



TECHNISCHE
UNIVERSITÄT
WIEN

DISSERTATION

Contributions to Statistical Modeling for Complex Longitudinal and Multivariate Time Series Data

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften unter der Leitung von

Prof. Dr. Peter Filzmoser

E105 – Institut für Statistik und Wirtschaftsmathematik, TU Wien

eingereicht an der Technischen Universität Wien

Fakultät für Mathematik und Geoinformation

von

Barbara Saskia Brune

Matrikelnummer: 12004643

Diese Dissertation haben begutachtet:

1. **Prof. Dr. Peter Filzmoser**
Institut für Stochastik und Wirtschaftsmathematik, TU Wien, Vienna, Austria
2. **Prof. RNDr. Karel Hron, Ph.D.**
Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science,
Palacký University, Olomouc, Czech Republic
3. **Assoc.-Prof. Dr. Gianna Serafina Monti**
Department of Economics, Management and Statistics, University of Milano-Bicocca, Milan,
Italy

Wien, im Oktober 2024

Kurzfassung

Datensätze mit wiederholten Messungen treten in vielen unterschiedlichen Bereichen auf, beispielsweise in der medizinischen Forschung, in wirtschaftlichen Anwendungen oder in den Materialwissenschaften. Während die statistische Analyse solcher Datensätze aufgrund ihrer komplexen Abhängigkeitsstrukturen herausfordernd sein kann, bieten sie auch Möglichkeiten für ein genaueres Verständnis der Beziehungen und zeitlichen Entwicklung sowohl zwischen als auch innerhalb von Beobachtungseinheiten. In der statistischen Literatur gibt es fundierte Methoden für die Analyse von Daten mit wiederholten Messungen.

Daten aus der realen Welt enthalten jedoch häufig Ausreißer oder ungewöhnliche Beobachtungen. Werden solche Abweichungen nicht berücksichtigt, können sie die Genauigkeit der statistischen Analysen erheblich beeinträchtigen und die Ergebnisse verfälschen oder sogar ungültig machen. Die Identifikation der abweichenden Beobachtungen ist wichtig und kann zu einem tieferen Verständnis der Daten und zugrunde liegenden Strukturen beitragen. Insbesondere bei langen Beobachtungszeiträumen können außerdem Veränderungen im Umfeld oder strukturelle Schocks die Beziehungen innerhalb der Beobachtungszeitreihen oder die Modellparameter verändern. Herkömmliche Modelle mit zeitkonstanten Parametern sind in diesen Fällen nicht geeignet, und es wird notwendig, die Variation der Parameter innerhalb des Modells zu berücksichtigen. Die Analyse der sich daraus ergebenden Parameterdynamik kann wertvolle Einblicke in die strukturellen Veränderungen liefern, die im Laufe der Zeit auftreten.

Diese Arbeit adressiert die oben genannten Problemen, indem sie Modelle entwickelt, die auf verschiedene Arten von Daten mit wiederholten Messungen zugeschnitten sind: multivariate Zeitreihen und (funktionale) Longitudinaldaten. Wir schlagen ein multivariates Regressionsmodell mit zeitvariablen Parametermatrizen von reduziertem Rang, eine robuste Schätzmethode für lineare Mixed-Effects-Modelle mit Ausreißern in der Zielvariable oder den erklärenden Variablen, sowie einen robusten Schätzalgorithmus für die marginale Hauptkomponentenanalyse von longitudinalen Funktionaldaten vor. Die Wirksamkeit der entwickelten Modelle und Schätzmethoden wird in Simulationsstudien demonstriert. Darüber hinaus werden die drei Methoden in realen Datenanwendungen aus verschiedenen Bereichen illustriert.

Abstract

Repeated measurements data are commonly encountered in numerous fields, including medical research, economics, material science, and engineering. While the statistical analysis of such datasets poses challenges due to their complex dependency structures, they also offer opportunities for more detailed understanding of relationships and temporal development both between and within observational units. There are well-founded statistical methods for the analysis of repeated measurements data in the statistical literature.

However, real-world data often contain outlying or unusual observations that do not align with the imposed model structure. If not properly accounted for, such deviations can significantly impact the accuracy of statistical analyses, bias or invalidate the results. Identifying the outlying observations is important and can contribute to a deeper understanding of the data and underlying structures. Furthermore, particularly in long observation periods, changes in the environment or structural shocks can alter model parameters and relationships. Traditional models with fixed parameters may therefore no longer be suitable, and it becomes necessary to account for parameter variation within the model. Analyzing the resulting parameter dynamics can yield valuable insights into the structural changes that occur over time.

This thesis addresses the issues mentioned above by developing models tailored to different types of repeated measurements data: multivariate time series and longitudinal (functional) data. We propose a time-varying parameter reduced-rank regression model for multivariate time series regression, a robust estimation method for mixed effects models in the presence of outlying responses and predictors, and a robust estimation algorithm for marginal principal components analysis of longitudinally observed functional datasets. The effectiveness of the developed models and estimation methods is demonstrated through simulation studies. Furthermore, the three methods are illustrated in real-world data applications from various fields.

Acknowledgement

Writing a PhD thesis next to working more or less full-time has turned out to be a bigger challenge than expected. Without the support of all the amazing people around me this would never have been possible.

First and foremost, I would like to thank my supervisor, Peter Filzmoser, for his support and encouragement at any time. You always asked the right questions, kept me on track, and without you I would have for sure given up on this project at some point. Special thanks go to the co-authors of my papers, Wolfgang Scherrer, Efstathia Bura, Irene Ortner and Una Radojičić, for their valuable inputs, supervision and support; and to my reviewers Karel Hron and Gianna Monti for kindly agreeing to review this thesis. This work received funding from the Austrian Science Fund FWF (P 30690-N35) and the Austrian Research Promotion Agency FFG (No. 881133).

I would like to thank all my friends for always being here for me, although many of you live a thousand (or more) kilometers away. I highly appreciate your support through long phone calls, weekend co-working sessions and pre-work walks. Thank you to my flatmates Azurra and Alex, for kitchen talks and cheering me up after long days. I am really happy to have you in my life. Big thanks go to my parents Eva and Rolf, my brother Johannes, and my uncle Jürgen, for always believing in me. Cheers to Brendan and Emma for providing me with shelter in the final stages of finishing this thesis – I really hope your house is not haunted now!

Another thank you to the whole CSTAT team and my fellow (no-longer) PhD students – Pia, Patricia, Lukas, Marcus, Roman and Jeremy – for always making me feel welcome at the office, for help and advice, and sharing your pasta with me. I am very grateful to my company, TÜV AUSTRIA Data Intelligence, and the whole team of immortal space explorers, for making finishing my thesis possible, and sometimes respecting my time off more than I did. Finally, thank you to Thomas and Irene Ortner, for always motivating me to go on.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 21. Oktober 2024

Barbara Brune

Contents

1. Introduction	1
1.1. Preliminaries	1
1.2. Modeling multivariate time series data	6
1.3. Mixed effects models for longitudinal data	14
1.4. Functional data analysis and models for longitudinal functional data	19
1.5. Overview of the three main chapters	26
1.6. Related software packages	28
2. A State-Space Approach to Time-Varying Reduced-Rank Regression	29
2.1. Introduction	29
2.2. Time-varying reduced-rank regression	31
2.3. Simulation studies	41
2.4. Data examples	49
2.5. Conclusions	55
3. Rank-Based Estimation of Mixed Effects Models	57
3.1. Introduction	57
3.2. Preliminaries	58
3.3. Methodology	61
3.4. Simulation	67
3.5. Applications	72
3.6. Summary and outlook	78
4. Robust modeling of repeated functional measurements	79
4.1. Introduction	79
4.2. Methodology	81
4.3. Robust estimation procedure	83
4.4. Methods for outlier analysis	89
4.5. Simulation	91
4.6. Real data applications	95
4.7. Discussion and conclusions	101
5. Conclusions and Discussion	103
Bibliography	107

A. Supplement: Chapter 2	117
A.1. Rank selection	117
A.2. Simulation results for fixed coefficients	120
A.3. Simulation results for deterministic coefficient transition	122
A.4. Simulation results for random walk coefficient evolution	127
B. Supplement: Chapter 3	133
B.1. Bias of the weighted rank-based estimator	134
B.2. Validity of the updating step	135
B.3. Response outliers	136
B.4. Outlying predictors	140
B.5. Leverage weights in the presence of dummy variables	144
C. Supplement: Chapter 4	145
C.1. Details on estimation procedure	145
C.2. Additional material for the simulation studies	150
C.3. Additional material for the data examples	157

1 | Introduction

1.1. Preliminaries

A common assumption in statistical analysis and modeling is that the data are independently and identically distributed (i.i.d.). However, in practice there are many data structures that do not fulfill this assumption. This thesis is concerned with datasets that have special dependency structures due to repeated measurements on different observational units. Such observational units can, among many other possibilities, be the subjects in a clinical study, machine parts subject to regular inspection, but also other entities such as countries or stock indices that are observed over time. Repeated measurements datasets arise in many different fields, including medical and biological science, material science and economics. The analysis of repeated measurements is complicated due to the complex dependency structures within and between the measurements from the different observational units. Especially in regression tasks, not accounting for such dependencies can result in correlated error structures and bias the results of the analysis (Fahrmeir et al. 2013, Ch. 7). However, sampling on multiple observational units also increases the total number of data points. This can lead to more accurate inference, and findings beyond those that can be obtained from taking measurements only on a single observational unit. Repeated measurements allow to simultaneously investigate global and individual changes along time (Singer and Andrade 2000). If the observed data series are correlated with each other, information can be reused across series. This can lead to overall improved models and forecasts, also for observational units with an initially limited number of available samples or missing values.

Assuming we have observed data from g observational units, the datasets of interest consist of measurements z_{ij} which are organized as follows,

$$\begin{pmatrix} z_{11}, & z_{12}, & \cdots, & z_{1n_1} \\ z_{21}, & z_{22}, & \cdots, & z_{2n_2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{g1}, & z_{g2}, & \cdots, & z_{gn_g} \end{pmatrix}, \tag{1.1}$$

where z_{ij} describes the j 'th measurement for the i 'th observational unit. For unit i , we observe n_i samples, where the n_i are not necessarily equal for all units. Each observation is associated with a measurement time point t_{ij} with $t_{ij} < t_{i(j+1)}$, thus the samples are assumed to have an inherent ordering in time. The samples z_{ij} can be of different shapes: They can represent simple scalar values, pairs of response and predictor variables in regression settings such that $z_{ij} = (y_{ij}, \mathbf{x}'_{ij})' \in \mathbb{R}^{p+1}$, or functional observations $z_{ij} = z_{ij}(s)$, $s \in \mathcal{S}$,

$$\begin{array}{ccc}
 \left(\begin{array}{cccc} z_{11}, & z_{12}, & \cdots, & z_{1n_1} \\ z_{21}, & z_{22}, & \cdots, & z_{2n_2} \\ \vdots & \vdots & & \vdots \\ z_{g1}, & z_{g2}, & \cdots, & z_{gn_g} \end{array} \right) & & \left(\begin{array}{cccc} z_{11}, & z_{12}, & \cdots, & z_{1n_1} \\ z_{21}, & z_{22}, & \cdots, & z_{2n_2} \\ \vdots & \vdots & & \vdots \\ z_{g1}, & z_{g2}, & \cdots, & z_{gn_g} \end{array} \right), \\
 \text{(a) Multivariate time series measurement} & & \text{(b) Longitudinal observation}
 \end{array}$$

Figure 1.1.: “Vertical” (left) and “horizontal” (right) measurements in the data structure (1.1), yielding multivariate time series or longitudinal measurements.

where $z_{ij} : \mathcal{S} \rightarrow \mathbb{R}$ is a real-valued function for each pair i, j .

If $n_i = T$ for all $i = 1, \dots, g$ and the samples are observed on a regular grid in time, they form a *multivariate time series* dataset. Commonly, the measurements in (1.1) are then represented “vertically” by a g -dimensional vector for each sampling time point t , denoted by

$$\mathbf{y}_t = (y_{1t}, \dots, y_{gt})' \in \mathbb{R}^g \text{ for } t = 1, \dots, T,$$

(also see left panel in Figure 1.1). In multivariate time series analysis, it is of interest to study the dynamic (cross-)relationships between the observed series y_{it} and y_{js} , and using this knowledge to improve the accuracy of prediction for future values of the series \mathbf{y}_{t+h} , $h \geq 1$ (Tsay 2014).

Another interpretation of (1.1) is that of *longitudinal data*, sometimes also referred to as *panel data* (Hsiao 2014). Here, the sampling points y_{ij} are not necessarily sampled on a regular grid, missing values are common. We usually observe multiple, rather short series of measurements on different observational units (Singer and Andrade 2000). In the setting of (1.1), the samples can be interpreted “horizontally”, i.e. we denote the i 'th vector of observations by

$$\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})' \in \mathbb{R}^{n_i} \text{ for } i = 1, \dots, g,$$

(also see right panel in Figure 1.1). In longitudinal data analysis, we aim to understand global patterns of change as well as individual-specific effects over time.

There are also other types of datasets that can be analyzed in the context of repeated measurements, beyond ordering in time. One common application is grouped data, where multiple samples are taken from different clusters. An example is sampling of multiple students from different schools. In this setting, a cluster structure is introduced since the data from the students from one school might be correlated. Another possibility is longitudinal data that is observed along another axis than time, e.g. in a spatial setting. However, we limit the following exposition to dependency in time, while keeping in mind that some of the proposed methodology is also applicable to other settings.

This thesis is concerned with the development of new methods to model and analyze complex longitudinal and multivariate time series data. We focus on methodology that can deal

with deviations from the assumptions of the “traditional” parametric models in classical statistics – on the one hand by the potential occurrence of structural breaks and changes in the relationships within the data of interest, on the other hand by the presence of outlying values in the datasets used for model fitting.

1.1.1. Structural breaks and time-varying parameter models

In classical multivariate time series analysis, the relationships within and between the time series are assumed to be constant. Usually, it is assumed that the mean and the variance and (auto)covariance structure of the time series do not change with time. This assumption is called *stationarity*. However, in practice we are often confronted with time series that cover a rather long time span or can be exposed to sudden shocks. In that case, there is a likelihood that the dynamics within and between the time series change over time. Examples for this can be climate data, where climate change slowly alters, e.g., the expected temperatures and amount of rainfall as well as their relationship; or economic data, where changes in government policies can cause changes in key variables such as unemployment and inflation rates. Furthermore, sudden shocks can cause *structural breaks* in the time series, thinking, for example, of the Covid crisis in 2020. Such structural changes can be addressed in the modeling process by applying so-called *time-varying parameter models*. Such models allow to incorporate gradual and sudden structural changes by not assuming a static relationship between the variables, but by allowing the model parameters to vary. This can lead to increased forecasting performance and a better understanding of the dynamics. There is a vast amount of literature on time-varying parameter models, both in theory and application. Early reviews of the relevant literature can already be found in Rosenberg (1973) and Beck (1983). The growing empirical evidence of parameter instability in economic applications led to wide recognition of the importance of time-varying coefficient models in the econometrics literature (Fu et al. 2023).

Time-variation can be modeled in many different ways. Commonly, the time-varying parameters are assumed to follow stochastic processes, usually random walks. In this case, the models can be parameterized using state-space representations, where the time-varying parameters are treated as latent, unobservable states driving the dynamics of the system (see e.g. Durbin and Koopman 2012, Ch. 6; Lubik and Matthes 2016). Another possibility is to treat the parameter dynamics as deterministic, e.g. through switching regression and smooth-transition models. Here, the model parameters change their regime as a function of certain covariates (see e.g. Hubrich and Teräsvirta 2013 for an overview).

1.1.2. The need for robust statistics

In any real-life application, there is the possibility of outlying samples within the set of observed values. Such observations can both be caused by measurement error (e.g. because of miscalibrated measurement devices) or occur naturally if the observation does not follow the structure imposed by the statistical model used. In classical statistical analysis, even

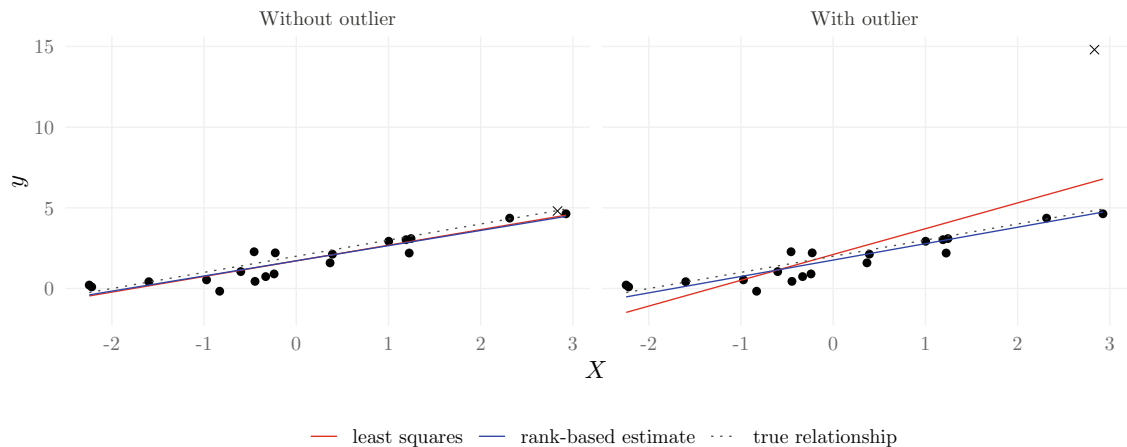


Figure 1.2.: Comparison of a non-robust (least-squares) estimate and a robust estimate (rank-based regression, R-estimate) in the presence of an outlying observation (marked by \times). True relationship: $y = 2 + 1 \cdot X + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, 0.25)$.

single outlying values can cause the results of the analysis to be biased. A good introduction to the effects of outliers on statistical analyses is given in Maronna (2019, Ch. 1). However, the outliers are not just “bad” data points, but can also contain valuable information that we can learn from. It is desirable to identify outliers using proper model diagnostic tools. However, when applying such methods to non-robust model fits, outlying values can easily be masked.

The field of *robust statistics* aims to address these issues. Early developments in this field go back to Tukey (1960), Huber (1964) and Hampel (1968). An important contribution of robust statistics is the recognition that all statistical models and methods are only an approximation of the truth. While classical statistics often rely on models that are optimal under certain strong assumptions (such as Gaussianity), their performance can deteriorate drastically if those assumptions are not fulfilled. In this case, the goal in robust statistics is to find estimators that are not affected by slight deviations from the assumptions and still provide reliable results in the presence of outlying or unusual observations. An example is displayed in Figure 1.2. It shows a simple regression setting where we have introduced one outlier to the response variable y . When comparing the setting with and without outliers, it is clearly visible how the outlying observation compromises the estimated regression relationship in the case of a non-robust model (red). Opposed to that, the robust (in this case rank-based) regression estimate (blue) is not affected by the outlying value.

The usual assumption in robust statistics is that the majority of the observations (at least roughly) follows the imposed model structure and assumptions, while there can be a small proportion of data points that deviate from those assumptions. There are various approaches to obtain robust estimates, which can be applied in different contexts: estimation of uni-

and multivariate location and scatter parameters, as well as the estimation of regression parameters. There are three main types of estimators in robust statistics: maximum likelihood type estimators (M-estimators), estimators based on linear combinations of order statistics (L-estimators), and estimators based on rank statistics (R-estimators). Some basic robustness properties of M-, L-, and R-estimators are outlined in Chapter 3 of Huber (1981, Ch. 3) and in Jurečková (1984).

M-estimators are a generalization of maximum likelihood estimation. By replacing the likelihood function in an ML estimator by a suitable loss function, commonly denoted by ρ , the influence of extreme or unusual observations (or residuals, in the case of regression) can be bounded. Common ρ -functions include the Huber loss $\rho_k(x) = x^2 \mathbf{1}\{|x| \leq k\} + (2k|x| - k^2) \mathbf{1}\{|x| > k\}$ and the biweight loss $\rho_k(x) = [1 - (1 - \frac{x}{k})^2]^3 \mathbf{1}\{|x| \leq k\} + \mathbf{1}\{|x| > k\}$ for some cutoff k . It is treated as a tuning parameter and regulates the trade-off between efficiency and robustness (Maronna 2019, Ch. 1). Commonly used in regression models are *S*- and *τ -scale estimators*, which minimize the residual scale using concepts from M-estimation, and MM-estimators, which combine M- and S-estimators to achieve high robustness while maintaining a high efficiency (Yohai 1987).

L-estimators are obtained from order statistics of the observations. In the simplest case, the resulting estimators are trimmed or winsorized means and standard deviations, or quantiles such as the median. For regression, a corresponding estimator is given by the least trimmed squares (LTS) estimator (Rousseeuw 1984). For multivariate data, the concept of L-estimators can be generalized to estimators that look for in some sense optimal subsets of the observations to calculate the estimate. One example is the minimum covariance determinant (MCD) estimator. It provides a robust estimate of the covariance matrix based on the subset of observations whose covariance matrix has the smallest determinant (Rousseeuw 1985; Rousseeuw and Driessen 1999).

The third class is given by *R-estimators*. Those estimators are based on score functions that utilize the ranks of the observations instead of the observations themselves. Precisely, R-estimators are derived from rank tests. Examples of such estimators include the (one- and two-sample) Hodges-Lehmann estimator of location (Hodges and Lehmann 1963) and the rank-based regression methods introduced later in this chapter. A unified overview of robust R-estimation in different contexts is given in Hettmansperger and McKean (2011).

1.1.3. Outline of the thesis

This thesis develops statistical models for complex longitudinal and multivariate time series data in the presence of outliers and structural changes. It is structured into three main chapters, each of them based on a separate article. Chapter 2 presents an approach to multivariate time series regression. It incorporates time-variation into the parameter matrices of the reduced-rank regression model. The following two chapters introduce and extend robust modeling approaches for longitudinal datasets. Chapter 3 presents a rank-based estimation method for linear mixed effects regression models. In Chapter 4, a model for repeated measurements functional data is presented and a robust estimation method is

proposed. Both methods provide reliable estimates in the presence of outlying observations and allow to identify outliers and analyze their causes. Finally, in Chapter 5, we summarize the contributions and discuss directions and topics for future research.

In the remainder of this chapter, important concepts and model classes relevant to this thesis are introduced. The three sections are oriented along the three data types analyzed: Multivariate time series, classic longitudinal data and repeated measurements functional data.

1.2. Modeling multivariate time series data

1.2.1. Definition and classification

Assume we observe a dataset consisting of measurements y_{it} , $i = 1, \dots, p$,¹ collected at different (usually equidistant) time points t according to the order of time. They can be written as a vector $\mathbf{y}_t = (y_{1t}, \dots, y_{pt})$, where y_{it} describes the i 'th component variable at time t . The series $\mathbf{y}_1, \mathbf{y}_2, \dots$ of p -variate observations is then called a *multivariate* or *vector time series*. The fundamental trait of such data is that there can be dependency structures both within the same component at different time points s and t , but also between different components i and j at different time points s and t . Capturing those complex dependency structures is one of the main topics in the analysis of multivariate time series.

For illustration, consider the time series of daily closing prices for seven stock indices: DAX Performance Index (DAX), Dow Jones Industrial Average (DJI), Hang Seng Index (HSI), Nikkei 225 (N225), NASDAQ Composite (NASDAQ), S+P 500 (S+P 500) and TSEC Weighted Index (TWII).² The time series are displayed in Figure 1.3 for the time period between June 1st, 2007 to June 1st, 2009. In this case, the index i corresponds to the stock index, while t represents the date the measurement was taken. Each index can be interpreted as an observational unit from which we take measurements on a regular grid in time. As can be seen, the indices differ in magnitude or level, but at the same time show interdependency structures: The courses of the time series are similar. Especially, all series show effects of the Great Recession in 2008. It can be assumed that the value of one stock index can (at least partly) be explained by the values of the other time series. A more detailed analysis of this dataset is given in Section 2.4 of Chapter 2, p. 49f.

Multivariate time series are characterized through their mean and covariance structure. A goal is often to build forecasting models that predict future values of the time series. We especially focus on two aspects:

- (1) Building a model that represents the multivariate time series data as a (linear) function of its lagged values.

¹While the number of observational units was denoted by g in Equation (1.1), we denote it by p in this section to avoid confusion in Chapter 2 which builds on this subsection.

²All time series retrieved from <https://finance.yahoo.com/>, accessed March 01, 2021.

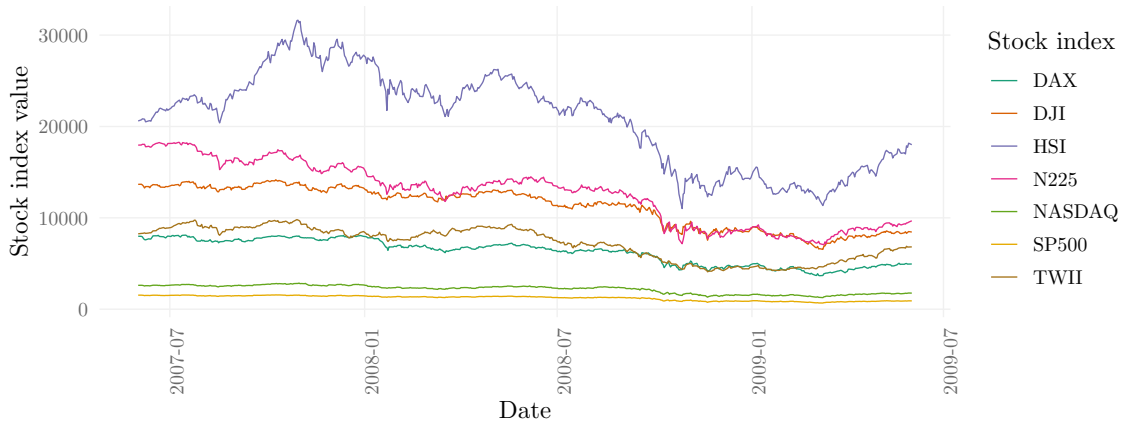


Figure 1.3.: Multivariate time series of daily closing prices for DAX Performance Index (DAX), Dow Jones Industrial Average (DJI), Hang Seng Index (HSI), Nikkei 225 (N225), NASDAQ Composite (NASDAQ), S+P 500 (S+P 500) and TSEC Weighted Index (TWII) from June 1st, 2007 to June 1st, 2009.

- (2) Performing dimension reduction, since the number of parameters for the relevant model classes increases with p^2 for a p -dimensional time series.

In order to be able to form meaningful predictive models, a common assumption is stationarity. We call a p -dimensional time series *stationary* in its first and second moment, if

$$\mathbb{E}(\mathbf{y}_t) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_p) \quad \forall t, \text{ and}$$

$$\text{Cov}(\mathbf{y}_t, \mathbf{y}_{t+s}) = \mathbb{E}((\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t+s} - \boldsymbol{\mu})') = \boldsymbol{\Gamma}(s) \quad \forall t, s.$$

$\boldsymbol{\Gamma}(s)$ is a matrix with entries $\gamma_{ij}(s) = \mathbb{E}((y_{it} - \mu_i)(y_{j(t+s)} - \mu_j))$ for $i, j = 1, \dots, p$. Stationarity means that the mean of the time series is time constant, and that the cross covariances at times t and $t + s$ are functions only of the time differences s . Thus, the relationships between the observations do not change over time (Lütkepohl 2005, Ch. 2).

A general linear time series regression model Given a multivariate time series $\mathbf{y}_t \in \mathbb{R}^p$ and a set of explanatory time series $\mathbf{x}_t \in \mathbb{R}^q$, a general linear time series regression model is given by

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \boldsymbol{\varepsilon}_t \quad (1.2)$$

where $\mathbf{C} \in \mathbb{R}^{p \times q}$ is a coefficient matrix and $\boldsymbol{\varepsilon}_t$ are independent and centered error terms. The predictor series \mathbf{x}_t can consist of lagged values of \mathbf{y}_t , e.g. $\mathbf{x}_t = \mathbf{y}_{t-1}$, but also of external variables. Those can be observable or unobservable. Due to the dependencies on time, the time series regression model needs to be set up carefully to account for potential trends and seasonalities, autocorrelation, non-stationarity and heteroscedasticity in the time series. The

models presented in the following subsection place different assumptions on the coefficient matrices C and observed or unobservable predictor time series \mathbf{x}_t .

1.2.2. Model classes of interest

In the following, we present different parameterizations of the linear time series regression model (1.2) that are relevant to this thesis. The presented model classes are commonly used in time series analysis and aim to provide accurate predictions while, if possible, reducing the number of model parameters. For simplicity of the exposition, we assume $\mathbb{E}(\mathbf{y}_t) = 0$ for all $t = 1, \dots, T$.

The Vector Autoregressive (VAR) model A p -dimensional vector autoregressive process or model of order m , also written as VAR(m), is given by

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_m \mathbf{y}_{t-m} + \boldsymbol{\varepsilon}_t \quad (1.3)$$

where $\boldsymbol{\varepsilon}_t$ is a p -dimensional white noise process (Wei 2019, Ch. 2). The matrices $\mathbf{A}_j \in \mathbb{R}^{p \times p}$, $j = 1, \dots, m$ are coefficient matrices that reflect the relationships between the current observation \mathbf{y}_t and its lags $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-m}$. For this thesis, the VAR(1) model

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad (1.4)$$

will be of special interest.

The model parameters in (1.3) can be estimated by multivariate ordinary least squares (OLS). However, under the assumption that the parameter matrices in the VAR model (1.3) and the general time series regression model (1.2) are of full rank, classic OLS estimation applied separately to the p equations in the models yields the same results as multivariate least squares and we can gain no further insight by estimating the equations jointly (Izenman 1975). Furthermore, due to the large number of parameters, VAR models are often overparameterized. We outline some common approaches to introduce more structure to the models (1.2), (1.3) and (1.4) in the following paragraphs.

Dimension reduction through reduced-rank regression An approach to reduce the number of parameters in multivariate regression models is offered by reduced-rank regression (Izenman 1975). For the general time series regression model (1.2), it is assumed that the coefficient matrix $C \in \mathbb{R}^{p \times q}$ is of reduced rank $d < \min\{p, q\}$. This implies that C decomposes as the product of two rank d matrices $\boldsymbol{\alpha} \in \mathbb{R}^{p \times d}$ and $\boldsymbol{\beta} \in \mathbb{R}^{q \times d}$ such that model (1.2) can be written as

$$\mathbf{y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t.$$

In the special case of a VAR(1) model as in (1.4), the model equation is then given by

$$\mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

(Velu et al. 1986). Instead of fitting separate regression models for each component of \mathbf{y}_t as in (1.2) and (1.3), the dimension is reduced by the assumption that the y_{it} can be explained by a total of d linear combinations of the explanatory series \mathbf{x}_t . Thus, this formulation introduces linear restrictions into the multivariate regression model. Solutions to the reduced-rank regression problem can be obtained by minimizing a criterion of form

$$\mathbb{E}(\text{tr}(\boldsymbol{\Gamma}^{1/2}(\mathbf{y}_t - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_t)(\mathbf{y}_t - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_t)'\boldsymbol{\Gamma}^{1/2})),$$

where $\boldsymbol{\Gamma}$ is a positive definite and symmetric matrix with matrix square root $\boldsymbol{\Gamma}^{1/2}$. Depending on the choice of $\boldsymbol{\Gamma}$, the solutions are closely related to other well-known concepts from multivariate statistics, e.g. principal component analysis ($\boldsymbol{\Gamma} = \mathbf{I}_{p \times p}$) or canonical correlation analysis ($\boldsymbol{\Gamma} = \text{Cov}(\mathbf{y}_t, \mathbf{y}_t)^{-1}$). Note that the decomposition into $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is not unique, since for any non-singular ($d \times d$)-matrix \mathbf{H} , we have

$$\mathbf{C} = \boldsymbol{\alpha}\boldsymbol{\beta}' = (\boldsymbol{\alpha}\mathbf{H})(\mathbf{H}^{-1}\boldsymbol{\beta}').$$

A detailed introduction to reduced rank-regression and its applications in multivariate time series analysis can be found in the book by Reinsel and Velu (1998).

Factor models for multivariate time series analysis Another important approach to dimension reduction in multivariate time series analysis is that of factor analysis. Here, we assume that the dynamics of \mathbf{y}_t are driven by a few, latent factors. Thus, we can define a model

$$\mathbf{y}_t = \mathbf{L}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \tag{1.5}$$

where $\mathbf{f}_t \in \mathbb{R}^k$, $k \ll p$, are the latent factors and $\mathbf{L} \in \mathbb{R}^{p \times k}$ is the matrix of loadings that maps the factors to each of the time series in \mathbf{y}_t . $\boldsymbol{\varepsilon}_t$ is the error term.

There are different approaches to the estimation of such models. One possibility for estimation of \mathbf{L} and the factors \mathbf{f}_t is using principal component analysis, and letting \mathbf{f}_t consist of the first $k \ll p$ principal components of the series \mathbf{y}_t . By assuming the errors and factors follow a normal distribution, we can also use maximum likelihood estimation to estimate \mathbf{L} . A third commonly observed situation is the case where the factors are observable. E.g., in economic studies macroeconomic variables such as inflation rate, employment and unemployment rates can be considered as factors driving an economic time series of interest. If the time series of factors are observable, estimating the loadings \mathbf{L} comes down to a multivariate regression problem where \mathbf{L} is the unknown coefficient matrix and the factors \mathbf{f}_t form the predictors. A detailed introduction to factor models for multivariate time series is given in Chapter 5 of Wei (2019).

The factor model in (1.5) does not impose a time-dependent structure on the factors. A

natural extension to this model is to assume that the factors follow a time series model. Such models are called *dynamic factor models* (DFM), and were first proposed and applied in Geweke (1977) and Sargent and Sims (1977). An example of a simple dynamic factor model is

$$\begin{aligned} \mathbf{y}_t &= \mathbf{L}\mathbf{f}_t + \boldsymbol{\varepsilon}_t \\ \mathbf{f}_t &= \boldsymbol{\Phi}\mathbf{f}_{t-1} + \mathbf{u}_t, \end{aligned} \tag{1.6}$$

where the factors \mathbf{f}_t are assumed to follow a VAR(1) model. In this case, the two model equations form a *state-space system* (see Section 1.2.3) where the factors are considered as the latent states driving the model. An introduction to DFMs can be found in Stock and Watson (2012).

In the stock index example (Figure 1.3), the observed time series exhibit very similar shapes, although at different magnitudes. Thus, they can likely be assumed to be driven by common factors. However, they are obviously non-stationary and exhibit trends. Thus, a more sophisticated model might be necessary.

Cointegration and the vector error correction model Oftentimes, the time series of interest are non-stationary, e.g. because they exhibit a trend, and thus a varying mean. A common approach to address this type of non-stationarity is to take the first differences of the time series, denoted by

$$\Delta \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}.$$

In the case of a trend in the mean, the first differences $\Delta \mathbf{y}_t$ are stationary. Such time series are sometimes called integrated at order 1 (or I(1), see Hamilton 1994, Ch. 15). However, in the case of multivariate time series, a commonly observed phenomenon is *cointegration* of the time series. In this case, there exist linear combinations

$$p_t = y_{1t} - \beta_1 y_{2t} - \dots - \beta_{p-1} y_{pt} \tag{1.7}$$

of the components $(y_{1t}, \dots, y_{pt})'$ that are stationary. Although the processes y_{it} might be individually non-stationary, they move together such that their linear combination p_t forms a stationary time series. This is the case if the time series are driven by a common stochastic trend. The coefficient vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p-1})'$ in (1.7) describes a so-called *cointegrating relationship* or *relation*. There can also exist multiple cointegrating relationships.

The *vector error correction model* (VECM) uses the information from the cointegrating relations to improve the forecasting performance and gain a more accurate model for the time series. The model equation is set up as

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \sum_{\ell=1}^m \boldsymbol{\Gamma}_\ell \Delta \mathbf{y}_{t-\ell} + \boldsymbol{\varepsilon}_t,$$

where the matrix $\boldsymbol{\beta} \in \mathbb{R}^{p \times d}$ contains the d cointegrating relationships. $\boldsymbol{\alpha} \in \mathbb{R}^{p \times d}$ maps the

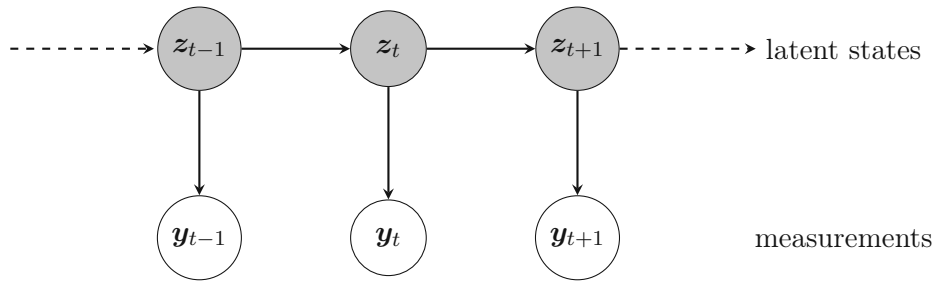


Figure 1.4.: State-space system: Latent states z_t influencing the measurements y_t

cointegrating relationships back to the first differences Δy_t . The matrices Γ_ℓ are coefficient matrices for the further lags of Δy_t . There are different estimation approaches for VECM, e.g. using least squares or maximum likelihood. The VECM can also be interpreted as a special case of reduced-rank regression as introduced above. A detailed introduction to the concept of cointegration and to vector error correction models is given in Part II of Lütkepohl (2005).

1.2.3. Models with time-varying parameters (TVP)

Classical time series analysis usually assumes that the relationships and parameter matrices are constant over time. Especially when the time series cover a long time span, this might not be realistic. For example, in the case of the Great Recession in 2008 seen in the stock index example (Figure 1.3), the relationships and time dynamics have likely changed during that time period.

This problem can be addressed by employing a time series regression model with time-varying parameters (TVP). Extending the model presented in Equation (1.2) with a time-varying coefficient matrix yields a model of form

$$y_t = C_t x_t + \varepsilon_t. \quad (1.8)$$

As already outlined in Section 1.1.1, there are many ways to parameterize the variation in C_t . We focus on the representation through *state-space models*.

State-space models State-space models can be seen as generalizations to regular time series models. They are used extensively in physical sciences and engineering, but are also very useful in the analysis of multivariate time series. In state-space modeling, the idea is that an observed multivariate time series y_1, \dots, y_T depends on a series of possibly unobserved states, denoted by z_t , $t = 1, \dots, T$ (see Figure 1.4). The model formulation was initially developed in the setting of spacecraft tracking. The exact location of the spacecraft, represented by the latent state z_t , is inferred from measurable quantities such as the velocity and azimuth of the spacecraft or its distance to certain reference points in space, represented by y_t (Shumway and Stoffer 2017, Ch. 6). Applications of state-space models can be found

in many different fields, including navigation, aerospace engineering, telecommunications, finance, medical, etc. (Särkkä 2013, Ch. 1). Detailed introductions to state-space models can e.g. be found in Lütkepohl (2005, Ch. 18), Shumway and Stoffer (2017, Ch. 6) and the book by Särkkä (2013).

We are concerned with linear Gaussian state-space models. In this case, the evolution of the states and their influence on the observed time series are structured into an equation system of the form

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{z}_t + \mathbf{v}_t \quad (1.9)$$

$$\mathbf{z}_t = \mathbf{B}_{t-1} \mathbf{z}_{t-1} + \mathbf{w}_{t-1}. \quad (1.10)$$

Equation (1.9) is called *measurement equation* and defines how the observed time series evolves from the unobserved states. Equation (1.10) describes how the states evolve in time and is commonly called *state equation*. The transition or coefficient matrices \mathbf{H}_t and \mathbf{B}_t can be dependent on time. The error terms \mathbf{v}_t and \mathbf{w}_t are assumed to be uncorrelated Gaussian white noise processes, with potentially time-varying covariance matrices. For simplicity, we assume the covariance matrices to be time constant, yielding

$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad \text{and} \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}). \quad (1.11)$$

The time-varying parameter model (1.8) can be represented as a state-space system as follows:

$$\mathbf{y}_t = \mathbf{C}_t \mathbf{x}_t + \boldsymbol{\varepsilon}_t = (\mathbf{x}'_t \otimes \mathbf{I}_{p \times p}) \text{vec}(\mathbf{C}_t) + \boldsymbol{\varepsilon}_t \quad (1.12)$$

$$\text{vec}(\mathbf{C}_t) = \mathbf{A} \text{vec}(\mathbf{C}_{t-1}) + \mathbf{v}_{t-1}. \quad (1.13)$$

$\text{vec}(\cdot)$ describes the vectorization operator which stacks the entries of the matrix columnwise into a vector. Thus, the state equation (1.13) describes how the (vectorized) coefficient matrices \mathbf{C}_t evolve over time (in this case as an AR(1) process with coefficient matrix $\mathbf{A} \in \mathbb{R}^{pq \times pq}$), and the measurement equation (1.12) connects the states with the observed data. The predictors \mathbf{x}_t correspond to the matrices \mathbf{H}_t in (1.9).

Other applications of state-space models in time series analysis include dynamic factor models as in Equation (1.6). Here, the factors \mathbf{f}_t can be treated as the latent states driving the observations.

State estimation through the Kalman filter The primary goal in the analysis of state-space models is to derive predictions of the latent states from the observed data. The *Kalman filter* offers the linear predictions of the state at time t with minimum mean squared error among all linear predictions, given the information up to time $s \leq t$. In the case of Gaussian errors, those predictions correspond to the expected values

$$\mathbf{z}_t^s = \mathbb{E}(\mathbf{z}_t | \mathbf{y}_t, \dots, \mathbf{y}_s).$$

The filter also provides estimates of the error covariances, i.e. the uncertainty about the state at a given time t :

$$\mathbf{P}_t^s = \text{Cov}(\mathbf{z}_t | \mathbf{y}_t, \dots, \mathbf{y}_s).$$

What makes the Kalman filter special is its ability to account for measurement errors. It uses a “feedback loop” based on the following two steps, denoted as *prediction step* and *updating step*:

Prediction step Based on the current estimate of the state and the state equation, a one-step ahead prediction of the state is obtained as \mathbf{z}_t^{t-1} . It is accompanied by an uncertainty estimate, given by the corresponding covariance matrix \mathbf{P}_t^{t-1} .

Updating step The predicted state from the prediction step is updated with the new measurement \mathbf{y}_t to obtain \mathbf{z}_t^t . The crucial information for the update is captured by the so-called *Kalman gain*, commonly denoted by \mathbf{K}_t . It is based on the uncertainty about the state estimate and the current measurement and thus determines how much the state prediction needs to be corrected given the current observation.

Given starting values \mathbf{z}_0^0 and \mathbf{P}_0^0 , we can obtain a series of estimated states \mathbf{z}_t^t and covariances \mathbf{P}_t^t by iterating between updating and predicting for each t .

Especially if the state-space system includes unknown parameters, e.g. the error covariance matrices \mathbf{R} and \mathbf{Q} for the terms \mathbf{v}_t and \mathbf{w}_t or the matrices \mathbf{H}_{t-1} and \mathbf{B}_t in the system (1.9) and (1.10), estimates of the states given the complete time series can be of interest. For this, the *Kalman smoother* can be employed, providing estimates of

$$\mathbf{z}_t^T = \mathbb{E}(\mathbf{z}_t | \mathbf{y}_t, \dots, \mathbf{y}_T) \quad \text{and} \quad \mathbf{P}_t^T = \text{Cov}(\mathbf{z}_t | \mathbf{y}_t, \dots, \mathbf{y}_T).$$

Those are obtained by processing the data backward in time to improve and smooth the estimated states obtained from processing the data forward with the Kalman filter. For details on both Kalman filter and smoother, see Shumway and Stoffer (2017, Ch. 6).

The state space systems considered in this section are assumed to be linear and Gaussian. However, there also exist extensions with non-linear state and measurement equations. For those, we can employ techniques such as the extended Kalman filter or simulation-based particle filters to estimate the latent states. A detailed overview can be found in Särkkä (2013).

Estimation of unknown parameters In the previous paragraph, we have introduced the Kalman filter and smoother as tools to derive predictions of the latent states given the observations. However, the state estimates also depend on the parameters of the state-space system – the transition matrices \mathbf{B}_t and \mathbf{H}_t , as well as the error covariances \mathbf{R} and \mathbf{Q} in (1.11). Given the normality assumption, the unknown parameters can be estimated using ML estimation. One possibility is to directly maximize the log-likelihood of the *innovations* $\boldsymbol{\varepsilon}_t^{t-1} = \mathbf{y}_t - \mathbf{H}_t \mathbf{z}_t^{t-1}$. Their likelihood is a complex, non-linear function of the unknown parameters. It can be optimized by employing a Newton-Raphson algorithm. A numerically

easier approach is provided by the *Expectation-Maximization (EM) algorithm* (Dempster et al. 1977). It is designed to obtain ML estimates in the presence of missing data. In our case, the latent states can be treated as missing data and we optimize the joint log-likelihood of the latent states $\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_T$ and the observations $\mathbf{y}_1, \dots, \mathbf{y}_T$. The EM algorithm consists of two estimation steps, the E-step and the M-step.

E-step In the E-step, we calculate the expected value of the log-likelihood with respect to the latent states, given the observed data and the current parameters.

M-step In the M-step, the parameter estimates are updated by maximizing the expected log-likelihood with respect to the unknown parameters.

By iterating between those two steps until convergence, we obtain estimates of the unknown parameters. The EM algorithm has optimality properties. It is guaranteed to increase the log-likelihood with every iteration. However, it cannot be guaranteed to converge to the ML estimate and might return a local optimum if the log-likelihood is not convex. In this case, it can be useful to restart the EM algorithm multiple times with different starting values. More information and technical details on the EM algorithm for linear state-space models, (1.9) and (1.10), can be found in Shumway and Stoffer (2017, Ch. 6).

1.3. Mixed effects models for longitudinal data

1.3.1. Definition and classification

Opposed to the multivariate time series data described in the previous section, “classic” longitudinal data can be characterized in a slightly more general way. Given a set of observational units or subjects $i = 1, \dots, g$, we repeatedly observe each subject i at time points t_{ij} , $j = 1, \dots, n_i$, yielding a time series of observations for each of them. The sampling time points t_{ij} and the number of samples per subject n_i can differ from subject to subject. Thus, the designs can be unbalanced, on irregular grids, or datasets can be partially incomplete. As in the case of multivariate time series datasets, we have dependencies within each of the observational units, but also some general effects driving the measurements for all subjects. Often, such datasets are analyzed in a regression context, however, they can also be analyzed in other settings or with different goals in mind (e.g., see Section 1.4.3 for a description of methodology to analyze longitudinally sampled functional data). A closely related situation is sampling from clusters or groups. Here, analogous data structures are introduced through similarities within observations of the same cluster or group. Examples for such datasets are samples of groups of patients from different hospitals, or students within classes from different school. In this case, the index ij corresponds to the j 'th sample taken from the i 'th cluster. However, for simplicity, we limit our exposition here to datasets with an inherent ordering in time.

Throughout this section, we assume a longitudinal dataset consists of measurements $(y_{ij}, \mathbf{x}'_{ij})'$, $y_{ij} \in \mathbb{R}$, $\mathbf{x}_{ij} \in \mathbb{R}^p$ for g units $i = 1, \dots, g$ with observations $j = 1, \dots, n_i$ at measurement time points t_{i1}, \dots, t_{in_i} for unit i . The goal is to describe the target variable y_{ij} as a function

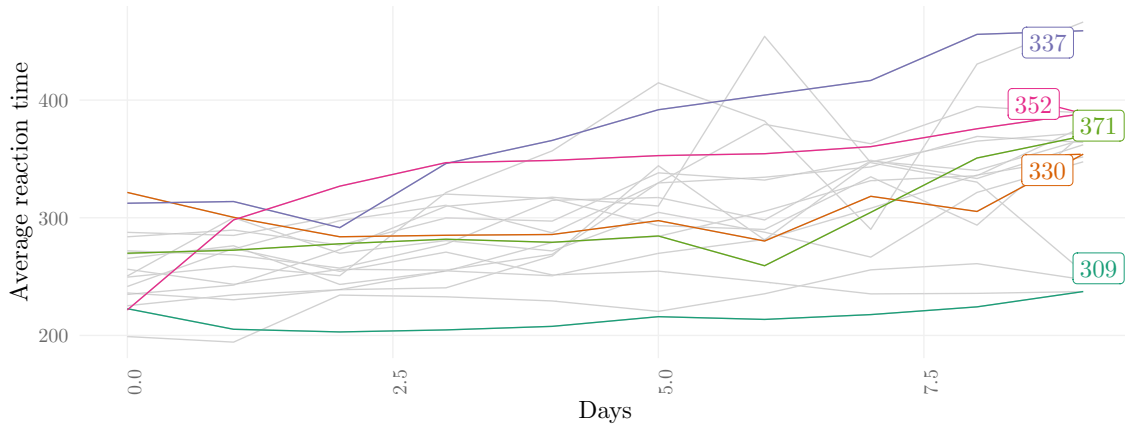


Figure 1.5.: Example for a longitudinal dataset: Average reaction times of 18 subjects after up to 10 days of sleep deprivation. Five randomly selected subjects are highlighted for clarity.

of the p -dimensional predictors \mathbf{x}_{ij} while accounting for the dependencies induced by the grouped structure of the data.

For illustration, consider the *sleepstudy* dataset (Belenky et al. 2003). The data were collected in a study about sleep deprivation. The subjects in the study were allowed to sleep only for a certain number of hours every night, and had to complete a reaction test every day. The reaction times for the different subjects were recorded. A subset of this dataset is displayed in Figure 1.5. The subjects react differently to the sleep deprivation, and also the baseline reaction time differs from subject to subject. In this case, y_{ij} is the j 'th reaction time for the i 'th subject, and $\mathbf{x}_{ij} \in \mathbb{R}$ is the number of days of exposure corresponding to this measurement. If we fitted a simple linear model to explain the average reaction time by the number of days of sleep deprivation, we would end up with systematically correlated errors for the different subjects of the study. A more detailed analysis of this dataset is given in Section 3.5 of Chapter 3, p. 72f.

Mixed effects models

To address the shortcomings of simple linear models in the presence of correlated errors, regression problems in longitudinal datasets are commonly solved using *mixed effects models*. The predictors $\mathbf{x}'_{ij}\boldsymbol{\beta}$ in the standard linear regression model

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i,$$

are amended to form a mixed predictor $\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$, resulting in a model equation

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i.$$

The predictors $\mathbf{z}_{ij} \in \mathbb{R}^k$ are usually a sub-vector of the covariates \mathbf{x}_{ij} , and the coefficients \mathbf{b}_i describe the subject-specific effects for subject i . While those can be modeled as fixed predictors, this can quickly lead to a very large number of parameters that need to be estimated. Alternatively, the coefficients \mathbf{b}_i can be assumed to be realizations of a random variable \mathbf{b} . Commonly, we assume $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Instead of having to estimate each of the effects \mathbf{b}_i separately, we only need to estimate $\mathbf{\Sigma}$. This saves parameters and allows to account for the information about the between-subject variation. This can especially be beneficial if the number of available samples per subject is small. The distributional assumptions for the random effects induce a certain regularization. Another useful aspect is that the random effects allow to correct for unobserved heterogeneity which can be induced by omitted covariates, and can absorb unexplained effects to a certain extent. Mixed effects models can handle balanced as well as unbalanced datasets. Thus, it is not necessary to observe the same number of samples per subject (Fahrmeir et al. 2013, Ch. 7).

For subject i , the resulting model equation is given by

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (1.14)$$

where $\mathbf{X}_i \in \mathbb{R}^{n_i \times p}$ is referred to as the *fixed effects design matrix* and $\mathbf{Z}_i \in \mathbb{R}^{n_i \times k}$ as the *random effects design matrix*. $\boldsymbol{\beta} \in \mathbb{R}^p$ are called the global regression coefficients or fixed effects, while $\mathbf{b}_i \in \mathbb{R}^k$ are the random effects, centered and with variance-covariance matrix $\mathbf{\Sigma}$. The error terms $\boldsymbol{\varepsilon}_i$ are assumed to be centered with variance-covariance matrix \mathbf{V}_i and independent of the random effects \mathbf{b}_i . The matrices \mathbf{V}_i are usually assumed to depend on i only through their dimension, e.g. by letting $\mathbf{V}_i = \sigma^2 \mathbf{I}_{n_i}$, where \mathbf{I}_{n_i} is the $n_i \times n_i$ dimensional identity matrix. Commonly, $\mathbf{\Sigma}$ and \mathbf{V}_i are parameterized as functions of a common parameter (vector) $\boldsymbol{\theta}$ and denoted by $\mathbf{\Sigma}(\boldsymbol{\theta})$ and $\mathbf{V}_i(\boldsymbol{\theta})$. If $\mathbf{z}_{ij} = 1$ for all $i = 1, \dots, n$ and $j = 1, \dots, k$, we call $\mathbf{b} \in \mathbb{R}^k$ a *random intercept*. If \mathbf{z}_{ij} (additionally) contains other, real-valued predictors, we talk about *random slopes*. Chapter 7 in Fahrmeir et al. (2013) gives a good introduction to different specifications of mixed effects models. Other standard references include Pinheiro and Bates (2000) and Davis (2003).

The model formulation presented in (1.14) is the simplest form of a mixed effects model where just one grouping factor is involved. If the observations are grouped by multiple factors, more complex types of models can be involved. Two types of effects are commonly analyzed: *nested effects*, where groups are formed within groups (e.g. classes within schools), and *crossed effects*, where multiple factors can be crossed. Crossed effects occur when all levels of one factor co-occur with the levels of a second factor. In both cases, the specification of the random effects and their covariances are more complex. Such models are out of scope for this work.

1.3.2. Estimation approaches

In this subsection, we will discuss some estimation approaches for mixed effects models. Opposed to the classic linear model, where we estimate the regression coefficients $\boldsymbol{\beta}$ and the error variance σ^2 , the mixed effect model additionally requires estimation of the variance

parameters in $\boldsymbol{\theta}$.

Estimation under the assumption of Gaussianity

The most common assumption is that the random effects \mathbf{b} and the independent error terms $\boldsymbol{\varepsilon}$ follow normal distributions, i.e.

$$\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})), \text{ and } \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta})). \quad (1.15)$$

Under Gaussianity, we thus have

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{Z}' + \mathbf{V}(\boldsymbol{\theta})), \text{ and} \quad (1.16)$$

$$\mathbf{y}|\mathbf{b} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{V}(\boldsymbol{\theta})), \quad (1.17)$$

respectively. For notational clarity, we have dropped the subject index i in Equations (1.15)–(1.17).

There are different ways of estimating the unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as well as predictors $\hat{\mathbf{b}}_i$ of the random effects in the mixed effects model (1.14) under Gaussianity. The three most commonly used methods are *maximum likelihood* (ML), *restricted maximum likelihood* (REML), and *Henderson's mixed model equations*. ML and REML optimize the joint log-likelihood of the *marginal* model (1.16) for the covariance parameter $\boldsymbol{\theta}$. This can either be achieved by direct optimization through profiling of $\boldsymbol{\beta}$ (ML), or by integrating $\boldsymbol{\beta}$ out of the likelihood (REML). REML has the advantage of achieving unbiased estimates of the variance components. However, ML estimates can have some advantages when it comes to hypothesis testing. In both cases, estimates of the random effects can be obtained by calculating posterior estimates. An overview of ML and REML is given in Chapter 7 of Fahrmeir et al. (2013). Another detailed reference, especially outlining the computational details of both ML and REML, is Chapter 3 in Pinheiro and Bates (2000). A third possibility is estimation based on Henderson's mixed model equations, where $\boldsymbol{\beta}$ and the random effects \mathbf{b}_i are estimated jointly from maximizing the joint log-likelihood of the *conditional* model (1.17) and of \mathbf{b} in (1.15). A detailed description is given in Searle et al. (1992, Ch. 7.6).

Robust modeling approaches

While the methodology based on the assumption of Gaussianity performs well in the case of clean, approximately normally distributed data, the results can be strongly affected by outlying observations (Pinheiro, Liu, et al. 2001). Thus, different methods to estimate mixed effects models robustly have been proposed in literature.

Estimation using rank-based regression In this thesis, we estimate mixed effects models using a notion of non-parametric regression, precisely rank-based regression models. For a regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ and centered error term $\boldsymbol{\varepsilon}$,

the least squares estimate is the solution to the minimization problem

$$\hat{\beta}_{\text{LS}} = \arg \min_{\beta \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\beta\|^2 = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - \mathbf{x}'_i \beta)^2,$$

where $\|\cdot\|$ denotes the Euclidean norm. By replacing the Euclidean norm by a loss function based on the ranks of the residuals from the regression, we can obtain an R-estimate of β as

$$\hat{\beta}_{\varphi} = \arg \min_{\beta \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\beta\|_{\varphi} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{j=1}^n a[R(y_i - \mathbf{x}'_i \beta)](y_i - \mathbf{x}'_i \beta). \quad (1.18)$$

$R(r_j)$ denotes the rank of the residual $r_j = y_i - \mathbf{x}'_i \beta$ among the components r_1, \dots, r_n of the residual vector $(r_1, \dots, r_n)'$. The function a is defined by $a[t] = \varphi(t/(n+1))$, where $\varphi(u)$ is a non-decreasing, bounded and square-integrable function such that $\sum_t a[t] = 0$. Popular choices for $\varphi(\cdot)$ are the *Wilcoxon score* $\varphi(u) = \sqrt{12} [u - \frac{1}{2}]$, and the *sign score* $\phi(u) = \text{sgn}(u - \frac{1}{2})$. $\|\cdot\|_{\varphi}$ in (1.18) can be shown to be a pseudo-norm and was proposed in Jureckova (1971). It is also referred to as *Jaekel's dispersion function* after Jaekel (1972). The R-estimator in (1.18) considers the signs of the residuals and implicitly tries to balance the number of positive and negative residuals. Thus, the influence the outliers can have on the resulting estimate is limited. This is also illustrated in Figure 1.2 in Section 1.1, where the regression line obtained from (1.18) is not affected by the outlying observation. Rank-based regression estimates based on the Wilcoxon score are almost as efficient as LS under normality (see Section 2.2 in McKean 2004), and can be a lot more efficient in cases where the data deviates from normality. A detailed introduction to rank-based regression in different contexts is given in the book by Hettmansperger and McKean (2011). A more compact overview can be found in McKean (2004).

While initially proposed to estimate classic linear models, the concept of rank-based regression has been extended to account for the more complex structures in mixed effects models as in Equation (1.14). Kloke et al. (2009) present an extension to models with random intercepts. Those approaches are further extended to accommodate nested effects (i.e. multiple levels of random intercepts) in the PhD thesis by Bilgic (2012). The author studies different methods to fit mixed effects models with nested effects. One approach is obtained by adapting concepts from iteratively reweighted least squares, and estimating the regression coefficients and variance components in an iterative manner. However, the presented methodology is also limited to models with random intercepts. Models with random slopes cannot be estimated in those frameworks. In Chapter 3, we introduce a novel algorithm to fit mixed effects models using rank-based regression. The proposed methodology can be applied in the presence of random slopes and is robust against outlying responses y_{ij} and predictor values \mathbf{x}_{ij} and \mathbf{z}_{ij} , respectively.

Other approaches In addition to the methodology based on rank-based regression outlined in the previous paragraph, other ways to obtain robust estimates of the parameters in mixed effects models have been proposed in the literature. One stream are modeling approaches

based on the concepts of M-, S- and τ -scale estimation. Copt and Victoria-Feser (2006) estimate the mixed effects model using high-breakdown S-estimators to obtain the covariance matrices. Koller (2013) proposes the SMDM estimator which is based on robust adaptations of Henderson's estimating equations using ρ -functions. A third estimation method is given in Agostinelli and Yohai (2016) and based on τ -scale estimators. It is not only robust to contamination of whole observations, but also to cellwise contamination where the different cells of an observational unit may be independently contaminated.

Another way to robustify the estimates is to replace the normal distributions in (1.15) by heavy-tailed distributions. Such distributions are more likely to incorporate extreme values due to their heavy tails, and can thus serve as an alternative to model extreme residuals and random effects. Pinheiro, Liu, et al. (2001) propose an efficient estimation procedure based on the multivariate t-distribution with ν_i degrees of freedom for subject i . Model estimation can be achieved by different EM algorithms that treat the random effects and the unknown degrees of freedom ν_i as missing data.

1.4. Functional data analysis and models for longitudinal functional data

This subsection starts by introducing the basic concepts of functional data analysis and functional principal components analysis in Section 1.4.1 and 1.4.2, respectively. Then, specific models for longitudinally observed functional datasets are introduced in Section 1.4.3.

1.4.1. Definition and classification

Functional data analysis (FDA) is the field of statistics that deals with data that appear in the form of functions, images and shapes, or even more general objects (Wang, Chiou, et al. 2016). In this thesis, we assume we observe a random sample of real-valued functions $X_1(s), \dots, X_m(s)$ for $s \in \mathcal{S}$, where \mathcal{S} is a compact subset of the real line. Each random function $X_i(s)$ is assumed to be a realization of a univariate stochastic process

$$X : \Omega \rightarrow \mathcal{L}^2(\mathcal{S}). \quad (1.19)$$

$\mathcal{L}^2(\mathcal{S})$ denotes the space of all square-integrable processes, i.e. those with $\mathbb{E} [\int_{\mathcal{S}} X^2(s) ds] < \infty$. The space $\mathcal{L}^2(\mathcal{S})$ is a Hilbert space with inner product

$$\langle f, g \rangle_{\mathcal{S}} = \int_{\mathcal{S}} f(s)g(s) ds$$

(Kokoszka and Reimherr 2017, Ch. 11). The analysis of functional data poses challenges, since the underlying data structure is inherently infinite-dimensional. However, in real data applications the functions usually cannot be observed continuously. Instead, they will

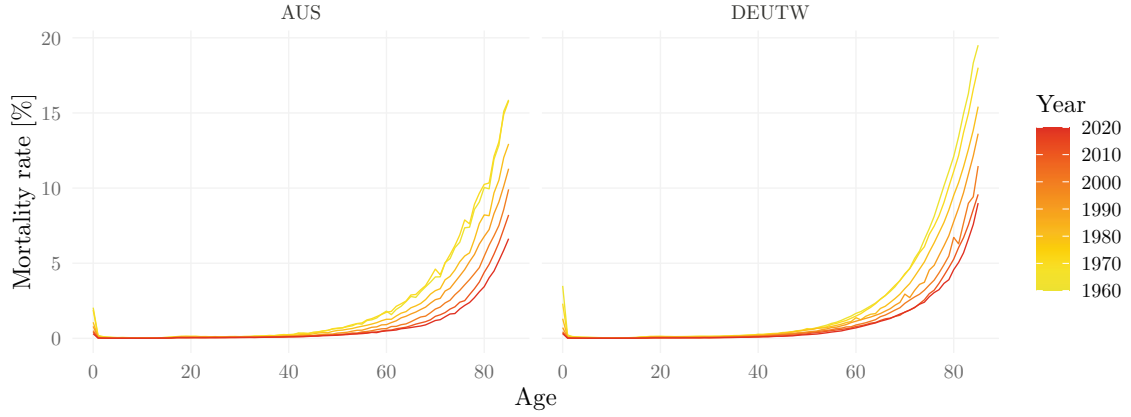


Figure 1.6.: Example for a functional dataset: Mortality rates for Austria (left) and (Western) Germany (right) between 1960 and 2020.

be observed on some discrete grid, denoted by s_{i1}, \dots, s_{ir_i} for observation i , $i = 1, \dots, m$. Depending on the application, this grid can be dense or sparse, regular or irregular. The number of sampling points r_i may be different for each curve.

Some examples of functional observations are displayed in Figure 1.6. It shows the mortality rate as a function of age in the time for Austria (left) and Germany (right). Each curve corresponds to a year between 1960 and 2020.³ The curves can be considered as realizations of a process like the one in (1.19). The sampling grid is discretized, each age j corresponds to a point s_{ij} on the sampling grid for the i 'th curve. The displayed data are a subset of a larger dataset which contains mortality curves for 36 countries. It is analyzed in detail in Section 4.6 of Chapter 4, p. 95f.

The processes X in (1.19) are characterized by their *mean function* and *covariance function*. The mean function of X is the expected value of the processes, i.e.

$$\mu(s) = \mathbb{E}(X(s)) \quad \forall s \in \mathcal{S}. \quad (1.20)$$

The covariance function of X is defined as a function $c : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ such that

$$\begin{aligned} c(s, t) &= \text{Cov}(X(s), X(t)) \\ &= \mathbb{E}[(X(s) - \mu(s))(X(t) - \mu(t))]. \end{aligned} \quad (1.21)$$

³Data obtained from the Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org (data downloaded on June 07, 2024).

Given a smooth representation of the samples, the above quantities can be estimated by

$$\hat{\mu}(s) = \frac{1}{m} \sum_{i=1}^m X_i(s), \text{ and } \hat{c}(s, t) = \frac{1}{m} \sum_{i=1}^m (X_i(s) - \hat{\mu}(s))(X_i(t) - \hat{\mu}(t)), \quad (1.22)$$

as described in Kokoszka and Reimherr (2017, Ch. 12). The estimators in (1.22) can be applied readily if the sampled data is available on a dense and regular grid. In cases where the samples are very sparse and irregular, we need to resort to more sophisticated estimation techniques and reuse information across the samples, for example by applying local linear smoothing estimators or bivariate smoothing algorithms [see for example Yao et al. (2005), Cai and Yuan (2011), Xiao et al. (2018), Boente and Salibián-Barrera (2021)].

1.4.2. Functional principal component analysis

Functional principal component analysis (FPCA) is one of the most prevalent tools in functional data analysis. It extends principal component analysis from multivariate data to functional data. It is a powerful tool to reduce the dimensionality of the infinite-dimensional functional observations. At the same time it captures the most significant features of the data.

The dimension reduction is achieved by an expansion of the random trajectories $X_i(s)$ into a functional basis that consists of the eigenfunctions of the *covariance operator* of the process X . Given the covariance function c as defined in Equation (1.21), the covariance operator is the kernel operator $C : \mathcal{L}^2(\mathcal{S}) \rightarrow \mathcal{L}^2(\mathcal{S})$, $\psi \rightarrow C\psi$, where

$$C\psi = \int_{\mathcal{S}} c(s, t)\psi(s)ds$$

for any $\psi \in \mathcal{L}^2(\mathcal{S})$. C is continuous and a compact Hilbert-Schmidt operator (Wang, Chiou, et al. 2016). Since it is symmetric and non-negative definite, it has real valued, non-negative eigenvalues, denoted by λ_k . Under mild conditions, Mercer's theorem (see e.g. Corollary 10.4.1 in Kokoszka and Reimherr 2017) implies that spectral decomposition of C allows to represent the covariance function c as

$$c(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t),$$

where $\{\lambda_k\}_k$ are the eigenvalues, and $\{\phi_k(\cdot)\}_k$ the corresponding orthonormal eigenfunctions of C . In 1946, Karhunen and Loève independently discovered the Karhunen-Loève (KL) expansion which allows to represent the processes $X_i(\cdot)$ as

$$X_i(s) = \mu(s) + \sum_{k=1}^{\infty} \alpha_{ik} \psi_k(s), \quad (1.23)$$

where $\psi_k(\cdot)$ for $k = 1, 2, \dots$ are the *functional principal components* (FPCs) and α_{ik} are the

corresponding *scores* for sample i . The scores are given by

$$\alpha_{ik} = \langle X_i - \mu, \phi_k \rangle_{\mathcal{S}} = \int_{\mathcal{S}} (X_i(s) - \mu(s)) \phi_k(s) ds. \quad (1.24)$$

Their expectation and variance are

$$\mathbb{E}(\alpha_{ik}) = 0, \quad \text{and} \quad \text{Var}(\alpha_{ik}) = \lambda_k.$$

Furthermore, they are uncorrelated across k and independent across i (Wang, Chiou, et al. 2016; Kokoszka and Reimherr 2017, Ch. 11.4). By truncating the representation (1.23) to the first $K \ll \infty$ components we can achieve dimension reduction and reduce the infinite-dimensional processes $X_i(s)$ to a finite-dimensional vector of scores for further analysis. The resulting decomposition is then given by

$$X_i(s) \approx \mu(s) + \sum_{k=1}^K \alpha_{ik} \psi_k(s). \quad (1.25)$$

Such dimension reduction can also be achieved by projecting onto other orthogonal basis functions such as B-splines or Fourier basis functions (see e.g. Chapter 3 in Ramsay and Silverman 2005). However, FPCA offers the best finite rank approximation of the functional data in terms of minimizing the mean squared error. The functional principal components (FPCs) explain the most variance among all finite-dimensional approximations of the processes. The number of components K can be chosen in different ways, usually by considering the sequence of eigenvalues $\{\lambda_k\}_k$. One possibility is the scree plot of the eigenvalues against the index. Another option is to analyze the proportion of variance explained (*pve*) and select components such that, e.g., 90% of the variance of the process are explained. The *pve* can be calculated from the cumulative sum of the eigenvalues λ_k , since by Parseval's identity, we have

$$\mathbb{E}(\|X - \mu, X - \mu\|_{\mathcal{S}}^2) = \mathbb{E}(\langle X - \mu, X - \mu \rangle_{\mathcal{S}}) = \sum_{k=1}^{\infty} \lambda_k,$$

i.e. the variance of the process X equals the sum of the variances of the projections of X onto the eigenfunctions or FPCs ψ_k (Kokoszka and Reimherr 2017, Ch. 11.4).

The estimation of (1.25) is achieved in two steps: From an estimate of the mean function $\mu(s)$ and the covariance function $c(s, t)$, we can obtain the spectral decomposition of the covariance operator by evaluating $c(s, t)$ on a fine grid and performing a standard spectral analysis of the resulting matrix. This gives us the eigenfunctions $\psi_k(s)$, $k = 1, \dots, K$. Depending on the sampling grid, estimating the corresponding scores α_{ik} is non-trivial, since it requires approximation of the integral in (1.24). This can be achieved easily if the functions $X_i(s)$ are available on a dense grid, however, if the functional data is sampled sparsely, the approximation is less straight forward. An alternative method that can be applied for both sparse and dense data is the *conditional expectation method* (PACE) proposed in Yao et al. (2005). Under the assumption of Gaussianity of the scores, the best linear prediction of the

scores is given by

$$\tilde{\alpha}_{ik} = \mathbb{E}(\alpha_{ik}|X_i(\cdot)) = \lambda_k \boldsymbol{\psi}'_{ik} \boldsymbol{\Sigma}_{\tilde{\mathbf{X}}_i}^{-1} (\tilde{\mathbf{X}}_i - \boldsymbol{\mu}_i), \quad (1.26)$$

where $\tilde{\mathbf{X}}_i = (X_i(s_{i1}), \dots, X_i(s_{ir_i}))'$, $\boldsymbol{\mu}_i = (\mu(s_{i1}), \dots, \mu(s_{ir_i}))'$, $\boldsymbol{\psi}_{ik} = (\psi_k(s_{i1}), \dots, \psi_k(s_{ir_i}))'$ and $\boldsymbol{\Sigma}_{\tilde{\mathbf{X}}_i} = \text{Cov}(\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_i)$. Estimates of the scores can be obtained by replacing the theoretical quantities in (1.26) by suitable estimators. Boente and Salibián-Barrera (2021) extend this result to the more general case of *elliptical distributions*.

1.4.3. Second-generation functional data

While classic FDA is concerned with the analysis of independently sampled functions, there are also more complex forms of functional data. Data can be collected as multivariate functional data or longitudinal functional data, but also in spatial and time series designs. Such samples require specific methodology to account for the more complex dependency structures in the data while at the same time keeping the computations feasible. Koner and Staicu (2023) call this type of functional data *second-generation functional data*, as opposed to the i.i.d. sampled *first-generation functional data* described in the previous subsection. We will focus on functional data from longitudinal designs in the following.

Longitudinal or repeated measurements functional data

Oftentimes, longitudinal studies do not collect data where the basic measurement is scalar or multivariate, but where it is a function instead. This procedure gives rise to what we call longitudinal functional data.

Formally, longitudinal functional data can be interpreted as realizations of bivariate stochastic processes. Opposed to the one-dimensional process in (1.19), we are concerned with random processes

$$Y : \Omega \rightarrow \mathcal{L}^2(\mathcal{S} \times \mathcal{T}), \quad (1.27)$$

where \mathcal{S} denotes the *functional* or *spatial domain* and \mathcal{T} the *time domain*. Another interpretation is to see Y as a function $Y : \mathcal{T} \rightarrow \mathcal{L}^2(\mathcal{S})$, $\mathcal{S}, \mathcal{T} \subset \mathbb{R}$, where a realization of Y at fixed time t is denoted by $Y(\cdot, t)$. Compared to the mean and covariance function in (1.20) and (1.21), the mean and covariance function of the bivariate process Y in (1.27) are more complex. They are given by

$$\begin{aligned} \mathbb{E}(Y(s, t)) &= \mu(s, t), \quad \text{and} \\ \text{Cov}(Y(s, t), Y(s', t')) &= c(\{s, t\}, \{s', t'\}) \\ &= \mathbb{E}[(Y(s, t) - \mu(s, t))(Y(s', t') - \mu(s', t'))]. \end{aligned} \quad (1.28)$$

Thus, the mean function is bivariate and the covariance function has four arguments. This underlines the increased complexity related to this type of data.

Consider observed data from n observational units.⁴ For each subject i , $i = 1, \dots, n$, we observe curves $Y_i(\cdot, t_{ij})$ in n_i random time points $t_{ij} \in \mathcal{T}$, $j = 1, \dots, n_i$. The sampling time points are assumed to be uniformly distributed on the interval \mathcal{T} and may be sparse. Since in practice the functional observations cannot be observed continuously, they are observed on a finite grid s_1, \dots, s_R . For simplicity, we assume this grid is the same for all functions. The whole dataset is then denoted by

$$\{(t_{ij}, Y_i(s_r, t_{ij})), r = 1, \dots, R\}, j = 1, \dots, n_i]_{i=1, \dots, n}.$$

The mortality data shown in Figure 1.6 can be interpreted as a longitudinal dataset. Each country corresponds to an observational unit i , and the sampling time points t_{ij} correspond to the years in which the curves were recorded. The age r corresponds to the index s_{r+1} , $r = 0, \dots, 86$. The mortality profile of country i for a fixed year t_{ij} is then given by the realization $Y_i(\cdot, t_{ij})$.

Goals of modeling and some example parameterizations

The analysis of longitudinal functional data can be focused on different aspects. First and foremost, the goal is to understand and parameterize the longitudinal dynamics of the trajectories. In the mortality data example (Figure 1.6), a question that can be asked could be: “What are the differences in development of the mortality curves over time between different European countries?”. Similarly, in the analysis of medical data, one could be interested in the differences between a treatment and a control group. Especially for processes that are observed on irregular grids, an aim is to fill in the missing time points and obtain predictions for the whole trajectories of the processes, also in time points t where data was not measured initially. Answers to those questions can be obtained by parameterizing the processes in a suitable way. In the following, we present different models that can be applied to answer these questions.

Since the processes (1.27) are bivariate, we can look at the multivariate KL expansion, given by

$$Y(s, t) = \mu(s, t) + \sum_{k=1}^{\infty} \alpha_k \gamma_k(s, t).$$

Here, $\{\gamma_k(s, t)\}_k$ is the orthonormal basis obtained from the spectral decomposition of the covariance operator induced by the covariance function (1.28) (Chen, Delicado, et al. 2017). This parameterization treats the dynamics in \mathcal{S} and \mathcal{T} symmetrically and does not allow for separate interpretation of the arguments. Instead, it requires the analysis of bivariate surfaces. However, often we are interested in a characterization that is more easily interpretable.

Greven et al. (2010) suggest to parameterize processes of form (1.27) in a functional mixed

⁴The index g in the overall data structure (1.1) is replaced by n in order to ensure consistency with the notation in Chapter 4.

effects model. In its easiest form, it is given by

$$Y_i(s, t) = \mu(s, t) + Z_{i,0}(s) + Z_{i,1}(s) \cdot t + \mathcal{E}_i(s, t),$$

where $Z_{i,0}(s)$ and $Z_{i,1}(s)$ are uncorrelated random processes and $\mathcal{E}_i(s, t)$ contains smooth and white noise error terms. In this model, the processes $Y_i(s, t)$ are assumed to evolve linearly in t , thus reducing the variation to univariate processes $Z_{i,j}(s)$, $i = 1, \dots, n$, $j = 0, 1$. The model estimation is carried out using what they call *longitudinal FPCA*. Each component $Z_{i,p}(s)$, $i = 1, \dots, n$, $p = 0, 1$, is represented by its own KL expansion, while allowing for correlations between the two processes. The model can also be generalized to allow for more complicated structures and additional linear covariates.

Chen and Müller (2012) offer another approach to obtain a more parsimonious and interpretable parameterization. They represent the processes $Y_i(s, t)$ as $Y_i(s|t)$, thus conditional on the measurement time. A model is set up using a *double FPCA method*. This yields the following representation of the processes

$$\begin{aligned} Y_i(s|t) &= \mu(s|t) + \sum_{k=1}^{\infty} \xi_{ik}(t) \phi_k(s|t) \\ &= \mu(s|t) + \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} \zeta_{ikp} \psi_{kp}(t) \phi_k(s|t) \\ &= \mu(s|t) + \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} \zeta_{ikp} \varphi_k(s|t), \end{aligned}$$

where $\varphi_k(s|t) = \psi_{kp}(t) \phi_k(s|t)$ and $i = 1, \dots, n$. This decomposition requires a separate set of eigenfunctions at each measurement time point t and is thus computationally quite demanding. The surfaces $\phi_k(s|t)$ are assumed to be smooth in both s and t .

An even more parsimonious parameterization is offered by Chen, Delicado, et al. (2017) and Park and Staicu (2015). Instead of estimating eigensurfaces as functions s and t , they separate the dynamics s and t in the decomposition of the process, yielding the representation

$$\begin{aligned} Y_i(s, t) &= \mu(s, t) + \sum_{k=1}^{\infty} \xi_{ik}(t) \phi_k(s) \\ &= \mu(s, t) + \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \zeta_{ik\ell} \psi_{k\ell}(t) \phi_k(s), \end{aligned} \tag{1.29}$$

where the dynamics in s and t are separated, $i = 1, \dots, n$. The representation (1.29) is purely data-driven and can be fit to datasets that are sampled on both sparse and dense grids in \mathcal{S} and \mathcal{T} . This model will be referred to as *marginal FPCA* in the following, and is topic of Chapter 4 of this thesis, where we propose a robust algorithm to estimate above decomposition.

1.5. Overview of the three main chapters

In the following, we give a short description of the content and novelty of the three main chapters of the thesis.

1.5.1. Chapter 2: Time-Varying Reduced Rank Regression

The chapter is based on the article B. Brune, W. Scherrer, et al. (2022). “A State-Space Approach to Time-Varying Reduced-Rank Regression”. In: *Econometric Reviews* 41.8, pp. 895–917. DOI: 10.1080/07474938.2022.2073743.

In the paper, a new approach to reduced-rank regression that allows for time-variation in the regression coefficients is introduced. Reduced-rank regression is a straight-forward way to reduce the number of parameters in multivariate regression models in an interpretable manner: The multivariate time series can be explained by a lower number of linear combinations of the predictor series. Reduced-rank regression contains useful model classes for multivariate time series as special cases, e.g. vector error correction models and special cases of (dynamic) factor models. We propose an algorithm for the regression model $\mathbf{y}_t = \mathbf{C}_t \mathbf{x}_t + \boldsymbol{\varepsilon}_t$, where the time-varying parameter matrices $\mathbf{C}_t \in \mathbb{R}^{p \times q}$ are assumed to decompose into two lower-rank matrices parameterized as either (A) $\mathbf{C}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}'$ or (B) $\mathbf{C}_t = \boldsymbol{\alpha} \boldsymbol{\beta}'_t$. Thus, time-variation is either in the “reductions” of the predictor series $\boldsymbol{\beta}'_t \mathbf{x}_{t-1}$, or in the way the reduced series $\boldsymbol{\beta}' \mathbf{x}_{t-1}$ are mapped onto the observed time series \mathbf{x}_t through the coefficient matrices $\boldsymbol{\alpha}_t$. The model is set up as a Gaussian state-space system where the time-varying parameter matrices are assumed to follow a random walk, and their covariances are specified in a way that incorporates the fact that the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are only identifiable up to rotations. In doing so, we resolve potential contradictions in the rotation-invariance of the imposed covariance structure for the time varying parameters pointed out in Yang and Bauwens (2018). The assumption of Gaussianity allows to use the well-established Kalman filter to estimate the time-varying parameter matrices. The unknown parameters of the system are estimated using an EM algorithm with closed-form updates. This closes a gap in the literature, since this is, to the best of our knowledge, the first method for time-varying reduced-rank regression that does not require knowledge of the unknown parameters of the state-space system, and can thus be fit to real datasets. We illustrate the proposed method with simulation studies and two real-data applications.

Contribution: B. Brune participated in discussions with the co-authors to come up with the initial ideas, came up with the EM-algorithm, implemented the methodology and carried out the simulation studies, and wrote the initial draft of the manuscript.

1.5.2. Chapter 3: Rank-Based Estimation of Mixed Effects Models

The chapter is based on the article B. Brune, I. Ortner, and P. Filzmoser (2024). *A Rank-Based Estimation Method for Mixed Effects Models in the Presence of Outlying Data*.

Conditionally accepted for publication in: *Journal of Data Science, Statistics, and Visualization*.

In the paper, we present a novel approach to the rank-based estimation of mixed effects models. While the existing methodology is limited to models with random intercepts, we propose an algorithm that furthermore allows to estimate random effects structures that contain random slope coefficients. Outlying observations can occur both in the response variables y_{ij} (response outliers) and the vectors of predictors \mathbf{x}_{ij} and \mathbf{z}_{ij} (leverage points). The proposed algorithm offers robustness against both types of outliers. This is achieved by adapting the GR procedure proposed in Bilgic (2012), and introducing a weighting procedure that limits the influence of the outliers. The proposed algorithm leverages the advantages of rank-based estimation in that it is (almost) tuning-parameter free and computationally very efficient. We propose different diagnostic tools and plots which allow for model diagnostics and especially help with identification of outlying observations or groups in the data. We study the performance of our estimators in extensive simulation studies, and the methodology is illustrated in three data applications.

Contribution: B. Brune came up with the initial idea, developed the methodology and algorithm, implemented the method and carried out the simulation studies, and wrote the initial draft of the manuscript.

1.5.3. Chapter 4: Robust modeling of longitudinal functional data

This chapter is based on joint work with Una Radojčić, Sonja Greven and Peter Filzmoser.

Longitudinal or repeated measurements functional data add a layer of complexity to standard i.i.d. functional data by introducing dependencies not only in terms of the functional argument s , but also between the realizations of the functions in different time points t . A common method to reduce the dimension of functional datasets and to understand the driving sources of variation is to apply functional principal component analysis (FPCA). Different parameterizations of FPCA models for functional data with longitudinal structure have been introduced in the literature. However, there is a lack of robust methodology to fit such models to observed datasets, although in practice, the observed data can be contaminated by erroneous observations or outliers. This paper proposes a robust estimation algorithm for the marginal FPCA model (1.29). It replaces the non-robust components of the algorithms proposed in Park and Staicu (2015) and Chen, Delicado, et al. (2017) by robust counterparts. Furthermore, we develop methodology to detect outlying observations on two levels: as single curves within an observational unit, and whole outlying observational units that do not follow the structure imposed by the model. The simulation studies presented in the paper show that the algorithm has excellent robustness properties and reliably detects outlying curves. Furthermore, it leads to interesting insights when applied to real datasets.

Contribution: B. Brune came up with the initial idea, developed the methodology and algorithm, implemented the method and carried out the simulation studies, and wrote the initial draft of the paper. The proofs were provided by U. Radojčić.

1.6. Related software packages

Implementations of the methodology for the three articles are available as R-Code from B. Brune's GitHub account (<https://github.com/b-brune>). The corresponding repositories are listed in the following:

tvRRR implements the methodology for time-varying reduced rank regression models as proposed in Chapter 2.

rankLME implements the rank-based mixed effects model from Chapter 3.

robLFDA contains code to fit the robust model for longitudinal functional data proposed in Chapter 4.

2 | A State-Space Approach to Time-Varying Reduced-Rank Regression

This chapter is joint work with Wolfgang Scherrer and Efstathia Bura and was published as B. Brune, W. Scherrer, et al. (2022). “A State-Space Approach to Time-Varying Reduced-Rank Regression”. In: *Econometric Reviews* 41.8, pp. 895–917. DOI: 10.1080/07474938.2022.2073743. The supplemental material can be found in Appendix A, p. 117ff.

2.1. Introduction

The simultaneous modeling and fitting of p time series when p is large is a challenging task as the number of parameters grows with p^2 and can often exceed the number of observations. Five main approaches to reduce the number of parameters in multivariate time series models have been proposed: (1) reduced-rank vector autoregressive (VAR) and multivariate autoregressive index models (MAI) models (Velu et al. 1986; Ahn and Reinsel 1988; Carriero et al. 2016), where the parameter matrix is assumed to have a lower rank structure; (2) shrinkage Bayesian VAR with shrinking prior distributions for controlling the number of parameters (Litterman 1986; Sims and Zha 1998; Ghosh et al. 2019); (3) sparse VAR estimation, where a low-dimensional analysis of groups of series is first carried out to identify rows and columns of parameter matrices that can be assumed to have zero values, as in Davis et al. (2016); (4) regularized estimation, where a penalty term is included in the estimation objective function (see, e.g., Nicholson et al. (2017) and Song and Bickel (2011)); and (5) the popular dynamic factor model (DFM), introduced by Geweke (1977) and Sargent and Sims (1977). Using the idea of risk diversification discussed in Chamberlain (1983), the so called generalized dynamic factor models (GDFM) have been introduced, extensively studied and developed by many researchers [see, e.g., Stock and Watson (2002), Stock and Watson (2005), Bai and Ng (2002), Forni, Hallin, Lippi, and Reichlin (2000), Forni, Hallin, Lippi, and Reichlin (2005), Forni, Hallin, Lippi, and Zaffaroni (2015), Peña and Yohai (2016), and Anderson and Deistler (2008)]. In GDFMs, the number of parameters typically grows linearly with the number of time series. A recent review of these five approaches can be found in Peña and Tsay (2020).

Classical linear time series models assume that the (auto-) covariance structure and hence

the serial (linear) dependency structure of the data is constant. This may not be a realistic assumption in many applications. For example, many economic data sets show significant structural changes.

Vector-autoregressive (VAR) models with time-varying parameters (TVP) form an important line of research for TVP time series models. They have been successfully applied to model time-varying relationships between multiple economic time series (e.g. in Lubik and Matthes 2016). Stock and Watson (2009) and Breitung and Eickmeier (2011) found empirical evidence of temporal instability of coefficients in dynamic factor models for large panels of economic data. DFMs that are able to deal with time-dependent parameters have been suggested in Del Negro and Otrok (2008) and Su and Wang (2017). Time-variation may also arise in the cointegrating parameter matrix of a vector error-correction model [see, e.g., Bierens and Martins (2010)].

Time-varying parameter is a special case of stochastic parameter models. An indication of the extent of interest in such models is the fact that reviews of related literature appeared as early as 1973 in Rosenberg (1973) followed by Johnson (1977). The estimation of TVP models for time series has been approached from different perspectives. Frequentist methodology includes state-space modeling [see e.g. Chapter 18 in Lütkepohl (2005), Eickmeier et al. (2015) and Chapter 13 of Hamilton (1994)] and non-parametric approaches [see e.g. Bierens and Martins (2010), Su and Wang (2017) or Coulombe (2021)]. Another research direction follows Bayesian estimation, often also in combination with state-space modeling [see, among many others, Doan et al. (1984), Primiceri (2005), Del Negro and Otrok (2008), Lubik and Matthes (2016) and Baştürk et al. (2017)].

Since its introduction by Izenman (1975), reduced-rank regression (RRR) has been successfully applied in multivariate regression modeling, mostly to independent and identically distributed (i.i.d.) data. In this paper, we model a vector of series jointly using reduced-rank regression with dynamics in the parameter matrix and apply classical tools from state-space models. The time-varying parameters are considered as latent states in a Gaussian state-space system. This modeling approach has been used successfully to fit multivariate time series models with time-varying coefficients [see, e.g., Hamilton (1994, Chapter 13); Durbin and Koopman (2012, Section 6.3.1); Guo (1990); Kilian and Lütkepohl (2017, Chapter 18); Lütkepohl (2005, Chapter 18)].

We model different types of temporal instability by decomposing the reduced-rank coefficient matrix in the multivariate time series regression model as the product of two full-rank time-varying matrices. To the best of our knowledge, the only paper presenting a similar methodology is Yang and Bauwens (2018), who assume the time-varying coefficient matrices live on the Stiefel manifold, the space of semi-orthogonal matrices (see e.g. Chikuse 2003). Yang and Bauwens (2018)'s approach does not lead to an estimate of all model parameters. In a different context, state-space models with states on the Stiefel manifold were considered by Tompkins and Wolfe (2007) and Bordin and Bruno (2019), where estimation was again not resolved. This restriction appears to impede the development of a feasible estimation algorithm for the parameters, and in consequence, prediction of the responses.

In our Gaussian state-space formulation, the time-varying coefficient matrices live in a

Euclidean space instead, which allows for evaluation of the likelihood via the Kalman filter (Shumway and Stoffer 2017, Ch. 6). Maximum likelihood estimation of all parameters is carried out using an EM-algorithm with closed form updates. We refer to this model, together with the corresponding estimation algorithm, as time-varying reduced-rank regression (TVRRR).

The rest of this paper is organized as follows. We review time-constant and time-varying reduced-rank regression in Section 2.2. We specify our models in Section 2.2.1 and present our estimation and model fitting algorithm in Section 2.2.2. In Section 2.2.3, we define the BIC criterion we use to estimate the rank of the parameter matrix. Section 2.3 contains our simulation studies and results, which confirm the usefulness of TVRRR in accurately modeling time series data whose relationship changes over time. There, we also show it is particularly robust to structural breaks in the coefficients. In Section 2.4, we showcase the adaptability of our model by analyzing two data sets. The first consists of seven financial indices, which are considered during two time periods, when international markets experienced extreme stress: the Great Recession in 2008 and the Covid-19 crisis of 2020. Here, we fit vector error correction models (VECM) with time-varying cointegrating relations. As a second example, we consider the number of weekly Covid-19 cases in 12 European countries and fit a TVP-VAR(1) with a reduced-rank coefficient matrix. Section 2.5 contains our concluding remarks.

2.2. Time-varying reduced-rank regression

The general reduced-rank regression (RRR) model is given by

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{\Gamma}\mathbf{u}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \dots, \quad (2.1)$$

where $\mathbf{y}_t \in \mathbb{R}^p$ is a p -dimensional target vector, $\mathbf{x}_t \in \mathbb{R}^q$ is a q -dimensional predictor, and $\mathbf{C} \in \mathbb{R}^{p \times q}$ is a matrix with $\text{rank}(\mathbf{C}) = d \leq \min(p, q)$. The latter implies that \mathbf{C} can be decomposed as the product of two rank d matrices $\boldsymbol{\alpha} \in \mathbb{R}^{p \times d}$ and $\boldsymbol{\beta} \in \mathbb{R}^{q \times d}$ such that $\mathbf{C} = \boldsymbol{\alpha}\boldsymbol{\beta}'$. The errors $\boldsymbol{\varepsilon}_t$ are normal white noise with covariance $\boldsymbol{\Omega}$. The vector $\mathbf{u}_t \in \mathbb{R}^k$ contains additional explanatory variables with unrestricted coefficient matrix $\mathbf{\Gamma} \in \mathbb{R}^{p \times k}$. Both \mathbf{x}_t and \mathbf{u}_t are observable. Throughout, we use boldfaced letters and symbols to denote vectors or matrices and non-boldfaced for scalars. \mathbf{A}' denotes the transpose of a matrix \mathbf{A} and \mathbf{I}_s stands for the $(s \times s)$ identity matrix.

Velu et al. (1986) were the first to use RRR in vector autoregressive (VAR) models. In a VAR(m) model, the explanatory variables \mathbf{x}_t are lags of the target variable \mathbf{y}_t , $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-m})' \in \mathbb{R}^{mp}$. While fitting full rank VAR models requires estimation of a large number of parameters if the number of time series p is large, RRR introduces sparsity in the relationship between the lagged variables and the response by assuming that the coefficient matrix $\mathbf{C} \in \mathbb{R}^{p \times mp}$ has rank $d \ll p$. Cointegration models are another popular application of RRR in the time series context. In the vector error correction model (VECM), the cointegrating relations of multiple non-stationary time series are represented by a

reduced-rank matrix [see e.g. Lütkepohl (2005), Part II].

RRR may also be simply used as a tool for constructing approximate yet parsimonious regression models. Reinsel and Velu (1998) provide a comprehensive review of reduced-rank regression models and their applications in both i.i.d. and time-series settings. The principle of RRR underlies several methods in sufficient dimension reduction based on inverse regression (Bura and Cook 2003). For example, parametric inverse regression (Bura and Cook 2001) and principal fitted components (Cook and Forzani 2008) rely on RRR methodology.

The most relevant interpretation to our development is of RRR as a *factor model*. Specifically, $\mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\varepsilon}_t$ can be interpreted as a factor model with factors $\mathbf{f}_t = \boldsymbol{\beta}'\mathbf{x}_t$ and factor loadings $\boldsymbol{\alpha}$, with the distinguishing feature that the factors, for given $\boldsymbol{\beta}$, are observable, as opposed to latent in “standard” factor models. The factors summarise all the information in \mathbf{x}_t that is necessary for approximating \mathbf{y}_t . If $\mathbf{x}_t = \mathbf{y}_{t-1}$, the model can be interpreted as a simple dynamic factor model, where the factors \mathbf{f}_t have a VAR(1) dynamic (Peña and Poncela 2006).

We incorporate time variation in the coefficients of model (2.1) by writing,

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \boldsymbol{\Gamma}\mathbf{u}_t + \boldsymbol{\varepsilon}_t. \quad (2.2)$$

We assume that the rank $d < \min(p, q)$ of \mathbf{C}_t is fixed and we model the time dependence of \mathbf{C}_t as

$$(A) \quad \mathbf{C}_t = \boldsymbol{\alpha}_t\boldsymbol{\beta}', \text{ where } \boldsymbol{\alpha}_t \in \mathbb{R}^{p \times d}, \boldsymbol{\beta} \in \mathbb{R}^{q \times d} \quad (2.3)$$

$$(B) \quad \mathbf{C}_t = \boldsymbol{\alpha}\boldsymbol{\beta}_t', \text{ where } \boldsymbol{\alpha} \in \mathbb{R}^{p \times d}, \boldsymbol{\beta}_t \in \mathbb{R}^{q \times d} \quad (2.4)$$

$$(C) \quad \mathbf{C}_t = \boldsymbol{\alpha}_t\boldsymbol{\beta}_t', \text{ where } \boldsymbol{\alpha}_t \in \mathbb{R}^{p \times d}, \boldsymbol{\beta}_t \in \mathbb{R}^{q \times d}. \quad (2.5)$$

In all three, the component matrices in the decomposition are of full rank d . In (A) and (B), one of the two factors in the decomposition of \mathbf{C}_t is assumed to be constant. Hence, the time-variation is attributed to the other. This simplifies the interpretation and estimation of the resulting model.

In model (A), it is assumed that the relevant information contained in \mathbf{x}_t may be extracted by non time-varying linear combinations of the regressors \mathbf{x}_t . However, the coefficients of the regression of \mathbf{y}_t on $\boldsymbol{\beta}'\mathbf{x}_t$ may vary with time. In this case the row space of \mathbf{C}_t , i.e. the row space of $\boldsymbol{\beta}'$, is independent of time. The case of the column space of \mathbf{C}_t being constant is modeled by (B) with an analogous interpretation. In model (C), both the column and the row space of \mathbf{C}_t are allowed to be time dependent.

Both matrices in the decompositions (A), (B) and (C) can only be identified up to invertible transformations. For model (C), the relation $\boldsymbol{\alpha}_t\boldsymbol{\beta}_t' = \boldsymbol{\alpha}_t\mathbf{H}_t\mathbf{H}_t^{-1}\boldsymbol{\beta}_t'$ holds for any invertible $\mathbf{H}_t \in \mathbb{R}^{d \times d}$, so that each pair $\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t$ can only be identified up to right multiplication with a non-singular matrix.

In models (A) and (B), the transformation $\mathbf{H}_t = \mathbf{H}$ has to be independent of time. For

example, in model (A),

$$\mathbf{C}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}' = \tilde{\boldsymbol{\alpha}}_t \tilde{\boldsymbol{\beta}}' \iff \tilde{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}_t \mathbf{H} \text{ and } \tilde{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_t (\mathbf{H}^{-1})'. \quad (2.6)$$

Ensuring identifiability requires additional restrictions, such as fixing a $d \times d$ dimensional sub-matrix of $\boldsymbol{\beta}$ to the identity matrix. We demonstrate at the end of this section that the estimates of the proposed algorithm are invariant to transformations as in (2.6).

In addition to the ambiguity of the component matrices introduced by the time-specific transformations \mathbf{H}_t , the observations \mathbf{y}_t in model (C) are a non-linear function of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}'_t$, rendering model fitting and parameter estimation non-trivial. This model cannot be represented as a linear state-space system and one would need to resort to non-linear filtering methods, such as the extended Kalman filter or particle filters, for estimation of the time-varying parameters $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ (see e.g. Särkkä 2013). On the other hand, models (A) and (B) are sufficiently flexible to capture various types of time-variation. A possible interpretation of model (A) could be a factor model with time-varying factor loadings: The factor composition remains constant, even though the factors' influence on different response variables can vary over time. With model (B), we can, among other applications, model time-varying cointegration spaces in vector error correction models. Moreover, time-varying factor compositions are covered by this formulation. We focus on models (A) and (B) in the rest of the paper.

2.2.1. Model specification

The reduced-rank regression model (2.2) with time-varying parameters \mathbf{C}_t as defined in (2.3) and (2.4), can be written as

$$\begin{aligned} \text{(A)} \quad y_t &= \boldsymbol{\alpha}_t \boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\Gamma} \mathbf{u}_t + \boldsymbol{\varepsilon}_t \\ &= (\mathbf{x}'_t \boldsymbol{\beta} \otimes \mathbf{I}_p) \text{vec}(\boldsymbol{\alpha}_t) + \boldsymbol{\Gamma} \mathbf{u}_t + \boldsymbol{\varepsilon}_t, \\ \text{(B)} \quad y_t &= \boldsymbol{\alpha} \boldsymbol{\beta}'_t \mathbf{x}_t + \boldsymbol{\Gamma} \mathbf{u}_t + \boldsymbol{\varepsilon}_t \\ &= (\mathbf{x}'_t \otimes \boldsymbol{\alpha}) \text{vec}(\boldsymbol{\beta}_t) + \boldsymbol{\Gamma} \mathbf{u}_t + \boldsymbol{\varepsilon}_t, \end{aligned}$$

where the errors are i.i.d. with $\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Omega}$. Both are regression models, where some of the coefficients are time-dependent. These time-varying parameters are modeled as (latent) states in a state space system, as in, e.g., Hamilton (1994).

Both cases (A) and (B) can be treated similarly. We therefore outline the estimation procedure in detail only for model (A). The derivation for model (B) is analogous. In order to simplify the exposition, we let $\boldsymbol{\Gamma} = \mathbf{0}$. However, the inclusion of this term is straightforward.

In absence of specific prior knowledge of the dynamics of the time-varying parameters, we assume that they follow a random walk. The time-varying parameters, i.e. states, are

assumed to evolve according to some density f that is centered at the previous state,

$$\text{vec}(\boldsymbol{\alpha}_{t+1}) \sim f(\text{vec}(\boldsymbol{\alpha}_t)), \quad (2.7)$$

as in, for example, Del Negro and Otrok (2008), Eickmeier et al. (2015) and Yang and Bauwens (2018).

A common approach is to let f be a multivariate normal density,

$$\text{vec}(\boldsymbol{\alpha}_{t+1}) \sim \mathcal{N}(\text{vec}(\boldsymbol{\alpha}_t), \boldsymbol{\Sigma}). \quad (2.8)$$

The number of parameters $pd(pd+1)/2$ in the state covariance matrix $\boldsymbol{\Sigma}$ is prohibitive for estimation, even for moderate sized models. This fact necessitates the imposition of further structure on $\boldsymbol{\Sigma}$. As has been noted by Yang and Bauwens (2018), this structure should be compatible with the inherent non-identifiability of the factorisation (2.6). For example, the widely used simplification of a diagonal $\boldsymbol{\Sigma}$ is not preserved under invertible transformations, since the transformed states $\boldsymbol{\alpha}_t \mathbf{H}$ evolve as $\text{vec}(\boldsymbol{\alpha}_{t+1} \mathbf{H}) \sim \mathcal{N}(\text{vec}(\boldsymbol{\alpha}_t \mathbf{H}), (\mathbf{H}' \otimes I) \boldsymbol{\Sigma} (\mathbf{H} \otimes I))$. In this case, the covariance matrix is *non-diagonal* in general.

To bypass this limitation and facilitate identifiability of the parameter matrices, Yang and Bauwens (2018) propose a non-linear state space model on the set of all semi-orthogonal matrices, the Stiefel manifold

$$\mathbb{V}_{p,d} = \{\mathbf{A} \in \mathbb{R}^{p \times d} : \mathbf{A}' \mathbf{A} = \mathbf{I}_d\}. \quad (2.9)$$

In their specification, the state matrices $\boldsymbol{\alpha}_t$ perform a random walk on the space (2.9) by requiring f in (2.7) to be the density of a matrix Langevin (or von Mises-Fisher) distribution. This distribution derives from restricting a specific matrix normal distribution to orthogonal matrices (see e.g. Chikuse 2003, Ch. 2). Although Yang and Bauwens (2018) derive closed form solutions for the predicted and updated states, the resulting posterior likelihoods involve intractable normalizing constants that depend on the model's time-constant parameters. This renders parameter estimation hard, as reflected in the absence of an estimation procedure. The model cannot be fitted to observed data and filtering the states is only possible if the unknown parameters are given a priori.

We do not require the time-varying parameter matrices $\boldsymbol{\alpha}_t$ be semi-orthogonal, but allow them to vary freely in $\mathbb{R}^{p \times d}$, which results in more flexible parameter estimates. We maintain the multivariate normality assumption (2.8) and further assume that the covariance matrix $\boldsymbol{\Sigma}$ is separable; i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ with $\boldsymbol{\Sigma}_c \in \mathbb{R}^{d \times d}$, $\boldsymbol{\Sigma}_r \in \mathbb{R}^{p \times p}$. This Kronecker product structure is invariant to transformations $\boldsymbol{\alpha}_t \rightarrow \boldsymbol{\alpha}_t \mathbf{H}$, since

$$\text{Cov}(\text{vec}(\boldsymbol{\alpha}_t \mathbf{H})) = \text{Cov}((\mathbf{H}' \otimes \mathbf{I}_p) \text{vec}(\boldsymbol{\alpha}_t)) = (\mathbf{H}' \boldsymbol{\Sigma}_c \mathbf{H}) \otimes \boldsymbol{\Sigma}_r.$$

Furthermore, this structure reduces the number of parameters from $pd(pd+1)/2$ to at most $(p(p+1) + d(d+1))/2 - 1$. The decomposition $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ corresponds to the assumption that the states follow a $(p \times d)$ -dimensional matrix normal distribution (Gupta and Nagar

2000),

$$\boldsymbol{\alpha}_t \sim \mathcal{N}_{p \times d}(\boldsymbol{\alpha}_{t-1}; \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c),$$

which agrees with the interpretation of the time-varying coefficients as rank d -matrices.

A further reduction of the number of parameters that need to be estimated is achieved by setting $\boldsymbol{\Sigma}_r = \mathbf{I}_p$, which further relieves the computational burden for large datasets. Due to this assumption the (changes of the) rows of $\boldsymbol{\alpha}_t$ are independent and identically distributed, so that we only need to estimate $d(d+1)/2$ parameters, a computationally non-prohibitive number when the rank of the matrix is small. The same applies to model (B). The limitation to this form of matrix normal distribution complies with the matrix Langevin distribution used in Yang and Bauwens (2018), as the latter derives from a matrix normal distribution with the above covariance structure (Chikuse 2003).

The resulting state-space model for model (A) is thus given by

$$\begin{aligned} \text{vec}(\boldsymbol{\alpha}_{t+1}) &= \text{vec}(\boldsymbol{\alpha}_t) + \boldsymbol{\eta}_t \\ \mathbf{y}_t &= (\mathbf{x}'_t \boldsymbol{\beta} \otimes \mathbf{I}_p) \text{vec}(\boldsymbol{\alpha}_t) + \boldsymbol{\Gamma} u_t + \boldsymbol{\varepsilon}_t, \end{aligned} \quad (2.10)$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ and $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_c \otimes \mathbf{I}_p)$ are independent i.i.d. sequences. Analogously, model (B) is of the form

$$\begin{aligned} \text{vec}(\boldsymbol{\beta}'_{t+1}) &= \text{vec}(\boldsymbol{\beta}'_t) + \boldsymbol{\zeta}_t \\ \mathbf{y}_t &= (\mathbf{x}'_t \otimes \boldsymbol{\alpha}) \text{vec}(\boldsymbol{\beta}'_t) + \boldsymbol{\Gamma} u_t + \boldsymbol{\varepsilon}_t \end{aligned} \quad (2.11)$$

with $\boldsymbol{\zeta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q \otimes \boldsymbol{\Sigma}_c)$.

2.2.2. State estimation and model fitting

The regressors $\mathbf{x}_t, \mathbf{u}_t, t = 1, \dots, T$, are understood as functions of explanatory variables \mathbf{z}_t and lagged dependent variables. That is,

$$(\mathbf{x}'_t, \mathbf{u}'_t)' = g(\mathbf{z}_t, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-m}).$$

We let \mathcal{F}_s denote the sigma-algebra generated by all the exogenous variables \mathbf{z}_t and the observed target variables up to time s , $\mathcal{F}_s = \sigma(\mathbf{z}_1, \dots, \mathbf{z}_T, \mathbf{y}_{-m+1}, \dots, \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_s)$. This setting includes as special cases VAR(m), VARX(m) and VECM(m) models, in which the predictors are either functions of the lagged variables, or of the exogenous variables, or both.

The time-varying parameters $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ in models (2.10), resp. (2.11), can be inferred using the Kalman filter. Again, we outline the procedure in detail only for model (A). Following Shumway and Stoffer (2017, Ch. 6), we use the following notation for the filtered and

smoothed coefficient matrices and the corresponding covariances:

$$\begin{aligned}\boldsymbol{\alpha}_t^s &= \mathbb{E}(\boldsymbol{\alpha}_t | \mathcal{F}_s) \\ \mathbf{P}_t^s &= \text{Cov}(\text{vec}(\boldsymbol{\alpha}_t) | \mathcal{F}_s) \\ \mathbf{P}_{t,t-1}^s &= \text{Cov}(\text{vec}(\boldsymbol{\alpha}_t), \text{vec}(\boldsymbol{\alpha}_{t-1}) | \mathcal{F}_s).\end{aligned}\tag{2.12}$$

In addition, we assume the noise vectors are jointly normally distributed with

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_c \otimes \mathbf{I}_p \end{pmatrix} \right)$$

and independent of \mathcal{F}_0 and the initial state $\boldsymbol{\alpha}_0$. Then, the conditional expectations and covariances in (2.12) can be computed by the usual Kalman filtering and smoothing recursions (Shumway and Stoffer 2017, Ch. 6). For the initial state, we assume $\text{vec}(\boldsymbol{\alpha}_0) | \mathcal{F}_0 \sim \mathcal{N}(\text{vec}(\mathbf{a}_0), \boldsymbol{\Psi}_0)$,¹ providing the starting values $\boldsymbol{\alpha}_0^0 = \mathbf{a}_0$ and $\mathbf{P}_0^0 = \boldsymbol{\Psi}_0$ for the Kalman filter and smoother.

The estimated states and the corresponding covariance matrices in (2.12) depend on the initial values $\mathbf{a}_0, \boldsymbol{\Psi}_0$ and the unknown parameters in system (2.10). We denote this set by²

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}_A = \{\mathbf{a}_0, \boldsymbol{\Psi}_0, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}_c, \boldsymbol{\Gamma}\}.\tag{2.13}$$

The negative log-likelihood of $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$ conditional on \mathcal{F}_0 is (up to constants) given by

$$\begin{aligned}-2 \log(L_{\mathbf{Y}}(\boldsymbol{\Theta})) &= \sum_{t=1}^T \log |\boldsymbol{\Sigma}_t| + \\ &\sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\alpha}_t^{t-1} \boldsymbol{\beta}' \mathbf{x}_t - \boldsymbol{\Gamma} \mathbf{u}_t) \boldsymbol{\Sigma}_t^{-1} (\mathbf{y}_t - \boldsymbol{\alpha}_t^{t-1} \boldsymbol{\beta}' \mathbf{x}_t - \boldsymbol{\Gamma} \mathbf{u}_t).\end{aligned}\tag{2.14}$$

where $\boldsymbol{\Sigma}_t = (\mathbf{x}_t' \boldsymbol{\beta} \otimes \mathbf{I}_p) \mathbf{P}_t^{t-1} (\mathbf{x}_t' \boldsymbol{\beta} \otimes \mathbf{I}_p) + \boldsymbol{\Omega}$ (Gibson and Ninness 2005; Shumway and Stoffer 2017). The log-likelihood (2.14) is a highly non-linear function of the parameters (2.13). We bypass its direct minimization using the expectation-maximization (EM) algorithm. Optimization is based on the (conditional) joint likelihood of the data \mathbf{Y} and the states $\mathbf{A} = (\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_T)$, denoted by $L_{\mathbf{Y}, \mathbf{A}}(\boldsymbol{\Theta})$.

Given a set of initial values $\boldsymbol{\Theta}^{(0)}$ for the model parameters (2.13), the EM-algorithm consists of two steps that are iterated in turn until certain stopping criteria are met.

E-Step: Evaluate the conditional expectation of the joint log-likelihood given \mathcal{F}_T as a

¹Analogously, we let $\text{vec}(\boldsymbol{\beta}_0') | \mathcal{F}_0 \sim \mathcal{N}(\text{vec}(\mathbf{b}_0'), \boldsymbol{\Psi}_0)$ for model (B).

²For model (B), $\boldsymbol{\Theta}_B = \{\mathbf{b}_0, \boldsymbol{\Psi}_0, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}_c, \boldsymbol{\Gamma}\}$

function of the parameters Θ , while assuming $\Theta^{(j)}$ are the “true” parameters:

$$Q(\Theta|\Theta^{(j)}) = \mathbb{E}_{\Theta^{(j)}} \left(-2 \ln(L_{Y, \mathbf{A}}(\Theta)) \middle| \mathcal{F}_T \right). \quad (2.15)$$

M-Step: Minimize the expected negative log-likelihood in Θ , i.e. find

$$\Theta^{(j+1)} = \arg \min_{\Theta} Q(\Theta|\Theta^{(j)}).$$

The E-step includes smoothing of the latent states given the current set of parameters and starting values. The expected log-likelihood (2.15) for model (A) is given by

$$\begin{aligned} Q(\Theta|\Theta^{(j)}) = & \ln |\Psi_0| + \text{tr} (\Psi_0^{-1} [\mathbf{P}_0^T + \text{vec}(\alpha_0^T - \mathbf{a}_0) \text{vec}(\alpha_0^T - \mathbf{a}_0)']) + \\ & Tp \ln |\Sigma_c| + \text{tr} \left((\Sigma_c \otimes \mathbf{I}_p)^{-1} \sum_{t=1}^T \left(\text{vec}(\alpha_t^T) \text{vec}(\alpha_t^T)' + \mathbf{P}_t^T + \right. \right. \\ & \left. \left. \text{vec}(\alpha_{t-1}^T) \text{vec}(\alpha_{t-1}^T)' + \mathbf{P}_{t-1}^T - (\text{vec}(\alpha_t^T) \text{vec}(\alpha_{t-1}^T)' + \mathbf{P}_{t,t-1}^T) - \right. \right. \\ & \left. \left. (\text{vec}(\alpha_{t-1}^T) \text{vec}(\alpha_t^T)' + \mathbf{P}_{t-1,t}^T) \right) \right) + T \ln |\Omega| + \\ & \text{tr} \left(\Omega^{-1} \sum_{t=1}^T (\mathbf{y}_t \mathbf{y}_t' - \mathbf{y}_t \mathbf{x}_t' \beta \alpha_t^{T'} - \alpha_t^T \beta' \mathbf{x}_t \mathbf{y}_t' + \right. \\ & \left. (\mathbf{x}_t' \beta \otimes \mathbf{I}_p) (\text{vec}(\alpha_t^T) \text{vec}(\alpha_t^T)' + \mathbf{P}_t^T) (\mathbf{x}_t' \beta \otimes \mathbf{I}_p)') \right). \end{aligned} \quad (2.16)$$

The smoothed states α_t^T and the covariance matrices \mathbf{P}_t^T and $\mathbf{P}_{t,t-1}^T$ in (2.16) depend on the current set of parameters $\Theta^{(j)}$. However, for better readability, this dependence is not reflected in the notation.

For the M-step we use a simple iterative scheme, where we, in turn, compute the minimizer of the expected log-likelihood with respect to one of the parameters from (2.13) while keeping the others fixed. The minimizers for \mathbf{a}_0 , Ψ_0 and Σ_c are independent of the other parameters. Inspection of (2.16) reveals that, for model (A), the updates for β , Ω and, if included, Γ , depend on each other. Nevertheless, it is straightforward to compute the minimizer for one of these parameters while keeping the other two fixed. The explicit formulas for the above minimizers are given in Table 2.1. The updates within each M-step are iterated until the changes of the parameter matrices are negligible, which is achieved fast in our simulations. The EM-algorithm terminates if either the change in the parameter estimates, or the relative change in the data log-likelihood (2.14) is negligible.

2. A State-Space Approach to Time-Varying Reduced-Rank Regression

Table 2.1.: Parameter updates for models (A) and (B), given all other parameters and the smoothed states and covariances. For better readability the notation does not show dependence on the current parameter estimates $\Theta^{(j)}$. In the update for β' , the subscript $i : j, k : l$ is used to refer to a submatrix with rows i to j and columns k to l . unvec denotes the inverse vec operation, i.e. $a = \text{vec}(\mathbf{A}) \Leftrightarrow \mathbf{A} = \text{unvec}(a)$.

(A)	
\mathbf{a}_0	$\leftarrow \mathbf{\alpha}_0^T$
Ψ_0	$\leftarrow \mathbf{P}_0^T$
Σ_c	$\leftarrow \frac{1}{T \cdot p} \sum_{i=1}^{\text{rank}(C)} \sigma_i U_i' U_i$ <p>where $C = \sum_{t=1}^T \left(\text{vec}(\mathbf{\alpha}_t^T) \text{vec}(\mathbf{\alpha}_t^T)' + \mathbf{P}_t^T + \text{vec}(\mathbf{\alpha}_{t-1}^T) \text{vec}(\mathbf{\alpha}_{t-1}^T)' + \mathbf{P}_{t-1}^T - \right.$ $\left. (\text{vec}(\mathbf{\alpha}_t^T) \text{vec}(\mathbf{\alpha}_{t-1}^T)' + \mathbf{P}_{t,t-1}^T) - (\text{vec}(\mathbf{\alpha}_{t-1}^T) \text{vec}(\mathbf{\alpha}_{t-1}^T)' + \mathbf{P}_{t-1,t}^T) \right)$ $= \sum_{i=1}^{\text{rank}(C)} \sigma_i u_i u_i'$ and $\text{vec}(U_i) = u_i, U_i \in \mathbb{R}^{p \times d}$</p>
β'	$\leftarrow \text{unvec} \left(\left[\sum_{t=1}^T (\mathbf{x}_t \mathbf{x}_t' \otimes \mathbf{\alpha}_t^T \Omega^{-1} \mathbf{\alpha}_t^T) + (\mathbf{x}_t \mathbf{x}_t' \otimes \bar{\mathbf{P}}_t) \right]^{-1} \left[\sum_{t=1}^T \text{vec}(\mathbf{\alpha}_t^T \Omega^{-1} (\mathbf{y}_t - \mathbf{\Gamma} \mathbf{u}_t) \mathbf{x}_t') \right] \right)$ <p>where $\bar{\mathbf{P}}_t \in \mathbb{R}^{p \times p}$ with $\bar{\mathbf{P}}_{t,ij} = \text{tr}(\Omega^{-1} (\mathbf{P}_t^T)_{((i-1)p+1):(ip), ((j-1)p+1):(jp)})$</p>
Ω	$\leftarrow T^{-1} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{\alpha}_t^T \beta' \mathbf{x}_t - \mathbf{\Gamma} \mathbf{u}_t) (\mathbf{y}_t - \mathbf{\alpha}_t^T \beta' \mathbf{x}_t - \mathbf{\Gamma} \mathbf{u}_t)' + (\mathbf{x}_t' \beta \otimes \mathbf{I}_p) \mathbf{P}_t^T (\mathbf{x}_t' \beta \otimes \mathbf{I}_p)'$
$\mathbf{\Gamma}$	$\leftarrow \left(\sum_{t=1}^T (\mathbf{y}_t - \mathbf{\alpha}_t^T \beta' \mathbf{x}_t) \mathbf{u}_t' \right) \left(\sum_{t=1}^T \mathbf{u}_t \mathbf{u}_t' \right)^{-1}$
(B)	
\mathbf{b}_0	$\leftarrow \beta_0^T$
Ψ_0	$\leftarrow \mathbf{P}_0^T$
Σ_c	$\leftarrow \frac{1}{T \cdot q} \sum_{i=1}^{\text{rank}(C)} \sigma_i U_i' U_i$ <p>where $C = \sum_{t=1}^T \left(\text{vec}((\beta_t^T)') \text{vec}((\beta_t^T)')' + \mathbf{P}_t^T + \text{vec}((\beta_{t-1}^T)') \text{vec}((\beta_{t-1}^T)')' + \mathbf{P}_{t-1}^T - \right.$ $\left. (\text{vec}((\beta_t^T)') \text{vec}((\beta_{t-1}^T)')' + \mathbf{P}_{t,t-1}^T) - (\text{vec}((\beta_{t-1}^T)') \text{vec}((\beta_{t-1}^T)')' + \mathbf{P}_{t-1,t}^T) \right)$ $= \sum_{i=1}^{\text{rank}(C)} \sigma_i u_i u_i'$ and $\text{vec}(U_i) = u_i, U_i \in \mathbb{R}^{d \times q}$</p>
α	$\leftarrow \left(\sum_{t=1}^T (\mathbf{y}_t - \mathbf{\Gamma} \mathbf{u}_t) \mathbf{x}_t' \beta_t^T \right) \left(\sum_{t=1}^T (\beta_t^T)' \mathbf{x}_t \mathbf{x}_t' \beta_t^T + (\mathbf{x}_t' \otimes \mathbf{I}_d) \mathbf{P}_t^T (\mathbf{x}_t \otimes \mathbf{I}_d) \right)^{-1}$
Ω	$\leftarrow T^{-1} \sum_{t=1}^T (\mathbf{y}_t - \alpha (\beta_t^T)' \mathbf{x}_t - \mathbf{\Gamma} \mathbf{u}_t) (\mathbf{y}_t - \alpha (\beta_t^T)' \mathbf{x}_t - \mathbf{\Gamma} \mathbf{u}_t)' + (\mathbf{x}_t' \otimes \alpha) \mathbf{P}_t^T (\mathbf{x}_t \otimes \alpha)'$
$\mathbf{\Gamma}$	$\leftarrow \left(\sum_{t=1}^T (\mathbf{y}_t - \alpha (\beta_t^T)' \mathbf{x}_t) \mathbf{u}_t' \right) \left(\sum_{t=1}^T \mathbf{u}_t \mathbf{u}_t' \right)^{-1}$

By construction, the EM-algorithm is guaranteed to increase the data likelihood in every step of the estimation procedure (see e.g. Casella and Berger 2002, Thm. 7.2.20.). If it converges, it reaches a stationary point; i.e., a local maximum of the data likelihood.

The non-identifiability of the decompositions (A) and (B) in (2.6) does not interfere with the estimation algorithm in the sense that the estimates for the matrices \mathbf{C}_t and the predictions for \mathbf{y}_{t+1} do not depend on the chosen factorization. For example, consider model (A) and an arbitrary non-singular matrix \mathbf{H} and define $\tilde{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}_t \mathbf{H}$ and $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} \mathbf{H}^{-1'}$. Then, the corresponding state-space system

$$\begin{aligned} \text{vec}(\tilde{\boldsymbol{\alpha}}_{t+1}) &= \text{vec}(\tilde{\boldsymbol{\alpha}}_t) + \tilde{\boldsymbol{\eta}}_t \\ \mathbf{y}_t &= (\mathbf{x}_t' \tilde{\boldsymbol{\beta}} \otimes \mathbf{I}_p) \text{vec}(\tilde{\boldsymbol{\alpha}}_t) + \boldsymbol{\varepsilon}_t \end{aligned} \quad (2.17)$$

is equivalent to the (original) state space system (2.10) if we transform the noise covariances and the starting values for the Kalman filter correspondingly. This means that the filtered and smoothed states and the corresponding covariances of these two systems are related to each other by the same transformation. That is, $\tilde{\boldsymbol{\alpha}}_t^s = \boldsymbol{\alpha}_t^s \mathbf{H}$ and $\tilde{P}_t^s = (\mathbf{H}' \otimes \mathbf{I}_p) P_t^s (\mathbf{H} \otimes \mathbf{I}_p)$ for $s \in \{t-1, t, T\}$ and all $t = 1, \dots, T$. This observation also carries over to the EM-updates in Table 2.1. For example, the relationships $\tilde{\boldsymbol{\beta}}^{(j)} = \boldsymbol{\beta}^{(j)} \mathbf{H}^{-1'}$, $\tilde{\boldsymbol{\Sigma}}_c^{(j)} = \mathbf{H}' \boldsymbol{\Sigma}_c^{(j)} \mathbf{H}$, and $\tilde{\boldsymbol{\Omega}}^{(j)} = \boldsymbol{\Omega}^{(j)}$ hold for the estimates of models (2.17) and (2.10).

Choice of starting values

The choice of starting values $\boldsymbol{\Theta}^{(0)}$ used for numerical optimization of the likelihood can have a strong influence on the resulting model fit due to local optima. However, our EM-algorithm is robust to the choice of starting values in our experiments.

We start with the initialization of the coefficient matrices \mathbf{a}_0 and $\boldsymbol{\beta}$ for model (A), and respectively $\boldsymbol{\alpha}$ and \mathbf{b}_0 for model (B). We denote the starting values for both cases by $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$, and determine them by time-constant RRR, as follows. We let $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\alpha}_{RR}$, and $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}_{RR}$ where $\boldsymbol{\alpha}_{RR}$ and $\boldsymbol{\beta}_{RR}$ are minimizers of the criterion

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_t)' \mathbf{W} (\mathbf{y}_t - \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_t) \quad (2.18)$$

for any positive definite, symmetric matrix \mathbf{W} . The solutions to (2.18) give a rank d -approximation of the ordinary least squares (OLS) regression matrix $\hat{\mathbf{C}}_{\text{OLS}}$ (Reinsel and Velu 1998, Ch. 2). A close look at the OLS estimate reveals why the above starting values are reasonable. Assuming that $\mathbb{E}(\boldsymbol{\varepsilon}_t | \mathbf{x}_1, \dots, \mathbf{x}_T) = 0$, the conditional expectation of the OLS estimate given the predictors is

$$\tilde{\mathbf{C}} = \mathbb{E}(\hat{\mathbf{C}}_{\text{OLS}} | \mathbf{x}_1, \dots, \mathbf{x}_T) = \sum_{t=1}^T \mathbf{C}_t \mathbf{x}_t \mathbf{x}_t' \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1}. \quad (2.19)$$

This conditional expectation is a weighted average of the time-varying matrices \mathbf{C}_t . Thus, its reduced-rank decomposition yields “averages” of the time-varying coefficients $\boldsymbol{\alpha}_t$ (resp. $\boldsymbol{\beta}_t$). In particular, for model (B), $\text{span}(\mathbf{C}_t) = \text{span}(\boldsymbol{\alpha})$ and thus also $\text{span}(\tilde{\mathbf{C}}) = \text{span}(\boldsymbol{\alpha}) \approx \text{span}(\boldsymbol{\alpha}_{RR})$, hinting that $\boldsymbol{\alpha}_{RR}$ is a reasonable starting value for $\boldsymbol{\alpha}$. If pronounced time-variation is suspected, it may be advisable to obtain starting values from a reduced-rank regression that only uses the first $\tau < T$ observations of the time series.

The initial value for the error covariance $\boldsymbol{\Omega}$ is set to the residual covariance of the time-constant RRR,

$$\boldsymbol{\Omega}^{(0)} = T^{-1} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\alpha}^{(0)} \boldsymbol{\beta}^{(0)' \prime} \mathbf{x}_t) (\mathbf{y}_t - \boldsymbol{\alpha}^{(0)} \boldsymbol{\beta}^{(0)' \prime} \mathbf{x}_t)'$$

If the number of parameters is large compared to the sample size, it can be restricted to be diagonal or to be a multiple of the identity $\boldsymbol{\Omega} = \sigma_\varepsilon^2 \mathbf{I}_p$. The updates for the EM-algorithm can be adjusted accordingly. Simulations imply that assuming a diagonal covariance matrix does not necessarily decrease the accuracy of the estimation of the states and the prediction performance of the model. If we wish to include additional predictors \mathbf{u}_t , the starting values for $\boldsymbol{\Gamma}$ and correspondingly adjusted starting values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ can be obtained from reduced-rank regression with an additional full rank coefficient matrix as derived in Reinsel and Velu (1998, Ch. 3). In our simulations, the choice of \mathbf{W} in (2.18) did not have a strong influence on the parameter estimates. We used either the identity \mathbf{I}_p or the least squares residual covariance.

The uncertainty about the initial states $\boldsymbol{\alpha}_0$, resp. $\boldsymbol{\beta}_0$, is represented in the choice of the initial covariance matrix $\boldsymbol{\Psi}_0$: The larger the covariance, the more flexibility and uncertainty about the initial state is allowed for.

Finally, there is no obvious choice for starting values of the state covariance matrix $\boldsymbol{\Sigma}_c$. Generally, the magnitude of the entries of $\boldsymbol{\Sigma}_c$ governs the flexibility of the resulting state estimates – the basic paths of the states will be similar regardless of the magnitude of $\boldsymbol{\Sigma}_c$. Our simulation studies indicate that different starting values for $\boldsymbol{\Sigma}_c$ only marginally influence the value of the likelihood and the residual variance at the time of convergence. Nevertheless, it is advisable to try different values and check for stability of the resulting estimates.

2.2.3. Rank selection

Rank selection for the tvRRR model can be achieved by information criteria. We use the Bayesian information criterion (BIC), first introduced by Schwarz (1978), and given by

$$\text{BIC}(d) = -2 \log(L_{\mathbf{Y}}(\hat{\boldsymbol{\Theta}})) + \log(T) K_d, \quad (2.20)$$

where $\hat{\boldsymbol{\Theta}}$ denotes the maximum likelihood estimates (computed with the EM algorithm) and K_d is the number of free parameters of the model. The estimate for the rank is the minimizer of this BIC criterion (2.20). For model (A), $\boldsymbol{\Sigma}_c$ has $d(d+1)/2$ free parameters, and $\boldsymbol{\beta}$ has

$q \cdot d$ free parameters, so that we set $K_d = d(d+1)/2 + q \cdot d$. Similarly, $K_d = d(d+1)/2 + p \cdot d$ for model (B). We do not count the parameters in the matrices $\mathbf{\Omega}$ and $\mathbf{\Gamma}$, since the dimension of these matrices does not depend on d . We also do not count the parameters for the distribution of the initial state \mathbf{a}_0 (resp. \mathbf{b}_0) and $\mathbf{\Psi}_0$.

Another popular criterion for model selection is Akaike's information criterion (AIC, Akaike 1973), where $\log(T)$ in (2.20) is replaced by 2. However, it has been shown in similar, time-constant settings [e.g. Aznar and Salvador (2002) for the choice of the cointegrating rank in vector error correction models; Bai and Ng (2002) for the number of factors in dynamic factor models] that AIC results in inconsistent rank estimates as the penalty term needs to tend to infinity at a certain rate to achieve consistency. Our simulation studies indicate that AIC does not lead to consistent estimates in our model. Specifically, its performance deteriorates with increasing sample size.

2.3. Simulation studies

In this section, we illustrate the filter performance via a toy example, and demonstrate the flexibility and usefulness of tvRRR in different settings in a simulation study.

2.3.1. Illustration of filter performance

We generate data \mathbf{y}_t from a five-dimensional VAR(1)-model with rank $d = 2$ coefficient matrix,

$$\mathbf{y}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}' \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \text{ where } \boldsymbol{\alpha}_t = \begin{cases} \boldsymbol{\alpha}_1, & \text{for } t \leq 50 \\ \boldsymbol{\alpha}_2 & \text{for } t > 50 \end{cases}, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_5(0, \mathbf{I}_5), \quad (2.21)$$

where $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ and $\boldsymbol{\beta}$ are random orthonormal matrices. Model (2.21) represents a structural break at time $t = 50$. The realizations of the five time series are displayed in the top panel of Figure 2.1. The structural break at time $t = 50$ induces a considerable change in the autocovariance structure of the time series.

We fit a tvRRR model using the estimation procedure in Section 2.2.2. The estimated coefficient matrices are compared to those obtained from time-constant RRR and "oracle" RRR. For the latter, we assume the location of the structural break is known and fit two separate RRR models. We assess the accuracy of the estimate $\widehat{\mathbf{C}}_t = \boldsymbol{\alpha}_t^t \widehat{\boldsymbol{\beta}}'$ of the coefficient matrix $\mathbf{C}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}'$ with the normalized Frobenius norm of the estimation error,

$$err_{C_t} := \frac{\|\mathbf{C}_t - \widehat{\mathbf{C}}_t\|_F}{\|\mathbf{C}_t\|_F}. \quad (2.22)$$

Its values are displayed in the bottom panel of Figure 2.1. The tvRRR coefficient estimates (---) are very close to those of oracle RRR (—). At the structural break, the error jumps

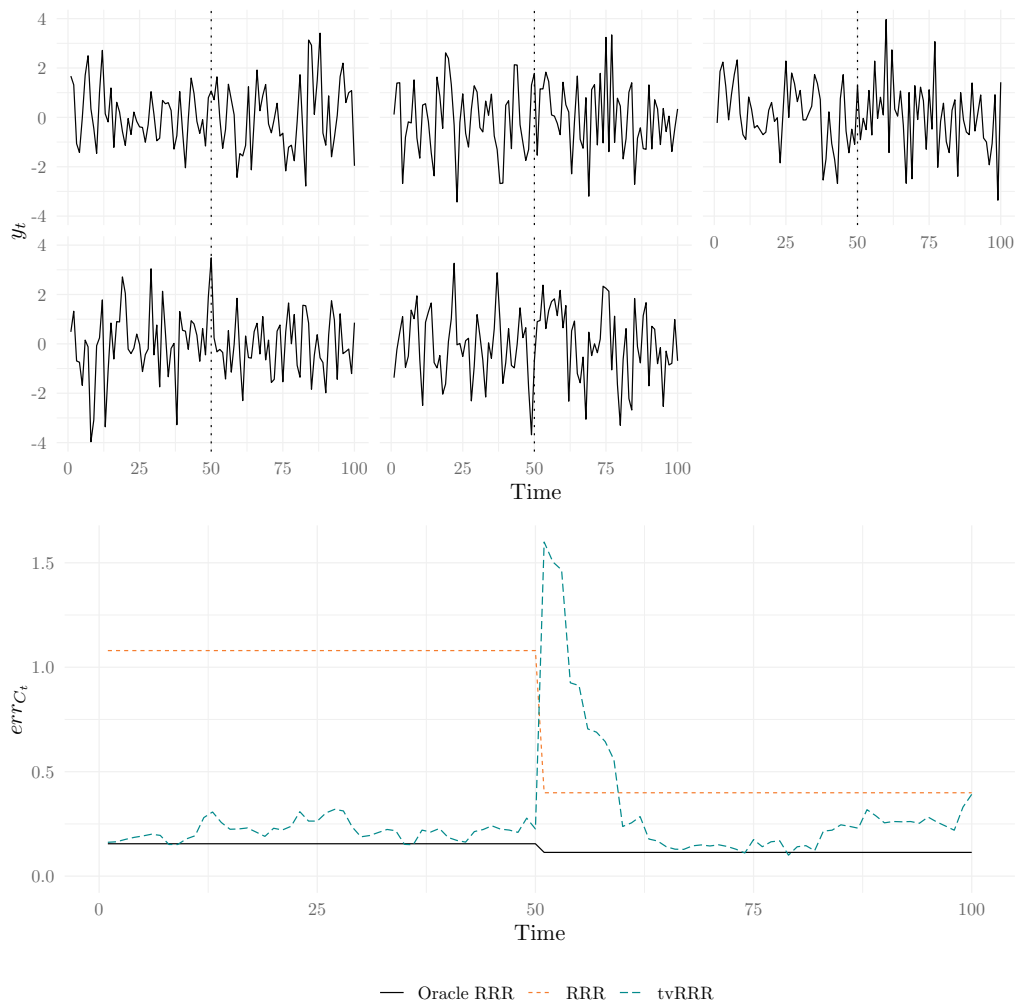


Figure 2.1.: The top panel(s) show a trajectory of the five time series (\mathbf{y}_{it}) , $i = 1, \dots, 5$ generated by the VAR(1) model (2.21) with a structural break of the coefficient matrix at $t = 50$. The bottom panel shows the corresponding estimation error $err_{C_t} = \|\mathbf{C}_t - \hat{\mathbf{C}}_t\|_F / \|\mathbf{C}_t\|_F$ for the three estimates considered: oracle RRR, time-constant RRR and tvRRR .

up rapidly, yet the coefficient estimates quickly recover and roughly line up with those of oracle RRR. As expected, time-constant RRR errors (- - -) are the largest as the method simply averages over the two matrices α_1 and α_2 (as can be seen from Equation (2.19)). This toy example also indicates the EM-algorithm we use to fit tvRRR is robust to the choice of the starting values, which were obtained from RRR, thus far from the truth.

2.3.2. Simulation

The data for our simulation study are drawn from the model equations

$$(A) \quad \mathbf{y}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad \text{and} \quad (B) \quad \mathbf{y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t. \quad (2.23)$$

The predictors $\mathbf{x}_t \in \mathbb{R}^q$ are specified as exogenous predictors, drawn from q independent MA(2) processes. The errors $\boldsymbol{\varepsilon}_t$ are normally distributed with mean zero and diagonal covariance matrix $\boldsymbol{\Omega}$. The diagonal entries of $\boldsymbol{\Omega}$ are drawn from the uniform distribution on the interval $[0.5, 1.5]$. Starting values for the EM-algorithm are obtained from time-constant reduced-rank regression as described in Section 2.2.2. The weight matrix \mathbf{W} in (2.18) is selected as the residual covariance from the ordinary least squares regression of \mathbf{y}_t on \mathbf{x}_t . The initial state covariance $\boldsymbol{\Psi}_0$ is set to $\boldsymbol{\Psi}_0 = 1000 \cdot \mathbf{I}_{pd}$ for model (A), and $\boldsymbol{\Psi}_0 = 1000 \cdot \mathbf{I}_{qd}$ for model (B). The large variances reflect the uncertainty about the initial values used for the Kalman filter. We also conducted experiments where the data follow a VAR(1) process with time-varying coefficients, i.e. we set $\mathbf{x}_t = \mathbf{y}_{t-1}$ in (2.23). Due to the random walk generation of the coefficient matrices, the time series often diverge. For this reason, we only report results for exogenous \mathbf{x}_t . However, when the VAR time series remain bounded, our method performs well.

To study the ability of the proposed method to capture different types of time variation, we generate datasets under three different types of time-varying coefficients. They are displayed in Table 2.2. The notation $\boldsymbol{\alpha} \sim \text{UStiefel}(p, d)$ indicates that $\boldsymbol{\alpha}$ is drawn from the uniform distribution on the Stiefel manifold $\mathbb{V}_{p,d}$ (see (2.9)) as defined in Chikuse (2003).

The fixed scenario was included in order to evaluate the robustness of TVRRR in the case of a true RRR model with time-invariant parameters. The *deterministic* transition produces a gradual change of the parameters from the starting to the terminal coefficient matrix. The *random walk* agrees with the modeling assumption that underlies our procedure and represents random fluctuations in the coefficients. The matrices $\boldsymbol{\Sigma}_d$ depend on the state dimension d and are selected as multiples of the matrices at the bottom of Table 2.2.

The dimension of the target time series vector takes values $p \in \{5, 10, 20\}$, the predictors are generated with dimension $q \in \{5, 10, 20\}$, and the rank takes values $d \in \{1, 2, 3\}$. Each time series has length $T = 200$. The validation time series of length $\tau = 200$ is generated by extrapolating the data generating processes in Table 2.2. For each combination of parameters, we draw 10 sets of coefficient matrices, and repeat the model fitting procedure 10 times, resulting in a total of $10 \cdot 10 = 100$ simulation runs per setting.

The in-sample MSE for model (A) in (2.10) is

$$MSE_{is} = \frac{1}{Tp} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\alpha}_t^t \widehat{\boldsymbol{\beta}}' \mathbf{x}_t)' (\mathbf{y}_t - \boldsymbol{\alpha}_t^t \widehat{\boldsymbol{\beta}}' \mathbf{x}_t). \quad (2.24)$$

The corresponding out-of-sample MSE is obtained from one-step-ahead predictions for the

2. A State-Space Approach to Time-Varying Reduced-Rank Regression

Table 2.2.: Coefficient matrices in the simulation studies for models (A) and (B). The time-constant parts are drawn as $\beta \sim \text{UStiefel}(q, d)$, resp. $\alpha \sim \text{UStiefel}(p, d)$.

Name	Model	Specification of matrices	Parameters
Fixed	A	$\alpha_t \equiv \alpha \forall t = 1, \dots, T$ $\alpha \sim \text{UStiefel}(p, d)$	–
	B	$\beta_t \equiv \beta \forall t = 1, \dots, T$, $\beta \sim \text{UStiefel}(q, d)$	–
Deterministic	A	Deterministic linear transition between two matrices $\alpha_1, \alpha_2 \sim \text{UStiefel}(p, d)$ such that $\alpha_t = s \frac{T-t}{T-1} \alpha_1 + s \frac{t-1}{T-1} \alpha_2$	shift $s \in \{1, 1.5, 2\}$
	B	Deterministic linear transition between two matrices $\beta_1, \beta_2 \sim \text{UStiefel}(q, d)$ such that $\beta_t = s \frac{T-t}{T-1} \beta_1 + s \frac{t-1}{T-1} \beta_2$	shift $s \in \{1, 1.5, 2\}$
Random walk	A	Matrices follow a random walk, i.e. $\alpha_t \sim \mathcal{N}_{p,d}(\alpha_{t-1}; \mathbf{I}_p, \Sigma_c)$	$\Sigma_c = s \cdot \Sigma_d$ where $s \in \{0.1, 1, 2\}$
	B	Matrices follow a random walk, i.e. $\beta_t \sim \mathcal{N}_{q,d}(\beta_{t-1}; \mathbf{I}_q, \Sigma_c)$ where $\Sigma_1 = 0.005, \quad \Sigma_2 = \begin{pmatrix} 0.005 & 0.0025 \\ 0.0025 & 0.005 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 0.005 & 0.0025 & 0.00125 \\ 0.0025 & 0.005 & 0.0025 \\ 0.00125 & 0.0025 & 0.005 \end{pmatrix}$	$\Sigma_c = s \cdot \Sigma_d$ where $s \in \{0.1, 1, 2\}$

validation data using the Kalman filter and the fitted parameter matrices,

$$MSE_{oos} = \frac{1}{\tau p} \sum_{t=T+1}^{T+\tau} (\mathbf{y}_t - \alpha_t^{t-1} \hat{\beta}' \mathbf{x}_t)' (\mathbf{y}_t - \alpha_t^{t-1} \hat{\beta}' \mathbf{x}_t). \quad (2.25)$$

We assess the accuracy of the estimate $\hat{\mathbf{C}}_t = \alpha_t^t \hat{\beta}'$ with the normalized Frobenius distances of the estimated coefficient matrices in (2.22). For model (B), these criteria are defined in an analogous manner. Throughout the simulation study, the performance is juxtaposed to that of the regular reduced-rank regression reference model.

Results

We report results for $p = q = 10$. The respective tables for other pairs of p and q are comparable and can be found in the supplementary material.

First, we present results for the *fixed* data generating process. Table 2.3 reports the in- and out-of-sample MSE in (2.24) and (2.25), respectively, and relative estimation errors (2.22) for RRR and tvRRR in models (A) and (B). Under this setting, simple RRR is more appropriate than tvRRR. This results in larger relative estimation errors err_{C_t} even though it seems that forecasting precision is only mildly affected. The in-sample MSE of tvRRR is smaller whereas the out-of-sample MSE is slightly larger than the respective MSEs of RRR in both models. tvRRR is more flexible than the time-constant RRR. As a consequence, it tends to slightly overfit the training data, resulting in a lower in-sample MSE and a higher out-of-sample MSE in the validation data.

Table 2.3.: Average MSE and relative parameter estimation error for fixed setting and $p = q = 10$ with exogenous predictors, standard deviation in parentheses.

	d	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)	
A	1	in sample	0.9731 (0.0467)	0.9194 (0.0502)	0.0461 (0.0067)	0.1445 (0.0124)
		out of sample	1.0112 (0.0527)	1.0269 (0.0507)	-	-
	2	in sample	0.9815 (0.1201)	0.8802 (0.1162)	0.0383 (0.0080)	0.1231 (0.0138)
		out of sample	1.0419 (0.1401)	1.0736 (0.1408)	-	-
	3	in sample	0.9224 (0.1054)	0.7973 (0.0983)	0.0386 (0.0091)	0.1189 (0.0172)
		out of sample	0.9919 (0.1105)	1.0316 (0.1144)	-	-
B	1	in sample	1.0077 (0.0758)	0.9749 (0.0750)	0.0408 (0.0070)	0.1571 (0.0153)
		out of sample	1.0386 (0.0786)	1.0567 (0.0799)	-	-
	2	in sample	0.9855 (0.0912)	0.9234 (0.0911)	0.0392 (0.0057)	0.1319 (0.0082)
		out of sample	1.0408 (0.0997)	1.0725 (0.0991)	-	-
	3	in sample	0.9998 (0.0755)	0.9014 (0.0706)	0.0402 (0.0046)	0.1394 (0.0094)
		out of sample	1.0666 (0.0816)	1.1142 (0.0822)	-	-

Next, we report the simulation results for the deterministic transition of the coefficients as specified in Table 2.2. The shift (s) is set to 1 (“small”), 1.5 (“medium”) and 2 (“large”). As can be seen in Table 2.4, this setting incurs a significant loss in estimation precision for RRR with increasing variation of the parameter matrices. While the in-sample fit is not much affected, out-of-sample prediction fails. In contrast, tvRRR is uniformly better than RRR. It has slightly worse out-of sample than in-sample MSE, while both remain roughly stable across all setting combinations. Moreover, the coefficient estimation accuracy of tvRRR improves as the shift between the starting and final coefficient matrices increases. This effect is probably partly due to the relative error criterion. The estimation accuracy of RRR stays roughly the same across all settings.

2. A State-Space Approach to Time-Varying Reduced-Rank Regression

Table 2.4.: Average MSE and relative parameter estimation error for deterministic coefficient transition and $p = q = 10$ with external predictors, standard deviation in parentheses.

	s	d		MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
A	1	1	in sample	1.0096 (0.1119)	0.9151 (0.1028)	0.3793 (0.1383)	0.2894 (0.0879)
			out of sample	1.6575 (0.2608)	1.0450 (0.1115)	-	-
		2	in sample	1.0801 (0.0607)	0.8848 (0.0680)	0.3770 (0.1202)	0.2831 (0.0741)
			out of sample	2.3940 (0.3604)	1.1295 (0.0783)	-	-
		3	in sample	1.0660 (0.0779)	0.8049 (0.0634)	0.3351 (0.0531)	0.2548 (0.0329)
			out of sample	2.9252 (0.4102)	1.1117 (0.0980)	-	-
	1.5	1	in sample	1.1022 (0.1054)	0.9334 (0.0590)	0.2608 (0.1483)	0.1502 (0.0646)
			out of sample	2.1126 (0.4480)	1.0875 (0.0734)	-	-
		2	in sample	1.0666 (0.1011)	0.7717 (0.0777)	0.2096 (0.0474)	0.1141 (0.0163)
			out of sample	3.4841 (0.6698)	1.0365 (0.0978)	-	-
		3	in sample	1.2539 (0.0764)	0.7955 (0.0802)	0.2670 (0.0665)	0.1405 (0.0246)
			out of sample	4.6271 (0.7045)	1.1954 (0.1063)	-	-
2	1	in sample	1.0805 (0.0891)	0.8659 (0.0791)	0.1671 (0.0493)	0.0785 (0.0180)	
		out of sample	2.4777 (0.4529)	1.0386 (0.0855)	-	-	
	2	in sample	1.3422 (0.1467)	0.8173 (0.0909)	0.2651 (0.0767)	0.0975 (0.0170)	
		out of sample	6.3858 (1.0800)	1.1556 (0.1121)	-	-	
	3	in sample	1.4983 (0.1434)	0.7842 (0.0646)	0.2706 (0.0411)	0.1079 (0.0150)	
		out of sample	8.3580 (1.3044)	1.2588 (0.0982)	-	-	
B	1	1	in sample	1.0672 (0.0998)	0.9962 (0.0904)	0.4490 (0.1771)	0.3457 (0.1020)
			out of sample	1.6924 (0.3180)	1.0948 (0.0925)	-	-
		2	in sample	1.0602 (0.0793)	0.9215 (0.0789)	0.3635 (0.0786)	0.2677 (0.0284)
			out of sample	2.1380 (0.2809)	1.1122 (0.0917)	-	-
		3	in sample	1.0636 (0.1344)	0.8826 (0.1219)	0.3315 (0.0644)	0.2677 (0.0467)
			out of sample	2.8633 (0.2484)	1.1558 (0.1437)	-	-
	1.5	1	in sample	1.0876 (0.0913)	0.9661 (0.0662)	0.3587 (0.2036)	0.2154 (0.1228)
			out of sample	2.4989 (0.8415)	1.0957 (0.0808)	-	-
		2	in sample	1.1106 (0.1144)	0.8594 (0.0984)	0.3262 (0.0976)	0.1807 (0.0381)
			out of sample	3.9858 (0.7443)	1.0933 (0.1091)	-	-
		3	in sample	1.2790 (0.0800)	0.8965 (0.0693)	0.3382 (0.0807)	0.1842 (0.0315)
			out of sample	5.3404 (0.7817)	1.2561 (0.0886)	-	-
2	1	in sample	1.0893 (0.0906)	0.9098 (0.0762)	0.3070 (0.1289)	0.1582 (0.0455)	
		out of sample	3.6538 (0.8978)	1.0592 (0.0930)	-	-	
	2	in sample	1.2554 (0.0794)	0.8661 (0.0715)	0.3119 (0.0762)	0.1303 (0.0258)	
		out of sample	6.5400 (1.0845)	1.1381 (0.0772)	-	-	
	3	in sample	1.3520 (0.0754)	0.8032 (0.0271)	0.2873 (0.0534)	0.1235 (0.0191)	
		out of sample	8.3886 (1.5862)	1.2044 (0.0295)	-	-	

The difference in performance between RRR and tvRRR can be explained by considering the errors of the coefficient matrix estimates over time in Figure 2.2. RRR averages over the coefficient matrices (see also the discussion on the starting values, Equation (2.19)), and thus the smallest distance is achieved at approximately $T/2$. This estimate is far off

for time points that are closer to the start and end of the time series, and deviate even further for the validation data set $t = 201, \dots, 400$. On the other hand, the estimation errors of tvRRR are consistently small throughout the time range. In particular, since the RRR estimates serve as starting values for the EM-algorithm, this further illustrates the robustness of tvRRR to “wrong” starting values.

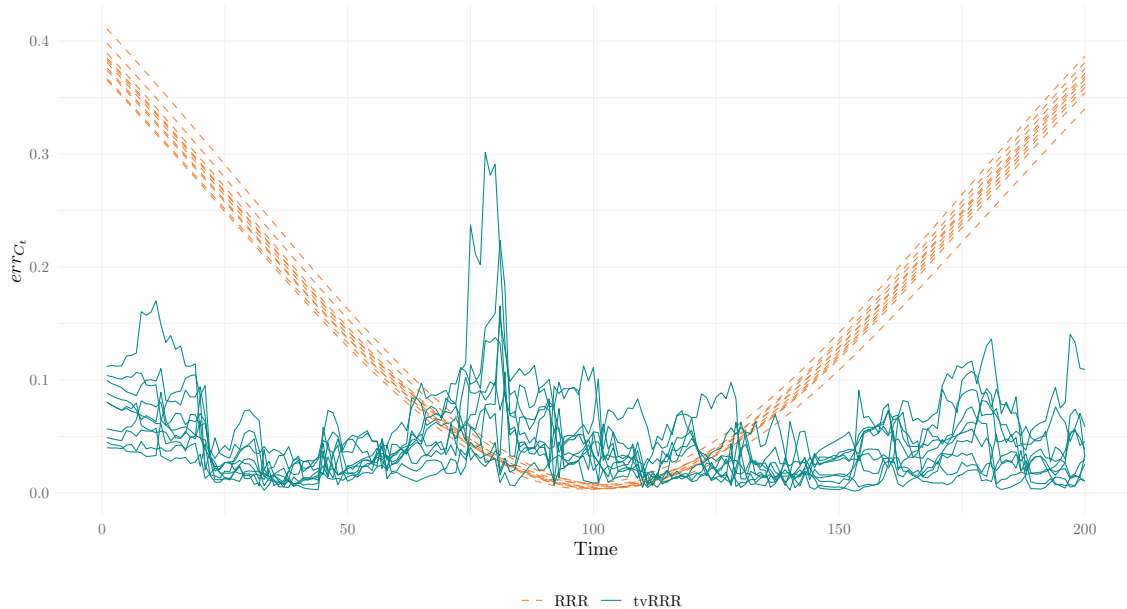


Figure 2.2.: err_{C_t} for ten simulation runs with the same set of coefficient matrices from the deterministic setting, RRR in orange (dashed), tvRRR in teal (solid), $p = q = 10$, $d = 1$, large variation, model A.

Table 2.5 displays the respective results for the random walk coefficient evolution in Table 2.2. Even for the low variation random walk, we observe the failure of RRR both in terms of MSE and of estimation accuracy. Both in- and out-of-sample MSEs are significantly larger than the corresponding values for tvRRR. As in the deterministic setting, tvRRR gains from increasing variation in the coefficients.

2. A State-Space Approach to Time-Varying Reduced-Rank Regression

Table 2.5.: Average MSE and relative parameter estimation error for random walk coefficient transition and $p = q = 10$ with external predictors, standard deviation in parentheses.

	s	d		MSE (RRR)	MSE (TVRRR)	err_{C_t} (RRR)	err_{C_t} (TVRRR)
A	0.1	1	in sample	0.9975 (0.0615)	0.9196 (0.0566)	1.8192 (0.5351)	0.7513 (0.2070)
			out of sample	1.2201 (0.1253)	1.0250 (0.0757)	-	-
		2	in sample	1.0620 (0.0877)	0.9199 (0.0906)	1.6212 (0.5203)	0.6609 (0.0871)
			out of sample	1.5204 (0.2758)	1.1018 (0.0931)	-	-
		3	in sample	1.0747 (0.1133)	0.8781 (0.1097)	1.6104 (0.3326)	0.6695 (0.0921)
			out of sample	1.7913 (0.2968)	1.1720 (0.1390)	-	-
	1.0	1	in sample	1.3507 (0.1150)	0.9233 (0.0829)	1.4125 (0.5407)	0.2529 (0.0693)
			out of sample	3.3552 (0.8473)	1.1837 (0.0850)	-	-
		2	in sample	1.7951 (0.2495)	0.7577 (0.0864)	1.2498 (0.2394)	0.2539 (0.0379)
			out of sample	7.0220 (2.6184)	1.2433 (0.1087)	-	-
		3	in sample	2.0290 (0.2905)	0.7326 (0.0562)	1.3983 (0.4371)	0.2747 (0.0488)
			out of sample	6.9571 (1.2448)	1.3938 (0.0802)	-	-
2.0	1	in sample	1.8004 (0.3165)	0.8448 (0.0962)	1.5344 (0.5307)	0.2084 (0.0364)	
		out of sample	5.3302 (2.1873)	1.2026 (0.1306)	-	-	
	2	in sample	2.6983 (0.5685)	0.7524 (0.0735)	1.4511 (0.3304)	0.2181 (0.0594)	
		out of sample	10.9765 (3.4839)	1.4067 (0.1048)	-	-	
	3	in sample	3.3075 (0.5509)	0.6729 (0.0930)	1.3349 (0.3448)	0.2017 (0.0280)	
		out of sample	13.6254 (2.7399)	1.6052 (0.1727)	-	-	
B	0.1	1	in sample	0.9685 (0.0723)	0.9237 (0.0742)	1.7297 (0.5701)	0.8691 (0.3079)
			out of sample	1.1945 (0.1383)	1.0065 (0.0724)	-	-
		2	in sample	1.0795 (0.1078)	0.9768 (0.1124)	1.8046 (0.3527)	0.7927 (0.0845)
			out of sample	1.5048 (0.2448)	1.1534 (0.1363)	-	-
		3	in sample	1.1224 (0.0787)	0.9658 (0.0802)	1.6299 (0.3092)	0.7298 (0.0844)
			out of sample	1.8261 (0.3654)	1.2150 (0.0818)	-	-
	1.0	1	in sample	1.3784 (0.1782)	0.9559 (0.0780)	1.3746 (0.4419)	0.4623 (0.1249)
			out of sample	3.0428 (0.8102)	1.1735 (0.0669)	-	-
		2	in sample	1.8141 (0.1512)	0.9056 (0.0599)	1.5669 (0.5451)	0.4259 (0.0524)
			out of sample	5.4969 (1.0081)	1.3393 (0.0796)	-	-
		3	in sample	1.9853 (0.2120)	0.8129 (0.0800)	1.4505 (0.3965)	0.4080 (0.0772)
			out of sample	8.4159 (3.2621)	1.4182 (0.1275)	-	-
2.0	1	in sample	1.6238 (0.2489)	0.9216 (0.0840)	1.9476 (0.8077)	0.4498 (0.1199)	
		out of sample	6.0949 (2.0447)	1.2117 (0.1063)	-	-	
	2	in sample	2.5481 (0.4328)	0.8555 (0.1051)	1.6949 (0.4620)	0.3317 (0.0482)	
		out of sample	10.5402 (2.9861)	1.4310 (0.1299)	-	-	
	3	in sample	2.9481 (0.7160)	0.8241 (0.0856)	1.4663 (0.5431)	0.3842 (0.0687)	
		out of sample	16.7759 (4.4155)	1.6799 (0.1570)	-	-	

2.3.3. Rank selection

We assess the performance of the BIC criterion (2.20) for rank selection in both models (A) and (B) and three scenarios for time varying parameters. In addition to the deterministic

transition (with $s = 1$) and the random walk transition (with $s = 0.1$), as defined in Table 2.2, we also consider a structural break setting similar to (2.21): For model (A), we draw two coefficient matrices $\alpha_1, \alpha_2 \sim \text{UStiefel}(p, d)$ and let $\alpha_t = \alpha_1$ for $t \leq \frac{T}{2}$ and $\alpha_t = \alpha_2$ for $t > \frac{T}{2}$. For model (B) we proceed analogously. We look at time series of length $T \in \{100, 200, 300\}$ and the same sets of values for p, q and d as before. Again, we carry out $10 \cdot 10 = 100$ repetitions of each setting.

The simulation results indicate that the BIC criterion is not sensitive to the coefficient transitions we consider. Table 2.6 reports the fraction of simulation runs where the rank was correctly estimated in both models for $p = q = 10$. Results for other combinations of p and q are comparable and can be found in the supplement. BIC rarely overestimates the true rank. For smaller samples, it frequently underestimates the rank, but drastically improves as the sample size increases. Rank estimation seems to be the least successful for the random walk coefficient transition, where the tendency to underestimate the rank of the model is the strongest. However, the performance of the criterion also in this scenario markedly improves with increasing the length of the time series.

Table 2.6.: Results of rank selection via BIC criterion for $p = q = 10$ and three different coefficient transitions, $10 \cdot 10 = 100$ repetitions each. “=” is the fraction of correct rank estimation, and “<” the fraction of underestimated ranks.

d	T	deterministic		random walk				str. break					
		Model (A)		Model (B)		Model (A)		Model (B)		Model (A)		Model (B)	
		=	<	=	<	=	<	=	<	=	<	=	<
1	100	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
1	200	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
1	300	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
2	100	0.890	0.110	0.750	0.250	0.020	0.980	0.040	0.960	0.810	0.190	0.870	0.130
2	200	0.990	0.010	1.000	-	0.570	0.430	0.480	0.520	1.000	-	1.000	-
2	300	1.000	-	1.000	-	0.968	0.031	1.000	-	1.000	-	1.000	-
3	100	0.430	0.570	0.250	0.750	-	1.000	-	1.000	0.570	0.430	0.580	0.420
3	200	0.980	0.020	0.970	0.030	0.110	0.890	0.110	0.890	1.000	-	1.000	-
3	300	1.000	-	1.000	-	0.710	0.290	0.787	0.212	1.000	-	1.000	-

2.4. Data examples

We apply tvRRR and an adaptive version of RRR that we define in this section on two data sets. The first consists of seven main stock indices, and the second contains Covid-19 cases for 12 European countries.

2.4.1. Stock index data

We perform an out-of-sample forecasting experiment with stock index data. We retrieved closing prices for DAX Performance Index (DAX), Dow Jones Industrial Average (DJI),

Hang Seng Index (HSI), Nikkei 225 (N225), NASDAQ Composite (NASDAQ), and S+P 500 (S+P 500) and TSEC Weighted Index (TWII) for two different time periods.³ The first period ranges from January 15, 2008 to May 18, 2009, and the second from June 05, 2019 to October 05, 2020. Both time series include abrupt changes in global economy. The first period contains the Great Recession and the second the Covid-19 pandemic. While it is thought that stock index data are cointegrated, the relationships of the time series might have changed during both crises, possibly due to markets recovering from the shocks at different rates.

We fit a lag-1 vector error correction model (VECM) with possibly time-varying cointegrating relations $\beta_t' \mathbf{y}_{t-1}$,

$$\Delta \mathbf{y}_t = \alpha \beta_t' \mathbf{y}_{t-1} + \Gamma \Delta \mathbf{y}_{t-1} + \varepsilon_t.$$

Prior to the analysis, the stock index time series are logarithmized, centered and scaled so that the sample mean equals zero and the variance is equal to one in the training data set. Missing values are imputed using `na_kalman` from the `imputeTS` R-package (Moritz and Bartz-Beielstein 2017). We fit the models to the first 150 observations of the time series, and, using the estimated coefficients, predict the following 200 observations.

We compare the following three model fitting approaches: tvRRR, time-constant RRR and time-constant updated RRR. For tvRRR, we compute the one-step ahead predictions with the Kalman filter. However, for the estimation of the model parameters, we only use the training data. These predictions are compared to two versions of the time-constant VECM. For the first, we fit the model to the first 150 observations and obtain predictions for the following 200 data points using the estimated coefficients (time-constant RRR). The second, fairer comparison model is a VECM that is refitted with every new observation of the prediction period to include all the past information in the prediction (updated RRR).

The MSEs from in-sample and out-of-sample prediction of the time series are reported relative to benchmark (“naive”) predictions obtained from the assumption that stock index data follow a random walk (Kendall 1953; Fama 1965):

$$\hat{\mathbf{y}}_{t+1} = \mathbf{y}_t \text{ and } \widehat{\Delta \mathbf{y}}_{t+1} = \hat{\mathbf{y}}_{t+1} - \mathbf{y}_t = 0. \quad (2.26)$$

We start with the analysis of the 2008 data. The seven time series before and after scaling are displayed in the left top and bottom panels of Figure 2.3. The structural break in September 2008 is evident in both, with the scaled data capturing it better. The decline of the prices was also accompanied by an increased volatility, which can be clearly seen in the plots of the returns (log first differences). The training period is January 15, 2008 to August 11, 2008. When fitting the tvRRR model, the rank selected by the BIC criterion is $d = 2$. The EM-algorithm converges after 54 iterations, and the likelihood reaches a plateau fairly quickly.

The relative in- and out-of-sample MSEs for both versions of time-constant RRR and

³All time series retrieved from <https://finance.yahoo.com/>, accessed March 01, 2021.

tvRRR are reported in the top panel of Table 2.7. The performance of RRR and tvRRR is comparable in-sample. However, we are more interested in the out-of-sample performance of the methods. We thus have applied the (modified) Diebold-Mariano test (Harvey et al. 1997) to check, whether there are significant differences in the forecasting performance of the naive method, updated RRR and tvRRR. For the 2008 data, tvRRR yields significantly better predictions than the naive prediction in four cases, and it beats updated RRR for three series. Conversely, the prediction performance of tvRRR is never significantly worse than that of the competitors. Its capability to adapt even to abrupt changes in the coefficient estimates while retaining estimation precision when the parameters are constant pays off: The coefficient estimates adapt to the structural break.

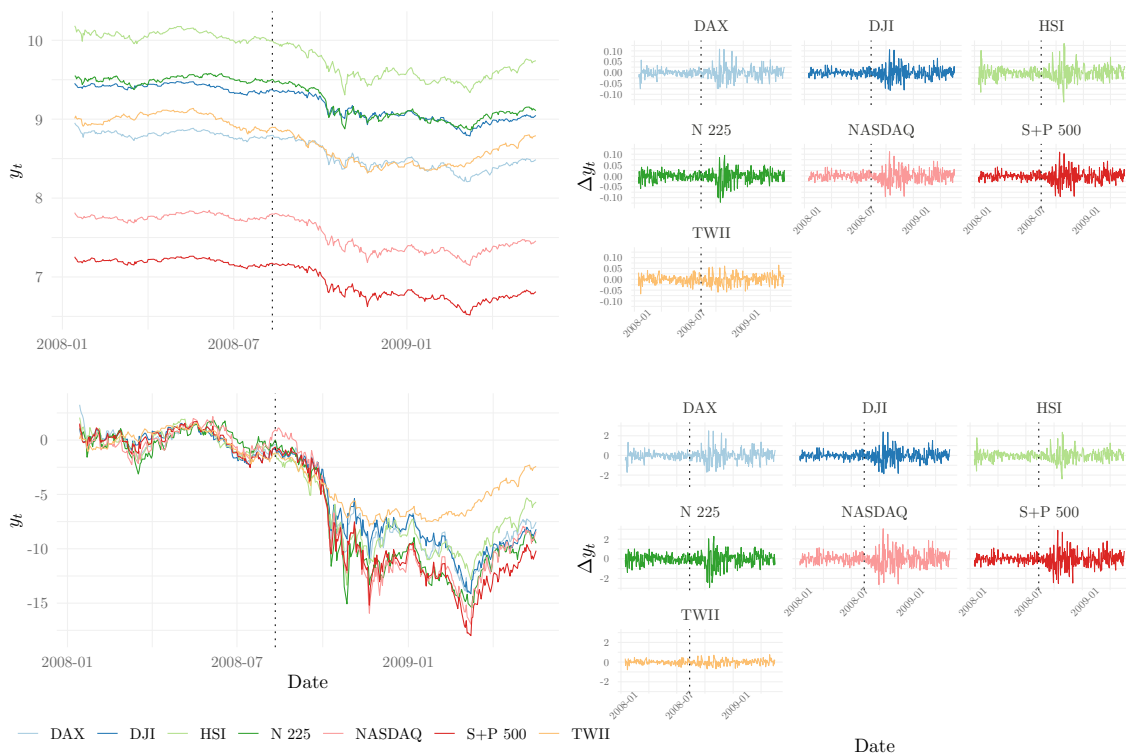


Figure 2.3.: 2008 stock index data before (top left) and after (bottom left) scaling and centering with values from the training period. The right panels show the corresponding log first-differences. The vertical dashed lines mark the end of the training period.

As expected, RRR fails completely at prediction. The different behavior of the methods may also be seen in the plot of the $d = 2$ estimated cointegrating relationships $\beta'_{1t}y_{t-1}$ and $\beta'_{2t}y_{t-1}$ in Figure 2.4, where β'_{it} denotes the i th row of β'_t . The RRR components (top) show a trending behavior. In contrast, the cointegrating relationships inferred using tvRRR (middle) only vary around zero, but they clearly display the increased volatility during the crisis. Compare also the plot of the log first-differences.

2. A State-Space Approach to Time-Varying Reduced-Rank Regression

Table 2.7.: In- and out-of-sample MSEs of tvRRR, RRR and updated RRR for the 2008 and 2020 stock data sets, relative to the MSEs for the naive predictions evolving from the random walk assumption (2.26). For the out-of-sample MSEs, boldface numbers indicate significantly better prediction performance than the naive predictions according to the Diebold-Mariano test, and numbers in italics indicate significantly better performance of tvRRR than updated RRR at 0.05 significance level.

		2008						
		DJI	DAX	HSI	N 225	NASDAQ	S+P 500	TWII
in-sample	TVRRR	0.9424	0.6555	0.4442	0.4641	0.9599	0.9509	0.6941
	RRR	0.8018	0.7035	0.5301	0.4938	0.8142	0.8101	0.7216
out-of-sample	TVRRR	<i>0.9439</i>	1.0503	1.0691	<i>0.6923</i>	0.9444	<i>0.9517</i>	0.9265
	RRR	3.3134	1.6374	2.0129	2.2012	3.8061	3.3488	1.6735
	RRR (upd.)	1.1229	1.0970	1.0014	0.6980	1.1339	1.1090	0.8960

		2020						
		DJI	DAX	HSI	N 225	NASDAQ	S+P 500	TWII
in-sample	TVRRR	0.8861	0.9691	0.7350	0.3340	0.7729	0.7269	0.5940
	RRR	0.8092	0.8819	0.8156	0.5048	0.7986	0.7749	0.6649
out-of-sample	TVRRR	0.9733	<i>0.9997</i>	1.0711	0.8085	<i>0.9555</i>	0.9648	0.9423
	RRR	1.9315	2.3440	3.5379	4.4262	2.4440	2.4903	2.5606
	RRR (upd.)	1.3869	1.4096	1.3999	1.1272	1.3891	1.4010	1.0368

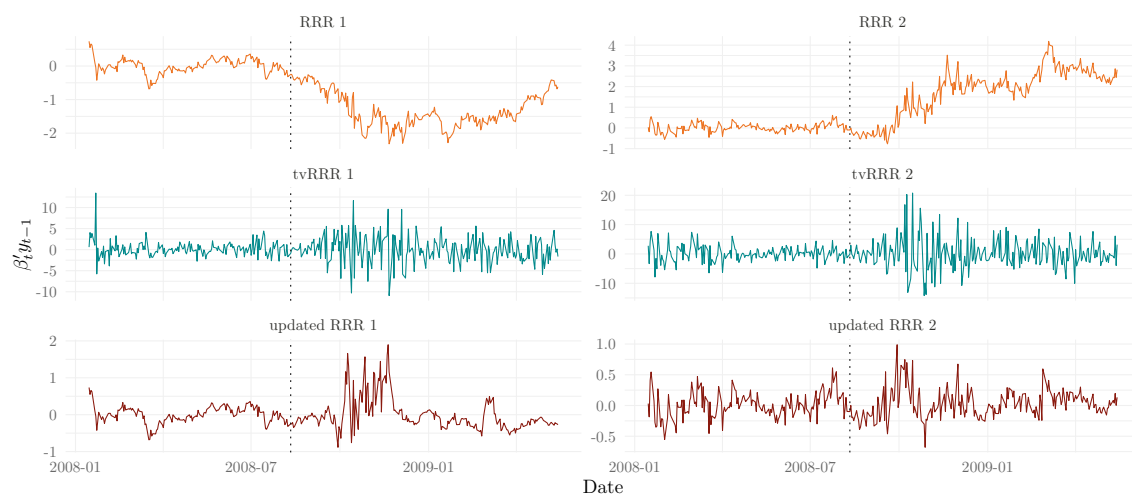


Figure 2.4.: Data reductions / stationary linear combinations of the time series $\beta_t^T y_{t-1} \in \mathbb{R}^2$, inferred from the three competing models for the 2008 data. The vertical dashed lines mark the end of the training period.

We next analyze the 2020 data. The corresponding time series are displayed in Figure 2.5. Here, the structural break is even more abrupt. Whereas the time series decreased slowly in 2008, the 2020 series exhibit a sudden drop around March 2020, with a fairly quick global recovery. The training period ranges from June 6, 2020 to December 30, 2020. When fitting tvRRR, the BIC criterion again selects $d = 2$ as the order of the model. The EM-algorithm converges after 104 iterations.

The corresponding MSEs are displayed in the bottom panel of Table 2.7. We observe a similar pattern as for the 2008 data set as regards the in-sample performance. According to the Diebold-Mariano test, the forecasting performance of tvRRR is significantly better than that of the naive prediction in one of the cases, and it is significantly better than updated RRR in three cases. In contrast, the forecasting performance of tvRRR is never significantly worse than that of the naive predictions and of updated RRR. Hence, also for the 2020 data, tvRRR shows significant performance improvement.

The cointegrating relations $\beta_t' y_{t-1}$ exhibit a similar pattern as those for 2008 shown in Figure 2.4. Therefore we do not include a corresponding figure for the 2020 data.

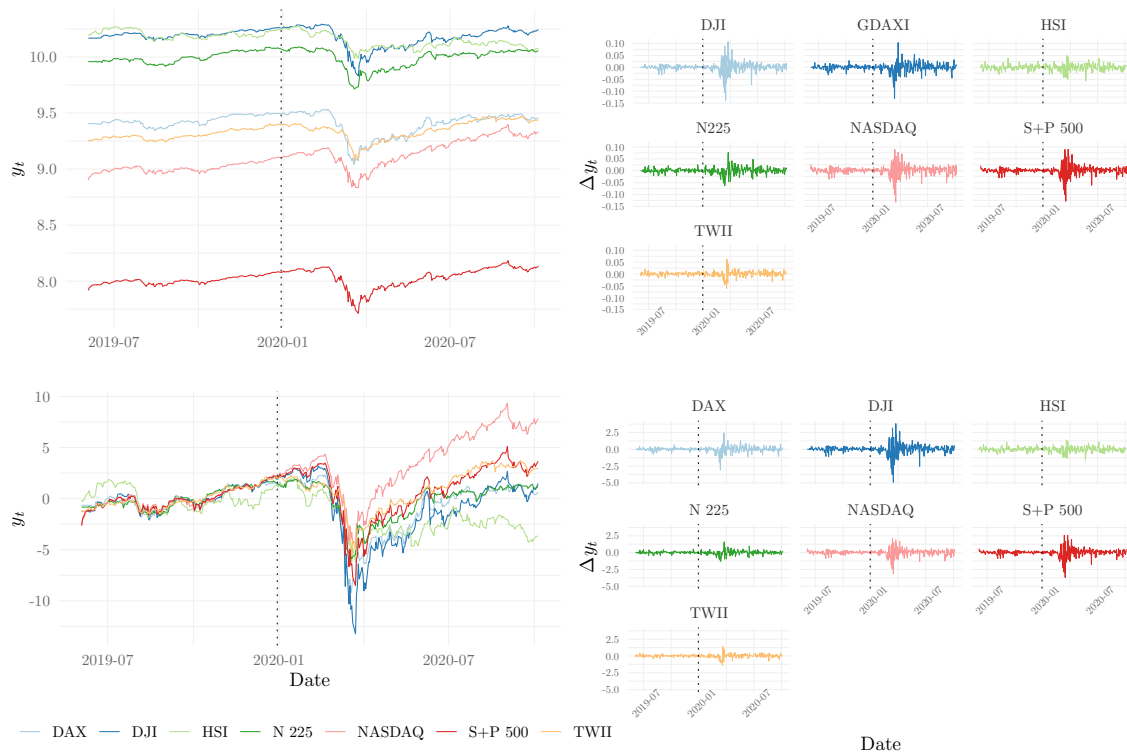


Figure 2.5.: 2020 stock index data before (top left) and after (bottom left) scaling and centering with values from the training period. The right panels show the corresponding log first-differences. The vertical dashed lines mark the end of the training period.

2.4.2. Covid-19 data

We analyze weekly Covid-19 cases per 100 000 capita for 12 European countries, namely Austria, Belgium, Czech Republic, France, Germany, Greece, Italy, Netherlands, Poland, Portugal, Sweden and Switzerland.⁴ Due to their proximity and frequent border crossings, the data can be assumed to be strongly correlated. Nevertheless, the relationships of the time series are likely to have changed during the pandemic.

We fit a time-varying VAR(1) model with reduced rank coefficient matrix of type (A),

$$\mathbf{y}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}' \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t.$$

We do not preprocess the data. We fit the model starting at the end of March 2020 until February 2021, and predict the remaining values of the time series up to mid-November 2021, using tvRRR and updated RRR, as described in Section 2.4.1. More precisely, for updated RRR we use a time-constant model for the estimation period which is then refitted with every new observation in the validation period. The data were roughly split in two halves. The training and prediction period consist of 50 and 40 observations, respectively. Both models use latent rank $d = 2$.

The time series and the corresponding predictions are displayed in Figure 2.6. During the

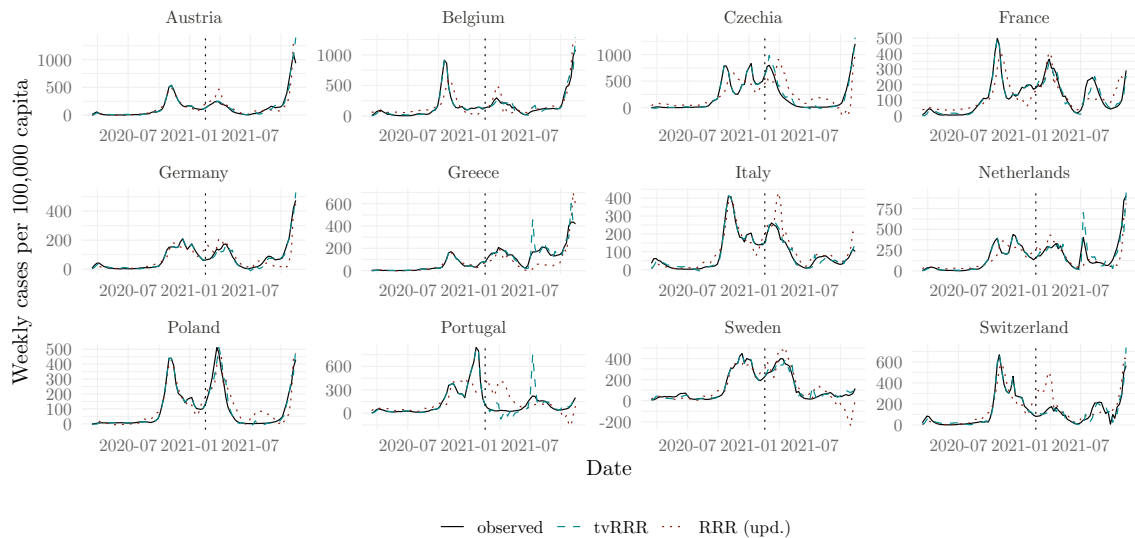


Figure 2.6.: Weekly Covid-19 cases during the pandemic for 12 European countries (—), with predictions obtained from tvRRR (— —) and updated RRR (···). The vertical dashed lines mark the end of the training period.

⁴Data retrieved from <https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19>, accessed December 5, 2021.

model fitting period, tvRRR captures the first and second wave of the pandemic nearly perfectly. For updated RRR, we can again observe the effect of averaging: The second peak leads to overestimation of the starting period. In the validation period from February 2021 to November 2021, the one-step-ahead predictions of tvRRR closely track the progression of the number of Covid-19 cases in all countries. In contrast, the updated RRR performs badly in the validation period. This can also be seen from the mean squared prediction errors shown in Figure 2.7. The overall MSE of tvRRR in the prediction period is less than one quarter of the MSE of updated RRR (3873 vs. 16761). The Diebold-Mariano test finds statistically significantly better prediction performance of tvRRR for all countries except for the Netherlands.

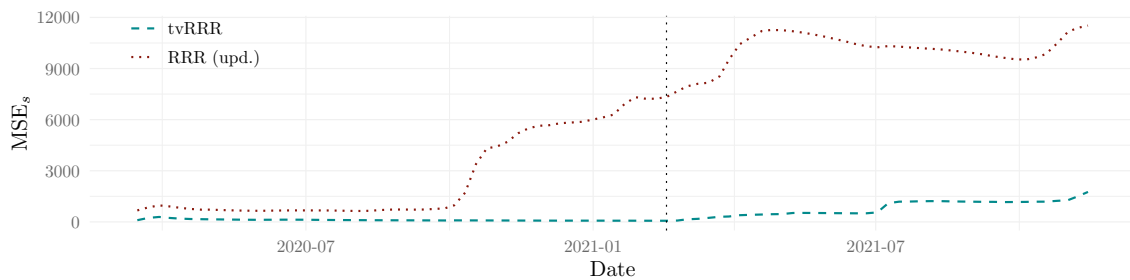


Figure 2.7.: Cumulative mean squared prediction errors $MSE_s = \sum_{t=1}^s (\mathbf{y}_t - \hat{\mathbf{y}}_t)'(\mathbf{y}_t - \hat{\mathbf{y}}_t)/12s$, $s = 1, \dots, 90$, (over all countries up to time s) for the weekly Covid-19 case data. The reported models are tvRRR (—) and updated RRR (···). The vertical dashed line marks the end of the training period.

We also tried model (B) for this data set. However, its performance is far worse than that of model (A), indicating that the time-invariant linear combinations $\beta' \mathbf{y}_t$ contain essential information for prediction. However, these linear combinations have a time-varying effect α_t on the Covid-19 cases in the different countries.

2.5. Conclusions

Reduced-rank regression forms the basis for various models frequently used in applications, such as multivariate time-series regressions, cointegration models and methods for sufficient dimension reduction based on inverse regression. We propose a new time-varying coefficient reduced-rank regression modeling approach (tvRRR) by modeling the dynamics of the observations and the time-varying coefficients as a Gaussian linear state-space system. The estimation algorithm relies on classical methods from state-space modeling. We estimate the time-varying parameters using the Kalman filter, and fit the model to observed data using an EM-algorithm which relies on analytic updates and is numerically stable. Our EM-algorithm converges quickly most of the time. It needs on average 40 iterations in our simulation studies. It takes longer to reach convergence in real data applications, as expected. With respect to computer time, the estimation of the model for the Covid data

takes about 10 seconds on a standard laptop with an Intel(R) i5 processor with 16 GB RAM. We use BIC to estimate the rank of the coefficient matrices with highly accurate results.

Our simulation studies and the data applications show the adaptive capabilities of the algorithm. It exhibits high estimation accuracy across different patterns of time-variation, as it accommodates both abrupt structural breaks and gradual changes in the time-varying coefficients. Moreover, tvRRR exhibits robustness to time-constant coefficients. An R-implementation of the model fitting algorithm is available on GitHub (<https://github.com/b-brune/tvRRR>). Additional simulation results are provided in the supplementary material.

Acknowledgements We thank the editor, the associate editor and three anonymous referees for their comments and suggestions that helped us improve the paper. This research was partially supported by the Austrian Science Fund (FWF P 30690-N35) and Vienna Science and Technology Fund (WWTF ICT19-018). The simulation study presented in this paper was carried out using the Vienna Scientific Cluster (VSC).

3 | Rank-Based Estimation of Mixed Effects Models

This chapter is based on joint work with Irene Ortner and Peter Filzmoser and is conditionally accepted as *A Robust Rank-Based Estimation Method For Mixed Effects Models in the Presence of Outlying Data* in the Journal of Data Science, Statistics, and Visualization (<https://jdssv.org/index.php/jdssv/index>). Supplemental information for this chapter can be found in Appendix B, p. 133ff.

3.1. Introduction

The mixed effects model is a popular extension of the standard linear regression model. It is able to deal with clustered error structures, e.g. in cases where multiple measurements are made on the same subject. In this case, the error terms are no longer independent and identically distributed.

In general, the model equation for a standard linear mixed effects model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta}$ are the so-called fixed effects, \mathbf{b} are the individual- or group-specific random effects, and $\boldsymbol{\varepsilon}$ are the independent errors. \mathbf{X} and \mathbf{Z} are matrices of predictors and \mathbf{y} is the response. If \mathbf{Z} is the vector of ones, the associated random effect is called a random intercept, otherwise we speak about random slopes. In the classical parametric setting, it is assumed that both the errors and random effects are homoscedastic and follow a normal distribution. Based on this assumption, the standard approaches of estimation are maximum likelihood (ML) and restricted maximum likelihood (REML). Pinheiro and Bates (2000) provide a comprehensive overview of both approaches.

In practice, we are often confronted with data that do not comply with the normality assumption imposed by the standard estimation approaches mentioned above. The data may contain measurement errors or unusual observations (so-called *outliers*) and the error distributions might be non-symmetric or heavy-tailed. Such deviations can cause the standard parametric estimation procedures to break down, resulting in heavily distorted coefficient estimates. The contaminated estimates make it difficult to identify the outlying observations which can be masked by wrong fits.

The fields of non-parametric and robust statistics aim to develop methodology that does not rely (or relies far less) on distributional assumptions and is not distorted by the presence of unusual measurements. Different approaches to estimate mixed effects models in a robust manner have been proposed in the literature. Various methods are based on down-weighting observations, i.e. M- or S-estimation (see e.g. Copt and Victoria-Feser 2006; Koller 2013; Agostinelli and Yohai 2016). These methods are often computationally very expensive. Other methods replace the normality assumption for the error term and the random effects with more heavy-tailed distributions such as the t-distribution (Pinheiro, Liu, et al. 2001). A further line of robust and non-parametric estimation methods is provided in the framework of rank-based regression (see e.g. Kloke et al. 2009; Bilgic 2012; Jung and Ying 2003; Wang and Zhu 2006; Wang and Zhao 2008). These methods have the advantage of being computationally cheap, robust against outlying responses and efficient, even if the modeling assumptions for ML and REML as mentioned above are violated.

There are two main shortcomings of the existing rank-based methods for the estimation of linear mixed effects models: (1) In the setting of experiments that are not designed, some of the predictors included in \mathbf{X} or \mathbf{Z} might be unusual or outlying (e.g. due to measurement errors). Such leverage points can still distort the coefficient estimates obtained from the existing models. (2) To the best of our knowledge, there does not exist a rank-based estimation method that is able to deal with more complex random effects structures: All rank-based methods are designed for random intercepts, but cannot account for random slopes of any type. However, we often do not want to limit our model estimates to random intercepts. Instead, we might be interested in the individual-specific response to a treatment, not only with regard to changes in the average response, but also regarding the slope in a certain covariate.

The present work extends the rank-based estimation approaches presented in Kloke et al. (2009) and Bilgic (2012) to allow for model structures that include random slopes. Furthermore, a weighting procedure is introduced that protects the estimates against the effects of leverage points. The results allow for model diagnostics both for the whole sample and on group-level. Unusual groups and observations can be identified, aiding interpretation and deeper understanding of the modeled data.

3.2. Preliminaries

3.2.1. Rank-based regression

Given a regular regression model

$$\mathbf{y} = \alpha \mathbf{1}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

with $\mathbf{y} \in \mathbb{R}^N$, $\mathbf{X} \in \mathbb{R}^{N \times p}$, intercept parameter $\alpha \in \mathbb{R}$, regression coefficients $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\mathbf{1}_N$ the N -dimensional all-ones vector, where N describes the number of observations. The error term $\mathbf{e} \in \mathbb{R}^N$ is assumed to be generated by a continuous distribution with a positive

definite covariance matrix given by $\sigma_e^2 \mathbf{I}_N$ with variance parameter $0 < \sigma_e^2 < \infty$. \mathbf{I}_N is the N -dimensional identity matrix.

A robust rank-based estimate of the regression coefficients $\boldsymbol{\beta}$ can be obtained by minimizing a pseudo norm of the form

$$\|\mathbf{r}\|_\varphi = \sum_{j=1}^N a[R(r_j)]r_j \quad (3.1)$$

in the residuals $\mathbf{r} = \mathbf{y} - \alpha \mathbf{1}_N - \mathbf{X}\boldsymbol{\beta} \in \mathbb{R}^N$. Here, $R(r_j)$ denotes the rank of the residual r_j among the components r_1, \dots, r_N of the residual vector \mathbf{r} . The scores are given by $a[t] = \varphi(t/(N+1))$ where $\varphi(u)$ is a non-decreasing, bounded and square-integrable function such that $\sum_t a[t] = 0$. Proposed in Jureckova (1971), the pseudo norm (3.1) is also called *Jaekel's dispersion function* after Jaekel (1972). The score function $\varphi(\cdot)$ can be specified in various ways. We use the *Wilcoxon score*

$$\varphi(u) = \sqrt{12} \left[u - \frac{1}{2} \right],$$

which provides a good trade-off between efficiency and robustness, if the error term \mathbf{e} is not normally distributed (McKean 2004). With this score function, the norm can be written equivalently as a scaled sum of pairwise differences (see McKean 2004):

$$\|\mathbf{r}\|_\varphi = \frac{\sqrt{3}(N+1)}{2} \sum_{j=1}^N \sum_{k=1}^N |r_j - r_k| \quad (3.2)$$

The ordinary rank-based regression estimate is then given by

$$\hat{\boldsymbol{\beta}}_\varphi = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_\varphi. \quad (3.3)$$

The criterion (3.3) does not allow to estimate the intercept parameter α directly, because the ranks of the residuals are invariant to shifts by a constant. The intercept can be estimated by applying a robust estimator of location $T: \mathbb{R}^N \rightarrow \mathbb{R}$ to the vector of residuals $\tilde{\mathbf{r}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi$:

$$\hat{\alpha} = T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi) = T(\tilde{\mathbf{r}}).$$

The error variance σ_e^2 is estimated by applying a robust scale estimator, denoted by $S: \mathbb{R}^N \rightarrow \mathbb{R}^+$, to the resulting vector of residuals $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi - \hat{\alpha}$. We suggest using the *one-sample Hodges-Lehmann estimator* (Hodges and Lehmann 1963) for intercept estimation, and the *Q_n estimator of scale* (Rousseeuw and Croux 1993) to estimate the error variance. These two estimators provide a good trade-off between robustness and efficiency. Other possible choices for $T(\cdot)$ and $S(\cdot)$ are e.g. the *sample median* and the *median absolute deviation (MAD)*

There is no closed form solution for the minimizer $\hat{\boldsymbol{\beta}}_\varphi$ in (3.3). However, the optimization

problem is convex and can be solved efficiently using numerical optimization techniques. Useful starting values for the optimization can be obtained from an initial ordinary least squares (OLS) fit or, if we suspect outlying values, from least trimmed squares (LTS, Rousseeuw and Van Driessen 2006).

A thorough overview of the theory regarding the rank-based regression model may be found in Chapter 3 of Hettmansperger and McKean (2011).

3.2.2. Terminology of mixed effects models

Given are multiple measurements from g independent groups. Then, for the n_i -dimensional response vector \mathbf{y}_i from group i , $i = 1, \dots, g$, we aim to fit a model of the form

$$\mathbf{y}_i = \alpha \mathbf{1}_{n_i} + \mathbf{X}_i \boldsymbol{\beta} + a_i \mathbf{1}_{n_i} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i \quad (3.4)$$

where $N = \sum_{i=1}^g n_i$ is the total sample size.

We call the coefficients α and $\boldsymbol{\beta}$ that are shared by all groups the *fixed effects*. $\mathbf{X}_i \in \mathbb{R}^{n_i \times (p-1)}$ contains the predictors associated with the fixed effects.

The group-specific *random effects* $a_i \in \mathbb{R}$ and $\mathbf{b}_i \in \mathbb{R}^{k-1}$ are treated as realizations of a centered random variable $(a, \mathbf{b}')' \in \mathbb{R}^k$ with positive definite covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \in \mathbb{R}^{k \times k}$ that is parametrized by a coefficient vector $\boldsymbol{\theta}$. We assume $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a diagonal matrix, and $\boldsymbol{\theta} \in \mathbb{R}^k$ contains the random effects' standard deviations. More generally, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ could also be specified as compound symmetric or unstructured with a $\boldsymbol{\theta}$ of suitable length. $\mathbf{Z}_i \in \mathbb{R}^{n_i \times (k-1)}$ is a matrix of predictors corresponding to the random slope(s) \mathbf{b}_i . The matrices \mathbf{Z}_i can be subsets of the columns of the matrices \mathbf{X}_i , but also contain further external predictors.

The errors $\boldsymbol{\varepsilon}_i$ are assumed to follow a distribution with finite second moment and covariance matrix $\sigma_\varepsilon^2 \mathbf{I}_{n_i}$. In addition, the errors and the random effects are independent from each other. The coefficients σ_ε and $\boldsymbol{\theta}$ are also referred to as *variance components*.

Given estimators $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$ and \hat{a}_i , $\hat{\mathbf{b}}_i$, $i = 1, \dots, g$, respectively, we denote the *marginal* and *conditional* residuals by

$$\begin{aligned} \mathbf{r}_{i,\text{marg}} &= \mathbf{y}_i - \mathbf{1}_{n_i} \hat{\alpha} - \mathbf{X}_i \hat{\boldsymbol{\beta}}, \text{ and} \\ \mathbf{r}_{i,\text{cond}} &= \mathbf{y}_i - \mathbf{1}_{n_i} \hat{\alpha} - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{1}_{n_i} \hat{a}_i - \mathbf{Z}_i \hat{\mathbf{b}}_i, \end{aligned}$$

accordingly. For ML and REML estimation (Pinheiro and Bates 2000), the random effects and the errors are assumed to be centered, independent and normally distributed.

3.2.3. Rank-based regression for dependent data

Kloke et al. (2009) extend the methodology of rank-based regression with independent errors to estimate simple mixed effects models with random intercept terms. Their methodology

is further extended by Bilgic (2012) to allow for nested random intercepts of arbitrary depth. Both methods rely on a two-step estimation procedure, where the fixed effects are estimated in the first step. The residuals from this first regression are then used to derive the random effects and the variance components. The authors give asymptotic properties of their estimators and use them to draw inference on the fixed effects.

Other rank-based estimation methods for repeated measurements data include the method proposed in Jung and Ying (2003), and its extensions in Wang and Zhu (2006) and Wang and Zhao (2008). Those methods follow a different model formulation and do not explicitly specify the dependency structure of the data beforehand. The authors consider a simple linear regression model for repeated measurements data, and propose estimation methods that are able to deal with possible dependencies and heteroscedasticity. Similarly, Abebe et al. (2016) present rank-based fits for generalized estimating equations.

All of the above mentioned papers focus on drawing inference on the fixed effects in the model. Contrarily, we focus on the accurate estimation of fixed and random effects in the presence of outliers. This especially allows for model diagnostics and identification of unusual (groups of) observations as well as interpretation of the realizations of the random effects. We do not consider inference for the parameters.

3.3. Methodology

In the following, we use the formulation of the mixed effects model as presented in Section 3.2.2. The proposed estimation procedure relies on the rank-based regression presented in Section 3.2.1.

3.3.1. Rank-based estimation of mixed effects models with random slopes

For simplicity of the exposition and without loss of generality, we drop the intercept terms α and a_i , $i = 1, \dots, g$, from model (3.4). Details for the estimation with intercept can be found in Algorithm 3.1.

By ignoring the dependency structure induced by the random effects, we can obtain an initial estimate $\hat{\beta}^{(0)}$ of β from a rank-based regression of the stacked models (3.4)

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_g \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_g \end{pmatrix} \beta + \tilde{\mathbf{e}}. \quad (3.5)$$

The random effect structure is absorbed into the error term $\tilde{\mathbf{e}} \in \mathbb{R}^N$. Then, we calculate the marginal residuals as

$$\mathbf{r}_{i,\text{marg}}^{(0)} = \mathbf{y}_i - \mathbf{X}_i \hat{\beta}^{(0)} \approx \mathbf{Z}_i \mathbf{b}_i + \varepsilon_i. \quad (3.6)$$

3. Rank-Based Estimation of Mixed Effects Models

for each $i = 1, \dots, g$, This gives us g further regression problems that can be solved for the most likely realizations of the random effects \mathbf{b}_i :

$$\hat{\mathbf{b}}_i^{(0)} = \arg \min_{\mathbf{b}} \|\mathbf{r}_{i,\text{marg}}^{(0)} - \mathbf{Z}_i \mathbf{b}\|_{\varphi}. \quad (3.7)$$

We obtain estimates of the variance components as $\hat{\sigma}_{\varepsilon}^{(0)}$ and $\hat{\boldsymbol{\theta}}^{(0)}$ by applying the Q_n estimator to the conditional residuals

$$\mathbf{r}_{i,\text{cond}}^{(0)} = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}^{(0)} - \mathbf{Z}_i \hat{\mathbf{b}}_i^{(0)}, \quad (3.8)$$

and the estimated random effects $\hat{\mathbf{b}}_i^{(0)}$.

This set of initial estimates $\hat{\boldsymbol{\beta}}^{(0)}$, $\hat{\mathbf{b}}_i^{(0)}$, $\hat{\sigma}_{\varepsilon}^{(0)}$ and $\hat{\boldsymbol{\theta}}^{(0)}$ can now be improved using a reweighting scheme in the manner of iteratively reweighted least squares. Similar methodology for the random intercept model is proposed as the *generalized rank procedure* in Bilgic (2012).

The basic idea is the following: From the model equation (3.4), we can directly obtain the covariance of each vector \mathbf{y}_i , $i = 1, \dots, g$. It is given by

$$\boldsymbol{\Sigma}_{\mathbf{y}_i} = \boldsymbol{\Sigma}_{\mathbf{y}_i}(\sigma_{\varepsilon}, \boldsymbol{\theta}) = \text{Cov}(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i) = \sigma_{\varepsilon}^2 \mathbf{I}_{n_i} + \mathbf{Z}_i \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{Z}_i'$$

and can be robustly estimated by plugging in the variance component estimates:

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{y}_i}^{(0)} = \boldsymbol{\Sigma}_{\mathbf{y}_i}(\hat{\sigma}_{\varepsilon}^{(0)}, \hat{\boldsymbol{\theta}}^{(0)}).$$

By multiplying the system of regression equations (3.5) with $\boldsymbol{\Sigma}_{\mathbf{y}_i}^{-1/2}$, the errors $\tilde{\mathbf{e}}$ are transformed to be homoscedastic and uncorrelated. One can then solve the rescaled equation system(s)

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{y}_i}^{(0)-1/2} \mathbf{y}_i = \hat{\boldsymbol{\Sigma}}_{\mathbf{y}_i}^{(0)-1/2} \mathbf{X}_i \boldsymbol{\beta} + \hat{\boldsymbol{\Sigma}}_{\mathbf{y}_i}^{(0)-1/2} \tilde{\mathbf{e}}_i \quad (3.9)$$

to achieve a better, more efficient estimate of $\boldsymbol{\beta}$.

However, this rescaling step can impose problems in the presence of outlying predictors or responses. If observations do not follow the regression line, the rotation (3.9) can amplify their influence: Due to the linear combinations built in the multiplication with the estimated covariance matrix, one outlying observation or predictor in group i is able to contaminate the whole group.

This can be circumvented by a preceding reweighting step: Based on the conditional residuals (3.8), we define *outlyingness weights*

$$\nu_{ij}^{(0)} = \min \left\{ 1, \frac{c \cdot \hat{\sigma}_{\varepsilon}^{(0)}}{|r_{ij,\text{cond}}^{(0)}|} \right\}, \quad (3.10)$$

for $i = 1, \dots, g$, $j = 1, \dots, n_i$, where $r_{ij,\text{cond}}^{(0)}$ are the elements of the conditional residuals in

(3.8). The tuning parameter c is the cutoff value that is used to classify the residuals as outlying. As it is a common choice in the literature we set $c = 2$. We reweigh the regression equations as follows,

$$\text{diag}(\boldsymbol{\nu}^{(0)}) \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_g \end{pmatrix} = \text{diag}(\boldsymbol{\nu}^{(0)}) \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_g \end{pmatrix} \boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}},$$

where $\boldsymbol{\nu}^{(0)} = (\nu_{11}^{(0)}, \dots, \nu_{gn_g}^{(0)})'$ are the stacked weights. We then proceed with the rescaling step as described in Equation (3.9) and re-estimate $\boldsymbol{\beta}$. The new estimate can then be used to recalculate the residuals $\mathbf{r}_{i,\text{marg}}$ as in (3.6), and improve the random effects estimates $\hat{\mathbf{b}}_i$ as in (3.7). We iterate between these steps until the estimates stabilize.

For error variance estimation, we suggest to correct for the degrees of freedom lost through the separate fitting of the g regression models that estimate the random effects. Thus, we apply the following finite-sample correction to the Q_n estimator of scale:

$$Q_{n,\text{corr}}(\mathbf{r}) = \sqrt{\frac{N}{N - (p + g)}} Q_n(\mathbf{r}).$$

This correction gives good results in simulation studies.

The model fitting procedure is summarized in Algorithm 3.1. In all our simulation experiments, the algorithm converged in less than 5 iterations.

Algorithm 3.1 (Rank-based fitting of mixed effects models with random slopes). *Starting with $l = 0$, the l 'th iteration of the algorithm is the following:*

1. Let

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{y}_i}^{(l)} = \begin{cases} \mathbf{I}_{n_i}, & \text{if } l = 0 \\ \boldsymbol{\Sigma}_{\mathbf{y}_i}(\hat{\sigma}_{\varepsilon}^{(l-1)}, \hat{\boldsymbol{\theta}}_b^{(l-1)}), & \text{else,} \end{cases}$$

and

$$\boldsymbol{\nu}^{(l)} = \begin{cases} \mathbf{1}_N, & \text{if } l = 0 \\ (\nu_{11}^{(l-1)}, \dots, \nu_{gn_g}^{(l-1)})', & \text{else.} \end{cases}$$

Obtain $\hat{\boldsymbol{\beta}}^{(l)}$ as a solution to the rank-based regression of

$$\mathbf{y}_*^{(l)} = (\hat{\boldsymbol{\Sigma}}_{\mathbf{y}_i}^{(l)})^{-\frac{1}{2}} \text{diag}(\boldsymbol{\nu}^{(l)}) \mathbf{y} \quad \text{on} \quad \mathbf{X}_*^{(l)} = (\hat{\boldsymbol{\Sigma}}_{\mathbf{y}_i}^{(l)})^{-\frac{1}{2}} \text{diag}(\boldsymbol{\nu}^{(l)}) \mathbf{X}.$$

If the model has an intercept α , estimate

$$\hat{\alpha}^{(l)} = T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(l)}).$$

3. Rank-Based Estimation of Mixed Effects Models

Otherwise, $\hat{\alpha}^{(l)} = 0$

2. Calculate the marginal residuals

$$\mathbf{r}_{i,marg}^{(l)} = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}^{(l)} - \hat{\alpha}^{(l)} \mathbf{1}_{n_i}.$$

3. Use $\mathbf{r}_{i,marg}^{(l)}$ to obtain predictions $\hat{\mathbf{b}}_i^{(l)}$ for the random effects \mathbf{b}_i from (3.7).

If the model has random intercepts a_i , estimate

$$\hat{a}_i^{(l)} = T(\mathbf{r}_{i,marg}^{(l)} - \mathbf{Z}_i \hat{\mathbf{b}}_i^{(l)}).$$

Otherwise, $\hat{a}_i^{(l)} = 0$.

4. Obtain conditional residuals

$$\mathbf{r}_{i,cond}^{(l)} = \mathbf{y}_i - \hat{\alpha}^{(l)} \mathbf{1}_{n_i} - \mathbf{X}_i \hat{\boldsymbol{\beta}}^{(l)} - \hat{a}_i^{(l)} \mathbf{1}_{n_i} - \mathbf{Z}_i \hat{\mathbf{b}}_i^{(l)}.$$

5. Use the $\mathbf{r}_{i,cond}^{(l)}$ and the realizations $\hat{a}_i^{(l)}$, $\hat{\mathbf{b}}_i^{(l)}$ to estimate the variance components $\boldsymbol{\theta}^{(l)}$ and $\sigma_\varepsilon^{(l)}$ with the Q_n estimator (Rousseeuw and Croux 1993, see).

6. Determine the outlyingness weights $\nu_{ij}^{(l)}$ from the conditional residuals using (3.10).

7. If $l > 0$, check for convergence using the criteria

$$\frac{\|\hat{\boldsymbol{\beta}}^{(l)} - \hat{\boldsymbol{\beta}}^{(l-1)}\|_2}{\|\hat{\boldsymbol{\beta}}^{(l-1)}\|_2} < tol, \quad \text{and} \quad \frac{\|(\hat{\sigma}_\varepsilon^{(l)}, \hat{\boldsymbol{\theta}}_b^{(l)})' - (\hat{\sigma}_\varepsilon^{(l-1)}, \hat{\boldsymbol{\theta}}_b^{(l-1)})'\|_2}{\|(\hat{\sigma}_\varepsilon^{(l-1)}, \hat{\boldsymbol{\theta}}_b^{(l-1)})'\|_2} < tol$$

for a prescribed tolerance tol .

If above criteria are not fulfilled or $l = 0$, increase l by one and return to step 1.

3.3.2. Further robustification against leverage points

The proposed method is robust against outlying values in the response space (*y-outliers*), but can be heavily affected by outliers in the predictor space (*X-outliers*, *leverage points*).

McKean (2004) suggests a method that additionally robustifies the rank-based regression against leverage points. This is achieved by introducing robustness weights to the Wilcoxon norm as given in (3.2). The weights are determined such that influential and outlying observations, either in terms of leverage (predictor space) or in terms of unusual responses, are downweighted.

To estimate mixed effects model (3.4) with a rank-based approach that is robust against both outliers in response and predictor space, we replace the two regression steps in Algorithm

3.1 points 1. and 3. by weighted versions. A weighted estimator for the fixed effects $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_w = \arg \min_{\boldsymbol{\beta}} \|\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|_{\varphi},$$

where $\mathbf{W} = \text{diag}(w_{11}, \dots, w_{gn_g}) \in \mathbb{R}^{N \times N}$ is a diagonal matrix of weights w_{ij} , $i = 1, \dots, g$, $j = 1, \dots, n_i$. Analogously, the weighted estimates of the random effects are given by

$$\hat{\mathbf{b}}_{i,w} = \arg \min_{\mathbf{b}} \|\tilde{\mathbf{W}}_i(\mathbf{r}_{i,\text{marg}} - \mathbf{Z}_i\mathbf{b})\|_{\varphi} \quad (3.11)$$

where $\tilde{\mathbf{W}}_i = \text{diag}(\tilde{w}_{i1}, \dots, \tilde{w}_{in_i})$ is a matrix of group specific robustness weights.

The weights w_{ij} can be specified in different ways. We robustify the estimates against leverage points and downweigh observations that are associated with influential predictors. The weights are determined based on a robust estimate of leverage. Let

$$w_{ik} = \min \left\{ 1, \frac{c_p}{\nu_{ik}} \right\}, \quad \text{where } \nu_{ik} = (\mathbf{x}_{ik} - \mathbf{v}_c)' \mathbf{V}^{-1} (\mathbf{x}_{ik} - \mathbf{v}_c), \quad (3.12)$$

for $i = 1, \dots, g$ and $k = 1, \dots, n_i$. The cutoff c_p is selected as the 95%-quantile of the $\chi^2(p-1)$ -distribution, where $p-1$ is the number of columns of \mathbf{X} . \mathbf{V} and \mathbf{v}_c are robust estimates of the covariance and mean of the predictor matrix \mathbf{X} . We use the (fast) minimum covariance determinant estimator (MCD, Hubert, Debruyne, et al. 2018). The robustification against leverage points using the MCD estimator assumes that the predictors are continuous and roughly elliptically distributed. In the case that the predictor matrices include dummy variables, an alternative way to construct weights is offered based on the hat matrix, see e.g. Cantoni and Ronchetti (2001). A short discussion can be found in Section B.5 of the supplement.

For the group-specific weights in (3.11), we proceed analogously, and calculate the leverage based on the matrices \mathbf{Z}_i . Note that the matrix \mathbf{W} needs to be recalculated based on the rescaled observations in every iteration. The group-specific weights in $\tilde{\mathbf{W}}_i$ only need to be determined once, as the matrices \mathbf{Z}_i are not modified in the iteration steps. The detailed estimation procedure is given in Algorithm 3.2.

Algorithm 3.2 (Rank-based fitting of mixed effects models with random slopes and leverage weights). *Starting with $l = 0$, the l 'th iteration of the algorithm is the following:*

1. Calculate the group-specific matrices of robustness weights

$$\tilde{\mathbf{W}}_i = \text{diag}(\tilde{w}_{i1}, \dots, \tilde{w}_{in_i})$$

by applying (3.12) to the matrices \mathbf{Z}_i , $i = 1, \dots, g$. These weights only need to be determined once.

2. Let

$$\hat{\Sigma}_{\mathbf{y}_i}^{(l)} = \begin{cases} \mathbf{I}_{n_i}, & \text{if } l = 0 \\ \Sigma_{\mathbf{y}_i}(\hat{\sigma}_{\varepsilon}^{(l-1)}, \hat{\boldsymbol{\theta}}_b^{(l-1)}), & \text{else,} \end{cases}$$

and

$$\boldsymbol{\nu}^{(l)} = \begin{cases} \mathbf{1}_N, & \text{if } l = 0, \\ \left(\nu_{11}^{(l-1)}, \dots, \nu_{g n_g}^{(l-1)} \right)', & \text{else.} \end{cases}$$

Determine the matrix of weights $\mathbf{W}^{(l)}$ as in (3.12) based on the rescaled matrix of covariates

$$\mathbf{X}_*^{(l)} = (\hat{\Sigma}_{\mathbf{y}_i}^{(l)})^{-\frac{1}{2}} \text{diag}(\boldsymbol{\nu}^{(l)}) \mathbf{X}.$$

Obtain $\hat{\boldsymbol{\beta}}_w^{(l)}$ as a solution to the rank-based regression of

$$\mathbf{y}_{*,w}^{(l)} = \mathbf{W}^{(l)} (\hat{\Sigma}_{\mathbf{y}_i}^{(l)})^{-\frac{1}{2}} \text{diag}(\boldsymbol{\nu}^{(l)}) \mathbf{y}$$

on

$$\mathbf{X}_{*,w}^{(l)} = \mathbf{W}^{(l)} (\hat{\Sigma}_{\mathbf{y}_i}^{(l)})^{-\frac{1}{2}} \text{diag}(\boldsymbol{\nu}^{(l)}) \mathbf{X}.$$

If the model has an intercept α , estimate

$$\hat{\alpha}^{(l)} = T(\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_w^{(l)}).$$

otherwise, $\hat{\alpha}^{(l)} = 0$.

3. Calculate the marginal residuals

$$\mathbf{r}_{i,marg}^{(l)} = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_w^{(l)} - \hat{\alpha}^{(l)} \mathbf{1}_{n_i}.$$

4. Use $\mathbf{r}_{i,marg}^{(l)}$ to obtain predictions $\hat{\mathbf{b}}_{i,w}^{(l)}$ for the random effects \mathbf{b}_i from (3.11) using the weight matrices $\tilde{\mathbf{W}}_i$.

If the model has intercepts a_i , estimate

$$\hat{a}_i^{(l)} = T(\mathbf{r}_{i,marg}^{(l)} - \mathbf{Z}_i \hat{\mathbf{b}}_{i,w}^{(l)}).$$

otherwise, $\hat{a}_i^{(l)} = 0$

5. Obtain the final (conditional) residuals

$$\mathbf{r}_{i,cond}^{(l)} = \mathbf{y}_i - \hat{\alpha}^{(l)} \mathbf{1}_{n_i} - \mathbf{X}_i \hat{\boldsymbol{\beta}}_w^{(l)} - \hat{a}_i^{(l)} \mathbf{1}_{n_i} - \mathbf{Z}_i \hat{\mathbf{b}}_{i,w}^{(l)}.$$

6. Use the $\mathbf{r}_{i,cond}^{(l)}$ and the realizations $\hat{\mathbf{b}}_i^{(l)}$ to estimate the variance components $\boldsymbol{\theta}^{(l)}$ and

$\sigma_\varepsilon^{(l)}$ with the Q_n estimator (see Rousseeuw and Croux 1993).

7. Determine the outlyingness weights $\nu_{ij}^{(l)}$ from the conditional residuals using (3.10).

8. If $l > 0$, check for convergence using the criteria

$$\frac{\|\hat{\beta}_w^{(l)} - \hat{\beta}_w^{(l-1)}\|_2}{\|\hat{\beta}_w^{(l-1)}\|_2} < tol, \quad \text{and} \quad \frac{\|(\hat{\sigma}_\varepsilon^{(l)}, \hat{\theta}_b^{(l)})' - (\hat{\sigma}_\varepsilon^{(l-1)}, \hat{\theta}_b^{(l-1)})'\|_2}{\|(\hat{\sigma}_\varepsilon^{(l-1)}, \hat{\theta}_b^{(l-1)})'\|_2} < tol$$

for a prescribed tolerance tol .

If above criteria are not fulfilled or $l = 0$, increase l by one and return to step 2.

9. Given the final estimates $\hat{\beta}_w$ and $\hat{\mathbf{b}}_{i,w}$, $i = 1, \dots, g$, and if \mathbf{Z}_i is a subset of \mathbf{X}_i , perform the adjustment step.

The approach of downweighting data points with high leverage can cause a loss of efficiency, as “good” leverage points might receive the same weights as “bad” leverage points. To further improve estimation, McKean (2004) suggests to use the so-called *Wilcoxon high-breakdown (HBR)* estimator as an alternative. Based on an initial *least trimmed squares (LTS)* fit, it determines weights that account for outlyingness in both predictor- and response space. The extension of our methodology to this estimator is straightforward.

3.4. Simulation

In the following, we assess the statistical properties and robustness of our two rank-based procedures in simulation studies. We refer to the estimator without additional leverage weights (Algorithm 3.1) as *Rank*, and to the estimator with leverage weights (Algorithm 3.2) as *Weighted Rank*. The proposed procedures are implemented in the package `rankLME` which is available on GitHub (Brune 2024). For comparison, we consider the classic REML estimator as implemented in the R-package `lme4` (Bates et al. 2015). Also, we compare with the SMDM estimator proposed by Koller (2013). This estimator is based on a different robustness concept, namely MM-estimation, and implemented in the R-package `robustlmm` (Koller 2016). All simulations are carried out on a Linux server using R, version 3.6.3 (R Core Team 2020). Replication files for the simulation studies can be found in the GitHub repository (Brune 2024).

For the simulation study, we start by generating an uncontaminated dataset according to model equation (3.4). The entries of the predictor matrices $\mathbf{X}_i \in \mathbb{R}^{n_i \times (p-1)}$ are drawn from the $\mathcal{N}(0, 4)$ distribution, and the random slope matrices $\mathbf{Z}_i \in \mathbb{R}^{n_i \times (k-1)}$ contain the first $(k-1)$ columns of \mathbf{X}_i . We add a fixed intercept α and set $(\alpha, \boldsymbol{\beta}')' = \mathbf{1}_p$; also random intercepts a_i , $i = 1, \dots, g$ are included. The random effects are distributed as $(a, \mathbf{b}')' \sim \mathcal{N}(\mathbf{0}, 0.5^2 \mathbf{I}_k)$. We limit the analysis to $p = 4$ and $k = 2$. Thus, we have regression coefficients $(\alpha, \boldsymbol{\beta}')' = (\alpha, \beta_1, \beta_2, \beta_3)$ and random effects (a, b_1) with scale parameters $\boldsymbol{\theta} = (\theta_0, \theta_1)$. The errors are drawn as $\varepsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_i})$.

In the next step, up to 50% of the observations are contaminated with one of the following outlier types:

- *y- or response outliers*: Observations y_{ij} are modified as either $y_{ij} + s$ (*additive*) or $y_{ij} \cdot s$ (*multiplicative*) for different outlier sizes s
- *leverage points or outlying predictors*: Rows of the predictor matrices are modified as $\mathbf{x}_{ij} + \ell$ (*additive*) or $\mathbf{x}_{ij} \cdot \ell$ (*multiplicative*) where $\mathbf{x}_{ij} = (x_{ij,1}, \dots, x_{ij,p-1}) \in \mathbb{R}^{p-1}$ denotes the j 'th row of the matrix \mathbf{X}_i , and ℓ is the leverage coefficient. The corresponding rows of the matrices \mathbf{Z}_i are adjusted accordingly.

The outlying observations can either be spread *randomly* over the g groups, or be located *sequentially*. In the sequential case, one observation at a time is replaced and we achieve groupwise contamination.

Each setting is repeated for $R = 200$ replications. We report the MSE of the estimates:

$$\text{MSE}(\hat{\beta}_i) = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_i^{(r)} - 1)^2,$$

where $\hat{\beta}_i^{(r)}$ is the estimate of the i 'th component of the coefficient vector $\boldsymbol{\beta}$, $i = 0, \dots, p-1$, obtained in the r 'th simulation run, $r = 1, \dots, R$. β_0 corresponds to the intercept parameter α .

In the interest of clarity, we only report results for randomly located multiplicative y -outliers and leverage points. Further simulation results for additive and sequentially located outliers can be found in the supplemental material (Sections B.3 and B.4).

3.4.1. Bias of the estimates

To explore the bias of the proposed estimators, we repeatedly draw datasets from the null model (i.e. without outliers), and fit the rank-based mixed effects model as described in Section 3.3.

The top row of Figure 3.1 shows boxplots of the coefficient estimates for different group sizes $n_i = \tilde{n}$, $i = 1, \dots, g$, and numbers of groups g . The boxplots are symmetrical and centered around the true values of the coefficients $\alpha = \beta_1 = \beta_2 = \beta_3 = 1$. Even for small total sample sizes, the estimates give no evidence for a bias, the median estimate resembles the true value. Increasing the number of groups g causes a decrease in the estimates' variance. The effect of increasing group sizes n_i is less pronounced. Having many groups seems to be more important than large groups. We observe a higher variance for the coefficients that are associated with random effects (α and β_1). This is the case as the same predictors are used for estimation of both the fixed and the random effects. Therefore, higher uncertainty is introduced to the estimation.

The respective estimates of the variance coefficients (or rather standard deviations) are

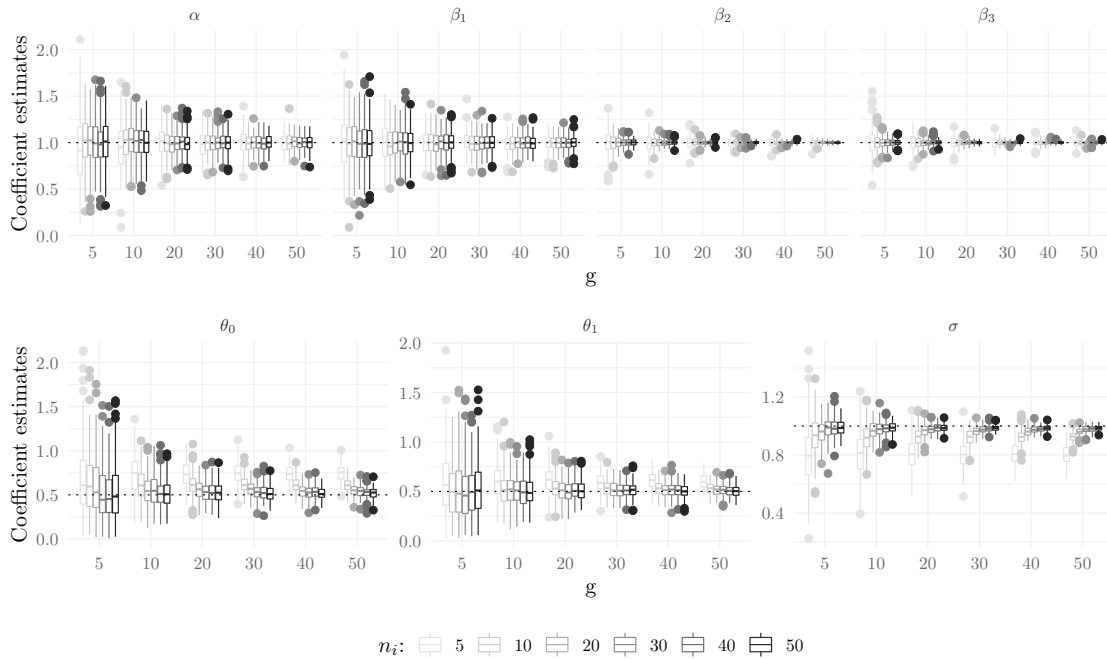


Figure 3.1.: Boxplots of regression coefficient and variance component estimates obtained from the rank-based estimator for different combinations of group sizes n_i and number of groups g (total sample size: $N = g \cdot n_i$)

displayed in the bottom row of Figure 3.1. The number of groups g and the number of observations per group n_i both have strong influence on the accuracy of the estimates. The number of groups g guides the variation of the coefficient estimates. The larger g , the more random effects are available for variance estimation, and thus the variation of the estimates decreases. Our method overestimates θ_0 and θ_1 if the group sizes n_i are small. Due to the two-stage approach for estimation, our method cannot make use of the self-regularizing properties that evolve from direct likelihood-based estimation of the variance parameters (with REML). Especially for small groups, the ratio of observations to parameters for the regression is comparably small. Thus, we tend to overfit the groups and the variance of the estimated random effects \mathbf{b}_i , $i = 1, \dots, g$, is overestimated. The contrary effect can be observed for the scale parameter of the errors, σ , which tends to be underestimated: The good fit in each group can reduce the variance of the conditional residuals. The larger the groups, the more degrees of freedom remain, and thus the better the variance is conserved.

However, our methodology yields accurate estimates of the random effects. These can help with the interpretation of the model and the identification of unusual observations. This is especially helpful if we want to run model diagnostics and interpret the realizations of the random effects (also see the applications in Section 3.5). As soon as n_i is large enough (this study indicates $n_i \geq 20$), all estimates seem to be consistent, regardless of the number of

groups g .

The results for the weighted rank-based estimator are similar and provided in the supplemental material (Section B.1).

3.4.2. Validity of the updating step and efficiency

It is well-known that non-parametric methods are outperformed by parametric methods if the sample sizes are very small and the assumptions of the parametric approaches are valid.

Table 3.1 reports the relative efficiency (Serfling 1980) of our rank-based estimator as opposed to the baseline REML, the SMDM and the weighted rank-based estimator, as well as the efficiency of the rank-based estimator to that of its initial value (i.e. the estimate $\beta^{(0)}$ obtained before iteration).

Table 3.1.: Efficiency of the rank-based estimator vs. the initial value, REML, SMDM and Weighted Rank for $g, n_i \in \{5, 20, 50\}$

g	n_i	initial value	REML	SMDM	Weighted Rank
5	5	0.858	1.159	1.112	0.978
5	20	0.730	1.048	1.046	0.994
5	50	0.712	1.085	1.048	1.003
20	5	0.821	1.134	1.084	0.966
20	20	0.728	1.041	1.017	0.987
20	50	0.726	1.080	1.013	0.988
50	5	0.796	1.163	1.097	0.969
50	20	0.718	1.068	1.031	0.985
50	50	0.768	1.084	1.021	1.000

The rank-based estimator is slightly less efficient than REML and SMDM, but its efficiency approaches that of REML (and SMDM) with increasing total sample size N . The weighted version of our estimator is slightly less efficient due to possible downweighting of useful observations / leverage points. The updating procedure yields a considerable improvement over the initial value. Further simulation studies shown in the supplemental material (Section B.2) confirm that the improvement induced by the updating step is even more pronounced in the presence of outliers.

3.4.3. Behavior in the presence of outliers

In the following, we examine the behavior of the estimates under contamination with randomly located multiplicative response outliers and leverage points. We limit our analyses to the setting with $g = 20$ and $n_i = 20$. Further simulation results are reported in the

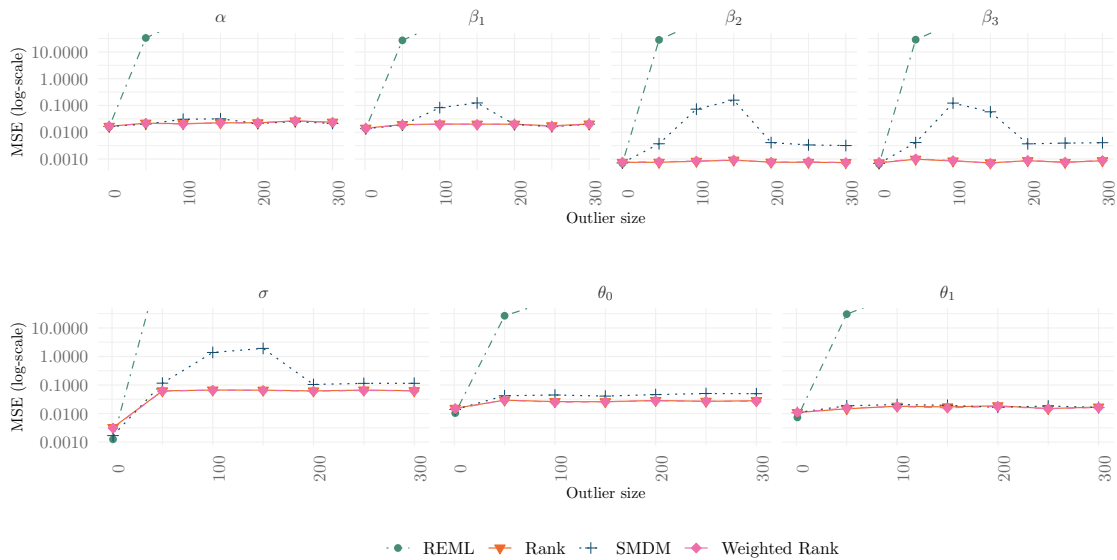


Figure 3.2.: MSE (log-scale) of the SMDM, REML, Rank and Weighted Rank estimators for a 10% proportion of multiplicative response outliers of increasing size, $n_i = g = 20$. The curves for Rank and Weighted Rank are overplotted.

supplemental material.

Response outliers

Figure 3.2 shows the MSEs of the four estimators for the situation of 10% response outliers of increasing size s . The MSEs of the two rank-based estimators and SMDM are fairly constant with increasing outlier size. However, SMDM seems to require a certain outlier size to make sure they are recognized and accounted for by the algorithm (explaining the higher error at medium outlier sizes). At the same time, the REML estimates are severely affected by even small contaminations. Thus, as soon as we suspect outliers to be present in the samples, it is definitely worth applying a robust estimator.

Figure B.3 in the supplemental material examines how the estimator reacts to an increasing proportion of response outliers of size $s = 1000$. The estimates of α and β provided by the SMDM estimator break down at an outlier proportion larger than 20%. The estimates from the rank-based and weighted rank-based estimator stay stable longer (up to 30% contamination).

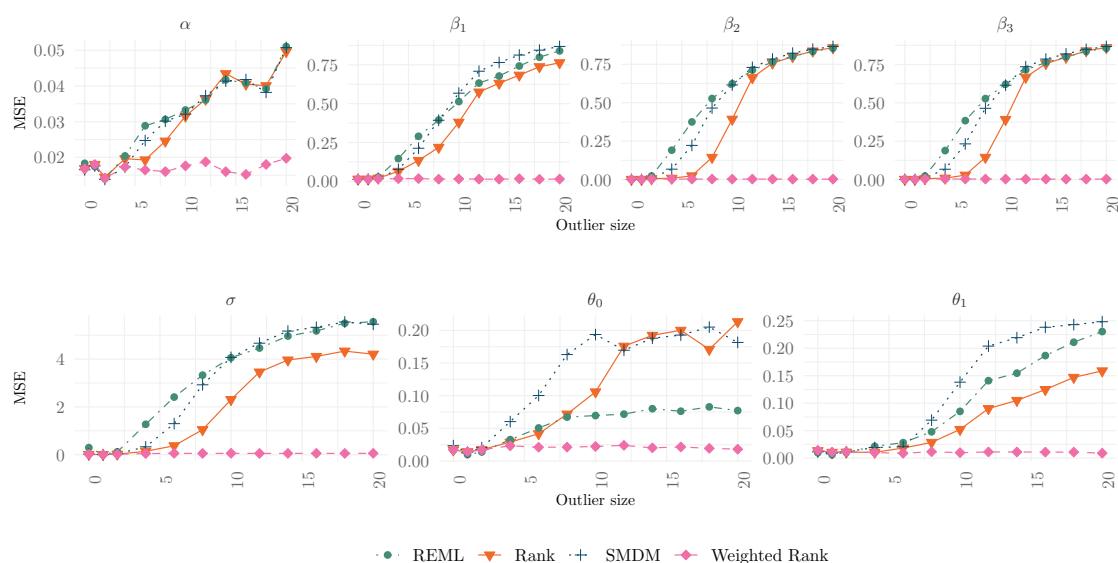


Figure 3.3.: MSE of the SMDM, REML, Rank and Weighted Rank estimators for a 10% proportion of multiplicative leverage points of increasing size, $g = n_i = 20$

Outlying predictors

The MSEs for the coefficient estimates under 10% contamination with leverage points are reported in Figure 3.3. This setting clearly shows the advantage of the leverage weighted rank-based estimator: While REML, SMDM and the non-weighted rank-estimator are severely distorted by the leverage points, the weighted rank-based estimator remains largely unaffected. Its coefficient estimates remain constant with increasing outlier size ℓ .

Figure B.7 in the supplemental material examines how the estimator reacts to an increasing proportion of leverage points with $\ell = 100$. The estimates of the leverage weighted estimator are stable up to an outlier proportion of at least 20%, while the other three estimators are distorted even for small outlier proportions.

3.5. Applications

In the following section we illustrate the performance of the (weighted) rank-based estimation method in three different applications, parts of which are mixtures of real data analysis and introduction of artificial outliers. Especially, we illustrate different diagnostic tools that may help with identifying outliers or outlying / unusual groups in the data. Replication files for the data analyses can be found in the GitHub repository (Brune 2024).

3.5.1. Sleepstudy data

We analyze the *sleepstudy* dataset (Belenky et al. 2003) as included in the `lme4` R-package (Bates et al. 2015). It consists of data from 18 subjects that were exposed to sleep deprivation and had to complete a reaction test each day. Thus, we expect the average reaction time to decrease throughout the experiment, but each subject might react differently to the sleep deprivation.

We fit a mixed effects model with both random and fixed intercept and slope in time. The model equation for the reaction time of subject i ($i = 1, \dots, 18$) after t ($t = 0, \dots, 9$) days of sleep deprivation is given by

$$\text{Reaction}_{i,t} = \alpha + \beta_1 \cdot t + a_i + \mathbf{b}_i \cdot t + \varepsilon_{i,t} \quad (3.13)$$

We contaminate the dataset in two different ways:

1. Modify the third observation of each individual by multiplying the reaction times by three (*outlying observations*). Outliers like this could e.g. be caused by an error in the measurement instrument that was used on that day.
2. Modify the first group (individual ‘308’) by multiplying the reaction times by three (*outlying group*). The subject here could happen to be very slow, but this fact should not affect the population average (i.e. fixed effects) too strongly.

Since the predictor in model (3.13) is exposure time, we do not expect leverage points. Thus, we compare the estimates from our procedure without leverage weights (*rank*, Algorithm 3.1) to the standard REML estimator. The original coefficient estimates and their ratios to the estimates from the two contamination settings are displayed in Table 3.2. For the rank-based model, the estimates of α and β_1 remain stable in all contamination settings, they deviate by at most 2% from the baseline estimates.

Table 3.2.: Coefficients estimated by REML and the rank-based estimator for the unmodified sleepstudy data and the ratio to coefficient estimates for the two outlier settings; boldface numbers indicate a relative change of more than 10%

Setting	Method	α	β_1	θ_0	θ_1	σ
baseline	rank	252.10	10.63	31.28	6.56	16.56
	REML	251.41	10.47	25.05	5.99	25.57
1.	rank	1.02	0.99	1.07	1.23	1.26
	REML	1.29	0.67	-	-	7.14
2.	rank	1.01	1.01	1.20	1.00	1.06
	REML	1.11	1.23	4.78	2.39	1.59

REML returns a singular fit in the *outlying observations* setting, the imposed random effects

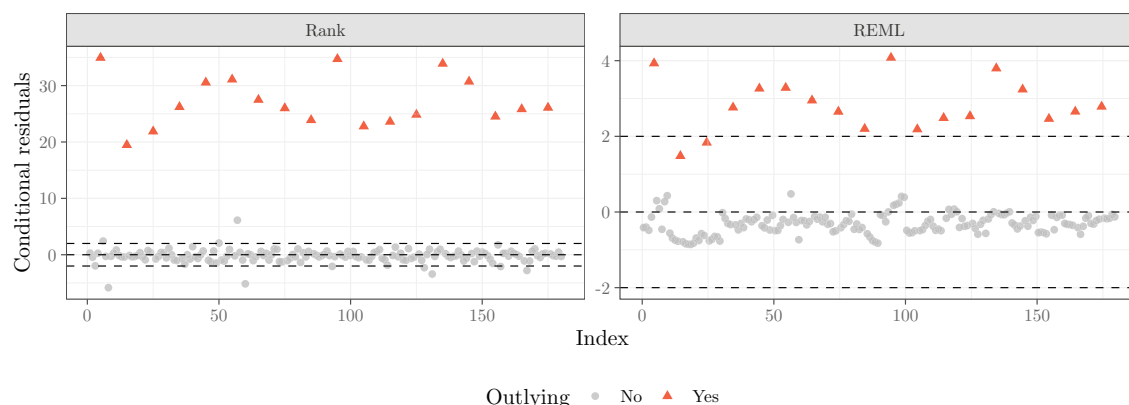


Figure 3.4.: Conditional residuals for the sleepstudy example in the outlying observations setting; REML and rank-based estimator

structure cannot be fitted with `lme4`. As a consequence, the error variance σ is overestimated. The estimate returned by the rank-based method remains more stable. The residuals from REML are strongly skewed (see Figure 3.4). Due to the outlying observations in each group, the REML regression lines are shifted upwards and a majority of the residuals is less than zero. Two of the outlying values are not recognized as such and fall within the cutoff of 2. For the rank-based method, the residuals are still centered around zero. The outlying values are recognized correctly. A few more observations are marked as conspicuous and could require further attention. However, they already fell outside of the cutoff area without the artificial outliers.

In the *outlying group* setting, we observe problems with the variance structure imposed by the model: REML severely overestimates the variance components due to the outlying random effect. As an indicator whether the imposed variance structure, i.e. the random effects structure, is reasonable, Lesaffre and Verbeke (1998) suggest to report the following norm

$$d_i = \left\| I_{n_i} - \hat{\Sigma}_{\mathbf{y}_i}^{-1/2} \mathbf{r}_{i,\text{marg}} \mathbf{r}'_{i,\text{marg}} \hat{\Sigma}_{\mathbf{y}_i}^{-1/2} \right\|^2 \quad (3.14)$$

for $i = 1, \dots, g$. d_i may be interpreted as a residual which measures how well the covariance structure of the observations \mathbf{y}_i is captured by the model. Thus, d_i should be small if the variance structure is captured well, and should produce large values if the variance structure in one of the groups does not match that imposed by the model. A plot of the group index against the variance diagnostic (3.14) yields Figure 3.5. The outlying group cannot be recognized in the case of REML, the magnitudes of the d_i 's are all fairly similar. The outlying individual (308) does not stand out. The d_i 's calculated from the rank-based fit clearly capture the outlying group.

The examples of contamination here show that outlying values can be masked by the outcome of the standard REML estimation approach. The standard diagnostic methods fail in this

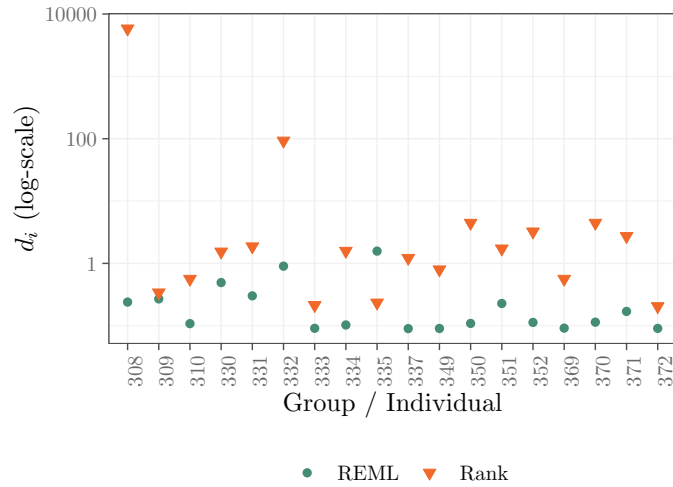


Figure 3.5.: Variance diagnostic (3.14) for the sleepstudy example in the outlying group setting; REML and rank-based estimator

case. We cannot reliably detect the outlying observations or groups. Opposed to this, our robust rank-based estimation method allows to clearly identify the unusual observations and thus enables further analysis.

3.5.2. Accelerated aging experiments for photovoltaic module data

The ADVANCE! project analyzes the aging behavior of photovoltaic (PV) modules (Berger et al. 2021). Different modules were exposed to various climatic conditions in accelerated aging experiments. Such experiments are used as the real-time assessment of material degradation is not possible due to the long time frames in real life. Throughout the experiments, chemical and electric measurements were taken. We're especially interested in modeling the *power at maximum power point* (P_{MPP}) over time, and as a function of other degradation indicators. The goal is to identify degradation pathways and understand the influence of treatment with different climatic settings on changes in material and power loss.

Pairwise comparison of degradation rates

We compare the results from the two climatic settings *Moderate 2* and *Moderate 5*. In both settings three modules were treated with damp-heat (i.e. high temperature and humidity) and irradiance for 1 200 hours. The modules in the Moderate 5-group were additionally exposed to temperature cycles. We fit the following model:

$$(P_{MPP})_{it} = 1 + \beta_1 \cdot \text{Ramp}(t) + \beta_2 \cdot t^2 + \beta_3 \cdot t^2 \cdot \mathbf{1}\{\text{Moderate 5}\} + b_i \cdot t^2 + \varepsilon_{i,t}.$$

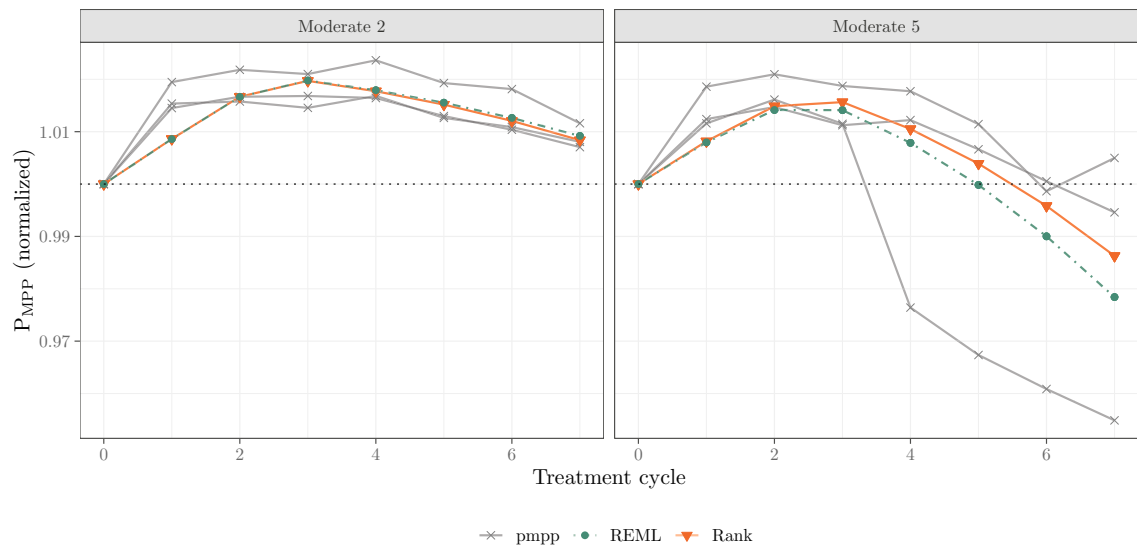


Figure 3.6.: Data and fitted values for pairwise comparison of the *Moderate 2* and *Moderate 5* aging settings

β_1 captures an initial power increase, β_2 the overall aging, and β_3 the additional aging induced by the treatment with temperature cycles. The intercept / initial power is fixed at 1. The random effects b_i account for the module specific degradation rates over time.

As can be seen from Figure 3.6, one module broke during the aging procedure. Thus we expect the REML estimate of β_3 to be biased downwards trying to account for the strong power loss of the faulty module. Since we are interested in modeling the continuous material wear, and not the abrupt degradation caused by material failures, the robust estimate gives a more reliable characterization of the degradation rate.

The resulting estimates for $(\beta_1, \beta_2, \beta_3)'$ are $(0.022, -0.014, -0.022)'$ for the rank-based method, and $(0.022, -0.013, -0.031)'$ for REML. Thus, the degradation for *Moderate 5* captured in β_3 is estimated to be about 40% larger by REML than by the rank-based method.

Degradation modeling with spectral measurements

In this analysis, we connect the power loss in P_{MPP} with the material aging measured using fitted spectra obtained from Fourier-transform infrared (FTIR) spectroscopy. The changes in the spectra over time indicate changes in the chemical composition of the materials. The spectral measurements are observed as functions which are fitted automatically to locate peaks and calculate their area. These fits can go wrong, causing outlying predictors / leverage points in the resulting data points. We model the degradation for two different

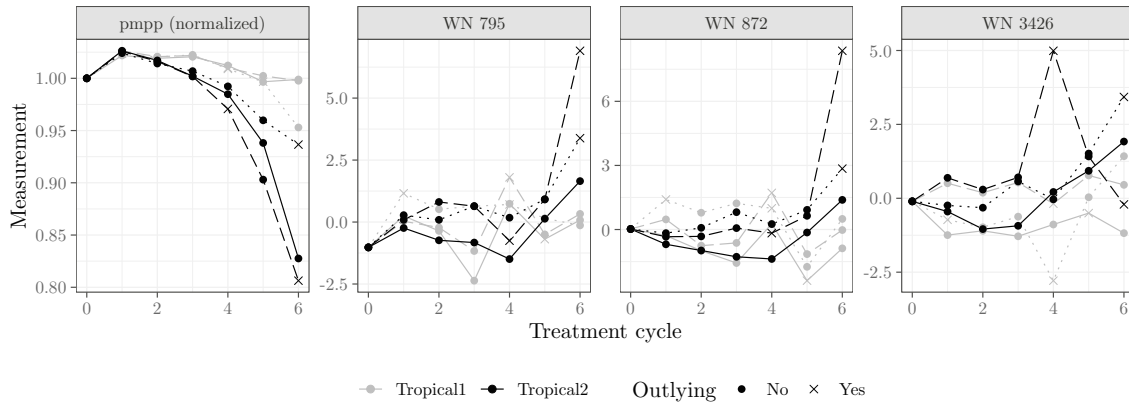


Figure 3.7.: Peaks from FTIR spectra and P_{MPP} for aging settings Tropical 1 and Tropical 2; points with leverage weight (3.12) < 1 are marked by \times

climatic settings: *Tropical 1* and *Tropical 2* treated with 85°C / 90°C and 85% / 90% humidity respectively. We aim to explain the power loss over time caused by the treatment with higher temperature and humidity for the Tropical 2-modules through the changes in three peaks areas extracted from the spectral measurements, namely those at wavenumbers (WN) 795, 872 and 3426 cm^{-1} . The standardized peaks and P_{MPP} are shown in Figure 3.7. As can be seen, the peak areas tend to increase with increasing exposition time. Thus, their influence on the decreasing P_{MPP} should be negative. We fit the model

$$(P_{MPP})_{it} = 1 + \beta_1 \cdot \text{Ramp}(t) + \beta_2 \cdot \text{WN } 795_{it} + \beta_3 \cdot \text{WN } 872_{it} + \beta_4 \cdot \text{WN } 3426_{it} + b_i \cdot t^2 + \varepsilon_{it}$$

using REML and our two rank-based methods. The observations that are downweighted by the weighted rank-based estimator are marked as crosses in Figure 3.7. Hence, the algorithm successfully recognizes the outlying predictors, especially for the peak area at wavenumber 3426 cm^{-1} .

The fitted coefficients $(\beta_1, \beta_2, \beta_3, \beta_4)'$ are $(0.0428, 0.0006, -0.0051, 0.0007)'$ for REML and $(0.0324, -0.0004, -0.0028, -0.0023)'$ for the weighted rank-based method. The signs of the estimates for β_2 and β_4 differ between the two estimation methods. The sign is positive for REML, but negative for our weighted rank-based method. As noticed before, the shapes of the curves indicate that a negative sign might be more realistic. For β_4 , this change in sign is probably caused by the module marked by dashed lines in the WN 3426 cm^{-1} series. The large leverage points lead the non-robust model to believe the peak has positive influence on the degradation. Accounting for the leverage points with additional weights as we propose for the weighted rank-based procedure is essential. Our rank-based procedure without weights has similar problems as REML.

3.6. Summary and outlook

The present work proposes a new approach to the rank-based estimation of mixed effects models. It extends existing methodology for mixed effects models with random intercepts with the possibility to model random slopes. The estimates obtained from the rank-based method are insensitive against outlying response values. In addition, the proposed method yields robustness in the presence of leverage points in the fixed or random effects predictor matrices. Our algorithm converges fast and can be implemented efficiently. An implementation in R is available on GitHub (Brune 2024).

The results of our simulation studies indicate that the bias of the estimates is insignificant: The estimates of the regression coefficients show no evidence of being biased. For small sample sizes, the variance of the random effects, especially of the random intercepts, tends to be slightly overestimated. This is inherent to the iterative nature of the algorithm. A possible mitigation is the derivation of correction factors for the estimation of the random intercepts. We plan to explore this in detail in future research.

Under normality, the estimates obtained from our methodology compare well with those obtained from the classic REML estimation. At the same time, the rank-based approach has advantages in situations where the error- or random effects distributions deviate from normality. Although we fit more parameters than for the standard estimation approaches based on REML, we still gain in efficiency due to our updating procedure. Accounting for the covariance structure reduces the variance of the fixed effects estimator $\hat{\beta}$. The applications show that the resulting estimates enable model diagnostics and help with the identification of unusual observations or groups in the data.

There are different lines of research that can be pursued from this point. One is to explore the consistency and asymptotic distribution of our estimators in theory. This would allow for formal inference on the fixed effects estimates. As the methodology is not yet able to estimate more complex random effects structures such as nested and crossed effects, it would be interesting to extend the model in this direction. An alternative to the rank-based estimators based on Jaeckel's dispersion function applied in this work could be offered by maximum rank correlation estimators, first proposed by Han (1987). The robustness properties of these estimators have been examined in detail in Alfons et al. (2017).

Acknowledgments

This work was conducted as part of the Austrian Energy Research Program project ADVANCE! (No. 881133), funded by the Austrian Climate and Energy Fund and the Austrian Research Promotion Agency (FFG). We thank the referees for their comments and suggestions that helped us improve the paper.

4 | Robust modeling of repeated functional measurements

This chapter is based on joint work with Una Radojčić, Sonja Greven and Peter Filzmoser. Supplemental information to this chapter can be found in Appendix C, p. 145ff.

4.1. Introduction

Classical functional data analysis (FDA) is concerned with the study of independently sampled random functions. However, there are various settings where functional data is sampled under assumptions that deviate from this independence, e.g. functional time series and longitudinal functional data. Koner and Staicu (2023) provide a comprehensive overview of the methodology for what they refer to as *second-generation functional data*.

This work is concerned with functional data sampled in longitudinal or repeated measurement designs, i.e. the functions are repeatedly observed on different observational units. Examples of such datasets include fertility or mortality curves for different countries over time (Chen and Müller 2012), spectral measurements capturing the aging of materials for different samples (Brune, Ortner, Eder, et al. 2023), but also various types of functional data gathered in medical studies over time (Greven et al. 2010). A specific trait of this type of data is that, while each function is usually sampled on a dense grid, the time component might be sampled sparsely and irregularly. An example of a longitudinal functional observation is given in Figure 4.1. It showcases the mortality rate as a function of age for Portugal from 1953 to 2023 (in 5-year steps). All curves show an inherent structure with increased mortality of infants and elderly people. The overall mortality rates show a clear trend and decrease with time.

This form of functional data can be interpreted as realization of a bivariate stochastic process $X(s, t)$, $s \in \mathcal{S}, t \in \mathcal{T}$, $\mathcal{S}, \mathcal{T} \subset \mathbb{R}$. For the sake of interpretability, we can also think of X as a function $X : \mathcal{T} \rightarrow \mathcal{L}^2(\mathcal{S})$, where the fixed realization of X at time t is denoted by $X(\cdot, t)$ and is a square-integrable random function, element of the set of all square-integrable functions $\mathcal{L}^2(\mathcal{S})$ on \mathcal{S} . We assume the process X has a mean function $\mu : \mathcal{S} \times \mathcal{T} \rightarrow \mathbb{R}$ and covariance

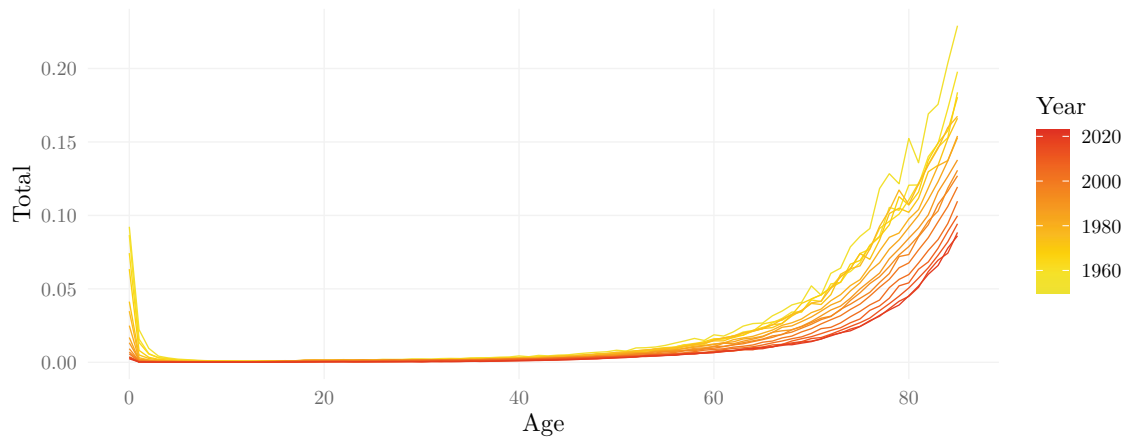


Figure 4.1.: Yearly age-specific mortality rates for Portugal between 1953 and 2023, ages 0 to 86.

function

$$c(\{s, t\}, \{s', t'\}) = \mathbb{E}(X(s, t)X(s', t')) - \mu(s, t)\mu(s', t'). \quad (4.1)$$

In the example in Figure 4.1, s would correspond to age, and t to the year.

Due to the infinite-dimensional nature of functional data, one of the primary goals in FDA is to represent the functional data in a lower-dimensional space. A common approach to achieve this is functional principal component analysis (FPCA, see e.g. Ramsay and Silverman 2005). FPCA leverages the Karhunen-Loève (KL) expansion, which allows the decomposition of the stochastic process into projections onto orthonormal basis functions. The KL expansion is particularly valuable because it offers the best finite rank approximation of the functional data in terms of minimizing the mean squared error. This optimality property makes the KL expansion a powerful tool for capturing the most significant features of the data while reducing dimensionality.

For two-dimensional processes, the KL expansion is given by

$$X(s, t) = \sum_{r=1}^{\infty} \alpha_r \gamma_r(s, t).$$

The bivariate functions $\{\gamma_r\}_{r \geq 1}$ are the eigenfunctions of the covariance operator of X , and form an orthonormal basis of $\mathcal{L}^2(\mathcal{S} \times \mathcal{T})$. $\{\alpha_r\}_{r \geq 1}$ are uncorrelated random scores given by $\alpha_r = \int_{\mathcal{S}} \int_{\mathcal{T}} \gamma_r(s, t) X(s, t) dt ds$ for $r \geq 1$. A downside of this two-dimensional representation is that it treats the dynamics in s and t symmetrically, and doesn't allow for separate analysis of the two domains. Furthermore, it requires estimation of the four-dimensional covariance surface (4.1), which is computationally very demanding.

Different approaches to avoid this complex parameterization have been proposed in the literature. Greven et al. (2010) propose a parameterization based on a longitudinally varying functional mixed model (*longitudinal FPCA*). Here, the dynamics in \mathcal{T} are treated as linear. Chen and Müller (2012) expand X using two layers of KL expansions (*double FPCA*) to allow for non-linearities in \mathcal{T} . This approach requires estimating the covariance functions $c(\{s, t\}, \{s', t\})$ in (4.1) for each t and can become computationally intensive as well. In order to overcome the computational limitations, Park and Staicu (2015) and Chen, Delicado, et al. (2017) propose to instead project the process X onto a basis that is invariant in \mathcal{T} . This results in a decomposition of the form

$$X(s, t) - \mu(s, t) = \sum_{k=1}^{\infty} \xi_k(t) \phi_k(s). \quad (4.2)$$

While being easier to handle from a computational point of view, this representation also offers advantages regarding interpretation. Instead of requiring the analysis of two-dimensional surfaces, it allows for a separate treatment and analysis of the dynamics in \mathcal{S} and \mathcal{T} . At the same time, many optimality traits of the standard KL expansion are preserved. Both Park and Staicu (2015) and Chen, Delicado, et al. (2017) provide an estimation method to fit the model under different assumptions about the error structure of the observed data.

In any real-world data collection, there is a possibility of encountering outlying values. These outliers may arise due to measurement errors, such as miscalibrated devices, or from atypical events, like unusual time periods (e.g., wars or crises). In robust statistics, a goal is to derive estimators that are insensitive to outlying observations in the data. The resulting estimates should also allow to identify and analyze the outliers. This paper aims to fill a gap in the literature by proposing a robust estimation method for longitudinal FDA based on the model (4.2). The method aims to reliably represent the data despite outliers, while also identifying outlying curves and subjects for further analysis. The resulting diagnostic plots help interpret the results and can provide valuable insights. The flexible estimation framework allows to incorporate prior structural knowledge into the analysis.

The paper is structured as follows: Section 4.2 introduces the methodology and provides an overview of the parameter estimation procedure. The details on the robust parameter estimation are given in Section 4.3. Section 4.4 introduces tools for outlier diagnostics. A performance evaluation of the proposed methodology is shown in simulations (Section 4.5) and in examples (Section 4.6), and the final Section 4.7 concludes.

4.2. Methodology

4.2.1. Marginal functional principal component analysis

Given $\mathcal{S}, \mathcal{T} \subset \mathbb{R}$, $X : \mathcal{T} \rightarrow \mathcal{L}^2(\mathcal{S})$ is a stochastic process, where the value at time t is denoted by $X(\cdot, t)$ and is a function of $s \in \mathcal{S}$. We refer to \mathcal{S} as *spatial domain* and to \mathcal{T} as *time domain*. As X will be later used to model deviations from the overall mean, we assume X

to have mean zero and continuous covariance function

$$c(\{s, t\}, \{s', t'\}) = \mathbb{E}(X(s, t)X(s', t')). \quad (4.3)$$

As a reminder, following Chen, Delicado, et al. (2017), the process X is represented by a decomposition $X(s, t) = \sum_{k=1}^{\infty} \xi_k(t)\phi_k(s)$. The basis functions ϕ_k capture the dynamics in the frequency domain \mathcal{S} and are obtained from an eigendecomposition of the *marginal covariance function*

$$\Sigma(s, s') = \int_{\mathcal{T}} c(\{s, t\}, \{s', t\}) dt. \quad (4.4)$$

The resulting eigenpairs $\{\lambda_k, \phi_k\}_k$ are later referred to as *marginal eigenvalues and functional principal components*. They are optimal in the sense that they minimize the mean squared average reconstruction error, where the averaging is with respect to the time domain, i.e.

$$\mathbb{E} \left(\int_{\mathcal{T}} \|X(\cdot, t) - \sum_{k=1}^K \langle X(\cdot, t), \phi_k \rangle_{\mathcal{S}} \phi_k\|_{\mathcal{S}}^2 dt \right).$$

$\langle \cdot, \cdot \rangle_{\mathcal{S}}$ denotes the inner product in $\mathcal{L}^2(s)$ (see Theorem 1 in Chen, Delicado, et al. 2017).

The time-dynamics are captured by functions $\{\xi_k\}_k$, formed by the scores on the corresponding components:

$$\xi_k(t) = \langle X(\cdot, t), \phi_k \rangle_{\mathcal{S}} = \int_{\mathcal{S}} \phi_k(s) X(s, t) ds. \quad (4.5)$$

Especially, the *score functions* ξ_k fulfill

$$\mathbb{E}(\xi_k(t)) = 0 \quad \forall t \in \mathcal{T}, \quad \text{and} \quad \mathbb{E}(\langle \xi_k, \xi_{\ell} \rangle_{\mathcal{T}}) = \int_{\mathcal{T}} \xi_k(t) \xi_{\ell}(t) dt = \lambda_k \delta_{k\ell}. \quad (4.6)$$

The score functions are random functions and thus can be treated as functional data as well. We denote their covariance functions by

$$\gamma_k(t, t') = \mathbb{E}(\xi_k(t)\xi_k(t')), \quad \text{for } k \geq 1,$$

with corresponding eigenpairs $\{\eta_{k\ell}, \psi_{k\ell}\}_{\ell}$ for each $k \geq 1$. The continuity of the covariance function c in (4.3) implies continuity of γ_k , $k \geq 1$. Thus, the score functions ξ_k admit KL expansions

$$\xi_k(t) = \sum_{\ell=1}^{\infty} \zeta_{k\ell} \psi_{k\ell}(t), \quad \text{for } k \geq 1.$$

4.2.2. Model estimation

Without loss of generality, let $\mathcal{S} = \mathcal{T} = [0, 1]$. Consider observed data from n subjects. For each subject i , $i = 1, \dots, n$, we observe n_i functions at random time points $t_{ij} \in \mathcal{T}$, $j = 1, \dots, n_i$. The sampling time points are assumed to be uniformly distributed on the interval $\mathcal{T} = [0, 1]$ and may be sparse. Each of the n_i functions is observed on a finite, high-resoluted grid s_1, \dots, s_R . The whole data set is then

$$[\{(t_{ij}, Y_i(s_r, t_{ij})), r = 1, \dots, R\}, j = 1, \dots, n_i]_{i=1, \dots, n}.$$

For simplicity of notation, we drop the index r from the frequency domain. Each functional observation $Y_i(\cdot, t_{ij})$ is represented as

$$Y_i(s, t_{ij}) = \mu(s, t_{ij}) + X_i(s, t_{ij}),$$

where $\mu : [0, 1]^2 \rightarrow \mathbb{R}$ is a fixed mean function and X_i is a realization of the (centered) process as in Section 4.2.1.

We aim to fit a decomposition of form (4.2) to the processes Y_i . In practice, the decomposition (4.2) is truncated to K components, resulting in the representation

$$Y_i(s, t_{ij}) \approx \mu(s, t_{ij}) + \sum_{k=1}^K \xi_{ik}(t_{ij}) \phi_k(s). \quad (4.7)$$

The cutoff K for the number of components in (4.7) can, e.g., be determined based on the proportion of variance explained (*pve*). Similar to the estimation procedures in Park and Staicu (2015) and Chen, Delicado, et al. (2017), the model estimation is performed in three steps:

- Step 1** Estimation of the mean function μ by employing a bivariate smoothing algorithm.
- Step 2** Estimation of the marginal covariance function (4.4) and the eigenpairs $\{\lambda_k, \phi_k\}_k$.
- Step 3** Estimation of the scores $\xi_{ik}(t_{ij})$, followed by smoothing to obtain estimates of the score functions $\xi_{ik}(t)$ at all points $t \in \mathcal{T}$ (non-parametrically or parametrically).

In order to achieve a robust estimate, we propose to replace the estimation components suggested in Park and Staicu (2015) and Chen, Delicado, et al. (2017) by robust counterparts. The proposed estimation procedure is presented in detail in the following Section 4.3.

4.3. Robust estimation procedure

This section outlines the novel robust model estimation procedure for steps 1–3 in detail.

4.3.1. Estimation of the mean function

The estimation of the mean function μ is not the main focus of the paper. The algorithm that can be employed is strongly dependent on the available data, on potential covariates, and the specified structure of the mean. Examples for specifications include

- (1) $\mu(s, t) = \delta(s)$ for a smooth function $\delta(s)$, i.e. the mean function is independent of the longitudinal dynamics.
- (2) $\mu(s, t) = \nu(s) + \delta(t)$ where $\nu(\cdot)$ and $\delta(\cdot)$ are smooth univariate functions. This structure implies that the dynamics in s and t are separable.
- (3) When lacking any further structural assumptions, the mean function can be fit using a bivariate smoothing algorithm,

(see also Koner and Staicu 2023). For the data applications in this paper, we use the parameterizations (1) and (2), i.e. additive models. An algorithm to estimate such models robustly from all observations under a working independence assumption is presented in Boente, Martínez, et al. (2017). For more complex mean functions requiring a bivariate fit, robust models like quantile smoothing splines (Koenker et al. 1994) or robust thin-plate splines (Kalogridis 2023) can be used.

Using the resulting estimate $\hat{\mu}$, we can center the samples $Y_i(\cdot, t_{ij})$. The resulting centered observations are denoted by

$$Y_i^C(s, t_{ij}) = Y_i(s, t_{ij}) - \hat{\mu}(s, t_{ij}), \quad i = 1, \dots, n, j = 1, \dots, n_i, \quad (4.8)$$

and are used in the further steps of the algorithm.

4.3.2. Estimation of the marginal covariance function and raw scores

Park and Staicu (2015) and Chen, Delicado, et al. (2017) estimate the marginal covariance function based on the sample covariance from all observed functions $Y_i(\cdot, t_{ij})$ for $i = 1, \dots, n, j = 1, \dots, n_i$, treating them as if they were i.i.d. data. As the sample covariance is prone to outliers, a robust alternative should be used in order to estimate the model in a way that is insensitive to outliers. Due to the functional form of the data, the estimator needs to be applicable in a high-dimensional setting.

The estimator of choice is the newly proposed Minimum Regularized Covariance Trace (MRCT) estimator which Oguamalam et al. (2024) designed primarily for the dense functional data. The MRCT covariance estimator is essentially a trimmed covariance estimator that favors subsets of fixed size h that are the most central w.r.t. the corresponding covariance. The centrality of the point is measured by the α -Mahalanobis distance, a generalization of the Mahalanobis distance to the \mathcal{L}^2 -space introduced by Berrendero et al. 2020. The approach results in a concentration-step (Rousseeuw and Driessen 1999) type algorithm, in which the traditional Mahalanobis distance is replaced by the α -Mahalanobis distance. Additionally, as

α -Mahalanobis distance is defined as the reproducing kernel Hilbert space (RKHS) distance between the smooth RKHS representatives of the given observation and the corresponding mean, a certain amount of internal smoothing defined by a regularization parameter $\alpha > 0$ is integrated within the procedure. Therefore, the estimator successfully handles datasets with a high number of observed time points without data preprocessing. This is particularly important in this context, since preprocessing via smoothing can mask the effect of outliers; for more details, see, e.g., Hubert, Rousseeuw, et al. 2015. For more details on the MRCT covariance estimator and the subsequent outlier detection, see Oguamalam et al. 2024.

Alternatives to MRCT include, among others, the minimum regularized covariance determinant estimator (MRCD, Boudt et al. 2020), or the pairwise spatial sign (PASS) covariance as suggested in Wang, Liu, et al. (2022). However, in an extensive simulation study from Oguamalam et al. 2024, both methods performed inferiorly to MRCT. Alternatively, an option is to use the estimator proposed in Boente and Salibián-Barrera (2021), which is also applied in step (3) of our estimation algorithm. However, this comes with a very high computational cost in high dimensions.

We denote the estimated marginal covariance function at the observed grid points by $\hat{\Sigma}(s_r, s_{r'})$, $r, r' = 1, \dots, R$. Estimates of the marginal functional principal components $\{\lambda_k, \phi_k\}_k$, denoted by $\{\hat{\lambda}_k, \hat{\phi}_k\}_k$, can then be obtained using standard methods for dense functional data (Ramsay and Silverman 2005).

4.3.3. Smoothing the score functions

Estimates of the scores are obtained by approximating the integral in (4.5). Since the index s is sampled densely, it is justified to use a mean as an approximation:

$$\tilde{\xi}_{k,ij} = \frac{1}{R} \sum_{r=1}^R Y_i^C(s_r, t_{ij}) \hat{\phi}_k(s_r). \quad (4.9)$$

We refer to those as *raw scores* in the following. The goal is to obtain robust estimates of the score functions from the raw scores. Outliers in the score functions can occur in two forms, as

- (A) *outlying subjects*, where the score functions ξ_k for the respective individual will deviate from the majority, and in the form of
- (B) *partial contamination*, where outlying observations cause spikes and structural deviations that differ from the actual functional form of the underlying ξ_k .

Partial contamination can be interpreted as a notion of cellwise outliers (Raymaekers and Rousseeuw 2024). The smoothing method that yields the score functions $\hat{\xi}_{ik}$ from the raw scores $\tilde{\xi}_{k,ij}$ should be robust against both outlier types (A) and (B).

Together with their measurement time points t_{ij} , the raw scores form a dataset of (potentially

sparse and irregularly sampled) functional data

$$\{(t_{ij}, \tilde{\xi}_{k,ij}); j = 1, \dots, n_i\}_{i=1, \dots, n},$$

with covariance function γ_k , $k = 1, \dots, K$. As outlined in Section 3.3, a non-parametric model can be obtained by representing the smooth score functions by truncated KL expansions. For $k = 1, \dots, K$ and each subject $i = 1, \dots, n$,

$$\hat{\xi}_{ik}(t) = \sum_{\ell=1}^{L_k} \hat{\zeta}_{i,k\ell} \hat{\psi}_{k\ell}(t), \quad (4.10)$$

where the cutoff L_k can again be selected using pve. In order to estimate the eigenfunctions and scores in (4.10) in a way that is insensitive to outliers of both type (A) and (B), while at the same time able to deal with sparse and irregular sampling grids, we apply a modified version of the estimation method proposed in Boente and Salibián-Barrera (2021). By construction the method is robust to outliers of type (A). By introducing a spike detection step and weights, we additionally robustify it against cellwise outlying scores $\tilde{\xi}_{k,ij}$, i.e. outliers of type (B). The methodology relies on Proposition 1 in their work (given as Proposition C.1 in Supplement C.1.1). Under the assumption that the score functions ξ_{ik} are realizations of an elliptically distributed process $\xi_k : \Omega \rightarrow \mathcal{L}^2(\mathcal{T})$, the proposition establishes the following relationship:

$$\mathbb{E}(\xi_k(t) | \xi_k(t')) = \frac{\gamma_k(t, t')}{\gamma_k(t', t')} \xi_k(t') =: \beta_k(t, t') \xi_k(t'). \quad (4.11)$$

Thus, the conditional expectation of $\xi_k(t)$ given $\xi_k(t')$ is a linear function of $\xi_k(t')$, and the coefficient of this regression is defined by the covariance function. This serves as a basis for a robust estimation method. It proceeds in two steps for each $k = 1, \dots, K$:

Step 3.1 Estimate the diagonal of the covariance function as $\hat{\gamma}_k(t, t)$ for each t on a grid $\{t_1, \dots, t_T\} \subset \mathcal{T}$ using a robust M-scale estimator for the variances of the raw scores $\tilde{\xi}_{k,ij}$, $i = 1, \dots, n$, $j = 1, \dots, n_i$.

Step 3.2 Estimate the off-diagonal elements as $\hat{\gamma}_k(t, t')$ for $t' \neq t$, $t, t' \in \{t_1, \dots, t_T\} \subset \mathcal{T}$, using locally weighted M-regressions that estimate the coefficients $\beta_k(t, t')$ in (4.11) by $\hat{\beta}_k(t, t')$. The off-diagonal elements are then given by $\hat{\gamma}_k(t, t') = \hat{\beta}_k(t, t') \hat{\gamma}_k(t', t')$. Symmetry of the resulting estimated covariance function is ensured by averaging over $\hat{\gamma}_k(t, t')$ and $\hat{\gamma}_k(t', t)$ for each pair $t \neq t'$.

Both estimation steps rely on local smoothing involving a bandwidth parameter h . In practice, h can be selected by cross-validation. The detailed estimation algorithm is described in Supplement C.1.1. We can then obtain estimates of the eigenfunctions $\psi_{k\ell}$ in (4.10) using standard eigenanalysis of the estimated covariance functions $\hat{\gamma}_k$. Estimates of the corresponding scores $\zeta_{i,k\ell}$ are then given by the best linear unbiased prediction estimator

based on Proposition C.1(c). We let

$$\hat{\zeta}_{i,k\ell} = \hat{\eta}_{k\ell} \hat{\psi}'_{i,k\ell} (\Gamma_{i,k} + \rho \mathbf{I}_{n_i})^{-1} \mathbf{W}_{i,k} \boldsymbol{\xi}_{i,k}, \quad (4.12)$$

where $\hat{\eta}_{k\ell}$ is the ℓ 'th eigenvalue of the k 'th estimated covariance function, $\hat{\psi}_{i,k\ell} = (\hat{\psi}_{k\ell}(t_{i1}), \dots, \hat{\psi}_{k\ell}(t_{in_i}))'$, $\Gamma_{i,k}$ is a matrix with (j, p) element $\hat{\gamma}_k(t_{ij}, t_{ip})$, and ρ is a small regularization constant ensuring non-singularity of the covariance function (see Equation (10) in Boente and Salibián-Barrera 2021). $\boldsymbol{\xi}_{i,k} = (\tilde{\xi}_{k,i1}, \dots, \tilde{\xi}_{k,in_i})'$ is the vector of raw scores for subject i . In contrast to Boente and Salibián-Barrera (2021), we additionally include weights $\mathbf{W}_{i,k} = \text{diag}(w_{k,i1}, \dots, w_{k,in_i})$, $w_{k,in_i} \in \{0, 1\}$, to ensure that outlying observations do not compromise the reconstruction of the score functions. If outlying curves within partially contaminated individuals are not downweighted, the estimated reconstruction can be influenced by the outliers although the covariance function estimate itself is robust against outliers. An additional robustification against outliers of type (A) is not necessary, since the reconstruction will still match the observed values, e.g. at a different magnitude, and not be drawn away from the majority of data points on the curve by a single outlying raw score. Mechanisms to recognize outliers of type (A) in the aftermath are proposed in Section 4.4. The weights are determined as follows: From (4.11), define residuals

$$r_{k,i}(t_{ij}, t_{ip}) = \tilde{\xi}_{k,ij} - \hat{\beta}_k(t_{ij}, t_{ip}) \tilde{\xi}_{k,ip}.$$

Standardization with their estimated variance (see Supplement C.1.2) then gives the standardized residuals

$$\tilde{r}_{k,i}(t_{ij}, t_{ip}) := \frac{|r_{k,i}(t_{ij}, t_{ip})|}{\sqrt{\hat{\gamma}_k(t_{ij}, t_{ij}) - \frac{\hat{\gamma}_k(t_{ij}, t_{ip})^2}{\hat{\gamma}_k(t_{ip}, t_{ip})}}}. \quad (4.13)$$

For fixed k and i and a pair of time points t_{ij} and t_{ip} , (4.13) measures how well the score at time t_{ij} can be predicted from the score t_{ip} . If $\tilde{\xi}_{k,ij}$ or $\tilde{\xi}_{k,ip}$ are outlying, the residual will be large. Taking the median of the standardized residuals over all $p \neq j$, we then obtain the ‘‘measure of outlyingness’’ at time t_{ij} for the i 'th observation in the k 'th score:

$$o(\tilde{\xi}_{k,ij}) = \text{med}(\tilde{r}_{k,i}(t_{ij}, t_{ip}), p \neq j). \quad (4.14)$$

The weights in $\mathbf{W}_{i,k}$ in (4.12) are defined based on comparing the expression in (4.14) to a fixed cutoff c , yielding

$$w_{k,ij} = \begin{cases} 1, & \text{if } o(\tilde{\xi}_{k,ij}) \leq c, \\ 0, & \text{else.} \end{cases} \quad (4.15)$$

The cutoff c is a tuning parameter. A good choice is $c = 2$, as is commonly used as a cutoff for standardized residuals in regression settings. An intuition on the construction of the outlier detection procedure is given in Supplement C.1.3.

If the number of available observations per subject is not sufficient to apply the methodology described above and estimate an expansion of form (4.10), a possible alternative is to smooth

4. Robust modeling of repeated functional measurements

using parametric approaches. One possibility is to impose a random effects model structure, for example a linear model

$$\xi_{ik}(t) = b_{ik,0} + b_{ik,1} \cdot t + \eta_{ikt},$$

for each subject i , where $b_{ik,0}$ and $b_{ik,1}$ are normally distributed, and $\eta_{ikt} \perp b_{ik,0}, b_{ik,1}$ is an error term. More complex functional forms are also easily implementable, e.g. polynomial structures. Such models can be fitted robustly using robust estimation approaches for mixed effects models, e.g. using the methodology proposed in Koller (2013) or Brune, Ortner, and Filzmoser (2024). Since such models are prone to numerical problems, especially in cases with only very few available observations, the model can also be fitted using simple robust linear models with separate coefficients for each subject. An example for this is shown in the photovoltaic (PV) data application in Section 4.6.

4.3.4. Summary of the estimation procedure

The estimation procedure is summarized in the following Algorithm 4.1.

Algorithm 4.1 (Robust marginal FPCA). *Given the dataset*

$$\{(t_{ij}, Y_i(s_r, t_{ij})), r = 1, \dots, R\}, j = 1, \dots, n_i\}_{i=1, \dots, n},$$

the estimation procedure consists of the following steps:

1. Obtain estimated mean function $\hat{\mu}(s, t)$ following Section 4.3.1.
2. Obtain centered observations $Y_i^C(s, t)$ as in (4.8).
3. Obtain estimated marginal covariance function $\hat{\Sigma}(s_r, s_{r'})$, $r, r' = 1, \dots, R$, following Section 4.3.2.
4. Eigenanalysis of $\hat{\Sigma}(s_r, s_{r'})$, $r, r' = 1, \dots, R$, yielding $\{\hat{\lambda}_k, \hat{\phi}_k(s_r) : r = 1, \dots, R\}_{k=1, \dots, K}$.
5. Calculate raw scores from (4.9), yielding datasets $\{(t_{ij}, \tilde{\xi}_{k,ij}); j = 1, \dots, n_i\}_{i=1, \dots, n}$ for $k = 1, \dots, K$.
6. For each $k = 1, \dots, K$:
 - a) **Non-parametric approach:** If enough observations (e.g. > 5) are available for each subject, fit the truncated KL expansion $\hat{\xi}_{ik}(t) = \sum_{\ell=1}^{L_k} \hat{\xi}_{i,k\ell} \hat{\psi}_{k\ell}(t)$ as described in Section 4.3.3.
 - b) **Parametric approach:** If not enough observations are available, approximate $\xi_{ik}(t)$ by a parametric function of t as described in Section 4.3.3.
7. Reconstruct trajectories for any s, t by $\hat{Y}_i(s, t) = \mu(s, t) + \sum_{k=1}^K \hat{\xi}_{ik}(t_{ij}) \hat{\phi}_k(s)$.

4.4. Methods for outlier analysis

In this section, we propose outlier detection methods for both partial contamination (B) and contamination of whole subjects (A).

4.4.1. Detecting outlying subjects

In order to detect outlying individuals, we construct distance measures that measure, on the one hand, how well the score functions are reconstructed in the estimated decomposition, and, on the other hand, if their shape and size match those of the other score functions. We define the score distance SD_i and the orthogonal distance OD_i as

$$SD_i = \sum_{k=1}^K \frac{\|\hat{\xi}_{ik}(\cdot)\|_{\mathcal{T}}^2}{\hat{\lambda}_k}, \quad \text{and} \quad OD_i = \sum_{k=1}^K \left\| \hat{\xi}_{k,ij} - \sum_{\ell=1}^{L_k} \hat{\zeta}_{i,k\ell} \hat{\psi}_{k\ell}(\cdot) \right\|_{\mathcal{T}}^2, \quad (4.16)$$

respectively. For more details on the score and orthogonal distance see e.g. Liu et al. 2013. The standardization of the score distance SD_i ensures that outliers in all components receive the same weight. Otherwise, the distribution of SD_i will be dominated by the first few components with high variance and deviations or outliers in the later components might be overlooked. Lemma 4.2 establishes theoretical properties of the above quantities under the assumption of Gaussianity. It allows for the construction of a cutoff value that can be used for the detection of outliers with respect to the score functions ξ_k .

Lemma 4.2. *Assume that for every $k = 1, \dots, K$, the k 'th score function ξ_k is a Gaussian random process. Denote by $\eta_{k\ell}$ the ℓ 'th eigenvalue of the covariance of ξ_k , and by $\psi_{k\ell}$ the corresponding ℓ 'th eigenfunction. Then, for every $k = 1, \dots, K$, the following holds:*

- (i) $\xi_k(t) \sim \mathcal{N}(0, \sigma_k(t))$, and $\sigma_k(t) = \sum_{\ell=1}^{\infty} \eta_{k\ell} \psi_{k\ell}^2(t)$,
- (ii) $\left\| \xi_k(\cdot) \right\|_{\mathcal{T}}^2 \sim \sum_{\ell=1}^{\infty} \eta_{k\ell} y_{k\ell}$, where the $y_{k\ell}$, $\ell \geq 1$, are mutually independent $\chi^2(1)$ random variables,
- (iii) $\left\| \sum_{\ell=1}^{L_k} \zeta_{k\ell} \psi_{k\ell}(\cdot) \right\|_{\mathcal{T}}^2 \sim \sum_{\ell=1}^{L_k} \eta_{k\ell} y_{k\ell}$, and $\left\| \xi_k(\cdot) - \sum_{\ell=1}^{L_k} \zeta_{k\ell} \psi_{k\ell}(\cdot) \right\|_{\mathcal{T}}^2 \sim \sum_{\ell=L_k+1}^{\infty} \eta_{k\ell} y_{k\ell}$, where the $y_{k\ell}$, $\ell \geq 1$, are mutually independent $\chi^2(1)$ random variables.

Proof. Assuming that the k 'th score function ξ_k is a Gaussian random process, ξ_k admits a Karhunen-Loève expansion $\xi_k(t) = \sum_{\ell=1}^{\infty} \zeta_{k\ell} \psi_{k\ell}(t)$, where the scores $\zeta_{k\ell} \sim \mathcal{N}(0, \eta_{k\ell})$ are mutually independent for $\ell \geq 1$ and the corresponding principal curves $\psi_{k\ell}$ are orthonormal for $\ell \geq 1$. From this, the claim (i) directly follows. Parseval's identity now implies that

$$\|\xi_k\|_{\mathcal{T}}^2 = \sum_{\ell=1}^{\infty} \zeta_{k\ell}^2 = \sum_{\ell=1}^{\infty} \eta_{k\ell} \left(\frac{\zeta_{k\ell}}{\sqrt{\eta_{k\ell}}} \right)^2 = \sum_{\ell=1}^{\infty} \eta_{k\ell} y_{k\ell},$$

where

$$y_{k\ell} := \left(\frac{\zeta_{k\ell}}{\sqrt{\eta_{k\ell}}} \right)^2 \sim \chi^2(1).$$

Additionally, $\zeta_{k1}, \zeta_{k2}, \dots$ are mutually uncorrelated as PC scores and are normally distributed, making them independent. Thus, y_{k1}, y_{k2}, \dots are independent as well, proving the claim (ii). (iii) follows directly from (ii). \square

The distribution of SD_i and OD_i in (4.16) as the sum of the norms of multiple score functions cannot be directly obtained since we cannot guarantee even uncorrelatedness of the score functions $\xi_k(t)$ across k ; see Chen, Delicado, et al. (2017, p. 181). However, Lemma 4.2 allows for the construction of upper bounds for the quantiles of the distance measures given in (4.16). The approach for constructing the cutoffs for both distance measures in (4.16), again under the assumption of Gaussianity, is summarized in the following Corollary 4.3.

Corollary 4.3. *Assume that for every $k = 1, \dots, K$, the k 'th score function ξ_k is a Gaussian random process, and let the score distance $SD(X)$ and orthogonal distance $OD(X)$ be as defined in (4.16). For $p \in (0, 1)$ let $q_{SD}(1-p)$ and $q_{OD}(1-p)$ be $(1-p)$ 'th quantiles of the distributions of $SD(X)$ and $OD(X)$, respectively. For $k = 1, \dots, K$, let further $\lambda_k \bar{q}_k > 0$ and \tilde{q}_k be $(1-p/K)$ 'th quantiles of $\sum_{\ell=1}^{\infty} \eta_{k\ell} y_{k\ell}$ and $\sum_{\ell=L_{k+1}}^{\infty} \eta_{k\ell} y_{k\ell}$, respectively, determined using Lemma 4.2. Then the following bounds hold:*

$$q_{SD}(1-p) \leq \sum_{k=1}^K \bar{q}_k, \quad \text{and} \quad q_{OD}(1-p) \leq \sum_{k=1}^K \tilde{q}_k.$$

Proof. We begin by proving the bound for the quantiles of the distribution of $SD(X)$. Based on Lemma 4.2, let $\bar{q}_k, k = 1, \dots, K$, be such that, such that $\mathbb{P}(\|\xi_k(t)\|_{\mathcal{T}}^2/\lambda_k > \bar{q}_k) \leq p/K, k = 1, \dots, K$. Then,

$$\begin{aligned} \mathbb{P} \left(SD(X) \leq \sum_{k=1}^K \bar{q}_k \right) &\geq \mathbb{P} \left(\frac{\|\xi_1(t)\|_{\mathcal{T}}^2}{\lambda_1} \leq \bar{q}_1, \dots, \frac{\|\xi_K(t)\|_{\mathcal{T}}^2}{\lambda_K} \leq \bar{q}_K \right) \\ &= 1 - \mathbb{P} \left(\left\{ \frac{\|\xi_1(t)\|_{\mathcal{T}}^2}{\lambda_1} > \bar{q}_1 \right\} \cup \dots \cup \left\{ \frac{\|\xi_K(t)\|_{\mathcal{T}}^2}{\lambda_K} > \bar{q}_K \right\} \right) \\ &\geq 1 - \sum_{k=1}^K \mathbb{P} \left(\frac{\|\xi_k(t)\|_{\mathcal{T}}^2}{\lambda_k} > \bar{q}_k \right) \geq 1 - p, \end{aligned}$$

giving that

$$\mathbb{P} \left(SD(X) > \sum_{k=1}^K \bar{q}_k \right) \leq p,$$

and implying that $\sum_{k=1}^K \bar{q}_k \geq q_{SD}(1-p)$ is a conservative p -value cutoff for score distance (4.16), under the assumption of Gaussianity. Observe that $\bar{q}_k, k = 1, \dots, K$, can be also

chosen in a more general way:

$$\mathbb{P}(\|\xi_k(t)\|_{\mathcal{T}}^2/\lambda_k > \bar{q}_k) \leq p_k, k = 1, \dots, K, \text{ for } \sum_{k=1}^K p_k \leq p.$$

The bound for the quantile of the distribution of orthogonal distance is proven analogously. \square

The bounds given in Corollary 4.3 can be thought of as a Bonferroni-type correction of the significance level for multiple hypothesis testing, where the hypotheses correspond to the individual scores exceeding the cutoff under Gaussianity, and can thus become very conservative if K is large. Alternatively, more sophisticated techniques from multiple testing can be employed to determine the outlying score functions.

4.4.2. Detecting partial contamination

In the case of non-parametric smoothing, single outlying curves within an individual are flagged based on the outlyingness (4.14). An overall outlyingness of individual i at time t_{ij} can be calculated by averaging $o(\xi_{k,ij})$ over the number of components K ,

$$o(Y_i(\cdot, t_{ij})) = \frac{1}{K} \sum_{k=1}^K o(\tilde{\xi}_{k,ij}), \quad (4.17)$$

and comparing the value to the cutoff value c as in (4.15).

If the score functions are estimated parametrically, standard outlier detection methods for parametric models can be used. A standard approach to identify outlying observations are standardized residuals. Outlying observational units can be identified using e.g. the variance diagnostics suggested in Lesaffre and Verbeke (1998).

4.5. Simulation

In the following, we assess the performance of the proposed robust estimation method and outlier detection framework in an extensive simulation study. The results are compared to those from a non-robust method which replaces the proposed robust components from our procedure by non-robust counterparts (see the description in Supplement C.2.1). The experiments are performed on a Windows Server in R, Version 3.4.1., using the `batchtools` framework (Lang et al. 2017). R-code implementing the method is made available on GitHub¹.

¹<https://github.com/b-brune/robLFDA>

4.5.1. Description of data generating settings

(Partly) following Park and Staicu (2015), the raw data for the simulation is generated as follows: The number of functional principal components with smooth score functions ξ_k is fixed at $K = 2$. In order to make the model estimation harder, we additionally include 8 more components with non-smooth score functions which are generated as Gaussian white noise with a low variance. They can be considered as smooth error terms. The basis functions ϕ_k are created from a Fourier basis with 11 basis functions, where the first basis function ($\phi_0(s) \equiv 1$) is dropped. The score functions for the meaningful components are given by

$$\xi_{ik}(t) = \zeta_{i,k1}\sqrt{2}\cos(2\pi t) + \zeta_{i,k2}\sqrt{2}\sin(2\pi t), \quad k = 1, 2, \quad i = 1, \dots, n, \quad (4.18)$$

with $\zeta_{i,k1} \sim N(0, \frac{3}{k^2})$ and $\zeta_{i,k2} \sim N(0, \frac{1.5}{k^2})$, $k = 1, 2$. The non-smooth score functions are generated as normal white noise:

$$\xi_{ik}(t) \sim \mathcal{N}(0, 0.05^2) \quad \forall t \in \mathcal{T}, k \geq 3, \quad i = 1, \dots, n.$$

The mean function μ is assumed to be zero, and its estimation was deliberately excluded from the simulation study so that the performance of the latter (novel) steps of the algorithm can be assessed more clearly. It should be noted that the robust mean estimation is not the main contribution of the paper, while the possible uncertainty involved with it is propagated further through the procedure, thus clouding the effect of the outliers on the outcome of the procedure.

The curves used in the simulation study are then constructed as

$$Y_i(s, t) = \sum_{k=1}^{10} \xi_{ik}(t)\phi_k(s), \quad i = 1, \dots, n.$$

Each Y_i is evaluated on a fine grid in (s_r, t_l) , $r, l = 1, \dots, R = 101$ such that $s_r = t_r = 0.01(r - 1)$ for $r = 1, \dots, R$, further giving the sample of discretely observed functions

$$\{Y_i(s_r, t_l) : r, l = 1, \dots, R\}_{i=1, \dots, n}.$$

For each $i = 1, \dots, n$, we then uniformly draw n_i time points $t_{ij} \in \{t_1, \dots, t_R\}$, $j = 1, \dots, n_i$, such that the subsample $\{Y_i(s_r, t_{ij}) : r = 1, \dots, R, j = 1, \dots, n_i\}_{i=1, \dots, n}$ is used for model training. The remaining subset serves for out-of-sample model validation.

We consider three outlier generation settings (OS) in which samples are contaminated such that a proportion c_1 of the data $Y_i(s, t_{ij})$, $i = 1, \dots, n$, is replaced by:

OS1 *Full amplitude outliers*, such that

$$\tilde{Y}_i(s_r, t_{ij}) = b \cdot Y_i(s_r, t_{ij})$$

for an amplitude parameter b and for every $j = 1, \dots, n_i$, $r = 1, \dots, R$.

OS2 *Partial amplitude outliers*, such that

$$\tilde{Y}_i(s_r, t_{ij}) = b \cdot Y_i(s_r, t_{ij}) \mathbb{I}(t_{ij} \in \mathcal{T}_{i,\text{out}}) + Y_i(s_r, t_{ij}) \mathbb{I}(t_{ij} \notin \mathcal{T}_{i,\text{out}}),$$

for $r = 1, \dots, R$, $j = 1, \dots, n_i$, where $\mathcal{T}_{i,\text{out}} \subset \{t_{i1}, \dots, t_{in_i}\}$ consists of time points t_{ij} with $\mathbb{P}(t_{ij} \in \mathcal{T}_{i,\text{out}}) = c_2$ for $j = 1, \dots, n_i$, specifying the outlying curves within the contaminated individual, and $\mathbb{I}(t_{ij} \in \mathcal{T}_{i,\text{out}}) = 1$ if $t_{ij} \in \mathcal{T}_{i,\text{out}}$, and 0 otherwise.

OS3 *Shape (structural) outliers*, such that

$$\tilde{Y}_i(s_r, t_{ij}) = \sum_{k=1}^{10} \bar{\xi}_{ik}(t) \phi_k(s_r),$$

where the structured score functions ξ_{ik} in (4.18) are replaced by $\bar{\xi}_{ik}(t) \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 3^2) \forall t \in \mathcal{T}$, for $k = 1, 2$, $j = 1 \dots, n_i$.

Some examples of the generated data and the effects of the three outlier generating settings are given in Supplement C.2.2. The performance of the proposed method is evaluated using the following metrics: The root mean squared curvewise reconstruction error for fixed $t \in \{t_1, \dots, t_R\}$ is defined as

$$\widehat{RMSE}(Y_i(\cdot, t)) = \sqrt{\frac{1}{R} \sum_{r=1}^R (Y_i(s_r, t) - \hat{Y}_i(s_r, t))^2}. \quad (4.19)$$

It can be evaluated in- and out-of-sample by either looking at $t \in \{t_{ij}, j = 1, \dots, n_i\}$ (observed time points) or at $t \notin \{t_{ij}, j = 1, \dots, n_i\}$ (non-observed time points) for each subject $i = 1, \dots, n$. The reconstruction error for the score functions is measured by

$$\widehat{RMSE}(\xi_{ik}, \hat{\xi}_{ik}) = \sqrt{\frac{1}{R} \sum_{l=1}^R (\xi_{ik}(t_l) - \hat{\xi}_{ik}(t_l))^2}. \quad (4.20)$$

The performance of the outlier detection mechanisms on both subject and curve level is evaluated with the true positive rate ($\text{TPR} = \frac{TP}{TP+FN}$), false positive rate ($\text{FPR} = \frac{FP}{FP+TN}$) and F1 score ($\text{F1} = \frac{2TP}{2TP+FP+FN}$), where TP denotes the number of correctly detected outliers, FN the number of outliers that were not detected, FP the number of falsely detected outliers, and TN the number of correctly recognized non-outlying observations.

Each experiment is replicated 100 times for

- sample sizes $n \in \{50, 100\}$,
- numbers of curves n_i uniformly drawn from $\{5, 6, \dots, 15\}$ (very sparse setting) and $\{15, 16, \dots, 30\}$ (sparse setting),
- outlying subject proportions $c_1 \in \{0.05, 0.1, 0.2\}$,

- outlying curve proportions $c_2 \in \{0.1, 0.2\}$ (for OS2), and
- amplitude parameters $b \in \{5, 10, 20\}$.

The bandwidth parameter for the non-parametric smoothing of the score functions described in Section 4.3.3 is fixed at $h = 0.1$ in all experiments.

For the sake of simplicity, we only report the findings from OS1 and OS2 for $n = 100$ in the sparse setting. Results for OS3 can be found in Supplement C.2. Generally, decreasing the number of available samples by decreasing n or n_i leads to slightly higher errors and variances of the estimates (see the results in Supplement C.2).

4.5.2. Contamination with amplitude outliers

Results for full contamination (OS1) Figure 4.2 (top) shows the in- and out-of-sample reconstruction errors (4.19) for $n = 100$ subjects with 15 to 30 randomly sampled measurements per subject, where a certain proportion of subjects is contaminated by amplitude outliers of increasing size. While the errors of the non-robust method increase with increasing outlier size, the errors of the robust method stay stable, even for large amounts of contamination. The corresponding reconstruction errors (4.20) for the score functions of the non-outlying subjects behave similarly (also see Figure C.8 in Supplement C.2). Finally, Figure 4.2 (bottom) shows TPR, FPR and F1 score for the outlying subject detection using the proposed Bonferroni-type procedure based on the score distance SD_i in (4.16). The procedure reliably detects most outlying subjects while keeping the nominal level $\alpha = 0.05$. The detection performance increases with increasing sample size and decreases with increasing total number of outlying subjects. An analogous plot for the orthogonal distance OD_i can be found in Figure C.13 in Supplement C.2. However, SD_i outperforms OD_i in OS1, since the altered amplitude of the outlying observations only affects the magnitude of the reconstructed score functions, and not the accuracy of the reconstruction itself.

Results for partial contamination (OS2) Figure 4.3 (top) shows the reconstruction errors (4.19) for partially contaminated subjects (right) and non-contaminated ones (left). The error rates stay stable for the non-contaminated samples. For the contaminated ones, the error variance increases with increasing outlier magnitude, but the median errors are still low even for large b . The errors of the non-robust method are affected for both contaminated and non-contaminated subjects. The corresponding reconstruction errors for the score functions, see Figure C.9 in Supplement C.2, are constant for the non-contaminated samples. For the contaminated samples, the median error stays low with increasing the outlier magnitude, but the error variance increases.

Finally, the results for the outlier detection in case of $c_1 = 10\%$ of partially contaminated observations are displayed in Figure 4.3 (bottom), and indicate that the procedure reliably detects the outlying curves with a FPR below 5%.

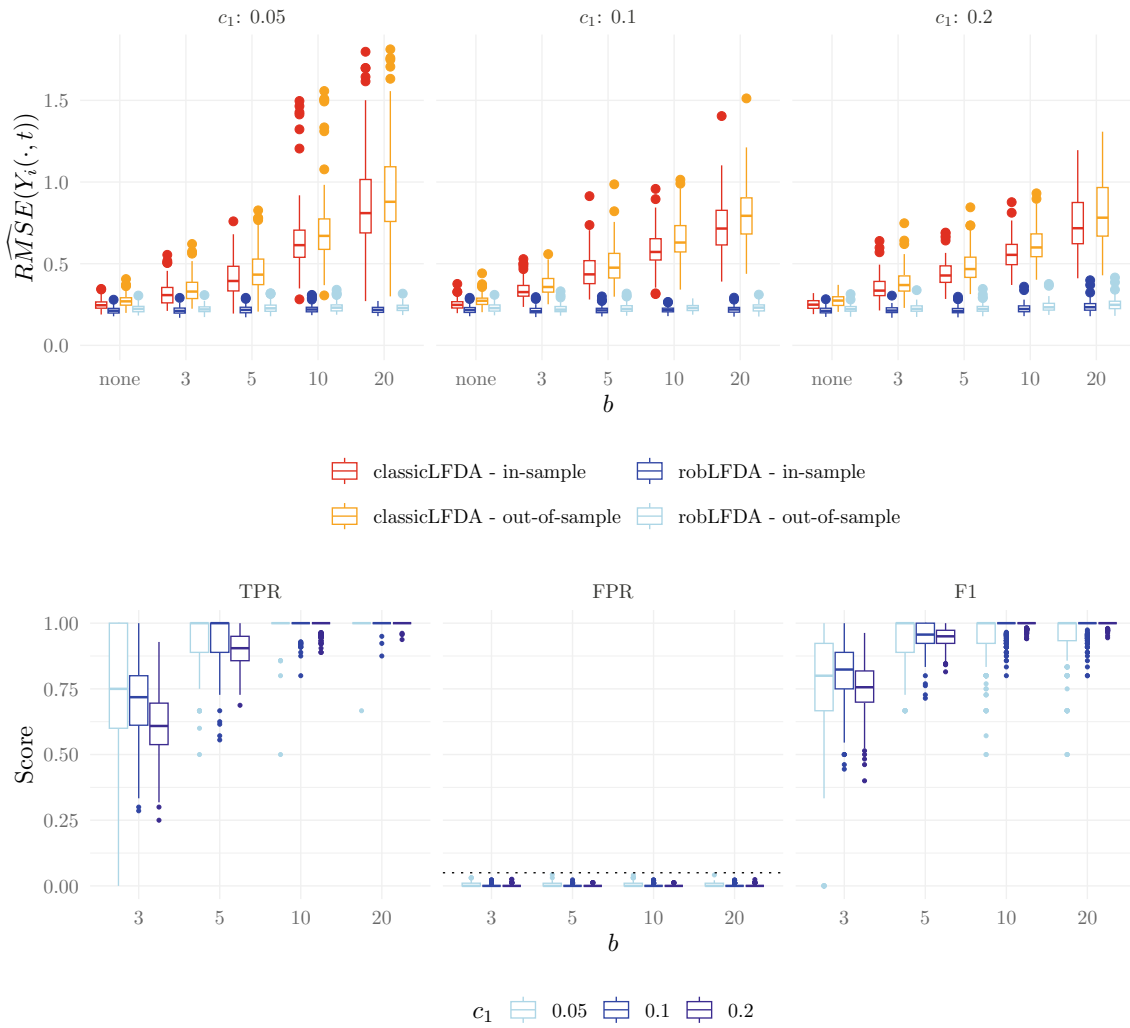


Figure 4.2.: Top: RMSE (4.19) for OS1 with $b \in \{3, 5, 10, 20\}$. Sample size $n = 100$, sparse setting, $c_1 \in \{0.05, 0.1, 0.2\}$. Bottom: Left to right: TPR, FPR, and F1 score for outlier detection based on the score distance SD_i with $\alpha = 0.05$ in (4.16) in OS1 with $b \in \{3, 5, 10, 20\}$ for sample size $n = 100$, sparse setting, $c_1 \in \{0.05, 0.1, 0.2\}$.

4.6. Real data applications

In the following, we show two data applications which illustrate different advantages of the proposed methodology. The first example is concerned with the analysis of mortality data profiles over time for different countries, with a focus on outlier detection and analysis. The second example provides a robust approach to dimension reduction for data on material degradation of photovoltaic module encapsulant materials.

4. Robust modeling of repeated functional measurements

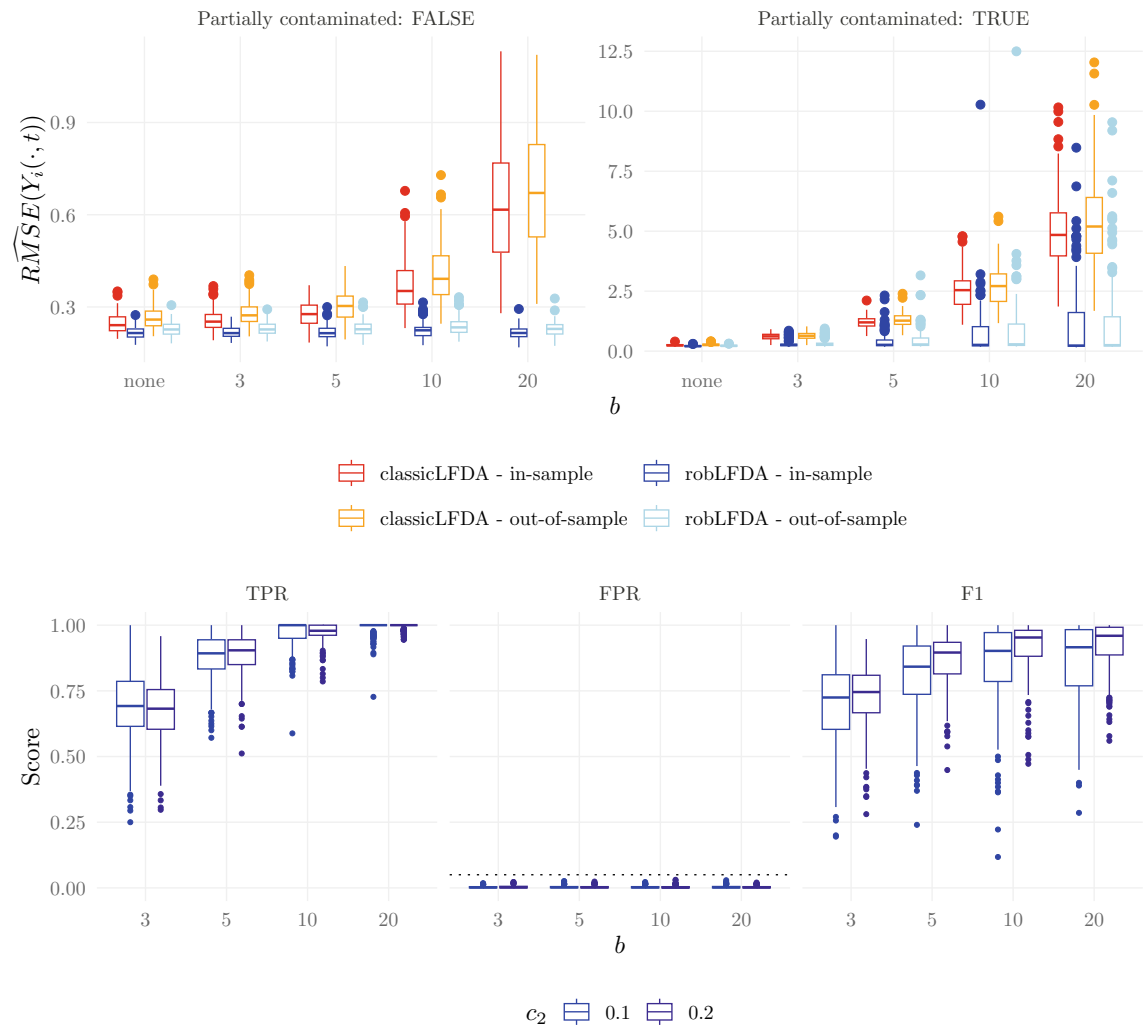


Figure 4.3.: Top: RMSE (4.19) for OS2 with $b \in \{3, 5, 10, 20\}$. Sample size $n = 100$, sparse setting, $c_1 = c_2 = 0.1$. Separated by contaminated (right) and non-contaminated (left) subjects. Bottom: Left to right: TPR, FPR, and F1 score for outlier detection based on (4.17) in OS2 with $b \in \{3, 5, 10, 20\}$ for sample size $n = 100$, $c_1 = 0.1$, sparse setting, $c_2 \in \{0.1, 0.2\}$.

4.6.1. Robust analysis of mortality data

To illustrate the outlier detection capabilities of the proposed methodology, we analyze mortality data from the Human Mortality Database (www.mortality.org²). The dataset of

²HMD. Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at

interest consists of age-specific mortality rates for 36 countries from 1953 to 2023, for ages 0 to 86. Age-specific mortality is calculated as the ratio of death counts to population exposure for a given year and age. For illustration, the mortality curves for Austria, Germany, Japan, and Russia are displayed in the following Figure 4.4 (top).

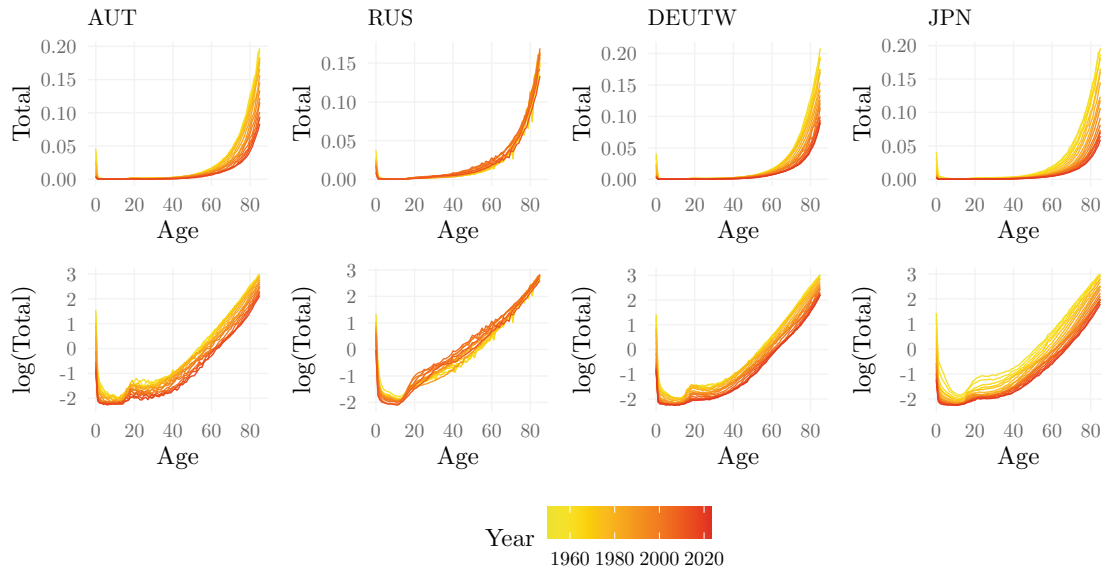


Figure 4.4.: Mortality curves for Austria, Russia, (Western) Germany and Japan from 1953 to 2023 (every fourth year). Raw data (top) and logarithmized data (bottom).

We aim to quantify the differences between the mortality profiles and detect outlying countries and time periods. For that, we fit the robust marginal FPCA model with non-parametric smoothing. Positivity of the estimates is ensured by working with the logarithmized mortality rates (see bottom panel of Figure 4.4).

The model is formulated as

$$\log(Y_{ij}(s) + 0.1) = \log(Y_i(s, t_{ij}) + 0.1) = \mu(s, t_{ij}) + \sum_{k=1}^K \xi_{ik}(t) \phi_k(s),$$

and we choose $K = 3$ marginal principal components. The data is centered using the robust backfitting algorithm from Boente, Martínez, et al. (2017) with additive effects in s and t , such that $\mu(s, t) = \mu_0 + \nu(s) + \delta(t)$, with bandwidths 0.1 for age and 0.15 for time, respectively. Due to the high computational complexity, the mean function is estimated from a subsample of the whole dataset. The resulting spline functions $\nu(s)$ and $\delta(t)$ are displayed in Figure C.14 in Supplement C.3.1. The three FPCs obtained from the marginal covariance matrix (estimated by MRCT) explain approximately 87% of the variance. The

www.mortality.org (data downloaded on June 07, 2024).

4. Robust modeling of repeated functional measurements

effects of the marginal eigenfunctions $\phi_k(s)$ on the component $\nu(s)$ of the mean function are displayed in Figure 4.5 (see Figure C.15 in Supplement C.3.1 for the raw marginal eigenfunctions). The first eigenfunction, ϕ_1 represents the differences in mortality for people above 30. ϕ_2 measures increased and decreased mortality of younger people (up to age 60). The third eigenfunction ϕ_3 mainly focuses on the infant mortality. Non-parametric

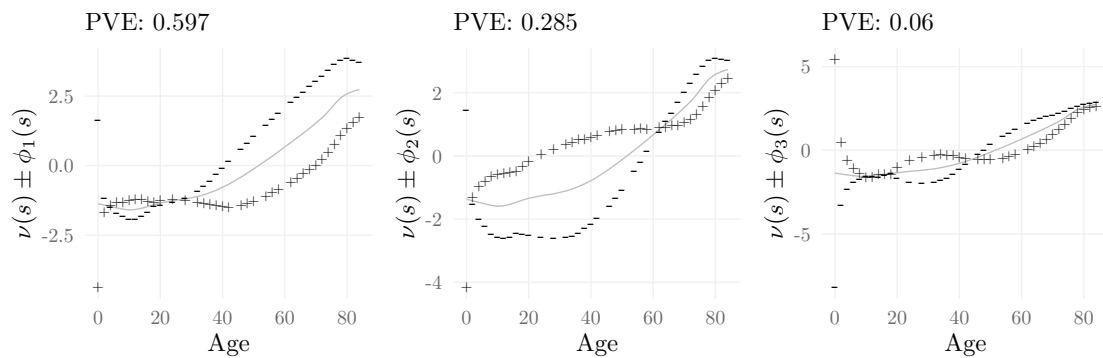


Figure 4.5.: Estimated age-wise average $\nu(s)$ from the additive model $\mu(s, t) = \mu_0 + \nu(s) + \delta(t)$ plus/ minus the estimated marginal eigenfunctions ($\nu(s) \pm \phi_k(s)$, $k = 1, 2, 3$) for the mortality dataset.

smoothing of the raw scores (calculated from the marginal eigenfunctions using (4.9)) with bandwidth $h = 0.03$ then yields the corresponding models for the time dynamics. The raw scores and the fitted score functions $\hat{\xi}_{ik}$ for the four countries from Figure 4.4 and Portugal are displayed in Figure 4.6. The years for which the mortality curves are flagged as outliers based on the weights (4.15) are marked in red. Furthermore, the countries that were flagged as outlying based on score distance and orthogonal distance using the cutoffs defined in Corollary 4.3 are displayed in a distance-distance (SD_i vs. OD_i) plot in Figure 4.7. Russia is flagged based on score distance and orthogonal distance, while the other outliers are only flagged based on the orthogonal distance OD_i . For illustration, we will interpret the outliers for two countries. For Russia, both the first and the second score function strongly deviate from the other countries displayed and show a large number of outlying scores. In contrast to most other countries, component ξ_1 seems to be decreasing for Russia, indicating an increase in overall mortality with time. At the same time, high positive values of the second component indicate that the mortality of the population below 60 years is higher than average. While the corresponding score functions of most inlying countries do not increase much, the fitted curve is very steep for Russia. The flagged outlying curves indicate a structural break in components 1 and 2 (also see C.16 in Supplement C.3.1). There seems to be a structural change in the mortality curves after 1990. Similar patterns can also be observed for the other former USSR countries. For Portugal we can observe a strong outlyingness in ξ_3 , indicating that the infant mortality was unusually high in Portugal until the 1980s.

Concluding, we compare the 20% trimmed mean squared reconstruction errors for the

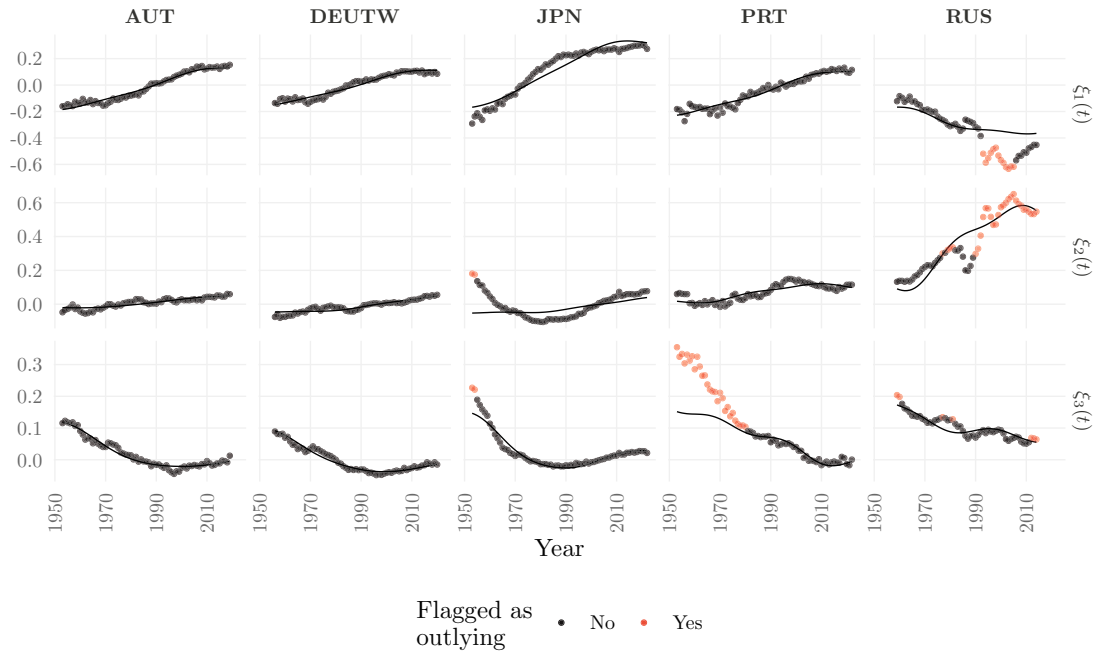


Figure 4.6.: Raw scores (points) and fitted score functions $\hat{\xi}_{ik}$, $k = 1, 2, 3$, from the mortality dataset (solid lines), for Austria, (Western) Germany, Japan, Portugal and Russia. Scores that are flagged as outlying based on the outlyingness (4.14) are marked in red.

mortality profiles of the robust model, and those of a corresponding non-robust fit (also see the country-wise errors in Figure C.17 in Supplement C.3.1). Use of the trimmed mean accounts for the potential outliers in the data. The robust estimates yield an overall reconstruction error of 0.0065, while the corresponding value for the classic estimate is 0.0078 and thus about 20% larger.

4.6.2. Dimension reduction for UV-F spectra from PV modules

We analyze data from accelerated aging experiments in material science for photovoltaic (PV) modules (Knoebl et al. 2024). In the accelerated aging experiments, the modules are exposed to different climatic stresses in a climate chamber.

The analyzed dataset consists of UV-F spectra measured on the modules' encapsulant (i.e. top) material after exposure to up to 3000 hours of 5 different climatic stresses (listed in Table C.2 in Supplement C.3.2). In total, data from 18 modules with 4 – 8 measurements each are available. Thus, this dataset poses an example of sparsely and irregularly sampled longitudinal functional data. Examples of the spectral measurements are shown in Figure 4.8. With increasing exposure time, the spectra form two peaks. The goal of the analysis is

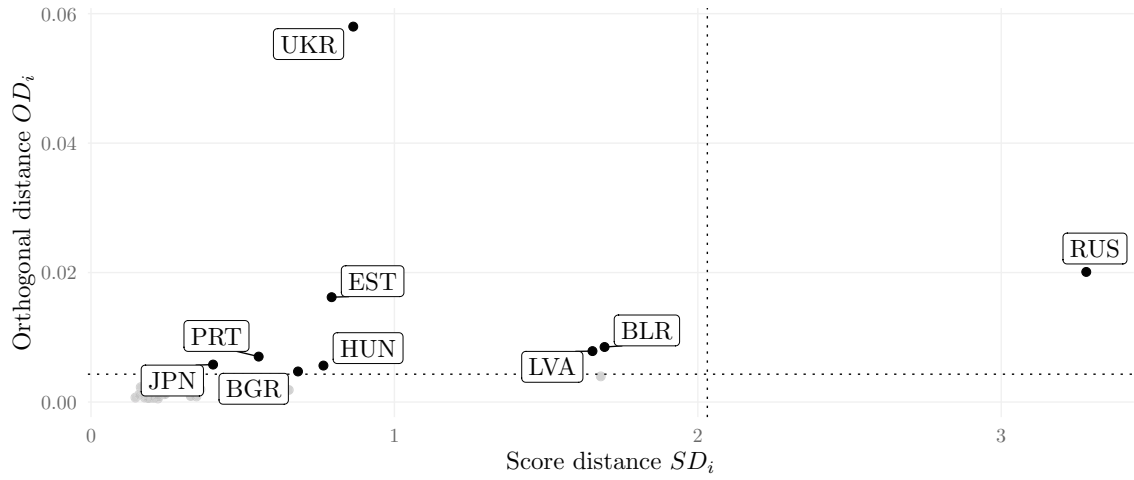


Figure 4.7.: Distance-distance plot of score distance SD_i vs. orthogonal distance OD_i from (4.16) for all countries in the mortality analysis.

to differentiate between the effects of the climatic stresses on the module materials. The dataset is known to contain some erroneous measurements caused by experimental problems (see Brune, Ortner, Eder, et al. 2023).

The spectra are again centered using the method from Boente, Martínez, et al. 2017, where we assume $\mu(s, t) = \mu_0 + \nu(s)$ (bandwidth 0.1 for the frequency domain; see Figure C.18 in Supplement C.3.2). We fit $K = 2$ marginal FPCs, explaining 97% of the variance in the data. The marginal eigenfunctions are displayed in Figures C.19 and C.20 in Supplement C.3.2. The first eigenfunction ϕ_1 captures the forming of the two peaks. The second component ϕ_2 allows to differentiate between the two peaks. Due to the low number of available samples in the time domain, the score functions are fitted parametrically using a robust linear model (function `lmrob` from R-package `robustbase`, Maechler et al. 2024) with separate intercepts and slopes in squared time for each module i and $k = 1, 2$, with model equation

$$\xi_{ik}(t) = \alpha_k \cdot \text{Ramp}(t) + \beta_{0,ki} + \beta_{1,ki} \cdot t^2.$$

It consists of a squared slope in time and the coefficient $\text{Ramp}(t) = \frac{t}{0.2}\mathbb{I}(t \leq 0.2) + \mathbb{I}(t > 0.2)$ accounting for an initial material change, where \mathbb{I} is the indicator function. The fitted score functions for the two FPCs are shown in Figure 4.9. The “average curve” for each climate setting is highlighted. For $\hat{\xi}_{i1}$ (Figure 4.9, left), the curves cluster clearly by climate setting, showing that there indeed are large differences between them. For $\hat{\xi}_{i2}$ (Figure 4.9, right), the structure is less pronounced. However, we can again observe trends: The score functions in the settings with strong irradiance (Arid 1 and Moderate 1) decrease with exposure time, while those with no / only little exposure to irradiance (Alpin 1, Moderate 5, Tropical 2) show a positive trend. Thus, for Arid 1 and Moderate 1, the first peak increases more strongly than the second peak. In contrast, for Alpin 1, Moderate 5 and Tropical 2,

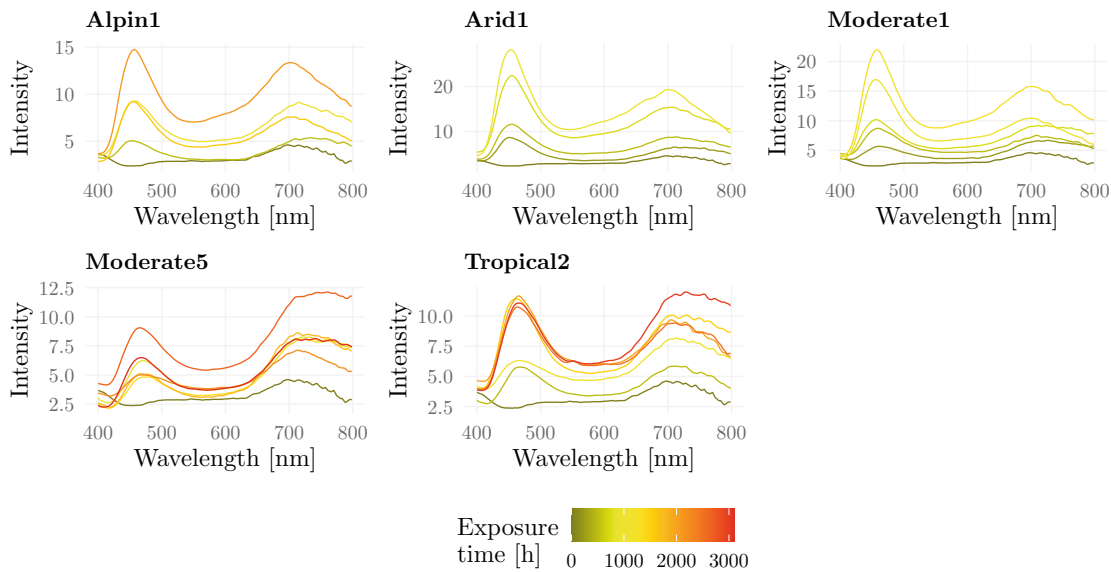


Figure 4.8.: Examples of UV-F spectra from modules exposed to the five climate settings.

the second peak increases relative to the first peak. The robust regression model provides weights for the outlyingness of each observed spectrum. In total, five curves are flagged as outlying based on the robustness weights returned by the function: Three from Moderate 5 aging action, and two from Tropical 2.

The 10% trimmed root mean squared reconstruction errors for the UV-F spectra are 0.736 for the robust model, and 0.786 for the corresponding non-robust fit.

4.7. Discussion and conclusions

In this paper, we propose a robust estimation approach for marginal functional principal component analysis, designed for functional data obtained in longitudinal or repeated measurements designs. The estimation algorithm consists of three steps, namely (1) mean function estimation, (2) estimation of the marginal covariance function, and (3) estimation of the smooth score functions. The algorithm is modular and built on a plug-in principle. Thus, it is very flexible and allows to replace components of the algorithm with suitable alternatives to the suggested robust estimators. Especially, steps (1) and (3) also allow to include prior knowledge into the estimation, e.g. in the specification of the functional forms of the mean function or parametrically smoothed score functions, or by letting those depend on additional covariates. We expect the approach of marginal FPCA to generalize to more complex longitudinal data, such as multivariate longitudinal data (Verbeke et al. 2014) or longitudinal image data (e.g. fMRI, Bowman et al. 2007), and plan to explore this in future

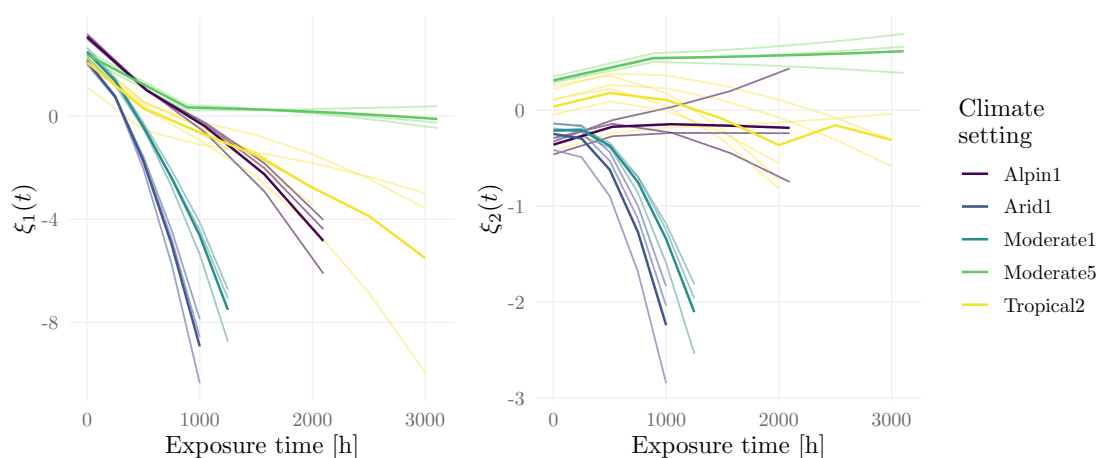


Figure 4.9.: Fitted score functions $\xi_{ik}(t)$, $k = 1, 2$, for the UV-F dataset. The mean curve for each climate setting is highlighted.

research.

Our simulation studies show that the robust procedure is stable with increasing outlier size and proportions. The proposed outlier detection techniques and theoretical cutoffs reliably identify outliers, both in the case of partial contamination (single curves within an individual), and in the case of whole outlying individuals. The data examples show two applications of the model and estimation algorithm. The identified outliers can give interesting insights into the data structure. At the same time, the models can reduce the dimension of the complex longitudinal functional datasets for further analysis.

The algorithm is designed to work with irregular and potentially sparse sampling grids for the repeated measurements (in the time domain), as well as with varying numbers of samples. However, by using a robust covariance estimation method for sparse functional data in Section 4.3.2, analogous to that applied to estimate the covariance functions of the scores, the proposed approach could be modified to accommodate sparse and irregularly observed data in the spatial domain as well. While the proposed non-parametric method to smooth the score functions works well in the presence of enough data points, it can be infeasible if the number of samples is very low. In this case, the parametric estimation provides a feasible alternative. However, reconstructing the score functions in very sparse settings can pose challenges that are even more pronounced when robust methods are employed. Although the pooling of information across subjects facilitates the model estimation, a very low number of available samples can be challenging for the robust models. Thus, in practice the parametric model (and robust algorithm) should be selected carefully.

5 | Conclusions and Discussion

In this thesis, we have presented three different modeling approaches for complex (functional) longitudinal and multivariate time series data. The proposed models and algorithms incorporate deviations from the standard assumptions in terms of stability and time-invariance of model parameters and the potential presence of outliers.

Chapter 2 presents an approach to time-varying reduced rank regression for multivariate time series regressions. The model is set up as a Gaussian linear state-space system. This allows to use well-established tools to estimate the model. We apply the Kalman filter and smoother together with an EM algorithm with analytic closed-form updates. The reduced rank of the coefficient matrices is selected with high accuracy using a Bayesian information criterion (BIC). The variation of the coefficient matrices is parameterized as a random walk which yields high flexibility and is able to deal with different types of time variation – gradual transitions of the parameters as well as abrupt structural breaks. At the same time, the algorithm gives reasonable results in the presence of time-constant coefficients..

One of the main motivations for the chosen approach was to develop a reduced-rank regression model with time-varying parameters which is relatively easy to understand and to handle. To achieve this, we introduced a number of simplifications which address the issues that were pointed out in the literature (e.g. in Yang and Bauwens 2018). The advantage of the derived model is that it can be handled with computationally efficient and well developed methods. The model may be generalized in many directions, e.g. by introducing time-variation to the error covariances. This comes with cost of significant extra efforts for estimation. An obvious question that could be asked is why we limited our analyses to what we refer to models (A) and (B) (see Equations (2.3) and (2.4), p. 32) and do not analyze model (C) more closely (see Equation (2.5), p. 32). In model (C), both of the parameter matrices in the decomposition of the regression coefficient matrix C_t vary in time. In this case, in addition to losing the linearity of the corresponding state-space system, the likelihood can no longer be computed analytically. Instead it can only be approximated by numerically intensive schemes (e.g. particle filtering), and we lose the possibility to apply the EM algorithm.

While allowing for time-variation can be considered as a “special case” of robustness, we should also consider the effect that single outlying observations can have on the algorithm. While the time-varying coefficients and updates should be able to at least partly absorb single outliers, the parameter estimation in the EM algorithm, which is based on least squares, could be strongly affected. Thus, an interesting topic for future research is to robustify the

proposed EM algorithm towards outlying response values, e.g. by modifying the updates in the EM algorithm as presented in Table 2.1 on p. 38. In the literature, different approaches to robust filtering in state-space models have been proposed, e.g. Meinhold and Singpurwalla (1989), where the normal densities are replaced by densities of t-distributions, or Cipra and Romera (1991), where concepts of M-estimation are applied. However, any modifications in this direction will again cause loss of the easy and straightforward techniques for model estimation.

Further topics for future research could include inference for the model. Although we have shown that the proposed model is rather robust in the case where no time-variation is present (see Table 2.3, p. 45), it could be of interest to formally test for the presence of time-variation in the parameters. The proposed rank detection procedure based on the Bayesian information criterion (see Section 2.2.3, p. 40f.) reliably detects the rank of the coefficient matrices and rarely overestimates the true rank (see Table 2.6, p. 49). In the context of time-varying relationships, another setting that could occur in practice is a change of the rank of the coefficient matrix. This would again render the proposed model formulation infeasible. However, it would be interesting to analyze the effects of changing ranks on the model estimation and rank selection.

In **Chapter 3**, we present an algorithm to fit mixed effects models using estimation techniques from rank-based regression (R-estimators). The proposed algorithm extends the existing methodology from Kloke et al. (2009) and Bilgic (2012) to incorporate random slope parameters. Furthermore, by introducing two types of weights, we ensure robustness of the algorithm in the presence of outlying observations and leverage points, respectively. An advantage of the proposed estimates towards other competing methods is that they can be computed very efficiently.

However, the two-stage iterative algorithm still poses some challenges. As can be seen in the analysis of the bias of the estimates (see Section 3.4.1, p. 68f), for small group sizes n_i , the error and random effects variances are under- and overestimated, respectively. It seems that we are overfitting the observations in each group. This effect is caused by the iterative construction of the algorithm. A similar effect can be observed if we replace the robust rank-based estimators by regular least squares estimation of the regression coefficients. The two-stage estimation method requires fitting a separate regression model to each of the groups or clusters, and the separate intercept estimation step (see Equation 3.2.1, p. 59) ensures the observations are centered exactly in each group. This is where our approach deviates from the classic estimation methods based on the assumption of Gaussianity and (restricted) maximum likelihood. Here, the distributional assumptions allow to circumvent the direct estimation of the regression coefficients and the random effects, and instead directly optimize for the variance parameters. Estimates of the regression parameters can then be obtained as posterior estimates given the observations and the estimated variances. This implicitly introduces a regularization into the estimates, described as a “shrinkage factor” in Fahrmeir et al. (2013, Ch. 7.1, p. 373f.). The larger the group sizes n_i , the lower the shrinkage and the closer the estimates are to those obtained without regularization. In this spirit, a potential solution for the overfitting problem could be offered by deriving analogous shrinkage or correction factors.

Generally, studying the asymptotic distributions of the proposed estimators by similar methodology as that of Bilgic (2012), or Chapter 3 in Hettmansperger and McKean (2011), could lead to new insights and allow for formal inference on the regression coefficients.

The leverage weights as proposed in Section 3.3.2, p. 64f, reliably robustify the proposed estimator against leverage points. However, downweighting leverage points can also cause losses in efficiency, since relevant or helpful data points might mistakenly be identified as outlying [see e.g. the example depicted in Figure 2 in McKean (2004)]. There is a differentiation between “good” and “bad” leverage points – while good leverage points are consistent with the regression line (but far away from the center of the data), bad leverage points are outliers that deviate from the actual regression relationship (Wilcox and Xu 2022). The *Wilcoxon high-breakdown estimator* described in McKean (2004) can lead to further improvements. Not only accounting for the distance in the predictor space, it also considers residual information when determining the weights for the observations. Furthermore, as discussed briefly in Section 3.3.2, the leverage weights in Equation (3.12) on p. 65 are based on the assumption that the predictors are continuously distributed. However, there could also be situations where we encounter dummy or discrete predictors. In those cases, the minimum covariance determinant estimator cannot be applied. Other approaches to downweighting could be based on estimates of leverage obtained from the hat-matrix. However, the naive weights we have explored so far (see Section B.5 in the appendix, p. 144) only offer a certain degree of protection against leverage points. They are not robust in the sense that a single observation can still contaminate the estimated leverage of the other observations. Exploring more robust alternatives in the presence of non-elliptical predictors could lead to an even more versatile estimator.

Finally, **Chapter 4** introduces a robust estimation algorithm for marginal functional principal component analysis as proposed in Park and Staicu (2015) and Chen, Delicado, et al. (2017). Marginal FPCA provides an interpretable, Karhunen-Loève type expansion for repeated functional measurements or longitudinally observed functional data. Data collected in such designs pose challenges since they exhibit complex dependency structures in the functional and the time domain. In practice, the marginal FPCA model can be fit to an observed dataset in three estimation steps: (1) mean function estimation, (2) estimation of the marginal eigenfunctions, and (3) smoothing of the score functions. The presented method is a robust estimation algorithm that is insensitive to outlying observations, both for single curves that are outlying within an observational unit, and whole observational units that behave differently from the majority of observations. The proposed outlier detection methods reliably identify the outlying curves and lead to interesting insights in the data applications. The estimation algorithm is modular and based on a plug-in principle that replaces the non-robust estimators from the initial papers by robust ones. This approach offers great flexibility and allows to incorporate prior knowledge into the model. The proposed extension of the method from Boente and Salibián-Barrera (2021) to robustify against spikes and cellwise outliers in the functional observations based on the “measure of outlyingness” in (4.14) on p. 87 could also prove useful in settings that require robust smoothing of functions or spike detection.

Within the reported simulation studies, we only considered outliers in the score functions:

Either in the form of whole outlying score functions (scenarios OS1 and OS3, see p. 91f), or as single curves within an observational unit (scenario OS2, see p. 91f). Obviously, those outlier settings are not exhaustive. One setting worth discussing is that of structural outliers which evolve from a model with modified marginal functional principal components. In this case, use of MRCT covariance estimator ensures that we still obtain a reliable estimate of the marginal covariance function. An outlier detection procedure based on the α -Mahalanobis distance will flag the unusual observations (see simulation study in Oguamalam et al. 2024). Another situation that could be encountered is that of outlying cells s_r within a functional observation $\{Y_i(s_r, t), r = 1, \dots, R\}$ for fixed t . This can be considered as a setting with spikes or cellwise outliers, similar to the outliers of type (B) described in Section 4.3.3, p. 85f. In this case, again, MRCT can be expected to provide at least a certain degree of robustness against the outlying cells. Another option could be to apply a cellwise outlier detection method such as DDC (Raymaekers and Rousseeuw 2024) for high-dimensional data to identify and smooth out the spikes before further processing the curves.

Throughout the paper, we have assumed that the functions are observed without error terms. However, in our simulation study we have included (orthogonal) smooth error terms into the model, which does not seem to affect the performance of the proposed methodology in practice. Similarly, we do not expect cases where the datasets are observed with white-noise type errors to drastically affect the model estimation performance. If the sampling grid s_1, \dots, s_R is sparse or irregular, our method can be adapted by replacing the MRCT estimator by the robust method for sparse functional data also applied in Step 3 of our estimation algorithm (see Section 4.3.3, p. 85f). However, since this cellwise robust method is based on solving many local regressions, the computational complexity will become prohibitive in the case of high-dimensional sampling grids.

The approach of (robust) marginal FPCA could be generalized to even more complex longitudinal data such as multivariate functional longitudinal data (Verbeke et al. 2014), spatial functional data (Ramsay, Ramsay, et al. 2011), or longitudinal image data (Bowman et al. 2007).

Concluding, this thesis addresses a broad range of problems in the modeling of repeated measurements data. The three methods that we have developed deal with various challenges in the statistical modeling of complex multivariate and longitudinal functional data. As described, all three papers leave open interesting questions that could be addressed in future research.

Bibliography

- Abebe, A., J. W. McKean, J. D. Kloke, and Y. K. Bilgic (2016). “Iterated Reweighted Rank-Based Estimates for GEE Models”. In: *Robust Rank-Based and Nonparametric Methods*. Ed. by R. Y. Liu and J. W. McKean. Vol. 168. Cham: Springer International Publishing, pp. 61–79. ISBN: 978-3-319-39063-5 978-3-319-39065-9. DOI: 10.1007/978-3-319-39065-9_4.
- Agostinelli, C. and V. J. Yohai (2016). “Composite Robust Estimators for Linear Mixed Models”. In: *Journal of the American Statistical Association* 111.516, pp. 1764–1774. DOI: 10.1080/01621459.2015.1115358.
- Ahn, S. K. and G. C. Reinsel (1988). “Nested Reduced-Rank Autoregressive Models for Multiple Time Series”. In: *Journal of the American Statistical Association* 83.403, pp. 849–856. DOI: 10.1080/01621459.1988.10478673.
- Akaike, H. (1973). “Information Theory as an Extension of the Maximum Likelihood Principle”. In: *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado. DOI: 10.1007/978-1-4612-1694-0_15.
- Alfons, A., C. Croux, and P. Filzmoser (2017). “Robust Maximum Association Estimators”. In: *Journal of the American Statistical Association* 112.517, pp. 436–445. DOI: 10.1080/01621459.2016.1148609.
- Anderson, B. D. O. and M. Deistler (2008). “Generalized Linear Dynamic Factor Models – a Structure Theory”. In: *Proceedings of the 47th IEEE Conference on Decision and Control*. DOI: 10.1109/CDC.2008.4739367.
- Aznar, A. and M. Salvador (2002). “Selecting the Rank of the Cointegration Space and the Form of the Intercept Using an Information Criterion”. In: *Econometric Theory* 18, pp. 926–947. DOI: 10.1017/S0266466602184064.
- Bai, J. and S. Ng (2002). “Determining the Number of Factors in Approximate Factor Models”. In: *Econometrica* 70.1, pp. 191–221. DOI: 10.1111/1468-0262.00273.
- Baştürk, N., L. Hoogerheide, and H. K. van Dijk (2017). “Bayesian Analysis of Boundary and Near-Boundary Evidence in Econometric Models with Reduced Rank”. In: *Bayesian Analysis* 12.3. DOI: 10.1214/17-BA1061.
- Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.1. DOI: 10.18637/jss.v067.i01.
- Beck, N. (1983). “Time-Varying Parameter Regression Models”. In: *American Journal of Political Science* 27.3, pp. 557–600. DOI: 10.2307/2110985.
- Belenky, G., N. J. Wesensten, D. R. Thorne, M. L. Thomas, H. C. Sing, D. P. Redmond, M. B. Russo, and T. J. Balkin (2003). “Patterns of Performance Degradation and Restoration during Sleep Restriction and Subsequent Recovery: A Sleep Dose-Response Study”. In: *Journal of Sleep Research* 12.1, pp. 1–12. DOI: 10.1046/j.1365-2869.2003.00337.x.

- Berger, K., K. Knöbl, B. Brune, I. Ortner, S. Grubner, G. Oreski, C. Barretta, B. Kubicek, L. Neumaier, M. Feichtner, Y. Voronko, and G. Eder (2021). “Advanced Degradation Modelling of Photovoltaic Modules and Materials”. In: *38th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 745–748. DOI: 10.4229/EUPVSEC20212021-4AV.1.19.
- Berrendero, J. R., B. Bueno-Larraz, and A. Cuevas (2020). “On Mahalanobis Distance in Functional Settings.” In: *Journal of Machine Learning Research* 21.9, pp. 1–33.
- Bierens, H. J. and L. F. Martins (2010). “Time-Varying Cointegration”. In: *Econometric Theory* 26, pp. 1453–1490. DOI: 10.1017/S0266466609990648.
- Bilgic, Y. K. (2012). *Rank-Based Estimation and Prediction for Mixed-Effects Models in Nested Designs*. Dissertation, Western Michigan University. URL: <https://scholarworks.wmich.edu/dissertations/40/>.
- Boente, G., A. Martínez, and M. Salibián-Barrera (2017). “Robust Estimators for Additive Models Using Backfitting”. In: *Journal of Nonparametric Statistics* 29.4, pp. 744–767. DOI: 10.1080/10485252.2017.1369077.
- Boente, G. and M. Salibián-Barrera (2021). “Robust Functional Principal Components for Sparse Longitudinal Data”. In: *METRON* 79.2, pp. 159–188. DOI: 10.1007/s40300-020-00193-3.
- Bordin, C. J. and M. G. S. Bruno (2019). “Nonlinear State Estimation Using Particle Filters on the Stiefel Manifold”. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, United Kingdom: IEEE. ISBN: 978-1-4799-8131-1. DOI: 10.1109/ICASSP.2019.8683866.
- Boudt, K., P. J. Rousseeuw, S. Vanduffel, and T. Verdonck (2020). “The Minimum Regularized Covariance Determinant Estimator”. In: *Statistics and Computing* 30.1, pp. 113–128. DOI: 10.1007/s11222-019-09869-x.
- Bowman, F. D., Y. Guo, and G. Derado (2007). “Statistical Approaches to Functional Neuroimaging Data”. In: *Neuroimaging Clinics of North America* 17.4, pp. 441–458. DOI: 10.1016/j.nic.2007.09.002.
- Breitung, J. and S. Eickmeier (2011). “Testing for Structural Breaks in Dynamic Factor Models”. In: *Journal of Econometrics* 163.1, pp. 71–84. DOI: 10.1016/j.jeconom.2010.11.008.
- Brune, B. (2024). *rankLME: Robust Rank-Based Estimation of Linear Mixed Effects Models*. DOI: 10.5281/zenodo.13843795. URL: <https://github.com/b-brune/rankLME>.
- Brune, B., I. Ortner, G. C. Eder, Y. Voronko, A. Gassner, K. A. Berger, G. Oreski, K. Knöbl, L. Neumaier, and M. Feichtner (2023). “Connecting Material Degradation and Power Loss of PV Modules using Advanced Statistical Methodology”. In: *Solar Energy Materials and Solar Cells* 260, p. 112485. DOI: 10.1016/j.solmat.2023.112485.
- Brune, B., I. Ortner, and P. Filzmoser (2024). *A Rank-Based Estimation Method for Mixed Effects Models in the Presence of Outlying Data*. Conditionally accepted for publication in: *Journal of Data Science, Statistics, and Visualization*.
- Brune, B., W. Scherrer, and E. Bura (2022). “A State-Space Approach to Time-Varying Reduced-Rank Regression”. In: *Econometric Reviews* 41.8, pp. 895–917. DOI: 10.1080/07474938.2022.2073743.

- Bura, E. and R. D. Cook (2001). “Estimating the Structural Dimension of Regressions via Parametric Inverse Regression”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 63.2, pp. 393–410. DOI: 10.1111/1467-9868.00292.
- (2003). “Rank Estimation in Reduced-Rank Regression”. In: *Journal of Multivariate Analysis* 87.1, pp. 159–176. DOI: 10.1016/S0047-259X(03)00029-0.
- Cai, T. T. and M. Yuan (2011). “Optimal Estimation of the Mean Function Based on Discretely Sampled Functional Data: Phase Transition”. In: *The Annals of Statistics* 39.5. DOI: 10.1214/11-AOS898.
- Cantoni, E. and E. Ronchetti (2001). “Robust Inference for Generalized Linear Models”. In: *Journal of the American Statistical Association* 96.455, pp. 1022–1030. DOI: 10.1198/016214501753209004.
- Carriero, A., G. Kapetanios, and M. Marcellino (2016). “Structural Analysis with Multivariate Autoregressive Index Models”. In: *Journal of Econometrics* 192.2, pp. 332–348. DOI: 10.1016/j.jeconom.2016.02.002.
- Casella, G. and R. L. Berger (2002). *Statistical Inference*. 2nd Edition. Pacific Grove, CA: Duxbury Press.
- Chamberlain, G. (1983). “Funds, Factors, and Diversification in Arbitrage Pricing Models”. In: *Econometrica* 51.5, pp. 1305–1323. DOI: 10.2307/1912276.
- Chen, K., P. Delicado, and H.-G. Müller (2017). “Modelling Function-Valued Stochastic Processes, with Applications to Fertility Dynamics”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 79.1, pp. 177–196. DOI: 10.1111/rssb.12160.
- Chen, K. and H.-G. Müller (2012). “Modeling Repeated Functional Observations”. In: *Journal of the American Statistical Association* 107.500, pp. 1599–1609. DOI: 10.1080/01621459.2012.734196.
- Chikuse, Y. (2003). *Statistics on Special Manifolds*. New York, NY: Springer New York. DOI: 10.1007/978-0-387-21540-2.
- Cipra, T. and R. Romera (1991). “Robust Kalman Filter and Its Application in Time Series Analysis”. In: *Kybernetika* 27.6, pp. 481–494.
- Cook, R. D. and L. Forzani (2008). “Principal Fitted Components for Dimension Reduction in Regression”. In: *Statistical Science* 23.4, pp. 485–501. DOI: 10.1214/08-STS275.
- Copt, S. and M.-P. Victoria-Feser (2006). “High-Breakdown Inference for Mixed Linear Models”. In: *Journal of the American Statistical Association* 101.473, pp. 292–300. DOI: 10.1198/016214505000000772.
- Coulombe, P. G. (2021). *Time-Varying Parameters as Ridge Regressions*. arXiv:2009.00401v2 [econ, stat]. URL: <https://arxiv.org/pdf/2009.00401v2.pdf>.
- Davis, C. S. (2003). *Statistical Methods for the Analysis of Repeated Measurements*. Corr. print. New York: Springer.
- Davis, R. A., P. Zang, and T. Zheng (2016). “Sparse Vector Autoregressive Modeling”. In: *Journal of Computational and Graphical Statistics* 25.4, pp. 1077–1096. DOI: 10.1080/10618600.2015.1092978.
- Del Negro, M. and C. M. Otrok (2008). “Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles”. In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1136163.

- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). “Maximum Likelihood from Incomplete Data Via the *EM* Algorithm”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 39.1, pp. 1–22. DOI: 10.1111/j.2517-6161.1977.tb01600.x.
- Doan, T., R. Litterman, and C. Sims (1984). “Forecasting and Conditional Projection Using Realistic Prior Distributions”. In: *Econometric Reviews* 3.1, pp. 1–100. DOI: 10.1080/07474938408800053.
- Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. 2nd Edition. Oxford: Oxford University Press.
- Eickmeier, S., W. Lemke, and M. Marcellino (2015). “Classical Time Varying Factor-Augmented Vector Auto-Regressive Models-Estimation, Forecasting and Structural Analysis”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178.3, pp. 493–533. DOI: 10.1111/rssa.12068.
- Fahrmeir, L., T. Kneib, S. Lang, and B. Marx (2013). *Regression: Models, Methods and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-34333-9.
- Fama, E. F. (1965). “The Behavior of Stock-Market Prices”. In: *Journal of Business* 38.1, pp. 34–105.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). “The Generalized Dynamic-Factor Model: Identification and Estimation”. In: *The Review of Economics and Statistics* 82.4, pp. 540–554. DOI: 10.1198/016214504000002050.
- (2005). “The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting”. In: *Journal of the American Statistical Association* 100.471, pp. 830–840. DOI: 10.1198/016214504000002050.
- Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015). “Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations”. In: *Journal of Econometrics* 185.2, pp. 359–371. DOI: 10.1016/j.jeconom.2013.10.017.
- Fu, Z., Y. Hong, L. Su, and X. Wang (2023). “Specification Tests for Time-Varying Coefficient Models”. In: *Journal of Econometrics* 235.2, pp. 720–744. DOI: 10.1016/j.jeconom.2022.08.001.
- Geweke, J. (1977). *The Dynamic Factor Analysis of Economic Time Series*. Amsterdam: North-Holland Publishing Company.
- Ghosh, S., K. Khare, and G. Michailidis (2019). “High-Dimensional Posterior Consistency in Bayesian Vector Autoregressive Models”. In: *Journal of the American Statistical Association* 114.526, pp. 735–748. DOI: 10.1080/01621459.2018.1437043.
- Gibson, S. and B. Ninness (2005). “Robust Maximum-Likelihood Estimation of Multivariable Dynamic Systems”. In: *Automatica* 41.10, pp. 1667–1682. DOI: 10.1016/j.automatica.2005.05.008.
- Goldsmith, J., F. Scheipl, L. Huang, J. Wrobel, C. Di, J. Gellar, J. Harezlak, M. W. McLean, B. Swihart, L. Xiao, C. Crainiceanu, P. T. Reiss, and E. Cui (2024). *refund: Regression with Functional Data*. URL: <https://CRAN.R-project.org/package=refund>.
- Greven, S., C. Crainiceanu, B. Caffo, and D. Reich (2010). “Longitudinal Functional Principal Component Analysis”. In: *Electronic Journal of Statistics* 4. DOI: 10.1214/10-EJS575.
- Guo, L. (1990). “Estimating Time-Varying Parameters by the Kalman Filter Based Algorithm: Stability and Convergence”. In: *IEEE Transactions on Automatic Control* 35.2, pp. 141–147. DOI: 10.1109/9.45169.

- Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distributions*. Boca Raton, FL: Chapman & Hall.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, N.J: Princeton University Press.
- Hampel, F. R. (1968). *Contributions to the Theory of Robust Estimation*. Dissertation, University of California, Berkeley.
- Han, A. K. (1987). “Non-Parametric Analysis of a Generalized Regression Model”. In: *Journal of Econometrics* 35.2-3, pp. 303–316. DOI: 10.1016/0304-4076(87)90030-3.
- Harvey, D., S. Leybourne, and P. Newbold (1997). “Testing the Equality of Prediction Mean Squared Errors”. In: *International Journal of Forecasting* 13, pp. 281–291.
- Hettmansperger, T. P. and J. W. McKean (2011). *Robust Nonparametric Statistical Methods*. 2nd Edition. Boca Raton, FL: CRC Press.
- Hodges, J. L. and E. L. Lehmann (1963). “Estimates of Location Based on Rank Tests”. In: *The Annals of Mathematical Statistics* 34.2, pp. 598–611. DOI: 10.1214/aoms/1177704172.
- Hsiao, C. (2014). *Analysis of Panel Data*. 3rd Edition. New York: Cambridge Univ. Press.
- Huber, P. J. (1964). “Robust Estimation of a Location Parameter”. In: *The Annals of Mathematical Statistics* 35.1, pp. 73–101. DOI: 10.1214/aoms/1177703732.
- (1981). *Robust Statistics*. New York: Wiley.
- Hubert, M., M. Debruyne, and P. J. Rousseeuw (2018). “Minimum Covariance Determinant and Extensions”. In: *WIREs Computational Statistics* 10.3. DOI: 10.1002/wics.1421.
- Hubert, M., P. J. Rousseeuw, and P. Segaert (2015). “Multivariate Functional Outlier Detection”. In: *Statistical Methods & Applications* 24.2, pp. 177–202. DOI: 10.1007/s10260-015-0297-8.
- Hubrich, K. and T. Teräsvirta (2013). “Thresholds and Smooth Transitions in Vector Autoregressive Models”. In: *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims*. Ed. by T. B. Fomby, L. Kilian, and A. Murphy. Vol. 32. Emerald Group Publishing Limited, pp. 273–326. DOI: 10.1108/S0731-9053(2013)0000031008.
- Izenman, A. J. (1975). “Reduced-Rank Regression for the Multivariate Linear Model”. In: *Journal of Multivariate Analysis* 5.2, pp. 248–264. DOI: 10.1016/0047-259X(75)90042-1.
- Jaekel, L. A. (1972). “Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals”. In: *The Annals of Mathematical Statistics* 43.5, pp. 1449–1458. DOI: 10.1214/aoms/1177692377.
- Johnson, L. W. (1977). “Stochastic Parameter Regression: An Annotated Bibliography”. In: *International Statistical Review / Revue Internationale de Statistique* 45.3, pp. 257–272.
- Jung, S.-H. and Z. Ying (2003). “Rank-Based Regression with Repeated Measurements Data”. In: *Biometrika* 90.3, pp. 732–740. DOI: 10.1093/biomet/90.3.732.
- Jureckova, J. (1971). “Nonparametric Estimate of Regression Coefficients”. In: *The Annals of Mathematical Statistics* 42.4, pp. 1328–1338. DOI: 10.1214/aoms/1177693245.
- Jurečková, J. (1984). “M-, L- and R-estimators”. In: *Handbook of Statistics*. Vol. 4. Elsevier, pp. 463–485. DOI: 10.1016/S0169-7161(84)04023-2.
- Kalogridis, I. (2023). “Robust Thin-Plate Splines for Multivariate Spatial Smoothing”. In: *Econometrics and Statistics*, S2452306223000394. DOI: 10.1016/j.ecosta.2023.06.002.
- Kendall, M. G. (1953). “The Analysis of Time-Series - Part I: Prices”. In: *Journal of the Royal Statistical Society: Series A (General)* 116, pp. 11–25. DOI: 10.2307/2980947.

- Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*: 1st Edition. Cambridge University Press. DOI: 10.1017/9781108164818.
- Kloke, J. D., J. W. McKean, and M. M. Rashid (2009). “Rank-Based Estimation and Associated Inferences for Linear Models With Cluster Correlated Errors”. In: *Journal of the American Statistical Association* 104.485, pp. 384–390. DOI: 10.1198/jasa.2009.0116.
- Knoebl, K., G. Oreski, L. Neumaier, and B. Brune (2024). *SQL Database with Detailed Characteristic Measurement Results of Photovoltaic Modules That Were Subjected to Accelerated Aging Sequences*. Zenodo. DOI: 10.5281/ZENODO.13354817.
- Koenker, R., P. Ng, and S. Portnoy (1994). “Quantile Smoothing Splines”. In: *Biometrika* 81.4, pp. 673–680. DOI: 10.1093/biomet/81.4.673.
- Kokoszka, P. and M. Reimherr (2017). *Introduction to Functional Data Analysis*. 1st Edition. Chapman and Hall/CRC. DOI: 10.1201/9781315117416.
- Koller, M. (2013). *Robust Estimation of Linear Mixed Models*. Dissertation, ETH Zurich.
- (2016). “robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models”. In: *Journal of Statistical Software* 75.6. DOI: 10.18637/jss.v075.i06.
- Koner, S. and A.-M. Staicu (2023). “Second-Generation Functional Data”. In: *Annual Review of Statistics and Its Application* 10.1, pp. 547–572. DOI: 10.1146/annurev-statistics-032921-033726.
- Lang, M., B. Bischl, and D. Surmann (2017). “Batchtools: Tools for R to Work on Batch Systems”. In: *The Journal of Open Source Software* 2.10, p. 135. DOI: 10.21105/joss.00135.
- Lesaffre, E. and G. Verbeke (1998). “Local Influence in Linear Mixed Models”. In: *Biometrics* 54, pp. 570–584.
- Litterman, R. B. (1986). “Forecasting with Bayesian Vector Autoregressions: Five Years of Experience”. In: *Journal of Business & Economic Statistics* 4.1, pp. 25–38.
- Liu, L., D. Zhang, H. Liu, and C. Arendt (2013). “Robust Methods for Population Stratification in Genome Wide Association Studies”. In: *BMC Bioinformatics* 14, p. 132. DOI: 10.1186/1471-2105-14-132.
- Lubik, T. and C. Matthes (2016). “Time-Varying Parameter Vector Autoregressions: Specification, Estimation, and an Application”. In: *Economic Quarterly* 101.04, pp. 323–352. DOI: 10.21144/eq1010403.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: New York : Springer. DOI: 10.1007/978-3-540-27752-1.
- Maechler, M., P. J. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, E. L. T. Conceicao, and M. Anna di Palma (2024). *robustbase: Basic Robust Statistics*. URL: <http://robustbase.r-forge.r-project.org/>.
- Maronna, R. A. (2019). *Robust Statistics: Theory and Methods (with R)*. 2nd Edition. Hoboken, NJ: Wiley.
- McKean, J. W. (2004). “Robust Analysis of Linear Models”. In: *Statistical Science* 19.4, pp. 562–570. DOI: 10.1214/088342304000000549.
- Meinhold, R. J. and N. D. Singpurwalla (1989). “Robustification of Kalman Filter Models”. In: *Journal of the American Statistical Association* 84.406, pp. 479–486. DOI: 10.1080/01621459.1989.10478794.
- Moritz, S. and T. Bartz-Beielstein (2017). “imputeTS: Time Series Missing Value Imputation in R”. In: *The R Journal* 9.1, p. 207. DOI: 10.32614/RJ-2017-009.

- Nicholson, W. B., D. S. Matteson, and J. Bien (2017). “VARX-L: Structured Regularization for Large Vector Autoregressions With Exogenous Variables”. In: *International Journal of Forecasting* 33.3, pp. 627–651. DOI: 10.1016/j.ijforecast.2017.01.003.
- Oguamalam, J., U. Radojičić, and P. Filzmoser (2024). “Minimum Regularized Covariance Trace Estimator and Outlier Detection for Functional Data”. In: *Technometrics*. DOI: 10.1080/00401706.2024.2336542.
- Park, S. Y. and A.-M. Staicu (2015). “Longitudinal Functional Data Analysis”. In: *STAT* 4.1, pp. 212–226. DOI: 10.1002/sta4.89.
- Peña, D. and P. Poncela (2006). “Dimension Reduction in Multivariate Time Series”. In: *Advances in Distribution Theory, Order Statistics, and Inference*. Ed. by N. Balakrishnan, J. M. Sarabia, and E. Castillo. Boston, MA: Birkhäuser, pp. 433–458. DOI: 10.1007/0-8176-4487-3_28.
- Peña, D. and R. S. Tsay (2020). *Statistical Learning with Big Dependent Data*. New York, NY: Wiley New York.
- Peña, D. and V. J. Yohai (2016). “Generalized Dynamic Principal Components”. In: *Journal of the American Statistical Association* 111.515, pp. 1121–1131. DOI: 10.1080/01621459.2015.1072542.
- Pinheiro, J. C., C. Liu, and Y. N. Wu (2001). “Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using the Multivariate t Distribution”. In: *Journal of Computational and Graphical Statistics* 10.2, pp. 249–276. DOI: 10.1198/10618600152628059.
- Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Primiceri, G. E. (2005). “Time Varying Structural Vector Autoregressions and Monetary Policy”. In: *Review of Economic Studies* 72.(3), pp. 821–852.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. New York, NY: Springer New York. DOI: 10.1007/b98888.
- Ramsay, J. O., T. Ramsay, and L. M. Sangalli (2011). “Spatial Functional Data Analysis”. In: *Recent Advances in Functional Data Analysis and Related Topics*. Ed. by F. Ferraty. Heidelberg: Physica-Verlag HD, pp. 269–275. DOI: 10.1007/978-3-7908-2736-1_42.
- Raymaekers, J. and P. J. Rousseeuw (2024). “Challenges of Cellwise Outliers”. In: *Econometrics and Statistics*. DOI: 10.1016/j.ecosta.2024.02.002.
- Reinsel, G. C. and R. P. Velu (1998). *Multivariate Reduced-Rank Regression*. New York, NY: Springer. DOI: 10.1007/978-1-4757-2853-8.
- Rosenberg, B. (1973). “A Survey of Stochastic Parameter Regression”. In: *Annals of Economic and Social Measurement, Volume 2, number 4*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 381–397. URL: <http://www.nber.org/chapters/c9933>.
- Rousseeuw, P. J. (1984). “Least Median of Squares Regression”. In: *Journal of the American Statistical Association* 79.388, pp. 871–880. DOI: 10.1080/01621459.1984.10477105.
- (1985). “Multivariate Estimation with High Breakdown Point”. In: *Mathematical Statistics and Applications*. Ed. by W. Grossmann, G. C. Pflug, I. Vincze, and W. Wertz. Dordrecht: Springer Netherlands, pp. 283–297. DOI: 10.1007/978-94-009-5438-0_20.

- Rousseeuw, P. J. and C. Croux (1993). “Alternatives to the Median Absolute Deviation”. In: *Journal of the American Statistical Association* 88.424, pp. 1273–1283. DOI: 10.1080/01621459.1993.10476408.
- Rousseeuw, P. J. and K. V. Driessen (1999). “A Fast Algorithm for the Minimum Covariance Determinant Estimator”. In: *Technometrics* 41.3, pp. 212–223. DOI: 10.1080/00401706.1999.10485670.
- Rousseeuw, P. J. and K. Van Driessen (2006). “Computing LTS Regression for Large Data Sets”. In: *Data Mining and Knowledge Discovery* 12.1, pp. 29–45. DOI: 10.1007/s10618-005-0024-4.
- Sargent, T. and C. Sims (1977). “Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory”. In: *New Methods in Business Cycle Research: Proceedings from a Conference*. Ed. by C. A. Sims. Minneapolis: Federal Reserve Bank of Minneapolis, pp. 45–109.
- Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge, U.K.; New York: Cambridge University Press. DOI: 10.1017/CB09781139344203.
- Schwarz, G. (1978). “Estimating the Dimension of a Model”. In: *The Annals of Statistics* 6.2, pp. 461–464. DOI: 10.1214/aos/1176344136.
- Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. 1st Edition. Wiley. DOI: 10.1002/9780470316856.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Hoboken, NJ, USA: John Wiley & Sons, Inc. DOI: 10.1002/9780470316481.
- Shumway, R. H. and D. S. Stoffer (2017). *Time Series Analysis and Its Applications: With R Examples*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-52452-8.
- Sims, C. and T. Zha (1998). “Bayesian Methods for Dynamic Multivariate Models”. In: *International Economic Review* 39.4, pp. 949–68. DOI: 10.2307/2527347.
- Singer, J. M. and D. F. Andrade (2000). “Analysis of Longitudinal Data”. In: *Handbook of Statistics*. Vol. 18. Elsevier, pp. 115–160. ISBN: 978-0-444-82900-9. DOI: 10.1016/S0169-7161(00)18007-1.
- Song, S. and P. J. Bickel (2011). *Large Vector Auto Regressions*. arXiv:1106.3915 [q-fin, stat]. URL: <https://arxiv.org/pdf/1106.3915v1.pdf>.
- Stock, J. H. and M. W. Watson (2002). “Forecasting Using Principal Components from a Large Number of Predictors”. In: *Journal of the American Statistical Association* 97.460, pp. 1167–1179. DOI: 10.1198/016214502388618960.
- (2005). *Implications of Dynamic Factor Models for VAR Analysis*. Working Paper 11467. National Bureau of Economic Research. DOI: 10.3386/w11467.
- (2009). “Forecasting in Dynamic Factor Models Subject to Structural Instability”. In: *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Edited by Jennifer Castle and Neil Shepard. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199237197.003.0007.
- (2012). “Dynamic Factor Models”. In: *The Oxford Handbook of Economic Forecasting*. Ed. by M. P. Clements and D. F. Hendry. 1st Edition. Oxford University Press, pp. 35–60. ISBN: 978-0-19-539864-9 978-0-19-994032-5. DOI: 10.1093/oxfordhb/9780195398649.013.0003.
- Su, L. and X. Wang (2017). “On Time-Varying Factor Models: Estimation and Testing”. In: *Journal of Econometrics* 198.1, pp. 84–101. DOI: 10.1016/j.jeconom.2016.12.004.

- Tompkins, F. and P. J. Wolfe (2007). “Bayesian Filtering on the Stiefel Manifold”. In: *2007 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*. St. Thomas, VI: IEEE. ISBN: 978-1-4244-1713-1. DOI: 10.1109/CAMSAP.2007.4498015.
- Tsay, R. S. (2014). *Multivariate Time Series Analysis: With R and Financial Applications*. Hoboken, New Jersey: John Wiley & Sons.
- Tukey, J. W. (1960). “A Survey of Sampling from Contaminated Distributions”. In: *Contributions to Probability and Statistics*, pp. 448–485.
- Velu, R. P., G. C. Reinsel, and D. W. Wichern (1986). “Reduced Rank Models for Multiple Time Series”. In: *Biometrika* 73.1, pp. 105–118. DOI: 10.1093/biomet/73.1.105.
- Verbeke, G., S. Fieuws, G. Molenberghs, and M. Davidian (2014). “The Analysis of Multivariate Longitudinal Data: A Review”. In: *Statistical Methods in Medical Research* 23.1, pp. 42–59.
- Wang, G., S. Liu, F. Han, and C.-Z. Di (2022). “Robust Functional Principal Component Analysis via a Functional Pairwise Spatial Sign Operator”. In: *Biometrics*. DOI: 10.1111/biom.13695.
- Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). “Functional Data Analysis”. In: *Annual Review of Statistics and Its Application* 3.1, pp. 257–295. DOI: 10.1146/annurev-statistics-041715-033624.
- Wang, Y.-G. and Y. Zhao (2008). “Weighted Rank Regression for Clustered Data Analysis”. In: *Biometrics* 64.1, pp. 39–45. DOI: 10.1111/j.1541-0420.2007.00842.x.
- Wang, Y.-G. and M. Zhu (2006). “Rank-based Regression for Analysis of Repeated Measures”. In: *Biometrika* 93.2, pp. 459–464. DOI: 10.1093/biomet/93.2.459.
- Wei, W. W. (2019). *Multivariate Time Series Analysis and Applications*. 1st Edition. Wiley. DOI: 10.1002/9781119502951.
- Wilcox, R. R. and L. Xu (2022). “Regression: Identifying Good and Bad Leverage Points”. In: *International Journal of Statistics and Probability* 12.1, p. 1. DOI: 10.5539/ijsp.v12n1p1.
- Xiao, L., C. Li, W. Checkley, and C. Crainiceanu (2018). “Fast Covariance Estimation for Sparse Functional Data”. In: 28.3, pp. 511–522. DOI: 10.1007/s11222-017-9744-8.
- Yang, Y. and L. Bauwens (2018). “State-Space Models on the Stiefel Manifold with a New Approach to Nonlinear Filtering”. In: *Econometrics* 6.4, p. 48. DOI: 10.3390/econometrics6040048.
- Yao, F., H.-G. Müller, and J.-L. Wang (2005). “Functional Data Analysis for Sparse Longitudinal Data”. In: *Journal of the American Statistical Association* 100.470, pp. 577–590. DOI: 10.1198/016214504000001745.
- Yohai, V. J. (1987). “High Breakdown-Point and High Efficiency Robust Estimates for Regression”. In: *The Annals of Statistics* 15.2. DOI: 10.1214/aos/1176350366.

A | Supplement: Chapter 2

Description of the content:

This appendix contains supplementary tables with additional simulation results for Chapter 2. Additional results for the rank-selection procedure based on BIC are given in Section A.1. We report the estimation performance for the fixed coefficients in Section A.2, for deterministic coefficient transitions in A.3, and for the random walk coefficient evolution in A.4. The material was published as a supplement to Brune, Scherrer, et al. (2022).

A.1. Rank selection

Table A.1.: Results of rank selection via BIC criterion for different combinations of p and q and the deterministic coefficient transition, $10 \cdot 10 = 100$ repetitions each. “=” is the fraction of correct rank estimation, and “<” the fraction of underestimated ranks, and analogously “>” the fraction of overestimated ranks.

p	d	q	t	Model (A)			Model (B)		
				=	<	>	=	<	>
10	1	5	100	1.000	0.000	0.000	1.000	0.000	0.000
10	1	5	200	0.950	0.000	0.050	1.000	0.000	0.000
10	1	5	300	0.939	0.000	0.061	1.000	0.000	0.000
10	2	5	100	0.970	0.030	0.000	0.700	0.300	0.000
10	2	5	200	0.980	0.000	0.020	1.000	0.000	0.000
10	2	5	300	0.889	0.000	0.111	1.000	0.000	0.000
10	1	10	100	1.000	0.000	0.000	1.000	0.000	0.000
10	1	10	200	1.000	0.000	0.000	1.000	0.000	0.000
10	1	10	300	1.000	0.000	0.000	1.000	0.000	0.000
10	2	10	100	0.890	0.110	0.000	0.750	0.250	0.000
10	2	10	200	0.990	0.010	0.000	1.000	0.000	0.000
10	2	10	300	1.000	0.000	0.000	1.000	0.000	0.000
10	3	10	100	0.430	0.570	0.000	0.250	0.750	0.000
10	3	10	200	0.980	0.020	0.000	0.970	0.030	0.000
10	3	10	300	1.000	0.000	0.000	1.000	0.000	0.000
20	1	10	100	0.990	0.000	0.010	1.000	0.000	0.000
20	1	10	200	0.990	0.000	0.010	1.000	0.000	0.000
20	1	10	300	0.988	0.000	0.012	1.000	0.000	0.000
20	2	10	100	0.940	0.060	0.000	0.590	0.410	0.000
20	2	10	200	0.970	0.000	0.030	0.990	0.010	0.000
20	2	10	300	0.927	0.000	0.073	1.000	0.000	0.000
20	3	10	100	0.650	0.340	0.010	0.290	0.710	0.000
20	3	10	200	0.970	0.010	0.020	0.914	0.086	0.000
20	3	10	300	0.802	0.000	0.198	1.000	0.000	0.000

Table A.2.: Results of rank selection via BIC criterion for different combinations of p and q and the random walk coefficient transition, $10 \cdot 10 = 100$ repetitions each. “=” is the fraction of correct rank estimation, and “<” the fraction of underestimated ranks, and analogously “>” the fraction of overestimated ranks.

p	d	q	t	Model (A)			Model (B)		
				=	<	>	=	<	>
10	1	5	100	1.000	0.000	0.000	1.000	0.000	0.000
10	1	5	200	1.000	0.000	0.000	1.000	0.000	0.000
10	1	5	300	0.990	0.000	0.010	1.000	0.000	0.000
10	2	5	100	0.140	0.860	0.000	0.000	1.000	0.000
10	2	5	200	0.830	0.170	0.000	0.130	0.870	0.000
10	2	5	300	1.000	0.000	0.000	0.490	0.510	0.000
10	1	10	100	1.000	0.000	0.000	1.000	0.000	0.000
10	1	10	200	1.000	0.000	0.000	1.000	0.000	0.000
10	1	10	300	1.000	0.000	0.000	1.000	0.000	0.000
10	2	10	100	0.020	0.980	0.000	0.040	0.960	0.000
10	2	10	200	0.570	0.430	0.000	0.480	0.520	0.000
10	2	10	300	0.969	0.031	0.000	1.000	0.000	0.000
10	3	10	100	0.000	1.000	0.000	0.000	1.000	0.000
10	3	10	200	0.110	0.890	0.000	0.110	0.890	0.000
10	3	10	300	0.710	0.290	0.000	0.788	0.212	0.000
20	1	10	100	1.000	0.000	0.000	1.000	0.000	0.000
20	1	10	200	1.000	0.000	0.000	1.000	0.000	0.000
20	1	10	300	1.000	0.000	0.000	1.000	0.000	0.000
20	2	10	100	0.440	0.560	0.000	0.010	0.990	0.000
20	2	10	200	1.000	0.000	0.000	0.406	0.594	0.000
20	2	10	300	1.000	0.000	0.000	0.900	0.100	0.000
20	3	10	100	0.000	1.000	0.000	0.000	1.000	0.000
20	3	10	200	0.860	0.140	0.000	0.100	0.900	0.000
20	3	10	300	0.990	0.010	0.000	0.640	0.360	0.000

Table A.3.: Results of rank selection via BIC criterion for different combinations of p and q and the structural break coefficient transition, $10 \cdot 10 = 100$ repetitions each. “=” is the fraction of correct rank estimation, and “<” the fraction of underestimated ranks, and analogously “>” the fraction of overestimated ranks.

p	d	q	t	Model (A)			Model (B)		
				=	<	>	=	<	>
10	1	5	100	1.000	0.000	0.000	1.000	0.000	0.000
10	1	5	200	1.000	0.000	0.000	1.000	0.000	0.000
10	1	5	300	0.990	0.000	0.010	1.000	0.000	0.000
10	2	5	100	1.000	0.000	0.000	0.850	0.150	0.000
10	2	5	200	1.000	0.000	0.000	1.000	0.000	0.000
10	2	5	300	1.000	0.000	0.000	1.000	0.000	0.000
10	1	10	100	1.000	0.000	0.000	1.000	0.000	0.000
10	1	10	200	1.000	0.000	0.000	1.000	0.000	0.000
10	1	10	300	1.000	0.000	0.000	1.000	0.000	0.000
10	2	10	100	0.810	0.190	0.000	0.870	0.130	0.000
10	2	10	200	1.000	0.000	0.000	1.000	0.000	0.000
10	2	10	300	1.000	0.000	0.000	1.000	0.000	0.000
10	3	10	100	0.570	0.430	0.000	0.580	0.420	0.000
10	3	10	200	1.000	0.000	0.000	1.000	0.000	0.000
10	3	10	300	1.000	0.000	0.000	1.000	0.000	0.000
20	1	10	100	1.000	0.000	0.000	1.000	0.000	0.000
20	1	10	200	0.990	0.000	0.010	1.000	0.000	0.000
20	1	10	300	1.000	0.000	0.000	1.000	0.000	0.000
20	2	10	100	0.980	0.020	0.000	0.880	0.120	0.000
20	2	10	200	0.990	0.000	0.010	1.000	0.000	0.000
20	2	10	300	0.990	0.000	0.010	1.000	0.000	0.000
20	3	10	100	0.860	0.140	0.000	0.730	0.270	0.000
20	3	10	200	1.000	0.000	0.000	1.000	0.000	0.000
20	3	10	300	0.990	0.000	0.010	1.000	0.000	0.000

A.2. Simulation results for fixed coefficients

Table A.4.: Average MSE and relative parameter estimation error for fixed coefficient setting in model A with exogenous predictors, standard deviation in parentheses.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	0.9807 (0.1102)	0.9046 (0.1052)	0.0214 (0.0038)	0.0887 (0.0111)
5	5	1	out of sample	1.0154 (0.0987)	1.0391 (0.1018)		
5	5	2	in sample	1.0056 (0.1438)	0.8734 (0.1251)	0.0161 (0.0021)	0.0760 (0.0085)
5	5	2	out of sample	1.0442 (0.1506)	1.0857 (0.1506)		
5	5	3	in sample	0.9019 (0.1379)	0.7525 (0.1154)	0.0140 (0.0034)	0.0638 (0.0154)
5	5	3	out of sample	0.9550 (0.1429)	1.0132 (0.1512)		
5	10	1	in sample	1.0764 (0.1131)	0.9941 (0.1080)	0.0329 (0.0069)	0.1016 (0.0176)
5	10	1	out of sample	1.1098 (0.1173)	1.1326 (0.1194)		
5	10	2	in sample	1.0119 (0.1637)	0.8833 (0.1494)	0.0311 (0.0061)	0.0910 (0.0142)
5	10	2	out of sample	1.0913 (0.1648)	1.1338 (0.1689)		
5	10	3	in sample	1.0225 (0.1632)	0.8527 (0.1346)	0.0296 (0.0052)	0.0871 (0.0131)
5	10	3	out of sample	1.1185 (0.1820)	1.1754 (0.1907)		
10	5	1	in sample	0.9769 (0.0887)	0.9190 (0.0852)	0.0272 (0.0057)	0.1219 (0.0269)
10	5	1	out of sample	0.9977 (0.0960)	1.0146 (0.0968)		
10	5	2	in sample	0.9666 (0.0704)	0.8699 (0.0657)	0.0267 (0.0029)	0.1110 (0.0102)
10	5	2	out of sample	0.9969 (0.0655)	1.0294 (0.0674)		
10	5	3	in sample	0.9858 (0.1042)	0.8540 (0.0942)	0.0263 (0.0047)	0.1064 (0.0093)
10	5	3	out of sample	1.0363 (0.1047)	1.0805 (0.1086)		
10	10	1	in sample	0.9731 (0.0467)	0.9194 (0.0502)	0.0461 (0.0067)	0.1445 (0.0124)
10	10	1	out of sample	1.0112 (0.0527)	1.0269 (0.0507)		
10	10	2	in sample	0.9815 (0.1201)	0.8802 (0.1162)	0.0383 (0.0080)	0.1231 (0.0138)
10	10	2	out of sample	1.0419 (0.1401)	1.0736 (0.1408)		
10	10	3	in sample	0.9224 (0.1054)	0.7973 (0.0983)	0.0386 (0.0091)	0.1189 (0.0172)
10	10	3	out of sample	0.9919 (0.1105)	1.0316 (0.1144)		
10	20	1	in sample	0.9946 (0.0715)	0.9402 (0.0699)	0.0820 (0.0144)	0.1763 (0.0193)
10	20	1	out of sample	1.0328 (0.0741)	1.0462 (0.0741)		
10	20	2	in sample	0.9969 (0.1004)	0.8957 (0.0962)	0.0812 (0.0178)	0.1737 (0.0206)
10	20	2	out of sample	1.0801 (0.1078)	1.1067 (0.1055)		
10	20	3	in sample	0.9829 (0.0697)	0.8472 (0.0628)	0.0727 (0.0064)	0.1625 (0.0100)
10	20	3	out of sample	1.0924 (0.0752)	1.1332 (0.0756)		
20	10	1	in sample	0.9906 (0.0570)	0.9529 (0.0587)	0.0645 (0.0087)	0.1831 (0.0173)
20	10	1	out of sample	1.0145 (0.0545)	1.0228 (0.0538)		
20	10	2	in sample	1.0100 (0.0628)	0.9398 (0.0601)	0.0638 (0.0084)	0.1811 (0.0110)
20	10	2	out of sample	1.0559 (0.0638)	1.0753 (0.0645)		
20	10	3	in sample	0.9525 (0.0787)	0.8568 (0.0781)	0.0613 (0.0093)	0.1719 (0.0160)
20	10	3	out of sample	1.0025 (0.0849)	1.0281 (0.0853)		
20	20	1	in sample	0.9734 (0.0402)	0.9362 (0.0384)	0.1003 (0.0149)	0.2172 (0.0295)
20	20	1	out of sample	1.0059 (0.0349)	1.0129 (0.0347)		
20	20	2	in sample	0.9629 (0.0446)	0.8942 (0.0428)	0.0972 (0.0097)	0.2143 (0.0181)
20	20	2	out of sample	1.0231 (0.0521)	1.0406 (0.0528)		
20	20	3	in sample	0.9673 (0.0424)	0.8689 (0.0434)	0.1028 (0.0107)	0.2220 (0.0150)
20	20	3	out of sample	1.0475 (0.0501)	1.0723 (0.0500)		

Table A.5.: Average MSE and relative parameter estimation error for fixed coefficient setting in model B with exogenous predictors, standard deviation in parentheses.

p	q	d	sample	MSE (RRR)	MSE (τ vRRR)	err_{C_t} (RRR)	err_{C_t} (τ vRRR)
5	5	1	in sample	0.9322 (0.0981)	0.8761 (0.0973)	0.0181 (0.0049)	0.0838 (0.0139)
5	5	1	out of sample	0.9558 (0.0870)	0.9812 (0.0882)		
5	5	2	in sample	0.9768 (0.1332)	0.8748 (0.1229)	0.0166 (0.0039)	0.0748 (0.0109)
5	5	2	out of sample	1.0050 (0.1364)	1.0478 (0.1422)		
5	5	3	in sample	0.9998 (0.0821)	0.8479 (0.0719)	0.0149 (0.0028)	0.0700 (0.0080)
5	5	3	out of sample	1.0538 (0.0971)	1.1161 (0.1029)		
5	10	1	in sample	0.9147 (0.1456)	0.8498 (0.1397)	0.0312 (0.0060)	0.1286 (0.0121)
5	10	1	out of sample	0.9493 (0.1446)	0.9794 (0.1452)		
5	10	2	in sample	0.9719 (0.1170)	0.8505 (0.1048)	0.0295 (0.0062)	0.1162 (0.0131)
5	10	2	out of sample	1.0313 (0.1144)	1.0828 (0.1125)		
5	10	3	in sample	0.9267 (0.1084)	0.7477 (0.0898)	0.0290 (0.0043)	0.1067 (0.0145)
5	10	3	out of sample	1.0271 (0.1250)	1.0987 (0.1289)		
10	5	1	in sample	0.9615 (0.0995)	0.9346 (0.0990)	0.0289 (0.0064)	0.1108 (0.0122)
10	5	1	out of sample	0.9895 (0.1006)	1.0018 (0.0993)		
10	5	2	in sample	0.9917 (0.0430)	0.9367 (0.0435)	0.0235 (0.0038)	0.0938 (0.0088)
10	5	2	out of sample	1.0285 (0.0443)	1.0542 (0.0420)		
10	5	3	in sample	0.9602 (0.1073)	0.8849 (0.1039)	0.0251 (0.0038)	0.0901 (0.0081)
10	5	3	out of sample	1.0070 (0.1096)	1.0417 (0.1074)		
10	10	1	in sample	1.0077 (0.0758)	0.9749 (0.0750)	0.0408 (0.0070)	0.1571 (0.0153)
10	10	1	out of sample	1.0386 (0.0786)	1.0567 (0.0799)		
10	10	2	in sample	0.9855 (0.0912)	0.9234 (0.0911)	0.0392 (0.0057)	0.1319 (0.0082)
10	10	2	out of sample	1.0408 (0.0997)	1.0725 (0.0991)		
10	10	3	in sample	0.9998 (0.0755)	0.9014 (0.0706)	0.0402 (0.0046)	0.1394 (0.0094)
10	10	3	out of sample	1.0666 (0.0816)	1.1142 (0.0822)		
10	20	1	in sample	0.9586 (0.1177)	0.9223 (0.1135)	0.0816 (0.0182)	0.2489 (0.0163)
10	20	1	out of sample	1.0072 (0.1329)	1.0253 (0.1303)		
10	20	2	in sample	1.0131 (0.0666)	0.9392 (0.0649)	0.0828 (0.0095)	0.2190 (0.0125)
10	20	2	out of sample	1.1058 (0.0745)	1.1392 (0.0752)		
10	20	3	in sample	0.9393 (0.0745)	0.8349 (0.0710)	0.0764 (0.0073)	0.2101 (0.0149)
10	20	3	out of sample	1.0543 (0.0984)	1.1045 (0.0950)		
20	10	1	in sample	0.9733 (0.0682)	0.9573 (0.0675)	0.0641 (0.0169)	0.1827 (0.0155)
20	10	1	out of sample	1.0003 (0.0681)	1.0071 (0.0692)		
20	10	2	in sample	0.9662 (0.0779)	0.9337 (0.0757)	0.0602 (0.0096)	0.1595 (0.0125)
20	10	2	out of sample	1.0033 (0.0792)	1.0144 (0.0799)		
20	10	3	in sample	0.9528 (0.0598)	0.9034 (0.0586)	0.0577 (0.0060)	0.1558 (0.0086)
20	10	3	out of sample	1.0090 (0.0593)	1.0299 (0.0597)		
20	20	1	in sample	0.9642 (0.0576)	0.9492 (0.0565)	0.0986 (0.0175)	0.2397 (0.0271)
20	20	1	out of sample	0.9902 (0.0568)	0.9951 (0.0567)		
20	20	2	in sample	0.9337 (0.0657)	0.9002 (0.0641)	0.1000 (0.0094)	0.2278 (0.0060)
20	20	2	out of sample	0.9932 (0.0666)	1.0023 (0.0663)		
20	20	3	in sample	0.9416 (0.0486)	0.8900 (0.0459)	0.0998 (0.0080)	0.2197 (0.0190)
20	20	3	out of sample	1.0232 (0.0523)	1.0361 (0.0489)		

A.3. Simulation results for deterministic coefficient transition

Table A.6.: Average MSE and relative parameter estimation error for deterministic coefficient transition in model A with external predictors, standard deviation in parentheses, **small** variation.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	1.0936 (0.0969)	0.9411 (0.0873)	0.2499 (0.1343)	0.1500 (0.0624)
5	5	1	out of sample	2.0812 (0.5109)	1.0764 (0.0827)		
5	5	2	in sample	1.1393 (0.1511)	0.8599 (0.1491)	0.3098 (0.1405)	0.1627 (0.0460)
5	5	2	out of sample	3.9652 (1.2012)	1.1188 (0.1750)		
5	5	3	in sample	1.2447 (0.1305)	0.8087 (0.0938)	0.3033 (0.1080)	0.1573 (0.0462)
5	5	3	out of sample	5.3751 (1.6562)	1.1696 (0.1236)		
5	10	1	in sample	1.0708 (0.1526)	0.9137 (0.1463)	0.3100 (0.1491)	0.1713 (0.0548)
5	10	1	out of sample	2.2180 (0.5064)	1.0837 (0.1664)		
5	10	2	in sample	1.0479 (0.1605)	0.7916 (0.1300)	0.2385 (0.1167)	0.1470 (0.0439)
5	10	2	out of sample	2.8148 (0.3714)	1.1358 (0.1697)		
5	10	3	in sample	1.2430 (0.1272)	0.7986 (0.0954)	0.3827 (0.1238)	0.1954 (0.0509)
5	10	3	out of sample	5.1495 (1.1131)	1.2845 (0.1850)		
10	5	1	in sample	1.0480 (0.1213)	0.9585 (0.1260)	0.3616 (0.1435)	0.2692 (0.0920)
10	5	1	out of sample	1.7730 (0.3276)	1.0632 (0.1308)		
10	5	2	in sample	1.0839 (0.1430)	0.8955 (0.1320)	0.3058 (0.0814)	0.2204 (0.0478)
10	5	2	out of sample	2.2815 (0.3659)	1.0926 (0.1573)		
10	5	3	in sample	1.1193 (0.0942)	0.8451 (0.0856)	0.3044 (0.0719)	0.2123 (0.0407)
10	5	3	out of sample	2.9828 (0.3925)	1.1154 (0.0972)		
10	10	1	in sample	1.0096 (0.1119)	0.9151 (0.1028)	0.3793 (0.1383)	0.2894 (0.0879)
10	10	1	out of sample	1.6575 (0.2608)	1.0450 (0.1115)		
10	10	2	in sample	1.0801 (0.0607)	0.8848 (0.0680)	0.3770 (0.1202)	0.2831 (0.0741)
10	10	2	out of sample	2.3940 (0.3604)	1.1295 (0.0783)		
10	10	3	in sample	1.0660 (0.0779)	0.8049 (0.0634)	0.3351 (0.0531)	0.2548 (0.0329)
10	10	3	out of sample	2.9252 (0.4102)	1.1117 (0.0980)		
10	20	1	in sample	0.9990 (0.0816)	0.9031 (0.0779)	0.3973 (0.1661)	0.3078 (0.0777)
10	20	1	out of sample	1.5679 (0.2783)	1.0685 (0.0846)		
10	20	2	in sample	1.0702 (0.1208)	0.8903 (0.1033)	0.3550 (0.0922)	0.2944 (0.0529)
10	20	2	out of sample	2.2421 (0.5078)	1.2079 (0.1526)		
10	20	3	in sample	1.0417 (0.1175)	0.7850 (0.1013)	0.4278 (0.1180)	0.3143 (0.0601)
10	20	3	out of sample	3.1284 (0.6008)	1.2371 (0.1716)		
20	10	1	in sample	1.0057 (0.0442)	0.9574 (0.0407)	0.3783 (0.0598)	0.3493 (0.0413)
20	10	1	out of sample	1.3823 (0.0875)	1.0393 (0.0448)		
20	10	2	in sample	1.0413 (0.0576)	0.9351 (0.0534)	0.3365 (0.0370)	0.3028 (0.0401)
20	10	2	out of sample	1.7926 (0.1430)	1.0933 (0.0692)		
20	10	3	in sample	1.0674 (0.0541)	0.9057 (0.0634)	0.3569 (0.0603)	0.3175 (0.0351)
20	10	3	out of sample	2.0777 (0.2764)	1.1366 (0.0677)		
20	20	1	in sample	0.9987 (0.0778)	0.9531 (0.0780)	0.4110 (0.1420)	0.3772 (0.0807)
20	20	1	out of sample	1.3770 (0.1701)	1.0590 (0.0829)		
20	20	2	in sample	1.0036 (0.0722)	0.8974 (0.0556)	0.3968 (0.0877)	0.3636 (0.0464)
20	20	2	out of sample	1.6642 (0.2339)	1.0875 (0.0687)		
20	20	3	in sample	1.0271 (0.0662)	0.8900 (0.0640)	0.3861 (0.0700)	0.3615 (0.0556)
20	20	3	out of sample	1.9989 (0.2183)	1.1689 (0.0876)		

Table A.7.: Average MSE and relative parameter estimation error for deterministic coefficient transition in model A with external predictors, standard deviation in parentheses, **medium** variation.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	1.1343 (0.1569)	0.8354 (0.1029)	0.2830 (0.2234)	0.0957 (0.0571)
5	5	1	out of sample	3.6796 (1.6489)	1.0202 (0.1195)		
5	5	2	in sample	1.2982 (0.1990)	0.8209 (0.1132)	0.2213 (0.1081)	0.0869 (0.0392)
5	5	2	out of sample	5.8776 (2.3212)	1.1471 (0.1668)		
5	5	3	in sample	1.4735 (0.2213)	0.6840 (0.0838)	0.2587 (0.1225)	0.0937 (0.0344)
5	5	3	out of sample	8.5910 (2.9719)	1.1278 (0.1342)		
5	10	1	in sample	1.1823 (0.1574)	0.9014 (0.1197)	0.3290 (0.1958)	0.1298 (0.0686)
5	10	1	out of sample	3.4023 (0.8751)	1.1421 (0.1557)		
5	10	2	in sample	1.2753 (0.0950)	0.8443 (0.0652)	0.1988 (0.0842)	0.0901 (0.0197)
5	10	2	out of sample	5.1298 (1.7455)	1.2300 (0.0747)		
5	10	3	in sample	1.5778 (0.2242)	0.7198 (0.0987)	0.3552 (0.1421)	0.1205 (0.0373)
5	10	3	out of sample	10.4495 (3.6514)	1.3272 (0.1795)		
10	5	1	in sample	1.0954 (0.0901)	0.9357 (0.0816)	0.2403 (0.0830)	0.1292 (0.0322)
10	5	1	out of sample	2.3112 (0.3045)	1.0850 (0.0891)		
10	5	2	in sample	1.1775 (0.1014)	0.8603 (0.0686)	0.2490 (0.0468)	0.1229 (0.0141)
10	5	2	out of sample	3.9848 (0.8962)	1.1184 (0.0843)		
10	5	3	in sample	1.2008 (0.0725)	0.7804 (0.0484)	0.2227 (0.0331)	0.1198 (0.0120)
10	5	3	out of sample	4.5429 (0.6976)	1.1182 (0.0607)		
10	10	1	in sample	1.1022 (0.1054)	0.9334 (0.0590)	0.2608 (0.1483)	0.1502 (0.0646)
10	10	1	out of sample	2.1126 (0.4480)	1.0875 (0.0734)		
10	10	2	in sample	1.0666 (0.1011)	0.7717 (0.0777)	0.2096 (0.0474)	0.1141 (0.0163)
10	10	2	out of sample	3.4841 (0.6698)	1.0365 (0.0978)		
10	10	3	in sample	1.2539 (0.0764)	0.7955 (0.0802)	0.2670 (0.0665)	0.1405 (0.0246)
10	10	3	out of sample	4.6271 (0.7045)	1.1954 (0.1063)		
10	20	1	in sample	1.0905 (0.0650)	0.9275 (0.0600)	0.3264 (0.0996)	0.1754 (0.0276)
10	20	1	out of sample	2.7496 (0.5880)	1.1409 (0.0686)		
10	20	2	in sample	1.1496 (0.0659)	0.8517 (0.0646)	0.2913 (0.0938)	0.1738 (0.0299)
10	20	2	out of sample	3.8083 (0.7630)	1.2286 (0.0749)		
10	20	3	in sample	1.2298 (0.1100)	0.7976 (0.0806)	0.3352 (0.0683)	0.1937 (0.0311)
10	20	3	out of sample	5.4292 (1.1519)	1.3674 (0.1649)		
20	10	1	in sample	0.9958 (0.0606)	0.8988 (0.0469)	0.3256 (0.0785)	0.2257 (0.0472)
20	10	1	out of sample	1.7255 (0.0962)	1.0114 (0.0527)		
20	10	2	in sample	1.0557 (0.0917)	0.8777 (0.0696)	0.2780 (0.0508)	0.1976 (0.0262)
20	10	2	out of sample	2.3299 (0.3639)	1.0775 (0.0867)		
20	10	3	in sample	1.1001 (0.0441)	0.8310 (0.0540)	0.2913 (0.0477)	0.2121 (0.0336)
20	10	3	out of sample	3.1460 (0.5151)	1.1107 (0.0609)		
20	20	1	in sample	1.0105 (0.0607)	0.9106 (0.0661)	0.3197 (0.0846)	0.2316 (0.0452)
20	20	1	out of sample	1.6623 (0.2273)	1.0423 (0.0689)		
20	20	2	in sample	1.0808 (0.0567)	0.8985 (0.0557)	0.2753 (0.0287)	0.2210 (0.0189)
20	20	2	out of sample	2.3861 (0.4300)	1.1432 (0.0643)		
20	20	3	in sample	1.0953 (0.0500)	0.8156 (0.0526)	0.2943 (0.0557)	0.2139 (0.0306)
20	20	3	out of sample	2.9818 (0.4579)	1.1651 (0.0541)		

Table A.8.: Average MSE and relative parameter estimation error for deterministic coefficient transition in model A with external predictors, standard deviation in parentheses, **large** variation.

p	q	d	sample	MSE (RRR)	MSE (TVRRR)	err_{C_t} (RRR)	err_{C_t} (TVRRR)
5	5	1	in sample	1.2952 (0.1756)	0.8317 (0.1322)	0.2475 (0.1078)	0.0619 (0.0222)
5	5	1	out of sample	4.9838 (1.4438)	1.0411 (0.1542)		
5	5	2	in sample	1.7864 (0.2795)	0.8333 (0.1117)	0.2758 (0.1011)	0.0671 (0.0143)
5	5	2	out of sample	12.1038 (2.7089)	1.2426 (0.1461)		
5	5	3	in sample	1.8064 (0.2215)	0.7037 (0.1205)	0.2218 (0.0831)	0.0601 (0.0173)
5	5	3	out of sample	14.8445 (4.5543)	1.1984 (0.1656)		
5	10	1	in sample	1.3251 (0.2550)	0.8211 (0.0684)	0.3076 (0.1781)	0.0795 (0.0317)
5	10	1	out of sample	6.7647 (3.4244)	1.0707 (0.1048)		
5	10	2	in sample	1.6996 (0.2718)	0.8105 (0.1042)	0.2763 (0.0816)	0.0723 (0.0193)
5	10	2	out of sample	11.0841 (3.6273)	1.2997 (0.1388)		
5	10	3	in sample	2.0122 (0.3715)	0.6394 (0.0952)	0.3229 (0.1074)	0.0738 (0.0221)
5	10	3	out of sample	16.6818 (4.0634)	1.2642 (0.1600)		
10	5	1	in sample	1.1279 (0.0683)	0.9207 (0.0484)	0.2057 (0.1011)	0.0971 (0.0245)
10	5	1	out of sample	3.2885 (1.2664)	1.0842 (0.0596)		
10	5	2	in sample	1.2791 (0.1982)	0.7959 (0.0934)	0.2130 (0.0617)	0.0815 (0.0156)
10	5	2	out of sample	5.7660 (1.7668)	1.0879 (0.1150)		
10	5	3	in sample	1.5001 (0.1213)	0.7761 (0.0569)	0.2642 (0.0693)	0.0991 (0.0269)
10	5	3	out of sample	8.3358 (1.4290)	1.2053 (0.0661)		
10	10	1	in sample	1.0805 (0.0891)	0.8659 (0.0791)	0.1671 (0.0493)	0.0785 (0.0180)
10	10	1	out of sample	2.4777 (0.4529)	1.0386 (0.0855)		
10	10	2	in sample	1.3422 (0.1467)	0.8173 (0.0909)	0.2651 (0.0767)	0.0975 (0.0170)
10	10	2	out of sample	6.3858 (1.0800)	1.1556 (0.1121)		
10	10	3	in sample	1.4983 (0.1434)	0.7842 (0.0646)	0.2706 (0.0411)	0.1079 (0.0150)
10	10	3	out of sample	8.3580 (1.3044)	1.2588 (0.0982)		
10	20	1	in sample	1.1499 (0.0660)	0.8691 (0.0862)	0.2663 (0.0888)	0.1125 (0.0247)
10	20	1	out of sample	3.4584 (0.7170)	1.0957 (0.0847)		
10	20	2	in sample	1.2210 (0.1714)	0.8044 (0.1002)	0.2318 (0.0727)	0.1055 (0.0241)
10	20	2	out of sample	5.3014 (1.2006)	1.2073 (0.1615)		
10	20	3	in sample	1.3910 (0.1351)	0.7163 (0.0505)	0.2490 (0.0862)	0.1070 (0.0212)
10	20	3	out of sample	7.6330 (2.0372)	1.2914 (0.0817)		
20	10	1	in sample	1.0838 (0.0824)	0.9299 (0.0751)	0.2483 (0.0339)	0.1413 (0.0130)
20	10	1	out of sample	2.2113 (0.2005)	1.0595 (0.0775)		
20	10	2	in sample	1.1964 (0.0979)	0.8523 (0.0663)	0.2685 (0.0547)	0.1387 (0.0204)
20	10	2	out of sample	3.5534 (0.4233)	1.1133 (0.0812)		
20	10	3	in sample	1.2257 (0.0932)	0.7898 (0.0701)	0.2764 (0.0498)	0.1438 (0.0191)
20	10	3	out of sample	4.6792 (0.7400)	1.1398 (0.0811)		
20	20	1	in sample	1.0737 (0.1029)	0.9252 (0.0845)	0.2753 (0.0852)	0.1641 (0.0356)
20	20	1	out of sample	2.2222 (0.2929)	1.0819 (0.0855)		
20	20	2	in sample	1.1406 (0.0904)	0.8463 (0.0490)	0.2735 (0.0632)	0.1552 (0.0192)
20	20	2	out of sample	3.3837 (0.6633)	1.1411 (0.0786)		
20	20	3	in sample	1.2113 (0.0740)	0.8042 (0.0509)	0.2862 (0.0508)	0.1639 (0.0174)
20	20	3	out of sample	5.1578 (0.7888)	1.2183 (0.0725)		

Table A.9.: Average MSE and relative parameter estimation error for deterministic coefficient transition in model B with external predictors, standard deviation in parentheses, **small** variation.

p	q	d	sample	MSE (RRR)	MSE (TVRRR)	err_{C_t} (RRR)	err_{C_t} (TVRRR)
5	5	1	in sample	1.1030 (0.1463)	0.9465 (0.1337)	0.4224 (0.1488)	0.2542 (0.0978)
5	5	1	out of sample	2.7040 (0.7976)	1.0989 (0.1579)		
5	5	2	in sample	1.1219 (0.1003)	0.8553 (0.0553)	0.2596 (0.1098)	0.1565 (0.0434)
5	5	2	out of sample	3.3282 (0.8737)	1.0903 (0.0573)		
5	5	3	in sample	1.2616 (0.1569)	0.8451 (0.1425)	0.3378 (0.1431)	0.1822 (0.0533)
5	5	3	out of sample	4.6659 (1.1897)	1.1927 (0.1827)		
10	10	1	in sample	1.0672 (0.0998)	0.9962 (0.0904)	0.4490 (0.1771)	0.3457 (0.1020)
10	10	1	out of sample	1.6924 (0.3180)	1.0948 (0.0925)		
10	10	2	in sample	1.0602 (0.0793)	0.9215 (0.0789)	0.3635 (0.0786)	0.2677 (0.0284)
10	10	2	out of sample	2.1380 (0.2809)	1.1122 (0.0917)		
10	10	3	in sample	1.0636 (0.1344)	0.8826 (0.1219)	0.3315 (0.0644)	0.2677 (0.0467)
10	10	3	out of sample	2.8633 (0.2484)	1.1558 (0.1437)		
20	20	1	in sample	1.0211 (0.0505)	0.9942 (0.0536)	0.5173 (0.1666)	0.4357 (0.0610)
20	20	1	out of sample	1.3302 (0.0856)	1.0667 (0.0543)		
20	20	2	in sample	1.0024 (0.0827)	0.9437 (0.0818)	0.5441 (0.0689)	0.4839 (0.0632)
20	20	2	out of sample	1.7334 (0.1324)	1.1010 (0.0870)		
20	20	3	in sample	1.0378 (0.0508)	0.9469 (0.0549)	0.5035 (0.0634)	0.4513 (0.0314)
20	20	3	out of sample	2.0718 (0.2278)	1.1828 (0.0579)		

Table A.10.: Average MSE and relative parameter estimation error for deterministic coefficient transition in model B with external predictors, standard deviation in parentheses, **medium** variation.

p	q	d	sample	MSE (RRR)	MSE (TVRRR)	err_{C_t} (RRR)	err_{C_t} (TVRRR)
5	5	1	in sample	1.1470 (0.1150)	0.8717 (0.1123)	0.3743 (0.1790)	0.1481 (0.0674)
5	5	1	out of sample	4.5132 (1.9689)	1.0350 (0.1192)		
5	5	2	in sample	1.2694 (0.1339)	0.8424 (0.0954)	0.2102 (0.0991)	0.0865 (0.0259)
5	5	2	out of sample	5.0793 (1.7092)	1.1315 (0.1062)		
5	5	3	in sample	1.5243 (0.2413)	0.7433 (0.1257)	0.3128 (0.0941)	0.1122 (0.0354)
5	5	3	out of sample	10.6418 (3.0691)	1.1761 (0.1744)		
10	10	1	in sample	1.0876 (0.0913)	0.9661 (0.0662)	0.3587 (0.2036)	0.2154 (0.1228)
10	10	1	out of sample	2.4989 (0.8415)	1.0957 (0.0808)		
10	10	2	in sample	1.1106 (0.1144)	0.8594 (0.0984)	0.3262 (0.0976)	0.1807 (0.0381)
10	10	2	out of sample	3.9858 (0.7443)	1.0933 (0.1091)		
10	10	3	in sample	1.2790 (0.0800)	0.8965 (0.0693)	0.3382 (0.0807)	0.1842 (0.0315)
10	10	3	out of sample	5.3404 (0.7817)	1.2561 (0.0886)		
20	20	1	in sample	1.0176 (0.0385)	0.9664 (0.0467)	0.3788 (0.1662)	0.3127 (0.0854)
20	20	1	out of sample	1.6726 (0.2232)	1.0524 (0.0420)		
20	20	2	in sample	1.0470 (0.0695)	0.9395 (0.0612)	0.3948 (0.0734)	0.2901 (0.0396)
20	20	2	out of sample	2.4409 (0.2605)	1.1176 (0.0749)		
20	20	3	in sample	1.0460 (0.0693)	0.8840 (0.0663)	0.4067 (0.0542)	0.2971 (0.0366)
20	20	3	out of sample	3.2673 (0.2981)	1.1514 (0.0823)		

Table A.11.: Average MSE and relative parameter estimation error for deterministic coefficient transition in model B with external predictors, standard deviation in parentheses, **large** variation.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	1.3742 (0.2608)	0.9036 (0.1409)	0.3370 (0.1794)	0.0929 (0.0425)
5	5	1	out of sample	6.6529 (2.6617)	1.1056 (0.1709)		
5	5	2	in sample	1.4422 (0.1903)	0.8151 (0.1309)	0.1873 (0.0742)	0.0546 (0.0124)
5	5	2	out of sample	7.9805 (2.5363)	1.1356 (0.1565)		
5	5	3	in sample	1.7888 (0.2922)	0.7745 (0.1095)	0.2385 (0.0712)	0.0630 (0.0153)
5	5	3	out of sample	15.840 (4.3341)	1.2641 (0.1621)		
10	10	1	in sample	1.0893 (0.0906)	0.9098 (0.0762)	0.3070 (0.1289)	0.1582 (0.0455)
10	10	1	out of sample	3.6538 (0.8978)	1.0592 (0.0930)		
10	10	2	in sample	1.2554 (0.0794)	0.8661 (0.0715)	0.3119 (0.0762)	0.1303 (0.0258)
10	10	2	out of sample	6.5400 (1.0845)	1.1381 (0.0772)		
10	10	3	in sample	1.3520 (0.0754)	0.8032 (0.0271)	0.2873 (0.0534)	0.1235 (0.0191)
10	10	3	out of sample	8.3886 (1.5862)	1.2044 (0.0295)		
20	20	1	in sample	1.0352 (0.0333)	0.9527 (0.0397)	0.3668 (0.1727)	0.2464 (0.0732)
20	20	1	out of sample	2.2384 (0.3899)	1.0558 (0.0394)		
20	20	2	in sample	1.0759 (0.1096)	0.9162 (0.0990)	0.3853 (0.0931)	0.2306 (0.0275)
20	20	2	out of sample	3.5873 (0.5231)	1.1134 (0.0962)		
20	20	3	in sample	1.1378 (0.0705)	0.8859 (0.0563)	0.3694 (0.0566)	0.2176 (0.0240)
20	20	3	out of sample	4.7772 (0.5271)	1.1842 (0.0623)		

A.4. Simulation results for random walk coefficient evolution

Table A.12.: Average MSE and relative parameter estimation error for random walk coefficient transition in model A with external predictors, standard deviation in parentheses, **small** variation.

p	q	d	sample	MSE (RRR)	MSE (TvRRR)	err_{C_t} (RRR)	err_{C_t} (TvRRR)
5	5	1	in sample	1.0539 (0.1546)	0.9869 (0.1507)	2.4283 (1.9799)	1.1320 (0.6406)
5	5	1	out of sample	1.3649 (0.2601)	1.1091 (0.1791)		
5	5	2	in sample	1.0995 (0.1422)	0.9582 (0.1251)	1.8945 (1.0490)	0.6964 (0.1696)
5	5	2	out of sample	1.4891 (0.2288)	1.1509 (0.1245)		
5	5	3	in sample	1.0443 (0.1307)	0.8509 (0.1221)	1.7470 (0.5805)	0.7164 (0.1534)
5	5	3	out of sample	1.7224 (0.2929)	1.1063 (0.1597)		
5	10	1	in sample	1.0697 (0.1282)	0.9921 (0.1185)	2.6767 (1.3680)	1.1640 (0.4239)
5	10	1	out of sample	1.3000 (0.1903)	1.1347 (0.1418)		
5	10	2	in sample	1.1088 (0.1749)	0.9679 (0.1734)	1.9995 (0.7037)	1.1870 (0.4516)
5	10	2	out of sample	1.6299 (0.2570)	1.2448 (0.2099)		
5	10	3	in sample	1.0346 (0.1432)	0.8323 (0.1072)	2.0970 (0.8251)	0.7475 (0.1425)
5	10	3	out of sample	1.7264 (0.3992)	1.1619 (0.1456)		
10	5	1	in sample	1.0045 (0.1110)	0.9425 (0.1133)	1.6369 (0.7760)	0.6471 (0.1688)
10	5	1	out of sample	1.2141 (0.1394)	1.0259 (0.1104)		
10	5	2	in sample	1.0843 (0.1325)	0.9287 (0.1215)	1.6102 (0.5515)	0.5984 (0.1478)
10	5	2	out of sample	1.5914 (0.2140)	1.1032 (0.1255)		
10	5	3	in sample	1.0618 (0.0597)	0.8468 (0.0437)	1.5831 (0.3825)	0.5521 (0.0892)
10	5	3	out of sample	1.7090 (0.2221)	1.0928 (0.0583)		
10	10	1	in sample	0.9975 (0.0615)	0.9196 (0.0566)	1.8192 (0.5351)	0.7513 (0.2070)
10	10	1	out of sample	1.2201 (0.1253)	1.0250 (0.0757)		
10	10	2	in sample	1.0620 (0.0877)	0.9199 (0.0906)	1.6212 (0.5203)	0.6609 (0.0871)
10	10	2	out of sample	1.5204 (0.2758)	1.1018 (0.0931)		
10	10	3	in sample	1.0747 (0.1133)	0.8781 (0.1097)	1.6104 (0.3326)	0.6695 (0.0921)
10	10	3	out of sample	1.7913 (0.2968)	1.1720 (0.1390)		
10	20	1	in sample	1.0114 (0.1091)	0.9391 (0.0994)	1.8115 (0.8649)	0.7763 (0.3395)
10	20	1	out of sample	1.2476 (0.1567)	1.0779 (0.1171)		
10	20	2	in sample	1.0341 (0.0677)	0.9197 (0.0714)	1.8680 (0.5941)	0.8219 (0.1309)
10	20	2	out of sample	1.5405 (0.1146)	1.1835 (0.0764)		
10	20	3	in sample	1.0702 (0.1080)	0.8785 (0.0869)	1.9953 (0.3921)	0.8410 (0.1694)
10	20	3	out of sample	1.8800 (0.2707)	1.2606 (0.1253)		
20	10	1	in sample	1.0253 (0.0643)	0.9553 (0.0656)	1.5031 (0.4612)	0.5988 (0.1096)
20	10	1	out of sample	1.2339 (0.0823)	1.0440 (0.0672)		
20	10	2	in sample	1.0676 (0.0752)	0.9232 (0.0558)	1.6542 (0.2843)	0.5555 (0.0965)
20	10	2	out of sample	1.4691 (0.1260)	1.0906 (0.0624)		
20	10	3	in sample	1.1017 (0.0836)	0.9078 (0.0856)	1.4174 (0.2580)	0.5726 (0.0694)
20	10	3	out of sample	1.8446 (0.2607)	1.1505 (0.1006)		
20	20	1	in sample	0.9993 (0.0864)	0.9390 (0.0858)	1.6709 (0.3452)	0.6744 (0.1478)
20	20	1	out of sample	1.2383 (0.0680)	1.0424 (0.0899)		
20	20	2	in sample	1.0496 (0.0783)	0.9319 (0.0653)	1.6315 (0.3272)	0.6871 (0.1074)
20	20	2	out of sample	1.5233 (0.1032)	1.1388 (0.0778)		
20	20	3	in sample	1.0576 (0.0695)	0.8710 (0.0592)	1.6133 (0.2372)	0.6748 (0.1026)
20	20	3	out of sample	1.7590 (0.1888)	1.1677 (0.0586)		

Table A.13.: Average MSE and relative parameter estimation error for random walk coefficient transition in model A with external predictors, standard deviation in parentheses, **medium** variation.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	1.3669 (0.1697)	0.8967 (0.1157)	1.4812 (0.7012)	0.3540 (0.0907)
5	5	1	out of sample	3.1365 (1.0535)	1.1623 (0.1355)		
5	5	2	in sample	1.6731 (0.2722)	0.7884 (0.1157)	1.9553 (0.9240)	0.3501 (0.0889)
5	5	2	out of sample	5.1565 (0.6617)	1.2420 (0.1303)		
5	5	3	in sample	1.9745 (0.2988)	0.7768 (0.1786)	1.3501 (0.6374)	0.2929 (0.1100)
5	5	3	out of sample	7.3190 (1.7056)	1.4499 (0.2607)		
5	10	1	in sample	1.3963 (0.2224)	0.8853 (0.1352)	2.1986 (2.4737)	0.3667 (0.1710)
5	10	1	out of sample	3.2028 (1.4025)	1.1729 (0.1542)		
5	10	2	in sample	1.7664 (0.4040)	0.8092 (0.1299)	1.6501 (1.0556)	0.2909 (0.0843)
5	10	2	out of sample	5.3507 (1.6193)	1.3593 (0.2073)		
5	10	3	in sample	2.1557 (0.3924)	0.7516 (0.1114)	1.5262 (0.3978)	0.3072 (0.0655)
5	10	3	out of sample	7.8642 (2.8489)	1.5255 (0.1195)		
10	5	1	in sample	1.3369 (0.1455)	0.8486 (0.0711)	1.2695 (0.3697)	0.2803 (0.0697)
10	5	1	out of sample	3.6856 (1.2229)	1.0993 (0.0774)		
10	5	2	in sample	1.7392 (0.1862)	0.7967 (0.0751)	1.3309 (0.2026)	0.2509 (0.0271)
10	5	2	out of sample	4.9991 (1.0956)	1.2319 (0.1210)		
10	5	3	in sample	2.0173 (0.2060)	0.7345 (0.0678)	1.1314 (0.2074)	0.2730 (0.0376)
10	5	3	out of sample	8.5322 (2.6074)	1.3581 (0.1015)		
10	10	1	in sample	1.3507 (0.1150)	0.9233 (0.0829)	1.4125 (0.5407)	0.2529 (0.0693)
10	10	1	out of sample	3.3552 (0.8473)	1.1837 (0.0850)		
10	10	2	in sample	1.7951 (0.2495)	0.7577 (0.0864)	1.2498 (0.2394)	0.2539 (0.0379)
10	10	2	out of sample	7.0220 (2.6184)	1.2433 (0.1087)		
10	10	3	in sample	2.0290 (0.2905)	0.7326 (0.0562)	1.3983 (0.4371)	0.2747 (0.0488)
10	10	3	out of sample	6.9571 (1.2448)	1.3938 (0.0802)		
10	20	1	in sample	1.5157 (0.2531)	0.9163 (0.0869)	1.5040 (0.5882)	0.2835 (0.1102)
10	20	1	out of sample	4.0216 (2.1033)	1.2409 (0.1196)		
10	20	2	in sample	1.7604 (0.3734)	0.7628 (0.0878)	1.6929 (0.4878)	0.2868 (0.0481)
10	20	2	out of sample	6.4552 (2.3907)	1.3305 (0.1457)		
10	20	3	in sample	1.9559 (0.3004)	0.7467 (0.0801)	1.3369 (0.3375)	0.3146 (0.0785)
10	20	3	out of sample	9.8396 (3.2152)	1.6262 (0.1126)		
20	10	1	in sample	1.4719 (0.1672)	0.8942 (0.0489)	1.4273 (0.5991)	0.2388 (0.0379)
20	10	1	out of sample	3.5353 (0.8641)	1.1543 (0.0762)		
20	10	2	in sample	1.8681 (0.2062)	0.7693 (0.0661)	1.5382 (0.3870)	0.2499 (0.0437)
20	10	2	out of sample	5.2747 (1.2133)	1.2100 (0.0777)		
20	10	3	in sample	2.1182 (0.2297)	0.7557 (0.0509)	1.3481 (0.2755)	0.2820 (0.0416)
20	10	3	out of sample	7.7612 (1.4595)	1.3835 (0.0518)		
20	20	1	in sample	1.4249 (0.1612)	0.8807 (0.0827)	1.2623 (0.4461)	0.2532 (0.0734)
20	20	1	out of sample	3.7199 (0.4589)	1.1664 (0.0935)		
20	20	2	in sample	1.7322 (0.2236)	0.7584 (0.0530)	1.3397 (0.2835)	0.2595 (0.0325)
20	20	2	out of sample	6.0643 (1.7542)	1.2454 (0.0510)		
20	20	3	in sample	2.0681 (0.1870)	0.7117 (0.0536)	1.4485 (0.2529)	0.2662 (0.0247)
20	20	3	out of sample	8.5483 (2.1350)	1.4063 (0.0884)		

Table A.14.: Average MSE and relative parameter estimation error for random walk coefficient transition in model A with external predictors, standard deviation in parentheses, **large** variation.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	2.3211 (1.3090)	0.8759 (0.0760)	2.6296 (1.8870)	0.2361 (0.0624)
5	5	1	out of sample	6.2194 (2.9533)	1.2429 (0.0821)		
5	5	2	in sample	2.4354 (0.5644)	0.7419 (0.1089)	1.6546 (0.6157)	0.2439 (0.0591)
5	5	2	out of sample	9.4583 (2.9142)	1.3573 (0.1524)		
5	5	3	in sample	3.5668 (0.6869)	0.6458 (0.0538)	1.8362 (0.5219)	0.2510 (0.0779)
5	5	3	out of sample	14.3819 (4.3991)	1.5452 (0.0931)		
5	10	1	in sample	1.8814 (0.2726)	0.8518 (0.1365)	2.7981 (3.1109)	0.2438 (0.1028)
5	10	1	out of sample	5.8419 (2.2981)	1.2294 (0.1705)		
5	10	2	in sample	2.8181 (0.7857)	0.7165 (0.0668)	1.6196 (0.6079)	0.2063 (0.0624)
5	10	2	out of sample	10.9122 (3.7625)	1.4300 (0.1396)		
5	10	3	in sample	3.5889 (0.9440)	0.6868 (0.0989)	1.7693 (0.6088)	0.2284 (0.0364)
5	10	3	out of sample	17.0521 (7.2372)	1.7260 (0.2001)		
10	5	1	in sample	1.8902 (0.2606)	0.8392 (0.0625)	1.5804 (0.6751)	0.1881 (0.0375)
10	5	1	out of sample	5.4851 (1.9907)	1.1729 (0.0697)		
10	5	2	in sample	2.4918 (0.2224)	0.7178 (0.0579)	1.3377 (0.6522)	0.2243 (0.0629)
10	5	2	out of sample	12.7689 (4.5673)	1.3380 (0.0920)		
10	5	3	in sample	3.1672 (0.7496)	0.6462 (0.0709)	1.4068 (0.3606)	0.2200 (0.0433)
10	5	3	out of sample	12.2643 (2.4434)	1.4750 (0.1189)		
10	10	1	in sample	1.8004 (0.3165)	0.8448 (0.0962)	1.5344 (0.5307)	0.2084 (0.0364)
10	10	1	out of sample	5.3302 (2.1873)	1.2026 (0.1306)		
10	10	2	in sample	2.6983 (0.5685)	0.7524 (0.0735)	1.4511 (0.3304)	0.2181 (0.0594)
10	10	2	out of sample	10.9765 (3.4839)	1.4067 (0.1048)		
10	10	3	in sample	3.3075 (0.5509)	0.6729 (0.0930)	1.3349 (0.3448)	0.2017 (0.0280)
10	10	3	out of sample	13.6254 (2.7399)	1.6052 (0.1727)		
10	20	1	in sample	1.6889 (0.2083)	0.8108 (0.0827)	1.5183 (0.4267)	0.1905 (0.0592)
10	20	1	out of sample	5.1668 (1.5427)	1.1996 (0.0952)		
10	20	2	in sample	2.4415 (0.3291)	0.7320 (0.1134)	1.2787 (0.3346)	0.2209 (0.0623)
10	20	2	out of sample	10.0634 (3.7046)	1.4560 (0.1899)		
10	20	3	in sample	2.9957 (0.4276)	0.6730 (0.0379)	1.2617 (0.3351)	0.2194 (0.0234)
10	20	3	out of sample	18.0626 (4.3625)	1.7634 (0.1117)		
20	10	1	in sample	1.7784 (0.2071)	0.8427 (0.0752)	1.2749 (0.4641)	0.1730 (0.0262)
20	10	1	out of sample	5.2187 (1.4097)	1.1824 (0.0906)		
20	10	2	in sample	2.4532 (0.3831)	0.7805 (0.0424)	1.2460 (0.3767)	0.2110 (0.0294)
20	10	2	out of sample	9.6725 (1.5779)	1.3876 (0.0886)		
20	10	3	in sample	3.2687 (0.5620)	0.6788 (0.0788)	1.3616 (0.2344)	0.2172 (0.0308)
20	10	3	out of sample	15.5335 (2.1083)	1.5651 (0.1086)		
20	20	1	in sample	1.8687 (0.4042)	0.8284 (0.0589)	1.5159 (0.3020)	0.2221 (0.0400)
20	20	1	out of sample	6.6437 (1.7356)	1.1951 (0.0599)		
20	20	2	in sample	2.7774 (0.4800)	0.7350 (0.0596)	1.3589 (0.4956)	0.1882 (0.0531)
20	20	2	out of sample	10.4826 (3.2061)	1.4019 (0.0877)		
20	20	3	in sample	3.1220 (0.2329)	0.6758 (0.0509)	1.3100 (0.2280)	0.2017 (0.0368)
20	20	3	out of sample	16.2295 (3.9574)	1.6078 (0.0592)		

Table A.15.: Average MSE and relative parameter estimation error for random walk coefficient transition in model B with external predictors, standard deviation in parentheses, **small** variation.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	1.0064 (0.1417)	0.9342 (0.1314)	2.9536 (1.6543)	1.0042 (0.3259)
5	5	1	out of sample	1.3150 (0.2196)	1.0429 (0.1323)		
5	5	2	in sample	1.0012 (0.1765)	0.8788 (0.1965)	1.7646 (0.6746)	0.7847 (0.2087)
5	5	2	out of sample	1.3787 (0.1650)	1.0431 (0.2053)		
5	5	3	in sample	1.1455 (0.1473)	0.9488 (0.1312)	1.7889 (0.6811)	0.6967 (0.1198)
5	5	3	out of sample	1.6290 (0.2167)	1.1858 (0.1637)		
10	10	1	in sample	0.9685 (0.0723)	0.9237 (0.0742)	1.7297 (0.5701)	0.8691 (0.3079)
10	10	1	out of sample	1.1945 (0.1383)	1.0065 (0.0724)		
10	10	2	in sample	1.0795 (0.1078)	0.9768 (0.1124)	1.8046 (0.3527)	0.7927 (0.0845)
10	10	2	out of sample	1.5048 (0.2448)	1.1534 (0.1363)		
10	10	3	in sample	1.1224 (0.0787)	0.9658 (0.0802)	1.6299 (0.3092)	0.7298 (0.0844)
10	10	3	out of sample	1.8261 (0.3654)	1.2150 (0.0818)		
20	20	1	in sample	1.0237 (0.0589)	0.9790 (0.0573)	1.9142 (0.6377)	0.8414 (0.1741)
20	20	1	out of sample	1.2945 (0.0562)	1.0671 (0.0613)		
20	20	2	in sample	1.0373 (0.0453)	0.9522 (0.0385)	1.8169 (0.2573)	0.7664 (0.0697)
20	20	2	out of sample	1.5086 (0.2695)	1.1113 (0.0512)		
20	20	3	in sample	1.0303 (0.0706)	0.9024 (0.0565)	1.8530 (0.2539)	0.8192 (0.0856)
20	20	3	out of sample	1.7465 (0.2695)	1.1267 (0.0593)		

Table A.16.: Average MSE and relative parameter estimation error for random walk coefficient transition in model B with external predictors, standard deviation in parentheses, **medium** variation.

p	q	d	sample	MSE (RRR)	MSE (tvRRR)	err_{C_t} (RRR)	err_{C_t} (tvRRR)
5	5	1	in sample	1.3363 (0.1509)	0.9500 (0.1121)	1.6290 (0.7477)	0.5007(0.1919)
5	5	1	out of sample	3.6626 (0.9100)	1.2015 (0.1235)		
5	5	2	in sample	2.0930 (0.3999)	0.8867 (0.1271)	2.8817 (1.7713)	0.4459(0.1370)
5	5	2	out of sample	5.7967 (2.2031)	1.3335 (0.1502)		
5	5	3	in sample	2.1292 (0.2600)	0.8026 (0.1014)	1.7387 (0.7621)	0.3605(0.0607)
5	5	3	out of sample	8.5882 (2.8413)	1.4776 (0.1423)		
10	10	1	in sample	1.3784 (0.1782)	0.9559 (0.0780)	1.3746 (0.4419)	0.4623(0.1249)
10	10	1	out of sample	3.0428 (0.8102)	1.1735 (0.0669)		
10	10	2	in sample	1.8141 (0.1512)	0.9056 (0.0599)	1.5669 (0.5451)	0.4259(0.0524)
10	10	2	out of sample	5.4969 (1.0081)	1.3393 (0.0796)		
10	10	3	in sample	1.9853 (0.2120)	0.8129 (0.0800)	1.4505 (0.3965)	0.4080(0.0772)
10	10	3	out of sample	8.4159 (3.2621)	1.4182 (0.1275)		
20	20	1	in sample	1.2823 (0.0900)	0.9620 (0.0456)	1.7533 (0.4822)	0.6138(0.1004)
20	20	1	out of sample	3.1985 (0.5285)	1.1805 (0.0511)		
20	20	2	in sample	1.5838 (0.2335)	0.9258 (0.0615)	1.6260 (0.3685)	0.5311(0.0721)
20	20	2	out of sample	6.4650 (2.2359)	1.3605 (0.0936)		
20	20	3	in sample	1.8476 (0.2129)	0.8613 (0.0480)	1.5922 (0.3344)	0.4968(0.0497)
20	20	3	out of sample	8.8800 (1.5822)	1.5001 (0.0825)		

Table A.17.: Average MSE and relative parameter estimation error for random walk coefficient transition in model B with external predictors, standard deviation in parentheses, **large** variation.

p	q	d	sample	MSE (RRR)	MSE (TVRRR)	err_{C_t} (RRR)	err_{C_t} (TVRRR)
5	5	1	in sample	1.9329 (0.5040)	0.9212 (0.1006)	2.7399 (1.1504)	0.4462 (0.2486)
5	5	1	out of sample	7.2173 (3.4577)	1.2282 (0.1121)		
5	5	2	in sample	2.1966 (0.4289)	0.7486 (0.1361)	1.3992 (0.7226)	0.2954 (0.0658)
5	5	2	out of sample	9.9445 (6.3352)	1.2755 (0.1653)		
5	5	3	in sample	3.4337 (0.6240)	0.7216 (0.1260)	1.5591 (0.3090)	0.2815 (0.0766)
5	5	3	out of sample	16.3863 (5.2964)	1.5432 (0.2162)		
10	10	1	in sample	1.6238 (0.2489)	0.9216 (0.0840)	1.9476 (0.8077)	0.4498 (0.1199)
10	10	1	out of sample	6.0949 (2.0447)	1.2117 (0.1063)		
10	10	2	in sample	2.5481 (0.4328)	0.8555 (0.1051)	1.6949 (0.4620)	0.3317 (0.0482)
10	10	2	out of sample	10.5402 (2.9861)	1.4310 (0.1299)		
10	10	3	in sample	2.9481 (0.7160)	0.8241 (0.0856)	1.4663 (0.5431)	0.3842 (0.0687)
10	10	3	out of sample	16.7759 (4.4155)	1.6799 (0.1570)		
20	20	1	in sample	1.4902 (0.0963)	0.9457 (0.0644)	1.4790 (0.3645)	0.4713 (0.0899)
20	20	1	out of sample	5.9178 (1.2608)	1.2828 (0.0871)		
20	20	2	in sample	2.1843 (0.1778)	0.8896 (0.0610)	1.7265 (0.4388)	0.4839 (0.0483)
20	20	2	out of sample	10.9872 (3.0140)	1.5191 (0.0777)		
20	20	3	in sample	2.5651 (0.2805)	0.8773 (0.0898)	1.6148 (0.3205)	0.4851 (0.0461)
20	20	3	out of sample	15.1047 (3.5140)	1.8201 (0.1070)		

B | Supplement: Chapter 3

Description of the content:

This appendix contains supplementary material for Chapter 3. We provide visualizations of the bias for the weighted rank-based estimator in Section B.1 and the improvements of the estimates due to the iterations in the algorithm in Section B.2. Sections B.3 and B.4 visualize the behavior of the proposed estimators under various contamination settings. Finally, Section B.5 contains a short discussion of alternatives to the proposed leverage weights.

B.1. Bias of the weighted rank-based estimator

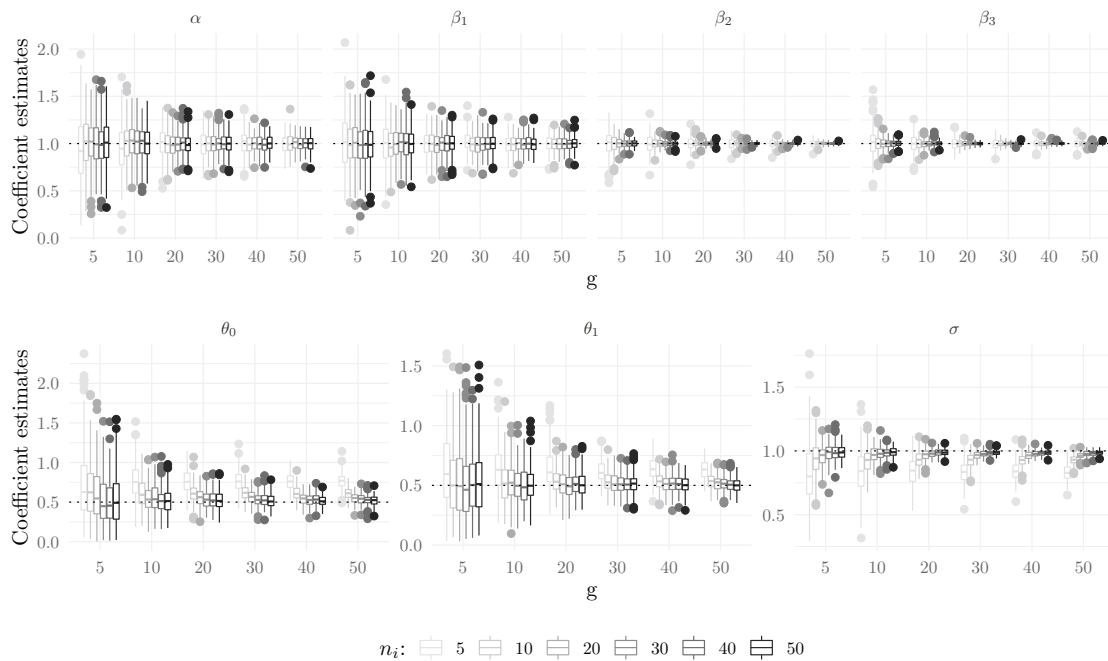


Figure B.1.: Boxplots of regression coefficient and variance component estimates obtained from the leverage-weighted rank-based estimator for different combinations of group sizes n_i and number of groups g (total sample size: $N = g \cdot n_i$).

B.2. Validity of the updating step

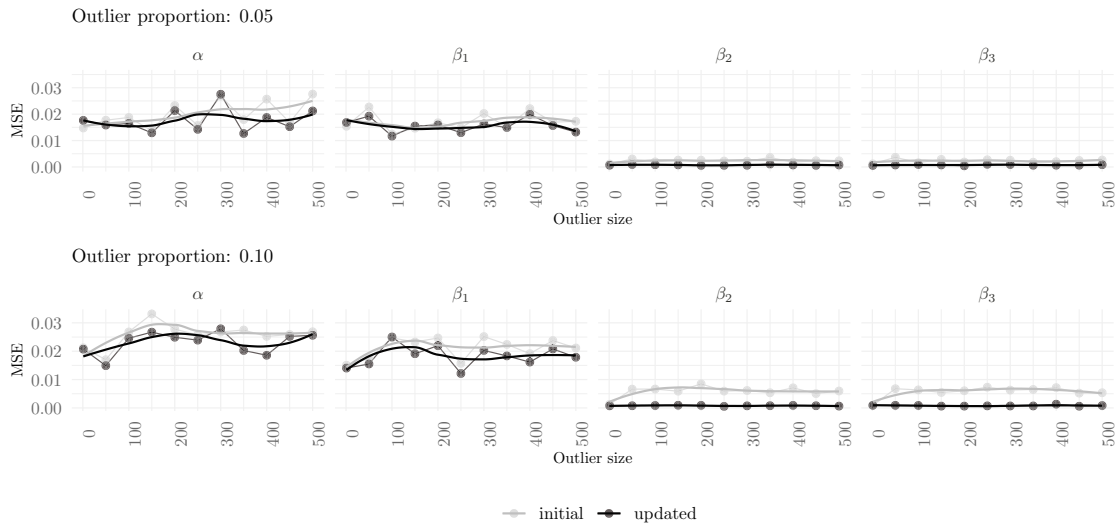


Figure B.2.: MSE of the initial and final coefficient estimates obtained from the rank-based estimator for 5% and 10% of multiplicative y-outliers of increasing size, $n_i = g = 20$.

B.3. Response outliers

B.3.1. Breakdown point under randomly located multiplicative outliers

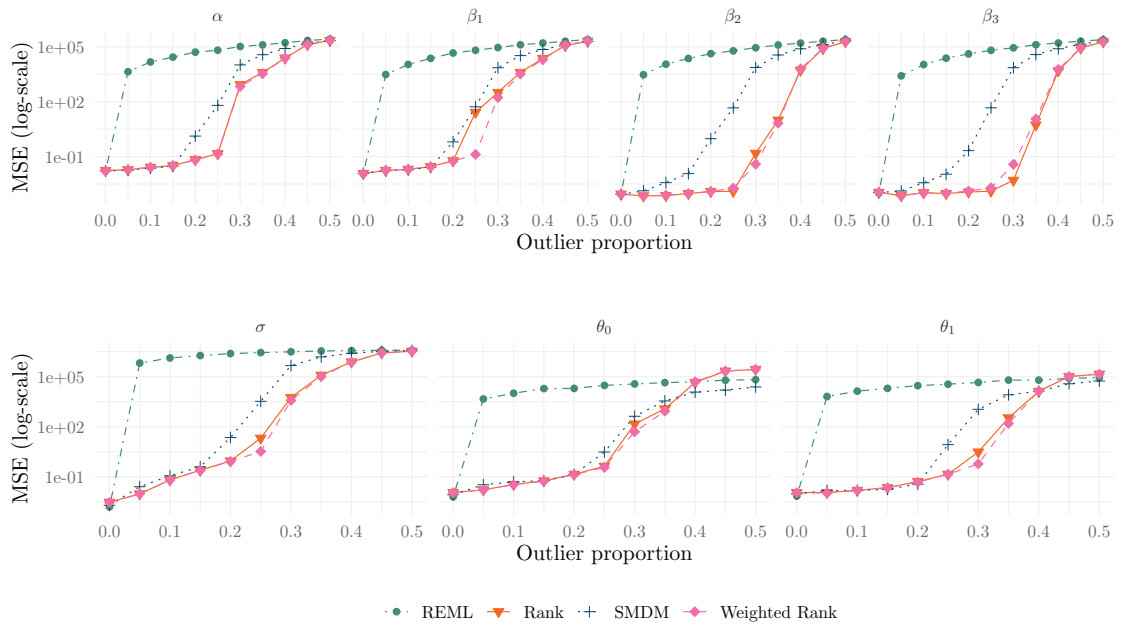


Figure B.3.: MSE (log-scale) of the SMDM, REML, Rank and Weighted Rank estimators for an increasing proportion of multiplicative response outliers of size 1 000, $n_i = g = 20$

B.3.2. Contamination with randomly located additive outliers

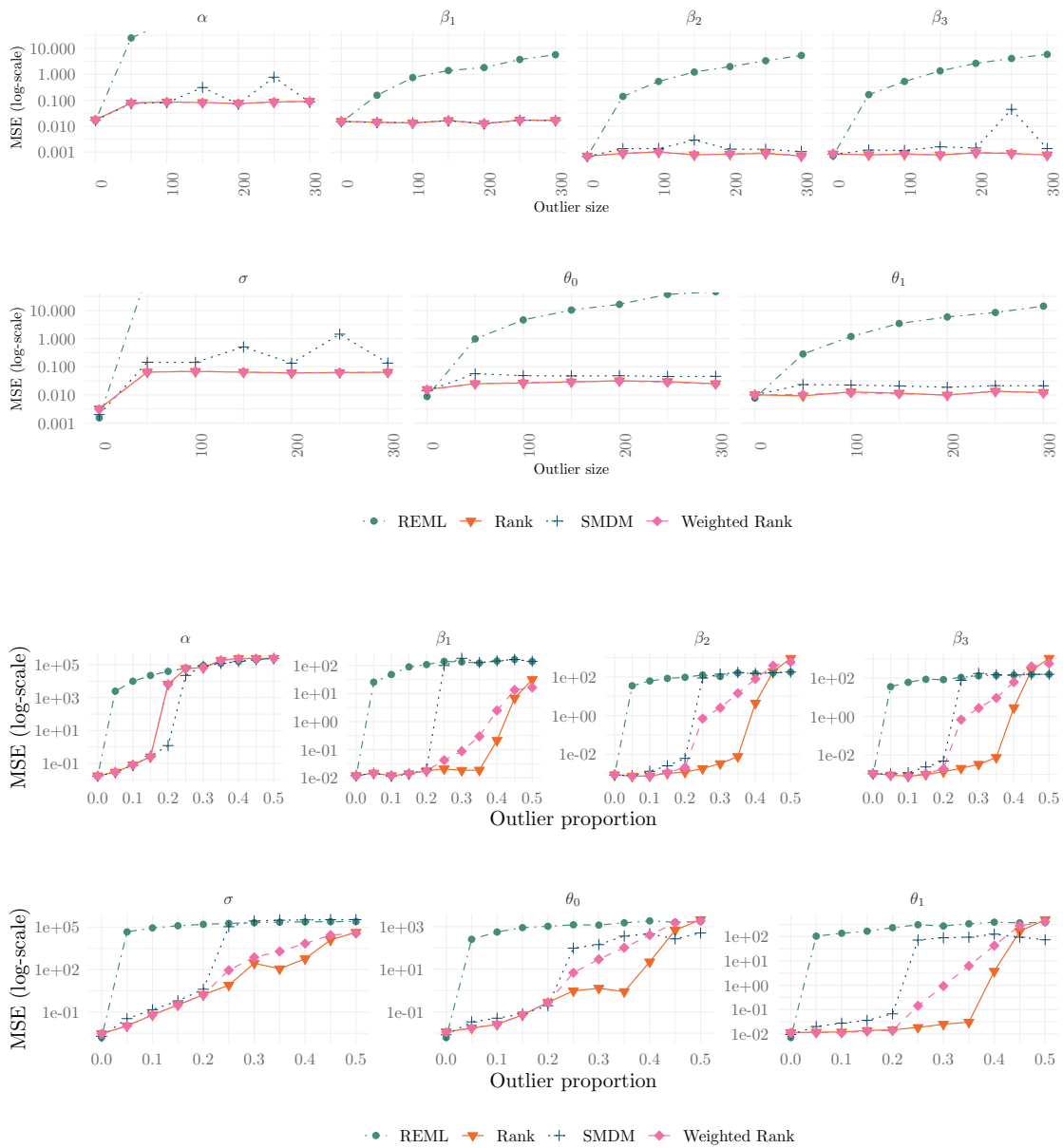


Figure B.4.: MSE (log-scale) of the SMDM, REML, Rank and Weighted Rank estimators for a ten percent proportion of randomly located additive response outliers of increasing size (top) and for an increasing proportion of randomly located additive response outliers of size 1000 (bottom), $n_i = g = 20$.

B.3.3. Groupwise contamination

Multiplicative outliers

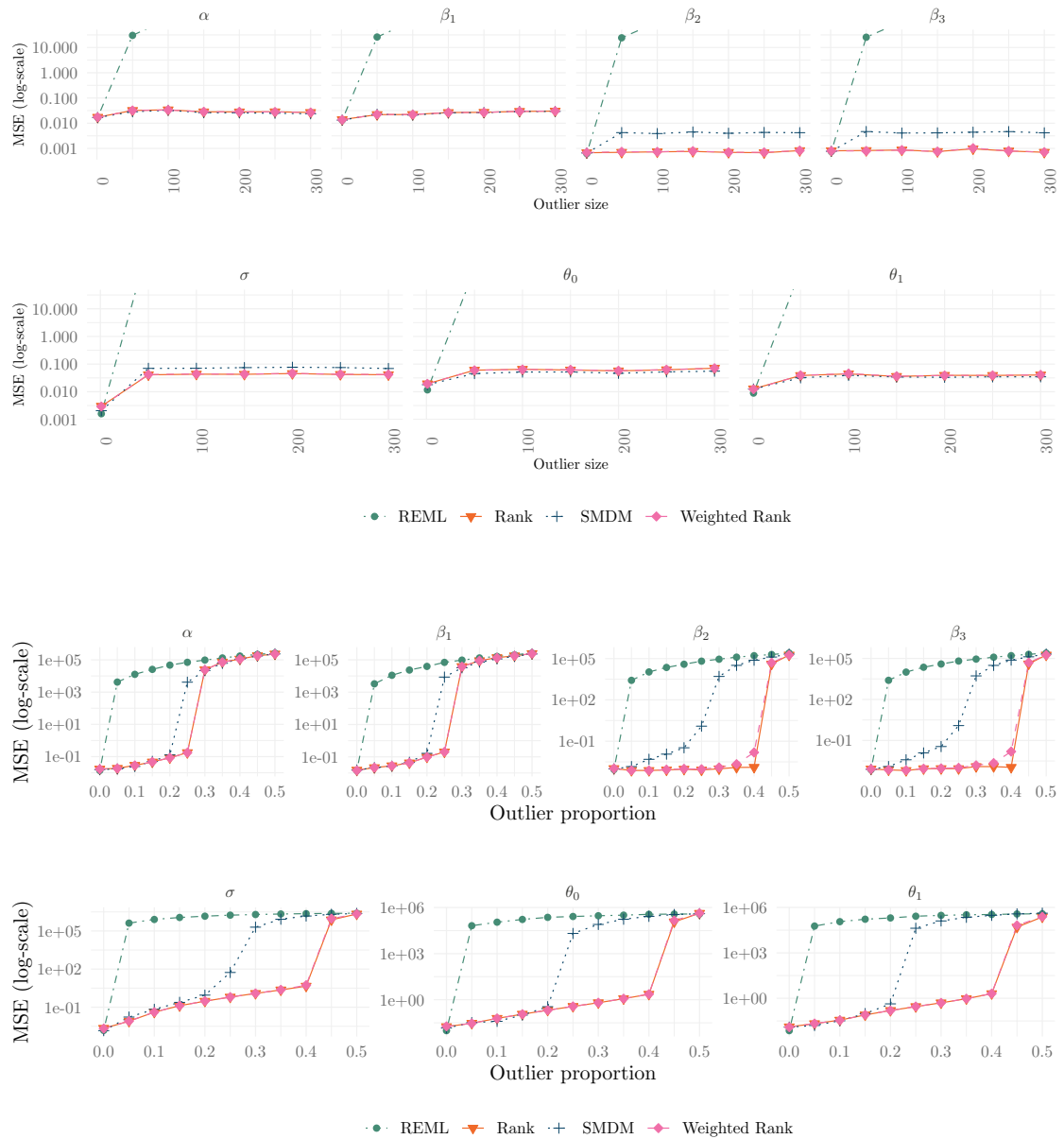


Figure B.5.: MSE (log-scale) of the SMDM, REML, Rank and Weighted Rank estimators for a ten percent proportion of sequentially located multiplicative response outliers of increasing size (top) and for an increasing proportion of sequentially located multiplicative response outliers of size 1000 (bottom), $n_i = g = 20$.

Additive outliers

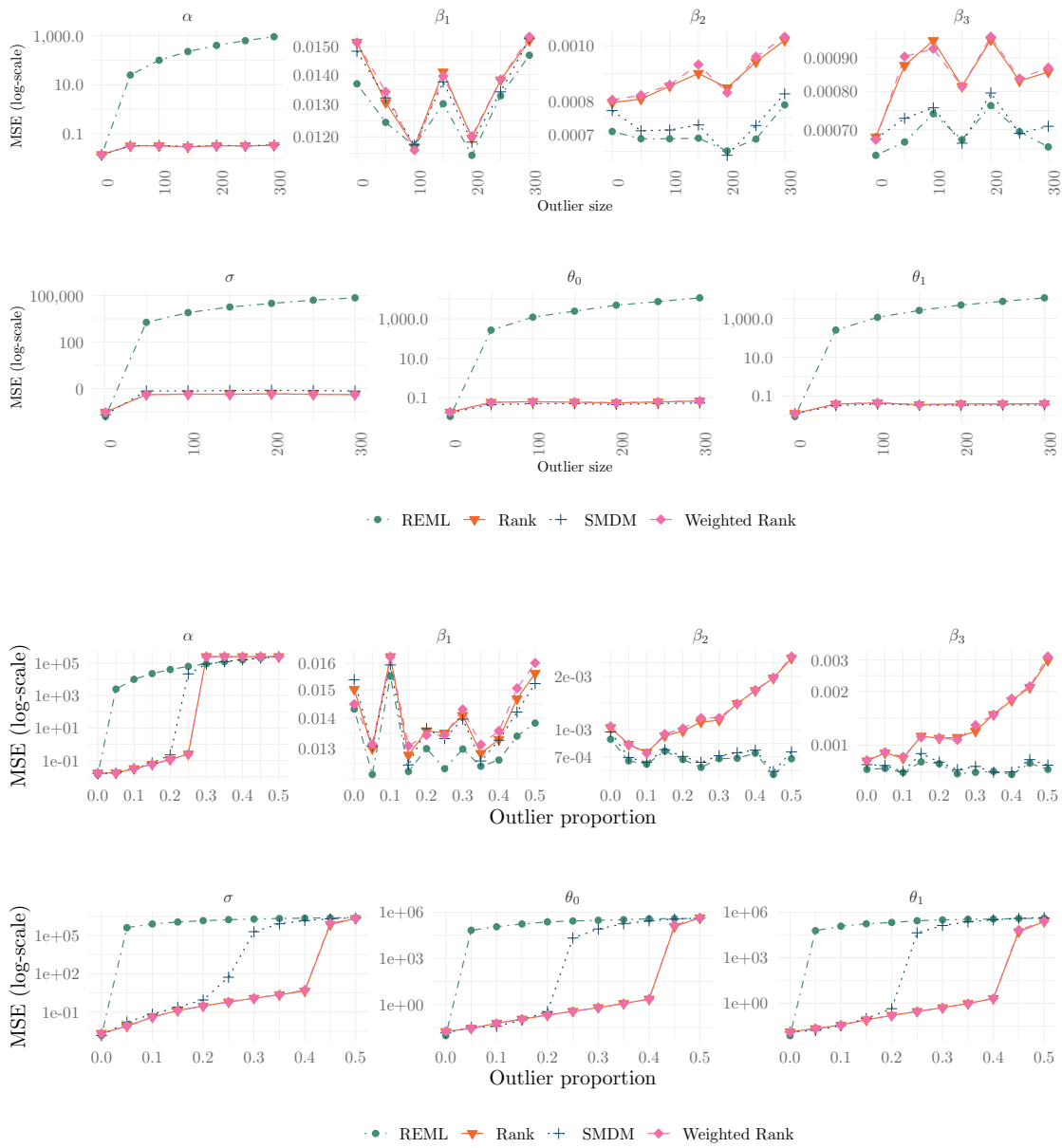


Figure B.6.: MSE (log-scale) of the SMDM, REML, Rank and Weighted Rank estimators for a ten percent proportion of sequentially located additive response outliers of increasing size (top) and for an increasing proportion of sequentially located additive response outliers of size 1 000 (bottom), $n_i = g = 20$.

B.4. Outlying predictors

B.4.1. Breakdown point under randomly located multiplicative leverage points

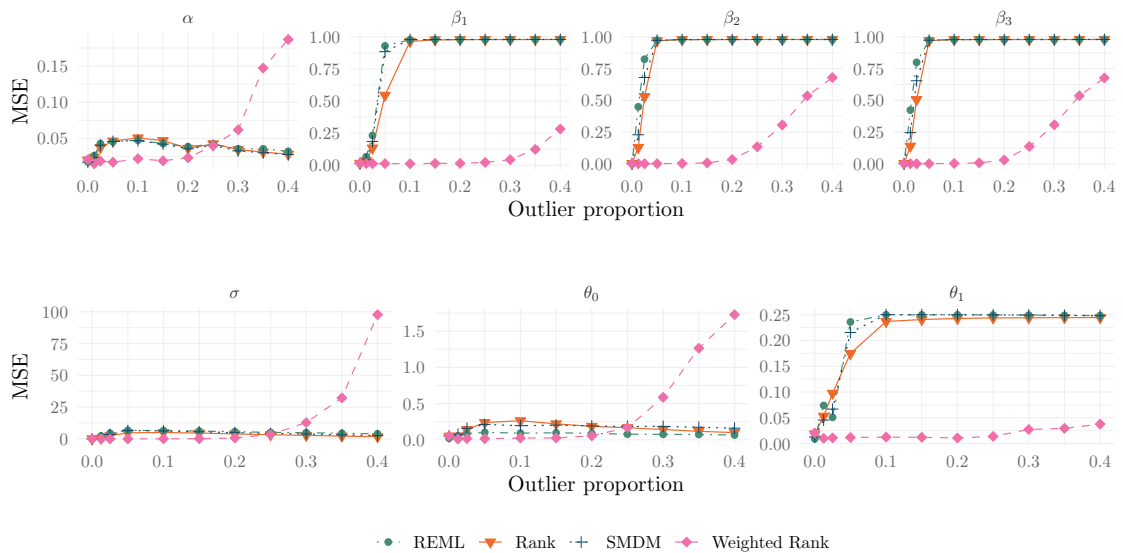


Figure B.7.: MSE of SMDM, REML, Rank and Weighted Rank estimators for an increasing proportion of multiplicative leverage points of size 100, $g = n_i = 20$

B.4.2. Contamination with randomly located additive leverage points

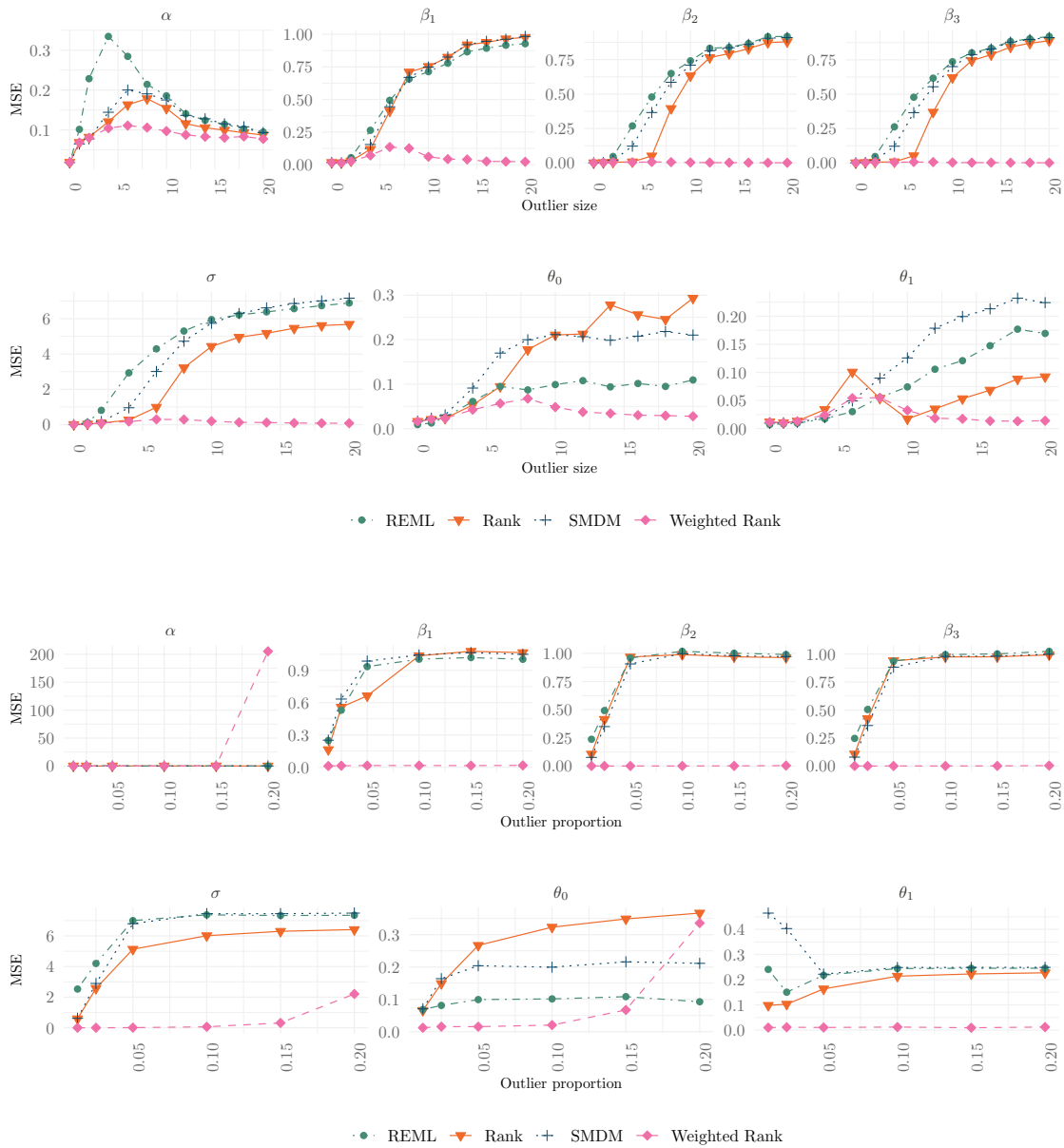


Figure B.8.: MSE of the SMDM, REML, Rank and Weighted Rank estimators for a ten percent proportion of randomly located additive leverage points of increasing size (top) and for an increasing proportion of randomly located additive leverage points of size 100 (bottom), $n_i = g = 20$.

B.4.3. Groupwise contamination

Multiplicative outliers



Figure B.9.: MSE of the SMDM, REML, Rank and Weighted Rank estimators for a ten percent proportion of sequentially located multiplicative leverage points of increasing size (top) and for an increasing proportion of sequentially located multiplicative leverage points of size 100 (bottom), $n_i = g = 20$.

Additive outliers

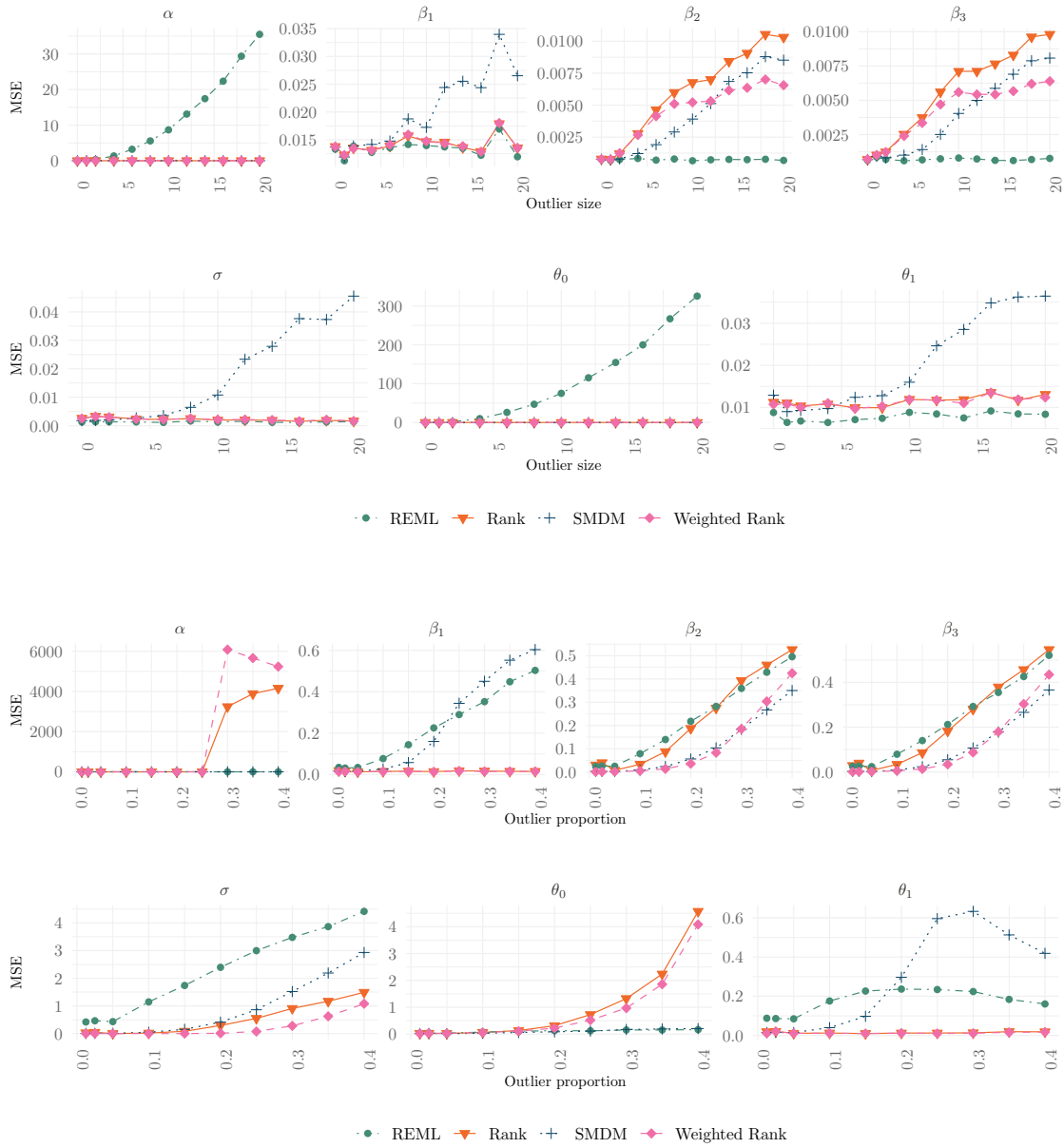


Figure B.10.: MSE of the SMDM, REML, Rank and Weighted Rank estimators for a ten percent proportion of sequentially located additive leverage points of increasing size (top) and for an increasing proportion of sequentially located additive leverage points of size 100 (bottom), $n_i = g = 20$.

B.5. Leverage weights in the presence of dummy variables

The leverage weights in Equation (23) require an estimate of the Mahalanobis distance. However, in the presence of discrete or dummy predictors, we can instead use weights based on the hat-matrix that do not place assumptions on the distribution of the predictors. A possibility is given as follows: Given the hat matrix $X(X'X)^{-1}X'$, its diagonal elements, denoted by $h_{11}, h_{12}, \dots, h_{1n_i}, h_{21}, \dots, h_{gn_g}$, are a measure of the leverage of each observation \mathbf{x}_{ik} . Common cutoffs to flag leverage points are $2 \cdot \frac{p}{n}$ and $3 \cdot \frac{p}{n}$. Thus, following the approach used in (23), we can define weights

$$w_{ik} = \min \left\{ 1, \frac{3 \frac{p}{n}}{h_{ik}} \right\} \quad (\text{B.1})$$

Alternatively, Cantoni and Ronchetti 2001 suggest weight

$$\tilde{w}_{ik} = \sqrt{1 - h_{ik}}.$$

However, weights based on the hat-matrix are not necessarily robust, since an outlying observation can “contaminate” the leverage of the other observations. This can also be seen in Figure B.11 – the weights (B.1) offer some protection against the leverage points, but do not protect fully.

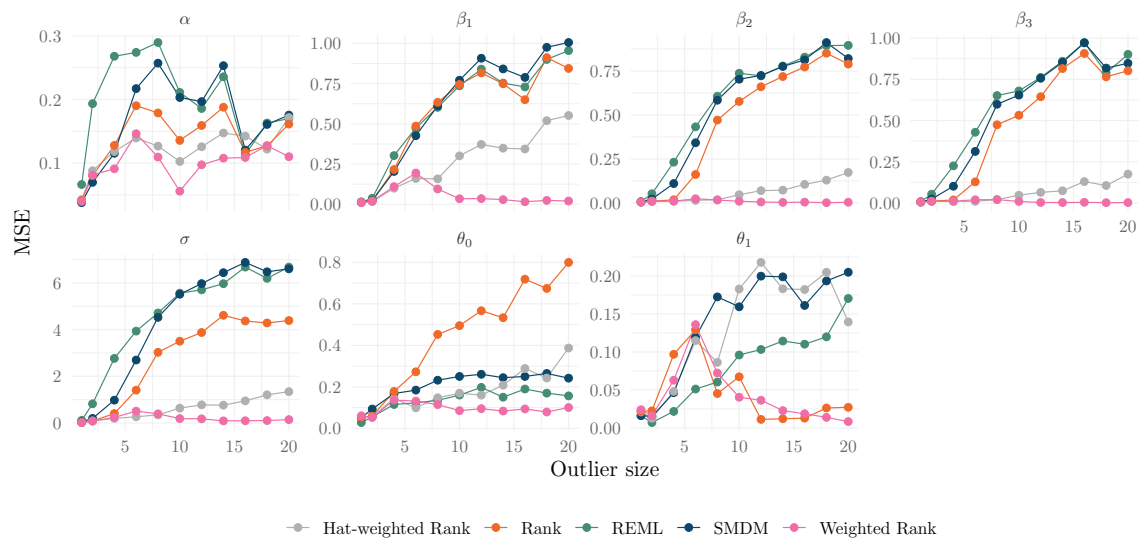


Figure B.11.: MSE of the SMDM, REML, Rank, Weighted Rank and Hat-weighted Rank (with weights (B.1)) estimators for a ten percent proportion of randomly located multiplicative leverage points of increasing size, $n_i = 10$, $g = 20$.

C | Supplement: Chapter 4

Description of the content:

This appendix contains supplemental material to Chapter 4. Section C.1 gives details on the estimation procedure, especially on the algorithm from Boente and Salibián-Barrera (2021) and our extension to robustify against cellwise outliers. Section C.2 contains additional material for the simulation studies: We provide a description of the non-robust method used for comparison, a visualization of the data and outlier scenarios, and additional simulation results complementing those reported in the paper. Finally, Section C.3 provides additional plots and tables for the data examples.

C.1. Details on estimation procedure

C.1.1. Estimation algorithm from Boente and Salibián-Barrera (2021)

The estimation algorithm in Boente and Salibián-Barrera (2021) is based on the following proposition, which especially establishes a linear relationship between different points $\xi(s)$ and $\xi(t)$ under the assumption of an elliptical distribution for the process ξ (see part (b)); for more details on elliptical processes see Boente and Salibián-Barrera (2021). In addition, it provides a way to estimate the scores of the KL decomposition without having to rely on integral approximation. This is especially useful in the case of sparsely observed functions.

Proposition C.1 (Boente and Salibián-Barrera 2021, Prop. 1). *Let $\xi \sim \mathcal{E}(\nu, \Gamma, \varphi)$ be a random element on $L^2(\mathcal{T})$ with $\mathcal{T} \subset \mathbb{R}$, and assume that the kernel γ associated with Γ is continuous. Let $\eta_1 \geq \eta_2 \geq \dots$ be the non-null eigenvalues of Γ and ψ_ℓ the eigenfunction of Γ associated with η_ℓ , chosen so that the set $\{\psi_\ell, \ell \in \mathbb{N}\}$ is an orthonormal set in $L^2(\mathcal{T})$. Then*

- (a) *For any fixed m and t_1, \dots, t_m in \mathcal{T} , the random vector $\boldsymbol{\xi}_m = (\xi(t_1), \dots, \xi(t_m))'$ has an elliptical distribution in \mathbb{R}^m with location $\boldsymbol{\nu}_m = (\nu(t_1), \dots, \nu(t_m))'$ and scatter matrix Σ with elements $\Sigma_{\ell,j} = \gamma(t_\ell, t_j)$, $1 \leq \ell, j \leq m$.*
- (b) *For any $t' \neq t \in \mathcal{T}$, we have that $\xi(t)|\xi(t') \sim \mathcal{E}(\nu_{t|t'}, \sigma_{t|t'}, \varphi_{t|t'})$ where the conditional location is given by*

$$\nu_{t|t'} = \nu(t) + \frac{\gamma(t, t')}{\gamma(t', t')}(\xi(t') - \nu(t')).$$

- (c) *For any fixed m and $t_1, \dots, t_m \in \mathcal{T}$, let $\boldsymbol{\xi}_m = (\xi(t_1), \dots, \xi(t_m))'$. We have that $\zeta_\ell | \boldsymbol{\xi}_m \sim$*

$\mathcal{E}(\nu_\ell, \sigma_\ell, \phi_{X_m}^*)$ where

$$\nu_\ell = \eta_\ell \psi'_{\ell,m} \Sigma_{\xi_m}^{-1} (\xi_m - \nu_m)$$

with $\psi_{\ell,m} = (\psi_\ell(t_1), \dots, \psi_\ell(t_m))'$, $\nu_m = (\nu(t_1), \dots, \nu(t_m))'$ and the (p, j) element of Σ_{ξ_m} equals $\gamma(t_p, t_j)$.

Given an observed data set $\{\tilde{\xi}_{k,ij}, t_{ij}\}_{i=1, \dots, n; j=1, \dots, n_i; k=1, \dots, K}$ obtained from estimation step 2, Equation (4.9), we aim to obtain a representation for the score functions $\xi_{i,k}$ of form (4.10) for each $k = 1, \dots, K$. Due to (4.6), we assume the estimates to be centered.

Boente and Salibián-Barrera (2021) propose the following estimation algorithm for the covariance functions γ_k , $k = 1, \dots, K$, based on Proposition C.1. Since the estimates are centered, the first step in their algorithm, the estimation of the mean function $\nu(t)$ can be skipped.

Diagonal of the covariance function In the first step of the algorithm, the diagonal of the covariance function $\gamma_k(t, t)$ is estimated using a robust M-scale estimator. Let $\tilde{\rho} : \mathbb{R} \rightarrow \mathbb{R}$ be a bounded ρ -function such that $\sup_t \tilde{\rho}(t) = 1$. Furthermore, let $b \in (0, 1)$ and define weights

$$w_{k,ij}(t) = \kappa \left(\frac{t_{ij} - t}{h} \right) \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \kappa \left(\frac{t_{ij} - t}{h} \right) \right\}^{-1}.$$

$\kappa : \mathbb{R} \rightarrow \mathbb{R}$ is a kernel function (e.g. Epanechnikov kernel) that assigns higher weight to observations close to t and h is a bandwidth parameter. In practice, it can e.g. be chosen by cross-validation. The estimate is then obtained by solving

$$\sum_{i=1}^n \sum_{j=1}^{n_i} w_{k,ij}(t) \tilde{\rho} \left(\frac{\tilde{\xi}_{k,ij}}{\hat{\gamma}_k(t, t)} \right) = b$$

for $\hat{\gamma}_k(t, t)$, where t are selected to form a fine grid in \mathcal{T} . Boente and Salibián-Barrera (2021) set $b = 1/2$, ensuring consistency and maximum breakdown point in the case of i.i.d. observations with Gaussian error.

Estimation of the off-diagonal elements Relationship (4.11) implies that the conditional mean of $\xi_k(t)$ given $\xi_k(t')$ is a linear function of $\xi_k(t')$. The slope in this linear relationship is given by $\gamma_k(t, t')/\gamma_k(t', t')$. Thus, the off-diagonal elements of the covariance function can be estimated using a local M-regression estimator.

$$\hat{\beta}_k(t, t') = \arg \min_{\beta \in \mathbb{R}} \sum_{i=1}^n \sum_{j \neq p} \rho \left(\frac{\tilde{\xi}_{k,ij} - \beta \tilde{\xi}_{k,ip}}{\hat{s}_k(t, t')} \right) \kappa \left(\frac{t_{ij} - t}{h} \right) \kappa \left(\frac{t_{ip} - t'}{h} \right), \quad (\text{C.1})$$

and letting

$$\tilde{\gamma}_k(t, t') = \hat{\gamma}_k(t', t') \cdot \hat{\beta}_k(t', t).$$

In (C.1), $\hat{s}(t, t')$ is an initial robust scale estimator and ρ is a bounded ρ function. The implementation uses Tukey's biweight function. Following Boente and Salibián-Barrera (2021), $\hat{s}(t, t')$ is obtained as the local MAD of the residuals from an initial robust estimator. Namely, let

$$\tilde{\beta}_k(t, t') = \text{med}_{|t_{ij}-t|\leq h, |t_{il}-t'|\leq h} \frac{\tilde{\xi}_{k,ij}}{\tilde{\xi}_{k,il}},$$

and let $\hat{s}_k(t, t')$ be the MAD of the residuals $r_{i,k}(t, t') = \tilde{\xi}_{k,ij} - \tilde{\beta}_k(t, t')\tilde{\xi}_{k,il}$, i.e.

$$\hat{s}_k(t, t') = \text{med}_{|t_{ij}-t|\leq h, |t_{il}-t'|\leq h} |r_{i,k}(t, t') - m_k(t, t')|,$$

where $m_k(t, t') = \text{med}_{|t_{ij}-t|\leq h, |t_{il}-t'|\leq h} r_{i,k}(t, t')$.

Postprocessing In order to ensure that the estimated covariance function $\hat{\gamma}(\cdot, \cdot)$ is smooth and symmetric, $\hat{\beta}_k(t', t)$ is calculated as well. A bivariate smoother (e.g. bivariate B-spline smoother) is applied to both of the estimated covariance surfaces, yielding $\tilde{\tilde{\gamma}}_k(t, t')$ and $\tilde{\tilde{\gamma}}_k(t', t)$, and the final estimated covariance function is obtained as

$$\hat{\gamma}(t, t') = (\tilde{\tilde{\gamma}}_k(t, t') + \tilde{\tilde{\gamma}}_k(t', t))/2 \quad (\text{C.2})$$

The eigenfunction-eigenvalue pairs, denoted by $\{\hat{\eta}_{k,\ell}, \hat{\psi}_{k,\ell}(\cdot)\}_\ell$ can be calculated by applying standard FPCA to the estimated covariance surface. If (C.2) is not positive definite, we set $\hat{\gamma}_k(s, t) = \sum_{\ell: \hat{\eta}_{k\ell} \geq 0} \hat{\eta}_{k\ell} \hat{\psi}_{k\ell}(s) \hat{\psi}_{k\ell}(t)$.

C.1.2. Variance of the standardized residuals

For $\xi_k(t')$ and $\xi_k(t)$, Proposition C.1 allows to calculate the theoretical variance of the residuals $r_{k,i}(t_{ij}, t_{ip}) = \tilde{\xi}_{k,ij} - \hat{\beta}(t_{ij}, t_{ip})$ as follows. For two time points $t, t' \in \mathcal{T}$, we have:

$$\begin{aligned} \text{Var}(\xi_k(t) - \beta_k(t, t')\xi_k(t')) &= \text{Var}(\xi_k(t)) + \text{Var}(\beta_k(t, t')\xi_k(t')) - 2\text{Cov}(\xi_k(t), \beta_k(t, t')\xi_k(t')) \\ &= \gamma_k(t, t) + \left(\frac{\gamma_k(t, t')}{\gamma_k(t', t')} \right)^2 \gamma_k(t', t') - 2 \frac{\gamma_k(t, t')}{\gamma_k(t', t')} \gamma_k(t, t') \\ &= \gamma_k(t, t) - \frac{\gamma_k(t, t')^2}{\gamma_k(t', t')}, \end{aligned}$$

making use of $\beta_k(t, t') = \frac{\gamma_k(t, t')}{\gamma_k(t', t')}$. An estimate is obtained by replacing the theoretical quantities with their corresponding estimators.

C.1.3. Intuition for the robustification against cellwise outliers

In the following, we give a short intuition for our robustification of the methodology in Boente and Salibián-Barrera (2021) against “cellwise” outliers. For this, we have a realization of a curve from the process

$$X(t) = \zeta_{i1}\sqrt{2}\cos(2\pi t) + \zeta_{i2}\sin(2\pi t), \quad (\text{C.3})$$

where $\zeta_{i1} \sim \mathcal{N}(0, 3)$ and $\zeta_{i2} \sim \mathcal{N}(0, 1.5)$. The theoretical covariance function is given by

$$\gamma(s, t) = 6\cos(2\pi t)\cos(2\pi s) + 3\sin(2\pi t)\sin(2\pi s).$$

We now contaminate our observation with 3 outliers by adding or subtracting a shift of size $s = 3$ to a few points. This yields the contaminated observation visualized in Figure C.1.

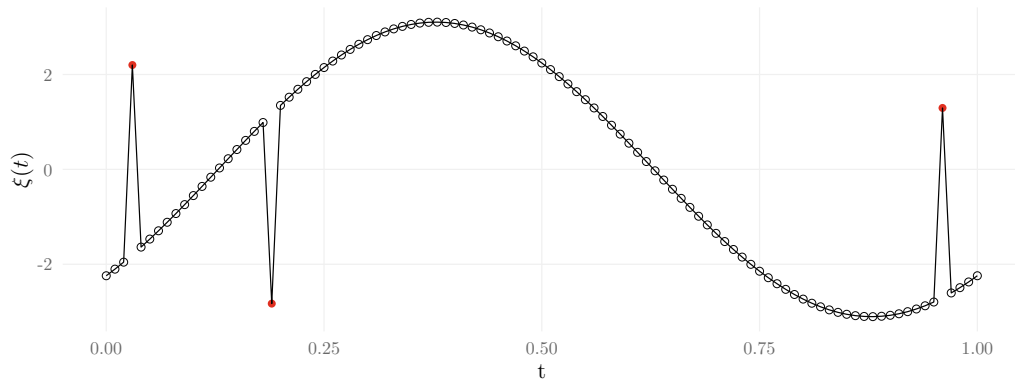


Figure C.1.: Contaminated functional observation from the process (C.3).

Now we can calculate the standardized residuals as in (4.13), given by

$$\tilde{r}_{k,i}(t, t') := \frac{|r_{k,i}(t, t')|}{\sqrt{\hat{\gamma}_k(t, t) - \frac{\hat{\gamma}_k(t, t')^2}{\hat{\gamma}_k(t', t')}}} \quad (\text{C.4})$$

for any pair t, t' . The results can be visualized in a matrix as displayed in C.2

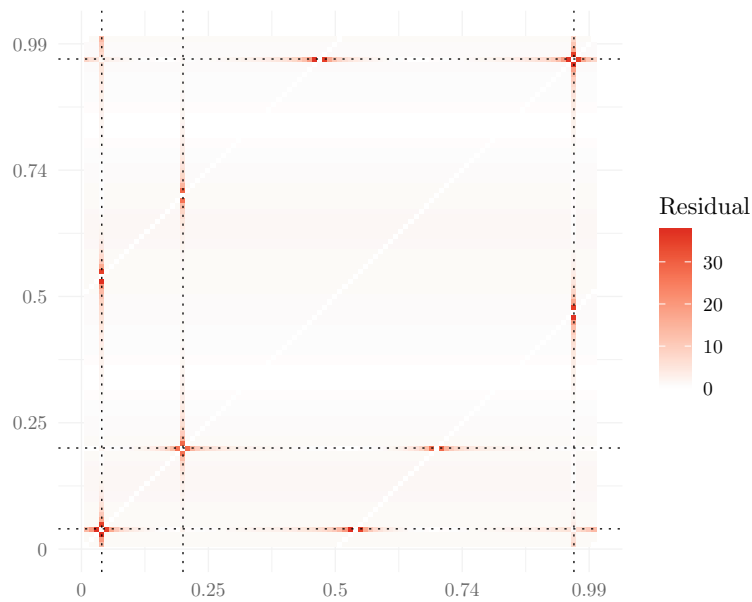


Figure C.2.: Standardized residuals (C.4) for the observation displayed in Figure C.1. The three outlying cells are marked by dotted lines.

The residuals for the outlying observations are very large. At the same time, the residuals of the neighboring, highly correlated observations are also high, since they are calculated in comparison with the outlying value. If we now construct a measure of “overall outlyingness” for each observed point t , an obvious solution is to average over the individual outlyingness of each data point. However, the estimator used for averaging should be robust to ensure that the time points highly correlated with the outlying ones are not accidentally flagged as outlying as well. Suitable choices are trimmed means and the median, as we use in Equation (4.14). The resulting average outlyingness for the standard mean (left) and the median (right) is depicted in Figure C.3. The mean outlyingness of the observations that are highly correlated with the outliers is drawn upwards. At the same time, median outlyingness of those observations remains largely unaffected.

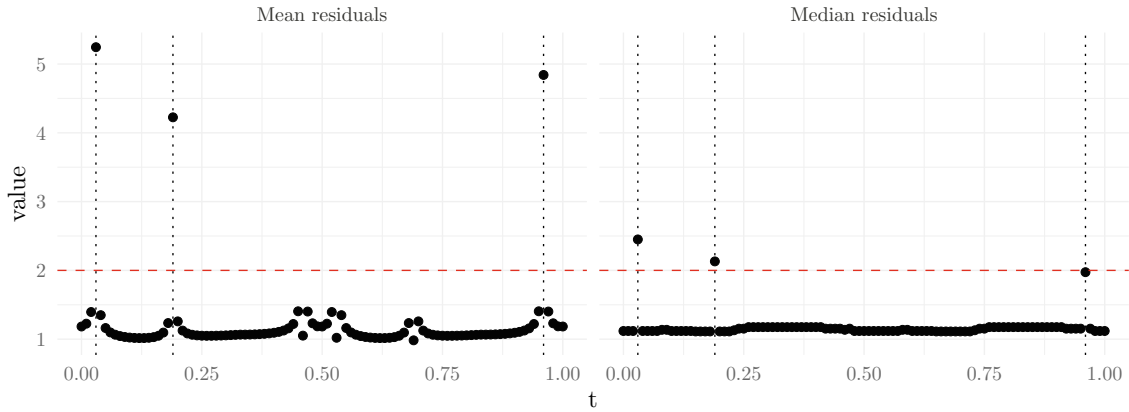


Figure C.3.: Mean and median standardized residuals calculated from the residuals C.2. The three outlying points are marked by dotted lines. The dashed red line depicts the recommended cutoff $c = 2$.

The cutoff c is a tuning parameter.

C.2. Additional material for the simulation studies

C.2.1. Description of the non-robust estimation method

The non-robust estimates in the simulation study are calculated using a method that replaces the proposed robust components with non-robust ones.

Estimation of marginal covariance function The marginal covariance function is obtained using the estimator used in Park and Staicu (2015) and Chen, Delicado, et al. (2017), i.e.

$$\hat{\Sigma}_{\text{classic}}(s_r, s_{r'}) = \frac{1}{(\sum_{i=1}^n n_i) - 1} \sum_{i=1}^n \sum_{j=1}^{n_i} Y_i^C(s_r, t_{ij}) Y_i^C(s_{r'}, t_{ij}).$$

We do not apply any smoothing to the raw covariance matrix and calculate the eigenfunctions and corresponding eigenvalues using a standard spectral decomposition of the matrix $\hat{\Sigma}_{\text{classic}}$ evaluated in all combinations $s_r, s_{r'}, r, r' \in \{1, \dots, R\}$.

Estimation of the score functions The score functions are estimated using the PACE method (Yao et al. 2005) implemented in `fpc.sc()` from the `refund` package (Goldsmith et al. 2024) with the default parameter choices.

C.2.2. Illustration of data in the simulation study

The following Figure C.4 displays a functional observation generated in the simulation study. The top-left panel shows a raw observation without outliers. The corresponding score functions are displayed in Figure C.5. The eight unstructured eigenfunction cause the “wigglyness” of the raw observation. This observation is modified in three ways to visualize the effects of the three outlier settings: For OS1 (top-right panel), the scale of all observations is increased, while the structure stays the same. For OS2 (bottom-left panel), there is one curve that deviates from the majority. For OS3 (bottom-right panel), no structure can be observed.

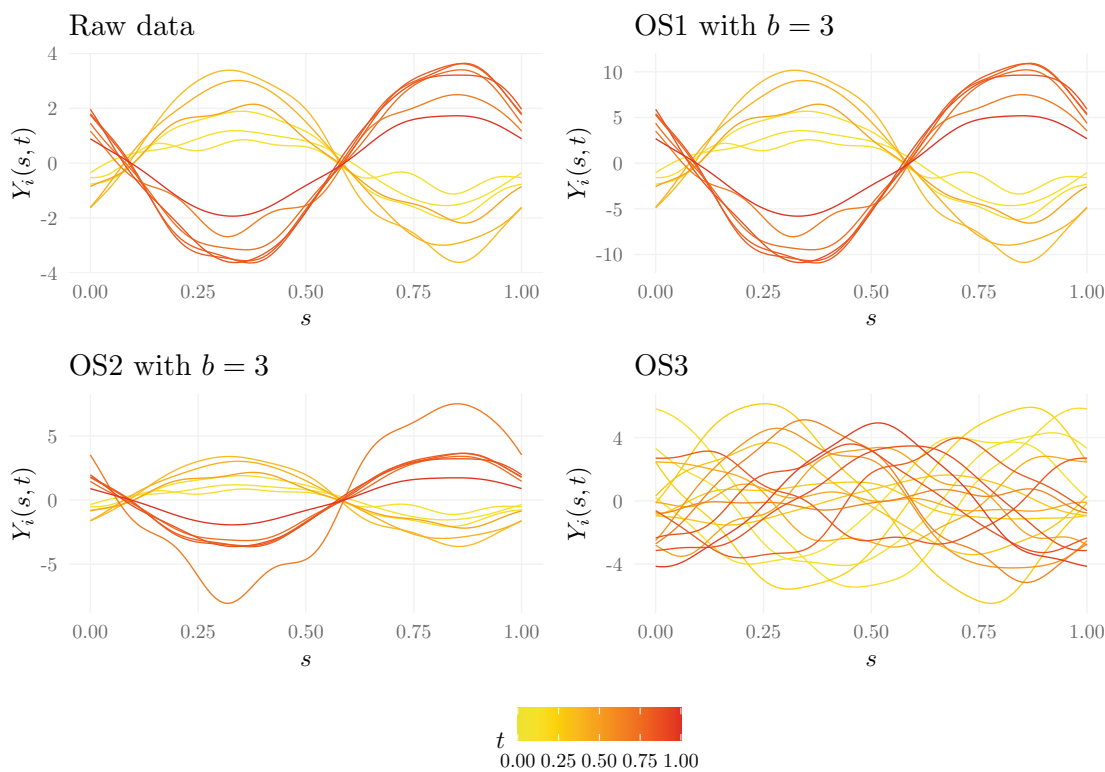


Figure C.4.: Visualization of the raw data and the three outlier types in the simulation study.

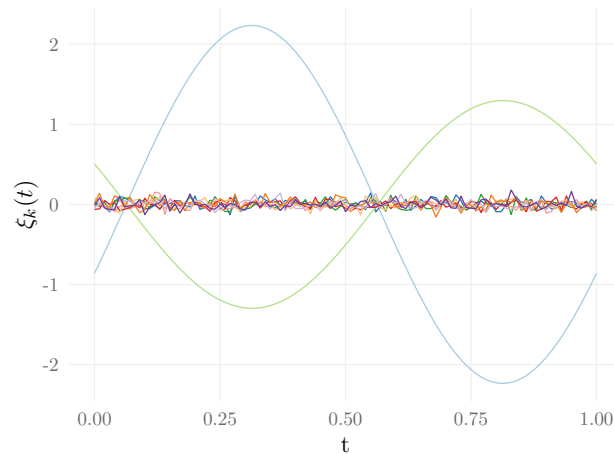


Figure C.5.: Score functions for the observation displayed in top-left panel of Figure C.4.

C.2.3. Additional simulation results

Simulation results for OS3

In the setting with shape or structural outliers OS3, the reconstruction errors for the robust method are unaffected by the outlying score functions, as shown in Figure C.6. The RMSE (4.19) stays stable with increasing proportion of contaminated subjects. The outlier detection based on the orthogonal distance OD_i in (4.16) reliably detects all outlying individuals (TPR=1) while preserving a false positive rate below 5%. In outlier setting OS3, score distance SD_i in (4.16) completely fails to recognize the outliers, as expected. However, it also keeps the nominal level of $\alpha = 0.05$.

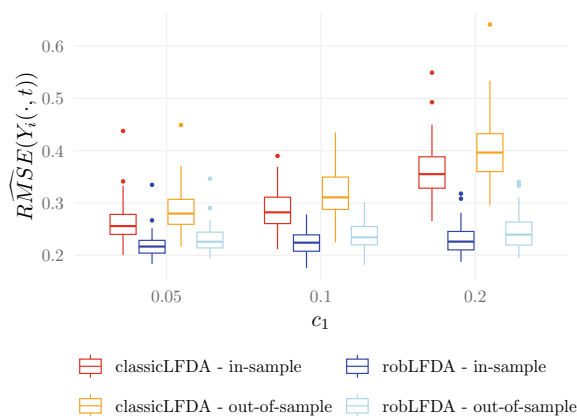


Figure C.6.: Comparison of RMSE (4.19) in OS3 for sample size $n = 100$, sparse setting, $c_1 \in \{0.05, 0.1, 0.2\}$.

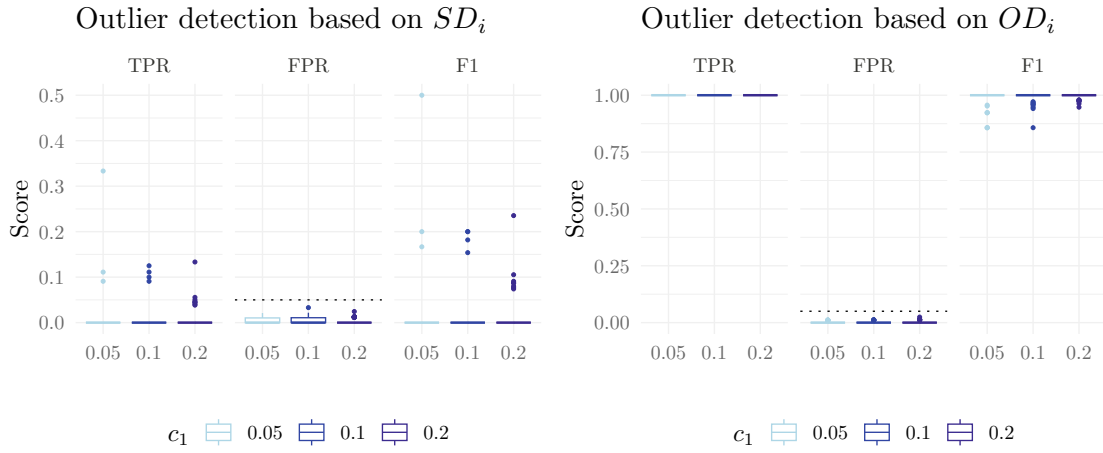


Figure C.7.: Left to right: TPR, FPR, and F1 score for outlier detection based on the score distance SD_i (left panel) orthogonal distance OD_i (right panel) in (4.16) in OS3 with $b \in \{3, 5, 10, 20\}$ for sample size $n = 100$, $c_1 = 0.1$, sparse setting.

Reconstruction errors in OS1 and OS2

Figures C.8 and C.9 show boxplots of the reconstruction errors for the score functions ξ_{ik} , for outlier settings OS1 and OS2, respectively. The reconstruction error is stable for the robust method, but strongly affected by the outlying values for the non-robust method.

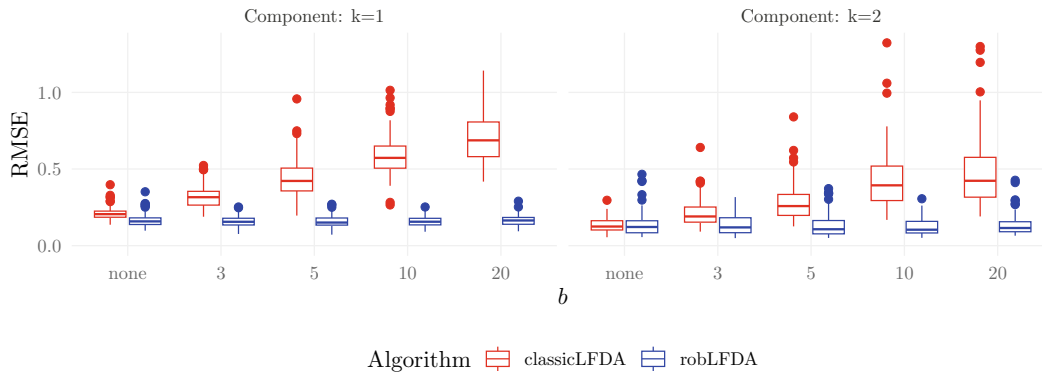


Figure C.8.: RMSE (4.20) in OS1 with $b \in \{3, 5, 10, 20\}$ for $n = 100$, sparse setting, $c_1 = 0.1$

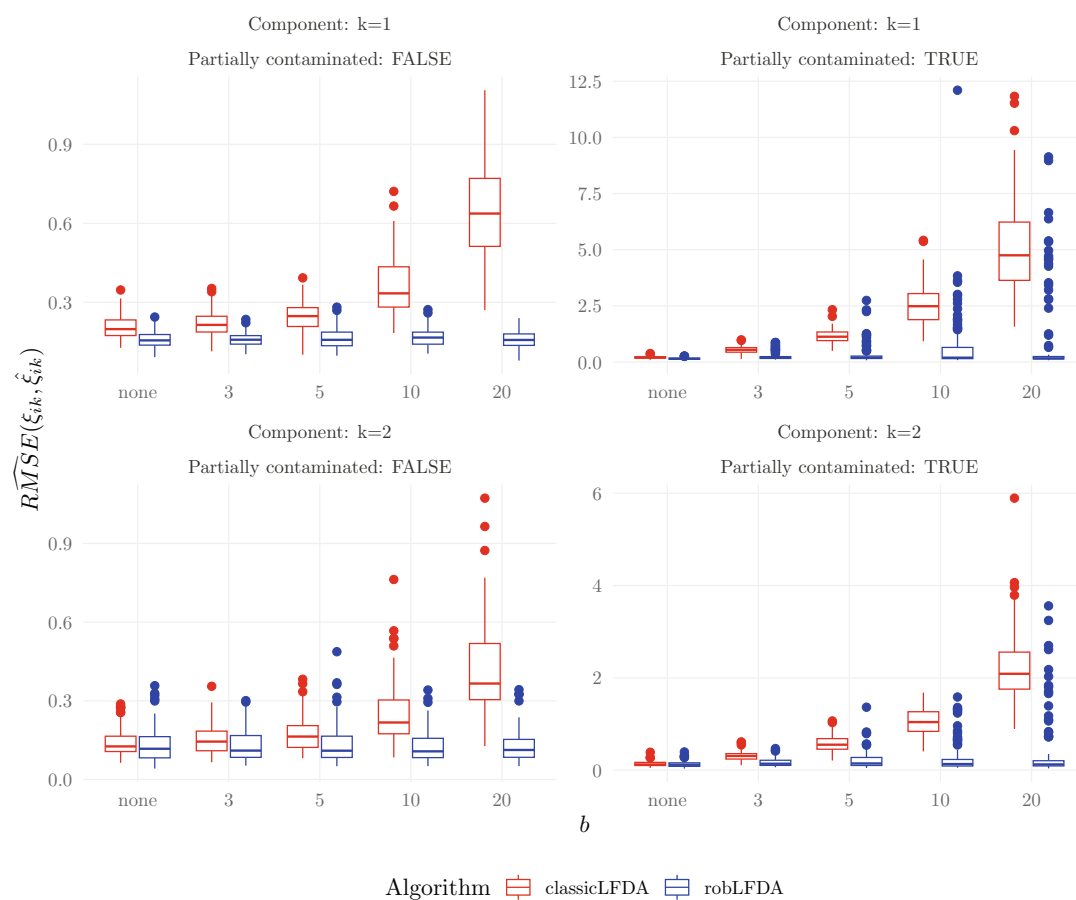


Figure C.9.: RMSE (4.20) for OS2 with $b \in \{3, 5, 10, 20\}$ for $n = 100$, sparse setting, $c_1 = c_2 = 0.1$. Separated by non-contaminated (left) and contaminated (right) subjects.

The following Figures C.10 and C.11 visualize the effect of decreasing the number of observational units or subjects n or and the number of curves n_i in outlier setting OS1, respectively. As expected, increasing the sample size in terms of n or n_i reduces the overall reconstruction errors (4.19) as well as their variances. The estimates stabilize with increasing number of available samples.

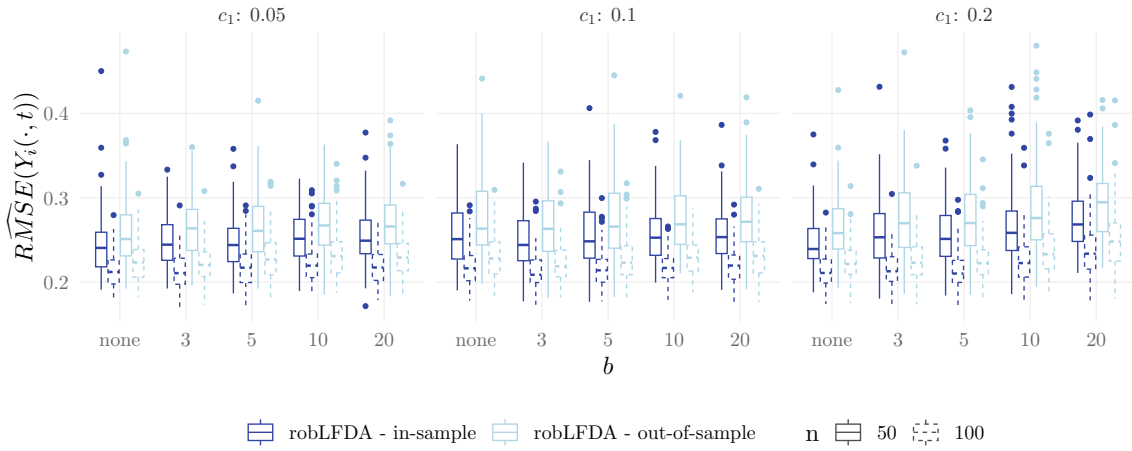


Figure C.10.: Comparison of RMSE (4.19) for $n \in \{50, 100\}$ in OS1, $c_1 = 0.1$, sparse setting.

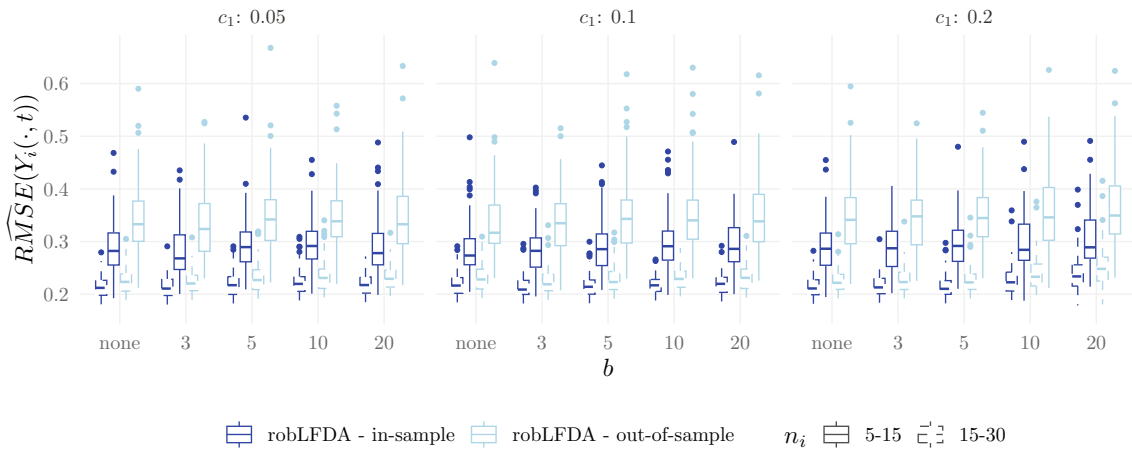


Figure C.11.: Comparison of RMSE (4.19) for sparse and very sparse setting in OS1, $n = 100$, $c_1 = 0.1$.

Furthermore, we compare the RMSE (4.19) in outlier setting OS2 for different proportions of outlying curves $c_2 \in \{0.1, 0.2\}$. The proportion of contaminated curves within the samples affects the reconstruction errors within the partially contaminated subjects. However, the errors stay stable within the non-contaminated subjects.

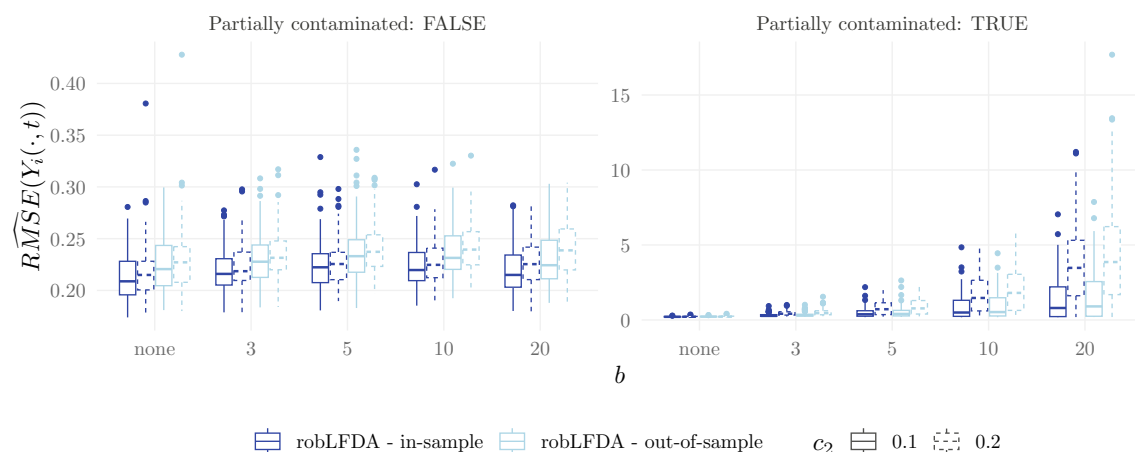


Figure C.12.: Comparison of RMSE (4.19) for proportions of contamination $c_2 \in \{0.1, 0.2\}$ in OS2, $n = 100$, $c_1 = 0.1$, sparse setting.

Outlier analysis results for OS1 and OS2

Figure C.13 shows the TPR, FPR and F1 score for outlier detection based on the orthogonal distance in (4.16). As expected, orthogonal distance does not recognize the amplitude outliers if they are small, however, if they are large enough, both orthogonal and score distance give reasonable results.

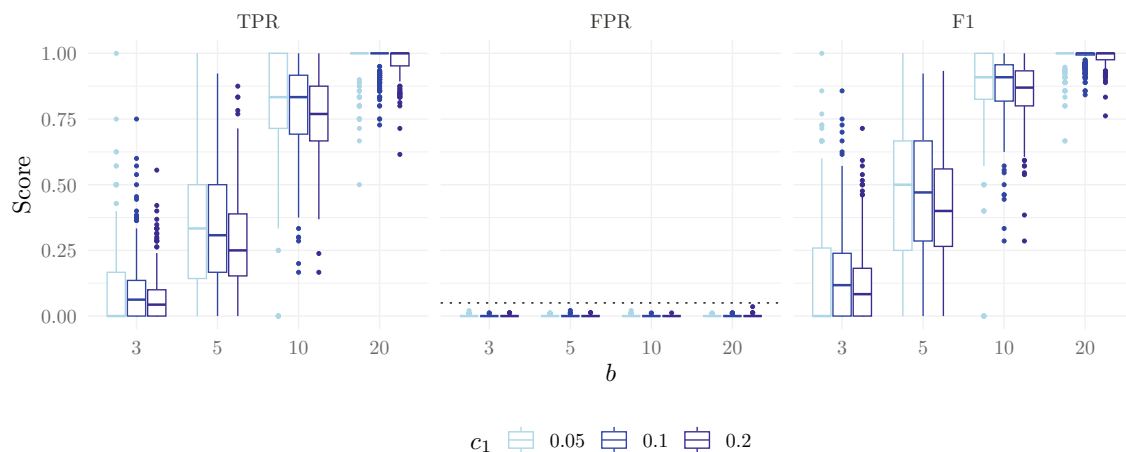


Figure C.13.: Left to right: TPR, FPR, and F1 score for outlier detection based on the orthogonal distance OD_i in (4.16) in OS1 with $b \in \{3, 5, 10, 20\}$ for sample size $n = 100$, $c_1 = 0.1$, sparse setting.

C.3. Additional material for the data examples

C.3.1. Analysis of mortality data

Table C.1.: Overview of countries and corresponding codes.

Code	Country	Code	Country
AUS	Australia	GBR-NP	United Kingdom
AUT	Austria	GBR-SCO	Scotland
BGR	Bulgaria	HUN	Hungary
BLR	Belarus	IRL	Ireland
CAN	Canada	ISL	Iceland
CHE	Switzerland	ITA	Italy
CZE	Czechia	JPN	Japan
DEUTE	East Germany	LTU	Lithuania
DEUTW	West Germany	LVA	Latvia
DNK	Denmark	NLD	Netherlands
ESP	Spain	NOR	Norway
EST	Estonia	POL	Poland
FIN	Finland	PRT	Portugal
FRACNP	France (civilian)	RUS	Russia
FRATNP	France (total)	SVK	Slovakia
GBRCENW	England and Wales (civilian)	SWE	Sweden
GBRTENW	England and Wales (total)	UKR	Ukraine
GBR-NIR	Northern Ireland	USA	United States of America

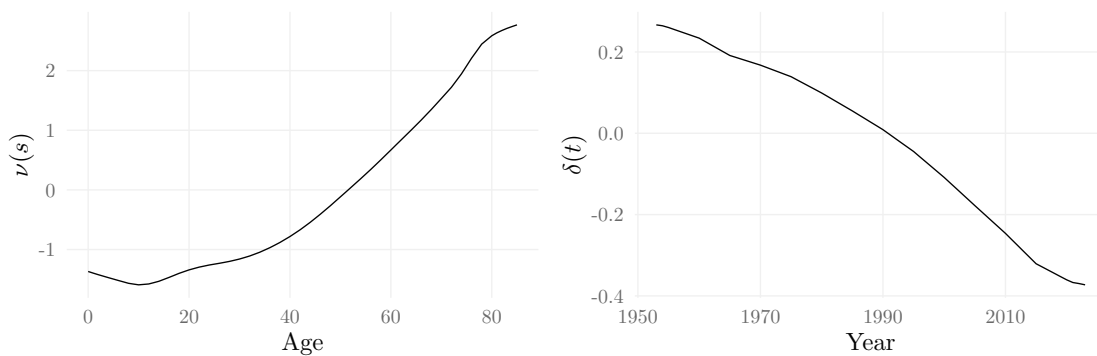


Figure C.14.: Components of the estimated mean function for the mortality dataset, parametrized as $\mu(s, t) = \mu_0 + \nu(s) + \delta(t)$.

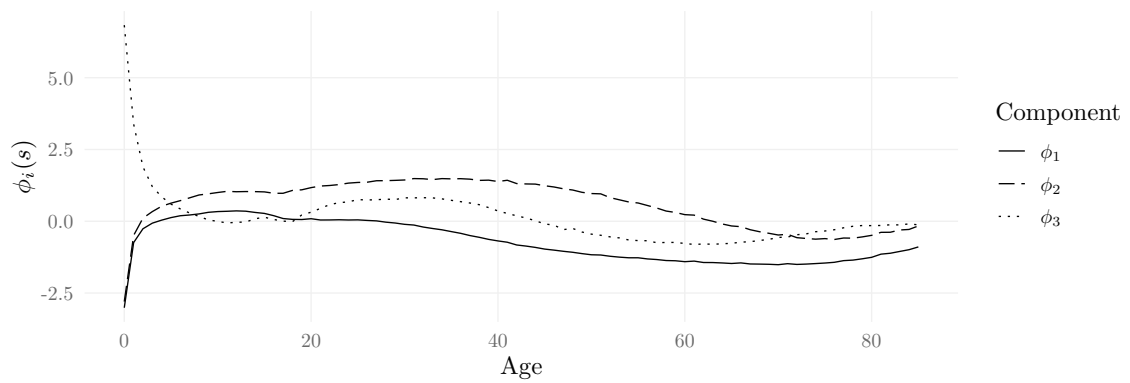


Figure C.15.: Estimated marginal eigenfunctions $\phi_k(s)$, $k = 1, \dots, 3$, for the mortality dataset.

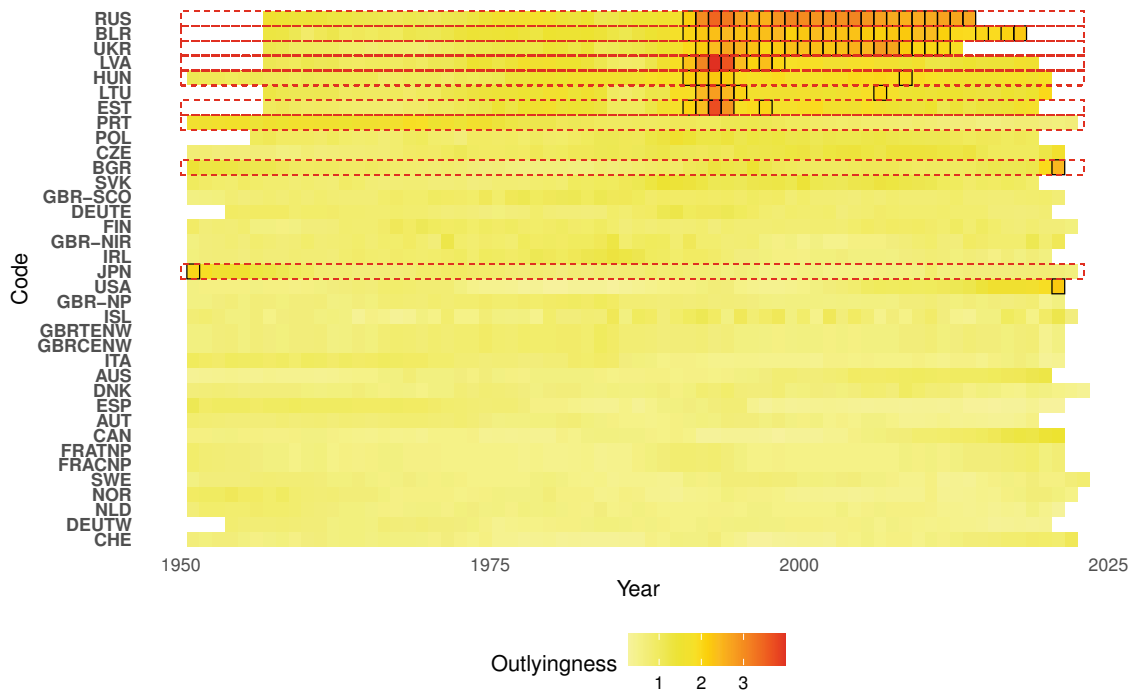


Figure C.16.: Outlyingness (4.17) by year for the 36 countries in the mortality data analysis. The subjects flagged by SD_i or OD_i using the Bonferroni procedure (Corollary 4.3) are marked by red dashed lines.

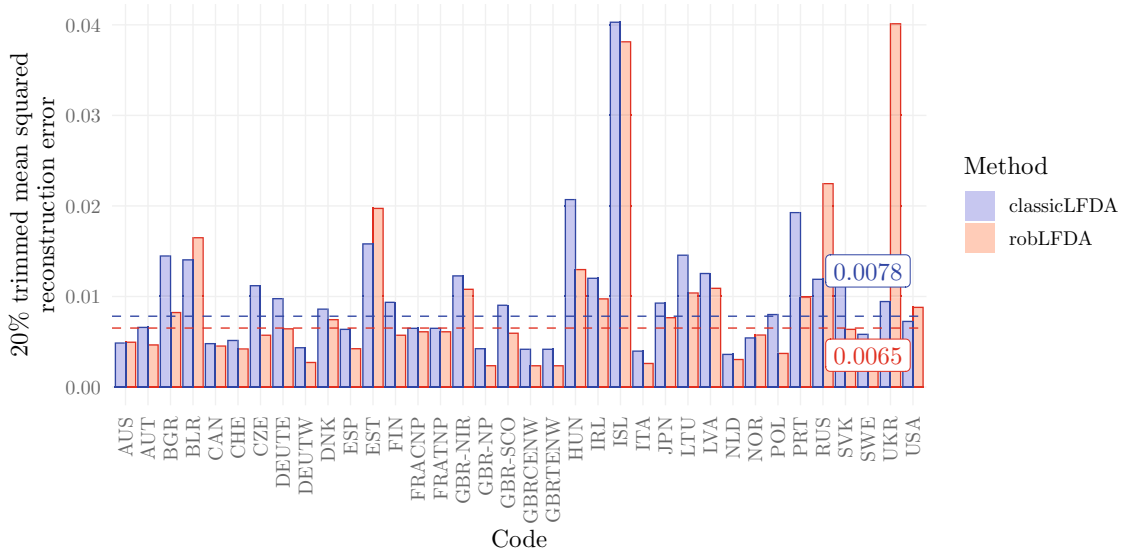


Figure C.17.: Comparison of 20% trimmed mean squared reconstruction errors on the test data; mortality dataset. The average error across all countries is represented by the dashed lines.

C.3.2. Analysis of UV-F spectra

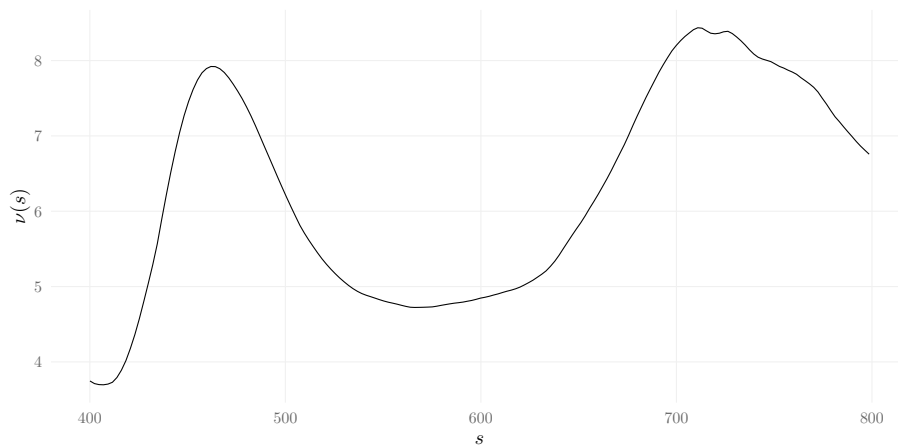


Figure C.18.: Estimated mean function for the UV-F dataset, parametrized as $\mu(s, t) = \mu_0 + \nu(s)$.

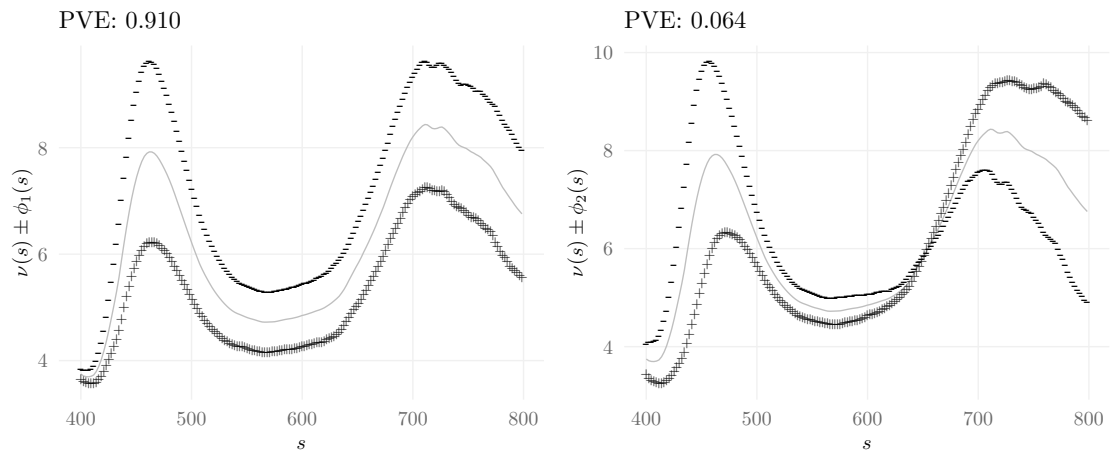


Figure C.19.: Estimated mean function $\nu(s)$ plus and minus the estimated marginal eigenfunctions $(\nu(s) \pm \phi_k(s), k = 1, 2)$ for the UV-F dataset.

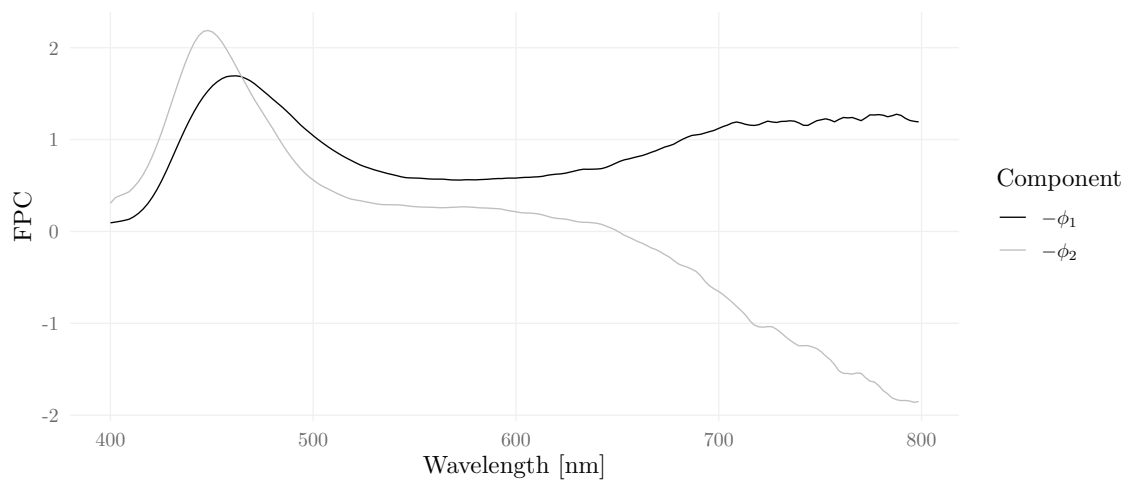


Figure C.20.: Marginal eigenfunctions $\phi_k(s), k = 1, 2$ for the UV-F dataset.

Table C.2.: Aging settings considered in the analysis of photovoltaic material degradation data. The module temperature and humidity differ from the climate chamber temperature and humidity in the presence of irradiance and are reported in parentheses. DML = Dynamic mechanical load (simulation of wind), TC = Temperature cycles

Setting	Total	Cycle	Duration	Temperature	Humidity	Irradiance	DML	TC	Number of modules
Moderate 1	1000 h	1	1000 h	85 °C (113 °C)	85% (30.4%)	1000 W/m ²	-	-	3
Alpin 1	2000 h	1	250 h	85 °C	85%	-	-	-	3
		2	250 h	85 °C (119 °C)	85%	1200 W/m ²	-	-	
		3	24 h	-	-	-	1000 c	-	
Arid 1	1200 h	1	1200 h	95 °C (129 °C)	50% (15.7%)	1200 W/m ²	-	-	3
Moderate 5	3000 h	1	48 h	60 °C (78 °C)	40% (18.1%)	1000 W/m ²	-	-	3
		2	96 h	85 °C	85%	-	-	-	
		3	300 h	-40/85 °C	-	-	-	50	
Tropical 2	3000 h	1	3000 h	90 °C	90%	-	-	-	6

Curriculum Vitae

Personal Data

Name	Barbara Saskia Brune
Date of birth	07.10.1995
Place of birth	Herdecke, Germany
Nationality	German
E-mail	barbara.brune@ymail.com

Education

since 02/2020	PhD studies at the Institute for Statistics and Mathematics in Economics, TU Wien, Vienna, Austria
11/2019	Master's studies: Statistics, minor mathematics, TU Dortmund University, Germany; Master thesis: <i>On Subgraph Counts and Goodness-Of-Fit Testing for Stochastic Network Models</i>
01/2017–08/2017	Visiting scholarship: Colorado School of Mines, Department of Mining Engineering, Golden, Colorado, US
11/2016	Bachelor's studies: Statistics, minor mathematics, TU Dortmund University, Germany; Bachelor thesis: <i>Vergleich von robusten Zweistichproben-Tests für diskrete Daten</i>
06/2013	Abitur, Helene-Lange-Gymnasium, Dortmund, Germany

Scientific publications:

- B. Brune, I. Ortner and P. Filzmoser (2024): *A Rank-Based Estimation Method for Mixed Effects Models in the Presence of Outlying Data*. Conditionally accepted for publication in: *Journal of Data Science, Statistics, and Visualization*.

- B. Brune, J. Floßdorf and C. Jentsch (2024): *Goodness-of-Fit Testing Based On Graph Functionals for Homogeneous Erdős-Rényi Graphs*. Scandinavian Journal of Statistics, <https://doi.org/10.1111/sjos.12750>.
- B. Brune, I. Ortner, G. C. Eder, Y. Voronko, A. Gassner, K. A. Berger, G. Oreski, K. Knöbl, L. Neumaier and M. Feichtner (2023): *Connecting Material Degradation and Power Loss of PV Modules Using Advanced Statistical Methodology*. Solar Energy Materials and Solar Cells, 260, <https://doi.org/10.1016/j.solmat.2023.112485>.
- B. Brune, I. Ortner, G. C. Eder, Y. Voronko, G. Oreski, K. A. Berger, K. Knöbl, L. Neumaier and M. Feichtner (2023): *Quantifying the Influence of Encapsulant and Backsheet Composition on PV-Power and Electrical Degradation*. Progress in Photovoltaics, 31(7), pp. 716–128, <https://doi.org/10.1002/pip.3679>.
- S. Abbas, B. Brune and R. Fried (2023): *robnptests – An R package for Robust Two-Sample Location and Dispersion Tests*. Journal of Open Source Software, 8(82), 4947, <https://doi.org/10.21105/joss.04947>

Conference talks:

- *Robust Rank-Based Estimation of Mixed Effects Models*, CMStatistics 2022, December 17–19, 2022, London, UK
- *Rank-Based Estimation of Mixed Effects Models*, Austrian and Slovenian Statistical Days 2022, April 20–22, 2022, Graz, Austria
- *A State-Space Approach to Time-Varying Reduced Rank Regression*, CMStatistics 2021, December 18–20, 2021, London, UK
- *The Use of Mixed Effects Models to Characterize the Degradation Behavior of PV Modules*, Workshop der Österreichischen Technologieplattform Photovoltaik (TPPV) – Machine Learning for Data Evaluation, Predictive Maintenance and Material-Design (S. Block I b), December 14, 2021, Bolzano, Italy.
- *A Comparison Study of Robust Mixed Effects Models for Analyzing Degradation of Photovoltaic Modules and a New Rank-Based Estimation Approach*, International Conference on Robust Statistics 2021, September 20–24, 2021, Vienna, Austria