Dynamics and Instabilities in Time Series and Panel Data

Michal Pešta

2019

CHARLES UNIVERSITY
Faculty of mathematics
and physics

Charles University

Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Habilitation Thesis

Dynamics and Instabilities in Time Series and Panel Data

Michal Pešta

Mathematics – Probability and Mathematical Statistics

September 1, 2019

Michal Pešta

Dynamics and Instabilities in Time Series and Panel Data

Habilitation Thesis

September 1, 2019

Charles University
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 49/83
186 75 Prague 8 – Karlín
Czech Republic

To my wife Barbora
and to my children Gréta, Samuel, and Michal

Abstract

Changing behavior of data—smooth or abrupt. That is our primary goal of investigation to be presented in this habilitation thesis. Several contributions of the author to the branch of mathematics called probability and mathematical statistics are demonstrated. Especially, the main focus lies in detection of instabilities, analysis of changes, and modeling of dynamics when allowing for some uncertainty or imprecision in the observations. The data structures considered are time series, panel data, and triangular data. All of them are contaminated by random fluctuations or unobservable disturbances.

Preface

> *Omnia mutantur, nihil interit.*
>
> — **Publius Ovidius Naso**
> (Metamorphoses, book XV, line 165)

An intensive research has sprung up for methods to handle dynamic systems containing randomness and, simultaneously, to analyze structural changes in data. This habilitation thesis is structured into three core chapters with respect to the type of data governed and the corresponding type of problems dealt.

Firstly, the introductory chapter sets up a framework of the thesis. There are main problems and goals described. Moreover, the corresponding scientific papers, on which this thesis is based, are presented.

The second chapter deals with detection and estimation of so-called changepoints in time series. Ratio type and self-normalized type test statistics for at most one change are introduced. The testing methods are enriched by resampling techniques, which serve as improving add-ons. All approaches are invented in a way that the whole testing procedure is supposed to be nuisance-parameter-free. Consistent estimators for a change in mean and a change in trend of time series are developed.

In the third chapter, structural breaks in panel data are covered, especially the cases when the panels are short. Hypothesis testing for a common change in panels is performed. Again, the resampling methods are utilized in order to overcome difficulties and imperfections of the asymptotic versions of the detection procedures. Furthermore, various competing estimators for the tested common change are introduced and shown to be consistent.

The fourth chapter is devoted to modeling of dynamics in triangular data, which can be considered as a special case of panel data. Justification of the traditional chain ladder method is given at long last. Moreover, modeling approaches capable

of handling dependencies, like generalized estimating equations or conditional least squares with copulae, are implemented within the triangular data framework.

Afterwards, our conclusions follow. Finally, the original versions of the foundation papers are attached in the appendix.

Even though this habilitation thesis was originally regarded as homeopathic enforcement of the author's inventiveness by himself, I have finally figured out that I felt a compulsion to babble on about what I was truly researching.

Acknowledgement

First of all, I would like to thank my whole family. My deepest and warmest thanks belong to Barborka for love and patience. I would also like to express my ecstatic gratitude to my kids—Grétka, Samko, and Miško—for inspiration. Without them this thesis could not have been completed. Furthermore, I warmly appreciate the care and love of my parents—Vierka and Zdeněk.

I would also like to thank to all my mentors and colleagues for introducing me into the world of higher mathematics, providing me important theoretical knowledge, discussing interesting topics, and collaborating on scientific papers. They have inspired me with new ideas and opened my mind for stochastic thinking. Also their practical advice and suggestions have directed me in my work whenever I have needed it. I warmly appreciate

Last, but not least, I would like to express my thanks to all my friends, whose encouragement have meant to me so much during the pursuit of my postdoctoral research and the composition of this thesis.

This work was supported by the Czech Science Foundation projects GAČR No. 13-12994P, P402/12/G097, 15-04774Y, 18-00522Y, 18-01781Y, and by the Charles University Progres Programme Q49.

Prague, September 1, 2019

Michal Pešta

Contents

1 Introduction

> *Where we're going, we don't need roads.*
>
> — **Christopher Lloyd**
> (As Dr. Emmett Brown in 'Back to the Future')

A character of a *change* can basically be *rapid* or *gradual*. We are going to stochastically inspect both types of the change in data. The considered data structures, which possibly possess some change, are *time series* and *panel data* (or their mutations and combinations). We will elaborate on them in the consequent section.

There are essentially two main statistical challenges with respect to existence of the change: *testing* and *estimation*. In the first case, one is interested in answering a question, whether there is some change present or not. This can be schematically represented by a decision between two statements: A null hypothesis

$$\mathcal{H}_0 : \text{ there is no change present;}$$

versus an alternative

$$\mathcal{H}_1 : \text{ there exists a change.}$$

This situation can be simply represented by two disjunctive sets as in Figure 1.1, where the true reality ($\mathcal{H}_0$ or $\mathcal{H}_1$) is and will remain unknown to us, whereas our decision is encumbered by an error. And this error is desired to be controlled.

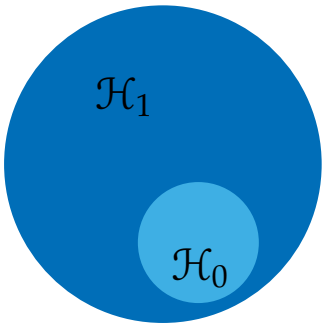

Fig. 1.1. Null hypothesis versus alternative.

For the second case, the aim is to predict the location or, alternatively, characteristics of the change based on the observed data. And again, the true change description is

and will stay unknown to us. However, imprecision of our estimation based purely on the obtained data is intended to be kept reasonably small.

1.1 Vague talk

Whatever the changes you may think (for instance, structural breaks or continuous dynamics), we try to investigate their presence in *various data structures*.

The first type of data considered are *time series*, where one observes finite part of a random sequence, i.e., $\{Y_i\}_{i=1}^n$ as shown in Figure 1.2.

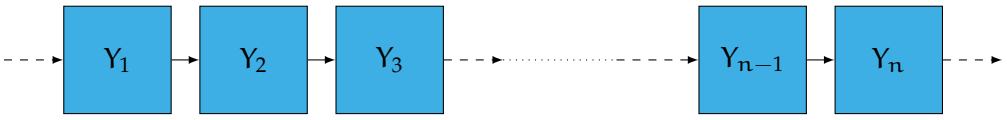

Fig. 1.2. Time series.

Nevertheless, we will consider a generalized structure of the time series that can be organized in so-called *triangular scheme*. Here, for the nth realization (or measurement), one possesses a vector $\{Y_i^{(n)}\}_{i=1}^{k_n}$ as visualized in Figure 1.3. The usual case is that $k_n \equiv n$, but it does not have to be.

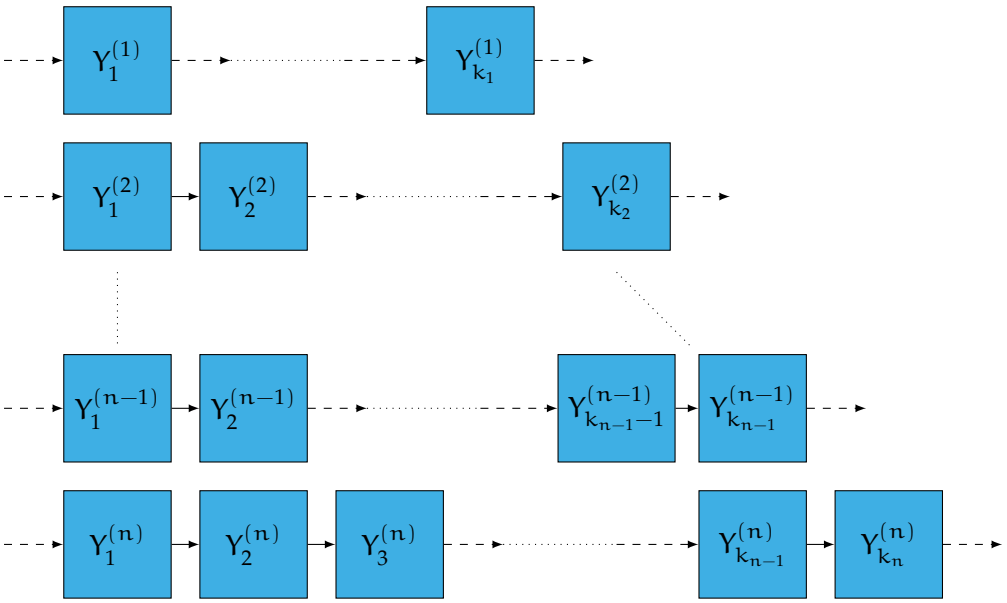

Fig. 1.3. Triangular scheme of a time series.

Panel data are the second type of data structures, where several finite or infinite dimensional vectors are recorded. In Figure 1.4, we have so-called *balanced* panel data $\{Y_j^{(i)}\}_{j=1,i=1}^{t,n}$ consisting of n panels such that the ith panel $[Y_1^{(i)}, \dots, Y_t^{(i)}]$ is an observed vector with a common length t. One can also think of *unbalanced* panel data, where the panels do not necessarily be of the same length.

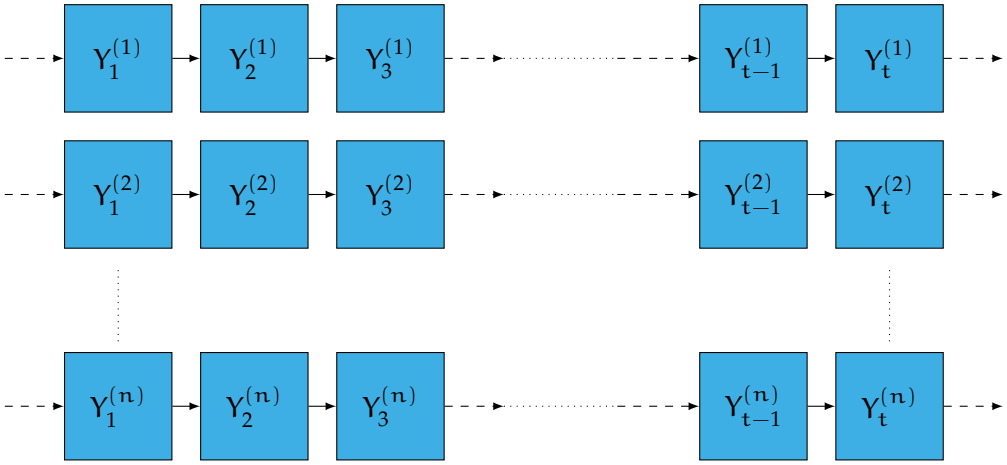

Fig. 1.4. Panel data (balanced).

The third type of data structure contemplated are *triangular data*, which can be considered as a special case of the unbalanced panel data. Here, one measures n panels such that the ith panel has a length $n - i + 1$. This property forms the data in the shape of a *right angled isosceles*, i.e., $\{Y_j^{(i)}\}_{j=1,i=1}^{n-i+1,n}$ as illustrated in Figure 1.5. Hence, each next panel is one observation shorter.

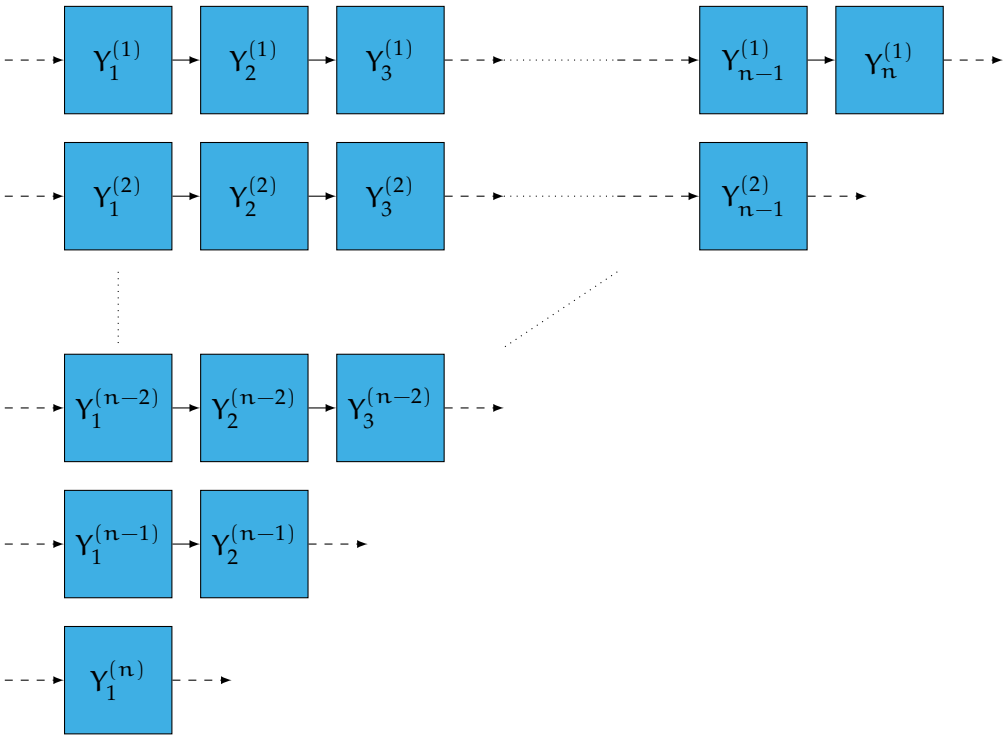

Fig. 1.5. Triangular data.

1.2 Motivation

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 finance, insurance, medicine, biology, hydrology, climatology, or ecology.

In recent years, an intensive research has sprung up for methods to handle *structural changes and developments* in various data structures. A prime motivation for this thesis is to introduce *stochastic models* capable of handling *dynamics and instabilities* of data. Our goal is to clearly summarize already developed properties of these models. Consequently, we want to derive additional important properties of such models possibly allowing for so-called *changepoints*, to extend their applicability, and to overcome some problems that exist up to now. We mainly take into account modern asymptotic approaches in order to finalize unsolved questions regarding usage of the changepoint models. Therefore, we will concentrate on and incorporate the following:

- finite sample properties and limiting behavior,
- computational feasibility,
- robust approaches,
- weak dependence of errors,
- simulation studies,
- real data analyses.

1.3 Thesis structure and results

This habilitation thesis is based on nine original papers published in scientific journals and one erratum. The mathematical and statistical problems contained in these papers are structured into three chapters. At the end of each chapter, very brief summaries in a form of *key contributions* are listed.

1.3.1 Time series

Chapter 2 deals with structural breaks in the time series framework (cf. Figure 1.2 and 1.3). The basis for this chapter is formed by the following papers:

- Peštová, B. and Pešta, M. (2018). 'Abrupt change in mean using block bootstrap and avoiding variance estimation'. *Computational Statistics* 33(1), 413–441.
  doi.org/10.1007/s00180-017-0785-4

- Pešta, M. and Wendler, M. (2019). 'Nuisance-parameter-free changepoint detection in non-stationary series'. *TEST*, Online First.
  doi.org/10.1007/s11749-019-00659-1

- Pešta, M. (2016). 'Unitarily invariant errors-in-variables estimation'. *Statistical Papers* 57(4), 1041–1057.
  doi.org/10.1007/s00362-016-0800-9

- Pešta, M. (2017). 'Block bootstrap for dependent errors-in-variables'. *Communications in Statistics – Theory and Methods* 46(4), 1871–1897.
  doi.org/10.1080/03610926.2015.1030423

1.3.2 Panel data

Changepoint problems within the panel data setup are elaborated in Chapter 3. The main aim lies in the changepoint testing and estimation, where the panel's length can be relatively short. The second group of papers, building up the results for this chapter, is as follows:

- Peštová, B. and Pešta, M. (2015). 'Testing structural changes in panel data with small fixed panel size and bootstrap'. *Metrika* 78(6), 665–689.
  doi.org/10.1007/s00184-014-0522-8

- Peštová, B. and Pešta, M. (2016). 'Erratum to: Testing structural changes in panel data with small fixed panel size and bootstrap'. *Metrika* 79(2), 237–238.
  doi.org/10.1007/s00184-015-0562-8

- Peštová, B. and Pešta, M. (2017) 'Change point estimation in panel data without boundary issue'. *Risks* 5(1), 7.
  doi.org/10.3390/risks5010007

1.3.3 Triangular data

Chapter 4 investigates dynamic behavior within the triangular data structures. The goal is to develop methods capable to capture dependence among observations structured in the right angled isosceles. The third group of papers summarizing these outcomes are predominantly having mathematical applications in insurance:

- Pešta, M. and Hudecová, Š. (2012). 'Asymptotic consistency and inconsistency of the chain ladder'. *Insurance: Mathematics and Economics* 51(2), 472–479. doi.org/10.1016/j.insmatheco.2012.07.004

- Hudecová, Š. and Pešta, M. (2013). 'Modeling dependencies in claims reserving with GEE'. *Insurance: Mathematics and Economics* 53(3), 786–794. doi.org/10.1016/j.insmatheco.2013.09.018

- Pešta, M. and Okhrin, O. (2014). 'Conditional least squares and copulae in claims reserving for a single line of business'. *Insurance: Mathematics and Economics* 56, 28–37. doi.org/10.1016/j.insmatheco.2014.02.007

Afterwards, our conclusions follows. The original published papers are put in the Appendix A. The main body of the thesis does not contain simulation studies, practical data analyses, or proofs. However, they can be found in the papers attached within the Appendix A.

2 Instabilities in Times Series

> *There is nothing stable in the world; uproar's your only music.*
>
> — **John Keats**
> (Poet of the English Romantic movement)

We deal with sequences of weakly *dependent* observations that are naturally ordered in time. Their mean is supposed to be constant or having linear trend. Although, we alternatively think of a possible *instability* of the unchanging mean or constant trend: The mean or trend is subject to change at most once at some unknown time point—occurrence of a *structural break*. The primary aim is to test whether such instability is present or not. The secondary goal is to estimate such changepoint, if it has been detected.

2.1 Change in mean using ratio statistics

The problem of an unknown change in mean of the time series is studied and procedures for detection of such change within the observed time ordered sequence are presented. The considered underlying stochastic model allows *at most one change*. Moreover, the change is rapid, also called abrupt, which happens suddenly at some unknown time point.

The ratio type test statistic elaborated in this section is derived from the non-ratio type test statistics (e.g., Csörgő and Horváth (1997)) based on partial sums of the residuals that are commonly used in the changepoint analysis. They do not need to be standardized by any variance estimate, which makes them a suitable alternative for the non-ratio type test 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. However, the variance estimators often do not perform well even in the i.i.d. case, especially under alternatives (Antoch et al., 1997).

We extend the ideas for the ratio type test statistics presented by Hušková (2007) and Horváth et al. (2008) in the way that *weakly dependent errors* of the model are considered together with incorporating a *general score function* in the test statistics. In order to obtain critical values for the studied test statistic not only from their asymptotic distributions, we focus on the *circular moving block bootstrap method* by Politis and Romano (1992). This type of resampling procedure was applied in a similar situation by Kirch (2006).

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 test 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 test statistics based on the cumulative sums of the 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 a change from the asymptotic stationary sequence into the 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. Antoch and Hušková (2001), Hušková (2004), and Kirch (2006) dealt with permutation principles and bootstrapping in the changepoint analysis.

Generally, there are two main distinguishable classes of changepoint detection procedures: parametric (Csörgő and Horváth, 1997:Chapter 1) and nonparametric (Csörgő and Horváth, 1997:Chapter 2). The parametric ones rely on an underlying known parametric form of the distribution, where only the distributional parameters are unknown. A typical representative of the parametric methods are the likelihood ratio (LR) based methods (Andrews, 1993). If one assumes additional (regularity) assumptions, for instance, regarding the Fisher information matrix, some optimality properties can be derived. E.g., a Wald-like type or a Bahadur-like type of optimality. Optimal detections within the changepoint framework were investigated by, e.g., Andrews and Ploberger (1994) or Andrews et al. (1996). Although, one has to pay for these properties by assuming a specific parametric form of the distribution and by adding regularity conditions, which can be sometimes hard to verify. On the other hand, a classical representative of the nonparametric class are cumulative sums (CUSUM) based approaches that do not assume a parametric distribution, which gives them wider applicability. Our proposed methods belong to the CUSUM-based techniques.

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

tification of the ratio type test statistic for change in mean detections was elaborated by Madurkayová (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. A similar ratio type statistic as introduced in this section is studied to analyse change in mean for heavy tailed distributions in the work of Wang et al. (2016). This leads to a more general asymptotic distribution under the null hypothesis, which is a functional of stable Lévy processes. A bootstrap approximation method to determine the critical values and changepoint estimation using the ratio method are also discussed. Bazarova et al. (2015) 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 test statistics are compared by simulations. Peštová and Pešta (2015) and Peštová and Pešta (2016) applied the ratio type test statistics for detections of the structural changes in panel data.

A background of the following work comes from Peštová and Pešta (2018a).

2.1.1 Abrupt changepoint model

Let us consider observations $Y_{1,n}, \dots, Y_{n,n}$ obtained at time ordered points. We are interested in testing the null hypothesis of all observations being random variables having equal mean values. Our goal is to test against the alternative of the first τ_n observations have mean value μ and the remaining $n - \tau_n$ observations come from distributions with mean values $\mu + \delta_n$, $\delta_n \neq 0$. We can describe the situation as a model with *at most one abrupt change in constant mean*

$$Y_{k,n} = \mu + \delta_n \, \mathbb{1}\{k > \tau_n\} + \varepsilon_k, \quad k = 1, \dots, n, \tag{2.1}$$

where μ, δ_n and τ_n are unknown parameters, $\varepsilon_1, \dots, \varepsilon_n$ are random errors, and $\mathbb{1}\{A\}$ denotes the indicator of set A.

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. 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 the variables depending on the latter) whenever possible. However, we have to keep in mind that in the asymptotic results below, as n increases over all bounds, both $\delta = \delta_n$ and

$\tau = \tau_n$ may be changing when n is increasing. Then, model (2.1) can be rewritten as the *abrupt changepoint* model

$$Y_k = \mu + \delta\, \mathbb{1}\{k > \tau\} + \varepsilon_k, \quad k = 1, \dots, n. \tag{2.2}$$

The time point τ is called the *changepoint*.

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

$$H_0 : \tau = n \tag{2.3}$$

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

$$H_1 : \tau < n,\ \delta \neq 0. \tag{2.4}$$

2.1.2 Ratio type test statistic based on M-residuals

The main focus is given to the test procedures based on the ratio type test statistics that are functionals of the partial sums of M-type residuals. We robustify the original ratio type test statistic described by Hušková (2007) and Horváth et al. (2008). A test statistic based on the M-residuals is considered

$$\mathcal{R}_n(\psi, \gamma) = \max_{n\gamma \leqslant k \leqslant n - n\gamma} \frac{\max\limits_{1 \leqslant i \leqslant k} \left| \sum_{j=1}^{i} \psi(Y_j - \widehat{\mu}_{1k}(\psi)) \right|}{\max\limits_{k \leqslant i \leqslant n-1} \left| \sum_{j=i+1}^{n} \psi(Y_j - \widehat{\mu}_{2k}(\psi)) \right|}, \tag{2.5}$$

where $0 < \gamma < 1/2$ is a given constant, $\widehat{\mu}_{1k}(\psi)$ is an M-estimate of parameter μ based on the observations $Y_1, \dots, Y_k$ (under H_0), and $\widehat{\mu}_{2k}(\psi)$ is an M-estimate of μ based on the observations $Y_{k+1}, \dots, Y_n$ (under H_0). That means, $\widehat{\mu}_{1k}(\psi)$ is a solution of the estimating equation

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

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

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

Let us remark that the test statistic does not require an estimate of the possible changepoint. 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.

Strong mixing dependence

Prior to deriving asymptotic properties of the test statistic, we summarize the notion of strong mixing (α-mixing) dependence in more detail, which will be imposed on the model's errors.

Prior to postulating an *errors' assumption*, we summarize the notion of *strong mixing* (α-mixing) dependence in more detail, which will be imposed on the model's errors.

Suppose that $\{\xi_n\}_{n=1}^{\infty}$ is a sequence of random elements on a probability space $(\Omega, \mathcal{F}, P)$. For σ-fields $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$, let

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

Intuitively, $\alpha(\cdot|\cdot)$ measures the dependence of the events in $\mathcal{B}$ on those in $\mathcal{A}$. There are many ways in which one can describe weak dependence or, in other words, *asymptotic independence* of random variables, see Bradley (2005). Considering a filtration $\mathcal{F}_m^n := \sigma\{\xi_i \in \mathcal{F}, m \leqslant i \leqslant n\}$, sequence $\{\xi_n\}_{n=1}^{\infty}$ of random variables is said to be *strong mixing* (α-mixing) if $\alpha(\xi_\circ, n) := \sup_{k \in \mathbb{N}} \alpha(\mathcal{F}_1^k | \mathcal{F}_{k+n}^{\infty}) \to 0$ as $n \to \infty$. The notion of α-mixing was introduced by Ibragimov (1959).

Anderson (1958) comprehensively analyzed a class of m-dependent processes. They are α-mixing, since they are finite order ARMA processes with innovations satisfying *Doeblin's condition* (Billingsley, 1968:p. 168). 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 processes just mentioned are mixing as well. This means that the class of the α-mixing processes is sufficiently large for the further practical applications and that is why we chose such a mixing condition.

Limit distribution under null hypothesis and alternative

We proceed to the assumptions that are needed for deriving asymptotic properties of the proposed test statistic. Before that, assumptions for the score function ψ and the distribution of the random errors $\varepsilon_1, \ldots, \varepsilon_n$ are formulated.

Assumption $\mathcal{A}$.2.1. 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

$$\sum_{h=0}^{\infty} (h+1)^{\chi/2} \alpha(\varepsilon_\circ, h)^{\chi'/(2+\chi+\chi')} \leqslant C_1(\chi, \chi'), \tag{2.6}$$

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

Assumption $\mathcal{A}$.2.2. The score function ψ is a non-decreasing and antisymmetric function.

Assumption $\mathcal{A}$.2.3. $\int |\psi(x)|^{2+\chi+\chi'} dF(x) < \infty$ and

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

for some constants $1 \leqslant \eta \leqslant 2 + \chi + \chi'$, $\chi > 0$, $\chi' > 0$ as in (2.6) and the constants $C_2(\chi, \chi') > 0$, $C_3(\chi, \chi') > 0$ both depending only on χ and χ'.

Assumption $\mathcal{A}$.2.4. Let us denote $\lambda(t) = -\int \psi(e - t) dF(e)$, 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 $\mathcal{A}$.2.5. $0 < \sigma^2(\psi) = \mathsf{E}\,\psi^2(\varepsilon_1) + 2\sum_{i=1}^{\infty} \mathsf{E}\,\psi(\varepsilon_1)\psi(\varepsilon_{i+1}) < \infty$.

Assumption $\mathcal{A}$.2.1 is satisfied for example for the ARMA processes with continuously distributed stationary innovations and bounded variance (Doukhan, 1994:Section 2.4). The conditions regarding ψ reduce to the moment restrictions for $\psi_{L_2}(x) = x$ (L_2 method) taking $\eta = 2 + \chi + \chi'$. For $\psi_{L_1}(x) = \operatorname{sgn}(x)$ (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.,

$$\psi_H(x) = x\ \mathbb{1}\{|x| \leqslant C\} + C \operatorname{sgn}(x)\ \mathbb{1}\{|x| > C\} \tag{2.7}$$

for some $C > 0$. In that case to satisfy Assumptions $\mathcal{A}$.2.2–$\mathcal{A}$.2.4, we need to assume F being a symmetric distribution function with the continuous density f in a neighborhood of C and $-C$ satisfying $f(C) > 0$ and $f(-C) > 0$ with $\eta = 2 + \chi + \chi'$.

For $\psi_{L_2}(x) = x$, $x \in \mathbb{R}$, the above stated Assumptions $\mathcal{A}$.2.2 and $\mathcal{A}$.2.4 are satisfied. We can also drop the requirement of symmetry of F in Assumption $\mathcal{A}$.2.1 and replace it by $\mathsf{E}\,\varepsilon_1 = 0$. Assumptions $\mathcal{A}$.2.3 and $\mathcal{A}$.2.5 reduce to the following two assumptions.

Assumption $\mathcal{A}$.2.6. $\mathsf{E}\,|\varepsilon_1|^{2+\beta} < \infty$ for some constant $\beta > 0$.

Assumption $\mathcal{A}$.2.7. $0 < \sigma^2(\psi_{L_2}) = \mathsf{E}\,\varepsilon_1^2 + 2\sum_{i=1}^{\infty} \mathsf{E}\,\varepsilon_1\varepsilon_{i+1} < \infty$.

Although we consider only symmetric distribution function F in our approach, there is also another common source of outlying observations which occur in practical applications—asymmetric random error distribution (i.e., $\nexists x_0 \in \mathbb{R} : F(x_0 - x) = 1 - F(x_0 + x), \forall x > 0$). To generalize our approach for asymmetric distribution function F, one needs to assure in modified Assumption $\mathcal{A}$.2.4 that there exists some unique $t_0 \in \mathbb{R}$ such that $\lambda(t_0) = 0$. Correspondingly, all other assumptions required to hold in the neighborhood of $t = 0$ need to be satisfied in the neighborhood of $t = t_0$ instead.

Henceforth, $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{D}$ convergence in distribution, $\xrightarrow[n\to\infty]{D[0,1]}$ weak convergence in the Skorokhod space $D[0,1]$ of càdlàg functions on $[0,1]$, and $[x]$ denotes the integer part of the real number x.

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

Theorem 2.1.1 (Under null)**.** *Suppose* $Y_1, \ldots, Y_n$ *follow model* (2.2) *and assume that Assumptions* $\mathcal{A}$.2.1–$\mathcal{A}$.2.5 *hold. Then, under null hypothesis* (2.3)

$$\mathcal{R}_n(\psi,\gamma) \xrightarrow[n\to\infty]{D} \sup_{\gamma\leqslant t\leqslant 1-\gamma} \frac{\sup\limits_{0\leqslant u\leqslant t} \left|\mathcal{W}(u) - u/t\mathcal{W}(t)\right|}{\sup\limits_{t\leqslant u\leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|}, \tag{2.8}$$

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

The null hypothesis is rejected for large values of $\mathcal{R}_n(\psi,\gamma)$. We reject H_0 at significance level α if $\mathcal{R}_n(\psi,\gamma) > a_{1-\alpha,\gamma}$, where $a_{1-\alpha,\gamma}$ is the $(1-\alpha)$-quantile of the asymptotic distribution from (2.8). Explicit form of the limit distribution (2.8) under the null hypothesis is not known. Therefore, in order to obtain the critical values, we have to use either simulation from the limit distribution or resampling methods.

The lower bound for k in (2.5) 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 of (2.8) may be replaced with $\sup_{0<t\leqslant 1-\gamma}$. However, 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 $\lim_{t\to 1-} \sup_{t\leqslant s\leqslant 1} \left|\mathcal{W}(s) - \frac{s}{t}\mathcal{W}(t)\right| = 0$ a.s. Nevertheless, one may redefine the test statistics $\mathcal{R}_n(\psi,\gamma)$ by interchanging its numerator and denominator in order to relax the upper bound for k up to n

(or $n-1$, since the reversed ratio becomes 0 for $k = n$) and, correspondingly, $\sup_{\gamma \leqslant t \leqslant 1-\gamma}$ in the new limit distribution may be replaced with $\sup_{\gamma < t \leqslant 1}$.

Next, we study the behavior of the test statistic under the alternative.

Theorem 2.1.2 (Under alternative). *Suppose* $Y_1, \ldots, Y_n$ *follow model* (2.2), *assume that* $\delta_n = O(n^\theta)$ *as* $n \to \infty$ *for* $\theta \in \left(-\frac{1}{2}, \frac{\eta}{3(2+\chi+\chi')} - \frac{1}{2}\right)$, *and* $\tau = [\zeta n]$ *for some* $\gamma < \zeta < 1-\gamma$. *Then, under Assumptions* $\mathcal{A}.2.1$–$\mathcal{A}.2.5$ *and alternative* (2.4)

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

Theorem 2.1.2 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 2.1.1 can be used to construct the test.

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

Asymptotic critical values

The explicit form of the limit distribution (2.8) is not known. The critical values may be determined by simulations from the limit distribution from Theorem 2.1.1. Theorem 2.1.2 ensures that we reject the null hypothesis for large values of the test statistic. We tried to simulate the asymptotic distribution (2.8) 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. In Table 2.1, we present several critical values for $\gamma = 0.1$ and $\gamma = 0.2$.

Tab. 2.1. Simulated critical values corresponding to the asymptotic distribution of the test statistic $\mathcal{R}_n(\psi, \gamma)$ under the null hypothesis

$100(1-\alpha)\%$	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

Note that the numerator and denominator in the test statistic $\mathcal{R}_n(\psi, \gamma)$ can be interchanged and such a modified test statistic can still be used for detection of the abrupt change in mean (but using different critical values).

2.1.3 Block bootstrap with replacement

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 on 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

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

With this method, there is equal probability for each observation to be included in the bootstrap sample.

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., a situation where the set of n observations can be divided in exactly L blocks of length K. It can be proved, as in 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 (standard) non-circular moving block bootstrap (Künsch, 1989), where one does not append the first few consequent observations from the original sequence after the last observation. This bootstrap version 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 $0 < \gamma < 1/2$

$$\begin{aligned}
\Pi_{l,k,L,K} = \big\{&(p, q) :\ p, q \in \mathbb{N},\\
&1 \leqslant p \leqslant l,\ 1 \leqslant q \leqslant K,\ (p-1)K + q \leqslant (l-1)K + k\big\},\\
\widetilde{\Pi}_{l,k,L,K} = \big\{&(p, q) :\ p, q \in \mathbb{N},
\end{aligned}$$

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

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

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

for $p,l=1,\dots,L$, $q,k=1,\dots,K$, $p\leqslant l$, $(p-1)K+q\leqslant (l-1)K+k$, where $m^{\boldsymbol{U}}_{L,K}(l,k)$ is an M-estimate that solves the estimating equation

$$\sum_{r=1}^{l-1}\sum_{s=1}^{K}\psi(Y_{U_r+s}-\mu)+\sum_{s=1}^{k}\psi(Y_{U_l+s}-\mu)=0$$

with respect to μ. Similarly, we define

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

for $p,l=1,\dots,L$, $q,k=1,\dots,K$ such that $p\geqslant l$, $(p-1)K+q\geqslant (l-1)K+k+1$ and $\tilde{m}^{\boldsymbol{U}}_{L,K}(l,k)$ is an M-estimate that solves the estimating equation

$$\sum_{s=k+1}^{K}\psi\left(Y_{U_l+s}-\mu\right)+\sum_{r=l+1}^{L}\sum_{s=1}^{K}\psi\left(Y_{U_r+s}-\mu\right)=0.$$

Now, define the block bootstrap version of $\mathcal{R}_n(\psi)$ from (2.5) by

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

Statistic $\mathcal{R}^{*}_{L,K}(\psi,\gamma)$ is constructed in a similar fashion as the original ratio type test statistic $\mathcal{R}_n(\psi,\gamma)$. The idea behind the bootstrap test statistic lies in indexing the randomly chosen (possibly overlapping) bootstrap blocks by $l=1,\dots,L$. The first l blocks are used in the numerator of the bootstrap statistic. The lth block is employed in the numerator as well as in the denominator. The last $L-l+1$

blocks are used in the denominator of the statistic $\mathcal{R}^*_{L,K}(\psi,\gamma)$. Regarding the lth block appearing in the numerator and denominator, this particular block is split into two continuous disjunctive parts: the first one contains the first elements from the lth block, has k elements, and is used for the numerator; the second part contains the last elements from the lth block, has $K-k$ elements, and is used for the denominator. So, there does not exist a bootstrapped observation appearing simultaneously in the numerator and denominator. An algorithm for the circular block bootstrap is illustratively shown in Procedure 2.1.1 and its validity will be proved in Theorem 2.1.3.

Procedure 2.1.1 Bootstrapping test statistic $\mathcal{R}_n(\psi,\gamma)$

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

1: determine number of blocks $L=[n/K]$
2: define set $\Omega_{L,K}(\gamma)$
3: **for** $b=1$ to B **do** // repeat in order to obtain the empirical distribution
4: generate random sample $\boldsymbol{U}=(U_1,\dots,U_L)$ from discrete uniform distribution on $\{0,\dots,n-1\}$
5: **for** $(l,k)\in\Omega_{L,K}(\gamma)$ **do**
6: define sets $\Pi_{l,k,L,K}$ and $\widetilde{\Pi}_{l,k,L,K}$
7: calculate ${}_{(b)}m^{\boldsymbol{U}}_{L,K}(l,k)$ and ${}_{(b)}\tilde{m}^{\boldsymbol{U}}_{L,K}(l,k)$
8: **for** $(p,q)\in\Pi_{l,k,L,K}$ **do**
9: calculate ${}_{(b)}S^{\boldsymbol{U}}_{L,K}(p,q,l,k)$
10: **end for**
11: compute $\max_{(p,q)\in\Pi_{l,k,L,K}}\left|{}_{(b)}S^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|$
12: **for** $(p,q)\in\widetilde{\Pi}_{l,k,L,K}$ **do**
13: calculate ${}_{(b)}\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)$
14: **end for**
15: compute $\max_{(p,q)\in\widetilde{\Pi}_{l,k,L,K}}\left|{}_{(b)}\widetilde{S}^{\boldsymbol{U}}_{L,K}(p,q,l,k)\right|$
16: **end for**
17: compute bootstrap test statistics ${}_{(b)}\mathcal{R}^*_{L,K}(\psi,\gamma)$
18: **end for**

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 $\mathcal{R}^*_{L,K}(\psi,\gamma)$ provides asymptotically correct critical values for the test based on $\mathcal{R}_n(\psi,\gamma)$, when the observations follow either the null hypothesis or the alternative.

Theorem 2.1.3 (Bootstrap consistency)**.** *Suppose* $Y_1,\dots,Y_n$ *follow model* (2.2) *and* $E|\psi(\varepsilon_1)|^\nu<\infty$ *for some* $\nu>2$. *Let Assumptions* $\mathcal{A}$.2.1 *and* $\mathcal{A}$.2.3 *be satisfied for* $\chi_1,\chi_1'>0$ *and for* $\chi_2,\chi_2'>0$ *such that* $\kappa<\chi_1<\nu-2$, $\chi_1'=\nu-2-\chi_1$, *and* $\chi_2>2+2\kappa$ *for some* $0<\kappa<\nu-2$. *Moreover, let Assumptions* $\mathcal{A}$.2.2, $\mathcal{A}$.2.4, *and* $\mathcal{A}$.2.5 *be satisfied. Under alternative* (2.4), *let* $\delta_n=O(n^\theta)$ *as* $n\to\infty$ *for* $\theta\in\left(-\frac{1}{2},\frac{\eta}{3(2+\chi+\chi')}-\frac{1}{2}\right)$ *and*

$\tau = [n\zeta]$ *for some* $\zeta : \gamma < \zeta < 1-\gamma$. *If* $K = O(L)$ *as* $L \to \infty$, *then we have for all* $y \in \mathbb{R}$, *as* $L \to \infty$,

$$P\left[\mathcal{R}^*_{L,K}(\psi,\gamma) \leqslant y | Y_1, \ldots, Y_n\right] \xrightarrow{P} P\left[\sup_{\gamma \leqslant t \leqslant 1-\gamma} \frac{\sup_{0 \leqslant u \leqslant t} \left|\mathcal{W}(u) - u/t\mathcal{W}(t)\right|}{\sup_{t \leqslant u \leqslant 1} \left|\widetilde{\mathcal{W}}(u) - (1-u)/(1-t)\widetilde{\mathcal{W}}(t)\right|} \leqslant y\right],$$

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

The conditional weak convergence in probability from the above stated Theorem 2.1.3 can be replaced by the conditional weak convergence almost surely using Theorem 3.5.1, Remark 3.5.4 from Kirch (2006), and strengthening some assumptions.

A choice of the block length K 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 of the circular moving block bootstrap variance estimate. Fitzenberger (1997) proved that this approach yields $K = O(n^{1/3})$ as $n \to \infty$ in case of the α-mixing random errors. In contrast to this theoretical 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).

2.2 Structural breaks and self-normalized statistics

Many changepoint detection procedures rely on the estimation of nuisance parameters (like long-run variance). If a change has occurred, estimators might be biased and data-adaptive rules for the choice of tuning parameters might not work as expected. If the data is not stationary, this becomes more challenging.

Under the assumption of finite expectations, changes in the location are typically detected by comparing sample means and the asymptotic distribution can be derived from an invariance principle for the partial sum process. However, one has to estimate the *long-run variance* to utilize the traditional CUSUM-statistic and

this involves some difficulties. For time series, the long-run variance includes the covariances, which have to be estimated and combined with, e.g., kernels. Under the alternative, the estimation of the covariances is biased, so the long-run covariance is typically overestimated, which results in a loss of power, see Hušková and Kirch (2010). In many applications, the observations do not seem to be stationary even under the hypothesis of no change in mean, because the amount of fluctuation is not constant. To estimate the time-varying long-run variance is even more difficult. In a recent article, Górecki et al. (2018) have followed this approach.

The main aim of this section is to develop tests for the hypothesis of a constant expectation against the alternative of at most one changepoint that avoid the problems of long-run variance estimation and heteroscedasticity. Our new test statistics will involve *neither nuisance parameters nor tuning constants* and will work for *heteroscedastic and dependent* time series under some mild mixing conditions. Additionally, we will give a consistent estimator for the time of the change.

Some authors proposed to use nonparametric resampling methods like bootstrap (e.g., Hušková and Kirch (2012) and Peštová and Pešta (2018a)) or subsampling (e.g., Betken and Wendler (2018)) to avoid the estimation of the long-run variance. However, these methods still involve the choice of tuning parameters like bandwidths or block sizes and only work for stationary time series. Other approaches are *ratio statistics* and *self-normalized statistics*, which do not rely on tuning parameters. Ratio tests have been introduced to detect changes in persistence by Kim (2000) and since have been studied for changes in mean (Horváth et al., 2008), for changes in variance (Zhao et al., 2011), for heavy-tailed sequences (Wang et al., 2016), for panel date framework (Peštová and Pešta, 2015), and for robust M-estimators (Peštová and Pešta, 2018a).

Self-normalized test statistics for changepoints were firstly proposed by Shao and Zhang (2010) and were generalized to long range dependent time series (Shao, 2011). Betken (2016) developed a robust self-normalized test based on the Wilcoxon-statistic, Zhang and Lavitas (2018) proposed a self-normalized test for multiple changepoints. Our approach is to combine new variants of the self-normalized test statistics with the *wild bootstrap*. The wild bootstrap was proposed by Wu (1986) and is consistent under heteroscedasticity. However, it does not reproduce the dependence of the data. We will show that under our model assumptions, it still gives the correct critical values for the self-normalized test statistics. In this way, we can avoid using the dependent wild bootstrap of Shao (2010), which involves the choice of a kernel and of a bandwidth parameter. One of our main contributions is to provide a *fully automatic and completely data driven changepoint detection procedure*. Let us note that in the context of time series regression with heteroscedastic and dependent errors, Rho and Shao (2015) have already observed that a self-normalized

test statistic is not pivotal. They also used the wild bootstrap for obtaining critical values.

From now on, we are going to present results published in Pešta and Wendler (2019).

2.2.1 Stochastic changepoint model and self-normalization

We tend to study time series with one abrupt change in the mean at an unknown point in time. Let us consider observations $Y_{1,n}, \dots, Y_{n,n}$ obtained at n time ordered points. We are interested in testing the null hypothesis of all observations being random variables having equal expectation. Our goal is to test against the alternative of the first τ_n observations have expectation μ and the remaining $n - \tau_n$ observations come from distributions with expectation $\mu + \delta_n$, where $\delta_n \neq 0$. More precisely, our model is

$$Y_{n,k} = \mu + \delta_n \, \mathbb{1}\{k > \tau_n\} + \sigma\left(k/n\right) \varepsilon_k, \quad k = 1, \dots, n, \tag{2.9}$$

where μ, δ_n, and τ_n are unknown parameters, $\{Y_{n,k}\}_{n=1,k=1}^{\infty,n}$ is a *triangular array* of random variables, $\{\varepsilon_n\}_{n=1}^{\infty}$ is a sequence of stationary centered disturbances, $\sigma(t)$ is a non-stochastic variance function, and $\mathbb{1}\{A\}$ denotes the indicator of set A. The time point τ_n is called the *changepoint*. This model was assumed by Górecki et al. (2018).

We are going to test the null hypothesis that no change occurred against the alternative that a change occurred at some unknown time point τ_n, i.e.,

$$\mathcal{H}_0 : \tau_n = n \quad \text{versus} \quad \mathcal{H}_1 : \tau_n < n, \ \delta_n \neq 0.$$

Test statistics

The *CUSUM-statistic* is frequently used to detect changes in the mean and it is based on the partial sums $\sum_{i=1}^{k} \left(Y_{n,i} - \bar{Y}_{n,1:n}\right)$, $k < n$ of the centered observations, where $\bar{Y}_{n,i:j} = \frac{1}{j-i+1} \sum_{k=i}^{j} Y_{n,k}$, $i \leqslant j$. To combine the values of the partial sums for different k's into a single test statistic, one can use the *supremum-type* CUSUM-statistic $\max_{k=1,\dots,n-1} \left|\sum_{i=1}^{k} \left(Y_{n,i} - \bar{Y}_{n,1:n}\right)\right|$ or, alternatively, the *integral-type* CUSUM-statistic $\sum_{k=1}^{n-1} \left(\sum_{i=1}^{k} \left(Y_{n,i} - \bar{Y}_{n,1:n}\right)\right)^2$. These test statistics need to be standardized by a variance of the series. However, it is practically *difficult to find a variance estimator* with satisfactory properties. Such difficulty can occur in situations with dependent or heteroscedastic random errors. Nonetheless, the vari-

ance estimators often do not perform well even in the i.i.d. case, especially under alternatives (Antoch et al., 1997).

To avoid the estimation of variance parameters, different ratios of such test statistics have been proposed. Horváth et al. (2008) divide the supremum-test statistic of the first part of the series by the supremum-test statistic of the second part of the data. Wenhua and Hao (2016) use a ratio test based on the integral-type statistic. Shao and Zhang (2010) introduced a self-normalized statistic, which uses the supremum-type CUSUM-statistic of the whole data set in the numerator, divided by the sum of two integral-type statistics of the data before k and after k.

Our idea is to use a self-normalization of the CUSUM-statistic by the same type: We divide the supremum-type statistic by two supremum-type statistics and the integral-type statistic by two integral-type statistics. Our test statistics can be expressed as functionals of the *cumulative sums* $V_n(k) := \sum_{i=1}^{k} Y_{n,i}$ and $\widetilde{V}_n(k) := V_n(n) - V_n(k)$. We define the *self-normalized test statistics* as

$$\begin{aligned}\mathscr{Q}(V_n) &:= \max_{1\leqslant k\leqslant n} \left| \frac{V_n(k) - \frac{k}{n}V_n(n)}{\max\limits_{1\leqslant i\leqslant k} \left|V_n(i) - \frac{i}{k}V_n(k)\right| + \max\limits_{k<i\leqslant n} \left|\widetilde{V}_n(i) - \frac{n-i}{n-k}\widetilde{V}_n(k)\right|} \right| \qquad (2.10)\\ &\equiv \max_{1\leqslant k\leqslant n} \frac{\left| \sum\limits_{i=1}^{k} \left(Y_{n,i} - \bar{Y}_{n,1:n}\right) \right|}{\max\limits_{1\leqslant i\leqslant k} \left| \sum\limits_{j=1}^{i} \left(Y_{n,j} - \bar{Y}_{n,1:k}\right) \right| + \max\limits_{k<i\leqslant n} \left| \sum\limits_{j=i}^{n} \left(Y_{n,j} - \bar{Y}_{n,(k+1):n}\right) \right|}\end{aligned}$$

and

$$\begin{aligned}\mathscr{R}(V_n) &:= \sum_{k=1}^{n} \frac{\left\{V_n(k) - \frac{k}{n}V_n(n)\right\}^2}{\sum\limits_{i=1}^{k} \left\{V_n(i) - \frac{i}{k}V_n(k)\right\}^2 + \sum\limits_{i=k+1}^{n} \left\{\widetilde{V}_n(i) - \frac{n-i}{n-k}\widetilde{V}_n(k)\right\}^2} \qquad (2.11)\\ &\equiv \sum_{k=1}^{n} \frac{\left\{ \sum\limits_{i=1}^{k} \left(Y_{n,i} - \bar{Y}_{n,1:n}\right) \right\}^2}{\sum\limits_{i=1}^{k} \left\{ \sum\limits_{j=1}^{i} \left(Y_{n,j} - \bar{Y}_{n,1:k}\right) \right\}^2 + \sum\limits_{i=k+1}^{n} \left\{ \sum\limits_{j=i}^{n} \left(Y_{n,j} - \bar{Y}_{n,(k+1):n}\right) \right\}^2}.\end{aligned}$$

For many changepoint tests, one has to skip, for instance, the first and the last 10% of observations as possible candidates for a changepoint, see, e.g., Shao and Zhang (2010). Moreover, the amount of trimming can be viewed as an additional tuning parameter. For our test statistics, we are able to consider all time points $k = 1, \ldots, n$. The limit distribution of our statistics is obtained with the help of the continuous mapping theorem, using limit theorems for the partial sum process

under weak dependence and heteroscedasticity by Cavaliere (2005) or Górecki et al. (2018).

2.2.2 Asymptotic results

We proceed to the assumptions that are needed for deriving asymptotic results for the proposed test statistics.

Assumptions

Assumption $\mathcal{A}$.2.8. $\{\varepsilon_n\}_{n=1}^{\infty}$ form a zero-mean sequence such that $\operatorname{Var}\varepsilon_n = 1$ and $\{\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}\varepsilon_i, 0 \leqslant t \leqslant 1\}$ converges weakly in $D[0,1]$ to $\sqrt{\lambda}W$, where $\lambda > 0$ and $\{W(t), 0 \leqslant t \leqslant 1\}$ is a standard Wiener process.

To control variability, we have an assumption regarding heteroscedasticity.

Assumption $\mathcal{A}$.2.9. $\sigma : [0,1] \to \mathbb{R}^+$ has at most a finite number of points of discontinuity. Restricted to any open interval without discontinuities, σ is Lipschitz-continuous.

For bootstrapping, we need an assumption controlling the dependence of the underlying errors.

Assumption $\mathcal{A}$.2.10. $\{\varepsilon_n\}_{n=1}^{\infty}$ form a zero-mean strictly stationary α-mixing sequence such that $\operatorname{Var}\varepsilon_n = 1$, $\mathsf{E}\,|\varepsilon_n|^p < \infty$ for some $p > 2$ with mixing coefficients $\alpha(\varepsilon_\circ, n)$ satisfying $\sum_{n=1}^{\infty}\{\alpha(\varepsilon_\circ, n)\}^{2(1/r-1/p)} < \infty$ for some $r \in (2, p]$ and, additionally, $\sum_{n=1}^{\infty} n\alpha(\varepsilon_\circ, n) < \infty$. Furthermore, for the long-run variance, it holds that $0 < \lambda := 1 + 2\sum_{n=1}^{\infty}\mathsf{E}\,\varepsilon_1\varepsilon_{n+1} < \infty$.

As usual, if the time series $\{\varepsilon_n\}_{n\in\mathbb{N}}$ has more moments, the mixing assumption becomes less restrictive, because the exponent $2(1/r - 1/p)$ is higher for higher values of p. Moreover, one can replace the convergences of sums containing $\alpha(\varepsilon_\circ, n)$ in Assumption $\mathcal{A}$.2.10 by assuming $\alpha(\varepsilon_\circ, n) = O(n^{-p/(p-2)-\gamma})$ for some $\gamma > \max\{0, (p-4)/(p-2)\}$, which is more restrictive on one hand, but less complicated on the other hand. The strong mixing properties will be inherited by $Y_{n,k} = \mu + \delta_n\,\mathbb{1}\{k > \tau_n\} + \sigma(k/n)\varepsilon_k$, but this process can additionally model heteroscedasticity, which is important for many applications.

Asymptotic distribution of the test statistics

The following functionals of the partial sum process $U_n(t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \sigma(i/n)\varepsilon_i$ can be regarded as continuous modifications of our test statistic: Let

$$\mathscr{S}(U_n) := \sup_{0\leqslant t\leqslant 1} \left| \frac{U_n(t) - tU_n(1)}{\sup_{0\leqslant s\leqslant t} \left|U_n(s) - s/tU_n(t)\right| + \sup_{t\leqslant u\leqslant 1} \left|\widetilde{U}_n(u) - (1-u)/(1-t)\widetilde{U}_n(t)\right|} \right|, \tag{2.12}$$

where $\widetilde{U}_n(t) := U_n(1) - U_n(t)$. Moreover,

$$\mathscr{T}(U_n) := \int_0^1 \frac{\left\{U_n(t) - tU_n(1)\right\}^2}{\int_0^t \left\{U_n(s) - s/tU_n(t)\right\}^2 ds + \int_t^1 \left\{\widetilde{U}_n(u) - (1-u)/(1-t)\widetilde{U}_n(t)\right\}^2 du} dt. \tag{2.13}$$

Under the null hypothesis and the technical assumptions from the previous subsection, the test statistics defined in (2.10) and (2.11) converge to *non-degenerate limit distributions*.

Theorem 2.2.1 (Under the null)**.** *Under Assumptions* $\mathcal{A}$.2.8, $\mathcal{A}$.2.9, *and under* $\mathcal{H}_0$,

$$\mathscr{Q}(V_n) \xrightarrow{D} \mathscr{S}(W_\eta) \quad \textit{and} \quad \mathscr{R}(V_n) \xrightarrow{D} \mathscr{T}(W_\eta), \quad n \to \infty, \tag{2.14}$$

where $W_\eta(t) := W(\eta(t))$, $\eta(t) := \frac{\int_0^t \sigma^2(s)ds}{\int_0^1 \sigma^2(s)ds}$, $\{W(t), 0 \leqslant t \leqslant 1\}$ *is a standard Wiener process, and the functionals* $\mathscr{S}$ *and* $\mathscr{T}$ *are defined in* (2.12) *and* (2.13).

One may replace very general and theoretical Assumption $\mathcal{A}$.2.8 by more practical Assumption $\mathcal{A}$.2.10.

Remark 2.2.1. Under Assumptions $\mathcal{A}$.2.10, $\mathcal{A}$.2.9, and under $\mathcal{H}_0$, relation (2.14) holds.

The null hypothesis is rejected at significance level α for large values of $\mathscr{Q}(V_n)$ and $\mathscr{R}(V_n)$. The critical values can be obtained as the $(1-\alpha)$-quantiles of the asymptotic distributions from (2.14), if η is known. Furthermore, the tests based on these two statistics are *consistent* (for fixed as well as for local alternatives), as the test statistics converge to infinity under the alternative, provided that the size of the change does not convergence to 0 to fast.

Theorem 2.2.2 (Under the alternative). *Suppose Assumptions A.2.8 and A.2.9 hold. Under $\mathcal{H}_1$ such that $|\delta_n|\sqrt{n} \to \infty$ as $n \to \infty$ and $\tau_n = [n\zeta]$ for some $\zeta \in (0,1)$, $\mathcal{Q}(V_n) \xrightarrow{P} \infty$ and $\mathcal{S}(V_n) \xrightarrow{P} \infty$ as $n \to \infty$.*

Theorem 2.2.2 says that in presence of the structural change in mean, the test statistics *explode above all bounds*. Hence, the asymptotic distributions from Theorem 2.2.1 can be used to construct the tests. Although, explicit forms of those distributions are unknown. Therefore in order to obtain the critical values, we have to use either simulations from the limit distributions or resampling methods. For the simulation purposes, one would need to know or to estimate the nuisance function $\eta(t)$. The resampling techniques will help us to avoid and overcome such an issue.

2.2.3 Wild bootstrap

Wild bootstrap *replications* are defined as

$$Y^{\star}_{n,k} := \left(Y_{n,k} - \bar{Y}_{n,1:n}\right) X_k, \quad k = 1, \ldots, n,$$

where $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables having *standard normal* $N(0,1)$ distribution. Moreover, $\{Y_{n,k}\}_{n=1,k=1}^{\infty,n}$ and $\{X_n\}_{n=1}^{\infty}$ are also *independent*. The schematic algorithm of the wild bootstrap can be seen as Procedure 2.2.1. In general, the wild bootstrap replications should be defined in the following way $Y^{*}_{n,k} := \bar{Y}_{n,1:n} + \left(Y_{n,k} - \bar{Y}_{n,1:n}\right) X_k$, $k = 1, \ldots, n$. However, in our case how the test statistics are defined, there would be no impact by adding $\bar{Y}_{n,1:n}$. We define $V^{\star}_n(k) := \sum_{i=1}^{k} Y^{\star}_{n,i}$ and $\widetilde{V}^{\star}_n(k) := V^{\star}_n(n) - V^{\star}_n(k)$.

Procedure 2.2.1 Wild bootstrap of the test statistic $\mathcal{Q}(V_n)$ and $\mathcal{R}(V_n)$

Input: Sequence of observations $Y_{1,n}, \ldots, Y_{n,n}$ and number of bootstrap replications B

Output: Bootstrap distributions of $\mathcal{Q}(V_n)$ and $\mathcal{R}(V_n)$, respectively; i.e., the empirical distributions where probability mass $1/B$ concentrates at each of ${}_{(1)}\mathcal{Q}(V^{\star}_n), \ldots, {}_{(B)}\mathcal{Q}(V^{\star}_n)$ and ${}_{(1)}\mathcal{R}(V^{\star}_n), \ldots, {}_{(B)}\mathcal{R}(V^{\star}_n)$, respectively

1: **for** $b = 1$ to B **do** // repeat in order to obtain the empirical distributions
2: generate random sample $[{}_{(b)}X_1, \ldots, {}_{(b)}X_n]$ from $N(0,1)$ independently for different b
3: calculate ${}_{(b)}Y^{\star}_{n,k} = \left(Y_{n,k} - \bar{Y}_{n,1:n}\right) \times {}_{(b)}X_k$ for all k
4: calculate ${}_{(b)}V^{\star}_n(k) = \sum_{i=1}^{k} {}_{(b)}Y^{\star}_{n,i}$ and ${}_{(b)}\widetilde{V}^{\star}_n(k) = {}_{(b)}V^{\star}_n(n) - {}_{(b)}V^{\star}_n(k)$ for all k
5: compute the bootstrap test statistics ${}_{(b)}\mathcal{Q}(V^{\star}_n)$ and ${}_{(b)}\mathcal{R}(V^{\star}_n)$
6: **end for**

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. It is not known

and it does not matter whether our observations come form the null hypothesis or the alternative. We are going to prove that $\mathscr{Q}(V_n^\star)$ and $\mathscr{R}(V_n^\star)$, respectively, provide asymptotically correct critical values for the test based on $\mathscr{Q}(V_n)$ and $\mathscr{R}(V_n)$, respectively.

Theorem 2.2.3 (Wild bootstrap validity)**.** *Suppose that Assumptions $\mathcal{A}$.2.10 and $\mathcal{A}$.2.9 hold. Under the null hypothesis $\mathcal{H}_0$ or under local alternatives $\mathcal{H}_1$ with $\delta_n \to 0$ as $n \to \infty$,*

$$\mathscr{Q}(V_n^\star) \xrightarrow{D} \mathscr{S}(W_\eta) \quad \textit{and} \quad \mathscr{R}(V_n^\star) \xrightarrow{D} \mathscr{T}(W_\eta), \quad n \to \infty$$

almost surely conditionally on $\{Y_{n,k}\}_{n=1,k=1}^{\infty,n}$. Under the alternative hypothesis $\mathcal{H}_1$ with $\tau_n = [n\zeta]$ for some $\zeta \in (0,1)$ and having $\delta_n \equiv \delta \neq 0$ fixed, let $\{B(t), 0 \leqslant t \leqslant 1\}$ be a standard Wiener processes independent of W. Then,

$$\mathscr{Q}(V_n^\star) \xrightarrow{D} \mathscr{S}\left(W_\eta - \frac{\delta}{\vartheta} B_\zeta\right) \quad \textit{and} \quad \mathscr{R}(V_n^\star) \xrightarrow{D} \mathscr{T}\left(W_\eta - \frac{\delta}{\vartheta} B_\zeta\right), \quad n \to \infty$$

almost surely conditionally on $\{Y_{n,k}\}_{n=1,k=1}^{\infty,n}$ with $\vartheta^2 = \int_0^1 \sigma^2(t)dt$ and

$$B_\zeta(t) = \begin{cases} (1-\zeta)B(t), & t \leqslant \zeta; \\ B(\zeta) - \zeta B(t), & t > \zeta. \end{cases}$$

Theorem 2.2.3 assures that the asymptotic distribution of the bootstrap test statistics and the limit distribution of the original test statistics *coincide* under the null hypothesis. Thus, the bootstrap tests approximately keep the same level as the original tests based on the asymptotics from Theorem 2.2.1 even without knowing or estimating the nuisance function $\eta(t)$. Moreover, the limit distribution of the bootstrap test is not changed under local alternatives, so we avoid the power loss that would be caused by overestimation of the long-run variance under the alternative.

Even under fixed alternatives, the distribution of the bootstrap statistics converge to an almost sure finite limit. In contrast, an uncorrected kernel estimator for the long-run variance would converge to infinity in this case. Depending on the function η and the time of the change represented by ζ, the quantiles $\mathscr{S}(W_\eta - \frac{\delta}{\vartheta} B_\zeta)$ and $\mathscr{T}(W_\eta - \frac{\delta}{\vartheta} B_\zeta)$, respectively, might be larger or smaller than the quantiles of $\mathscr{S}(W_\eta)$ and $\mathscr{T}(W_\eta)$, respectively, resulting in a loss or gain of power compared to the use of the asymptotic quantiles from Theorem 2.2.1 (which is only feasible when η is known).

Now, the simulated (empirical) distributions of the bootstrap test statistics can be used to calculate the bootstrap critical values, which will be compared to the values of the original test statistics in order to reject the null or not.

2.2.4 Consistent changepoint estimator

If a change is *detected*, it is of interest to estimate the time of the change. It is sensible to use

$$\hat{\tau}_n := \underset{1\leqslant k\leqslant n}{\operatorname{argmax}} \frac{\left|\sum_{i=1}^{k}\left(Y_{n,i}-\bar{Y}_{n,1:n}\right)\right|+\left|\sum_{i=n-k+1}^{n}\left(Y_{n,i}-\bar{Y}_{n,1:n}\right)\right|}{\max\limits_{1\leqslant i\leqslant k}\left|\sum_{j=1}^{i}\left(Y_{n,j}-\bar{Y}_{n,1:k}\right)\right|+\max\limits_{k<i\leqslant n}\left|\sum_{j=i}^{n}\left(Y_{n,j}-\bar{Y}_{n,(k+1):n}\right)\right|}$$
$$\equiv \underset{1\leqslant k\leqslant n}{\operatorname{argmax}} \frac{\left|V_n(k)-k/nV_n(n)\right|+\left|\widetilde{V}_n(n-k)-k/nV_n(n)\right|}{\max\limits_{1\leqslant i\leqslant k}\left|V_n(i)-i/kV_n(k)\right|+\max\limits_{k<i\leqslant n}\left|\widetilde{V}_n(i)-(n-i)/(n-k)\widetilde{V}_n(k)\right|}$$

as a *changepoint estimator*. Our next theorem shows that under the alternative, the changepoint τ_n is consistently estimated by the estimator $\hat{\tau}_n$.

Theorem 2.2.4 (Estimator's consistency)**.** *Suppose Assumptions A.2.8 and A.2.9. Under $\mathcal{H}_1$ such that $|\delta_n|\sqrt{n}\to\infty$ as $n\to\infty$ and $\tau_n=[n\zeta]$ for some $\zeta\in(0,1)$, it holds $\hat{\tau}_n/n\xrightarrow[n\to\infty]{P}\zeta$.*

2.3 Change in trend for randomly spaced series

Traditional notion of time series covers regularly spaced time series, which are also called evenly or equally spaced time series. In contrast to that, there are situations, where one observes outcomes Y_i's at some unevenly located time points Z_i's. The time series $\{Z_i, Y_i\}_{i=1}^n$ is, therefore, called an *irregularly (unevenly or unequally) spaced time series*. The time moments Z_i's are usually deterministic and measured precisely, whereas the outcomes Y_i's are considered to be random due to some output fluctuations (e.g., measurement imprecision).

Change in mean detection can be basically performed in the same way for regularly and irregularly series. Therefore, we move our attention to a more general type of structural break—*change in trend*.

Now, it may happen that our outcome observations Y_i's are supposed to be measured at some *unknown time points* Z_i's. However, due to some measurement imprecision, the actual observation Y_i, which should correspond to Z_i, is indeed observed at the time point Z_i, but the time point X_i is recorded instead, where $X_i = Z_i + \theta_i$ and $\{\theta_i\}_{i=1}^n$ are some *random errors*. The unobservable sequence $\{Z_i\}_{i=1}^n$ can be regularly or irregularly spaced. A series of the couples

$$\{[X_i, Y_i]\}_{i=1}^n$$

is called a *randomly spaced time series* and, in the presented context, it can be considered as an extension of the irregularly spaced time series.

We immediately generalize the framework of the randomly spaced time series approach by leaving the purely time series content and by allowing to handle multivariate $\boldsymbol{X}_{i,\bullet}$'s. On the other hand, we loose natural ordering of the one-dimensional X_i's. Although, one still possess the basic (time) arrangement of $\boldsymbol{X}_{i,\bullet}$'s through their underlying indices i's.

Henceforth, we are going to deal with *changepoints in linear relations* encapsulating the problem of change in trend for randomly spaced time series. We rely mainly on results from Pešta (2016) and Pešta (2017) with partial contribution by Pešta (2011), Pešta (2013a), Pešta (2013b), and Peštová and Pešta (2018b).

2.3.1 Changepoints in linear relations

If measured input and output data are in some linear relations, then it is of particular interest to detect whether impact of the input characteristics has changed over time on the output observables. Despite the fact that the relations and, consequently, suitable underlying stochastic models are linearly defined, the possible estimates and the corresponding inference may be highly non-linear (Gleser, 1981). It becomes even more challenging to handle measurement errors in input and output data simultaneously, when the linear relations are subject to change at some unknown time point—*changepoint*.

There is a vast literature aimed at linear relations modeled through so-called *measurement error models* or *errors-in-variables models* (for an overview, see Fuller (1987), Van Huffel and Vandewalle (1991), Carroll et al. (2006), or Buonaccorsi (2010)), but very little has been explored in the changepoint analysis for these models yet. A change in regression has been explored thoroughly, cf. Horváth (1995) or Aue et al. (2008). However, such a framework does not cover the case of measurement error models. Maximum likelihood approach (Chang and Huang, 1997) and Bayesian approach (Carroll et al., 1999) to the changepoint estimation in the measurement error models were applied, both requiring parametric distributional assumptions on the errors. Kukush et al. (2007) estimated the changepoint in the input data only. A change in the variance parameter of the normally distributed errors within the measurement error models was investigated by Dong et al. (2016). All of these mentioned contributions dealt with the changepoint estimation solely. Our *main goal* is to test for a possible change in the parameters relating the input and output data, both encumbered by some errors. Consequently, if a change is detected, we aim to estimate it. By our best knowledge, we are not aware of any similar results

even for the independent and identically distributed errors. Additionally to that, our changepoint tests are supposed to be *nuisance-parameter-free*, *distributional-free*, and to allow for a very general error structures.

2.3.2 Errors-in-variables model with changepoint

Errors-in-variables (EIV) or also called *measurement error* model

$$\boldsymbol{X} = \boldsymbol{Z} + \boldsymbol{\Theta} \tag{$\mathcal{M}$}$$

and

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{$\mathcal{H}_0$}$$

is considered, where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown *regression parameters* possibly subject to change, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{Y} \in \mathbb{R}^{n\times 1}$ consist of *observable random* variables ($\boldsymbol{X}$ are covariates and $\boldsymbol{Y}$ is a response), $\boldsymbol{Z} \in \mathbb{R}^{n\times p}$ consists of *unknown constants* and has full rank, $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{\Theta} \in \mathbb{R}^{n\times p}$ are *random errors*. This setup can be extended to a *multivariate case*, where $\boldsymbol{\beta} \in \mathbb{R}^{p\times q}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times q}$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times q}$, $q \geqslant 1$, see Subsection 2.3.3.

The EIV model ($\mathcal{M}$)–($\mathcal{H}_0$) with non-random unknown constants $\boldsymbol{Z}$ is sometimes called *functional EIV* model (Booth and Hall, 1993). On the other hand, a different approach may handle $\boldsymbol{Z}$ as random covariates, which is called *structural EIV* model (Chang and Huang, 1997). Here, we will concentrate on the first mentioned one, i.e., the functional EIV model.

To estimate the unknown parameter $\boldsymbol{\beta}$, one usually minimizes the *Frobenius matrix norm* of the errors $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$, see Golub and Van Loan (1980). This approach leads to a *total least squares* (TLS) estimate $\boldsymbol{b} = (\boldsymbol{X}^\top\boldsymbol{X} - \lambda_{min}([\boldsymbol{X},\boldsymbol{Y}]^\top[\boldsymbol{X},\boldsymbol{Y}])\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$, where $\lambda_{min}(\boldsymbol{M})$ is the smallest eigenvalue of the matrix $\boldsymbol{M}$ and $\boldsymbol{I}_p$ is a $(p \times p)$ identity matrix. Geometrically speaking, the Frobenius norm tries to minimize the *orthogonal distance* between the observations and the fitted hyperplane. Therefore, the TLS are usually known as *orthogonal regression*. One can generalize this method by replacing the Frobenius norm by *any unitary invariance matrix norm*, which surprisingly yields the same TLS estimate, having interesting invariance and equivariance properties (Pešta, 2016). The TLS estimate is shown to be strongly and weakly consistent (Gleser, 1981; Gallo, 1982a; Pešta, 2011) as well as to be asymptotically normal (Gallo, 1982b; Pešta, 2013b; Pešta, 2017) under various conditions.

We aim to detect a possible change in the linear relation parameter $\boldsymbol{\beta}$. The interest lies in testing the *null hypothesis* ($\mathcal{H}_0$) of all observations Y_i's being random variables having expectations $\boldsymbol{Z}_{\bullet,i}\boldsymbol{\beta}$'s. Our goal is to test against the alternative of the first τ

observations have expectations $\boldsymbol{Z}_{\bullet,i}\beta$'s and the remaining $n-\tau$ observations come from distributions with expectations $\boldsymbol{Z}_{\bullet,i}(\beta+\boldsymbol{\delta})$'s, where $\boldsymbol{\delta}\neq\mathbf{0}$. A 'row-column' notation for a matrix $\boldsymbol{M}$ is used in this manner: $\boldsymbol{M}_{i,\bullet}$ denotes the ith row of $\boldsymbol{M}$ and $\boldsymbol{M}_{\bullet,j}$ corresponds to the jth column of $\boldsymbol{M}$. Furthermore, $\boldsymbol{M}_i$ stays for the first i rows of $\boldsymbol{M}$ and $\boldsymbol{M}_{-i}$ represents the remaining $n-i$ rows of $\boldsymbol{M}$, when the first i rows are deleted. Now more precisely, our *alternative hypothesis* is

$$\boldsymbol{Y}_\tau=\boldsymbol{Z}_\tau\beta+\boldsymbol{\varepsilon}_\tau\quad\text{and}\quad\boldsymbol{Y}_{-\tau}=\boldsymbol{Z}_{-\tau}(\beta+\boldsymbol{\delta})+\boldsymbol{\varepsilon}_{-\tau}. \tag{$\mathcal{H}_A$}$$

Here, $\boldsymbol{\delta}\equiv\boldsymbol{\delta}(n)\neq\mathbf{0}$ is an unknown vector parameter representing the size of change and is possibly depending on n. The *changepoint* $\tau\equiv\tau(n)<n$ is also an unknown scalar parameter, which depends on n as well. Although, β is considered to be independent of n.

Intercept and fixed regressors

Note that the EIV model $(\mathcal{M})$–$(\mathcal{H}_0)$ has no intercept and all the covariates are encumbered by some errors. To overcome such a restriction, one can think of an extended regression model, where some explanatory variables are *subject to error* and some are measured *precisely*. I.e., $\boldsymbol{Y}=\boldsymbol{W}\boldsymbol{\alpha}+\boldsymbol{Z}\beta+\boldsymbol{\varepsilon}$, where $\boldsymbol{W}$ are *observable true* and $\boldsymbol{Z}$ are *unobservable true* constants, both having full rank. Regression parameters $\boldsymbol{\alpha}$ and β remain unknown. Then, the non-random (fixed) *intercept* can be incorporated into the regression model by setting one column of the matrix $\boldsymbol{W}$ equal to $[1,\ldots,1]^\top$. Consequently, we may *project out* exact observations using projection matrix $\boldsymbol{R}:=\boldsymbol{I}_n-\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top$. Notice that $\boldsymbol{R}$ is symmetric and idempotent. Finally, one may work with $\boldsymbol{R}\boldsymbol{Y}=\boldsymbol{R}\boldsymbol{Z}\beta+\boldsymbol{R}\boldsymbol{\varepsilon}$ instead of $(\mathcal{H}_0)$.

2.3.3 Spectral weak invariance principle

A theoretical device is going to be developed in order to construct the changepoint tests. The *smallest eigenvalue* of $[\boldsymbol{X},\boldsymbol{Y}]^\top[\boldsymbol{X},\boldsymbol{Y}]$—the squared smallest singular value of the data matrix $[\boldsymbol{X},\boldsymbol{Y}]$—plays a key role. We proceed to the assumptions that are needed for deriving forthcoming asymptotic results.

Assumptions

Firstly, a *design assumption* on the unobservable regressors is needed.

Assumption $\mathcal{A}$.2.11. For every $\zeta \in (0,1)$, there exist positive definite

$$\boldsymbol{\Delta}_{\zeta} := \lim_{n\to\infty} n^{-1} \boldsymbol{Z}_{\tau}^{\top} \boldsymbol{Z}_{\tau} \quad \text{and} \quad \boldsymbol{\Delta}_{-\zeta} := \lim_{n\to\infty} n^{-1} \boldsymbol{Z}_{-\tau}^{\top} \boldsymbol{Z}_{-\tau},$$

where $\tau = [n\zeta]$. Moreover, $\boldsymbol{\Delta} := \lim_{n\to\infty} n^{-1} \boldsymbol{Z}^{\top} \boldsymbol{Z}$ is positive definite.

For example in one-dimensional case (i.e., $p = 1$), a simple design, where $Z_{i,1} = i/(n+1)$, provides $\boldsymbol{\Delta}_{\zeta} = \zeta^3/3$ and $\boldsymbol{\Delta} = 1/3$.

Assumption $\mathcal{A}$.2.12. $\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}$, and $\{\varepsilon_n\}_{n=1}^{\infty}$ are pairwise independent sequences of α-mixing absolutely continuous random variables having zero mean and variance equal $\sigma^2 > 0$ such that

$$\alpha(\Theta_{\circ,j}, n) = \mathcal{O}(n^{-1-\varpi_j}),\ j = 1, \ldots, p \quad \text{and} \quad \alpha(\varepsilon_{\circ}, n) = \mathcal{O}(n^{-1-\varpi_{p+1}}),$$

as $n \to \infty$ for some $\varpi_j > 0$, $j \in \{1, \ldots, p+1\}$. Moreover,

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty,\ j \in \{1, \ldots, p\},$$

$$\sup_{n\in\mathbb{N}} \mathsf{E}\,|\Theta_{n,j}|^{4+\omega_j} < \infty,\ j \in \{1, \ldots, p\}, \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mathsf{E}\,|\varepsilon_n|^{4+\omega_{p+1}} < \infty$$

for some $\omega_j > 0$, $j \in \{1, \ldots, p+1\}$ such that $\min_{j=1,\ldots,p+1} \varpi_j \min_{j=1,\ldots,p+1} \omega_j > 2$.

Let us emphasize that the sequences of the errors *do not have to be stationary*. A homoscedastic covariance structure of the within-individual errors $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]$ can be generalized by knowing the *heteroscedastic covariance matrix* $\boldsymbol{\Gamma} > \mathbf{0}$ in advance. Mathematically speaking, the homoscedastic covariance matrix $\sigma^2 \boldsymbol{I}_{p+1}$ can be replaced by a general one $\boldsymbol{\Gamma} \in \mathbb{R}^{(p+1)\times(p+1)}$. Then, the observation data are just multiplied by its square root as already discussed in Van Huffel and Vandewalle ((1991):Section 8.4) or Gleser ((1981):Section 5), i.e., the new transformed data are $[\boldsymbol{X}, \boldsymbol{Y}]\boldsymbol{\Gamma}^{-1/2}$. This transformation of the original data is purely linear, which is not restrictive at all in our situation. The whole asymptotic inference remain also valid even for the heteroscedastic case. The only property that needs to be satisfied is pairwise independence of the linearly transformed errors $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]\boldsymbol{\Gamma}^{-1/2}$. If the covariance matrix $\boldsymbol{\Gamma}$ is unknown, it can be estimated when possessing *repeated observations*, cf. Pešta (2013b).

Furthermore, a *variance assumption* for the misfit disturbances is stated. It can be considered as an assumption for the *long-run variance* of residuals.

Assumption $\mathcal{A}$.2.13. There exists

$$\upsilon := \lim_{n\to\infty} n^{-1} \mathsf{Var}\left\{\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2\right\} > 0.$$

Finally, the spectral weak invariance principle for the smallest eigenvalues is provided. Let us denote $\lambda_i := \lambda_{\min}([\boldsymbol{X}_i, \boldsymbol{Y}_i]^\top [\boldsymbol{X}_i, \boldsymbol{Y}_i])$ for $2 \leqslant i \leqslant n$, $\lambda_0 := \lambda_1 := 1$ and $\widetilde{\lambda}_i := \lambda_{\min}([\boldsymbol{X}_{-(i-1)}, \boldsymbol{Y}_{-(i-1)}]^\top [\boldsymbol{X}_{-(i-1)}, \boldsymbol{Y}_{-(i-1)}])$ for $1 \leqslant i \leqslant n-1$, $\widetilde{\lambda}_n := 0$, $\widetilde{\lambda}_0 := \widetilde{\lambda}_1$.

Proposition 2.3.1 (SWIP). *Let $\mathcal{M}$ and $\mathcal{H}_0$ hold. If Assumptions $\mathcal{A}$.2.11, $\mathcal{A}$.2.12, and $\mathcal{A}$.2.13 are satisfied, then*

$$\left\{ \frac{1}{\sqrt{n}} \left(\lambda_{[nt]} - [nt]\sigma^2 \right) \right\}_{t \in [0,1]} \xrightarrow[n \to \infty]{D[0,1]} \left\{ \frac{\upsilon}{1 + \|\boldsymbol{\beta}\|_2^2} \mathcal{W}(t) \right\}_{t \in [0,1]}$$

and

$$\left\{ \frac{1}{\sqrt{n}} \left(\widetilde{\lambda}_{[n(1-t)]} - [n(1-t)]\sigma^2 \right) \right\}_{t \in [0,1]} \xrightarrow[n \to \infty]{D[0,1]} \left\{ \frac{\upsilon}{1 + \|\boldsymbol{\beta}\|_2^2} \widetilde{\mathcal{W}}(t) \right\}_{t \in [0,1]},$$

where $\{\mathcal{W}(t)\}_{t \in [0,1]}$ is a standard Wiener process and $\widetilde{\mathcal{W}}(t) = \mathcal{W}(1) - \mathcal{W}(t)$.

Extension to multivariate case

Suppose that $\boldsymbol{\beta} \in \mathbb{R}^{p \times q}$, $\boldsymbol{Y} \in \mathbb{R}^{n \times q}$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{n \times q}$, $q \geqslant 1$. Let the singular value decomposition (SVD) of the partial data be

$$[\boldsymbol{X}_{[nt]}, \boldsymbol{Y}_{[nt]}] = \boldsymbol{U}_{[nt]} \boldsymbol{\Sigma}_{[nt]} \boldsymbol{V}_{[nt]}^\top = \sum_{i=1}^{p+q} \vartheta_{[nt]}^{(i)} \boldsymbol{u}_{[nt]}^{(i)} \boldsymbol{v}_{[nt]}^{(i)\top},$$

where $\boldsymbol{u}_{[nt]}^{(i)}$'s are the left-singular vectors, $\boldsymbol{v}_{[nt]}^{(i)}$'s are the right-singular vectors, and $\vartheta_{[nt]}^{(i)}$'s are the singular values in the non-increasing order. One may replace $\lambda_{[nt]}$ by

$$\Lambda_{[nt]} := \sum_{j=1}^{q} \left(\vartheta_{[nt]}^{(p+j)} \right)^2$$

in Proposition 2.3.1 (and analogously for $\widetilde{\lambda}_{[n(1-t)]}$). Then, the SWIP can be derived again (see the proof of Proposition 2.3.1), provided adequately extended assumptions on the errors $\{\varepsilon_{n,1}\}_{n=1}^{\infty}, \ldots, \{\varepsilon_{n,q}\}_{n=1}^{\infty}$ instead of the original ones $\{\varepsilon_n\}_{n=1}^{\infty}$. However, the consequent proofs would become more technical.

2.3.4 Nuisance-parameter-free detection

Estimating $\boldsymbol{\beta}$ via the TLS approach can be viewed as solving optimizing problem

$$[\boldsymbol{b}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\varepsilon}}] := \underset{[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}] \in \mathbb{R}^{n \times (p+1)}, \boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \|[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|_F \quad \text{s.t.} \quad \boldsymbol{Y} - \boldsymbol{\varepsilon} = (\boldsymbol{X} - \boldsymbol{\Theta})\boldsymbol{\beta}, \tag{2.15}$$

where $\|\cdot\|_F$ stands for the Frobenius matrix norm. Part of the solution (2.15) are fitted errors $[\hat{\boldsymbol{\Theta}}, \hat{\varepsilon}]$ such that

$$\left\|[\hat{\boldsymbol{\Theta}}, \hat{\varepsilon}]\right\|_F^2 = \lambda_n \tag{2.16}$$

due to Golub and Van Loan (1980). And we construct the changepoint test statistics based on property (2.16).

Changepoint test statistics

Let us think of two TLS estimates of β: The first one based on the first i data lines $[\boldsymbol{X}_i, \boldsymbol{Y}_i]$ and the second one based on the first k data lines $[\boldsymbol{X}_k, \boldsymbol{Y}_k]$ such that $1 \leqslant i \leqslant k \leqslant n$. Under the null $\mathcal{H}_0$, these two TLS estimates should be close to each other. On the other hand, under the alternative $\mathcal{H}_A$ such that $\tau \in \{i, \ldots, k\}$, they should be somehow different. A similar conclusion can be made for the *goodness-of-fit* statistics coming from (2.16). It means that

$$\lambda_i - \frac{i}{k}\lambda_k$$

should be reasonably small under the null $\mathcal{H}_0$. Under the alternative $\mathcal{H}_A$ such that $\tau \in \{i, \ldots, k\}$, it should be relatively large. For the multivariate case described in previous Subsection 2.3.3, one has to replace λ_k by $\Lambda_k = \sum_{j=1}^{q} \left(\vartheta_k^{(p+j)}\right)^2$.

We rely on *self-normalized test statistics* introduced by Shao and Zhang (2010), because the unknown quantity $\upsilon/(1 + \|\beta\|_2^2)$ from Proposition 2.3.1 cancels out in the test statistics. Our *supremum-type self-normalized test statistic* based on the *goodness-of-fit* is defined as

$$\mathscr{S}_n := \max_{1 \leqslant k < n} \frac{\left|\lambda_k - \frac{k}{n}\lambda_n\right|}{\max_{1 \leqslant i < k}\left|\lambda_i - \frac{i}{k}\lambda_k\right| + \max_{k < i \leqslant n}\left|\widetilde{\lambda}_i - \frac{n-i}{n-k}\widetilde{\lambda}_{k+1}\right|} \tag{2.17}$$

and the *integral-type self-normalized test statistic* is defined as

$$\mathscr{T}_n := \sum_{k=1}^{n-1} \frac{\left(\lambda_k - \frac{k}{n}\lambda_n\right)^2}{\sum_{i=1}^{k-1}\left(\lambda_i - \frac{i}{k}\lambda_k\right)^2 + \sum_{i=k+1}^{n}\left(\widetilde{\lambda}_i - \frac{n-i}{n-k}\widetilde{\lambda}_{k+1}\right)^2}. \tag{2.18}$$

Let us note that evaluations of the above defined test statistics require just several singular value decompositions, which is reasonably *quick*. Our new test statistics involve *neither nuisance parameters nor tuning constants* and will work for non-stationary and weakly dependent data.

Under the null hypothesis and the technical assumptions from Subsection 2.3.3, the test statistics defined in (2.17) and (2.18) converge to *non-degenerate limit distributions* (their quantiles can be found in Subsection 2.3.4).

Theorem 2.3.2 (Under the null)**.** *Let $\mathcal{M}$ and $\mathcal{H}_0$ hold. Suppose Assumptions $\mathcal{A}$.2.11, $\mathcal{A}$.2.12, and $\mathcal{A}$.2.13 are satisfied. Then,*

$$\mathscr{S}_n \xrightarrow[n\to\infty]{D} \sup_{t\in[0,1]} \frac{\left|W(t) - tW(1)\right|}{\sup_{s\in[0,t]}\left|W(s) - \frac{s}{t}W(t)\right| + \sup_{s\in[t,1]}\left|\widetilde{W}(s) - \frac{1-s}{1-t}\widetilde{W}(t)\right|} \tag{2.19}$$

and

$$\mathscr{T}_n \xrightarrow[n\to\infty]{D} \int_0^1 \frac{\left\{W(t) - tW(1)\right\}^2}{\int_0^t \left\{W(s) - \frac{s}{t}W(t)\right\}^2 ds + \int_t^1 \left\{\widetilde{W}(s) - \frac{1-s}{1-t}\widetilde{W}(t)\right\}^2 ds} dt, \tag{2.20}$$

where $\{W(t)\}_{t\in[0,1]}$ is a standard Wiener process and $\widetilde{W}(t) = W(1) - W(t)$.

The null hypothesis is rejected at significance level α for large values of $\mathscr{S}_n$ and $\mathscr{T}_n$. The critical values can be obtained as the $(1-\alpha)$-quantiles of the asymptotic distributions from (2.19) and (2.20). In order to describe limit behavior of the test statistics under the alternative, an additional *changepoint assumption* is required.

Assumption $\mathcal{A}$.2.14. If $n \to \infty$, then

$$\|\delta\|_2 \to 0 \quad \text{and} \quad \left\{\eta(\beta^\top \boldsymbol{\Delta}\beta + \boldsymbol{\delta}\boldsymbol{\Delta}_{-\zeta}\boldsymbol{\delta}) - \beta^\top \boldsymbol{\Delta}^2\beta\right\}\sqrt{n} \to \infty, \tag{2.21}$$

where $\sigma^2 + \eta = \lambda_{min}(\boldsymbol{\Delta} + \sigma^2 \boldsymbol{I}_p)$.

This assumption may be considered as a changepoint *detectability requirement*, because it manages the relationship between the size of the change, the location of the change, and the noisiness of the data in order to be able to detect the changepoint.

Now, the tests based on $\mathscr{S}_n$ and $\mathscr{T}_n$ are shown to be *consistent*, as the test statistics converge to infinity under some local alternatives, provided that the size of the change does not convergence to zero too fast.

Theorem 2.3.3 (Under local alternatives)**.** *Let $\mathcal{M}$ and $\mathcal{H}_A$ hold. Suppose Assumptions $\mathcal{A}$.2.14, $\mathcal{A}$.2.11, $\mathcal{A}$.2.12, and $\mathcal{A}$.2.13 are satisfied. If $\tau = [n\zeta]$ for some $\zeta \in (0,1)$, then*

$$\mathscr{S}_n \xrightarrow[n\to\infty]{P} \infty \xleftarrow[n\to\infty]{P} \mathscr{T}_n. \tag{2.22}$$

Assumption $\mathcal{A}$.2.14 can be sharpened as remarked below (Dembo, 1988; Ma and Zarowski, 1995).

Remark 2.3.1. The second part of relation (2.21) can be replaced by

$$\sqrt{n}\Big\{\kappa+\eta - \sqrt{(\kappa+2\sigma^2+\eta)^2-4(\kappa+\sigma^2-\beta^\top \boldsymbol{\Delta}(\boldsymbol{\Delta}+\sigma^2\boldsymbol{I}_p)^{-1}\boldsymbol{\Delta}\beta)(\sigma^2+\eta)}\Big\} \to \infty, \quad (2.23)$$

where $\kappa := \beta^\top \boldsymbol{\Delta}\beta + \boldsymbol{\delta}\boldsymbol{\Delta}_{-\zeta}\boldsymbol{\delta}$, and the assertion of Theorem 2.3.3 still holds.

Basically, Theorem 2.3.3 says that in presence of the structural change in linear relations, the test statistics *explode above all bounds*. Hence, the asymptotic distributions from Theorem 2.3.2 can be used to construct the tests. Although, explicit forms of those distributions stated in (2.19) and (2.20) are unknown.

Asymptotic critical values

The critical values may be determined by simulations from the limit distributions $\mathscr{S}_n$ and $\mathscr{T}_n$ from Theorem 2.3.2. Theorem 2.3.3 ensures that we reject the null hypothesis for large values of the test statistics. We have simulated the asymptotic distributions (2.19) and (2.20) by *discretizing* the standard Wiener process and using the relationship of a random walk to the standard Wiener process. We considered 1000 as the number of discretization points within $[0,1]$ interval and the number of simulation runs equals to 100000. In Table 2.2, we present several critical values for the test statistics $\mathscr{S}_n$ and $\mathscr{T}_n$.

Tab. 2.2. Simulated critical values corresponding to the asymptotic distributions of the test statistics $\mathscr{S}_n$ and $\mathscr{T}_n$ under the null hypothesis

$100(1-\alpha)\%$	90%	95%	97.5%	99%	99.5%
$\mathscr{S}_n$-based	1.209008	1.393566	1.571462	1.782524	1.966223
$\mathscr{T}_n$-based	5.700222	7.165705	8.807070	10.597625	11.755233

Changepoint estimator

If a change is *detected*, it is of interest to estimate the time of the change. It is sensible to use

$$\hat{\tau}_n := \underset{1\leqslant k\leqslant n-1}{\operatorname{argmax}} \frac{\left|\lambda_k - \frac{k}{n}\lambda_n\right|}{\max_{1\leqslant i<k}\left|\lambda_i - \frac{i}{k}\lambda_k\right| + \max_{k<i\leqslant n}\left|\widetilde{\lambda}_i - \frac{n-i}{n-k}\widetilde{\lambda}_{k+1}\right|}$$

as a *changepoint estimator*. Our next theorem shows that under the alternative, the changepoint τ is consistently estimated by the estimator $\hat{\tau}_n$.

Corollary 2.3.4 (Consistency)**.** *Let the assumptions of Theorem 2.3.3 hold. If for every* $t \in (\zeta, 1)$

$$\left\{\eta(t)\left(\beta^\top \boldsymbol{\Delta}_t \beta + \delta(\boldsymbol{\Delta}_t - \boldsymbol{\Delta}_\zeta)\delta\right) - \beta^\top \boldsymbol{\Delta}_t^2 \beta\right\} \sqrt{n} \xrightarrow{n\to\infty} \infty \tag{2.24}$$

and for every $t \in (0, \zeta)$

$$\left\{\tilde{\eta}(t)\left(\beta^\top \boldsymbol{\Delta}_{-t} \beta + \delta(\boldsymbol{\Delta}_\zeta - \boldsymbol{\Delta}_t)\delta\right) - \beta^\top \boldsymbol{\Delta}_{-t}^2 \beta\right\} \sqrt{n} \xrightarrow{n\to\infty} \infty, \tag{2.25}$$

where $\eta(t) = \lambda_{min}(\boldsymbol{\Delta}_t + t\sigma^2 \boldsymbol{I}_p) - t\sigma^2$ *and* $\tilde{\eta}(t) = \lambda_{min}(\boldsymbol{\Delta}_{-t} + (1-t)\sigma^2 \boldsymbol{I}_p) - (1-t)\sigma^2$, *then*

$$\frac{\hat{\tau}_n}{n} \xrightarrow[n\to\infty]{P} \zeta.$$

In order to estimate more than one changepoint, it is possible to use an arbitrary 'divide-and-estimate' *multiple changepoints* method relying on our changepoint estimator, for instance, wild binary segmentation by Fryzlewicz (2014).

2.4 Key contributions

- Change in mean and in trend of time series are investigated, allowing for heteroscedasticity and non-stationarity.
- Ratio and self-normalized test statistics are used for the changepoint detection in order to avoid long-run variance estimation.
- General score functions are introduced for the test statistics, making the approaches more robust against outliers and more suitable for heavy tailed distributions.
- Circular moving block bootstrap and wild bootstrap extensions are incorporated. Their validity is formally and computationally justified.
- Spectral weak invariance principle is proved.
- A nuisance-parameters-free methods are developed and their asymptotic behavior is derived under the null as well as under the alternative hypothesis.
- Changepoint estimators are proposed and their consistency is shown.

3 Changepoint in Panel Data

> *I don't need a friend who changes when I change and who nods when I nod; my shadow does that much better.*
>
> — **Plutarch**
> (Greek philosopher, biographer, and essayist)

To this end, we have considered a changepoint in a sequence of observations, where only one stochastic copy of the sequence is available. We focus now on a changepoint problem such that several sequences are subject to change simultaneously. Panel data of our interest consist of a moderate or relatively large number of panels, while the panels contain a small number of observations.

3.1 Testing for a common change

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. Foundations for the forthcoming results can be found in Peštová and Pešta (2015) and Peštová and Pešta (2016).

Tests for changepoint detection in the panel data have been proposed usually 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 changepoint 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 changepoint 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.

Motivation from insurance industry

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 $i = 1, \dots, N$ provides the total claim amount $Y_{i,t}$ paid in year $t = 1, \dots, T$ 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 section shows a way how to test for a possible historical instability.

3.1.1 Panel changepoint model

Let us consider the panel changepoint model

$$Y_{i,t} = \mu_i + \delta_i \mathcal{I}\{t > \tau\} + \sigma \varepsilon_{i,t}, \quad 1 \leqslant i \leqslant N,\ 1 \leqslant t \leqslant T; \tag{3.1}$$

where $\sigma > 0$ is an unknown variance-scaling parameter and T is fixed, not depending on N. The possible *common changepoint time* is denoted by $\tau \in \{1, \dots, T\}$. 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 $\mathcal{A}$.3.15. The vectors $[\varepsilon_{i,1}, \dots, \varepsilon_{i,T}]^\top$ existing on a probability space $(\Omega, \mathcal{F}, P)$ are iid for $i = 1, \dots, N$ with $E\,\varepsilon_{i,t} = 0$ and $\mathrm{Var}\,\varepsilon_{i,t} = 1$, having the autocorrelation function

$$\rho_t = \mathrm{Corr}\,(\varepsilon_{i,s}, \varepsilon_{i,s+t}) = \mathrm{Cov}\,(\varepsilon_{i,s}, \varepsilon_{i,s+t}), \quad \forall s \in \{1, \dots, T-t\},$$

which is independent of the lag s, the cumulative autocorrelation function

$$r(t) = \mathrm{Var} \sum_{s=1}^{t} \varepsilon_{i,s} = \sum_{|s|<t} (t - |s|)\rho_s,$$

and the shifted cumulative correlation function

$$R(t,v) = \mathrm{Cov}\left(\sum_{s=1}^{t} \varepsilon_{i,s}, \sum_{u=t+1}^{v} \varepsilon_{i,u} \right) = \sum_{s=1}^{t} \sum_{u=t+1}^{v} \rho_{u-s}, \quad t < v$$

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

$$H_0 : \tau = T$$

against the *alternative* that at least one panel has a change in mean

$$H_1 : \tau < T \quad \text{and} \quad \exists i \in \{1, \ldots, N\} : \delta_i \neq 0.$$

3.1.2 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 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 changepoint test statistic is

$$\mathcal{R}_N(T) = \max_{t=2,\dots,T-2} \frac{\max_{s=1,\dots,t} \left| \sum_{i=1}^{N} \left[\sum_{r=1}^{s} \left(Y_{i,r} - \bar{Y}_{i,t} \right) \right] \right|}{\max_{s=t,\dots,T-1} \left| \sum_{i=1}^{N} \left[\sum_{r=s+1}^{T} \left(Y_{i,r} - \tilde{Y}_{i,t} \right) \right] \right|},$$

where $\bar{Y}_{i,t}$ is the average of the first t observations in panel i and $\tilde{Y}_{i,t}$ is the average of the last $T-t$ observations in panel i, i.e.,

$$\bar{Y}_{i,t} = \frac{1}{t} \sum_{s=1}^{t} Y_{i,s} \quad \text{and} \quad \tilde{Y}_{i,t} = \frac{1}{T-t} \sum_{s=t+1}^{T} Y_{i,s}.$$

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 $\mathcal{R}_N(T)$. 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 3.1.1 (Under null). *Under hypothesis* H_0 *and Assumption* $\mathcal{A}$.3.15

$$\mathcal{R}_N(T) \xrightarrow[N\to\infty]{D} \max_{t=2,\dots,T-2} \frac{\max_{s=1,\dots,t} \left| X_s - \frac{s}{t} X_t \right|}{\max_{s=t,\dots,T-1} \left| Z_s - \frac{T-s}{T-t} Z_t \right|},$$

where $Z_t := X_T - X_t$ *and* $[X_1, \dots, X_T]^\top$ *is a multivariate normal random vector with zero mean and covariance matrix* $\boldsymbol{\Lambda} = \{\lambda_{t,\nu}\}_{t,\nu=1}^{T,T}$ *such that*

$$\lambda_{t,t} = r(t) \quad \textit{and} \quad \lambda_{t,\nu} = r(t) + R(t,\nu), \ t < \nu.$$

The limiting distribution does not depend on the variance nuisance parameter σ, but it depends on the unknown correlation structure of the panel changepoint model, which has to be estimated for testing purposes. The way of its estimation is shown in Subsection 3.1.4. Furthermore, Theorem 3.1.1 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 $\boldsymbol{\Lambda}$ is simplified such that $r(t) = t$ and $R(t, v) = 0$.

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

Assumption $\mathcal{A}$.3.16. $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \left| \sum_{i=1}^{N} \delta_i \right| = \infty$.

Theorem 3.1.2 (Under alternative). *If* $\tau \leqslant T - 3$, *then under Assumptions* $\mathcal{A}$.3.15, $\mathcal{A}$.3.16 *and alternative* H_1

$$\mathcal{R}_N(T) \xrightarrow[N\to\infty]{P} \infty. \tag{3.2}$$

Assumption $\mathcal{A}$.3.16 is satisfied, for instance, if $0 < \delta \leqslant \delta_i \, \forall i$ (a common lower changepoint threshold) and $\delta\sqrt{N} \to \infty$, $N \to \infty$. Another suitable example of δ_is for the condition in Assumption $\mathcal{A}$.3.16, can be $0 < \delta_i = KN^{-1/2+\eta}$ for some $K > 0$ and $\eta > 0$. Or $\delta_i = Ci^{\alpha-1}\sqrt{N}$ 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 changepoint. It is also possible to redefine the test statistic by interchanging the nominator and the denominator of $\mathcal{R}_N(T)$. Afterwards, Theorem 3.1.2 for the modified test statistic would require three observations before the changepoint, i.e., $\tau \geqslant 3$.

Theorem 3.1.2 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 3.1.1 can be used to construct the test.

3.1.3 Changepoint estimation

Despite the fact that the aim of the section is to establish testing procedures for detection of a panel mean change, it is necessary to construct a *consistent estimate* for a possible changepoint. There are two reasons for that: Firstly, the estimation of the covariance matrix $\boldsymbol{\Lambda}$ from Theorem 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 changepoint 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 changepoint 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 as

$$\widehat{\tau}_N := \arg \min_{t=2,\dots,T} \frac{1}{w(t)} \sum_{i=1}^{N} \sum_{s=1}^{t} (Y_{i,s} - \bar{Y}_{i,t})^2, \tag{3.3}$$

where $\{w(t)\}_{t=2}^{T}$ is a sequence of weights specified later on.

Assumption $\mathcal{A}$.3.17. The sequence $\left\{ \frac{t}{w(t)} \left(1 - \frac{r(t)}{t^2}\right) \right\}_{t=2}^{T}$ is decreasing.

Assumption $\mathcal{A}$.3.18. There exist constants $L > 0$ and $N_0 \in \mathbb{N}$ such that

$$L < \sigma^2 \left[\frac{t}{w(t)} \left(1 - \frac{r(t)}{t^2}\right) - \frac{\tau}{w(\tau)} \left(1 - \frac{r(\tau)}{\tau^2}\right) \right] + \frac{\tau(t-\tau)}{t w(t)} \frac{1}{N} \sum_{i=1}^{N} \delta_i^2,$$

for each $t = \tau + 1, \dots, T$ and $N \geqslant N_0$.

Assumption $\mathcal{A}$.3.19. $\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \delta_i^2 = 0$.

Assumption $\mathcal{A}$.3.20. $\mathsf{E}\, \varepsilon_{1,t}^4 < \infty$, $t \in \{1, \dots, T\}$.

Theorem 3.1.3 (Changepoint estimate consistency)**.** *Suppose that* $\tau \neq 1$. *Then under Assumptions A1,* $\mathcal{A}$.3.17, $\mathcal{A}$.3.18, $\mathcal{A}$.3.19, *and* $\mathcal{A}$.3.20

$$\lim_{N\to\infty} \mathsf{P}[\widehat{\tau}_N = \tau] = 1.$$

Assumption $\mathcal{A}$.3.18 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. Assumption $\mathcal{A}$.3.19 is needed to control the asymptotic boundedness of the variability of $\frac{1}{w(t)} \sum_{i=1}^{N} \sum_{s=1}^{t} (Y_{i,s} - \bar{Y}_{i,t})^2$, because a finite T cannot do that.

Assumptions $\mathcal{A}$.3.18 and $\mathcal{A}$.3.19 are satisfied for $0 < \delta \leqslant \delta_i < \Delta, \forall i$ (a common lower and upper bound for the change amount) and suitable σ, $r(t)$, and $w(t)$. The monotonicity Assumption $\mathcal{A}$.3.17 in not very restrictive at all. For example in case of independent observations within the panel (i.e., $r(t) = t$) and weight function $w(t) = t^q$, $q \geqslant 2$, this assumption is automatically fulfilled, since sequence $\{t^{1-q} - t^{-q}\}_{t=2}^{T}$ is decreasing. This also gives us an idea how to choose weights $w(t)$.

If one is interested in sensitivity of the changepoint estimate (i.e., what is the size of the change that can be estimated), let us consider the following model scenario: $T = 10$, $\tau = 5$, $\sigma = 0.1$, independent observations within the panel, and $w(t) = t^2$.

Then, Assumption $\mathcal{A}$.3.18 is satisfied if $\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 > 0.029$ for all $N \geqslant N_0$. In case of a common value of $\delta = \delta_i$ for all i, we need $\delta > \sqrt{0.029} \approx 0.170$.

Assumption $\mathcal{A}$.3.18 can be considered as too complicated. Therefore, one can replace it by the following simpler, but more restrictive assumption.

Assumption $\mathcal{A}$.3.21.

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 = \infty.$$

On one hand, this assumption might be considered as too strong, because a common fixed (not depending on N) value of $\delta = \delta_i$ for all i does not fulfill Assumption $\mathcal{A}$.3.21. On the other hand, Assumption $\mathcal{A}$.3.21 is satisfied when $\delta_j^2/N \to \infty$ as $N \to \infty$ for some $j \in \mathbb{N}$ and $\delta_i = 0$ for all $i \neq j$. This stands for a situation when all the panels do not change in mean except one panel having a sufficiently large change in mean with respect to the number of panels.

3.1.4 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 $[\varepsilon_{1,1},\dots,\varepsilon_{1,T}]^\top$ empirically. We base the errors' estimates on *residuals*

$$\widehat{e}_{i,t} := \begin{cases} Y_{i,t} - \bar{Y}_{i,\widehat{\tau}_N}, & t \leqslant \widehat{\tau}_N, \\ Y_{i,t} - \tilde{Y}_{i,\widehat{\tau}_N}, & t > \widehat{\tau}_N. \end{cases} \tag{3.4}$$

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):

$$\widehat{\tau}(t) = \sum_{|s|<t}(t-|s|)\kappa\left(\frac{s}{h}\right)\widehat{\rho}_s,$$

$$\widehat{R}(t,v) = \sum_{s=1}^{t}\sum_{u=t+1}^{v}\kappa\left(\frac{u-s}{h}\right)\widehat{\rho}_{u-s}, \quad t < v;$$

where $h > 0$ stands for the window size and κ belongs to a class of kernels given by

$$\Big\{\kappa(\cdot):\mathbb{R}\to[-1,1]\,\Big|\,\kappa(0)=1,\ \kappa(x)=\kappa(-x),\ \forall x,\ \int_{-\infty}^{+\infty}\kappa^2(x)dx<\infty,$$
$$\kappa(\cdot)\ \text{is continuos at 0 and at all but a finite number of other points}\Big\}.$$

Since the variance parameter σ is not present in the limiting distribution of Theorem 3.1.1, it neither has to be estimated nor known. Nevertheless, one can use $\widehat{\sigma}^2 := \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\widehat{e}_{i,s}^2$.

3.1.5 Residual bootstrap and hypothesis testing

A wide range of literature has been published on bootstrapping in the changepoint 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 $\{[\widehat{e}_{i,1},\dots,\widehat{e}_{i,T}]\}_{i=1,\dots,N}$ corresponding to the panels. This provides bootstrapped row vectors $\{[\widehat{e}^*_{i,1},\dots,\widehat{e}^*_{i,T}]\}_{i=1,\dots,N}$. 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 $\mathcal{R}_N(T)$, where the original observations $Y_{i,t}$ are replaced by their bootstrap counterparts $\widehat{Y}^*_{i,t}$:

$$\mathcal{R}^*_N(T) = \max_{t=2,\dots,T-2}\frac{\max_{s=1,\dots,t}\left|\sum_{i=1}^{N}\left[\sum_{r=1}^{s}\left(\widehat{Y}^*_{i,r}-\tilde{\widehat{Y}}^*_{i,t}\right)\right]\right|}{\max_{s=t,\dots,T-1}\left|\sum_{i=1}^{N}\left[\sum_{r=s+1}^{T}\left(\widehat{Y}^*_{i,r}-\tilde{\widehat{Y}}^*_{i,t}\right)\right]\right|},$$

such that

$$\tilde{\widehat{Y}}^*_{i,t} = \frac{1}{t}\sum_{s=1}^{t}\widehat{Y}^*_{i,s} \quad\text{and}\quad \tilde{\widehat{Y}}^*_{i,t} = \frac{1}{T-t}\sum_{s=t+1}^{T}\widehat{Y}^*_{i,s}.$$

An *algorithm* for the bootstrap is illustratively shown in Procedure 3.1.1 and its validity will be proved in Theorem 3.1.4.

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 $\mathbb{Y} \equiv \{Y_{i,t}\}_{i,t=1}^{N,T}$.

Procedure 3.1.1 Bootstrapping test statistic $\mathcal{R}_N(T)$.

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 $\mathcal{R}_N(T)$, i.e., the empirical distribution where probability mass $1/B$ concentrates at each of ${}_{(1)}\mathcal{R}^*_N(T), \ldots, {}_{(B)}\mathcal{R}^*_N(T)$.

1: estimate the changepoint by calculating $\widehat{\tau}_N$
2: compute residuals $\widehat{e}_{i,t}$
3: **for** $b = 1$ to B **do** // repeat in order to obtain the empirical distribution
4: $\{[\widehat{e}^*_{i,1}, \ldots, \widehat{e}^*_{i,T}]\}_{i=1}^N$ resampled with replacement from original rows $\{[\widehat{e}_{i,1}, \ldots, \widehat{e}_{i,T}]\}_{i=1}^N$
5: calculate bootstrap panel data $\widehat{Y}^*_{i,t}$
6: compute bootstrap test statistics ${}_{(b)}\mathcal{R}^*_N(T)$
7: **end for**

First of all, two simple and just technical assumptions are needed.

Assumption $\mathcal{A}$.3.22. $\{\varepsilon_{i,t}\}_t$ possesses the lagged cumulative correlation function

$$S(t, v, d) = \mathrm{Cov}\left(\sum_{s=1}^{t} \varepsilon_{i,s}, \sum_{u=t+d}^{v} \varepsilon_{i,u}\right) = \sum_{s=1}^{t} \sum_{u=t+d}^{v} \rho_{u-s}, \quad \forall i \in \mathbb{N}.$$

Assumption $\mathcal{A}$.3.23. $\lim_{N\to\infty} P[\widehat{\tau}_N = \tau] = 1$.

Assumption $\mathcal{A}$.3.22 is not really an assumption, actually it is only a notation. Notice that $S(t, v, 1) \equiv R(t, v)$. Assumption $\mathcal{A}$.3.23 is satisfied for our estimate proposed in (3.3), if the assumptions of Theorem 3.1.2 hold. Assumption $\mathcal{A}$.3.23 is postulated in a rather broader sense, because we want to allow any other consistent estimate of τ to be used instead.

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 3.1.4 (Bootstrap justification). *Under Assumptions* $\mathcal{A}$.3.15, $\mathcal{A}$.3.22, $\mathcal{A}$.3.23, *and* $\mathcal{A}$.3.20

$$\mathcal{R}^*_N(T)|\mathbb{Y} \xrightarrow[N\to\infty]{D} \max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t} \left|\mathcal{X}_s - \frac{s}{t}\mathcal{X}_t\right|}{\max_{s=t,\ldots,T-1} \left|\mathcal{Z}_s - \frac{T-s}{T-t}\mathcal{Z}_t\right|} \quad \textit{in probability } P,$$

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* $\boldsymbol{\Gamma} = \{\gamma_{t,\nu}(\tau)\}_{t,\nu=1}^{T,T}$ *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

$$\gamma_{t,\nu}(\tau) = \begin{cases} 0, \quad t = \tau \text{ or } \nu = \tau, \\ r(t) + R(t,\nu) + \frac{t\nu}{\tau^2} r(\tau) - \frac{\nu}{\tau}[r(t) + R(t,\tau)] \\ \quad -\frac{t}{\tau}[r(\nu) + R(\nu,\tau)], \quad t < \nu < \tau; \\ S(t,\nu,\tau+1-t) + \frac{t(\nu-\tau)}{\tau(T-\tau)} R(\tau,T) \\ \quad -\frac{\nu-\tau}{T-\tau} S(t,T,\tau+1-t) - \frac{t}{\tau} R(\tau,\nu), \quad t < \tau < \nu; \\ r(t-\tau) + R(t-\tau,\nu-\tau) + \frac{(t-\tau)(\nu-\tau)}{(T-\tau)^2} r(T-\tau) \\ \quad -\frac{\nu-\tau}{T-\tau}[r(t-\tau) + R(t-\tau, T-\tau)] \\ \quad -\frac{t-\tau}{T-\tau}[r(\nu-\tau) + R(\nu-\tau, T-\tau)], \quad \tau < t < \nu. \end{cases}$$

The validity of the bootstrap test is assured by Theorem 3.1.4. 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 3.1.1.

Theorem 3.1.5 (Bootstrap test consistency)**.** *Under Assumptions* $\mathcal{A}$.3.15, $\mathcal{A}$.3.23, $\mathcal{A}$.3.20 *and hypothesis* H_0*, the asymptotic distribution of* $\mathcal{R}_N(T)$ *from Theorem 3.1.1 and the asymptotic distribution of* $\mathcal{R}_N^*(T)|\mathbb{Y}$ *from Theorem 3.1.4 coincide.*

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.

Note that one cannot think about any local alternative in this setup, because τ has a discrete and finite support. Finally, the above elaborated model with the corresponding results can be extended for a setup assuming *dependence within panels*, see Maciak et al. (2018).

3.2 Estimation of common breaks

We now introduce another improved common break point estimator, which relies on less restrictive assumptions and has better finite sample performance compared to the estimator presented in the previous section. Here, the theoretical summarized results are obtained from Peštová and Pešta (2017).

Tests for changepoint detection in panel data have been proposed by Horváth and Hušková (2012) for sufficiently large panel sizes T, i.e., the limiting results were derived under the assumption that T increases over all limits. Testing procedures for the change in panel means with T fixed, that can be relatively small or moderate, were considered in Peštová and Pešta (2015). The changepoint estimation in panel data for fixed as well as for unbounded T was studied by Bai (2010). However, the panel changepoint estimator in Bai (2010) is derived only for a situation, where one knows for sure, that the change in means occurred within the given time period. This restriction can become insurmountable for some further utilizations of the changepoint estimator, as it will be demonstrated later in this thesis. In Peštová and Pešta (2016), a consistent changepoint estimator was introduced, requiring no definite knowledge about existence of the changepoint in the given panel data. In the case of no change being present, the estimator picks the last observation, which means that no structural break is identified. However, this estimator has several disadvantages. It assumes a certain kind of homoscedasticity in the panels. Further, it does not take into account the possibility that the change may occur right after the first time point. It also assumes conditions that may be viewed as too complicated with regard to verification and model checking. The remaining task is, therefore, to develop a changepoint estimator that is consistent regardless of the change's presence/absence. Moreover, such estimator would gain from allowing *heteroscedasticity* in the panels, having broader scope of applications. Besides that, the applicability of the estimator is enhanced by simple consistency conditions and no *boundary issue*. The boundary issue means that the changepoint can neither be detected nor estimated when being close to the beginning or to the end of the observation regime.

Further on, Kim (2011) and Kim (2014) dealt with the changepoint estimator under cross sectional dependence in the panels modeled by a common factor, and expanded the estimation problem for more complicated types of structural changes. The first and second order asymptotics that can be used to derive consistent confidence intervals for the time of change in panel data was established by Horváth et al. (2015). The panel length T was considered as unbounded and depending on the number of panels N. However, there is some literature on the short panel

changepoint framework where also weighting functions, as we employ later on, are suggested, cf. Baltagi et al. (2016).

3.2.1 Abrupt change in panel data

Let us consider the panel changepoint model

$$Y_{i,t} = \mu_i + \delta_i \mathcal{J}\{t > \tau\} + \sigma_i \varepsilon_{i,t}, \quad 1 \leqslant i \leqslant N,\ 1 \leqslant t \leqslant T; \tag{3.5}$$

where $\sigma_i > 0$ are unknown variance-scaling panel-specific parameters and T is fixed, not depending on N. The possible *common changepoint time* is denoted by $\tau \in \{1, \dots, T\}$. 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. There is at most one change per panel in model (3.5) and the type of change in the panel mean is *abrupt*.

Furthermore, it is assumed that the sequences of panel disturbances $\{\varepsilon_{i,t}\}_t$ are independent. At the same time, the errors within each panel form a weakly stationary sequence with a common correlation structure. This can be formalized in the following assumption.

Assumption $\mathcal{A}$.3.24. The vectors $[\varepsilon_{i,1}, \dots, \varepsilon_{i,T}]^\top$ existing on a probability space $(\Omega, \mathcal{F}, P)$ are iid for $i = 1, \dots, N$ with $E\,\varepsilon_{i,t} = 0$ and $\operatorname{Var} \varepsilon_{i,t} = 1$, having the autocorrelation function

$$\rho_t = \operatorname{Corr}(\varepsilon_{i,s}, \varepsilon_{i,s+t}) = \operatorname{Cov}(\varepsilon_{i,s}, \varepsilon_{i,s+t}), \quad \forall s \in \{1, \dots, T-t\},$$

which is independent of the lag s, the cumulative autocorrelation function

$$r(t) = \operatorname{Var} \sum_{s=1}^{t} \varepsilon_{i,s} = \sum_{|s|<t} (t - |s|)\rho_s,$$

and the shifted cumulative correlation function

$$R(t, \nu) = \operatorname{Cov}\left(\sum_{s=1}^{t} \varepsilon_{i,s}, \sum_{u=t+1}^{\nu} \varepsilon_{i,u} \right) = \sum_{s=1}^{t} \sum_{u=t+1}^{\nu} \rho_{u-s}, \quad t < \nu$$

for all $i = 1, \dots, N$ and $t, \nu = 1, \dots, T$. The covariance matrix

$$\boldsymbol{\Lambda} := \operatorname{Var} \left[\sum_{s=1}^{1} \varepsilon_{1,s}, \dots, \sum_{s=1}^{T} \varepsilon_{1,s} \right]^\top$$

is non-singular.

The sequence $\{\varepsilon_{i,t}\}_{t=1}^{T}$ can be viewed as a part of a *weakly stationary* process. Note that the within-panel dependent errors do not necessarily need to be linear processes. GARCH processes are a plausible alternative, for instance.

The assumption of independent panels can be relaxed. It would, however, make the setup much more complex, cf. Kim (2011). 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 with regard to real life experience.

Assumption $\mathcal{A}$.3.25. There exist constants $\underline{\sigma}, \overline{\sigma} > 0$ not depending on N, such that

$$\underline{\sigma} \leqslant \sigma_i \leqslant \overline{\sigma}, \quad 1 \leqslant i \leqslant N.$$

The assumption of the bounded panel variances from both below and above allows for heteroscedasticity between the panels. In case, when the equiboundedness cannot be satisfied, the panel model can be generalized by introducing weights $w_{i,t}$, which are supposed to be known. Subsequently, claim ratios $Y_{i,t}/w_{i,t}$ can be modeled. For instance in actuarial practice (see the insurance motivation in the previous section), it would mean to normalize the total claim amount by the premium received (considered as the weight), since bigger insurance companies are expected to have higher variability in total claim amounts paid.

3.2.2 Another changepoint estimator

A consistent estimator of the changepoint in 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 has occurred or not*. Therefore, we modify the estimate proposed by Bai (2010) in the following way. If the panel means change somewhere inside $\{1, \dots, T-1\}$, let the estimate select this break point. If there is no change in panel means, the estimator points out the very last time point T with probability going to one. In other words, the value of the changepoint estimate can be T meaning no change. This is in contrast to Bai (2010), where T is not reachable.

Our estimator of the time of change τ in panel data is defined as

$$\widehat{\tau}_N := \arg \min_{t=1,\dots,T} \sum_{i=1}^{N} \left\{ \frac{1}{w(t)} \sum_{s=1}^{t} (Y_{i,s} - \bar{Y}_{i,t})^2 + \frac{1}{w(T-t)} \sum_{s=t+1}^{T} (Y_{i,s} - \widetilde{Y}_{i,t})^2 \right\}, \tag{3.6}$$

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.,

$$\bar{Y}_{i,t} = \frac{1}{t}\sum_{s=1}^{t} Y_{i,s} \quad \text{and} \quad \widetilde{Y}_{i,t} = \frac{1}{T-t}\sum_{s=t+1}^{T} Y_{i,s}.$$

By convention, the value of an empty sum is zero. A sequence of positive weights $\{w(t)\}_{t=0}^{T}$ is specified later on.

Consistency

We postulate additional assumptions on the panel changepoint model (3.5) in order to derive the estimator's consistency. The following conditions take into account that the length T of the observation regime is fixed; that the length T does not depend on the number of panels N; and that the length T can even be relatively small.

Assumption $\mathcal{A}$.3.26. Let $g(t) := \frac{t}{w(t)}\left(1 - \frac{r(t)}{t^2}\right)$ for $t \in \{1,\dots,T\}$, $g(0) \equiv 0$, and

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \left\{ \frac{\tau}{\tau+1}\sum_{i=1}^{N} \delta_i^2 - (g(\tau)+g(T-\tau)) \max_{t=1,\dots,T} w(t) \sum_{i=1}^{N} \sigma_i^2 \right\} = \infty,$$

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \left\{ \frac{T-\tau}{T-\tau+1}\sum_{i=1}^{N} \delta_i^2 - (g(\tau)+g(T-\tau)) \max_{t=1,\dots,T} w(t) \sum_{i=1}^{N} \sigma_i^2 \right\} = \infty.$$

Assumption $\mathcal{A}$.3.27. $\lim_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^{N} \delta_i^2 = 0$.

Assumption $\mathcal{A}$.3.28. $\mathsf{E}\,\varepsilon_{1,t}^4 < \infty$, $t \in \{1,\dots,T\}$.

Theorem 3.2.1 (Changepoint estimator consistency)**.** *Under Assumptions* $\mathcal{A}$.3.24–$\mathcal{A}$.3.28

$$\lim_{N\to\infty} \mathsf{P}[\widehat{\tau}_N = \tau] = 1.$$

The formally postulated estimator's consistency in Theorem 3.2.1 can be practically interpreted: as one observes more and more panels, the probability that the proposed estimator is different from the true unknown changepoint gets smaller and smaller.

Assumption $\mathcal{A}$.3.26 is not restrictive at all, although it may be seen as a complicated one. For example in case of independent observations within the panel (i.e., $r(t) = t$) and the weight function $w(t) = t^q$, $q \geqslant 2$ for $t \in \{1,\dots,T\}$, $w(0) = 1$, the sequence $\{g(t)\}_{t=2}^{T}$ becomes $\{t^{1-q} - t^{-q}\}_{t=2}^{T}$ and is non-increasing. Then Assumption $\mathcal{A}$.3.26 is automatically fulfilled, if $q = 2$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\delta_i^2 - T^2\sigma_i^2\right) \to \infty$ as $N \to \infty$.

This also gives us an idea how to choose the weights $w(t)$. Conditions $\mathcal{A}.3.24$ and $\mathcal{A}.3.26$, we impose on the model errors, only pertain to the correlation structure. Hence, our results hold for nearly all stationary time series models of interest, including nonlinear time series like the ARCH and GARCH processes. Moreover, Assumption $\mathcal{A}.3.26$ controls trade-off between the size of breaks and the variability of errors. It may be considered as a *detectability assumption*, because it specifies the value of signal-noise ratio for finding the consistent estimator.

Assumptions $\mathcal{A}.3.26$ and $\mathcal{A}.3.27$ are satisfied, for instance, if $0 < \delta \leqslant \delta_i \leqslant \Delta$ for all i's (a common lower and upper threshold for the means' shifts), $\delta^2 = \mathcal{O}\left(N^{\zeta}\right)$, $\zeta > 0$ and $\Delta^2/N \to 0$ as $N \to \infty$ (bearing in mind Assumption $\mathcal{A}.3.25$). Another suitable example of δ_i's for the conditions in Assumptions $\mathcal{A}.3.26$ and $\mathcal{A}.3.27$, can be $\delta_i = Ki^{\eta}$ for some $K > 0$ and $0 < \eta < 1/2$. Conditions $\mathcal{A}.3.26$ and $\mathcal{A}.3.27$ do not require each panel to have a break. Sometimes, a *more restrictive* assumption can be assumed instead of Assumptions $\mathcal{A}.3.26$ and $\mathcal{A}.3.27$, e.g.,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i^2 = \infty. \tag{3.7}$$

On one hand, this assumption might be considered as too strong, because a common fixed (not depending on N) value of $\delta = \delta_i$ for all i's does not fulfill (3.7). On the other hand, (3.7) is satisfied when $\delta_j^2/N \to \infty$ as $N \to \infty$ for some $j \in \mathbb{N}$ and $\delta_i = 0$ for all $i \neq j$. This stands for a situation when all the panels do not change in mean except one panel having a sufficiently large change in mean with respect to the number of panels. Let us notice that one could replace Assumption $\mathcal{A}.3.26$ with a stronger assumption from (3.7), but it would mean disappearing the detectability relation between the size of breaks and the variability of errors. One would also lose an idea how to choose the weights. Furthermore, Assumptions E1 and E2 from Peštová and Pešta (2016) are more restrictive than Assumption $\mathcal{A}.3.26$, which makes the presented approach even more general.

Various competing consistent estimators of a possible changepoint can be suggested, e.g., the maximizer of $\sum_{i=1}^{N} \left[\sum_{s=1}^{t} (Y_{i,s} - t\bar{Y}_{i,T})\right]^2$ as in Horváth et al. (2015). To show consistency of this estimator, one needs to postulate different assumptions on the cumulative autocorrelation function and this may be rather complex.

In our opinion, it is erroneously assumed in Bai (2010) that only the second moment of the errors is sufficient to prove the consistency result. In particular, Lemma A.1 from Bai (2010) has to require the finite fourth errors' moments, which coincides with Assumption $\mathcal{A}.3.28$.

3.3 Key contributions

- Detection of structural changes in panel data, which consist of a moderate or relatively large number of panels, while the panels contain a small number of observations, is studied.
- Variance function estimation is avoided by utilizing the ratio test statistics.
- There is no need to estimate the correlation structure due to residual bootstrapping.
- A boundary issue is overcome, allowing to detect very early or late changes.
- Various consistent changepoint estimators are developed.
- In the case of no change being present, the estimator picks the last observation, which means that no structural break is identified.

4 Dynamics in Triangles

> *I ran into Isosceles. He had a great idea for a new triangle!*
>
> — **Woody Allen**
> (American director, writer, and actor)

If data can be somehow organized into the triangles, then one naturally thinks of the triangular schemes as discussed in Section 2. One of the typical properties of these data structures is that each consecutive row has more and more data entries compared to the previous row. What happens if it is vice-versa? Thus, we are going to model dynamic behavior of panel data, where each next row is one observation shorter.

4.1 Consistency and inconsistency of the chain ladder

Claims reserving is a classical problem in general insurance. A number of various methods has been invented in this field, see England and Verrall (2002) or Wüthrich and Merz (2008) for an overview. Among them, the chain ladder method is probably the most popular and frequently used one for estimating outstanding claims reserves. Besides its simplicity, this approach leads to reasonable estimates of the outstanding loss liabilities under quite mild assumptions on the mean structure and under the assumption of independence of the observations in different accident years.

In recent years, many authors investigated the relationship between various stochastic models and the chain ladder technique, see for instance Mack (1994b) or Renshaw and Verrall (1998). A number of different properties of the estimated ultimate claims amount have been studied. The distribution-free approach introduced by Mack (1993) is probably the most famous one.

In this section we deal with some *asymptotic properties* of the estimators of development factors within this distribution-free framework. Various types of the

conditional asymptotic consistency are defined. Necessary and sufficient conditions for being the development factors' estimates conditionally consistent are provided and discussed as originally done in Pešta and Hudecová (2012).

4.1.1 Chain ladder

We introduce the classical claims reserving notation and terminology. Outstanding loss liabilities are structured in so-called claims development triangles. Let us denote $X_{i,j}$ all the claim amounts in development year j with accident year i. Therefore, $X_{i,j}$ stands for the *incremental claims* in accident year i made in accounting year $i+j$. The current year is n, which corresponds to the most recent accident year and development period as well. That is, our data history consists of right-angled isosceles triangles $X_{i,j}$, where $i = 1, \dots, n$ and $j = 1, \dots, n+1-i$.

Suppose that $C_{i,j}$ are *cumulative payments* or *cumulative claims* in origin year i after j development periods, i.e., $C_{i,j} = \sum_{k=1}^{j} X_{i,k}$. Hence, C_{ij} is a random variable of which we have an observation if $i+j < n+1$ (run-off triangle, see Table 4.1). The aim is to estimate the ultimate claims amount $C_{i,n}$ and the outstanding claims reserve $R_i = C_{i,n} - C_{i,n+1-i}$ for all $i = 2, \dots, n$.

Tab. 4.1. Run-off triangle for cumulative claims $C_{i,j}$.

Accident	Development year j				
year i	1	2	$\cdots$	$n-1$	n
1	$C_{1,1}$	$C_{1,2}$	$\cdots$	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	$\cdots$	$C_{2,n-1}$	
		$\ddots$			
$\vdots$	$\vdots$	$\vdots$	$C_{i,n+1-i}$		
$n-1$	$C_{n-1,1}$	$C_{n-1,2}$			
n	$C_{n,1}$				

Distribution-free approach

The *distribution-free chain ladder* reserving technique is still one of the most frequently used approaches in non-life reserving.

Suppose that $\{C_{i,j}\}_{i,j=1}^{n,n}$ are random variables on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Assume the following stochastic assumptions:

[1] $\mathsf{E}[C_{i,j+1}|C_{i,1},\dots,C_{i,j}] = f_j C_{i,j}, \quad 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant n-1;$

[2] $\mathsf{Var}[C_{i,j+1}|C_{i,1},\dots,C_{i,j}] = \sigma_j^2 C_{i,j}, \quad 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant n-1;$

[3] accident years $[C_{i,1},\dots,C_{i,n}], \quad 1 \leqslant i \leqslant n$ are independent vectors.

These stochastic assumptions correspond to the distribution-free approach proposed by Mack (1993). The parameters f_j are referred to as *development factors*. If n years of the claims history are available then the *estimates of the development factors* based on the chain ladder method are given as

$$\widehat{f}_j^{(n)} = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}, \quad 1 \leqslant j \leqslant n-1; \qquad \widehat{f}_n^{(n)} \equiv 1 \quad \text{(assuming no tail)}. \tag{4.1}$$

The upper index in (4.1) is used in order to emphasize that the estimate of development factor f_j depends on n years of history, i.e., we prefer $\widehat{f}_j^{(n)}$ more than $\widehat{f}_j$ from the formal point of view. The ultimate claims amounts C_{in} are estimated by

$$\widehat{C}_{in} = C_{i,n+1-i} \times \widehat{f}_{n+1-i}^{(n)} \times \cdots \times \widehat{f}_{n-1}^{(n)}.$$

Mack (1993) proved that the estimators $\widehat{f}_j^{(n)}$ are unbiased and mutually uncorrelated under the assumptions [1] and [3] together with an additional assumption

[4] $\sum_{i=1}^{n-j} C_{i,j} > 0, \quad 1 \leqslant j \leqslant n-1.$

Furthermore, the assumption [2] is essential for the calculation of the mean squared error and the standard error of $\widehat{C}_{in}$. It has to be remarked that assumption [2] straightforwardly postulates a condition that $C_{i,j} \geqslant 0$ [P]-a.s. for all $i, j \in \mathbb{N}$.

If the cumulative claim $C_{i,n+1-i}$ is equal to zero for some particular accident year $i \in \{1,\dots,n\}$, then all the consequent predictions $\widehat{C}_{i,j}$, where $j > n+1-i$, are zeros as well. But this situation occurs very exceptionally. Independence assumption [3] can sometimes be viewed as slightly unrealistic. In these cases, the

chain ladder does not seem to be a suitable choice in reserving. Nevertheless, the assumptions of distribution-free chain ladder were thoroughly discussed many times and we refer the reader for completeness to relevant articles, e.g., Mack (1994a) or Mack (1994b).

Properties of development factors' estimators

Unbiasedness of the development factors estimators $\widehat{f}_j^{(n)}$ is often stressed out as an important advantageous property of the chain ladder method. However, unbiasedness of an estimator as such is from the statistical point of view of less importance compared to the consistency. A simple example illustrates why. Suppose that $Y_1, \dots, Y_n$ are iid variables sampled from a distribution with finite mean $\mathsf{E}\,Y$. One can use $T_1(Y_1, \dots, Y_n) = Y_1$ as an estimator of the unknown mean $\mathsf{E}\,Y$. This would be, of course, very naive in practice as only the first observation from the sample is used and directly taken as the estimator. However, we take this example because of its simplicity. The estimator T_1 is obviously unbiased, but surely inconsistent. On the other hand, the estimator $T_2(Y_1, \dots, Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i + \frac{1}{n}$ is biased, but *consistent*. It is easy to see that T_2 approaches $\mathsf{E}\,Y$ with probability one as n tends to infinity. Alternatively speaking, for n large enough, T_2 is very close to the sample average and, hence, provides a reasonable estimate for the mean.

The latter simple example illustrates that the unbiasedness of the development factors' estimators does not guarantee reasonable estimates of the ultimate claims. From an actuarial point of view, the consistency of $\widehat{f}_j^{(n)}$ might be more tempting property of the method. It ensures that a sufficiently large number of observations leads to estimates close to the true quantity. In the claims reserving problem this means that the method provides accurate estimates of the outstanding loss of liabilities.

However, the consistency of an estimator is important property for other reasons as well. Let us present one of them. If $\widehat{f}_j^{(n)}$ is an unbiased estimator of f_j, then this does not imply (and in the majority of cases it is not true) that $[\widehat{f}_j^{(n)}]^{-1}$ is an unbiased estimate of f_j^{-1}. In general, $[\widehat{f}_j^{(n)}]^{-1}$ can behave quite unpredictably. On the other hand, if $\widehat{f}_j^{(n)}$ is a consistent estimator of f_j, then $[\widehat{f}_j^{(n)}]^{-1}$ is a consistent estimator of f_j^{-1}. In general, a *continuous transformation* preserves the property of being consistent estimator (continuous mapping theorem). This is very useful in many applications. For instance, consider the Bornhuetter-Ferguson method (BF),

see Wüthrich and Merz (2008), for reserves estimates. The claims development pattern β_j is sometimes estimated using $\widehat{f}_j^{(n)}$ as

$$\widehat{\beta}_j^{(n)} = \prod_{k=j}^{n-1} \frac{1}{\widehat{f}_k^{(n)}}.$$

Hence, the consistency of $\widehat{f}_j^{(n)}$ implies the consistency of $\widehat{\beta}_j^{(n)}$. On the other hand, the unbiasedness of $\widehat{f}_j^{(n)}$ does not "transfer" to $\widehat{\beta}_j^{(n)}$ in any sense.

Furthermore, an estimate of the *mean squared error (MSE) of reserves* depends on the estimates of development factors and the dependence is not linear (Mack, 1993:Theorem 3). The same holds for the *MSE of prediction*. Therefore, the unbiasedness of $\widehat{f}_j^{(n)}$ does not preserve the unbiasedness for estimates of the MSE of reserves or prediction. Contrary to unbiasedness, the consistency of development factors' estimates $\widehat{f}_j^{(n)}$ also *guarantees* the consistency of the reserves' or prediction's MSE.

Finally, the estimator not only has to stay on target asymptotically but its *variability* (usually measured by variance) has also to shrink, leading to better accuracy. Since the consistency is only a qualitative property of the estimate, it is needed to characterize the consistency of the development factors' estimates from a quantitative point of view. Indeed, the variance of the estimates will provide us a *rate of convergence* of the estimates, as will be pointed out later.

4.1.2 Consistency of the estimates

The *consistency* of the development factors' estimators discussed above is of indisputable importance. In spite of that, it has not been investigated or discussed in the literature according to our best knowledge.

Conditional convergence

In order to formulate and prove the consistency of the chain ladder correctly, we introduce *conditional convergence almost surely*, *in probability* and *in mean square*. For detailed information see, e.g., Belyaev (1995). Suppose that ζ and χ are random variables, and $\{\xi_n\}_{n=1}^{\infty}, \{\zeta_n\}_{n=1}^{\infty}$ are sequences of random variables with a finite mean on a probability space $(\Omega, \mathcal{F}, P)$. Let us define a *conditional probability* given some random variable ζ

$$P_\zeta[\cdot] := E_P[\mathcal{I}(\cdot)|\zeta],$$

where $\mathcal{I}(\cdot)$ is an indicator function.

Definition 4.1.1 (Convergence in conditional probability)**.** To say that ξ_n converges to χ $[P_\zeta]$-almost surely as n tends to infinity with probability one, i.e., $\xi_n \xrightarrow[n\to\infty]{[P_\zeta]\text{-a.s.}} \chi$, $[P]$-a.s. means

$$P\left[P_\zeta\left\{\lim_{n\to\infty}\xi_n = \chi\right\} = 1\right] = 1.$$

To say that ξ_n converges to χ in probability P_{ζ_n} as n tends to infinity with probability one, i.e., $\xi_n \xrightarrow[n\to\infty]{P_{\zeta_n}} \chi$, $[P]$-a.s. means

$$\forall \varepsilon > 0 : P\left[\lim_{n\to\infty} P_{\zeta_n}\{|\xi_n - \chi| \geqslant \varepsilon\} = 0\right] = 1.$$

For $p \geqslant 1$, to say that ξ_n converges to χ in $L_p(P_{\zeta_n})$ as n tends to infinity with probability one, i.e., $\xi_n \xrightarrow[n\to\infty]{L_p(P_{\zeta_n})} \chi$, $[P]$-a.s. means

$$P\left[\lim_{n\to\infty} E_{\zeta_n}|\xi_n - \chi|^p = 0\right] = 1.$$

The conditional convergence in probability and in L_p *along* some sequence of random variables $\{\zeta_n\}_{n=1}^{\infty}$ can be defined, because the concept of these two types of convergence comes from a topology. Despite of that, the almost sure convergence does not correspond to a convergence with respect to any topology and, hence, it is not metrizable. Thereafter, the conditional convergence almost surely cannot be defined along a sequence of random variables, but only *given* one random variable.

Consistency of the chain ladder

Let us define variables ε_{ij} as

$$\varepsilon_{i,j} := C_{i,j+1} - f_j C_{i,j}, \quad 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant n-1. \tag{4.2}$$

If we look at the assumption [1] from a "regression" point of view, then $C_{i,j}$s play the role of the "regressors", and $\varepsilon_{i,j}$ are the "disturbances". Considering $C_{i,j}$ as fixed, the variance of the "responses" $C_{i,j+1}$ as well as the variance of the "errors" $\varepsilon_{i,j}$ equals to $\sigma_j^2 C_{i,j}$. More formally, we have

$$E[\varepsilon_{i,j+1}|C_{i,1},\ldots,C_{i,j}] = 0, \quad \mathrm{Var}[\varepsilon_{i,j+1}|C_{i,1},\ldots,C_{i,j}] = \sigma_j^2 C_{i,j};$$

for all $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant n-1$.

In the following, we investigate the asymptotic properties of the estimated development factors $\widehat{f}_j^{(n)}$. Hence, we assume that the number of accident years n tends to

infinity. For each $n \in \mathbb{N}$, we observe variables $C_{i,j}$, $i = 1, 2, \ldots, n$, $j = 1, \ldots, n+1-i$, and the estimate $\widehat{f}_j^{(n)}$ is calculated from (4.1). The consistency of $\widehat{f}_j^{(n)}$ means that $\widehat{f}_j^{(n)}$ approaches (in some sense) the true value f_j as $n \to \infty$. Equivalently, the difference $\widehat{f}_j^{(n)} - f_j$ converges to zero (in some sense) as $n \to \infty$. According to (4.2), we can express this difference as

$$\widehat{f}_j^{(n)} - f_j = \frac{\sum_{i=1}^{n-j} \left(C_{i,j+1} - f_j C_{i,j} \right)}{\sum_{i=1}^{n-j} C_{i,j}} = \frac{\sum_{i=1}^{n-j} \varepsilon_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}}. \tag{4.3}$$

The term on the right hand side of the equation (4.3) is a random variable and, therefore, several different kinds of convergence can be considered. We will deal with the almost sure convergence, convergence in probability, and L_2 convergence. It is common to work with the conditional probability (and expectation) given the observed data $C_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, n+1-i$ in the claims reserving. Hence, we follow this convention (as it is reasonable from the practical point of view) and we investigate the conditional convergences, as defined in the previous section, instead of the unconditional ones. In particular, for each $n \in \mathbb{N}$ we define $D_j^{(n)} = \{C_{i,k} : k \leqslant j, i \leqslant n-j+1\}$ the set of cumulative claims $C_{i,k}$ observed so far such that the development year k is less or equal to j. Subsequently, we investigate the limiting behavior of (4.3) conditional on $D_j^{(n)}$. However, it is mentioned at the end of Section 4.1.2 that the conditional convergence almost surely cannot be defined along a sequence of random variables, but only given a single random variable. For this reason, the almost sure consistency is studied conditional on the set $D_j = \{C_{i,k} : k \leqslant j, i \in \mathbb{N}\}$ of all the past and future variables $C_{i,k}$ such that $k \leqslant j$.

The following theorem provides the sufficient as well as the necessary condition for the *weak* (in probability) and *strong* (almost sure) *consistency* of the development factors' estimate $\widehat{f}_j^{(n)}$. This condition distinguishes whether the usage of the chain ladder is a consistent approach or not. Furthermore, we show that if this condition holds, then all the mentioned "kinds of consistency" (according to the kind of convergence) are equivalent.

In practice, of course, we always deal only with finite data sets. However, the results of the following theorem apply to the finite data as well ensuring that the number of accident years n is large enough.

Theorem 4.1.1. *Let us consider the chain ladder with assumptions [1]–[3]. Suppose that* $\sum_{i=1}^{\infty} C_{i,j} > 0$ [P]-a.s. *for all* $j \in \mathbb{N}$, *and denote* $D_j^{(n)} = \{C_{i,k} : k \leqslant j, i \leqslant n-j+1\}$ *and* $D_j = \{C_{i,k} : k \leqslant j, i \in \mathbb{N}\}$. *For every* $j \in \mathbb{N}$, *the consequent statements are equivalent:*

(i)

$$\widehat{f}_j^{(n)} \xrightarrow[n\to\infty]{[P_{D_j}]\text{-a.s.}} f_j, \quad [P]\text{-a.s.};$$

(ii)

$$\widehat{f}_j^{(n)} \xrightarrow[n\to\infty]{P_{D_j^{(n)}}} f_j, \quad [P]\text{-a.s.};$$

(iii)

$$\widehat{f}_j^{(n)} \xrightarrow[n\to\infty]{L_2\left(P_{D_j^{(n)}}\right)} f_j, \quad [P]\text{-a.s.};$$

(iv)

$$\sum_{i=1}^{n-j} C_{i,j} \xrightarrow[n\to\infty]{} \infty, \quad [P]\text{-a.s.}$$

Remark 4.1.1. Due to the independence of the different accident years (assumption [3]), the statements *(ii)* and *(iii)* of Theorem 4.1.1 can be equivalently replaced by $\widehat{f}_j^{(n)} \xrightarrow[n\to\infty]{P_{D_j}} f_j$, [P]-a.s. and $\widehat{f}_j^{(n)} \xrightarrow[n\to\infty]{L_2(P_{D_j})} f_j$, [P]-a.s., respectively.

The previous Theorem 4.1.1 postulates a *complete* characterization of the conditional convergence of development factors' estimate, because it gives *necessary* and *sufficient condition (iv)* for the convergence almost surely, in probability, and in mean square. If the condition *(iv)* of Theorem 4.1.1 holds, we refer to the corresponding estimator $\widehat{f}_j^{(n)}$ as *consistent*, dropping the term *conditionally*. As the three types of convergence are equivalent, the type of consistency (strong, weak, L_2) is left out as well.

Note that the assumptions of Theorem 4.1.1 consist exactly of the same assumptions as the classical distribution-free chain ladder approach. The condition $\sum_{i=1}^{\infty} C_{i,j} > 0$ is not restrictive at all, because the assumption [2] implies $C_{i,j} \geqslant 0$ [P]-a.s. for all $i, j \in \mathbb{N}$. It ensures that the estimators $\widehat{f}_j^{(n)}$ are well defined for all $j \in \mathbb{N}$ and all $n \in \mathbb{N}$.

The necessary and sufficient condition for the consistency are studied in more detail in Subsection 4.1.3.

Remark 4.1.2. We have explained above that the conditional consistency is natural to be considered in the claims reserving. However, for the sake of completeness we briefly discuss the unconditional consistency as well. Consider, for instance, the L_2

convergence. It follows from the proof of Theorem 4.1.1 that $\widehat{f}_j^{(n)}$ converges to f_j in L_2 (unconditionally) as $n \to \infty$ if and only if

$$E\left[\frac{1}{\sum_{i=1}^{n-j} C_{i,j}}\right] \to 0, \quad n \to \infty.$$

However, this condition is obviously more complicated than the condition *(iv)* in Theorem 4.1.1, and it is practically unverifiable. Thus, the conditional convergence is not only more natural one in this case, but even more convenient one.

4.1.3 Rate of convergence

Consistency of an estimator is a very important but only qualitative property. If we want to measure consistency, it is necessary to evaluate its rate. This quantitative attribute can be obtained directly from the proof of Theorem 4.1.1 via the *mean square error of the development factor's estimate*.

Proposition 4.1.2. *Consider the assumptions of Theorem 4.1.1 and let $j \in \mathbb{N}$ be fixed. Denote the conditional mean square error of the estimate of development factor f_j as*

$$\mathrm{MSE}\left(\widehat{f}_j^{(n)}\right) := E\left\{\left[\widehat{f}_j^{(n)} - E\left(\widehat{f}_j^{(n)}\right)\right]^2 \middle| D_j^{(n)}\right\}.$$

Then, with probability one holds

$$\mathrm{MSE}\left(\widehat{f}_j^{(n)}\right) = \mathcal{O}\left(\left[\sum_{i=1}^{n-j} C_{i,j}\right]^{-1}\right), \quad n \to \infty.$$

Crucial interpretation of previous Proposition 4.1.2 is as follows: the slower (faster) divergence of $\sum_{i=1}^{\infty} C_{i,j}$ implies the slower (faster) realization of consistency of the development factors' estimates.

On the necessary and sufficient condition

Let us investigate the necessary and sufficient condition *(iv)* from Theorem 4.1.1

$$\sum_{i=1}^{n-j} C_{i,j} \xrightarrow[n\to\infty]{} \infty, \quad [P]\text{-a.s.} \tag{4.4}$$

in more detail. First, recall that variables $C_{i,j}$ are called *summable* [P]-a.s. if $\sum_{i=1}^{n} C_{i,j}$ converges [P]-a.s. to a finite variable as $n \to \infty$. The assumption [2] of the

distribution-free approach implies that $C_{i,j} \geqslant 0$ [P]-a.s. for all $i, j \in \mathbb{N}$ and, therefore, it is not restrictive to assume that $C_{i,j} \geqslant 0$ for all $i, j \in \mathbb{N}$. In this case, $C_{i,j}$ are either summable [P]-a.s. or condition (4.4) holds.

Proposition 4.1.3. *Consider the chain ladder with assumptions [1]–[3]. Suppose that* $C_{i,j} \geqslant 0$ *for all* $i, j \in \mathbb{N}$, *and* $\sum_{i=1}^{\infty} C_{i,j} > 0$ [P]-a.s. *for all* $j \in \mathbb{N}$. *Let* $j \in \mathbb{N}$ *be fixed. Then the following conditions are equivalent.*

(i) The condition (4.4) *holds.*

(ii)
$$\sum_{i=1}^{\infty} \mathsf{E}\, C_{i,1} = \infty. \tag{4.5}$$

(iii) The condition (4.4) *holds for* $j_0 \in \mathbb{N}$, $j \neq j_0$.

Corollary 4.1.4. *Consider the assumptions of Proposition 4.1.3 and let* $j \in \mathbb{N}$ *be fixed.*

(i) The estimator $\widehat{f}_j^{(n)}$ *is consistent for* f_j *if and only if* $\sum_{i=1}^{\infty} \mathsf{E}\, C_{i,1} = \infty$.

(ii) The estimator $\widehat{f}_j^{(n)}$ *is consistent for* f_j *if and only if* $\sum_{i=1}^{n} C_{i,1} \to \infty$ [P]-a.s. *as* $n \to \infty$.

It follows from Corollary 4.1.4 that either $\widehat{f}_j^{(n)}$ is consistent for f_j for all $j \in \mathbb{N}$, or none of them is consistent. Furthermore, the consistency of $\widehat{f}_j$ depends only on the asymptotic behavior of the cumulative sums of $\mathsf{E}\, C_{i,1}$ (or $C_{i,1}$) and there is no effect of the true values of f_j, $j \in \mathbb{N}$ on the property of being a consistent estimator or not.

The condition *(i)* of Corollary 4.1.4 is better understandable compared to (4.4), because it is formulated in terms of divergence of series of real numbers instead of divergence of random variables. The condition *(ii)* of Corollary 4.1.4 is useful for practice. Here, we observe just one data set and, therefore, condition (4.5) cannot be verified. On the other hand, we can deduce conclusions about behavior of $\sum_{i=1}^{n} C_{i,1}$ from the observations so far.

4.2 Generalized estimating equations in triangles

A common approach to the claims reserving problem is based on generalized linear models (GLM), where the claims in different origin and development years are assumed to be independent variables. If this is violated, the classical techniques may

provide incorrect predictions of the claims reserves or even misleading estimates of the prediction error. In this section, the application of generalized estimating equations (GEE) for estimation of the claims reserves is shown as already demonstrated by Hudecová and Pešta (2013).

All the classical approaches are based on the assumption that the claim amounts in different years are independent variables. However, this assumption can be sometimes unrealistic or at least questionable. It has been pointed out that methods, which enable *modeling the dependencies*, are needed, cf. Antonio et al. (2006) or Antonio and Beirlant (2007). The mentioned papers suggest the generalized linear mixed models (GLMM) to handle the possible dependence among the incremental claims in successive development years. This approach extends the classical GLM and is frequently used in panel (longitudinal) data analyses. The GLMM approach was proposed in a way that more granular data are required (i.e., claim-by-claim data). In this section, we present the use of another possible extension of GLM, namely the *generalized estimating equations* (GEE) method. On the contrary to the above mentioned GLMM method, the GEE require just a simple triangle of paid losses.

GEE were introduced by Liang and Zeger (1986) as a method for estimating model parameters if the independence assumption is violated. The primary interest of the analysis is to model the marginal expectation of the response variable given the covariates. In contrast to GLMM, this method does not explicitly model the correlation structure. The associations are treated as nuisance parameters and modeled using so called "working correlation matrices". The method yields consistent and asymptotically normal parameter estimates even though the correlation structure is misspecified (Ziegler, 2011:Sec. 5.2). In addition, *no additional distributional assumptions* are required, compared to a specific probability distribution for the outcome in the GLM (or even GLMM) framework. The GEE solely assume that the distribution belongs to the exponential family.

Outstanding loss liabilities are again structured in the claims development triangles. Let us denote $X_{i,j}$ all the claim amounts in development year j with accident year i. Therefore, $X_{i,j}$ stands for the *incremental claims* in accident year i made in accounting year $i+j$. The current year is n, which corresponds to the most recent accident year and development period as well. That is, our data history consists of right-angled isosceles triangle $\{X_{i,j}\}$, where $i = 1, \ldots, n$ and $j = 1, \ldots, n+1-i$.

Suppose that $Y_{i,j}$ are *cumulative payments* or *cumulative claims* in origin year i after j development periods, i.e., $Y_{i,j} = \sum_{k=1}^{j} X_{i,k}$ (cf. Table 4.1, where the cumulative claims are denoted by $C_{i,j}$). Hence, $Y_{i,j}$ is a random variable of which we have an observation if $i+j < n+1$ (a run-off triangle). The aim is to estimate the ultimate

claims amount $Y_{i,n}$ and the outstanding *claims reserve* $R_i^{(n)} = Y_{i,n} - Y_{i,n+1-i}$ for all $i = 2, \dots, n$.

4.2.1 GEE

Run-off triangles are comprised by observations which are ordered in time. It is therefore natural to suspect the observations to be correlated. Probably the most natural approach is to assume that the observations of a common accident year are correlated—they form a *cluster*. On the other hand, observations of different accident years are supposed to be independent. This assumption is similar to those of the Mack's chain ladder model, cf. Mack (1993).

Let us consider that the incremental claims for the accident year $i \in \{1, \dots, n\}$ create an $(n-i+1) \times 1$ vector $\boldsymbol{X}_i = [X_{i,1}, \dots, X_{i,n-i+1}]^\top$. It is assumed that the vectors $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are independent, but the components of $\boldsymbol{X}_i$ are allowed to be correlated. Hence, the claim triangle can be considered as a specific type of *panel data* with accident year (row) clusters.

In the next subsections we explain the main principles of GEE and give some recommendation for the use within the claims reserving. We refer to Hardin and Hilbe (2003) and Ziegler (2011) for further reading on this topic.

Three pillars of GEE

Denote the expectation of $\boldsymbol{X}_i$ as

$$\mathsf{E}\,\boldsymbol{X}_i = \boldsymbol{\mu}_i = [\mu_{i,1}, \dots, \mu_{i,n-i+1}]^\top.$$

Suppose that accident year i and development year j influence the expectation of claim amount via so-called *link function* g in the following manner:

$$\mu_{i,j} = g^{-1}(\eta_{i,j}) = g^{-1}(\boldsymbol{z}_{i,j}^\top \boldsymbol{\theta}), \tag{4.6}$$

where g^{-1} is the inverse of scalar link function g and $\boldsymbol{z}_{i,j}$ is a $p \times 1$ vector of dummy covariates that arranges the impact of accident and development year on the claim amounts through model parameters $\boldsymbol{\theta} \in \mathbb{R}^{p \times 1}$. The relation (4.6) defines the *linear predictor* $\eta_i = \boldsymbol{z}_{i,j}^\top \boldsymbol{\theta}$, which together with the link function g fully specifies the *mean structure* $\boldsymbol{\mu}_i$.

Besides the mean structure, one needs to specify the *variance* of claim amounts. Assume that the variance of the incremental claim amount $X_{i,j}$ can be expressed as a *known function* h of its expectations $\mu_{i,j}$:

$$\operatorname{Var} X_{i,j} = \phi h(\mu_{i,j}), \tag{4.7}$$

where $\phi > 0$ is a *scale* or a *dispersion parameter*. In connection to the GLM, if $X_{i,j}$ followed the Poisson distribution (or the overdispersed Poisson distribution), then h would be an identity, i.e., $h(x) = x$. For the gamma distribution, $h(x) = x^2$, etc. The relation (4.7) defines so-called *variance function*.

In the GEE framework, it is *not necessary* to specify the *whole distribution* of the data, because the method is quasi-likelihood based. Only the mean structure and the mean-variance relationship need to be defined. Furthermore, the correlation between the components of $\boldsymbol{X}_i$ is modeled using a *working correlation matrix* $\boldsymbol{C}_i(\boldsymbol{\vartheta}) \in \mathbb{R}^{(n-i+1)\times(n-i+1)}$, which depends only on an $s \times 1$ vector of unknown parameters $\boldsymbol{\vartheta}$, which is the same for all the accident years i. Consequently, the *working covariance matrix* of the incremental claims is

$$\boldsymbol{V}_i = \operatorname{Cov} \boldsymbol{X}_i = \phi \boldsymbol{A}_i^{1/2} \boldsymbol{C}_i(\boldsymbol{\vartheta}) \boldsymbol{A}_i^{1/2}, \tag{4.8}$$

where $\boldsymbol{A}_i$ is an $(n-i+1) \times (n-i+1)$ diagonal matrix with $h(\mu_{i,j})$ as the jth diagonal element. The name "working" comes from the fact that the structure of $\boldsymbol{C}_i$ does not need to be correctly specified. Some commonly used correlation structures are described in Section 4.2.1.

To sum up, the GEE framework has two pillars common with the GLM framework (linear predictor and link function). However, the *third pillar is different*: The GEE approach does not require any specification of the whole distribution for the outcome as this is the case for the GLM. On contrary, the GEE only assume that the (unknown) distribution belongs to the exponential family of probability distributions and the third pillar consists of specification of the *variance-covariance structure* (variance function and working correlation matrix).

Estimation in GEE

The generalized estimating equations are formed via quasi-score vector

$$\boldsymbol{u}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_i),$$

where $\boldsymbol{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\theta} \equiv \{\partial\mu_{i,j}/\partial\theta_k\}_{j,k=1}^{n-i+1,p}$. For given estimates $(\widehat{\phi}, \widehat{\boldsymbol{\vartheta}})$ of $(\phi, \boldsymbol{\vartheta})$, the estimate of parameter $\boldsymbol{\theta}$ solves the equation $\boldsymbol{u}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{0}$. The parameters $(\phi, \boldsymbol{\vartheta})$ are usually estimated by the moment estimates. The fitting algorithm is therefore iterative: updating the estimate of $\boldsymbol{\theta}$ in one step and re-estimating $(\phi, \boldsymbol{\vartheta})$ in the second step.

The procedure yields a consistent and asymptotically normal estimate of $\boldsymbol{\theta}$ even though the correlation matrix $\boldsymbol{C}_i(\boldsymbol{\vartheta})$ is misspecified, see Liang and Zeger (1986). The empirically corrected variance estimates for $\widehat{\boldsymbol{\theta}}$ can be obtained using the so-called *sandwich estimate*

$$\boldsymbol{\Sigma}_{\widehat{\theta}} \equiv \widehat{\mathrm{Cov}}\widehat{\boldsymbol{\theta}} = \boldsymbol{B}^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{S}(\widehat{\boldsymbol{\theta}})\boldsymbol{B}^{-1}(\widehat{\boldsymbol{\theta}}), \tag{4.9}$$

where

$$\boldsymbol{B} = \sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{D}_i, \quad \boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_i)(\boldsymbol{X}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \tag{4.10}$$

are evaluated at $\widehat{\boldsymbol{\theta}}$. The matrix $\boldsymbol{B}^{-1}(\widehat{\boldsymbol{\theta}})$ is referred to as a *model based* estimator of the variance matrix of $\widehat{\boldsymbol{\theta}}$. The estimator $\boldsymbol{\Sigma}_{\widehat{\theta}}$ is consistent for $\mathrm{Cov}\,\widehat{\boldsymbol{\theta}}$ even if the correlation matrix $\boldsymbol{C}_i$ is misspecified. However, it can be slightly biased in small samples.

Covariance structure

Although the GEE method is robust to a misspecification of the correlation structure, selection of the working correlation structure, which is closer to the true one, may lead to more efficient estimates of $\boldsymbol{\theta}$.

There exist several common choices for the working correlation matrix. The simplest case is to assume *uncorrelated* (or *independent*) incremental claims, i.e., $\boldsymbol{C}_i(\boldsymbol{\vartheta}) = \boldsymbol{I}_{n-i+1} = \{\delta_{j,k}\}_{j,k=1}^{n-i+1,n-i+1}$, where $\delta_{j,k}$ symbolizes the Kronecker's delta being 1 for $j = k$ and 0 otherwise. The opposite extreme case is an *unstructured* correlation matrix $\boldsymbol{C}_i(\boldsymbol{\vartheta}) = \{\vartheta_{j,k}\}_{j,k=1}^{n-i+1,n-i+1}$ such that $\vartheta_{j,j} = 1$ for $j = 1, \ldots, n-1+1$ and $\boldsymbol{C}_i(\boldsymbol{\vartheta})$ is positive definite. As a compromise to these two extreme cases, one can consider an *exchangeable* correlation structure

$$\boldsymbol{C}_i(\boldsymbol{\vartheta}) = \{\delta_{j,k} + (1-\delta_{j,k})\vartheta\}_{j,k=1}^{n-i+1,n-i+1}, \quad \boldsymbol{\vartheta} = [\vartheta, \ldots, \vartheta]^\top;$$

an m-*dependent* correlation structure $\boldsymbol{C}_i(\boldsymbol{\vartheta}) = \{c_{j,k}\}_{j,k=1}^{n-i+1,n-i+1}$,

$$c_{j,k} = \begin{cases} 1, & j = k, \\ \vartheta_{|j-k|}, & 0 < |j-k| \leqslant m, \quad \boldsymbol{\vartheta} = \{\vartheta_l\}_{l=1}^{m}, \\ 0, & |j-k| > m; \end{cases}$$

or an *autoregressive AR(1)* correlation structure

$$\boldsymbol{C}_i(\boldsymbol{\vartheta}) = \{\vartheta^{|j-k|}\}_{j,k=1}^{n-i+1,n-i+1}, \quad \boldsymbol{\vartheta} = [\vartheta, \ldots, \vartheta]^\top.$$

4.2.2 Application of the GEE to claims reserving

In the claims reserving, the link function is usually chosen as the logarithm, see, e.g., Wüthrich and Merz (2008). The most common mean structure assumes that

$$\log(\mu_{i,j}) = \gamma + \alpha_i + \beta_j, \tag{4.11}$$

where α_i stands for the effect of accident year i, β_j represents the effect of the development year j, and γ is so-called baseline parameter corresponding a value for the first accident and development year (taking $\alpha_1 = 0 = \beta_1$). In this case $\boldsymbol{\theta} = [\gamma, \alpha_2, \ldots, \alpha_n, \beta_2, \ldots, \beta_n]$ and

$$\boldsymbol{z}_{i,j} = [1, \delta_{2,i}, \ldots, \delta_{n,i}, \delta_{2,j}, \ldots, \delta_{n,j}]^\top.$$

Another common model, the Hoerl curve with the logarithmic link function, can be coded by design matrix

$$\boldsymbol{z}_{i,j} = [1, \delta_{2,i}, \ldots, \delta_{n,i}, 1 \times \delta_{2,j}, \ldots, n \times \delta_{n,j}, \delta_{2,j} \times \log 2, \ldots, \delta_{n,j} \times \log n]^\top$$

and parameters of interest $\boldsymbol{\theta} = [\gamma, \alpha_2, \ldots, \alpha_n, \beta_2, \ldots, \beta_n, \lambda_2, \ldots, \lambda_n]^\top$. Afterwards, $\log(\mu_{i,j}) = \gamma + \alpha_i + j\beta_j + \lambda_j \log j$, where again $\alpha_1 = \beta_1 = \lambda_1 = 0$. For some other possible mean structures in claims reserving, see Björkwall et al. (2011).

The choice of the variance function is somehow analogous to the specification of the distribution in the GLM. Hence, suitable variance functions for the claims reserving purposes are: linear (its quasi-score vector corresponds to the score vector of the overdispersed Poisson distribution) or quadratic (gamma distribution). The variance function as a non-integer power of the mean (multiplied by the scaling parameter) can also be a practical choice if one realizes the concordance with the Tweedie distribution (Tweedie, 1984). This distribution has been recently proven as suitable one for the claims reserving (Wütrich, 2003) and can be considered within the GEE as well.

Finally, one needs to choose an appropriate working correlation structure. The most feasible choice might be *AR(1)* since the observations within an accident year are *ordered in time* and, in such situations, it is natural that the correlation between two observations decays with their time distance. However, in situations, where the observations are strongly dependent—that is the decay of the correlations is

slower than it is in AR(1)—the *exchangeable structure* could be considered as a good guess as well. Finally, the *independence structure* should always be considered for a comparison. For more details regarding the choice of particular working correlation matrix, we refer to, e.g., (Hardin and Hilbe, 2003:Subsection 3.2.1). This approach combined with the sandwich estimate of the covariance matrix of parameter estimates $\widehat{\boldsymbol{\theta}}$ may lead to satisfactory results as well, for small data sets in particular (Hardin and Hilbe, 2003:Chap. 4).

4.2.3 Model selection

Similarly as in the GLM setting, two nested models (nested in the mean structure) can be compared using Wald tests, see Hardin and Hilbe ((2003):Sec. 4.5.2). A comparison of two non-nested models in the GLM framework can be based on information criteria as AIC or BIC, see, e.g., Björkwall et al. (2011). However, since the GEE method is only quasi-likelihood based (and not full likelihood), these criteria cannot be used within the GEE.

Pan (2001) suggested an analogy of the AIC for GEE, namely *quasi-likelihood under the independence model criterion* (QIC). The QIC is defined as

$$\text{QIC} = -2\text{Q}(\widehat{\boldsymbol{\theta}}, \boldsymbol{I}) + 2\text{trace}(\widehat{\boldsymbol{\Omega}}_{\text{I}}(\widehat{\boldsymbol{\theta}})\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}),$$

where $\text{Q}(\cdot, \boldsymbol{I})$ is the quasi-likelihood under working independence model, see McCullagh and Nelder ((1989):p. 325), and $\widehat{\boldsymbol{\Omega}}_{\text{I}}(\boldsymbol{\theta}) = \sum_{\text{i}=1}^{\text{n}} \boldsymbol{D}_{\text{i}}^{\top} \boldsymbol{A}_{\text{i}}^{-1} \boldsymbol{D}_{\text{i}}$. A model with a *smaller QIC* value indicates a *better fit* to the data. The QIC equals to AIC (up to a constant) under the independence in cases when the model implies the proper likelihood.

Hardin and Hilbe (2003) considered a modified version of QIC,

$$\text{QIC}_{\text{HH}} = -2\text{Q}(\widehat{\boldsymbol{\theta}}, \boldsymbol{I}) + 2\text{trace}(\widetilde{\boldsymbol{\Omega}}_{\text{I}}(\widehat{\boldsymbol{\theta}}(\boldsymbol{I}))\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}),$$

where $\widetilde{\boldsymbol{\Omega}}_{\text{I}}(\boldsymbol{\theta}) = \sum_{\text{i}=1}^{\text{n}} \boldsymbol{D}_{\text{i}}^{\top} \boldsymbol{A}_{\text{i}}^{-1} \boldsymbol{D}_{\text{i}}$ is evaluated at the estimate $\widehat{\boldsymbol{\theta}}(\boldsymbol{I})$ obtained by GEE with the independence working correlation structure. The main advantage of this modification is that QIC_{HH} can be easily computed, because matrices $\widetilde{\boldsymbol{\Omega}}_{\text{I}}(\widehat{\boldsymbol{\theta}}(\boldsymbol{I}))$ and $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ are provided by standard software packages for the GEE estimation.

The two criteria can be used for choosing the appropriate mean structure as well as the working correlation matrix. However, simulations have shown that QIC tends to be more sensitive to changes in the mean structure than changes in the covariance structure, see Hin and Wang (2009). For this reason, Hin and Wang (2009) suggested

a *correlation information criterion* (CIC), which improves the performance of QIC for selecting the appropriate working correlation structure. The CIC is defined as

$$\mathrm{CIC} = \mathsf{trace}(\widehat{\boldsymbol{\Omega}}_{\mathrm{I}}(\widehat{\boldsymbol{\theta}})\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}).$$

Analogously, its modification defined as

$$\mathrm{CIC}_{\mathrm{HH}} = \mathsf{trace}(\widetilde{\boldsymbol{\Omega}}_{\mathrm{I}}(\widehat{\boldsymbol{\theta}}(\boldsymbol{I}))\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}})$$

can be used for the comparison of working correlation structures as well.

4.2.4 Mean square error of prediction

In order to quantify the precision of the estimates and predictions, let us define the *mean square error* (MSE) of prediction for the ith claims reserve

$$\begin{aligned}\mathsf{MSE}\left[\widehat{R}_i^{(n)}\right] &:= \mathsf{E}\left[\widehat{R}_i^{(n)} - R_i^{(n)}\right]^2 = \mathsf{E}\left[\sum_{j=n+2-i}^{n}\left(\widehat{X}_{i,j}^{(n)} - X_{i,j}\right)\right]^2 \\ &= \sum_{j=n+2-i}^{n}\mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]^2 + \sum_{\substack{j,k=n+2-i \\ j\neq k}}^{n}\mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]\left[\widehat{X}_{i,k}^{(n)} - X_{i,k}\right], \qquad (4.12)\end{aligned}$$

where $\widehat{X}_{i,j}^{(n)} = g^{-1}\left(\boldsymbol{z}_{i,j}^{\top}\widehat{\boldsymbol{\theta}}\right)$ is the *plug-in prediction* of the incremental claim amounts $X_{i,j}$ based on the GEE estimate $\widehat{\boldsymbol{\theta}}$.

Mean square error in the GEE

Our aim is to derive the MSE of prediction for the claims reserves within the GEE framework. Elaborating the expected value from the first sum in (4.12) yields

$$\begin{aligned}\mathsf{MSE}\left[\widehat{X}_{i,j}^{(n)}\right] &:= \mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]^2 = \mathsf{Var}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right] + \left(\mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]\right)^2 \\ &= \mathsf{Var}\,\widehat{X}_{i,j}^{(n)} - 2\,\mathsf{Cov}\left(\widehat{X}_{i,j}^{(n)}, X_{i,j}\right) + \mathsf{Var}\,X_{i,j} + \left(\mathsf{E}\,\widehat{X}_{i,j}^{(n)} - \mathsf{E}\,X_{i,j}\right)^2. \qquad (4.13)\end{aligned}$$

Many authors directly assume the unbiasedness or approximate unbiasedness of estimator $\widehat{X}_{i,j}^{(n)}$ for $\mathsf{E}\,X_{i,j}$, that is $\mathsf{E}\,\widehat{X}_{i,j}^{(n)} = \mathsf{E}\,X_{i,j}$ or $\mathsf{E}\,\widehat{X}_{i,j}^{(n)} \approx \mathsf{E}\,X_{i,j}$. See, for instance, Renshaw (1994), England and Verrall ((2002):Subsec. 7.1.2), or Wüthrich and Merz ((2008):Sec. 3.1). Nevertheless, this is neither the case for the GLM nor the GEE, because a non-linear link function (e.g., logarithm) makes *biased prediction* from approximately unbiased parameter estimates—especially in small samples—due to the non-exchangeability of the expectation operator and the link function.

The conjecture of the (approximately) unbiased predictor then implies that the MSE of prediction for incremental claims is given as $\mathsf{MSE}\left[\widehat{X}_{i,j}^{(n)}\right] \equiv \mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]^2 \approx \mathsf{Var}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]$. That is, the MSE is reduced to the variance of the difference between observation and its prediction. However, the unbiasedness of $\widehat{X}_{i,j}^{(n)}$ can be arguable (or even unrealistic), for smaller samples in particular.

If the prediction is really unbiased and the incremental claim amounts are independent, then the MSE of prediction is equal to the process variance plus the estimation variance, see, e.g., Wüthrich and Merz ((2008):Sec. 3.1). On the other hand, violation of such strict assumptions could provide incorrect MSE of prediction, because it simply ignores the covariance or the squared bias term in (4.13).

Nevertheless, such simplification cannot be applied in the GEE framework, because *the incremental claim amounts* are *not independent*. Hence, the covariance among a future observation and its predictor in (4.13) is not zero anymore, because the predictor is a function of the past observations, which are not independent of the future observation. And this needs to be taken into account in the calculation of the MSE.

For $i = 2, \dots, n$ define $\vec{\boldsymbol{X}}_i = [X_{i,n+2-i}, \dots, X_{i,n}]^\top$ as the vector of the unobserved claim amounts of accident year i. Similarly, an arrow above a vector/matrix stands for its complement for the unobserved data $\{X_{i,j}\}$, $i = 2, \dots, n$ and $j = n+2-i, \dots, n$ (bottom-right right-angled isosceles triangle). For instance, $\vec{\boldsymbol{\mu}}_i = \mathsf{E}\,\vec{\boldsymbol{X}}_i$ stands for the expectation of the future claims of accident year i, $\widehat{\vec{\boldsymbol{X}}}_i^{(n)}$ is the prediction of $\vec{\boldsymbol{X}}_i$, $\vec{\boldsymbol{D}}_i = \partial\vec{\boldsymbol{\mu}}_i/\partial\boldsymbol{\theta}$, etc.

Consider a first-order stochastic Taylor expansion (Brockwell and Davis, 2006:Proposition 6.1.6), for the residual vector $\vec{\boldsymbol{r}}_i := \vec{\boldsymbol{X}}_i - \widehat{\vec{\boldsymbol{X}}}_i^{(n)}$ around $\boldsymbol{\theta}$. It gives

$$\vec{\boldsymbol{r}}_i = \vec{\boldsymbol{e}}_i + \left[\frac{\partial\vec{\boldsymbol{e}}_i}{\partial\boldsymbol{\theta}}\right](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|), \tag{4.14}$$

where $\vec{\boldsymbol{e}}_i \equiv \vec{\boldsymbol{e}}_i(\boldsymbol{\theta}) = \vec{\boldsymbol{X}}_i - \vec{\boldsymbol{\mu}}_i$. Notice that $\partial\vec{\boldsymbol{e}}_i/\partial\boldsymbol{\theta} = -\vec{\boldsymbol{D}}_i$. Previous linearization is reasonable, because under the regularity conditions for quasi-likelihood estimation in the GEE framework postulated by White (1982) and Ziegler ((2011):Sec. 5.2), the quasi-likelihood GEE estimate $\widehat{\boldsymbol{\theta}}$ is strongly consistent for the parameter $\boldsymbol{\theta}$.

The MSE of $\widehat{\vec{\boldsymbol{X}}}_i^{(n)}$ can be calculated using residuals $\vec{\boldsymbol{r}}_i$ and (4.14) as

$$\mathsf{MSE}\left[\widehat{\vec{\boldsymbol{X}}}_i^{(n)}\right] = \mathsf{E}\left[\vec{\boldsymbol{r}}_i\vec{\boldsymbol{r}}_i^\top\right] \approx \mathsf{E}\left[\vec{\boldsymbol{e}}_i\vec{\boldsymbol{e}}_i^\top\right] - \mathsf{E}\left[\vec{\boldsymbol{e}}_i(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top\vec{\boldsymbol{D}}_i^\top\right] - \mathsf{E}\left[\vec{\boldsymbol{D}}_i(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\vec{\boldsymbol{e}}_i^\top\right]$$

$$+ \mathsf{E}\left[\vec{\boldsymbol{D}}_i(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^\top \vec{\boldsymbol{D}}_i^\top\right]. \tag{4.15}$$

The Taylor expansion applied on $\boldsymbol{u}(\widehat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$ together with the chain rule provide a first-order approximation

$$\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta} \approx \left(\sum_{l=1}^{n} \boldsymbol{D}_l^\top \boldsymbol{V}_l^{-1}\boldsymbol{D}_l\right)^{-1} \sum_{j=1}^{n} \boldsymbol{D}_j^\top \boldsymbol{V}_j^{-1}\boldsymbol{e}_j.$$

For a detailed derivation see Ziegler ((2011):Sec. 5.2). This approximation together with the independence of accident years i and j, $i \neq j$, imply that (4.15) can be further expressed as

$$\begin{aligned}
\mathsf{MSE}\left[\widehat{\vec{\boldsymbol{X}}}_i^{(n)}\right] &\approx \mathsf{Cov}\,\vec{\boldsymbol{X}}_i - \mathsf{Cov}\left(\vec{\boldsymbol{X}}_i, \boldsymbol{X}_i\right)\boldsymbol{H}_{ii}^\top - \boldsymbol{H}_{ii}\,\mathsf{Cov}\left(\boldsymbol{X}_i, \vec{\boldsymbol{X}}_i\right) \\
&\quad + \sum_{j=1}^{n} \boldsymbol{H}_{ij}\,\mathsf{Cov}\,\boldsymbol{X}_j \boldsymbol{H}_{ij}^\top \qquad (4.16)\\
&= \mathsf{Cov}\,\vec{\boldsymbol{X}}_i - 2\,\mathsf{Cov}\left(\vec{\boldsymbol{X}}_i, \boldsymbol{X}_i\right)\boldsymbol{H}_{ii}^\top + \vec{\boldsymbol{D}}_i \boldsymbol{B}^{-1}[\mathsf{E}\,\boldsymbol{S}]\boldsymbol{B}^{-1}\vec{\boldsymbol{D}}_i^\top, \qquad (4.17)
\end{aligned}$$

where $\boldsymbol{H}_{ij} = \vec{\boldsymbol{D}}_i \left(\sum_{l=1}^{n} \boldsymbol{D}_l^\top \boldsymbol{V}_l^{-1}\boldsymbol{D}_l\right)^{-1} \boldsymbol{D}_j^\top \boldsymbol{V}_j^{-1}$ and the matrices $\boldsymbol{B}$ and $\boldsymbol{S}$ are defined in (4.10).

Estimate for the MSE of prediction

One of the main goals is to estimate the theoretical MSE of prediction for claims reserves. This means to find a proper estimate for the left hand side of approximation (4.16). Indeed, comparing relations (4.12) and (4.15) gives

$$\mathsf{MSE}\left[\widehat{\mathsf{R}}_i^{(n)}\right] = \underbrace{[1,\ldots,1]}_{(i-1)\times 1}\,\mathsf{MSE}\left[\widehat{\vec{\boldsymbol{X}}}_i^{(n)}\right][1,\ldots,1]^\top.$$

The core problem lies in the estimation of the covariances in (4.17). The remaining terms from (4.17) can be estimated straightforwardly using the plug-in estimates, i.e.,

$$\begin{aligned}
&\widehat{\boldsymbol{A}}_i := \boldsymbol{A}_i(\widehat{\boldsymbol{\theta}}), \quad \widehat{\boldsymbol{C}}_i := \boldsymbol{C}_i(\widehat{\boldsymbol{\vartheta}}), \quad \widehat{\boldsymbol{V}}_i := \widehat{\phi}\,\widehat{\boldsymbol{A}}_i^{1/2}\widehat{\boldsymbol{C}}_i\widehat{\boldsymbol{A}}_i^{1/2}, \\
&\widehat{\boldsymbol{D}}_i := \boldsymbol{D}_i(\widehat{\boldsymbol{\theta}}), \quad \widehat{\vec{\boldsymbol{D}}}_i := \vec{\boldsymbol{D}}_i(\widehat{\boldsymbol{\theta}}).
\end{aligned}$$

The covariance of the observed incremental claim amounts can be estimated as suggested by Liang and Zeger (1986):

$$\widehat{\mathrm{Cov}}\boldsymbol{X}_i = (\boldsymbol{X}_i - \widehat{\boldsymbol{X}}_i)(\boldsymbol{X}_i - \widehat{\boldsymbol{X}}_i)^\top.$$

It follows from (4.9), that the last term of (4.17) can be estimated by $\widehat{\vec{\boldsymbol{D}}}_i \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}} \widehat{\vec{\boldsymbol{D}}}_i^\top$. Furthermore, the variance structure in (4.8) implies that the covariance of the unobserved (future) incremental claim amounts may be estimated as

$$\widehat{\mathrm{Cov}}\vec{\boldsymbol{X}}_i = \widehat{\phi}\,\widehat{\vec{\boldsymbol{A}}}_i^{1/2}\widehat{\vec{\boldsymbol{C}}}_i\widehat{\vec{\boldsymbol{A}}}_i^{1/2},$$

where $\widehat{\vec{\boldsymbol{A}}}_i = \mathrm{diag}\{h(\mu_{i,j}(\widehat{\boldsymbol{\theta}})),\, j = n+2-i,\ldots,n\}$ and $\widehat{\vec{\boldsymbol{C}}}_i = \boldsymbol{C}_{n+2-i}(\widehat{\boldsymbol{\vartheta}})$ for the standard correlation structures as AR(1), MA(1), independence, or exchangeable (i.e., correlation structures with the translation symmetry property). If a different correlation structure is used, then the future correlations $\vec{\boldsymbol{C}}_i$ have to be predefined in advance and estimated according to that.

Another way how to look at the covariances from (4.16) is to consider a joint vector of the past and future incremental claim amounts for a particular accident year. Hence,

$$\mathrm{Cov}[\boldsymbol{X}_i^\top, \vec{\boldsymbol{X}}_i^\top]^\top = \begin{bmatrix} \mathrm{Cov}\,\boldsymbol{X}_i & \mathrm{Cov}\left(\boldsymbol{X}_i, \vec{\boldsymbol{X}}_i\right) \\ \mathrm{Cov}\left(\vec{\boldsymbol{X}}_i, \boldsymbol{X}_i\right) & \mathrm{Cov}\,\vec{\boldsymbol{X}}_i \end{bmatrix} = \phi\tilde{\boldsymbol{A}}_i^{1/2}\tilde{\boldsymbol{C}}_i\tilde{\boldsymbol{A}}_i^{1/2}, \quad (4.18)$$

where $\tilde{\boldsymbol{A}}_i = \mathrm{diag}\{h(\mu_{i,j}(\boldsymbol{\theta})),\, j = 1,\ldots,n\} \equiv \mathrm{diag}\{\mathrm{diag}(\boldsymbol{A}_i), \mathrm{diag}(\vec{\boldsymbol{A}}_i)\}$, i.e., joint diagonals from $\boldsymbol{A}_i$ and $\vec{\boldsymbol{A}}_i$ are placed on the diagonal of matrix $\tilde{\boldsymbol{A}}_i$. The correlation matrix $\tilde{\boldsymbol{C}}_i$ is an extension of the original correlation matrix $\boldsymbol{C}_i$ for the standard correlation structures as above, or needs to be known in advance.

Henceforth, the estimate of covariance among the past and future incremental claim amounts can easily be taken from (4.18), i.e.,

$$\widehat{\mathrm{Cov}}\left(\vec{\boldsymbol{X}}_i, \boldsymbol{X}_i\right) = \widehat{\phi}\,\widehat{\vec{\boldsymbol{A}}}_i^{1/2}\widehat{\vec{\boldsymbol{C}}}_i\widehat{\boldsymbol{A}}_i^{1/2},$$

where $\bar{\boldsymbol{C}}_i$ is a lower-left segment of the correlation matrix $\tilde{\boldsymbol{C}}_i$ corresponding to $\mathrm{Cov}\left(\vec{\boldsymbol{X}}_i, \boldsymbol{X}_i\right)$, i.e., $\bar{\boldsymbol{C}}_i = \{\tilde{\boldsymbol{C}}_{i;j,k}\}_{j=n+2-i,k=1}^{n,n+1-i}$.

In order to calculate the estimate for the total claims reserve $R^{(n)}$, one just needs to sum up the claims reserve's estimates for each accident year due to the fact that the claim amounts in different accident years are independent. Hence,

$$\widehat{\mathsf{MSE}}\left[\widehat{R}^{(n)}\right] = \sum_{i=2}^{n} \underbrace{[1,\ldots,1]}_{(i-1)\times 1} \widehat{\mathsf{MSE}}\left[\widehat{\vec{X}}_i^{(n)}\right] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{4.19a}$$

$$\widehat{\mathsf{MSE}}\left[\widehat{\vec{X}}_i^{(n)}\right] = \widehat{\phi}\widehat{\vec{A}}_i^{1/2}\widehat{\vec{C}}_i\widehat{\vec{A}}_i^{1/2} - 2\widehat{\phi}\widehat{\vec{A}}_i^{1/2}\widehat{\vec{C}}_i\widehat{A}_i^{1/2}\widehat{H}_{ii}^{\top} + \widehat{\vec{D}}_i\Sigma_{\widehat{\theta}}\widehat{\vec{D}}_i^{\top}, \tag{4.19b}$$

and

$$\widehat{H}_{ij} = \widehat{\vec{D}}_i\left(\sum_{l=1}^{n}\widehat{D}_l^{\top}\widehat{V}_l^{-1}\widehat{D}_l\right)^{-1}\widehat{D}_j^{\top}\widehat{V}_j^{-1}. \tag{4.19c}$$

4.3 Conditional least squares and copulae

One of the main goals in non-life insurance is to estimate the claims reserve distribution. A generalized time series model, that allows for modeling the conditional mean and variance of the claim amounts, is proposed for the claims development. On contrary to the classical stochastic reserving techniques, the number of model parameters does not depend on the number of development periods, which leads to a more precise forecasting. The demonstrated work relies on Pešta and Okhrin (2014).

The main aim of this section is to deal with serious issues in contemporary reserving techniques, which are quite often set aside, but cause serious problems in the actuarial estimation and prediction. Such pitfalls are assumptions of independent claims, independent stochastic errors (or residuals) in the corresponding claims reserving model, and considering large number of parameters often depending on the number of observations.

The majority of the classical approaches are based on the assumption that the claim amounts in different years are independent. However, this assumption can sometimes be unrealistic or at least questionable. It has been pointed out that methods, enabling *modeling the dependencies*, are needed, cf. Antonio and Beirlant (2007) or Hudecová and Pešta (2013). Mentioned papers suggest the generalized linear mixed models (GLMM) or generalized estimating equations (GEE) to handle the possible dependence among the incremental claims in successive development years. These approaches extend the classical GLM and are frequently used in panel

(longitudinal) data analyses. In this section, we present another possible attitude, namely the *conditional mean-variance* model with a *copula* function.

On the one hand, time series model by Buchwalder et al. (2006) nicely and simply allow to model conditional mean and variance of the claim amounts. On the other hand, that model possesses two disadvantages, which are common for a huge majority of the reserving methods: infinite number of parameters (i.e., depending on the number of observation) and independent errors. Generally, large number of parameters decreases the precision of estimation, because of not sufficient amount of data for the estimation. Furthermore, the classical statistical inference is not valid anymore when the number of parameters depends on the number of observation. To overcome such difficulties, we consider a generalized time series model with *a finite number of parameters not depending on the number of development periods* and, additionally, the *model errors* belonging to the same accident period are *not independent*.

Moreover, all the currently used bootstrap methods in claims reserving require independent residuals in order to estimate the distribution of the reserve and, consequently, calculate some distributional quantities, e.g., VaR at 99.5%. Assumption of independent residuals can be quite unrealistic in the claims reserving setup. Hence, an alternative and more suitable resampling method needs to be proposed in order to sensibly estimate the reserves distribution.

Copulae have already been utilized in the claims reserving to model dependencies between different lines of business, e.g., Shi and Frees (2011). On the contrary, it has to be emphasized that in our approach, only one line of business is taken into account. Copulae are therefore used to model dependencies *within claims* corresponding to that *single line of business*. For sure, our approach can be generalized for several lines of business in the way that a second level of dependence (for instance, modeled again by the copulae) is introduced between the claim amounts from different lines of business.

Outstanding loss liabilities are again structured in the claims development triangles. Let us denote $Y_{i,j}$ all the claim amounts up to development year $j \in \{1,\dots,n\}$ with accident year $i \in \{1,\dots,n\}$. Therefore, $Y_{i,j}$ stands for the *cumulative claims* in accident year i after j development periods. Thus, our data history consists of right-angled isosceles triangle $\{Y_{i,j}\}$, where $i = 1,\dots,n$ and $j = 1,\dots,n+1-i$ (cf. Table 4.1, where the cumulative claims are denoted by $C_{i,j}$).

The aim is to predict the ultimate claims amount $Y_{i,n}$ and the outstanding *claims reserve* $R_i^{(n)} = Y_{i,n} - Y_{i,n+1-i}$ for all $i = 2,\dots,n$. Additional to that, *estimation of the whole distribution of the reserves* is needed in order to provide important distri-

butional quantities for the *Solvency II purposes*, e.g., quantiles for the value at risk calculation.

4.3.1 Conditional mean and variance model

Run-off triangles are comprised of observations which are ordered in time. It is therefore natural to suspect the observations to be dependent. On one hand, the most natural approach is to assume that the observations of a common accident year are dependent. On the other hand, observations of different accident years are supposed to be independent. This assumption is similar to those of the Mack's chain ladder model, cf. Mack (1993).

$\mathcal{F}_{i,j}$ denotes the information set generated by trapezoid $\{Y_{k,l} : l \leqslant j, k \leqslant i+1-j\}$, i.e., $\mathcal{F}_{i,j} = \sigma(Y_{k,l} : l \leqslant j, k \leqslant i+1-j)$ is a filtration corresponding to the smallest σ-algebra containing historical claims with at most j development periods paid in accounting period i or earlier. This notation allows for a zero or even negative index in filtration despite the fact that the claims corresponding to a zero or negative development of accident years are not observed.

Let us define a *nonlinear generalized semiparametric regression* type of model. It can be considered as a generalization of the model proposed by Buchwalder et al. (2006). The first level of generalization is in the mean and variance structure, which was inspired by Patton (2012). The second level of generalization regarding the dependence structure will be introduced in Section 4.3.2.

Definition 4.3.1 (CMV model)**.** The Conditional Mean and Variance (CMV) model assumes

$$Y_{i,j} = \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) + \sigma(Y_{i,j-1}, \boldsymbol{\beta}, j)\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{4.20}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown parameters, whose dimensions do not depend on n, μ is a continuous function in $\boldsymbol{\alpha}$ and σ is a positive and continuous function in $\boldsymbol{\beta}$. Disturbances $\{\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})\}_{j=1}^{n+1-i}$ are independent sample copies of a stationary first-order Markov process for all i. $\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ have the common true invariant distribution $G_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, which is absolutely continuous with respect to Lebesgue measure on the real line. Suppose that

$$\mathsf{E}[\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})|\mathcal{F}_{i,j-1}] = 0, \tag{4.21a}$$

$$\mathsf{Var}[\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})|\mathcal{F}_{i,j-1}] = s(\boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{4.21b}$$

for all i and j. Moreover for the unknown true values $[\boldsymbol{\alpha}^{*\top}, \boldsymbol{\beta}^{*\top}]^\top$ of parameters $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$, the conditional variance of errors equals one, due to identifiability purposes, i.e., $s(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = 1$.

The name of the model comes from the fact that the conditional mean and variance can be expressed as

$$\begin{aligned}\mathsf{E}[Y_{i,j}|\mathcal{F}_{i,j-1}] &= \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j),\\ \mathsf{Var}[Y_{i,j}|\mathcal{F}_{i,j-1}] &= \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j)s(\boldsymbol{\alpha}, \boldsymbol{\beta}).\end{aligned}$$

This property allows for a wide variety of models for the conditional mean: types of ARMA models, vector autoregressions, linear and nonlinear regressions, and others. It also allows for a variety of models for the conditional variance: ARCH and any of its numerous parametric extensions (GARCH, EGARCH, GJR-GARCH, etc., see Bollerslev (2010)), stochastic volatility models, and others.

Patton (2012) considered a similar model, but the dependence was assumed in a different way: dependent copies of the time series (dependence between rows) were supposed, not dependent errors within each time series as we propose. Here, independent rows of errors $[\varepsilon_{i,1}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \ldots, \varepsilon_{i,n+1-i}(\boldsymbol{\alpha}, \boldsymbol{\beta})]$ imply independent rows of claims $[Y_{i,1}, \ldots, Y_{i,n+1-i}]$. Moreover, the unconditional mean and variance of the CMV model's errors equal the conditional ones: $\mathsf{E}\,\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$ and $\mathsf{Var}\,\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = s(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

To provide an insight into possible candidates for the mean function μ and variance function σ, one may propose $\mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) = \eta(\boldsymbol{\alpha}, j)Y_{i,j-1}$ and $\sigma(Y_{i,j-1}, \boldsymbol{\beta}, j) = \nu(\boldsymbol{\beta}, j)\sqrt{Y_{i,j-1}}$ or $\sigma(Y_{i,j-1}, \boldsymbol{\beta}, j) = \nu(\boldsymbol{\beta}, j)Y_{i,j-1}$. Sherman (1984) investigated decays $\eta(\boldsymbol{\alpha}, j)$, which should correspond to the link ratios. Hence, $\eta(\boldsymbol{\alpha}, j)$ should be decreasing in j with limit 1 as j tends to infinity: $1 + \alpha_1 \exp\{-\alpha_2 j\}$, $1 + \alpha_1 \exp\{-j^{\alpha_2}\}$, $1 + \alpha_1 j^{-\alpha_2}$, $1 + \alpha_1(\alpha_2 + j)^{-\alpha_3}$, $1 + \alpha_1^{-\alpha_2^j}$, $\alpha_1^{\alpha_2^{-j}}$, $(1 - \exp\{-\alpha_1 j^{\alpha_2}\})^{-1}$, $\exp\{\alpha_1 j^{-\alpha_2}\}$, $1 + \alpha_1\alpha_2 j^{-1-\alpha_2}\exp\{\alpha_1 j^{-\alpha_2}\}$, $1 + \alpha_1/(j + \alpha_2)$, $1 + \alpha_1/\log(j + \alpha_2)$, etc., where $\alpha_1, \alpha_2, \alpha_3 > 0$. On the other hand, the decay $\nu(\boldsymbol{\beta}, j)$ should be decreasing in j with limit 0: $\beta_1 \exp\{-\beta_2 j\}$, $\beta_1 j^{-\beta_2}$, $\beta_1 \log j \exp\{-\beta_2 j\}$, $\beta_1 j^{-\beta_2}\exp\{-\beta_3 j\}$, $\beta_1 \exp\{-\beta_2 j^2\}$, $\beta_1/(j + \beta_2)$, $\beta_1/\log(j + \beta_2)$, $\beta_1 j^{-\beta_2}\exp\{-\beta_3 j^2\}$, etc., where $\beta_1, \beta_2, \beta_3 > 0$.

In actuarial praxis, these decays are quite often used, mainly for *projecting the development* (forecasting of the claim amounts after n development periods). Despite that, the parameters of the decay curves are not estimated directly from the triangle, but the chain ladder estimates $\widehat{f}_j$ and $\widehat{\sigma}_j^2$ of the development factors f_j and the nuisance variance parameters σ_j^2 (Mack, 1993) are smoothed and used for the decay parameters' estimation. This two-step procedure does not guarantee that the estimated decay parameters will be at least asymptotically unbiased. Unlike that, we will estimate the parameters *directly from the data triangle* and prove the estimates' consistency.

Sherman (1984) assumed independence of individual link ratios (development factors) when estimating the decay parameters by parametric curve fitting. We relax the independence assumption and model the link ratios conditionally having dependent errors.

When comparing the CMV model with the model investigated by Buchwalder et al. (2006), two main differences arise. The CMV model allows for dependent errors and assumes finite number of parameters not depending on the number of development periods n. Indeed, the CMV model requires known functions with unknown finite dimensional parameters. Parameters of the time series model from Buchwalder et al. (2006) are $\{f_j\}_{j=1}^{n}$ and $\{\sigma_j\}_{j=1}^{n}$, which play the role of $\eta(\alpha, j)$ and $\nu(\beta, j)$, respectively. It is important to note that the classical stochastic inference is not valid in the case, when the number of parameters depends on the number of observations. Thus, the legitimacy of the bootstrap procedure in that case is questionable.

Furthermore in the chain ladder, the estimate for f_{n-1} is just a pure ratio of two random variables. Moreover, to estimate σ_{n-1}^2, only doubtful ad-hoc estimates were proposed due to the fact that the claims triangle simply does not contain data for a reasonable estimate (e.g., a consistent one).

4.3.2 Dependence modeling by copulae

Following the 2-dimensional version of Sklar's theorem: Let H be a bivariate distribution function with margins G_1, G_2, then a copula C exists such that

$$H(e_1, e_2) = C\{G_1(e_1), G_2(e_2)\}, \quad e_1, e_2 \in \overline{\mathbb{R}}. \tag{4.22}$$

If G_i are continuous for $i = 1, 2$ then C is unique. Otherwise C is uniquely determined on $Ran G_1 \times Ran G_2$. Conversely, if C is a copula and G_1, G_2 are univariate distribution functions, then the function H defined above is a bivariate distribution function with margins G_1, G_2.

The representation in (4.22) can be used in construction of new bivariate distributions by changing either the copula or marginal distributions. For an arbitrary continuous bivariate distribution we determine its copula from

$$C(u_1, u_2) = H\{G_1^{-1}(u_1), G_2^{-1}(u_2)\}, \quad u_1, u_2 \in [0, 1], \tag{4.23}$$

where G_i^{-1} are inverse marginal distribution functions. The continuous random variables are assumed in order to assure that the inverse functions are well defined.

As the bivariate copula is a bivariate distribution with uniform margins, the copula density is then given through

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}, \quad u_1, u_2 \in [0, 1]. \tag{4.24}$$

Being armed with Sklar's theorem and (4.24), we may write the density h of the bivariate distribution H as follows

$$h(e_1, e_2) = c(G_1(e_1), G_2(e_2)) \prod_{i=1}^{2} g_i(e_i), \quad e_1, e_2 \in \overline{\mathbb{R}}.$$

For a nice copula overview we refer reader to Joe (1997) or Nelsen (2006). There is a huge list of different copula families satisfying different properties, and being constructed in different ways. In the empirical study of this section we concentrate on two members from the elliptical family, namely Gaussian and t-copula and on three members of the Archimedean family, namely Clayton, Frank, and Gumbel. Gaussian and t-copulae contain the dependency structure from the multivariate Normal and multivariate t-distribution respectively, and they are directly derived through the Sklar's theorem. Archimedean copulae on their side are not constructed using this theorem, but are related to Laplace transforms of univariate distribution functions, and are defined as

$$C(u_1, u_2) = \phi\{\phi^{-1}(u_1) + \phi^{-1}(u_2)\}, \quad u_1, u_2 \in [0, 1];$$

where ϕ is the 2-monotone generator function with $\phi : [0; \infty) \to [0, 1]$, $\phi(0) = 1$, $\phi(\infty) = 0$. In particular generator functions for Clayton, Frank, and Gumbel copulae are given respectively through

$$\begin{aligned}
\phi_{\mathrm{Clayton}}(x; \gamma) &= (\gamma x + 1)^{-\frac{1}{\gamma}}, \quad -1 \leqslant \gamma < \infty, \gamma \neq 0, x \in [0, \infty); \\
\phi_{\mathrm{Frank}}(x; \gamma) &= \gamma^{-1} \log\{1 - (1 - e^{-\gamma}) e^{-x}\}, \quad 0 \leqslant \gamma < \infty, \ x \in [0, \infty); \\
\phi_{\mathrm{Gumbel}}(x; \gamma) &= \exp\{-x^{1/\gamma}\}, \quad 1 \leqslant \gamma < \infty, \ x \in [0, \infty).
\end{aligned}$$

These three members represent lower tail but no upper tail for Clayton copula, elliptically contoured form for Frank copula, and no lower tail but upper tail for Gumbel copula.

Dependent errors

Since the mean and variance trends are removed by the CMV model, the rest of the relationship among claim amounts can be additionally captured by modeling de-

pendent errors. The inspiration for the dependence structure was taken from Chen and Fan (2006).

Assumption $\mathcal{A}$.4.29. $\{\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})\}_{j=1}^{n+1-i}$ are independent sample copies of a stationary first-order Markov process for all i generated from $(G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\cdot), C(\cdot,\cdot;\boldsymbol{\gamma}))$, where $C(\cdot,\cdot;\boldsymbol{\gamma})$ is the true parametric copula for $[\varepsilon_{i,j-1}(\boldsymbol{\alpha},\boldsymbol{\beta}), \varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})]$, which is given and fixed up to unknown parameter $\boldsymbol{\gamma}$ and is absolutely continuous with respect to Lebesgue measure on $[0,1]^2$.

It is believed that there exists a kind of *information overlap* between the claims from consecutive development periods, which corresponds to the dependence between the CMV model's errors modeled by copulae.

Assumption $\mathcal{A}$.4.29 together with the CMV model yield a copula-based model, where the joint bivariate distribution of errors $[\varepsilon_{i,j-1}(\boldsymbol{\alpha},\boldsymbol{\beta}), \varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})]$ has the following distribution function

$$H(e_1,e_2) = C(G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_1), G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_2);\boldsymbol{\gamma}).$$

Then, the conditional copula density can be derived as

$$h(e_2|e_1) = g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_2)c(G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_1), G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_2);\boldsymbol{\gamma}), \tag{4.25}$$

where c is the copula density and $g_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is the marginal density corresponding to the univariate distribution function $G_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. The latter relation (4.25) will play an important role in "making" the dependent errors *conditionally independent* during the forthcoming estimation and prediction process.

4.3.3 Parameter estimation

The CMV model from Definition 4.3.1 together with the copula Assumption $\mathcal{A}$.4.29 contain three vector parameters, which need to be estimated. The estimation process consists of two stages. In the first one, mean and variance parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are estimated in a distribution-free fashion, since no specific distributional assumptions are proposed nor required for the claims. The second stage concerns the estimation of the dependence structure, mainly the copula parameter $\boldsymbol{\gamma}$, in a likelihood based way.

Estimation in CMV model

Since the CMV model is defined in a conditional style, *conditional least squares* (CLS) of the sample centered conditional moments of the claims are minimized in order to obtain estimates of the CMV model parameters.

Definition 4.3.2 (Conditional least squares estimates)**.** Let us denote

$$M_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{n-1} \sum_{j=2}^{n} \frac{1}{n+1-j} \sum_{i=1}^{n+1-j} \frac{\left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j)\right]^2}{\sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j)}$$

and

$$V_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{n-1} \sum_{j=2}^{n} \frac{1}{n+1-j} \sum_{i=1}^{n+1-j} \left\{ \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j)\right]^2 - \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j) \right\}^2,$$

where parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ belong to parameter spaces $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$. The conditional least squares estimate of the mean parameter $\boldsymbol{\alpha}$ for a fixed value of parameter $\boldsymbol{\beta} \in \boldsymbol{\Theta}_2$ is defined as

$$\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\alpha} \in \boldsymbol{\Theta}_1} M_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

and the conditional least squares estimate of the variance parameter $\boldsymbol{\beta}$ for a fixed value of parameter $\boldsymbol{\alpha} \in \boldsymbol{\Theta}_1$ is defined as

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \arg \min_{\boldsymbol{\beta} \in \boldsymbol{\Theta}_2} V_n(\boldsymbol{\alpha}, \boldsymbol{\beta}).$$

The reason, why the parameter estimates for the CMV model are defined as above, lies in the fact that it is computationally not feasible to find the global minimum of M_n and V_n with respect to $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$ simultaneously.

The forthcoming theory (Theorems 4.3.1, 4.3.2, and Corollary 4.3.3) assures that the CLS estimates are reasonable and, moreover, consequent Algorithm 4.3.1 provides a computational way for obtaining the CMV parameter estimates.

Theorem 4.3.1 (Conditional least squares consistency for the mean)**.** *Let the CMV model hold and $\boldsymbol{\beta} \in \boldsymbol{\Theta}_2$ be fixed. Assume that*

(i) $\left\{ \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j)\right]^2 / \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j) \right\}_{i,j \in \mathbb{N}}$ *is uniformly integrable,*

(ii) for all $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \boldsymbol{\Theta}_1$ *and* $n \in \mathbb{N}$,

$$\left| \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) \right]^2 - \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}', j) \right]^2 \right| \leqslant C_j \, g(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|) \quad (4.26)$$

almost surely, where $\{C_j\}_{j \in \mathbb{N}}$ *is a stochastic sequence not depending on* $\boldsymbol{\alpha}$ *such that* $C_j = \mathcal{O}_P(\sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j))$, $j \to \infty$ *for all* $i \in \mathbb{N}$ *and* g *is nonstochastic such that* $g(t) \downarrow 0$ *as* $t \downarrow 0$,

(iii) $s(\cdot, \boldsymbol{\beta})$ *is a Lipschitz function on the compact parameter space* $\boldsymbol{\Theta}_1$ *such that the true unknown parameter* $\boldsymbol{\alpha}^*(\boldsymbol{\beta})$ *is its unique global minimum.*

Then, $\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) \xrightarrow[n \to \infty]{P} \boldsymbol{\alpha}^*(\boldsymbol{\beta})$.

Weak consistency (in probability) of the mean parameter estimate is shown, but also the strong version (almost sure convergence) can be provided. It would require C_j to be bounded almost surely, which is less feasible.

The Lipschitz kind of assumption (ii) can be replaced by a stronger one: uniform equiboundedness in probability. In that case, it suffices to assume

$$E \sup_{\boldsymbol{\alpha} \in \boldsymbol{\Theta}_1} \|\nabla_{\boldsymbol{\alpha}} M_n(\boldsymbol{\alpha}, \boldsymbol{\beta})\| < \Delta_1$$

for all n and $\boldsymbol{\beta}$, and convexity of the compact parameter space $\boldsymbol{\Theta}_1$ for applying the stochastic mean-value theorem. The compactness of the parameter space can even be relaxed to its total boundedness.

A similar theorem as above is going to be postulated for the CLS variance parameter estimate to ensure its appropriateness. Firstly, let us define

$$B_{n,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \frac{1}{(n-1)(n+1-j)} \sum_{i=1}^{n+1-j} \left\{ \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) \right]^2 - \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j) \right\}^2$$

and

$$v(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \lim_{n \to \infty} \sum_{j=2}^{n} E \, B_{n,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}).$$

Theorem 4.3.2 (Conditional least squares consistency for the variance)**.** *Let the CMV model hold and* $\boldsymbol{\alpha} \in \boldsymbol{\Theta}_1$ *be fixed. Assume that*

(i) random array

$$\left\{ \left\{ \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) \right]^2 - \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j) \right\}^2 \right.$$
$$\left. - \mathsf{E} \left\{ \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) \right]^2 - \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j) \right\}^2 \right\}_{i,j \in \mathbb{N}}$$

is uniformly integrable,

(ii) for all $\boldsymbol{\beta}, \boldsymbol{\beta}' \in \boldsymbol{\Theta}_2$ *and* $n \in \mathbb{N}$,

$$\left| \left\{ \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) \right]^2 - \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j) \right\}^2 \right.$$
$$\left. - \left\{ \left[Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) \right]^2 - \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}', j) \right\}^2 \right| \leqslant D_j h(\|\boldsymbol{\beta} - \boldsymbol{\beta}'\|) \quad (4.27)$$

almost surely, where $\{D_j\}_{j \in \mathbb{N}}$ *is a stochastic sequence not depending on* $\boldsymbol{\beta}$ *such that* $D_j = \mathcal{O}_P(1)$, $j \to \infty$ *for all* $i \in \mathbb{N}$ *and* h *is nonstochastic such that* $h(t) \downarrow 0$ *as* $t \downarrow 0$,

(iii) for all $j \leqslant n$, $n \in \mathbb{N}$, *and* $m \in \mathbb{N}_0$,

$$\mathsf{E} \left| \mathsf{E} \left[B_{n,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathsf{E}\, B_{n,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}) | \mathcal{F}_{n,j-m} \right] \right| \leqslant c_{n,j}\, d_m, \quad (4.28)$$

where $\{c_{n,j}\}_{j \leqslant n, n \in \mathbb{N}}$ *and* $\{d_m\}_{m \in \mathbb{N}_0}$ *are constants such that* $d_m \downarrow 0$ *as* $m \to \infty$,

(iv) $\nu(\boldsymbol{\alpha}, \cdot)$ *is a Lipschitz function on the compact parameter space* $\boldsymbol{\Theta}_2$ *such that the true unknown parameter* $\boldsymbol{\beta}^*(\boldsymbol{\alpha})$ *is its unique global minimum.*

Then, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \xrightarrow[n \to \infty]{P} \boldsymbol{\beta}^*(\boldsymbol{\alpha})$.

A natural question arises: What is the connection between the true unknown parameter values $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ of the CMV model and the true unknown parameter values $\boldsymbol{\alpha}^*(\boldsymbol{\beta})$ and $\boldsymbol{\beta}^*(\boldsymbol{\alpha})$ from Theorems 4.3.1 and 4.3.2? The intuition behind the CMV model is that the function μ should mimic the conditional mean of the claims and the function σ^2 should model their conditional variance. Mathematically speaking, $\mathsf{Var}[Y_{i,j}/\sigma(Y_{i,j-1}, \boldsymbol{\beta}, j) | \mathcal{F}_{i,j-1}]$ and, similarly, $\mathsf{Var}[(Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j))^2 | \mathcal{F}_{i,j-1}]$ should be as small as possible. Taking into account that the data triangle does not possess the same number of claim amounts entries for each development period j, it is reasonable to assume that if the CMV model holds, then both discrepancy measures

$$\lim_{n \to \infty} \mathsf{E} \left\{ \frac{1}{n-1} \sum_{j=2}^{n} \frac{1}{n+1-j} \sum_{i=1}^{n+1-j} \mathsf{Var} \left[\frac{Y_{i,j}}{\sigma(Y_{i,j-1}, \boldsymbol{\beta}, j)} \middle| \mathcal{F}_{i,j-1} \right] \right\} \quad (4.29)$$

and

$$\lim_{n\to\infty} \mathsf{E}\left\{ \frac{1}{n-1} \sum_{j=2}^{n} \frac{1}{n+1-j} \sum_{i=1}^{n+1-j} \mathsf{Var}\left[\left(Y_{i,j} - \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) \right)^2 | \mathcal{F}_{i,j-1} \right] \right\} \tag{4.30}$$

reach their global minimum at the same true unknown parameter values $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ of the CMV model. However, measures (4.29) and (4.30) are nothing else than $s(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $v(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Now, let us define the interior of set $\boldsymbol{\Theta}$ by $\boldsymbol{\Theta}^\circ$.

Corollary 4.3.3 (Consistency of the CLS estimates)**.** *Suppose that the assumptions of Theorems 4.3.1 and 4.3.2 hold. Let* $s \in \mathcal{C}^2(\boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2)$, $v \in \mathcal{C}^2(\boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2)$, *and both functions* s *and* v *have their unique global minimum on compact set* $\boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2$ *at* $[\boldsymbol{\alpha}^{*\top}, \boldsymbol{\beta}^{*\top}]^\top \in \boldsymbol{\Theta}_1^\circ \times \boldsymbol{\Theta}_2^\circ$. *If* $\det[\partial^2 s(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top] \neq 0$ *and* $\det[\partial^2 v(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top] \neq 0$ *for all* $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top \in \boldsymbol{\Theta}_1^\circ \times \boldsymbol{\Theta}_2^\circ$, *then*

$$\begin{bmatrix} \widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}^*) \\ \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}^*) \end{bmatrix} \xrightarrow[n\to\infty]{\mathsf{P}} \begin{bmatrix} \boldsymbol{\alpha}^* \\ \boldsymbol{\beta}^* \end{bmatrix}.$$

The variance parameter $\boldsymbol{\beta}$ can be viewed as a nuisance parameter when estimating the mean parameter $\boldsymbol{\alpha}$ and vice-versa. The idea of joint estimation of $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$ is to alternately perform partial optimizations from Definition 4.3.2. In fact, we iteratively estimate $\boldsymbol{\alpha}$, given the fixed value of $\boldsymbol{\beta}$. Consequently, we estimate $\boldsymbol{\beta}$, given the fixed value of $\boldsymbol{\alpha}$ (obtained from the previous step). This two steps are repeated in turns until almost no change in the consecutive estimates of $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$, see Algorithm 4.3.1. Based on Corollary 4.3.3, it is believed that each turn will bring our iterated estimates closer to the true unknown parameter values. Moreover, Algorithm 4.3.1 can be modified: the initial value of $\boldsymbol{\alpha}^{(0)}$ could be required on the input instead of $\boldsymbol{\beta}^{(0)}$ and the whole iteration procedure would start with the estimation of $\boldsymbol{\beta}^{(1)}$ instead of $\boldsymbol{\alpha}^{(1)}$.

Estimation of dependence structure

The second stage of the parameter estimation process involves the estimation of the whole dependence structure in the claims triangle. The strict stationarity of the first order Markov process imposed on the CMV model errors (Assumption $\mathcal{A}$.4.29) arranges that only the bivariate distribution of two in row neighbouring errors are necessary to know. Using Sklar's theorem, this bivariate distribution is obtained by specifying a proper parametric copula and a marginal distribution of the errors. Hence, the second stage consists of the estimation of the copula parameter and the marginal errors' distribution.

Procedure 4.3.1 Iterative conditional least squares estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Input: Cumulative claims triangle $\{Y_{i,j}\}_{i,j=1}^{n,n+1-i}$, mean and variance functions μ and σ, initial (starting) parameter value $\boldsymbol{\beta}^{(0)}$, maximum number of iterations M, and convergence precision ϵ.

Output: CLS parameter estimates $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, fitted residuals $\{\widehat{\varepsilon}_{i,j}\}_{i=1,j=2}^{n-1,n+1-i}$.

$m \leftarrow 1$ and $\boldsymbol{\alpha}^{(0)} \leftarrow \mathbf{0}$
$\boldsymbol{\alpha}^{(1)} \leftarrow \arg\min_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}_1} M_n(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(0)})$
$\boldsymbol{\beta}^{(1)} \leftarrow \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}_2} V_n(\boldsymbol{\alpha}^{(1)}, \boldsymbol{\beta})$
while $m \leqslant M$ **and** $\|[\boldsymbol{\alpha}^{(m)\top}, \boldsymbol{\beta}^{(m)\top}]^\top - [\boldsymbol{\alpha}^{(m-1)\top}, \boldsymbol{\beta}^{(m-1)\top}]^\top\| > \epsilon$ **do**
  $\boldsymbol{\alpha}^{(m+1)} \leftarrow \arg\min_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}_1} M_n(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(m)})$
  $\boldsymbol{\beta}^{(m+1)} \leftarrow \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}_2} V_n(\boldsymbol{\alpha}^{(m+1)}, \boldsymbol{\beta})$
  $m \leftarrow m+1$
end while
$\widehat{\boldsymbol{\alpha}} \leftarrow \boldsymbol{\alpha}^{(m)}$ and $\widehat{\boldsymbol{\beta}} \leftarrow \boldsymbol{\beta}^{(m)}$
for $i = 1$ **to** $n-1$ **do**
  for $j = 2$ **to** $n+1-i$ **do**
    $\widehat{\varepsilon}_{i,j} \leftarrow [Y_{i,j} - \mu(Y_{i,j-1}, \widehat{\boldsymbol{\alpha}}, j)]/\sigma(Y_{i,j-1}\widehat{\boldsymbol{\beta}}, j)$
  end for
end for

Since the estimates of the CMV model parameters are already available, one can estimate the unknown *marginal distribution* function $G_{\alpha,\beta}$ of the CMV model errors $\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ non-parametrically by the empirical distribution function

$$\widehat{G}_n(e) = \frac{1}{n(n-1)/2+1} \sum_{i=1}^{n-1} \sum_{j=2}^{n+1-i} \mathcal{I}\{\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) \leqslant e\},$$

of the fitted residuals

$$\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \frac{Y_{i,j} - \mu(Y_{i,j-1}, \widehat{\boldsymbol{\alpha}}, j)}{\sigma(Y_{i,j-1}, \widehat{\boldsymbol{\beta}}, j)}.$$

The consistency of the CMV model parameter estimates $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ ensures that the fitted residuals $\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ are reasonable predictors of the unknown non-observable errors $\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Algorithm 4.3.1 also provides the fitted residuals as a side product.

Assumption $\mathcal{A}$.4.29 demands a prior knowledge of the parametric copula up to its unknown parameter γ. Nevertheless in practical applications, one needs to perform a copula goodness-of-fit in order to choose a suitable copula. Assuming that we know the bivariate copula function $C(\cdot, \cdot; \gamma)$, the copula parameter γ is estimated by the *quasi-likelihood* method.

Having in mind that rows of errors $[\varepsilon_{i,1}(\boldsymbol{\alpha},\boldsymbol{\beta}),\ldots,\varepsilon_{i,n+1-i}(\boldsymbol{\alpha},\boldsymbol{\beta})]$ are independent for all i, the full log-likelihood for the copula parameter γ with respect to (4.25) is

$$\mathcal{L}(\gamma) = \sum_{i=1}^{n-2}\sum_{j=2}^{n+1-i} \log g_{\alpha,\beta}(\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})) + \sum_{i=1}^{n-2}\sum_{j=3}^{n+1-i} \log c(G_{\alpha,\beta}(\varepsilon_{i,j-1}(\boldsymbol{\alpha},\boldsymbol{\beta})), G_{\alpha,\beta}(\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta}));\gamma).$$

Ignoring the first term in $\mathcal{L}(\gamma)$ and replacing ε's and $G_{\alpha,\beta}$ by their estimated counterparts $\widehat{\varepsilon}$'s and $\widehat{G}_n$, the parameter γ can be estimated by the so-called canonical maximum likelihood, i.e., maximizing the *partial (pseudo) log-likelihood*:

$$\widehat{\gamma} = \arg\max_{\gamma} \widetilde{\mathcal{L}}(\gamma),$$
$$\widetilde{\mathcal{L}}(\gamma) = \sum_{i=1}^{n-2}\sum_{j=3}^{n+1-i} \log c(\widehat{G}_n(\widehat{\varepsilon}_{i,j-1}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})), \widehat{G}_n(\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}));\gamma).$$

The correctness of this approach was shown by Chen and Fan (2006), where the consistency of the canonical likelihood estimate of copula parameter γ was proved under Assumption $\mathcal{A}$.4.29. Here, the unknown unobservable CMV model errors are just replaced by the fitted residuals based on the consistent CMV model parameter estimates. Algorithm 4.3.2 encapsulates the way of getting copula parameter estimate.

Procedure 4.3.2 Copula parameter γ estimation by pseudo-likelihood.

Input: Fitted residuals $\{\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})\}_{i=1,j=2}^{n-1,n+1-i}$ and copula density $c(\cdot,\cdot;\gamma)$.
Output: Copula parameter estimate $\widehat{\gamma}$.

1: marginal ecdf $\widehat{G}_n(e) \leftarrow \frac{1}{n(n-1)/2+1}\sum_{i=1}^{n-1}\sum_{j=2}^{n+1-i}\mathcal{I}\{\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}) \leqslant e\}$
2: $\widehat{\gamma} \leftarrow \arg\max_{\gamma}\sum_{i=1}^{n-2}\sum_{j=3}^{n+1-i}\log c(\widehat{G}_n(\widehat{\varepsilon}_{i,j-1}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})), \widehat{G}_n(\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}));\gamma)$

Chen and Fan (2006) also remarked that the empirical distribution function $\widehat{G}_n$ can be *smoothed by kernels* as an alternative estimate of the marginal distribution of errors. This can be especially helpful in case of a smaller number of residuals available.

Now, the whole estimation procedure becomes a *semiparametric* one.

4.3.4 Prediction of reserves and estimating their distribution

The main goals in actuarial reserving are prediction of reserves $R_i^{(n)}$ and, consequently, estimation of the reserves' distribution, e.g., in order to obtain quantiles—99.5% quantile for the Solvency II purposes (VaR calculation).

A predictor for the reserve $R_i^{(n)}$ can be defined as

$$\widehat{R}_i^{(n)} = \widehat{Y}_{i,n} - Y_{i,n+1-i}.$$

Therefore, finding a predictor $\widehat{Y}_{i,n}$ for $Y_{i,n}$ is crucial. One naive proposal is to predict $Y_{i,j}$ as the conditional mean of previous claim amount $Y_{i,j-1}$ with the plugged-in CMV model parameter estimates

$$\check{Y}_{i,j} := E_{\alpha,\beta}[Y_{i,j}|Y_{i,n+1-i}]|_{\alpha=\widehat{\alpha},\beta=\widehat{\beta}}, \quad i = 2,\dots,n,\ j = n+2-i,\dots,n. \tag{4.31}$$

This approach gives $\check{Y}_{i,j} = \mu(\check{Y}_{i,j-1}, \widehat{\alpha}, j)$, $i+j > n+1$. In spite of that, such an approach would only be eligible if prediction (4.31) was unbiased or at least asymptotically unbiased, which is not justified.

Semiparametric bootstrap

The prediction of unobserved claims may be done in a *telescopic* way based on the CMV model formulation: start with the diagonal element $Y_{i,n+1-i}$ and predict $Y_{i,j}$, $j > n+1-i$ stepwise in each row

$$\widehat{Y}_{i,j} = Y_{i,j}, \quad i+j \leqslant n+1; \tag{4.32a}$$
$$\widehat{Y}_{i,j} = \mu(\widehat{Y}_{i,j-1}, \widehat{\alpha}, j) + \sigma(\widehat{Y}_{i,j-1}, \widehat{\beta}, j)\widetilde{\varepsilon}_j, \quad i+j > n+1. \tag{4.32b}$$

Errors $\widetilde{\varepsilon}_j$ are simulated from the fitted residuals. One convenient resampling procedure is the semiparametric bootstrap which takes advantage of the fact that $\varepsilon_{i,j}(\alpha,\beta) = G^{-1}_{\alpha,\beta}(X_j)$ for all i (due to the independent rows), where $\{X_j\}_{j=2}^n$ is a stationary first-order Markov process with the copula $C(x_1, x_2; \gamma)$ being the joint distribution of $[X_{j-1}, X_j]$.

When the errors $\widetilde{\varepsilon}_j$ are simulated sufficiently many times, the empirical (bootstrap) distribution of $\widehat{Y}_{i,n}$ is obtained, which should mimic the true unknown distribution of $Y_{i,n}$. Thereafter, an estimate of the reserves' distribution is acquired and some imported quantities of the reserves can be easily calculated, e.g., mean, variance, or quantiles of the reserves.

4.4 Key contributions

- Consistency of the distribution-free chain ladder method is examined.
- Necessary and sufficient conditions for the conditional consistency of the development factors' estimates are derived for the first time.
- Generalized estimating equations (GEE) are applied within the triangular data setup, allowing for dependencies in the claims reserving framework.
- Model selection criteria within the GEE reserving method are proposed.
- Mean square error of prediction is calculated and its estimate is provided.
- A generalized time series model for the triangular data, that allows for modeling of the conditional mean and variance of the claims, is assumed.
- Conditional least squares are used to estimate model parameters and consistency of these estimates is proved.
- A copula approach is used for modeling the dependence structure.
- A semiparametric bootstrap is utilized in order to estimate the whole distribution of the predicted reserves more precisely compared to the traditional approaches.

5 Conclusions

> *Progress is impossible without change, and those who cannot change their minds cannot change anything.*
>
> **— George Bernard Shaw**
> (Irish playwright, critic, and polemicist)

This habilitation thesis studies dynamics and instabilities of complex data structures like time series, panel data, and triangular data. The main goal lies in modeling of the time development and detection of the structural breaks.

In the statistical analysis, it is of particular interest to be able to detect systematic changes—so-called changepoints—in the underlying structure despite the random fluctuations and to estimate the time of these changes. The analyses of the changepoint problems are conducted within the time series setup as well as within the panel data framework. Various testing procedures are proposed, their asymptotic behavior is derived, consistency is proved, and bootstrap extensions are developed. Moreover, the changepoint estimators are invented, their performance is investigated, and again consistency is shown. The merit of the presented approaches was to avoid usage of the tuning constants or nuisance parameters involved in the detection procedures, which makes these methods effortlessly applicable.

Prediction of the forthcoming expenses coming, for instance, from insurance vigorously requires its stochastic solutions. Therefore, dynamics of the actuarial data structured in the triangles is modeled in various ways. Dependencies between claims are allowed in order to increase the precision of statistical prediction. After that, resampling techniques are introduced for the predictive models aiming to provide probabilistic distributions of the future losses.

Bibliography

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons: New York, NY (cit. on p. 11).

Andrews, D. W. K. (1991). „Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation“. *Econometrica* 59(3), 817–858 (cit. on p. 43).

Andrews, D. W. K. (1993). „Tests for parameter instability and structural change with unknown change point“. *Econometrica* 61(4), 821–858 (cit. on p. 8).

Andrews, D. W. K. and Ploberger, W. (1994). „Optimal tests when a nuisance parameter is present only under the alternative“. *Econometrica* 62(6), 1383–1414 (cit. on p. 8).

Andrews, D. W. K., Lee, I., and Ploberger, W. (1996). „Optimal changepoint tests for normal linear regressions“. *J. Econometrics* 70(1), 9–380 (cit. on p. 8).

Antoch, J. and Hušková, M. (2001). „Permutation tests in change point analysis“. *Stat. Probabil. Lett.* 53(1), 37–46 (cit. on p. 8).

Antoch, J., Hušková, M., and Prášková, Z. (1997). „Effect of dependence on statistics for determination of change“. *J. Stat. Plan. Infer.* 60(2), 291–310 (cit. on pp. 7, 21).

Antonio, K. and Beirlant, J. (2007). „Actuarial statistics with generalized linear mixed models“. *Insur. Math. Econ.* 40, 58–76 (cit. on pp. 63, 73).

Antonio, K., Beirlant, J., Hoedemakers, T., and Verlaak, R. (2006). „Lognormal mixed models for reported claim reserve“. *North American Actuarial Journal* 10(1), 30–48 (cit. on p. 63).

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2008). „Testing for changes in polynomial regression“. *Bernoulli* 14(3), 637–660 (cit. on p. 27).

Bai, J. (2010). „Common Breaks in Means and Variances for Panel Data“. *J. Econometrics* 157(1), 78–92 (cit. on pp. 37, 41, 42, 47, 49, 51).

Baltagi, B. H., Feng, Q., and Kao, C. (2016). „Estimation of heterogeneous panels with structural breaks“. *J. Econometrics* 191(2016), 176–195 (cit. on p. 48).

Bazarova, A., Berkes, I., and Horváth, L. (2015). „Change point detection with stable AR(1) errors“. In: *Asymptotic Laws and Methods in Stochastics*. Ed. by D. Dawson, R. Kulik, M. O. Haye, B. Szyszkowicz, and Y. Zhao. Vol. 76. Fields Inst. Commun. Springer, pp. 179–193 (cit. on p. 9).

Belyaev, Y. K. (1995). *Bootstrap, Resampling and Mallows Metric*. Lecture Notes 1. Institute of Mathematical Statistics, Umeå University (cit. on p. 57).

Betken, A. (2016). „Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test“. *J. Time Ser. Anal.* 37(6), 785–809 (cit. on p. 19).

Betken, A and Wendler, M (2018). „Subsampling for general statistics under long range dependence“. *Statist. Sinica* 28(3), 1199–1224 (cit. on p. 19).

Billingsley, P. (1968). *Convergence of Probability Measures*. 1st. John Wiley & Sons: New York, NY (cit. on p. 11).

Björkwall, S., Hössjer, O., Ohlsson, E., and Verrall, R. (2011). „A generalized linear model with smoothing effects for claims reserving“. *Insur. Math. Econ.* 49(1), 27–37 (cit. on pp. 67, 68).

Bollerslev, T. (2010). „Glossary to ARCH (GARCH)“. In: *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Ed. by T. Bollerslev, J. R. Russell, and M. W. Watson. Oxford University Press, Oxford, pp. 137–163 (cit. on p. 76).

Booth, J. G. and Hall, P. (1993). „Bootstrap confidence regions for functional relationships in errors-in-variables models“. *Ann. Stat.* 21(4), 1780–1791 (cit. on p. 28).

Bradley, R. C. (2005). „Basic properties of strong mixing conditions. A survey and some open questions“. *Probab. Surveys* 2, 107–144 (cit. on p. 11).

Brockwell, P. J. and Davis, R. A. (2006). *Time Series: Theory and Methods*. 2nd. Springer: New York, NY (cit. on p. 70).

Buchwalder, M., Bühlmann, H., Merz, M., and Wüthrich, M. V. (2006). „The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited)“. *ASTIN Bull.* 36(2), 521–542 (cit. on pp. 74, 75, 77).

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. Chapman and Hall/CRC: Boca Raton, FL (cit. on p. 27).

Carroll, R. J., Roeder, K., and Wasserman, L. (1999). „Flexible parametric measurement error models“. *Biometrics* 55(1), 44–54 (cit. on p. 27).

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd. Chapman and Hall/CRC: Boca Raton, FL (cit. on p. 27).

Cavaliere, G. (2005). „Unit root tests under time-varying variances“. *Economet. Rev.* 23(3), 259–292 (cit. on p. 22).

Chan, J., Horváth, L., and Hušková, M. (2013). „Change-point Detection in Panel Data“. *J. Stat. Plan. Infer.* 143(5), 955–970 (cit. on p. 37).

Chang, Y.-P. and Huang, W.-T. (1997). „Inferences for the linear errors-in-variables with changepoint models“. *J. Am. Stat. Assoc.* 92(437), 171–178 (cit. on pp. 27, 28).

Chen, X. and Fan, Y. (2006). „Estimation of copula-based semiparametric time series models“. *J. Econometrics* 130(2), 307–335 (cit. on pp. 79, 85).

Chen, Z. and Tian, Z. (2014). „Ratio tests for variance change in nonparametric regression“. *Statistics* 48(1), 1–16 (cit. on pp. 9, 40).

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley: Chichester, UK (cit. on pp. 7, 8, 40).

Dembo, A. (1988). „Bounds on the extreme eigenvalues of positive-definite Toeplitz matrices“. *IEEE T. Inform. Theory* 34(2), 352–355 (cit. on p. 33).

Dong, C., Tan, C., Jin, B., and Miao, B. (2016). „Inference on the change point estimator of variance in measurement error models“. *Lith. Math. J.* 56(4), 474–491 (cit. on p. 27).

Doukhan, P. (1994). *Mixing: Properties and Examples*. Vol. 85. Lecture Notes in Statistics. Springer: New York, NY (cit. on p. 12).

England, P. D. and Verrall, R. J. (2002). „Stochastic claims reserving in general insurance (with discussion)“. *British Actuarial Journal* 8(3), 443–518 (cit. on pp. 53, 69).

Fitzenberger, B. (1997). „The moving block bootstrap and robust inference for linear least squares and quantile regression“. *J. Econometrics* 82, 235–287 (cit. on p. 18).

Fryzlewicz, P. (2014). „Wild binary segmentation for multiple change-point detection“. *Ann. Stat.* 42(6), 2243–2281 (cit. on p. 35).

Fuller, W. A. (1987). *Measurement Error Models*. Wiley: New York, NY (cit. on p. 27).

Gallo, P. P. (1982a). „Consistency of Regression Estimates When Some Variables Are Subject to Error“. *Commun. Stat. A-Theor.* 11, 973–983 (cit. on p. 28).

Gallo, P. P. (1982b). „Properties of Estimators in Errors-in-Variables Models“. PhD thesis. Chapel Hill, NC: University of North Carolina (cit. on p. 28).

Gleser, L. J. (1981). „Estimation in a multivariate “errors in variables” regression model: Large sample results“. *Ann. Stat.* 9, 24–44 (cit. on pp. 27, 28, 30).

Golub, G. H. and Van Loan, C. F. (1980). „An analysis of the total least squares problem“. *SIAM J. Numer. Anal.* 17(6), 883–893 (cit. on pp. 28, 32).

Górecki, T., Horváth, L., and Kokoszka, P. (2018). „Change point detection in heteroscedastic time series“. *Econom. Stat.* 7, 63–88 (cit. on pp. 19, 20, 22).

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). „On blocking rules for the bootstrap with dependent data“. *Biometrika* 82(3), 561–574 (cit. on p. 18).

Hardin, J. W. and Hilbe, J. (2003). *Generalized Estimating Equations*. Chapman & Hall/CRC: Boca Raton, FL (cit. on pp. 64, 68).

Hin, L.-Y. and Wang, Y.-G. (2009). „Working-correlation-structure identification in generalized estimating equations“. *Stat. Med.* 28, 642–658 (cit. on p. 68).

Horváth, L. (1995). „Detecting changes in linear regressions“. *J. Am. Stat. Assoc.* 26(3), 189–208 (cit. on p. 27).

Horváth, L., Hušková, M., Rice, G., and Wang, J. (2015). „Estimation of the time of change in panel data“. *ArXiv e-prints*. arXiv: `1503.04455 [math.ST]` (cit. on pp. 47, 51).

Horváth, L. and Hušková, M. (2012). „Change-point Detection in Panel Data“. *J. Time Ser. Anal.* 33(4), 631–648 (cit. on pp. 37, 47).

Horváth, L., Horváth, Z., and Hušková, M. (2008). „Ratio tests for change point detection“. In: *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*. Ed. by N. Balakrishnan, E. A. Peña, and M. J. Silvapulle. Vol. 1. Institute of Mathematical Statistics: Beachwood, Ohio, pp. 293–304 (cit. on pp. 8, 10, 14, 19, 21, 40).

Hudecová, Š. and Pešta, M. (2013). „Modeling Dependencies in Claims Reserving with GEE“. *Insur. Math. Econ.* 53(3), 786–794 (cit. on pp. 63, 73).

Hušková, M. (2004). „Permutation principle and bootstrap in change point analysis“. *Fields Inst. Commun.* 44, 273–291 (cit. on p. 8).

Hušková, M. (2007). „Ratio type test statistics for detection of changes in time series“. In: *Bulletin of the International Statistical Institute, Proceedings of the 56th Session*. Ed. by M. I. Gomes, J. A. Pinto Martins, and J. A. Silva. Vol. 976. Lisboa, pp. 3934–3937 (cit. on pp. 8, 10, 14).

Hušková, M. and Kirch, C. (2010). „A note on studentized confidence intervals for the change-point“. *Comput. Statist.* 25(2), 269–289 (cit. on p. 19).

Hušková, M. and Marušiaková, M. (2012). „M-procedures for detection of changes for dependent observations“. *Commun. Stat. B-Simul.* 41(7), 1032–1050 (cit. on p. 14).

Hušková, M. and Kirch, C. (2012). „Bootstrapping sequential change-point tests for linear regression“. *Metrika* 75(5), 673–708 (cit. on pp. 19, 44).

Hušková, M., Kirch, C., Prášková, Z., and Steinebach, J. (2008). „On the detection of changes in autoregressive time series, II. Resampling procedures“. *J. Stat. Plan. Infer.* 138(6), 1697–1721 (cit. on p. 44).

Ibragimov, I. A. (1959). „Some limit theorems for stochastic processes stationary in the strict sense“. *Doklady Akademii Nauk SSSR (In Russian)* 125, 711–714 (cit. on p. 11).

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff: Groningen, The Netherlands (cit. on p. 11).

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall: London, UK (cit. on p. 78).

Kim, D. (2011). „Estimating a common deterministic time trend break in large panels with cross sectional dependence“. *J. Econometrics* 164(2), 310–330 (cit. on pp. 47, 49).

Kim, D. (2014). „Common breaks in time trends for large panel data with a factor structure“. *Economet. J.* 17(3), 301–337 (cit. on p. 47).

Kim, J.-Y. (2000). „Detection of change in persistence of a linear time series“. *J. Econometrics* 95(1), 97–116 (cit. on pp. 8, 19).

Kim, J.-Y. and Amador, R. B. (2002). „Corrigendum to Detection of change in persistence of a linear time series“. *J. Econometrics* 109(2), 389–392 (cit. on p. 8).

Kirch, C. (2006). „Resampling Methods for the Change Analysis of Dependent Data“. PhD thesis. Germany: University of Cologne (cit. on pp. 8, 15, 18).

Kukush, A., Markovsky, I., and Van Huffel, S. (2007). „Estimation in a linear multivariate measurement error model with a change point in the data“. *Comput. Stat. Data An.* 52(2), 1167–1182 (cit. on p. 27).

Künsch, H. R. (1989). „The jacknife and the bootstrap for general stationary observations“. *Ann. Stat.* 17, 1217–1241 (cit. on p. 15).

Lahiri, S., Furukawa, K., and Lee, Y.-D. (2007). „A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods“. *Stat. Methodol.* 4(3), 292–321 (cit. on p. 18).

Leybourne, S. J. and Taylor, R. (2006). „Persistence change tests and shifting stable autoregressions“. *Econ. Lett.* 91(1), 44–49 (cit. on p. 8).

Liang, K. and Zeger, S. L. (1986). „Longitudinal data analysis using generalized linear models“. *Biometrika* 73(1), 13–22 (cit. on pp. 63, 66, 72).

Liu, Y., Zou, C., and Zhang, R. (2008). „Empirical Likelihood Ratio Test for a Change-Point in Linear Regression Model“. *Commun. Stat. A-Theor.* 37(16), 2551–2563 (cit. on p. 40).

Ma, E. M. and Zarowski, C. J. (1995). „On lower bounds for the smallest eigenvalue of a Hermitian positive-definite matrix“. *IEEE T. Inform. Theory* 41(2), 539–540 (cit. on p. 33).

Maciak, M., Peštová, B., and Pešta, M. (2018). „Structural breaks in dependent, heteroscedastic, and extremal panel data“. *Kybernetika* 54(6), 1106–1121 (cit. on p. 46).

Mack, T. (1993). „Distribution-free calculation of the standard error of chain ladder reserve estimates“. *ASTIN Bull.* 23(2), 213–225 (cit. on pp. 53, 55, 57, 64, 75, 76).

Mack, T. (1994a). „Measuring the variability of chain ladder reserve estimates“. In: *Proceedings of the Casualty Actuarial Society Forum, Spring 1994*, pp. 101–182 (cit. on p. 56).

Mack, T. (1994b). „Which stochastic model is underlying the chain ladder model?“ *Insur. Math. Econ.* 15, 133–138 (cit. on pp. 53, 56).

Madurkayová, B. (2011). „Ratio type statistics for detection of changes in mean“. *Acta Universitatis Carolinae: Mathematica et Physica* 52(1), 47–58 (cit. on pp. 9, 40).

McCullagh, P. and Nelder, P. (1989). *Generalized Linear Models*. 2nd. Chapman and Hall/CRC: London, UK (cit. on p. 68).

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer: New York, NY (cit. on p. 78).

Pan, W. (2001). „Akaike's information criterion in generalized estimating equations“. *Biometrics* 57(1), 120–125 (cit. on p. 68).

Patton, A. J. (2012). „A review of copula models for economic time series“. *J. Multivariate Anal.* 110, 4–18 (cit. on pp. 75, 76).

Pešta, M. (2011). „Strongly consistent estimation in dependent errors-in-variables“. *Acta Universitatis Carolinae. Mathematica et Physica* 52(1), 69–79 (cit. on pp. 27, 28).

Pešta, M. (2013a). „Asymptotics for weakly dependent errors-in-variables“. *Kybernetika* 49(5), 692–704 (cit. on p. 27).

Pešta, M. (2013b). „Total least squares and bootstrapping with application in calibration“. *Statistics* 47(5), 966–991 (cit. on pp. 27, 28, 30).

Pešta, M. (2016). „Unitarily invariant errors-in-variables estimation“. *Stat. Pap.* 57(4), 1041–1057 (cit. on pp. 27, 28).

Pešta, M. (2017). „Block bootstrap for dependent errors-in-variables“. *Commun. Stat. A-Theor.* 46(4), 1871–1897 (cit. on pp. 27, 28).

Pešta, M. and Hudecová, Š. (2012). „Asymptotic consistency and inconsistency of the chain ladder“. *Insur. Math. Econ.* 51(2), 472–479 (cit. on pp. 38, 54).

Pešta, M. and Okhrin, O. (2014). „Conditional least squares and copulae in claims reserving for a single line of business“. *Insur. Math. Econ.* 56, 28–37 (cit. on p. 73).

Pešta, M. and Wendler, M. (2019). „Nuisance-parameter-free changepoint detection in non-stationary series“. *TEST* doi.org/10.1007/s11749-019-00659-1, Online First (cit. on p. 20).

Peštová, B. and Pešta, M. (2015). „Testing structural changes in panel data with small fixed panel size and bootstrap“. *Metrika* 78(6), 665–689 (cit. on pp. 9, 19, 37, 47).

Peštová, B. and Pešta, M. (2016). „Erratum to: Testing Structural Changes in Panel Data with Small Fixed Panel Size and Bootstrap“. *Metrika* 79(2), 237–238 (cit. on pp. 9, 37, 47, 51).

Peštová, B. and Pešta, M. (2017). „Change point estimation in panel data without boundary issue“. *Risks* 5(1), 7 (cit. on p. 47).

Peštová, B. and Pešta, M. (2018a). „Abrupt change in mean using block bootstrap and avoiding variance estimation“. *Comput. Stat.* 33(1), 413–441 (cit. on pp. 9, 19).

Peštová, B. and Pešta, M. (2018b). „Variance Estimation Free Tests for Structural Changes in Regression“. In: *Nonparametric Statistics*. Ed. by P. Bertail, D. Blanke, P.-A. Cornillon, and E. Matzner-Løber. Vol. 250. Springer: Cham, Switzerland, pp. 357–373 (cit. on p. 27).

Politis, D. N. and Romano, J. P. (1992). „A circular block-resampling procedure for stationary data“. In: *Exploring the Limits of Bootstrap*. Ed. by R. LePage and L. Billard. Wiley: New York, NY, pp. 263–270 (cit. on pp. 8, 15).

Politis, D. N. and White, H. (2004). „Automatic Block-Length Selection for the Dependent Bootstrap“. *Economet. Rev.* 23, 53–70 (cit. on p. 18).

Renshaw, A. E. (1994). *On the second moment properties and the implementation of certain GLIM based stochastic claims reserving models*. Actuarial Research Paper 65. Department of Actuarial Science and Statistics, City University, London (cit. on p. 69).

Renshaw, A. E. and Verrall, R. J. (1998). „A stochastic model underlying the chain-ladder technique“. *British Actuarial Journal* 4, 903–923 (cit. on p. 53).

Rho, Y. and Shao, X. (2015). „Inference for time series regression models with weakly dependent and heteroscedastic errors“. *J. Bus. Econ. Stat.* 33(2), 444–457 (cit. on p. 19).

Rosenblatt, M. (1971). *Markov Processes: Structure and Asymptotic Behavior*. Springer-Verlag: Berlin (cit. on p. 11).

Shao, X. (2011). „A simple test of changes in mean in the possible presence of long-range dependence“. *J. Time Ser. Anal.* 32(6), 598–606 (cit. on p. 19).

Shao, X. (2010). „The dependent wild bootstrap“. *J. Am. Stat. Assoc.* 105(489), 218–235 (cit. on p. 19).

Shao, X. and Zhang, X. (2010). „Testing for change points in time series“. *J. Am. Stat. Assoc.* 105(491), 1228–1240 (cit. on pp. 19, 21, 32).

Sherman, R. E. (1984). „Extrapolating, Smoothing, and Interpolating Development Factors“. *Proceedings of the Casualty Actuarial Society* 72, 122–155 (cit. on pp. 76, 77).

Shi, P. and Frees, E. W. (2011). „Dependent loss reserving using copulas“. *ASTIN Bull.* 41(2), 449–486 (cit. on p. 74).

Tweedie, M. C. K. (1984). „An index which distinguishes between some important exponential families“. In: *Statistics: Applications and New Directions*. Ed. by J. K. Ghosh and J. Roy. Calcutta: Indian Statistical Institute. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, pp. 579–604 (cit. on p. 67).

Van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM: Philadelphia, PA (cit. on pp. 27, 30).

Wang, D., Guo, P., and Xia, Z. (2016). „Detection and estimation of structural change in heavy-tailed sequence“. *Commun. Stat. A-Theor.* 46(2), 815–827 (cit. on pp. 9, 19).

Wenhua, C. and Hao, J. (2016). „Ratio testing for changes in the long memory indexes“. *International Business and Management* 12(3), 62–70 (cit. on p. 21).

White, H. (1982). „Maximum likelihood estimation of misspecified models“. *Econometrica* 50, 1–25 (cit. on p. 70).

Wu, C.-F. J. (1986). „Jackknife, bootstrap and other resampling methods in regression analysis“. *Ann. Stat.* 14(4), 1261–1295 (cit. on p. 19).

Wüthrich, M. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*. Wiley: New York, NY (cit. on pp. 53, 57, 67, 69, 70).

Wütrich, M. V. (2003). „Claims reserving using Tweedie's compound Poisson model“. *ASTIN Bull.* 33(2), 331–346 (cit. on p. 67).

Zhang, T. and Lavitas, L. (2018). „Unsupervised self-normalized change-point testing for time series“. *J. Am. Stat. Assoc.* 113(522), 637–648 (cit. on p. 19).

Zhao, W., Xia, Z., and Tian, Z. (2011). „Ratio test to detect change in the variance of linear process“. *Statistics* 45(2), 189–198 (cit. on pp. 8, 19).

Zhao, W., Tian, Z., and Xia, Z. (2010). „Ratio test for variance change point in linear process with long memory“. *Stat. Pap.* 51, 397–407 (cit. on p. 8).

Ziegler, A. (2011). *Generalized Estimating Equations*. Vol. 204. Lecture notes in statistics. Springer: New York, NY (cit. on pp. 63, 64, 70, 71).

Appendix: Scientific Papers

A

> *Change is the end result of all true learning.*
>
> **— Leo Buscaglia**
> (American writer and motivational speaker)

[1] Peštová, B. and Pešta, M. (2018). 'Abrupt change in mean using block bootstrap and avoiding variance estimation'. *Computational Statistics* 33(1), 413–441.
doi.org/10.1007/s00180-017-0785-4

[2] Pešta, M. and Wendler, M. (2019). 'Nuisance-parameter-free changepoint detection in non-stationary series'. *TEST*, Online First.
doi.org/10.1007/s11749-019-00659-1

[3] Pešta, M. (2016). 'Unitarily invariant errors-in-variables estimation'. *Statistical Papers* 57(4), 1041–1057.
doi.org/10.1007/s00362-016-0800-9

[4] Pešta, M. (2017). 'Block bootstrap for dependent errors-in-variables'. *Communications in Statistics – Theory and Methods* 46(4), 1871–1897.
doi.org/10.1080/03610926.2015.1030423

[5] Peštová, B. and Pešta, M. (2015). 'Testing structural changes in panel data with small fixed panel size and bootstrap'. *Metrika* 78(6), 665–689.
doi.org/10.1007/s00184-014-0522-8

[6] Peštová, B. and Pešta, M. (2016). 'Erratum to: Testing structural changes in panel data with small fixed panel size and bootstrap'. *Metrika* 79(2), 237–238.
doi.org/10.1007/s00184-015-0562-8

[7] Peštová, B. and Pešta, M. (2017) 'Change point estimation in panel data without boundary issue'. *Risks* 5(1), 7.
doi.org/10.3390/risks5010007

[8] Pešta, M. and Hudecová, Š. (2012). 'Asymptotic consistency and inconsistency of the chain ladder'. *Insurance: Mathematics and Economics* 51(2), 472–479. doi.org/10.1016/j.insmatheco.2012.07.004

[9] Hudecová, Š. and Pešta, M. (2013). 'Modeling dependencies in claims reserving with GEE'. *Insurance: Mathematics and Economics* 53(3), 786–794. doi.org/10.1016/j.insmatheco.2013.09.018

[10] Pešta, M. and Okhrin, O. (2014). 'Conditional least squares and copulae in claims reserving for a single line of business'. *Insurance: Mathematics and Economics* 56, 28–37. doi.org/10.1016/j.insmatheco.2014.02.007