Estimation and Detection of Model Instabilities using Regularization and Self-normalization

Matúš Maciak

2022

CHARLES UNIVERSITY
Faculty of mathematics
and physics

Charles University

Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Habilitation Thesis

Estimation and Detection of Model Instabilities using Regularization and Self-normalization

Matúš Maciak

Mathematics – Probability and Mathematical Statistics

January 21, 2022

Matúš Maciak

Habilitation Thesis
January 21, 2022

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

... to all I love
to all who care ...

Abstract

This habilitation thesis summarizes some of the author's contributions in mathematics—probability and mathematical statistics in particular—related to the research area well-known among mathematical statisticians as the *changepoint problem*.

Changepoints—or model instabilities alternatively—introduce a very important and substantial direction in advanced statistical modeling requiring sophisticated concepts and complex methodological backgrounds with many open problems. In this thesis, the changepoints are approached from various theoretical and practical perspectives using different stochastic models and generally very non-restrictive technical assumptions beyond an existing changepoint literature. There is always a formal mathematical theory, proper stochastic model, and a complex methodological framework postulated throughout the whole thesis. All this together—based on rigorous mathematical proofs—allows for an advanced changepoint analysis involving an implicit *changepoint detection* and an explicit *changepoint estimation*—all within a well formulated underlying probabilistic model—thus, under the presence of uncertainty, random fluctuations, and unobserved disturbances.

Preface

> *Most people use statistics like a drunk man uses a lamppost; more for support than for illumination.*
>
> — **Andrew Lang**
> (Scottish poet, novelist, and literary critic)

Model instabilities and the changepoint problem in general attract a huge attention with an outstanding research going on in this area over the last years. However, this is not only the case of scientists—mathematicians and mathematical statisticians in particular—but, taken a much wider perspective, the changepoint problem affects—roughly speaking—everyone and the whole population all at once.

Government authorities around the world try to instantly react to changes due to the most recent Covid-19 infection spread, financial experts deal with unprecedented volatility of asset markets when managing portfolios, all kinds of companies and institutions want to control their daily updated data, varying resources, or uncertain employee staff, and basically every living individual does its best to adapt to sudden changes and the most recent (random) outcomes. This all can be mathematically—and more importantly statistically—formulated in terms of some changepoint problem or, alternatively but equivalently, it can be all described and addressed by some underlying stochastic model with instabilities.

If such mathematical approach is adopted and the changepoint problem and instabilities are analyzed within a proper stochastic framework then the mathematical statistic itself becomes a lamppost that indeed serves for illumination as it can provide very useful, relevant, and most importantly valid conclusions. Moreover, such conclusions can be further used is a straightforward way to govern practical decisions basically related to almost all areas of human lives.

This was also an indigenous intention, later percolated as the main aim, of this habilitation thesis and the overall author's research: to ensure that the lamppost

is truly used for illumination purposes and, thus, the conclusion drawn from an advanced mathematical changepoint analysis of models with instabilities are stochastically valid and practically useful.

The first chapter provides a rather vague motivation of the changepoint problem itself and some important related mathematical research is sketched out. The main problems are described and the main goals of this thesis are formulated there as well. The central part of the thesis—the second and the third chapter—focuses on two very particular aspects of mathematical statistics when dealing with models with instabilities: The estimation of changepoint effects (change magnitudes respectively) is considered in different models in Chapter 2 and the detection of the changepoint presence (change existence respectively) is formally addressed in Chapter 3. Both of these aspect are, however, closely related and they can be even considered together in an omnibus model briefly mentioned at the end of the third chapter.

The most important facts are again gathered together with some final remarks in Chapter 4 and all (seven) scientific papers which serve as the theoretical core of this thesis are provided in full at the end, in Appendix[1].

[1] The appendix itself may or may not be the part of this habilitation thesis—depending on the thesis version (long/short) and specific requirements of the Department of Research and International Affairs of the Faculty of Mathematics and Physics, Charles University.

Acknowledgement

I would like to express my deepest thanks to all my family and all my friends. I am very grateful for your kind words, your support and encouragement, for any help when needed. It was always highly appreciated. Thank you very much for being around and thank you very much for everything.

Many thanks are also addressed to my colleagues at Charles University or other universities around the world who I had a chance to visit or collaborate with. Thank you for introducing me into the world of real mathematics. I am thankful for your attitudes, many pieces of advice, and inspiring ideas I could learn from.

The research comprehended in this thesis was partially supported by the GAČR grants (No.18-00522Y, No.21-10768S, No.18-01781Y, and No.21-13323S) and the Mobility grant of the Ministry of Educations, Youth, and Sports in the Czech Republic (No.7AMB17FR030). The author would like to also express thanks to Banff International Research Station (BIRS) in Banff, Canada and Mathematisches Forschungsinstitut Oberwolfach (MFO) in Oberwolfach, Germany for their great support during the '*research in teams*' (BIRS) and '*research in pairs*' (MFO) programs (14rit206 and RiP-R2057).

London, December 24, 2021

Matúš Maciak

Contents

1 Introduction

> *The statistics of yesterday should not be used to solve the problems of tomorrow.*
>
> **— John Tukey (1915 – 2000)**
> (American mathematician and statistician)

Instabilities of any types are usually considered to be system flaws, imperfections in general. This may seem to become even more serious issue when such instabilities come into scene in sophisticated theoretical models—methodological frameworks or data structures used to base complex decisions on—especially when trying to govern some underlying non-deterministic mechanism.

On the other hand, when the instabilities are approached with more caution, using a proper mathematical approach, rigorous stochastic background, and valid statistical inference, imperfections may turn into valuable assets—undeniable pieces of precious information with huge relevance with respect to the underlying process.

In this thesis, formal mathematical and stochastic theory is advocated in order to achieve such goals. Instabilities of various types—abrupt changes, gradual structural breaks, or changepoints—are all treated using sophisticated statistical models while focusing at two specific methodological aspects which are typical in practical situations: a) estimation of the changepoint magnitude(s) and b) detection of the changepoint presence. The first one tries to correctly quantify the effect caused by the underlying change while the second one aims to answer a relatively simple question whether some change indeed occurs in the system or not.

Both of these aspects are handled from a formal statistical—data-based and data-driven—perspective: the mathematical theory and the underlying data generated by some stochastic mechanism (which is the primary target of our interest) are used together in conjunction with the indigenous intention to pursuit the given goal.

1.1 Motivation

Instabilities, or changepoints in general, attract a lot of attention in recent years. Continuous and even smooth character of theoretical models used to describe real data generating processes turns out to be very often rather unrealistic and various model improvements have been proposed over the last decades in order to more closely adapt the models that are used with the data which are observed.

Some of the first initiatives in this direction appeared already in late 60's and 70's when Hinkley (1969) and Brown et al. (1975) assumed simple parametric regression lines in two separate stages—before and after a possible changepoint—and, after that, the resulting change was incorporated into the final model in terms of a jump in an intercept parameter, or a jump in a slope parameter respectively. In 90's, an analogous idea was also elaborated for more complex scenarios by Müller (1992) and Loader (1996) as they proposed specific generalizations of nonparametric models to account for a changepoint in location and a changepoint in location and direction simultaneously (see also Wu and Chu (1993) and Eubank and Speckman (1994)).

In the following years, a huge effort resulting in many different proposals has been dedicated to this problem—the statistical changepoint analysis in particular—while considering a whole range of realistic or even rather unrealistic models and assumptions. From the overall statistical perspective, the existing changepoint literature could fall into three main categories: It either refers to a *changepoint estimation* within some underlying regression framework, or it deals with the problem of the *changepoint detection* in terms of a formal statistical test, or both together—performed either simultaneously or in a step-by-step manner. This thesis deals with all aforementioned situations.

Considering the theoretical/methodological point of view, relatively straightforward techniques for the changepoint detection and estimation are based on standard approaches and two-stage procedures where the unknown changepoint location is firstly detected and, later, the underlying model is estimated separately before and after the detected change using already the knowledge about the estimated changepoint location learned from the first phase (Csörgő and Horváth (1988); Yao (1988); Bai and Perron (1998); Qiu and Yandell (1998); Kim et al. (2009b), and many others). If more changepoints with unknown locations are assumed, the detection and estimation is typically performed in multiple stages and the final model is usually selected using the likelihood framework (Hinkley (1971); Feder (1975)), statistical (permutation) tests (Kim et al. (2000); Bosetti et al. (2008); Qui et al. (2009)), or vari-

ous selection criteria and Bayesian framework (Carlin et al. (1992); Liu et al. (1997); Tiwari et al. (2005); Zhang and Siegmund (2007); Martinez-Beneito et al. (2011)).

Furthermore, advanced (usually data-driven and mostly just one stage) strategies typically rely on standard theoretical assumptions—independent observations, Gaussian (or sub-Gaussian respectively) error distributions, and linearity induced by the L_2-norm based objective function in particular (Chu et al. (1996); Horváth and Kokoszka (2002); Horváth et al. (2004); Antoch et al. (2006); Friedrich et al. (2008); Jeng et al. (2010); Chan and Walther (2013); Frick et al. (2014), etc.). Non-standard but, on the other hand, also way more realistic assumptions (for instance, dependent or extreme data, heavy-tailed error distributions, outlying observations, heteroscedasticity, multidimensionality, nonstationarity, or even nonlinearity) are imposed rather rarely and only partially and such model assumptions only appeared recently over the last years. This is mainly due to the fact that a) theoretical properties are challenging to derive under non-linearity and non-Gaussian error distributions and b) practical utilization of such models is almost impossible from the computational point of view without efficient algorithms enabling intensive computations and advanced IT equipment. The primary focus in this thesis is particularly given to the models under such atypical, non-standard, and generally non-restrictive assumptions.

Finally, there are two specific conceptual approaches always adopted and further elaborated in the scientific papers summarized in this thesis when dealing with the *changepoint estimation* and the *changepoint detection* problem under various model/data scenarios. Both of these concepts only percolated into the statistical methodology in the most recent years mainly due to an enormous growth of publicly available data and, also, a huge burst of machine learning algorithms and artificial intelligence applications. The first concept—*regularization*—allows for an effective filtering of relevant information out of huge amount of mostly irrelevant data. The second concept—*self-normalization*—implements a convenient definition of fundamental quantities avoiding some rather redundant calculations while introducing a fully data-driven statistical testing approach.

The *changepoint estimation* based on the *regularization* principle is described in details in Chapter 2 and some relevant theoretical aspects and statistical properties are elaborated in four scientific papers comprehended in this chapter. Similarly, the *changepoint detection* with the *self-normalization* concept plays the key role in Chapter 3 where another three original scientific papers are used as the main theoretical pivots of this chapter.

1.2 Research rudiments

The research results summarized exclusively in this thesis were all fully obtained during the tenure track period at Charles University (Prague, Czech Republic) which started in January 2015.

As already evoked above, the theoretical and methodological core of the thesis is grounded in two seemingly unrelated ideas, both with very strong individual mathematical background and rich personal experience. The first idea comes from the theory of *estimation and detection* of changepoints in advanced regression models and complex data structures—contemplated and studied, for instance, in Maciak (2010), Maciak and Hušková (2017), or Maciak (2018a)—all during the author's Ph.D. studies at Charles University (Czech Republic). The second idea arose later, during the author's postdoctoral stay at University of Alberta (Edmonton, Canada), and it takes an advantage of modern concepts in mathematical statistics, recent achievements in atomic pursuit methods, and algorithmic improvements in computer intensive calculations (Maciak and Mizera (2016); Maciak (2018b); Maciak (2018c), or Maciak and Mizera (2022)), all considered with respect to a popular phenomena of so called "big data" which typically requires some form of *regularization and self-normalization* in order to effectively handle finite sample problems.

The research was also partially driven by emergent problems occurring over time in various real life situations (epidemiology, ecology, econometrics & insurance, pharmacology & medicine, engineering, etc.) and it was catalyzed by numerous collaborations with many experts from top universities and research institutions around the World (University of British Columbia, McGill, University of Alberta, University of Lyon, Hamburg University, Cologne University, Dresden University of Technology, Czech University of Life Sciences Prague, Mayo Clinic, European Commission Joint Research Centre, Scottish Environment Protection Agency, T.G.Masaryk Water Research Institute, etc.).

Some of the results were also successfully applied and rich experience was gained in practical collaborations with, for instance, Škoda automobile manufacturer, Generali ČP insurance company, Erst & Young, F.D. Roosevelt Teaching Hospital, Motol University Hospital, Central Military Hospital Ružomberok, and others.

1.3 Thesis structure & results

The thesis is based on seven original theoretical (mathematical and statistical) papers. All seven papers are published in well recognized international impact

factor journals. The papers are structured into two intuitive groups (four and three papers) and for each group there is one thesis chapter entirely devoted to a detailed description of the main results derived and proved in the given papers.

Different *regularization concepts* utilized for the *changepoint estimation* problem within various models and data structures are described in Chapter 2. In particular, Chapter 2 is based on the following four manuscripts:

- ❏ Ciuperca, G. and Maciak, M. (2020). 'Changepoint Detection by the Quantile LASSO Method'. *Journal of Statistical Theory and Practice* 14(11), 1–38.
  DOI: 10.1007/s42519-019-0078-z
- ❏ Ciuperca, G. and Maciak, M. (2019). 'Change-point detection in a linear model by adaptive fused quantile method'. *Scandinavian Journal of Statistics* 47(1), 425–463.
  DOI: 10.1111/sjos.12412
- ❏ Ciuperca, G., Maciak, M., and Wahl, F. (2020). 'Detection of similar successive groups in a model with diverging number of variable groups'. *Sequential Analysis* 39(1), 92–114.
  DOI: 10.1080/07474946.2020.1726687
- ❏ Maciak, M. (2019). 'Quantile LASSO with changepoints in panel data models applied to option pricing'. *Econometrics and Statistics* 20(2021), 166–175.
  DOI: 10.1016/j.ecosta.2019.12.005

In Chapter 3, *self-normalization concepts* employed for the *changepoint detection* problem in terms of formal statistical tests are introduced and briefly described. In particular, Chapter 3 is based on the following three papers:

- ❏ Maciak, M., Peštová, B., and Pešta, M. (2018). 'Structural breaks in dependent, heteroscedastic, and extremal panel data'. *Kybernetika* 54(2018), 1106–1121.
  DOI: 10.14736/kyb-2018-6-1106
- ❏ Maciak, M., Pešta, M., and Peštová, B. (2020). 'Changepoint in dependent and non-stationary panels'. *Statistical Papers* 61(2020), 1385–1407.
  DOI: 10.1007/s00362-020-01180-6
- ❏ Pešta, M., Peštová, B., and Maciak, M. (2020). 'Changepoint estimation for dependent and non-stationary panels'. *Applications of Mathematics* 65(3), 299–310.
  DOI: 10.21136/AM.2020.0296-19

Some additional comments and important remarks are again highlighted in the conclusion in Chapter 4. Finally, all seven scientific papers are attached in Appendix[1], at the end of the thesis. The key theoretical contributions are always summarized at the end of the following two chapters and, also, in a paper-wise manner, in the appendix part.

[1]The appendix itself may or may not be the part of this habilitation thesis—depending on the thesis version (long/short) and specific requirements of the Department of Research and International Affairs of the Faculty of Mathematics and Physics, Charles University.

2 Estimation via Regularization

> *I never guess. It is a shocking habit destructive to the logical faculty.*
>
> **— Sir Arthur Conan Doyle**
> (British writer and creator of Sherlock Holmes)

In this chapter, we describe four different changepoint models being considered in four underlying scientific papers which frame the main theoretical core of this section. The models are formulated in terms of their increasing mathematical and structural complexity and, also, somehow more general and less restrictive practical applicability. On the other hand, the main theoretical pivots remain the same for all considered models throughout the chapter and they always rely on six fundamental cornerstones.

Firstly, hypothetical (uknown) changepoints are always implemented in the model by using some conveniently formulated *sparsity principle*. Second, the final model is obtained by minimizing an objective function together with some appropriate regularization penalty while performing the changepoint estimation and the changepoint detection both at once. Third, very general and relatively non-restrictive theoretical assumptions are considered and all theoretical proofs are exclusively derived under such conditions. Fourth, additional robustness and a complex characterization of the underlying data-generating mechanism is acquired by estimating arbitrary conditional quantiles rather than just the conditional mean. Fifth, important theoretical properties are derived and the proposed methodology is proved to be consistent in some desirable asymptotic sense. Last but not least, straightforward applicability is guaranteed due to a convex formulation and the overall changepoint estimation and detection is performed within a fully data-driven framework.

The sparsity principle and regularization are both used simultaneously with a proper mathematical and statistical theory in order to construct valid estimates for the unknown changepoints in the underlying data generating model. The number of true changepoints, their corresponding magnitudes, and the changepoint locations are all supposed to be apriori unknown and appropriate finite sample surrogates are to be found for all.

2.1 Penalized changepoint estimation

The most crucial problem in stochastic modeling and statistics in general lies in a proper trade-off between the precision of a constructed estimate and its volatility. Flexible models tend to be very precise when adapting to some underlying data but uncertainty increases correspondingly. On the other hand, simple models can be preferred because of their generally smaller volatility but the overall precision with respect to the observed data may be relatively low.

It is not an easy task to decide which model from a set of all plausible models should be used. Various selection criteria can be adopted, expert knowledge can be brought into the decision process, or, alternatively, complex models handling the so-called "bias-variance trade-off" automatically, on a data-driven basis, can be used instead. This is also the case of the penalized model estimation—the changepoint *estimation via regularization* addressed on the following lines.

The first ideas of utilizing the regularized (changepoint) estimation can be already tracked back to the work of Logan (1965) and Breiman (1995) who proposed an estimation approach for finding a (sparse) subset regression model while shrinking and zeroing irrelevant coefficients. Thus, initial models are defined by relatively many coefficients (i.e., unknown parameters) but the final solutions always contain—by principle—many zeros. This is known as the sparsity principle. Initially, such regularization techniques and sparsity principles were especially popular in signal processing and filtering problems (Donoho and Stark (1989); Donoho and Logan (1992); Chen et al. (1998)) and, only later, they were also popularized in mathematical statistics—mostly by Tibshirani (1996).

The main idea is to use a model which is overparametrized (too flexible) in some sense (having an ability to easily adapt itself to the given data) while some additional penalty is implemented in the estimation algorithm to automatically balance the overall bias-variance trade-off—not allowing the final model to become too rugged. In a very vague and general form, the model estimated via regularization can be expressed by the minimization problem

$$\textbf{Estimated model} \quad = \quad \operatorname{Arg\,min} \Big[\text{Objective function} \; + \; \text{Penalty term} \Big] \qquad (2.1)$$

where the *objective function* depends on the underlying model formulation and the given data while the *penalty term* usually depends on some tuning parameter, the model formulation, and possibly the data as well. The right-hand side is minimized with respect to a class of all plausible models and the outcome of this minimization—the left-hand side—provides the final (regularized) model.

In the following, we firstly briefly describe three specific forms of the penalty term in (2.1) which will be further used for the regularized changepoint estimation in the models introduced later. Similarly, the objective function will be also specified later when providing explicit model formulations in Section 2.2.

2.1.1 Total variation penalty

Vaguely speaking, the total variation penalty is essentially a standard L_1-norm of some form of a model based derivative (Dodge, 1987) derived from the total variation concept of an infinitesimal L_1-norm (Jordan, 1881). Historically, the L_1-norm methods in mathematical statistics go back to Galileo and Laplace where, in comparison with more common least squares with a closed form linear solution and relatively simple calculations, the L_1-based estimation is nonlinear and computationally more challenging.

In terms of a sparse estimation via regularization, the L_1-norm plays its main role in the penalty term in (2.1)—-enforcing sparsity in the estimated model. For simplicity, if the model is defined in a way that some of its derivative is piece-wise constant in terms of some consecutive parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)^\top \in \mathbb{R}^n$ then the corresponding total variation penalty can be expressed as

$$\lambda \sum_{i=1}^{n-1} |\beta_{i+1} - \beta_i|, \tag{2.2}$$

for some tuning parameter $\lambda > 0$. Such penalty directly penalizes differences between each two consecutive coefficients and the overall sparsity is, therefore, interpreted in terms of zero differences $\beta_{i+1} - \beta_i$, which should ideally hold for many $i \in \{1, \dots, n-1\}$ but some few exceptions—changepoints. This type of the penalty term is implemented in the first theoretical paper summarized in this thesis (Ciuperca and Maciak, 2019a). Such penalty can be also found in, for instance, Rudin et al. (1992), Künsch (1994), Mammen and Geer (1997), Tibshirani et al. (2005), Kim et al. (2009a), Tibshirani (2014), or Sadhanala and Tibshirani (2019).

2.1.2 Lasso type penalty

The lasso type penalty firstly appeared only implicitly in a constrained minimization proposed in Breiman (1995) while an explicit formulation was given later in Tibshirani (1996). The idea is, again, based on the L_1-norm and, despite the fact that the theoretical principle behind is different, the lasso penalty can be seen,

in some (computational) sense, as an analogy of the total variation penalty. Using a reparametrization $\theta_i = \beta_{i+1} - \beta_i$, for $i = 1, \ldots, n-1$, it is easy to see that

$$\lambda \sum_{i=1}^{n-1} |\beta_{i+1} - \beta_i| = \lambda \sum_{i=1}^{n-1} |\theta_i|,$$

where the right-hand side of the equation is typically known as a lasso type penalty being typically expressed as

$$\lambda \|\boldsymbol{\theta}\|_1, \tag{2.3}$$

where $\|\cdot\|_1$ stands for a standard L_1-norm and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n-1})^\top \in \mathbb{R}^{n-1}$. The lasso type penalty is also applied and further elaborated for the changepoint estimation problem in the first theoretical paper of this thesis (Ciuperca and Maciak, 2019a) where some similarities with the total variation type penalty are discussed and additional theoretical results are proved. The lasso type penalty can be also found in the work of Zou and Hastie (2005), Tibshirani (2011), Tibshirani and Taylor (2011), Tibshirani and Taylor (2012), Lee et al. (2016), and many others.

2.1.3 Group lasso penalty

Finally, the group lasso penalty originates in the previous two versions: By considering a model being determined by a set of consecutive multivariate parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n \in \mathbb{R}^p$ it may be of interest, similarly as in (2.2), to again compare two consecutive quantities—vectors in this case—which can be straightforwardly done using the L_2-norm and the expression

$$\lambda \sum_{i=1}^{n-1} \|\boldsymbol{\beta}_{i+1} - \boldsymbol{\beta}_i\|_2, \tag{2.4}$$

again for some $\lambda > 0$ where $\|\cdot\|_2$ stands for the L_2-norm (i.e., $\|\boldsymbol{u}\|_2 = \sqrt{\sum_{i=1}^p u_i^2}$ for some $\boldsymbol{u} = (u_1, \ldots, u_p)^\top \in \mathbb{R}^p$). The expression in (2.4) can be seen a multivariate generalization of the total variation penalty in (2.2). Despite the fact that the L_2-norm appears in its definition, the nature of the whole penalty term is still grounded within the L_1-norm framework. Indeed, it is easy to see, that for univariate parameters $\beta_1, \ldots, \beta_n \in \mathbb{R}$ the penalty in (2.4) reduces to (2.2) due to the L_1 and L_2-norm definitions, where

$$\lambda \sum_{i=1}^{n-1} \|\beta_{i+1} - \beta_i\|_2 = \lambda \sum_{i=1}^{n-1} \sqrt{(\beta_{i+1} - \beta_i)^2} = \lambda \sum_{i=1}^{n-1} |\beta_{i+1} - \beta_i|.$$

Going back to the multivariate case, an analogous reparametrization as before can be used (e.g., $\boldsymbol{\theta}_i = \boldsymbol{\beta}_{i+1} - \boldsymbol{\beta}_i$, for $\boldsymbol{\theta}_i \in \mathbb{R}^p$ and $i = 1, \ldots, n-1$) to obtain a multivariate analogy of the lasso type penalty in (2.3), as it holds that

$$\lambda \sum_{i=1}^{n-1} \|\boldsymbol{\beta}_{i+1} - \boldsymbol{\beta}_i\|_2 = \lambda \sum_{i=1}^{n-1} \|\boldsymbol{\theta}_i\|_2 = \lambda \sum_{i=1}^{n-1} \sqrt{\theta_{i,1}^2 + \cdots + \theta_{i,p}^2}, \tag{2.5}$$

where $\boldsymbol{\theta}_i = (\theta_{i,1}, \ldots, \theta_{i,p})^\top$. The last expression above again simply reduces to (2.3) in a univariate case (if $\theta_1, \ldots, \theta_n \in \mathbb{R}$). The expression on the right-hand side of (2.5) is known as a group lasso type penalty.

The intuition behind the group lasso type penalty can be, however, also seen from another perspective. For Assuming again univariate parameters $\beta_1, \ldots, \beta_n \in \mathbb{R}$ it may be sometimes appropriate to define disjoint groups

$$\underbrace{\beta_1, \ldots, \beta_{i_1}}_{\text{group 1}}, \underbrace{\beta_{i_1+1}, \ldots, \beta_{i_2}}_{\text{group 2}}, \ldots, \ldots, \beta_{i_{K-1}}, \underbrace{\beta_{i_{K-1}+1}, \ldots, \beta_n}_{\text{group K}},$$

for $K \in \mathbb{N}$ groups being defined by indexes $0 = i_0 < i_1 < \cdots < i_K = n$. Such groups can be considered separately in the group lasso type penalty of the form

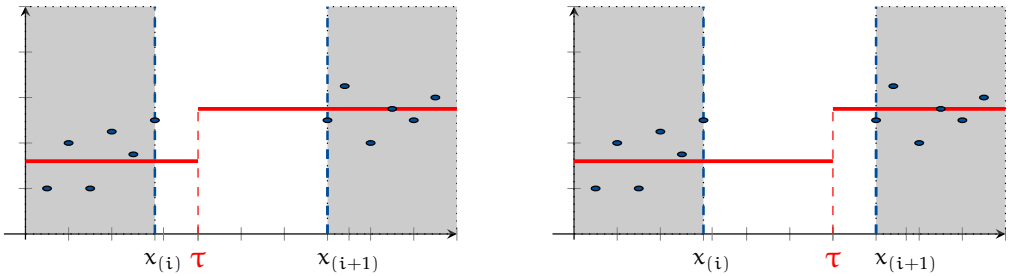
$$\lambda \sum_{k=0}^{K-1} \sqrt{\beta_{i_k+1}^2 + \cdots + \beta_{i_{k+1}}^2},$$

achieving sparsity with respect to the underlying groups in the final model. The whole principle of the parameter groups can be easily generalized for multivariate parameters as well. The group lasso type penalties are implemented in the models proposed in the second, third, and fourth theoretical paper of this thesis (Ciuperca and Maciak (2019b), Ciuperca et al. (2020), and Maciak (2021b)) where advanced theoretical properties are investigated and proved.

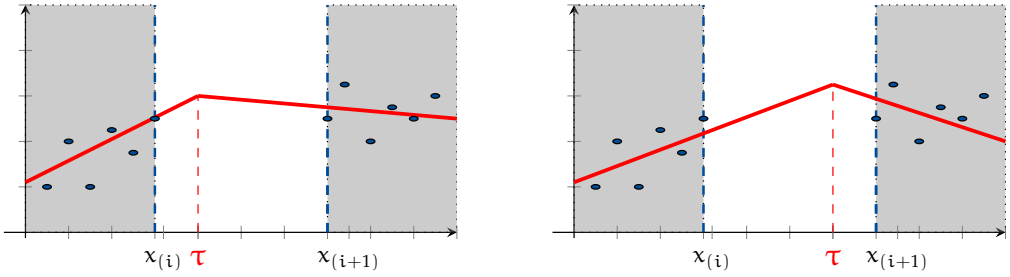
Similar penalties are also addressed, for instance, in Tibshirani and Saunders (2005), Yuan and Lin (2006), Friedman et al. (2010), Simon and Tibshirani (2012), Lim and Hastie (2015), and others.

2.2 Changepoints models via sparsity

The key idea in this section (and all four underlying theoretical papers considered in this chapter) is to use the *sparsity principle* (model over-parametrization) for incorporating hypothetical changepoints—various types of instabilities—into the underlying model. Of course, different models can be considered and different changepoint hierarchy structures can be implemented within such models (sudden



(a) Same data points, different changepoint location τ, but equivalent models



(b) Same data points, different changepoint location τ, and different models

Fig. 2.1. Illustration of the effect of the imposed smoothness assumption: For a piece-wise constant model (top two panels) there is no information about the true changepoint location given in the data if there are no more observations given between $x_{(i)}$ and $x_{(i+1)}$ (white area) and both models (left and right panels) are equivalent (in gray shaded areas). In contrary, for any higher order of smoothness (e.g., piece-wise linear models in the bottom panels), there is already some implicit information about the true changepoint location available in the (same) data and both models (left and right again) are clearly different (in the gray shaded areas).

breaks, jumps in higher order derivatives, or some combinations of both). However, different smoothness assumptions imposed on the underlying model induce also different restrictions regarding the ability to estimate the true changepoint locations. This—rather philosophical problem—is illustrated in Figure 2.1. From the empirical point of view the problem will be practically disregarded by assuming changepoint locations occurring only at the available observations.

Pioneering attempts to combine the changepoint estimation in some underlying model and the sparsity principle with regularization can be tracked back to Tibshirani and Wang (2008) and Harchaoui and Lévy-Leduc (2010) where the underlying model is conveniently over-parametrized such that unknown changepoints may occur—hypothetically speaking—at each available observation. The regularization penalty enforces the overall sparsity in terms of picking up only a few relevant changepoints while the remaining ones are eliminated from the final model—the corresponding parameter estimates are shrunk to zero (see also Rinaldo (2009), Ciuperca (2014), Tibshirani (2014), Leonardi and Bühlmann (2016), or Qian and Su (2016b)). All these approaches, however, involve simple (linear) model structure and standard L_2-norm based minimization. Thus, the resulting model is interpreted

in terms of a standard conditional mean. This is convenient for interpretation purposes but it only offers a very limited insight about the underlying stochastic nature of the model itself and, more importantly, it strictly determines the set of imposed model assumptions required for the theoretical proofs of the model validity.

Different techniques can be used to "over-parametrize" the underlying stochastic model in order to allow for changepoint occurrences. In the following four papers, there are four specific models considered under the four specific data structures being all motivated by four real-life problems. All four models and the corresponding data types generated by the assumed models are illustrated in Figure 2.2.

2.2.1 Location model

A relatively very simple model is considered in the first paper (Ciuperca and Maciak, 2019a). However, the strength of the proposed model does not rely on its explicit formulation but the main advantage of the model is rather brought in by a complex estimation framework and a set of simple and non-restrictive assumptions considered for the mathematical proofs of the model validity.

Independent observations $Y_1, \dots, Y_n$, for $n \in \mathbb{N}$, are given together with a specific location structure (i.e., stochastic model)

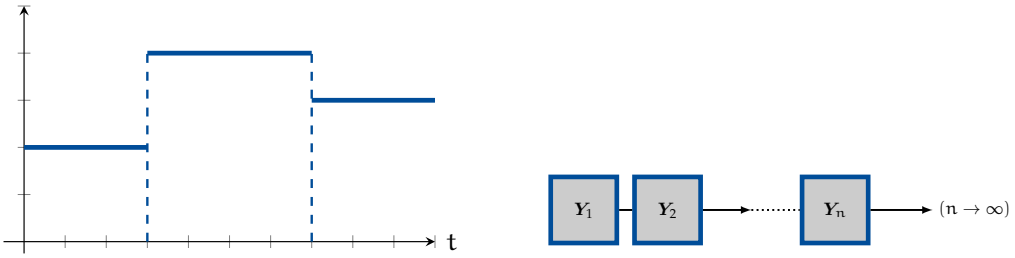
$$Y_t = \mu_t + \varepsilon_t, \qquad t = 1, \dots, n, \tag{2.6}$$

where, hypothetically, each random variable Y_t may have its own mean parameter $\mu_t \in \mathbb{R}$ while no specific restrictions are imposed on random error terms $\{\varepsilon_t\}_{t=1}^n$. The sparsity principle is introduced by an additional assumption that not necessarily all μ_t parameters are different. More precisely, it is assumed that the observations are somehow naturally ordered and $\mu_{t+1} - \mu_t \neq 0$ holds only for some small subset of indexes $t \in \{1, \dots n-1\}$.

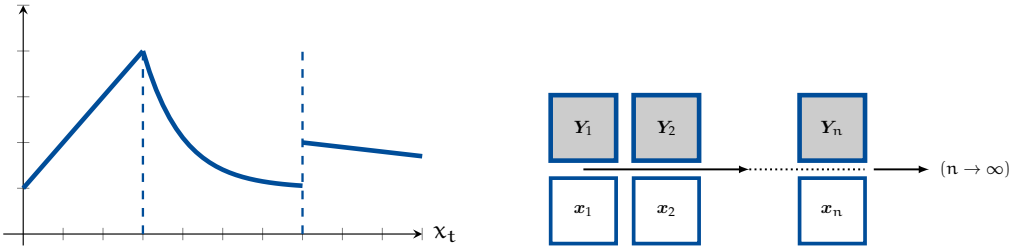
In other words, for some unknown changepoint locations—indexes $t_1^* \dots t_{K^*}^* \in \{1, \dots, n\}$, such that $1 < t_1^* < t_2^* < \cdots < t_{K^*}^* < n$ with an unknown number of total changepoints $K^* \in \mathbb{N}$, the model in (2.6) can be also expressed as

$$Y_t = \mu_k^* + \varepsilon_t, \quad \text{for} \quad t = 1, \cdots, n, \quad k = 1, \cdots, K^* + 1, \quad t_{k-1}^* \leqslant t \leqslant t_k^* - 1, \tag{2.7}$$

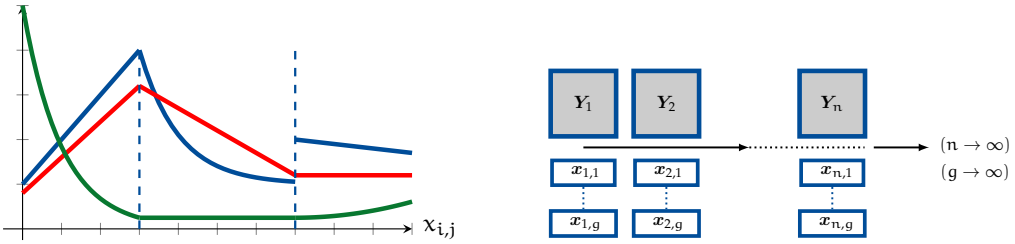
where $t_0^* = 1$ and $t_{K^*+1}^* = n + 1$. The key goal is to estimate the overall number of changepoints $K^* \in \mathbb{N}$, the corresponding changepoints locations (indexes $t_1^* \dots t_{K^*}^* \in \{1, \dots, n\}$), and, also, the set of the true mean parameters $\mu_k^* \in \mathbb{R}$, for $k = 1, \dots, K^* + 1$. As far as $K^* \in \mathbb{N}$ is fixed and $n \to \infty$, it is obvious that mostly it will hold that $\mu_{t+1} - \mu_t = 0$ (sparsity in terms of many zero differences of



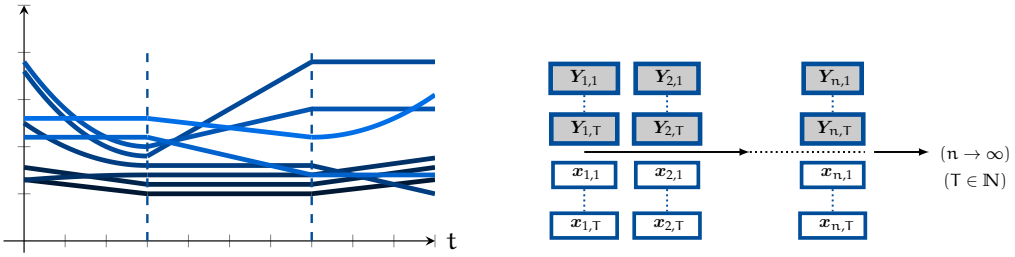
(a) piece-wise constant profile



(b) non-constant, non-smooth, and possibly discontinuous profile



(c) finite set of non-constant, non-smooth, and possibly discontinuous profiles



(d) infinitely many non-constant, non-smooth, and possibly discontinuous profiles

Fig. 2.2. Illustration of four underlying models (piece-wise constant (location) model (top row), linear regression model (second row), group linear model (third row), and panel data model in the last row) on left panels and the corresponding data structures generated by the assumed models on the right panels. Dashed vertical lines (in left panels) denote the locations of two true changepoints. Different smoothness assumptions, changepoint hierarchy structures, and the overall data complexity are assumed across all four models.

the consecutive parameters). The underlying model and the corresponding data structure are both illustrated in Figure 2.2a. Analogous model formulations were considered by many authors (see, for instance, Yao and Au (1989); Mammen and Geer (1997); Lavielle and Moulines (2000); Boysen et al. (2009); Harchaoui and Lévy-Leduc (2010); Frick et al. (2014); Fryzlewicz (2014); Hyun et al. (2016); Li et al. (2019); Yu and Chatterjee (2020)).

The key difference in our approach is the assumption regarding the error distribution. In all the aforementioned papers the authors consider Gaussian (sub-Gaussian respectively) error distributions and a standard L_2-norm objective function. Our approach is, therefore, more robust, free of restrictive distributional assumptions, complex in terms of the overall insight given by the model (estimating any conditional quantile rather than just the conditional mean), and consistent in terms of the changepoint detection and the model estimation when the sample size increases ($n \to \infty$). In addition, the proposed estimation approach offers a fully data-driven methodological framework and effective algorithms for obtaining the solution.

The estimation of the model in (2.6) is based on the minimization problem

$$\widehat{\boldsymbol{u}} = \underset{u_1,\ldots,u_n \in \mathbb{R}^n}{\operatorname{Arg\,min}} \left[\sum_{i=1}^{n} \rho_\tau(Y_i - u_i) + n\lambda_n \sum_{i=1}^{n-1} |u_{i+1} - u_i| \right], \tag{2.8}$$

where $\widehat{\boldsymbol{u}} = (\widehat{u}_1, \ldots, \widehat{u}_n)^\top \in \mathbb{R}^n$ are the given estimates and $\rho_\tau(x) = x(\tau - \mathbb{I}_{(x<0)})$ is so-called quantile check function (see Koenker (1998) for more details) defined for any $\tau \in (0,1)$ and $x \in \mathbb{R}$.

The total variation type penalty in (2.8) penalizes for too flexible models in terms of too many changepoints and the regularization parameter $\lambda_n > 0$ controls for the overall number of changepoints appearing in the final model: for $\lambda_n \to 0$ the minimization in (2.8) results in $\widehat{\boldsymbol{u}}$ where $\widehat{u}_t \neq \widehat{u}_{t-1}$, for each $t = 2, \ldots, n$, while for $\lambda_n \to \infty$ one gets $\widehat{u}_t = \widehat{u}_{t-1}$, for all $t = 2, \ldots, n$. The estimated changepoint locations are defined as indexes $\widehat{t} \in \{1, \ldots, n-1\}$, for which $\widehat{u}_{\widehat{t}+1} \neq \widehat{u}_{\widehat{t}}$, and the true number of changepoints $K^* \in \mathbb{N}$ is estimated as $\widehat{K} = \#\{\widehat{t} \in \{1, \ldots, n-1\};\ \widehat{u}_{\widehat{t}+1} \neq \widehat{u}_{\widehat{t}}\}$. All together, there are $\widehat{K}$ estimated changepoints at the locations $\widehat{t}_1 < \cdots < \widehat{t}_{\widehat{K}}$ which specify the final (regularized) piece-wise constant model with $\widehat{K}+1$ pieces $\widehat{\mu}_k \in \mathbb{R}$, for $k = 1, \ldots, \widehat{K}+1$. The minimization problem in (2.8) is convex and standard optimization toolboxes can be used to obtain the solution.

The theoretical results derived and proved in Ciuperca and Maciak (2019a) are based on a careful inspection of the Karush-Kuhn-Tucker (KKT) optimality conditions determined by the minimization problem formulated in (2.8)—see Kuhn and Tucker (1951) for details.

Specifically, the following statistical properties are derived and proved in the paper:

(a) efficiency in terms of all true changepoints being detected

$$\mathbb{P}[\widehat{K} < K^*] \to 0, \qquad \text{as } n \to \infty;$$

(b) consistency with respect to the true changepoint location estimation if $\widehat{K} = K^*$

$$\mathbb{P}\left[\max_{1 \leqslant k \leqslant K^*} |\widehat{t}_k - t_k^*| \geqslant n\delta_n\right] \to 0, \qquad \text{for } n \to \infty \text{ and some } \delta_n \to 0;$$

(c) consistency with respect to the true changepoint location estimation if $\widehat{K} > K^*$

$$\mathbb{P}\left[\sup_{1 \leqslant k^* \leqslant K^*} \inf_{1 \leqslant k \leqslant \widehat{K}} |\widehat{t}_k - t_{k^*}^*| \geqslant n\delta_n\right] \to 0, \qquad \text{for } n \to \infty \text{ and } \delta_n \to 0;$$

(d) consistency with respect to the true model parameters estimation

$$|\widehat{\mu}_k - \mu_k^*| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right), \qquad \text{for any } k = 1, \cdots, K^* + 1.$$

In addition, some further technical details are provided and finite sample comparisons with respect to various competitive approaches are performed using an extensive simulation study. Practical applicability is illustrated on a hot spot region detection in CGH genomic hybridization data. The full paper is in Appendix A.1.

2.2.2 Multiple regression model

In the second paper (Ciuperca and Maciak, 2019b), the location model from Section 2.2.1 is generalized for a multiple regression framework and the overall performance of the estimation algorithm is further improved by implementing an adaptive version of the penalty term. Analogous properties as for the model in (2.6) are investigated from the stochastic point of view but the more general model formulation and the adaptive penalty in addition require some more advanced mathematical theory and more complex derivations in general.

The underlying data are assumed to be of the form $\{(Y_t, \boldsymbol{x}_t);\ t = 1, \ldots, n\}$, where, in addition to random variables $\{Y_t\}_{t=1}^n$, there are also some subject specific vectors of covariates $\boldsymbol{x}_t \in \mathbb{R}^p$. The formal mathematical model can be expressed as

$$Y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta}_t + \varepsilon_t, \qquad t = 1, \cdots, n, \tag{2.9}$$

where $\boldsymbol{\beta}_t = (\beta_{t,1}, \ldots, \beta_{t,p})^\top \in \mathbb{R}^p$ for $t = 1, \ldots, n$ are some unknown p-dimensional vectors of unknown parameters which may change from one observation to another. Nevertheless, there is again a specific sparsity structure imposed as it is supposed to hold that $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t+1}$ for almost all indexes $t \in \{1, \ldots, n-1\}$ but a few exceptions—changepoints—while some underlying (natural) ordering is implicitly expected. Similarly as before, the key task is to identify the true number of changepoints, the corresponding indexes where $\boldsymbol{\beta}_t \neq \boldsymbol{\beta}_{t+1}$ (changepoint locations detection), and to find finite-sample surrogates for the unknown vector parameters (model estimation). The model and the corresponding data are visualized in Figure 2.2b.

Similar models were considered, for instance, in Zhang and Siegmund (2012), Zheng et al. (2013), Zheng et al. (2015), Lin et al. (2016), Lee et al. (2016), and others but always under the standard L_2-norm minimization. In our approach, we offer a complex insight by estimating arbitrary conditional quantiles—similarly as in Lee et al. (2018) where, however, the authors only considered one possible changepoint in the model. No specific distributional assumptions are postulated and the final model can be again obtained in a fully data-driven manner.

Moreover, in order to improve the overall performance of the total variation penalty (in terms of both, the changepoint detection and the model estimation) there is a two stage approach proposed in the paper. The first step is analogous with the estimation performed in Section 2.2.1 as the parameter estimates for the model in (2.9) are obtained as

$$(\overset{\vee}{\boldsymbol{\beta}}_1, \ldots, \overset{\vee}{\boldsymbol{\beta}}_n) = \underset{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n \in \mathbb{R}^p}{\operatorname{Arg\,min}} \left[\sum_{i=1}^{n} \rho_\tau(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_i) + n\lambda_n \sum_{i=2}^{n} \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i-1}\|_2 \right], \tag{2.10}$$

with the quantile check function $\rho_\tau(x) = x(\tau - \mathbb{I}_{\{x<0\}})$ where $\tau \in (0,1)$ defines the corresponding conditional quantile to be estimated and $\|\boldsymbol{v}\|_2 = \sqrt{v_1^2 + \cdots + v_p^2}$ is a standard L_2-norm for some p-dimensional real vector $\boldsymbol{v} = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$. The minimization formulation in (2.10) is again convex and the solution can be effectively obtained by standard optimization algorithms.

Depending on the choice of the regularization parameter $\lambda_n > 0$ one can control the overall number of changepoints occurring in the final model. For $\lambda_n \to 0$ there will be $\overset{\vee}{\boldsymbol{\beta}}_t \neq \overset{\vee}{\boldsymbol{\beta}}_{t-1}$ for each $t \in \{2, \ldots n\}$ while for $\lambda_n \to \infty$ the model reduces to a simple linear regression model with some $\overset{\vee}{\boldsymbol{\beta}} \in \mathbb{R}^p$, such that $\overset{\vee}{\boldsymbol{\beta}}_t = \overset{\vee}{\boldsymbol{\beta}}$ for all $t \in \{1, \ldots, n\}$.

The estimated changepoint locations are again identified by those indexes $t \in \{2, \ldots, n\}$ for which it holds that $\overset{\vee}{\boldsymbol{\beta}}_t \neq \overset{\vee}{\boldsymbol{\beta}}_{t-1}$. The true number of changepoints, $K^* \in \mathbb{N}$, is estimated by $\overset{\vee}{K}$ which is, similarly as before, the cardinality of the

set $\{t \in \{2,\dots,n\};\ \overset{\vee}{\beta}_t \neq \overset{\vee}{\beta}_{t-1}\}$. Finally, the true model parameters $\beta_t \in \mathbb{R}^p$, for $t = 1,\dots,n$, are estimated by the corresponding finite sample surrogates—regularized vector parameter estimates $(\overset{\vee}{\beta}_1,\dots,\overset{\vee}{\beta}_n) \in \mathbb{R}^{n\times p}$.

Analogous statistical properties as for the location model in Section 2.2.1 are also proved in Ciuperca and Maciak (2019b) for the model in (2.9) and the estimation procedure defined by (2.10). In addition, proper conference rates for the vector parameter estimates are derived. However, all these results only serve as a preliminary tool for improving the model performance and defining the adaptive penalty.

Estimation by the adaptive fused penalty

The modification in the model formulation is very minor. Differences in the theoretical proofs are substantial. Instead of the minimization problem in (2.10) with the standard group lasso type penalty we assume a modified version of the minimization problem where the final model is now estimated by solving a slightly modified minimization problem

$$(\widehat{\beta}_1,\dots,\widehat{\beta}_n) = \underset{\beta_1,\dots,\beta_n\in\mathbb{R}^p}{\operatorname{Arg\,min}} \left[\sum_{i=1}^{n} \rho_\tau(Y_i - x_i^\top \beta_i) + n\lambda_n \sum_{i=2}^{n} \omega_i \|\beta_i - \beta_{i-1}\|_2\right], \quad (2.11)$$

with some additional weights $\omega_i > 0$, for $i = 2,\dots,n$. The penalty term above is known as an adaptive version of the penalty used in (2.10). The weights depend on the differences between two consecutive parameter estimates obtained from the minimization problem in (2.10). Briefly, the weights are defined as

$$\omega_i = \left[\max\left(\|\overset{\vee}{\beta}_i - \overset{\vee}{\beta}_{i-1}\|_\infty;\ d_n\right)\right]^{-\gamma}$$

for $\|\cdot\|_\infty$ denoting the supremum norm, $d_n \to 0$ being some deterministic sequence, and $\gamma > 0$ where some additional reduction of the changepoints detected by (2.10) can be applied (more details can be found in the underlying paper—Ciuperca and Maciak (2019b)). Thus, the estimation with the adaptive penalty in (2.11) can not be performed without some initial pre-estimation step (in this case in terms of (2.10)). Analogously as before, the true changepoint locations are estimated by indexes $t \in \{2,\dots,n\}$ for which $\widehat{\beta}_t \neq \widehat{\beta}_{t-1}$ and for the overall number of the estimated changepoints we have $\widehat{K} = \#\{t \in \{2,\dots,n\};\ \widehat{\beta}_t \neq \widehat{\beta}_{t-1}\}$.

The main advantage of using the adaptive penalty is an additional improvement with respect to the changepoint detection performance. Standard total variation and lasso based penalties have a well-known tendency to overfit the final model.

This can be seen also from the results mentioned in Section 2.2.1 and explicitly derived in Ciuperca and Maciak (2019a) where it is proved that

$$\mathbb{P}[\widehat{K} < K^*] \to 0, \qquad \text{as } n \to \infty$$

but also

$$\mathbb{P}[\widehat{K} \leqslant CK^*] \to 1, \qquad \text{as } n \to \infty$$

for some constant $C < \infty$, usually greater than one. Thus, the number of the estimated changepoints using the standard total variation penalty—as in (2.8) or (2.10) respectively—overestimates the true number of the unknown changepoints. The adaptive penalty in (2.11) can effectively correct for this drawback while achieving—under some reasonable assumptions—a consistent detection in a sense

$$\mathbb{P}\Big[\{t \in \{2,\ldots,n\};\ \widehat{\beta}_t \neq \widehat{\beta}_{t-1}\} = \{t \in \{2,\ldots,n\};\ \beta_t \neq \beta_{t-1}\}\Big] \to 1, \qquad \text{as } n \to \infty.$$

To conclude, the theoretical results (based on the KKT optimality conditions and the theory of stochastic (quantile) processes) derived in the second theoretical paper of this chapter, include the following:

(a) all results analogous to those being derived for the location model in Section 2.2.1 are also derived and proved for the model in (2.9) considering both estimation approaches—the group lasso type penalty formulated in (2.10) and the adaptive version of the penalty given in (2.11);

(b) rates of convergence for the parameter estimates $\overset{\vee}{\beta}_1, \ldots, \overset{\vee}{\beta}_n$ obtained in (2.10);

(c) rates of convergence for the parameter estimates $\widehat{\beta}_1, \ldots, \widehat{\beta}_n$ obtained in (2.11);

(d) the true changepoints recovery with probability tending to one

$$\mathbb{P}\Big[\{t \in \{2,\ldots,n\};\ \widehat{\beta}_t \neq \widehat{\beta}_{t-1}\} = \{t \in \{2,\ldots,n\};\ \beta_t \neq \beta_{t-1}\}\Big] \to 1, \quad \text{as } n \to \infty.$$

In addition, some further technical derivations are provided in the paper and a detailed discussion is dedicated to a theoretical and empirical comparison of the model being estimated in terms of (2.10) and its adaptive version in (2.11). Extensive simulations and real data applications accompany the paper as well. The paper is given in full in Appendix A.2.

2.2.3 Group linear model

The third paper (Ciuperca et al., 2020) discusses a specific generalization of the model from Section 2.2.2 where, in addition, multiple covariates may form some consecutive groups—which are, however, unknown. Such models are frequently applied for econometric and financial data, market analysis, meteorological predictions, and other similar situations.

The main difference lies in the domain where the changepoints are expected to occur now. In the previous two models (Section 2.2.1 and Section 2.2.3) the observations were assumed to be somehow naturally ordered (for instance, over time $t = 1, \dots n$) and the changepoints occurred after some specific fraction of the data. In the model formulation used in this section, the changepoints are not implemented with respect to the consecutive observations but rather in terms of the given subject specific covariates (with an increasing/diverging dimension as $n \to \infty$).

From the mathematical point of view, the model can be formalized as

$$Y_i = \sum_{j=1}^{g} \boldsymbol{x}_{i,j}^\top \boldsymbol{\beta}_j + \varepsilon_i, \qquad i = 1, \dots, n, \tag{2.12}$$

where $(\boldsymbol{x}_{i,1}^\top, \dots, \boldsymbol{x}_{i,g}^\top)^\top \in \mathbb{R}^{p \times g}$ is a vector of subject specific covariates (for subject $i \in \{1, \dots, n\}$) being grouped into $g \in \mathbb{N}$ same sized groups where $\boldsymbol{\beta}_j \in \mathbb{R}^p$ for $j = 1, \dots, g$ are p-dimensional vectors of unknown parameters which correspond to the groups. Visual illustration of the model and the underlying data of the form $\{(Y_i, (\boldsymbol{x}_{i,1}^\top, \dots, \boldsymbol{x}_{i,g}^\top));\ i = 1, \dots, n\}$ is given in Figure 2.2c.

The key task is to identify consecutive groups for $j = 2, \dots, g$, such that the estimated effect of two consecutive groups on the response Y_i will be different. In other words, we want to identify two consecutive groups for which $\boldsymbol{\beta}_j \neq \boldsymbol{\beta}_{j-1}$. The sparsity and the regularization principle are again both used to achieve the goal. This time, however, more competitive models are proposed and their statistical properties are studied in the paper (standard L_2-norm and the quantile check function are used for the objective function and they are both combined either with the total variation type penalty or its adaptive modification).

Similar model also appears in Liu et al. (2018) for non-grouped variables, in Jiang et al. (2013) and Jiang et al. (2014) for a finite number of groups, in Wei and Huang (2010), Guo et al. (2015), or Campbell and Allen (2017) for Gaussian errors with the least squares approach, or Zhang and Xiang (2015) for non-Gaussian errors. Unlike all the aforementioned works, the primary focus in our work is given on a situation where the number of the consecutive groups diverges for $n \to \infty$

(thus, $g \to \infty$ as well). Moreover, no specific assumptions on the error distribution are postulated (distribution free approach) and the theoretical results are derived for the model estimation performed in terms of the standard least squares sense and, also, within a more complex conditional quantile estimation framework.

Quantile loss function

Similarly as in the models discussed in the previous sections, the minimization problem related to the model in (2.12) can be expressed as

$$(\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_g)_Q = \underset{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_g \in \mathbb{R}^p}{\operatorname{Arg\,min}} \left[\sum_{i=1}^{n} \rho_\tau\Big(Y_i - \sum_{j=1}^{g} \boldsymbol{x}_{i,j}^\top \boldsymbol{\beta}_j\Big) + n\lambda_n \sum_{j=2}^{g} \|\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j-1}\|_q \right], \tag{2.13}$$

for some $q \in \{1, 2\}$. It should be obvious from the form of the penalty term that the regularization parameter $\lambda_n > 0$ enforces now sparsity with respect to differences between two consecutive groups of subject specific covariates—unknown vector parameters $\{\boldsymbol{\beta}_j\}_{j=1}^g$. In general, it is expected that $\widehat{\boldsymbol{\beta}}_j = \widehat{\boldsymbol{\beta}}_{j-1}$ should hold for most of the indexes $j \in \{2, \ldots, g\}$.

The consistency (including proper convergence rates) of the parameter estimates $(\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_g)_Q$ in (2.13) is proved for three different regularization scenarios:

- consistency of the parameter estimates when $\lambda_n = 0$ (thus, no penalty term is considered in the minimization problem and no regularization is involved)

  $$\|\widehat{\boldsymbol{\beta}}^g - \boldsymbol{\beta}^0\|_1 = O_{\mathbb{P}}(b_n),$$

  for $\widehat{\boldsymbol{\beta}}^g = (\widehat{\boldsymbol{\beta}}_1^\top, \ldots, \widehat{\boldsymbol{\beta}}_g^\top)^\top \in \mathbb{R}^{p \times g}$, the true values $\boldsymbol{\beta}^0 \in \mathbb{R}^{p \times g}$ and some deterministic sequence $b_n \to 0$, as $n \to \infty$;

- consistency of the parameter estimates in (2.13) for some $\lambda_n > 0$ (estimation via regularization) in the same sense as before, i.e.

  $$\|\widehat{\boldsymbol{\beta}}^g - \boldsymbol{\beta}^0\|_1 = O_{\mathbb{P}}(b_n),$$

  however, for a generally different sequence $b_n \to 0$, which now also depends on $\lambda_n > 0$;

- consistency of the parameter estimates when the penalty term in (2.13) is replaced by an adaptive version with weights $\omega_j > 0$ being defined with

respect to some preliminary estimates from a pre-estimation step (in our case, the minimization in terms of (2.13)), similarly as in Section 2.2.2.

In addition, detailed theoretical derivations of explicit boundaries for the number of the estimated changepoints are provided for both scenarios—the regularized estimation in terms of (2.13) and, also, its adaptive version implemented in order to improve the overall finite sample performance.

Least squares loss function

For the least squares approach, the quantile check function $\rho_\tau(\cdot)$ in (2.13) is replaced with the standard L_2-norm, resulting in the minimization problem

$$(\widehat{\boldsymbol{\beta}}_1,\ldots,\widehat{\boldsymbol{\beta}}_g)_{LS} = \underset{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_g\in\mathbb{R}^p}{\operatorname{Arg\,min}} \left[\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{g} \boldsymbol{x}_{i,j}^{\top}\boldsymbol{\beta}_j\right)^2 + n\lambda_n \sum_{j=2}^{g} \|\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j-1}\|_q\right], \tag{2.14}$$

where $q \in \{1,2\}$. Similarly as before—when considering the quantile check function in (2.13)—three specific regularization scenarios are also considered for the least squares estimation approach and analogous results are derived and proved.

The consistency of the parameter estimates $(\widehat{\boldsymbol{\beta}}_1,\ldots,\widehat{\boldsymbol{\beta}}_g)_{LS}$ in (2.14) with the corresponding convergence rates are provided for

- $\lambda_n = 0$ (thus, no penalty term is considered in the minimization problem and no regularization is involved);
- some $\lambda_n > 0$ (estimation via regularization);
- the adaptive estimation when the penalty in (2.14) is replaced by an adaptive version with some weights $\omega_j > 0$ being defined analogously as in Section 2.2.2.

From the overall point of view, the proposed group model allows for very general modeling situations. The novelty of the model also lies within a unique combination of the objective function and the penalty term. In addition, non-restrictive assumptions imposed for the quantile estimation in particular introduce a very robust estimation framework and fully data-driven changepoint detection and estimation approach. The paper also contains some additional theoretical details and extensive finite sample comparisons with different competitive methods. The full paper is provided in Appendix A.3.

2.2.4 Panel data model

Finally, the last paper of this chapter (Maciak, 2021b) introduces a stochastic change-point model for partially dependent data usually known as *panel data*. Such types of data are also considered in the next chapter where some slightly different perspective is adopted.

From the structural point of view, the underlying model and the corresponding data can be considered to be the most complex from all four models postulated in this chapter so far. The overall illustrative comparison is given in Figure 2.2 (with the panel data structure and the panel data model in Fig. 2.2d).

Let us assume the underlying data of the form $\{(Y_{t,i}, \boldsymbol{x}_{t,i});\ t = 1, \dots, T, i = 1, \dots, n\}$, where $T \in \mathbb{N}$ is assumed to be fixed and $n \to \infty$. This is known as a (relatively complex) panel data scheme, where for each panel—individual $i \in \{1, \dots, n\}$—the response variable $Y_{t,i}$ is observed for some specific follow-up period $t \in \{1, \dots T\}$ (common for all subjects) while some additional subject specific and time specific information $\{\boldsymbol{x}_{t,i}\}_{i=1,t=1}^{n,T}$ may be provided (for p-dimensional vectors $\boldsymbol{x}_{t,i} \in \mathbb{R}^p$). As far as we are dealing with repeated observations here (within panel observations) it is reasonable to assume some form of dependence among $\{Y_{t,i}\}_{t=1}^{T}$. Individual panels are assumed to be independent for $i = 1, \dots, n$.

Such and similar models (usually with a simpler structure in terms of the overall model flexibility, the available information, and the data complexity) are very frequent in empirical econometric, for instance, for modeling a financial development of a set of companies, economic growth of some specific countries, or assessing a qualitative performance of various industrial businesses or indexes (Qian and Su (2014); Qian and Su (2016a); Qian and Su (2016b); Maciak et al. (2020b); Maciak (2021a); Maciak and Vitali (2022); Drábek et al. (2022)).

For simplicity and for the purposes of the underlying theoretical paper and this section correspondingly it is assumed that $\boldsymbol{x}_{t,i} = \boldsymbol{x}_i$ for all $i = 1, \dots, n$. Thus the subject specific information is stable over time. This can be nicely motivated, for instance, by using some option market and a riskiness analysis of some specific option contract in particular. The risk assessment is usually performed in terms of an implied volatility which is repeatedly observed over time for a certain set of the underlying strikes—panels $i = 1, \dots, n$. The strike value—panel label $x_i \in \mathcal{D} \subseteq \mathbb{R}$ represents the price of the underlying asset and $\boldsymbol{x}_i = (\varphi_1(x_i), \dots, \varphi_p(x_i))^\top \in \mathbb{R}^p$ can bee seen as some well-defined functional basis expansion over the domain $\mathcal{D}$ evaluated at the given strike price $x_i \in \mathcal{D}$.

The underlying stochastic model, for some independent random error vectors $\boldsymbol{\varepsilon}_i = [\varepsilon_{1,i}, \ldots, \varepsilon_{T,i}]$ (where the elements within each vector may form some unspecified dependent structure), can be formalized as

$$Y_{t,i} = \boldsymbol{x}_i^\top \boldsymbol{\beta}_t + \varepsilon_{t,i}, \quad i = 1, \ldots, n; \;\; t = 1, \ldots, T, \tag{2.15}$$

where the unknown vector parameters $\boldsymbol{\beta}_t \in \mathbb{R}^p$ may change at different time points $t \in \{1, \ldots, T\}$ in order to reflect the development of the underlying time dependence of $\{Y_{t,i}\}_{i=1}^n$ on $\{\boldsymbol{x}_i\}_{i=1}^n$. The sparsity principle is again imposed with respect to the consecutive differences of the unknown vector parameters $\boldsymbol{\beta}_t \in \mathbb{R}^p$ over time: the underlying dependence of $\{Y_{t,i}\}_{i=1}^n$ on $\{\boldsymbol{x}_i\}_{i=1}^n$ is supposed to be relatively stable over time—reflected by the fact that $\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t$ for most of the time points $t \in \{1, \ldots, T-1\}$ and, occasionally, the underlying dependence may change to adapt for the situation on the market—thus, $\boldsymbol{\beta}_{t+1} \neq \boldsymbol{\beta}_t$ for some unknown changepoint location—time point $t \in \{1, \ldots, T\}$.

The model formulated in (2.15) can be directly estimated by solving the convex minimization problem

$$(\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_T) = \underset{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T \in \mathbb{R}^p}{\operatorname{Arg\,min}} \left[\sum_{t=1}^{T} \sum_{i=1}^{n} \rho_\tau \left(Y_{ti} - \boldsymbol{x}_i^\top \boldsymbol{\beta}_t \right) + n\lambda_n \sum_{t=2}^{T} \|\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}\|_2 \right], \tag{2.16}$$

using again the quantile check function to estimate conditional quantiles for an arbitrary $\tau \in (0, 1)$ and the group lasso type penalty to achieve the final sparse (regularized) solution. The estimated changepoint locations are determined by the indexes $t \in \{2, \ldots, T\}$ where $\widehat{\boldsymbol{\beta}}_t \neq \widehat{\boldsymbol{\beta}}_{t-1}$ and the number of true changepoints is estimated by the cardinality of the set $\{t \in \{2, \ldots, T\};\; \widehat{\boldsymbol{\beta}}_t \neq \widehat{\boldsymbol{\beta}}_{t-1}\}$. Recall, that the follow-up period $T \in \mathbb{N}$ is fixed.

Additional restrictions—shape constraints—may be implemented to guarantee some specific qualitative properties. Illustrated again on the implied volatility and the financial theory on arbitrage free markets, it is desired to obtain a final model with is convex with respect to the strikes. Non-increasing property can be sometimes imposed as well. Both can be easily implemented into the minimization problem in (2.16) by solving the original minimization problem with respect to the linear constraints

$$\begin{aligned} &\boldsymbol{D}\boldsymbol{\beta}_t \leqslant \boldsymbol{0}, \; t = 1, \ldots, T; \quad \text{(non-increasing in the strike)} \quad \text{(C1)} \\ &\boldsymbol{C}\boldsymbol{\beta}_t \geqslant \boldsymbol{0}, \; t = 1, \ldots, T; \qquad \text{(convexity in the strike)} \quad \text{(C2)} \end{aligned} \tag{2.17}$$

where both inequality signs are meant in an element-wise manner and $\boldsymbol{D}$ and $\boldsymbol{C}$ are some well-structured matrices—derivatives of the basis expansion over the domain $\mathcal{D} \subseteq \mathbb{R}$ (see Maciak (2021b) for further technical details).

Different regularization sources

Beside the overall model complexity already mentioned above there might be another theoretical and computational issue involved: the *estimation via regularization* in terms of (2.16) under the constraints in (2.17) is simultaneously governed by two regularization sources which are, by their main principle, very different.

If the minimization problem in (2.16) is considered together with some optional shape restrictions in (2.17) then the final estimate is regularized with respect to its both domains. Firstly, same as in the previous models, the first regularization is provided by the penalty term which penalizes for too many changepoints occurring in the final model. This is controlled by the regularization parameter $\lambda_n > 0$. By the model definition in (2.15) the changepoints may only occur over time $t \in \{1, \ldots, T\}$ and, therefore, the first regularization source plays its role with respect to the time domain. However, there is also another regularization source implicitly present within the given (optional) shape restrictions in (2.17). Indeed, the final model can not be too rough with respect to the subject specific covariates $\{\boldsymbol{x}_i\}_{i=1}^n$ if, for instance, some convexity or monotonicity is required. Thus, the second regularization source—the linear constraints in (2.17)—plays its role with respect to the spatial domain $\mathcal{D} \subseteq \mathbb{R}$.

Of course, not necessarily both regularization sources must be considered simultaneously. For instance, for $\lambda_n \to 0$ there is no regularization with respect to the time domain and the underlying dependence of $\{Y_{t,i}\}_{i=1}^n$ on $\{\boldsymbol{x}_i\}_{i=1}^n$ changes at each time point $t \in \{1, \ldots, T\}$. On the other hand, when no shape restrictions are considered the final estimate becomes rough with respect to the spatial domain $\mathcal{D}$ and sometimes it may be useful to considered additional penalty terms to control for the overall bias-variance trade-off in the final model. This is, however, already beyond the scope of the underlying paper (Maciak, 2021b) and it is further discussed, for instance, in Maciak (2021a).

An extreme situation occurs when none of the two regularization sources is in effect during the estimation phase. In such scenarios the final model may even interpolate all available observation $\{Y_{t,i}\}_{t=1,i=1}^{T,n}$. Such model becomes useless from the practical point of view and, therefore, some caution is always needed when handling such data and such complex model.

The last paper of this chapter (Maciak, 2021b) provides the consistency result for the estimated vector parameters $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_T \in \mathbb{R}^p$ obtained in (2.16). Some additional theoretical details are given as well together with the finite sample investigation and a real option market application. The consistency proof derived in the paper

is based on the KKT optimality conditions and a detailed investigation of the theoretical properties of two specific stochastic (quantile) processes defined as

$$G_n(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T) = \sum_{t=1}^{T} \sum_{i=1}^{n} \rho_\tau \left(Y_{t,i} - \boldsymbol{x}_i^\top \boldsymbol{\beta}_t \right)$$

and

$$G_n^\star(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T) = G_n(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T) + n\lambda_n \sum_{t=2}^{T} \|\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}\|_2.$$

In particular, the novelty of the paper Maciak (2021b), which is attached in Appendix A.4., relies on the following:

- ❏ complex panel data structure with subject specific information being estimated in terms of the conditional quantiles—achieving robustness with respect to the error distributions and outlying observations while providing a complex insight about the underlying stochastic data generating mechanism;

- ❏ regularized changepoint estimation and detection with respect to the time domain performed by a convex minimization obtaining a fully data-driven methodological framework;

- ❏ additional shape properties of the final model (e.g., convexity, monotonicity) achieved automatically by implementing explicit linear constraints;

- ❏ asymptotic consistency with respect to an increasing number of the given panels ($n \to \infty$) derived for the unknown model parameter estimates

$$\|\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^*\|_1 = O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{n}} \right), \quad t = 1, \ldots, T.$$

In addition, a finite sample performance of the proposed model is investigated and compared with a competitive approach in terms of a simulation study and a practical application on options' implied volatility surface estimation under the arbitrage free conditions is illustrated.

2.3 Key contributions

- ❏ All proposed models allow for the changepoint detection and estimation under very general, non-restrictive, and widely applicable assumptions—in particular, no distributional restrictions, heavy-tailed or asymmetric error distributions, or outlying observations;
- ❏ A complex insight about the underlying probabilistic model—the data generating mechanism—is analyzed from the theoretical, as well as the empirical point of view—all by adopting the conditional quantile estimation;
- ❏ Detailed rigorous mathematical proofs of consistency with respect to various perspectives (such as the model parameters estimation, changepoints detection, changepoint locations estimation, or the estimation of the true number of changepoints) are provided;
- ❏ The final solutions are always obtained effectively and uniquely—all guaranteed by the convex minimization problem formulations where, in addition, the solution is obtained within a fully data-driven methodological framework;
- ❏ In all four papers, the formulated models and the proposed estimation approaches seem to empirically outperform standard estimation techniques which are usually based on more restrictive assumptions and less complex model definitions.

3 Detection via Self-normalization

> *The aim of science is to seek*
> *the simplest explanation of complex facts.*
>
> **— Alfred North Whitehead (1861-1947)**
> (English mathematician, author of Principia Mathematica)

Unlike the previous chapter, which was rather devoted to the problem of the *changepoint estimation* in some underlying model, the key interest of this chapter does not lie on the quantification of the change magnitude but, instead, it provides a statistically valid answer to a relatively very simple question: *"Did the change indeed occur in the model, or not?"* This is known as a *changepoint detection* problem.

The theoretical core of this chapter is based on three original theoretical papers. The ultimate goal of all three of them is to develop a fully data-driven methodological approach for testing a changepoint significance within a standard panel data structure—similar to the one used in the last model of the previous chapter—however, assuming a set of even less restrictive (universal) assumptions with respect to the given panels and the overall errors' dependence form. Firstly, formal statistical tests based on the *self-normalization* principle are constructed in a step-by-step manner. Second, various competitive test statistics are proposed and their theoretical and finite sample properties are investigated. Third, a so-called bootstrap add-on is suggested to finally achieve the given goal of having a fully data-driven detection method. Last but not least, the whole *changepoint problem* can be also considered within an omnibus model where the *changepoint estimation* and the *changepoint detection* play both their substantial roles simultaneously in one overall mathematical/statistical approach.

A complex *changepoint detection* problem—statistical changepoint tests in particular—developed from simpler two sample problems and some related statistical tests being applied for various real-life situations (Quandt (1960); Chow (1960); Bai (2006)). However, unlike the aforementioned situations where the changepoint locations are a priori known, for the changepoint detection problem the locations are supposed to be generally unknown. This of course introduces some additional complexity from both—the theoretical point of view as well as the computational one.

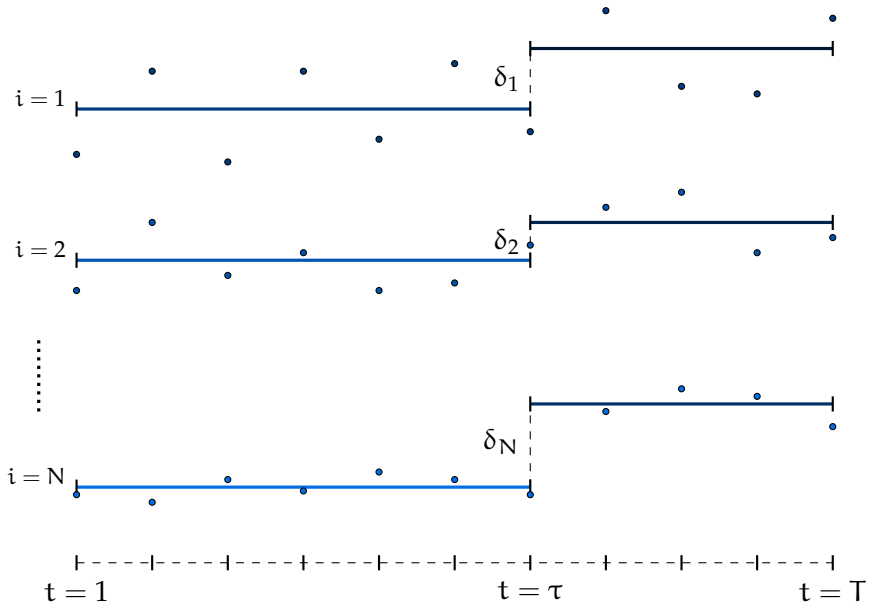


Fig. 3.1. Illustrative example of a simple but very general panel data structure. The location panels are allowed to be dependent, heteroscedastic, non-stationary, with a common changepoint location $\tau \in \{1, \ldots, T\}$ but different jump magnitudes and possibly an extremely short follow-up period $T \in \mathbb{N}$.

Restricting our attention to changepoint detection problems within simple panel data structures as illustrated in Figure 3.1, the underlying data, represented by $\{Y_{i,t}\}_{i=1,t=1}^{N,T}$, for $N \in \mathbb{N}$ and $T \in \mathbb{N}$, are assumed to satisfy

$$Y_{i,t} = \mu_i + \delta_i \mathbb{I}_{\{t>\tau\}} + \varepsilon_{i,t}, \tag{3.1}$$

for some panel specific mean parameters $\mu_i \in \mathbb{R}$, some common (but unknown) changepoint location $\tau \in \{1, \ldots, T\}$, with the corresponding panel specific jump magnitudes $\delta_i \in \mathbb{R}$, and some random error vectors $\boldsymbol{\varepsilon}_i = [\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]$. All quantities on the right-hand side of (3.1) are generally unknown and they are supposed to be estimated using the available data—observations $\{Y_{i,t}\}_{i=1,t=1}^{N,T}$. From the theoretical point of view, the number of subjects $N \in \mathbb{N}$ is allowed to tend to infinity while the follow-up period $T \in \mathbb{N}$ is supposed to be fixed and it can be even extremely short (referring mainly to the situations being explicitly discussed in this chapter).

The *changepoint detection* problem related to the panel data structure in (3.1) can be reformulated in terms of a formal statistical test where the null hypothesis of no changepoint presence is formally expressed as

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

while a general alternative hypothesis specifies the following:

$$H_0 : \tau < T \quad \wedge \quad \exists i \in \{1, \ldots, N\} \text{ such that } \delta_i \neq 0. \tag{3.3}$$

Referring to the existing literature, statistical tests in terms of the null hypothesis in (3.2) and the alternative hypothesis in (3.3) are performed under various scenarios. Typical assumptions mostly involve independent panels—either with homoscedastic errors (De Wachter and Tzavalis, 2012) or heteroscedastic ones (Pesaran (2006); Kim (2011); Baltagi et al. (2016)). In terms of the follow-up period specification, one can assume that $T \to \infty$, same as $N \to \infty$ (Chan et al. (2013b); Chan et al. (2013a); Horváth and Hušková (2012)), or, alternatively, the follow-up period may be considered to be fixed (Bai (2010); Peštová and Pešta (2015); Antoch et al. (2019), and others).

In what follows, we rather impose very general—non-restrictive—conditions, where the panels are allowed to be mutually dependent and the errors terms within each panel may again form some dependent sequence of random variables. The follow-up period is assumed to be fixed, $T \in \mathbb{N}$ but no further restrictions are given and it can be even also extremely short—as short as, for instance, 2–10 observations.

3.1 Panel data dependence

Dependent panel data structures appeared in statistical literature quite recently while there are mainly two different concepts being used to implement the mutual panel data dependence. Theoretically simpler idea is based on a common factor model investigated, for instance, in Kim (2014), Barigozzi et al. (2018), or Westerlund (2019). More complex dependence structures can be imposed by relaxing the random error assumptions—allowing for almost any arbitrary dependence form—similarly as in Cho (2016) or Bhattacharjee et al. (2019). Both of these approaches are subsequently considered in this chapter.

3.1.1 Common factors

The first theoretical paper of this chapter (Maciak et al., 2018) specifically deals with the changepoint detection within a dependent panel data structure where the dependence is imposed in terms of some common factors and the corresponding panel specific loadings. Assuming the observations $\{Y_{i,t}\}_{i=1,t=1}^{N,T}$, the underlying panel data structure can be expressed as

$$Y_{i,t} = \mu_i + \delta_i \mathbb{I}_{\{t>\tau\}} + \zeta_i \xi_t + \sigma_i \varepsilon_{i,t}, \quad i = 1, \dots, N, \;\; t = 1, \dots, T, \tag{3.4}$$

where $\mu_i \in \mathbb{R}$, for $i = 1, \dots, N$, are the unknown panel specific mean parameters, the true changepoint location is represented by an unknown time point $\tau \in \{1, \dots, T\}$, and the corresponding panel specific change magnitudes are $\delta_i \in \mathbb{R}$, again for

$i = 1, \ldots, N$. Note, the the formulation also allows for situations where only some panels are subjected to the change and for the remaining panels it may hold that $\delta_i = 0$. Random error vectors $\boldsymbol{\varepsilon}_i = [\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]^\top$ are independent and identically distributed for $i = 1, \ldots, N$ with a zero mean vector and some finite, positive-definite variance-covariance matrix. The random elements within each random vector $\boldsymbol{\varepsilon}_i$ are not assumed to be independent and they may form some fragment of a dependent sequence. Different variance parameters $\sigma_i > 0$, for $i = 1, \ldots, N$, may incorporate heteroscedasticity.

In addition, there is a mutual dependence between the panels introduced by implementing random factors $\{\xi_t\}_{t=1}^T$ and some panel specific (deterministic) loading parameters $\{\zeta_i\}_{i=1}^N$, which are usually left unknown. The level of the overall dependence between the panels is modeled by the magnitudes of the unknown loading parameters. If $\zeta_i = 0$ for some $i \in \{1, \ldots N\}$ then the panel i is independent of the remaining panels. The *changepoint detection* problem within the panel data structure in (3.4) reduces to a formal statistical test of the null hypothesis in (3.2) against the alternative hypothesis in (3.3).

For a more coherent text formulation, the proposed test statistics and some further theoretical details continue in Section 3.2.1. The paper itself (Maciak et al., 2018) is provided in Appendix A.5.

3.1.2 Dependent errors

Another way how to introduce a mutual dependence between the panels is more complex but it also allows for more general variance-covariance assumptions which may turn out to be far more realistic from the practical point of view. The underlying panel data structure is now formulated as

$$Y_{i,t} = \mu_i + \delta_i \mathbb{I}_{\{t>\tau\}} + \varepsilon_{i,t}, \quad i = 1, \ldots, N, \;\; t = 1, \ldots, T, \tag{3.5}$$

which may rather look as a simpler model than the one formulated in (3.4) however, the panel specific disturbances $\boldsymbol{\varepsilon}_i = [\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]$ are now allowed to form more general random sequences. Particularly, there is no stationarity assumed within the panels and, also, no stationarity among the panels. Again, the *changepoint detection* problem reduces to a formal statistical test of the same null hypothesis as in (3.2), against the alternative (3.3). The theoretical justification of the test must be, however, different and it is much more complex. The proposed test statistics and the corresponding statistical properties are addressed in the second paper of this chapter (Maciak et al., 2020a). The paper is given in full in Appendix A.6. Some more details are further provided in Section 3.2.2.

3.2 Formal statistical tests

A mathematical framework needed to perform a statistical test is typically based on some test statistic—a random variable—which is defined in a way that it becomes sensitive/large in situations described by the alternative hypothesis. More formally, if the null hypothesis in (3.2) is true, the test statistic should be rather small (it actually follows some specific probabilistic distribution usually concentrated around zero). On the other hand, if the alternative hypothesis (3.3) is valid, the test statistic should reflect this by diverging to infinity (in probability, as the sample size increases). In the following, there are four such test statistics proposed. Two ratio type statistics described in Section 3.2.1 are designed to handle the *changepoint detection* problem in the panel data structure in (3.4). The theoretical properties and the test consistency are proved in Maciak et al. (2018). Another two test statistics, specified in Section 3.2.2, are intended to hande the *changepoint detection* problem in the panel data structure in (3.5) and the theoretical details and the test validity are proved in Maciak et al. (2020a).

3.2.1 Ratio type statistics

Various types of the test statistic can be of course formulated to handle the changepoint detection problem in terms of the null hypothesis in (3.2) and the alternative hypothesis in (3.3). The most common approaches involve cumulative sum statistics, maximum type statistics, or some Cramér von Mises type statistic. In our approach, so-called ratio type statistics are used as they do not require any estimation of the nuisance variance parameters (see, for instance, Csörgő and Horváth (1997), Horváth et al. (2008), Liu et al. (2008), Chen and Tian (2014), or Pešta and Wendler (2019)) which will turn out to be very useful when approaching the ultimate goal of having a fully data driven changepoint detection framework.

In our particular situation, the following two competitive test statistics are proposed and they are in detail investigated in the underlying paper (Maciak et al., 2018)

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

and

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

where $\overline{Y}_{i,t}$ and $\widetilde{Y}_{i,t}$ respectively are the sample averages of the first t observations in the panel $i \in \{1,\ldots,N\}$ or the last $T-t$ observations respectively, i.e.,

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

The main reason for two different test statistics is the following: The first one performs more robustly with respect to possible outliers while the second one may have relatively more power to reject the null hypothesis. Otherwise, both test statistics are equivalent in terms of the underlying test which is formally performed by comparing the observed value of the test statistic with the theoretical quantile of the asymptotic distribution valid under the null hypothesis. In particular, it is derived in Maciak et al. (2020a) that

$$\mathcal{R}_N(T) \quad \xrightarrow[N\to\infty]{\mathcal{D}} \quad \max_{t=2,\ldots,T-2} \quad \frac{\max_{s=1,\ldots,t}\left|X_s - \frac{s}{t}X_t\right|}{\max_{s=1,\ldots,T-1}\left|Z_s - \frac{T-s}{T-t}Z_t\right|} \tag{3.6}$$

and

$$\mathcal{S}_N(T) \quad \xrightarrow[N\to\infty]{\mathcal{D}} \quad \max_{t=2,\ldots,T-2} \quad \frac{\sum_{s=1}^{t}\left(X_s - \frac{s}{t}X_t\right)^2}{\sum_{s=1}^{T-1}\left(Z_s - \frac{T-s}{T-t}Z_t\right)}, \tag{3.7}$$

where $Z_t = X_T - X_t$ and $[X_1,\ldots,X_T]^\top$ is a multivariate normal random vector with a zero mean vector and some specific variance covariance matrix $\boldsymbol{\Lambda} = (\lambda_{t,\nu})_{t=1,\nu=1}^{T,T}$, where $\lambda_{t,t}$ is the so-called cumulative autocorrelation function

$$\lambda_{t,t} = \mathsf{Var}\left[\sum_{s=1}^{t}\varepsilon_{i,s}\right],$$

for $t \in \{1,\ldots,T\}$, and $\lambda_{t,\nu}$ is a shifted cumulative correlation function defined as

$$\lambda_{t,\nu} = \lambda_{t,t} + \mathsf{Cov}\left(\sum_{s=1}^{t}\varepsilon_{i,s}, \sum_{u=t+1}^{\nu}\varepsilon_{i,u}\right),$$

for $t,\nu \in \{1,\ldots,T\}$ such that $t < \nu$. Statistical tests based either on the test statistic defined in (3.6) or the test statistic given in (3.7) are both proved to be consistent in a sense that the corresponding test statistics converge to infinity in probability as the number of panels increases ($N \to \infty$) and the alternative hypothesis holds true. Technical details, theoretical proofs, finite sample performance, and empirical comparisons of both ratio type test statistics under various scenarios are provided in Maciak et al. (2018), given in Appendix A.5.

3.2.2 Self-normalized statistics

More complex dependence structure are assumed in the second paper of this chapter (Maciak et al., 2020a) while the same principle of using the ratio-type statistics with *self-normalization* is advocated again.

Recall, that the underlying dependence between and within the panels is now fully determined by the structure of the panel specific error vectors $\boldsymbol{\varepsilon}_i = [\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]$. The corresponding test statistics, proposed in the paper, are defined as

$$\mathcal{Q}_N(T) = \max_{t=1,\ldots,T-1} \frac{|\mathcal{L}_N(t,T)|}{\max_{s=1,\ldots,t} |\mathcal{L}_N(s,t)| + \max_{s=t,\ldots,T-1} |\mathcal{R}_N(s,t)|} \tag{3.8}$$

and

$$\mathcal{S}_N(T) = \sum_{i=1}^{T-1} \frac{\mathcal{L}_N^2(t,T)}{\sum_{s=1}^{t} \mathcal{L}_N^2(s,t) + \sum_{s=t}^{T-1} \mathcal{R}_N^2(s,t)}, \tag{3.9}$$

for the cumulative sums of the partial residuals $\mathcal{L}_N(s,t) = \sum_{i=1}^{N} \sum_{r=1}^{s} (Y_{i,r} - \overline{Y}_{i,t})$ and $\mathcal{R}_N(s,t) = \sum_{i=1}^{N} \sum_{r=s+1}^{T} (Y_{i,r} - \widetilde{Y}_{i,t})$, where $\overline{Y}_{i,t}$ denotes the average of the first t observations in the panel $i \in \{1, \ldots, N\}$ and $\widetilde{Y}_{i,t}$ is the average of the last $T - t$ observations within the same panel, i.e., $\overline{Y}_{i,t} = \frac{1}{t} \sum_{s=1}^{t} Y_{i,s}$ and $\widetilde{Y}_{i,t} = \frac{1}{T-t} \sum_{s=t+1}^{T} Y_{i,s}$.

The asymptotic behavior of the test statistics in (3.8) and (3.9) under the null hypothesis in (3.2) is proven to be of the form

$$\mathcal{Q}_N(T) \xrightarrow[N\to\infty]{\mathcal{D}} \max_{t=1,\ldots,T-1} \frac{\left|X_t - \frac{t}{T} X_T\right|}{\max_{s=1,\ldots,t} \left|X_s - \frac{s}{t} X_t\right| + \max_{s=t,\ldots,T-1} \left|Z_s - \frac{T-s}{T-t} Z_t\right|} \tag{3.10}$$

and

$$\mathcal{S}_N(T) \xrightarrow[N\to\infty]{\mathcal{D}} \sum_{t=1}^{T-1} \frac{\left(X_t - \frac{t}{T} X_T\right)^2}{\sum_{s=1}^{t} \left(X_s - \frac{s}{t} X_t\right)^2 + \sum_{s=t}^{T-1} \left(Z_s - \frac{T-s}{T-t} Z_t\right)^2}, \tag{3.11}$$

where, again, $Z_t = X_T - X_t$ and $[X_1, \ldots, X_T]^\top$ is a multivariate normal random vector with a zero mean vector and the unknown (theoretical) variance covariance matrix

$$\boldsymbol{\Lambda} = \lim_{N\to\infty} \frac{1}{N} \mathrm{Var} \left[\sum_{i=1}^{N} \left(\sum_{s=1}^{1} \varepsilon_{i,s}, \ldots, \sum_{s=1}^{T} \varepsilon_{i,s} \right)^\top \right]. \tag{3.12}$$

The errors vectors $\boldsymbol{\varepsilon}_i = [\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]^\top$ are neither independent nor identically distributed. Similarly as before, the test of the null hypothesis in (3.2) against the

alternative in (3.3) based on the test statistic in (3.8) or (3.9) is proved to be consistent in a sense that both test statistics converge to infinity in probability if the alternative hypothesis holds true and the number of panels increases ($N \to \infty$).

Despite the fact that the ratio type statistics and the self-normalization principle are used, there are still some nuisance parameters that do not cancel out automatically when performing the test. This involves the covariance parameters reflecting not the true variance but, instead, the underlying dependence structure between and within the given panels. This can be, however, effectively handled by a proper resampling—a bootstrap add-on—which can avoid estimation of the covariance structure (the matrix $\boldsymbol{\Lambda}$ occurring in the asymptotic distribution of the ratio type statistics and the self-normalized test statistics).

3.3 Data-driven approach

The next step towards the ultimate goal of this chapter (a fully data driven changepoint detection approach) is to avoid the estimation of any nuisance parameters and tuning constants and to eliminate any non-expert—user-based—intervention during the changepoint detection process itself. As already mentioned above, this can be effectively handled by a proper bootstrap resampling algorithm but there is another challenge that arises with it: A consistent changepoint estimator $\widehat{\tau} \in \{1, \ldots, T-1\}$ is needed for bootstrap resampling. The tricky part of this step lies within the fact that it is usually not known whether the changepoint really occurs in the model (thus, some changepoint location estimate should be indeed given) or not (thus, no changepoint location estimate is needed).

This problem is closely studied in the third theoretical paper in this chapter and the last (seventh) paper of this habilitation thesis (Pešta et al., 2020). The paper is given in full in Appendix A.7.

3.3.1 Location estimation

In most situations considered in the literature, the changepoint estimator is constructed under the situation where it is apriori known that the changepoint occurs for sure (see, for instance, Pesaran (2006), Bai (2010), or Baltagi et al. (2016)). This is, however, clearly not the case in our situations, therefore, some different approach must be adopted. In the paper, we propose the estimator of the form

$$\widehat{\tau}_N = \operatorname*{Arg\,max}_{t=1,\ldots,T} \mathcal{U}_N(t), \tag{3.13}$$

where

$$\mathcal{U}_N(t) = \begin{cases} \frac{1}{t(T-t)} \sum_{i=1}^{N} \sum_{u=1}^{t} \sum_{v=t+1}^{T} (Y_{i,u} - Y_{i,v})^2 & t < T; \\ \frac{2}{(T-t)^2} \sum_{i=1}^{N} \sum_{v=2}^{T} \sum_{u=1}^{v-1} (Y_{i,u} - Y_{i,v})^2 & t = T. \end{cases}$$

This estimator does not suffer from any boundary issues (very common for other types of the estimators proposed in the literature), similarly as in Peštová and Pešta (2017) or Bardwell et al. (2019) but, in addition to the previous authors, our estimator is proved to be consistent under the same non-restrictive assumptions as those assumed for the panel data structure in (3.5) in Sections 3.1.2 and 3.2.2.

The consistency of the proposed nuisance parameters free estimator is proved regardless of the presence or absence of the changepoint in (3.5). In other words, if there is no changepoint present in in the panel data structure in (3.5), the proposed estimator points out at the very last time point $T \in \mathbb{T}$ with probability tending to one, as $N \to \infty$. An existence of such changepoint estimator allows us to finally build a fully data driven changepoint detection methodological framework—the ultimate goal of this chapter.

3.3.2 Bootstrap add-on

Because there are dependent panels considered in general, the Moving Block Bootstrap (MBB) algorithm is proposed to handle the underlying dependence properly (see, for instance, Künsch (1989), Dehling et al. (2015), or Betken and Wendler (2018)). MMB is based on resampled residuals however, the residuals are supposed to be resampled under the validity of the null hypothesis. This is where the proposed changepoint estimator from the previous section comes into play.

Using the proposed changepoint estimator, the residuals can be defined as

$$\widehat{e}_{i,t} = \begin{cases} Y_{i,t} - \overline{Y}_{i,\widehat{\tau}_N}, & t \leqslant \widehat{\tau}_N; \\ Y_{i,t} - \widetilde{Y}_{i,\widehat{\tau}_N}, & t > \widehat{\tau}_N; \end{cases} \tag{3.14}$$

where $\widehat{\tau}_N \in \{1, \ldots, T\}$ is the proposed (consistent) changepoint location estimator from Section 3.3.1 and $\overline{Y}_{i,\widehat{\tau}_N}$ and $\widetilde{Y}_{i,\widehat{\tau}_N}$ are the partial averages. The bootstrap is built on resampling with replacement of the blocks of residuals from $\{\widehat{e}_i\}_{i=1}^N$, where $\widehat{e}_i = [\widehat{e}_{i,1}, \ldots, \widehat{e}_{i,T}]$. The bootstrapped residuals are centered by their conditional expectation yielding the bootstrapped panel data structure of the form

$$\widehat{Y}_{i,t}^{(b)} = \widehat{e}_{i,t}^{(b)} - \frac{1}{N} \sum_{i=1}^{N} \widehat{e}_{i,t}. \tag{3.15}$$

Formal steps of the bootstrap algorithm are described below.

Procedure 3.3.1 Moving block bootstrap for the test statistics (3.8) and (3.9)

Input: Block-size b, number of bootstrap replications M, and panel data consisting of $N = bn$ panels with length T, i.e., N row vectors of the observations $[Y_{i,1}, \ldots, Y_{i,T}]$

Output: Bootstrap distribution of $\mathcal{Q}_N^{(b)}(T)$ and $\mathcal{S}_N^{(b)}(T)$, i.e., the empirical distribution where the probability mass $1/M$ concentrates at each of ${}_{(1)}\mathcal{Q}_N^{(b)}(T), \ldots, {}_{(M)}\mathcal{Q}_N^{(b)}(T)$ and ${}_{(1)}\mathcal{S}_N^{(b)}(T), \ldots, {}_{(M)}\mathcal{S}_N^{(b)}(T)$, respectively

estimate the changepoint by calculating $\widehat{\tau}_N$
compute residuals $\widehat{e}_{i,t}$ as in (3.14)
for $j = 1$ to $N - b + 1$ **do** // construct the blocks
    define a block of sub-panels $\mathbb{B}_j$, which is a $(b \times T)$ matrix having rows $\widehat{\boldsymbol{e}}_j, \ldots, \widehat{\boldsymbol{e}}_{j+b-1}$
end for
for $m = 1$ to M **do** // repeat in order to obtain the empirical distributions
    independently resample with replacement $(b \times T)$-dimensional blocks ${}_{(m)}\mathbb{B}_1^{(b)}, \ldots, {}_{(m)}\mathbb{B}_n^{(b)}$ from $\{\mathbb{B}_1, \ldots, \mathbb{B}_{N-b+1}\}$ with equal probability $1/(N - b + 1)$
    the MBB of resample of size N, denoted by $\{{}_{(m)}\widehat{e}_{i,t}^{(b)}\}_{i,t=1}^{N,T}$, is formed by joining (stacking) the resampled blocks ${}_{(m)}\mathbb{B}_1^{(b)}, \ldots, {}_{(m)}\mathbb{B}_n^{(b)}$ to one big block, i.e., $[{}_{(m)}\mathbb{B}_1^{(b)\top}, \ldots, {}_{(m)}\mathbb{B}_n^{(b)\top}]^\top = [{}_{(m)}\widehat{\boldsymbol{e}}_1^{(b)\top}, \ldots, {}_{(m)}\widehat{\boldsymbol{e}}_N^{(b)\top}]^\top$
    calculate bootstrap panel data ${}_{(m)}\widehat{Y}_{i,t}^{(b)}$ as in (3.15)
    compute bootstrap test statistics ${}_{(m)}\mathcal{Q}_N^{(b)}(T)$ and ${}_{(m)}\mathcal{S}_N^{(b)}(T)$ as in (3.8) and (3.9), where $Y_{i,t}$'s are replaced by $\widehat{Y}_{i,t}^{(b)}$'s
end for

The data-driven *changepoint detection* approach is built up in three consecutive steps and the theoretical results obtained for the corresponding statistical tests and the bootstrap add-on validity are all summarized and proved in the three theoretical papers of this chapter (Maciak et al. (2018), Maciak et al. (2020a), and Maciak et al. (2020a)). In particular, the following theoretical proofs are provided:

- ❏ asymptotic distribution under the null hypothesis for all four test statistics defined by (3.6), (3.7), (3.8), and (3.9);
- ❏ consistency of the statistical tests based on the proposed statistics which converge, under the alternative hypothesis in (3.3), to infinity in probability;
- ❏ consistency of the nuisance parameter free changepoint estimator defined by (3.13) given regardless of the presence/absence of the true changepoint;
- ❏ validity of the MBB bootstrap add-on as described in Algorithm 3.3.1 and the overall validity of the fully data-driven approach.

All three papers also contain an extensive empirical based finite sample comparisons under various scenarios (e.g., different within and between panel dependence forms, error distributions, various proportions of panels being subjected to the change, and others). Practical illustrations based on real data are provided as well.

To conclude, let us consider the whole *changepoint problem* and both of its aspects explicitly mentioned in Chapter 2 and Chapter 3—the *changepoint estimation* and the *changepoint detection*—together in one omnibus model where both aspects play their substantial roles within just one stochastic framework. Considering the basic principles of the changepoint problem there are two different methodological attitudes that were not mentioned yet. All the situations described so far concerned the changepoint problem where all the data were already available. However, it maybe also of some practical interest to design a so called online procedure where the data arrive in time (usually in an observation-by-observation manner) and the model estimation process and the changepoint detection algorithm run concurrently as the new observations appear. Such procedures are especially important in situations where an immediate changepoint detection in some data generating mechanism may trigger some retraining process or, even more frequently, it may help to govern important decisions effecting specific subjects of even the whole populations—such as different pandemic restrictions related to the very recent SARS-CoV-2 virus spread.

This is a particular situation where both—the *changepoint estimation* and the *changepoints detection* may go hand-in-hand in order to provide mathematically valid real-time conclusions. This also show that both may seem to be conceptually different but, at the end, the *changepoint estimation* and the *changepoints detection* are both just two sides of the same coin. Such online procedures—which formally bring together the ideas summarized in Chapter 2 and Chapter 3—are further elaborated in Ciuperca et al. (2022) but they are already beyond the scope of this thesis.

3.4 Key contributions

- ❏ All changepoint detection tests considered in all three theoretical papers in this chapter are derived under the very general and non-restrictive assumptions that go beyond typically imposed conditions in existing literature (e.g., generally dependent panels, hetereroscedasticity, non-stationarity, extreme follow-up period);
- ❏ The proposed changepoint tests based on the designed competitive test statistics (bearing in mind, for instance, different robustness assumptions or a variable power of the test) are all proved to be consistent and, moreover, with a closed form asymptotic distribution under the null hypothesis;
- ❏ The asymptotic distribution of the test statistics can be always easily obtained either in terms of some resampling method (e.g., moving block bootstrap), or some simulations (e.g., the multivariate normal process with the given variance-covariance matrix);
- ❏ All the proposed changepoint tests are closely investigated from the theoretical as well as the empirical perspective—detailed comparisons are always provided and practical pieces of advice for a real problem utilization are complemented.

4 Conclusion

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

This habilitation thesis summarizes some important mathematical and statistical contributions to the well-known *changepoint problem* while the theoretical results rely only on very non-restrictive, generally applicable, and easily verifiable theoretical assumptions. The stochastic framework of the changepoint analysis postulated in this thesis apriori assumes some inconsiderable presence of unobserved fluctuations and random error disturbances with no further specifications. The key goal is to provide (practical) conclusions which are valid from the mathematical/statistical perspective—despite any uncertainty involved.

This is all achieved by adopting two seemingly different but mutually closely related methodological frameworks: the *changepoint estimation* performed under the sparsity principle and *regularization* and the *changepoint detection* utilized via the ratio based test statistics and *self-normalization*. Both approaches are equally important and, moreover, their importance is eventually even illustrated within just one ominibus model elaborated in Ciuperca et al. (2022), where both—the *changepoint estimation* and the *changepoint detection*—play their key roles simultaneously in an online regime model providing instant conclusions and real-time decisions.

Advanced mathematical and statistical theory is applied to derive the underlying theoretical results briefly summarized in Chapter 2 and Chapter 3. Similarly, complex algorithms and sophisticated numerical techniques are needed to obtain the finite sample solutions. All theoretical details, non-trivial mathematical proofs, and empirical comparisons are given in the seven original manuscripts (all published in well recognized international impact factor journals) comprehended in this habilitation thesis and provided in full in Appendix[1].

[1] The appendix itself may or may not be the part of this habilitation thesis—depending on the thesis version (long/short) and specific requirements of the Department of Research and International Affairs of the Faculty of Mathematics and Physics, Charles University.

All seven original theoretical papers significantly contribute to both—the mathematical part of the changepoint problem as well as its finite sample—empirical counterpart. However, the papers do not represent a full, complex, and complete methodological framework, nor the author's overall research. They are only meant to serve as a brief illustration of the authors research interests and some of the author's research contributions achieved over the last few years. There is always a room for some new ideas, there is always a way for some new progress—for instance, trying to relax the theoretical assumptions being used or improving the finite sample performance, or both simultaneously. This was actually also the author's primary inspiration that drove the research summarized in this thesis.

Bibliography

Antoch, J., Gregoire, G., and Hušková, M. (2006). „Test for Continuity of Regression Function“. *Journal for Statistical Planning and Inference* 1(137), 753 –777 (cit. on p. 3).

Antoch, J., Hanousek, J., Horváth, L., Hušková, M., and Wang, S. (2019). „Structural breaks in panel data: Large number of panels and short length time series“. *Econometric Reviews* 38(7), 828 –855 (cit. on p. 31).

Bai, J. (2006). „Least squares estimation of a shift in linear processes“. *Journal of Time Series Analysis* 15(5), 453 –472 (cit. on p. 29).

Bai, J. and Perron, P. (1998). „Estimating and testing linear models with multiple structural changes“. *Econometrics* 70, 9 –38 (cit. on p. 2).

Bai, J. (2010). „Common Breaks in Means and Variances for Panel Data“. *Journal of Econometrics* 157(1), 78–92 (cit. on pp. 31, 36).

Baltagi, B. H., Feng, Q., and Kao, C. (2016). „Estimation of heterogeneous panels with structural breaks“. *Journal of Econometrics* 191(2016), 176–195 (cit. on pp. 31, 36).

Bardwell, L., Fearnhead, P., Eckley, I. A., Smith, S., and Spott, M. (2019). „Most Recent Changepoint Detection in Panel Data“. *Technometrics* 61(1), 88 –98 (cit. on p. 37).

Barigozzi, M., Cho, H., and Fryzlewicz, P. (2018). „Simultaneous multiple change-point and factor analysisfor high-dimensional time series“. *Journal of Econometrics* 206, 187 –225 (cit. on p. 31).

Betken, A and Wendler, M (2018). „Subsampling for general statistics under long range dependence“. *Statist. Sinica* 28(3), 1199–1224 (cit. on p. 37).

Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2019). „Change Point Estimation in Panel Data with Temporal and Cross-sectional Dependence“. *arXiv: Statistics Theory* (cit. on p. 31).

Bosetti, C., Bertuccio, P., Levi, F., et al. (2008). „Cancer Mortality in the European Union, 1970 – 2003, with a Joinpoint Analysis“. *Annals of Oncology* 16(1), 631–640 (cit. on p. 2).

Boysen, L., Kempe, A., Munk, A., Liebscher, V., and Wittich, O. (2009). „Consistencies and rates of conference of jump penalized least squares estimators“. *Annals of Statistics* 137(1), 157–183 (cit. on p. 15).

Breiman, L. (1995). „Better Subset Regression Using the Nonnegative Garrote“. *Technometrics* 37(4), 373–384 (cit. on pp. 8, 9).

Brown, R., Durbin, J., and Evans, J. (1975). „Techniques for testing constancy of regression relationships over time (with discussion)“. *Journal of the Royal Statistical Society. Series B* 37, 149–192 (cit. on p. 2).

Campbell, F. and Allen, G. (2017). „Within group variable selection through the Exclusive Lasso“. *Electronic Journal of Statistics* 11(2), 4220 –4257 (cit. on p. 20).

Carlin, B., Gelfand, A., and Smith, A. (1992). „Hierarchical Bayesian Analysis of Change-point Problems“. *Applied Statistics* 41, 389–405 (cit. on p. 3).

Chan, H. P. and Walther, G. (2013). „Detection with the scan and the average likelihood ratio“. *Statistica Sinica* 23(1), 409 –428 (cit. on p. 3).

Chan, J., Horváth, L., and Hušková, M. (2013a). „Change-point Detection in Panel Data“. *Journal of Statistical Planning and Inference* 143(5), 955–970 (cit. on p. 31).

Chan, J., Horváth, L., and Hušková, M. (2013b). „Darling–Erdős limit results for change-point detection in panel data“. *Journal of Statistical Planning and Inference* 143(5), 955–970 (cit. on p. 31).

Chen, S., Donoho, D., and Saunders, M. (1998). „Atomic Decomposition by Basis Pursuit“. *SIAM Journal of the Scientic Computing* 20, 33 –61 (cit. on p. 8).

Chen, Z. and Tian, Z. (2014). „Ratio tests for variance change in nonparametric regression“. *Statistics* 48(1), 1–16 (cit. on p. 33).

Cho, H. (2016). „Change-point detection in panel data via double cusum statistic“. *Electronic Journal of Statistics* 10(2), 2000 –2038 (cit. on p. 31).

Chow, G. (1960). „Tests of Equality Between Sets of Coefficients in Two Linear Regressions“. *Econometrica* 28(3), 591 –605 (cit. on p. 29).

Chu, C.-S., Maxwell, S., and White, H. (1996). „Monitoring Structural Change“. *Econometrica* 64(5), 1045 –1065 (cit. on p. 3).

Ciuperca, G. (2014). „Model selection by LASSO methods in a changepoint model“. *Statistical Papers* 55(1), 349–374 (cit. on p. 12).

Ciuperca, G. and Maciak, M. (2019a). „Change-point Detection by the Quantile LASSO Method“. *Journal of Statistical Theory and Practice* 14, 1 –18 (cit. on pp. 9, 10, 13, 15, 19).

Ciuperca, G. and Maciak, M. (2019b). „Change-point detection in a linear model by adaptive fused quantile method“. *Scandinavian Journal of Statistics* 47, 425 –463 (cit. on pp. 11, 16, 18).

Ciuperca, G., Maciak, M., and Wahl, F. (2020). „Detection of similar successive groups in a model with diverging number of variable groups“. *Sequential Analysis* 39(1), 92 –114 (cit. on pp. 11, 20).

Ciuperca, G., Maciak, M., and Pešta, M. (2022). „Real-time changepoint detection in a nonlinear expectile model“. *Journal of Statistical Planning and Inference*. (submitted) (cit. on pp. 39, 41).

Csörgő, M. and Horváth, L. (1988). „20 Nonparametric methods for changepoint problems“. *Handbook of Statistics* 7, 403 –425 (cit. on p. 2).

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley Series in Probability & Statistics: Chichester, England (cit. on p. 33).

De Wachter, S. and Tzavalis, E. (2012). „Detection of structural breaks in linear dynamic panel data models“. *Computational Statistics and Data Analysis* 56, 3020 –3034 (cit. on p. 31).

Dehling, H., Fried, R., and Wendler, M. (2015). „A robust method for shift detection in time series“. *arXiv:1506.03345* (cit. on p. 37).

Dodge, Y. (1987). *Statistical data analysis based on the* L_1*-norm and related methods*. North-Holland, Amsterdam (cit. on p. 9).

Donoho, D. L. and Logan, B. F. (1992). „Signal Recovery and the Large Sieve“. *SIAM Journal on Applied Mathematics* 52(2), 577 –591 (cit. on p. 8).

Donoho, D. L. and Stark, P. (1989). „Uncertainty principles and signal recovery“. *Journal of Applied Mathematics* 49, 906 –931 (cit. on p. 8).

Drábek, Z., Kopa, M., Maciak, M., Pešta, M., and Vitali, S. (2022). „Do investment disputes affect implied volatility ofoptions?“ *Journal of Applied Economics* (submitted) (cit. on p. 23).

Carlstein, E., Müeller, H., and Siegmund, D., eds. (1994). *Nonparametric estimation of functions with jump discontinuities*. IMS, pp. 130 –144 (cit. on p. 2).

Feder, P. (1975). „The Log Likelihood Ratio in Segmented Regression“. *The Annals of Statistics* 3, 84–97 (cit. on p. 2).

Frick, K., Munk, A., and Sieling, H. (2014). „Multiscale change point inference“. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 76(3), 495 –580 (cit. on pp. 3, 15).

Friedman, J. H., Hastie, T. J., and Tibshirani, R. (2010). „A note on the group lasso and a sparse group lasso“. *arXiv: Statistics Theory* (cit. on p. 11).

Friedrich, F., Kempe, A., Liebscher, V., and Winkler, G. (2008). „Complexity Penalized M-Estimation“. *Journal of Computational and Graphical Statistics* 17(1) (cit. on p. 3).

Fryzlewicz, P. (2014). „Wild binary segmentation for multiple change-point detection“. *Annals of Statistics* 742(6), 2243 –2281 (cit. on p. 15).

Guo, X., Zhang, H., Wang, Y., and Wu, J.-L. (2015). „Model selection and estimation in high dimensional regression models with group SCAD“. *Statistics and Probability Letters* 103, 86–92 (cit. on p. 20).

Harchaoui, Z. and Lévy-Leduc, C. (2010). „Multiple changepoint estimation with a total variation penalty“. *Journal of the American Statistical Association* 105(492), 1480 –1493 (cit. on pp. 12, 15).

Hinkley, D. (1969). „Inference about the intersection in two-phase regression“. *Biometrika* 56, 495 –504 (cit. on p. 2).

Hinkley, D. (1971). „Inference in Two-Phase Regression“. *Journal of American Statistical Association* 66(336), 736–743 (cit. on p. 2).

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). „Monitoring changes in linear models“. *Journal of Statistical Planning and Inference* 126(1), 225–251 (cit. on p. 3).

Horváth, L. and Hušková, M. (2012). „Change-point detection in panel data“. *J. Time Ser. Anal.* 33(4), 631–648 (cit. on p. 31).

Horváth, L. and Kokoszka, P (2002). „Change-point detection in non-parametric regression“. *Statistics* 36, 9 –31 (cit. on p. 3).

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 p. 33).

Hyun, S., G'Sell, M., and Tibshirani, R. (2016). „Exact Post-Selection Inference for the Generalized Lasso Path“. *Electronic Journal of Statistics* 12, 1053–1097 (cit. on p. 15).

Jeng, X., Cai, T., and Li, H. (2010). „Optimal Sparse Segment Identification With Application in Copy Number Variation Analysis“. *Journal of the American Statistical Association* 105(491), 1156–1166 (cit. on p. 3).

Jiang, L., Wanga, H., and Bondell, H. (2013). „Interquantile Shrinkage in Regression Models“. *Journal of Computational and Graphical Statistics* 22(4), 970 –986 (cit. on p. 20).

Jiang, L., Bondell, H., and Wanga, H. (2014). „Interquantile Shrinkage and Variable Selection in Quantile Regression“. *Computational statistics and Data Analysis* 69, 208 –219 (cit. on p. 20).

Jordan, C. (1881). „Sur la série de Fourier“. *Comptes rendus hebdomadaires des séances de l'Académie des sciences (in French)* 92, 228 –230 (cit. on p. 9).

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. 31).

Kim, D. (2011). „Estimating a common deterministic time trend break in large panels with cross sectional dependence“. *Journal of Econometrics* 164(2), 310–330 (cit. on p. 31).

Kim, H., Fay, M., Feuer, E., and Midthune, D. (2000). „Permutation Tests for Joinpoint Regression with Application to Cancer Rates“. *Statistics in Medicine* 19, 335–351 (cit. on p. 2).

Kim, H., Koh, K., Boyd, S., and Gorinevsky, D. (2009a). „l_1 trend filtering“. *SIAM Reviews* 51, 339 –360 (cit. on p. 9).

Kim, H., Yu, B., and Feuer, E. (2009b). „Selecting the Number of Change-poins in Segmented Line Regression“. *Statistica Sinica* 19, 597 –609 (cit. on p. 2).

Koenker, R. (1998). *Quantile Regression*. Cambridge University Press: Cambridge, United Kingdom (cit. on p. 15).

Kuhn, H. and Tucker, A. (1951). „Nonlinear programming“. *Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press*, 481 –492 (cit. on p. 15).

Künsch, H. (1989). „The Jackknife and the Bootstrap for General Stationary Observations“. *Annals of Statistics* 17(3), 1217 –1241 (cit. on p. 37).

Künsch, H. (1994). „Robust priors for smoothing and image restoration“. *Annals of the Institute of Statistical Mathematics* 46, 1 –19 (cit. on p. 9).

Lavielle, M. and Moulines, E. (2000). „Least-squares estimation of an unknown number of shifts in a time series“. *Journal Time Series Analysis* 21, 33 –59 (cit. on p. 15).

Lee, S., Seo, M., and Shin, Y. (2016). „The lasso for high dimensional regression with a possible change point“. *Journal of the Royal Statistical Society: Series B* 78(1), 193 –210 (cit. on pp. 10, 17).

Lee, S., Liao, Y., Seo, M., and Shin, Y. (2018). „Oracle estimation of a change point in high-dimensional quantile regression“. *J. Amer. Statist. Assoc.* 113(523), 1184 –1194 (cit. on p. 17).

Leonardi, F. and Bühlmann, P. (2016). „Computationally efficient change point detection for high-dimensional regression“. *arXiv:1601.03704* (cit. on p. 12).

Li, H., Guo, Q., and Munk, A. (2019). „Multiscale change-point segmentation: beyond step functions“. *Electronical Journal of Statistics* 13(2) (cit. on p. 15).

Lim, M. and Hastie, T. (2015). „Learning Interactions via Hierarchical Group-Lasso Regularization“. *Journal of Computational and Graphical Statistics* 24(3), 627–654 (cit. on p. 11).

Lin, K., Sharpnack, J., Rinaldo, A., and Tibshirani, R. (2016). „Approximate Recovery in Changepoint Problems, from l_2 Estimation Error Rates“. *arxiv.org/abs/1606.06746* (cit. on p. 17).

Liu, J., Wu, S., and Zidek, J. (1997). „On segmented multivariate regression“. *Statistica Sinica* 7, 497 –525 (cit. on p. 3).

Liu, Y., Tao, J., Zhang, H., Xiu, X., and Kong, L. (2018). „Fused LASSO penalized least absolute deviation estimator for high dimensional linear regression“. *Numerical Algebra, Control and Optimization* 8(1), 97–117 (cit. on p. 20).

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. 33).

Loader, C. (1996). „Change Point Estimation Using Nonparametric Regression“. *Annals of Statistics*(24), 1667 –1678 (cit. on p. 2).

Logan, B. (1965). *Properties of high-pass signals*. Ph.D. thesis. Columbia University: New York (cit. on p. 8).

Maciak, M. (2010). „Bootstrapping of M-smoothers“. *Acta Universitatis Carolinae - Mathematica et Physica* 52(1), 37 –46 (cit. on p. 4).

Maciak, M. (2018a). „Bootstrapping Nonparametric M-Smoothers with Independent Error Terms“. In: *Nonparametric Statistics*. Ed. by P. Bertail, D. Blanke, P. Cornillon, and E. Matzner-Løber. 3rd ISNPS. Springer, pp. 233 –250 (cit. on p. 4).

Maciak, M. (2018b). „Linear Trend Filtering via Adaptive LASSO“. In: *Time Series Analysis and Forecasting, ITISE 2017*. Ed. by I. Rojas, H. Pomares, and O. Valenzuela. Proceedings ITISE 2017, Granada, 18-20 September 2017, pp. 131 –145 (cit. on p. 4).

Maciak, M. (2018c). „Testing Shape Constrains in Lasso Regularized Joinpoint Regression“. In: *Analytical Methods in Statistics*. Ed. by J. Antoch, J. Jurečková, M. Maciak, and M. Pešta. AMISTAT, Prague, November 2015. Springer, pp. 105 –122 (cit. on p. 4).

Maciak, M. (2021a). „Quantile LASSO in arbitrage-free option markets“. *Econometrics and Statistics* 18, 106 –116 (cit. on pp. 23, 25).

Maciak, M. (2021b). „Quantile LASSO with changepoints in panel data models applied to option pricing“. *Econometrics and Statistics* 20, 166 –175 (cit. on pp. 11, 23–26).

Maciak, M. and Hušková, H. (2017). „Discontinuities in Robust Nonparametric Regression with α-mixing Dependence“. *Journal of Nonparametric Statistics* 29(2), 447 –475 (cit. on p. 4).

Maciak, M. and Mizera, I. (2016). „Regularization Techniques in Joinpoint Regression“. *Statistical Papers* 57(4), 939 –955 (cit. on p. 4).

Maciak, M. and Mizera, I. (2022). „Splines with changepoints: Additive models for functional data“. (forthcoming) (cit. on p. 4).

Maciak, M. and Vitali, S. (2022). „Detection of changepoints due to exogenous effects analyzed via artificial options volatility“. *Decisions in Economics and Finance* (submitted) (cit. on p. 23).

Maciak, M., Pešta, M., and Peštová, B. (2020a). „Changepoint in dependent and non-stationary panels“. *Statistical Papers* 61, 1385 –1407 (cit. on pp. 32–35, 38).

Maciak, M., Pešta, M., and Vitali, S. (2020b). „Implied volatility surface estimation via quantile regularization“. In: *Analytical Methods in Statistics*. Ed. by M. Maciak, M. Pešta, and M. Schindler. Springer Proceedings in Mathematics & Statistics 329. Springer Nature Switzerland AG, pp. 131 –145 (cit. on p. 23).

Maciak, M., Peštová, B., and Pešta, M. (2018). „Structural breaks in dependent, heteroscedastic, and extremal panel data“. *Kybernetika* 54(6), 1106–1121 (cit. on pp. 31–34, 38).

Mammen, E. and Geer, S. V. D. (1997). „Locally Adaptive Regression Splines“. *Annals of Statistics* 25(1), 387 –413 (cit. on pp. 9, 15).

Martinez-Beneito, M., García-Donato, G., and Salmerón, D. (2011). „A Bayesian Joinpoint Regression Model with an Unknown Number of Break-Points“. *The Annals of Applied Statistics* 5(3), 2150–2168 (cit. on p. 3).

Müller, H. (1992). „Change Points in Nonparametric Regression Analysis“. *Annals of Statistics*(20), 737 –715 (cit. on p. 2).

Pesaran, M. (2006). „Estimation and inference in large heterogeneous panelswith a miltifactor error structure“. *Econometrica* 74, 967 –1012 (cit. on pp. 31, 36).

Pešta, M., Peštová, B., and Maciak, M. (2020). „Changepoint estimation for dependent and non-stationary panels“. *Applications of Mathematics* 65(3), 299 –310 (cit. on p. 36).

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. 33).

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 p. 31).

Peštová, B. and Pešta, M. (2017). „Change point estimation in panel data without boundary issue“. *Risks* 5(1), 7 (cit. on p. 37).

Qian, J. and Su, L. (2014). „Structural change estimation in time series regression with endogenous variables“. *Economic Letters* 125, 415 –421 (cit. on p. 23).

Qian, J. and Su, L. (2016a). „Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso“. *Journal of Econometrics* 191(1), 86 –109 (cit. on p. 23).

Qian, J. and Su, L. (2016b). „Shrinkage estimation of regression models with multiple structural changes“. *Econometric Theory* 32(6), 1376 –1433 (cit. on pp. 12, 23).

Qiu, P. and Yandell, B. (1998). „A Local Polynomial Jump Detection Algorithm in Nonparametric Regression“. *Technometrics* 40, 141 –152 (cit. on p. 2).

Quandt, R. (1960). „Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes“. *Journal of the American Statistical Association* 55(290), 324 –330 (cit. on p. 29).

Qui, D., Katanoda, K., Tomomi, M., and Tomotaka, S. (2009). „A Joinpoint Regression Analysis of Long-term Trends in Cancer Mortality in Japan (1958 – 2004)“. *International Journal of Cancer* 24, 443–448 (cit. on p. 2).

Rinaldo, A. (2009). „Properties and refinements of the fused Lasso“. *Annals of Statistics* 37(5B), 2922–2952 (cit. on p. 12).

Rudin, L., Osher, S., and Fatemi, E. (1992). „Nonlinear total variation based noise removal algorithms“. *Physica D: Nonlinear Phenomena* 60(1), 259 –268 (cit. on p. 9).

Sadhanala, V. and Tibshirani, R. (2019). „Additive models with trend filtering“. *Annals of Statistics* 47(6), 3032 –3068 (cit. on p. 9).

Simon, N. and Tibshirani, R. (July 2012). „Standardization and the Group Lasso Penalty“. *Statistica Sinica* 22 (cit. on p. 11).

Tibshirani, R. (1996). „Regression shrinkage and selection via the LASSO“. *Journal of the Royal Statistical Society: Series B* 58, 267 –288 (cit. on pp. 8, 9).

Tibshirani, R. (2011). „Regression shrinkage and selection via the lasso: a retrospective“. *Journal of the Royal Statistical Society: Series B* 73, 273 –282 (cit. on p. 10).

Tibshirani, R. and Saunders, M. (2005). „Sparsity and smoothness via the fused lass“. *Journal of the Royal Statistical Society: Series B* 67(1), 91 –108 (cit. on p. 11).

Tibshirani, R. and Wang, P. (2008). „Spatial smoothing and hot spot detection for CGH data using the fused lasso“. *Biostatistics* 9(1), 18 –29 (cit. on p. 12).

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). „Sparsity and smoothness via the fused lasso“. *Journal of the Royal Statistical Society: Series B* 67, 91 –108 (cit. on p. 9).

Tibshirani, R. (2014). „Adaptive Piecewise Polynomial Estimation via Trend Filtering“. *Annals of Statistics* 42(1), 285 –323 (cit. on pp. 9, 12).

Tibshirani, R. J. and Taylor, J. (2011). „The solution path of the generalized lasso“. *The Annals of Statistics* 39(3), 1335 –1371 (cit. on p. 10).

Tibshirani, R. J. and Taylor, J. (2012). „Degrees of freedom in lasso problems“. *The Annals of Statistics* 40(2), 1198 –1232 (cit. on p. 10).

Tiwari, R., Cronin, K., Davies, W., et al. (2005). „Bayesian model selection for joinpoint regression with application to age-adjusted cancer rates“. *Journal of the Royal Statistical Society: Series C* 54, 919 –939 (cit. on p. 3).

Wei, F. and Huang, J. (2010). „Consistent group selection in high-dimensional linear regression“. *Bernoulli* 16(4), 1369 –1384 (cit. on p. 20).

Westerlund, J. (2019). „Common breaks in means for cross-correlated fixed-t panel data“. *Journal ofTime Series Analysis* 40, 248 –255 (cit. on p. 31).

Wu, J. and Chu, C. (1993). „Kernel-type estimators of jump points and values of a regression function“. *Annals of Statistics* 21, 1545 –1566 (cit. on p. 2).

Yao, Y.-C. (1988). „Estimating the number of change-points via Schwarz' criterion". *Statistics and Probability Letters* 6, 181 –189 (cit. on p. 2).

Yao, Y.-C. and Au, S. (1989). „Least-Squares Estimation of a Step Function". *The Indian Journal of Statistics* 51(3), 370 –381 (cit. on p. 15).

Yu, Y. and Chatterjee, S. (2020). „Localising change points in piecewise polynomials of general degrees". *arXiv: Statistics Theory* (cit. on p. 15).

Yuan, M. and Lin, Y. (Feb. 2006). „Model Selection and Estimation in Regression With Grouped Variables". *Journal of the Royal Statistical Society Series B* 68, 49–67 (cit. on p. 11).

Zhang, C. and Xiang, Y. (2015). „On the oracle property of adaptive group LASSO in high-dimensional linear models". *Statistical Papers* 57(1), 249–265 (cit. on p. 20).

Zhang, N. and Siegmund, D. (2007). „A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data". *Biometrics* 63(1), 22 –32 (cit. on p. 3).

Zhang, N. and Siegmund, D. (2012). „Model selection for high dimensional multi-sequence change-point problems". *Statistica Sinica* 22, 1507–1538 (cit. on p. 17).

Zheng, Q., Gallagher, C., and Kulasekera, K. (2013). „Adaptive penalized quantile regression for high dimensional data". *J. Statist. Plann. Inference* 143, 1029–1038 (cit. on p. 17).

Zheng, Q., Peng, L., and He, X. (2015). „Globally adaptive quantile regression with ultra-high dimensional data". *Ann. Statist.* 43(5), 2225–2258 (cit. on p. 17).

Zou, H. and Hastie, T. (2005). „Regularization and variable selection via the elastic net". *Journal of the Royal Statistical Society: Series B* 67(2), 301 –320 (cit. on p. 10).

Appendix: List of Scientific Papers

A

[1] Ciuperca, G. and Maciak, M. (2020). 'Changepoint Detection by the Quantile LASSO Method'. *Journal of Statistical Theory and Practice* 14(11), 1–38.
DOI: 10.1007/s42519-019-0078-z

[2] Ciuperca, G. and Maciak, M. (2019). 'Change-point detection in a linear model by adaptive fused quantile method'. *Scandinavian Journal of Statistics* 47(1), 425–463.
DOI: 10.1111/sjos.12412

[3] Ciuperca, G., Maciak, M., and Wahl, F. (2020). 'Detection of similar successive groups in a model with diverging number of variable groups'. *Sequential Analysis* 39(1), 92–114.
DOI: 10.1080/07474946.2020.1726687

[4] Maciak, M. (2019). 'Quantile LASSO with changepoints in panel data models applied to option pricing'. *Econometrics and Statistics* 20(2021), 166–175.
DOI: 10.1016/j.ecosta.2019.12.005

[5] Maciak, M., Peštová, B., and Pešta, M. (2018). 'Structural breaks in dependent, heteroscedastic, and extremal panel data'. *Kybernetika* 54(2018), 1106–1121.
DOI: 10.14736/kyb-2018-6-1106

[6] Maciak, M., Pešta, M., and Peštová, B. (2020). 'Changepoint in dependent and non-stationary panels'. *Statistical Papers* 61(2020), 1385–1407.
DOI: 10.1007/s00362-020-01180-6

[7] Pešta, M., Peštová, B., and Maciak, M. (2020). 'Changepoint estimation for dependent and non-stationary panels'. *Applications of Mathematics* 65(3), 299–310.
DOI: 10.21136/AM.2020.0296-19