Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS

Barbora Peštová

Testing Structural Changes Using Ratio Type Statistics

Department of Probability and Mathematical Statistics

Supervised by
Prof. RNDr. Marie Hušková, DrSc.

Study program
Mathematics

Study branch
Probability and Mathematical Statistics

Dedicated to my mother, grandmother,

to my little boys and husband

Acknowledgements

First of all, I would like to express my deepest thanks to my supervisor Prof. RNDr. Marie Hušková, DrSc. for her patient guidance throughout my studies. Without her endless support and encouragement this thesis could not have been completed. Her advice and suggestions have directed me in my work whenever I have needed it. Providing me such an interesting topic and her broad range of knowledge helped me to advance in statistical thinking.

I would also like to thank to all my teachers from Charles University in Prague and from Free University in Amsterdam for guiding me through the world of probability theory and mathematical statistics.

Then I would like to thank to all of my friends and family, for their support during my studies and for many joyful moments that gave me strength and energy.

Last but not least, my thanks belong to Michal—not only for his encouragement, patience and love, but also for many inspiring discussions that have helped me to clear my views and to see particular topics in broader context. I am also grateful for his advise on LaTeX, R, English, and for pointing my attention to several interesting references.

This work was supported by the grants GAUK B-MAT 116209, GAČR 15-09663S, and GAČR P201/12/1277.

Statement of Honesty

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

Prague, May 15, 2015

Barbora Peštová

Signature

Annotations

Title

Testing Structural Changes Using Ratio Type Statistics

Author

Mgr. Barbora Peštová, M.Sc.

Department

Department of Probability and Mathematical Statistics
Faculty of Mathematics and Physics
Charles University in Prague
Czech Republic

Supervisor

Prof. RNDr. Marie Hušková, DrSc.

Mailing Address

KPMS MFF UK
Sokolovská 83
186 75 Prague 8
Czech Republic

Anotace

Název

Testování strukturálních změn pomocí statistik podílového typu

Autor

Mgr. Barbora Peštová, M.Sc.

Katedra

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta
Univerzita Karlova v Praze
Česká republika

Vedoucí

Prof. RNDr. Marie Hušková, DrSc.

Adresa

KPMS MFF UK
Sokolovská 83
186 75 Praha 8
Česká republika

Abstract

We deal with sequences of observations that are naturally ordered in time and assume various underlying stochastic models. These models are parametric and some of the parameters are possibly subject to change at some unknown time point. The main goal of this thesis is to test whether such an unknown change has occurred or not. The core of the change point methods presented here is in ratio type statistics based on maxima of cumulative sums.

Firstly, an overview of thesis' starting points is given. Then we focus on methods for detecting a gradual change in mean. Consequently, procedures for detection of an abrupt change in mean are generalized by considering a score function. We explore the possibility of applying the bootstrap methods for obtaining critical values, while disturbances of the change point model are considered as weakly dependent.

Procedures for detection of changes in parameters of linear regression models are shown as well and a permutation version of the test is derived. Then, a related problem of testing a change in autoregression parameter is studied. Finally, our interest lies in panel data of a moderate or relatively large number of panels, while the panels contain a small number of observations. Asymptotic and bootstrap testing procedures to detect a possible common change in means of the panels are established.

All the theoretical results are illustrated through simulations. Several practical applications of the developed procedures are presented on real data as well.

Keywords:
change point, maximum type statistics, ratio type statistics, hypothesis testing, change in mean, abrupt change, gradual change, change in regression, change in autoregression, panel data, asymptotic distribution, robustness, bootstrap, weak dependence, block bootstrap

Abstrakt

Budeme se zabývat posloupnostmi pozorování, která jsou přirozeně uspořádána v čase a současně pro ně uvažujeme různé stochastické modely. Tyto modely jsou parametrické a některé z parametrů mohou podléhat změně v předem neznámém čase. Hlavní cíl této disertace spočívá v testování, zda taková změna nastala nebo ne. Jádrem zde prezentovaných metod detekce okamžiku změny jsou statistiky podílového typu založené na maximech kumulativních součtů.

Nejdřív jsou prezentována východiska disertační práce. Pak se zaměříme na metody detekce postupné změny ve střední hodnotě. Následně zobecníme procedury pro detekci náhlé změny ve střední hodnotě pomocí skórové funkce. Budeme studovat možnosti použití metody bootstrap pro získání kritických hodnot v případě, že náhodné chyby modelu mohou být slabě závislé.

Představíme také procedury pro detekci změny v parametrech lineárního regresního modelu a odvodíme permutační verzi testu. Dále budeme studovat příbuzný problém testování změny v parametru autoregrese. Na závěr se zaměříme na panelová data se středně velkým až velkým počtem panelů, kde panely obsahují malý počet pozorování. Odvodíme asymptotické a bootstrapové procedury pro detekci možné společné změny v panelech.

Všechny teoretické výsledky jsou ilustrovány pomocí simulací. Navržené metody jsou taktéž aplikovány na reálných datech.

Klíčová slova:
okamžik změny, statistika maximálního typu, statistika podílového typu, testování hypotéz, změna ve střední hodnotě, náhlá změna, postupná změna, změna v regresi, změna v autoregresi, panelová data, asymptotické rozdělení, robustnost, bootstrap, slabá závislost, blokový bootstrap

Preface

To know whether a change has happened is a task that is not only interesting, but also desirable for many scientific fields, e.g., in econometrics, biology, or climatology. Our approach to detect the unknown change lies in usage of *ratio type test statistics*. When computing adequately constructed test statistics, it is not necessary to estimate variance of the underlying stochastic model. This may be considered as the most remarkable advantage of ratio type test statistics. Such property makes them a reasonable alternative to classical (non-ratio) statistics—most of all in situations, *when it is difficult to find a suitable variance estimate*.

Our main interest—ratio type test statistics for detecting the unknown change—is introduced in the first chapter of the thesis. A summary of previous results on model of abrupt change in mean is given. An overview of the recent results concerning the change point problem is provided.

One can think of a change in mean in the way that the change from a constant mean is not rapid but rather continuous. Therefore, a *gradual change* in mean is investigated in the second chapter and testing procedures based on the ratio type test statistics are derived.

In the third chapter, the simplest change point model is considered—an *abrupt change* in mean, where at most one change in mean of the observed sequence could happen. Our contribution to this setup lies in considering α-mixing model errors in combination with robust ratio type test statistics. Besides that, a block bootstrap resampling testing procedure is implemented for the abrupt change in mean model.

The fourth chapter deals with detection of a *change in regression* parameters in a trending regression model with independent random errors. Again, a resampling testing procedure is derived and its properties are studied in simulations.

Moreover, tests for a *change in autoregression* based on the ratio type statistics are developed in the fifth chapter as a special case of the regression change point detection procedures.

On the top of that, the sixth chapter concentrates on the change point problem for *panel* *data*. In the considered scenario, it is assumed that there can be at most one *common change* point for all the panels that have fixed length. The number of panels is sufficiently large and this fact is used to obtain asymptotic results. A bootstrap testing technique in this setup is proposed and its validity proved.

Finally, several *simulation studies* and *real data examples* through the whole thesis illustrate the theoretical results presented here. Some of the well-known and frequently used definitions and theorems are recapitulated in the appendix at the end.

Notation

$a.s.$	...	almost surely
$\mathcal{B}$	...	Brownian bridge
$\xrightarrow{a.s.}$	...	convergence almost surely
$\xrightarrow{\mathscr{D}}$	...	convergence in distribution
$\xrightarrow{\mathsf{P}}$	...	convergence in probability P
$\xrightarrow{\mathscr{D}[a,b]}$	...	convergence in the Skorokhod topology on $[a,b]$
O, o	...	deterministic Landau symbols, confer Appendix
E	...	expectation
iid	...	independent and identically distributed
$\mathcal{I}$	...	indicator function
$\mathbb{Z}$	...	integers, i.e., $\{\ldots, -2-1, 0, 1, 2, \ldots\}$
$\mathbb{N}$	...	natural numbers, i.e., $\{1, 2, \ldots\}$
$\mathbb{N}_0$	...	natural numbers with zero, i.e., $\{0, 1, 2, \ldots\}$
P	...	probability
$\mathbb{R}$	...	real numbers
sgn	...	signum function
$\mathscr{D}[a,b]$	...	Skorokhod space on interval $[a,b]$

$\mathcal{W}$	...	standard Wiener process
O_{P} ,o_{P}	...	stochastic Landau symbols, confer Appendix
$[\cdot]$	...	truncated number to zero decimal digits (rounding down the absolute value of the number while maintaining the sign)
Var	...	variance

Contents

Chapter 1

Introduction to Ratio Type Tests In Change Point Analysis

Ratio type statistics studied in this thesis are derived from non-ratio type statistics based on partial sums of residuals. They do not need to be standardized by any variance estimate, which makes them a suitable alternative for non-ratio type statistics, most of all in situations, when it is difficult to find a variance estimate with satisfactory properties. Such difficulty can occur in situations with dependent random errors. Although, variance estimators often do not perform well even in the iid case, especially under alternative.

1.1 Main goals

We aim to utilize the idea of ratio type statistic for testing structural changes in different setups, which include gradual change, abrupt change, robust procedures suitable for heavy-tailed distributions, change in regression parameters, change in autoregression of first order, and change in mean for panel data with fixed panel size. An important part of this task is to derive *asymptotic properties under null hypothesis as well as under alternative.* Limiting distributions can be theoretically derived, but the corresponding critical values are typically not available in an explicit form. Therefore, it is advantageous to use *resampling methods* in order to determine the rejection region. Such methods also need to be justified theoretically by proving asymptotic equivalency of both the original and the resampling statistic. Computer simulations using R software and real data applications give us the idea about performance of the proposed tests.

Our main concern are changes in the mean values of stochastic processes, while the particular stochastic process can have a very general error structure. The only chapter, where the change in mean is not of the main interest, concerns the change of autoregression parameter. In all of the studied models, the random deviations from the mean structure are assumed to possess common unknown variance. Hence, the change in the variance structure is not in the center of our interest. Although similarly constructed ratio type statistics were used to test for a change in variance in Zhao et al. (2010), Zhao et al. (2011) and Chen and Tian (2014).

1.2 General location model

Let us begin with the simplest model of a *single change in location parameter*. For a fixed $n \in \mathbb{N}$, we consider a set of observations $Y_{1,n}, \ldots, Y_{n,n}$ obtained at time ordered points. We are interested in testing the null hypothesis of all observations being random variables with distributions having equal mean values. Our goal is to test against the alternative of the first τ_n observations having distributions with mean value μ_n and the remaining $n - \tau_n$ observations coming from distributions with mean values $\mu_n + \delta_n$, $\delta_n \neq 0$. We suppose that τ_n, μ_n and δ_n are unknown parameters. We can describe the situation as a special case of the *general location model*

$$Y_{k,n} = \mu_{k,n} + \varepsilon_k, \quad k = 1, \ldots, n, \tag{1.1}$$

where $\mu_{1,n}, \ldots, \mu_{n,n}$ are unknown mean values of the original observations $Y_{1,n}, \ldots, Y_{n,n}$ and $\varepsilon_1, \ldots, \varepsilon_n$ are random error terms. The forthcoming results for the performing statistical tests will be of asymptotic character. It means that we consider that the number of observations n is increasing over all limits.

1.3 At most one change in mean

When considering a model with *at most one change in constant mean*, we can further specify the above described model (1.1) as

$$Y_{k,n} = \mu + \delta_n \mathcal{I}\{k > \tau_n\} + \varepsilon_k, \quad k = 1, \ldots, n, \tag{1.2}$$

where μ, δ_n and τ_n are unknown parameters.

Despite the fact, that the observed data $\{Y_{k,n}\}_{k=1,n=1}^{n,\infty}$ form a stochastic *triangular array*, the random disturbances $\{\varepsilon_n\}_{n=1}^{\infty}$ are just a single sequence of random variables. So, the errors remain the same for each row of the triangular array of the observed variables.

Remark 1.1 ($\triangle$-scheme). For the sake of convenience, we suppress the index n in the observations $Y_{k,n}$ as well as in the parameters δ_n and τ_n (and in variables depending on the latter) whenever possible. However, we have to keep in mind that in the asymptotic results below, as $n \to \infty$, both δ_n and τ_n may be changing when n is increasing.

With respect to previous Remark 1.1 about the triangular scheme, model (1.2) can be rewritten as

$$Y_k = \mu + \delta\mathcal{I}\{k > \tau\} + \varepsilon_k, \quad k = 1, \ldots, n.$$

Assumption T1. Random errors are assumed to have zero mean and satisfy the functional central limit theorem (CLT), i.e., they satisfy

$$\exists\, \sigma > 0 \text{ such that } \frac{1}{\sqrt{n}} \sum_{1 \leqslant i \leqslant nt} \varepsilon_i \xrightarrow[n\to\infty]{\mathscr{D}[0,1]} \sigma\mathcal{W}(t),$$

where $\{\mathcal{W}(t), t \in [0,1]\}$ denotes a standard Wiener process and the symbol $\xrightarrow[n\to\infty]{\mathscr{D}[0,1]}$ stands for weak convergence in space $\mathscr{D}[0,1]$.

Remark 1.2. We are using simplified notation for the weak convergence of sequences of stochastic processes every time when it is clear from the context. That is, let us consider a sequence of stochastic processes $\{\mathcal{U}_n(t), t \in [a,b]\}_{n=1}^{\infty}$. We write

$$\mathcal{U}_n(t) \xrightarrow[n\to\infty]{\mathscr{D}[a,b]} \mathcal{U}(t)$$

instead of

$$\{\mathcal{U}_n(t), t \in [a,b]\} \xrightarrow[n\to\infty]{\mathscr{D}[a,b]} \{\mathcal{U}(t), t \in [a,b]\},$$

which means that the sequence of stochastic processes $\{\mathcal{U}_n(t), t \in [a,b]\}_{n=1}^{\infty}$ converges in the Skorokhod topology (Billingsley, 1968) to a stochastic process $\{\mathcal{U}(t), t \in [a,b]\}$. We also say, that the sequence of processes weakly converges to the process.

Assumption T1 is satisfied, e.g., for iid errors having finite $(2+\Delta)$-th moment for some $\Delta > 0$. In case of α-mixing random errors, the functional CLT holds, when the assumptions of Theorem 1 by Doukhan (1994, Section 1.5.1) hold. The functional central limit theorem can also be applied to martingale differences, when the assumptions of Theorem 27.14 by Davidson (1994) are satisfied.

Using the above introduced notation, the null and the alternative hypothesis can be expressed as

$$H_0 : \tau = n \tag{1.3}$$

and

$$H_1 : \tau < n, \ \delta \neq 0. \tag{1.4}$$

A graphical illustration of the change point model (1.2) in mean under the alternative, can be seen in Figure 1.1.

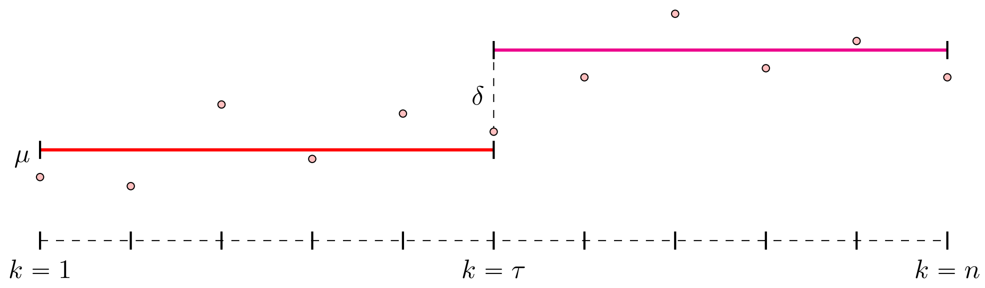


Figure 1.1: Illustration of the change point in location problem – abrupt change in mean.

For this situation, a broad range of test statistics has been developed. Many of them are functionals of the *cumulative sums (CUSUM) statistic*

$$S_{k,n} = \sum_{i=1}^{k} (Y_i - \bar{Y}_n).$$

where $\bar{Y}_n = 1/n \sum_{j=1}^{n} Y_j$ is the sample mean. This is due to the fact that such test statistics naturally arise as a result of the likelihood approach (Csörgő and Horváth, 1997, Chapter 1).

For example the test statistic

$$T_n = \frac{1}{\sqrt{\widehat{\sigma}_n^2}} \max_{1 \leqslant k \leqslant n-1} \left| \sum_{i=1}^{k} (Y_i - \bar{Y}_n) \right| \tag{1.5}$$

with a suitable variance estimator $\widehat{\sigma}_n^2$ may be used. The null hypothesis is rejected for large values of T_n.

1.4 Variance estimation

In order to ensure that a test statistic is asymptotically distribution-free under the null hypothesis, it is necessary to use a suitable estimator of variance for the random error terms. The minimal requirement for $\widehat{\sigma}_n^2$ would be consistency (i.e., $\widehat{\sigma}_n^2 \xrightarrow{\mathsf{P}} \sigma^2$) under H_0 and boundedness (in probability) under H_1.

Often, the *Bartlett estimator* is used to estimate the variance

$$\hat{\sigma}_n^2(L) = \hat{R}(0) + 2 \sum_{1 \leqslant k \leqslant L} \left(1 - \frac{k}{L}\right) \hat{R}(k), \quad L < n,$$

where

$$\hat{R}(k) = \frac{1}{n} \sum_{1 \leqslant i \leqslant n-k} (Y_i - \bar{Y}_n)(Y_{i+k} - \bar{Y}_n), \quad 0 \leqslant k < n.$$

However, it does not always provide satisfactory results and finding a proper estimate may be troublesome. The rate of convergence is small even under the null hypothesis and $\hat{\sigma}_n^2(L)$ might go to infinity under the alternative (Horváth et al., 2008).

The consistency properties of the above described Bartlett estimator and of its modification are studied in Antoch et al. (1997). The authors also describe difficulties of long run variance estimation when detecting a change in the mean of a linear process in more detail. A simulation study shows that it is not easy to find a variance estimate that would work well both under null hypothesis and under alternative. Furthermore, such estimators are often very sensitive to the *choice of the window length* L.

1.5 Ratio type test statistic based on CUSUM statistic

In Horváth et al. (2008) the authors introduced and studied several ratio CUSUM-type test statistics. When using these statistic, estimating variance is not necessary. We take a closer look on one of the proposed ratio type statistics

$$\mathcal{T}_n = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \frac{\max_{1 \leqslant i \leqslant k} \left| \sum_{1 \leqslant j \leqslant i} (Y_j - \bar{Y}_k) \right|}{\max_{k \leqslant i \leqslant n-1} \left| \sum_{i+1 \leqslant j \leqslant n} (Y_j - \widetilde{Y}_k) \right|}, \tag{1.6}$$

where $0 < \gamma < 1/2$ is a given constant and

$$\bar{Y}_k = \frac{1}{k} \sum_{i=1}^{k} Y_i \qquad \text{and} \qquad \widetilde{Y}_k = \frac{1}{n-k} \sum_{i=k+1}^{n} Y_i. \tag{1.7}$$

The asymptotic properties under the null hypothesis and under the alternative are described in following two theorems.

Theorem 1.1. *Suppose that* $Y_1, \ldots, Y_n$ *follow model* (1.2) *and that the null hypothesis* (1.3) *is true. Then, if Assumption T1 holds*

$$\mathcal{T}_n \xrightarrow[n\to\infty]{\mathscr{D}} \sup_{\gamma \leqslant t \leqslant 1-\gamma} \frac{\sup_{0 \leqslant u \leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup_{t \leqslant u \leqslant 1} \left| \widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t) \right|}, \tag{1.8}$$

where $\{\mathcal{W}(x), 0 \leqslant x \leqslant 1\}$ *is a standard Wiener process and* $\widetilde{\mathcal{W}}(x) = \mathcal{W}(1) - \mathcal{W}(x)$.

Theorem 1.2. *Suppose that* $Y_1, \ldots, Y_n$ *follow model* (1.2) *and that the alternative hypothesis* (1.4) *is true. Further suppose that* $\tau = [nt]$ *with some* $0 < t < 1$ *and*

$$n^{1/2} |\delta_n| \xrightarrow[n\to\infty]{} \infty.$$

Then, for $\gamma < t < 1 - \gamma$ *holds*

$$\mathcal{T}_n \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Proof. The proofs of both theorems may be found in Horváth et al. (2008). □

Let us consider the asymptotic distribution from (1.8). A Brownian bridge $\{\mathcal{B}(t),\, t \in [0,1]\}$ is defined as

$$\mathcal{B}(t) = \mathcal{W}(t) - t\mathcal{W}(1), \quad t \in [0,1].$$

Then, the change of variable and the scale transformation of $\mathcal{W}$ give that

$$\sup_{0\leqslant u\leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)| = \sup_{0\leqslant s\leqslant 1} |\mathcal{W}(st) - s\mathcal{W}(t)| \overset{\mathscr{D}}{=} \sqrt{t} \sup_{0\leqslant s\leqslant 1} |\mathcal{B}(s)|, \quad \forall\, t \in (0,1).$$

Similarly,

$$\sup_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right| \overset{\mathscr{D}}{=} \sqrt{1-t} \sup_{0\leqslant s\leqslant 1} \left|\widetilde{\mathcal{B}}(s)\right|, \quad \forall\, t \in (0,1),$$

where $\{\widetilde{\mathcal{B}}(s),\, s \in [0,1]\}$ is also a Brownian bridge. The Wiener process $\mathcal{W}$ has independent increments and, therefore, for any $0 < t < 1$ we have that $\{\mathcal{W}(u) - u/t\mathcal{W}(t), 0 \leqslant u \leqslant t\}$ and $\{\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t), t \leqslant u \leqslant 1\}$ are independent. Hence, $\{\mathcal{B}(s),\, s \in [0,1]\}$ and $\{\widetilde{\mathcal{B}}(s),\, s \in [0,1]\}$ are two independent Brownian bridges.

It follows that the ratio from the limit distribution (1.8) can be equivalently expressed as a functional of Brownian bridges

$$\frac{\sup_{0\leqslant u\leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|} \overset{\mathscr{D}}{=} \left(\frac{t}{1-t}\right)^{1/2} \frac{\sup_{0\leqslant s\leqslant 1} |\mathcal{B}(s)|}{\sup_{0\leqslant s\leqslant 1} \left|\widetilde{\mathcal{B}}(s)\right|} \tag{1.9}$$

for any $t \in (0,1)$.

The lower bound for k in (1.6) may be relaxed to 1 (or 2, since the ratio is equal to 0 for $k = 1$) and, correspondingly, $\sup_{\gamma\leqslant t\leqslant 1-\gamma}$ in the limit distribution may be replaced with $\sup_{0<t\leqslant 1-\gamma}$. However, as noted in Hušková (2007), this does not remain true for the supremum of the upper bound, i.e., $\sup_{\gamma\leqslant t\leqslant 1-\gamma}$ cannot be replaced by $\sup_{0<t<1}$ nor can it be replaced by $\sup_{\gamma \leqslant t<1}$, since

$$\lim_{t\to 1-} \sup_{t\leqslant s\leqslant 1} \left| \mathcal{W}(s) - \frac{s}{t}\mathcal{W}(t) \right| = 0, \quad a.s.$$

In the original article, we can also find modifications of V_n and other ratio type statistics for different kinds of alternatives, particularly for detecting changes from asymptotic stationary sequence into an asymptotic difference stationary sequence.

1.6 General form of the ratio type test statistic

A general form of a test statistic studied in this thesis is

$$\max_{n\gamma\leqslant k\leqslant n-n\gamma} \frac{\max\limits_{1\leqslant j\leqslant k} V_{j,k}}{\max\limits_{k\leqslant j\leqslant n-1} \widetilde{V}_{j,k}}, \tag{1.10}$$

where $V_{j,k}$ for $j = 1,\ldots,k$ denotes a statistic based on observations $Y_1,\ldots,Y_k$ and $\widetilde{V}_{j,k}$ for $j = k+1,\ldots,n$ is a similar statistic based on observations $Y_{k+1},\ldots,Y_n$. Constant $0 < \gamma < 1/2$ is a known (predetermined) parameter. A reversed ratio, i.e., numerator based on the last $n-k$ observation and denominator based on the first k observations can be also considered. Using a reversed ratio type statistic is possible and all the properties and asymptotic results for the test statistic remain the same, when the same model structure before and after the change is considered under alternative and when the random errors are iid (under certain conditions, they can even be weakly dependent). A situation, where both statistics cannot be symmetrically interchanged is studied in Chapter 2.

The basic motivation for studying such ratio type test statistics lies in the fact that when computing such test statistic, it is *not necessary to estimate variance of the underlying model.* This property makes the ratio type statistics a suitable alternative of classical (non-ratio type) statistic—most of all in situations, when variance estimation is not a straightforward task.

1.7 Change point estimation

Until now, we mainly dealt with task of hypotheses testing. Particularly, we tried to answer the question whether a change in mean of a time series occurred or not. However, hypothesis testing in change point analysis goes hand in hand with change point estimation and after a change is detected, it is a very important task to determine the exact time point, when the change occurred. The ratio type test statistics are typically not suitable to do this, since the ratio values in (1.10) usually achieve their maximum towards one of the end points of the observed interval (i.e., for k close to $n\gamma$ or $n(1-\gamma)$).

To estimate the actual change point, several methods may be used. In the situation of stable variance, common change point estimators generally do not require variance estimation. Therefore, the ratio type statistic may be used in combination with such methods. For example, a natural choice in the previously discussed case would be to employ the related non-ratio type statistic, which corresponds to the numerator of the ratio type statistic for $k = n$ in (1.6), i.e.,

$$\widehat{\tau}_n = \arg \max_{1 \leqslant i \leqslant n} \left| \sum_{1 \leqslant j \leqslant i} (Y_j - \bar{Y}_n) \right|.$$

1.8 State of art and preliminary work on the ratio type test statistics

Kim (2000) and, consequently, Kim and Amador (2002) studied how to detect a structural change characterized by a shift in persistence of linear time series using a similarly constructed ratio type statistic. This work is further continued by the simulation study of Leybourne and Taylor (2006), where both asymptotic and finite sample properties of the proposed test are studied.

In Hušková (2007) two ratio type statistics based on cumulative sums of residuals are briefly introduced. These are intended to be used when testing whether the mean has changed at an unknown time and also when testing for change from asymptotic stationary sequence into asymptotic difference stationary sequence. In Horváth et al. (2008), more details on the topic are given and other ratio type test statistics are introduced. The applicability of the method is demonstrated through a simulation study.

Zhao et al. (2010) and Zhao et al. (2011) propose ratio type tests to detect change in the variance of linear processes. Furthermore, a new change point estimation method based on the ratio type statistics is introduced by Zhao et al. (2011).

Chen and Tian (2014) studied a ratio type test to detect the variance change in the nonparametric regression models under both fixed and random design cases.

The same ratio type statistic as introduced in this chapter is studied to analyse change in mean for heavy tailed distributions in the work of Wang et al. (2014). This leads to a more general asymptotic distribution under the null hypothesis, which is a functional of stable Levy processes. A bootstrap approximation method to determine the critical values and change point estimation using the ratio method are also discussed.

Bazarova et al. (2014) develop a ratio type test to detect changes in the location parameters of dependent observations with infinite variances. The ratio type statistic based on the cumulative sums' process is adjusted by trimming off a set of observations that are the largest in magnitude. Moreover, the ratio and non-ratio type statistics are compared by simulations.

1.9 Groundwork for the thesis

This thesis focuses on the ratio type statistics based on cumulative sums that where derived from maximum type test statistics for detection of changes. Part of the work has already been published.

In Madurkayová (2007) a ratio type statistic for testing the gradual change in mean is derived. The random errors are considered to be iid with finite variance. Ratio type statistics for a trending regression model with iid random errors are investigated in Madurkayová (2009a). Madurkayová (2009b) and Madurkayová (2011) deal with a robust version of the ratio type statistic for the abrupt change in location parameter, while α-mixing random errors are assumed. Furthermore, a block bootstrap method for obtaining critical values of the test statistic is studied. Peštová and Pešta (2015) focus on panel data that consist of a moderate or relatively large number of panels, while the panels contain a small number of observations. Testing procedures based on the ratio type statistics to detect a possible common change in means of the panels are established. A bootstrap testing technique in this setup is proposed and its validity proved.

Chapter 2

Least Squares Procedures For Gradual Change In Mean

In this chapter, we deal with statistical methods for detection of a gradual change in mean after an unknown time point, where the change is *no more rapid but rather continuous*. We report on some recent results related to the topic and, consequently, we try to extend the ideas by incorporating the ratio type test statistics for the detection of a gradual change.

We outline the possibility of extending the idea of the ratio CUSUM type statistic introduced in the previous chapter for testing simple shift in mean to the case of testing for the gradual change. A demonstration of the proposed method on simulated data is also included. The text is based on the paper by Madurkayová (2007).

2.1 Introduction

In the previous chapter, a ratio type test statistic for detecting a single change in the mean was described. Now, we handle the problem of testing for the gradual change. We describe a non-ratio type test statistic based on partial sums of weighted residuals, which can be considered as an analogue of the CUSUM statistic. Then, we take it as a basis for the ratio type test statistic, similar to the one for testing against one abrupt change in the mean. Moreover, we study asymptotic properties of the proposed test statistic.

The problem of testing the null hypothesis of no change against the alternative of the gradual change after some unknown time point τ is related to the case of testing a change in regression parameters. Several methods for handling such types of alternatives are generally discussed for example in Csörgő and Horváth (1997) or Hušková and Steinebach (2000). Particularly, Jarušková (1998) and Albin and Jarušková (2003) studied and proposed testing appearance of a linear trend. Additionally to that, Jarušková (1999) dealt with appearance of a polynomial trend.

2.2 Basic assumptions and notation

First of all, let us introduce the notation. Similarly as in the previous chapter, we suppose to have a set of observations $Y_1, \ldots, Y_n$ obtained at time ordered points, which follow the general location model (1.1) taking into account Remark 1.1. The model with the *gradual change after an unknown time point* τ can be further specified as

$$Y_k = \mu + \delta \left(\frac{k-\tau}{n} \right)_+^\alpha + \varepsilon_k, \quad k = 1, \ldots, n, \tag{2.1}$$

where μ, $\delta = \delta_n$, and $\tau = \tau_n$ are unknown parameters. The symbol a_+ denotes the positive part of a real number a, i.e.,

$$a_+ = \begin{cases} a & \text{if } a \geqslant 0, \\ 0 & \text{if } a < 0. \end{cases}$$

The parameter $\alpha > 0$ is supposed to be known. We assume the following model assumption.

Assumption G1. Errors $\varepsilon_1, \ldots, \varepsilon_n$ are iid such that $\mathsf{E}\,\varepsilon_k = 0$, $\mathsf{Var}\,\varepsilon_k = \sigma^2 > 0$, and $\mathsf{E}\,|\varepsilon_k|^{2+\Delta} < \infty$, for $k = 1, \ldots, n$ and some $\Delta > 0$.

Note that Assumption G1 is postulated in the way that the functional central limit theorem holds for the errors of model (2.1).

We are again interested in testing the null hypothesis of no change in mean

$$H_0 : \tau = n \tag{2.2}$$

against the alternative of the gradual change in mean

$$H_1 : \tau < n, \ \delta \neq 0. \tag{2.3}$$

A graphical illustration of the change point model (2.1) for the gradual change in mean under the alternative can be seen in Figure 2.1.

2.3 Non-ratio type test statistic

The above described model is considered in Hušková and Steinebach (2000). The authors studied the properties of a class of test procedures based on partial sums of weighted residuals

$$\bar{S}_{k,n} = \sum_{i=1}^{n} (x_{ik} - \bar{x}_{k,n})\, Y_i, \qquad k = 1, \ldots, n,$$

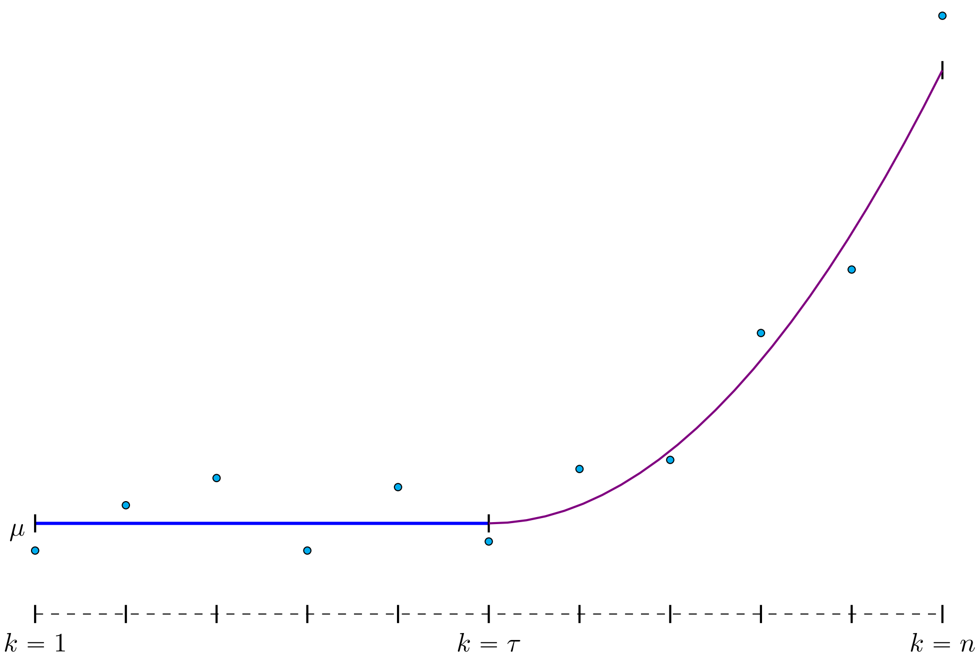


Figure 2.1: Illustration of the gradual change point problem.

where

$$x_{ik} = x_{ik}(n, \alpha) = \left(\frac{i-k}{n}\right)_+^\alpha, \quad i, k = 1, \ldots, n;$$
$$\bar{x}_{k,n} = \bar{x}_{k,n}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} x_{ik}, \quad k = 1, \ldots, n.$$

The (non-ratio type) test statistic has the form

$$G_n = \frac{1}{\widehat{\sigma}_n \sqrt{n}} \max_{1 \leqslant k \leqslant n-1} \left|\bar{S}_{k,n}\right|,$$

where $\widehat{\sigma}_n$ is some general consistent estimator of σ, i.e., it satisfies the condition

$$\widehat{\sigma}_n - \sigma = o_{\mathsf{P}}(1), \quad n \to \infty. \tag{2.4}$$

Under the assumptions for the model of the gradual change, asymptotic behavior of G_n under the null hypothesis is given by the following theorem.

Theorem 2.1. *Suppose that $Y_1, \ldots, Y_n$ follow model* (2.1)*, Assumption G1 is satisfied, and* *condition* (2.4) *holds. Then, under null hypothesis* (2.2)

$$G_n \xrightarrow[n\to\infty]{\mathscr{D}} \sup_{0\leqslant t\leqslant 1}\left|\int_t^1 \alpha(x-t)^{\alpha-1}\mathcal{W}(1-x)dx - \frac{(1-t)^{\alpha+1}}{\alpha+1}\mathcal{W}(1)\right|,$$

where $\{\mathcal{W}(t), 0 \leqslant t \leqslant 1\}$ *is a standard Wiener process.*

Proof. Theorem 2.2 by Hušková and Steinebach (2000) for $g(t) = t_+^{\alpha}$, $t \in \mathbb{R}$ is applied together with the Slutsky's theorem and consistency assumption (2.4). □

The next theorem describes the limit behavior of the non-ratio type test statistic under the alternative.

Theorem 2.2. *Suppose that* $Y_1, \ldots, Y_n$ *follow model* (2.1) *and Assumption G1 is satisfied. Assume that* $\sqrt{n}|\delta_n| \to \infty$ *as* $n \to \infty$ *and* $\tau = [nt]$ *for some* $t \in (0,1)$. *Then, under alternative* (2.3)

$$G_n \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Proof. Without loss of generality suppose $\mu = 0$. Let us take $k = \tau$. Then for the mean value of $n^{-1/2}|\bar{S}_{\tau,n}|$, we have that

$$\mathsf{E}\,\frac{|\bar{S}_{\tau,n}|}{n^{1/2}} \geqslant \left|\mathsf{E}\,\frac{\bar{S}_{\tau,n}}{n^{1/2}}\right|,$$

where the right side is equal to

$$\left|\sum_{j=\tau+1}^{n}\frac{\delta_n}{n^{1/2}}\left(\frac{j-\tau}{n}\right)^{2\alpha} - \sum_{j=\tau+1}^{n}\left(\frac{j-\tau}{n}\right)^{\alpha}\frac{1}{n}\sum_{m=\tau+1}^{n}\frac{\delta_n}{n^{1/2}}\left(\frac{m-\tau}{n}\right)^{\alpha}\right|$$
$$= n^{1/2}|\delta_n|\left[\frac{(n-[nt])^{2\alpha+1}}{n^{2\alpha+1}}\left(\frac{\alpha^2}{(2\alpha+1)(\alpha+1)^2} + \frac{[nt]}{n(\alpha+1)^2} + o\left(1\right)\right)\right]$$

as $n \to \infty$. The expression in square brackets tends to a positive non-zero limit. Hence by assumption $\sqrt{n}|\delta_n| \to \infty$ as $n \to \infty$, we have $\mathsf{E}\,|n^{-1/2}\bar{S}_{\tau,n}| \to \infty$ as $n \to \infty$. Subsequently, $\max_{1\leqslant k\leqslant n-1}|n^{-1/2}\bar{S}_{k,n}| \xrightarrow{\mathsf{P}} \infty$ as $n \to \infty$. □

Remark 2.1. The case of $\alpha = 1$ was discussed and limiting extreme value distributions of Gumbel type were derived for the test statistic in Jarušková (1998). These results were extended to the polynomial trend alternative in Jarušková (1999).

2.4 Ratio type test statistic for gradual change

Now, a natural question arises, whether it is possible to generalize the concept of ratio type statistics and obtain a test for the gradual change alternative using a ratio type test statistic that does not depend on the choice of estimate of σ. Being particular, it is demanded to avoid variance estimation and, hence, omit assumption (2.4). Following the analogy to the ratio type statistic (1.6), we arrive at the idea to study the statistic of the following form

$$\mathcal{G}_n = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \frac{\max\limits_{k \leqslant i \leqslant n-1} |\widetilde{S}_{i,k}|}{\max\limits_{1 \leqslant i \leqslant k} |\bar{S}_{i,k}|},$$

where $0 < \gamma < 1/2$ is a given constant and

$$\begin{aligned}
\bar{S}_{i,k} &= \sum_{j=1}^{k} (x_{ji} - \bar{x}_{i,k})\, Y_j, \quad i,k = 1,\dots,n, \\
\widetilde{S}_{i,k} &= \sum_{j=k+1}^{n} (x_{ji} - \tilde{x}_{i,k})\, Y_j, \quad i = 1,\dots,n;\ k = 1,\dots,n-1, \\
\bar{x}_{i,k} &= \frac{1}{k}\sum_{j=1}^{k} x_{ji}, \quad i,k = 1,\dots,n, \\
\tilde{x}_{i,k} &= \frac{1}{n-k}\sum_{j=k+1}^{n} x_{ji}, \quad i = 1,\dots,n;\ k = 1,\dots,n-1.
\end{aligned}$$

Similarly as in the case of statistics described in Chapter 1, the numerator and the denominator in the ratio are based on different subsets of subsequent observations. However, the subsets are not the same. For a fixed k the numerator in $\mathcal{G}_n$ is based on observations $Y_{k+1},\dots,Y_n$, while in $\mathcal{T}_n$ from (1.6), it is the denominator that is based on $Y_{k+1},\dots,Y_n$. The reason for the subsets being different is that otherwise the statistic $\mathcal{G}_n$ would not have desirable asymptotic properties under alternative (2.3).

Let us discuss the asymptotic properties of the ratio type test statistic $\mathcal{G}_n$ under the null hypothesis.

Theorem 2.3 (Under null)**.** *Suppose that* $Y_1,\dots,Y_n$ *follow model* (2.1) *and Assumption G1 is satisfied. Then, under null hypothesis* (2.2)

$$\mathcal{G}_n \xrightarrow[n\to\infty]{\mathscr{D}} \sup_{\gamma \leqslant t \leqslant 1-\gamma} \frac{\sup\limits_{t \leqslant s \leqslant 1} \left| \int_s^1 \alpha(x-s)^{\alpha-1}\widetilde{\mathcal{W}}(x)\,dx - \frac{1}{1-t}\frac{(s-t)^{\alpha+1}}{\alpha+1}\widetilde{\mathcal{W}}(t) \right|}{\sup\limits_{0 \leqslant s \leqslant t} \left| \int_s^t \alpha(x-s)^{\alpha-1}\mathcal{W}(t-x)\,dx - \frac{1}{t}\frac{(t-s)^{\alpha+1}}{\alpha+1}\mathcal{W}(t) \right|}, \tag{2.5}$$

where $\{\mathcal{W}(u), 0 \leqslant u \leqslant 1\}$ *is a standard Wiener process and* $\widetilde{\mathcal{W}}(u) = \mathcal{W}(1) - \mathcal{W}(u)$.

Proof. The theorem can be proved by the same means as the asymptotic results for T_n. We give the proof along the lines of the proof of Theorem 2.1 by Hušková and Steinebach (2000, p. 61–63). First, we note that by the Komlós-Major-Tusnády strong approximations

(Csörgő and Révész, 1981), there exists a Wiener process $\{\mathcal{W}(y) : 0 \leqslant y < \infty\}$ such that

$$\max_{1 \leqslant i \leqslant k} \left| \sum_{j=1}^{i} \frac{Y_j - \mu}{\sigma} - \mathcal{W}(i) \right| = o_{\mathsf{P}}\left(k^{\frac{1}{2+\Delta}}\right), \quad k \to \infty. \tag{2.6}$$

Moreover, Theorem 1.2.1 by Csörgő and Révész (1981) provides that for any Wiener process $\{\mathcal{W}(y) : 0 \leqslant y < \infty\}$:

$$\sup_{0 \leqslant t \leqslant T} \sup_{0 \leqslant y \leqslant 1} |\mathcal{W}(t+y) - \mathcal{W}(t)| \overset{a.s.}{=} O\left((\log T)^{1/2}\right), \quad T \to \infty. \tag{2.7}$$

Let us denote $i' = k - i + 1$ and $Y'_j = Y_{k-j+1}$, $j = 1, \ldots, k$. Then, after some calculations, by (2.6) we get

$$\begin{aligned} \bar{S}_{i,k} &= \sigma \left(\sum_{m=1}^{i'-1} \left[\left(\frac{m}{n}\right)^{\alpha} - \left(\frac{m-1}{n}\right)^{\alpha} \right] \sum_{l=1}^{i'-m} \frac{Y'_l - \mu}{\sigma} - \frac{1}{k} \sum_{m=1}^{i'-1} \left(\frac{i'-m}{n}\right)^{\alpha} \sum_{l=1}^{k} \frac{Y'_l - \mu}{\sigma} \right) \\ &= \sigma \left(\sum_{m=1}^{i'-1} \left[\left(\frac{m}{n}\right)^{\alpha} - \left(\frac{m-1}{n}\right)^{\alpha} \right] \mathcal{W}(i'-m) - \frac{1}{k} \sum_{m=i+1}^{k} \left(\frac{m-i}{n}\right)^{\alpha} \mathcal{W}(k) \right) \\ &\quad + o_{\mathsf{P}}\left(\left(\frac{k-i}{n}\right)^{\alpha} \left((k-i)^{\frac{1}{2+\Delta}} + \frac{k-i}{k} k^{\frac{1}{2+\Delta}} \right) \right), \quad k \to \infty, \end{aligned} \tag{2.8}$$

uniformly in $i = 1, \ldots, k-1$. Now, let us take $k = [nt]$ and $i = [ns]$ for some $0 < s < t < 1$. Similarly as in Hušková and Steinebach (2000), by (2.7) and considering properties of the Wiener process, it can be shown that as $n \to \infty$

$$\begin{aligned} &\sum_{m=[ns]+1}^{[nt]} \left[\left(\frac{m-[ns]}{n}\right)^{\alpha} - \left(\frac{m-[ns]-1}{n}\right)^{\alpha} \right] \mathcal{W}([nt]-m+1) \\ &= \int_{\frac{[ns]+1}{n}}^{\frac{[nt]}{n}} \alpha \left(y - \frac{[ns]}{n}\right)_{+}^{\alpha-1} \mathcal{W}(n(t-y)) \mathrm{d}y \\ &\quad + O_{\mathsf{P}}\left(\left(\frac{[nt]-[ns]}{n}\right)^{\alpha} \sqrt{\log([nt]-[ns])} \right) \\ &= \frac{1}{t} \frac{(t-s)^{\alpha+1}}{(\alpha+1)} \mathcal{W}(nt) + O_{\mathsf{P}}\left(\left(\frac{[nt]-[ns]}{n}\right)^{\alpha} \sqrt{\log([nt]-[ns])} \right) \end{aligned} \tag{2.9}$$

and

$$\begin{aligned}
&\frac{1}{[nt]}\sum_{m=1}^{[nt]}\left(\frac{m-[ns]}{n}\right)_+^{\alpha}\mathcal{W}([nt])\\
&=\frac{1}{[nt]}\sum_{m=1}^{[nt]}\left(\frac{m-[ns]}{n}\right)_+^{\alpha}\mathcal{W}(nt)+O_{\mathsf{P}}\left(\frac{([nt]-[ns])^{\alpha+1}\sqrt{\log([nt])}}{[nt]n^{\alpha}}\right)\\
&=\frac{1}{t}\frac{(t-s)^{\alpha+1}}{(\alpha+1)}\mathcal{W}(nt)+O_{\mathsf{P}}\left(\frac{([nt]-[ns])^{\alpha+1}\sqrt{\log([nt])}}{[nt]n^{\alpha}}\right) \qquad (2.10)
\end{aligned}$$

uniformly in $s,t\in[0,1]$, $s\leqslant t-\frac{1}{n}$. Furthermore,

$$\begin{aligned}
&\int_{\frac{[ns]+1}{n}}^{\frac{[nt]}{n}}\alpha\left(y-\frac{[ns]}{n}\right)_+^{\alpha-1}\mathcal{W}(n(t-y))\mathrm{d}y\\
&\qquad=\int_s^t\alpha\,(y-s)_+^{\alpha-1}\,\mathcal{W}(n(t-y))\mathrm{d}y+O_{\mathsf{P}}\left(\left(\frac{[nt]-[ns]}{n}\right)_+^{\alpha}\right) \qquad (2.11)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{[nt]}\sum_{m=1}^{[nt]}\left(\frac{m-[ns]}{n}\right)_+^{\alpha}\mathcal{W}(nt)\\
&\qquad=\frac{1}{t}\frac{(t-s)^{\alpha+1}}{(\alpha+1)}\mathcal{W}(nt)+O_{\mathsf{P}}\left(\left(\frac{1}{\sqrt{nt}}\frac{[nt]-[ns]}{n}\right)_+^{\alpha}\right) \qquad (2.12)
\end{aligned}$$

uniformly in $s,t\in[0,1]$, $s\leqslant t-\frac{1}{n}$. It follows that

$$\begin{aligned}
\frac{\bar{S}_{i,k}}{\sqrt{n}}&=\sigma\int_0^t\left((x-s)_+^{\alpha}-\frac{1}{t}\int_0^t(y-s)_+^{\alpha}\mathrm{d}y\right)\mathrm{d}\mathcal{W}(x)+O_{\mathsf{P}}\left(n^{\frac{1}{2+\Delta}-\frac{1}{2}}\right) \qquad (2.13)\\
&=\sigma\int_0^t\left((x-s)_+^{\alpha}-\frac{1}{t}\frac{(t-s)^{\alpha+1}}{\alpha+1}\right)\mathrm{d}\mathcal{W}(x)+O_{\mathsf{P}}\left(n^{\frac{1}{2+\Delta}-\frac{1}{2}}\right),\quad n\to\infty. \qquad (2.14)
\end{aligned}$$

Then combining (2.8)–(2.14) for a given $k=[nt]$ and integrating by parts, we get

$$\begin{aligned}
&\sup_{0\leqslant s\leqslant t}\left|n^{-1/2}\bar{S}_{i,[nt]}-\sigma\left(\int_s^t\alpha(x-s)^{\alpha-1}\mathcal{W}(t-x)\mathrm{d}x-\frac{1}{t}\frac{(t-s)^{\alpha+1}}{\alpha+1}\mathcal{W}(t)\right)\right|\\
&\qquad=O_{\mathsf{P}}\left(n^{\frac{1}{2+\Delta}-\frac{1}{2}}\right),\quad n\to\infty. \qquad (2.15)
\end{aligned}$$

An analogous statement for $\max_{k \leqslant i \leqslant n} |\widetilde{S}_{i,k}|$

$$\sup_{t \leqslant s \leqslant 1} \left| n^{-1/2} \widetilde{S}_{i,[nt]} - \sigma \left(\int_s^1 \alpha(x-s)^{\alpha-1} \widetilde{\mathcal{W}}(x) \mathrm{d}x - \frac{1}{1-t} \frac{(s-t)^{\alpha+1}}{\alpha+1} \widetilde{\mathcal{W}}(t) \right) \right| \\ = O_{\mathsf{P}}\left(n^{\frac{1}{2+\Delta} - \frac{1}{2}} \right), \quad n \to \infty \quad (2.16)$$

can be proved by the same arguments and by realizing the property of independent increments of the Wiener process.

Assumption G1 yields that

$$\left(\frac{1}{\sqrt{n}} \sum_{1 \leqslant i \leqslant nt} \frac{Y_i - \mu}{\sigma}, \frac{1}{\sqrt{n}} \sum_{nt < i \leqslant n} \frac{Y_i - \mu}{\sigma} \right) \xrightarrow[n \to \infty]{\mathscr{D}^2[0,1]} \left(\mathcal{W}(t), \widetilde{\mathcal{W}}(t) \right). \quad (2.17)$$

Relations (2.15)–(2.17) under hypothesis H_0 give

$$\left(\frac{1}{\sigma\sqrt{n}} \max_{1 \leqslant i \leqslant [nt]} |\bar{S}_{i,[nt]}|, \frac{1}{\sigma\sqrt{n}} \max_{[nt] \leqslant i \leqslant n} |\widetilde{S}_{i,[nt]}| \right) \\ \xrightarrow[n \to \infty]{\mathscr{D}^2[\gamma, 1-\gamma]} \left(\sup_{0 \leqslant s \leqslant t} \left| \int_s^t \alpha(x-s)^{\alpha-1} \mathcal{W}(t-x) \mathrm{d}x - \frac{1}{t} \frac{(t-s)^{\alpha+1}}{\alpha+1} \mathcal{W}(t) \right|, \right. \\ \left. \sup_{t \leqslant s \leqslant 1} \left| \int_s^1 \alpha(x-s)^{\alpha-1} \widetilde{\mathcal{W}}(x) \mathrm{d}x - \frac{1}{1-t} \frac{(s-t)^{\alpha+1}}{\alpha+1} \widetilde{\mathcal{W}}(t) \right| \right)$$

for all $0 < \gamma < 1/2$. The assertion of the theorem is a direct consequence of the previous statement and the continuous mapping theorem. □

A consistent test is a test, which power for a fixed untrue null hypothesis increases to one as the number of observations increases. Therefore, we need to show how the test statistic behaves under the alternative.

Theorem 2.4 (Under alternative). *Suppose that $Y_1, \ldots, Y_n$ follow model* (2.1) *and Assumption G1 is satisfied. Assume that*

$$n^{1/2} |\delta_n| \to \infty, \quad n \to \infty \quad (2.18)$$

and $\tau = [nt]$ for some $\gamma < t < 1 - \gamma$. Then under alternative (2.3)

$$\mathcal{G}_n \xrightarrow[n \to \infty]{\mathsf{P}} \infty.$$

Proof. Without loss of generality suppose $\mu = 0$. Let us take $k = \tau$ and $i = \tau$. Then for the mean value of $n^{-1/2}|\widetilde{S}_{\tau,\tau}|$ we have by the Jensen's inequality

$$\mathsf{E}\,\frac{|\widetilde{S}_{\tau,\tau}|}{n^{1/2}} \geqslant \left|\mathsf{E}\,\frac{\widetilde{S}_{\tau,\tau}}{n^{1/2}}\right|,$$

where the right side is equal to

$$\left|\sum_{j=\tau+1}^{n}\frac{\delta_n}{n^{1/2}}\left(\frac{j-\tau}{n}\right)^{2\alpha} - \sum_{j=\tau+1}^{n}\left(\frac{j-\tau}{n}\right)^{\alpha}\frac{1}{n-\tau}\sum_{m=\tau+1}^{n}\frac{\delta_n}{n^{1/2}}\left(\frac{m-\tau}{n}\right)^{\alpha}\right|$$
$$= n^{1/2}|\delta_n|\left[\frac{(n-[nt])^{2\alpha+1}}{n^{2\alpha+1}}\left(\frac{\alpha^2}{(2\alpha+1)(\alpha+1)^2}+\frac{[nt]}{n(\alpha+1)^2}+o\,(1)\right)\right],$$

as $n \to \infty$. The expression in square brackets tends to a positive non-zero limit. Hence by (2.18), we have that $\mathsf{E}\,|n^{-1/2}\widetilde{S}_{\tau,\tau}| \to \infty$ as $n \to \infty$. Subsequently,

$$\max_{\tau\leqslant i\leqslant n-1}|n^{-1/2}\widetilde{S}_{i,\tau}| \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Since we have taken $\tau = k$, then $n^{-1/2}\max_{1\leqslant i\leqslant k}|\bar{S}_{i,k}|$ has the same distribution as it has under hypothesis H_0 (see the proof of Theorem 2.3). Therefore, it is bounded in probability, which implies that $\mathcal{G}_n \xrightarrow{\mathsf{P}} \infty$ as $n \to \infty$. □

Theorem 2.4 says that in presence of the gradual change in mean, the test statistic explodes above all bounds. Hence, the testing procedure is *consistent* and the asymptotic distribution from Theorem 2.3 can be used to construct the test.

The null hypothesis is rejected for large values of $\mathcal{G}_n$. I.e., we reject H_0 at significance level α if and only if $\mathcal{G}_n > g_{1-\alpha,\gamma}$, where $g_{1-\alpha,\gamma}$ is the $(1-\alpha)$-quantile of the asymptotic distribution (2.5), However, an explicit form of the limit distribution (2.5) under the null hypothesis is not known. Therefore, in order to obtain critical values, we have to use, for example, simulations from the limit distribution.

2.5 Application on simulated data

We present some applications of the proposed ratio type test statistic to simulated data from normal and Laplace distributions. When simulating the gradual change, we took $\alpha = 1$, i.e, a constant mean changes into a linear one. The difference between the behavior under the null hypothesis and under the alternative becomes apparent approximately for $\delta = 3$. In Figures 2.2 and 2.3, we can see the values of ratio

$$Q_k = \frac{\max_{k\leqslant i\leqslant n-1}|\widetilde{S}_{i,k}|}{\max_{1\leqslant i\leqslant k}|\bar{S}_{i,k}|},$$

computed for $k: n\gamma \leqslant k \leqslant n - n\gamma$ with $\gamma = 0.1$. Simulated 95% critical values for each of the two distributions are depicted by a horizontal line.

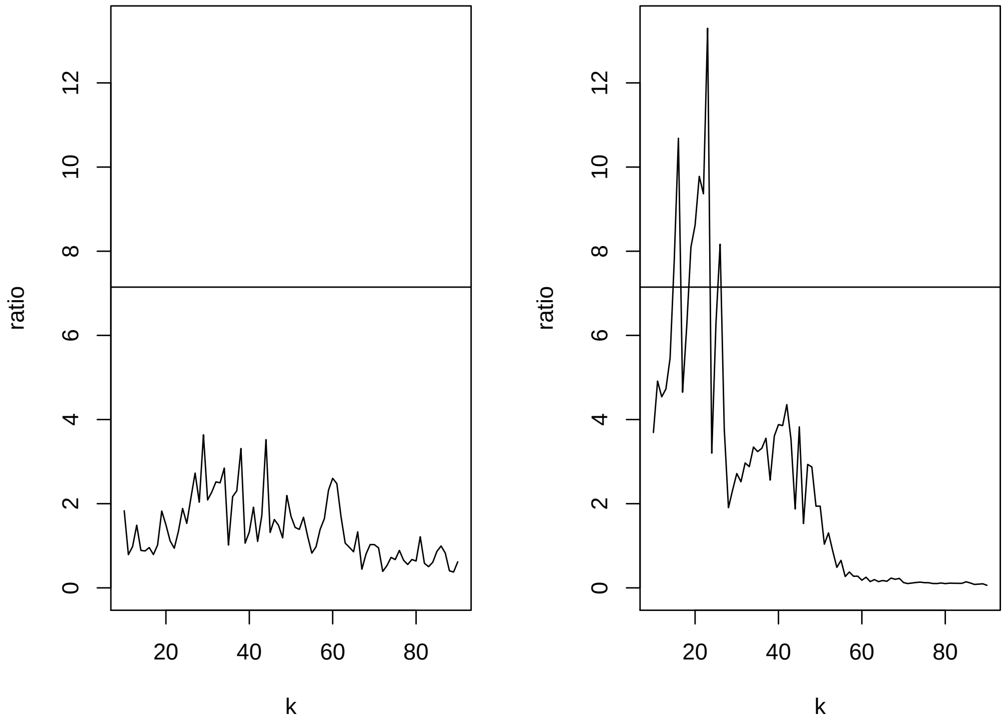


Figure 2.2: The values of Q_k from $\mathcal{G}_n$ for simulated normal distribution samples with parameters $\mu = 0$, $\sigma = 1$, $n = 100$, and $\gamma = 0.1$. The figure on the left side refers to the null hypothesis. Figure on the right refers to the alternative with $\tau = n/2 = 50$, $\alpha = 1$, and $\delta = 4$.

On one hand, one can observe that in both Figures 2.2 and 2.3, the curves from left subfigures—corresponding to the ratios under the null—hardly come closer to the straight horizontal line (critical value). This means that the value of ratio type test statistic $\mathcal{G}_n$ is not sufficiently large to reject hypothesis H_0.

On the other hand, the curves from the right subfigures—corresponding to the ratios under the alternative—clearly cross to the critical value depicted by the straight horizontal line. Therefore, the value of ratio type test statistic $\mathcal{G}_n$ is sufficiently large to reject hypothesis H_0 against alternative H_1.

A deeper simulation study concerning also a sensitivity analysis, but for a different test

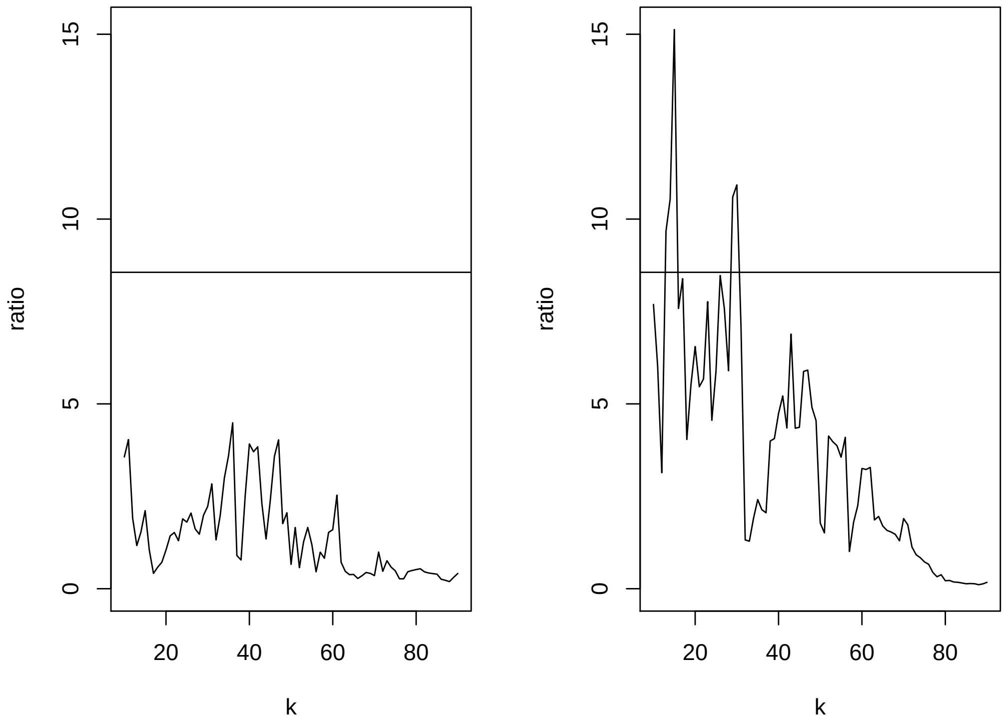


Figure 2.3: The values of Q_k from $\mathcal{G}_n$ for simulated Laplace distribution samples with parameters $\mu = 0$, $b = 1$, $n = 100$, $\delta = 0$, and $\gamma = 0.1$. The figure on the left side refers to the null hypothesis. Figure on the right refers to the alternative with $\tau = n/2 = 50$, $\alpha = 1$, and $\delta = 4$.

statistic, will be performed in the next Chapter 3. The reason for not showing a similar simulation experiment here is that the model and setup in Chapter 3 allow more complex design of the simulations, which is more interesting.

2.6 Summary

The detection of the gradual change in mean with at most one change at some unknown time point is studied. When a suitable variance estimate is not available or problematic, the ratio type test statistics provide an advantageous alternative to the non-ratio type statistics. Therefore, we extend the usage of non-ratio type test statistics for the ratio type ones in the model of gradual change in mean. Asymptotic behavior of the ratio type test statistic is elaborated under the null hypothesis of no change and under the alternative of one gradual change as well.

Chapter 3

Robust Procedures For Abrupt Change In Mean

This chapter presents procedures for detection of a change in mean of the observed time ordered sequence. The considered underlying stochastic model allows *at most one change*. Moreover, the character of the change is that it is a rapid—also called abrupt—change, which happens suddenly at some unknown time point. Main focus is given to the test procedures based on ratio type test statistics that are functionals of the partial sums of M-type residuals. We explore the possibility of applying the bootstrap method for obtaining critical values of the proposed test statistics and derive the limit behavior of the circular block bootstrap test statistic. The core of this chapter comes from Madurkayová (2009b) and Madurkayová (2011).

3.1 Introduction

We describe basic properties of statistics for detection of a change in the location model with at most one abrupt change in mean. Asymptotic behavior of the ratio type test statistics is studied under the null hypothesis of no change as well as under the alternative of a change occurring at some unknown time point. We extend the ideas presented by Hušková (2007) and Horváth et al. (2008) in the way that *weakly dependent errors* of the model are supposed together with incorporating *general score function* in the test statistics.

In order to obtain critical values for the studied test statistics not only from their asymptotic distributions, we focus on the *circular moving block bootstrap method* (Politis and Romano, 1992) in case of L_2 score function. This type of resampling method was applied in a similar situation by Kirch (2006).

3.2 Abrupt change point model

Let us consider observations $Y_1, \dots, Y_n$ that were obtained at n time ordered points. We study the location model with at most one abrupt change in mean

$$Y_k = \mu + \delta\, \mathcal{I}\{k > \tau\} + \varepsilon_k, \quad k = 1, \dots, n, \tag{3.1}$$

where μ, $\delta = \delta_n$, and $\tau = \tau_n$ are unknown parameters. Time point τ is called the change point. By $\varepsilon_1, \dots, \varepsilon_n$, we denote the random error terms.

We are going to test the null hypothesis that no change occurred

$$H_0 : \tau = n \tag{3.2}$$

against the alternative that change occurred at some unknown time point τ

$$H_1 : \tau < n,\ \delta \neq 0. \tag{3.3}$$

3.3 M-estimates of a location parameter

To estimate a location parameter of the distribution, the maximum likelihood approach is traditionally used. Huber et al. (1964) proposed a generalization of this method.

An *M-estimate of location parameter* $\theta \in \mathbb{R}$ is defined as

$$\hat{\theta} = \arg\min_{t \in \mathbb{R}} \sum_{i=1}^{m} \rho(X_i - t),$$

where $X_1, \dots, X_m$ is the random sample, θ is an unknown parameter of interest, ρ is a loss function and $\hat{\theta}$ is the so-called M-estimate. If function $\rho(x - t)$ is differentiable with respect to t, the M-estimate is a solution of equation

$$\sum_{i=1}^{m} \psi(X_i - t) = 0,$$

where $\psi(u) = \frac{\partial}{\partial u}\rho(u)$ denotes a *score function*. More on M-estimates can be found in Jurečková et al. (2012, Chapters 3 and 5) or Serfling (1980, p. 243).

Alternatively, the M-estimate may be viewed as generalization of the least squares estimate. Considering loss function $\rho(x - t) = (x - t)^2/2$ results in the sample mean as the least squares (and also empirical) estimate of the mean value (theoretical mean). The corresponding score function is $\psi(x - t) = x - t$. Therefore, M-estimates may be viewed as generalized estimates of the location parameter. The main advantage of using M-estimates is that they are more robust with respect to outliers and to heavy tailed distributions when comparing to the least squares estimate.

3.4 Ratio type statistic based on M-residuals

Let us move to the ratio type test statistic. We robustify the original ratio type test statistic (1.6) from Chapter 1. Following the ideas described in Hušková (2007), Horváth et al. (2008), and Hušková and Marušiaková (2012), a test statistic based on M-residuals is considered

$$\mathcal{A}_n(\psi) = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \frac{\max_{1\leqslant i\leqslant k} \left| \sum_{1\leqslant j\leqslant i} \psi(Y_j - \widehat{\mu}_{1k}(\psi)) \right|}{\max_{k\leqslant i\leqslant n-1} \left| \sum_{i+1\leqslant j\leqslant n} \psi(Y_j - \widehat{\mu}_{2k}(\psi)) \right|}, \tag{3.4}$$

where $0 < \gamma < 1/2$ is a given constant, $\widehat{\mu}_{1k}(\psi)$ is an M-estimate of parameter μ based on observations $Y_1, \ldots, Y_k$, and $\widehat{\mu}_{2k}(\psi)$ is an M-estimate of μ based on observations $Y_{k+1}, \ldots, Y_n$. That means, $\widehat{\mu}_{1k}(\psi)$ is a solution of estimating equation

$$\sum_{i=1}^{k} \psi(Y_i - \mu) = 0$$

and, similarly, $\widehat{\mu}_{2k}(\psi)$ is a solution of estimating equation

$$\sum_{i=k+1}^{n} \psi(Y_i - \mu) = 0.$$

For the choice of $\psi_{L_2}(x) = x$, we get one of the statistics studied in Horváth et al. (2008). By considering different score functions, we may construct similar statistics, but more robust against outliers and more suitable for heavy tailed distributions.

3.5 Strong mixing dependence

Prior to deriving asymptotic properties of the test statistic, we need to formulate assumptions for the score function ψ and the distribution of random errors $\varepsilon_1, \ldots, \varepsilon_n$. Before that, we explain the notion of strong mixing (α-mixing) dependence in more detail.

Suppose that $\{\varepsilon_n\}_{n=1}^{\infty}$ is a sequence of random elements on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. For sub-$\sigma$-fields $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$, we define

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A\in\mathcal{A}, B\in\mathcal{B}} \left| \mathsf{P}(A \cap B) - \mathsf{P}(A)\mathsf{P}(B) \right|.$$

Intuitively, $\alpha(\cdot, \cdot)$ measures the dependence of the events in $\mathcal{B}$ on those in $\mathcal{A}$. There are many ways how to describe weak dependence or, in other words, *asymptotic independence* of random variables (Bradley, 2005). Considering a filtration $\mathcal{F}_m^n := \sigma\{\varepsilon_i \in \mathcal{F}, m \leqslant i \leqslant n\}$, sequence $\{\varepsilon_n\}_{n=1}^{\infty}$ of random variables is said to be *strong mixing* (α-mixing) if

$$\alpha(n) := \sup_{k\in\mathbb{N}} \alpha(\mathcal{F}_1^k, \mathcal{F}_{k+n}^{\infty}) \to 0, \quad n \to \infty. \tag{3.5}$$

This notion was introduced by Ibragimov (1959). Coefficients of dependence $\alpha(n)$ measure how much dependence exists between events separated by at least n observations or time periods. Note that in case of a strictly stationary sequence, the $\sup_{k\in\mathbb{N}}$ in the definition becomes redundant.

Finite order processes, which do not satisfy Doeblin's condition, can be shown to be α-mixing (Ibragimov and Linnik, 1971, pp. 312–313). Rosenblatt (1971) provides general conditions under which stationary Markov processes are α-mixing. Since functions of mixing processes are themselves mixing (Bradley, 2005), time-varying functions of any of the above mentioned processes just mentioned are mixing as well.

3.6 Limit distribution under null hypothesis

Now we proceed to the assumptions that are needed for deriving asymptotic properties of the proposed test statistic.

Assumption A1. The random error terms $\{\varepsilon_i, i \in \mathbb{N}\}$ form a strictly stationary α-mixing sequence with marginal distribution function F, that is symmetric around zero, and for some $\chi > 0$, $\chi' > 0$ there exists a constant $C_1(\chi,\chi') > 0$ such that

$$\sum_{h=0}^{\infty} (h+1)^{\chi/2} \alpha(h)^{\chi'/(2+\chi+\chi')} \leqslant C_1(\chi,\chi'), \tag{3.6}$$

where $\alpha(k)$, $k = 0, 1, \ldots$ are the α-mixing coefficients.

Assumption A2. The score function ψ is a non-decreasing and antisymmetric function.

Assumption A3.

$$\int |\psi(x)|^{2+\chi+\chi'} \mathrm{d}F(x) < \infty$$

and

$$\int |\psi(x+t_2) - \psi(x+t_1)|^{2+\chi+\chi'} \mathrm{d}F(x) \leqslant C_2(\chi,\chi')|t_2 - t_1|^{\eta},$$
$$|t_j| \leqslant C_3(\chi,\chi'), \quad j = 1, 2$$

for some constants $1 \leqslant \eta \leqslant 2 + \chi + \chi'$, $\chi > 0$, $\chi' > 0$ as in (3.6) and constants $C_2(\chi,\chi') > 0$, $C_3(\chi,\chi') > 0$ both depending only on χ and χ'.

Assumption A4. Let us denote $\lambda(t) = -\int \psi(e-t)\mathrm{d}F(e)$, for $t \in \mathbb{R}$. We assume that $\lambda(0) = 0$ and that there exists a first derivative $\lambda'(\cdot)$ that is Lipschitz in the neighborhood of 0 and satisfies $\lambda'(0) > 0$.

Assumption A5. Let

$$0 < \sigma^2(\psi) = \mathsf{E}\,\psi^2(\varepsilon_1) + 2\sum_{i=1}^{\infty} \mathsf{E}\,\psi(\varepsilon_1)\psi(\varepsilon_{i+1}) < \infty.$$

Assumption A1 is satisfied for example for ARMA processes with continuously distributed stationary innovations and bounded variance (Doukhan, 1994, Section 2.4).

The conditions regarding ψ reduce to moment restrictions for $\psi_{L_2}(x) = x$ (L_2 method) taking $\eta = 2 + \chi + \chi'$. For $\psi_{L_1}(x) = \mathrm{sgn}(x)$ (L_1 method), the conditions reduce to F being a symmetric distribution, having continuous density f in a neighborhood of 0 with $f(0) > 0$, and $\eta = 1$ for any $\chi > 0$ and $\chi' > 0$. Similarly, we may consider the derivative of the Huber loss function, i.e.,

$$\psi_H(x) = x\,\mathcal{I}\{|x| \leqslant C\} + C\,\mathrm{sgn}(x)\,\mathcal{I}\{|x| > C\} \tag{3.7}$$

for some $C > 0$. In that case to satisfy Assumptions A2–A4, we need to assume F being a symmetric distribution function with continuous density f in a neighborhood of C and $-C$ satisfying $f(C) > 0$ and $f(-C) > 0$ with $\eta = 2 + \chi + \chi'$.

The following theorem states the asymptotic behavior of the studied ratio type test statistic under the null hypothesis.

Theorem 3.1 (Under null)**.** *Suppose that $Y_1, \ldots, Y_n$ follow model* (3.1) *and assume that Assumptions A1–A5 hold. Then, under null hypothesis* (3.2)

$$\mathcal{A}_n(\psi) \xrightarrow[n\to\infty]{\mathscr{D}} \sup_{\gamma \leqslant t \leqslant 1-\gamma} \frac{\sup\limits_{0 \leqslant u \leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup\limits_{t \leqslant u \leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|}, \tag{3.8}$$

where $\{\mathcal{W}(u), 0 \leqslant u \leqslant 1\}$ is a standard Wiener process and $\widetilde{\mathcal{W}}(u) = \mathcal{W}(1) - \mathcal{W}(u)$.

Proof. The proof is inspired by several steps from the proof of Theorem 1.1 in Horváth et al. (2008). Without loss on generality, we assume that $\mu = 0$. Let

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{1 \leqslant j \leqslant nt} \psi(\varepsilon_j) \quad \text{and} \quad \widetilde{Z}_n(t) = \frac{1}{\sqrt{n}} \sum_{nt < j \leqslant n} \psi(\varepsilon_j).$$

Then, by applying Theorem 1 from Doukhan (1994, Section 1.5.1) with the consequent remark, we get

$$\left(Z_n(t), \widetilde{Z}_n(t)\right) \xrightarrow[n\to\infty]{\mathscr{D}^2[0,1]} \sigma(\psi)\,\left(\mathcal{W}(t), \widetilde{\mathcal{W}}(t)\right), \tag{3.9}$$

where $\widetilde{\mathcal{W}}(t) = \mathcal{W}(1) - \mathcal{W}(t)$. Lemma 4.3 and Lemma 4.4 by Hušková and Marušiaková (2012) and Assumptions A1–A5 lead to

$$\sup_{1\leqslant i\leqslant nt}\left\{n^{\kappa}\sqrt{\frac{[nt]}{i([nt]-i)}}\left|\sum_{1\leqslant j\leqslant i}\psi\left(Y_j-\widehat{\mu}_{1,[nt]}(\psi)\right)\right.\right.$$
$$\left.\left.-\left(\sum_{1\leqslant j\leqslant i}\psi(\varepsilon_j)-\frac{i}{[nt]}\sum_{1\leqslant j\leqslant nt}\psi(\varepsilon_j)\right)\right|\right\}\xrightarrow[n\to\infty]{\mathsf{P}}0,$$

for some $\kappa > 0$, where $[a]$ denotes the integer part of $a \in \mathbb{R}$. Hence,

$$\frac{1}{\sqrt{n}}\sup_{1<i\leqslant nt}\left|\sum_{1\leqslant j\leqslant i}\psi\left(Y_j-\widehat{\mu}_{1,[nt]}(\psi)\right)\right| = \sup_{1\leqslant i\leqslant nt}\left|Z_n\left(\frac{i}{n}\right)-\frac{i}{[nt]}Z_n(t)\right|+o_{\mathsf{P}}(1),\quad n\to\infty.$$

Similarly, we get

$$\frac{1}{\sqrt{n}}\sup_{nt<i\leqslant n}\left|\sum_{i\leqslant j\leqslant n}\psi\left(Y_j-\widehat{\mu}_{2,[nt]}(\psi)\right)\right|$$
$$=\sup_{nt<i\leqslant n}\left|\widetilde{Z}_n\left(\frac{i}{n}\right)-\frac{n-i}{n-[nt]}\widetilde{Z}_n(t)\right|+o_{\mathsf{P}}(1),\quad n\to\infty.$$

With respect to (3.9), we get for all $0 < \gamma < 1/2$

$$\left(\frac{1}{\sqrt{n}}\sup_{1<i\leqslant nt}\left|\sum_{1\leqslant j\leqslant i}\psi\left(Y_j-\widehat{\mu}_{1,[nt]}(\psi)\right)\right|,\frac{1}{\sqrt{n}}\sup_{nt-1<i\leqslant n-1}\left|\sum_{i+1\leqslant j\leqslant n}\psi\left(Y_j-\widehat{\mu}_{2,[nt]}(\psi)\right)\right|\right)$$
$$\xrightarrow[n\to\infty]{\mathscr{D}^2[\gamma,1-\gamma]}\sigma(\psi)\left(\sup_{0\leqslant u\leqslant t}\left|\mathcal{W}(u)-\frac{u}{t}\mathcal{W}(t)\right|,\sup_{t\leqslant u\leqslant 1}\left|\widetilde{\mathcal{W}}(u)-\frac{1-u}{1-t}\widetilde{\mathcal{W}}(t)\right|\right).$$

Finally, the continuous mapping theorem completes the proof. □

The null hypothesis is rejected for large values of $\mathcal{A}_n(\psi)$. Hence, we reject H_0 at significance level α if $\mathcal{A}_n(\psi) > a_{1-\alpha,\gamma}$, where $a_{1-\alpha,\gamma}$ is the $(1-\alpha)$-quantile of the asymptotic distribution (3.8). Explicit form of the limit distribution (3.8) under the null hypothesis is not known. Therefore, in order to obtain critical values, we have to use either simulation from the limit distribution or resampling methods.

For $\psi_{L_2} : \psi_{L_2}(x) = x$, $x \in \mathbb{R}$, the above stated Assumptions A2 and A4 are satisfied. We can also drop the requirement of symmetry of F in Assumption A1 and replace it by $\mathsf{E}\,\varepsilon_1 = 0$. Assumptions A3 and A5 reduce to the following two assumptions.

Assumption D1.

$$\mathsf{E}\,|\varepsilon_1|^{2+\beta} < \infty.$$

for some constant $\beta > 0$.

Assumption D2.

$$0 < \sigma^2(\psi_{L_2}) = \mathsf{E}\,\varepsilon_1^2 + 2\sum_{i=1}^{\infty} \mathsf{E}\,\varepsilon_1\varepsilon_{i+1} < \infty.$$

Next, we show how the test statistic behaves under the alternative.

Theorem 3.2 (Under alternative)**.** *Suppose that* $Y_1, \dots, Y_n$ *follow model* (3.1)*, assume that* $\sqrt{n}|\delta_n| \to \infty$ *as* $n \to \infty$*, and* $\tau = [\zeta n]$ *for some* $\gamma < \zeta < 1-\gamma$*. Then, under Assumptions A1–A5 and alternative* (3.3)

$$\mathcal{A}_n(\psi) \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Proof. Let $k > \tau+1$ and $k = [\xi n]$ for some $\zeta < \xi < 1-\gamma$. Note that $\tau = O(n)$ and $k = O(n)$ as $n \to \infty$. Lemma 4.3 and Lemma 4.4 by Hušková and Marušiaková (2012) provide

$$\sqrt{k}\left(\widehat{\mu}_{1k}(\psi) - \mu\right) = \frac{1}{\sqrt{k}\lambda'(0)}\sum_{i=1}^{k}\psi(\varepsilon_i) + \sqrt{k}\delta_n(1-\zeta) + o_{\mathsf{P}}(1), \quad n \to \infty.$$

Consequently, applying Lemma 4.3 by Hušková and Marušiaková (2012) again, we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{k}}\left|\sum_{j=1}^{\tau+1}\psi\left(Y_j - \widehat{\mu}_{1k}(\psi)\right) - \left(\sum_{j=1}^{\tau+1}\psi(\varepsilon_j) - \frac{\tau+1}{k}\sum_{l=1}^{k}\left(\psi(\varepsilon_l) + \lambda'(0)\delta_n g_\zeta((\tau+1)/k)\right)\right)\right| \\
&\leqslant \max_{1\leqslant i\leqslant k}\frac{1}{\sqrt{k}}\left|\sum_{j=1}^{i}\psi\left(Y_j - \widehat{\mu}_{1k}(\psi)\right) - \left(\sum_{j=1}^{i}\psi(\varepsilon_j) - \frac{i}{k}\sum_{l=1}^{k}\left(\psi(\varepsilon_l) + \lambda'(0)\delta_n g_\zeta(i/k)\right)\right)\right| \\
&\xrightarrow[n\to\infty]{\mathsf{P}} 0,
\end{aligned}$$

where $g_\zeta(x) = \min(\zeta, x)[1 - \max(\zeta, x)]$ for $x \in (0,1)$. Since $\sqrt{n}|\delta_n| \to \infty$ as $n \to \infty$ and

$$\frac{1}{\sqrt{k}}\left|\sum_{j=1}^{\tau+1}\psi(\varepsilon_j) - \frac{\tau+1}{k}\sum_{l=1}^{k}\psi(\varepsilon_l)\right| = O_{\mathsf{P}}(1), \quad n \to \infty$$

according to the proof of Theorem 3.1, we get

$$\max_{1\leqslant i\leqslant k}\frac{1}{\sqrt{k}}\left|\sum_{j=1}^{i}\psi\left(Y_j - \widehat{\mu}_{1k}(\psi)\right)\right| \geqslant \frac{1}{\sqrt{k}}\left|\sum_{j=1}^{\tau+1}\psi\left(Y_j - \widehat{\mu}_{1k}(\psi)\right)\right| \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Note that there is no change in the means of $Y_k, \dots, Y_n$. Again from the proof of Theorem 3.1, we have

$$\max_{[\xi n] \leqslant i \leqslant n-1} \frac{1}{\sqrt{n}} \left| \sum_{j=i+1}^{n} \psi\left(Y_j - \widehat{\mu}_{2k}(\psi)\right) \right| \xrightarrow[n\to\infty]{\mathscr{D}[\gamma,1-\gamma]} \sigma(\psi) \sup_{\xi \leqslant u \leqslant 1} \left| \widetilde{\mathcal{W}}(u) - \frac{1-u}{1-\xi} \widetilde{\mathcal{W}}(\xi) \right|,$$

which completes the proof. □

Theorem 3.2 says that in presence of the structural change in mean, the test statistic explodes above all bounds. Hence, the procedure is consistent and the asymptotic distribution from Theorem 3.1 can be used to construct the test.

The limit results for $\mathcal{A}_n(\psi_{L_2})$ were derived in Hušková (2007) and Horváth et al. (2008) under less restrictive assumptions regarding the random errors (cf. Chapter 1). For other score functions ψ, results regarding limit behavior under fixed as well as under local alternatives for the related non-ratio type statistic are presented in Hušková and Marušiaková (2012). The result for the ratio type statistic under fixed alternative may be derived by a modification of the proof therein.

3.7 Block bootstrap with replacement

In the following section, we are going to study only the case of $\psi_{L_2}(x) = x$. Extension to the case of general score function ψ from the previous sections is straightforward, but the proofs are much more complex.

There are several different approaches that may be used when resampling dependent observations. Classical resampling methods are not suitable, since they do not take into account the underlying dependency structure. Here, we focus our attention to a so-called *circular moving block bootstrap method*, which was introduced by Politis and Romano (1992). Overlapping blocks of consequent observations are formed from the original observations. The first few consequent observations from the original sequence are appended after the last observation, so that for a sequence of length n, we always have n possible blocks of subsequent observations to choose from

$$\{(Y_{j+1}, \dots, Y_{j+K}),\ j = 0, \dots, n-1\}; \quad \text{where } Y_i = Y_{i-n},\ i > n.$$

With this method, there is equal probability for each observation to be included in the bootstrap sample. For more details on the method, we also refer to Kirch (2006).

Let L denote the number of blocks and let K be the block length. In order to keep the notation as simple as possible, we restrict ourselves to situation, where $n = KL$, i.e., if the set of n observations can be divided in exactly L blocks of length K. It can be proved (Kirch, 2006) that the limit results remain the same after omitting the last K_1 observations, if $n = KL + K_1$, $1 \leqslant K_1 \leqslant K - 1$. We will assume that K and n are both functions of L

such that $n = KL$. Moreover, we will suppose that

$$L \to \infty \quad \text{and} \quad K \to \infty \quad \text{as} \quad n \to \infty.$$

It is also possible to use the non-circular moving block bootstrap, where one does not append the first few consequent observations from the original sequence after the last observation. This bootstrap version effectively gives $n - K$ blocks to choose from (instead of n blocks), but we will not concentrate on this approach here.

First, let us define the following subsets of $\mathbb{N} \times \mathbb{N}$ for integer numbers l, k, L, K and real number $0 < \gamma < 1/2$

$$\Pi_{l,k,L,K} = \{(p,q) : \ p,q \in \mathbb{N}, \\ 1 \leqslant p \leqslant l, \ 1 \leqslant q \leqslant K, \ (p-1)K + q \leqslant (l-1)K + k\},$$

$$\widetilde{\Pi}_{l,k,L,K} = \{(p,q) : \ p,q \in \mathbb{N}, \\ l \leqslant p \leqslant L, \ 1 \leqslant q \leqslant K, \ (p-1)K + q \geqslant (l-1)K + k + 1\},$$

$$\Omega_{L,K}(\gamma) = \{(l,k) : \ l,k \in \mathbb{N}, \\ 1 \leqslant l \leqslant L, \ 1 \leqslant k \leqslant K, \ KL\gamma \leqslant (l-1)K + k \leqslant KL(1-\gamma)\}.$$

For a set of iid random variables $\boldsymbol{U} = (U_1, \ldots, U_L)$, uniformly distributed on the set $\{0, \ldots, n-1\}$, we define the following block bootstrap statistic

$$S^{\boldsymbol{U}}_{L,K}(p,q,l,k) = \sum_{i=1}^{p-1} \sum_{j=1}^{K} \left(Y_{U_i+j} - m^{\boldsymbol{U}}_{L,K}(l,k) \right) + \sum_{j=1}^{q} \left(Y_{U_p+j} - m^{\boldsymbol{U}}_{L,K}(l,k) \right),$$

where

$$m^{\boldsymbol{U}}_{L,K}(l,k) = \frac{1}{(l-1)K + k} \left(\sum_{r=1}^{l-1} \sum_{s=1}^{K} Y_{U_r+s} + \sum_{s=1}^{k} Y_{U_l+s} \right)$$

for $p, l = 1, \ldots, L$, $q, k = 1, \ldots, K$, $p \leqslant l$, $(p-1)K + q \leqslant (l-1)K + k$. Similarly, we define

$$\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k) = \sum_{j=k+1}^{K} \left(Y_{U_l+j} - \tilde{m}^{\boldsymbol{U}}_{L,K}(l,k) \right) \mathcal{I}\{p \geqslant l+1\} \\ + \sum_{i=l+1}^{p-1} \sum_{j=1}^{K} \left(Y_{U_i+j} - \tilde{m}^{\boldsymbol{U}}_{L,K}(l,k) \right) \mathcal{I}\{p \geqslant l+2\} + \sum_{j=1}^{q} \left(Y_{U_p+j} - \tilde{m}^{\boldsymbol{U}}_{L,K}(l,k) \right),$$

where $\mathcal{I}\{A\}$ denotes the indicator of set A and

$$\tilde{m}^{\boldsymbol{U}}_{L,K}(l,k) = \frac{1}{(L-l+1)K-k}\left(\sum_{s=k+1}^{K} Y_{U_l+s} + \sum_{s=l+1}^{L}\sum_{r=1}^{K} Y_{U_r+s}\right)$$

for $p,l = 1,\ldots,L$, $q,k = 1,\ldots,K$ such that $p \geqslant l$, $(p-1)K+q \geqslant (l-1)K+k+1$.

Now, define the block bootstrap version of $\mathcal{A}_n(\psi_{L_2})$ from (3.4) by

$$\mathcal{A}^*_{L,K}(\psi_{L_2}) = \max_{(l,k)\in\Omega_{L,K}(\gamma)} \frac{\max_{(p,q)\in\Pi_{l,k,L,K}} \left|S^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \left|\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}.$$

Statistic $\mathcal{A}^*_{L,K}(\psi_{L_2})$ is constructed in a similar fashion as the original ratio type test statistic $\mathcal{A}_n(\psi_{L_2})$. The idea behind the bootstrap test statistic lies in indexing the randomly chosen (possibly overlapping) bootstrap blocks by $l = 1,\ldots,L$. The first l blocks are used in the nominator of the bootstrap statistic. The lth block is employed in the nominator as well as in the denominator. The last $L-l+1$ blocks are used in the denominator of the statistic $\mathcal{A}^*_{L,K}(\psi_{L_2})$. Regarding the lth block appearing in the nominator and denominator, this particular block is split into two continuous disjunctive parts: the first one contains the first element from the lth block, has up to k elements, and is used for the nominator; the second part contains the last elements from the lth block, has up to $K-k$ elements, and is used for the denominator. So, there does not exist an observation appearing simultaneously in the nominator and denominator.

An algorithm for the circular block bootstrap is illustratively shown in Procedure 3.1 and its validity will be proved in Theorem 3.3. We are going to show that the bootstrapped ratio type test statistic, conditioned on the original observations, has exactly the *same limit behavior* as the original test statistic under the null. It does not matter whether our observations come form the null hypothesis or the alternative. In other words, we are going to prove that $\mathcal{A}^*_{L,K}(\psi_{L_2})$ provides asymptotically correct critical values for the test based on $\mathcal{A}_n(\psi_{L_2})$, when observations follow either the null hypothesis or the alternative.

Theorem 3.3 (Bootstrap consistency)**.** *Suppose that $Y_1,\ldots,Y_n$ follow model* (3.1)*. Let $\mathsf{E}\,|\varepsilon_1|^\nu < \infty$ for some $\nu > 4$. Let Assumption A1 be satisfied for $\chi_1,\chi'_1 > 0$ and for $\chi_2,\chi'_2 > 0$ such that $2+2\kappa < \chi_1 < \nu-2$, $\chi'_1 = \nu-2-\chi_1$ and $0 < \chi_2 < (\chi_1-2-2\kappa)/(2+\kappa)$, $\chi'_2 = (\chi_1-2-2\kappa)/(2+\kappa)-\chi_2$ for some $0<\kappa<(\nu-4)/2$. Moreover, let Assumption D2 be satisfied, $K = O(L)$ as $L\to\infty$, and let*

$$K \leqslant L^{\chi_2/2-\epsilon} \tag{3.10}$$

for some $0<\epsilon<\frac{\chi_2}{2}$. Under alternative, let $\tau = [n\zeta]$ for some $\zeta:\ \gamma<\zeta<1-\gamma$. Then we

Procedure 3.1 Bootstrapping test statistic $\mathcal{A}_n(\psi_{L_2})$.

Input: Sequence of observations $Y_1, \ldots, Y_n$, block length K and $0 < \gamma < 1/2$.
Output: Bootstrap distribution of $\mathcal{A}_n(\psi_{L_2})$, i.e., the empirical distribution where probability mass $1/B$ concentrates at each of ${}_{(1)}\mathcal{A}^*_{L,K}(\psi_{L_2}), \ldots, {}_{(B)}\mathcal{A}^*_{L,K}(\psi_{L_2})$.

determine number of blocks $L = [n/K]$
define set $\Omega_{L,K}(\gamma)$
for $b = 1$ to B **do** // repeat in order to obtain the empirical distribution
  generate random sample $\boldsymbol{U} = (U_1, \ldots, U_L)$ from discrete uniform distribution on $\{0, \ldots, n-1\}$
  for $(l,k) \in \Omega_{L,K}(\gamma)$ **do**
    define sets $\Pi_{l,k,L,K}$ and $\widetilde{\Pi}_{l,k,L,K}$
    calculate ${}_{(b)}m^{\boldsymbol{U}}_{L,K}(l,k)$ and ${}_{(b)}\tilde{m}^{\boldsymbol{U}}_{L,K}(l,k)$
    for $(p,q) \in \Pi_{l,k,L,K}$ **do**
      calculate ${}_{(b)}S^{\boldsymbol{U}}_{L,K}(p,q,l,k)$
    end for
    compute $\max_{(p,q)\in\Pi_{l,k,L,K}} \left|{}_{(b)}S^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|$
    for $(p,q) \in \widetilde{\Pi}_{l,k,L,K}$ **do**
      calculate ${}_{(b)}\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)$
    end for
    compute $\max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \left|{}_{(b)}\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|$
  end for
  compute bootstrap test statistics ${}_{(b)}\mathcal{A}^*_{L,K}(\psi_{L_2})$
end for

have for all $y \in \mathbb{R}$, as $L \to \infty$,

$$\mathsf{P}\left(\mathcal{A}^*_{L,K}(\psi_{L_2}) \leqslant y | Y_1, \ldots, Y_n\right) \xrightarrow{a.s.} \mathsf{P}\left(\sup_{\gamma\leqslant t\leqslant 1-\gamma} \frac{\sup_{0\leqslant u\leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|} \leqslant y\right),$$

where $\{\mathcal{W}(u), 0 \leqslant u \leqslant 1\}$ is a standard Wiener process and $\widetilde{\mathcal{W}}(u) = \mathcal{W}(1) - \mathcal{W}(u)$.

Proof. The proof goes along the lines of proof of Theorem 3.6.2 in Kirch (2006) for the case of $q(t) = 1$, $t \in (0,1)$ and uses several results derived there. In contrast to Kirch (2006), we drop the assumption of random errors being a linear process, because this is only assumed in order to show that the original (not bootstrapped) statistic weakly converges under the null hypothesis. Assumptions of Theorem 3.3 and null hypothesis (3.2) provide us the asymptotic distribution of $\mathcal{A}_n(\psi_{L_2})$ from Theorem 3.1.

Note that we have (possibly) overlapping blocks. If $\{\varepsilon_i, i \in \mathbb{N}\}$ is an α-mixing sequence, then

$$\left\{ \frac{1}{\sqrt{K}} \left(\sum_{j=1}^{n-r} \varepsilon_{r+j} + \sum_{j=1}^{K-(n-r)} \varepsilon_j \right), n \in \mathbb{N} \right\}$$

is also α-mixing for all $K \geqslant 2$ and $r = n-K+1, \ldots, n-1$, but with smaller or equal mixing coefficients than $\{\varepsilon_i, i \in \mathbb{N}\}$ (Bradley, 2005, Theorem 5.2). Hence, assumption (3.6) is uniformly fulfilled in K and r. Moreover, the above considered sequence remains stationary for stationary $\{\varepsilon_i, i \in \mathbb{N}\}$.

Theorem 1 by Yokoyama (1980) says that there exists a constant $D = D(\boldsymbol{\alpha}, \chi, \chi', C_1) > 0$ depending only on the constants χ, χ', sequence of α-mixing coefficients $\boldsymbol{\alpha} \equiv \{\alpha(k)\}_{k \in \mathbb{N}}$, and $C_1(\chi, \chi')$, such that

$$\mathsf{E} \left| \frac{1}{\sqrt{K}} \left(\sum_{j=1}^{n-r} \varepsilon_{r+j} + \sum_{j=1}^{K-(n-r)} \varepsilon_j \right) \right|^{2+\chi_1} \leqslant D, \quad a.s.,$$

uniformly in K and r. Then according to the de la Vallée-Poussin theorem (Meyer, 1966, p. 19, Theorem T22),

$$\frac{1}{n} \sum_{r=0}^{n-1} \max_{k=0,\ldots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} \varepsilon_{r+j} \right|^{2+\kappa} \leqslant \mu_{2+\kappa} + o(1) \leqslant D_2, \quad a.s.,$$

for some constant $D_2 > 0$, where here $\varepsilon_j = \varepsilon_{j-n}$ for $j > n$ and

$$\mu_{2+\kappa} := \mathsf{E} \max_{k=0,\ldots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} \varepsilon_{r+j} \right|^{2+\kappa}$$

uniformly in K and r.

The Markov inequality and Yokoyama (1980, Theorem 1) give for any $\omega > 0$

$$\begin{aligned} &\mathsf{P} \left[\frac{1}{n} \sum_{j=1}^{n} |Z_j| \geqslant \omega \right] = \mathsf{P} \left[\sum_{k=1}^{K} \frac{1}{L} \sum_{l=0}^{L-1} |Z_{Kl+k}| \geqslant K\omega \right] \\ &\leqslant \mathsf{P} \left[\max_{k=1,\ldots,K} \frac{1}{L} \sum_{l=0}^{L-1} |Z_{Kl+k}| \geqslant \omega \right] \leqslant \sum_{k=1}^{K} \mathsf{P} \left[\frac{1}{L} \sum_{l=0}^{L-1} |Z_{Kl+k}| \geqslant \omega \right] \\ &= O\left(\frac{1}{\omega^{2+\chi_2}} \frac{K}{L^{1+\chi_2/2}} \right), \quad L \to \infty, \end{aligned} \tag{3.11}$$

where

$$Z_s := \max_{k=0,\ldots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} \varepsilon_{s-1+j} \right|^{2+\kappa} - \mu_{2+\kappa}.$$

By condition (3.10), we have

$$\sum_{L=1}^{\infty} \frac{K}{L^{1+\chi_2/2}} \leqslant \sum_{L=1}^{\infty} \frac{1}{L^{1+\epsilon}} < \infty.$$

This yields that (3.11) converges sufficiently fast to zero, which implies convergence almost surely. Consequently,

$$\frac{1}{n}\sum_{s=0}^{n-1}\left(\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\varepsilon_{s+j}\right|^{2+\kappa} - \mu_{2+\kappa}\right) + \mu_{2+\kappa} \leqslant D_2, \quad a.s.$$

This is going to be used for verifying the assumptions of Theorem 3.6.1 in Kirch (2006), i.e.,

$$\frac{1}{n}\sum_{i=0}^{n-1}\left(\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n(i+j)-\bar{a}_n)\right|^{\rho}\right) \leqslant D_1, \quad a.s., \tag{3.12}$$

for some $2 < \rho \leqslant 4$ with appropriately chosen scores

$$\boldsymbol{a}_n = (a_n(1), \dots, a_n(n)),$$

such that $\bar{a}_n = \frac{1}{n}\sum_{i=1}^{n} a_n(i)$.

Let us consider three different situations and for each of them we choose the appropriate scores $\boldsymbol{a}_n$:

(i) $K\delta_n^2 = O(1)$. This case also includes the null hypothesis (with $\delta_n = 0$). We choose $a_n(i) = Y_i$.

(ii) $1/(K\delta_n^2) = O(1)$. In this case, we let $a_n(i) = Y_i/(\sqrt{K}\delta_n)$.

(iii) Both $K\delta_n^2 \leqslant 1$ and $K\delta_n^2 > 1$ is true for infinitely many $n \in \mathbb{N}$. Note that $K \equiv K_n$. In this case, we use a combination of both score choices, i.e.,

$$a_n(i) = \begin{cases} Y_i, & K\delta_n^2 \leqslant 1; \\ \frac{Y_i}{\sqrt{K}\delta_n}, & K\delta_n^2 > 1. \end{cases}$$

The proof of Theorem 3.6.2 and Theorem 3.5.1 by Kirch (2006) provide the verification of (3.12).

Using the above chosen scores $\boldsymbol{a}_n$, the proof of Theorem 3.6.1 (c) (or Theorem 3.4.1 (c)) for the nominator of the considered statistic gives that conditionally on the observations

$Y_1, \ldots, Y_n$

$$\max_{(p,q)\in\Pi_{l,k,L,K}} \frac{\left|S^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\sigma(\psi_{L_2})\sqrt{LK}} \overset{\mathscr{D}}{=} \max_{(p,q)\in\Pi_{l,k,L,K}} \left|\mathcal{W}\left(\frac{p}{L}\right) - \frac{p}{l}\mathcal{W}\left(\frac{l}{L}\right)\right| + o_{\mathsf{P}}(1), \quad n \to \infty. \quad (3.13)$$

Here, we understand that $V_n = o_{\mathsf{P}}(1)$ as $n \to \infty$, conditionally on the $Y_1, \ldots, Y_n$, if

$$\forall \phi > 0 : \mathsf{P}\left(|V_n| \geqslant \phi | Y_1, \ldots, Y_n\right) \xrightarrow[n\to\infty]{a.s.} 0.$$

Sometimes one writes $V_n = o_{\mathsf{P}^*}(1)$ as $n \to \infty$, where $\mathsf{P}^*(\cdot) \equiv \mathsf{P}(\cdot|Y_1, \ldots, Y_n)$, but this is not necessary here in this proof, since the meaning is clear from the context. Moreover, the same proof gives (conditionally on $Y_1, \ldots, Y_n$)

$$\begin{aligned}
&\max_{(p,q)\in\Pi_{l,k,L,K}} \left|\mathcal{W}\left(\frac{p}{L}\right) - \frac{p}{l}\mathcal{W}\left(\frac{l}{L}\right)\right| \\
&= \max_{(p,q)\in\Pi_{l,k,L,K}} \left|\mathcal{W}\left(\frac{(p-1)K+q}{KL}\right) \right. \\
&\qquad \left. - \frac{(p-1)K+q}{(l-1)K+k}\mathcal{W}\left(\frac{(l-1)K+k}{KL}\right)\right| + o_{\mathsf{P}}(1), \quad n \to \infty. \quad (3.14)
\end{aligned}$$

Since $l = [Lt]$ for some $t \in [0,1]$, equations (3.13) and (3.14) imply that it suffices to consider the asymptotic behavior of

$$\max_{1\leqslant i\leqslant nt} \frac{1}{\sqrt{n}} \left|\sum_{1\leqslant j\leqslant i} \left(X_j - \bar{X}_{[nt]}\right)\right|,$$

where $\{X_j,\, j \in \mathbb{N}\}$ are iid standard normal variables and $\bar{X}_i = \frac{1}{i}\sum_{1\leqslant j\leqslant i} X_j$. I.e., conditionally on $Y_1, \ldots, Y_n$,

$$\max_{(p,q)\in\Pi_{l,k,L,K}} \frac{\left|S^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\sigma(\psi_{L_2})\sqrt{LK}} \overset{\mathscr{D}}{=} \max_{1\leqslant i\leqslant nt} \frac{1}{\sqrt{n}} \left|\sum_{1\leqslant j\leqslant i} \left(X_j - \bar{X}_{[nt]}\right)\right| + o_{\mathsf{P}}(1), \quad n \to \infty$$

for all $t \in [\gamma, 1-\gamma]$.

Similarly, conditionally on $Y_1, \ldots, Y_n$, it holds that

$$\max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \frac{\left|\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\sigma(\psi_{L_2})\sqrt{LK}} \overset{\mathscr{D}}{=} \max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \left|\widetilde{\mathcal{W}}\left(\frac{p}{L}\right) - \frac{L-p}{L-l}\widetilde{\mathcal{W}}\left(\frac{l}{L}\right)\right| + o_{\mathsf{P}}(1), \quad n \to \infty$$

and

$$\begin{aligned}
&\max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \left| \widetilde{\mathcal{W}}\left(\frac{p}{L}\right) - \frac{L-p}{L-l}\widetilde{\mathcal{W}}\left(\frac{l}{L}\right) \right| \\
&= \max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \left| \widetilde{\mathcal{W}}\left(\frac{(p-1)K+q}{KL}\right) \right. \\
&\qquad \left. - \frac{L-((p-1)K+q)}{L-((l-1)K+k+1)}\widetilde{\mathcal{W}}\left(\frac{(l-1)K+k+1}{KL}\right) \right| + o_{\mathsf{P}}(1), \quad n\to\infty.
\end{aligned}$$

Thus, conditionally on $Y_1,\ldots,Y_n$,

$$\max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \frac{\left|\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\sigma(\psi_{L_2})\sqrt{LK}} \overset{\mathscr{D}}{=} \max_{nt<i\leqslant n} \frac{1}{\sqrt{n}} \left| \sum_{i\leqslant j\leqslant n} \left(X_j - \widetilde{X}_{[nt]}\right) \right| + o_{\mathsf{P}}(1), \quad n\to\infty$$

for all $t\in[\gamma,1-\gamma]$, where $\widetilde{X}_i = \frac{1}{n-i}\sum_{i<j\leqslant n} X_j$.

Now, the assumptions of Theorem 3.1 are satisfied for $\{X_j,\ j\in\mathbb{N}\}$ under the null hypothesis as well as under the alternative. Hence, we get that along almost all samples $Y_1,\ldots,Y_n$, it holds that

$$\begin{aligned}
&\left(\max_{(p,q)\in\Pi_{l,k,L,K}} \frac{\left|S^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\sigma(\psi_{L_2})\sqrt{LK}}, \max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}} \frac{\left|\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\sigma(\psi_{L_2})\sqrt{LK}} \right) \\
&\qquad \xrightarrow[n\to\infty]{\mathscr{D}^2[\gamma,1-\gamma]} \left(\sup_{0\leqslant u\leqslant t} \left|\mathcal{W}(u) - u/t\mathcal{W}(t)\right|, \sup_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right| \right),
\end{aligned}$$

conditionally on $Y_1,\ldots,Y_n$.

Finally, the assertion of Theorem 3.3 is straightforward, since the considered bootstrap statistic is a continuous function of the above vector of statistics. □

Remark 3.1. Condition (3.10) can be omitted, but the almost sure convergence in the assertion of Theorem 3.3 needs to be changed into convergence in probability (Kirch, 2006, Theorem 3.6.2). After that, the condition $\nu > 4$ can be relaxed to $\nu > 2$ (Kirch, 2006, Remark 3.5.4), although the remaining assumptions concerning ν have to be changed accordingly.

A choice of the block length L in the circular moving block bootstrap is an important decision. It will affect the bootstrapped version of the test statistic. Therefore, the block length can be viewed as a tuning parameter in the circular moving block bootstrap procedure. One possibility, how to make such optimal choice, is to minimize the asymptotic mean square error (MSE) of the circular moving block bootstrap variance estimate. Fitzenberger (1997) proved that this approach yields $K = \mathcal{O}(n^{1/3})$ as $n\to\infty$ in case of α-mixing random errors. In contrast to this asymptotic result, the practical choice of the block length usually needs to be made based on one finite sample consisting of n observations. Several finite sample approaches for choosing the block length K were proposed by Hall et al. (1995), Politis and White (2004), and Lahiri et al. (2007).

3.8 Modification of the test statistic

As it will be further seen in simulations in Subsection 3.9, α-errors (i.e., probabilities of the first type error) are sometimes not sufficiently close to the theoretical α-errors for test procedures based on the ratio type test statistic $\mathcal{A}_n(\psi)$. Therefore, based on the numerical results, we suggest a *modification of the original statistic* $\mathcal{A}_n(\psi)$ in the following form

$$\widetilde{\mathcal{A}}_n(\psi) = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \sqrt{\frac{n-k}{k}} \frac{\max_{1\leqslant i\leqslant k} \left| \sum_{1\leqslant j\leqslant i} \psi(Y_j - \widehat{\mu}_{1k}(\psi)) \right|}{\max_{k\leqslant i\leqslant n-1} \left| \sum_{i+1\leqslant j\leqslant n} \psi(Y_j - \widehat{\mu}_{2k}(\psi)) \right|}, \tag{3.15}$$

where $0 < \gamma < 1/2$ is a given constant. The way, how the original test statistics is modified, is that the ratio in the original test statistic is standardized by $\sqrt{(n-k)/k}$. This makes the ratio more constant (flatter) with respect to k and, hence, the test based on the modified test statistic keeps the theoretical α-level better (more firmly).

Similarly as in Theorem 3.1, one can derive the asymptotic behavior of the modified test statistics under the null hypothesis.

Theorem 3.4. *Suppose that $Y_1, \ldots, Y_n$ follow model* (3.1) *and assume that Assumptions A1–A5 hold. Then, under null hypothesis* (3.2)

$$\widetilde{\mathcal{A}}_n(\psi) \xrightarrow[n\to\infty]{\mathscr{D}} \sup_{\gamma\leqslant t\leqslant 1-\gamma} \left(\frac{1-t}{t}\right)^{1/2} \frac{\sup_{0\leqslant u\leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|}, \tag{3.16}$$

where $\{\mathcal{W}(u), 0 \leqslant u \leqslant 1\}$ is a standard Wiener process and $\widetilde{\mathcal{W}}(u) = \mathcal{W}(1) - \mathcal{W}(u)$.

Proof. See proof of Theorem 3.1. □

A theoretical argument for the modification suggested above comes from distributional relation (1.9) and asymptotic distribution (3.16). The asymptotic distribution of the original test statistic $\mathcal{A}_n(\psi)$ from (3.8) can be consequently rewritten as

$$\sup_{\gamma\leqslant t\leqslant 1-\gamma} \frac{\sup_{0\leqslant u\leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|} \stackrel{\mathscr{D}}{=} \sup_{\gamma\leqslant t\leqslant 1-\gamma} \left(\frac{t}{1-t}\right)^{1/2} \frac{\sup_{0\leqslant s\leqslant 1} |\mathcal{B}(s)|}{\sup_{0\leqslant s\leqslant 1} \left|\widetilde{\mathcal{B}}(s)\right|},$$

where $\{\mathcal{B}(s),\, s \in [0,1]\}$ and $\{\widetilde{\mathcal{B}}(s),\, s \in [0,1]\}$ are two independent Brownian bridges. Function $\sqrt{t/(1-t)}$ is strictly increasing, which offers the idea of multiplying asymptotic distribution of the original test statistic $\mathcal{A}_n(\psi)$ by the reciprocal function, i.e., $\sqrt{(1-t)/t}$. Thus, the asymptotic distribution of the original test statistic $\widetilde{\mathcal{A}}_n(\psi)$ from (3.16) can simply be rewritten as

$$\frac{\sup_{0\leqslant s\leqslant 1}|\mathcal{B}(s)|}{\sup_{0\leqslant s\leqslant 1}\left|\widetilde{\mathcal{B}}(s)\right|}.$$

This leads to our modification of the original test statistic, where the ratio of maxima in $\mathcal{A}_n(\psi)$ is multiplied by $\sqrt{\frac{1-k/n}{k/n}} = \sqrt{(n-k)/k}$.

The asymptotic behavior of the modified test statistic under the alternative can be derived as well.

Theorem 3.5. *Suppose that $Y_1,\ldots,Y_n$ follow model* (3.1)*, assume that $\sqrt{n}|\delta| \to \infty$ as $n\to\infty$, and $\tau = [\zeta n]$ for some $\gamma < \zeta < 1-\gamma$. Then, under Assumptions A1–A5 and alternative* (3.3)

$$\widetilde{\mathcal{A}}_n(\psi) \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Proof. See proof of Theorem 3.2. □

The block bootstrap version of the modified ratio type test statistic $\widetilde{\mathcal{A}}_n(\psi_{L_2})$ can be defined as

$$\widetilde{\mathcal{A}}^{*}_{L,K}(\psi_{L_2}) = \max_{(l,k)\in\Omega_{L,K}(\gamma)} \sqrt{\frac{KL-((l-1)K+k)}{(l-1)K+k}}\,\frac{\max_{(p,q)\in\Pi_{l,k,L,K}}\left|S^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}{\max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}}\left|\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|}.$$

The block bootstrap version of the modified statistic $\widetilde{\mathcal{A}}^{*}_{L,K}(\psi_{L_2})$ also provides asymptotically correct critical values for the test based on $\widetilde{\mathcal{A}}_n(\psi_{L_2})$, when observations follow either the null hypothesis or the alternative as it is formally stated in the following theorem.

Theorem 3.6. *Suppose that $Y_1,\ldots,Y_n$ follow model* (3.1)*. Let $\mathsf{E}\,|\varepsilon_1|^\nu < \infty$ for some $\nu > 4$. Let Assumption A1 be satisfied for $\chi_1,\chi'_1 > 0$ and for $\chi_2,\chi'_2 > 0$ such that $2+2\kappa < \chi_1 < \nu-2$, $\chi'_1 = \nu-2-\chi_1$ and $0<\chi_2<(\chi_1-2-2\kappa)/(2+\kappa)$, $\chi'_2 = (\chi_1-2-2\kappa)/(2+\kappa)-\chi_2$ for some $0<\kappa<(\nu-4)/2$. Moreover, let Assumption D2 be satisfied, $K = O(L)$ as $L\to\infty$, and let*

$$K \leqslant L^{\chi_2/2-\epsilon} \tag{3.17}$$

for some $0<\epsilon<\frac{\chi_2}{2}$. Under alternative, let $\tau = [n\zeta]$ for some $\zeta:\ \gamma<\zeta<1-\gamma$. Then we *have for all $y \in \mathbb{R}$, as $L \to \infty$,*

$$\mathsf{P}\left(\widetilde{\mathcal{A}}^{*}_{L,K}(\psi_{L_2}) \leqslant y | Y_1, \ldots, Y_n\right) \xrightarrow{a.s.} \mathsf{P}\left(\sup_{\gamma \leqslant t \leqslant 1-\gamma} \left(\frac{1-t}{t}\right)^{1/2} \frac{\sup_{0 \leqslant u \leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup_{t \leqslant u \leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|} \leqslant y \right),$$

where $\{\mathcal{W}(u), 0 \leqslant u \leqslant 1\}$ is a standard Wiener process and $\widetilde{\mathcal{W}}(u) = \mathcal{W}(1) - \mathcal{W}(u)$.

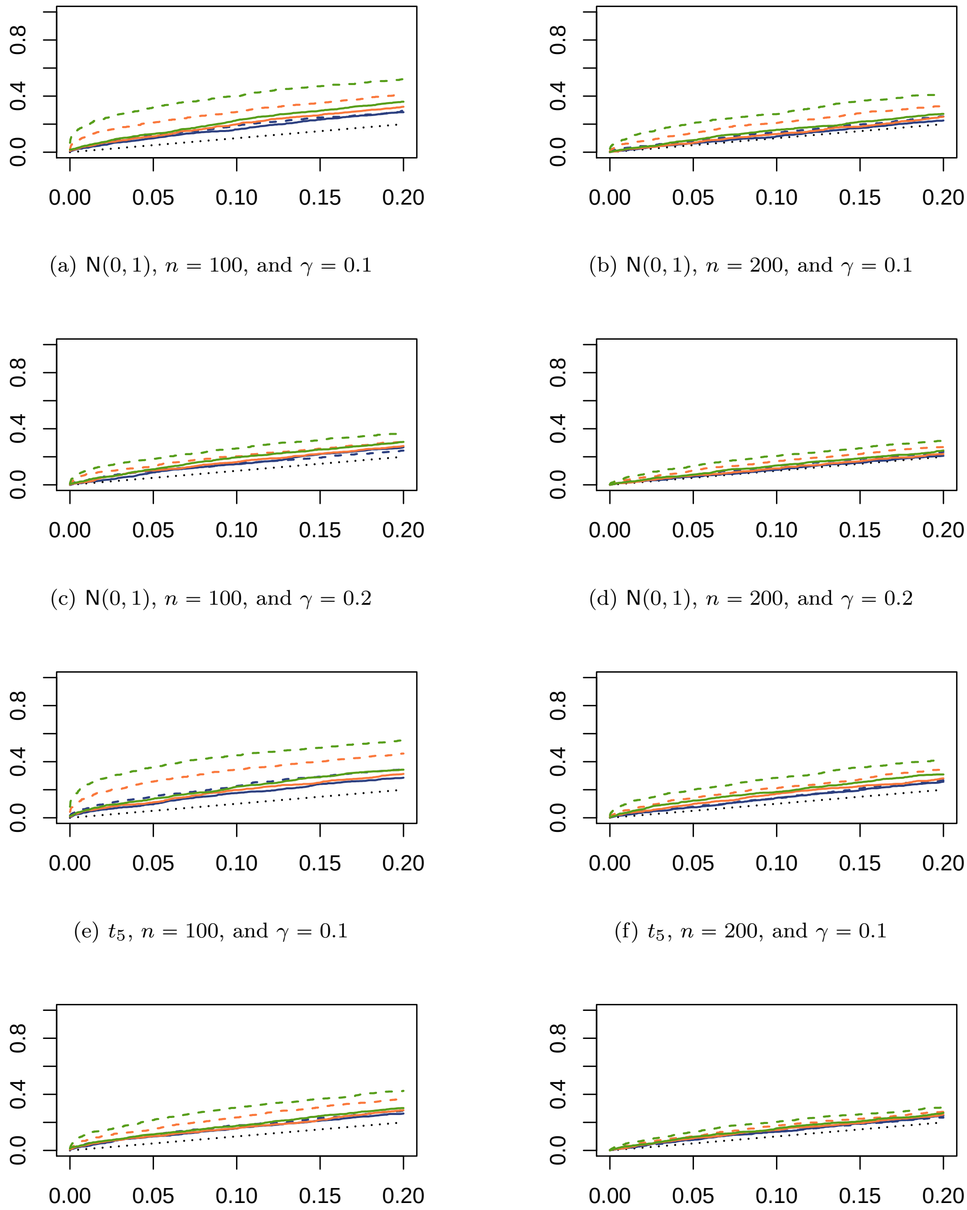
Proof. See the proof of Theorem 3.3. □

3.9 Simulations

We were interested in the performance of the tests based on the ratio type test statistics $\mathcal{A}_n(\psi)$ and $\widetilde{\mathcal{A}}_n(\psi)$ with $\psi_{L_2}(x) = x$ and $\psi_{L_1}(x) = \text{sgn}(x)$. We focused on comparison of the accuracy of critical values obtained by circular moving block bootstrap method with the accuracy of critical values obtained by simulation from the limit distribution. Some simulation results concerning the test based on asymptotic critical values for the studied type of test statistic can be also found in Horváth et al. (2008) (the L_2 method and asymptotic critical values) and Madurkayová (2009b).

In Figures 3.1–3.8, one may see *size-power plots* or size-power curves (SPC) for choices of $n = 100$ or 200, $\gamma = 0.1$ or 0.2, $\tau = 0.5$, and $\delta = 1$ considering test statistics $\mathcal{A}_n(\psi)$ and $\widetilde{\mathcal{A}}_n(\psi)$ in case of L_2 and L_1 score function both under the null hypothesis and under the alternative. The size-power plots illustrate the power of a test. The empirical distribution function of the p-values of the test statistic for the null hypothesis or a given alternative is plotted with respect to the distribution used to determine the critical values of the test. What we get is a plot that shows the actual α-errors resp. $1-(\beta$-errors) on the y-axis for the chosen quantiles on the x-axis. The ideal situation under the null hypothesis is depicted by the straight diagonal dotted line. Under the alternative, the desired situation would be a steep function with values close to 1. For more details on size-power plots we may refer, e.g., to Kirch (2006). The random errors were simulated as an AR(1) process with autoregression coefficients 0.3 (orange) and 0.5 (green), and as a set of iid random errors (blue). Rejection rates based on simulated asymptotic critical values are depicted by a dashed line, rejection rates based on block bootstrap with block length $K = 5$ are depicted by a solid line. Standard normal distribution and Student t-distribution with 5 degrees of freedom are used for generating the innovations of models' errors.

We generated 10000 independent samples in order to compute the asymptotic critical values. When bootstrapping, for each sample we used 1000 bootstrap samples to compute the bootstrap critical values. In simulations of rejection rates, we used 1000 repetitions.



(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$

(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$

(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$

(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$

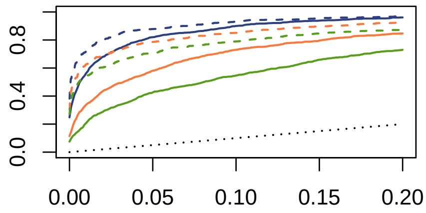
(e) t_5, $n = 100$, and $\gamma = 0.1$

(f) t_5, $n = 200$, and $\gamma = 0.1$

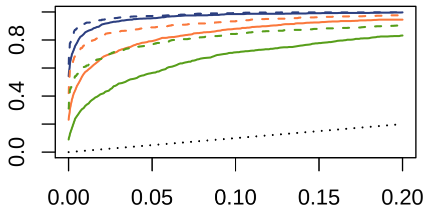
(g) t_5, $n = 100$, and $\gamma = 0.2$

(h) t_5, $n = 200$, and $\gamma = 0.2$

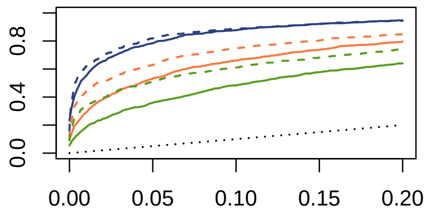
Figure 3.1: Size-power plots for $\mathcal{A}_n(\psi_{L_2})$ under null hypothesis H_0. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.



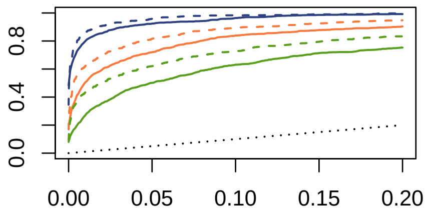
(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$



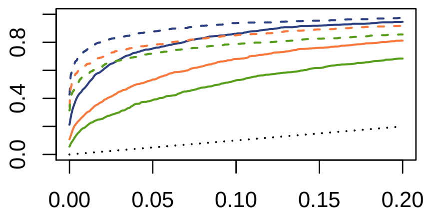
(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$



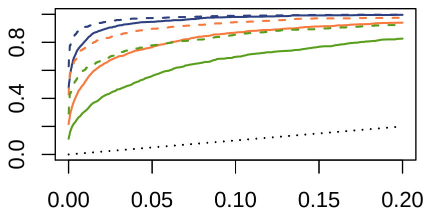
(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$



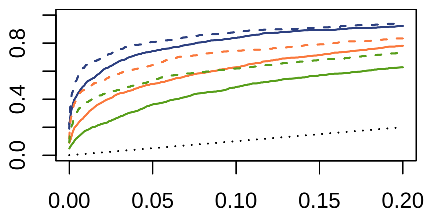
(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$



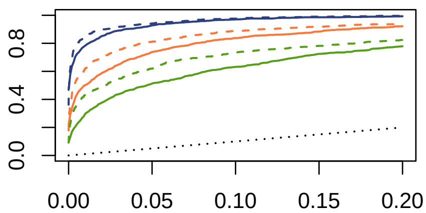
(e) t_5, $n = 100$, and $\gamma = 0.1$



(f) t_5, $n = 200$, and $\gamma = 0.1$

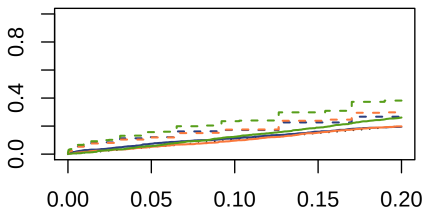


(g) t_5, $n = 100$, and $\gamma = 0.2$

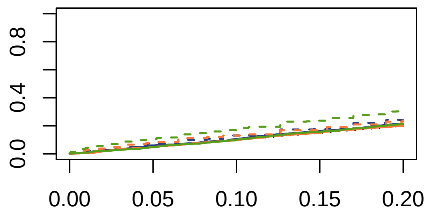


(h) t_5, $n = 200$, and $\gamma = 0.2$

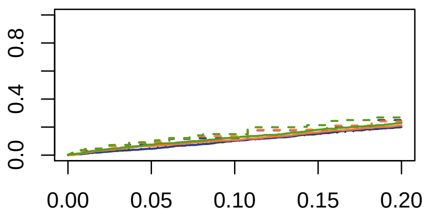
Figure 3.2: Size-power plots for $\mathcal{A}_n(\psi_{L_2})$ under alternative H_1. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.



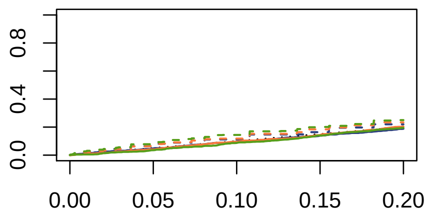
(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$



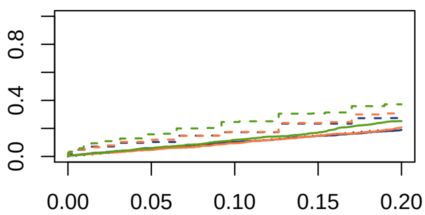
(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$



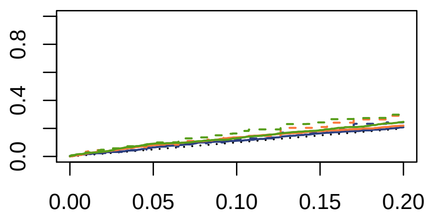
(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$



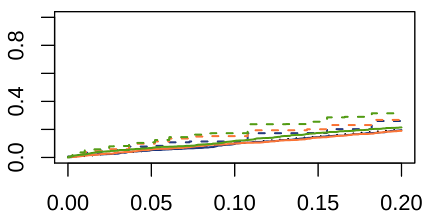
(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$



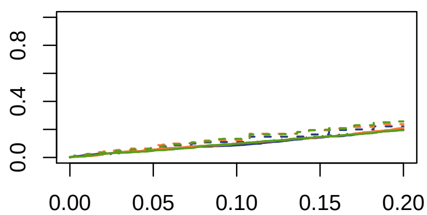
(e) t_5, $n = 100$, and $\gamma = 0.1$



(f) t_5, $n = 200$, and $\gamma = 0.1$

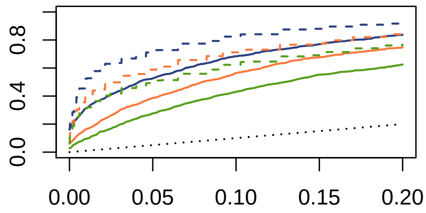


(g) t_5, $n = 100$, and $\gamma = 0.2$

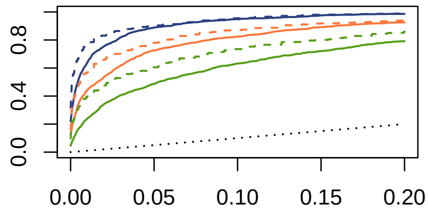


(h) t_5, $n = 200$, and $\gamma = 0.2$

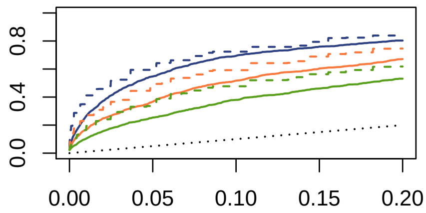
Figure 3.3: Size-power plots for $\mathcal{A}_n(\psi_{L_1})$ under null hypothesis H_0. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.



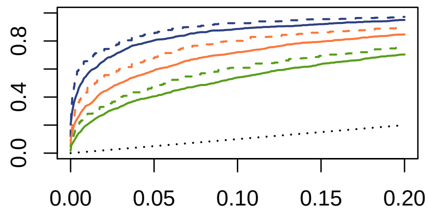
(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$



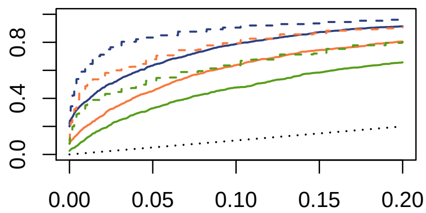
(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$



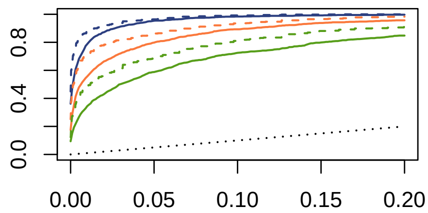
(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$



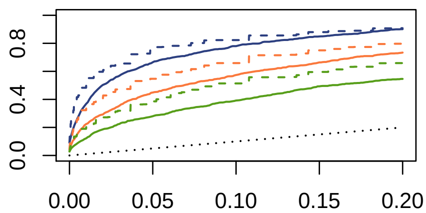
(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$



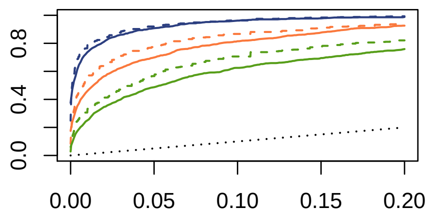
(e) t_5, $n = 100$, and $\gamma = 0.1$



(f) t_5, $n = 200$, and $\gamma = 0.1$

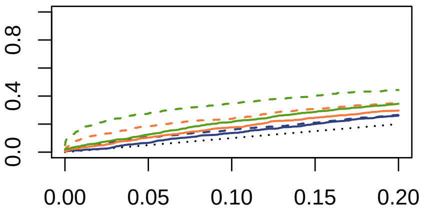


(g) t_5, $n = 100$, and $\gamma = 0.2$

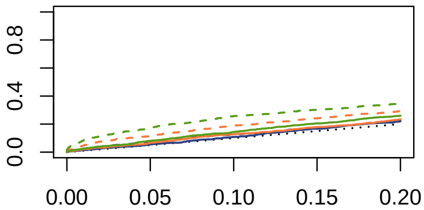


(h) t_5, $n = 200$, and $\gamma = 0.2$

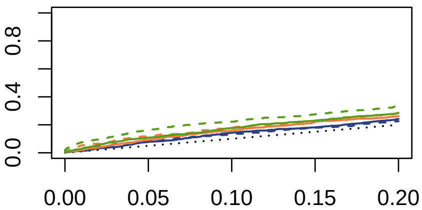
Figure 3.4: Size-power plots for $\mathcal{A}_n(\psi_{L_1})$ under alternative H_1. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.



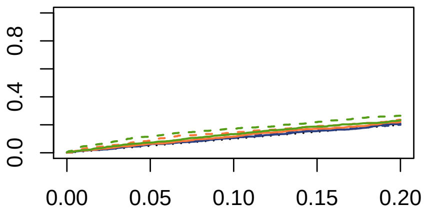
(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$



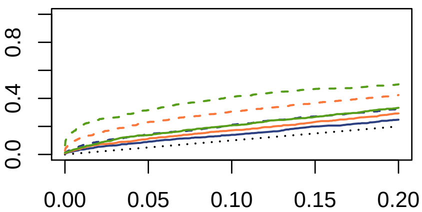
(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$



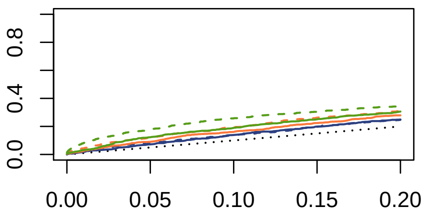
(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$



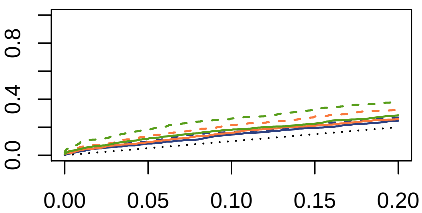
(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$



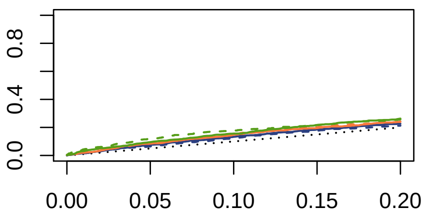
(e) t_5, $n = 100$, and $\gamma = 0.1$



(f) t_5, $n = 200$, and $\gamma = 0.1$

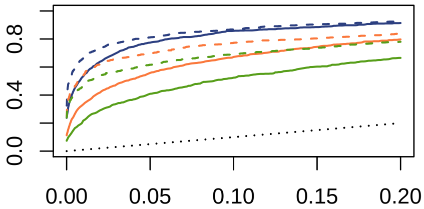


(g) t_5, $n = 100$, and $\gamma = 0.2$

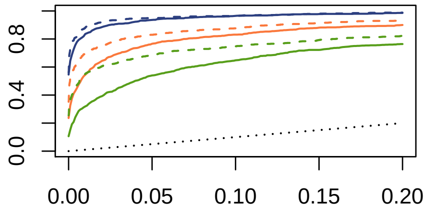


(h) t_5, $n = 200$, and $\gamma = 0.2$

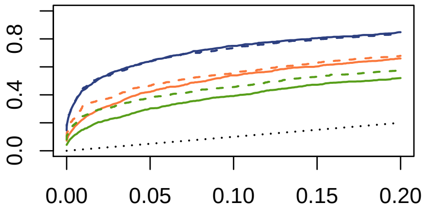
Figure 3.5: Size-power plots for $\widetilde{\mathcal{A}}_n(\psi_{L_2})$ under null hypothesis H_0. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.



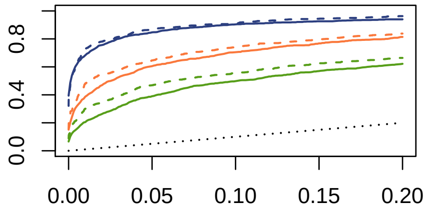
(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$



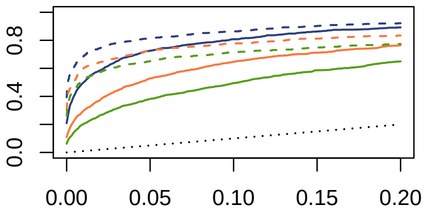
(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$



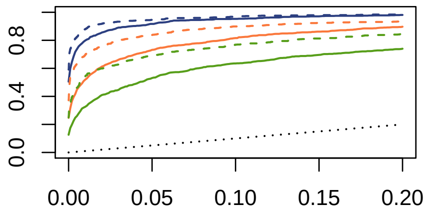
(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$



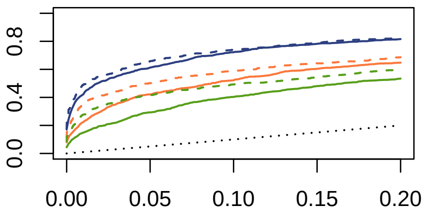
(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$



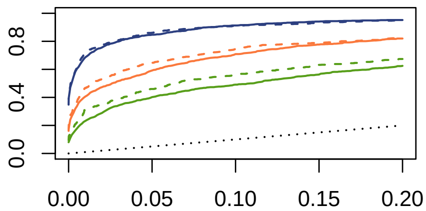
(e) t_5, $n = 100$, and $\gamma = 0.1$



(f) t_5, $n = 200$, and $\gamma = 0.1$

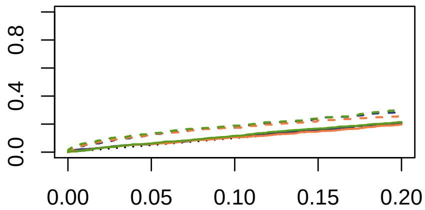


(g) t_5, $n = 100$, and $\gamma = 0.2$

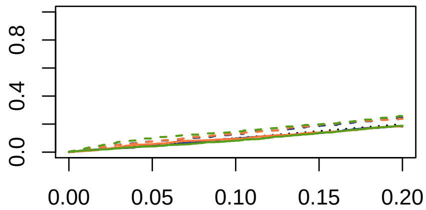


(h) t_5, $n = 200$, and $\gamma = 0.2$

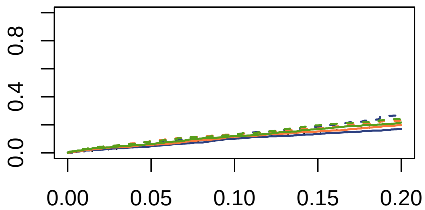
Figure 3.6: Size-power plots for $\widetilde{\mathcal{A}}_n(\psi_{L_2})$ under alternative H_1. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.



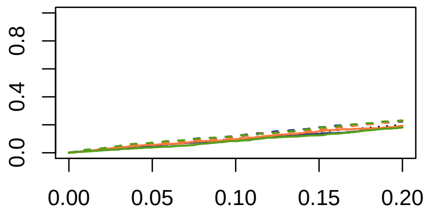
(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$



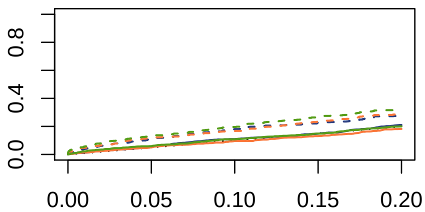
(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$



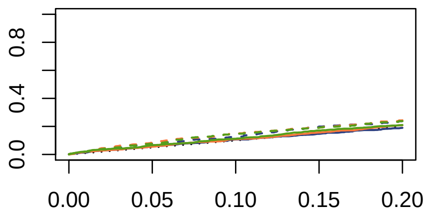
(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$



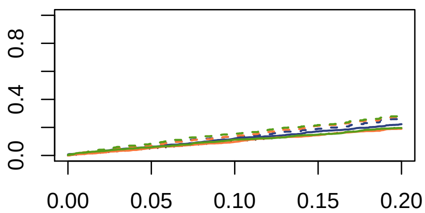
(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$



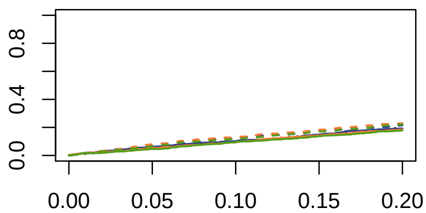
(e) t_5, $n = 100$, and $\gamma = 0.1$



(f) t_5, $n = 200$, and $\gamma = 0.1$

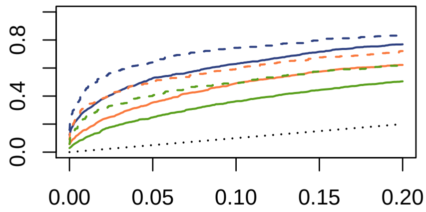


(g) t_5, $n = 100$, and $\gamma = 0.2$

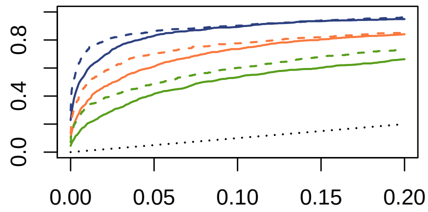


(h) t_5, $n = 200$, and $\gamma = 0.2$

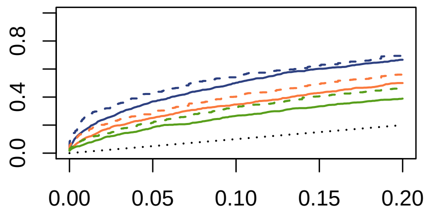
Figure 3.7: Size-power plots for $\widetilde{\mathcal{A}}_n(\psi_{L_1})$ under null hypothesis H_0. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.



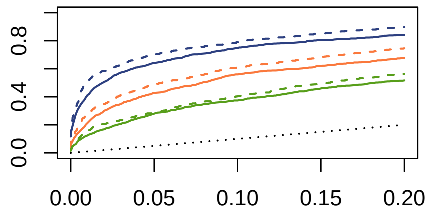
(a) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.1$



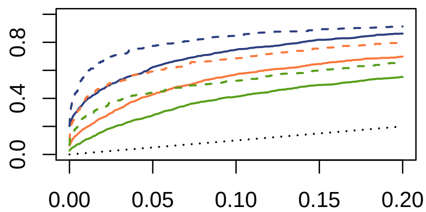
(b) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.1$



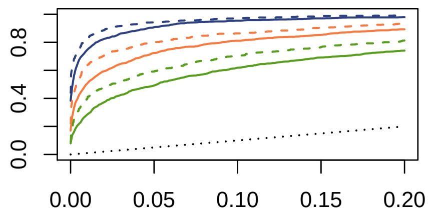
(c) $\mathsf{N}(0,1)$, $n = 100$, and $\gamma = 0.2$



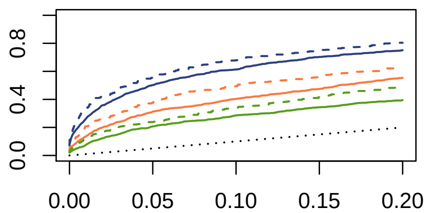
(d) $\mathsf{N}(0,1)$, $n = 200$, and $\gamma = 0.2$



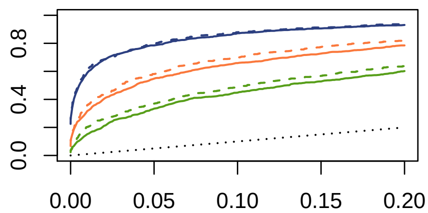
(e) t_5, $n = 100$, and $\gamma = 0.1$



(f) t_5, $n = 200$, and $\gamma = 0.1$



(g) t_5, $n = 100$, and $\gamma = 0.2$



(h) t_5, $n = 200$, and $\gamma = 0.2$

Figure 3.8: Size-power plots for $\widetilde{\mathcal{A}}_n(\psi_{L_1})$ under alternative H_1. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid.

In all of 64 figures depicting a situation under the null hypothesis, we may see that comparing to the critical values obtained by simulations from the asymptotic distribution, the critical values obtained by bootstrapping are more accurate, especially for AR(1) sequences. When comparing the accuracy of α-errors for different choices of score function ψ, the L_1 method seems to perform better than the L_2 method. However, when using the L_1 method, power of the test slightly decreases, as we may also see in Table 3.1. Similarly, the choice of $\gamma = 0.2$ seems to provide more accurate critical values than the choice of $\gamma = 0.1$, but the test power is larger in the latter case.

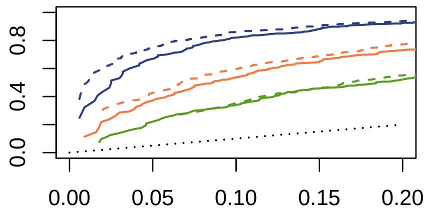
	Asymptotics				Bootstrap			
	L_2 statistic		L_1 statistic		L_2 statistic		L_1 statistic	
	$\mathsf{N}(0,1)$	t_5	$\mathsf{N}(0,1)$	t_5	$\mathsf{N}(0,1)$	t_5	$\mathsf{N}(0,1)$	t_5
$\varphi = 0$	0.924	0.906	0.769	0.866	0.922	0.925	0.801	0.906
$\varphi = 0.3$	0.726	0.729	0.577	0.682	0.718	0.733	0.585	0.696
$\varphi = 0.5$	0.531	0.523	0.365	0.448	0.500	0.508	0.404	0.485

Table 3.1: Simulated rejection rates for the abrupt change in mean based on the asymptotic and bootstrap version of $\mathcal{A}_n(\psi)$ for the L_1 and L_2 method with $n = 200$, $\gamma = 0.2$ under alternative H_1 with $\delta = 1$ and $\tau = n/2$, considering significance level $\alpha = 0.05$. Random errors were simulated either as $\mathsf{N}(0,1)$ or as t_5 distributed AR(1) sequences with several values of autoregression coefficient φ.

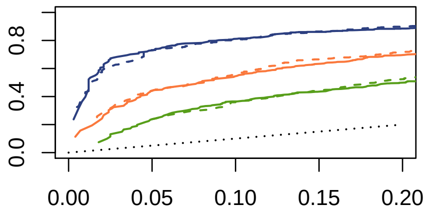
Furthermore, with the choice of ψ_{L_2}, the simulated rejection rates under H_0 are higher than the corresponding theoretical α-levels for larger values of the autoregression coefficient, while for the L_1 method they remain much more stable. Comparing the case of $\mathsf{N}(0,1)$ innovations with the case of t_5 innovations, rejection rates for the L_1 version of the test statistic tend to be slightly higher for t_5 distribution, while they remain more or less the same for the L_2 version. As expected, the accuracy of the critical values tends to be better for larger n.

On one hand, simulations showed that the actual α-errors are closer to the theoretical α-errors (significance levels) for the test procedures based on $\widetilde{\mathcal{A}}_n$. On the other hand, the test procedures based on $\mathcal{A}_n$ provide slightly higher power than the procedures based on the modified test statistic.

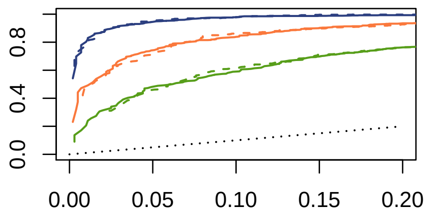
Additionally, one can use a size-power plot with adjusted (empirical) α-errors to compare the performance of $\mathcal{A}_n$ against $\widetilde{\mathcal{A}}_n$. The empirical size-power plots in Figure 3.9 display empirical size of the test (i.e., 1−sensitivity) on the x-axis versus empirical power of the test (i.e., specificity) on the y-axis. The ideal shape of the curve is as steep as possible. The empirical size-power plots demonstrate that the modified ratio type test statistic $\widetilde{\mathcal{A}}_n$ (or its bootstrap counterpart) gives approximately same (only slightly smaller) empirical powers for adjusted empirical sizes comparing to the original test statistic $\mathcal{A}_n$. This is due to two opposing facts: $\widetilde{\mathcal{A}}_n$ keeps the significance level of the test better, but $\mathcal{A}_n$ gives higher power of the test.



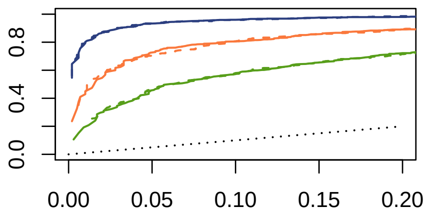
(a) $\mathcal{A}_n(\psi_{L_2})$ and $n = 100$



(b) $\widetilde{\mathcal{A}}_n(\psi_{L_2})$ and $n = 100$



(c) $\mathcal{A}_n(\psi_{L_2})$ and $n = 200$



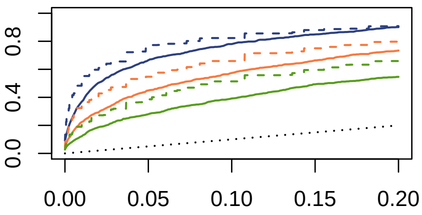
(d) $\widetilde{\mathcal{A}}_n(\psi_{L_2})$ and $n = 200$

Figure 3.9: Empirical size-power plots for $\mathcal{A}_n(\psi_{L_2})$ and $\widetilde{\mathcal{A}}_n(\psi_{L_2})$ with $\delta = 1$. Dashed lines come from the asymptotic distributions, solid correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid. Innovations are $\mathsf{N}(0, 1)$ distributed and $\gamma = 0.1$.

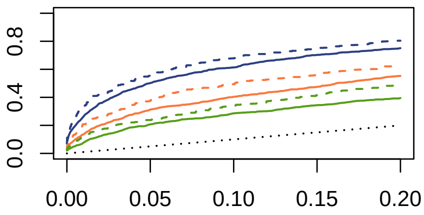
Finally, we may conclude that with larger abrupt change, the power of the test increases. Plots in Figure 3.10 show the powers of the test for alternatives with $\delta = 0.1$ and $\delta = 0.2$. Besides that, one can again see the previously mentioned fact that the tests based on the modified test statistic lack power compared to the tests based on the original test statistic.

3.10 Summary

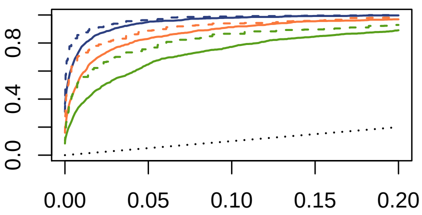
Ratio type statistics provide an alternative to non-ratio type statistics in situations, in which variance estimation is problematic. The change point detection in the location model with



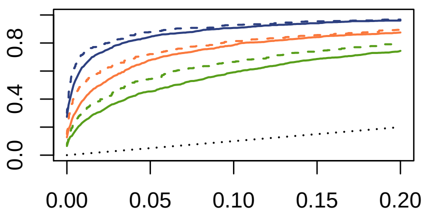
(a) $\mathcal{A}_n(\psi_{L_1})$ and $\delta = 1$



(b) $\widetilde{\mathcal{A}}_n(\psi_{L_1})$ and $\delta = 1$



(c) $\mathcal{A}_n(\psi_{L_1})$ and $\delta = 2$



(d) $\widetilde{\mathcal{A}}_n(\psi_{L_1})$ and $\delta = 2$

Figure 3.10: Size-power plots for $\mathcal{A}_n(\psi_{L_1})$ and $\widetilde{\mathcal{A}}_n(\psi_{L_1})$ under various alternatives. Dashed lines are rejection rates based on the asymptotic critical values, solid ones correspond to the block bootstrap. Orange lines stand for AR(1) with coefficient 0.3, green lines for AR(1) with coefficient 0.5, and blue ones for iid. Innovations possess Student t_5 distribution, $n = 100$, and $\gamma = 0.2$.

at most one abrupt change in mean is discussed. Asymptotic behavior of the ratio type test statistics is studied under the null hypothesis of no change and under the alternative of a change occurring at some unknown time point. We generalize testing procedures by assuming *weakly dependent errors* of the model together with incorporating *general score function* in the test statistics. The circular block bootstrap method is investigated. We prove that the *block bootstrap method provides asymptotically correct critical values* for the studied ratio type statistics in the location model with α-mixing random errors. Simulations show that critical values obtained by block bootstrapping seem to be more accurate than critical values obtained by simulation from the limiting distribution, especially for AR(1) sequences.

Chapter 4

Change In Regression Parameters

Linear regression models are relatively frequently used models in statistical analysis, with many possible applications. Difficulties with variance estimation in such models lead to the idea of avoiding the necessity of standardizing the test statistic by a variance estimate. Therefore, it seems reasonable to use ratio type test statistics.

In this chapter, we focus on the asymptotic properties of the *robust ratio type test statistics* for detection of changes in linear regression models, particularly the trending regression models, and demonstrate these properties both on simulated and real data. Moreover, a *permutation bootstrap* is proposed to overcome computational issues for obtaining the critical values for the test. The chapter is partially based on work by Madurkayová (2009a).

4.1 Introduction and regression model description

We assume to have a set of observations $Y_1, \ldots, Y_n$ obtained at time ordered points and that these data follow a linear regression model. Particularly, we are interested in studying a situation, where a *change in regression parameters* may occur at some unknown time point τ. We may formally describe such situation as

$$Y_k = \boldsymbol{h}^\top(k/n)\boldsymbol{\beta} + \boldsymbol{h}^\top(k/n)\boldsymbol{\delta}\mathcal{I}\{k > \tau\} + \varepsilon_k, \quad k = 1, \ldots, n, \tag{4.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\boldsymbol{\delta} = \boldsymbol{\delta}_n = (\delta_1, \ldots, \delta_p)^\top$, and $\tau = \tau_n$ are unknown parameters. Functions $\boldsymbol{h}(t) = (h_1(t), \ldots, h_p(t))^\top$ are such that $h_1(t) = 1$ for $t \in [0,1]$ and $h_j(t)$, $j = 2, \ldots, p$ are continuously differentiable functions on $[0,1]$. We are going to assume that the error terms $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed (iid) random variables, satisfying $\mathsf{E}\,\varepsilon_k = 0$ and $\mathsf{Var}\,\varepsilon_k = \sigma^2 > 0$ for $k = 1, \ldots, n$.

Model (4.1) corresponds to the situation where the first τ observations follow the linear model with the regression parameter $\boldsymbol{\beta}$ and the remaining $n - \tau$ observations follow the linear regression model with the changed regression parameter $\boldsymbol{\beta} + \boldsymbol{\delta}$. The parameter τ is again called the change point.

The basic question, we are trying to answer, is whether a change in regression parameters occurred at some unknown time point τ or not. Using the above introduced notation, the null hypothesis of no change can be expressed as

$$H_0 : \tau = n. \tag{4.2}$$

We are going to test this null hypothesis against the alternative hypothesis that the change occurred at some time point τ prior to the latest observed time n, i.e.,

$$H_1 : \tau < n, \ \boldsymbol{\delta} \neq \mathbf{0}. \tag{4.3}$$

A graphical illustration of the change point model (4.1) for regression parameters under the alternative can be seen in Figure 4.1.

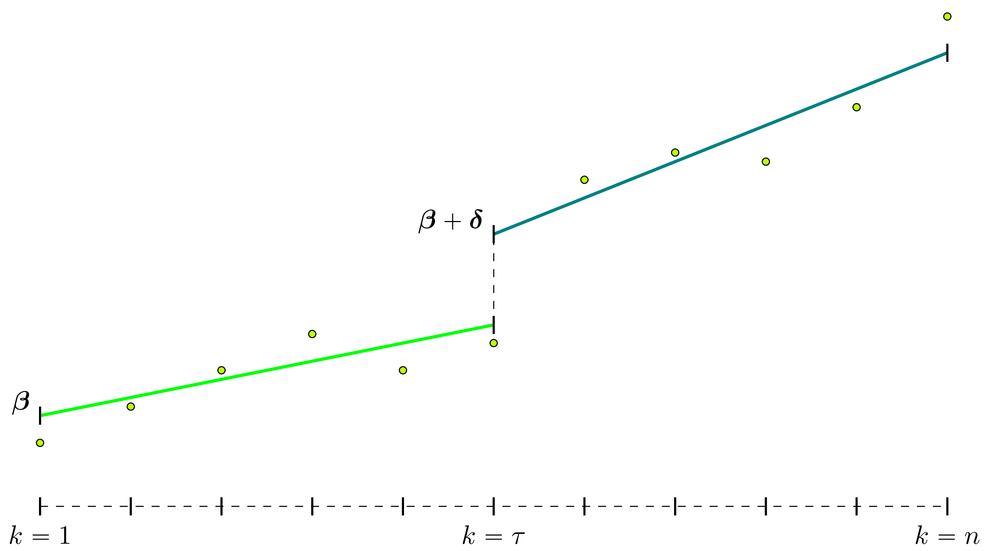


Figure 4.1: Illustration of the change point problem in regression.

A procedure for testing the change in linear regression with equidistant design was considered by Jarušková (2003). Limit distribution for over-all maximum type test statistics under assumption of no change was given. Antoch and Hušková (2003) described detection of structural changes in a general regression setup. Nonlinear polynomial regression model from the change point perspective was studied by Aue et al. (2008). M-tests for detection of changes in the linear models are presented by Hušková and Picek (2002). The paper focuses on the application of modified permutational arguments in order to obtain approximations for critical values. Furthermore, Hušková and Picek (2004) performed permutation type tests in the linear models. Bootstrap with and without replacement in change point analysis in the linear regression models is discussed in Hušková and Picek (2005). Bai and Perron (1998) give an extension into multiple structural changes, occurring at unknown time points, in the linear regression model estimated by least squares. Lately, Prášková and Chochola (2014) considered procedures for detecting a change of regression parameters in the linear model when both the regressors and the errors are weakly dependent in the sense of L_p-m approximability. M-estimators and weighted M-residuals are used to construct the test statistics.

4.2 Test statistic for detection of a change in regression parameters

For the situation described above, test statistics based on the *weighted partial sums of residuals* are often used, i.e., statistics of the form

$$\boldsymbol{S}_{j,k}(\psi) = \sum_{i=1}^{j} \boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}_k(\psi)\right), \quad j,k = p+1,\ldots,n,\ j \leqslant k, \tag{4.4}$$

which can be rewritten also elementwise (for the lth element of $\boldsymbol{S}_k$)

$$S_{j,k}^{(l)}(\psi) = \sum_{i=1}^{j} h_l(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}_k(\psi)\right), \quad l = 1,\ldots,p;\ j,k = p+1,\ldots,n,\ j \leqslant k.$$

Here, ψ is a score function and $\boldsymbol{b}_k(\psi)$ is an M-estimate of the regression parameter $\boldsymbol{\beta}$ based on observations $Y_1,\ldots,Y_k$ from model (4.1) with $\tau = n$ (under the null), i.e., it is the solution of the equation

$$\sum_{i=1}^{k} \boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}\right) = \boldsymbol{0}$$

with respect to $\boldsymbol{b}$. Let us similarly denote

$$\widetilde{\boldsymbol{S}}_{j,k}(\psi) = \sum_{i=j+1}^{n} \boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\widetilde{\boldsymbol{b}}_k(\psi)\right), \quad j,k = 1,\ldots,n-p-1,\ k \leqslant j, \tag{4.5}$$

where $\widetilde{\boldsymbol{b}}_k(\psi)$ is an M-estimate of the parameter $\boldsymbol{\beta}$ based on observations $Y_{k+1}, \ldots, Y_n$. That means, it is a solution of the equation

$$\sum_{i=k+1}^{n} \boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}\right) = \mathbf{0}$$

with respect to $\boldsymbol{b}$. Further, we denote

$$\boldsymbol{C}_{j,k} = \sum_{i=j}^{k} \boldsymbol{h}(i/n)\boldsymbol{h}^\top(i/n), \quad j,k = 1,\ldots,n,\ j \leqslant k. \tag{4.6}$$

Using this notation, we may now define the ratio type test statistic

$$\mathcal{R}_n(\psi) = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \frac{\max_{1\leqslant j\leqslant k} \boldsymbol{S}_{j,k}^\top(\psi)\boldsymbol{C}_{1,k}^{-1}\boldsymbol{S}_{j,k}(\psi)}{\max_{k\leqslant j\leqslant n-1} \widetilde{\boldsymbol{S}}_{j,k}^\top(\psi)\boldsymbol{C}_{k+1,n}^{-1}\widetilde{\boldsymbol{S}}_{j,k}(\psi)}, \tag{4.7}$$

where $0 < \gamma < 1/2$ is a given constant.

Remark 4.1. Let us note that the matrices $\boldsymbol{C}_{1,k}$ and $\boldsymbol{C}_{k+1,n}$ become regular after adding Assumption M2 (see below) and considering k and $n-k$ sufficiently large. Being particular, $k-1$ and $n-k-1$ have to be at least as large as p, i.e., the dimension of $\boldsymbol{h}(\cdot)$. Inverses of these matrices in (4.7) exist, because γ is a fixed constant known in advance and the test statistic $\mathcal{R}_n(\psi)$ is mainly studied from the asymptotic point of view. This ensures that the number of summands in (4.6) is larger than fixed p.

The idea behind the construction of the test statistic $\mathcal{R}_n(\psi)$ in (4.7) lies in comparing two total distances of weighted residuals from their center of gravity (by evaluating the ratio of the nominator and the denominator). This view comes from the fact that $\boldsymbol{S}_{j,k}(\psi)$ from (4.4) is a sum of weighted residuals and $\boldsymbol{C}_{1,k}$ from (4.6) acts as a distance measure in the Mahalanobis sense. Similarly for the denominator of (4.7).

4.3 Asymptotic properties of the robust test statistic

We proceed with deriving asymptotic properties of the robust ratio type test statistic $\mathcal{R}_n(\psi)$, under the null hypothesis as well as under the alternative. Before stating the main asymptotic results, we introduce several model assumptions. The following four assumptions apply to the model's errors $\varepsilon_1, \ldots, \varepsilon_n$ and the score function ψ.

Assumption R1. The random error terms $\{\varepsilon_i, i \in \mathbb{N}\}$ are iid random variables with a distribution function F, that is symmetric around zero.

Assumption R2. The score function ψ is a non-decreasing and antisymmetric function.

Assumption R3.

$$0 < \int \psi^2(x)\mathrm{d}F(x) < \infty$$

and

$$\int |\psi(x+t_2) - \psi(x+t_1)|^2 \mathrm{d}F(x) \leqslant C_1|t_2 - t_1|^\eta, \quad |t_j| \leqslant C_2,\ j = 1,2$$

for some constants $\eta > 0$ and $C_1, C_2 > 0$.

Assumption R4. Let us denote $\lambda(t) = -\int \psi(e-t)\mathrm{d}F(e)$, for $t \in \mathbb{R}$. We assume that $\lambda(0) = 0$ and that there exists a first derivative $\lambda'(\cdot)$ that is Lipschitz in the neighborhood of 0 and satisfies $\lambda'(0) > 0$.

The choice of the score function ψ has already been discussed in previous Chapter 3. To recapitulate, the commonly used score functions are the L_2 score function $\psi_{L_2}(x) = x$, the L_1 score function $\psi_{L_1}(x) = \mathrm{sgn}(x)$, and the Huber score function (3.7). Besides that, the use of score function

$$\psi_\beta(x) = \beta - \mathcal{I}\{x < 0\}, \quad x \in \mathbb{R},\ \beta \in (0,1)$$

results in test procedures related to the β-regression quantiles.

The next pair of assumptions refer to the system of covariate functions $\boldsymbol{h} = (h_1, \ldots, h_p)^\top$, which represent the model design.

Assumption M1. $h_1(t) = 1$, $t \in [0,1]$.

Assumption M2. $h_2(\cdot), \ldots, h_p(\cdot)$ are continuously differentiable functions on $[0,1]$ such that

$$\int_0^1 h_j(t)\mathrm{d}t = 0, \quad j = 2, \ldots, p.$$

The $p \times p$ matrix functions

$$\boldsymbol{C}(t) = \left(\int_0^t h_j(x)h_l(x)\mathrm{d}x\right)_{j,l=1,\ldots,p}, \quad t \in [0,1]$$

and $\widetilde{\boldsymbol{C}}(t) = \boldsymbol{C}(1) - \boldsymbol{C}(t)$ are regular for each $t \in (0,1]$ and $t \in [0,1)$, respectively.

Concerning the design points $\boldsymbol{h}(i/n)$, $i = 1, \ldots, n$, quite often one assumes that, as $s \to \infty$, $(\boldsymbol{C}_{k+s} - \boldsymbol{C}_k)/s$ is close to a regular matrix C uniformly in k which is not generally satisfied under Assumption M2 (Hušková and Picek, 2005). Assumption M2 covers important situations like polynomial and harmonic polynomial regression.

Now, we may characterize the limit behavior of the test statistic under the null hypothesis.

Theorem 4.1 (Under null). *Suppose that $Y_1, \ldots, Y_n$ follow model* (4.1) *and assume that Assumptions R1–R4 and M1–M2 hold. Then, under null hypothesis* (4.2)

$$\mathcal{R}_n(\psi) \xrightarrow[n\to\infty]{\mathscr{D}} \sup_{\gamma\leqslant t\leqslant 1-\gamma} \frac{\sup_{0\leqslant s\leqslant t} \boldsymbol{S}^\top(s,t)\boldsymbol{C}^{-1}(t)\boldsymbol{S}(s,t)}{\sup_{t\leqslant s\leqslant 1} \widetilde{\boldsymbol{S}}^\top(s,t)\widetilde{\boldsymbol{C}}^{-1}(t)\widetilde{\boldsymbol{S}}(s,t)}, \tag{4.8}$$

such that

$$\boldsymbol{S}(s,t) = \int_s^t \boldsymbol{h}(x)d\mathcal{W}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t)\int_0^t \boldsymbol{h}(x)d\mathcal{W}(x), \quad 0\leqslant s\leqslant t\leqslant 1,\ t\neq 0$$

and

$$\widetilde{\boldsymbol{S}}(s,t) = \int_t^s \boldsymbol{h}(x)d\widetilde{\mathcal{W}}(x) - \widetilde{\boldsymbol{C}}(s)\widetilde{\boldsymbol{C}}^{-1}(t)\int_t^1 \boldsymbol{h}(x)d\widetilde{\mathcal{W}}(x), \quad 0\leqslant t\leqslant s\leqslant 1,\ t\neq 1,$$

where $\{\mathcal{W}(x), 0\leqslant x\leqslant 1\}$ is a standard Wiener process and $\widetilde{\mathcal{W}}(x) = \mathcal{W}(1) - \mathcal{W}(x)$.

Proof. The proof goes along the lines of proof of Theorem 2.1 in Hušková and Picek (2005). Asymptotic representation for the M-estimate of regression parameter $\boldsymbol{\beta}$ can be obtained by Jurečková et al. (2012, Section 5.5)

$$\boldsymbol{b}_k(\psi) - \boldsymbol{\beta} = \boldsymbol{C}_{1,k}^{-1}\frac{1}{\lambda'(0)}\sum_{i=1}^k \boldsymbol{h}(i/n)\psi(\varepsilon_i) + O_{\mathsf{P}}\left(k^{-1}\right) \tag{4.9}$$

as $k\to\infty$ and $n\gamma\leqslant k\leqslant n(1-\gamma)$. Moreover, by the Hájek-Rényi-Chow inequality (Chow and Teicher, 2003) for each $A>0$, $\varphi\in(0,1/2]$, and $\boldsymbol{t}\in\mathbb{R}^p$

$$\begin{aligned}
&\mathsf{P}\Bigg[\max_{1\leqslant l\leqslant k/2} k^{-1/2+\varphi}l^{-\varphi}\bigg|\sum_{i=1}^l h_j(i/n)\Big(\psi(\varepsilon_i - \boldsymbol{h}^\top(i/n)\boldsymbol{t}k^{-1/2}) - \psi(\varepsilon_i) \\
&\qquad\qquad + \lambda(\boldsymbol{h}^\top(i/n)\boldsymbol{t}k^{-1/2})\Big)\bigg| \geqslant A\Bigg] \\
&\leqslant D_1A^{-2}k^{-1+2\varphi}\sum_{i=1}^{[k/2]} l^{-2\varphi}\int\Big(\psi(\varepsilon - \boldsymbol{h}^\top(i/n)\boldsymbol{t}k^{-1/2}) - \psi(\varepsilon)\Big)^2 \mathrm{d}F(\varepsilon) \\
&\leqslant D_2A^{-2}\left(k^{-1/2}\|\boldsymbol{t}\|\right)^\eta, \quad j=1,\ldots,p,
\end{aligned} \tag{4.10}$$

with some constants $D_1, D_2 > 0$, where η is the constant from Assumption R3. Similarly,

$$\mathsf{P}\left[\max_{k/2\leqslant l\leqslant k-1} k^{-1/2+\varphi}(k-l)^{-\gamma}\left|\sum_{i=l+1}^{k} h_j(i/n)\Big(\psi(\varepsilon_i - \boldsymbol{h}^\top(i/n)\boldsymbol{t}k^{-1/2}) - \psi(\varepsilon_i) + \lambda(\boldsymbol{h}^\top(i/n)\boldsymbol{t}k^{-1/2})\Big)\right| \geqslant A\right]$$
$$\leqslant D_3 A^{-2}\left(k^{-1/2}\|\boldsymbol{t}\|\right)^{\eta}, \quad j=1,\ldots,p, \tag{4.11}$$

with some constant $D_3 > 0$. Combining (4.9)–(4.11), we get

$$\max_{1\leqslant j\leqslant k-1}\frac{1}{\sqrt{k}}\left(\frac{j(k-j)}{k^2}\right)^{-\varphi}\left\|\sum_{i=1}^{j}\boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}_k(\psi)\right) - \left(\sum_{i=1}^{j}\boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{1,j}\boldsymbol{C}_{1,k}^{-1}\sum_{i=1}^{k}\boldsymbol{h}(i/n)\psi(\varepsilon_i)\right)\right\| = o_{\mathsf{P}}(1), \tag{4.12}$$

as $k\to\infty$. Using again the same arguments, we also have as $(n-k)\to\infty$

$$\max_{k+1\leqslant j\leqslant n-1}\frac{1}{\sqrt{n-k}}\left(\frac{(n-j)(j-k)}{(n-k)^2}\right)^{-\varphi}\left\|\sum_{i=j+1}^{n}\boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\widetilde{\boldsymbol{b}}_k(\psi)\right) - \left(\sum_{i=j+1}^{n}\boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{j+1,n}\boldsymbol{C}_{k+1,n}^{-1}\sum_{i=k+1}^{n}\boldsymbol{h}(i/n)\psi(\varepsilon_i)\right)\right\| = o_{\mathsf{P}}(1). \tag{4.13}$$

Hence with respect to (4.12) and (4.13), the limit distribution of

$$\left(\max_{1\leqslant j\leqslant k}\boldsymbol{S}_{j,k}^\top(\psi)\boldsymbol{C}_{1,k}^{-1}\boldsymbol{S}_{j,k}(\psi),\ \max_{k\leqslant j\leqslant n-1}\widetilde{\boldsymbol{S}}_{j,k}^\top(\psi)\boldsymbol{C}_{k+1,n}^{-1}\widetilde{\boldsymbol{S}}_{j,k}(\psi)\right)$$

is the same as that of

$$\Bigg(\max_{1\leqslant j\leqslant k}\left\{\frac{1}{\sqrt{k}}\left(\sum_{i=1}^{j}\boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{1,j}\boldsymbol{C}_{1,k}^{-1}\sum_{i=1}^{k}\boldsymbol{h}(i/n)\psi(\varepsilon_i)\right)^\top\left(\frac{1}{k}\boldsymbol{C}_{1,k}\right)^{-1}\frac{1}{\sqrt{k}}\left(\sum_{i=1}^{j}\boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{1,j}\boldsymbol{C}_{1,k}^{-1}\sum_{i=1}^{k}\boldsymbol{h}(i/n)\psi(\varepsilon_i)\right)\right\},$$
$$\max_{k\leqslant j\leqslant n-1}\left\{\left(\sum_{i=j+1}^{n}\boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{j+1,n}\boldsymbol{C}_{k+1,n}^{-1}\sum_{i=k+1}^{n}\boldsymbol{h}(i/n)\psi(\varepsilon_i)\right)^\top\boldsymbol{C}_{k+1,n}^{-1}\left(\sum_{i=j+1}^{n}\boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{j+1,n}\boldsymbol{C}_{k+1,n}^{-1}\sum_{i=k+1}^{n}\boldsymbol{h}(i/n)\psi(\varepsilon_i)\right)\right\}\Bigg),$$

which by denoting $k = [nt]$ for $t \in (0,1)$ weakly converges in $\mathscr{D}^2[\gamma, 1-\gamma]$ to

$$\Bigg(\sup_{0 \leqslant s \leqslant t} \left\{ \left(\int_0^s \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t) \int_0^t \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) \right)^\top \boldsymbol{C}^{-1}(t) \left(\int_0^s \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t) \int_0^t \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) \right) \right\},$$
$$\sup_{t \leqslant s \leqslant 1} \left\{ \left(\int_s^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) - \widetilde{\boldsymbol{C}}(s)\widetilde{\boldsymbol{C}}^{-1}(t) \int_t^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) \right)^\top \widetilde{\boldsymbol{C}}^{-1}(t) \left(\int_s^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) - \widetilde{\boldsymbol{C}}(s)\widetilde{\boldsymbol{C}}^{-1}(t) \int_t^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) \right) \right\} \Bigg),$$

as $n \to \infty$. The weak distributional convergence holds due to Jandhyala and MacNeill (1997, Theorem 1), Assumption M2, the fact that

$$\begin{aligned} &\sup_{0 \leqslant s \leqslant t} \left\{ \left(\int_0^s \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) - \int_0^s \left[\boldsymbol{h}(x) \int_0^t \boldsymbol{h}^\top(x) \boldsymbol{C}^{-1}(t) \boldsymbol{h}(y) \mathrm{d}\mathcal{W}(y) \right] \mathrm{d}x \right)^\top \boldsymbol{C}^{-1}(t) \right. \\ &\qquad \left. \left(\int_0^s \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) - \int_0^s \left[\boldsymbol{h}(x) \int_0^t \boldsymbol{h}^\top(x) \boldsymbol{C}^{-1}(t) \boldsymbol{h}(y) \mathrm{d}\mathcal{W}(y) \right] \mathrm{d}x \right) \right\} \\ &= \sup_{0 \leqslant s \leqslant t} \left\{ \left(\int_0^s \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t) \int_0^t \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) \right)^\top \boldsymbol{C}^{-1}(t) \right. \\ &\qquad \left. \left(\int_0^s \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t) \int_0^t \boldsymbol{h}(x) \mathrm{d}\mathcal{W}(x) \right) \right\}, \end{aligned}$$

and that

$$\begin{aligned} &\sup_{t \leqslant s \leqslant 1} \left\{ \left(\int_s^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) - \int_s^1 \left[\boldsymbol{h}(x) \int_t^1 \boldsymbol{h}^\top(x) \widetilde{\boldsymbol{C}}^{-1}(t) \boldsymbol{h}(y) \mathrm{d}\widetilde{\mathcal{W}}(y) \right] \mathrm{d}x \right)^\top \widetilde{\boldsymbol{C}}^{-1}(t) \right. \\ &\qquad \left. \left(\int_s^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) - \int_s^1 \left[\boldsymbol{h}(x) \int_t^1 \boldsymbol{h}^\top(x) \widetilde{\boldsymbol{C}}^{-1}(t) \boldsymbol{h}(y) \mathrm{d}\widetilde{\mathcal{W}}(y) \right] \mathrm{d}x \right) \right\} \\ &= \sup_{t \leqslant s \leqslant 1} \left\{ \left(\int_s^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) - \widetilde{\boldsymbol{C}}(s)\widetilde{\boldsymbol{C}}^{-1}(t) \int_t^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) \right)^\top \widetilde{\boldsymbol{C}}^{-1}(t) \right. \\ &\qquad \left. \left(\int_s^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) - \widetilde{\boldsymbol{C}}(s)\widetilde{\boldsymbol{C}}^{-1}(t) \int_t^1 \boldsymbol{h}(x) \mathrm{d}\widetilde{\mathcal{W}}(x) \right) \right\}. \end{aligned}$$

Then, the assertion of the theorem directly follows by the continuous mapping theorem. ☐

Remark 4.2. Realizing the property of a standard Wiener process, the definition of a Brownian bridge $\mathcal{B}(x) = \mathcal{W}(x) - x\mathcal{W}(1)$, $x \in [0,1]$, and using stochastic calculus together with

Assumption M2, we end up with

$$\int_0^s \boldsymbol{h}(x)\mathrm{d}\mathcal{W}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t)\int_0^t \boldsymbol{h}(x)\mathrm{d}\mathcal{W}(x) \\ = \int_0^s \boldsymbol{h}(x)\mathrm{d}\mathcal{B}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t)\int_0^t \boldsymbol{h}(x)\mathrm{d}\mathcal{B}(x).$$

Therefore, one can still have the same limit distribution when $\mathrm{d}\mathcal{W}(x)$ is replaced by $\mathrm{d}\mathcal{B}(x)$ and $\mathrm{d}\widetilde{\mathcal{W}}(x)$ is replaced by $\mathrm{d}\widetilde{\mathcal{B}}(x)$, where $\{\mathcal{B}(x), 0 \leqslant x \leqslant 1\}$ and $\{\widetilde{\mathcal{B}}(x), 0 \leqslant x \leqslant 1\}$ are independent Brownian bridges.

The next theorem describes situation under some local alternatives.

Theorem 4.2 (Under local alternatives)**.** *Suppose that $Y_1, \ldots, Y_n$ follow model* (4.1)*, assume that*

$$\|\boldsymbol{\delta}_n\| \to 0 \quad \textit{and} \quad \sqrt{n}\|\boldsymbol{\delta}_n\| \to \infty, \quad \textit{as} \quad n \to \infty, \tag{4.14}$$

and $\tau = [\zeta n]$ for some $\gamma < \zeta < 1-\gamma$ (alternative (4.3) *holds). Then, under Assumptions R1–R4 and M1–M2*

$$\mathcal{R}_n(\psi) \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Proof. Let us choose $k > \tau + 1$ and $k = [\xi n]$ for some $\zeta < \xi < 1 - \gamma$. Moreover, let us take into account assumption (4.14). Using the same arguments as in (4.12) and due to the fact that the local alternatives hold, we have as $n \to \infty$,

$$\max_{1\leqslant j\leqslant k} \boldsymbol{S}_{j,k}^\top(\psi)\boldsymbol{C}_{1,k}^{-1}\boldsymbol{S}_{j,k}(\psi) \geqslant \boldsymbol{S}_{\tau,k}^\top(\psi)\boldsymbol{C}_{1,k}^{-1}\boldsymbol{S}_{\tau,k}(\psi) \\ = \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}_k(\psi)\right)\right)^\top \boldsymbol{C}_{1,k}^{-1} \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}_k(\psi)\right)\right) \\ = A_{k1} + 2A_{k2} + A_{k3} + o_{\mathsf{P}}(1),$$

where

$$A_{k1} = \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\sum_{i=1}^{k} \boldsymbol{h}(i/n)\psi(\varepsilon_i)\right)^\top \boldsymbol{C}_{1,k}^{-1} \\ \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\sum_{i=1}^{k} \boldsymbol{h}(i/n)\psi(\varepsilon_i)\right),$$

$$A_{k2} = \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\psi(\varepsilon_i) - \boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1} \sum_{i=1}^{k} \boldsymbol{h}(i/n)\psi(\varepsilon_i) \right)^{\top} \boldsymbol{C}_{1,k}^{-1} \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\boldsymbol{h}^{\top}(i/n) - \boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{\tau+1,k} \right) \boldsymbol{\delta},$$

$$A_{k3} = \boldsymbol{\delta}^{\top} \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\boldsymbol{h}^{\top}(i/n) - \boldsymbol{C}_{\tau+1,k}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{1,\tau} \right) \boldsymbol{C}_{1,k}^{-1} \left(\sum_{i=1}^{\tau} \boldsymbol{h}(i/n)\boldsymbol{h}^{\top}(i/n) - \boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{\tau+1,k} \right) \boldsymbol{\delta}.$$

Then from the proof of Theorem 4.1 we get, as $n \to \infty$,

$$A_{k1} = O_{\mathsf{P}}(1).$$

Furthermore, with respect to assumption (4.14),

$$\begin{aligned} A_{k3} &= \boldsymbol{\delta}^{\top} \left(\boldsymbol{C}_{1,\tau} - \boldsymbol{C}_{\tau+1,k}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{1,\tau} \right) \boldsymbol{C}_{1,k}^{-1} \left(\boldsymbol{C}_{1,\tau} - \boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{\tau+1,k} \right) \boldsymbol{\delta} \\ &= \boldsymbol{\delta}^{\top}\boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{1,\tau}\boldsymbol{C}_{1,k}^{-1}\boldsymbol{C}_{1,\tau}\boldsymbol{\delta} \xrightarrow[n\to\infty]{\mathsf{P}} \infty. \end{aligned}$$

Finally,

$$|A_{k2}| \leqslant \sqrt{A_{k1}A_{k3}}.$$

Therefore, under the considered assumptions, the term A_{k3} is asymptotically dominant over the remaining terms. It follows that

$$\max_{1\leqslant j\leqslant k} \boldsymbol{S}_{j,k}^{\top}(\psi)\boldsymbol{C}_{1,k}^{-1}\boldsymbol{S}_{j,k}(\psi) \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

For $\tau + 1 < k = [\xi n]$, the denominator in (4.7) has the same distribution as under the null hypothesis and it is, therefore, bounded in probability. It follows that the maximum of the ratio has to tend to infinity as well, as $n \to \infty$. □

The previous theorem provides asymptotic consistency of the studied test statistic under the given assumptions. By Theorem 4.2, the statistic $\mathcal{R}_n(\psi)$ converges in probability to infinity, the null hypothesis is rejected for large values of the ratio type test statistic. Being more formal, we reject H_0 at significance level α if $\mathcal{R}_n(\psi) > r_{1-\alpha,\gamma}$, where $r_{1-\alpha,\gamma}$ is the $(1-\alpha)$-quantile of the asymptotic distribution (4.8).

4.4 Asymptotic critical values for the change in regression

The explicit form of the limit distribution (4.8) is not known. The critical values may be determined by *simulation of the limit distribution* from Theorem 4.1. Theorem 4.2 ensures that we reject the null hypothesis for large values of the test statistic. We tried to simulate the asymptotic distribution (4.8) by discretizing the stochastic integrals present in $\boldsymbol{S}(s,t)$ and $\widetilde{\boldsymbol{S}}(s,t)$ and using the relationship of a random walk to a Wiener process. We considered 1000 as the number of discretization points within $[0,1]$ interval and the number of simulations equal to 1000. We also tried to use higher numbers of discretization points, but only small differences in the critical values were acquired. In Table 4.1, we present several critical values for covariate functions $h_1(t) = 1$ and $h_2(t) = t - 1/2$.

	90%	95%	97.5%	99%	99.5%
$\gamma = 0.1$	7.629223	9.923813	13.114384	17.339711	20.891231
$\gamma = 0.2$	4.351720	5.9810256	7.583841	11.048126	14.371703

Table 4.1: Simulated critical values corresponding to the asymptotic distribution of the test statistic $\mathcal{R}_n(\psi)$ under the null hypothesis and to the covariate functions $h_1(t) = 1$ and $h_2(t) = t - 1/2$.

Moreover, Table 4.2 shows critical values for covariate functions $h_1(t) = 1$, $h_2(t) = t-1/2$, and $h_3(t) = 4t^2 - 4t + 2/3$ with $\gamma = 0.1$. The system of covariate functions was chosen in the way that those functions are orthogonal in the $L_2([0,1])$ sense.

	90%	95%	97.5%	99%	99.5%
$\gamma = 0.1$	5.638486	7.062320	8.633249	12.225500	13.631059

Table 4.2: Simulated critical values corresponding to the asymptotic distribution of the test statistic $\mathcal{R}_n(\psi)$ under the null hypothesis and to the covariate functions $h_1(t) = 1$, $h_2(t) = t - 1/2$, and $h_3(t) = 4t^2 - 4t + 2/3$.

To illustrate the applicability of the asymptotic critical values, a random sample ($n = 100$) from the regression change point model model (4.1) with the quadratic covariate system $h_1(t) = 1$, $h_2(t) = t - 1/2$, and $h_3(t) = 4t^2 - 4t + 2/3$ is simulated for particular choices of the errors' distribution (standard normal or Student t_5), τ, and $\boldsymbol{\delta}$. Considering $\gamma = 0.1$, we plot ratio

$$Q_k = \frac{\max\limits_{1 \leqslant j \leqslant k} \boldsymbol{S}_{j,k}^\top(\psi)\boldsymbol{C}_{1,k}^{-1}\boldsymbol{S}_{j,k}(\psi)}{\max\limits_{k \leqslant j \leqslant n-1} \widetilde{\boldsymbol{S}}_{j,k}^\top(\psi)\boldsymbol{C}_{k+1,n}^{-1}\widetilde{\boldsymbol{S}}_{j,k}(\psi)}, \quad n\gamma \leqslant k \leqslant n - n\gamma \tag{4.15}$$

from the ratio type test statistic $\mathcal{R}_n(\psi)$ for $\psi(x) = x$ in Figure 4.2 and for $\psi(x) = \text{sgn}(x)$ in Figure 4.3. The null hypothesis is rejected when the curve corresponding to Q_k goes above the colored horizontal line depicting the critical value. One may also notice, that under the alternative, the values of Q_k tend to increase with k in most cases and the maximum value is obtained for k close to n.

As it will be noticed further in the simulation study (cf. Section 4.7), the simulated critical values from the asymptotic distribution seem to be too liberal, meaning that the simulated versions of the critical values are smaller than they should be. In other words, the tests based on simulated asymptotic critical values reject more often than they should.

4.5 Permutation bootstrap

In the previous section, we dealt with simulation from the limit distribution as a way for approximation of the critical values of the proposed test statistic. Similarly as in Chapter 3, an alternative way to construct the test is to use resampling methods. These may even provide computationally better results when comparing to simulations of the asymptotic distribution.

Here, we propose *resampling procedure without replacement.* By permutation principle, we would like to resample the iid random errors $\varepsilon_1, \ldots, \varepsilon_n$. Then, the permutation version of our test statistics can be obtained by replacing original errors $\varepsilon_1, \ldots, \varepsilon_n$ by $\varepsilon_{R_1}, \ldots, \varepsilon_{R_n}$, where $R_1, \ldots, R_n$ is a random permutation of $1, \ldots, n$, which is independent of observations $Y_1, \ldots, Y_n$. However, the random errors $\varepsilon_1, \ldots, \varepsilon_n$ are not observed and, hence, unknown. Therefore, we *permute predicted errors*—the M-residuals. We apply the permutation principle on the nominator and on the denominator of the ratio type test statistic *separately.*

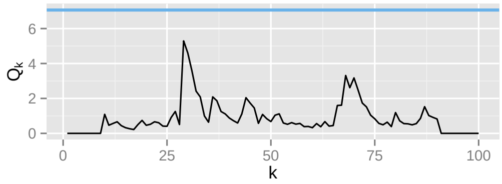
For more detailed explanation of the permutation principles in the change point analysis, we refer to Antoch and Hušková (2001), Hušková and Picek (2002), Antoch and Hušková (2003), Hušková (2004), Hušková and Picek (2004), or Hušková and Picek (2005).

First of all, we define *residuals* for the first k observations

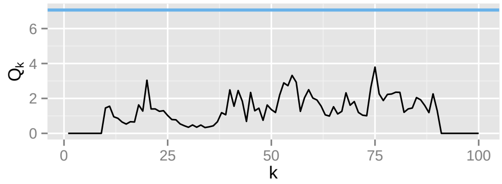
$$\widehat{\varepsilon}_i(\psi) := \psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}_k(\psi)\right), \quad i = 1, \ldots, k$$

and, similarly, residuals for the last $n - k$ observations

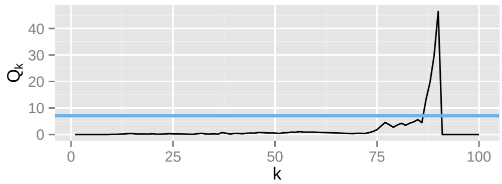
$$\widehat{\widetilde{\varepsilon}}_i(\psi) := \psi\left(Y_i - \boldsymbol{h}^\top(i/n)\widetilde{\boldsymbol{b}}_k(\psi)\right), \quad i = k+1, \ldots, n.$$



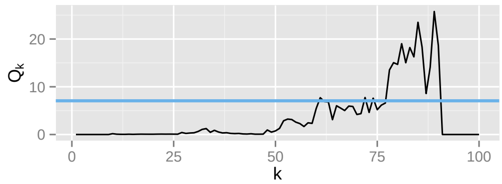
(a) H_0, $\mathsf{N}(0,1)$



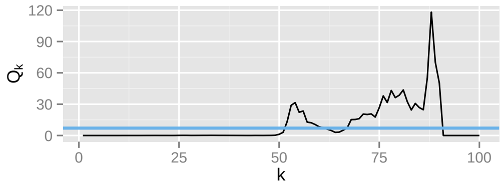
(b) H_0, t_5



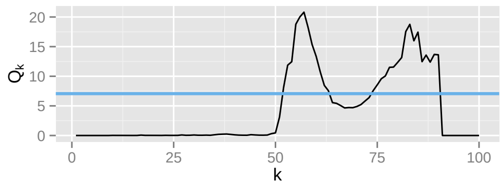
(c) H_1, $\boldsymbol{\delta} = (0, 0, \sqrt{27})^\top$, $\mathsf{N}(0,1)$



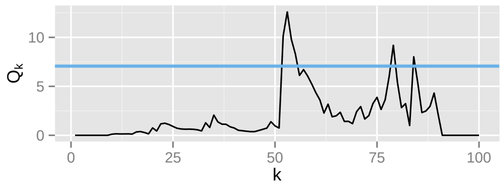
(d) H_1, $\boldsymbol{\delta} = (0, 0, \sqrt{27})^\top$, t_5



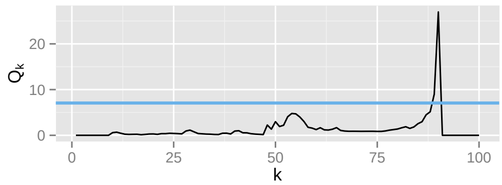
(e) H_1, $\boldsymbol{\delta} = (\sqrt{27}, 0, 0)^\top$, $\mathsf{N}(0,1)$



(f) H_1, $\boldsymbol{\delta} = (\sqrt{27}, 0, 0)^\top$, t_5

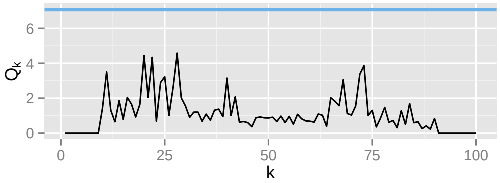


(g) H_1, $\boldsymbol{\delta} = (3, 3, 3)^\top$, $\mathsf{N}(0,1)$

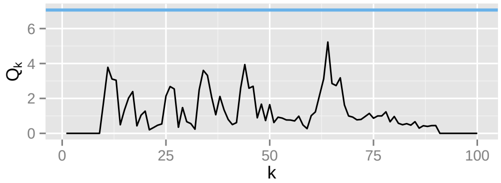


(h) H_1, $\boldsymbol{\delta} = (3, 3, 3)^\top$, t_5

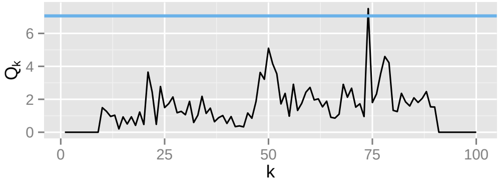
Figure 4.2: The values of Q_k from the $\mathcal{R}_n(\psi_{L_2})$ test statistic with $\gamma = 0.1$ for the simulated normal distribution samples with parameters $\mu = 0$ and $\sigma = 1$ (left hand side) and for the simulated Student t_5 distribution samples (right hand side), where $n = 100$ in both cases. The upper figures refer to the null hypothesis. The other figures refer to the alternatives with $\tau = n/2 = 50$ and $\boldsymbol{\delta} = (0, 0, \sqrt{27})^\top$, $\boldsymbol{\delta} = (\sqrt{27}, 0, 0)^\top$, $\boldsymbol{\delta} = (3, 3, 3)^\top$.



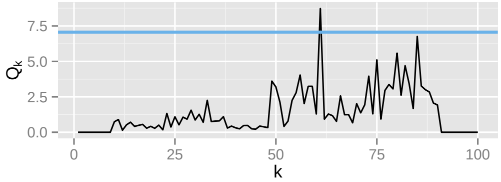
(a) H_0, $\mathsf{N}(0,1)$



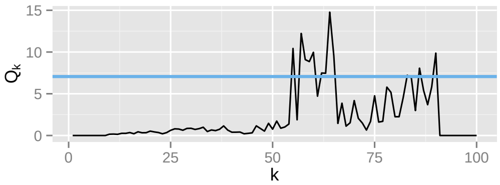
(b) H_0, t_5



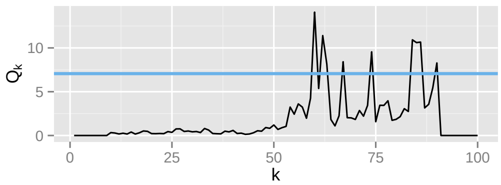
(c) H_1, $\boldsymbol{\delta} = (0, 0, \sqrt{27})^\top$, $\mathsf{N}(0,1)$



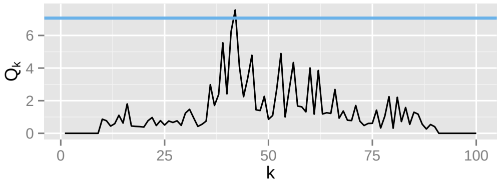
(d) H_1, $\boldsymbol{\delta} = (0, 0, \sqrt{27})^\top$, t_5



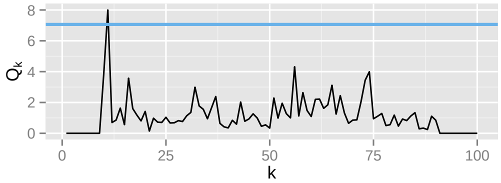
(e) H_1, $\boldsymbol{\delta} = (\sqrt{27}, 0, 0)^\top$, $\mathsf{N}(0,1)$



(f) H_1, $\boldsymbol{\delta} = (\sqrt{27}, 0, 0)^\top$, t_5



(g) H_1, $\boldsymbol{\delta} = (3, 3, 3)^\top$, $\mathsf{N}(0,1)$



(h) H_1, $\boldsymbol{\delta} = (3, 3, 3)^\top$, t_5

Figure 4.3: The values of Q_k from the $\mathcal{R}_n(\psi_{L_1})$ test statistic with $\gamma = 0.1$ for the simulated normal distribution samples with parameters $\mu = 0$ and $\sigma = 1$ (left hand side) and for the simulated Student t_5 distribution samples (right hand side), where $n = 100$ in both cases. The upper figures refer to the null hypothesis. The other figures refer to the alternatives with $\tau = n/2 = 50$ and $\boldsymbol{\delta} = (0, 0, \sqrt{27})^\top$, $\boldsymbol{\delta} = (\sqrt{27}, 0, 0)^\top$, $\boldsymbol{\delta} = (3, 3, 3)^\top$.

Consequently, for each $n\gamma \leqslant k \leqslant n(1-\gamma)$, we generate permutations $\boldsymbol{R}_k = (R_1, \ldots, R_k)$ of sequence $(1, \ldots, k)$ and $\widetilde{\boldsymbol{R}}_{n-k} = (\widetilde{R}_{k+1}, \ldots, \widetilde{R}_n)$ of $(k+1, \ldots, n)$. Now, the residuals are bootstrapped without replacement, which leads to two sets of permuted residuals

$$\left(\widehat{\varepsilon}_{R_1}(\psi), \ldots, \widehat{\varepsilon}_{R_k}(\psi)\right) \quad \text{and} \quad \left(\widehat{\widetilde{\varepsilon}}_{\widetilde{R}_{k+1}}(\psi), \ldots, \widehat{\widetilde{\varepsilon}}_{\widetilde{R}_n}(\psi)\right).$$

The bootstrapped version of $\boldsymbol{S}_{j,k}(\psi)$ is *centered by a weighted average of the residuals.*

$$\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k) = \sum_{i=1}^{j} \boldsymbol{h}(i/n)\widehat{\varepsilon}_{R_i} - \boldsymbol{C}_{1,j}\boldsymbol{C}_{1,k}^{-1} \sum_{l=1}^{k} \boldsymbol{h}(l/n)\widehat{\varepsilon}_{R_l}, \quad j, k = 1, \ldots, n,\ j \leqslant k$$

and, analogously, for $\widetilde{\boldsymbol{S}}_{j,k}(\psi)$

$$\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k}) = \sum_{i=j+1}^{n} \boldsymbol{h}(i/n)\widehat{\widetilde{\varepsilon}}_{\widetilde{R}_i} - \boldsymbol{C}_{j+1,n}\boldsymbol{C}_{k+1,n}^{-1} \sum_{l=k+1}^{n} \boldsymbol{h}(l/n)\widehat{\widetilde{\varepsilon}}_{\widetilde{R}_l}, \quad j, k = 1, \ldots, n,\ k \leqslant j.$$

Finally, the bootstrapped version of the original ratio type test statistic $\mathcal{R}_n(\psi)$ is obtained by replacing the original statistics by their permuted counterparts, i.e.,

$$\mathcal{R}_n^*(\psi) = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \frac{\max_{1 \leqslant j \leqslant k} \boldsymbol{S}_{j,k}^\top(\psi; \boldsymbol{R}_k)\boldsymbol{C}_{1,k}^{-1}\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k)}{\max_{k \leqslant j \leqslant n-1} \widetilde{\boldsymbol{S}}_{j,k}^\top(\psi; \widetilde{\boldsymbol{R}}_{n-k})\boldsymbol{C}_{k+1,n}^{-1}\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k})}, \tag{4.16}$$

where $0 < \gamma < 1/2$ is the same given constant from the definition of the original test statistic $\mathcal{R}_n(\psi)$.

An algorithm for the permutation bootstrap is illustratively shown in Procedure 4.1 and its validity is proved in Theorem 4.3. It is necessary to show that the proposed permutation test statistic is at least asymptotically correct when data follow either the null hypothesis or some alternative. Toward this, it suffices to show that given the observed data, the asymptotic conditional distribution of the original ratio type test statistic asymptotically coincides with the unconditional limit distribution of $\mathcal{R}_n(\psi)$ under the null hypothesis.

Theorem 4.3 (Permutation bootstrap validity)**.** *Suppose that $Y_1, \ldots, Y_n$ follow model* (4.1) *with $\|\boldsymbol{\delta}_n\| \to 0$ as $n \to \infty$. Suppose that Assumptions R1–R4 and M1–M2 hold. Under alternative, let $\tau = [n\zeta]$ for some $\zeta: \ \gamma < \zeta < 1-\gamma$. Then, for all $y \in \mathbb{R}$,*

$$\mathsf{P}\left(\mathcal{R}_n^*(\psi) \leqslant y | Y_1, \ldots Y_n\right) \xrightarrow[n\to\infty]{\mathsf{P}} \mathsf{P}\left(\sup_{\gamma \leqslant t \leqslant 1-\gamma} \frac{\sup_{0 \leqslant s \leqslant t} \boldsymbol{S}^\top(s,t)\boldsymbol{C}^{-1}(t)\boldsymbol{S}(s,t)}{\sup_{t \leqslant s \leqslant 1} \widetilde{\boldsymbol{S}}^\top(s,t)\widetilde{\boldsymbol{C}}^{-1}(t)\widetilde{\boldsymbol{S}}(s,t)} \leqslant y\right),$$

such that

$$\boldsymbol{S}(s,t) = \int_s^t \boldsymbol{h}(x)d\mathcal{W}(x) - \boldsymbol{C}(s)\boldsymbol{C}^{-1}(t)\int_0^t \boldsymbol{h}(x)d\mathcal{W}(x), \quad 0 \leqslant s \leqslant t \leqslant 1,\ t \neq 0,$$

Procedure 4.1 Bootstrapping test statistic $\mathcal{R}_n(\psi)$ without replacement.

Input: Sequence of observations $Y_i, \ldots, Y_n$, score function ψ and $0 < \gamma < 1/2$.
Output: Bootstrap distribution of $\mathcal{R}_n(\psi)$, i.e., the empirical distribution where probability mass $1/B$ concentrates at each of ${}_{(1)}\mathcal{R}_n^*(\psi), \ldots, {}_{(B)}\mathcal{R}_n^*(\psi)$.

for $n\gamma \leqslant k \leqslant n(1-\gamma)$ **do** // permute separately in nominator and denominator
  calculate $\boldsymbol{b}_k(\psi)$ and $\widetilde{\boldsymbol{b}}_k(\psi)$
  compute $\boldsymbol{C}_{1,k}^{-1}$ and $\boldsymbol{C}_{k+1,n}^{-1}$
  calculate residuals $\big(\widehat{\varepsilon}_1(\psi), \ldots, \widehat{\varepsilon}_k(\psi)\big)$ and $\big(\widehat{\widetilde{\varepsilon}}_{k+1}(\psi), \ldots, \widehat{\widetilde{\varepsilon}}_n(\psi)\big)$
  for $b = 1$ to B **do** // repeat in order to obtain the empirical distribution
    generate $({}_{(b)}R_1, \ldots, {}_{(b)}R_k)$ as a random permutation of $(1, \ldots, k)$
    generate $({}_{(b)}\widetilde{R}_{k+1}, \ldots, {}_{(b)}\widetilde{R}_n)$ as a random permutation of $(k+1, \ldots, n)$
    for $1 \leqslant j \leqslant k$ **do** // evaluate for the nominator
      construct permuted residuals $\big({}_{(b)}\widehat{\varepsilon}_{R_1}(\psi), \ldots, {}_{(b)}\widehat{\varepsilon}_{R_k}(\psi)\big)$
      calculate ${}_{(b)}\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k)$
    end for
    calculate $\max_{1\leqslant j\leqslant k}\, {}_{(b)}\boldsymbol{S}_{j,k}^\top(\psi; \boldsymbol{R}_k)\boldsymbol{C}_{1,k}^{-1}\, {}_{(b)}\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k)$
    for $k \leqslant j \leqslant n-1$ **do** // evaluate for the denominator
      construct permuted residuals $\big({}_{(b)}\widehat{\widetilde{\varepsilon}}_{\widetilde{R}_{k+1}}(\psi), \ldots, {}_{(b)}\widehat{\widetilde{\varepsilon}}_{\widetilde{R}_n}(\psi)\big)$
      calculate ${}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k})$
    end for
    calculate $\max_{k\leqslant j\leqslant n-1}\, {}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}^\top(\psi; \widetilde{\boldsymbol{R}}_{n-k})\boldsymbol{C}_{k+1,n}^{-1}\, {}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k})$
    evaluate

$$
{}_{(b)}Q_k^{(n)}(\psi) := \frac{\max_{1\leqslant j\leqslant k}\, {}_{(b)}\boldsymbol{S}_{j,k}^\top(\psi; \boldsymbol{R}_k)\boldsymbol{C}_{1,k}^{-1}\, {}_{(b)}\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k)}{\max_{k\leqslant j\leqslant n-1}\, {}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}^\top(\psi; \widetilde{\boldsymbol{R}}_{n-k})\boldsymbol{C}_{k+1,n}^{-1}\, {}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k})}
$$

  end for
end for
for $b = 1$ to B **do** // pick the highest bootstrapped ratio
  compute bootstrap test statistics ${}_{(b)}\mathcal{R}_n^*(\psi) = \max_{n\gamma\leqslant k\leqslant n(1-\gamma)}\, {}_{(b)}Q_k^{(n)}(\psi)$
end for

$$
\widetilde{\boldsymbol{S}}(s,t) = \int_t^s \boldsymbol{h}(x)\,d\widetilde{\mathcal{W}}(x) - \widetilde{\boldsymbol{C}}(s)\widetilde{\boldsymbol{C}}^{-1}(t)\int_t^1 \boldsymbol{h}(x)\,d\widetilde{\mathcal{W}}(x), \quad 0 \leqslant t \leqslant s \leqslant 1,\ t \neq 1,
$$

where $\{\mathcal{W}(x), 0 \leqslant x \leqslant 1\}$ *is a standard Wiener process and* $\widetilde{\mathcal{W}}(x) = \mathcal{W}(1) - \mathcal{W}(x)$.

Proof. According to Hušková and Picek (2005, Theorem 2.3) for all $(y,z)^\top \in \mathbb{R}^2$,

$$
\begin{aligned}
&\mathsf{P}\Bigg(\max_{1\leqslant j\leqslant nt} \boldsymbol{S}_{j,[nt]}^\top(\psi; \boldsymbol{R}_{[nt]})\boldsymbol{C}_{1,[nt]}^{-1}\boldsymbol{S}_{j,[nt]}(\psi; \boldsymbol{R}_{[nt]}) \leqslant y, \\
&\qquad \max_{nt<j\leqslant n-1} \widetilde{\boldsymbol{S}}_{j,[nt]}^\top(\psi; \widetilde{\boldsymbol{R}}_{n-[nt]})\boldsymbol{C}_{[nt]+1,n}^{-1}\widetilde{\boldsymbol{S}}_{j,[nt]}(\psi; \widetilde{\boldsymbol{R}}_{n-[nt]}) \leqslant z \,\Big|\, Y_1, \ldots Y_n\Bigg) \\
&\xrightarrow[n\to\infty]{\mathsf{P}} \mathsf{P}\left(\sup_{0\leqslant s\leqslant t} \boldsymbol{S}^\top(s,t)\boldsymbol{C}^{-1}(t)\boldsymbol{S}(s,t) \leqslant y,\ \sup_{t\leqslant s\leqslant 1} \widetilde{\boldsymbol{S}}^\top(s,t)\widetilde{\boldsymbol{C}}^{-1}(t)\widetilde{\boldsymbol{S}}(s,t) \leqslant z\right), \qquad (4.17)
\end{aligned}
$$

for each $t \in (0, 1)$. Since (4.17) holds also for all $t \in [\gamma, 1-\gamma]$, the assertion of the theorem directly follows. □

We have already shown that the resampled ratio type test statistic $\mathcal{R}_n^*(\psi)$, conditioned on the original observations $Y_1, \ldots, Y_n$, has exactly the same limit behavior as the original test statistic $\mathcal{R}_n(\psi)$ under the null hypothesis. This remains true both under the null hypothesis and the local alternatives. Hence, the bootstrap distribution of $\mathcal{R}_n(\psi)$ gives critical values as empirical quantiles. This means that we have proved that $\mathcal{R}_n^*(\psi)$ provides asymptotically correct critical values for the test based on $\mathcal{R}_n(\psi)$, when observations follow either the null hypothesis or the alternative. We reject the null hypothesis for large values of the test statistic $\mathcal{R}_n(\psi)$ due to Theorem 4.2.

4.6 Extension for weakly dependent random errors

In the previous sections, properties of the ratio type statistics for detection of changes in the linear regression models were only described for the case of independent random errors. We would also like to study the possibility of extending these methods for the case of dependent random errors, including ARMA processes.

In case of independent random errors, the variance of the studied model is usually estimated by sums of squared residuals. In case, when the random errors form, e.g, a linear process, it is more appropriate to use an estimate that respects the underlying dependency structure. Therefore, the Bartlett estimator or one of its modifications is often used to estimate the long run variance. Especially for the case of dependent random errors, it may be difficult to find variance estimates that have satisfactory behavior both under null hypothesis and under alternative. Hence, the ratio type test statistics can become even more applicable when model's errors are no more independent.

One can indeed extend the results from this chapter concerning the change in regression parameters into the setup, where the *weakly dependent errors* are assumed. Hence the independence assumption for model's errors can be dropped out. This extension could follow the same steps as the proof of Theorem 4.1 and use similar arguments to justify the results. However, other assumptions on the weak dependency structure of the errors $\varepsilon_1, \ldots, \varepsilon_n$ would need to be added. These additional assumptions have to assure that one can apply the Bahadur-Kiefer asymptotic representation (4.9) of the regression M-estimates, the Hájek-Rényi type inequality (4.10), and two-dimensional functional central limit theorem on $\left(n^{-1/2}\sum_{1\leqslant i\leqslant nt}\varepsilon_i, n^{-1/2}\sum_{nt<i\leqslant n}\varepsilon_i\right)$ for dependent errors. Then the asymptotic results would remain the same for the independent and for the weakly dependent errors. The only difference would be more restrictive assumptions on the errors' structure in the case of weakly dependent errors. An alternative way of extending the presented results from this chapter for weakly dependent errors is to use the Abel type of summability proposed

by Hušková and Steinebach (2000, p. 61–62) and combine it with the results derived in previous Chapter 3.

4.7 Simulation study

A simulation experiment was conducted to study the *finite sample properties* of the asymptotic and permutation bootstrap test for an unknown change in the regression parameters. Performance of the tests based on ratio type test statistic $\mathcal{R}_n(\psi)$ with $\psi_{L_2}(x) = x$ and $\psi_{L_1}(x) = \text{sgn}(x)$ is studied from a numerical point of view. In particular, the interest lies in the *empirical size* of the proposed tests under the null hypothesis and in the empirical *rejection rate* (power) under the alternative. Random samples of data (1000 repetitions) are generated from the linear regression change point model (4.1) with $h_1(t) = 1$ and $h_2(t) = t - 1/2$. The number of observations considered is mainly $n = 100$. Higher sample sizes were also tried and the effect of number of observations will be discussed at the end of this section. Parameter γ is set to 0.1.

The innovations are obtained as iid random variables from a standard normal $\mathsf{N}(0, 1)$ or Student t_5 distribution. The regression parameters $\boldsymbol{\beta}$ is chosen as $(2, 3)^\top$. Simulation scenarios are produced by varying all possible combinations of these settings. The number of bootstrap replications used is 1000. Table 4.3 provides the empirical size of the tests for both the asymptotic and bootstrap version of the regression change point test, where the theoretical significance level is α.

score	innovations	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
L_2	$\mathsf{N}(0, 1)$	0.023	0.005	0.111	0.023	0.214	0.042
	t_5	0.050	0.008	0.191	0.028	0.294	0.052
L_1	$\mathsf{N}(0, 1)$	0.003	0.000	0.045	0.012	0.136	0.024
	t_5	0.003	0.002	0.039	0.010	0.112	0.026

Table 4.3: Empirical size of the test for the change in regression under H_0 using the asymptotic critical values and the permutation bootstrap from $\mathcal{R}_n(\psi)$, considering various significance levels α and $n = 100$.

Additionally, a graphical illustration of the performance of the test under the null hypothesis for significance level varying from 0.01 to 0.20 is provided by the *size-power plots* (for a more detailed description see Section 3.9) in Figure 4.4. The theoretical rejection rate (i.e., the significance level) under the null hypothesis is depicted by a straight dotted line.

Generally, the empirical sizes in case of asymptotic test are higher than they should be. That means the test rejects the null hypothesis more often than it should. Possible explanation of this difficulty can be that the test statistics converge only very slowly to

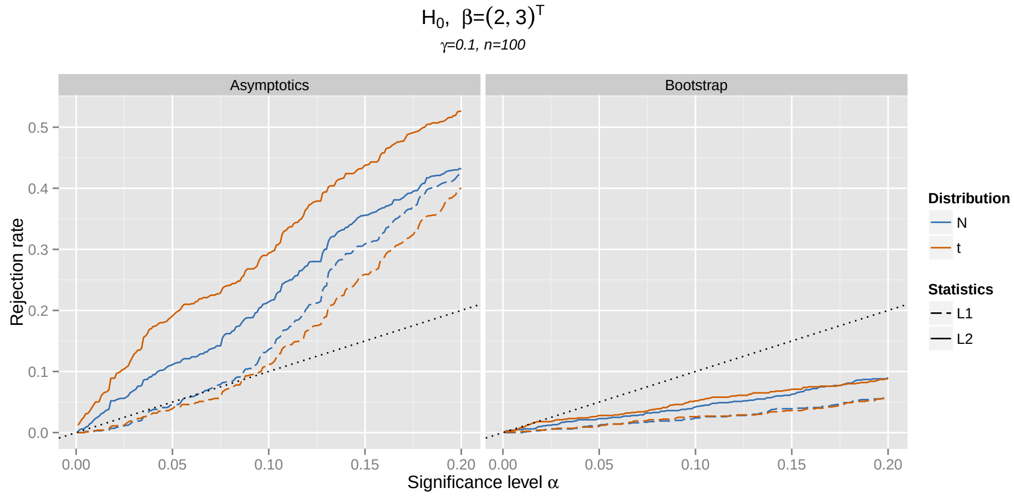


Figure 4.4: Rejection rates for the asymptotic and permutation bootstrap tests for change in regression based on $\mathcal{R}_n(\psi)$ under null hypothesis H_0.

the theoretical asymptotic distribution under the null hypothesis. On the other hand, the permutation bootstrap keeps the theoretical significance level. Nevertheless, the bootstrap method for testing change in regression is too conservative meaning that it rejects the null hypothesis more often than it should.

Better performance of the asymptotic test under the null hypothesis is achieved, when the L_1 score function is chosen for $\mathcal{R}_n(\psi)$ compared to the L_2 method. The empirical significance levels are approximately the same as the theoretical significance levels, especially for values of α less or equal to 0.1. The L_2 method seems to be too liberal in rejecting the null hypothesis. However, the L_2 method works better in case of bootstrapping. There is no significant effect of the errors' distribution on the empirical rejection rates based on this simulation study.

The performance of the testing procedure under H_1 in terms of the empirical rejection rates is shown in Table 4.4, where the change point is set to $\tau = n/2$ or $\tau = n/4$. The values of $\boldsymbol{\delta}$ are chosen as $\boldsymbol{\delta} = (1,1)^\top$ and $\boldsymbol{\delta} = (2,3)^\top$. The performance of the tests under the alternatives for significance level varying from 0.01 to 0.20 is visualized by the size-power plots in Figures 4.5–4.8.

The size power plots under the alternatives show that the power is generally higher in case of the asymptotic test compared to the bootstrap one. However, we should keep in mind that the test based on asymptotic critical values does not keep the theoretical significance level under the null hypothesis.

The test power drops when switching from a change point located in the middle of the

score	innovations	δ	τ	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
L_2	$\mathsf{N}(0,1)$	$(1,1)^\top$	$n/2$	0.110	0.036	0.332	0.114	0.476	0.176
			$n/4$	0.065	0.011	0.197	0.042	0.330	0.069
		$(2,3)^\top$	$n/2$	0.600	0.299	0.855	0.546	0.922	0.674
			$n/4$	0.096	0.023	0.296	0.084	0.465	0.137
	t_5	$(1,1)^\top$	$n/2$	0.108	0.016	0.319	0.060	0.469	0.099
			$n/4$	0.087	0.007	0.247	0.037	0.348	0.065
		$(2,3)^\top$	$n/2$	0.429	0.162	0.690	0.346	0.801	0.452
			$n/4$	0.106	0.013	0.281	0.044	0.426	0.077
L_1	$\mathsf{N}(0,1)$	$(1,1)^\top$	$n/2$	0.017	0.008	0.145	0.042	0.291	0.077
			$n/4$	0.002	0.001	0.086	0.017	0.211	0.047
		$(2,3)^\top$	$n/2$	0.105	0.076	0.508	0.279	0.715	0.386
			$n/4$	0.005	0.005	0.115	0.035	0.241	0.065
	t_5	$(1,1)^\top$	$n/2$	0.009	0.006	0.112	0.050	0.237	0.082
			$n/4$	0.005	0.000	0.072	0.008	0.190	0.031
		$(2,3)^\top$	$n/2$	0.091	0.060	0.436	0.220	0.651	0.342
			$n/4$	0.003	0.002	0.102	0.022	0.233	0.051

Table 4.4: Empirical power of the test for the change in regression under H_1 using the asymptotic critical values and the permutation bootstrap from $\mathcal{R}_n(\psi)$, considering various significance levels α and $n = 100$.

time series to a change point closer to the beginning or the end of the time series. The errors with heavier tails (i.e., t_5) yield slightly smaller power than the errors with lighter tails. When using the L_1 method, power of the test decreases compared to the L_2 method. Although, it keeps the theoretical significance level under the null hypothesis better in case of the asymptotic test.

We have shown power plots only for one choice of $\boldsymbol{\beta}$, however, several other values were tried in simulations. There was no visible effect of the value of regression parameter on the power of the tests. Naturally, the higher the value of change in the regression parameter is, the higher the power is achieved.

For several simulation scenarios, the proposed methods do not seem very satisfactory. Better results may be obtained by considering larger sample size (cf. Figure 4.10) or an alternative more far away from the null hypothesis. The results for the choice of $\boldsymbol{\delta} = (5,5)^\top$ are shown in Figure 4.9. It may be concluded that in case of relatively large change in the regression parameter, the power of the tests (both asymptotic and bootstrap) increases.

Note that the length of each time series considered until this point in the simulation

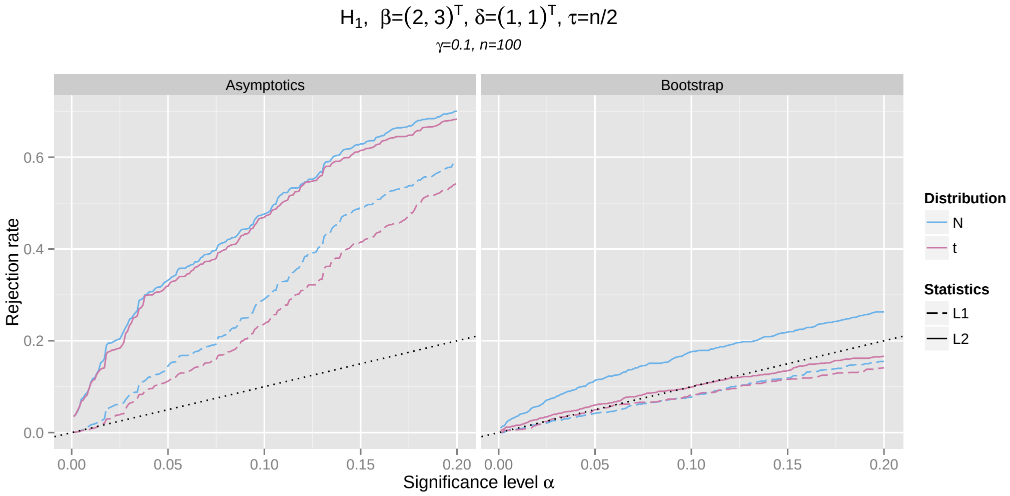


Figure 4.5: Rejection rates for the asymptotic and permutation bootstrap tests for change in regression based on $\mathcal{R}_n(\psi)$ under alternative H_1 with $\tau = n/2$ and $\boldsymbol{\delta} = (1, 1)^\top$.

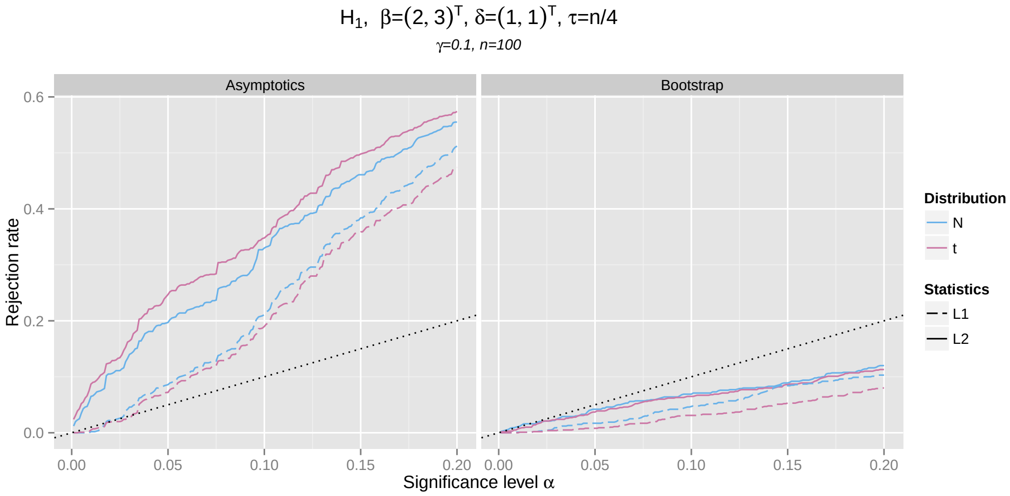


Figure 4.6: Rejection rates for the asymptotic and permutation bootstrap tests for change in regression based on $\mathcal{R}_n(\psi)$ under alternative H_1 with $\tau = n/4$ and $\boldsymbol{\delta} = (1, 1)^\top$.

study was only 100. Let us now consider 250 observations and investigate the performance of the tests under the null hypothesis and under the alternative, see Figure 4.10.

One can conclude that the power of the tests (both asymptotic and bootstrap) increases as the number of observations increases, which is expected. For the bootstrap test, even the

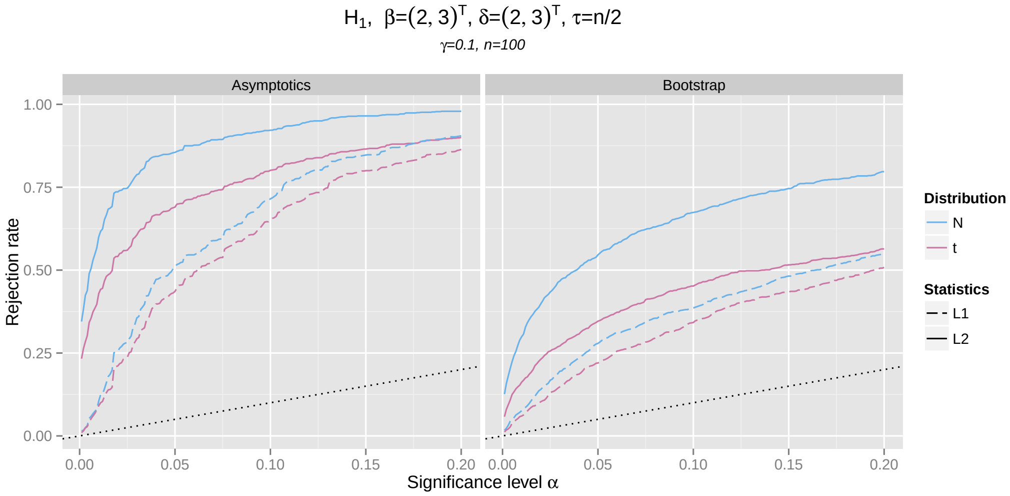


Figure 4.7: Rejection rates for the asymptotic and permutation bootstrap tests for change in regression based on $\mathcal{R}_n(\psi)$ under alternative H_1 with $\tau = n/2$ and $\boldsymbol{\delta} = (2,3)^\top$.

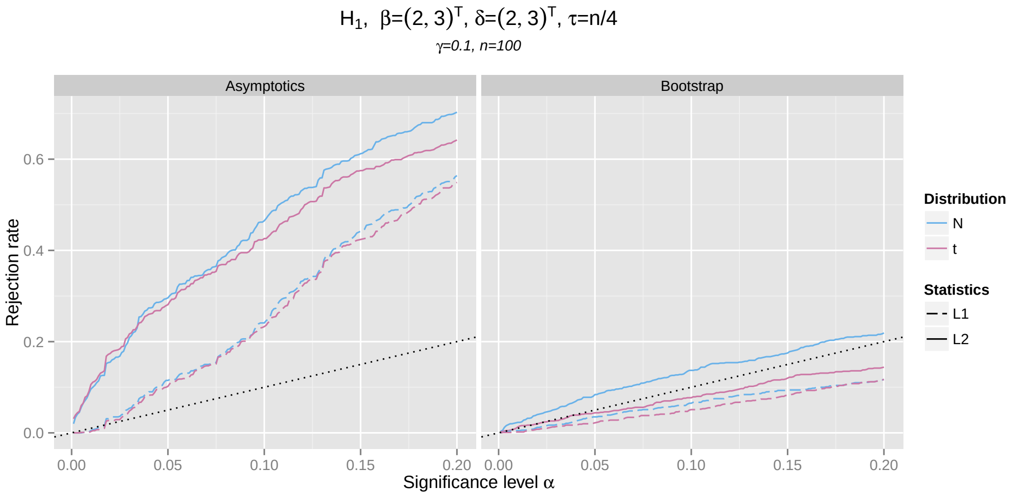


Figure 4.8: Rejection rates for the asymptotic and permutation bootstrap tests for change in regression based on $\mathcal{R}_n(\psi)$ under alternative H_1 with $\tau = n/4$ and $\boldsymbol{\delta} = (2,3)^\top$.

size of the test increases and comes closer to the theoretical significance level. It seems that bootstrapping starts to work satisfactory for 250 observations and more. On the other hand, bootstrapping longer time series becomes computationally very intensive even when using parallel cluster computing. When comparing the detection procedures of abrupt change

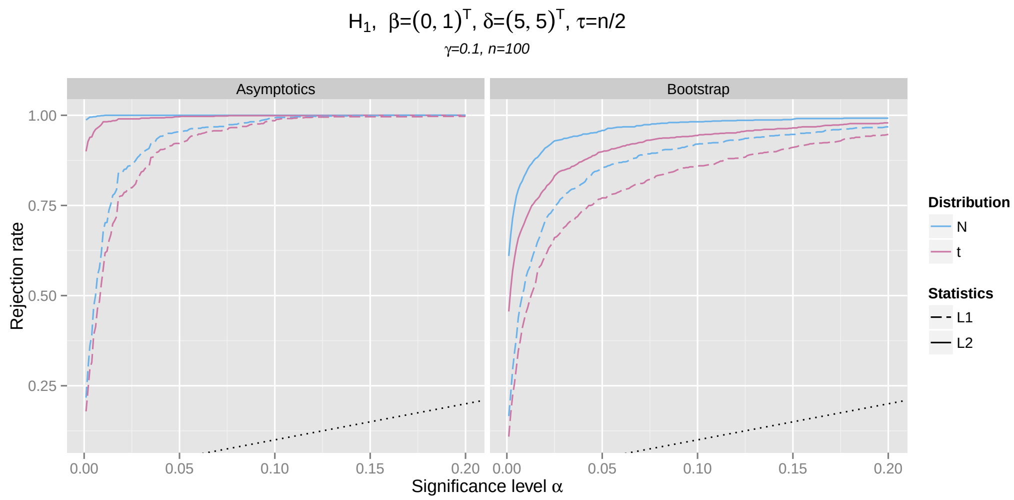


Figure 4.9: Rejection rates for the asymptotic and permutation bootstrap tests for change in regression based on $\mathcal{R}_n(\psi)$ under alternative H_1 with $\tau = n/2$ and $\boldsymbol{\delta} = (5,5)^\top$.

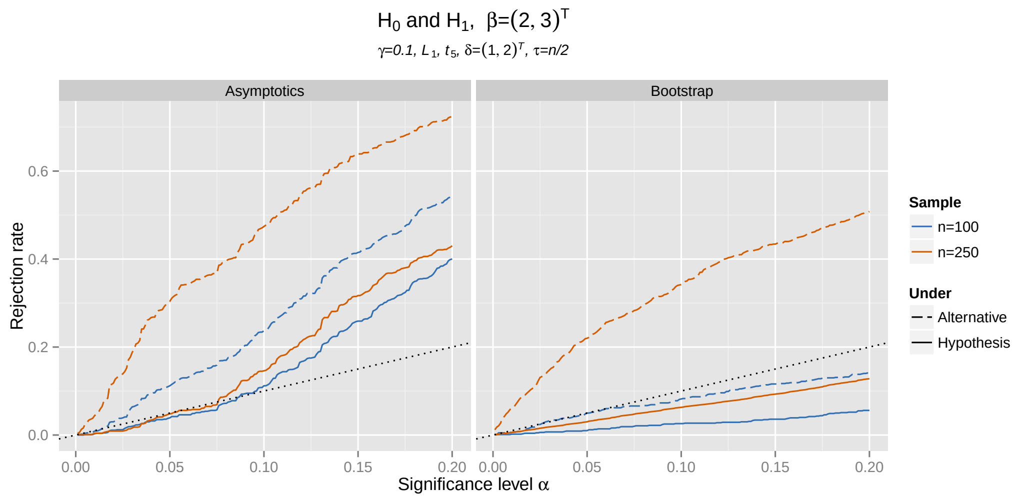


Figure 4.10: Rejection rates for the asymptotic and permutation bootstrap tests for change in regression based on $\mathcal{R}_n(\psi)$ under H_0 and under H_1 for different sample sizes.

in mean from Chapter 3 and the detection procedures of change in regression parameters presented in this chapter, the tests for the simpler model of abrupt change perform generally better.

Further, we compare the results to an alternative algorithm to the permutation bootstrap (without replacement). The residuals are now going to be resampled *with replacement* and, moreover, the resampling of residuals is done from the whole vector of residuals. Hence, the residuals from the nominator and the denominator of the test statistic are mixed (interlaced) after resampling. This is different from the previously described permutation bootstrap, where the resampled residuals remain separated—those from the nominator are used for the nominator of the bootstrap version of the test statistics and similarly for those from the denominator. The whole regression bootstrap algorithm can be formalized in Procedure 4.2.

Procedure 4.2 Bootstrapping test statistic $\mathcal{R}_n(\psi)$ with replacement.

Input: Sequence of observations $Y_i, \ldots, Y_n$, score function ψ and $0 < \gamma < 1/2$.
Output: Bootstrap distribution of $\mathcal{R}_n(\psi)$, i.e., the empirical distribution where probability mass $1/B$ concentrates at each of ${}_{(1)}\mathcal{R}_n^*(\psi), \ldots, {}_{(B)}\mathcal{R}_n^*(\psi)$.

for $n\gamma \leqslant k \leqslant n(1-\gamma)$ **do** // permute separately in nominator and denominator
  calculate $\boldsymbol{b}_k(\psi)$ and $\widetilde{\boldsymbol{b}}_k(\psi)$
  compute $\boldsymbol{C}_{1,k}^{-1}$ and $\boldsymbol{C}_{k+1,n}^{-1}$
  calculate residuals $\big(\widehat{\varepsilon}_1(\psi), \ldots, \widehat{\varepsilon}_k(\psi)\big)$ and $\big(\widehat{\widetilde{\varepsilon}}_{k+1}(\psi), \ldots, \widehat{\widetilde{\varepsilon}}_n(\psi)\big)$
  merge residuals together $\big(\breve{\varepsilon}_1(\psi), \ldots, \breve{\varepsilon}_n(\psi)\big) = \big(\widehat{\varepsilon}_1(\psi), \ldots, \widehat{\varepsilon}_k(\psi), \widehat{\widetilde{\varepsilon}}_{k+1}(\psi), \ldots, \widehat{\widetilde{\varepsilon}}_n(\psi)\big)$
  for $b = 1$ to B **do** // repeat in order to obtain the empirical distribution
    generate $\big({}_{(b)}R_1, \ldots, {}_{(b)}R_n\big)$ randomly from $(1, \ldots, n)$ with replacement
    for $1 \leqslant j \leqslant k$ **do** // evaluate for the nominator
      construct permuted residuals $\big({}_{(b)}\breve{\varepsilon}_{R_1}(\psi), \ldots, {}_{(b)}\breve{\varepsilon}_{R_k}(\psi)\big)$
      replace ${}_{(b)}\widehat{\varepsilon}_{R_i}(\psi)$ by ${}_{(b)}\breve{\varepsilon}_{R_i}(\psi)$ and calculate ${}_{(b)}\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k)$
    end for
    calculate $\max_{1 \leqslant j \leqslant k} {}_{(b)}\boldsymbol{S}_{j,k}^\top(\psi; \boldsymbol{R}_k)\boldsymbol{C}_{1,k}^{-1}{}_{(b)}\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k)$
    for $k \leqslant j \leqslant n-1$ **do** // evaluate for the denominator
      construct permuted residuals $\big({}_{(b)}\breve{\varepsilon}_{R_{k+1}}(\psi), \ldots, {}_{(b)}\breve{\varepsilon}_{R_n}(\psi)\big)$
      replace ${}_{(b)}\widehat{\widetilde{\varepsilon}}_{\widetilde{R}_i}(\psi)$ by ${}_{(b)}\breve{\varepsilon}_{R_i}(\psi)$ and calculate ${}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k})$
    end for
    calculate $\max_{k \leqslant j \leqslant n-1} {}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}^\top(\psi; \widetilde{\boldsymbol{R}}_{n-k})\boldsymbol{C}_{k+1,n}^{-1}{}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k})$
    evaluate

$$ {}_{(b)}Q_k^{(n)}(\psi) := \frac{\max_{1 \leqslant j \leqslant k} {}_{(b)}\boldsymbol{S}_{j,k}^\top(\psi; \boldsymbol{R}_k)\boldsymbol{C}_{1,k}^{-1}{}_{(b)}\boldsymbol{S}_{j,k}(\psi; \boldsymbol{R}_k)}{\max_{k \leqslant j \leqslant n-1} {}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}^\top(\psi; \widetilde{\boldsymbol{R}}_{n-k})\boldsymbol{C}_{k+1,n}^{-1}{}_{(b)}\widetilde{\boldsymbol{S}}_{j,k}(\psi; \widetilde{\boldsymbol{R}}_{n-k})} $$

  end for
end for
for $b = 1$ to B **do** // pick the highest bootstrapped ratio
  compute bootstrap test statistics ${}_{(b)}\mathcal{R}_n^*(\psi) = \max_{n\gamma \leqslant k \leqslant n(1-\gamma)} {}_{(b)}Q_k^{(n)}(\psi)$
end for

Despite the original expectations when proposing the alternative version of the permutation bootstrap without replacement, the regression bootstrap with replacement does not gain significantly higher power compared to the permutation bootstrap without replacement. To demonstrate this numerically using size-power plots, the rejection rates of the separated permutation bootstrap and the interlaced regression bootstrap under the null hypothesis are shown in Figure 4.11. Moreover, the comparison of rejection rates of the bootstrap without replacement (Procedure 4.1) and bootstrap with replacement (Procedure 4.2) under the alternative is displayed in Figure 4.12.

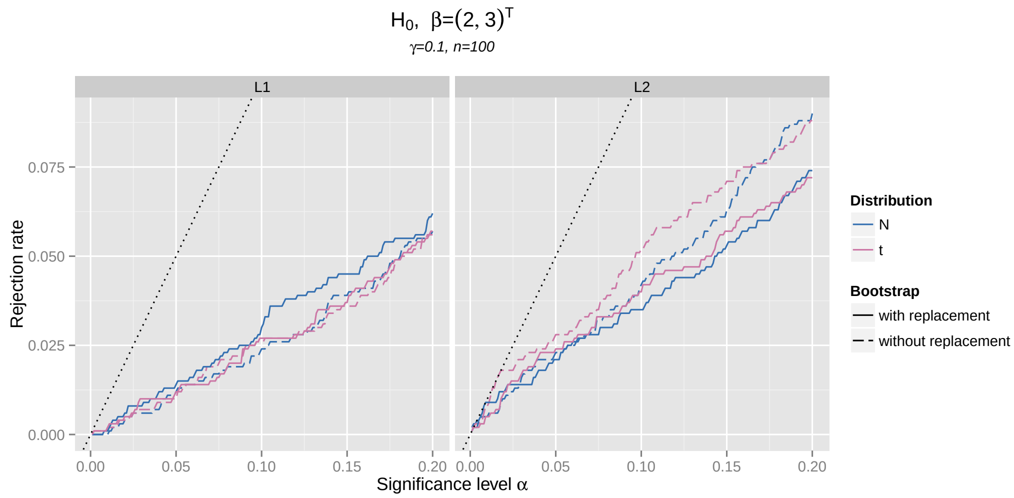


Figure 4.11: Rejection rates for the permutation bootstrap test without replacement and the regression bootstrap test with replacement based on $\mathcal{R}_n(\psi)$ under null hypothesis H_0.

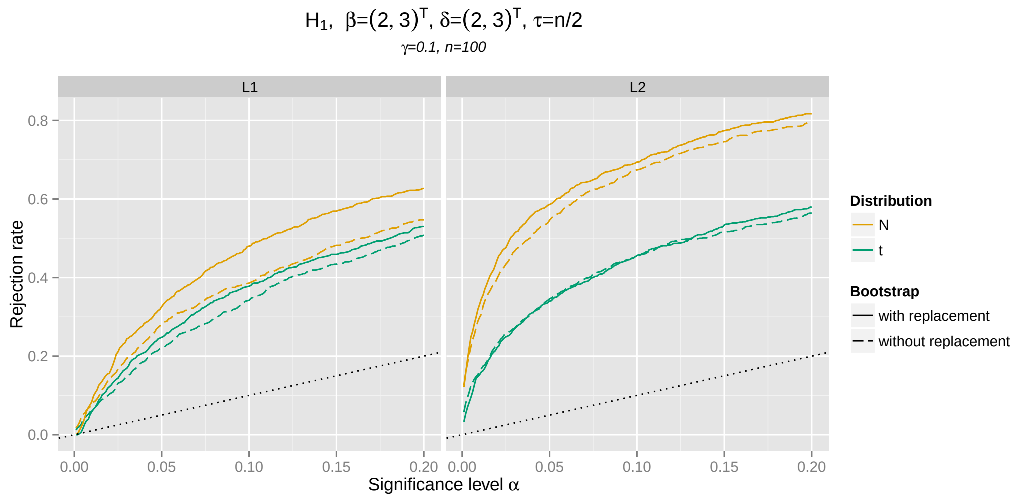


Figure 4.12: Rejection rates for the permutation bootstrap test without replacement and the regression bootstrap test with replacement based on $\mathcal{R}_n(\psi)$ under alternative H_1.

4.8 Application to surface temperature data

The analysed data come from a large data set based on long term surface temperature measurements at several meteorological stations around the world (for more details see Met Office Hadley Centre (2008), data set HadCRUT3). In Figure 4.15, we may see the data together with already estimated regression curves. The data represent *temperature anomalies*, i.e., differences from what is expected to be measured in some particular area at some particular time of the year. Each observation corresponds to monthly measurements at the chosen area located in the South Pacific Ocean, close to New Zealand (the center of the 5×5 degree area is located at 177.5W and 32.5S). The data covers the period of years 1947–1987 including 485 months.

We took L_2 score function with $p = 3$, $h_1(x) = 1$, $h_2(x) = x - 1/2$, $h_3(x) = 4x^2 - 4x + 2/3$, and $\gamma = 0.1$. We apply the methods described above. The values of the ratio Q_k defined by (4.15) are shown in Figure 4.13 for the L_2 method. We also show the ratio Q_k for the L_1 method in Figure 4.14.

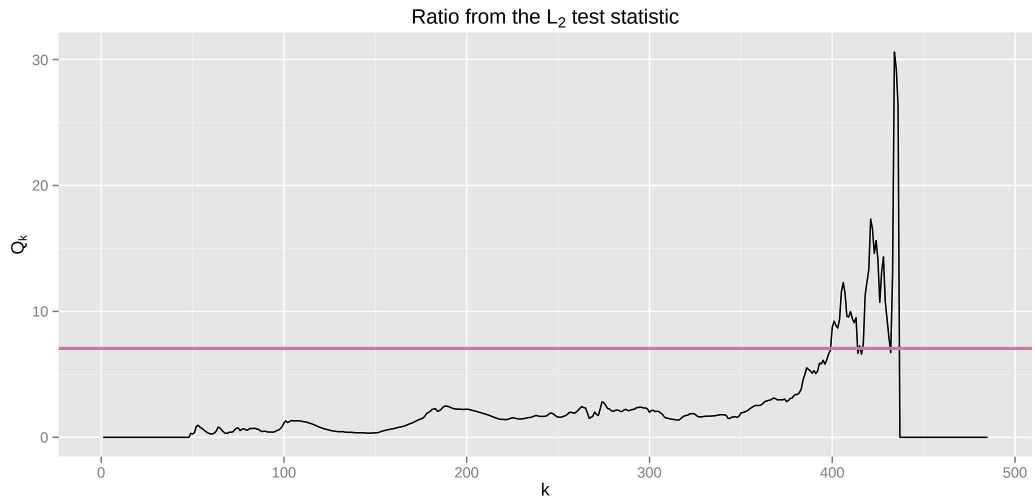


Figure 4.13: The values of Q_k from the L_2 test statistic for the surface temperature data. The simulated 95% critical value is depicted by the colored horizontal line.

We reject the null hypothesis of no change in the parameters of the quadratic regression model based on the asymptotic test (95% critical value equals 7.06232) both for the L_2 method and the L_1 method, since $\mathcal{R}_{485}(\psi_{L_2}) = 30.59436$ and $\mathcal{R}_{485}(\psi_{L_1}) = 8.477089$. We also reject the null hypothesis of no change according to the permutation bootstrap, because the bootstrap critical values are even smaller than the asymptotic ones.

We estimate the time of change τ by maximizing the nominator in (4.7) when using all

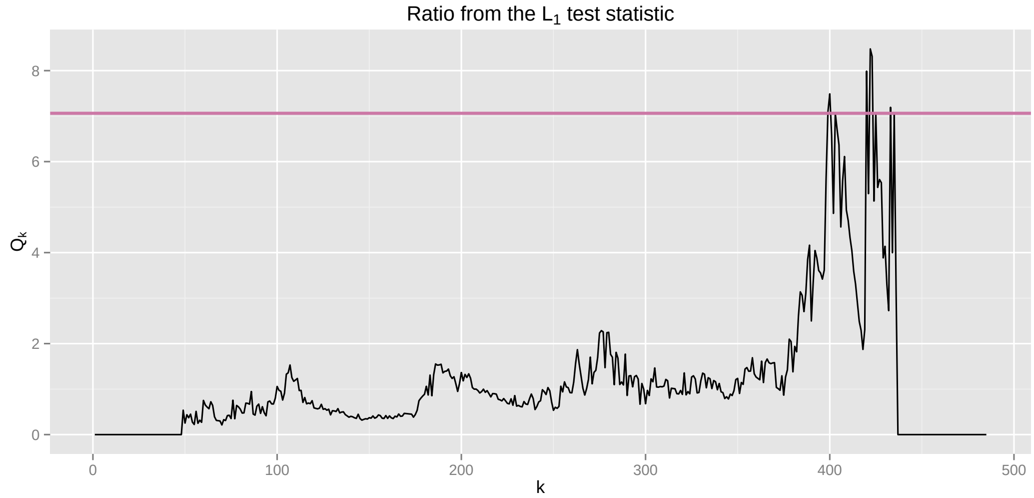


Figure 4.14: The values of Q_k from the L_1 test statistic for the surface temperature data. The simulated 95% critical value is depicted by the colored horizontal line.

the time series' observations for the statistic in the nominator, i.e.,

$$\widehat{\tau} = \arg\max_k \boldsymbol{S}_{k,n}^{\top}(\psi)\boldsymbol{C}_{1,n}^{-1}\boldsymbol{S}_{k,n}(\psi). \tag{4.18}$$

For the L_2 score function, we get $\widehat{\tau} = 171$. Using L_1 approach, we obtain $\widehat{\tau} = 212$.

The estimates of the regression parameters can be then obtained as

$$\boldsymbol{b}_{\widehat{\tau}} = \boldsymbol{C}_{1,\widehat{\tau}}^{-1}\sum_{i=1}^{\widehat{\tau}} \boldsymbol{h}\,(i/n)\,Y_i \quad \text{and} \quad \widetilde{\boldsymbol{b}}_{\widehat{\tau}} = \boldsymbol{C}_{\widehat{\tau}+1,n}^{-1}\sum_{i=\widehat{\tau}+1}^{n} \boldsymbol{h}\,(i/n)\,Y_i. \tag{4.19}$$

The fitted quadratic curves for the surface temperature data before and after the estimated change point are shown in Figure 4.15 for the L_2 method and in Figure 4.16 for the L_1 method.

Note that the estimated change points using the L_2 and L_1 method are not very close to each other. As a consequence, the estimated quadratic regression parameter corresponding to the fitted curve before the estimated change point using the L_2 method possesses the opposite sign compared to the estimated quadratic regression parameter corresponding to the fitted curve before the estimated change point using the L_1 method. Similarly for the estimated quadratic regression parameter corresponding to the fitted curve after the estimated change point. One of the possible reasons is that there exist more change points in such a long observation history.

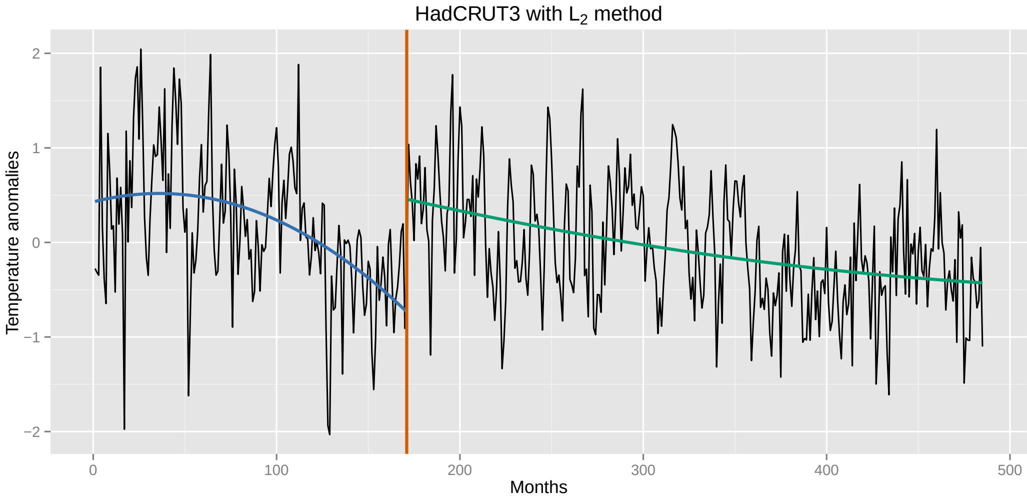


Figure 4.15: The surface temperature data analysed by the L_2 method. Estimated change point is depicted by the orange vertical line and estimated regression curves are drawn by the blue and green lines.

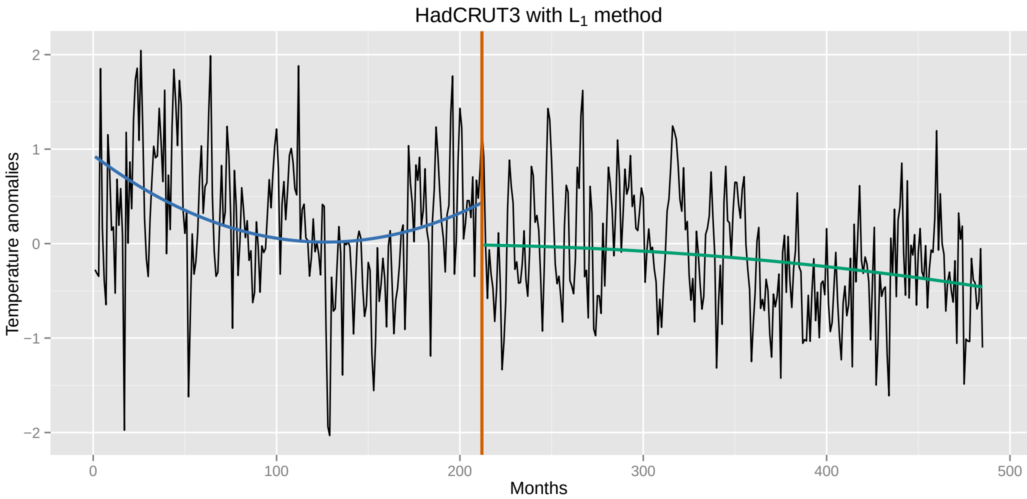


Figure 4.16: The surface temperature data analysed by the L_1 method. Estimated change point is depicted by the orange vertical line and estimated regression curves are drawn by the blue and green lines.

4.9 Application to ratings of clients data

We consider two time series obtained from a Czech loan providing company. Note that the demonstrated approaches shown below should not be considered as a complete data analysis, but only to show practical application of the derived methods and results.

Each time series contains daily averages of ratings of clients applying for a specific loan product. The first time series includes ratings of existing clients, the second one covers new clients. By ratings we mean assessments of clients' creditworthiness, based on credit scoring. Assessments are typically made shortly after filling an application (on the same day), using both data provided by the client and data that are already at the company's disposal. There are 377 business days available for both time series. Our question is whether the population of clients applying for this product changed in the last 18 months. We use only the L_2 approach for the detection of change in regression, since the L_1 method performs very similarly.

A linear regression model is assumed for the newcomers. The values of the ratio Q_k defined by (4.15) are plotted in Figure 4.17. The value of the ratio type test statistic for

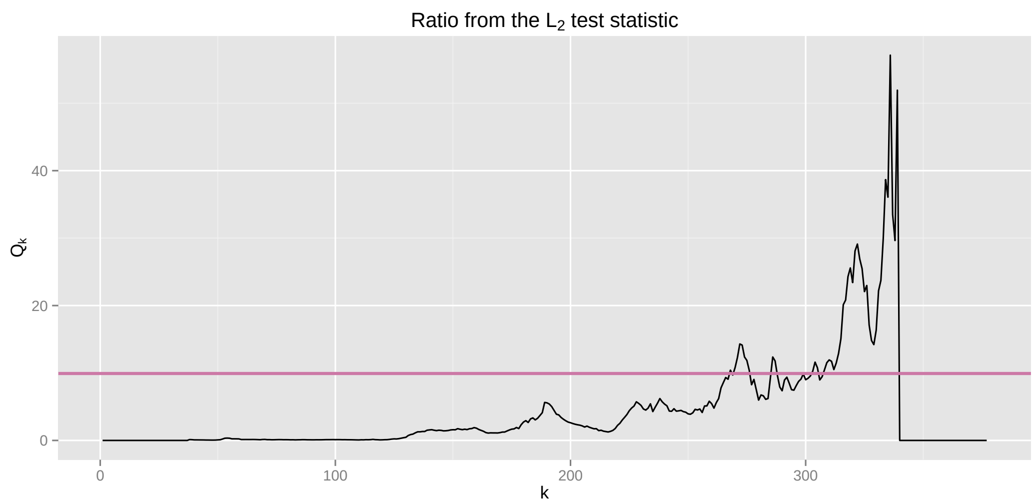


Figure 4.17: The values of Q_k from the L_2 test statistic for the rating of new clients data. The simulated 95% critical value is depicted by the colored horizontal line.

the change in regression is 57.12984, which is larger than the simulated 95% critical value of 9.923813. Therefore, we reject the null hypothesis of no change in the regression parameters. Moreover, we estimate the time of change τ as in (4.18), which yields $\widehat{\tau} = 234$. The estimates of the regression parameters can be then obtained from (4.19). The fitted linear lines for the rating of new clients data before and after the estimated change point are shown in Figure 4.18.

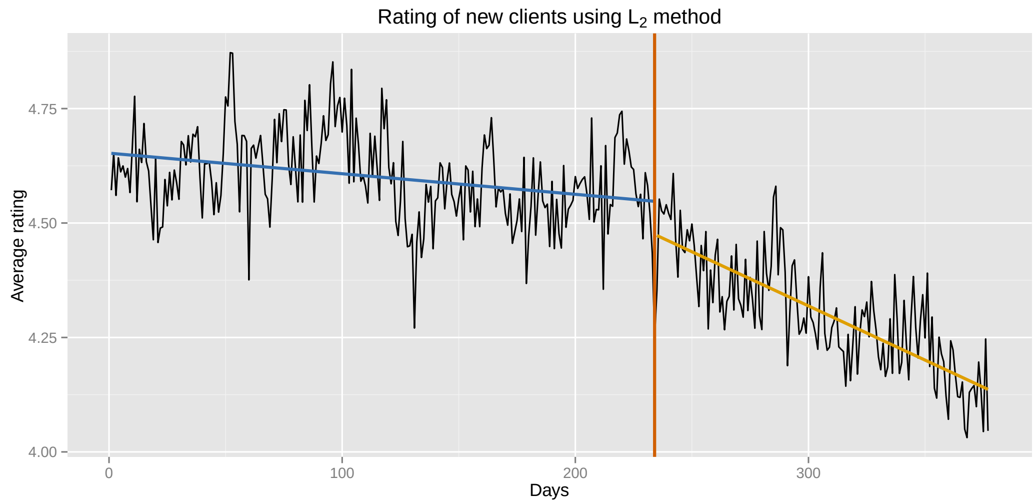


Figure 4.18: The rating of new clients data analysed by the L_2 method. Estimated change point is depicted by the orange vertical line and estimated regression lines are drawn by the blue and yellow lines.

A quadratic regression model is considered for data about the existing clients. The values of the ratio Q_k are visualized in Figure 4.19.

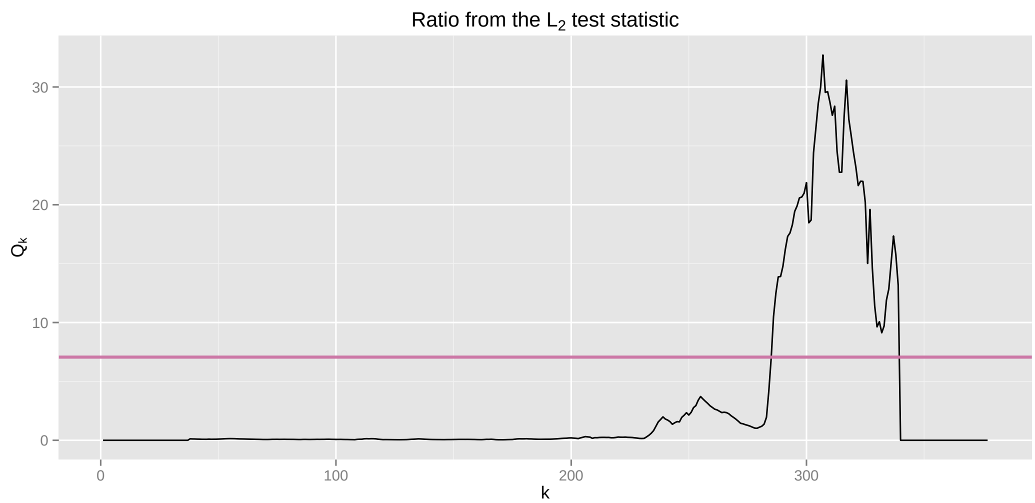


Figure 4.19: The values of Q_k from the L_2 test statistic for the rating of existing clients data. The simulated 95% critical value is depicted by the colored horizontal line.

The test statistic for the change in regression is $\mathcal{R}_{377}(\psi_{L_2}) = 32.71579$, which is larger than the simulated 95% critical value of 7.062320. Hence, we reject the null hypothesis of no change in the regression parameters. Besides that, we estimate the time of change τ as $\widehat{\tau} = 288$. The fitted quadratic curves for the rating of existing clients data before and after the estimated change point are shown in Figure 4.20. Permutation bootstrap extensions of the asymptotic techniques presented on these time series give the same conclusions.

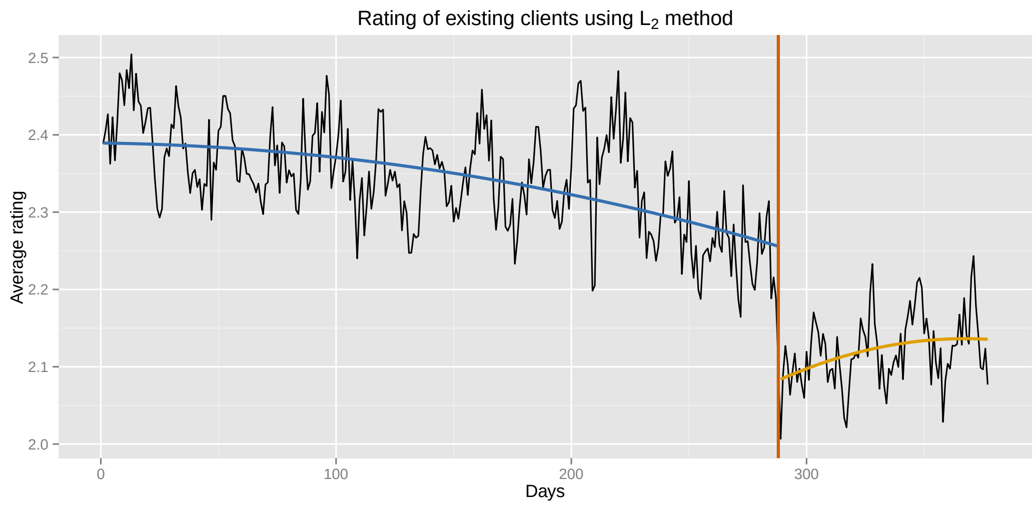


Figure 4.20: The rating of existing clients data analysed by the L_2 method. Estimated change point is depicted by the orange vertical line and estimated regression lines are drawn by the blue and yellow lines.

The detected change point in the existing clients' average ratings corresponds to a change in the company's business strategy. However, the detected change point for newcomers does not have any straightforward explanation. Maybe considering multiple changes would bring some deeper insight.

4.10 Future research

As it has already been pointed out, one of the possible extensions of the presented results is to derive testing procedures based on the asymptotics as well as on the bootstrapping for a change in the regression parameters in case of weakly dependent errors with a general score function.

Besides that, it would be preferable to increase the power of the bootstrap tests (with or without replacement) in case of moderate sample size (in this case, e.g., $n = 100$). A possible solution to this issue of low power can be a two-step testing procedure for detecting the change in regression: The first step would be a consistent estimation of the possible change point, i.e., to obtain $\widehat{\tau}_n$ such that $\widehat{\tau}_n - \tau_n = o_{\mathsf{P}}(1)$ as $n \to \infty$. The second step would consist of bootstrapping residuals

$$\widehat{e}_i := \begin{cases} \psi\left(Y_i - \boldsymbol{h}^\top(i/n)\boldsymbol{b}_{\widehat{\tau}_n}(\psi)\right), & i \leqslant \widehat{\tau}_n, \\ \psi\left(Y_i - \boldsymbol{h}^\top(i/n)\widetilde{\boldsymbol{b}}_{\widehat{\tau}_n}(\psi)\right), & i > \widehat{\tau}_n. \end{cases}$$

This approach will be implemented in forthcoming Section 6.7 (cf. Procedure 6.1) when testing a common change in panel means. The one-step bootstrap procedure derived in the similar manner as in this chapter, but used for the panel data problem introduced later on, gave very low power.

4.11 Summary

Procedures for detection of at most one *change in the regression parameters* of the regression model are considered. In particular, the test procedures based on the ratio type test statistics—that are functionals of partial sums of the residuals—are studied. Ratio type statistics are interesting for the fact, that in order to compute such statistics, there in no requirement to estimate the variance of the underlying model.

The presented methods generalize the traditional L_2 approach in constructions of the test statistics for the change detection by incorporating a general score function. To approximate the critical values for testing procedures, either *approximations of the limit distribution or resampling methods* are used. We concentrate on the permutation bootstrap. The asymptotic behavior of the proposed ratio type test statistics is studied under the null hypothesis as well as under local alternatives. Additionally, the justification for the permutation bootstrap method is given.

Chapter 5

Change in Autoregression Parameter

In the present chapter, we focus on *autoregressive time series of order one*, i.e., AR(1) series. We try to detect a possible change of the scalar parameter from a stationary autoregressive model using ratio type test statistics, which allows us to avoid estimating the unknown nuisance dispersion parameter of the time series.

The results are inspired by a paper by Hušková et al. (2007), where an autoregressive time series model of order p is taken into account and the whole vector of autoregression parameters is subject to change. The authors proposed to detect such change by computing partial sums of weighted residuals based on maximum type CUSUM test statistics. The results were consequently extended by the bootstrap approach in Hušková et al. (2008).

5.1 Autoregressive model with possibly changed parameter

We consider the time series model with a possible change in parameter after an unknown time point τ

$$Y_t = \beta Y_{t-1} + \delta Y_{t-1}\mathcal{I}\{t > \tau\} + \varepsilon_t, \quad t = 2, \dots n, \tag{5.1}$$

where β and $\delta \neq 0$ are fixed (not depending on n) unknown parameters, $1 < \tau = \tau_n \leqslant n$ is the unknown change point, and $\varepsilon_1, \dots, \varepsilon_n$ are iid random errors satisfying further conditions specified below.

We are going to test the null hypothesis that the autoregression parameter remained constant for the whole observation period

$$H_0 : \tau = n \tag{5.2}$$

against the alternative that a change of the autoregression parameter occurred at some unknown time point τ

$$H_1 : \tau < n, \ \delta \neq 0. \tag{5.3}$$

5.2 Test statistic for change in autoregression

We propose the following ratio type test statistic to detect the change in autoregression of order one

$$\mathcal{V}_n = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \frac{\max_{2\leqslant i\leqslant k} \left| \sum_{j=1}^{i-1} Y_j (Y_{j+1} - \widehat{\beta}_{1k} Y_j) \right|}{\max_{k+1\leqslant i\leqslant n-1} \left| \sum_{j=i}^{n-1} Y_j (Y_{j+1} - \widehat{\beta}_{2k} Y_j) \right|}, \tag{5.4}$$

where $0 < \gamma < 1/2$ is a given constant, $\widehat{\beta}_{1k}$ is an ordinary least squares estimate of parameter β based on observations $Y_1, \ldots, Y_k$ and $\widehat{\beta}_{2k}$ is an ordinary least squares estimate of β based on observations $Y_{k+1}, \ldots, Y_n$. Being more formal, estimate $\widehat{\beta}_{1k}$ is obtained when regressing vector of responses $\mathbf{y}_{1,k} := (Y_2, \ldots, Y_k)^\top$ on the vector of covariates $\mathbf{x}_{1,k} := (Y_1, \ldots, Y_{k-1})^\top$. Analogously, estimate $\widehat{\beta}_{2k}$ is obtained when regressing vector of responses $\mathbf{y}_{k+1,n} := (Y_{k+2}, \ldots, Y_n)^\top$ on the vector of regressors $\mathbf{x}_{k+1,n} := (Y_{k+1}, \ldots, Y_{n-1})^\top$.

The motivation for constructing the ratio type test statistic $\mathcal{V}_n$ comes from the linear regression setup (so-called normal equations). Estimate $\widehat{\beta}_{1k}$ is a solution of

$$\mathbf{x}_{1,k}^\top (\mathbf{y}_{1,k} - \mathbf{x}_{1,k} b) = 0$$

with respect to $b \in \mathbb{R}$ and estimate $\widehat{\beta}_{2k}$ is a solution of

$$\mathbf{x}_{k+1,n}^\top (\mathbf{y}_{k+1,n} - \mathbf{x}_{k+1,n} b) = 0$$

with respect to $b \in \mathbb{R}$. Therefore, we may define partial sums of weighted residuals as

$$\mathbf{x}_{1,i}^\top \left(\mathbf{y}_{1,i} - \mathbf{x}_{1,i} \widehat{\beta}_{1k} \right), \quad i = 2, \ldots, k$$

and

$$\mathbf{x}_{i,n}^\top \left(\mathbf{y}_{,n} - \mathbf{x}_{i,n} \widehat{\beta}_{2k} \right), \quad i = k+1, \ldots, n.$$

Consequently, these partial sums can be used as basis for the maxima of partial sums in the nominator and the denominator of $\mathcal{V}_n$.

Note that this approach—usage of the ratio type test statistics—can be generalized for the change of a vector autoregression parameter of the stationary autoregressive AR(p) process, when $p \geqslant 2$, using the notation from Hušková et al. (2007).

Before deriving asymptotic properties of the ratio type test statistic, we formulate several stochastic assumptions on time series model (5.1):

Assumption F1. $\beta \in (-1, 1)\backslash\{0\}$.

Assumption F2. $\beta + \delta \in (-1, 1)\backslash\{0\}$.

Assumption I1. $\{\varepsilon_i,\ i = 0, \pm 1, \ldots\}$ are iid random variables having $\mathsf{E}\,\varepsilon_i = 0$, $\mathsf{Var}\,\varepsilon_i = \sigma^2 > 0$, and $\mathsf{E}\,\varepsilon_i^4 < \infty$ for all i. Observation Y_1 is independent of $\{\varepsilon_2, \varepsilon_3, \ldots\}$.

Assumptions F1, F2, and I1 ensure that the time series is a stationary autoregressive sequence of order one (and not an iid sequence) before and even after the possible change point.

The limit behavior of the test statistic under the null hypothesis is characterized by the following theorem.

Theorem 5.1 (Under null)**.** *Suppose that* $Y_1, \ldots, Y_n$ *follow model* (5.1)*, assume that Assumptions F1 and I1 hold. Then, under null hypothesis* (5.2)

$$\mathcal{V}_n \xrightarrow[n\to\infty]{\mathscr{D}} \sup_{\gamma \leqslant t \leqslant 1-\gamma} \frac{\sup_{0 \leqslant u \leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)|}{\sup_{t \leqslant u \leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|}, \tag{5.5}$$

where $\{\mathcal{W}(x), 0 \leqslant x \leqslant 1\}$ *is a standard Wiener process and* $\widetilde{\mathcal{W}}(x) = \mathcal{W}(1) - \mathcal{W}(x)$.

Proof. Let us consider an array

$$U_{n,i} = \frac{\sqrt{1-\beta^2}}{\sigma^2\sqrt{n-1}} Y_{i-1}\varepsilon_i, \quad i = 2, \ldots, n$$

and a filtration $\mathcal{F}_{n,i} = \sigma\{\varepsilon_j,\ j \leqslant i\}$, $i = 2, \ldots, n$ and $n \in \mathbb{N}$. Then, $\{U_{n,i}, \mathcal{F}_{n,i}\}$ is a martingale difference array such that

$$\mathsf{E}\,U_{n,i}^2 = \frac{1-\beta^2}{\sigma^4(n-1)} \mathsf{E}\,Y_{i-1}^2\varepsilon_i^2 = \frac{1}{n-1}.$$

Moreover,

$$\sum_{i=2}^{n} U_{n,i}^2 - \sum_{i=2}^{n} \mathsf{E}\,U_{n,i}^2 = \frac{1-\beta^2}{\sigma^4(n-1)} \sum_{i=2}^{n} (Y_{i-1}^2\varepsilon_i^2 - \mathsf{E}\,Y_{i-1}^2\varepsilon_i^2).$$

Furthermore,

$$\frac{1}{n-1}\sum_{i=2}^{n}(Y_{i-1}^2\varepsilon_i^2-\mathsf{E}\,Y_{i-1}^2\varepsilon_i^2)=\frac{1}{n-1}\sum_{i=2}^{n}[Y_{i-1}^2(\varepsilon_i^2-\sigma^2)]+\frac{1}{n-1}\sum_{i=2}^{n}(Y_{i-1}^2-\mathsf{E}\,Y_{i-1}^2)\sigma^2.$$

Since $\{Y_{i-1}^2(\varepsilon_i^2-\sigma^2)\}$ is a martingale difference array again with respect to $\mathcal{F}_{n,i}$, we have under Assumption I1 from the Chebyshev's inequality that

$$\frac{1}{n-1}\sum_{i=2}^{n}[Y_{i-1}^2(\varepsilon_i^2-\sigma^2)]\xrightarrow[n\to\infty]{\mathsf{P}}0.$$

Similarly, as a consequence of Lemma 4.2 by Hušková et al. (2007),

$$\frac{1}{n-1}\sum_{i=2}^{n}(Y_{i-1}^2-\mathsf{E}\,Y_{i-1}^2)\xrightarrow[n\to\infty]{\mathsf{P}}0.$$

Thus,

$$\sum_{i=2}^{n}U_{n,i}^2\xrightarrow[n\to\infty]{\mathsf{P}}1. \tag{5.6}$$

Next, for any $\epsilon>0$,

$$\begin{aligned}\mathsf{P}\left(\max_{2\leqslant i\leqslant n}U_{n,i}^2>\epsilon\right)&\leqslant\sum_{i=2}^{n}\mathsf{P}\left(\frac{1-\beta^2}{\sigma^4(n-1)}Y_{i-1}^2\varepsilon_i^2>\epsilon\right)\\&\leqslant\frac{(1-\beta^2)^2}{\epsilon^2\sigma^8(n-1)^2}\sum_{i=2}^{n}\mathsf{E}\,Y_{i-1}^4\mathsf{E}\,\varepsilon_i^4\xrightarrow[n\to\infty]{}0.\end{aligned} \tag{5.7}$$

Additionally,

$$\lim_{n\to\infty}\sum_{i=2}^{[nt]}\mathsf{E}\,U_{n,i}^2=\lim_{n\to\infty}\frac{[nt]-1}{n-1}=t \tag{5.8}$$

for all $t\in[0,1]$.

According to Theorem 27.14 by Davidson (1994) for the martingale difference array $\{U_{n,i},\mathcal{F}_{n,i}\}$, where the assumptions of this theorem are satisfied due to (5.6), (5.7), and (5.8), we get

$$\sum_{i=2}^{[nt]}U_{n,i}\xrightarrow[n\to\infty]{\mathscr{D}[0,1]}\mathcal{W}(t).$$

Therefore,

$$\frac{1}{\sqrt{n-1}}\left(\sum_{i=2}^{[nt]} Y_{i-1}\varepsilon_i, \sum_{i=[nt]+2}^{n} Y_{i-1}\varepsilon_i\right) \xrightarrow[n\to\infty]{\mathscr{D}^2[0,1]} \frac{\sigma^2}{\sqrt{1-\beta^2}}\left(\mathcal{W}(t), \widetilde{\mathcal{W}}(t)\right), \tag{5.9}$$

where $\widetilde{\mathcal{W}}(t) = \mathcal{W}(1) - \mathcal{W}(t)$.

Let us define $\boldsymbol{Y}_{j,l} = (Y_j, \ldots, Y_l)^\top$ and $\boldsymbol{\varepsilon}_{j,l} = (\varepsilon_j, \ldots, \varepsilon_l)^\top$. Hence, for the expression from the nominator of $\mathcal{V}_n$ holds

$$\begin{aligned}
&\sum_{j=1}^{i-1} Y_j(Y_{j+1} - \widehat{\beta}_{1k} Y_j) = \boldsymbol{Y}_{1,i-1}^\top \left(\boldsymbol{Y}_{2,i} - \boldsymbol{Y}_{1,i-1}\widehat{\beta}_{1k}\right)\\
&= \boldsymbol{Y}_{1,i-1}^\top \left(\boldsymbol{Y}_{1,i-1}\beta + \boldsymbol{\varepsilon}_{2,i} - \boldsymbol{Y}_{1,i-1}\beta - \boldsymbol{Y}_{1,i-1}\left(\boldsymbol{Y}_{1,k-1}^\top \boldsymbol{Y}_{1,k-1}\right)^{-1} \boldsymbol{Y}_{1,k-1}^\top \boldsymbol{\varepsilon}_{2,k}\right)\\
&= \boldsymbol{Y}_{1,i-1}^\top \boldsymbol{\varepsilon}_{2,i} - \boldsymbol{Y}_{1,i-1}^\top \boldsymbol{Y}_{1,i-1}\left(\boldsymbol{Y}_{1,k-1}^\top \boldsymbol{Y}_{1,k-1}\right)^{-1} \boldsymbol{Y}_{1,k-1}^\top \boldsymbol{\varepsilon}_{2,k}.
\end{aligned} \tag{5.10}$$

Similarly for the expression from the denominator of $\mathcal{V}_n$

$$\begin{aligned}
&\sum_{j=i}^{n-1} Y_j(Y_{j+1} - \widehat{\beta}_{2k} Y_j)\\
&= \boldsymbol{Y}_{i,n-1}^\top \boldsymbol{\varepsilon}_{i+1,n} - \boldsymbol{Y}_{i,n-1}^\top \boldsymbol{Y}_{i,n-1}\left(\boldsymbol{Y}_{k+1,n-1}^\top \boldsymbol{Y}_{k+1,n-1}\right)^{-1} \boldsymbol{Y}_{k+1,n-1}^\top \boldsymbol{\varepsilon}_{k+2,n}.
\end{aligned} \tag{5.11}$$

Lemma 4.2 by Hušková et al. (2007) gives

$$\sup_{\gamma \leqslant t < 1} \frac{1}{[nt]} \left|\sum_{s=1}^{[nt]} (Y_s^2 - \mathsf{E}\, Y_s^2)\right| = o_{\mathsf{P}}(1) \tag{5.12}$$

and

$$\sup_{0 < t \leqslant 1-\gamma} \frac{1}{[n(1-t)]} \left|\sum_{s=[nt]+1}^{n-1} (Y_s^2 - \mathsf{E}\, Y_s^2)\right| = o_{\mathsf{P}}(1). \tag{5.13}$$

Finally, (5.9) together with (5.10), (5.11), (5.12), and (5.13) implies

$$\begin{aligned}
&\frac{1}{\sqrt{n-1}}\left(\begin{array}{c} \sup_{0\leqslant u\leqslant t} \left|\sum_{j=1}^{[nu]-1} Y_j(Y_{j+1} - \widehat{\beta}_{1[nt]} Y_j)\right| \\ \sup_{t\leqslant u\leqslant 1} \left|\sum_{j=[nu]+1}^{n-1} Y_j(Y_{j+1} - \widehat{\beta}_{2[nt]} Y_j)\right| \end{array}\right)\\
&\qquad \xrightarrow[n\to\infty]{\mathscr{D}^2[\gamma,1-\gamma]} \frac{\sigma^2}{\sqrt{1-\beta^2}}\left(\begin{array}{c} \sup_{0\leqslant u\leqslant t} |\mathcal{W}(u) - u/t\mathcal{W}(t)| \\ \sup_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right| \end{array}\right).
\end{aligned}$$

Then, the assertion of the theorem directly follows. □

The next theorem describes the test statistic's behavior under a fixed alternative.

Theorem 5.2 (Under alternative)**.** *Suppose that* $Y_1, \ldots, Y_n$ *follow model* (5.1)*, assume that alternative* (5.3) *holds for some fixed* $\delta \neq 0$*, and* $\tau = [\zeta n]$ *for some* $\gamma < \zeta < 1 - \gamma$*. Then, under Assumptions F1, F2, and I1*

$$\mathcal{V}_n \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Proof. Let us take $k = \tau + 2$, $k = [\xi n]$ for some $\zeta < \xi < 1 - \gamma$ and $i = \tau + 1$. Then,

$$\sum_{j=1}^{\tau} Y_j (Y_{j+1} - \widehat{\beta}_{1(\tau+2)} Y_j)$$
$$= \boldsymbol{Y}_{1,\tau}^{\top} \boldsymbol{\varepsilon}_{2,\tau+1} - \boldsymbol{Y}_{1,\tau}^{\top} \boldsymbol{Y}_{1,\tau} \left(\boldsymbol{Y}_{1,\tau+1}^{\top} \boldsymbol{Y}_{1,\tau+1}\right)^{-1} \boldsymbol{Y}_{1,\tau+1}^{\top} \boldsymbol{\varepsilon}_{2,\tau+2} - \boldsymbol{Y}_{1,\tau}^{\top} \boldsymbol{Y}_{1,\tau} \delta.$$

According to the proof of Theorem 5.1, as $n \to \infty$,

$$\frac{1}{\sqrt{n-1}} \left(\boldsymbol{Y}_{1,\tau}^{\top} \boldsymbol{\varepsilon}_{2,\tau+1} - \boldsymbol{Y}_{1,\tau}^{\top} \boldsymbol{Y}_{1,\tau} \left(\boldsymbol{Y}_{1,\tau+1}^{\top} \boldsymbol{Y}_{1,\tau+1}\right)^{-1} \boldsymbol{Y}_{1,\tau+1}^{\top} \boldsymbol{\varepsilon}_{2,\tau+2}\right) = O_{\mathsf{P}}(1).$$

Lemma 4.2 by Hušková et al. (2007) gives

$$\frac{1}{\sqrt{n-1}} \left|\boldsymbol{Y}_{1,\tau}^{\top} \boldsymbol{Y}_{1,\tau} \delta\right| \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

Now,

$$\frac{1}{\sqrt{n-1}} \max_{2 \leqslant i \leqslant k} \left|\sum_{j=1}^{i-1} Y_j (Y_{j+1} - \widehat{\beta}_{1k} Y_j)\right| \xrightarrow[n\to\infty]{\mathsf{P}} \infty.$$

For $\tau < k = [\xi n]$, the denominator in (5.4) divided by $\sqrt{n-1}$ has the same distribution as under the null hypothesis and it is, therefore, bounded in probability. It follows that the maximum of the ratio has to tend in probability to infinity as well, while $n \to \infty$. □

The previous theorem provides consistency of the studied test statistic under the given assumptions. The null hypothesis is rejected for large values of the ratio type test statistic. Being more formal, we reject H_0 at significance level α if $\mathcal{V}_n > v_{1-\alpha,\gamma}$, where $v_{1-\alpha,\gamma}$ is the $(1-\alpha)$-quantile of the asymptotic distribution (5.5).

5.3 Asymptotic critical values for the change in AR parameter

The explicit form of the limit distribution (5.5) is not known. However, the asymptotic distribution of $\mathcal{V}_n$ is the same as the asymptotic distribution from (3.8) for the test statistic

$\mathcal{A}_n(\psi)$, because the order of the considered autoregression time series model under the null hypothesis is just one. The critical values may be determined by simulation from the limit distribution from Theorem 5.1. Theorem 5.2 ensures that we reject the null hypothesis for large values of the test statistic. We tried to simulate the asymptotic distribution (5.5) by discretizing the Wiener process and using the relationship of a random walk to the Wiener process. We considered 1000 as the number of discretization points within $[0, 1]$ interval and the number of simulation runs equals to 100000. Higher numbers of discretization points and simulations were tried as well, but only negligible differences in the critical values were acquired. In Table 5.1, we present several critical values for $\gamma = 0.1$ and $\gamma = 0.2$.

	90%	95%	97.5%	99%
$\gamma = 0.1$	6.298815	7.293031	8.283429	9.589896
$\gamma = 0.2$	4.117010	4.745884	5.368286	6.159252

Table 5.1: Simulated critical values corresponding to the asymptotic distribution of the test statistic $\mathcal{V}_n$ under the null hypothesis.

5.4 Connection between test statistics for the change in regression and in autoregression

As it was already mentioned above when constructing the test statistic (5.4) for the change in autoregression, one can see $\mathcal{V}_n$ as an analogy to the ratio type test statistic (4.7) for the change in regression with the L_2 score function. To point out this informal similarity, let us imagine that the regression parameter from Chapter 4 is only one-dimensional corresponding to a linear trend without intercept. The vectors of weighted partial sums of residuals $\boldsymbol{S}_{j,k}(\psi)$ and $\widetilde{\boldsymbol{S}}_{j,k}(\psi)$ are just scalars and, furthermore, matrix $\boldsymbol{C}_{j,k}$ becomes a scalar as well. Now, let us replace the covariate functions $\boldsymbol{h}(i/n)$, $i = 1, \ldots k-1$ with the lagged time series observations $\mathbf{x}_{1,k}$. The regression setup with a system of covariate functions $\boldsymbol{h}$ on a compact interval is usually called *trending regression*, whereas we proceed to the *non-trending regression* setup satisfying that $\mathbf{x}_{1,k}^\top \mathbf{x}_{1,k}/k$ converges to a positive constant as $k \to \infty$.

After that, $\boldsymbol{S}_{j-1,k}(\psi_{L_2})$ can be seen as $\mathbf{x}_{1,j}^\top \left(\mathbf{y}_{1,j} - \mathbf{x}_{1,j}\widehat{\beta}_{1k}\right)$ and $\boldsymbol{C}_{1,k-1}$ becomes $\mathbf{x}_{1,k}^\top \mathbf{x}_{1,k}$. Similarly for the terms from the denominator of $\mathcal{R}_n(\psi_{L_2})$. Being informal, $\mathcal{R}_n(\psi_{L_2})$ becomes equivalent to

$$\widetilde{\mathcal{V}}_n = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \frac{\max\limits_{2 \leqslant i \leqslant k} \left[\mathbf{x}_{1,i}^\top \left(\mathbf{y}_{1,i} - \mathbf{x}_{1,i}\widehat{\beta}_{1k}\right)\right]^2 \left(\mathbf{x}_{1,k}^\top \mathbf{x}_{1,k}\right)^{-1}}{\max\limits_{k+1 \leqslant i \leqslant n-1} \left[\mathbf{x}_{i,n}^\top \left(\mathbf{y}_{i,n} - \mathbf{x}_{i,n}\widehat{\beta}_{2k}\right)\right]^2 \left(\mathbf{x}_{k+1,n}^\top \mathbf{x}_{k+1,n}\right)^{-1}}.$$

Being again informal, $\mathbf{x}_{1,k}^{\top}\mathbf{x}_{1,k} = \sum_{t=1}^{k-1} Y_t^2 \approx (k-1)\mathsf{E}\, Y_1^2$ under H_0. Hence, instead of $\sqrt{\widetilde{\mathcal{V}}_n}$, one may use

$$\breve{\mathcal{V}}_n = \max_{n\gamma \leqslant k \leqslant n-n\gamma} \sqrt{\frac{n-k-1}{k-1}} \frac{\max_{2\leqslant i\leqslant k} \left| \mathbf{x}_{1,i}^{\top} \left(\mathbf{y}_{1,i} - \mathbf{x}_{1,i}\widehat{\beta}_{1k} \right) \right|}{\max_{k+1\leqslant i\leqslant n-1} \left| \mathbf{x}_{i,n}^{\top} \left(\mathbf{y}_{i,n} - \mathbf{x}_{i,n}\widehat{\beta}_{2k} \right) \right|},$$

which is just a modification of the original test statistics $\mathcal{V}_n$ for the change in autoregression. The modification is in the same manner as in (3.15).

Note that the nominator and denominator both in the original test statistic $\mathcal{V}_n$ and in the modified one $\breve{\mathcal{V}}_n$ can be interchanged and can still be used for detection of the change in autoregression (but using different critical values).

5.5 Brief simulation study

A simulation experiment was performed to study the *finite sample properties* of the asymptotic test for the change in the AR(1) parameter. In particular, the interest lies in the *empirical size* of the proposed test under the null hypothesis and in the empirical *rejection rate* (power) under the alternative. Random samples (1000 each time) are generated from the time series change point model (5.1). The number of observations is set to $n = 150$ and $n = 300$ in order to demonstrate the performance of the testing approaches in case of different sample sizes. Two values of the autoregression parameter are taken into consideration, i.e., $\beta = 0.2$ and $\beta = 0.5$. The innovations are obtained as iid random variables from a standard normal $\mathsf{N}(0,1)$ or Student t_5 distribution. Simulation scenarios are produced as all possible combinations of the above mentioned settings.

To assess the theoretical results under H_0 numerically, Table 5.2 provides the empirical sizes (empirical probabilities of the type I error) of the test for change in the autoregression parameter, where the significance level is $\alpha = 0.05$. The proportion of rejecting the null

	α	0.01		0.05		0.10	
	innovations	$\mathsf{N}(0,1)$	t_5	$\mathsf{N}(0,1)$	t_5	$\mathsf{N}(0,1)$	t_5
$\beta = 0.2$	$n = 150$	0.113	0.220	0.212	0.341	0.292	0.413
	$n = 300$	0.062	0.097	0.155	0.221	0.235	0.294
$\beta = 0.5$	$n = 150$	0.119	0.173	0.231	0.296	0.302	0.385
	$n = 300$	0.067	0.121	0.160	0.231	0.231	0.300

Table 5.2: Empirical size of the test for the change in autoregression under H_0 using the asymptotic critical values of $\mathcal{V}_n$ with $\gamma = 0.1$, considering a significance level α. Innovations are iid having $\mathsf{N}(0,1)$ or t_5 distribution.

hypothesis is getting closer to the theoretical significance level as the number of time series' observations increases. Better performance of the test under the null hypothesis is observed, when the innovations have lighter tails. Note that the test statistic $\mathcal{V}_n$ is based on the L_2 regression approach. There is no visible direct effect of the value of the autoregression parameter (considering $\mathsf{N}(0,1)$ distributed innovations) on the empirical rejection rates based on this particular simulation study. Generally, the empirical sizes are higher than they should be. The same effect—the test rejects the null hypothesis more often than it should—can be seen in the detection of change in regression. Possible explanation of this difficulty can be slow convergence of the test statistics under the null hypothesis (see Section 4.7).

The performance of the testing procedure under H_1 in terms of the empirical rejection rates is shown in Table 5.3, where the change point is set to $\tau = n/2$ or $\tau = n/3$. The parameter δ is chosen as $\delta = 0.4$. We may conclude that the power of the test increases

		α	0.01		0.05		0.10	
		innovations	$\mathsf{N}(0,1)$	t_5	$\mathsf{N}(0,1)$	t_5	$\mathsf{N}(0,1)$	t_5
$\beta = 0.2$	$n = 150$	$\tau = n/2$	0.427	0.372	0.589	0.535	0.676	0.610
		$\tau = n/3$	0.409	0.325	0.571	0.485	0.662	0.564
	$n = 300$	$\tau = n/2$	0.515	0.456	0.691	0.655	0.773	0.729
		$\tau = n/3$	0.470	0.411	0.630	0.620	0.725	0.727
$\beta = 0.5$	$n = 150$	$\tau = n/2$	0.392	0.347	0.531	0.491	0.585	0.582
		$\tau = n/3$	0.311	0.257	0.453	0.392	0.536	0.485
	$n = 300$	$\tau = n/2$	0.477	0.433	0.632	0.623	0.734	0.717
		$\tau = n/3$	0.377	0.336	0.540	0.516	0.637	0.607

Table 5.3: Empirical power of the test for the change in autoregression under H_1 using the asymptotic critical values of $\mathcal{V}_n$ with $\gamma = 0.1$, considering a significance level α and $\delta = 0.4$. Innovations are iid having $\mathsf{N}(0,1)$ or t_5 distribution.

as the number of observations increases, which was expected. The test power drops when switching from a change point located in the middle of the time series to a change point closer to the beginning or the end of the time series. Innovations with heavier tails (i.e., t_5) yield slightly smaller power than innovations with lighter tails. There is again no visible effect of the value of autoregression parameter β on the power of the test.

In contrast to the slightly lower power in case of relatively small sample size and moderate change in the autoregression parameter, one may try to consider larger change in β from -0.8 to 0.8 in case of $n = 150$. Here, the simulated power reaches 0.994 (for $\alpha = 0.05$). Hence, in case of a large change in autoregression, the test achieves high power.

To improve the computational performance of the test for detecting the change in autoregression, longer time series of observations are a general solution. Moreover, a suitable bootstrap extension of the developed procedures could be helpful from a numerical and computational point of view.

5.6 Application to stock exchange index

As an illustrative example of the proposed technique for detecting of the change in autoregression, we concentrate on the Prague Stock Exchange index called *PX Index (formerly PX50)*. It is a capitalization-weighted index of major stocks that trade on the Prague Stock Exchange.

The starting exchange day for the Index PX50 was April 5, 1994. We consider a time series consisting of daily PX50 values starting from November 16, 1994 up to September 27, 2001. Only business days were taken into account, providing 1850 observations. The starting date of the observation period was chosen later than the starting day of the exchange, since only weekly (not daily) values of PX50 records were available at the beginning. Moreover, the market after opening the exchange was not as stable as later on. The last observation date was chosen in order to avoid effects of the attacks on September 11, 2001. The considered time series can be seen in Figure 5.1.

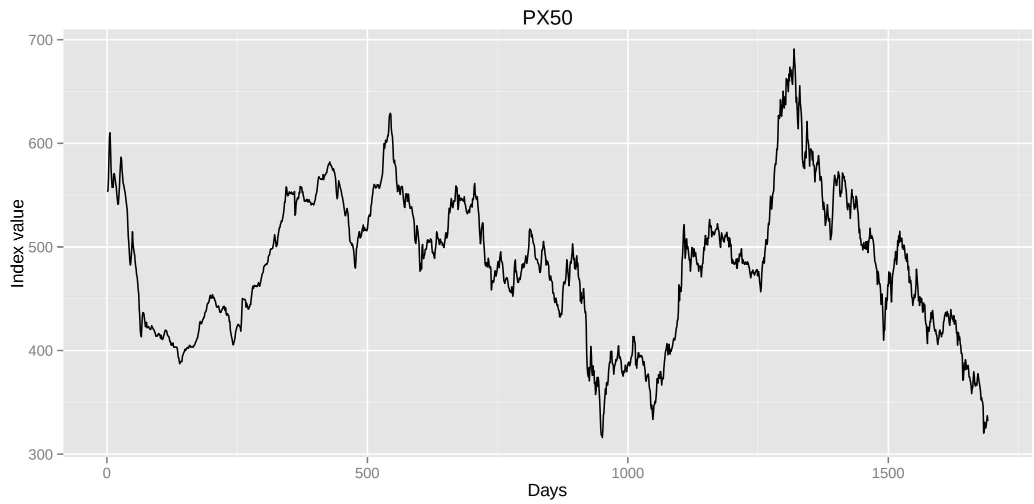


Figure 5.1: Daily Prague Stock Exchange index (PX50) values from November 16, 1994 to September 27, 2001.

I would like to thank doc. RNDr. Zuzana Prášková, CSc. from Charles University in Prague for pointing out the interesting nature of this data set and providing the data. The PX50 data can also be downloaded from the Prague Stock Exchange (2015) webpage.

We denote the original data of the PX50 index as $\{X_t\}_t$. Firstly, we transform the

PX50 index by taking into account the differences of logarithms, i.e, $Y_t = \log(X_t/X_{t-1})$. This transformation can be interpreted as considering logarithms of daily returns of the PX50 index. Besides that, using this approach stationary time series before and even after a possible change point are obtained. The transformed index values are shown in Figure 5.3.

Let us assume that $Y_1, \ldots, Y_n$ follow autoregressive change point model (5.1). We are going to decide whether the change in the AR(1) parameter occurred or not based on the proposed asymptotic test. The value of the test statistic $\mathcal{V}_n$ for $\gamma = 0.1$ is 7.321143, which is larger than the 95%-critical value 7.293031 simulated from the limit distribution under the null hypothesis. Therefore, we reject the null hypothesis of no change in the autoregressive parameter. The progress of the ratio of the test statistic

$$Q_k = \frac{\max\limits_{2\leqslant i\leqslant k} \left|\sum_{j=1}^{i-1} Y_j(Y_{j+1} - \widehat{\beta}_{1k}Y_j)\right|}{\max\limits_{k+1\leqslant i\leqslant n-1} \left|\sum_{j=i}^{n-1} Y_j(Y_{j+1} - \widehat{\beta}_{2k}Y_j)\right|}, \quad n\gamma \leqslant k \leqslant n - n\gamma$$

is depicted in Figure 5.2.

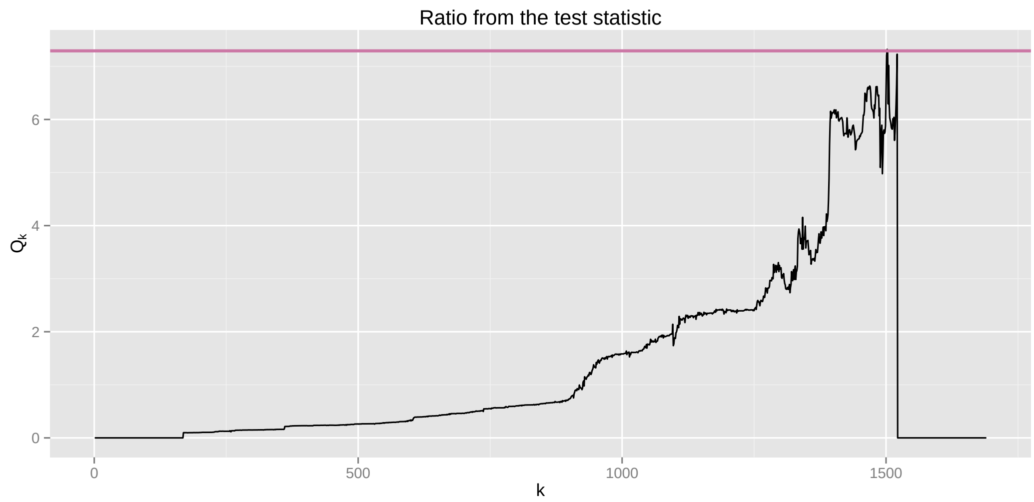


Figure 5.2: The values of Q_k for the PX50 index data with $\gamma = 0.1$. The colored horizontal line represents the 95%-critical value.

We may also estimate the unknown change point τ in a similar fashion as described in Chapter 1 by

$$\widehat{\tau} = \arg\max_{2\leqslant k\leqslant n} \left|\sum_{j=1}^{k-1} Y_j(Y_{j+1} - \widehat{\beta}_{1n}Y_j)\right|.$$

This leads to $\widehat{\tau} = 949$, which corresponds to October 7, 1998. The log returns of PX50 together with the depicted change point for the change in autoregression are displayed in Figure 5.3.

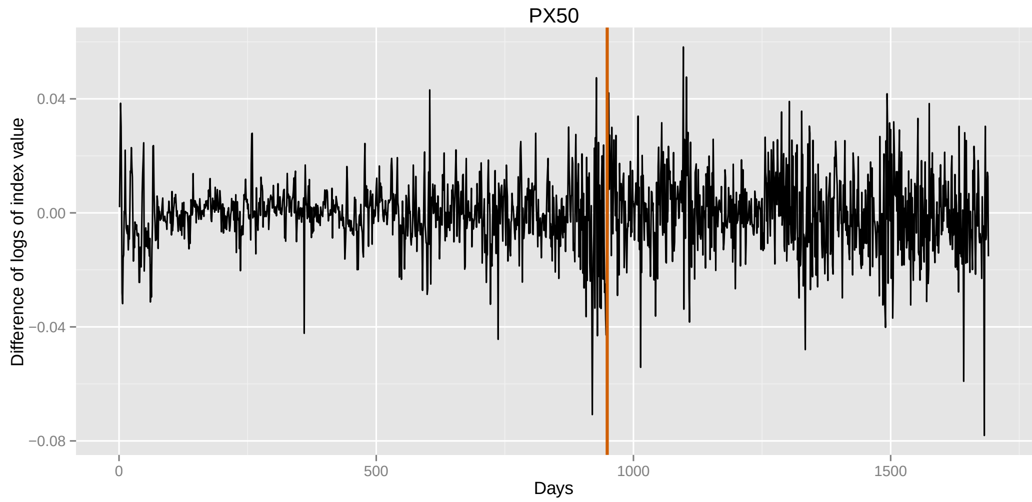


Figure 5.3: Logs of ratios of index PX50 values.

The explanation of the detected change in autoregression is possibly connected to the Russian financial crisis (also called Ruble crisis) that hit Russia on August 17, 1998. It resulted in the Russian government and the Russian Central Bank devaluing the ruble and defaulting on its debt. In 1998 influenced by Russian financial crisis, the index reached its historical bottom on October 8 with 316 points, which is the first day after the detected change in autoregression of the PX50 log returns.

Finally, we investigated the ACF (autocorrelation function) and PACF (partial autocorrelation function) plots of the time series before and after the estimated change point. Both ACF plots go to zero at an exponential rate, while both PACF plots become zero immediately after the first lag. We applied the Ljung-Box test on the residuals of the fitted AR(1) models (before and after the change). The hypothesis that the residuals in each AR(1) model have no autocorrelation is rejected in both cases, which suggests that the two series are stationary.

5.7 Summary

We investigate a possible change in the time series model, where the situation of no change corresponds to the fact that the considered sequence is a stationary AR(1) process. The alternative situation of the change present means that the time series is an AR(1) process up to some unknown time point and it is again an AR(1) process after that unknown time point, but the autoregression parameter is different.

The testing procedure for the change in autoregression can be viewed as an *analogy to the testing procedure for the change in regression.* The asymptotic behavior of the ratio type test statistic for the change in autoregression is investigated under the null hypothesis as well as under the alternative. The theoretical limiting distribution under the null hypothesis provides critical values for the test, which are obtained by simulations. A brief simulation study is conducted to perform the numerical performance of the proposed testing approach. Finally, an application of the developed procedure on the stock exchange index data is performed.

Chapter 6

Common Change in Panel Data

Until this moment, we have considered a change point in a sequence of observations, where only one stochastic copy of the sequence is available. We focus now on a change point problem, where several sequences are subject to change simultaneously.

Our interest lies in panel data that consist of a moderate or relatively large number of panels, while each of the panels contain a small number of observations. In this chapter, we establish testing procedures to detect a possible common change in means of the panels. We again consider a ratio type test statistic and derive its asymptotic distribution under the no change null hypothesis for the panel change point model. Moreover, we prove the consistency of the test under the alternative. A bootstrap technique is proposed as an add-on to the testing procedure based on asymptotics in order to make our approach completely data driven without any tuning parameters. The validity of the bootstrap algorithm is shown. As a by-product of the developed tests, we introduce a common break point estimate and prove its consistency. This chapter is based on paper Peštová and Pešta (2015).

6.1 Introduction

The problem of an unknown common change in means of the panels is studied here, where the panel data consist of N panels and each panel contains T observations over time. Various values of the change are possible for each panel at some unknown common time $\tau = 1, \ldots, N$. The panels are considered to be independent, but this restriction can be weakened. In spite of that, observations within the panel are usually not independent. It is supposed that a common unknown dependence structure is present over the panels.

Tests for change point detection in the panel data have been proposed only in case when the panel size T is sufficiently large, i.e., T increases over all limits from an asymptotic point of view, cf. Chan et al. (2013) or Horváth and Hušková (2012). However, the change point estimation has already been studied for finite T not depending on the number of panels

N, see Bai (2010). The remaining task is to develop testing procedures to decide whether a common change point is present or not in the panels, while taking into account that the length T of each observation regime is fixed and can be relatively small.

6.2 Motivation

Structural changes in panel data—especially *common breaks in means*—are wide spread phenomena. Our primary motivation comes from non-life insurance business, where associations in many countries uniting several insurance companies collect claim amounts paid by every insurance company each year. Such a database of cumulative claim payments can be viewed as panel data, where insurance company $i = 1, \ldots, N$ provides the total claim amount $Y_{i,t}$ paid in year $t = 1, \ldots, T$ into the common database. The members of the association can consequently profit from the joint database.

For the whole association it is important to know, whether a possible change in the claim amounts occurred during the observed time horizon. Usually, the time period is relatively short, e.g., 10–15 years. To be more specific, a widely used and very standard actuarial method for predicting future claim amounts—called chain ladder—assumes a kind of stability of the historical claim amounts. The formal necessary and sufficient condition is derived in Pešta and Hudecová (2012). This chapter shows a way how to test for a possible historical instability.

6.3 Panel change point model

Let us consider the panel change point model

$$Y_{i,t} = \mu_i + \delta_i \mathcal{I}\{t > \tau\} + \sigma\varepsilon_{i,t}, \quad 1 \leqslant i \leqslant N, 1 \leqslant t \leqslant T; \tag{6.1}$$

where $\sigma > 0$ is an unknown variance-scaling parameter and T is fixed, not depending on N. The possible *common change point time* is denoted by $\tau \in \{1, \ldots, T\}$. A situation where $\tau = T$ corresponds to *no change* in means of the panels. The means μ_i are panel-individual. The amount of the break in mean, which can also differ for every panel, is denoted by δ_i. Furthermore, it is assumed that the sequences of panel disturbances $\{\varepsilon_{i,t}\}_t$ are independent and within each panel the errors form a weakly stationary sequence with a common correlation structure. This can be formalized in the following assumption.

Assumption P1. The vectors $[\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]^\top$ existing on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ are *iid* for $i = 1, \ldots, N$ with $\mathsf{E}\,\varepsilon_{i,t} = 0$ and $\mathsf{Var}\,\varepsilon_{i,t} = 1$, having the autocorrelation function

$$\rho_t = \mathsf{Corr}\,(\varepsilon_{i,s}, \varepsilon_{i,s+t}) = \mathsf{Cov}\,(\varepsilon_{i,s}, \varepsilon_{i,s+t}), \quad \forall s \in \{1, \ldots, T-t\},$$

which is independent of the lag s, the cumulative autocorrelation function

$$r(t) = \mathsf{Var} \sum_{s=1}^{t} \varepsilon_{i,s} = \sum_{|s|<t} (t - |s|)\rho_s,$$

and the shifted cumulative correlation function

$$R(t, v) = \mathsf{Cov} \left(\sum_{s=1}^{t} \varepsilon_{i,s}, \sum_{u=t+1}^{v} \varepsilon_{i,u} \right) = \sum_{s=1}^{t} \sum_{u=t+1}^{v} \rho_{u-s}, \quad t < v$$

for all $i = 1, \ldots, N$ and $t, v = 1, \ldots, T$.

The sequence $\{\varepsilon_{i,t}\}_{t=1}^{T}$ can be viewed as a part of a *weakly stationary* process. Note that the dependent errors within each panel do not necessarily need to be linear processes. For example, GARCH processes as error sequences are allowed as well. The assumption of independent panels can indeed be relaxed, but it would make the setup much more complex. Consequently, probabilistic tools for dependent data need to be used (e.g., suitable versions of the central limit theorem). Nevertheless, assuming, that the claim amounts for different insurance companies are independent, is reasonable. Moreover, the assumption of a common homoscedastic variance parameter σ can be generalized by introducing weights $w_{i,t}$, which are supposed to be known. Being particular in actuarial practice, it would mean to normalize the total claim amount by the premium received, since bigger insurance companies are expected to have higher variability in total claim amounts paid.

It is required to test the *null hypothesis* of no change in the means

$$H_0 : \tau = T \tag{6.2}$$

against the *alternative* that at least one panel has a change in mean

$$H_1 : \tau < T \quad \text{and} \quad \exists i \in \{1, \ldots, N\} : \delta_i \neq 0. \tag{6.3}$$

A graphical illustration of the change point model (6.1) in panel data under the alternative, where the means change, can be seen in Figure 6.1.

6.4 Test statistic and asymptotic results

We propose a *ratio type statistic* to test H_0 against H_1, because this type of statistic does not require estimation of the nuisance parameter for the variance. Generally, this is due to the fact that the variance parameter simply cancels out from the nominator and denominator of the statistic. In spite of that, the common variance could be estimated from all the panels, of which we possess a sufficient number. Nevertheless, we aim to construct a valid and completely data driven testing procedure without interfering estimation and plug-in

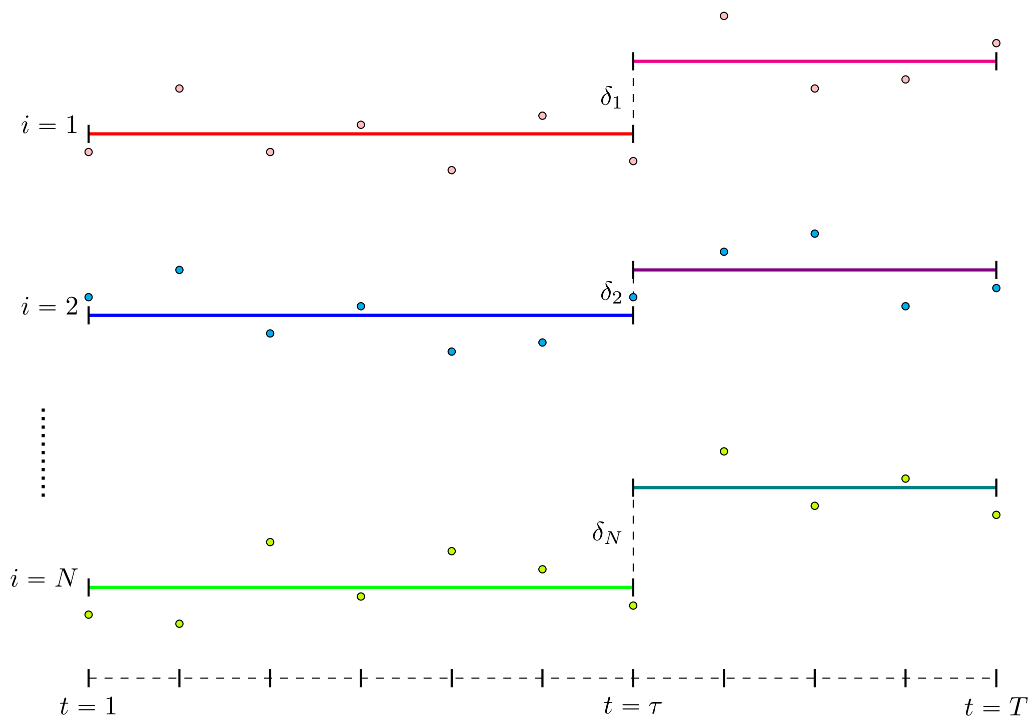


Figure 6.1: Illustration of the common change point problem in panel data.

estimates instead of nuisance parameters. A bootstrap add-on is going to serve this purpose as it is seen later on.

For surveys on ratio type test statistics, we refer to Chen and Tian (2014), Csörgő and Horváth (1997), Horváth et al. (2008), Liu et al. (2008), and Madurkayová (2011). Our particular panel change point test statistic is

$$\mathcal{P}_N(T) = \max_{t=2,\dots,T-2} \frac{\max_{s=1,\dots,t} \left| \sum_{i=1}^N \left[\sum_{r=1}^s \left(Y_{i,r} - \bar{Y}_{i,t} \right) \right] \right|}{\max_{s=t,\dots,T-1} \left| \sum_{i=1}^N \left[\sum_{r=s+1}^T \left(Y_{i,r} - \widetilde{Y}_{i,t} \right) \right] \right|},$$

where $\bar{Y}_{i,t}$ is the average of the first t observations in panel i and $\widetilde{Y}_{i,t}$ is the average of the last $T-t$ observations in panel i, i.e.,

$$\bar{Y}_{i,t} = \frac{1}{t} \sum_{s=1}^t Y_{i,s} \quad \text{and} \quad \widetilde{Y}_{i,t} = \frac{1}{T-t} \sum_{s=t+1}^T Y_{i,s}.$$

An alternative way for testing the change in panel means could be a usage of CUSUM type statistics. For example, a maximum or minimum of a sum (not a ratio) of properly standardized or modified sums from our test statistic $\mathcal{P}_N(T)$. The theory, which follows, can be appropriately rewritten for such cases.

Firstly, we derive the behavior of the test statistics under the null hypothesis.

Theorem 6.1 (Under null)**.** *Suppose that panel data* $\{Y_{i,t}\}_{i,t=1}^{N,T}$ *follow model* (6.1)*. Under null hypothesis* (6.2) *and Assumption P1*

$$\mathcal{P}_N(T) \xrightarrow[N\to\infty]{\mathscr{D}} \max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t}\left|X_s - \frac{s}{t}X_t\right|}{\max_{s=t,\ldots,T-1}\left|Z_s - \frac{T-s}{T-t}Z_t\right|}, \tag{6.4}$$

where $Z_t := X_T - X_t$ *and* $[X_1,\ldots,X_T]^\top$ *is a multivariate normal random vector with zero mean and covariance matrix* $\mathbf{\Lambda} = \{\lambda_{t,v}\}_{t,v=1}^{T,T}$ *such that*

$$\lambda_{t,t} = r(t) \quad \textit{and} \quad \lambda_{t,v} = r(t) + R(t,v),\ t < v.$$

Proof. Let us define

$$U_N(t) := \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\sum_{s=1}^{t}(Y_{i,s}-\mu_i).$$

Using the multivariate Lindeberg-Lévy CLT for a sequence of T-dimensional iid random vectors $\{[\sum_{s=1}^{1}\varepsilon_{i,s},\ldots,\sum_{s=1}^{T}\varepsilon_{i,s}]^\top\}_{i\in\mathbb{N}}$, we have under H_0

$$[U_N(1),\ldots,U_N(T)]^\top \xrightarrow[N\to\infty]{\mathscr{D}} [X_1,\ldots,X_T]^\top,$$

since $\mathsf{Var}\,[\sum_{s=1}^{1}\varepsilon_{1,s},\ldots,\sum_{s=1}^{T}\varepsilon_{1,s}]^\top = \mathbf{\Lambda}$. Indeed, the t-th diagonal element of the covariance matrix $\mathbf{\Lambda}$ is

$$\mathsf{Var}\sum_{s=1}^{t}\varepsilon_{1,s} = r(t)$$

and the upper off-diagonal element on position (t,v) is

$$\begin{aligned}\mathsf{Cov}\left(\sum_{s=1}^{t}\varepsilon_{1,s},\sum_{u=1}^{v}\varepsilon_{1,u}\right) &= \mathsf{Var}\sum_{s=1}^{t}\varepsilon_{1,s} + \mathsf{Cov}\left(\sum_{s=1}^{t}\varepsilon_{1,s},\sum_{u=t+1}^{v}\varepsilon_{1,u}\right)\\ &= r(t)+R(t,v), \quad t<v.\end{aligned}$$

Moreover, let us define the reverse analogue to $U_N(t)$, i.e.,

$$V_N(t) := \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\sum_{s=t+1}^{T}(Y_{i,s}-\mu_i) = U_N(T)-U_N(t).$$

Hence,

$$U_N(s) - \frac{s}{t}U_N(t) = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\left\{\sum_{r=1}^{s}\left[(Y_{i,r}-\mu_i) - \frac{1}{t}\sum_{v=1}^{t}(Y_{i,v}-\mu_i)\right]\right\}$$
$$= \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\sum_{r=1}^{s}\left(Y_{i,r}-\bar{Y}_{i,t}\right)$$

and, consequently,

$$V_N(s) - \frac{T-s}{T-t}V_N(t) = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\left\{\sum_{r=s+1}^{T}\left[(Y_{i,r}-\mu_i) - \frac{1}{T-t}\sum_{v=t+1}^{T}(Y_{i,v}-\mu_i)\right]\right\}$$
$$= \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\sum_{r=s+1}^{T}\left(Y_{i,r}-\widetilde{Y}_{i,t}\right).$$

Using the continuous mapping theorem, we end up with

$$\max_{t=2,\ldots,T-2}\frac{\max_{s=1,\ldots,t}\left|U_N(s)-\frac{s}{t}U_N(t)\right|}{\max_{s=t,\ldots,T-1}\left|V_N(s)-\frac{T-s}{T-t}V_N(t)\right|}$$
$$\xrightarrow[N\to\infty]{\mathscr{D}}\max_{t=2,\ldots,T-2}\frac{\max_{s=1,\ldots,t}\left|X_s-\frac{s}{t}X_t\right|}{\max_{s=t,\ldots,T-1}\left|(X_T-X_s)-\frac{T-s}{T-t}(X_T-X_t)\right|}.$$

□

The limiting distribution does not depend on the variance nuisance parameter σ, but it depends on the unknown correlation structure of the panel change point model, which has to be estimated for testing purposes. The way of its estimation is shown in Section 6.6. Furthermore, Theorem 6.1 is just a theoretical mid-step for the bootstrap test, where the correlation structure need not to be known. That is why the presence of unknown quantities in the asymptotic distribution is not troublesome.

Note that in case of independent observations within the panel, the correlation structure and, hence, the covariance matrix $\mathbf{\Lambda}$ is simplified such that $r(t) = t$ and $R(t, v) = 0$.

Next, we show how the test statistic behaves under the alternative.

Assumption P2. $\lim_{N\to\infty}\frac{1}{\sqrt{N}}\left|\sum_{i=1}^{N}\delta_i\right| = \infty$.

Theorem 6.2 (Under alternative)**.** *Suppose that panel data* $\{Y_{i,t}\}_{i,t=1}^{N,T}$ *follow model* (6.1). *If* $\tau \leqslant T-3$, *then under Assumptions P1, P2 and alternative* (6.3)

$$\mathcal{P}_N(T)\xrightarrow[N\to\infty]{\mathsf{P}}\infty. \tag{6.5}$$

Proof. Let $t = \tau + 1$. Then, under alternative H_1

$$\begin{aligned}
&\frac{1}{\sigma\sqrt{N}} \max_{s=1,\ldots,\tau+1} \left| \sum_{i=1}^{N} \left[\sum_{r=1}^{s} \left(Y_{i,r} - \bar{Y}_{i,\tau+1} \right) \right] \right| \\
&\geqslant \frac{1}{\sigma\sqrt{N}} \left| \sum_{i=1}^{N} \sum_{r=1}^{\tau} \left(Y_{i,r} - \bar{Y}_{i,\tau+1} \right) \right| \\
&= \frac{1}{\sigma\sqrt{N}} \left| \sum_{i=1}^{N} \sum_{r=1}^{\tau} \left(\mu_i + \sigma\varepsilon_{i,r} - \frac{1}{\tau+1} \sum_{v=1}^{\tau+1} (\mu_i + \sigma\varepsilon_{i,v}) - \frac{1}{\tau+1}\delta_i \right) \right| \\
&= \frac{1}{\sqrt{N}} \left| \sum_{i=1}^{N} \sum_{r=1}^{\tau} (\varepsilon_{i,r} - \bar{\varepsilon}_{i,\tau+1}) - \frac{\tau}{\sigma(\tau+1)} \sum_{i=1}^{N} \delta_i \right| \\
&= O_{\mathsf{P}}(1) + \frac{\tau}{\sigma(\tau+1)\sqrt{N}} \left| \sum_{i=1}^{N} \delta_i \right| \xrightarrow{\mathsf{P}} \infty, \quad N \to \infty,
\end{aligned}$$

where $\bar{\varepsilon}_{i,\tau+1} = \frac{1}{\tau} \sum_{v=1}^{\tau+1} \varepsilon_{i,v}$.

Since there is no change after $\tau + 1$ and $\tau \leqslant T - 3$, then by Theorem 6.1 we have

$$\frac{1}{\sigma\sqrt{N}} \max_{s=\tau+1,\ldots,T-1} \left| \sum_{i=1}^{N} \sum_{r=s+1}^{T} \left(Y_{i,r} - \widetilde{Y}_{i,\tau+1} \right) \right| \xrightarrow[N\to\infty]{\mathscr{D}} \max_{s=\tau+1,\ldots,T-1} \left| Z_s - \frac{T-s}{T-\tau} Z_{\tau+1} \right|.$$

□

Assumption P2 is satisfied, for instance, if $0 < \delta \leqslant \delta_i$, $\forall i$ (a common lower change point threshold) and $\delta\sqrt{N} \to \infty$, $N \to \infty$. Another suitable example of δ_is for the condition in Assumption P2, can be $0 < \delta_i = KN^{-1/2+\eta}$, $\forall i$ for some $K > 0$ and $\eta > 0$. Or $\delta_i = Ci^{\alpha-1}\sqrt{N}$, $\forall i$ may be used as well, where $\alpha \geqslant 0$ and $C > 0$. The assumption $\tau \leqslant T - 3$ means that there are at least three observations in the panel after the change point. It is also possible to redefine the test statistic by interchanging the nominator and the denominator of $\mathcal{P}_N(T)$. Afterwards, Theorem 6.2 for the modified test statistic would require three observations before the change point, i.e., $\tau \geqslant 3$.

Theorem 6.2 says that in presence of a structural change in the panel means, the test statistic explodes above all bounds. Hence, the procedure is consistent and the asymptotic distribution from Theorem 6.1 can be used to construct the test. The null hypothesis is rejected for large values of $\mathcal{P}_N(T)$. Hence, we reject H_0 at significance level α if and only if $\mathcal{P}_N(T) > p_{1-\alpha}$, where $p_{1-\alpha}$ is the $(1-\alpha)$-quantile of the asymptotic distribution (6.4).

6.5 Change point estimation

Despite the fact that the aim of the chapter is to establish testing procedures for detection of a panel mean change, it is necessary to construct a *consistent estimate* for a possible change point. There are two reasons for that: Firstly, the estimation of the covariance matrix

$\mathbf{\Lambda}$ from Theorem 6.1 requires panels as vectors with elements having common mean (i.e., without a jump). Secondly, the bootstrap procedure, introduced later on, requires centered residuals to be resampled.

A consistent estimate of the change point in the panel data is proposed in Bai (2010), but under circumstances that the change occurred for sure. In our situation, we do not know whether a change occurs or not. Therefore, we modify the estimate proposed by Bai (2010) in the following way. If the panel means change somewhere inside $\{2,\ldots,T-1\}$, let the estimate consistently select this change. If there is no change in panel means, the estimate points out the very last time point T with probability going to one. In other words, the value of the change point estimate can be T meaning no change. This is in contrast with Bai (2010), where T is not reachable.

Let us define the estimate of τ:

$$\widehat{\tau}_N := \arg\max_{t=2,\ldots,T} \frac{1}{t}\sum_{i=1}^{N}\sum_{s=1}^{t}(Y_{i,s}-\bar{Y}_{i,t})^2. \tag{6.6}$$

Now, we show the desired property of consistency for the proposed change point estimate under the following assumptions.

Assumption C1. $L < \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 < \infty$, where $L=-\infty$ if $\tau=T$ and $L=\max_{t=\tau+1,\ldots,T}\frac{\sigma^2 t^2}{\tau(t-\tau)}\left(\frac{r(\tau)}{\tau^2}-\frac{r(t)}{t^2}\right)$ otherwise.

Assumption C2. $\mathsf{E}\,\varepsilon_{1,t}^4<\infty$, $t\in\{1,\ldots,T\}$.

Theorem 6.3 (Change point estimate consistency)**.** *Suppose that panel data* $\{Y_{i,t}\}_{i,t=1}^{N,T}$ *follow model* (6.1)*. Assume that* $\tau\neq 1$ *and the sequence* $\{r(t)/t^2\}_{t=2}^{T}$ *is decreasing. Then under Assumptions P1, C1, and C2*

$$\lim_{N\to\infty}\mathsf{P}\left[\widehat{\tau}_N=\tau\right]=1.$$

Proof. Let us define

$$S_N^{(i)}(t) := \frac{1}{t}\sum_{s=1}^{t}(Y_{i,s}-\bar{Y}_{i,t})^2$$

and, consequently, $S_N(t) := \frac{1}{N}\sum_{i=1}^{N}S_N^{(i)}(t)$. Then,

$$S_N^{(i)}(t) = \begin{cases} \frac{\sigma^2}{t}\sum_{s=1}^{t}(\varepsilon_{i,s}-\bar{\varepsilon}_{i,t})^2, & t\leqslant\tau,\\ \frac{1}{t}\left[\sum_{s=1}^{\tau}(\sigma\varepsilon_{i,s}-\sigma\bar{\varepsilon}_{i,t}-\frac{t-\tau}{t}\delta_i)^2+\sum_{s=\tau+1}^{t}(\sigma\varepsilon_{i,s}-\sigma\bar{\varepsilon}_{i,t}+\frac{\tau}{t}\delta_i)^2\right], & t>\tau; \end{cases}$$

where $\bar{\varepsilon}_{i,t}=\frac{1}{t}\sum_{s=1}^{t}\varepsilon_{i,s}$. By the definition of the cumulative autocorrelation function, we have for $2 \leqslant t \leqslant \tau$

$$\begin{aligned}\mathsf{E}\, S_N^{(i)}(t) &= \frac{\sigma^2}{t}\sum_{s=1}^{t}\mathsf{E}\,(\varepsilon_{i,s}-\bar{\varepsilon}_{i,t})^2 = \frac{\sigma^2}{t}\sum_{s=1}^{t}\left[1-\frac{2}{t}\sum_{r=1}^{t}\mathsf{E}\,\varepsilon_{i,s}\varepsilon_{i,r}+\frac{1}{t^2}r(t)\right]\\ &= \sigma^2\left(1-\frac{r(t)}{t^2}\right).\end{aligned}$$

In the other case when $t > \tau$, one can calculate

$$\begin{aligned}\mathsf{E}\, S_N^{(i)}(t) &= \sigma^2\left(1-\frac{r(t)}{t^2}\right)+\frac{\tau}{t}\left(\frac{t-\tau}{t}\right)^2\delta_i^2+\frac{t-\tau}{t}\left(\frac{\tau}{t}\right)^2\delta_i^2\\ &= \sigma^2\left(1-\frac{r(t)}{t^2}\right)+\frac{\tau(t-\tau)}{t^2}\delta_i^2.\end{aligned}$$

Realize that $S_N^{(i)}(t) - \mathsf{E}\, S_N^{(i)}(t)$ are independent with zero mean for fixed t and $i = 1, \ldots, N$. Due to Assumption C2, for $2 \leqslant t \leqslant \tau$ it holds

$$\mathsf{Var}\, S_N(t) = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\sigma^4}{t^2}\mathsf{Var}\left[\sum_{s=1}^{t}(\varepsilon_{i,s}-\bar{\varepsilon}_{i,t})^2\right] = \frac{1}{N}C_1(t,\sigma),$$

where $C_1(t, \sigma) > 0$ is some constant not depending on N. If $t > \tau$, then

$$\begin{aligned}\mathsf{Var}\, S_N(t) = \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{t^2}\mathsf{Var}\Bigg[&\sigma^2\sum_{s=1}^{\tau}(\varepsilon_{i,s}-\bar{\varepsilon}_{i,t})^2 - 2\frac{t-\tau}{t}\sigma\delta_i\sum_{s=1}^{\tau}(\varepsilon_{i,s}-\bar{\varepsilon}_{i,t})\\ &+\left(\frac{t-\tau}{t}\right)^2\delta_i^2+\sigma^2\sum_{s=\tau+1}^{t}(\varepsilon_{i,s}-\bar{\varepsilon}_{i,t})^2\\ &+2\frac{\tau}{t}\sigma\delta_i\sum_{s=\tau+1}^{t}(\varepsilon_{i,s}-\bar{\varepsilon}_{i,t})+\left(\frac{\tau}{t}\right)^2\delta_i^2\Bigg]\\ \leqslant \frac{1}{N}C_2(t,\tau,\sigma)+\frac{1}{N^2}C_3(t,\tau,\sigma)&\sum_{i=1}^{N}\delta_i^2+\frac{1}{N^2}C_4(t,\tau,\sigma)\left|\sum_{i=1}^{N}\delta_i\right|,\end{aligned}$$

where $C_j(t, \tau, \sigma) > 0$ does not depend on N for $j = 2, 3, 4$.

The Chebyshev inequality provides $S_N(t) - \mathsf{E}\, S_N(t) = O_{\mathsf{P}}\left(\sqrt{\mathsf{Var}\, S_N(t)}\right)$ as $N \to \infty$. According to Assumption C1 and the Cauchy-Schwarz inequality, we have

$$\frac{1}{N^2}\left|\sum_{i=1}^{N}\delta_i\right| \leqslant \frac{1}{N}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \to 0, \quad N\to\infty.$$

Since the index set $\{1, \ldots, T\}$ is finite and τ is finite as well, then

$$\max_{1 \leqslant t \leqslant T} \mathsf{Var}\, S_N(t) \leqslant \frac{1}{N} K_1(\sigma) + K_2(\sigma) \frac{1}{N^2} \sum_{i=1}^{N} \delta_i^2 + K_3(\sigma) \frac{1}{N^2} \left| \sum_{i=1}^{N} \delta_i \right| \leqslant \frac{1}{N} K_4(\sigma),$$

where $K_j(\sigma) > 0$ are constants not depending on N for $j = 1, 2, 3, 4$. Thus, we also have uniform stochastic boundedness, i.e.,

$$\max_{1 \leqslant t \leqslant T} |S_N(t) - \mathsf{E}\, S_N(t)| = O_{\mathsf{P}}\left(\frac{1}{\sqrt{N}}\right), \quad N \to \infty.$$

Adding and subtracting, one has

$$\begin{aligned} S_N(\tau) - S_N(t) &= S_N(\tau) - \mathsf{E}\, S_N(\tau) - [S_N(t) - \mathsf{E}\, S_N(t)] + \mathsf{E}\, S_N(\tau) - \mathsf{E}\, S_N(t) \\ &\geqslant -2 \max_{1 \leqslant r \leqslant T} |S_N(r) - \mathsf{E}\, S_N(r)| + \mathsf{E}\, S_N(\tau) - \mathsf{E}\, S_N(t) \\ &= -2 \max_{1 \leqslant r \leqslant T} |S_N(r) - \mathsf{E}\, S_N(r)| + \sigma^2 \left(\frac{r(t)}{t^2} - \frac{r(\tau)}{\tau^2} \right) + \mathcal{I}\{t > \tau\} \frac{\tau(t-\tau)}{t^2} \frac{1}{N} \sum_{i=1}^{N} \delta_i^2. \end{aligned}$$

The above inequality holds for each $t \in \{2, \ldots, T\}$ and, particularly, it holds for $\widehat{\tau}_N$. Note that $\widehat{\tau}_N = \arg\max_t S_N(t)$. Hence, $S_N(\tau) - S_N(\widehat{\tau}_N) \leqslant 0$. Therefore,

$$\begin{aligned} 2\sqrt{N} &\max_{1 \leqslant r \leqslant T} |S_N(r) - \mathsf{E}\, S_N(r)| \\ &\geqslant \sqrt{N} \left[\sigma^2 \left(\frac{r(\widehat{\tau}_N)}{\widehat{\tau}_N^2} - \frac{r(\tau)}{\tau^2} \right) + \mathcal{I}\{\widehat{\tau}_N > \tau\} \frac{\tau(\widehat{\tau}_N - \tau)}{\widehat{\tau}_N^2} \frac{1}{N} \sum_{i=1}^{N} \delta_i^2 \right]. \end{aligned} \tag{6.7}$$

If $\widehat{\tau}_N > \tau$, then the left hand side of (6.7) is $O_{\mathsf{P}}(1)$ as $N \to \infty$, but the right hand side is unbounded because of Assumption C1. So, if $\widehat{\tau}_N \leqslant \tau$, then

$$0 \xleftarrow[N \to \infty]{\mathsf{P}} 2 \max_{1 \leqslant r \leqslant T} |S_N(r) - \mathsf{E}\, S_N(r)| \geqslant \sigma^2 \left(\frac{r(\widehat{\tau}_N)}{\widehat{\tau}_N^2} - \frac{r(\tau)}{\tau^2} \right),$$

which yields due to the monotonicity of $r(t)/t^2$ that $\mathsf{P}[\widehat{\tau}_N = \tau] \to 1$ as $N \to \infty$. □

Assumption C1 assures that the values of changes have to be large enough compared to the variability of the random noise in the panels and to the strength of dependencies within the panels as well. On one hand, Assumption C1 implies the usual assumption $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \delta_i^2 = \infty$ in change point analysis, cf. Bai (2010) or Horváth and Hušková (2012). On the other hand, Assumption C1 assures that $\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \delta_i^2 = 0$, which is not present when the panel size T is considered as unbounded, i.e., $T \to \infty$. Here, this second part is needed to control the asymptotic boundedness of the variability of $\frac{1}{t} \sum_{i=1}^{N} \sum_{s=1}^{t} (Y_{i,s} - \bar{Y}_{i,t})^2$, because a finite T cannot do that.

Similarly as in the previous section, Assumption C1 is satisfied for $0 < \delta \leqslant \delta_i < \Delta, \forall i$

(a common lower and upper bound for the change amount) and suitable σ and $r(t)$. Assumptions P2 and C1 are generally incomparable. The monotonicity assumption from Theorem 6.3 in not very restrictive at all. For example in case of independent observations within the panel, this assumption is automatically fulfilled, since $\{1/t\}_{t=2}^{T}$ is decreasing. Moreover, the weaker the dependency within the panel, the faster the decrease of $r(t)/t^2$.

One can check the proof of Theorem 6.3 and see that Assumption C1 can be replaced by more restrictive assumptions $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\delta_i^2 = \infty$ and $\lim_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^{N}\delta_i^2 = 0$. This first assumption might be considered as too strong, because a common value of $\delta = \delta_i$ for all i does not fulfill it.

Various competing consistent estimates of a possible change point can be suggested, e.g., the maximizer of $\sum_{i=1}^{N}\left[\sum_{s=1}^{t}(Y_{i,s} - \bar{Y}_{i,T})\right]^2$. To show the consistency, one needs to postulate different assumptions on the cumulative autocorrelation function and shifted cumulative correlation function compared to Theorem 6.3 and this may be rather complex.

6.6 Estimation of the correlation structure

Since the panels are considered to be independent and the number of panels may be sufficiently large, one can estimate the correlation structure of the errors $[\varepsilon_{1,1}, \ldots, \varepsilon_{1,T}]^\top$ empirically. We base the errors' estimates on *residuals*

$$\widehat{e}_{i,t} := \begin{cases} Y_{i,t} - \bar{Y}_{i,\widehat{\tau}_N}, & t \leqslant \widehat{\tau}_N, \\ Y_{i,t} - \widetilde{Y}_{i,\widehat{\tau}_N}, & t > \widehat{\tau}_N. \end{cases} \tag{6.8}$$

Then, the empirical version of the autocorrelation function is

$$\widehat{\rho}_t := \frac{1}{\widehat{\sigma}^2 NT}\sum_{i=1}^{N}\sum_{s=1}^{T-t}\widehat{e}_{i,s}\widehat{e}_{i,s+t}.$$

Consequently, the kernel estimation of the cumulative autocorrelation function and shifted cumulative correlation function is adopted in lines with Andrews (1991):

$$\widehat{r}(t) = \sum_{|s|<t}(t-|s|)\kappa\left(\frac{s}{h}\right)\widehat{\rho}_s,$$

$$\widehat{R}(t,v) = \sum_{s=1}^{t}\sum_{u=t+1}^{v}\kappa\left(\frac{u-s}{h}\right)\widehat{\rho}_{u-s}, \quad t < v;$$

where $h > 0$ stands for the window size and κ belongs to a class of kernels given by

$$\Big\{\kappa(\cdot) : \mathbb{R} \to [-1,1] \,\big|\, \kappa(0) = 1,\, \kappa(x) = \kappa(-x),\, \forall x,\, \int_{-\infty}^{+\infty} \kappa^2(x)\mathrm{d}x < \infty,$$
$$\kappa(\cdot) \text{ is continuos at 0 and at all but a finite number of other points}\Big\}.$$

Since the variance parameter σ is not present in the limiting distribution of Theorem 6.1, it neither has to be estimated nor known. Nevertheless, one can use $\widehat{\sigma}^2 := \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{e}_{i,s}^2$.

6.7 Bootstrap and hypothesis testing

A wide range of literature has been published on bootstrapping in the change point problem, e.g., Hušková and Kirch (2012) or Hušková et al. (2008). We build up the bootstrap test on the resampling with replacement of row vectors $\{[\widehat{e}_{i,1},\ldots,\widehat{e}_{i,T}]\}_{i=1,\ldots,N}$ corresponding to the panels. This provides bootstrapped row vectors $\{[\widehat{e}_{i,1}^*,\ldots,\widehat{e}_{i,T}^*]\}_{i=1,\ldots,N}$. Then, the bootstrapped residuals $\widehat{e}_{i,t}^*$ are *centered* by their conditional expectation $\frac{1}{N}\sum_{i=1}^{N}\widehat{e}_{i,t}$ yielding

$$\widehat{Y}_{i,t}^* := \widehat{e}_{i,t}^* - \frac{1}{N}\sum_{i=1}^{N}\widehat{e}_{i,t}.$$

The bootstrap test statistic is just a modification of the original statistic $\mathcal{P}_N(T)$, where the original observations $Y_{i,t}$ are replaced by their bootstrap counterparts $\widehat{Y}_{i,t}^*$:

$$\mathcal{P}_N^*(T) = \max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t}\left|\sum_{i=1}^{N}\left[\sum_{r=1}^{s}\left(\widehat{Y}_{i,r}^* - \bar{\widehat{Y}}_{i,t}^*\right)\right]\right|}{\max_{s=t,\ldots,T-1}\left|\sum_{i=1}^{N}\left[\sum_{r=s+1}^{T}\left(\widehat{Y}_{i,r}^* - \widetilde{\widehat{Y}}_{i,t}^*\right)\right]\right|},$$

such that

$$\bar{\widehat{Y}}_{i,t}^* = \frac{1}{t}\sum_{s=1}^{t}\widehat{Y}_{i,s}^* \quad \text{and} \quad \widetilde{\widehat{Y}}_{i,t}^* = \frac{1}{T-t}\sum_{s=t+1}^{T}\widehat{Y}_{i,s}^*.$$

An *algorithm* for the bootstrap is illustratively shown in Procedure 6.1 and its validity will be proved in Theorem 6.6.

6.8 Validity of the resampling procedure

The idea behind bootstrapping is to *mimic the original distribution* of the test statistic in some sense with the distribution of the bootstrap test statistic, conditionally on the original data denoted by $\mathbb{Y} \equiv \{Y_{i,t}\}_{i,t=1}^{N,T}$.

First of all, two simple and just technical assumptions are needed.

Procedure 6.1 Bootstrapping test statistic $\mathcal{P}_N(T)$.

Input: Panel data consisting of N panels with length T, i.e., N row vectors of observations $[Y_{i,1}, \ldots, Y_{i,T}]$.

Output: Bootstrap distribution of $\mathcal{P}_N(T)$, i.e., the empirical distribution where probability mass $1/B$ concentrates at each of ${}_{(1)}\mathcal{P}_N^*(T), \ldots, {}_{(B)}\mathcal{P}_N^*(T)$.

estimate the change point by calculating $\widehat{\tau}_N$
compute residuals $\widehat{e}_{i,t}$
for $b = 1$ to B **do** // repeat in order to obtain the empirical distribution
  $\{[\widehat{e}_{i,1}^*, \ldots, \widehat{e}_{i,T}^*]\}_{i=1}^N$ resampled with replacement from the original rows of residuals $\{[\widehat{e}_{i,1}, \ldots, \widehat{e}_{i,T}]\}_{i=1}^N$
  calculate bootstrap panel data $\widehat{Y}_{i,t}^*$
  compute bootstrap test statistics ${}_{(b)}\mathcal{P}_N^*(T)$
end for

Assumption B1. $\{\varepsilon_{i,t}\}_t$ possesses the lagged cumulative correlation function

$$S(t, v, d) = \mathsf{Cov}\left(\sum_{s=1}^{t} \varepsilon_{i,s}, \sum_{u=t+d}^{v} \varepsilon_{i,u}\right) = \sum_{s=1}^{t} \sum_{u=t+d}^{v} \rho_{u-s}, \quad \forall i \in \mathbb{N}.$$

Assumption B2. $\lim_{N \to \infty} \mathsf{P}\left[\widehat{\tau}_N = \tau\right] = 1$.

Assumption B1 is not really an assumption, actually it is only a notation. Notice that $S(t, v, 1) \equiv R(t, v)$. Assumption B2 is satisfied for our estimate proposed in (6.6), if the assumptions of Theorem 6.2 hold. Assumption B2 is postulated in a rather broader sense, because we want to allow any other consistent estimate of τ to be used instead.

Let us now introduce two supporting theorems, in order to be able to justify the bootstrap method in our setup.

Suppose that $\{\boldsymbol{\xi}_n\}_{n=1}^{\infty}$ is a sequence of random variables/vectors existing on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. A *bootstrap version* of $\boldsymbol{\xi} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]^\top$ is its (randomly) resampled sequence with replacement—denoted by $\boldsymbol{\xi}^* \equiv [\boldsymbol{\xi}_1^*, \ldots, \boldsymbol{\xi}_n^*]^\top$—with the same length, where for each $i \in \{1, \ldots, n\}$ it holds that $\mathsf{P}_{\boldsymbol{\xi}}^*[\boldsymbol{\xi}_i^* = \boldsymbol{\xi}_j] \equiv \mathsf{P}[\boldsymbol{\xi}_i^* = \boldsymbol{\xi}_j | \boldsymbol{\xi}] = 1/n$, $j = 1, \ldots, n$. In the sequel, $\mathsf{P}_{\boldsymbol{\xi}}^*$ denotes the conditional probability given $\boldsymbol{\xi}$. So, $\boldsymbol{\xi}_i^*$ has a discrete uniform distribution on $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}$ for every $i = 1, \ldots, n$. The conditional expectation and variance given $\boldsymbol{\xi}$ are denoted by $\mathsf{E}_{\mathsf{P}_{\boldsymbol{\xi}}^*}$ and $\mathsf{Var}_{\mathsf{P}_{\boldsymbol{\xi}}^*}$.

If a statistic has an approximate normal distribution, one may be interested in the asymptotic comparison of the bootstrap distribution with the original one. A tool for assessing such an approximate closeness can be a *bootstrap central limit theorem* for triangular arrays.

Theorem 6.4 (Bootstrap CLT for triangular arrays)**.** *Let* $\{\xi_{n,k_n}\}_{n=1}^{\infty}$ *be a triangular array of zero mean random variables on the same probability space such that the elements of the vector* $[\xi_{n,1}, \ldots, \xi_{n,k_n}]^\top$ *are iid for every* $n \in \mathbb{N}$ *satisfying*

$$\sup_{n \in \mathbb{N}} \mathsf{E}_{\mathsf{P}}\, \xi_{n,1}^4 < \infty \tag{6.9}$$

and $k_n \to \infty$ as $n \to \infty$. Suppose that $\boldsymbol{\xi}^* \equiv [\xi^*_{n,1}, \ldots, \xi^*_{n,k_n}]^\top$ is the bootstrapped version of $\boldsymbol{\xi} \equiv [\xi_{n,1}, \ldots, \xi_{n,k_n}]^\top$ and denote

$$\bar{\xi}_n := k_n^{-1} \sum_{i=1}^{k_n} \xi_{n,i}, \quad \bar{\xi}^*_n := k_n^{-1} \sum_{i=1}^{k_n} \xi^*_{n,i}, \quad \text{and} \quad \varsigma_n^2 := \mathsf{Var}_{\mathsf{P}}\, \xi_{n,1}.$$

If

$$\liminf_{n\to\infty} \varsigma_n^2 = \varsigma^2 > 0, \tag{6.10}$$

then

$$\sup_{x\in\mathbb{R}} \left| \mathsf{P}^*_{\boldsymbol{\xi}} \left[\frac{\sqrt{k_n}}{\sqrt{\varsigma_n^2}} \left(\bar{\xi}^*_n - \bar{\xi}_n \right) \leqslant x \right] - \mathsf{P} \left[\frac{\sqrt{k_n}}{\sqrt{\varsigma_n^2}} \bar{\xi}_n \leqslant x \right] \right| \xrightarrow[n\to\infty]{\mathsf{P}} 0.$$

Proof. The Lyapunov condition (Billingsley, 1986, p. 371) for a triangular array of random variables $\{\xi_{n,k_n}\}_{n=1}^\infty$ is satisfied due to (6.9) and (6.10), i.e., for $\omega = 2$:

$$\frac{1}{\sqrt{k_n \varsigma_n^2}^{2+\omega}} \sum_{i=1}^{k_n} \mathsf{E}\, |\xi_{n,i}|^{2+\omega} \leqslant \frac{k_n^{-\omega/2}}{\varsigma_n^{2+\omega}} \sup_{\iota\in\mathbb{N}} \mathsf{E}\, |\xi_{\iota,1}|^{2+\omega} \to 0, \quad n \to \infty.$$

Therefore, the CLT for $\{\xi_{n,k_n}\}_{n=1}^\infty$ holds and

$$\sup_{x\in\mathbb{R}} \left| \mathsf{P} \left[\frac{\sqrt{k_n}}{\sqrt{\varsigma_n^2}} \bar{\xi}_n \leqslant x \right] - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{t^2}{2} \right\} \mathrm{d}t \right| \xrightarrow[n\to\infty]{} 0.$$

Now, to prove the theorem, it suffices to show the following three statements:

(i)

$$\sup_{x\in\mathbb{R}} \left| \mathsf{P}^*_{\boldsymbol{\xi}} \left[\frac{\sqrt{k_n}}{\sqrt{\mathsf{Var}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \xi^*_{n,1}}} \left(\bar{\xi}^*_n - \mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \bar{\xi}^*_n \right) \leqslant x \right] - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{t^2}{2} \right\} \mathrm{d}t \right| \xrightarrow[n\to\infty]{\mathsf{P}} 0;$$

(ii) $\mathsf{Var}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \xi^*_{n,1} - \varsigma_n^2 \xrightarrow[n\to\infty]{\mathsf{P}} 0$;

(iii) $\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \bar{\xi}^*_n = \bar{\xi}_n$, $[\mathsf{P}]$-*a.s.*

Proving *(iii)* is trivial, because $\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \bar{\xi}^*_n = \mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \xi^*_{n,1} = k_n^{-1} \sum_{i=1}^{k_n} \xi_{n,i} = \bar{\xi}_n$, $[\mathsf{P}]$-*a.s.*

Let us calculate the conditional variance of the bootstrapped variable $\xi^*_{n,1}$: $\mathsf{Var}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \xi^*_{n,1} = \mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \xi^{*2}_{n,1} - (\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \xi^*_{n,1})^2 = k_n^{-1} \sum_{i=1}^{k_n} \xi^2_{n,i} - \left(k_n^{-1} \sum_{i=1}^{k_n} \xi_{n,i} \right)^2$, $[\mathsf{P}]$-*a.s.* The weak law of large numbers together with (6.9) provides

$$\bar{\xi}_n - n^{-1}\sum_{i=i}^{k_n} \mathsf{E}_{\mathsf{P}}\,\xi_{n,i} = \bar{\xi}_n \xrightarrow[n\to\infty]{\mathsf{P}} 0$$

and

$$0 \xleftarrow[n\to\infty]{\mathsf{P}} k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}^2 - \left(k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}\right)^2 - \mathsf{E}_{\mathsf{P}}\,\xi_{n,1}^2 = \mathsf{Var}_{\mathsf{P}^*_{\boldsymbol{\xi}}}\xi_{n,1}^* - \varsigma_n^2.$$

The last result of the WLLN is true, because (6.9) implies

$$k_n^{-2}\sum_{i=1}^{k_n}\mathsf{Var}_{\mathsf{P}}\,\xi_{n,i}^2 \leqslant k_n^{-2}\sum_{i=1}^{k_n}\mathsf{E}_{\mathsf{P}}\,\xi_{n,i}^4 \leqslant k_n^{-1}\sup_{\iota\in\mathbb{N}}\mathsf{E}_{\mathsf{P}}\,\xi_{\iota,1}^4 \xrightarrow[n\to\infty]{} 0.$$

Thus *(ii)* is proved.

The Berry-Esseen-Katz theorem (see Katz (1963)) with $g(x) = |x|^\epsilon$, $\epsilon > 0$ for the bootstrapped sequence of *iid* (with respect to P^*) random variables $\{\xi_{n,i}^*\}_{i=1}^{k_n}$ results in

$$\sup_{x\in\mathbb{R}}\left|\mathsf{P}^*_{\boldsymbol{\xi}}\left[\frac{\sqrt{k_n}}{\sqrt{\mathsf{Var}_{\mathsf{P}^*_{\boldsymbol{\xi}}}\xi_{n,1}^*}}\left(\bar{\xi}_n^* - \mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}}\bar{\xi}_n^*\right) \leqslant x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}\mathrm{d}t\right|$$
$$\leqslant Ck_n^{-\epsilon/2}\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}}\left|\frac{\xi_{n,1}^* - \mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}}\xi_{n,1}^*}{\mathsf{Var}_{\mathsf{P}^*_{\boldsymbol{\xi}}}\xi_{n,1}^*}\right|^{2+\epsilon} \quad [\mathsf{P}]\text{-}a.s., \qquad (6.11)$$

for all $n\in\mathbb{N}$ where $C > 0$ is an absolute constant.

The Jensen inequality and Minkowski inequality provide an upper bound for the nominator from the right-hand side of (6.11):

$$\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}}|\xi_{n,1}^* - \mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}}\xi_{n,1}^*|^{2+\epsilon} = k_n^{-1}\sum_{i=1}^{k_n}\left|\xi_{n,i} - k_n^{-1}\sum_{j=1}^{k_n}\xi_{n,j}\right|^{2+\epsilon}$$
$$\leqslant k_n^{-1}\left\{\left(\sum_{i=1}^{k_n}|\xi_{n,i}|^{2+\epsilon}\right)^{1/(2+\epsilon)} + k_n^{-(1+\epsilon)/(2+\epsilon)}\left|\sum_{j=1}^{k_n}\xi_{n,j}\right|\right\}^{2+\epsilon}$$
$$\leqslant 2^{1+\epsilon}k_n^{-1}\sum_{i=1}^{k_n}|\xi_{n,i}|^{2+\epsilon} + 2^{1+\epsilon}\left|k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}\right|^{2+\epsilon} \quad [\mathsf{P}]\text{-}a.s.$$

The right-hand side of the previously derived upper bound is uniformly bounded in probability P, because of Markov's inequality and (6.9). Indeed, for fixed $\eta > 0$

$$\mathsf{P}\left[k_n^{-1}\sum_{i=1}^{k_n}|\xi_{n,i}|^{2+\epsilon} \geqslant \eta\right] \leqslant \eta^{-1}k_n^{-1}\sum_{i=1}^{k_n}\mathsf{E}_{\mathsf{P}}|\xi_{n,i}|^{2+\epsilon} \leqslant \eta^{-1}\sup_{\iota\in\mathbb{N}}\mathsf{E}_{\mathsf{P}}|\xi_{\iota,1}|^{2+\epsilon} < \infty, \quad \forall n\in\mathbb{N}$$

and

$$\mathsf{P}\left[\left|k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}\right|\geqslant\eta\right]\leqslant\eta^{-1}k_n^{-1}\mathsf{E}_{\mathsf{P}}\left|\sum_{i=1}^{k_n}\xi_{n,i}\right|\leqslant\eta^{-1}\sup_{\iota\in\mathbb{N}}\mathsf{E}_{\mathsf{P}}|\xi_{\iota,1}|<\infty,\quad\forall n\in\mathbb{N}.$$

Since $\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}}|\xi^*_{n,1}-\mathsf{E}_{\mathsf{P}^*}\xi^*_{n,1}|^{2+\epsilon}$ is bounded in probability P uniformly over n and the denominator of the right-hand side of (6.11) is uniformly bounded away from zero due to (6.10), then the left-hand side of (6.11) converges in probability P to zero as n tends to infinity. So, *(i)* is proved as well. □

Theorem 6.5 (Bootstrap multivariate CLT for triangular arrays)**.** *Let $\{\boldsymbol{\xi}_{n,k_n}\}_{n=1}^{\infty}$ be a triangular array of zero mean q-dimensional random vectors on the same probability space such that the elements of the vector sequence $\{\boldsymbol{\xi}_{n,1},\ldots,\boldsymbol{\xi}_{n,k_n}\}$ are iid for every $n\in\mathbb{N}$ satisfying*

$$\sup_{n\in\mathbb{N}}\mathsf{E}_{\mathsf{P}}|\xi_{n,1}^{(j)}|^4<\infty,\quad j\in\{1,\ldots,q\},\tag{6.12}$$

where $\boldsymbol{\xi}_{n,1}\equiv[\xi_{n,1}^{(1)},\ldots,\xi_{n,1}^{(q)}]^\top\in\mathbb{R}^q$, $n\in\mathbb{N}$ and $k_n\to\infty$ as $n\to\infty$. Assume that $\boldsymbol{\Xi}^*\equiv[\boldsymbol{\xi}^*_{n,1},\ldots,\boldsymbol{\xi}^*_{n,k_n}]^\top$ is the bootstrapped version of $\boldsymbol{\Xi}\equiv[\boldsymbol{\xi}_{n,1},\ldots,\boldsymbol{\xi}_{n,k_n}]^\top$. Denote

$$\bar{\boldsymbol{\xi}}_n:=k_n^{-1}\sum_{i=1}^{k_n}\boldsymbol{\xi}_{n,i},\quad\bar{\boldsymbol{\xi}}_n^*:=k_n^{-1}\sum_{i=1}^{k_n}\boldsymbol{\xi}^*_{n,i},\quad\textit{and}\quad\boldsymbol{\Gamma}_n:=\mathsf{Var}_{\mathsf{P}}\,\boldsymbol{\xi}_{n,1}.$$

If

$$\liminf_{n\to\infty}\boldsymbol{\Gamma}_n=\boldsymbol{\Gamma}>\mathbf{0},\tag{6.13}$$

then

$$\mathsf{P}^*_{\boldsymbol{\Xi}}\left[\sqrt{k_n}\boldsymbol{\Gamma}_n^{-1/2}\left(\bar{\boldsymbol{\xi}}_n^*-\bar{\boldsymbol{\xi}}_n\right)\leqslant\mathbf{x}\right]-\mathsf{P}\left[\sqrt{k_n}\boldsymbol{\Gamma}_n^{-1/2}\bar{\boldsymbol{\xi}}_n\leqslant\mathbf{x}\right]\xrightarrow[n\to\infty]{\mathsf{P}}0,\quad\forall\mathbf{x}\in\mathbb{R}^q.$$

Proof. According to the Cramér-Wold theorem, it is sufficient to ensure that all assumptions of one-dimensional bootstrap CLT 6.4 for triangular arrays are valid for any linear combination of the elements of the random vector $\boldsymbol{\xi}_{n,1}$, $n\in\mathbb{N}$.

For arbitrary fixed $\mathbf{t}\in\mathbb{R}^q$ using the Jensen inequality, we get

$$\sup_{n\in\mathbb{N}}\mathsf{E}_{\mathsf{P}}|\mathbf{t}^\top\boldsymbol{\xi}_{n,1}|^4\leqslant q^3\sup_{n\in\mathbb{N}}\sum_{j=1}^{q}t_j^4\mathsf{E}_{\mathsf{P}}|\xi_{n,1}^{(j)}|^4\leqslant q^4\max_{j=1,\ldots,q}t_j^4\sup_{n\in\mathbb{N}}\mathsf{E}_{\mathsf{P}}|\xi_{n,1}^{(j)}|^4<\infty.$$

Hence, assumption (6.12) implies assumption (6.9) for the random variables $\{\mathbf{t}^\top\boldsymbol{\xi}_{n,k_n}\}_{n\in\mathbb{N}}$.

Similarly, assumption (6.13) implies assumption (6.10) for such an arbitrary linear combination, i.e., positive definiteness of the matrix $\mathbf{\Gamma}$ yields

$$\liminf_{n\to\infty} \mathsf{Var}_{\mathsf{P}}\, \mathbf{t}^\top \boldsymbol{\xi}_{n,1} = \liminf_{n\to\infty} \mathbf{t}^\top \left(\mathsf{Var}_{\mathsf{P}}\, \boldsymbol{\xi}_{n,1}\right) \mathbf{t} \geqslant \mathbf{t}^\top \left(\liminf_{n\to\infty} \mathbf{\Gamma}_n\right) \mathbf{t} = \mathbf{t}^\top \mathbf{\Gamma} \mathbf{t} > 0.$$

□

Now we may proceed to the bootstrap justification theorem. Realize that it is not known, whether the common panel means' change occurred or not. In other words, one does not know *whether the data come from the null or the alternative* hypothesis. Therefore, the following theorem holds under H_0 as well as H_1.

Theorem 6.6 (Bootstrap justification)**.** *Suppose that panel data* $\mathbb{Y} = \{Y_{i,t}\}_{i,t=1}^{N,T}$ *follow model* (6.1). *Under Assumptions P1, B1, B2, and C2*

$$\mathcal{P}_N^*(T)|\mathbb{Y} \xrightarrow[N\to\infty]{\mathscr{D}} \max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t} \left|\mathcal{X}_s - \frac{s}{t}\mathcal{X}_t\right|}{\max_{s=t,\ldots,T-1} \left|\mathcal{Z}_s - \frac{T-s}{T-t}\mathcal{Z}_t\right|} \quad \textit{in probability } \mathsf{P},$$

where $\mathcal{Z}_t := \mathcal{X}_T - \mathcal{X}_t$ *and* $[\mathcal{X}_1, \ldots, \mathcal{X}_T]^\top$ *is a multivariate normal random vector with zero mean and covariance matrix* $\mathbf{\Gamma} = \{\gamma_{t,v}(\tau)\}_{t,v=1}^{T,T}$ *such that*

$$\gamma_{t,t}(\tau) = \begin{cases} r(t) + \frac{t^2}{\tau^2} r(\tau) - \frac{2t}{\tau}[r(t) + R(t,\tau)], \quad t < \tau; \\ 0, \quad t = \tau; \\ r(t-\tau) + \frac{(t-\tau)^2}{(T-\tau)^2} r(T-\tau) - \frac{2(t-\tau)}{T-\tau}\left[r(t-\tau) + R(t-\tau, T-\tau)\right], \quad t > \tau; \end{cases}$$

and

$$\gamma_{t,v}(\tau) = \begin{cases} 0, \quad t = \tau \textit{ or } v = \tau, \\ r(t) + R(t,v) + \frac{tv}{\tau^2} r(\tau) - \frac{v}{\tau}[r(t) + R(t,\tau)] - \frac{t}{\tau}[r(v) + R(v,\tau)], \quad t < v < \tau; \\ S(t,v,\tau+1-t) + \frac{t(v-\tau)}{\tau(T-\tau)} R(\tau,T) \\ \quad -\frac{v-\tau}{T-\tau} S(t,T,\tau+1-t) - \frac{t}{\tau} R(\tau,v), \quad t < \tau < v; \\ r(t-\tau) + R(t-\tau, v-\tau) + \frac{(t-\tau)(v-\tau)}{(T-\tau)^2} r(T-\tau) \\ \quad -\frac{v-\tau}{T-\tau}[r(t-\tau) + R(t-\tau, T-\tau)] \\ \quad -\frac{t-\tau}{T-\tau}[r(v-\tau) + R(v-\tau, T-\tau)], \quad \tau < t < v. \end{cases}$$

Proof. Let us define $\widehat{\epsilon}_{i,t} := \sigma^{-1} \sum_{s=1}^t \widehat{e}_{i,s}$, $\widehat{\epsilon}_{i,t}^* := \sigma^{-1} \sum_{s=1}^t \widehat{e}_{i,s}^*$,

$$\widehat{U}_N(t) := \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^N \sum_{s=1}^t \widehat{e}_{i,s} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \widehat{\epsilon}_{i,t},$$

and

$$\widehat{U}_N^*(t) := \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} \sum_{s=1}^{t} \widehat{Y}_{i,s}^* = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} \sum_{s=1}^{t} \left(\widehat{e}_{i,s}^* - \frac{1}{N} \sum_{i=1}^{N} \widehat{e}_{i,s} \right)$$
$$= \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} \sum_{s=1}^{t} \left(\widehat{e}_{i,s}^* - \widehat{e}_{i,s} \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\widehat{\epsilon}_{i,t}^* - \widehat{\epsilon}_{i,t} \right).$$

Realize that $\widehat{\epsilon}_{i,t}$ depends on $\widehat{\tau}_N$ and, hence, it depends on N. Thus, $\widehat{\epsilon}_{i,t} \equiv \widehat{\epsilon}_{i,t}(N)$. Since Assumption C2 holds, then according to the bootstrap multivariate CLT for triangular arrays (Theorem 6.5) of T-dimensional vectors $\boldsymbol{\xi}_{N,i} = [\widehat{\epsilon}_{i,1}(N), \ldots, \widehat{\epsilon}_{i,T}(N)]^\top$ with $k_N = N$, we have

$$\mathsf{P}\left[\boldsymbol{\Gamma}_N^{-1/2}[\widehat{U}_N^*(1), \ldots, \widehat{U}_N^*(T)]^\top \leqslant \mathbf{x} \middle| \mathbb{Y}\right] - \mathsf{P}\left[\boldsymbol{\Gamma}_N^{-1/2}[\widehat{U}_N(1), \ldots, \widehat{U}_N(T)]^\top \leqslant \mathbf{x}\right]$$
$$\xrightarrow[N\to\infty]{\mathsf{P}} 0, \quad \forall \mathbf{x} \in \mathbb{R}^T,$$

where $\boldsymbol{\Gamma}_N = \mathsf{Var}\,[\widehat{\epsilon}_{i,1}, \ldots, \widehat{\epsilon}_{i,T}]^\top$.

Now, it is sufficient to realize that $[\widehat{U}_N(1), \ldots, \widehat{U}_N(T)]^\top$ has an approximate multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Gamma} = \lim_{N\to\infty} \boldsymbol{\Gamma}_N$. Using the law of total variance,

$$\mathsf{Var}\,\widehat{\epsilon}_{i,t} = \mathsf{E}\,[\mathsf{Var}\,\{\widehat{\epsilon}_{i,t}|\widehat{\tau}_N\}] + \mathsf{Var}\,[\mathsf{E}\,\{\widehat{\epsilon}_{i,t}|\widehat{\tau}_N\}]$$
$$= \sum_{\pi=1}^{T} \mathsf{P}\,[\widehat{\tau}_N = \pi]\mathsf{Var}\,[\widehat{\epsilon}_{i,t}|\widehat{\tau}_N = \pi] + \sum_{\pi=1}^{T} \mathsf{P}\,[\widehat{\tau}_N = \pi]\{\mathsf{E}\,[\widehat{\epsilon}_{i,t}|\widehat{\tau}_N = \pi]\}^2$$
$$- \left\{ \sum_{\pi=1}^{T} \mathsf{P}\,[\widehat{\tau}_N = \pi]\mathsf{E}\,[\widehat{\epsilon}_{i,t}|\widehat{\tau}_N = \pi] \right\}^2.$$

Since $\lim_{N\to\infty} \mathsf{P}\,[\widehat{\tau}_N = \tau] = 1$ and $\mathsf{E}\,[\widehat{e}_{i,t}|\widehat{\tau}_N = \tau] = 0$, then

$$\lim_{N\to\infty} \mathsf{Var}\,\widehat{\epsilon}_{i,t} = \lim_{N\to\infty} \mathsf{Var}\,[\widehat{\epsilon}_{i,t}|\widehat{\tau}_N = \tau].$$

Similarly with the covariance, i.e., after applying the law of total covariance, we have

$$\lim_{N\to\infty} \mathsf{Cov}\,(\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v}) = \lim_{N\to\infty} \mathsf{Cov}\,(\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v}|\widehat{\tau}_N = \tau).$$

Note that

$$(\widehat{e}_{i,t}|\widehat{\tau}_N = \tau) = \begin{cases} \sigma(\varepsilon_{i,t} - \bar{\varepsilon}_{i,\tau}), & t \leqslant \tau; \\ \sigma(\varepsilon_{i,t} - \widetilde{\varepsilon}_{i,\tau}), & t > \tau; \end{cases}$$

where

$$\bar{\varepsilon}_{i,t} = \frac{1}{t}\sum_{s=1}^{t}\varepsilon_{i,s} \quad \text{and} \quad \widetilde{\varepsilon}_{i,t} = \frac{1}{T-t}\sum_{s=t+1}^{T}\varepsilon_{i,s}.$$

Taking into account the definitions of $r(t)$, $R(t,v)$, and $S(t,v,d)$ together with some simple algebra, we obtain that $\mathsf{Var}\,[\widehat{\epsilon}_{i,s}|\widehat{\tau}_N = \tau] = \gamma_{t,t}(\tau)$ and $\mathsf{Cov}\,(\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v}|\widehat{\tau}_N = \tau) = \gamma_{t,v}(\tau)$ for $t < v$, where the elements $\gamma_{t,t}(\tau)$ and $\gamma_{t,v}(\tau)$ are as in the statement of Theorem 6.6.

Then the sum in the nominator of $\mathcal{P}_N^*(T)$ can be alternatively rewritten as

$$\frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\sum_{r=1}^{s}\left(\widehat{Y}_{i,r}^{*} - \bar{\widehat{Y}}_{i,t}^{*}\right) = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\left\{\left[\sum_{r=1}^{s}\widehat{Y}_{i,r}^{*}\right] - \frac{s}{t}\sum_{v=1}^{t}\widehat{Y}_{i,v}^{*}\right\} = \widehat{U}_N^*(s) - \frac{s}{t}\widehat{U}_N^*(t).$$

Concerning the denominator of $\mathcal{P}_N^*(T)$, one needs to perform a similar calculation as in the proof of Theorem 6.1 with $V_N(t)$, i.e., to define $\widehat{V}_N(t)$ and $\widehat{V}_N^*(t)$ analogously to $\widehat{U}_N(t)$ and $\widehat{U}_N^*(t)$ as $V_N(t)$ is to $U_N(t)$. Applying the Cramér-Wold theorem completes the proof. □

The validity of the bootstrap test is assured by Theorem 6.6. Indeed, the conditional asymptotic distribution of the bootstrap test statistic is a functional of a multivariate normal distribution under the null as well as under the alternative. It does not converge to infinity (in probability) under the alternative. That is why it can be used for correctly rejecting the null in favor of the alternative, having sufficiently large N. Moreover, the following theorem states that the conditional distribution of the bootstrap test statistic and the unconditional distribution of the original test statistic *coincide*. And that is the reason why the bootstrap test should approximately keep the same level as the original test based on the asymptotics from Theorem 6.1.

Theorem 6.7 (Bootstrap test consistency)**.** *Suppose that panel data* $\mathbb{Y} = \{Y_{i,t}\}_{i,t=1}^{N,T}$ *follow model* (6.1)*. Under Assumptions P1, B2, C2 and null hypothesis* (6.2)*, the asymptotic distribution of* $\mathcal{P}_N(T)$ *from Theorem 6.1 and the asymptotic distribution of* $\mathcal{P}_N^*(T)|\mathbb{Y}$ *from Theorem 6.6 coincide.*

Proof. Recall the notation from the proof of Theorem 6.6. Under H_0, B2, and C2 it holds

$$\lim_{N\to\infty}\mathsf{P}\left[\widehat{\tau}_N = T\right] = 1.$$

Then in view of (6.8),

$$\lim_{N\to\infty}\mathsf{P}\left[\widehat{U}_N(s) - \frac{s}{t}\widehat{U}_N(t) = U_N(s) - \frac{s}{t}U_N(t)\right] = 1, \quad 1 \leqslant s \leqslant t \leqslant T.$$

□

Now, the simulated (empirical) distribution of the bootstrap test statistic can be used to calculate the bootstrap critical value, which will be compared to the value of the original test statistic in order to reject the null or not.

Assuming common variance parameter σ for each panel in panel model (6.1) might seem restrictive for some practical applications. To generalize the panel model, it is possible to consider a panel specific unknown variance parameters σ_i such that there exist a lower and an upper bound for all the σ_i's, i.e., $0 < \sigma_{min} \leqslant \sigma_i \leqslant \sigma_{max} < +\infty$ for all $i = 1, \ldots, N$. For this model's generalization, one may also show the bootstrap validity similar to Theorem 6.6, but the corresponding proof becomes more technical.

Finally, note that one cannot think about any local alternative in this setup, because τ has a discrete and finite support.

6.9 Simulations

A simulation experiment was performed to study the *finite sample properties* of the asymptotic and bootstrap test for a common change in panel means. In particular, the interest lies in the empirical *sizes* of the proposed tests under the null hypothesis and in the empirical *rejection* rate (power) under the alternative. Random samples of panel data (5000 each time) are generated from the panel change point model (6.1). The panel size is set to $T = 10$ and $T = 25$ in order to demonstrate the performance of the testing approaches in case of small and intermediate panel length. The number of panels considered is $N = 50$ and $N = 200$.

The correlation structure within each panel is modeled via random vectors generated from iid, AR(1), and GARCH(1,1) sequences. To recall the notation for the Generalized Autoregressive Conditional Heteroskedasticity processes (Bollerslev, 1986), a GARCH(1,1) process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with volatility process $\{\sigma_t\}_{t\in\mathbb{Z}}$ is a solution to the equations

$$\varepsilon_t = \sigma_t \epsilon_t, \quad t \in \mathbb{Z};$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{Z};$$

where the process $\{\sigma_t\}_{t\in\mathbb{Z}}$ is non-negative and the driving noise sequence $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of iid random variables.

The considered AR(1) process has coefficient $\phi = 0.3$. In case of GARCH(1,1) process, we use coefficients $\alpha_0 = 1$, $\alpha_1 = 0.1$, and $\beta_1 = 0.2$, which according to Lindner (2009, Example 1) gives a strictly stationary process. In all three sequences, the innovations are obtained as iid random variables from a standard normal $\mathsf{N}(0, 1)$ or Student t_5 distribution. Simulation scenarios are produced as all possible combinations of the above mentioned settings.

When using the asymptotic distribution from Theorem 6.1, the covariance matrix is estimated as proposed in Section 6.6 using the Parzen kernel

$$\kappa_P(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & 0 \leqslant |x| \leqslant 1/2; \\ 2(1-|x|)^3, & 1/2 \leqslant |x| \leqslant 1; \\ 0, & \text{otherwise}. \end{cases}$$

Several values of the smoothing window width h are tried from the interval $[2, 5]$ and all of them work fine providing comparable results. To simulate the asymptotic distribution of the test statistics, 2000 multivariate random vectors are generated using the pre-estimated covariance matrix.

The bootstrap approach does not need to estimate the covariance structure. The number of bootstrap replications used is 2000. To access the theoretical results under H_0 numerically, Table 6.1 provides the empirical specificity (one minus size) of the tests for both the asymptotic and bootstrap version of the panel change point test, where the significance level is $\alpha = 5\%$.

T	N	innovations	IID		AR(1)		GARCH(1,1)	
10	50	$\mathsf{N}(0,1)$	0.948	0.949	0.943	0.955	0.949	0.955
		t_5	0.949	0.954	0.941	0.956	0.946	0.953
	200	$\mathsf{N}(0,1)$	0.952	0.951	0.937	0.954	0.942	0.952
		t_5	0.948	0.953	0.935	0.960	0.944	0.953
25	50	$\mathsf{N}(0,1)$	0.948	0.951	0.929	0.952	0.954	0.959
		t_5	0.946	0.951	0.932	0.954	0.944	0.958
	200	$\mathsf{N}(0,1)$	0.950	0.950	0.927	0.951	0.947	0.949
		t_5	0.948	0.953	0.931	0.952	0.952	0.952

Table 6.1: Empirical specificity (1−size) of the test for the common change in panel means based on $\mathcal{P}_N(T)$ under H_0 using the asymptotic and the bootstrap critical values, considering a significance level of 0.05.

It may be seen that both approaches (using asymptotic and bootstrap distribution) are close to the theoretical value of specificity 0.95. As expected, the best results are achieved in case of independence within the panel, because there is no information overlap between two consecutive observations. The precision of not rejecting the null is increasing as the number of panels is getting higher and the panel is getting longer as well.

The performance of both testing procedures under H_1 in terms of the empirical rejection rates is shown in Table 6.2, where the change point is set to $\tau = [T/2]$ and the change sizes δ_i are independently uniform on $[1, 3]$ in 33%, 66% or in all panels.

One can conclude that the power of both tests increases as the panel size and the number

H_1	T	N	innovations	IID		AR(1)		GARCH(1,1)	
33%	10	50	$\mathsf{N}(0,1)$	0.25	0.29	0.26	0.28	0.21	0.22
			t_5	0.18	0.18	0.19	0.21	0.20	0.24
		200	$\mathsf{N}(0,1)$	0.46	0.50	0.48	0.51	0.40	0.45
			t_5	0.37	0.38	0.39	0.41	0.40	0.47
	25	50	$\mathsf{N}(0,1)$	0.38	0.43	0.31	0.37	0.30	0.36
			t_5	0.34	0.32	0.27	0.30	0.32	0.37
		200	$\mathsf{N}(0,1)$	0.72	0.79	0.69	0.66	0.60	0.67
			t_5	0.53	0.54	0.51	0.52	0.58	0.59
66%	10	50	$\mathsf{N}(0,1)$	0.45	0.52	0.47	0.50	0.39	0.41
			t_5	0.36	0.35	0.38	0.39	0.39	0.42
		200	$\mathsf{N}(0,1)$	0.76	0.84	0.81	0.81	0.70	0.77
			t_5	0.65	0.66	0.68	0.64	0.69	0.76
	25	50	$\mathsf{N}(0,1)$	0.71	0.79	0.63	0.66	0.60	0.70
			t_5	0.60	0.63	0.44	0.45	0.60	0.72
		200	$\mathsf{N}(0,1)$	0.96	0.98	0.95	0.95	0.90	0.94
			t_5	0.84	0.82	0.86	0.85	0.90	0.92
100%	10	50	$\mathsf{N}(0,1)$	0.64	0.69	0.67	0.71	0.56	0.64
			t_5	0.52	0.48	0.49	0.51	0.55	0.62
		200	$\mathsf{N}(0,1)$	0.93	0.96	0.94	0.95	0.86	0.92
			t_5	0.84	0.83	0.86	0.83	0.87	0.92
	25	50	$\mathsf{N}(0,1)$	0.86	0.91	0.83	0.86	0.80	0.85
			t_5	0.76	0.79	0.66	0.67	0.79	0.85
		200	$\mathsf{N}(0,1)$	1.00	1.00	0.99	0.99	0.98	0.97
			t_5	0.98	0.97	0.98	0.97	0.99	0.99

Table 6.2: Empirical sensitivity (power) of the test for the common change in panel means based on $\mathcal{P}_N(T)$ under H_1 using the asymptotic and the bootstrap critical values, considering a significance level of 0.05.

of panels increase, which is straightforward and expected. It should be noticed that numerical instability issues may appear for larger T, when generating from a T-variate normal distribution. Moreover, higher power is obtained when a larger portion of panels is subject to have a change in mean. The test power drops when switching from independent observations within the panel to dependent ones. Innovations with heavier tails (i.e., t_5) yield smaller power than innovations with lighter tails. Generally, the bootstrap outperforms the classical asymptotics in all scenarios.

Let us mention that for finite sections of processes with a stronger dependence structure than taken into account in the simulation scenarios, Assumption C1 does not have to be fulfilled. For example, Assumption C1 is violated for AR(1) with coefficient $\phi = 0.9$, $\delta_i = 2$, $\sigma = 1$, standard normal or Student t_5 innovations, and $\tau = 5$ for $T = 10$ or $\tau = 12$ for $T = 25$. Here, the dependency under the considered variability is too strong compared to the change size. It is rather difficult to detect possible changes in such a setup.

Finally, an early change point is discussed very briefly. We stay with standard normal innovations, iid observations within the panel, the size of changes δ_i being independently uniform on $[1, 3]$ in all panels, and the change point is $\tau = 3$ in case of $T = 10$ and $\tau = 5$ for $T = 25$. The empirical sensitivities of both tests for small values of τ are shown in Table 6.3.

T	N	τ	H_1, iid, $\mathsf{N}(0,1)$	
10	50	3	0.59	0.63
	200	3	0.89	0.91
25	50	5	0.66	0.68
	200	5	0.94	0.96

Table 6.3: Empirical sensitivity of the test for the common change in panel means based on $\mathcal{P}_N(T)$ for small values of τ under H_1 using the asymptotic and the bootstrap critical values, considering a significance level of 0.05.

When the change point is not in the middle of the panel, the power of the test generally falls down. The source of such decrease is that the left or right part of the panel possesses less observations with constant mean, which leads to a decrease of precision in the correlation estimation in case of the asymptotic test and in the change point estimation in case of the bootstrap test. Nevertheless, the bootstrap test again outperforms the asymptotic version and, moreover, provides solid results even for early or late change points (the late change points are not numerically demonstrated here).

6.10 Real data analysis

As mentioned in the introduction, our primary motivation for testing the panel mean change comes from the *insurance business*. The data set is provided by the National Association of Insurance Commissioners (NAIC) database, see Meyers and Shi (2011). We concentrate on the 'Private passenger auto liability/medical' insurance line of business. The data collect records from $N = 136$ insurance companies (having positive earned premium every year). Each insurance company provides $T = 10$ yearly total claim amounts—denoted by

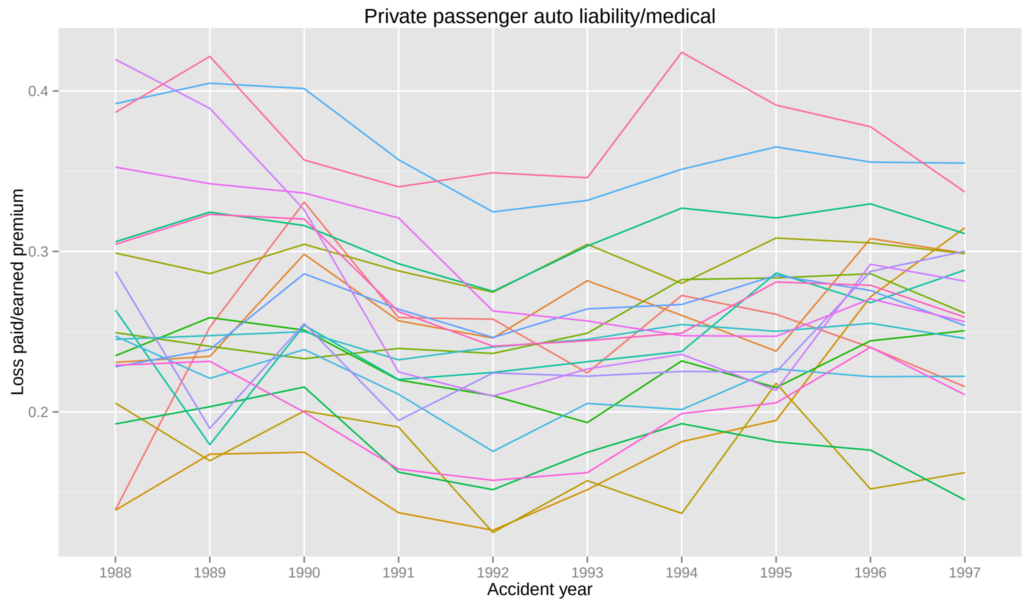


Figure 6.2: Developments of the yearly total claim amounts standardized by the earned premium for 20 selected insurance companies.

$X_{i,t}$—and earned premiums $p_{i,t}$ starting from year 1988 up to year 1997. One can consider standardizing (normalizing) the claim amounts by the premium received by company i in year t. That is thinking of panel data $Y_{i,t} = X_{i,t}/p_{i,t}$ as loss ratios. Such a standardization is reasonable, because it puts the claim amounts of different insurance companies in different years on the 'same level' (same magnitude). Besides that, this may also yield a stabilization of series' variability, which corresponds to the assumption of a common variance.

Figure 6.2 graphically shows series of standardized claim amounts (loss ratios) for 20 selected insurance companies (a plot with all 136 panels would be cluttered).

The data are considered as panel data in the way that each insurance company corresponds to one panel, which is formed by the company's yearly total standardized claim amounts. The length of the panel is quite short. This is very typical in insurance business, because considering longer panels may invoke incomparability between the early claim amounts and the late ones due to changing market or policies' conditions over time.

We want to test whether or not a change in the standardized claim amounts occurred in a common year, assuming that the standardized claim amounts are approximately constant in the years before and after the possible change for every insurance company. Our ratio type test statistic gives $\mathcal{P}_{136}(10) = 15.6$. The asymptotic critical value is 57.3 and the bootstrap critical value equals 56.8. These values mean that we do not reject the hypothesis of no change in panel means in both cases. If a striking difference occurred between the two critical values (asymptotic and bootstrap), it would mean inefficient correlation structure estimation or violation of the model assumptions (e.g., not common volatility of the panels).

We also try to take the logarithms of loss ratios and to consider log standardized amounts as the panel data observations. Nevertheless, we do not reject the hypothesis of no change in the panel means (i.e., means of log standardized amounts) again. For the sake of completeness, we may reveal that our estimate of the panel change point provides value $\widehat{\tau}_N = 10$ meaning no change in panels.

6.11 Summary

In this chapter, we consider the change point problem in panel data with fixed panel size. Occurrence of common breaks in panel means is tested. We introduce a ratio type test statistic and derive its asymptotic properties. Under the null hypothesis of no change, the test statistic weakly converges to a functional of the multivariate normal random vector with zero mean and covariance structure depending on the intra-panel covariances. As shown in this chapter, these covariances can be estimated and, consequently, used for testing whether a change in means occurred or not. This is indeed feasible, because the test statistic under the alternative converges to infinity in probability.

The secondary aim of the chapter lies in proposing a consistent change point estimate, which is straightforwardly used for bootstrapping the test statistic. We establish the asymptotic behavior of the bootstrap version of the test statistic, regardless of the fact whether the data come from the null or the alternative hypothesis. Moreover, the asymptotic distribution of the bootstrap test statistic coincides with the original test statistic's limiting distribution. This provides justification for the bootstrap method. One of the main goals is to obtain a completely data driven testing approach whether the means remain the same during the observation period or not. The ratio type test statistic allows us to omit a variance estimation and the bootstrap technique overcomes estimation of the correlation structure. Hence, neither nuisance nor smoothing parameters are present in the whole testing process, which makes it very simple for practical use. Furthermore, the whole stochastic theory behind requires relatively simple assumptions, which are not too restrictive.

A simulation study illustrates that even for small panel size, both presented approaches—based on traditional asymptotics and on bootstrapping—work fine. One may judge that both methods keep the significance level under the null, while various simulation scenarios are considered. Besides that, the power of the test is slightly higher in case of the bootstrap. Finally, the proposed methods are applied to insurance data, for which the change point analysis in panel data provides an appealing approach.

6.12 Discussion

First of all, it has to be noted that the non-ratio CUSUM type test statistic can be used instead of the ratio type, but this requires to estimate the variance of the observations. The statements of theorems and proofs would become even less complicated. Omitting the usage of the bootstrap test statistic can especially be unreliable in short panels from a computational point of view. This is due to the fact that the bootstrap overcomes the issue of estimating the correlation structure.

Furthermore, our setup can be modified by considering large panel size, i.e., $T \to \infty$. Consequently, the whole theory leads to convergences to functionals of Gaussian processes with a covariance structure derived in a very similar fashion as for fixed T. However, our motivation is to develop techniques for fixed and relatively small panel size.

Dependent panels may be taken into account and the presented work might be generalized for some kind of asymptotic independence over the panels or prescribed dependence among the panels. Nevertheless, our incentive is determined by a problem from non-life insurance, where the association of insurance companies consists of a relatively high number of insurance companies. Thus, the portfolio of yearly claims is so diversified, that the panels corresponding to insurance companies' yearly claims may be viewed as independent and neither natural ordering nor clustering has to be assumed.

Chapter 7

Conclusions

Various parametric models for sequences of ordered observations, where some parameters can change at unknown time point, are considered in this thesis. The aim is to develop stochastic approaches for testing whether a change occurred at some unknown time or not. These testing procedures rely on maximum ratio type statistics based on cumulative sums. Generally, the main advantage of the ratio type statistics in hypotheses testing is that they provide an alternative to non-ratio type statistics mainly in situations, in which variance estimation is problematic or cumbersome.

Asymptotic distributional behavior of the test statistics is derived under the null hypothesis for each change point model. Consequently, large sample properties of the test statistics are studied under alternatives. In many cases, it is not possible to calculate critical values for the test directly from the derived asymptotic distribution. However, to overcome this issue, one can use *simulations and resampling methods*. Validity of such approaches is shown and their appropriateness is justified. Moreover, simulation studies showed that the critical values obtained by resampling methods seem to be more accurate than critical values obtained by simulations from the limiting distributions.

One of the simplest model for a structural change is the model with a possible single abrupt change in mean. The idea of ratio type statistics was firstly described for this setup in existing literature. We study the possibility of extending this idea to the situation of testing the no-change hypothesis against the alternative of a *gradual change in mean*. Being specific, the means of the observations are constant for a while and, after reaching some time point, the means slowly start to change. It means that the changes occur gradually rather than abruptly, which can be considered as a smooth change point.

Then we focus our attention back on the testing null hypothesis of no change in mean against the alternative of one *abrupt shift in mean*. A testing procedure based on ratio type statistic for detection of this type of change is generalized for *α-mixing model disturbances* with heavy tails. Hence, the traditional ratio type statistic is robustified by *considering* *general loss function* instead of the traditional quadratic one. A block bootstrap method is also proposed for the testing purposes to handle observations that are not necessarily independent. We prove that the *block bootstrap method* provides asymptotically correct critical values for the studied ratio type statistic in the location model with α-mixing random errors.

Subsequently, ratio type statistics with a general score function for detection of *changes in linear regression models* are investigated. Their asymptotic behavior under the null hypothesis and under local alternatives is proved. Application of the permutation bootstrap technique is elaborated and its justification is given. As an *analogy* of the change point problem in regression, a testing procedure for a possible *change in the autoregression parameter is demonstrated*. It detects whether the observed sequence is an AR(1) process, or the time series is an AR(1) process up to some unknown time point and it is again an AR(1) process after this unknown time point with a different autoregression parameter.

Finally, we deal with the *change point problem in panel data* with *fixed panel size*, where occurrence of common breaks in panel means is tested. Besides that, a consistent change point estimate is proposed. A bootstrap version of the ratio type test statistic is defined in order to obtain a completely data driven approach to test whether the means remain the same during the observation period or not.

Appendix A

Useful Definitions and Theorems

Definition A.1 (Deterministic Landau symbols)**.** Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be two sequences of real numbers. One writes

$$a_n = O(b_n), \quad n \to \infty;$$

if and only if there exists a positive real number $M > 0$ and an integer $n_0 \in \mathbb{N}$ such that

$$|a_n| \leqslant M|b_n|, \quad \forall n \geqslant n_0.$$

One writes

$$a_n = o(b_n), \quad n \to \infty;$$

if and only if, for every positive real number $\tau > 0$, there exists an integer $n_0 \in \mathbb{N}$ such that

$$|a_n| \leqslant \tau|b_n|, \quad \forall n \geqslant n_0.$$

Definition A.2 (Stochastic Landau symbols)**.** Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and $\{a_n\}_{n=1}^{\infty}$ be a sequence of constants. One writes

$$X_n = O_{\mathsf{P}}(a_n), \quad n \to \infty;$$

if and only if, for every positive real number $\epsilon > 0$, there exists a positive real number $M > 0$ and an integer $n_0 \in \mathbb{N}$ such that

$$\mathsf{P}\left[\left|\frac{X_n}{a_n}\right| > M\right] < \epsilon, \quad \forall n \geqslant n_0.$$

One writes

$$X_n = o_{\mathsf{P}}(a_n), \quad n \to \infty;$$

if and only if, for all positive real numbers $\epsilon > 0$ and $\tau > 0$, there exists an integer $n_0 \in \mathbb{N}$ such that

$$\mathsf{P}\left[\left|\frac{X_n}{a_n}\right| > \tau\right] < \epsilon, \quad \forall n \geqslant n_0.$$

Theorem A.1 (Slutsky's)**.** *Let* $\{\mathbf{X}_n\}_{n=1}^{\infty}$ *and* $\{\mathbf{Y}_n\}_{n=1}^{\infty}$ *be sequences of scalar or vector or matrix random elements. If*

$$\mathbf{X}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{X}$$

and

$$\mathbf{Y}_n \xrightarrow[n\to\infty]{\mathsf{P}} \mathbf{c},$$

where $\mathbf{c}$ *is a constant element, then (for suitable dimensions)*

(i) $\mathbf{X}_n + \mathbf{Y}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{X} + \mathbf{c}$*;*

(ii) $\mathbf{Y}_n\mathbf{X}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{c}\mathbf{X}$*;*

(iii) $\mathbf{Y}_n^{-1}\mathbf{X}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{c}^{-1}\mathbf{X}$*, provided that* $\mathbf{Y}_n$ *and* $\mathbf{c}$ *are invertible.*

Proof. See van der Vaart (1998, Lemma 2.8). □

Theorem A.2 (Continuous mapping)**.** *Let* $\{\mathbf{X}_n\}_{n=1}^{\infty}$ *be a sequence of random vectors in* $\mathbb{R}^k$*,* $\mathbf{X}$ *be a random vector in* $\mathbb{R}^k$*, and* $g : \mathbb{R}^k \to \mathbb{R}^m$ *be continuous at every point of a set* C *such that* $\mathsf{P}[\mathbf{X} \in C] = 1$*.*

(i) If $\mathbf{X}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{X}$*, then* $g(\mathbf{X}_n) \xrightarrow[n\to\infty]{\mathscr{D}} g(\mathbf{X})$*.*

(ii) If $\mathbf{X}_n \xrightarrow[n\to\infty]{\mathsf{P}} \mathbf{X}$*, then* $g(\mathbf{X}_n) \xrightarrow[n\to\infty]{\mathsf{P}} g(\mathbf{X})$*.*

(iii) If $\mathbf{X}_n \xrightarrow[n\to\infty]{a.s.} \mathbf{X}$*, then* $g(\mathbf{X}_n) \xrightarrow[n\to\infty]{a.s.} g(\mathbf{X})$*.*

Proof. See van der Vaart (1998, Theorem 2.3). □

Theorem A.3 (Cramér-Wold)**.** *Let* $\{\mathbf{X}_n\}_{n=1}^{\infty}$ *be a sequence of random vectors in* $\mathbb{R}^k$ *and* $\mathbf{X}$ *be a random vector in* $\mathbb{R}^k$*. Then,*

$$\mathbf{X}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{X}$$

if and only if

$$\mathbf{t}^\top \mathbf{X}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{t}^\top \mathbf{X}$$

for all $\mathbf{t} \in \mathbb{R}^k$.

Proof. A consequence of Lévy's continuity theorem (van der Vaart, 1998, Theorem 2.13). □

Theorem A.4 (Hájek-Rényi-Chow inequality). *If* $\{Y_n = \sum_{j=1}^n X_j, \mathcal{F}_n, n \geqslant 1\}$ *is an* L_2 *martingale and* $\{b_n, n \geqslant 1\}$ *is a positive, non-decreasing real sequence, then for any* $\lambda > 0$

$$\mathsf{P}\left[\max_{1\leqslant j\leqslant n}\left|\frac{Y_j}{b_j}\right| \geqslant \lambda\right] \leqslant \frac{1}{\lambda^2}\sum_{j=1}^n \frac{\mathsf{E}\,X_j^2}{b_j^2}.$$

Proof. See Chow and Teicher (2003, Theorem 8(iii)). □

List of Procedures

List of Figures

List of Tables

Bibliography

Albin, J. and Jarušková, D. (2003). On a test statistic for linear trend. *Extremes*, 6(3):247–258.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Antoch, J. and Hušková (2001). Permutation tests in change point analysis. *Statistics & Probability Letters*, 53(1):37–46.

Antoch, J. and Hušková, M. (2003). Detection of structural changes in regression. *Tatra Mountains Mathematical Publications*, 26:201–215.

Antoch, J., Hušková, M., and Prášková, Z. (1997). Effect of dependence on statistics for determination of change. *Journal of Statistical Planing and Inference*, 60(2):291–310.

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2008). Testing for changes in polynomial regression. *Bernoulli*, 14(3):637–660.

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, 157(1):78–92.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Bazarova, A., Berkes, I., and Horváth, L. (2014). Change point detection with stable AR(1) errors. In Kulik, R., editor, *Asymptotic methods in stochastics*, Fields Institute Communications. Springer. To appear.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons, New York, 1st edition.

Billingsley, P. (1986). *Probability and Measure.* John Wiley & Sons, New York, 2nd edition.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144.

Chan, J., Horváth, L., and Hušková, M. (2013). Change-point detection in panel data. *Journal of Statistical Planning and Inference*, 143(5):955–970.

Chen, Z. and Tian, Z. (2014). Ratio tests for variance change in nonparametric regression. *Statistics*, 48(1):1–16.

Chow, Y. S. and Teicher, H. (2003). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York, 3rd edition.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press Inc.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction For Econometricians*. Oxford University Press, New York.

Doukhan, P. (1994). *Mixing: Properties and Examples, Lecture Notes in Statistics, vol. 85*. Springer, New York.

Fitzenberger, B. (1997). The moving block bootstrap and robust inference for linear least squares and quantile regression. *Journal of Econometrics*, 82:235–287.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Horváth, L., Horváth, Z., and Hušková, M. (2008). Ratio tests for change point detection. In Balakrishnan, N., Peña, E. A., and Silvapulle, M. J., editors, *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, volume 1, pages 293–304, Beachwood, Ohio. IMS Collections.

Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648.

Huber, P. J. et al. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Hušková, M. (2004). Permutation principle and bootstrap in change point analysis. *Fields Institute Communications*, 44:273–291.

Hušková, M. (2007). Ratio type test statistics for detection of changes in time series. In Gomes, M. I., Pinto Martins, J. A., and Silva, J. A., editors, *Bulletin of the International Statistical Institute, Proceedings of the 56th Session*, volume 976, Lisboa.

Hušková, M., Kirch, C., Prášková, Z., and Steinebach, J. (2008). On the detection of changes in autoregressive time series, ii. resampling procedures. *Journal of Statistical Planning and Inference*, 138(6):1697–1721.

Hušková, M. and Marušiaková, M. (2012). M-procedures for detection of changes for dependent observations. *Communications in Statistics – Simulation and Computation*, 41(7):1032–1050.

Hušková, M. and Picek, J. (2002). M-tests for detection of structural changes in regression. In *Statistical data analysis based on the L_1-norm and related methods*, pages 213–227. Springer.

Hušková, M. and Picek, J. (2004). Some remarks on permutation type tests in linear models. *Discuss. Math. Probab. Stat*, 24:151–182.

Hušková, M. and Picek, J. (2005). Bootstrap in detection of changes in linear regression. *Sankhyā: The Indian Journal of Statistics*, 67(2):200–226.

Hušková, M., Prášková, Z., and Steinebach, J. (2007). On the detection of changes in autoregressive time series i. asymptotics. *Journal of Statistical Planning and Inference*, 137(4):1243–1259.

Hušková, M. and Steinebach, J. (2000). Limit theorems for a class of tests of gradual changes. *Journal of Statistical Planning and Inference*, 89(1-2):57–77.

Hušková, M. and Kirch, C. (2012). Bootstrapping sequential change-point tests for linear regression. *Metrika*, 75(5):673–708.

Hušková, M., Kirch, C., Prášková, Z., and Steinebach, J. (2008). On the detection of changes in autoregressive time series, II. Resampling procedures. *Journal of Statistical Planning and Inference*, 138(6):1697–1721.

Ibragimov, I. A. (1959). Some limit theorems for stochastic processes stationary in the strict sense. *Doklady Akademii Nauk SSSR (In Russian)*, 125:711–714.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, The Netherlands.

Jandhyala, V. and MacNeill, I. (1997). Iterated partial sum sequences of regression residuals and tests for changepoints with continuity constraints. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):147–156.

Jarušková, D. (1998). Testing appearance of linear trend. *Journal of Statistical Planning and Inference*, 70:263–276.

Jarušková, D. (1999). Testing appearance of polynomial trend. *Extremes*, 2(1):25–37.

Jarušková, D. (2003). Asymptotic distribution of a statistic testing a change in simple linear regression with equidistant design. *Statistics & Probability Letters*, 64(1):89–95.

Jurečková, J., Sen, P. K., and Picek, J. (2012). *Methodology in Robust and Nonparametric Statistics*. CRC Press, Boca Raton, FL.

Katz, M. L. (1963). Note on the Berry-Esseen theorem. *Annals Mathematical Statistics*, 34(3):1107–1108.

Kim, J.-Y. (2000). Detection of change in persistence of a linear time series. *Journal of Econometrics*, 95(1):97–116.

Kim, J.-Y. and Amador, R. B. (2002). Corrigendum to detection of change in persistence of a linear time series. *Journal of Econometrics*, 109(2):389–392.

Kirch, C. (2006). *Resampling Methods for the Change Analysis of Dependent Data*. PhD thesis, University of Cologne, Germany.

Lahiri, S., Furukawa, K., and Lee, Y.-D. (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Statistical Methodology*, 4(3):292–321.

Leybourne, S. J. and Taylor, A. R. (2006). Persistence change tests and shifting stable autoregressions. *Economics Letters*, 91(1):44–49.

Lindner, A. M. (2009). Stationarity, mixing, distributional properties and moments of GARCH(p,q)-processes. In Andersen, T. G., Davis, R. A., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 481–496, Berlin. Springer.

Liu, Y., Zou, C., and Zhang, R. (2008). Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics – Theory and Methods*, 37(16):2551–2563.

Madurkayová, B. (2007). Ratio tests for gradual changes. In *WDS'07 Proceedings of Contributed Papers: Part I - Mathematics and Computer Sciences*, pages 175–180, Prague. MATFYZPRESS.

Madurkayová, B. (2009a). Ratio type statistics for detection of changes in linear regression models. In Antoch, J. and Dohnal, G., editors, *Sborník prací 15. letní školy JČMF ROBUST 2008*, pages 247–254.

Madurkayová, B. (2009b). Ratio type statistics for detection of changes in location. In *Proceedings of the 16th European Young Statisticians Meeting*, pages 144–148.

Madurkayová, B. (2011). Ratio type statistics for detection of changes in mean. *Acta Universitatis Carolinae: Mathematica et Physica*, 52(1):47–58.

Met Office Hadley Centre (2008). Met Office Hadley Centre observations datasets. HadCRUT3. [Online; Available from http://www.metoffice.gov.uk/hadobs/; Updated February 2, 2008; Accessed October 16, 2008].

Meyer, P. A. (1966). *Probability and Potentials*. Blaisdell Publishing, New York.

Meyers, G. G. and Shi, P. (2011). Loss reserving data pulled from NAIC Schedule P. [Online; Available from http://www.casact.org/research/index.cfm?fa=loss_reserves_data; Updated September 01, 2011; Accessed June 10, 2014].

Peštová, B. and Pešta, M. (2015). Testing structural changes in panel data with small fixed panel size and bootstrap. *Metrika*, DOI 10.1007/s00184-014-0522-8.

Pešta, M. and Hudecová, Š. (2012). Asymptotic consistency and inconsistency of the chain ladder. *Insurance: Mathematics and Economics*, 51(2):472–479.

Politis, D. N. and Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In LePage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*. Wiley, New York.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23:53–70.

Prague Stock Exchange (2015). PX Index. [Online; Available from https://www.pse.cz/dokument.aspx?k=Burzovni-Indexy; Updated April 30, 2015; Accessed April 30, 2015].

Prášková, Z. and Chochola, O. (2014). M-procedures for detection of a change under weak dependence. *Journal of Statistical Planning and Inference*, 149:60–76.

Rosenblatt, M. (1971). *Markov Processes: Structure and Asymptotic Behavior*. Springer-Verlag, Berlin.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, New York.

Wang, D., Guo, P., and Xia, Z. (2014). Detection and estimation of structural change in heavy-tailed sequence. *Communications in Statistics – Theory and Methods*. Submitted.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 52(1):45–57.

Zhao, W., Tin, Z., and Xia, Z. (2010). Ratio test for variance change point in linear process with long memory. *Statictical Papers*, 51:397–407.

Zhao, W., Xia, Z., and Tian, Z. (2011). Ratio test to detect change in the variance of linear process. *Statistics*, 45:189–198.