemma C.2.3.*

Proof:
The result a) follows by Theorem B.3 in Kirch [2006].
For b) first note that for $k > m$ it holds $b_k|S_k| \leq b_k(|S_m| + |S_{(m+1),k}|) \leq b_m|S_m| + b_k|S_{(m+1),k}|$, where $S_{(m+1),k} = \sum_{t=m+1}^{k} Z_t$. Therefore, we get

$$\begin{aligned} P(\max_{m\leq k\leq n} b_k|S_k| > \sigma) &\leq P\left(b_m|S_m| > \frac{\sigma}{2}\right) + P\left(\max_{m+1\leq k\leq n} b_k|S_{(m+1),k}| > \frac{\sigma}{2}\right) \\ &\leq \left(\frac{2b_m}{\sigma}\right)^{2+\delta} \mathbb{E}|S_m|^{2+\delta} + \left(\frac{2}{\sigma}\right)^{2+\delta} \mathbb{E}\left(\max_{1\leq l\leq n-m} b_{l+m}|S_l'|\right)^{2+\delta} \\ &\leq \left(\frac{2b_m}{\sigma}\right)^{2+\delta} O\left(m^{\frac{2+\delta}{2}}\right) + \left(\frac{2}{\sigma}\right)^{2+\delta} \mathbb{E}\left(\max_{1\leq l\leq n-m} b_{l+m}|S_l'|\right)^{2+\delta} \end{aligned}$$

with $S_l' = \sum_{t=1}^{l} Z_{t+m}$. The second term in the last line can be estimated by the result from a).

■

The result derived in b) is called the Hájek-Rényi-Typ inequality for stationary α-mixing time series.

As we also need the LIL (Law of iterated logarithm) we just want to state that due to the invariance principle this also holds true (see Berkes and Philipp [1979] or for the one-dimensional case Oodaira and Yoshihara [1971]), i.e. for some Λ, $0 \leq \Lambda < \infty$

$$P\left(\sup_n \frac{\|\sum_{t=1}^{n} X_t\|}{a_n} < \infty\right) = P\left(\overline{\lim}_n \frac{\|\sum_{t=1}^{n} X_t\|}{a_n} = \Lambda\right) = 1\,. \tag{C.3}$$

with $a_n := (n \log(\log(n)))^{\frac{1}{2}}$.

D. Tables - Critical value

We give tables for the critical value of $\sup_{\eta<s<(1-\eta)} \frac{\|B(s)\|}{(s(1-s))^{\gamma}}$, where B is a d-dimensional Brownian bridge.

		d		
γ	η	1	2	3
0	0	1.212256	1.443812	1.609381
0.5	0.01	2.99186	3.45851	3.79555
0.4901	0	2.995043	3.44158	3.764023
0.5	0.05	2.85296	3.32581	3.66867
0.4525	0	2.673823	3.088289	3.389556
0.5	0.1	2.75516	3.23421	3.5791
0.41	0	2.417517	2.810113	3.096722

Table D.1.: Critical value of weighted norm of a d-dimensional Brownian bridge, with $\alpha = 0.1$

		d		
γ	η	1	2	3
0	0	1.347522	1.574039	1.736181
0.5	0.01	3.24195	3.70026	4.02713
0.4901	0	3.231012	3.662986	3.978833
0.5	0.05	3.1168	3.57791	3.9108
0.4525	0	2.897949	3.303564	3.602494
0.5	0.1	3.02738	3.49251	3.8292
0.41	0	2.632796	3.022263	3.301614

Table D.2.: Critical value of weighted norm of a d-dimensional Brownian bridge, with $\alpha = 0.05$

		d		
γ	η	1	2	3
0	0	1.468787	1.689353	1.851499
0.5	0.01	3.47012	3.91041	4.23665
0.4901	0	3.440831	3.865758	4.176037
0.5	0.05	3.3493	3.80231	4.12883
0.4525	0	3.102743	3.501203	3.793509
0.5	0.1	3.26821	3.72102	4.05426
0.41	0	2.830868	3.211655	3.491204

Table D.3.: Critical value of weighted norm of a d-dimensional Brownian bridge, with $\alpha = 0.025$

		d		
γ	η	1	2	3
0	0	1.615893	1.830308	1.992486
0.5	0.01	3.74589	4.17324	4.48813
0.4901	0	3.702681	4.113801	4.416754
0.5	0.05	3.63846	4.07303	4.39301
0.4525	0	3.351194	3.741036	4.027234
0.5	0.1	3.55567	4.003	4.31947
0.41	0	3.075082	3.441805	3.719283

Table D.4.: Critical value of weighted norm of a d-dimensional Brownian bridge, with $\alpha = 0.01$

Bibliography

Ulrich Anders. *Statistische neuronale Netze*. Vahlen GmbH, München, Germany, 1997. ISBN 3-8006-2184-3.

Jaromir Antoch, Marie Hušková, and Noël Veraverbeke. Change-point problem and bootstrap. *Journal of Nonparametric Statistics*, 5(2):123–144, 1995.

Jushan Bai. Estimation of a change point in multiple regression models. *The Review of Economics and Statistics*, (4):551–563, 1997.

István Berkes and Walter Philipp. Approximation theorems for independent and weakly dependent random vectors. *The Annals of Probability*, 7(1):29–54, 1979.

Matthias Bigler, Regine Röthlisberger, Fabrice Lambert, Eric W. Wolff, Emiliano Castellano, Roberto Udisti, Thomas F. Stocker, and Hubertus Fischer. Atmospheric decadal variability from high-resolution Dome C ice core records of aerosol constituents beyond the Last Interglacial. *Quaternary Science Reviews*, 29:324–337, 2010.

Patrick Billingsley. *Probability and measure*. John Wiley & Sons, Inc, 605 Third-Avenue, New York, NY, 1995. ISBN 0-471-00710-2.

Richard C. Bradley. *Introduction to strong mixing conditions*. Kendrick Press, Heber City, Utah, USA, 2007. ISBN 0-9740427-9-X.

Leo Breiman. *Probability*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992. ISBN 0-89871-296-3.

Gabriela Ciupera. Two tests for sequential detection of a change-point in a nonlinear model. *Journal of Statistical Planning and Inference*, 143:1719–1743, 2013.

Miklós Csörgő and Lajos Horváth. *Weighted Approximations in Probability and Statistics*. John Wiley & Sons Ldt, Baffins Lane, Chichester, west Sussex PO19 IUD, England, 1993. ISBN 0-471-93635-9.

Miklós Csörgő and Lajos Horváth. *Limit Theorems in Change-point Analysis*. John Wiley & Sons Ldt, Baffins Lane, Chichester, west Sussex PO19 IUD, England, 1997. ISBN 0-471-95522-1.

Donald A. Darling and Paul Erdös. A limit theorem for the maximum of normalized sums of independent random variables. *Duke Mathematical Journal*, 23(1):143–155, 1956.

Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events for insurance and finance*. Applications of mathematics. Springer, Berlin, Heidelberg, New York, 1997. ISBN 978-3-540-60931-5.

Paul Embrechts, Filip Lindskog, and Alexander Mcneil. Chapter 8 - modelling dependence with copulas and applications to risk management. In Svetlozar T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 329 – 384. North-Holland, Amsterdam, 2003.

Jürgen Franke and Michael H. Neumann. Bootstrapping neural networks. *Neural Computation*, 12(8):1929–49, 1998.

Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014.

Wayne A. Fuller. *Inroduction to Statistical Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, 1996. ISBN 978-0-471-55239-0.

Peter Gänssler and Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer, 1977. ISBN 9783540084181.

Vidyadhar P. Godambe. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31:1208–1211, 1960.

Edit Gombay. Change detection in linear regression with time series errors. *The Canadian Journal of Statistics*, 38(1):65–79, 2010.

Edit Gombay, Lajos Horváth, and Marie Hušková. Estiamtors and tests for change in variances. *Statistics & Decisions*, 14, 1996.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359 – 366, 1989.

Marie Hušková. Tests and estimators for the change point problem based on M-statistics. *Statistics & Risk Modeling*, 14(2):115–136, 1996.

Marie Hušková and Miriam Marušiaková. M-procedures for detection of changes for dependent observations. *Communication in Statistics- Simulation and Computation*, 41:1032–1050, 2012.

Marie Hušková, Zuzana Prášková, and Josef Steinebach. On the detection of changes in autoregressive time series i. asymptotics. *Journal of Statistical Planning and Inference*, 137:1243–1259, 2007.

J. T. Gene Hwang and A. Adam Ding. Prediction intervals for artificial neural networks. *Journal of American Statistical Association*, 92(438):748–757, 1997.

Barry James, Kang L. James, and David Siegmund. Tests for a change-point. *Biometrika*, 74 (1):71–83, 1987.

Claudia Kirch. *Resampling Methods for the Change Analysis of Dependent Data*. PhD thesis, Universität zu Köln, Juni 2006.

Claudia Kirch and Joseph Tadjuidje Kamgaing. A uniform central limit theorem for neural network-based autoregressive processes with applications to change-point analysis. *Statistics*, 48(6):1187–1201, 2014.

Claudia Kirch and Joseph Tadjuidje Kamgaing. Testing for parameter stability in nonlinear autoregressive models. *Journal of Time Series Analysis*, 33(3):365–385, 2012.

Claudia Kirch and Joseph Tadjuidje Kamgaing. On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference*, 161, 2014.

Claudia Kirch, Birta Muhsal, and Hernando Ombao. Detection of changes in multivariate time series with application to EEG data. *Journal of American Statistical Association*, 110 (511):1197–1216, 2015.

Dirk P. Kroese, Thomas Taimre, and Zdravko I. Botev. *Handbook of Monte Carlo methods.* Wiley series in probability and statistics. Hoboken, N.J. Wiley, 2011. ISBN 978-0-470-17793-8.

James D. Kuelbs and Walter Philipp. Almost sure invariance principles for partial sums of mixing b-valued random variables. *The Annals of Probability*, 8(6):1003–1036, 1980.

Roger B. Nelsen. *An introduction to copulas.* Lecture notes in statistics. Springer, New York, Berlin, Heidelberg, 1999. ISBN 0-387-98623-5.

Hiroshi Oodaira and Ken-ichi Yoshihara. The law of the iterated logarithm for stationary processes satisfying mixing conditions. *Kodai Mathematical Seminar Reports*, 23(3):311–334, 1971.

E. S. Page. On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44(1/2):248–252, 1957.

Benedikt M. Pötscher and Ingmar R. Prucha. *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Springer, 1997. ISBN 9783540628576.

Zuzana Prášková and Ondřej Chochola. M-procedures for detection of a change under weak dependence. *Journal of Statistical Planning and Inference*, 149:60 – 76, 2014.

Ramaswamy Ranga Rao. Relation between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962.

Ashish Sen and Muni S. Srivaastava. On tests for detecting change in Mean. *The Annals of Statistics*, 3(1):98–108, 1975.

Qi-Man Shao, Miklós Csörgő, and Barbara Szyszokowicz. A note on local and global functions of a wiener process and some renyi-type statistics. *Studia Scientiarum Mathematicarum Hungaric*, 26:239–259, 1991.

Michael Sørensen. On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics*, 13(2):111–136, 1999.

Cătălin Stărică and Clive Granger. Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87(3):503–522, 2005.

Aad W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 1998. ISBN 0521496039.

Chengguo Weng and Yi Zhang. Characterization of mulivariate heavy-tailed distribution families via copula. *Journal of Multivariate Analysis*, 106:176–186, 2012.

Ryozo Yokoyama. Moment bounds for stationary mixing sequences. *Probability Theory and Related Fields*, 52(1):45–57, 1980.

Jie Zhou and San Y. Liu. Inference for mean change-point in infinite variance AR(p) process. *Statistics and Probability Letters*, 79:6–15, 2008.

Curriculum Vitae – education

Since 06/2011	Ph.D. student, Technische Universität Kaiserslautern, Kaiserslautern.
04/2006 - 08/2009	Diploma in mathematics, Humboldt Universität zu Berlin, Berlin.
10/2002 - 03/2006	Study mathematics, Brandenburgische Technische Universität, Cottbus.
09/1997 - 06/2002	Abitur, Fürst-Pückler-Gymnasium, Cottbus.

Curriculum Vitae – Ausbildung

Seit 06/2011	Ph.D. Studentin, Technische Universität Kaiserslautern, Kaiserslautern.
04/2006 - 08/2009	Diplom in Mathematik, Humboldt Universität zu Berlin, Berlin.
10/2002 - 03/2006	Studium der Mathematik, Brandenburgische Technische Universität, Cottbus.
09/1997 - 06/2002	Abitur, Fürst-Pückler-Gymnasium, Cottbus.