Statistical Modeling of Regime Changes, Structural Breaks, and Heterogeneity in Financial data

Zhenni Tan

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN STATISTIC
YORK UNIVERSITY
TORONTO, ONTARIO

June 2026

© Zhenni Tan, 2026

Abstract

Financial time series exhibit complex characteristics, including structural breaks, regime changes, heavy tails, cross-sectional heterogeneity, and high dimensionality, all of which pose significant challenges to traditional statistical models. Developing robust statistical methodologies that account for these features is essential for accurate modeling, inference, and risk assessment. This dissertation presents three complementary methodological contributions that address these challenges from the perspectives of regime-switching modeling, change-point detection, and multi-platform information integration.

The first part develops a robust hidden semi-Markov regime-switching (rHSMS) model for financial time series with heavy-tailed returns and persistent market regimes. Unlike conventional hidden Markov models, the proposed approach relaxes the Markovian assumption on regime durations by allowing more flexible state sojourn distributions. To improve robustness against outliers and extreme observations, Huber M-estimation is incorporated into the emission distributions within a likelihood-based estimation framework. The resulting model provides a robust and interpretable representation of market dynamics while preserving efficient statistical inference.

The second part addresses the problem of change-point detection in high-dimensional multivariate time series. We propose PCA-uCPD, an integrated framework that combines principal component analysis with univariate change-point detection methods. By projecting high-dimensional data onto a lower-dimensional subspace and refining the estimated change points, the proposed method effectively detects structural changes while maintaining computational efficiency and statistical accuracy, making it suitable for complex financial datasets.

The third part investigates high-dimensional quantile regression for integrating information across multiple data platforms. Recognizing that financial data collected from different platforms may exhibit platform-specific scales and noise while sharing common underlying risk factors, we develop multi-platform quantile regression methods that borrow strength across platforms while accommodating heterogeneity. By modeling conditional quantiles instead of conditional means, the proposed methods provide a robust characterization of tail

behavior and improve estimation in high-dimensional settings with limited sample sizes.

Collectively, these contributions establish a unified framework for robust financial time series analysis. The proposed methodologies enhance the modeling of regime dynamics, the detection of structural changes, and the integration of heterogeneous data sources, with broad applications in financial econometrics, risk management, and quantitative market analysis.

Acknowledgements

I am deeply grateful to my supervisor, Professor Yuehua Wu, and to Shanshan Qin for their invaluable mentorship and guidance throughout my research. With their support, I began my PhD studies at York University and received excellent training in statistical research. This dissertation would not have been possible without their continuous encouragement and support.

I would also like to thank Professors Xin Gao and Cindy Fu, who served as members of my PhD thesis examining committee, for their valuable feedback and insightful suggestions, which helped improve this dissertation. I am also grateful to Professor Yishu Zeng, the internal examiner from the Department of Economics at York University, and Professor Grace Yi, the external examiner from Western University, for their insightful comments and thoughtful recommendations.

Finally, I would like to express my heartfelt thanks to my family for their unwavering love and support.

Contents

Abstract **ii**

Acknowledgements **iv**

Contents **v**

1 Introduction **1**

1.1 Background 1

1.2 Literature Review 4

1.3 Outline of the thesis 7

2 Review on Regime switching model **10**

2.1 Introduction 10

2.2 Threshold Models 11

2.3 Hidden Markov Regime-Switching Models 13

2.3.1 The Maximum Likelihood Estimation (MLE) Algorithm 15

2.3.2 The Expectation-Maximization(EM) Algorithm 17

2.3.3 The Markov Chain Monte Carlo (MCMC) Algorithm 18

2.3.4 Other Algorithms 20

2.4 Hidden Semi-Markov State Transition Models 20

2.5 Smooth Transition Models 22

2.6 Some Other Switching Mechanisms 25

2.7 Empirical Examples 25

2.7.1 Example 1: The daily returns of the S&P 500 27

2.7.2 Example 2: The weekly returns of the EURO STOXX 50 31

2.8 Conclusion 35

3 On Robust Estimation of Hidden semi-Markov Regime-switching Model **36**

3.1 Introduction 36

3.2 The formulation of rHSMS . . . 38
3.3 Model estimation and goodness-of-fit . . . 40
3.3.1 EM algorithm . . . 41
3.3.2 Forward-backward algorithm . . . 44
3.3.3 Evaluation metrics for goodness-of-fit . . . 46
3.4 Simulation studies . . . 47
3.4.1 Two-state hidden semi-Markov RS model . . . 48
3.4.2 Three-state Hidden Semi-Markov RS model . . . 53
3.4.3 The impacts of atypical values on the transitions and durations of rHSMS . . . 53
3.4.4 Comparison of rHSMS, HSMM-LP and HSMM-t . . . 56
3.5 Real data applications . . . 61
3.5.1 S&P500 monthly returns . . . 61
3.5.2 HSI monthly returns . . . 64
3.6 Conclusion . . . 68

4 PCA-uCPD: An ensemble method for multiple change-point detection in moderately high-dimensional data 70
4.1 Introduction . . . 70
4.2 Proposed methods . . . 71
4.2.1 PCA . . . 71
4.2.2 Univariate change-point detection (uCPD) methods . . . 73
4.2.3 Proposed PCA-uCPD methods . . . 74
4.3 Theoretical justification . . . 75
4.4 Statistical test on single CPD . . . 78
4.5 Simulation studies on multiple CPD . . . 80
4.5.1 Evaluation Metrics . . . 81
4.5.2 Performance comparison on multivariate data . . . 82
4.5.3 Performance comparison on moderately high-dimensional data . . . 86
4.5.4 Performance comparison of RPCA-CP and 'Geometry' for heavy-tail distributed data . . . 89
4.6 Experiments on Real Data . . . 90
4.6.1 Example 1: Bladder tumor microarray data (ACGH) . . . 90
4.6.2 Example 2: Dow Jones Industrial Average (DJIA) . . . 92
4.7 Conclusion . . . 93

5 High-Dimensional Multivariate Quantile Regression with Multi-Platform Data Integration **95**
5.1 Introduction . . . 95
5.2 Methodology . . . 96
5.2.1 Model setup . . . 96
5.2.2 Parameter estimation . . . 98
5.3 Simulation Study . . . 103
5.3.1 Evaluation metrics . . . 103
5.3.2 Multivariate quantile regression for multi-platform data . . . 104
5.3.3 Unvariate quantile regression for multi-platform data . . . 108
5.4 Real data applications . . . 108
5.4.1 U.S. stock market indices . . . 108
5.4.2 Multiple experiments on liver toxicity study . . . 111
5.5 Conclusion . . . 112

6 Conclusion **114**

Appendix **117**

Bibliography **119**

Chapter 1

Introduction

1.1 Background

Statistical models play a central role in modern finance. Financial econometrics is a specialized field that applies mathematical and statistical methods to analyze and address a wide range of problems in financial markets, including the pricing of assets such as stocks, bonds, interest rates, and exchange rates (Hardy, 2001; Jegadeesh et al., 2019), the valuation of financial derivatives such as options, futures, and swaps (Buffington and Elliott, 2002), portfolio analysis (Markovitz , 1959), risk management (Beder , 1995), and the interpretation of market dynamics (Chkili and Nguyen, 2014). Most financial data are collected as time series and exhibit complex temporal and cross-sectional dependence structures.

Among these data, stock market indices are particularly important because they summarize the behavior of a broad set of assets and serve as key indicators of overall market conditions. Moreover, the prices of many financial derivatives are closely linked to the performance of underlying stocks, and stock indices can represent the overall movements of these stocks. Consequently, statistical modeling and analysis of stock market indices are essential for capturing market dynamics, identifying structural changes, and quantifying financial risk. These challenges have motivated extensive research into the development of new statistical methodologies for forecasting and interpreting financial time series. Stock market indices are characterized by regime changes, structural breaks, heavy-tailed distributions, and heterogeneous behavior across markets and platforms; as such, they naturally give rise to high-dimensional data structures.

One of the most prominent challenges in financial time series modeling is the presence of structural breaks. Structural breaks refer to changes in the underlying data-generating process, often induced by major economic or financial events such as crises or shifts in economic structure. These breaks may be permanent, temporary, or recurrent, with business cycles

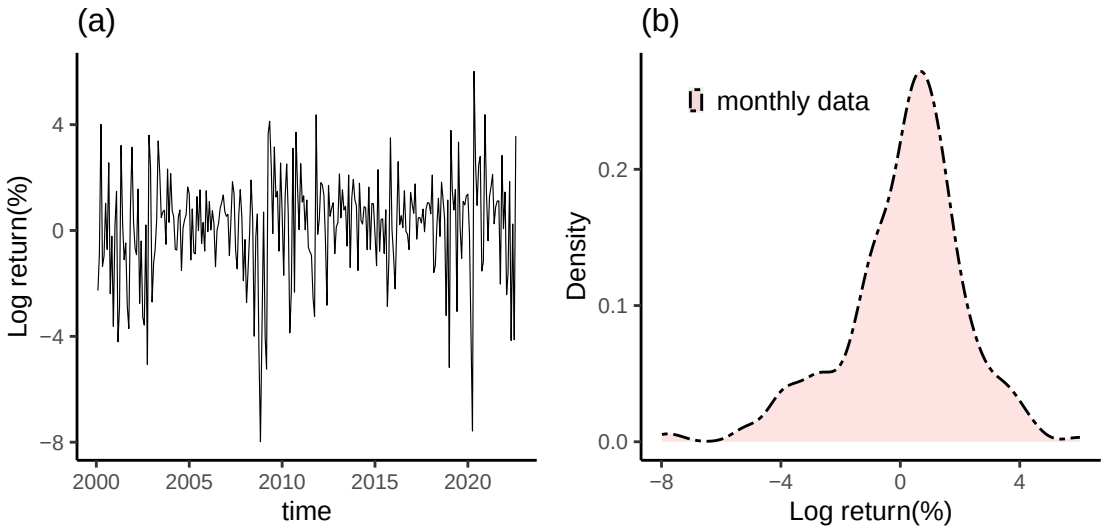

Figure 1.1: Log returns of S&P500, the return $y_t = (\log P_t - \log P_{t-1}) \times 100$, where P_t is the closing price in week or month t. (a) monthly returns (b) density of monthly returns.

providing a canonical example of the latter. Stock market indices commonly exhibit such structural breaks, as extreme events are reflected through abnormally large price movements and heightened volatility. Figure 1.1, panel (a), presents the monthly returns of the S&P 500 from 2000 to 2022 and illustrates these stylized facts. Consistent with the existing literature, index volatility is generally negatively correlated with returns, reflecting periods of market stress during which prices tend to decline. Major events over the past two decades, including the 2008 global financial crisis and the COVID-19 pandemic, have therefore resulted in pronounced volatility in asset returns.

Motivated by these empirical challenges, a substantial body of research has focused on developing models that explicitly account for structural breaks. Regime-switching (RS) models are widely used in finance, allowing model parameters to take different values across a finite number of latent regimes or states. This switching mechanism provides an effective way to capture the dynamic behavior of macroeconomic and financial variables observed over time. RS models have been employed to study business cycle phases, "bull" and "bear" markets in equity returns, high- and low-volatility regimes in asset prices, and periods of inflation and deflation (Meyers, 2011; Zheng et al., 2021). Traditional hidden Markov regime-switching (HMS) models assume that regime durations follow a geometric distribution and that the current state depends only on the most recent past state. Although these assumptions offer mathematical tractability, they often limit the applicability of HMS models to real financial data, where regime durations vary substantially and returns frequently exhibit heavy tails and extreme values (Bulla and Bulla, 2006; Maruotti et al., 2019).

Hidden semi-Markov regime-switching (HSMS) models provide a more flexible alternative by allowing arbitrary sojourn (state duration) distributions while retaining the core regime-switching structure. This additional flexibility enables HSMS models to better capture the temporal persistence of market regimes, such as expansions, recessions, and periods of high or low volatility. Recent advances in robust estimation techniques further enhance the HSMS framework by improving its ability to handle heavy-tailed returns and outliers, making it particularly well suited for modeling financial time series characterized by extreme events and irregular behavior.

Another important line of research related to structural breaks is change-point detection (CPD). Change-point detection provides a complementary and widely used framework for identifying structural changes in time series data. CPD aims to detect and localize distributional changes in a sequence of observations ordered by time, space, or other indexing variables (Li and Zhang, 2024). When one or more change-points (CPs) are present but ignored, statistical inference or estimation based on the full data sequence may become biased or misleading. Consequently, CPD has attracted sustained research interest due to its importance in identifying structural changes across a wide range of application domains.

While classical CPD methods were primarily developed for univariate or low-dimensional settings, many modern applications involve high-dimensional data structures in which the number of variables is comparable to or exceeds the available sample size. Such settings are common in financial markets with many correlated assets, as well as in network data, image and text streams, transportation systems, and high-resolution neurophysiological recordings (Chen and Chu, 2023). In these environments, observations often exhibit strong dependence, heterogeneity, and non-Gaussian behavior.

These features pose substantial challenges for CPD in high-dimensional settings. Existing methods frequently focus on detecting changes in specific distributional components, such as the mean or covariance, and often rely on sparsity assumptions or Gaussian approximations. Moreover, relatively few CPD procedures are explicitly designed to accommodate heavy-tailed distributions or outliers, which are pervasive in financial data. As a result, effective CPD for moderately high-dimensional financial time series remains a challenging problem, particularly when cross-sectional dependence and extreme events play a central role.

Beyond temporal and distributional complexities, stock market data also exhibit substantial heterogeneity across markets and platforms. Indices such as the S&P 500, NASDAQ-100, and Dow Jones Industrial Average operate at different scales and respond differently to market conditions. Furthermore, stock market indices are inherently high-dimensional, as they are constructed from hundreds of constituent assets; for example, the S&P 500 comprises 500 individual stocks. When combined with limited sample sizes, this structure often leads

to high-dimensional modeling scenarios where $p \gg n$. In the era of big data, financial information is increasingly collected from multiple sources or platforms, making data integration a significant methodological challenge. Effective data integration seeks to exploit shared information across platforms while accommodating platform-specific heterogeneity, thereby improving statistical efficiency and inferential accuracy.

Finally, the complex features of financial data—including structural breaks, heavy tails, high dimensionality, cross-platform heterogeneity, and extreme events—highlight the limitations of traditional mean-based modeling approaches. Methods that focus solely on conditional means may fail to adequately capture tail behavior, regime persistence, and heterogeneous effects that are central to financial risk analysis. These considerations motivate the development of flexible and robust statistical frameworks, such as HSMS models with robust estimation procedures, that are capable of accommodating the full complexity of financial time series and providing more accurate risk assessment and predictive modeling.

1.2 Literature Review

As a result, RS models have found widespread application across a broad range of financial problems. Examples include modeling and identifying the phases of the business cycle (Chauvet, 1998; Hamilton, 1989; Hamilton and Lin, 1996; Contreras-Reyes, 2023), regime shifts in stock or equity returns (Hardy, 2001; Hamilton and Susmel, 1994; Dueker, 1997; Choi and Hammoudeh, 2010; Zheng et al., 2021; Guo et al., 2011a,b), or interest and exchange rates (Ang and Bekaert, 2002; Garcia and Perron, 1996; Chkili and Nguyen, 2014). Other applications include derivative pricing (Buffington and Elliott, 2002) and portfolio optimization (Yin and Zhou, 2004), among others. A comprehensive review of RS models is provided in Chapter 2.

CPD has been successfully applied in numerous fields, including finance, medicine, bioinformatics, business analytics, and signal processing. In medical and biological studies, CPD has been used to analyze DNA sequences (Vert and Bleakley, 2010; Muggeo and Adelfio, 2011; Frick et al., 2014; Zou et al., 2014) and physiological signals such as heart rate variability (Haynes et al., 2017). In financial time series, CPD methods have been employed to identify market regime shifts and structural changes in stock indices, exchange rates, and cryptocurrency prices (Lavielle, 1999; Wen et al., 2018; Li and Zhang, 2024). Beyond finance, CPD has been applied to credit card fraud detection (Bolton and Hand, 2002), quality control in industrial processes (Lai, 1995), the assessment of intervention effects during the COVID-19 pandemic (Dehning et al., 2020), speech segmentation (Angelosante and Giannakis, 2012), image stream analysis (Kim et al., 2009), and network traffic monitoring for cyber-attack

detection (Lévy-Leduc and Roueff, 2009).

Change-point detection (CPD) has been extensively studied for univariate and multivariate (low-dimensional) data (Chen and Chu, 2023). A wealth of literature has explored both single (Csörgö and Horváth, 1997; Chen et al., 2000) and multiple *cps* detection in univariate data (Inclan and Tiao, 1994; Ombao et al., 2001; Killick et al., 2012; Jin et al., 2013; Aminikhanghahi and Cool, 2017; Fearnhead and Rigaill , 2020), and references therein. Moreover, CPD for multivariate data has received increasing attention. The methods encompass a wide range of approaches, including classical techniques such as multivariate CUSUM (Crosier, 1988; Qiu and Hawkins, 2001, 2003; Li et al., 2013), likelihood ratio tests (Srivastava and Worsley, 1986; James et al., 1992; Chen et al., 2000), and Bayesian methods (Chen et al., 2000; Corradin et al., 2022). More advanced models account for heteroscedasticity (Górecki et al., 2018), spatial dependencies (Moradi et al., 2023), and deep learning architectures (Du and Duan, 2022). These methods address various challenges inherent in multivariate data analysis, such as detecting distributional changes in mean and covariance structures, dealing with heterogeneity and spatial dependence, and capturing complex patterns of changes.

However, many applications of CPD involve moderately high-dimensional and complex data, including network evolution, image analysis, sequence or text streams, multi-sensor monitoring systems, transportation data for disruption detection, and neuropixel recordings (Chen and Chu, 2023). Numerous methods have been developed for change-point estimation in high-dimensional settings. For example, Xie et al. (2012) proposed an online CPD method for high-dimensional time series with missing values, while Cho and Fryzlewicz (2015) introduced a sparsified binary segmentation (SBS) algorithm that aggregates CUSUM statistics. The two-stage "Inspect" procedure of Wang and Samworth (2018) focuses on sparse mean shifts, whereas Avanesov and Buzun (2018) studied change-point detection in covariance structures of high-dimensional processes. Other approaches include HDcp for dependent data (Li et al., 2019), distance-based CPD methods (Li, 2020), geometric mappings for joint mean and covariance changes (Grundy et al., 2020), sparse mean change detection (Liu et al., 2021), and aggregation-based homogeneity tests (Pilliat et al., 2023).

Owing to its broad applicability, CPD has received substantial research attention across numerous fields, including business, finance, medicine, agriculture, human activity analysis, and bioinformatics. In the medical domain, for instance, CPD plays a vital role in examining DNA sequences (Vert and Bleakley, 2010; Muggeo and Adelfio, 2011; Frick et al., 2014; Zou et al., 2014) and analyzing physiological signals such as heart rate (Haynes et al., 2017). Its utility is equally evident in financial time series analysis, where it is employed to identify structural changes in stock indices, U.S. dollar exchange rates, and Bitcoin prices (Lavielle, 1999; Wen et al., 2018; Li and Zhang, 2024). Beyond these areas, CPD has been applied

to detect changes in human behavior and organizational behavior, ranging from credit card fraud detection (Bolton and Hand, 2002) and business quality control (Lai, 1995) to identifying change points in the spread of the COVID-19 pandemic to assess the effectiveness of public health interventions (Dehning et al., 2020). Furthermore, CPD has been used to recognize boundaries between silence, sentences, words, and noise in speech recognition (Angelosante and Giannakis, 2012), to detect abrupt events in image streams (Kim et al., 2009), and to identify internet attacks in network traffic data (Lévy-Leduc and Roueff, 2009).

In addition to the challenges of structural breaks and change-point detection, a growing concern in modern financial and economic analysis arises from the integration of data collected across multiple platforms or sources. With the increasing availability of large-scale datasets generated by different markets, experiments, or institutions, data integration has emerged as an important methodological topic. A substantial body of research has focused on developing statistical frameworks that exploit shared information across related datasets while allowing for platform-specific heterogeneity. Early and influential work includes Ando et al. (2005), who proposed a framework for joint estimation across related regression models, followed by numerous extensions in high-dimensional settings (Gao and Carroll., 2017; Gao and Zhong, 2019; Zhang and Yang., 2021; Maity et al., 2019; Dai et al., 2023). These approaches have found applications in a wide range of fields, including biostatistics and finance, where multi-market or multi-experiment data structures are common.

Despite these advances, the majority of existing data integration methods have focused on conditional mean regression. While mean-based approaches are computationally convenient, they often rely on restrictive distributional assumptions and can be highly sensitive to outliers, heavy-tailed errors, and heterogeneous effects—features that are pervasive in financial data. These limitations have motivated increasing interest in quantile regression as a robust and flexible alternative. Since its introduction by Koenker and Bassett (1978), quantile regression has become a widely used tool for modeling conditional distributions beyond the mean, offering robustness to misspecification and the ability to capture heterogeneous covariate effects across different parts of the response distribution. Comprehensive reviews of univariate quantile regression and its extensions are provided by Koenker (2017).

Recent methodological developments have further expanded the scope of quantile regression to accommodate complex data structures. For example, Romano et al. (2019) proposed a conformalized quantile regression approach that adapts to heteroscedasticity, while Chernozhukov et al. (2022) developed fast algorithms for estimating a large number of quantile regressions. In financial applications, quantile regression has been employed to study tail risk, volatility dynamics, and regime-dependent behavior (Maruotti et al., 2021; Petrella and Raponi, 2019). However, much of this literature remains focused on univariate responses.

In many modern applications, including finance and multi-platform studies, responses are inherently multivariate and often exhibit strong dependence structures. Ignoring these dependencies may result in inefficient estimation and an incomplete understanding of joint behavior. Multivariate quantile regression has therefore received increasing attention in recent years (Petrella and Raponi, 2019; Carlier et al., 2022; Dai et al., 2023; Tian et al., 2024). A central challenge in this setting is the absence of a natural ordering for multivariate responses, which complicates the definition and estimation of multivariate quantiles (Chakraborty., 2003; Hallin et al., 2010; Kong and Mizera., 2012). Existing approaches can be broadly classified into likelihood-based methods, often relying on multivariate asymmetric Laplace distributions (Petrella and Raponi, 2019), and loss-based methods that directly minimize empirical quantile loss functions (Chen and Chen, 2024). While likelihood-based approaches provide a probabilistic interpretation, they can become computationally prohibitive in high-dimensional settings, whereas loss-based approaches offer greater flexibility but introduce non-smooth optimization challenges.

More recently, quantile regression has been extended to multi-platform integration settings. For instance, Dai et al. (2023) proposed a univariate quantile regression framework for identifying predictors that exert common effects across multiple platforms. Nevertheless, existing methods largely focus on univariate responses and do not explicitly model dependence among multiple outcomes. As a result, there remains a gap in the literature for multivariate quantile regression models that simultaneously accommodate high-dimensional predictors, cross-platform heterogeneity, and dependence across multiple responses.

1.3 Outline of the thesis

Regime-switching models provide a flexible and powerful framework for modeling time series that alternate between distinct dynamic behaviors. Chapter 2 presents a comprehensive review of regime-switching models with various switching mechanisms, including threshold models, hidden Markov regime-switching models, hidden semi-Markov regime-switching models, and smooth transition models. For each class of models, we summarize the underlying framework, representative specifications, and commonly used estimation methods. This review lays the foundation for and motivates the robust estimation framework developed in Chapter 3.

Building on this review, Chapter 3 focuses on hidden semi-Markov regime-switching (HSMS) models, which generalize hidden Markov regime-switching (HMS) models that have attracted increasing attention in recent years. HSMS models address a key limitation of HMS models by allowing more flexible state duration distributions, thereby relaxing the

strong Markovian assumptions imposed on regime durations. To further accommodate the heavy-tailed nature of stock index returns, we propose a robust estimation method for HSMS models. Specifically, we incorporate a Huber ρ-function from M-estimation into the emission distribution of a K-state HSMS model to mitigate the influence of heavy-tailed errors and outliers. This constitutes the main methodological contribution of this chapter, referred to as the robust HSMS (rHSMS) model. The proposed method is developed within a likelihood-based estimation framework.

Chapter 4 continues the investigation of structural changes in stock data, but from the perspective of change-point detection (CPD). Stock index data typically form complex multivariate time series characterized by moderate to high dimensionality, strong cross-correlations, and the presence of outliers or heavy-tailed distributions. Detecting change points in such settings remains challenging for existing CPD methods, which are often designed for univariate series. To address this challenge, we propose an integrated change-point detection method, termed PCA-uCPD. This approach leverages principal component analysis (PCA) to project the original multivariate series onto a set of uncorrelated principal components (PCs). Existing univariate CPD methods are then applied to the transformed PCs, followed by a refinement procedure to recover the final change-point estimates for the original data. The proposed method features a flexible architecture, making it well suited for complex, high-dimensional data structures.

Finally, Chapter 5 addresses structural heterogeneity and high dimensionality from the perspective of multi-platform data integration. In modern financial and economic applications, data are increasingly collected across multiple platforms or markets, where responses may be multivariate, strongly correlated, and influenced by a large number of predictors. Existing data integration methods predominantly focus on mean regression and are often sensitive to outliers and distributional misspecification. To overcome these limitations, we develop a multivariate quantile regression framework for integrating high-dimensional data across multiple platforms. The proposed method extends recent advances in multivariate quantile regression by explicitly modeling vector-valued responses within a unified loss-based optimization framework, thereby capturing dependence across response dimensions. To exploit shared information across platforms while allowing for platform-specific heterogeneity, we incorporate an ℓ_1/ℓ_2 cross-platform group penalty that encourages the identification of predictors with common effects across platforms. The resulting framework generalizes several existing models, including univariate multi-platform quantile regression, multivariate quantile regression, and high-dimensional penalized quantile regression. This chapter also develops an efficient smoothing-based optimization algorithm, establishes its convergence properties, and demonstrates the effectiveness of the proposed approach through simulation

studies and real data applications.

Chapter 6 concludes the thesis by summarizing the main findings and highlighting the key methodological contributions.

Chapter 2

Review on Regime switching model

2.1 Introduction

Tong (1978) first proposed the threshold autoregression (TAR) model in 1978 to capture the dynamics of financial time series data by switching between different regimes. The general form of Tong's TAR model is given by:

$$X_t = \phi_0^{(S_t)} + \sum_{i=1}^{p} \phi_i^{(S_t)} + \theta^{(S_t)} a_t, \tag{2.1.1}$$

where a_t is a sequence of independent and identically distributed (IID) random variables with mean 0 and variance σ^2, and S_t is an indicator variable representing the regimes or states. The value of S_t determined by a threshold variable s_t, can be either observable or hidden.

Threshold models are a class of non-linear process designed to provide local linear approximations to underlying nonlinear relationships rather than to describe regime-switching behavior. To capture regime-switching behavior like business cycles, Hamilton (1989) extended the regime-switching model by introducing a regime-switching mechanism based on hidden Markov chains (HMCs) with an unobservable stochastic process S_t, called hidden Markov switching (HMS). In this framework, the transitions between regimes are not directly observable; instead, probabilistic inference is used to determine whether and when a regime change is likely to occur based on the time series. In Hamilton (1989) and Zakoian (1994), an autoregressive conditional heteroskedasticity (ARCH) type models were applied in the regimes.

Ferguson (1980) extended the hidden Markov chain (HMC) model to the hidden semi-Markov state transition (HSMS) model, introducing more flexible sojourn distributions $d_k(u)$

that represent the probability of spending time u in state k. These sojourn distributions $d_k(u)$ were generalized from the geometric distributions used in HMC to a wider range of discrete and continuous distributions with positive support, such as Poisson, negative binomial, logarithmic, nonparametric, gamma, Weibull, and lognormal distributions (Amini et al., 2022).

Teräsvirta (1994) proposed another regime-switching model, the smooth transition autoregressive (STAR) time series model. In this model, regime changes occur gradually rather than abruptly. The smooth transition function $G(s_t; \gamma, c)$ determines the degree of influence exerted by each regime. $G(s_t; \gamma, c)$ is continuous and bounded between 0 and 1. The transition variable s_t is usually considered to be a lagged endogenous variable and is defined as $s_t = y_{t-d}$, where $d > 0$ is an integer. The parameter γ controls the smoothness of the transition, while c represents the threshold between two regimes. A common example of $G(s_t; \gamma, c)$ is a first-order logistic function:

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp\{-\gamma(s_t - c)\}}, \gamma > 0 \tag{2.1.2}$$

which defines the logistic STAR (LSTAR) model.

This Chapter 2 reviews various regime-switching models with different switching mechanisms. Threshold models that depend on observable variables are discussed in Section 2.2, hidden Markov regime-switching models are covered in Section 2.3, hidden semi-Markov models in Section 2.4, smooth transition models in Section 2.5, some new switching mechanisms will be introduced in Section 2.6.

2.2 Threshold Models

The autoregressive (AR) model is a popular choice for modeling the time series y_t, $t = 1, 2, \ldots, T$, and its formula is as follows:

$$y_t = \alpha + \sum_{j=1}^{k} \phi_j y_{t-j} + \epsilon_t,$$

where k is the order of the model, α and ϕ_j, $j = 1, 2, \ldots, k$, are parameters, and $\epsilon_t \sim N(0, \sigma^2)$ represents white noise.

The regime-switching version of the autoregressive model of order k has the following formulation

$$y_t = \alpha^{(S_t)} + \sum_{j=1}^{k} \phi_j^{(S_t)} y_{t-j} + \epsilon_t, \tag{2.2.1}$$

where $\epsilon_t \sim N(0, (\sigma^{(S_t)})^2)$. In addition to continuous-valued models like the autoregressive model, discrete-valued models can also be used to model a given regime S_t. Many regime-switching discrete-valued models have been proposed, such as those in Liu et al. (2019, 2020); Yang et al (2022, 2023). The parameters of the model depend on $S_t = i$, where $i = 1, 2, \ldots, K$, which is an indicator variable representing the regime or state at time t. The parameters vary between K different regimes. The threshold model assumes that S_t is determined by a deterministic function of the observed variable s_t. Different types of threshold models may have different deterministic functions of the observed variable s_t or different model specifications within regimes.

Tong and Lim (1980) introduced the self-exciting TAR (SETAR) model. In this model, s_t is the lagged endogenous variable y_{t-d}. For the SETAR model, S_t is determined as follows:

$$S_t = \begin{cases} 1, & y_{t-d} < \tau_1 \\ 2, & \tau_1 \leq y_{t-d} \\ \vdots & \\ K & \tau_{K-1} \leq y_{t-d} \end{cases}$$

The parameters d and τ_i in the SETAR model represent the "delay" parameter and the thresholds, respectively, both of which are unobservable. S_t determined by d and τ_i is unobservable and needs to be estimated from the data.

The threshold variable s_t can also be an exogenous time series X_t (Chen, 1998), a linear combination of lagged or exogenous variables (Chen et al., 2003; Chen and So, 2006; Gerlach et al., 2006), or a nonlinear combination (Chen, 1995; Wu and Chen, 2007).

There are different types of threshold models for the financial time series. Chen et al. (2011b) provided a comprehensive summary of threshold models categorized into three classes: nonlinearity in mean, nonlinearity in volatility, and 'double' threshold dynamics (in both mean and volatility). Nonlinearity models in mean include the threshold autoregressive (TAR) model (Petruccelli and Woolford, 1984; Chen and Tsay, 1991; Tong, 2012; Chan et al., 1985; Chan and Tong, 1985), the threshold moving-average (TMA) model (Ling and Tong, 2005; Ling et al., 2007), and the threshold autoregressive moving-average (TARMA) model (Brockwell et al., 1992; Ling, 1999). Within each regime, the autoregressive, moving-average (MA), and autoregressive moving-average (ARMA) models are employed.

Nonlinearities in volatility models include GJR-GARCH (Glosten et al., 1993), ST-GARCH (Hagerud, 1997; Gerlach and Chen, 2008), DTGARCH (Chen and So, 2006), Threshold GJR-GARCH (Chen et al., 2005), TARR (Chen et al., 2008b), TGARCH(Zakoian, 1994) and DTGARCH (Chen et al., 2010). These models are extensions of one of the most

popular volatility models, the GARCH-type model.

To account for nonlinearity in the mean and volatility models, Li and Li (1996) proposed a DTARCH model to handle asymmetric mean responses in the presence of nonlinearities in both mean and volatility. In addition, Brooks (2001) extended the DTARCH model to a double threshold GARCH (DT-GARCH) model to assess asymmetries in market indices, stock returns, and exchange rates. Chen and So (2006) allowed s_t to be a weighted average of important auxiliary variables, defined as:

$$s_t = \sum_{i=1}^{l} w_i s_{it}, \quad 0 \leq w_i \leq 1, \quad \sum_{i=1}^{l} w_i = 1. \tag{2.2.2}$$

Here, s_{it} can represent any function of endogenous variables $y_t, \ldots, y_1$ and exogenous variables. Therefore, information about both endogenous variables and exogenous variables can be used to govern the switching mechanism. Chen and So (2006) found that domestic returns are the main factor determining the threshold compared with international returns. Chen et al. (2010, 2011a) proposed a three-regime DTGARCH model to capture the asymmetry in both mean and volatility. Audrino and Bühlmann (2001) proposed a tree-based GARCH model. In addition to GARCH-type models, So et al. (2002) proposed threshold stochastic volatility (THSV) models. Chen et al. (2008a) introduced a generalization of the THSV model, incorporating error innovations that follow a standardized t-distribution.

The challenge in estimating threshold models is that the likelihood function is not differentiable with respect to the thresholds and is often multimodal. Various approaches have been used to estimate the parameters of threshold models. Two-stage methods are commonly used, as described in Tong and Lim (1980) and Ghaddar and Tong (1981). These methods specify delay parameters and thresholds and then use maximum likelihood estimation (MLE) to estimate the parameters of the model. Researchers are increasingly using Bayesian methods via Markov chain Monte Carlo (MCMC) methods to estimate more complex models, benefiting from improvements in computational speed. Subsequently, all parameters, including the delay parameters and thresholds, can be estimated simultaneously.

2.3 Hidden Markov Regime-Switching Models

Hamilton (1989) and Hamilton (1994) introduced the hidden Markov regime-switching model (also called hidden Markov switching model, abbreviated by HMS model) to analyze financial time series and Hamilton (1994) incorporated the ARCH model within regimes. The regimes are unobservable and are modeled using probabilistic reasoning with a Markov

chains, rather than deterministic functions. In contrast to the threshold models, which are used to model nonlinearities in data, the Markov regime-switching model is used to capture regime-switching behaviors, such as business cycles. Since then, hidden Markov regime-switching models have been widely used in financial modeling. The HMS model can also be viewed as a type of hidden Markov model (HMM) (Baum and Eagon, 1967) with continuous observations. HMMs are statistical models designed using a Markov process with hidden states. These models are widely used in speech recognition, bioinformatics, human activity recognition, and other fields.

In the HMS model, S_t represents the latent regime or state at time t. The observations y_t, $t = 1, \ldots, T$; depend on the emission probabilities $f(y_t \mid S_t = k, \Theta_k)$, which depends on the hidden state S_t. Here, $f(y_t \mid S_t = k, \Theta_k)$ is the conditional distribution of the observations given the hidden state S_t, where Θ_k denotes the model parameters under the regime $S_t = k$, $k = 1, 2, \ldots, K$. The initial probability $\pi_k = Pr(S_1 = k)$ represents the probability of starting from state k, and the transition probability p_{ij} represents the probability of transitioning from state i to state j.

The two-state HMS model (where $K = 2$) remains a popular specification in financial applications. The two states can usually correspond to depreciation and appreciation regimes. The three-state HMS model ($K = 3$) is also used in financial applications. The three states can correspond to three different growth rate phases: recession regime, medium growth regime, and high growth regime, as described in Medhioub (2015); Ayodeji (2016).

In contrast to the threshold model, the hidden state in the Markov regime-switching model is unobservable and is modeled by probabilistic reasoning with a Markov chain rather than a deterministic function. The Markov regime-switching chain is defined by the transition matrix $\boldsymbol{P} = \{p_{ij}\}$. For the two-state model, the transition matrix $\boldsymbol{P}_{2\times 2}$ is given by:

$$\boldsymbol{P}_{2\times 2} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

where

$$p_{ij} = Pr(S_t = j \mid S_{t-1} = i, S_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots) = Pr(S_t = j \mid S_{t-1} = i).$$

This reflects the Markov property, in contrast to the behavior of the threshold model. In the Markov switching model, regimes evolve exogenously with respect to the time series, whereas the threshold model shows that the regimes evolve endogenously based on the behavior of the observable variables. Markov switching models do not tie regime-switching to the behavior of particular observations.

In finance, there are several popular examples of different types of Markov switching models. Hamilton (1994) and Gray (1996) discussed the Markov switching GARCH (MS-GARCH) model to capture the volatility dynamics of financial time series. Krolzig (1998) developed the Markov switching vector autoregression (MS-VAR) model using the traditional vector autoregression (VAR) model as the emission distribution. So et al. (1998) combined the stochastic volatility model with Markov regime switching to model time-varying volatility. Hardy (2001) considered a simplified Markov regime-switching model for analyzing complex monthly stock price return data using log-normal distributions within regimes. Shi and Ho (2015) first modeled long memory and regime-switching via the autoregressive fractional moving average (ARFIMA) and Markov regime-switching models. Ma et al. (2017) introduced Markov regime switching into the heterogeneous autoregressive model of realized volatility (HAR-RV type) models and evaluated the forecasting performance.

2.3.1 The Maximum Likelihood Estimation (MLE) Algorithm

Among the estimation methods of the HMS model, likelihood method is one of the main approaches for estimating the model parameters. For example, Hamilton (1989) used maximum likelihood estimation (MLE) to estimate the parameters in the HMS model.

Denote by Θ the vector of all the parameters in an HMS model. For a given Θ, the log-likelihood function is given by:

$$L(\Theta) = \sum_{t=1}^{T} \log f(y_t \mid \Omega_t; \Theta),$$

where $f(\cdot)$ denotes the density function of y_t, and $\Omega_t = (y_t, y_{t-1}, \ldots, y_0)$ represents the observations up to time t.

For the simplest case with $K = 2$ regimes, where $S_t = 1$ or 2, $Pr(S_t = i \mid \Omega_t; \Theta)$ is given as follows:

$$\begin{aligned} Pr(S_t = i \mid \Omega_t; \Theta) &= \sum_{j=1}^{2} Pr(S_t = i \mid S_{t-1} = j, \Omega_{t-1}; \Theta) Pr(S_{t-1} = j \mid \Omega_{t-1}; \Theta) \\ &= \sum_{j=1}^{2} p_{ji} Pr(S_{t-1} = j \mid \Omega_{t-1}; \Theta), \end{aligned} \tag{2.3.1}$$

where p_{ji} denotes the transition probabilities between regimes. The likelihood of the obser-

vations, $f(y_t \mid \Omega_t; \Theta)$, is then computed as:

$$\begin{aligned} f(y_t \mid \Omega_t; \Theta) &= \sum_{i=1}^{2} f(y_t \mid S_t = i, \Omega_t; \Theta) Pr(S_t = i \mid \Omega_t; \Theta) \\ &= \sum_{i=1}^{2} f(y_t \mid S_t = i, \Omega_t; \Theta) \sum_{j=1}^{2} p_{ji} Pr(S_{t-1} = j \mid \Omega_{t-1}; \Theta) \\ &= f(y_t \mid S_t = 1, \Omega_t; \Theta) p_{11} Pr(S_{t-1} = 1 \mid \Omega_{t-1}; \Theta) \\ &\quad + f(y_t \mid S_t = 1, \Omega_t; \Theta) p_{21} Pr(S_{t-1} = 2 \mid \Omega_{t-1}; \Theta) \\ &\quad + f(y_t \mid S_t = 2, \Omega_t; \Theta) p_{12} Pr(S_{t-1} = 1 \mid \Omega_{t-1}; \Theta) \\ &\quad + f(y_t \mid S_t = 2, \Omega_t; \Theta) p_{22} Pr(S_{t-1} = 2 \mid \Omega_{t-1}; \Theta). \end{aligned} \tag{2.3.2}$$

If we use the regime-switching AR(1) model (Equation (2.2.1) with $k = 1$) as the state-dependent model, the conditional likelihood function in Equation (2.3.2) is expressed as:

$$f(y_t \mid S_t = i, \Omega_t; \Theta) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ -\frac{(y_t - \alpha_i - \phi_{1,i} y_{t-1})^2}{2\sigma_i^2} \right\}, \tag{2.3.3}$$

where $\Theta = (\alpha_1, \alpha_2, \phi_{1,1}, \phi_{1,2}, \sigma_1, \sigma_2)$. By Bayes' rule, the posterior probability of regime i at time t is:

$$Pr(S_t = i \mid \Omega_t; \Theta) = \frac{f(y_t \mid S_t = i, \Omega_{t-1}; \Theta) Pr(S_t = i \mid \Omega_{t-1})}{f(y_t \mid \Omega_{t-1}; \Theta)}. \tag{2.3.4}$$

The right-hand side of (2.3.4) is obtained using Equations (2.3.1) and (2.3.3).

To initialize the model, we calculate the initial probabilities as follows:

$$\begin{aligned} Pr(S_1 = 1 \mid y_1, \Theta) &= \frac{f(S_1 = 1, y_1 \mid \Theta)}{f(y_1 \mid \Omega_0, \Theta)} \\ &= \frac{f(S_1 = 1, y_1 \mid \Theta)}{\sum_{i=1}^{2} f(S_1 = i, y_1 \mid \Theta)} \\ &= \frac{\pi_1 f(y_1 \mid S_1 = 1, \Omega_0; \Theta)}{\sum_{i=1}^{2} \pi_i f(y_1 \mid S_1 = i, \Omega_0; \Theta)}, \end{aligned}$$

where π_1 and π_2 are the initial probabilities. These probabilities can be derived by solving

the following stationary condition:

$$\boldsymbol{\pi P} = \boldsymbol{\pi},$$
$$\begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix}, \tag{2.3.5}$$

where $\boldsymbol{P}$ cannot be the identity matrix. Otherwise, regime switching would not occur, and they will depend solely on the initial probabilities, even in the presence of stationary distributions. Solving Equation (2.3.5) yields:

$$\pi_1 = \frac{p_{21}}{p_{12} + p_{21}}, \quad \pi_2 = 1 - \pi_1 = \frac{p_{12}}{p_{12} + p_{21}}. \tag{2.3.6}$$

Using these initial probabilities, we can recursively update $Pr(S_t = i \mid \Omega_t; \Theta)$ via Equation (2.3.4) if $Pr(S_{t-1} = i \mid \Omega_{t-1}; \Theta)$,which follows an AR(1) model here, is known. Finally, the likelihood function $L(\Theta)$ can be constructed recursively, allowing the parameters to be estimated by maximizing the likelihood function.

2.3.2 The Expectation-Maximization(EM) Algorithm

In some cases, directly maximizing the likelihood of the observed data can be challenging. For example, Augustyniak (2014) applied the EM algorithm and an importance sampling method to improve the performance of MLE in the Markov-switching GARCH model. The EM algorithm provides an alternative technique to obtain the MLE of the likelihood of the observed data through an iterative process without directly calculating the likelihood function.

The EM algorithm consists of two steps, the E-step and the M-step, as described below:

1. E-step: Compute the expected complete data log-likelihood to obtain the Q-function, $Q(\boldsymbol{\Phi} \mid \boldsymbol{\Phi}^{(\ell)})$, based on the observed data Ω_T and the current estimate of the parameters, $\boldsymbol{\Phi}^{(\ell)}$, at the ℓ-th iteration:

$$\begin{aligned} Q(\boldsymbol{\Phi} \mid \boldsymbol{\Phi}^{(\ell)}) &= E_{\boldsymbol{\Phi}^{(\ell)}} \left[\log L(\boldsymbol{\Phi}) \mid \Omega_T \right] \\ &= \sum_{j=1}^{K} \sum_{t=1}^{T} \log f(y_t \mid S_t = j, \Omega_t; \boldsymbol{\Phi}^{(\ell)}) Pr(S_t = j \mid \Omega_t; \boldsymbol{\Phi}^{(\ell)}). \end{aligned}$$

   If we use the regime-switching AR(1) model (Equation (3) with k = 1) as the state-dependent model, $f(y_t \mid S_t = j, \Omega_t; \boldsymbol{\Phi}^{(\ell)})$ is same as the conditional likelihood function in Equation (2.3.3).

2. M-step: Maximize the Q-function with respect to $\boldsymbol{\Phi}$, i.e.,

$$\hat{\boldsymbol{\Phi}} = \arg\max_{\boldsymbol{\Phi}} Q(\boldsymbol{\Phi} \mid \boldsymbol{\Phi}^{(\ell)}).$$

Here, $\boldsymbol{\Phi} = \{\Theta, \boldsymbol{\pi}, \boldsymbol{P}\}$ denotes the set of model parameters. The above two steps are iterated until convergence, at which point the parameter estimates $\hat{\boldsymbol{\Phi}}$ stabilize.

2.3.3 The Markov Chain Monte Carlo (MCMC) Algorithm

The Bayesian Markov chain Monte Carlo (MCMC) algorithm is a commonly used approach for HMS estimation. For instance, Amisano and Fagan (2013) estimated parameters via the MCMC algorithm because their model exhibited a highly irregular likelihood surface, which makes it challenging to estimate parameters via maximum likelihood estimation. A joint posterior distribution for the parameters and latent variables is obtained by the Bayesian approach. This is done through Gibbs sampling-based posterior simulation. Kim and Nelson (1999) described in detail in Chapter 9 how to use the Gibbs sampling method in the HMS model. Within the HMS framework, the model's unknown parameters Θ are estimated first. Then, the latent regime S_t conditional on parameters Θ is estimated. In the Bayesian analysis, both parameters Θ and latent regime S_t are treated as random variables. The key to the Bayesian approach is that the unknown parameters Θ are treated as random variables. The joint posterior density is derived as follows:

$$\begin{aligned} f(\tilde{S}_T, \Theta, \mathcal{P}, \mid \Omega_T) &= f(\Theta, \mathcal{P} \mid \Omega_T, \tilde{S}_T) Pr(\tilde{S}_T, \mid \Omega_T) \\ &= f(\Theta \mid \Omega_T, \tilde{S}_T) f(\mathcal{P} \mid \Omega_T, \tilde{S}_T) Pr(\tilde{S}_T, \mid \Omega_T) \\ &= f(\Theta \mid \Omega_T, \tilde{S}_T) f(\mathcal{P} \mid \tilde{S}_T) Pr(\tilde{S}_T, \mid \Omega_T), \end{aligned} \tag{2.3.7}$$

where $\tilde{S}_T = (S_1, S_2, \ldots, S_T)^\top, \mathcal{P} = \{p_{11}, p_{22}\}$. Equation (2.3.7) assumes that the transition probabilities in $\mathcal{P}$ given $\tilde{S}_T$ are independent of other parameters Θ and the observed data Ω_T so that $f(\mathcal{P} \mid \Omega_T, \tilde{S}_T) = f(\mathcal{P} \mid \tilde{S}_T)$. We then apply the Gibbs sampling as follows:

1. Generate S_t from $f(S_t \mid \tilde{S}_{\neq t}, \Theta, \mathcal{P}, \Omega_T), t = 1, 2, \ldots, T$, where
   $\tilde{S}_{\neq t} = (S_1, \ldots, S_{t-1}, S_{t+1}, \ldots, S_T)^\top$ is a vector of S variables that exclude S_t;
   or generate $\tilde{S}_T$ from $f(\tilde{S}_T \mid \Theta, p_{11}, p_{22}, \Omega_T), t = 1, 2, \ldots, T$.

2. Generate transition probabilities $\mathcal{P} = \{p_{11}, p_{22}\}$ from $f(\mathcal{P} \mid \tilde{S}_T)$.

3. Generate $\Theta = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ from $f(\Theta \mid \Omega_T, \tilde{S}_T)$.

These three steps are iterated until convergence. The detailed implementation is given

below.

Step 1: There are two generation methods. The first is single-step Gibbs sampling, which generates S_t based on $f(S_t \mid \tilde{S}_{\neq t}, \Theta, \mathcal{P}, \Omega_T)$, $t = 1, 2, \ldots, T$. By suppressing the conditioning on the parameters, S_t is generated by $f(S_t \mid \tilde{S}_{\neq t}, \Omega_T)$, and the following results can be derived:

$$f(S_t \mid \tilde{S}_{\neq t}, \Omega_T) \propto f(S_t \mid S_{t-1}) f(y_t \mid S_t) f(S_{t+1} \mid S_t) \tag{2.3.8}$$

For the derivation process, see Section 9.1.1 of Kim and Nelson (1999). Using the equation (2.3.8), $Pr(S_t = j \mid \tilde{S}_{\neq t}, \Omega_T)$ can be calculated as

$$Pr(S_t = j \mid \tilde{S}_{\neq t}, \Omega_T) = \frac{f(S_t = j \mid \tilde{S}_{\neq t}, \Omega_T)}{\sum_{j=1}^{2} f(S_t = j \mid \tilde{S}_{\neq t}, \Omega_T)}.$$

Then, S_t can be generated using a uniform distribution. If the random number generated from $U(0, 1)$ is less than $Pr(S_t = j \mid \tilde{S}_{\neq t}, \Omega_T)$, then $S_t = 1$; otherwise, $S_t = 2$.
Similarly, the second method is multi-move Gibbs sampling, which generates S_t based on $f(\tilde{S}_T \mid \Theta, \mathcal{P}, \Omega_T)$, $t = 1, 2, \ldots, T$. Finally, $Pr(S_t = j \mid \tilde{S}_{t+1}, \Omega_T)$ can be calculated by the following formula,

$$Pr(S_t = j \mid S_{t+1}, \Omega_T) = \frac{f(S_{t+1} \mid S_t = j) f(S_t = j \mid \Omega_T)}{\sum_{j=1}^{2} f(S_{t+1} \mid S_t = j) f(S_t = j \mid \Omega_T)}.$$

The derivation is given in Section 9.1.1 of Kim and Nelson (1999). Then, S_t can be generated by a random variable drawn from $U(0, 1)$ based on $Pr(S_t = j \mid S_{t+1}, \Omega_T)$, which is similar to the single-step Gibbs sampling method.

Step 2: Kim and Nelson (1999) used the beta distribution as a conjugate prior for the transition probabilities. The proposed posterior distribution is given by the following two independent beta distributions:

$$p_{11} \mid \tilde{S}_T \sim \text{Beta}(u_{11} + n_{11}, u_{12} + n_{12})$$
$$p_{22} \mid \tilde{S}_T \sim \text{Beta}(u_{22} + n_{22}, u_{21} + n_{21}),$$

where u_{ij} is a known prior hyper-parameter and n_{ij} represents the transitions from state i to j, which can be easily calculated for a given $\tilde{S}_T$.

Step 3: The last step is based on the parameters in the state-dependent model, which need to be derived from different models. An example in Kim and Nelson (1999) is about generating μ_1 and μ_2, conditional on $\sigma_1^2, \sigma_2^2, \Omega_T, \tilde{S}_T$. They assumed a normal prior for μ as follows:

$$\tilde{\mu} \mid \sigma_1^2, \sigma_2^2 \sim N(b_0, B_0),$$

where $\tilde{\mu} = (\mu_1, \mu_2)^\top$. The posterior distribution is given by

$$\tilde{\mu} \mid \sigma_1^2, \sigma_2^2, \Omega_T, \tilde{S}_T \sim N(b_1, B_1),$$

where $b_1 = (B_0^{-1} + X^\top X)^{-1}(B_0^{-1} b_0 + X^\top (X\tilde{\mu} + V))$ and $B_1 = (B_0^{-1} + X^\top X)^{-1}$, $X = (\frac{1}{\sigma_{S_t}}, \frac{S_t}{\sigma_{S_t}})$, and $V \sim N(0, I_T)$. The parameters μ_1 and μ_2 can be generated based on the posterior distribution.

2.3.4 Other Algorithms

In addition, Francq and Zakoı (2008) employed the generalized method of moments (GMM) to estimate the MS-GARCH parameters. GMM provides an alternative to simulation-based methods such as MCMC, which typically require more computational power. For moderately large samples, GMM performs well.

Furthermore, Zheng et al. (2021) proposed a spectral clustering hidden Markov model that extracts information using spectral methods. Spectral methods aim to reveal the relationship between hidden states and observations through singular value decomposition (SVD) to relate past and future observations. Zheng et al. (2021) extended the spectral method of discrete HMMs with discrete observations to accommodate continuous observations, which can be used for HMS models. The advantages of spectral methods include scalability, strong theoretical properties, and the fact that no likelihood function is required.

2.4 Hidden Semi-Markov State Transition Models

HMS has been extended and adapted into more complex models, such as hidden semi-Markov models (HSMMs), for tasks where HMS faces limitations. Note that the HMS model has two strong assumptions. First, the current latent state depends only on the most recent state in the past, i.e., $Pr(S_t \mid S_{t-1}, \ldots, S_1) = Pr(S_t \mid S_{t-1})$. Second, the time spent in a state follows a geometric distribution. Let $d_k(u)$ be the probability that time u is spent in state k, called the sojourn distribution or duration distribution. In the HMS model, $d_k(u)$ is simply shown to be geometrically distributed (Maruotti et al., 2021).

In fact, the two strong assumptions mentioned above limit the scope and flexibility of

possible applications of the HMS model. If the underlying process does not follow a geometric sojourn distribution, the HMS model may not be appropriate. More drawbacks of the HMS model can be found in Bulla and Bulla (2006) and Maruotti et al. (2019). To overcome the inflexibility, Ferguson (1980) proposed a hidden semi-Markov state switching (HSMS) model that allows for arbitrary sojourn distributions with great flexibility, including discrete and continuous distributions with positive support, such as Poisson, negative binomial, logarithmic, nonparametric, gamma, Weibull, lognormal (Amini et al., 2022).

In the hidden semi-Markov switching model, we still use the following notations as in HMS models:

1. the initial probabilities $\pi_k = Pr(S_1 = k)$
2. the emission probabilities $f(y_t \mid S_t = k, \Theta_k)$
3. the transition matrix $\boldsymbol{P} = \{p_{ij}\}$, where $p_{ij} = Pr[S_t = j | S_{t-1} = i]$

In addition, the probability $d_k(u)$ of the hidden process $\{S_t\}$ staying in the kth state for a time u, is defined in HSMS model as follows:

$$d_k(u) = Pr(S_{t+u} \neq k, S_{t+u-1} = k, \ldots, S_{t+1} = k \mid S_t = k, S_{t-1} \neq k).$$

Ferguson (1980) pointed out that speech recognition is one of the earliest and most successful implementations of HSMS. Over time, this approach has been extended to different domains, including the recognition of printed text, handwriting, human DNA sequences, film events, symbolic plan, and language. Its versatility has also inspired many other applications, such as ground target tracking, mobility tracking in cellular networks, protein structure prediction, air particle prediction, speech synthesis, music classification, image segmentation, electrocardiogram segmentation, and financial time series modeling. Yu (2010) gave an overview of these applications, while Balcilar et al. (2015) used the HSMS model to study various stages of human information processing. Based on the HSMS framework, Xiao and Dong (2015) introduced an innovative reputation management system for the online-to-offline (O2O) e-commerce market. There are also some applications of the HSMS model in the financial field. Bernardi et al. (2018) constructed an HSMS model with Student-t distribution for tail risk interdependence. Maruotti et al. (2019) introduced an HSMS model with a multivariate leptokurtic-normal distribution. Maruotti et al. (2021) combined the HSMS model with quantile regression for time series. Qin et al. (2024) proposed a robust HSMS model with Huber's least favorable distribution.

The expectation-maximization (EM) algorithm and forward-backward algorithm are commonly used to estimate parameters $\boldsymbol{\Phi} = \{\Theta, \boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{D}\}$ in the HSMS models, where $\boldsymbol{D} =$

$\{d_k(u) : u = 1, \ldots, U, k = 1, \ldots, K-1\}$. If we continue with the AR(1) model in Equation (2.2.1) as a simple state-dependent example, the parameter set is given by $\Theta = (\alpha_1, \alpha_2, \phi_{1,1}, \phi_{1,2}, \sigma_1, \sigma_2)$. Through simple calculation, the Q function in the HSMS model becomes the following formula,

$$Q(\boldsymbol{\Phi} \mid \boldsymbol{\Phi}^{(\ell)}) = Q_1(\boldsymbol{\pi}) + Q_2(\boldsymbol{P}) + Q_3(\boldsymbol{D}) + Q_4(\Theta), \tag{2.4.1}$$

where

$$\begin{aligned}
Q_1(\boldsymbol{\pi}) &= \sum_{k=1}^{K} \log \pi_k Pr(S_1 = k \mid \Omega_T; \boldsymbol{\Phi}^{(\ell)}), \\
Q_2(\boldsymbol{P}) &= \sum_{t=1}^{T-1} \sum_{k=1}^{K} \sum_{j \neq k} \log p_{jk} Pr(S_{t+1} = k, S_t = j \mid \Omega_T; \boldsymbol{\Phi}^{(\ell)}), \\
Q_3(\boldsymbol{D}) &= \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{u} \log d_k(u) Pr(S_{t+u+1} = k, \ldots, S_t = k \mid \Omega_T; \boldsymbol{\Phi}^{(\ell)}), \\
Q_4(\Theta) &= \sum_{t=1}^{T} \sum_{k=1}^{K} \log f(y_t \mid S_t = k) Pr(S_t = k \mid \Omega_T; \boldsymbol{\Phi}^{(\ell)}).
\end{aligned}$$

2.5 Smooth Transition Models

The autoregressive model that allows for regime-switching in Equation (2.2.1) can also be viewed as:

$$y_t = \phi_0(S_t) + \sum_{i=1}^{p} \phi_i(S_t) y_{t_i} + \varepsilon_t, t = 1, \ldots, T, \tag{2.5.1}$$

where $\phi_0(S_t), \phi_1(S_t), \ldots, \phi_p(S_t)$ is a function of the state S_t. The regime S_t can be represented by the vector $\boldsymbol{x}_t = (1, y_{t-1}, \ldots, y_{t-p}, s_{1t}, \ldots, s_{kt})^\top$, where $s_{1t}, \ldots, s_{kt}$ are observable variables. Then, the equation (2.5.1) can be written as:

$$y_t = \phi_0(\boldsymbol{x}_t) + \sum_{i=1}^{p} \phi_i(\boldsymbol{x}_t) y_{t_i} + \varepsilon_t, t = 1, \ldots, T, \tag{2.5.2}$$

where $\phi_j(\boldsymbol{x}_t), j = 0, \ldots, p$ can be estimated by nonparametric estimation methods.

For the observed data y_t, where $t = 1, 2, \ldots, T$, the smooth transition model is as follows:

$$y_t = \boldsymbol{\phi}_1^\top(\boldsymbol{x}_t)(1 - G(s_t; \gamma, c)) + \boldsymbol{\phi}_2^\top(\boldsymbol{x}_t) G(s_t; \gamma, c) + \varepsilon_t, t = 1, \ldots, T, \tag{2.5.3}$$

where $\boldsymbol{x}_t = (1, y_{t-1}, \ldots, y_{t-p}, s_{1t}, \ldots, s_{kt})^\top$ consists of lagged endogenous and exogenous variables. The coefficients are denoted as $\boldsymbol{\phi}_i = (\phi_{i,0}, \phi_{i,1}, \ldots, \phi_{i,m})^\top$, where $i = 1, 2$ and $m = p+k$. The variables $s_{1t}, \ldots, s_{kt}$ are observable and exogenous. The transition function $G(s_t; \gamma, c)$ is continuous and bounded between 0 and 1. The transition variable s_t can be a lagged endogenous variable ($s_t = y_{t-d}$ for a certain integer $d > 0$), an exogenous variable ($s_t = z_t$, where z_t is exogenous), or a function of lagged endogenous and exogenous variables. A popular choice for s_t is the lagged endogenous variable y_{t-d}. The threshold between the two regimes is denoted by c, and the smoothness of the transition is controlled by the parameter γ.

We consider the AR(1) model in Equation (2.2.1) as a simple state-dependent example with the number of states being $K = 2$. The smooth transition autoregressive (STAR) model is given as follows:

$$y_t = (\alpha_1 + \phi_{1,1} y_{t-1})(1 - G(s_t; \gamma, c)) + (\alpha_2 + \phi_{1,2} y_{t-1}) G(s_t; \gamma, c) + \epsilon_t.$$

The STAR model is a regime-switching model with two regimes, corresponding to $G(s_t; \gamma, c) = 0$ and $G(s_t; \gamma, c) = 1$. Moreover, the transition between the two regimes is gradual because it depends on the smooth variation of the transition function $G(s_t; \gamma, c)$. As an example, we use the logistic STAR (LSTAR) model, where the transition function $G(s_t; \gamma, c)$ is defined in Equation (2.1.2) and discussed. We further consider $s_t = y_{t-d}$, which leads to the self-excited TAR model as described in Section 2.2. When $\gamma \to 0$, the value of the transition function $G(s_t; \gamma, c)$ approaches 0.5. Therefore, when $\gamma = 0$, the LSTAR model degenerates into a linear model.

In the LSTAR model, the two regimes are usually associated with smaller and larger values of s_t associated with a threshold c. This type of transition function is particularly suitable for modeling business cycles. The LSTAR model with $s_t = y_{t-d}$ can describe periods of positive and negative growth. However, there are other types of regime-switching behavior. If regime transitions are associated with smaller and larger absolute values of s_t, then the exponential STAR (ESTAR) model with the following $G(s_t; \gamma, c)$ is more appropriate:

$$G(s_t; \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\}, \quad \gamma > 0$$

For example, Michael et al. (1997) and Baum et al (2001) applied the ESTAR model to real exchange rates. The ESTAR model becomes linear for both $\gamma \to 0$ and $\gamma \to \infty$. Unlike the LSTAR model, the ESTAR model does not nest SETAR as a special case.

In order to overcome the limitations of SETAR, Jansen and Teräsvirta (1996) proposed

a quadratic logistic function

$$G(s_t;\gamma,c)=\frac{1}{1+\exp\{-\gamma(s_t-c_1)(s_t-c_2)\}},\quad c_1\leq c_2,\gamma>0.$$

This STAR model can nest a three-state SETAR model as a special case.

The smooth transition model mentioned above is a basic smooth transition model that only allows the transition function $G(s_t;\gamma,c)$ to handle two regimes. To overcome this limitation, the model has also been extended to incorporate multiple regimes, accounting for three business cycle phases: depreciation period, slow appreciation, and strong appreciation.

The equation (2.5.3) can be rewritten as:

$$y_t=\boldsymbol{\phi}_1^\top(x_t)+(\boldsymbol{\phi}_2(x_t)-\boldsymbol{\phi}_1(x_t))^\top G(s_t;\gamma,c)+\varepsilon_t. \tag{2.5.4}$$

This can then be extended to a three-regime model by adding a second nonlinear component as follows:

$$y_t=\boldsymbol{\phi}_1^\top(x_t)+(\boldsymbol{\phi}_2(x_t)-\boldsymbol{\phi}_1(x_t))^\top G_1(s_t;\gamma_1,c_1)+(\boldsymbol{\phi}_3(x_t)-\boldsymbol{\phi}_2(x_t))^\top G_2(s_t;\gamma_2,c_2)+\varepsilon_t. \tag{2.5.5}$$

Assuming that $c_1<c_2$, the autoregressive parameters in this model change smoothly from ϕ_1 to ϕ_3 as s_t increases.

More generally, it can be extended to an m-regime model with $m-1$ smoothing parameters $\gamma_1,\ldots,\gamma_{m-1}$ and thresholds $c_1,\ldots,c_{m-1}$, as shown below:

$$\begin{aligned}y_t=\boldsymbol{\phi}_1^\top(x_t)+(\boldsymbol{\phi}_2(x_t)-\boldsymbol{\phi}_1(x_t))^\top G_1(s_t;\gamma_1,c_1)+(\boldsymbol{\phi}_3(x_t)-\boldsymbol{\phi}_2(x_t))^\top G_2(s_t;\gamma_2,c_2)\\+\cdots+(\boldsymbol{\phi}_m(x_t)-\boldsymbol{\phi}_{m-1}(x_t))^\top G_{m-1}(s_t;\gamma_{m-1},c_{m-1})+\varepsilon_t.\end{aligned}$$

Another way to extend the basic model to a four-regime model is to nest two different two-regime smooth transition models, as shown in the following equation:

$$\begin{aligned}y_t=&[\boldsymbol{\phi}_1^\top(x_t)(1-G_1(s_{1t};\gamma_1,c_1))+\boldsymbol{\phi}_2^\top(x_t)G_1(s_{1t};\gamma_1,c_1)](1-G_2(s_{2t};\gamma_2,c_2))+\\&+[\boldsymbol{\phi}_3^\top(x_t)(1-G_1(s_{1t};\gamma_1,c_1))+\boldsymbol{\phi}_4^\top(x_t)G_1(s_{1t};\gamma_1,c_1)]G_2(s_{2t};\gamma_2,c_2)+\varepsilon_t.\end{aligned} \tag{2.5.6}$$

A further extension of the multi-regime model discussed in the previous paragraphs is the time-varying smooth transition autoregressive (TV-STAR) model, which allows for both nonlinear dynamics and time-varying parameters. The TV-STAR model is derived from Equation (2.5.6) by setting $s_{1t}=y_{t-d}$ and $s_{2t}=y_t$, using a simplified notation for the

transition functions, as shown below:

$$\begin{aligned} y_t = &[\boldsymbol{\phi}_1^\top(x_t)(1 - G_1(y_{t-d})) + \boldsymbol{\phi}_2^\top(x_t)G_1(y_{t-d})](1 - G_2(y_t)) + \\ &+ [\boldsymbol{\phi}_3^\top(x_t)(1 - G_1(y_{t-d})) + \boldsymbol{\phi}_4^\top(x_t)G_1(y_{t-d})]G_2(y_t) + \varepsilon_t. \end{aligned} \tag{2.5.7}$$

The switching variables in Equation (2.5.7) include the lagged endogenous variable y_{t-d} and time t.

2.6 Some Other Switching Mechanisms

The previous sections review switching mechanisms in regime-switching models, which have attracted the attention of many researchers. There also some other regime-switching models with different switching mechanisms.

Chang et al. (2017) introduced a new regime-switching model with autoregressive latent factors, which combined the concepts of both threshold and Markov switching. In the HMS models, the Markov chain determines the regimes independent from other parts of the model, which is unrealistic in some cases. In the model proposed by Chang et al. (2017), the mean and volatility processes switch depending on whether the autoregressive latent factor is above or below a certain threshold. The innovation in the latent factor is correlated with the previous innovation. Chang et al. (2017) also allowed the autoregressive latent factor to have a unit root and to accommodate a strongly persistent regime change. If the autoregressive latent factor is exogenous and stationary, the regime-switching model proposed by Chang et al. (2017) simplifies to a conventional Markov-switching model. This new model can take advantage of HMS models and overcome their shortcomings effectively.

Bazzi et al. (2017) proposed a time-varying Markov regime-switching model. In the basic HMS model, the transition probability matrix $\boldsymbol{P}$ is constant over time. Bazzi et al. (2017) let the transition probability matrix vary over time as specific transformations of lagged dependent observations. The update of the time-varying parameters is based on the probability of the regimes, given information up to time $t - 1$.

2.7 Empirical Examples

In this section, we compare the performance of threshold models, hidden Markov regime-switching models, hidden semi-Markov models, and smooth transition models when applied to real data. These models are assumed to have the same AR(1) state-dependent structure as Equation (2.3.3) but are different in the switching mechanism. In addition, we consider

the hidden Markov regime-switching model and the hidden semi-Markov model, where the state-dependent conditional distributions of observations are

$$f(y_t \mid S_t = i, \Omega_t; \Theta) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ -\frac{(y_t - \mu_i)^2}{2\sigma_i^2} \right\}. \tag{2.7.1}$$

For convenience, we denote these models as follows: the threshold model with an AR(1) state-dependent structure (TH-ar), the hidden Markov regime-switching model with an AR(1) state-dependent structure (HMS-ar), the hidden semi-Markov model with an AR(1) state-dependent structure (HSMS-ar), the smooth transition model with an AR(1) state-dependent structure (ST-ar), the hidden Markov regime-switching model with a normal distribution state-dependent structure (HMS-norm), and the hidden semi-Markov model with a normal distribution state-dependent structure (HSMS-norm). The number of regimes here is set to two.

For the TH-ar model, we only employ a simplest SETAR model, where the state variable S_t is given by

$$S_t = \begin{cases} 1, & y_{t-1} < \tau \\ 2, & \tau \leq y_{t-1}. \end{cases}$$

We estimate the model parameters $\alpha_1, \alpha_2, \phi_{1,1}, \phi_{1,2}$, and τ using the least squares estimation (LSE) method (Li and Ling, 2012) by minimizing the following equation:

$$\begin{aligned} L_n(\Theta) &= \sum_{t=1}^{T} (y_t - E_\Theta(y_t \mid y_{t-1}))^2 \\ &= \sum_{t=1}^{T} (y_t - \alpha_1 + \phi_{1,1} y_{t-1})^2 I(y_{t-1} < \tau) \\ &+ \sum_{t=1}^{T} (y_t - \alpha_2 + \phi_{1,2} y_{t-1})^2 I(y_{t-1} \geq \tau). \end{aligned}$$

After obtaining the estimates $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\phi}_{1,1}, \hat{\phi}_{1,2}$, and $\hat{\tau}$, we estimate σ_1^2 and σ_2^2 as follows:

$$\hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{t=1}^{T} (y_t - \hat{\alpha}_1 + \hat{\phi}_{1,1} y_{t-1})^2 I(y_{t-1} < \hat{\tau}),$$

$$\hat{\sigma}_2^2 = \frac{1}{n_2} \sum_{t=1}^{T} (y_t - \hat{\alpha}_2 + \hat{\phi}_{1,2} y_{t-1})^2 I(y_{t-1} \geq \hat{\tau}),$$

where $n_1 = \sum_{t=1}^{T} I(y_{t-1} < \hat{\tau})$ and $n_2 = \sum_{t=1}^{T} I(y_{t-1} \geq \hat{\tau})$.

For the HMS models, we estimate the model parameters by applying the maximum likelihood estimation(MLE)(Specify in this section where MLE is used.) method as outlined in Section 2.3. Details of the MLE algorithm are provided in Hardy (2001). For the HSMS models, we estimate the model parameters by utilizing the expectation-maximization(EM)(Specify in this section where EM is used.) algorithm described in Section 2.4. Specifically, a nonparametric distribution is used as the sojourn distribution. Details of the EM algorithm are provided in Qin et al. (2024).

For the ST-ar model, we use the logistic function in Equation (2.1.2) as the transition function, with $s_t = y_{t-1}$. We estimate the parameters $\alpha_1, \alpha_2, \phi_{1,1}, \phi_{1,2}, \gamma$, and c by minimizing the following function (van Dijk, 1999):

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{t=1}^{T} (y_t - F(y_{t-1}; \Theta))^2, \tag{2.7.2}$$

where

$$F(y_{t-1}; \Theta) = (\alpha_1 + \phi_{1,1} y_{t-1})(1 - G(s_t; \gamma, c)) + (\alpha_2 + \phi_{1,2} y_{t-1}) G(s_t; \gamma, c). \tag{2.7.3}$$

2.7.1 Example 1: The daily returns of the S&P 500

This subsection presents a real example comparing above six different regime-switching models. We consider the daily log returns of the S&P 500 from January 1, 2020 to December 27, 2024, which contain 1255 observations. The original dataset is available at `https://ca.investing.com/indices/us-spx-500-historical-data`. This example illustrates the switching processes modeled by these six different regime-switching approaches.

Figures 2.1 and 2.2 visually depict the estimated regimes or states of the daily log returns of the S&P 500 via these six models. The TH-ar model aims to provide a local linear approximation to the underlying nonlinear relationship, but it fails to capture regime-switching behavior. The TH-ar model partitions the data based on an estimated threshold value of $\hat{\tau} = -0.5226036$ that does not convey any information about the business cycle. In contrast, the HMS and HSMS models effectively identify bear and bull markets as States 1 and 2, respectively. The ST-ar model provides a smooth transition between states, capturing the intermediate dynamics. In addition, the ST-ar model can represent the business cycle, with dark blue dots appearing around recession periods in the time series, indicating that the model aligns more closely with State 1 during these time periods.

Table 2.1 provides the estimated parameters for each regime or state. State 1 represents

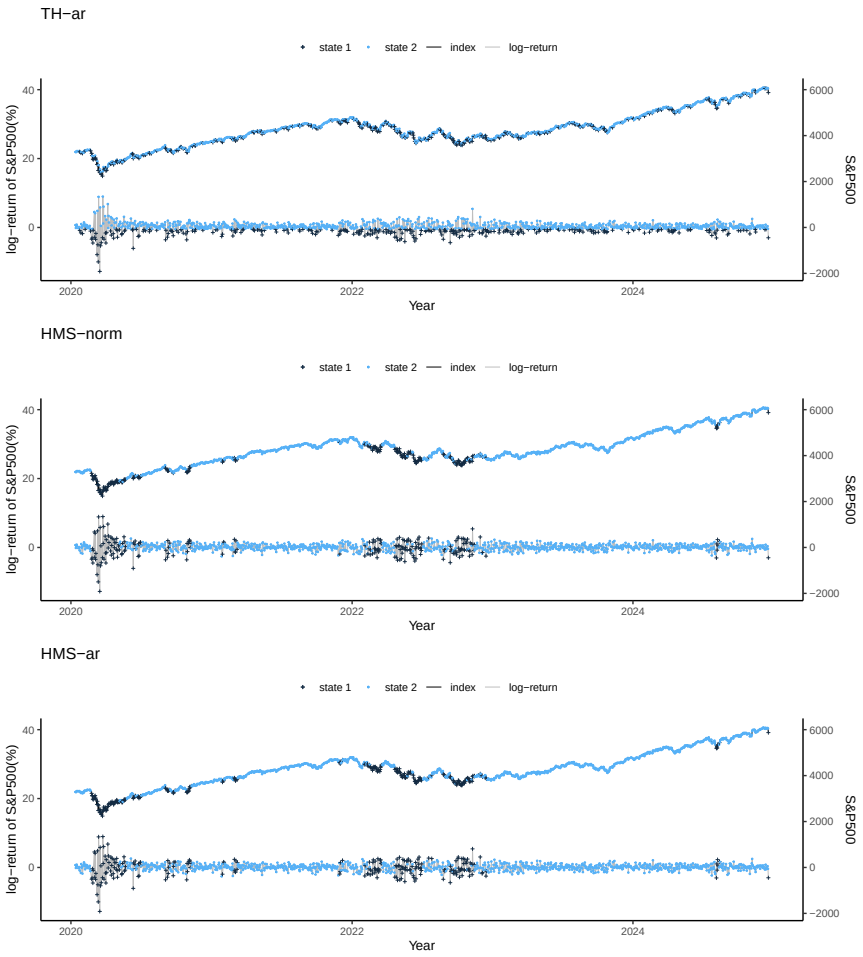

Figure 2.1: Estimation of the two states of the daily S&P500 returns by TH-ar, HMS-norm, and HMS-ar: State 1 (dark blue "+") and State 2 (light blue "•"). In each panel, the upper graph is the daily S&P 500 returns with the estimated states (right axis), and the lower graph is the daily log-return of the S&P500 with the estimated states (left axis).

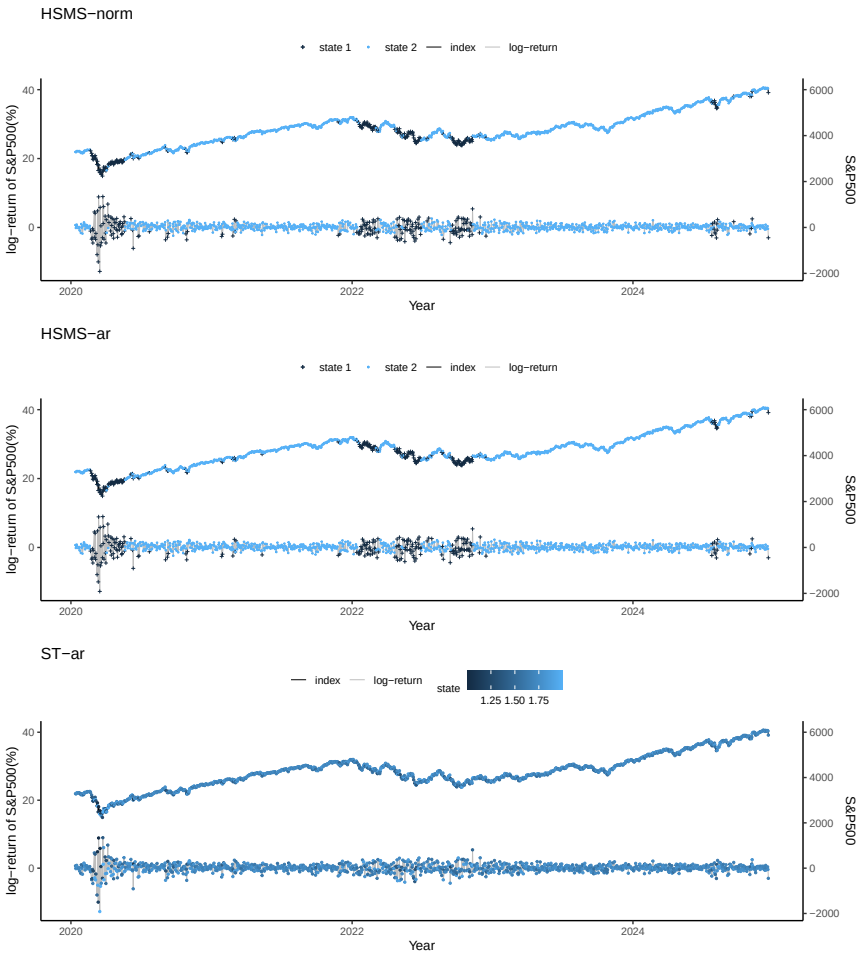

Figure 2.2: Estimation of the two states of the daily S&P500 returns by the HSMS-norm, HSMS-ar, and ST-ar: State 1 (dark blue "+") and State 2 (light blue "•"). In each panel, the upper graph is the daily S&P 500 returns with the estimated states (right axis), and the lower graph is the daily log-return of the S&P500 with the estimated states (left axis).

depreciation periods with larger standard deviations and negative means. The TH-ar model exhibits slight differences in estimated parameters compared to the HMS and HSMS models. The HMS and HSMS exhibit more similar estimated parameters because HSMS model extends the HMS model by incorporating more flexible sojourn distributions. Both models are able to effectively capture regime-switching behavior. The histogram of the state estimates by fitting the ST-ar model is displayed in Figure 2.3. Most estimated states fall within the range of 1.6 to 1.8. The state-dependent parameters of the ST-ar framework exhibit greater variability across states than those of the other models.

Table 2.1: Estimated parameters ($\hat{\mu}_i$) and standard deviations ($\hat{\sigma}_i$) of two-state regime-switching models with state-dependent structure (2.7.1) for the daily S&P500 returns (%). Estimated parameters $\hat{\alpha}_i$ and $\hat{\phi}_{1,i}$ of two-state regime-switching models with state-dependent structure (2.3.3) for the daily S&P500 returns (%).

Method	$\hat{\mu}_1$	$\hat{\mu}_2$			$\hat{\sigma}_1$	$\hat{\sigma}_2$
HMS-norm	-0.26082	0.127168			2.517356	0.79785
HSMS-norm	-0.19689	0.108279			2.514641	0.835683
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\phi}_{1,1}$	$\hat{\phi}_{1,2}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
TH-ar	-0.28359	0.185356	-0.27389	-0.28502	3.123607	1.272196
HMS-ar	-0.25524	0.113606	-0.27559	-0.01371	2.381922	0.827336
HSMS-ar	-0.29157	0.132293	-0.18612	-0.01109	2.409252	0.792376
ST-ar	-8.15227	3.774604	-1.27783	-1.01425		

Akaikei. (1973) introduced the Akaike Information Criterion (AIC), one of the most popular model selection criteria, which is defined as follows:

$$AIC = -2\sum_{i=1}^{n} \ln f(x_i|\hat{\theta}) + 2k,$$

where k is the number of parameters, and $\sum_{i=1}^{n} \ln f(x_i|\hat{\theta})$ is the log-likelihood of the model with estimated parameters. Some studies have found that AIC is not consistent, especially in large samples.

Shibata. (1976) found that AIC tends to select over-parameterized models. To address this issue, Schwarz. (1978) introduced a consistent criterion for large samples, the Bayesian Information Criterion (BIC), which is based on a Bayesian framework and is given by:

$$BIC = -2\sum_{i=1}^{n} \ln f(x_i|\hat{\theta}) + 2k\ln(n),$$

where k is the number of parameters, and $\sum_{i=1}^{n} \ln f(x_i|\hat{\theta})$ is the log-likelihood of the model

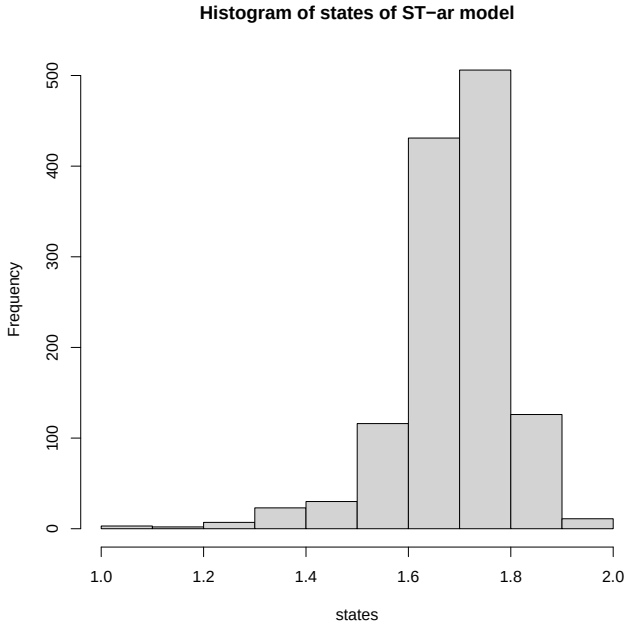

Figure 2.3: Histogram of the state estimates by fitting the ST-ar model for the daily S&P500 returns (%) .

with estimated parameters.

The AIC and BIC of two-state regime-switching models for the daily S&P 500 returns are compared in Table 2.2. The HMS-ar model has the best performance for both AIC and BIC. The HSMS models performs better under AIC but worse than the TH-ar model under BIC, as BIC imposes a stronger penalty on model complexity.

Table 2.2: AIC and BIC of two-state regime-switching models for the daily S&P500 returns (%).

Method	**AIC**	**BIC**
HMS-norm	1693.650	1744.999
HSMS-norm	1812.655	2172.098
TH-ar	1923.529	1959.473
HMS-ar	**1678.032**	**1739.651**
HSMS-ar	1800.965	2170.677
ST-ar	5421.071	5457.015

2.7.2 Example 2: The weekly returns of the EURO STOXX 50

We estimate the two-state regime-switching model for the weekly returns of EURO STOXX 50 spanning January 1, 2015, to January 1, 2025, using data available as of January

2025. The original dataset is from `https://ca.investing.com/indices/eu-stoxx50-historical-data`. Figures 2.4 and 2.5 illustrate the states estimated using six different approaches. The TH-ar model fails to clearly distinguish between depreciation and appreciation periods, as evidenced by the overlap in the state assignments for States 1 and 2.

In contrast, the HMS and HSMS models successfully separated the two business phases, with the states showing more pronounced differences. These differences are even more distinct than those for the daily returns of the S&P 500. It is noteworthy that the HSMS model identifies a greater number of State 1, which can be attributed to its more flexible sojourn distribution. This flexibility suggests that, unlike the states estimated for the daily returns of the S&P 500, the sojourn distribution for the weekly returns of EURO STOXX 50 is not close to a geometric distribution.

The ST-ar model performs poorly, resulting in overlapping state estimates and thus no clear state separation can be achieved. The estimated parameters for all models are summarized in Table 2.3, while the histogram of the state estimates by fitting the ST-ar model to the EURO STOXX 50 returns is presented in Figure 2.6. The parameter estimates for HMS and HSMS models are similar, closely aligning with the results observed for the daily returns of the S&P 500. However, the state distribution by the ST-ar model for the weekly returns of EURO STOXX 50 is less concentrated than that of the daily returns of the S&P 500, where states are mainly located around 1 and 2.

Table 2.3: Estimated parameters ($\hat{\mu}_i$) and standard deviations ($\hat{\sigma}_i$) of two-state regime-switching models with state-dependent structure (2.7.1) for the weekly EURO STOXX 50 returns (%). Estimated parameters $\hat{\alpha}_i$ and $\hat{\phi}_{1,i}$ of two-state regime-switching models with state-dependent structure (2.3.3) for the weekly EURO STOXX 50 index (%).

Method	$\hat{\mu}_1$	$\hat{\mu}_2$			$\hat{\sigma}_1$	$\hat{\sigma}_2$
HMS-norm	-0.41524	0.2549			4.498138	1.685982
HSMS-norm	-0.81259	0.162807			6.395816	2.095628
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\phi}_{1,1}$	$\hat{\phi}_{1,2}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
TH-ar	-0.02741	0.23181	-0.0826	0.022965	7.404448	6.017856
HMS-ar	-0.8359	0.203724	0.080186	-0.11253	6.119401	2.042276
HSMS-ar	-0.65223	0.312221	0.148738	-0.19128	4.841668	1.735389
ST-ar	0.074356	0.120602	-0.15609	0.105752		

The AIC and BIC of two-state regime-switching models for the weekly EURO STOXX 50 returns are compared in Table 2.4. The HMS-ar model performs best according to AIC, while the HMS-Norm model performs best according to BIC. The HSMS model performs better under AIC but worse than the TH-ar model under BIC, as BIC imposes a stronger penalty on model complexity.

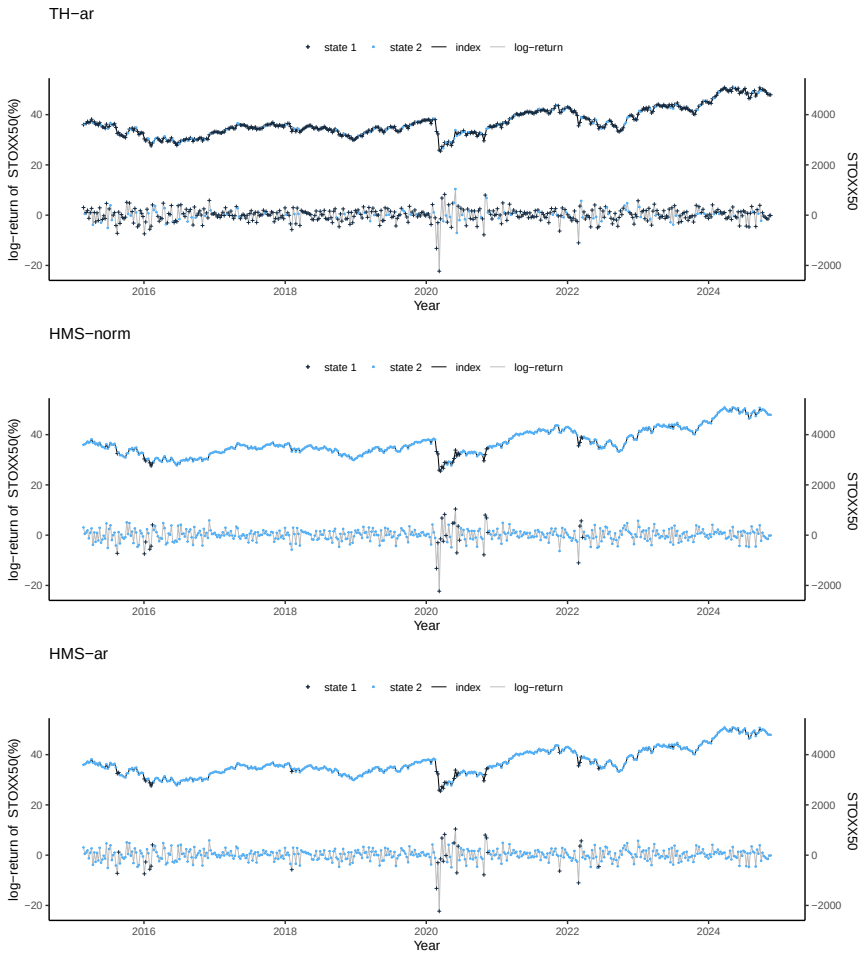

Figure 2.4: Estimation of the two states of the weekly EURO STOXX 50 index by TH-ar, HMS-norm, and HMS-ar: State 1 (dark blue "+") and State 2 (light blue "•"). In each panel, the upper graph is the weekly EURO STOXX 50 returns with the estimated states (right axis), and the lower graph is the weekly log-return of the EURO STOXX 50 with the estimated states (left axis).

Table 2.4: AIC and BIC of two-state regime-switching models for the weekly EURO STOXX 50 returns (%).

Method	**AIC**	**BIC**
HMS-norm	2395.212	**2447.77**
HSMS-norm	2446.582	2814.484
TH-ar	2506.803	2543.593
HMS-ar	**2392.059**	2455.128
HSMS-ar	2433.276	2811.690
ST-ar	2511.391	2548.181

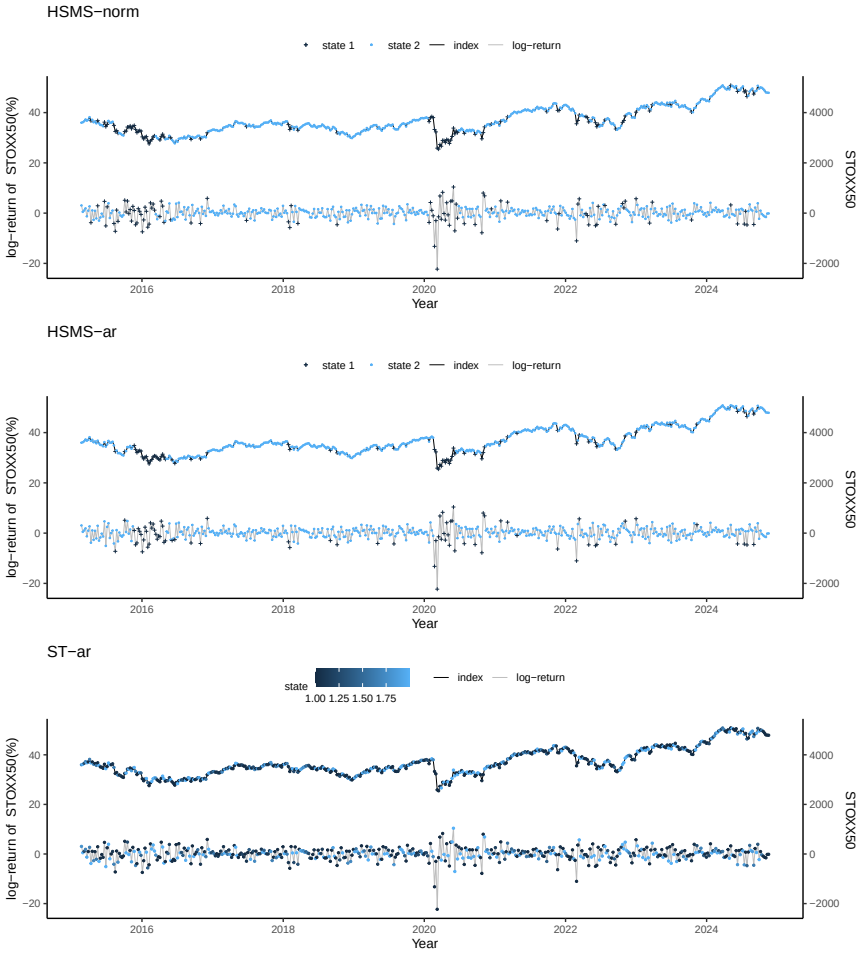

Figure 2.5: Estimation of the two states of the weekly EURO STOXX 50 index by the HSMS-norm, HSMS-ar, and ST-ar: State 1 (dark blue "+") and State 2 (light blue "•"). In each panel, the upper graph is the weekly EURO STOXX 50 returns with the estimated states (right axis), and the lower graph is the weekly log-return of the EURO STOXX 50 with the estimated states (left axis).

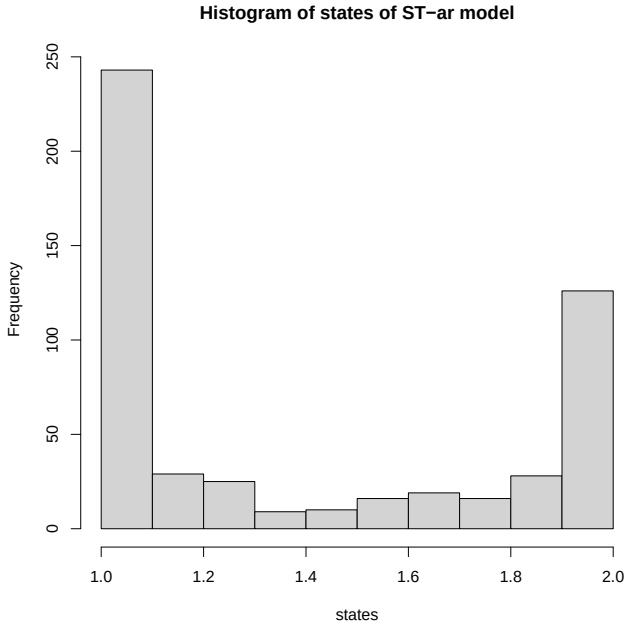

Figure 2.6: Histogram of the state estimates by fitting the ST-ar model to the EURO STOXX 50 returns (%) .

2.8 Conclusion

In this chapter, we review four popular regime-switching models, each defined by a different mechanism: threshold models, hidden Markov regime switching models, hidden semi-Markov models, and smooth transition models. For each type of switching mechanism, we review its relevant framework, popular models, and commonly used estimation methods. In addition, in Section 2.6, we introduce several emerging switching models that deserve further theoretical development and empirical investigation.

We further conduct a comparative analysis of six regime-switching models using two real data examples, focusing on differences in switching mechanisms and state-dependent structures. Threshold models aim to provide a local linear approximation of the underlying nonlinear relationship. HMS and HSMS can better capture the business cycle dynamics compared to other models. In addition, HSMS features a more flexible sojourn distribution. Smooth transition models help achieve a smooth regime-switching process. However, in some data applications, they may have difficulty in clearly separating the states.

Chapter 3

On Robust Estimation of Hidden semi-Markov Regime-switching Model

3.1 Introduction

The general hidden Markov regime-switching (HMS) model is a discrete-time stochastic process $\{y_t, s_t\}_{t\in R^+}$, where s_t represents an unobserved latent state at time t and y_t is a sequence of conditionally independent random variables given s_t. There are two strong assumptions on HMS models. First, the current latent state depends only on the most recent state in the past, $Pr(s_t \mid s_{t-1}, \ldots, s_1) = Pr(s_t \mid s_{t-1})$. Second, the time spent in a state follows a geometric distribution. Let $d_k(u)$ be the probability of time u spending in a state k, named as sojourn distribution or duration distribution. In an HMS model, $d_k(u)$ is simply proved to be a geometric distribution (Maruotti et al., 2021). While these assumptions simplify model formulation and estimation, they limit flexibility in practical applications, particularly for financial time series where state durations often deviate from the geometric assumption and data exhibit heavy tails, skewness, or extreme values(Bulla and Bulla, 2006; Maruotti et al., 2019).

To overcome the inflexibility, a hidden semi-Markov regime-switching (HSMS) model was proposed by Ferguson (1980), which allows great flexibility of arbitrary sojourn distributions, including discrete and continuous distributions with positive support, such as Poisson, negative binomial, logarithmic, non-parametric, gamma, Weibull, log-normal (Amini et al., 2022). This flexibility enables HSMS models to better capture the temporal structure of complex time series. Since their inception, HSMS models have found broad applications across multiple domains.

The first and most successful application of the HSMS is in speech recognition made by

Ferguson (1980), and was further applied in the recognition of printed text, handwritten words, human genes in DNA, events in videos, symbolic plan recognition, language identification, etc. HSMS has received more attention from vast other applications, such as ground target tracking and mobility tracking in cellular networks, protein structure prediction, prediction of particulate matter in the air, speech synthesis, classification of music, image segmentation, ECG segmentation, financial time series modelling, etc. Yu (2010) drew an outline of the applications of HSMS. After that, Balcilar et al. (2015) identified processing stages in human information processing by HSMS models. Xiao and Dong (2015) developed a new reputation management system for online to offline (O2O) e-commerce markets by the HSMS model. Pimentel et al. (2015) applied the HSMS model to detect heartbeat by analyzing the electrocardiogram (ECG) signal. Zhu and Liu (2018) estimated tool wear conditions and predicted the remaining useful life of automated machining in precision manufacturing by HSMS models. Recently, HSMS is also applied in the segmentation of temporal sensory data (Cardot et al., 2019) and reading phases from eye movements (Olivier et al., 2022), modelling financial time series (Zheng et al., 2021; Lin et al., 2020), among others.

The main approach for estimating HMS and HSMS models is likelihood-based methods. Hamilton (1989) used the maximum likelihood estimation (MLE) to estimate the model's parameters of HMS. Gray (1996) adopted the MLE to their Markov regime-switching GARCH model. Lin et al. (2020) also used the MLE to estimate the parameters of their segmented generalized Markov regime-switching model. A general assumption for model estimation via MLE is that the observations under each regime follow normal distributions, which may not hold exactly, especially in financial markets where financial and economic data always have stylized characteristics, such as heavy tails, skewness, and extreme values. Maruotti et al. (2019) developed models under hidden Markov and semi-Markov settings to describe the temporal evolution of returns, whereas the marginal distribution of returns is described by a mixture of multivariate leptokurtic-normal (LN) distributions. Lin et al. (2020) proposed a generalized HMS (GHMS) method with a general class of density function of $y_t|s_t$, a Huber ρ-function for M-estimation, to rectify the problems of heavy-tailed error distribution or data containing outliers. In addition, another robust spectral clustering-based approach was also developed for the estimation of hidden Markov regime-switching models (SC-HMM), targeting discrete (Hsu et al., 2012) and continuous (Zheng et al., 2021) variables, which does not need any knowledge of conditional distributions under latent states. However, both the GHMS and SC-HMM are limited to a geometric sojourn time distribution.

In this chapter, therefore, we propose a robust HSMS estimation method (abbreviated as rHSMS), where the estimation is based on the likelihood-based procedure. We apply two robust benchmark methods for methods comparison, i.e., GHMS and SC-HMM. Lin

et al. (2020) proposed the GHMS using Huber's least favourable distribution to estimate two-state RS models, which improved the robustness from Hardy (2001)'s model. Zheng et al. (2021) proposed the SC-HMM, showing a more robustness than Hardy (2001)'s method. These methods' performance is illustrated in terms of parameter estimation and latent state prediction via simulations where four types of sojourn distributions and error distributions are considered. We also investigate the impacts of atypical values on the transitions matrix $\boldsymbol{\Pi}$ and durations $\boldsymbol{D}$ of the rHSMS. At last, we use the monthly returns of S&P500 and HSI to further probe into the performance of these methods in real data analysis.

There are some main contributions of this study. First, the proposed rHSMS method not only permits an arbitrary sojourn distribution but also relaxes the normality assumption on observations under each regime. We take the ρ-function for M-estimation as the emission distribution in the K-state HSMS model to rectify the problems of heavy-tailed error distribution or data containing outliers, which is the main methodological novelty. Second, we extend technically the estimation procedures of two-state GHMS and SC-HMM to the three-state ones for methods' comparison. Third, the investigations on the impacts of atypical values show that the estimations of transitions and sojourn distributions under the HSMS setting are more easily affected by the heavy-tailed data than the other three scenarios, such as normality, outliers, and contaminated normality. Finally, it is demonstrated that the rHSMS achieves superior performance in robustness and accuracy of parameter estimation and latent state prediction via simulations. Real data analysis on S&P500 and HSI also exhibits the outperformance of rHSMS in extreme values fitting, robustness, interpretability and rationality and risk measure of market behaviours, and the capability of risk measure.

The remainder of Chapter 3 is organized as follows. Section 3.2 presents the formulation of the rHSMS model. Details of the estimation procedure are provided in Section 3.3. Section 5.3 reports simulation studies comparing the performance of the rHSMS, GHMS, and SC-HMM models. Finally, Section 5.4 investigates real market behavior using data from the S&P 500 and the Hang Seng Index (HSI).

3.2 The formulation of rHSMS

Let $\{y_t, t = 1, \ldots, T\}$ be a sequence of T observations. For the hidden semi-Markov RS model, the observed sequence $\{y_t\}$ is related to the unobserved semi-Markov state process $\{s_t\}$ by the conditional distributions (Bulla et al., 2010), i.e.,

$$y_t = \mu_{s_t} + \sigma_{s_t}\epsilon_t,\ s_t = 1, \ldots, K,\ t = 1, \ldots, T, \tag{3.2.1}$$

where μ_{s_t} and σ_{s_t} represent respectively the mean and standard deviation of y_t at state s_t, ϵ_t are error terms following an independent and identical distribution (I.I.D.) with density g. The state process $\{s_t\}$ is a K-state semi-Markov chain with initial probabilities $\pi_k = Pr(s_1 = k)$ and transition probabilities

$$\pi_{k|j} = Pr(s_{t+1} = k \mid s_t = j, s_{t+1} \neq j),$$

satisfying $\pi_{j|j} = 0$ and $\sum_{k=1}^{K} \pi_{k|j} = 1$. One should note that π_k is the probability of starting states, which differs from the HMS model's stationary probability. $\pi_{k|j}$ refers to the transition probability from the previously visited state j to the current state k, $j \neq k$. Denote the K initial probabilities by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^\top$ and the $K \times K$ transition matrix by $\mathbf{\Pi}$. Clearly, the diagonal elements of $\mathbf{\Pi}$ are zeros. In the HSMS model, the duration time of staying in the same state is modelled by the sojourn time distribution. Define $d_k(u)$ the probability of the hidden process $\{s_t\}$ staying time u in the k-th state, i.e.,

$$d_k(u) = Pr(s_{t+u} \neq k, s_{t+u-1} = k, \ldots, s_{t+1} = k \mid s_t = k, s_{t-1} \neq k).$$

In fact, the HMS model is a special case of the hidden semi-Markov RS model, where $d_k(u)$ admits greater flexibility of arbitrary sojourn distribution, such as non-parametric, Poisson, negative binomial, gamma, etc. While in the HMS model, $d_k(u)$ is simply assumed to be a geometric distribution. Maruotti et al. (2021) argued that it is a complicated problem to select an appropriate sojourn distribution for the latent state. Especially for real applications, the sojourn distribution is usually unknown. To allow for the most flexibility, in light of Langrock et al. (2011), we assume the sojourn densities to be discrete non-parametric distributions, by assigning arbitrary point mass to the feasible duration values, which allows us to approximate any parametric sojourn distributions accurately if the maximum times spent in state k, U_k, is sufficiently large. In most empirical situations, $U_k = 30$ is large enough (Maruotti et al., 2021).

Usually, it assumes that the error term ϵ_t follows a standard normal distribution, resulting in unstable model estimation when ϵ_t are heavy-tailed or observations contain outliers. In light of Lin et al. (2020), we assume that ϵ_t follow a general density function $g(\cdot)$,

$$g(x) = \frac{1}{c} e^{-\rho(x)}, \tag{3.2.2}$$

where c is a constant satisfying $c = \int e^{-\rho(x)} dx$ and $\rho(\cdot)$ is a discrepancy function. Herein,

we take $\rho(\cdot)$ is the Huber's $\rho(\cdot)$ function for M-estimation, i.e.,

$$\rho_\tau(x) = \begin{cases} x^2, & |x| \leq \tau, \\ 2\tau|x| - \tau^2, & |x| > \tau, \end{cases} \tag{3.2.3}$$

where the parameter τ shows where the transitions from quadratics to linear take place. Huber's ρ function becomes more similar to the L_1 function for small values of τ while it becomes more similar to the L_2 function for large values of τ (Qin and Wu, 2020). Therefore, Huber's least favourable distribution is unimodal and symmetric, which is equivalent to a double exponential distribution in the tails and a Gaussian distribution in the middle (Lin et al., 2020). Usually, τ=1.345 with an asymptotic relative efficiency of 0.95 for least squares estimation under normal errors. By Eqs.(3.2.1)-(3.2.3), the emission distribution of y_t, $f(y_t \mid s_t = k, \mu_k, \sigma_k)$, is thus given by

$$f(y_t \mid s_t = k, \mu_k, \sigma_k) = \frac{1}{\sigma_k} g\left(\frac{y_t - \mu_k}{\sigma_k}\right) = \frac{1}{c\sigma_k} \exp\left\{-\rho_\tau\left(\frac{y_t - \mu_k}{\sigma_k}\right)\right\}. \tag{3.2.4}$$

The hidden semi-Markov RS model, with the emission distribution $f(y_t \mid s_t = k, \mu_k, \sigma_k)$ as defined by Eq.(3.2.4), can be estimated using Huber's least favourable distribution-based MLE. This method denoted as rHSMS achieves robustness against outliers and accommodates heavy-tailed data. The rHSMS method offers flexibility not only in admiting arbitrary sojourn distributions, but also in the error term. An alternative discrepancy function, such as the L_1 loss, can be employed to construct the general density function $g(\cdot)$, extending beyond the confines of the Huber loss. Consequently, the method does not require a priori knowledge of the error types associated with each latent state.

3.3 Model estimation and goodness-of-fit

We use the MLE-based method to estimate model parameters. For a K-regime hidden semi-Markov RS model in (3.2.1), denote the means and standard deviations for K states respectively by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)^\top$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_K)^\top$. Denote the sojourn distribution probabilities by $\boldsymbol{D} = \{d_k(u) : u = 1, \ldots, U, k = 1, \ldots, K-1\}$. Designate $\boldsymbol{\Theta}$ the set of all parameters, $\boldsymbol{\Theta} = \{\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\Pi}, \boldsymbol{D}\}$. Let $\mathcal{I}_t = \{y_1, \ldots, y_t\}$ represent the observed data sequence through time t and $I_0 = \emptyset$, and let $\mathcal{S}_t = \{s_1, \ldots, s_t\}$ be the corresponding hidden

states sequence of $\mathcal{I}_t$. Then the complete likelihood based on $\mathcal{I}_T$ and $\mathcal{S}_T$ is given by

$$\mathcal{L}_c(\boldsymbol{\Theta} \mid \mathcal{I}_T, \mathcal{S}_T) = Pr(s_1, \ldots, s_T) \prod_{t=1}^{T=1} f(y_t \mid s_t; \boldsymbol{\Theta}). \tag{3.3.1}$$

Let s_r^* represent the rth visited state, and u_r represent the cumulative staying time at state s_r^* (Guedon, 2003). The state starts with s_1^* and stays at the same state with length u_1, then jump to a different state s_2^* at time $u_1 + 1$ and stays there from time $u_1 + 1$ to $u_1 + u_2$. For example, if the staying time of the first visited state is from 1 to 3, then $s_1 = s_1^*, s_2 = s_1^*, s_3 = s_1^*$. Suppose the number of state changes is $R-1$ during the observation period of one to T. The similar process repeats until time $u_1 + \cdots + u_{R-1}$. Whereas the last staying time u_R is right-censored since the information we have is that the last sojourn time is at least u_R. For simplicity, assume that we observe more states sequence until it jumps out from the last visited state s_R^*. We thus have the states probability function, i.e.,

$$P(s_1, \ldots, s_T) = \pi_{s_1^*} d_{s_1^*}(u_1) \prod_{r=2}^{R} \pi_{s_r^* | s_{r-1}^*} d_{s_r^*}(u_r). \tag{3.3.2}$$

Putting (3.3.1) and (3.3.2) together, it follows that

$$\mathcal{L}_c(\boldsymbol{\Theta}) = \pi_{s_1^*} d_{s_1^*}(u_1) \prod_{r=2}^{R} \pi_{s_r^* | s_{r-1}^*} d_{s_r^*}(u_r) \prod_{t=1}^{T} f(y_t \mid s_t), \tag{3.3.3}$$

where $f(y_t \mid s_t)$ is given by Eq.(3.2.4).

We use the well-known Expectation-Maximization (EM) algorithm and Forward-backward algorithm to estimate $\boldsymbol{\Theta}$ in (3.3.3) since the visited states $s_r^*,\ r = 1, \ldots, R$, are unobserved. We breifly descibe the algorithms, and more details can be referenced to Zucchini et al. (2016) or other works.

3.3.1 EM algorithm

EM algorithm proceeds in two steps, E-step and M-step, until convergence. We follow a similar procedure of Maruotti et al. (2021) to illustrate the E-step and M-step.

E-step: compute the expected complete data log-likelihood to obtain the Q-function, $Q(\boldsymbol{\Theta} \mid$

$\boldsymbol{\Theta}^{(\ell)}$), based on $\mathcal{I}_T$ and $\boldsymbol{\Theta}^{(\ell)}$, where $\boldsymbol{\Theta}^{(\ell)}$ denotes an estimate of $\boldsymbol{\Theta}$ at the ℓ-th iteration,

$$
\begin{aligned}
Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(\ell)}) &= E_{\boldsymbol{\Theta}^{(\ell)}}[\log(\mathcal{L}_c(\boldsymbol{\Theta}) \mid \mathcal{I}_T)] && (3.3.4)\\
&= \sum_{s_1=1}^{K} \cdots \sum_{s_T=1}^{K} \log(\mathcal{L}_c(\boldsymbol{\Theta})) Pr(\mathcal{S}_T|\mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}) \\
&= \sum_{s_1=1}^{K} \cdots \sum_{s_T=1}^{K} \log \pi_{s_1^*} Pr(\mathcal{S}_T|\mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}) && (3.3.5)\\
&\quad + \sum_{s_1=1}^{K} \cdots \sum_{s_T=1}^{K} \left(\sum_{r=2}^{R} \log \pi_{s_r^*|s_{r-1}^*} \right) Pr(\mathcal{S}_T|\mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}) && (3.3.6)\\
&\quad + \sum_{s_1=1}^{K} \cdots \sum_{s_T=1}^{K} \left(\sum_{r=1}^{R} \log d_{s_r^*}(u_r) \right) Pr(\mathcal{S}_T|\mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}) && (3.3.7)\\
&\quad + \sum_{s_1=1}^{K} \cdots \sum_{s_T=1}^{K} \left(\sum_{t=1}^{T} \log f(y_t|s_t) \right) Pr(\mathcal{S}_T|\mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}). && (3.3.8)
\end{aligned}
$$

By simple computations, the Q-function becomes that

$$
Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(\ell)}) = Q_1(\boldsymbol{\pi}) + Q_2(\boldsymbol{\Pi}) + Q_3(\boldsymbol{D}) + Q_4(\boldsymbol{\mu}, \boldsymbol{\sigma}), \tag{3.3.9}
$$

where $Q_1(\boldsymbol{\pi}) = \sum_{k=1}^{K} \log \pi_k Pr(s_1 = k \mid \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$, $Q_2(\boldsymbol{\Pi}) = \sum_{t=1}^{T-1} \sum_{k=1}^{K} \sum_{j \neq k} \log \pi_{k|j} Pr(s_{t+1} = k, s_t = j \mid \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$, $Q_3(\boldsymbol{D}) = \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{u} \log d_k(u) Pr(s_{t+u+1} = k, \ldots, s_t = k \mid \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$, $Q_4(\boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{t=1}^{T} \sum_{k=1}^{K} \log f(y_t \mid s_t = k) Pr(s_t = k \mid \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$.

M-step: maximize $Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(\ell)})$ with respect to (w.r.t.) $\boldsymbol{\Theta}$, i.e.,

$$
\hat{\Theta} = \arg\max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(\ell)}). \tag{3.3.10}
$$

Thus, maximizing Q-function is equivalent to maximizing $Q_1(\boldsymbol{\pi})$, $Q_2(\boldsymbol{\Pi})$, $Q_3(\boldsymbol{D})$ and $Q_4(\boldsymbol{\mu}, \boldsymbol{\sigma})$ individually. By the Lagrange method and basic computations, we have that Now we proceed to estimate parameters $\boldsymbol{\Theta}$.

- Initial parameters $\boldsymbol{\pi}$. Maximize $Q_1(\boldsymbol{\pi})$ with the constraint $\sum_{k=1}^{K} \pi_k = 1$. The Lagrange function is formulated as

$$
Q_1(\boldsymbol{\pi}, \lambda) = Q_1(\boldsymbol{\pi}) - \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right), \tag{3.3.11}
$$

  where $\lambda > 0$. Taking the derivative to $Q_1(\boldsymbol{\pi}, \lambda)$ w.r.t. π_k, $k = 1, \ldots, K$, and putting

it equal to zero,

$$\frac{\partial}{\partial \pi_k} Q_1(\boldsymbol{\pi}, \lambda) = \frac{\partial}{\partial \pi_k} \left[\sum_{k=1}^{K} \log \pi_k Pr(s_1 = k | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)}) - \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) \right] = 0, \quad (3.3.12)$$

we have that

$$\pi_k = \frac{\gamma_1^{(\ell)}(k)}{\sum_{k=1}^{K} \gamma_1^{(\ell)}(k)}, \quad (3.3.13)$$

where $\gamma_1^{(\ell)}(k) = Pr(s_1 = k | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$.

- Transition probabilities $\boldsymbol{\Pi}$. Maximize $Q_2(\boldsymbol{\Pi})$ with the constraint $\sum_{k=1, j \neq k}^{K} \pi_{k|j} = 1$. The Lagrange function is formulated as

$$Q_2(\boldsymbol{\Pi}, \lambda) = Q_2(\boldsymbol{\Pi}) - \lambda \left(\sum_{k=1, j \neq k}^{K} \pi_{k|j} - 1 \right), \quad (3.3.14)$$

where $\lambda > 0$. Taking the derivative to $Q_2(\boldsymbol{\Pi}, \lambda)$ w.r.t. $\pi_{k|j}$, and putting it equal to zero,

$$\begin{aligned} \frac{\partial}{\partial \pi_{k|j}} Q_2(\boldsymbol{\Pi}, \lambda) =& \frac{\partial}{\partial \pi_{k|j}} \sum_{t=1}^{T-1} \sum_{k=1}^{K} \sum_{j \neq k} \log \pi_{k|j} Pr(s_{t+1} = k, s_t = j | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)}) \\ & - \frac{\partial}{\partial \pi_{k|j}} \lambda \left(\sum_{k=1, j \neq k}^{K} \pi_{k|j} - 1 \right) = 0, \end{aligned} \quad (3.3.15)$$

we thus have that

$$\pi_{k|j} = \frac{\sum_{t=1}^{T-1} \xi_t^{(\ell)}(j, k)}{\sum_{k=1, j \neq k}^{K} \sum_{t=1}^{T-1} \xi_t^{(\ell)}(j, k)}, \quad (3.3.16)$$

where $\xi_t^{(\ell)}(j, k) = Pr(s_{t+1} = k, s_t = j | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$.

- Sojourn distributions $\boldsymbol{D}$. Maximize $Q_3(\boldsymbol{D})$ with the constraint $\sum_u d_k(u) = 1$. The Lagrange function can be formulated as

$$Q_3(\boldsymbol{D}, \lambda) = Q_3(\boldsymbol{D}) - \lambda \left(\sum_u d_k(u) - 1 \right), \quad (3.3.17)$$

where $\lambda > 0$. Taking the derivative to $Q_3(\boldsymbol{D}, \lambda)$ w.r.t. $d_k(u)$, and putting it equal to

zero,

$$\frac{\partial}{\partial d_k(u)} Q_3(\boldsymbol{D}, \lambda) = \frac{\partial}{\partial d_k(u)} \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{u} \log d_k(u) Pr(s_{t+u+1} = k, \ldots, s_t = k | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)}) - \frac{\partial}{\partial d_k(u)} \lambda \left(\sum_{u} d_k(u) - 1 \right) = 0, \tag{3.3.18}$$

we thus have that

$$d_k(u) = \frac{\sum_{t=1}^{T} \eta_t^{(\ell)}(k, u)}{\sum_u \sum_{t=1}^{T} \eta_t^{(\ell)}(k, u)}, \tag{3.3.19}$$

where $\eta_t^{(\ell)}(k, u) = Pr(s_{t-u+1} = k, \ldots, s_t = k | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$.

- Mean ($\boldsymbol{\mu}$) and sigma ($\boldsymbol{\sigma}$). Maximize $Q_4(\boldsymbol{\mu}, \boldsymbol{\sigma})$ to obtain an optimum of $\boldsymbol{\mu}, \boldsymbol{\sigma}$, i.e.,

$$\begin{aligned} (\boldsymbol{\mu}, \boldsymbol{\sigma}) &= \arg\max_{\mu_k, \sigma_k} \sum_{t=1}^{T} \sum_{k=1}^{K} \log f(y_t | s_t = k) Pr(s_t = k | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)}) \\ &= \arg\max_{\mu_k, \sigma_k} \sum_{t=1}^{T} \sum_{k=1}^{K} \log \left\{ \frac{1}{c\sigma_k} \exp \left[-\rho_\tau \left(\frac{y_t - \mu_k}{\sigma_k} \right) \right] \right\} \gamma_t^{(\ell)}(k) \\ &= \arg\min_{\mu_k, \sigma_k} \sum_{t=1}^{T} \sum_{k=1}^{K} \left[\log(\sigma_j) + \rho_\tau \left(\frac{y_t - \mu_k}{\sigma_k} \right) \right] \gamma_t^{(\ell)}(k). \end{aligned} \tag{3.3.20}$$

where $\rho_\tau(\cdot)$ is given by Eq.(3.2.3), and $\gamma_t^{(\ell)}(k) = Pr(s_t = k | \mathcal{I}_T; \boldsymbol{\Theta}^{(\ell)})$. Obviously, there is no explicit solution to (3.3.20) w.r.t. $\mu_k, \sigma_k,\ k = 1, \ldots, K$. Many optimizing algorithms can be used, such as `optim()` function in R package `stats`, `nloptr()` function in the R package `nloptr` (Ypma et al., 2020).

3.3.2 Forward-backward algorithm

Define forward variables, $\alpha_t(k, u) = Pr(s_{t-u+1} = k, \ldots, s_t = k, \mathcal{I}_t)$, where $k = 1, \ldots, K;\ t = 1, \ldots, T;\ u = 1, \ldots, \min\{U_k, t\}$. Then,

$$\alpha_t(k, u) = \sum_{j=1}^{K} \sum_{u'=1}^{\min(U_k, t-u)} \alpha_{t-u}(j, u') \pi_{k|j} d_k(u) f(y_{t-u+1}^{t} \mid s_{t-u+1}^{t} = k). \tag{3.3.21}$$

Specially, $\alpha_1(k,u) = \pi_k f(y_1 \mid s_1 = k)$ for $t = 1$, and

$$\text{Likelihood} = \sum_{j=1}^{K} \sum_{d=1}^{\min(U_k,T)} \alpha_T(j,d) \quad \text{for} \quad t = T. \tag{3.3.22}$$

Define backward variables, $\beta_t(k,u) = Pr(Y_{t+1}^T = y_{t+1}^T \mid s_{t-u+1}^t = k)$. Then,

$$\beta_t(k,u) = \sum_{j=1}^{K} \sum_{u'=1}^{\min(U_k,T-t)} \pi_{j|k} d_j(u') f(y_{t+1}^{t+u'} \mid s_{t+1}^{t+u'} = j) \beta_{t+u'}(j,u'), \tag{3.3.23}$$

and $\beta_T(k,u) = 1$. It follows that

$$\eta_t^{(\ell)}(k,u) = Pr(s_{t-u+1}^t = k \mid \mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}) = \frac{\alpha_t(k,u)\beta_t(k,u)}{\text{Likelihood}}, \tag{3.3.24}$$

$$\gamma_t^{(\ell)}(k) = Pr(s_t = k \mid \mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}) = \sum_{u=1}^{\min(U_k,t)} \eta_t^{(\ell)}(k,u), \tag{3.3.25}$$

$$\begin{aligned} \xi_t^{(\ell)}(j,k) &= Pr(s_{t+1} = k, s_t = j \mid \mathcal{I}_T, \boldsymbol{\Theta}^{(\ell)}) \\ &= \frac{\sum_{u'=1}^{\min(U_k,t)} \sum_{u=1}^{\min(U_k,T-t)} \alpha_t(j,u') \pi_{k|j} d_k(u') f(y_{t+1}^{t+u} \mid s_{t+1}^{t+u} = k) \beta_{t+u}(k,u)}{\text{Likelihood}}. \end{aligned} \tag{3.3.26}$$

Based on the above quantities, the estimation procedure for the rHSMS is summarized in algorithm 1. We remark that the termination condition is $\log(\mathcal{L}_c(\boldsymbol{\Theta}^{(\ell+1)})) - \log(\mathcal{L}_c(\boldsymbol{\Theta}^{(\ell)}) \leq$ 1e-4 or the iteration reaches the maximum, say, 1000. Moreover, we briefly investigate and discuss the convergence of the algorithm in section 3.4.1.

Algorithm 1 Estimation procedure of the rHSMS.

Input: data sequence $\{y_1, \ldots, y_T\}$, the number of states K.

Output: $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\Pi}}, \hat{\boldsymbol{D}}\}$.

Step 1: initialization, $\boldsymbol{\Theta}^{(0)} := \{\boldsymbol{\mu}^{(0)}, \boldsymbol{\sigma}^{(0)}, \boldsymbol{\pi}^{(0)}, \boldsymbol{\Pi}^{(0)}, \boldsymbol{D}^{(0)}\}$;

Step 2: calculate the forward variable $\alpha_t(k,u)$, Likelihood and backward variable $\beta_t(k,u)$ via Eq.(3.3.21), Eq.(3.3.22) and Eq.(3.3.23), respectively;

Step 3: calculate the quantities $\eta_t^{(\ell)}(k,u)$, $\gamma_t^{(\ell+1)}(k)$, $\xi_t^{(\ell+1)}(j,k)$ by Eq.(3.3.24), Eq.(3.3.25) and Eq.(3.3.26), respectively;

Step 4: update state means $\boldsymbol{\mu}^{(\ell+1)}$ and standard deviations $\boldsymbol{\sigma}^{(\ell+1)}$ by Eq.(3.3.20);

Step 5: update transition probabilities $\boldsymbol{\Pi}^{(\ell+1)}$ by Eq.(3.3.16);

Step 6: update the sojourn distributions $\boldsymbol{D}^{(\ell+1)}$ by Eq.(3.3.19);

Step 7: update the initial probabilities $\boldsymbol{\pi}^{(\ell+1)}$ by Eq.(3.3.13);

Step 8: Repeat **Steps 2-7** until termination.

Remark 1 *For the EM algorithm, it is rather important to properly initialize the parameters to avoid spurious maxima or degenerate solutions. The initialization is stated in the following way. Firstly, we obtain an initial state partition* $\mathcal{S}_0$ *by dividing the ordered sequence of* $\{y_1, \ldots, y_T\}$ *into* K *parts evenly, where* K *is prespecified. Thus the initial value of* π_k *is* $1/K, k = 1, \ldots, K$*. The initials of* μ_k *and* σ_k *(*$k = 1, \ldots, K$*) are computed by the within sample means and sample standard deviations, respectively. For* $\mathbf{\Pi}^{(0)}$*, the diagonal elements are zeros and the off-diagonal elements are equal to* $1/(K-1)$*. Finally, the initial parameter of sojourn times in state* k*,* $d_k(u)$*, is set to be* $1/U_k$*, where* U_k *is a given maximum times spent in state* k*. Some other initialization strategies can be found in (Maruotti et al., 2021; Maruotti and Punzo, 2021) or reference therein.*

Remark 2 *The estimation procedure stated in Algorithm 1 does not produce standard errors in the estimates. To measure the stability of estimators in simulations, we use the Monte Carlo error (MCE) which is defined to be the standard deviation of the simulated estimators, taken across 1000 repetitions. In the real data analysis, we compute standard errors of estimates through a parametric bootstrap-based technique (Maruotti et al., 2019). Firstly, we obtain the fitted model where the parameters are estimated via Algorithm 1 and determine the number of state changes,* $R-1$*. Then, we repeat the following two steps over 1000 times, i.e.,*

> *Step 1: generate a bootstrap time series with* R*–*1 *state changes from the fitted model and record them.*
>
> *Step 2: estimate the parameters for the bootstrap time series through Algorithm 1 and record the bootstrap estimates.*

Finally, the standard errors of estimates (called bootstrap SEs) are approximated by the sample standard deviation of the bootstrap estimates.

3.3.3 Evaluation metrics for goodness-of-fit

One common evaluation metric for a K-dimensional parameter vector $\boldsymbol{\theta}$ is measured by the absolute error (AE) between the estimator $\hat{\boldsymbol{\theta}}$ and the true one $\boldsymbol{\theta}^*$, i.e.,

$$\text{AE}(\theta)_k = \mid \hat{\theta}_k - \theta_k^* \mid, \ k = 1, \ldots, K. \tag{3.3.27}$$

For the RS model, one may be interested in the mean and standard deviation for each regime or state. The smaller the AE values, the better the estimation of the parameter. We thus report the $\text{AE}(\mu)_k$ and $\text{AE}(\sigma)_k$, $k = 1, \ldots, K$. Another evaluation metric that one may also

be interested in is the overall prediction accuracy (abbreviated as Acc) of latent states, which is equal to computing the Acc for a classification problem. Similar to Zheng et al. (2021), the predicted states are obtained via a simple maximum posterior rule, also known as local decoding. Suppose that $\mathcal{A}_k^* = \{t : s_t = k, t = 1, \ldots, T\}$ is the true label index of the regime or state $k, k = 1, \ldots, K$. Denote $\hat{\mathcal{A}}_k = \{t : \hat{s}_t = k, t = 1, \ldots, T\}$ the predicted label index of the regime k, where $\hat{s}_t$ is the predicted state of y_t. Then the overall average accuracy is defined as the fraction of instances that are correctly identified for each state, i.e.,

$$\text{Acc} = \sum_{k=1}^{K} \sum_{t=1}^{T} I(t \in \mathcal{A}_k^* \ \& \ t \in \hat{\mathcal{A}}_k)/KT, \tag{3.3.28}$$

where $I(A) = 1$ if A is true, otherwise $I(A) = 0$. Clearly, Acc $\in [0, 1]$. The larger the Acc values, the better the state's prediction. Ideally, Acc equals one.

Since our goal is to analyze stock market returns, it is important to assess the models' capability of capturing extreme values. We use the test statistic (Bulla, 2011; Breunig et al., 2003; Maruotti et al., 2019), i.e.

$$W = \frac{(\phi - \hat{\phi})^2}{Var(\phi)},$$

where ϕ denotes the empirical proportions of observations lying outside the interval $(-r, r)$, and $\hat{\phi}$ is an estimate of ϕ. The null hypothesis is

$$H0 : \phi = \hat{\phi}$$

meaning that the fitted model correctly reproduces the probability of extreme observations beyond the threshold r. Under H_0, W is is asymptotically χ_1^2-distributed. The higher the p-value, the better the performance in modelling extreme values (Maruotti et al., 2019). In addition, we evaluate the risk measures, the value-at-risk (VaR) and conditional tail expectation (CTE).

3.4 Simulation studies

In this section, we explore four distinct sojourn distributions and examine four types of error terms or outliers. Initially, we provide a concise investigation into the convergence of the EM algorithm. Subsequently, we conduct a comparative analysis involving the SC-HMM, GHMS, and the newly proposed rHSMS, evaluating their performance in terms of parameter

estimation accuracy and latent state prediction for both two- and three-state hidden semi-Markov Regime-Switching (RS) models. Furthermore, we delve into the examination of the impact of atypical values on the transition matrix $\mathbf{\Pi}$ and durations $\boldsymbol{D}$ within the context of the rHSMS. Lastly, we broaden our comparison scope to include rHSMS, HSMM-LP, and HSMM-t, aiming to validate and assess the overall performance of these methods.

3.4.1 Two-state hidden semi-Markov RS model

We first consider a two-state hidden semi-Markov RS model, i.e., $y_t = \mu_{s_t} + \sigma_{s_t}\epsilon_t,\ s_t = 1, 2$, where $\boldsymbol{\mu} = (-1, 1)^\top, \boldsymbol{\sigma} = (1, 0.5)^\top, \boldsymbol{\pi} = (0.7, 0.3)^\top$ and

$$\mathbf{\Pi}_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

For the error terms or observations, we consider the following four scenarios.

- Scenario I: Normal random errors, i.e., $\epsilon_t \sim N(0, 1), t = 1, \ldots, T$.
- Scenario II: There exist outliers in the observations. Randomly take $[0.01n]$ observations from $\{y_1, \ldots, y_T\}$, and magnifying their values by 10 times, where $[x]$ denotes the integer part of x.
- Scenario III: Heavy-tailed random errors, i.e., $\epsilon_t \sim t(3), t = 1, \ldots, T$.
- Scenario IV: Contaminated-normal random errors, i.e., $\epsilon_t \sim 0.99N(0, 1) + 0.01N(1, 0.5), t = 1, \ldots, T$.

For the sojourn distributions, similar to Maruotti et al. (2021), we consider the following four cases.

- Non-parametric distribution,

$$d_k(u) = \begin{cases} (0.1, 0.2, 0.3, 0.4), & u = 1, 2, 3, 4 \text{ and } k = 1, \\ (0.25, 0.25, 0.5), & u = 1, 2, 3 \text{ and } k = 2. \end{cases} \tag{3.4.1}$$

- Shifted-Poisson distribution,

$$d_k(u) = \exp(-\lambda_k)\frac{\lambda_k^{u-1}}{(u-1)!},\ k = 1, 2 \text{ and } u = 1, 2, \ldots, \tag{3.4.2}$$

  where $\lambda_1 = 5, \lambda_2 = 10$.

- Shifted-negative binomial distribution,

$$d_k(u) = \frac{\Gamma(u+\lambda_k-1)}{(u-1)!\Gamma(\lambda_k)} p_k^{\lambda_k}(1-p_k)^{u-1},\ k=1,2 \text{ and } u=1,2,\ldots, \tag{3.4.3}$$

  where $\lambda_1 = 10,\ \lambda_2 = 5$, $p_1 = p_2 = 0.5$.

- Geometric distribution,

$$d_k(u) = p_k(1-p_k)^{u-1},\ k=1,2 \text{ and } u=1,2,\ldots, \tag{3.4.4}$$

  where $p_1 = 0.1, p_2 = 0.2$.

We remark that the maximum staying time is 10 for the parametric distributions.

Table 3.1: Average number of iterations until convergence over 100 replications, $T = 500$.

Sojourn Distribution	Scenario I	Scenario II	Scenario III	Scenario IV
Non-parametric	18.98	19.16	41.11	25.29
Shifted-Poisson	40.18	44.33	69.27	43.49
Shifted-negative binomial	50.85	51.65	81.84	55.38
Geometric	57.77	65.95	110.15	64.27

We initiate our investigation by examining the convergence of Algorithm 1 across various sojourn distributions and scenarios with $K = 2$ and $T = 500$. The results, presented in Table 3.1, disclose the average number of iterations required for convergence over 100 replications. Notably, Scenario III exhibits a significantly higher convergence iterations compared to the other scenarios, suggesting that heavy-tailed data demands more iterations for convergence. The comparable iteration numbers observed between Scenario I and Scenario II imply that outliers exert minimal influence on the convergence rate. Among the four considered sojourn distributions, the hidden semi-Markov Regime-Switching (RS) models with non-parametric sojourn distributions need the fewest iteration times in each scenario. In contrast, the geometric sojourn distribution (HMS) requires the highest number of iterations, implying that the convergence rate of the EM algorithm is marginally superior under the hidden semi-Markov RS setting compared to the HMS.

In our comparative analysis, we proceed to assess different methods by generating data sequences from the two-state RS model. Figures 3.1-3.2 display box plots illustrating the overall accuracy (Acc) over 1000 simulations for two-state Regime-Switching (RS) models with T=500 and T=1000. Clearly, the SC-HMM exhibits the lowest accuracy among the three methods across various scenarios and sojourn distributions. Additionally, the rHSMS

consistently demonstrates significantly better accuracy than the GHMS overall. Notably, when the sojourn distribution is geometric (refer to panels (D) in Figures 3.1-3.2), the discrepancies in state accuracy between the rHSMS and the GHMS are relatively small, as the hidden Markov model is equivalent to the hidden semi-Markov model under the geometric sojourn distribution. Tables 3.2-3.3 present the average Absolute Errors (AEs) of means and standard deviations, with the smallest AEs highlighted in bold. These results align with the prediction accuracy of states depicted in Figures 3.1-3.2. Specifically, the SC-HMM exhibits the poorest performance in terms of parameter estimation, while the rHSMS consistently outperforms the GHMS in the majority of cases.

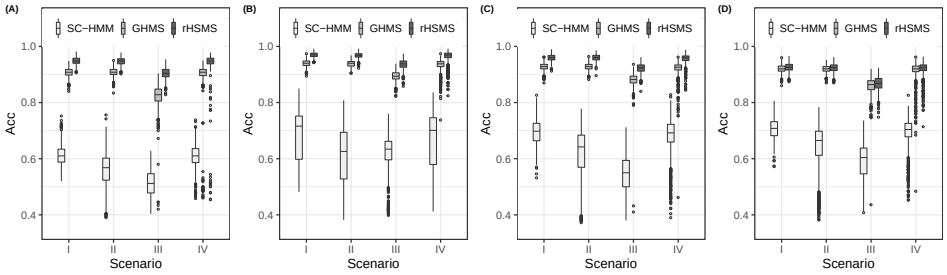

Figure 3.1: Boxplots of overall accuracy (Acc) for two-state RS models over 1000 simulations with $T = 500$ under four different sojourn distributions: (A) non-parametric (B) shifted-Poisson (C) shifted-negative binomial and (D) geometric.

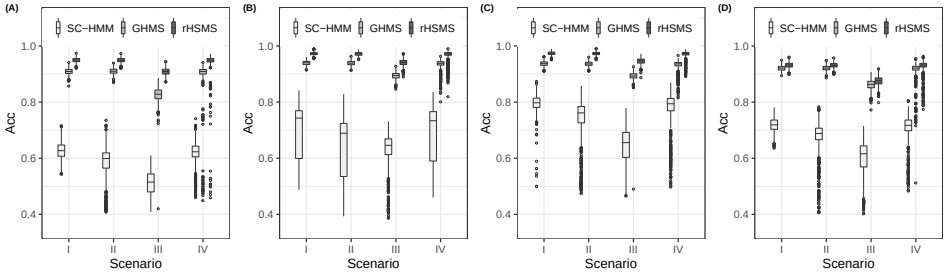

Figure 3.2: Boxplots of overall accuracy (Acc) for two-state RS models over 1000 simulations with $T = 1000$ under four different sojourn distributions: (A) non-parametric (B) shifted-Poisson (C) shifted-negative binomial and (D) geometric.

Table 3.2: Average absolute errors of means (AE(μ_k)) and standard deviations (AE(σ_k)) ($k = 1, 2$) for two-state RS models over 1000 simulations (standard errors in parentheses) with $T = 500$ under four different sojourn distributions and scenarios.

Method	State 1		State 2		State 1		State 2	
	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)
	non-parametric - Scenario I				shifted-Poisson - Scenario I			
SC-HMM	0.32(.13)	0.38(.07)	0.64(.11)	0.44(.05)	0.59(.33)	0.31(.07)	0.35(.20)	0.34(.18)
GHMS	0.10(.07)	0.17(.05)	0.06(.04)	**0.04**(.03)	0.09 (.06)	0.17(.05)	**0.03**(.02)	**0.06**(.02)
rHSMS	**0.05**(.04)	**0.13**(.04)	**0.03**(.03)	0.06(.03)	**0.06**(.05)	**0.14**(.05)	**0.03**(.02)	**0.06**(.02)
	non-parametric - Scenario II				shifted-Poisson - Scenario II			
SC-HMM	0.61(.60)	1.16(1.9)	0.91(.65)	1.25(1.7)	0.83(.56)	0.93(1.7)	0.56(.50)	0.91(1.4)
GHMS	0.09(.07)	0.10(.06)	0.05(.04)	**0.05**(.03)	0.08(.06)	0.11(.06)	**0.03**(.02)	**0.06**(.02)
rHSMS	**0.06**(.04)	**0.09**(.05)	**0.03**(.03)	**0.06**(.03)	**0.07**(.05)	**0.09**(.06)	**0.03**(.02)	**0.06**(.02)
	non-parametric - Scenario III				shifted-Poisson - Scenario III			
SC-HMM	0.62(1.1)	1.62(2.6)	0.94(.48)	1.20(1.5)	0.65(.55)	1.31(1.3)	0.55(.46)	0.86(1.2)
GHMS	0.38(.84)	0.43(2.0)	0.09(.65)	0.15(1.8)	0.14(.11)	0.25(.13)	0.04(.03)	**0.04**(.03)
rHSMS	**0.09**(.09)	**0.24**(.10)	**0.04**(.07)	**0.06**(.06)	**0.10**(.08)	**0.24**(.12)	**0.03**(.02)	0.06(.04)
	non-parametric - Scenario IV				shifted-Poisson - Scenario IV			
SC-HMM	0.32(.15)	0.4(.36)	0.65(.12)	0.47(.18)	0.61(.33)	0.40(.33)	0.38(.22)	0.40(.27)
GHMS	0.12(.12)	0.23(.42)	0.08(.14)	**0.07**(.20)	0.10(.09)	0.20(.11)	**0.03**(.02)	**0.06**(.02)
rHSMS	**0.07**(.10)	**0.17**(.26)	**0.05**(.11)	0.08(.17)	**0.07**(.07)	**0.16**(.11)	**0.03**(.02)	**0.06**(.02)
	shifted-negative binomial - Scenario I				geometric - Scenario I			
SC-HMM	0.22(.12)	0.28(.08)	0.49(.13)	0.38(.07)	0.32(.13)	0.29(.06)	0.40(.11)	0.35(.07)
GHMS	0.07(.05)	0.16(.04)	**0.04**(.03)	**0.06**(.03)	0.08(.06)	0.15(.05)	**0.04**(.03)	**0.05**(.03)
rHSMS	**0.06**(.04)	**0.14**(.04)	**0.04**(.03)	**0.06**(.03)	**0.07**(.05)	**0.14**(.05)	**0.04**(.03)	**0.05**(.03)
	shifted-negative binomial - Scenario II				geometric - Scenario II			
SC-HMM	0.52(.62)	1.02(1.8)	0.78(.56)	1.03(1.5)	0.51(.51)	0.88(1.7)	0.64(.47)	0.87(1.2)
GHMS	0.07(.05)	0.12(.05)	0.04(.03)	**0.06**(.03)	**0.07**(.05)	0.10(.06)	**0.04**(.03)	**0.06**(.03)
rHSMS	**0.06**(.04)	**0.10**(.05)	**0.03**(.03)	**0.06**(.03)	**0.07**(.05)	**0.09**(.06)	**0.04**(.03)	**0.06**(.03)
	shifted-negative binomial - Scenario III				geometric - Scenario III			
SC-HMM	0.51(1.3)	1.44(2.4)	0.95(.96)	1.19(2.5)	0.46(.76)	1.29(1.8)	0.72(.37)	0.97(1.2)
GHMS	0.13(.49)	0.35(3.8)	0.08(.80)	0.22(5.4)	0.21(.12)	0.29(.11)	0.06(.34)	0.14(2.9)
rHSMS	**0.08**(.06)	**0.22**(.10)	**0.04**(.03)	**0.06**(.05)	**0.15**(.11)	**0.27**(.11)	**0.05**(.04)	**0.05**(.05)
	Shifted-negative binomial - Scenario IV				geometric - Scenario IV			
SC-HMM	0.24(.18)	0.38(.40)	0.52(.19)	0.43(.25)	0.33(.16)	0.39(.38)	0.44(.18)	0.40(.28)
GHMS	0.08(.07)	0.18(.14)	0.04(.04)	**0.06**(.04)	0.09(.09)	0.18(.17)	**0.04**(.04)	**0.05**(.04)
rHSMS	**0.06**(.05)	**0.15**(.09)	**0.03**(.02)	**0.06**(.03)	**0.08**(.07)	**0.16**(.12)	**0.04**(.03)	**0.05**(.03)

Table 3.3: Average absolute errors of means (AE(μ_k)) and standard deviations (AE(σ_k)) ($k = 1, 2$) for two-state RS models over 1000 simulations (standard errors in parentheses) with $T = 1000$ under four different sojourn distributions and scenarios.

Method	State 1		State 2		State 1		State 2	
	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)
	non-parametric - Scenario I				shifted-Poisson - Scenario I			
SC-HMM	0.28(.11)	0.38(.06)	0.63(.08)	0.44(.04)	0.56(.33)	0.30(.05)	0.35(.20)	0.34(.18)
GHMS	0.09(.06)	0.17(.04)	0.05(.03)	**0.04**(.02)	0.07(.05)	0.17(.04)	**0.02**(.02)	**0.06**(.01)
rHSMS	**0.04**(.03)	**0.13**(.03)	**0.02**(.02)	0.06(.02)	**0.05**(.03)	**0.13**(.04)	**0.02**(.01)	**0.06**(.01)
	non-parametric - Scenario II				shifted-Poisson - Scenario II			
SC-HMM	0.45(.40)	0.71(1.3)	0.82(.48)	0.85(1.2)	0.75(.50)	0.70(1.6)	0.51(.35)	0.63(.87)
GHMS	0.07(.05)	0.13(.04)	0.05(.03)	**0.04**(.02)	0.06(.04)	0.14(.04)	**0.02**(.02)	**0.06**(.02)
rHSMS	**0.04**(.03)	**0.11**(.03)	**0.03**(.02)	0.06(.02)	**0.04**(.03)	**0.11**(.04)	**0.02**(.01)	**0.06**(.02)
	non-parametric - Scenario III				shifted-Poisson - Scenario III			
SC-HMM	0.66(1.8)	1.75(2.3)	0.97(.43)	1.20(1.4)	0.63(1.2)	1.35(1.3)	0.55(.38)	0.85(.96)
GHMS	0.35(.16)	0.38(1.3)	0.06(.05)	0.08(.05)	0.11(.09)	0.26(.10)	0.03(.02)	**0.03**(.02)
rHSMS	**0.07**(.05)	**0.25**(.07)	**0.03**(.02)	**0.06**(.03)	**0.08**(.06)	**0.25**(.08)	**0.02**(.02)	0.06(.03)
	non-parametric - Scenario IV				shifted-Poisson - Scenario IV			
SC-HMM	0.29(.12)	0.46(.43)	0.64(.10)	0.47(.19)	0.57(.33)	0.39(.35)	0.37(.22)	0.39(.23)
GHMS	0.11(.12)	0.24(.46)	0.08(.14)	0.07(.23)	0.08(.07)	0.19(.10)	**0.02**(.02)	**0.06**(.02)
rHSMS	**0.05**(.09)	**0.17**(.26)	**0.04**(.11)	**0.08**(.19)	**0.05**(.06)	**0.15**(.10)	**0.02**(.01)	**0.06**(.02)
	shifted-negative binomial - Scenario I				geometric - Scenario I			
SC-HMM	0.16(.12)	0.21(.06)	0.25(.08)	0.24(.06)	0.32(.10)	0.29(.04)	0.39(.08)	0.34(.05)
GHMS	0.06(.04)	0.16(.03)	**0.02**(.02)	**0.06**(.02)	0.06(.04)	0.15(.03)	**0.03**(.02)	**0.05**(.02)
rHSMS	**0.04**(.03)	**0.13**(.03)	**0.02**(.02)	**0.06**(.02)	**0.05**(.04)	**0.13**(.04)	**0.03**(.02)	**0.05**(.02)
	shifted-negative binomial - Scenario II				geometric - Scenario II			
SC-HMM	0.35(.44)	0.53(1.3)	0.40(.27)	0.48(.51)	0.40(.35)	0.54(1.1)	0.56(.28)	0.59(.61)
GHMS	0.06(.04)	0.14(.03)	**0.02**(.02)	**0.06**(.02)	**0.05**(.04)	0.12(.04)	**0.03**(.02)	**0.05**(.02)
rHSMS	**0.04**(.03)	**0.11**(.03)	**0.02**(.02)	**0.06**(.02)	**0.05**(.04)	**0.10**(.04)	**0.03**(.02)	**0.05**(.02)
	shifted-negative binomial - Scenario III				geometric - Scenario III			
SC-HMM	0.54(1.8)	1.73(4.6)	0.60(.32)	0.83(.78)	0.50(1.6)	1.31(1.9)	0.76(1.0)	0.99(1.1)
GHMS	0.08(.07)	**0.24**(.08)	0.07(1.2)	0.26(7.2)	0.21(.09)	0.30(.08)	0.04(.03)	**0.03**(.02)
rHSMS	**0.06**(.05)	**0.24**(.07)	**0.02**(.02)	**0.06**(.03)	**0.15**(.08)	**0.28**(.08)	**0.03**(.03)	0.04(.03)
	shifted-negative binomial - Scenario IV				geometric - Scenario IV			
SC-HMM	0.18(.15)	0.33(.40)	0.29(.16)	0.30(.19)	0.32(.11)	0.39(.37)	0.42(.14)	0.39(.20)
GHMS	0.07(.06)	0.18(.08)	**0.02**(.02)	**0.06**(.02)	0.08(.09)	0.17(.14)	**0.03**(.04)	**0.05**(.03)
rHSMS	**0.05**(.04)	**0.15**(.08)	**0.02**(.02)	**0.06**(.02)	**0.07**(.07)	**0.16**(.13)	**0.03**(.02)	**0.05**(.02)

3.4.2 Three-state Hidden Semi-Markov RS model

We now consider a three-state hidden semi-Markov RS model, i.e., $y_t = \mu_{s_t} + \sigma_{s_t}\epsilon_t,\ s_t = 1, 2, 3$, where $\boldsymbol{\mu} = (-1, 0, 1)^\top, \boldsymbol{\sigma} = (1, 0.5, 0.5)^\top, \boldsymbol{\pi} = (0.3, 0.3, 0.4)^\top$ and

$$\boldsymbol{\Pi}_{3\times 3} = \begin{bmatrix} 0 & 0.3 & 0.7 \\ 0.4 & 0 & 0.6 \\ 0.1 & 0.9 & 0 \end{bmatrix}.$$

The error terms distribution or outliers also follow the four scenarios given in subsection 3.4.1. For the sojourn distributions, we consider the non-parametric distribution, i.e.,

$$d_k(u) = \begin{cases} (0.1, 0.2, 0.3, 0.4), & u = 1, 2, 3, 4 \text{ and } k = 1, \\ (0.25, 0.25, 0.5), & u = 1, 2, 3 \text{ and } k = 2, \\ (0.2, 0.8), & u = 1, 2 \text{ and }, k = 3, \end{cases} \tag{3.4.5}$$

and three parametric distributions, such as, shifted-Poisson with three parameters $\lambda_1 = 5, \lambda_2 = 10, \lambda_3 = 13$, shifted-negative binomial with parameters $\lambda_1 = 10, \lambda_2 = 5, \lambda_3 = 10, p_1 = 0.5, p_2 = 0.2, p_3 = 0.6$, geometric distribution with parameters $p_1 = 0.1, p_2 = 0.2, p_3 = 0.3$. For the parametric distributions, we assume that the maximum staying time is 10. Generate data sequence from the assumed model with $T = 500, 1000$, and repeat 1000 simulations.

Figures 3.3-3.4 show the boxplots of Accs for three-state RS models over 1000 simulations with T=500 and T=1000, respectively. Similarly, the SC-HMM exhibits lower accuracy compared to the other two methods, while the rHSMS consistently outperforms the GHMS in the majority of scenarios. An exception is observed for cases involving geometric sojourn distribution, where both the rHSMS and GHMS methods demonstrate similar performance in states prediction accuracy. Tables 3.4-3.5 present the average AEs of means and standard deviations, with the smallest AEs highlighted in boldface. The results align with the state predictions in Figures 3.3-3.4 for the most part. However, it is found that under geometric sojourn distributions, the rHSMS method outperforms the GHMS in terms of parameter estimation, while they perform similarly in both the state's predictions and parameter estimations for two-state RS models.

3.4.3 The impacts of atypical values on the transitions and durations of rHSMS

We now investigate the impacts of atypical values on the transition matrix $\boldsymbol{\Pi}$ and dura-

Table 3.4: Average absolute errors of means (AE(μ_k)) and standard deviations (AE(σ_k)) ($k = 1, 2, 3$) for three-state RS models over 1000 simulations (standard errors in parentheses) with $T = 500$ under four different sojourn distributions and scenarios.

Method	State 1		State 2		State 3		State 1		State 2		State 3	
	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)	AE(μ_3)	AE(σ_3)	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)	AE(μ_3)	AE(σ_3)
	non-parametric - Scenario I						shifted-Poisson - Scenario I					
SC-HMM	0.64(.17)	0.19(.11)	0.25(.12)	0.33(.17)	0.73(.19)	0.40(.20)	0.53(.28)	**0.14**(.12)	0.45(.14)	0.22(.16)	0.36(.18)	0.34(.20)
GHMS	0.97(.76)	0.40(.18)	0.82(.52)	0.16(.11)	0.90(.68)	0.12(.12)	0.53(.43)	0.35(.21)	0.57(.56)	0.18(.12)	0.08(.20)	0.08(.03)
rHSMS	**0.13**(.10)	**0.17**(.07)	**0.04**(.03)	**0.05**(.03)	**0.05**(.04)	**0.06**(.03)	**0.14**(.10)	0.17(.08)	**0.03**(.02)	**0.07**(.03)	**0.03**(.02)	**0.07**(.02)
	non-parametric - Scenario II						shifted-Poisson - Scenario II					
SC-HMM	0.69(.28)	0.28(.82)	0.48(.53)	1.10(1.4)	0.75(.40)	0.63(.97)	0.61(.38)	0.25(.91)	0.72(.60)	1.03(1.4)	0.40(.22)	0.38(.47)
GHMS	1.00(.77)	0.39(.22)	0.77(.49)	0.21(.18)	0.89(.67)	0.15(.18)	0.54(.42)	0.34(.24)	0.53(.52)	0.23(.20)	0.09(.23)	0.08(.05)
rHSMS	**0.12**(.09)	**0.11**(.08)	**0.04**(.03)	**0.05**(.03)	**0.05**(.04)	**0.06**(.04)	**0.13**(.11)	**0.14**(.09)	**0.03**(.03)	**0.07**(.03)	**0.03**(.03)	**0.07**(.03)
	non-parametric - Scenario III						shifted-Poisson - Scenario III					
SC-HMM	0.66(.48)	0.67(1.3)	0.49(.62)	1.44(1.5)	0.72(.49)	0.93(1.0)	0.60(.42)	0.49(.85)	0.56(.50)	1.13(1.3)	0.39(.31)	0.61(.69)
GHMS	0.79(.59)	0.47(.39)	0.66(.44)	0.65(.65)	0.55(.50)	0.25(.37)	0.72(.35)	0.48(.25)	0.53(.44)	0.53(.45)	0.07(.14)	**0.07**(.13)
rHSMS	**0.34**(3.1)	**0.36**(.19)	**0.06**(.05)	**0.07**(.06)	**0.07**(.06)	**0.08**(.10)	**0.22**(.22)	**0.32**(.26)	**0.04**(.04)	**0.06**(.04)	**0.04**(.03)	**0.07**(.05)
	non-parametric - Scenario IV						shifted-Poisson - Scenario IV					
SC-HMM	0.64(.20)	0.25(.37)	0.25(.15)	0.38(.27)	0.72(.21)	0.43(.32)	0.56(.32)	**0.19**(.25)	0.44(.16)	0.30(.35)	0.37(.20)	0.38(.29)
GHMS	0.95(.75)	0.43(.31)	0.81(.53)	0.17(.20)	0.92(.70)	**0.14**(.21)	0.56(.43)	0.38(.26)	0.55(.54)	0.23(.28)	0.10(.26)	0.09(.13)
rHSMS	**0.15**(.16)	**0.20**(.23)	**0.05**(.07)	**0.06**(.05)	**0.06**(.14)	**0.08**(.21)	**0.15**(.14)	0.22(.20)	**0.03**(.03)	**0.07**(.03)	**0.03**(.02)	**0.07**(.03)
	shifted-negative binomial - Scenario I						geometric - Scenario I					
SC-HMM	0.42(.23)	**0.15**(.13)	0.41(.11)	0.23(.16)	0.58(.25)	0.46(.25)	0.46(.22)	**0.16**(.12)	0.33(.14)	0.25(.18)	0.61(.24)	0.41(.23)
GHMS	0.65(.59)	0.37(.21)	0.58(.54)	0.18(.12)	0.29(.58)	0.09(.06)	0.90(.68)	0.41(.19)	0.84(.48)	0.20(.13)	0.63(.69)	0.10(.10)
rHSMS	**0.11**(.09)	0.16(.07)	**0.04**(.03)	**0.07**(.03)	**0.04**(.03)	**0.07**(.03)	**0.17**(.14)	0.18(.08)	**0.06**(.05)	**0.05**(.04)	**0.08**(.07)	**0.07**(.04)
	shifted-negative binomial - Scenario II						geometric - Scenario II					
SC-HMM	0.48(.30)	0.20(.65)	0.67(.64)	1.03(1.6)	0.58(.33)	0.51(.65)	0.53(.30)	0.28(.91)	0.62(.70)	1.19(1.8)	0.63(.32)	0.56(.82)
GHMS	0.64(.57)	0.35(.23)	0.59(.52)	0.22(.19)	0.29(.58)	0.10(.10)	0.94(.67)	0.41(.21)	0.85(.43)	0.26(.23)	0.65(.70)	0.12(.14)
rHSMS	**0.11**(.09)	**0.13**(.08)	**0.03**(.03)	**0.07**(.03)	**0.04**(.03)	**0.07**(.03)	**0.15**(.12)	**0.13**(.09)	**0.06**(.05)	**0.05**(.04)	**0.08**(.07)	**0.07**(.04)
	shifted-negative binomial - Scenario III						geometric - Scenario III					
SC-HMM	0.48(.74)	0.49(1.1)	0.55(.57)	1.29(1.4)	0.54(.31)	0.70(.64)	0.48(.29)	0.49(.60)	0.54(.70)	1.44(1.7)	0.61(.40)	0.79(.81)
GHMS	0.87(.40)	0.49(.21)	0.74(.37)	0.62(.38)	0.19(.38)	0.11(.25)	0.85(.47)	0.49(.27)	0.72(.37)	0.67(.46)	0.35(.47)	0.19(.35)
rHSMS	**0.16**(.13)	**0.28**(.20)	**0.05**(.04)	**0.05**(.04)	**0.05**(.04)	**0.08**(.06)	**0.34**(2.2)	**0.37**(1.0)	**0.10**(.09)	**0.09**(.12)	**0.12**(.10)	**0.10**(.08)
	shifted-negative binomial - Scenario IV						geometric dist. - Scenario IV					
SC-HMM	0.43(.25)	**0.20**(.32)	0.43(.23)	0.34(.59)	0.57(.25)	0.50(.31)	0.46(.23)	0.21(.24)	0.33(.16)	0.35(.44)	0.60(.24)	0.45(.32)
GHMS	0.69(.57)	0.40(.24)	0.61(.52)	0.25(.30)	0.32(.60)	0.12(.24)	0.90(.68)	0.43(.27)	0.82(.47)	0.24(.28)	0.62(.68)	0.13(.25)
rHSMS	**0.12**(.13)	**0.20**(.18)	**0.04**(.03)	**0.07**(.03)	**0.04**(.03)	**0.07**(.04)	**0.18**(.15)	**0.20**(.20)	**0.06**(.06)	**0.06**(.08)	**0.08**(.07)	**0.08**(.10)

Table 3.5: Average absolute errors of means (AE(μ_k)) and standard deviations (AE(σ_k)) ($k = 1, 2, 3$) for three-state RS models over 1000 simulations (standard errors in parentheses) with $T = 1000$ under four different sojourn distributions and scenarios.

Method	State 1		State 2		State 3		State 1		State 2		State 3	
	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)	AE(μ_3)	AE(σ_3)	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)	AE(μ_3)	AE(σ_3)
	non-parametric - Scenario I						shifted-Poisson - Scenario I					
SC-HMM	0.66(.14)	0.17(.10)	0.26(.12)	0.32(.16)	0.77(.18)	0.43(.20)	0.54(.26)	0.12(.11)	0.47(.11)	0.23(.16)	0.40(.17)	0.40(.21)
GHMS	1.06(.79)	0.41(.18)	0.84(.48)	0.17(.11)	1.01(.66)	0.13(.13)	0.46(.44)	0.32(.20)	0.50(.55)	0.17(.12)	0.06(.20)	0.08(.02)
rHSMS	**0.09**(.07)	**0.16**(.05)	**0.03**(.02)	**0.05**(.02)	**0.04**(.03)	**0.06**(.02)	**0.10**(.07)	**0.15**(.06)	**0.02**(.02)	**0.06**(.02)	**0.02**(.02)	**0.06**(.02)
	non-parametric - Scenario II						shifted-Poisson - Scenario II					
SC-HMM	0.71(.20)	0.22(.60)	0.45(.58)	1.01(1.6)	0.77(.41)	0.54(.86)	0.57(.24)	0.15(.41)	0.63(.49)	0.71(1.2)	0.41(.16)	0.41(.24)
GHMS	1.08(.80)	0.40(.20)	0.80(.45)	0.20(.14)	1.05(.66)	0.15(.17)	0.47(.44)	0.30(.23)	0.47(.53)	0.20(.15)	0.05(.17)	0.07(.02)
rHSMS	**0.08**(.07)	**0.12**(.06)	**0.03**(.02)	**0.05**(.02)	**0.03**(.03)	**0.06**(.03)	**0.09**(.07)	**0.11**(.07)	**0.02**(.02)	**0.06**(.02)	**0.02**(.02)	**0.06**(.02)
	non-parametric - Scenario III						shifted-Poisson - Scenario III					
SC-HMM	0.62(.37)	0.54(1.0)	0.55(.69)	1.75(1.9)	0.71(.53)	0.93(1.2)	0.59(.40)	0.41(.91)	0.65(.87)	1.53(1.8)	0.39(.29)	0.57(.90)
GHMS	0.85(.59)	0.48(.32)	0.64(.37)	0.62(.57)	0.56(.53)	0.26(.43)	0.71(.35)	0.46(.17)	0.52(.39)	0.56(.43)	0.05(.14)	**0.06**(.12)
rHSMS	**0.22**(.13)	**0.34**(.13)	**0.04**(.03)	**0.05**(.04)	**0.05**(.04)	**0.06**(.05)	**0.17**(.47)	**0.38**(1.3)	**0.03**(.02)	**0.06**(.04)	**0.03**(.02)	0.07(.04)
	non-parametric - Scenario IV						shifted-Poisson - Scenario IV					
SC-HMM	0.66(.16)	0.22(.29)	0.27(.15)	0.35(.28)	0.77(.20)	0.48(.31)	0.54(.26)	**0.16**(.22)	0.47(.15)	0.32(.43)	0.40(.17)	0.44(.26)
GHMS	1.00(.77)	0.42(.25)	0.80(.48)	0.17(.20)	1.04(.69)	0.15(.22)	0.50(.45)	0.35(.22)	0.47(.53)	0.23(.30)	0.05(.17)	0.09(.15)
rHSMS	**0.11**(.14)	**0.19**(.20)	**0.04**(.08)	**0.05**(.03)	**0.05**(.16)	**0.07**(.16)	**0.11**(.11)	0.20(.19)	**0.02**(.02)	**0.06**(.02)	**0.02**(.02)	**0.07**(.03)
	shifted-negative binomial - Scenario I						geometric - Scenario I					
SC-HMM	0.43(.20)	0.12(.12)	0.42(.09)	0.21(.13)	0.65(.24)	0.56(.25)	0.46(.21)	**0.15**(.12)	0.36(.11)	0.23(.15)	0.65(.23)	0.48(.24)
GHMS	0.53(.57)	0.34(.20)	0.50(.53)	0.17(.12)	0.24(.55)	0.09(.06)	0.98(.71)	0.43(.18)	0.88(.43)	0.20(.13)	0.71(.72)	0.10(.10)
rHSMS	**0.08**(.06)	**0.15**(.05)	**0.02**(.02)	**0.06**(.02)	**0.03**(.02)	**0.07**(.02)	**0.12**(.10)	0.16(.06)	**0.04**(.03)	**0.04**(.03)	**0.05**(.04)	**0.06**(.03)
	shifted-negative binomial - Scenario II						geometric - Scenario II					
SC-HMM	0.47(.23)	0.16(.61)	0.62(.59)	0.84(1.5)	0.61(.24)	0.52(.36)	0.51(.24)	0.19(.59)	0.53(.56)	0.87(1.5)	0.62(.22)	0.46(.33)
GHMS	0.50(.54)	0.31(.22)	0.47(.52)	0.18(.15)	0.22(.53)	0.09(.08)	1.00(.71)	0.42(.20)	0.87(.41)	0.23(.16)	0.75(.73)	0.11(.13)
rHSMS	**0.08**(.06)	**0.12**(.06)	**0.02**(.02)	**0.06**(.02)	**0.03**(.02)	**0.06**(.02)	**0.11**(.09)	**0.13**(.07)	**0.04**(.03)	**0.04**(.03)	**0.05**(.04)	**0.06**(.03)
	shifted-negative binomial - Scenario III						geometric - Scenario III					
SC-HMM	0.45(.26)	0.38(.81)	0.62(.78)	1.67(.17)	0.53(.36)	0.70(.93)	0.50(.35)	0.48(.99)	0.61(.74)	1.83(1.8)	0.57(.38)	0.71(.98)
GHMS	0.90(.28)	0.52(.19)	0.80(.27)	0.69(.27)	0.13(.31)	0.09(.18)	0.92(.42)	0.50(.22)	0.74(.28)	0.73(.42)	0.27(.42)	0.15(.31)
rHSMS	**0.12**(.10)	**0.29**(.13)	**0.03**(.03)	**0.05**(.04)	**0.04**(.03)	**0.07**(.04)	**0.20**(.14)	**0.33**(.27)	**0.07**(.06)	**0.07**(.06)	**0.09**(.07)	**0.09**(.06)
	shifted-negative binomial - Scenario IV						geometric - Scenario IV					
SC-HMM	0.43(.21)	0.17(.22)	0.44(.33)	0.30(.47)	0.63(.24)	0.57(.28)	0.46(.21)	0.21(.28)	0.37(.18)	0.35(.54)	0.63(.24)	0.51(.30)
GHMS	0.58(.57)	0.36(.24)	0.51(.52)	0.23(.28)	0.25(.55)	0.09(.11)	1.00(.65)	0.47(.27)	0.90(.41)	0.27(.29)	0.67(.73)	0.13(.29)
rHSMS	**0.09**(.08)	**0.18**(.16)	**0.02**(.02)	**0.06**(.02)	**0.03**(.02)	**0.07**(.03)	**0.15**(.16)	**0.22**(.29)	**0.05**(.10)	**0.05**(.05)	**0.06**(.10)	**0.07**(.06)

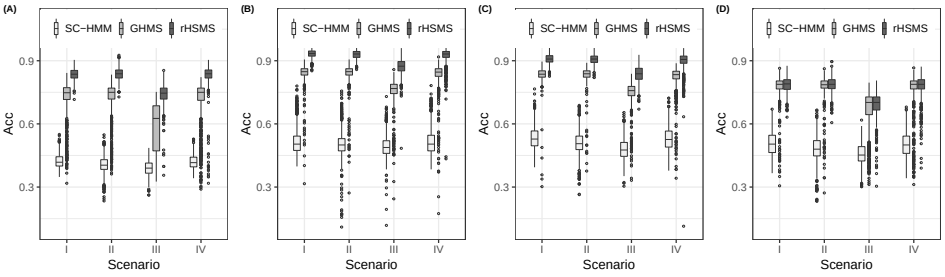

Figure 3.3: Boxplots of overall accuracy (Acc) for three-state RS models over 1000 simulations with $T = 500$ under four different sojourn distributions: (A) non-parametric (B) shifted-Poisson (C) shifted-negative binomial and (D) geometric.

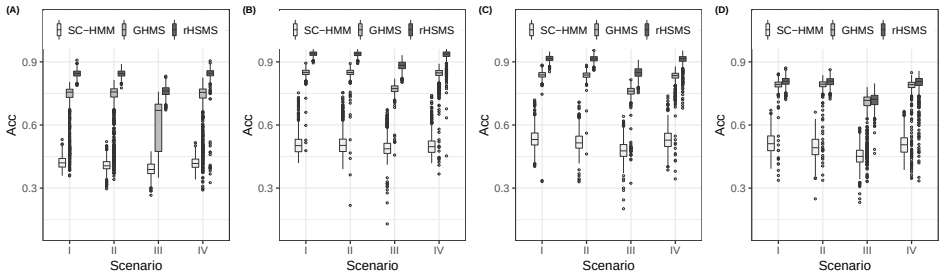

Figure 3.4: Boxplots of overall accuracy (Acc) for two-state RS models over 1000 simulations with $T = 1000$ under four different sojourn distributions: (A) non-parametric (B) shifted-Poisson (C) shifted-negative binomial and (D) geometric.

tions $\boldsymbol{D}$ of the rHSMS. Figure 3.5 shows the boxplots of bias of transition probabilities for three-statex models under four scenarios and four different sojourn distributions (see section 3.4.1). One should note that the diagonals of the transition probabilities $\pi_{k|k}(k = 1, 2, 3)$ are zeros. We only show the off-diagonals of $\mathbf{\Pi}$, $\pi_{k|j}, (j \neq k, j, k = 1, 2, 3)$. Generally, rHSMS overestimates $\pi_{2|1}, \pi_{3|1}, \pi_{1|2}$, while underestimates $\pi_{3|2}, \pi_{1|3}, \pi_{2|3}$ for most cases. It visually shows that the transition probabilities are easily affected by heavy-tailed ($t(3)$) data. As for the impacts of atypical values on durations $\boldsymbol{D}$, Figure 3.6 gives boxplots of probabilities of sojourn times (1 to 10) under four scenarios when sojourn distribution is geometric. Obviously, the durations are affected by the heavy-tailed random error (Scenario III) more easily than the other three scenarios. Similar results can be found when sojourn times follow the other three distributions, say, non-parametric, shifted-Poisson and shifted-negative binomial. Due to the limited space, we only show the geometric one herein.

3.4.4 Comparison of rHSMS, HSMM-LP and HSMM-t

For a broader comparison, we evaluate the proposed rHSMS method against two alternative methods that relax normality assumptions for the multivariate hidden semi-Markov

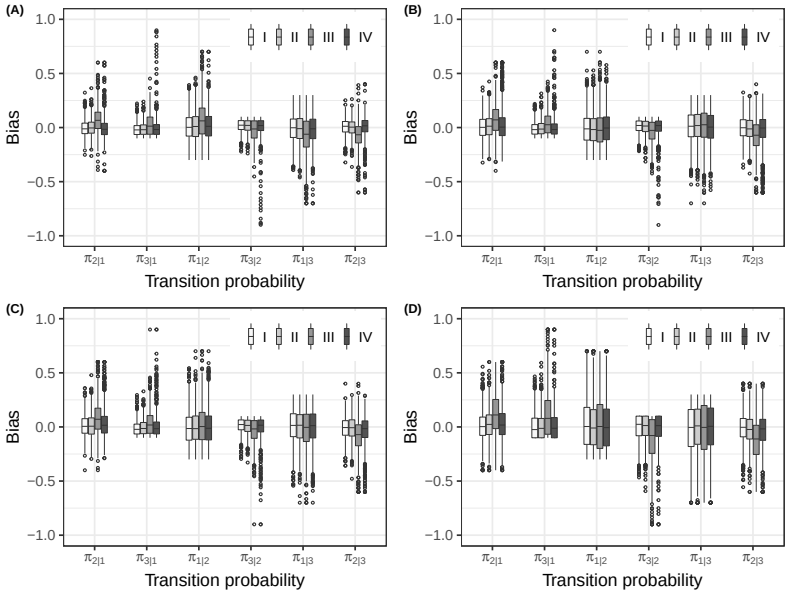

Figure 3.5: Bias of transition probabilities for three-state rHSMS models under four scenarios and four different sojourn distributions: (A) non-parametric (B) shifted-Poisson (C) shifted-negative binomial and (D) geometric.

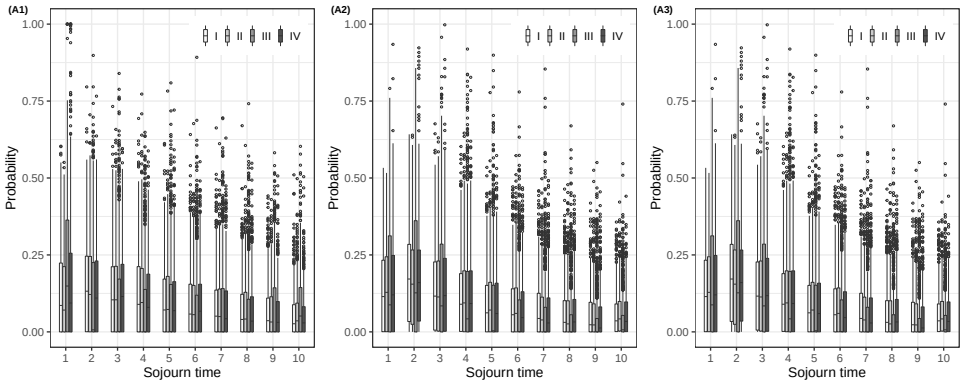

Figure 3.6: Probabilities of sojourn times under four scenarios when sojourn distribution is geometric for three-state rHSMS model (A1) State 1 (A2) State 2 and (A3) State 3.

model: multivariate t distribution (HSMM-t) by Bernardi et al. (2018) and multivariate leptokurtic normal distribution (HSMM-LP) by Maruotti et al. (2019), applied to their univariate versions. We generate data using a three-state hidden semi-Markov model, as detailed in subsection 3.4.2. Tables 3.6-3.7 report the average absolute errors of means (AE(μ_k)) and standard deviations (AE(σ_k)) ($k = 1, 2, 3$) for each state with $T = 500$ under four types of regime assumptions and sojourn distributions. Additionally, Figure 3.7 illustrates the state prediction accuracy for all cases.

Regarding parameter estimation, both rHSMS and HSMM-LP outperform HSMM-t, as evidenced by the smallest absolute errors. Specifically, the proposed rHSMS performs better than the other two in mean estimation, and HSMM-LP excels in standard deviation estimation across all regimes. However, it's challenging to conclusively determine one method's dominance over the others in all cases. In terms of state prediction, when the sojourn distribution is geometric, both rHSMS and HSMM-t outperform HSMM-LP in achieving higher median Acc values. As expected, the overall state prediction accuracy (D) is lower for geometric sojourn distribution compared to the other three types (A-C). As depicted in Figure 3.7(A-C), the top two highest accuracies are consistently obtained by either rHSMS or HSMM-t when dealing with outliers (Scenario II) and under normal (Scenario I) and contaminated normal (Scenario IV) assumptions. In the scenario where data in each regime follows a t-distribution (Scenario III), HSMM-LP achieves the highest Acc values. Importantly, rHSMS outperforms the other two methods in terms of robustness, as evidenced by the presence of more outliers beyond the boxes.

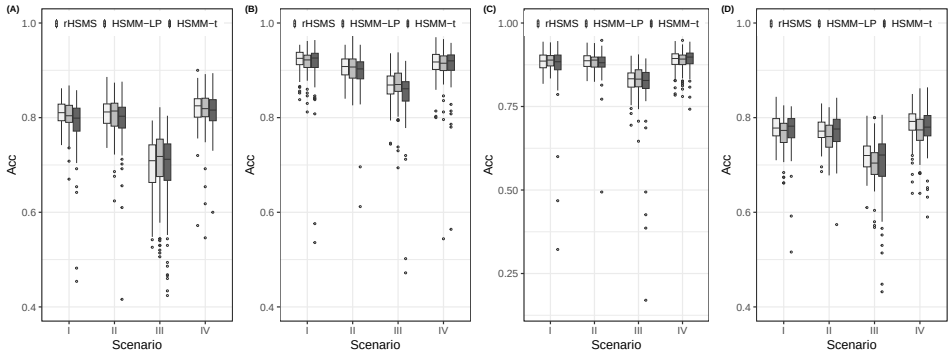

Figure 3.7: Boxplots of overall state prediction accuracy (Acc) for three-state hidden semi-Markov RS models under four different sojourn distributions: (A) non-parametric (B) shifted-Poisson (C) shifted-negative binomial and (D) geometric.

Table 3.6: Average (standard errors in parentheses) absolute errors of means (AE(μ_k)) and standard deviations (AE(σ_k)) ($k = 1, 2, 3$) for three-state hidden semi-Markov RS models (rHSMS, HSMS-LP, HSMM-t) with $T = 500$ under four scenarios where the sojourn distributions are non-parametric and shifted-nagative binomial.

Method	State 1		State 2		State 3	
	AE(μ_1)	AE(σ_1)	AE(μ_2)	AE(σ_2)	AE(μ_3)	AE(σ_3)
	non-parametric - Scenario I					
rHSMS	**0.1228**(.11)	**0.1191**(.08)	**0.0428**(.03)	0.0693(.04)	**0.0451**(.04)	0.0785(.04)
HSMM-LP	0.1654(.14)	0.1537(.10)	0.0484(.04)	**0.040**(.03)	0.0542(.05)	**0.0668**(.04)
HSMM-t	0.1890(.35)	0.7648(1.20)	0.0718(.09)	0.1638(.63)	0.0584(.07)	0.1005(.15)
	non-parametric - Scenario II					
rHSMS	**0.1556**(.12)	**0.1183**(.09)	**0.0469**(.04)	0.0712(.04)	**0.0501**(.04)	0.0720(.04)
HSMM-LP	0.1638(.13)	0.1564(.11)	0.0498(.04)	**0.0532**(.04)	0.0533(.06)	**0.0605**(.04)
HSMM-t	0.2108(.39)	0.7773(.88)	0.0645(.08)	0.1000(.09)	0.0601(.08)	0.0845(.11)
	non-parametric - Scenario III					
rHSMS	0.2701(.20)	0.3110(.16)	**0.0834**(.07)	0.0886(.08)	**0.1016**(.11)	0.1032(.10)
HSMM-LP	**0.2102**(.16)	**0.1197**(.09)	0.0875(.09)	**0.0773**(.07)	0.1040(.12)	**0.1013**(.11)
HSMM-t	0.2293(.20)	0.8845(.66)	0.1461(.47)	0.3639(.84)	0.1045(.11)	0.3207(.75)
	non-parametric - Scenario IV					
rHSMS	0.1455(.13)	0.2040(.14)	**0.0424**(.03)	0.0805(.04)	**0.0536**(.05)	0.0881(.05)
HSMM-LP	**0.1367**(.13)	**0.0972**(.09)	0.0491(.04)	**0.0469**(.03)	0.0622(.06)	**0.0523**(.06)
HSMM-t	0.1610(.14)	0.0980(.22)	0.0498(.04)	0.0617(.05)	0.0553(.05)	0.0525(.09)
	shifted-nagative binomial - Scenario I					
rHSMS	**0.1117**(.09)	**0.1208**(.11)	**0.0350**(.03)	0.0613(.03)	0.0416(.03)	**0.0500**(.03)
HSMM-LP	0.1193(.09)	**0.1208**(.09)	0.0373(.03)	**0.0492**(.03)	0.0422(.03)	0.0683(.04)
HSMM-t	0.1247(.13)	0.6236(.33)	0.0455(.08)	0.1144(.28)	**0.0403**(.03)	0.1120(.13)
	shifted-nagative binomial - Scenario II					
rHSMS	**0.0954**(.07)	**0.1124**(.08)	0.0337(.03)	0.0572(.03)	**0.0463**(.03)	**0.0559**(.04)
HSMM-LP	0.0983(.08)	0.1461(.08)	**0.0329**(.03)	**0.0458**(.03)	0.0472(.03)	0.0689(.04)
HSMM-t	0.1296(.10)	0.5903(.36)	0.0367(.03)	0.0840(.13)	0.0471(.03)	0.1380(.17)
	shifted-nagative binomial - Scenario III					
rHSMS	**0.1397**(.13)	0.2973(.21)	0.0504(.04)	0.0782(.06)	0.0490(.04)	0.0885(.06)
HSMM-LP	0.1421(.13)	**0.1214**(.09)	**0.0445**(.04)	**0.0464**(.03)	**0.0447**(.03)	**0.0440**(.03)
HSMM-t	0.1828(.17)	0.7559(.37)	0.0513(.05)	0.2257(.35)	0.0598(.11)	0.2758(.72)
	shifted-nagative binomial - Scenario IV					
rHSMS	0.1105(.10)	0.1920(.13)	0.0395(.03)	0.0725(.03)	0.0416(.04)	0.0781(.04)
HSMM-LP	**0.1091**(.09)	**0.0938**(.07)	**0.0379**(.03)	**0.0295**(.02)	**0.0374**(.03)	**0.0381**(.03)
HSMM-t	0.1097(.09)	0.1914(.29)	0.0396(.03)	0.0523(.04)	0.0380(.03)	0.0526(.06)

Table 3.7: Average (standard errors in parentheses) of absolute errors of means ($\text{AE}(\mu_k)$) and standard deviations ($\text{AE}(\sigma_k)$) ($k = 1, 2, 3$) for three-states hidden semi-Markov RS models (rHSMS, HSMS-LP, HSMM-t) with $T = 500$ under four scenarios where the sojourn distributions are shifted-Poisson and geometric.

Method	State 1		State 2		State 3	
	$\text{AE}(\mu_1)$	$\text{AE}(\sigma_1)$	$\text{AE}(\mu_2)$	$\text{AE}(\sigma_2)$	$\text{AE}(\mu_3)$	$\text{AE}(\sigma_3)$
	shifted-Poisson - Scenario I					
rHSMS	0.1164(.09)	0.1677(.07)	0.0376(.03)	0.0750(.03)	**0.0319**(.03)	0.0685(.02)
HSMM-LP	**0.1151**(.08)	0.0985(.08)	**0.0356**(.03)	**0.0311**(.03)	0.0346(.03)	**0.0244**(.02)
HSMM-t	0.1400(.14)	**0.0857**(.07)	0.0499(.08)	0.0436(.05)	0.0332(.03)	0.0317(.02)
	shifted-Poisson - Scenario II					
rHSMS	**0.1236**(.10)	**0.1575**(.11)	**0.0366**(.03)	0.0556(.03)	0.0341(.02)	**0.0449**(.03)
HSMM-LP	0.1354(.11)	0.1599(.12)	0.0386(.03)	**0.0506**(.03)	0.0337(.02)	0.0622(.03)
HSMM-t	0.1752(.13)	0.6701(.43)	0.0410(.04)	0.0899(.16)	**0.0350**(.02)	0.1369(.15)
	shifted-Poisson - Scenario III					
rHSMS	0.2004(.15)	0.3348(.22)	0.0460(.03)	0.0826(.06)	0.0430(.03)	0.0807(.05)
HSMM-LP	**0.2001**(.14)	**0.1603**(.11)	**0.0397**(.03)	**0.0433**(.03)	**0.0402**(.03)	**0.0380**(.03)
HSMM-t	0.2453(.17)	0.8955(.46)	0.0617(.14)	0.2473(.36)	0.0439(.03)	0.2437(.24)
	shifted-Poisson - Scenario IV					
rHSMS	**0.1377**(.11)	0.1824(.13)	**0.0340**(.03)	0.0724(.03)	**0.0281**(.02)	0.0698(.03)
HSMM-LP	0.1485(.13)	**0.1279**(.09)	0.0376(.05)	**0.0310**(.03)	0.0299(.02)	**0.0268**(.02)
HSMM-t	0.1548(.14)	0.1454(.23)	0.0395(.05)	0.0419(.04)	0.0307(.02)	0.0427(.05)
	geometric - Scenario I					
rHSMS	**0.1370**(.10)	**0.1067**(.08)	**0.0572**(.05)	0.0927(.05)	0.0687(.06)	0.0732(.05)
HSMM-LP	0.1504(.12)	0.1462(.10)	0.0598(.05)	**0.0489**(.04)	0.0742(.06)	**0.0636**(.05)
HSMM-t	0.1426(.12)	0.6402(.31)	0.0581(.05)	0.1105(.20)	**0.0667**(.06)	0.1121(.14)
	geometric - Scenario II					
rHSMS	0.1568(.10)	**0.1088**(.08)	0.0541(.05)	0.0882(.05)	0.0730(.06)	0.0877(.05)
HSMM-LP	**0.1520**(.12)	0.1488(.09)	0.0679(.05)	**0.0524**(.04)	0.0764(.06)	**0.0639**(.05)
HSMM-t	0.1635(.12)	0.6677(.34)	**0.0540**(.04)	0.1055(.14)	**0.0615**(.05)	0.1026(.12)
	geometric - Scenario III					
rHSMS	**0.1733**(.13)	0.3038(.19)	0.0953(.06)	0.1410(.08)	0.1393(.09)	0.1779(.10)
HSMM-LP	0.1819(.13)	**0.1476**(.12)	**0.0803**(.07)	**0.0825**(.06)	**0.0936**(.08)	**0.0771**(.06)
HSMM-t	0.2270(.25)	0.8386(.83)	0.1679(.70)	0.3608(.90)	0.1002(.10)	0.2340(.16)
	geometric - Scenario IV					
rHSMS	**0.1402**(.10)	0.2306(.17)	**0.0583**(.05)	0.1055(.05)	**0.0610**(.06)	0.1038(.10)
HSMM-LP	0.1567(.13)	**0.1102**(.08)	0.0670(.05)	**0.0472**(.03)	0.0765(.06)	**0.0504**(.04)
HSMM-t	0.1777(.14)	0.1718(.32)	0.066(.06)	0.0830(.06)	0.0688(.06)	0.06130(.09)

3.5 Real data applications

In the financial market, two-state and three-state RS models are commonly used to describe economic and business cycle states. For two-state models, state 1 stands for depreciation periods in the Bear market, and state 2 corresponds to appreciation periods in the Bull market. While three-state RS models encompass depreciation, strong appreciation, and slow appreciation. We undertake the construction of both two-state and three-state RS models to investigate the market behaviors of monthly returns for S&P500 and HSI, respectively. This analysis employs four robust methods, namely the proposed rHSMS, HSMM-LP, HSMM-t, and the GHMS. We remark that the SC-HMM is excluded from consideration due to its poor performance in parameter and state estimation.

3.5.1 S&P500 monthly returns

Table 3.8: Estimated means ($\hat{\mu}_k$), standard deviations ($\hat{\sigma}_k$) and bootstrap standard errors in parentheses of two- or three-state RS models for S&P500 monthly returns (%).

State	Method	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_3$
$K=2$	rHSMS	-0.5455$_{(.44)}$	0.7021$_{(.13)}$	–	2.3593 $_{(.29)}$	0.8247$_{(.14)}$	–
	HSMM-LP	-0.5082$_{(.43)}$	0.7421$_{(.13)}$	–	2.8646 $_{(.38)}$	1.1026$_{(.22)}$	–
	HSMM-t	-0.6806$_{(.52)}$	0.7047$_{(.15)}$	–	2.2160 $_{(.40)}$	0.8794$_{(.17)}$	–
	GHMS	-0.1631$_{(.54)}$	0.5672$_{(.21)}$	–	2.3145$_{(.28)}$	0.9005$_{(.22)}$	–
$K=3$	rHSMS	-1.4014$_{(.59)}$	0.5527$_{(.14)}$	1.7746$_{(.55)}$	2.0246$_{(.32)}$	0.7943$_{(.18)}$	1.5400$_{(.36)}$
	HSMM-LP	-1.2997$_{(.69)}$	0.4742$_{(.14)}$	1.5848$_{(.49)}$	2.4687$_{(.42)}$	0.9661$_{(.20)}$	1.6670$_{(.35)}$
	HSMM-t	-1.4774$_{(.55)}$	0.5204$_{(.13)}$	1.8831$_{(.50)}$	1.9698$_{(.38)}$	0.8143$_{(.12)}$	1.4681$_{(.39)}$
	GHMS	-1.3812$_{(.47)}$	0.5658$_{(.12)}$	3.5417$_{(.46)}$	2.1702$_{(.24)}$	0.9524$_{(.09)}$	0.4734$_{(.23)}$

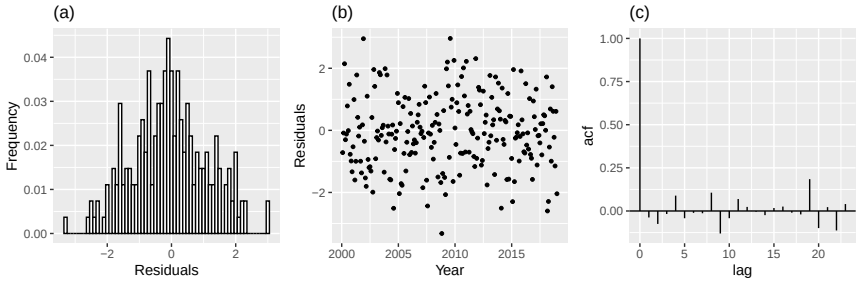

Figure 3.8: Residuas of S&P500 monthly returns estimated via the two-state rHSMS (a) residuals plot (b) ACF of residuals.

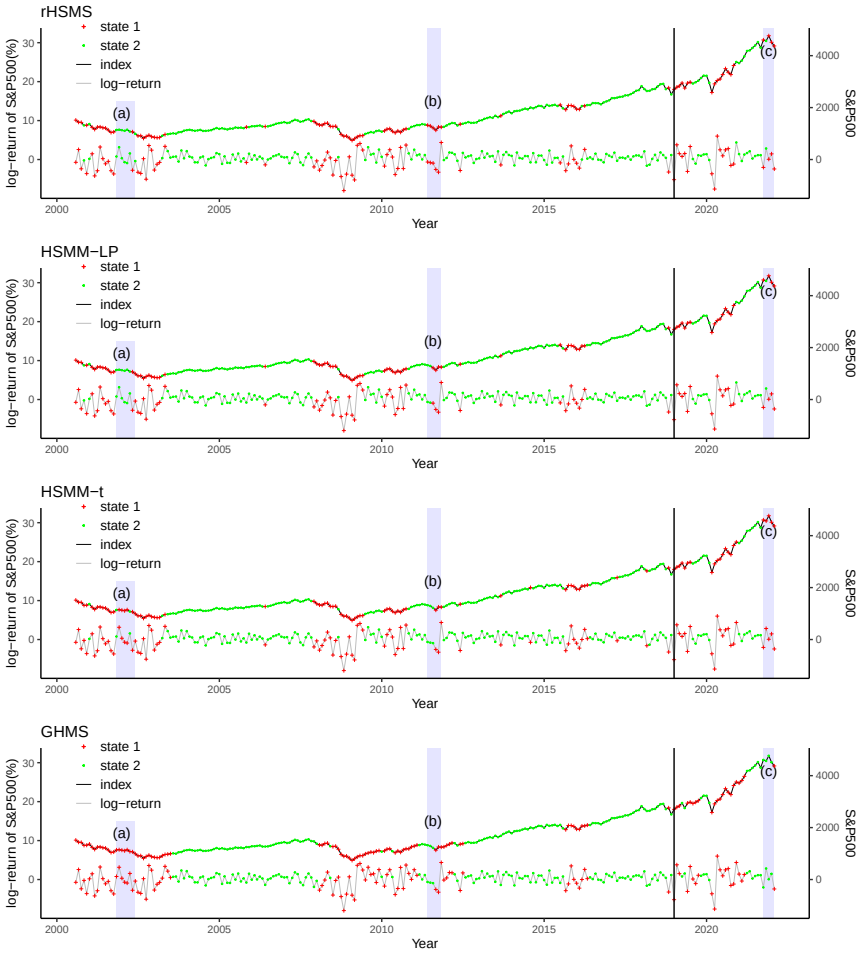

Figure 3.9: Two states estimation for monthly S&P500, state 1 (red "+") and state 2 (green "•") via the rHSMS, HSMS-LP and HSMS-t, and GHMS in the training set and test set, where the vertical line is the cut point at Jan. 1,2019. In each panel, the upper plot is the monthly S&P 500 index with estimated states (right axis), and the lower one is the log-returns of monthly S&P500 with estimated states (left axis).

We estimate both two- and three-state RS models using S&P500 monthly returns spanning from August 27, 2000, to August 21, 2022. The data is divided into a training set, covering the period from August 27, 2000, to December 31, 2018, and a test set, spanning from January 1, 2019, to August 21, 2022. The original dataset can be accessed at `https://ca.investing.com/indices/us-spx-500-historical-data`. Before presenting the estimation results, it is customary to assess the modeling adequacy of the proposed rHSMS method. We begin by applying the Ljung–Box (LB) test, a widely used test for independence in econometrics and other time series analysis applications, to the residuals. The LB test evaluates the null hypothesis that the elements of a series are independently distributed. Rejection of this null hypothesis occurs when the test statistic ($\mathcal{X}$-square) is sufficiently large or when the associated p-value falls below a predetermined significance level, typically 0.05. A rejection implies the presence of residual dependence, indicating modeling inadequacy. We conduct LB tests for both two- and three-state RS models, yielding $\mathcal{X}$-square statistics of 0.3239 and 0.5054, respectively. The corresponding p-values are 0.5693 and 0.4771, indicating strong support for the independence of the residual series. The higher p-value in the two-state RS model suggests that a two-state model adequately captures the S&P500 monthly returns. Furthermore, we visually assess the adequacy of the two-state rHSMS through the residuals plot and auto-correlation function (ACF) of residuals, depicted in Figure 3.8. The left panel of the figure illustrates a trend of constant variance in the residual series, while the ACF plot on the right panel indicates randomness in the residuals. These findings align with the LB test results, confirming the adequacy of regime switching models.

Table 3.8 provides the estimated means and standard deviations (with bootstrap standard errors in parentheses) for each state. In the context of two- or three-state RS models, state 1 denotes the depreciation period, as indicated by the high volatility of returns, negative means, and larger standard deviations presented in Table 3.8. In the two-state RS model, state 2 corresponds to the appreciation period, characterized by low volatility, positive means, and smaller standard deviations of returns. In the three-state RS model, state 2 signifies slow appreciation, featuring positive means, low volatility, and the smallest standard deviations of returns in all cases with one exception. Notably, in the estimation of standard deviations using GHMS, $\hat{\sigma}_2$ is larger than $\hat{\sigma}_3$, which poses a challenge for explanation. State 3 represents strong appreciation, showcasing larger means, higher volatility, and larger standard deviations of returns compared to state 2. It is noteworthy that rHSMS, HSMM-LP, and HSMM-t exhibit similar behaviors, as evidenced by the parameter estimates for means and standard deviations being closely aligned in most cases.

Henceforth, we employ two-state RS models for further analysis due to their modeling

adequacy. Regarding state predictions, Figure 3.9 visually illustrates the estimated two states for the monthly S&P500. The results from three hidden semi-Markov methods and GHMS exhibit a notable degree of proximity. This convergence can be attributed to the fact that the sojourn time distribution follows a geometric distribution, supported by the simulation results for state prediction accuracy in Section 5.3. However, there are distinct periods, such as (a)-(c) in Figure 3.9, where these methods yield disparate estimations of regimes. In period (a), rHSMS and HSMM-LP successfully capture the regime change, while HSMM-t and GHMS fail to identify the state switching. For period (b), rHSMS detects the change in regimes three months earlier than the other three methods. In the test set, all three hidden semi-Markov methods exhibit sensitivity to the changing of state 1 in period (c), whereas GHMS displays a delayed switching. Overall, rHSMS demonstrates greater sensitivity than the other methods in capturing state switching, a crucial aspect in the financial markets.

In addition, our investigation involves assessing and comparing the methods' proficiency in fitting extreme values, which hold significant importance for risk measures in the financial market. Table 3.9 presents the empirical ϕ, along with the estimates $\hat{\phi}$ and corresponding p-values associated with the statistics W given various r. A higher p-value indicates superior performance in modeling extreme values. Among all the methods, the estimated $\hat{\phi}$s are close to the empirical ones upon different r values. Both rHSMS and HSMM-LP fail to reject the null hypothesis at the 0.05 significance level, suggesting their effective performance in modeling extreme values. However, GHMS rejects the null when $r = 1.5, 2$, and HSMM-t rejects the null at $r = 1.5$ while showing an edge-cutting tendency at $r = 2$. This indicates their inadequacy in reproducing extreme values in specific cases.

Finally, we present the risk measures of VaR and CTE in Table 3.10. Assuming an investment of one dollar in such stocks during the study period with a zero interest rate, we calculate VaR and CTE. The largest values of VaR and CTE are obtained either by HSMM-t or GHSM for different quantiles, leading to an overestimation of risk. Moreover, GHSM provides the smallest VaR and CTE for some quantiles, potentially resulting in an underestimation of risk. In contrast, the VaR and CTE values based on rHSMS fall in the middle for most cases, offering a balanced estimation of risk without significant under or overestimation.

3.5.2 HSI monthly returns

We also estimate two- and three-state RS models for HSI monthly returns spanning from Sep. 1st, 2000 to Aug. 1st, 2022. The original data is from `https://ca.finance.yahoo.com/quote/%5EHSI/`. We carry out the BL test on the residuals of rHSMS. The p-

Table 3.9: Empirical (ϕ) and estimated ($\hat{\phi}$) proportions of observations lying outside $(-r, r)$, and p-values from the test statistic W for S&P500 monthly returns using two-state RS models.

	Emperical		rHSMS		HSMM-LP		HSMM-t		GHMS	
r	ϕ	p-value	$\hat{\phi}$	p-value	$\hat{\phi}$	p-value	$\hat{\phi}$	p-value	$\hat{\phi}$	p-value
1	0.5498	–	0.5455	0.1111	0.5432	0.1699	0.5452	0.1220	0.5398	**0.1716**
1.5	0.3764	–	0.3717	**0.1176**	0.3732	0.0796	0.3775	0.0292	0.3746	0.0260
2	0.2657	–	0.2576	0.2177	0.2578	**0.2330**	0.2643	0.0413	0.2620	0.0561
2.5	0.1956	–	0.1866	0.2761	0.1850	**0.3306**	0.1877	0.2631	0.1857	0.1847

Table 3.10: Estimated VaR and CTE for S&P500 monthly returns given different quantiles.

Method	VaR				CTE			
	0.900	0.950	0.975	0.995	0.900	0.950	0.975	0.995
rHSMS	0.0942	0.2379	0.3237	0.4145	0.2219	0.3351	0.4095	0.4983
HSMM-LP	0.0839	0.2087	0.2905	0.4012	0.2381	0.3169	0.3910	0.4770
HSMM-t	0.1365	0.2382	0.3174	0.4388	0.2695	0.3428	0.4045	0.5231
GHSM	0.0493	0.1953	0.3534	0.4263	0.2308	0.3304	0.4658	0.4734

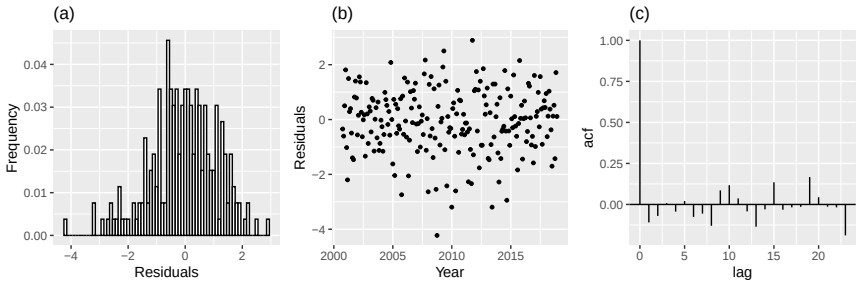

Figure 3.10: Residuals of HSI monthly returns estimated via the three-state rHSMS (a) residuals plot (b) ACF of residuals.

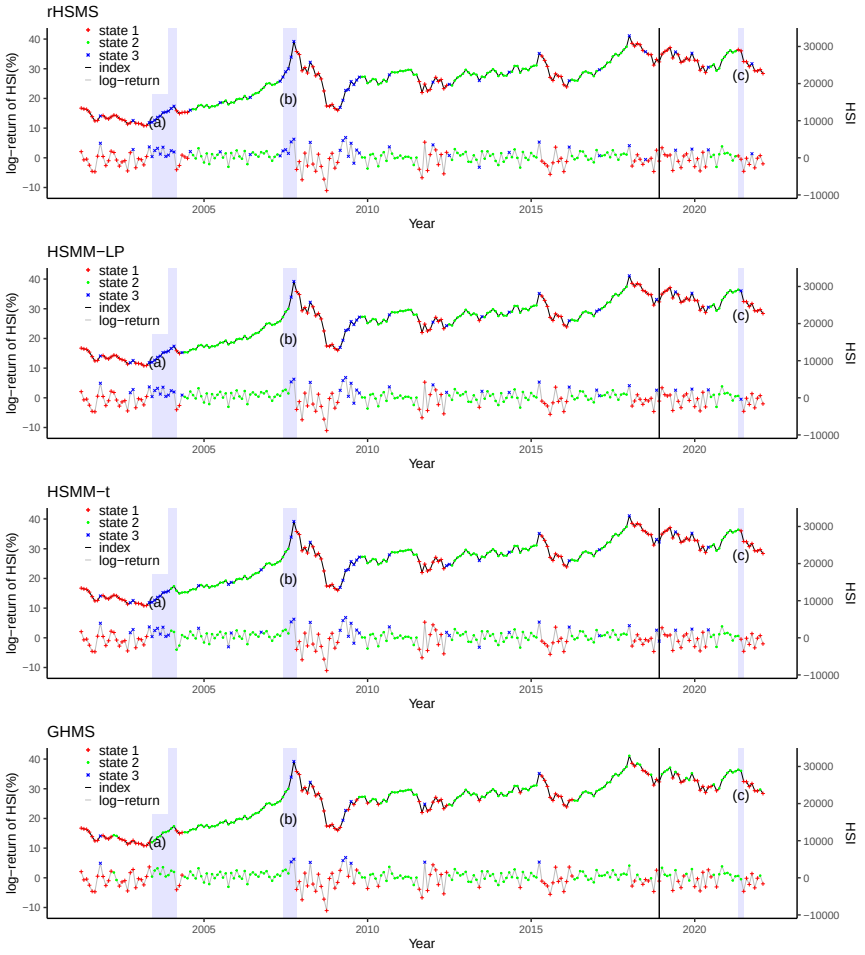

Figure 3.11: Three states estimation for monthly HSI, state 1 (red "+"), state 2 (green "•") and state 3 (blue "×") via the rHSMS, HSMS-LP and HSMS-t, and GHMS in the training set and test set, where the vertical line is the cut point at Jan. 1, 2019. In each panel, the upper plot is the monthly HSI index with estimated states (right axis), and the lower one is the log-returns of monthly HSI with estimated states (left axis).

Table 3.11: Estimated means ($\hat{\mu}_k$), standard deviations ($\hat{\sigma}_k$) and bootstrap standard errors in parentheses of two- or three-state RS models for HSI monthly returns (%).

State	Method	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_3$
$K=2$	rHSMS	-0.9739$_{(.59)}$	0.8429$_{(.20)}$	–	2.6857$_{(.36)}$	1.1970$_{(.23)}$	–
	HSMM-LP	-0.8400$_{(.53)}$	0.9159$_{(.22)}$	–	3.2072$_{(.41)}$	1.5339$_{(.30)}$	–
	HSMM-t	-1.0231$_{(.61)}$	0.8837$_{(.21)}$	–	2.5720$_{(.41)}$	1.2231$_{(.22)}$	–
	GHMS	-0.7581$_{(.66)}$	0.7508$_{(.25)}$	–	2.6047$_{(.37)}$	1.3514$_{(.29)}$	–
$K=3$	rHSMS	-1.3123$_{(.55)}$	0.5320$_{(.26)}$	2.1099$_{(.94)}$	2.4415$_{(.39)}$	1.2023$_{(.35)}$	1.7968$_{(.52)}$
	HSMM-LP	-1.1011$_{(.51)}$	0.6047$_{(.21)}$	1.5647$_{(.40)}$	3.1585$_{(.46)}$	1.2110$_{(.28)}$	1.6098$_{(.35)}$
	HSMM-t	-1.3156$_{(.64)}$	0.5648$_{(.25)}$	2.2249$_{(.94)}$	2.3951$_{(.47)}$	1.2117$_{(.30)}$	1.7727$_{(.51)}$
	GHMS	-0.7605$_{(.18)}$	0.6053$_{(.25)}$	2.0234$_{(.25)}$	2.4483$_{(.12)}$	0.9867$_{(.16)}$	0.9083$_{(.32)}$

values are 0.4706 and 0.7261 for two- and three-state rHSMS, respectively, implying that the independance nature of the residuals is well supported. Moreover, the three-state rHSMS is more adequacy in modeling HSI monthly returns due to the larger p-value. Residual plots and the ACF of residuals estimated by the three-state rHSMS are also given in Figure 3.10, providing supports for the results of BL test. Table 3.11 lists the estimated means ($\hat{\mu}_k$), standard deviations ($\hat{\sigma}_k$) and bootstrap standard errors in parentheses. The findings are similar to those for S&P500.

Hereafter, we opt for three-state RS models for further analysis, given its notably stronger modeling adequacy. Figure 3.11 illustrates the estimated three states for HSI monthly returns via these methods, with distinctive states highlighted in periods (a)-(c). During period (a), rHSMS, HSMM-LP, and HSMM-t successfully capture a sharp increase, while GHMS fails to do so. In period (b), rHSMS anticipates the regime change four months earlier than the other three methods. For the test set, the predicted states by rHSMS, HSMM-LP, and HSMM-t are nearly identical, except for period (c), which appears to represent a regime-switching point from state 2 to state 1. However, HSMM-LP fails to detect this transition. The state estimation performance of these methods is consistent with the two-state RS modeling for S&P500 monthly returns. This underscores that rHSMS demonstrates higher sensitivity than the other three methods in identifying changes in regimes.

Finally, we present the results of extreme values fitting and risk measures in Tables 3.12-3.13. Although GHMS has the largest p-value among different r values, its estimated $\hat{\phi}$ deviates the most from the empirical values. Conversely, $\hat{\phi}$ obtained via rHSMS and HSMM-t align more closely with the empirical values, accompanied by similar p-values for each r, except for $r = 2$ when rHSMS rejects the null at the significant level of 0.05, showing its inadequacy in capturing extreme values at this case. Concerning risk measures, akin

to S&P500 returns, the VaR and CTE values obtained by rHSMS consistently fall in the middle for all cases, providing a well-balanced estimation of risk without significant under or overestimation. In contrast, HSMM-t tends to underestimate the risk, while GHSM tends to overestimate it, as evidenced by exhibiting the smallest and largest VaR and CTE values at different quantiles.

Table 3.12: Empirical (ϕ) and estimated ($\hat{\phi}$) proportions of observations lying outside $(-r, r)$, and p-values from the test statistic W for HSI monthly returns using three-state RS models.

	Emperical		rHSMS		HSMM-LP		HSMM-t		GHMS	
r	ϕ	p-value	$\hat{\phi}$	p-value	$\hat{\phi}$	p-value	$\hat{\phi}$	p-value	$\hat{\phi}$	p-value
1	0.6692	–	0.6532	0.3857	0.6307	0.7143	0.6548	0.3473	0.7821	**1.0000**
1.5	0.5019	–	0.5113	0.2101	0.4797	0.4700	0.5132	0.2538	0.6835	**1.0000**
2	0.3954	–	0.3973	0.0445	0.3571	0.7313	0.3985	0.0707	0.5906	**1.0000**
2.5	0.2928	–	0.3058	0.2992	0.2631	0.6781	0.3055	0.2999	0.5051	**1.0000**

Table 3.13: Estimated VaR and CTE for HSI monthly returns given different quantiles.

Method	VaR				CTE			
	0.900	0.950	0.975	0.995	0.900	0.950	0.975	0.995
rHSMS	0.1964	0.3255	0.4540	0.5228	0.3724	0.4483	0.5444	0.5875
HSMM-LP	0.1691	0.2721	0.3588	0.4780	0.3154	0.3679	0.4495	0.5674
HSMM-t	-0.1518	0.0593	0.1733	0.2653	0.0813	0.2217	0.3042	0.4131
GHSM	0.9951	0.9962	0.9967	0.9973	0.9960	0.9969	0.9973	0.9977

3.6 Conclusion

In this chapter, we propose a robust hidden semi-Markov switching (rHSMS) model that relaxes the normality assumption on regime-dependent errors and accommodates potential outliers through the use of a Huber ρ-function within an M-estimation framework. The rHSMS model further extends the classical HMS framework by allowing flexible, non-geometric sojourn distributions. A likelihood-based estimation procedure is developed using the EM algorithm, and simulation results indicate that the EM algorithm exhibits slightly faster convergence under the HSMS setting than under the HMS setting.

Extensive simulation studies demonstrate that the proposed rHSMS model achieves superior performance in terms of parameter estimation accuracy and latent state prediction across

various sojourn distributions and error scenarios, including heavy-tailed and contaminated distributions. Among the competing methods, SC-HMM consistently shows the poorest performance, while rHSMS outperforms GHMS in most settings. When the sojourn distribution is geometric, the difference in state prediction accuracy between rHSMS and GHMS is small for both two- and three-state regime-switching models; however, in the three-state case, rHSMS still achieves higher parameter estimation accuracy.

Empirical applications to monthly returns of the S&P 500 and Hang Seng Index further validate the advantages of the proposed method. The similar estimation results obtained by rHSMS and GHMS for the two-state S&P 500 model suggest that the underlying hidden process may follow a geometric sojourn distribution. More importantly, rHSMS produces smaller bootstrap standard errors and provides a better fit to extreme observations in most cases, highlighting its robustness and interpretability in modeling market behavior. While rHSMS tends to underestimate risk for HSI monthly returns relative to GHMS, no definitive conclusion can be drawn for the S&P 500 risk measures.

Chapter 4

PCA-uCPD: An ensemble method for multiple change-point detection in moderately high-dimensional data

4.1 Introduction

As a matter of fact, the CPD methods in the above literature mainly focus on changes either in mean or covariance. Furthermore, few methods have been designed specifically for, or capable of handling, data with outliers or heavy-tailed distributions. Therefore, CPD is still a challenging problem for moderately high-dimensional complex data. A naive way is to consider the p-dimensional dataset as p univariate data sequences; thus, one can perform CPD on each of them. Proceeding in such a way, however, may dismiss some important information in the data, such as the correlation between any two variables.

This chapter proposes an integrated CPD method for complex data with medium to moderately high dimensions, high correlation, outliers, or heavy-tailed distribution. The method named PCA-uCPD, admits a flexible architecture by leveraging existing PCA and univariate CPD methods. Firstly, p-dimensional data sequences with correlations are transformed into uncorrelated sequences (principal components, PCs) with lower dimensions using PCA, which preserves the most statistical variability within the data. Subsequently, some existing univariate CPD methods are applied to these PCs individually, to identify potential *cps*. Moreover, we elaborate on a refining process and obtain the ultimate *cps*. For the proposed PCA-uCPD method, this chapter also provides theoretical insights into the method's guarantees. Numerically, we carry out extensive simulations under different scenarios and two real data analyses to examine the performance of our method.

Although our method looks similar to the 'Inspect' and 'Geometry' methods, their differences are given below. The 'Inspect' first projects a high-dimensional time series into a one-dimensional series by maximizing the difference in means. Then, it applies an existing univariate CPD method to the projected series. The 'Inspect' focuses on mean changes, which may lose information about changes in covariance. The 'Geometry' method reduces the p dimensions of sequences to two dimensions through distance and angle mappings that are respectively designed to identify distributional changes in mean and covariance. However, the proposed PCA-uCPD method projects a p-dimensional sequence to a low-dimensional one via PCA, without being restricted to one or two dimensions, as is in 'Inspect' or 'Geometry'. Our method can obtain more comprehensive insights into the data, which can simultaneously detect mixed changes both in mean and covariance. The flexible architecture of PCA-uCPD allows an arbitrary combination of any two PCA and univariate CPD methods, resulting in the capability to deal with complex data, such as moderately high dimensions, high correlations, outliers, and heavy-tailed distribution.

The rest of the Chapter 4 is organized as follows. In Section 4.2, we present technical preliminaries and the integrated PCA-uCPD method. We give theoretical justification in Section 4.3. We provide extensive simulation studies in Section 4.5 and two real data analyses in Section 4.6.

4.2 Proposed methods

We propose a two-step CPD method that integrates existing PCA with univariate change-point detection (uCPD) methods, abbreviated as PCA-uCPD. First, to address the high-dimensional problem, PCA transforms the p-dimensional data sequence with correlations into lower-dimensional, uncorrelated PCs, thereby preserving most of the statistical variability in the data. Subsequently, existing univariate CPD methods are applied to these PCs individually to identify potential *cps*. This section first reviews some classic PCA and uCPD methods. Then, we present a systematic description of different hybrid PCA-uCPD methods for detecting multiple *cps* in p-dimensional data sequences.

4.2.1 PCA

PCA is a key statistical technique for multivariate or moderately high-dimensional data analysis, which is widely used for dimension reduction and data interpretation (Pearson, 1901). Herein, we revisit the classic PCA and some well-known extended PCA methods.

CPCA Classic PCA (CPCA) applies the eigenvalue decomposition (EVD) algorithm to the covariance matrix of $\boldsymbol{X}$ (Jackson, 1991; Jolliffe, 2002). Suppose $\boldsymbol{X}$ is an $n \times p$ data matrix, where n is the number of observations, and p is the number of variables. Let $\boldsymbol{x}_j \in R^n$ be the j-th column of $\boldsymbol{X}$, $j = 1, \ldots, p$. For convenient presentation, suppose $\boldsymbol{x}_j$s are centred at zeros, i.e., $\boldsymbol{x}_j^\top \mathbf{1}_n = 0$. Denote $\boldsymbol{\Sigma}$ the covariance matrix of $\boldsymbol{X}$, λ_j and $\boldsymbol{a}_j \in R^p$ respectively the j-th eigenvalue and eigenvector of $\boldsymbol{\Sigma}$ such that $\lambda_1 \geq \cdots \geq \lambda_p$. Then the j-th PCs ($\boldsymbol{PC}_j$) can be obtained by a linear combination of the original variables, i.e., $\boldsymbol{PC}_j = \boldsymbol{X}\boldsymbol{a}_j$, $j = 1, \ldots, p$, where $\text{Var}(\boldsymbol{PC}_j) = \lambda_j$ and $\text{Cov}(\boldsymbol{PC}_j, \boldsymbol{PC}_k) = 0$ for $k \neq j$. PCA transforms the columns of $\boldsymbol{X}$ into another set of uncorrelated PCs, i.e., $\boldsymbol{P} = \boldsymbol{X}\boldsymbol{A}$, where $\boldsymbol{P} = [\boldsymbol{PC}_1, \ldots, \boldsymbol{PC}_p] \in R^{n \times p}$ and $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p] \in R^{p \times p}$, an orthogonal matrix. This transformation has the property that captures as much of the variation of the data as possible in the first few PCs, which provides a statistically optimal way of dimensionality reduction by projecting the original data onto a lower-dimensional orthogonal subspace, simply by discarding the last PCs (Andrecut, 2009). In addition to the EVD algorithm that uses the covariance matrix $\boldsymbol{\Sigma}$, some PCA algorithms use the data matrix $\boldsymbol{X}$ directly, such as Singular Value Decomposition (SVD), and Nonlinear Iterative Partial Least Squares (nipals). nipals computes the components sequentially, while it suffers from loss of orthogonality for large data matrices or matrices that have a high degree of column collinearity.

SPCA However, CPCA suffers from the fact that each PC is a linear combination of all original variables; thus it is often difficult to interpret the results. Zou et al. (2006) introduced a sparse PCA (SPCA) using the lasso or elastic net to produce modified PCs with sparse loadings. The authors showed that SPCA can be formulated as a regression-type optimization problem; sparse loadings are obtained by imposing the lasso or elastic net constraints on the regression coefficients. Consequently, SPCA can reduce the number of explicitly used variables, resulting in balanced variance explained (dimension reduction) and sparse loadings (interpretability) (Zou and Xue, 2018). SPCA can be implemented by the R package `nsprcomp`.

RPCA CPCA and its applicability in real scenarios are limited by a lack of robustness to outlying or corrupted observations (Wright et al., 2009). Wright et al. (2009) and Candès et al. (2011) proposed idealized versions of robust PCA (RPCA) for $\boldsymbol{X}$, even when a positive fraction of its entries are outliers or arbitrarily corrupted. Assume that $\boldsymbol{X} = \boldsymbol{L} + \boldsymbol{S}$, where $\boldsymbol{L} \in R^{n \times p}$ is a low-rank matrix with some corrupted entries, and $\boldsymbol{S} \in R^{n \times p}$ is a sparse matrix represented as an additive error. Candès et al. (2011) proposed to recover both the low-rank and the sparse components exactly by solving a very convenient convex program called principal component pursuit. Solving this problem suggests recovering the principal

components of the data matrix $\boldsymbol{X}$ even if some of the entries in the matrix are corrupt or outliers. RPCA can be implemented by the R package `rpca`.

RSPCA RPCA yields robust analysis resistant to outlying observations, while SPCA delivers PCs with sparse loadings on a small number of variables, thereby enhancing interpretability. Building upon these advancements, Croux et al. (2013) proposed a modified PCA method called RSPCA, which yields both sparse and robust results simultaneously. The authors introduced an algorithm to compute sparse and robust PCs that correspond to directions maximizing a robust measure of variance, with an additional penalty term to account for sparsity. The algorithm computes the components sequentially, enabling it to handle high-dimensional datasets. RSPCA can be implemented using the R package `pcaPP`.

4.2.2 Univariate change-point detection (uCPD) methods

Given a collection of time-ordered observations, CPD aims to detect whether a change occurs in the underlying distribution of the observations, and where it is if it occurs. Let $x_i \sim F_i(x)$, $i = 1, 2, \ldots, n$, be a sequence of n independent univariate random observations, where $F_i(\cdot)$ denotes the distribution function of X_i. A single CPD problem can be defined as a hypothesis test with H_0: there is no change-point, and H_1: there is a single change-point, i.e.,

$$
\begin{aligned}
H_0 &: F_1(x) = F_2(x) = \cdots = F_n(x), \\
H_1 &: F_1(x) = \cdots = F_{k^*}(x) \neq F_{k^*+1}(x) = \cdots = F_n(x),
\end{aligned}
$$

where $1 < k^* < n$ is an unknown location that needs to be identified. For a multiple CPD problem, there is more than one *cp*. Let K be the unknown number of *cps* and k_i, be their unknown locations, $i = 1, 2, \ldots, K, K > 1$. Define $k_0 = 1$ and $k_{K+1} = n + 1$, and $k_i < k_j$ for $i < j$.

There are various mainstream and specialized CPD methods for single and multiple *cps* within data, such as binary segmentation (Scott and Knott, 1974; Sen and Srivastava, 1975), segment neighbourhoods (Auger and Lawrence, 1989; Bai and Perron, 1998), and pruned exact linear time (PELT) (Killick et al., 2012), etc. These methods can be implemented to identify *cps* in a given set of summary statistics for a specified cost function and penalty, using the R package `changepoint`. For multiple CPD problems, it is critical to estimate the number of *cps* appropriately. Zou et al. (2020) proposed a data-driven selection criterion to select the number of *cps*, which is integrated with multiple change-searching algorithms for a variety of frequently considered parametric change-point models. The improved algorithm

can be implemented using the R package `cpss`.

4.2.3 Proposed PCA-uCPD methods

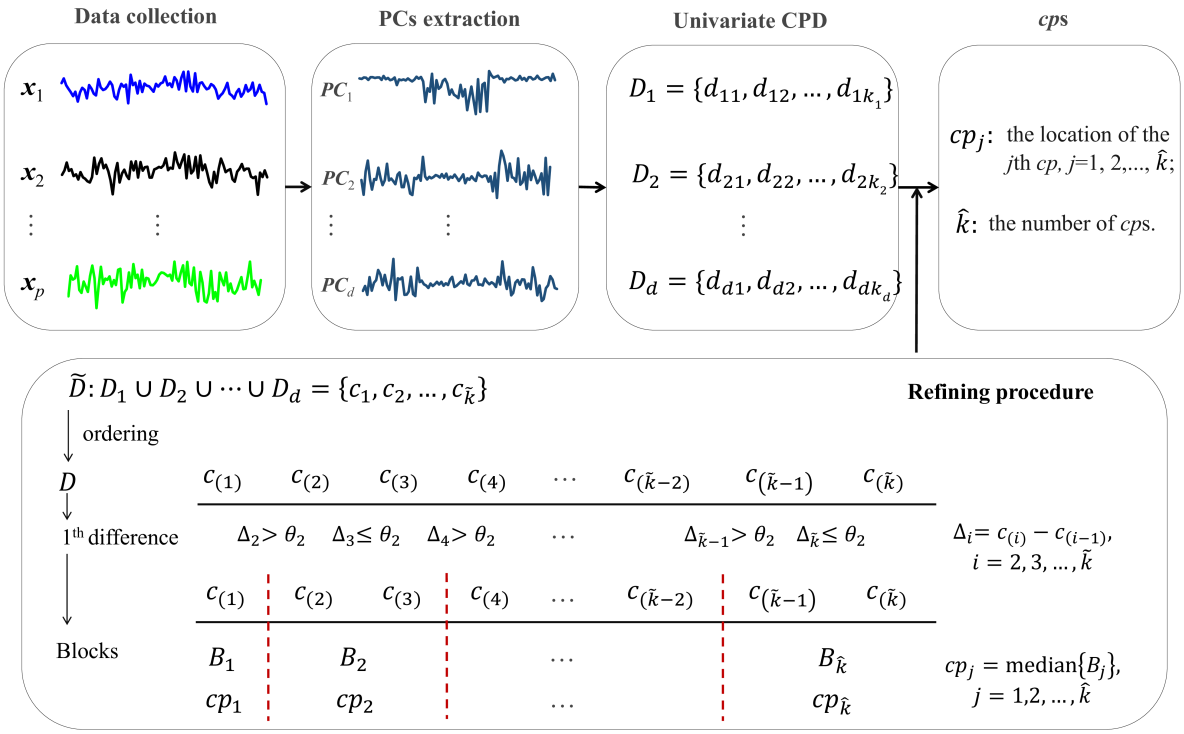

Figure 4.1: The illustrated procedure of the PCA-uCPD method.

The proposed PCA-uCPD methods integrate PCA and univariate CPD techniques, yielding a flexible architecture that allows data to involve more intricate circumstances, such as outliers. The procedure for identifying *cps* is shown in Figure 4.1, which intuitively summarizes the following four steps.

Step 1: Given a collection of dataset $\boldsymbol{X} \in R^{n \times p}$, project $\boldsymbol{X}$ using a PCA method, such that the proportion of variance explained by PCs is larger than a prespecified threshold θ_1 $(0 < \theta_1 < 1)$. Then extract d PCs, $\boldsymbol{PC}_j, j = 1, \dots, d$.

Step 2: Perform multiple CPD on PCs individually by some existing methods to obtain k_j potential *cps* for each $\boldsymbol{PC}_j$, denoted as $D_j = \{d_{j1}, \dots, d_{jk_j}\}$, $j = 1, \dots, d$.

Step 3: Take the union of D_j for $j = 1, \dots, d$, i.e., $\tilde{D} = D_1 \cup \cdots \cup D_d = \{c_1, \dots, c_{\tilde{k}}\}$, where $\tilde{k}$ denotes the number of elements of $\tilde{D}$. Then sort all the elements in $\tilde{D}$ in an increasing order, that is $D = \{c_{(1)}, \dots, c_{(\tilde{k})}\}$.

Step 4: Refine D. If two adjacent points in D satisfy $c_{(j)} - c_{(j-1)} > \theta_2 \in N^+$ (a pre-specified threshold), then set $c_{(j-1)}$ as the right endpoint of one block and set $c_{(j)}$ as the left endpoint of the next block. Subsequently, D is separated into different blocks, say $\{B_1, \ldots, B_{\hat{k}}\}$, and $\hat{k}$ is the estimated number of *cps*. We use the median of the jth block as the final jth change point estimate, i.e., $cp_j = \text{median}\{B_j\}$ for $j = 1, \ldots, \hat{k}$.

Remark 3 *In the proposed method, two thresholds need to be prespecified. θ_1 is the proportion of total variance explained by the extracted PCs, which controls the number of PCs. We take $\theta_1 = 0.8$ in the simulations and $\theta_1 = 0.5$ in the real data analysis. The refining process removes duplicate cps detected from different PCs, and combines those cps that are close enough. θ_2 defines the minimum distance between two neighbouring cps. Herein, we choose $\theta_2 = 10$, the same way used in the method 'Geometry' Grundy et al. (2020).*

This chapter presents a systematic evaluation of different ensemble (i.e., hybrid) methods to identify *cps*. In Step 1, we choose several well-known PCA methods, such as CPCA, RPCA, SPCA, and RSPCA. These methods are designed to deal with complex data having moderately high dimensions, high correlations, outliers or heavy-tailed distributions. In Step 2, we apply some existing multiple CPD methods for univariate data sequences. These methods can be implemented by the R packages `changepoint` (abbreviated as CP) and `cpss`. According to Van and Williams (2020), the binary segmentation method in `changepoint` performed best on average out of the 14 algorithms that were examined with 37 time series from different application domains. For convenience, we utilize the notation of A-B to denote a series of ensemble methods, where A and B are respectively the PCA and uCPD methods. Specifically, we evaluate the following hybrid methods: CPCA-CP, CPCA-CPSS, SPCA-CP, SPCA-CPSS, RPCA-CP, RPCA-CPSS, RSPCA-CP, and RSPCA-CPSS.

4.3 Theoretical justification

We provide the theoretical justification of the proposed method. Let $\boldsymbol{X}_i \sim F_i(\boldsymbol{x})$, $i = 1, 2, \ldots, n$, be a p-dimensional sequence of independently distributed random vectors, where $F_i(\cdot)$ denotes the distribution function. If there exists a k^* such that

$$F_i(\boldsymbol{x}) = \begin{cases} F_1(\boldsymbol{x}) & \text{if } 1 \leq i \leq k^*, \\ F_n(\boldsymbol{x}) & \text{if } k^* < i \leq n \end{cases}$$

with $F_1(\boldsymbol{x}) \neq F_n(\boldsymbol{x})$, then we say k^* is a change-point. The corresponding hypotheses are:

$$\begin{aligned} &H_0 : F_1(\boldsymbol{x}) = F_2(\boldsymbol{x}) = \cdots = F_n(\boldsymbol{x}), \\ &\quad vs \\ &H_1 : F_1(\boldsymbol{x}) = \cdots = F_{k^*}(\boldsymbol{x}) \neq F_{k^*+1}(\boldsymbol{x}) = \cdots = F_n(\boldsymbol{x}), \end{aligned}$$

where $1 < k^* < n$ is unknown. We need to find out from the data sequence whether or not there is a change-point and where it is if it exists.

Theorem 4.3.1 *Suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independently distributed p-dimensional random vectors. Denote $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)^\top$. Let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$ be a $p \times p$ matrix. Then the following statements hold true:*

(a) Under H_0, there is no change-point in $\{\boldsymbol{a}_j^\top \boldsymbol{X}_1, \ldots, \boldsymbol{a}_j^\top \boldsymbol{X}_n\}$ for any given j, $j = 1, \ldots, p$.

(b) If there is a j such that there exits a change-point $k^* \in \{1, \ldots, n\}$ in the data sequence $\{\boldsymbol{a}_j^\top \boldsymbol{X}_1, \ldots, \boldsymbol{a}_j^\top \boldsymbol{X}_n\}$, i.e., the common distribution of $\{\boldsymbol{a}_j^\top \boldsymbol{X}_1, \ldots, \boldsymbol{a}_j^\top \boldsymbol{X}_k\}$ is different from the common distribution of $\{\boldsymbol{a}_j^\top \boldsymbol{X}_{k+1}, \ldots, \boldsymbol{a}_j^\top \boldsymbol{X}_n\}$, then H_1 holds true.

Proof. The jth column of $\boldsymbol{XA}$ is

$$(\boldsymbol{a}_j^\top \boldsymbol{X}_1, \ldots, \boldsymbol{a}_j^\top \boldsymbol{X}_n)^\top.$$

Under H_0, denote the characteristic function of $\boldsymbol{X}_\ell$ by $\varphi(\boldsymbol{t})$. The characteristic function of $\boldsymbol{a}_j^\top \boldsymbol{X}_\ell$ is then

$$Ee^{\mathrm{i}t\boldsymbol{a}_j^\top \boldsymbol{X}_\ell} = \varphi(t\boldsymbol{a}_j),$$

where $\mathrm{i} = \sqrt{-1}$. We conclude (a) $\boldsymbol{a}_j^\top \boldsymbol{X}_\ell, \ell = 1, \ldots, n$, have the same characteristic function. Clearly, (b) can be obtained from (a) straightforwardly. We complete the proof. ■

Remark 4 *The condition that there exists a $p \times p$ matrix $\boldsymbol{A}$ such that all the elements in each column of $\boldsymbol{XA}$ have the same distribution cannot warrant that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ have the same distribution.*

Theorem 4.3.2 *Suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independently normally distributed p-dimensional random vectors. Assume that H_1 holds true, i.e., there exists a change-point $k^* \in \{1, \ldots, n-1\}$ such that the common distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k^*}$ are different from the common distribution of $\boldsymbol{X}_{k^*+1}, \ldots, \boldsymbol{X}_n$. Let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$ be a $p \times p$ full rank matrix such that all the elements in each row of $\boldsymbol{XA}$ are uncorrelated. Then there exists a column of $\boldsymbol{XA}$ such that the common distribution of its first k^* elements are different from the common distribution of its remaining $n - k^*$ elements.*

Proof. Under H_1, there is one change-point at k^*. Denote the characteristic function of $\boldsymbol{X}_\ell$, $\ell = 1, \ldots, k^*$, by $\varphi_1(\boldsymbol{t})$ and the characteristic function of $\boldsymbol{X}_\ell$, $\ell = k^* + 1, \ldots, n$, by $\varphi_n(\boldsymbol{t})$. Then $\varphi_1(\boldsymbol{t}) \neq \varphi_n(\boldsymbol{t})$.

We now prove the theorem by contradiction method. Assume that all the elements of any column of $\boldsymbol{X}\boldsymbol{A}$ have the same distribution, i.e., there is no change-point in any column of $\boldsymbol{X}\boldsymbol{A}$. By the definitions of $\varphi_1(\boldsymbol{t})$ and $\varphi_n(\boldsymbol{t})$, the characteristic function of the ℓth element of the jth column of $\boldsymbol{X}\boldsymbol{A}$ is

$$E\left(\exp\left\{\mathrm{i}t(\boldsymbol{a}_j^\top \boldsymbol{X}_\ell)\right\}\right) = \begin{cases} \varphi_1(t\boldsymbol{a}_j), & \text{for } 1 \leq \ell \leq k^*, \\ \varphi_n(t\boldsymbol{a}_j), & \text{for } k^* + 1 \leq \ell \leq n. \end{cases} \tag{4.3.1}$$

Since it is assumed that all the elements of any column of $\boldsymbol{X}\boldsymbol{A}$ have the same distribution, it follows that for any t,

$$\varphi_1(t\boldsymbol{a}_j) = \varphi_n(t\boldsymbol{a}_j), \quad 1 \leq j \leq n. \tag{4.3.2}$$

As $\boldsymbol{A}$ is of full rank, any fixed p-dimensional vector $\boldsymbol{b}$ can be expressed as

$$\boldsymbol{b} = (b_1, \ldots, b_p)^\top = c_1\boldsymbol{a}_1 + \ldots + c_p\boldsymbol{a}_p.$$

Note that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independently normally distributed. Thus, by the assumption that all the elements in each row of $\boldsymbol{X}\boldsymbol{A}$ are uncorrelated, they are in fact independent. Thus,

$$\varphi_1(\boldsymbol{b}) \;=\; \varphi_1\left(\sum_{j=1}^{p} c_j\boldsymbol{a}_j\right) = E\left(\exp\left\{\mathrm{i}\sum_{j=1}^{p} c_j\boldsymbol{a}_j^\top \boldsymbol{X}_1\right\}\right) = \prod_{j=1}^{p} \varphi_1(c_j\boldsymbol{a}_j).$$

Similarly, it can be shown that $\varphi_n(\boldsymbol{b}) = \prod_{j=1}^{p} \varphi_n(c_j\boldsymbol{a}_j)$. By (4.3.2), it follows that $\varphi_1(\boldsymbol{b}) = \varphi_n(\boldsymbol{b})$ for any p-dimensional vector $\boldsymbol{b}$. It is a contradiction because $\varphi_1 \neq \varphi_n$. Therefore, there exists a column of $\boldsymbol{X}\boldsymbol{A}$ such that the distribution of its first k^* elements are different from the distribution of its remaining $n - k^*$ elements.

■

Remark 5 *Theorem 4.3.2 may not hold if the normality assumption is removed. The proof relies on the property that, for multivariate normal distributions, uncorrelated random variables are independent. This property justifies the factorization of the objective function used in the proof. However, for multivariate Student's (t) distributions and other non-normal distributions, uncorrelatedness does not imply independence. Consequently, the factorization argument is no longer valid, and the theoretical guarantees of Theorem 4.3.2 may fail. Al-*

though the normality assumption is essential for the theoretical development, our numerical studies indicate that the proposed method remains robust in practice and performs well for complex data, including heavy-tailed distributions and datasets contaminated by outliers.

4.4 Statistical test on single CPD

The single CPD problem can be regarded as a hypothesis test with H_0: there is no change-point, and H_1: there is a single change-point. In the following, we conduct simulations to assess our method's performance in terms of type I error under H_0 and power under H_1. Under H_0, we generate data matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^\top$ under four types of settings as follows.

- Type A: $\boldsymbol{X}_i \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ with $\boldsymbol{\mu}_1 = \boldsymbol{0}$, $\boldsymbol{\Sigma}_1 = \boldsymbol{D}_1 \boldsymbol{S}_{0.9} \boldsymbol{D}_1$, $i = 1, \ldots, 300$, where $\boldsymbol{D}_{\rho_1} = \rho_1 \boldsymbol{I}$ and $\boldsymbol{S}_{\rho_2}$ is a positive matrix with principal diagonals being ones and the rests being ρ_2 $(0 < \rho_2 < 1)$. The covariance matrix can be constructed from $\boldsymbol{D}_{\rho_1}$ and $\boldsymbol{S}_{\rho_2}$, i.e., $\boldsymbol{\Sigma} = \boldsymbol{D}_{\rho_1} \boldsymbol{S}_{\rho_2} \boldsymbol{D}_{\rho_1}$. Thus, $\boldsymbol{D}_{\rho_1} = \sqrt{\text{diag}(\boldsymbol{\Sigma})}$ is the standard deviation of all variables, and $\boldsymbol{S}_{\rho_2}$ is the correlation matrix.

- Type B: contaminate the data matrix $\boldsymbol{X}$ (Type A) by randomly sampling 200 entries and replacing them by multiplying the original entries with five.

- Type C: $\boldsymbol{X}_i \sim \boldsymbol{t}_\nu(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2)$ with $\boldsymbol{\mu}_1 = \boldsymbol{0}$, $\boldsymbol{\Sigma}_2 = \boldsymbol{D}_{0.5} \boldsymbol{S}_{0.9} \boldsymbol{D}_{0.5} \frac{\nu-2}{\nu}$, $\nu = 3$, $i = 1, \ldots, 300$.

- Type D: contaminate the data matrix $\boldsymbol{X}$ (Type C) by randomly sampling 200 entries and replacing them by multiplying the original entries with five.

Table 4.1: Simulated Type I errors via PCA-CP and RPCA-CP for four types of data with different dimensions under H_0.

Type	Method	$p = 5$	$p = 10$	$p = 30$	$p = 50$
A	PCA-CP	0.001	0.001	0.001	0.007
	RPCA-CP	0.000	0.000	0.000	0.001
B	PCA-CP	0.914	0.998	1.000	1.000
	RPCA-CP	0.000	0.000	0.000	0.000
C	PCA-CP	0.695	0.899	0.998	1.000
	RPCA-CP	0.001	0.002	0.052	0.093
D	PCA-CP	0.990	1.000	1.000	1.000
	RPCA-CP	0.000	0.001	0.035	0.089

Given the presence of atypical values in the generated data, such as heavy-tailed distributions and/or outliers, we apply two representative methods: PCA-CP and RPCA-CP. Table 4.1 presents their simulated type I errors, which are estimated by the proportion of rejecting H_0 (false detections of *cps*) across 1,000 trials. For multivariate normal data (Type A), both PCA-CP and RPCA-CP exhibit excellent performance, with type I errors approaching zero, indicating that neither method falsely rejects the null hypothesis (H_0) when there is no change point. However, when dealing with data that contain atypical values (Types B-D), PCA-CP shows a significant increase in type I errors, highlighting its sensitivity to outliers and/or heavy-tailed distributions. In contrast, RPCA-CP demonstrates robustness, with type I errors below or near 0.05 in the majority of cases (10/12). Yet, when confronted with heavy-tailed distributions (Types C-D), particularly in moderately high-dimensional settings ($p = 50$), the type I errors of RPCA-CP exceed 0.05 but remain below 0.1. This indicates a decline in its performance when dealing with high-dimensional data that are heavy-tail dsitributed.

To investigate powers under H_1, we consider two scenarios of *cps*, involving a single distributional change in mean (Scenario 1) or variance (Scenario 2). Synthetic data are generated as follows.

- Scenario 1: there is a single distributional change in mean.
  - Type A: $\boldsymbol{X}_i \sim \begin{cases} N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), i = 1, \ldots, 150, \\ N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}), i = 151, \ldots, 300, \end{cases}$ where $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\mu}_2 = \mathbf{0.5}$, $\boldsymbol{\Sigma} = \boldsymbol{D}_{0.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{0.5}$.
  - Type B: randomly generate 200 corrupted entries of $\boldsymbol{X}$ for Type A.
  - Type C: $\boldsymbol{X}_i \sim \begin{cases} \boldsymbol{t}_\nu(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), i = 1, \ldots, 150, \\ \boldsymbol{t}_\nu(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}), i = 151, \ldots, 300, \end{cases}$ where $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\mu}_2 = \mathbf{0.5}$, $\boldsymbol{\Sigma} = \boldsymbol{D}_{0.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{0.5}\frac{\nu-2}{\nu}$, $\nu = 3$.
  - Type D: randomly generate 200 corrupted entries of $\boldsymbol{X}$ for Type C.
- Scenario 2: there is a single distributional change in variance.
  - Type A: $\boldsymbol{X}_i \sim \begin{cases} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1), i = 1, \ldots, 150, \\ N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2), i = 151, \ldots, 300, \end{cases}$ where $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma}_1 = \boldsymbol{D}_{0.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{0.5}$, $\boldsymbol{\Sigma}_2 = \boldsymbol{D}_{1.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{1.5}$.
  - Type B: randomly generate 200 corrupted entries of $\boldsymbol{X}$ for Type A.
  - Type C: $\boldsymbol{X}_i \sim \begin{cases} \boldsymbol{t}_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1), i = 1, \ldots, 150, \\ \boldsymbol{t}_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2), i = 151, \ldots, 300, \end{cases}$ where $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma}_1 = \boldsymbol{D}_{0.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{0.5}\frac{\nu-2}{\nu}$, $\boldsymbol{\Sigma}_2 = \boldsymbol{D}_{1.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{1.5}\frac{\nu-2}{\nu}$, $\nu = 3$.

- Type D: randomly generate 200 corrupted entries of $\boldsymbol{X}$ for Type C.

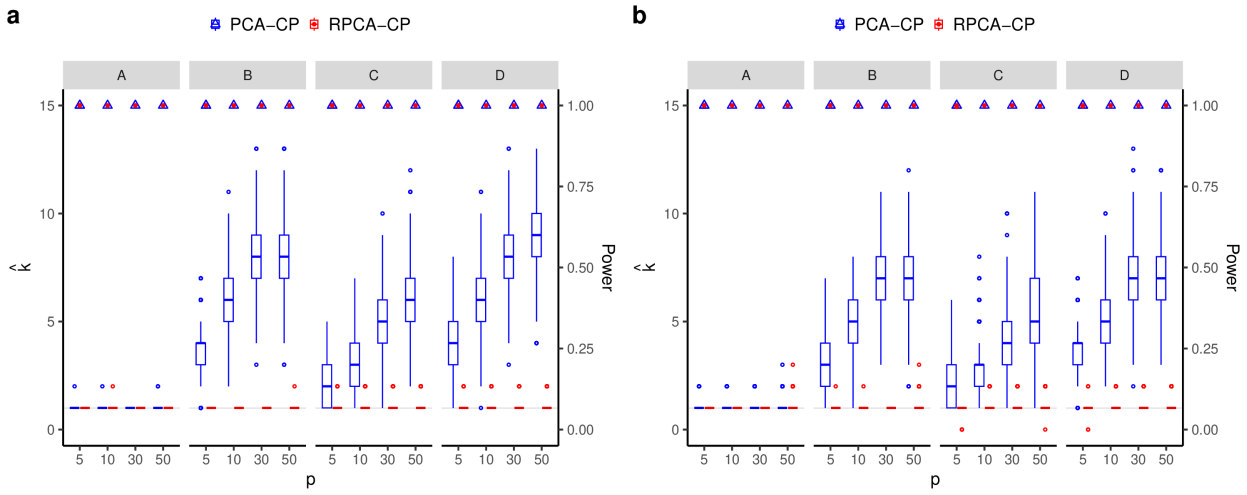

Figure 4.2: The simulated powers (right-axis) and the estimated number of *cps* (left-axis) via PCA-CP and RPCA-CP for four types of data with different dimensions under H_1 (a) Scenario 1: there is a single distributional change in mean; (b) Scenario 2: there is a single distributional change in variance.

Figure 4.2 illustrates the simulated powers and the estimated number of *cps*, $\hat{k}$, obtained through PCA-CP and RPCA-CP for both distributional changes. The simulated powers are calculated as the proportion of false detections of *cps* across 1,000 trials under the alternative hypothesis (H_1). It is evident that both methods exhibit high power in CPD across all data types and scenarios, with values equal to or approaching one. However, PCA-CP tends to overestimate the number of *cps* when data include atypical values (Types B-D), resulting in $\hat{k}$ values greater than one. In contrast, RPCA-CP provides more accurate estimates of the number of *cps*, with the estimated $\hat{k}$ being close to one in the majority of cases.

4.5 Simulation studies on multiple CPD

In this section, we carry out simulation studies to examine the CPD performance of our proposed methods for multivariate and high-dimensional data. Recently, Li and Zhang (2024) reviewed existing R packages for CPD. For low-dimensional data, we select benchmark methods, such as 'E.divisive' (Matteson and James, 2014), 'KCP' (Arlot et al., 2019), and 'changepoints' (Xu et al., 2021), which are used to detect multiple *cps* for multivariate data with both mean and variance changes and are outperformance in their study. The Binary Segmentation (BS) method, one of the most popular CPD algorithms, is adopted by the R package `changepoints`. For high-dimensional data, Zhang and Drikvandi (2023) reviewed

the five most recent CPD methods, among which 'E.division' and 'Geometry' have good performance. We thus compare our methods with 'E.divisive' (Matteson and James, 2014) and 'Geometry' (Grundy et al., 2020). 'E.divisive' can be implemented by the R package `ecp`, while the 'Geometry' can be implemented by the R package `changepoint.geo`.

4.5.1 Evaluation Metrics

Before proceeding, we briefly introduce some common metrics that are used to evaluate the performance of CPD methods, such as the percentages of successful detection of the locations of *cps* (Acc) and the number of *cps* ($Count$)(Arbelaez et al., 2010). Denote the set of true *cps* as $\mathcal{L} = \{l_1, \ldots, l_K\}$ with $l_i \in [2, n-1]$ and $l_i < l_j$ for $i < j$, $i, j = 1, \ldots, K$, where K is the true number of *cps*. We denote the set of estimated *cps* in the mth simulation by $\mathcal{L}^{(m)}$ and define, similar to Shi et al. (2016),

$$\mathcal{A}_{l_i}^{(m)} = \mathcal{L}^{(m)} \cap [l_i - \theta_2, l_i + \theta_2], m = 1, \ldots, M, \tag{4.5.1}$$

as the set of estimated *cps* lying in the neighbourhood of the ith true one. Moreover, we define the function $J(\cdot)$ as

$$J(\mathcal{A}) = \begin{cases} 0, & \text{if } \mathcal{A} = \emptyset, \\ 1, & \text{otherwise.} \end{cases} \tag{4.5.2}$$

Furthermore, we define $B^{(m)} = 1$ if $J\left(\mathcal{A}_{l_i}^{(m)}\right) = 1$ for all $i = 1, \ldots, K$; $B^{(m)} = 0$, otherwise. Note that $B^{(m)} = 1$ if and only if the mth simulation is successful in the sense that in each neighbourhood of a true *cp*, there is an estimated *cp*. Then, we define

$$Acc = \sum_{m=1}^{M} B^{(m)}/M,$$

the percentage of successful detection of the true *cps*. Moreover, an overestimation method that estimates too many *cps* including all the true ones will result in a high Acc value. We thus use a second evaluation metric to measure the number of *cps*. Define $Q(.)$ as

$$Q(\hat{k}_m) = \begin{cases} 1, & \text{if } \hat{k}_m = K, \\ 0, & \text{otherwise,} \end{cases} \tag{4.5.3}$$

where $\hat{k}_m$ is the estimated number of *cps* in the mth simulation. Then, $Count$ can be defined as

$$Count = \sum_{m=1}^{M} Q(\hat{k}_m)/M.$$

Herein, $M = 500$.

CPD inherently divides a data sequence into a partition (Van and Williams, 2020), where *cps* serve as the cutting points. For an ordered sequence with a length of n, let $\mathcal{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_m\}$ be the partition of $[1, n]$ associated with cutting points being the underlying true *cps*, and $S = \{S_1, \ldots, S_k\}$ be a partition where the cutting points are estimated by a CPD method. The covering metric of $\mathcal{G}$ and $\mathcal{S}$ can be defined as (Arbelaez et al., 2010),

$$Cover(\mathcal{G}, \mathcal{S}) = \frac{1}{n} \sum_{\mathcal{B}_j \in \mathcal{G}} \mid \mathcal{B}_j \mid \max_{\mathcal{S}_\ell \in \mathcal{S}} J(\mathcal{B}_j, \mathcal{S}_\ell),$$

where the Jaccard index $J(\mathcal{B}_j, \mathcal{S}_\ell)$ is also known as Intersection over Union, i.e., $\frac{|\mathcal{B}_j \cap \mathcal{S}_\ell|}{|\mathcal{B}_j \cup \mathcal{S}_\ell|}$. All three evaluation metrics range from zero to one, with higher values indicating better performance.

4.5.2 Performance comparison on multivariate data

We carry out methods comparison on multivariate data that are generated via the following model,

$$\boldsymbol{x}_i \sim \begin{cases} N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), & i = 1, \ldots, 100, \\ N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), & i = 101, \ldots, 200, \\ N_p(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3), & i = 201, \ldots, 300. \end{cases} \tag{4.5.4}$$

Three types of distributional parameters are given as follows,

- Type A: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 = \mathbf{0}$, $\boldsymbol{\mu}_2 = 0.5$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_3 = \boldsymbol{D}_1$, $\boldsymbol{\Sigma}_2 = \boldsymbol{S}_{0.9}$.
- Type B: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \mathbf{0}$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_3 = \boldsymbol{D}_1$, $\boldsymbol{\Sigma}_2 = \boldsymbol{D}_{1.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{1.5}$.
- Type C: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$, $\boldsymbol{\mu}_3 = \mathbf{0}.5$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_3 = \boldsymbol{D}_1$, $\boldsymbol{\Sigma}_2 = \boldsymbol{D}_{1.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{1.5}$.

According to the three types of generation mechanisms, it is easy to see that there are two *cps* located at 100 and 200 for each scenario. For multivariate data, we take $p = 5, 10$, and

consider these six ensemble methods, such as Naive-CP, Naive-CPSS, CPCA-CP, CPCA-CPSS, RPCA-CP, RPCA-CPSS, and three existing CPD methods designed for multivariate data, i.e., 'E.divisive', 'KCP' and 'chengepoints'.

Table 4.2: Comparison of evaluation metrics for multivariate data under different scenarios.

Evaluation Metrices	Method	Scenario (p, Type)					
		(5, A)	(5, B)	(5, C)	(10, A)	(10, B)	(10, C)
Acc	Naive-CP	0.0100	0.0200	0.2300	0.0000	0.0900	0.4000
	Naive-CPSS	0.5000	0.6500	0.6500	0.6600	0.7900	0.7500
	CPCA-CP	0.9800	0.9900	**1.0000**	**1.0000**	**0.9900**	**1.0000**
	CPCA-CPSS	**1.0000**	**1.0000**	0.9700	**1.0000**	**0.9900**	**1.0000**
	RPCA-CP	0.9400	0.9600	0.9800	0.9900	0.9800	**1.0000**
	RPCA-CPSS	0.8400	0.8700	0.9100	0.9900	0.9800	**1.0000**
	E.divisive	0.8100	0.6100	0.6800	0.8500	0.7900	0.7500
	KCP	0.9100	0.6100	0.7900	0.0500	0.0000	0.1500
	changepoints	0.8500	0.8900	0.9400	0.9000	0.8500	0.9000
Count	Naive-CP	0.0100	0.0100	0.3000	0.0100	0.0800	0.3300
	Naive-CPSS	0.0300	0.0900	0.4000	0.0100	0.0500	0.1300
	CPCA-CP	0.8000	0.9200	0.9000	0.6700	0.7800	0.8100
	CPCA-CPSS	0.2700	0.2100	0.2200	0.1100	0.0700	0.0300
	RPCA-CP	**0.9900**	**0.9700**	**0.9800**	**1.0000**	**1.0000**	0.9700
	RPCA-CPSS	0.8800	0.9000	0.9300	**1.0000**	0.9800	**0.9900**
	E.divisive	0.9400	0.7900	0.8300	0.9000	0.8400	0.9200
	KCP	0.9300	0.6400	0.8300	0.0500	0.0000	0.1600
	changepoints	0.6300	0.0200	0.0500	0.9800	0.9600	0.9600
Cover	Naive-CP	0.0250	0.0610	0.5825	0.0396	0.1772	0.7296
	Naive-CPSS	0.7336	0.7768	0.8395	0.6579	0.7338	0.8026
	CPCA-CP	0.9649	0.9746	0.9752	0.9517	0.9683	0.9700
	CPCA-CPSS	0.8955	0.8868	0.8848	0.8272	0.8035	0.7815
	RPCA-CP	**0.9796**	**0.9816**	**0.9809**	**0.9865**	**0.9896**	0.9885
	RPCA-CPSS	0.9458	0.9529	0.9596	**0.9865**	0.9885	**0.9888**
	E.divisive	0.9416	0.8727	0.8978	0.9437	0.9275	0.9246
	KCP	0.9308	0.6638	0.9130	0.0488	0.0000	0.2224
	changepoints	0.9126	0.8181	0.8291	0.9683	0.9632	0.9652

The largest values for each evaluation metric are in boldface

The results are shown in Table 4.2 and Figure 4.3. Table 4.2 lists the *Acc*, *Count*, and *Cover* metrics with different types of *cps* for different p. Among nine comparative methods, CPCA-CP, CPCA-CPSS, and RPCA-CP show superior performance in detecting all true *cps*, reflected by their larger *Acc*s for different scenarios. In terms of *Count* and *Cover*, RPCA-CP outperforms other methods in most cases. However, CPCA-CP and CPCA-CPSS

Table 4.3: Comparison of average evaluation metrics for multivariate data with outliers under different scenarios.

Evaluation Metrices	Method	Scenario (p, Type)					
		(5, A)	(5, B)	(5, C)	(10, A)	(10, B)	(10, C)
Acc	Naive-CP	0.0800	0.2200	0.3300	0.4100	0.5300	0.6500
	Naive-CPSS	0.0900	0.2700	0.2100	0.4800	0.6100	0.6600
	CPCA-CP	0.4000	0.7700	0.7900	0.9000	1.0000	0.9900
	CPCA-CPSS	0.4100	0.7200	0.7300	0.9000	0.9600	0.9600
	RPCA-CP	**0.9800**	**0.9900**	**0.9700**	**1.0000**	**0.9900**	**1.0000**
	RPCA-CPSS	0.8600	0.8900	0.9300	1.0000	0.9900	1.0000
	E.divisive	0.5900	0.2200	0.1800	0.7900	0.4300	0.4300
	KCP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	changepoints	0.4900	0.6800	0.7500	0.7200	0.7200	0.6600
Count	Naive-CP	0.1400	0.0600	0.0500	0.0100	0.0000	0.0000
	Naive-CPSS	0.0100	0.0100	0.0000	0.0000	0.0000	0.0000
	CPCA-CP	0.4300	0.6200	0.6500	0.2100	0.3800	0.2900
	CPCA-CPSS	0.0300	0.0000	0.0000	0.0000	0.0000	0.0000
	RPCA-CP	**0.9900**	**0.9900**	**1.0000**	**1.0000**	**0.9900**	**1.0000**
	RPCA-CPSS	0.8700	0.8900	0.9600	0.9800	0.9900	0.9900
	E.divisive	0.9000	0.5200	0.3600	0.9500	0.7100	0.7200
	KCP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	changepoints	0.0000	0.0000	0.0000	0.8000	0.8200	0.7900
Cover	Naive-CP	0.5607	0.5721	0.5792	0.5528	0.5555	0.5165
	Naive-CPSS	0.5642	0.6013	0.6204	0.4954	0.5203	0.5339
	CPCA-CP	0.7321	0.9056	0.9215	0.8535	0.8993	0.8637
	CPCA-CPSS	0.7107	0.7393	0.7907	0.6825	0.6443	0.6271
	RPCA-CP	**0.9796**	**0.9803**	**0.9809**	**0.9872**	0.9859	**0.9902**
	RPCA-CPSS	0.9407	0.9477	0.9688	0.9859	**0.9868**	0.9895
	E.divisive	0.8748	0.5384	0.7283	0.9371	0.7938	0.8348
	KCP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	changepoints	0.4338	0.4398	0.4330	0.9267	0.9313	0.9272

The largest values for each evaluation metric are in boldface

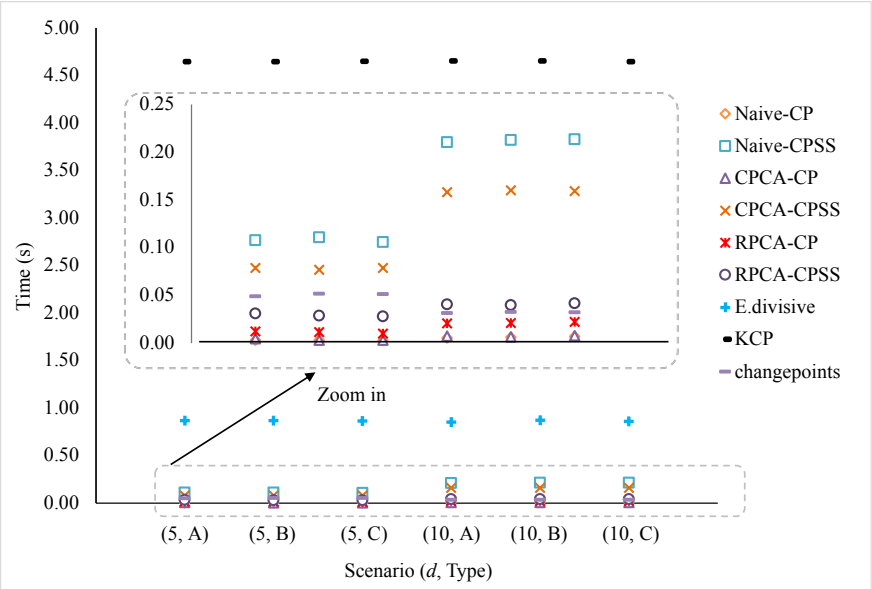

Figure 4.3: Comparison of average computation time for multivariate data under different scenarios.

perform worse, implying that these two methods are sensitive to changes in distribution and overestimate the number of *cps*. Moreover, as the dimension p increases, the measurements of 'KCP' decrease obviously, while other methods are not influenced by the increase of p for different scenarios. Generally, CPCA-CP performs better than the benchmarks, such as 'E.divisive', 'KCP', and 'changepoints'. As expected, naive methods exhibit the poorest performance measured by *Acc*, *Count*, and *Cover* due to their intrinsic nature in dismissing critical information. Figure 4.3 shows the average computation time for each method under different scenarios. Among the nine methods, 'E.divisive' and 'KCP' are time-consuming compared with others, while CPCA-CP, RPCA-CP, RPCA-CPSS show more time-efficiency than the others. As the dimension p increases, the computation time for CPCA-CPSS and Naive-CP increases significantly. To sum up, RPCA-CP is the most efficient method, balancing low computation time with high accuracy in change-point detection.

To examine the method's performance in resistance to outliers, we randomly sample 200 entries from the data matrix $\boldsymbol{X}$, and replace them by multiplying the original entries with five. The results of evaluation metrics are shown in Table 4.3. Undoubtedly, these evaluation metrics appear to be smaller than those in Table 4.2. However, both RPCA-CP and RPCA-CPSS exhibit strong capability in resistance to outliers, reflected by their similar values of *Acc*, *Count*, and *Cover* in the presence of outliers or not. Among all the methods, RPCA-CP dominates the others in most cases as outliers present.

4.5.3 Performance comparison on moderately high-dimensional data

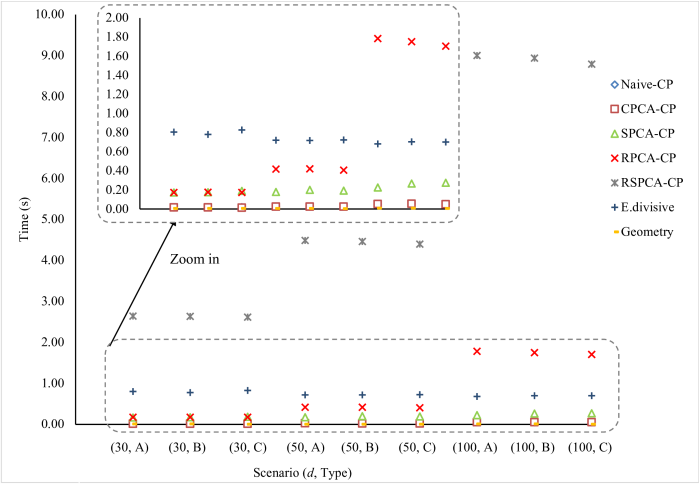

Figure 4.4: Comparison of average computation time for moderately high-dimensional data under different scenarios.

In this subsection, we examine the performance of the proposed methods, Naive-CP, CPCA-CP, SPCA-CP, RPCA-CP, SRPCA-CP, and two existing CPD methods designed for high-dimensional data, i.e., 'E.divisive' and 'Geometry'. The data are generated from Model (4.5.4) with different types of *cps*, whereas we consider larger dimensions, $p = 30, 50, 100$. The evaluation metrics are shown in Table 4.4. As can be seen from Table 4.4, CPCA-CP, RPCA-CP and 'Geometry' can detect all the true *cps* since their *Acc* values are equal to one generally. When it comes to *Count* and *Cover*, the performance of CPCA-CP decreases. For $p = 30$, RPCA-CP and 'Geometry' exhibit similar performance in CPD across all cases in the truth of their close evaluation metrics. As the dimension p increases, 'Geometry' becomes even better than RPCA-CP. Figure 4.4 compares the average computation time for moderately high-dimensional data under different scenarios. 'Geometry' performs best in time efficiency.

To examine the method's capabilities in terms of resistance to outliers for moderately high-dimensional data, we randomly generate 200 corrupted entries of $\boldsymbol{X}$, as is in the above subsection. The computed evaluation metrics are given in Table 4.5. For high-dimensional data with outliers, it cannot be concluded that one method dominates the other for different scenarios. However, one can easily find that both RPCA-CP and 'Geometry' outperform the other methods in most cases. As for the case of $p = 100$ and Type C *cps*, RSPCA-CP yields excellent performance, showing the largest *Acc*, *Count* and *Cover*.

Table 4.4: Comparison of average evaluation metrics for moderately high-dimensional data under different scenarios.

Scenario	Evaluation	Methods						
(p, Type)	Metrices	Naive-CP	CPCA-CP	SPCA-CP	RPCA-CP	RSPCA-CP	E.divisive	Geometry
(30, A)	*Acc*	0.0100	**1.0000**	0.9900	**1.0000**	0.2000	0.8700	**1.0000**
	Count	0.0500	0.2400	0.8200	**0.9900**	0.2400	0.9600	0.9600
	Cover	0.1269	0.8933	0.9705	**0.9920**	0.2341	0.9564	0.9916
(30, B)	*Acc*	0.1200	**1.0000**	**1.0000**	**1.0000**	0.4800	0.8300	**1.0000**
	Count	0.0800	0.5400	0.9500	**1.0000**	0.5200	0.9200	0.9600
	Cover	0.2821	0.9433	0.9855	0.9943	0.5152	0.9414	**0.9955**
(30, C)	*Acc*	0.5400	**1.0000**	0.9900	**1.0000**	0.9700	0.7400	**1.0000**
	Count	0.3300	0.5500	0.9200	**0.9900**	**0.9900**	0.8200	0.9700
	Cover	0.7771	0.9370	0.9800	0.9940	0.9885	0.9222	**0.9957**
(50, A)	*Acc*	0.0400	**1.0000**	0.9800	**1.0000**	0.4000	0.8700	**1.0000**
	Count	0.0700	0.2200	0.7800	0.9300	0.4000	0.8800	**0.9600**
	Cover	0.1595	0.8768	0.9676	0.9909	0.4220	0.9428	**0.9930**
(50, B)	*Acc*	0.2500	**1.0000**	**1.0000**	**1.0000**	0.7600	0.8500	**1.0000**
	Count	0.1900	0.3200	0.9200	0.9700	0.7900	0.8500	**0.9800**
	Cover	0.4760	0.9016	0.9811	**0.9957**	0.7679	0.9348	0.9952
(50, C)	*Acc*	0.6700	**1.0000**	**1.0000**	**1.0000**	0.9900	0.8600	**1.0000**
	Count	0.2400	0.2500	0.9700	0.9500	**0.9900**	0.8400	0.9700
	Cover	0.7929	0.8956	0.9849	0.9939	0.9918	0.9319	**0.9963**
(100, A)	*Acc*	0.0400	**1.0000**	0.9600	**1.0000**	0.6800	0.8300	**1.0000**
	Count	0.1000	0.0900	0.8300	0.6800	0.7600	**0.9500**	**0.9500**
	Cover	0.2267	0.8181	0.9680	0.9608	0.7340	0.9476	**0.9958**
(100, B)	*Acc*	0.3200	**1.0000**	0.9900	**1.0000**	0.8900	0.8100	**1.0000**
	Count	0.2900	0.2300	0.9100	0.9400	0.9300	0.8900	**0.9600**
	Cover	0.5905	0.8677	0.9827	0.9905	0.9157	0.9355	**0.9976**
(100, C)	*Acc*	0.8400	**1.0000**	0.9900	**1.0000**	**1.0000**	0.7800	**1.0000**
	Count	0.1400	0.2000	0.9400	0.9300	**1.0000**	0.9100	0.9800
	Cover	0.8079	0.8681	0.9817	0.9910	0.9972	0.9363	**0.9973**

The largest values for each evaluation metric are in boldface

Table 4.5: Comparison of evaluation metrics for moderate high-dimensional data with outliers under different scenarios.

Scenario	Evaluation	Methods						
(p, Type)	Metrices	Naive-CP	CPCA-CP	SPCA-CP	RPCA-CP	RSPCA-CP	E.divisive	Geometry
(30, A)	*Acc*	0.0300	0.9900	0.9800	**1.0000**	0.4600	0.8800	**1.0000**
	Count	0.2200	0.2000	0.8600	**1.0000**	0.5200	0.9700	0.9300
	Cover	0.3564	0.8714	0.9667	**0.9936**	0.5053	0.9540	0.9909
(30, B)	*Acc*	0.5800	**1.0000**	**1.0000**	**1.0000**	0.7500	0.8000	**1.0000**
	Count	0.1900	0.5000	0.8200	**0.9800**	0.8700	0.8500	0.9500
	Cover	0.7231	0.9297	0.9715	**0.9945**	0.8549	0.9339	0.9928
(30, C)	*Acc*	0.8500	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.7100	**1.0000**
	Count	0.0800	0.4800	0.8100	0.9900	**1.0000**	0.9200	0.9400
	Cover	0.7607	0.9334	0.9695	**0.9934**	0.9923	0.9271	0.9930
(50, A)	*Acc*	0.0600	**1.0000**	0.9900	**1.0000**	0.6200	0.8500	**1.0000**
	Count	0.1900	0.1500	0.8500	0.9300	0.7100	0.9500	**0.9600**
	Cover	0.3699	0.8482	0.9705	0.9898	0.6938	0.9567	**0.9927**
(50, B)	*Acc*	0.5500	1.0000	1.0000	**1.0000**	0.8000	0.7300	1.0000
	Count	0.1700	0.3500	0.8300	**1.0000**	0.9100	0.8400	0.9800
	Cover	0.7177	0.9019	0.9754	**0.9965**	0.8766	0.9210	0.9965
(50, C)	*Acc*	0.8400	0.9900	1.0000	**1.0000**	1.0000	0.7500	1.0000
	Count	0.0700	0.3500	0.7900	**0.9900**	1.0000	0.8900	0.9800
	Cover	0.7438	0.9121	0.9713	0.9948	0.9947	0.9291	**0.9951**
(100, A)	*Acc*	0.3933	0.9933	0.9767	**1.0000**	0.9033	0.8633	**1.0000**
	Count	0.0000	0.0300	0.6200	0.7100	0.9533	0.9467	**0.9667**
	Cover	0.4873	0.7405	0.9437	0.9650	0.9303	0.9512	**0.9936**
(100,B)	*Acc*	0.7733	0.9800	0.9967	**1.0000**	0.9467	0.8333	**1.0000**
	Count	0.0000	0.1000	0.7300	0.9133	**0.9833**	0.8467	0.9733
	Cover	0.5566	0.8161	0.9641	0.9914	0.9754	0.9320	**0.9954**
(100, C)	*Acc*	0.8233	0.9867	0.9933	**1.0000**	**1.0000**	0.7433	**1.0000**
	Count	0.0000	0.0867	0.7033	0.9167	**0.9967**	0.8833	0.9700
	Cover	0.5633	0.7986	0.9630	0.9893	**0.9967**	0.9245	0.9959

The largest values for each evaluation metric are in boldface

4.5.4 Performance comparison of RPCA-CP and 'Geometry' for heavy-tail distributed data

Table 4.6: Comparison of evaluation metrics for multivariate t distributed data with outliers under different p.

p	RPCA-CP			Geometry		
	Acc	*Count*	*Cover*	*Acc*	*Count*	*Cover*
5	**0.7667**	**0.9267**	**0.9344**	0.7300	0.1067	0.7777
10	0.6570	**0.8650**	**0.9010**	**0.7220**	0.1000	0.7931
30	0.7500	**0.8833**	**0.9294**	**0.8700**	0.2000	0.8482
50	0.7733	**0.8200**	**0.9245**	**0.8533**	0.1633	0.8378
100	0.7300	**0.8067**	**0.9179**	**0.8367**	0.1533	0.8293

The largest values for each evaluation metric are in boldface

As discussed in Section 4.5.2, RPCA-CP outperforms most other methods when applied to multivariate data. In the case of moderately high-dimensional data, both RPCA-CP and 'Geometry' exhibit comparable performance. In reality, it usually encounters heavy-tailed data, such as financial data. We further compare their performance in CPD for more complex data with heavy-tailed distributions and/or corrupted entries. The data are generated from multivariate t distributions, i.e.,

$$\boldsymbol{x}_i \sim \begin{cases} \boldsymbol{t}_\nu(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), & i = 1, \ldots, 100, \\ \boldsymbol{t}_\nu(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), & i = 101, \ldots, 200, \\ \boldsymbol{t}_\nu(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3), & i = 201, \ldots, 300, \end{cases} \tag{4.5.5}$$

where $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \mathbf{0}.5, \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{D}_{0.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{0.5}\frac{\nu-2}{\nu}$, $\boldsymbol{\Sigma}_3 = \boldsymbol{D}_{1.5}\boldsymbol{S}_{0.9}\boldsymbol{D}_{1.5}\frac{\nu-2}{\nu}$, and the degree of freedom $\nu = 3$. Additionally, we randomly generate a sample of 200 from the entries of $\boldsymbol{X}$, and replace them by multiplying the original entries with five.

The results of evaluation metrics are shown in Table 4.6. Clearly, 'Geometry' has smaller $Acc, Count$, and $Cover$ metrics than those of the RPCA-CP method when analyzing multivariate data ($p = 5$). As p increases, the Acc becomes better for moderately high-dimensional ones for which the method is designed. 'Geometry' performs better in terms of Acc, while RPCA-CP shows higher efficiency in detecting the number of $cps(Count)$ and the covering metrics($Cover$). In general, RPCA-CP exhibits robustness when data are heavy-tailed and contain corrupted entries.

Table 4.7: The estimated number of *cps* ($\hat{k}$) and elapsed time in seconds (ERT.S) by SPCA-CP, RPCA-CP, RSPCA-CP, 'E.divisive', and 'Geometry' for the ACGH dataset.

Measures	SPCA-CP	RPCA-CP	RSPCA-CP	'E.divisive'	'Geometry'
$\hat{k}$	17	32	63	96	32
ERT.S	0.69	7.36	47.51	1968.28	0.06

4.6 Experiments on Real Data

We carry out two real data analyses to further compare the CPD efficiency of these methods, such as SPCA-CP, RPCA-CP, RSPCA-CP, and two benchmarks, 'E.divisive' and 'Geometry'. The two real datasets are ACGH, the bladder tumor microarray Data in the R package `ecp` (James et al., 2023), and DJIA, the daily log returns of these companies that compose the Dow Jones Industrial Average (DJIA). Please note that we have excluded those methods that exhibited poor performance in the previous section or were deemed inapplicable to the two datasets.

4.6.1 Example 1: Bladder tumor microarray data (ACGH)

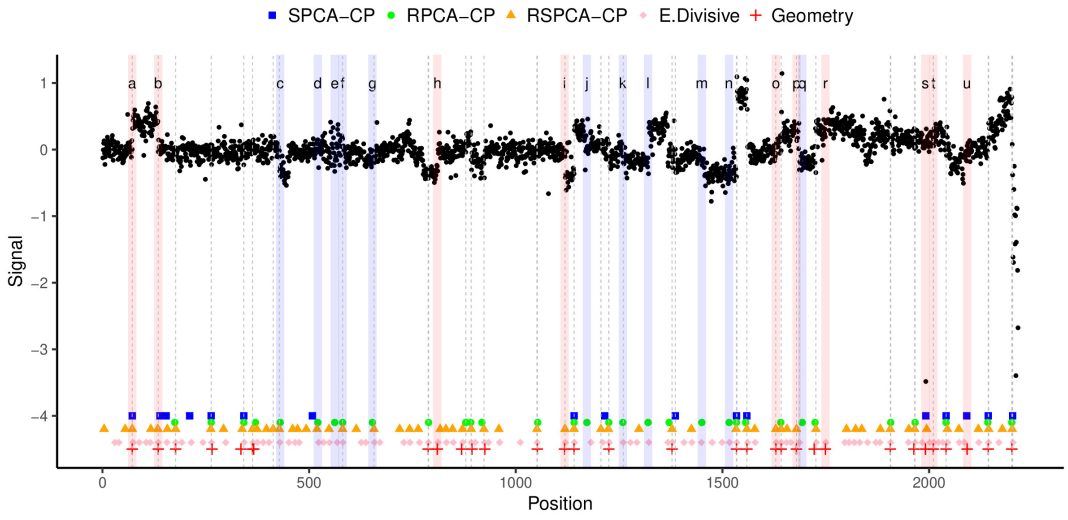

Figure 4.5: The estimated *cps* of mico-array data sequence for individual 10 with a bladder tumor by using SPCA-CP, RPCA-CP, RSPCA-CP, 'E.divisive', and 'Geometry', where dashed vertical lines indicate the *cps* detected by more than one methods ('E.divisive' excluded).

The ACGH dataset contains 43 individuals with bladder tumors, where a numeric vector of mico-array data for each individual is presented. This information is stored in a 2215 by

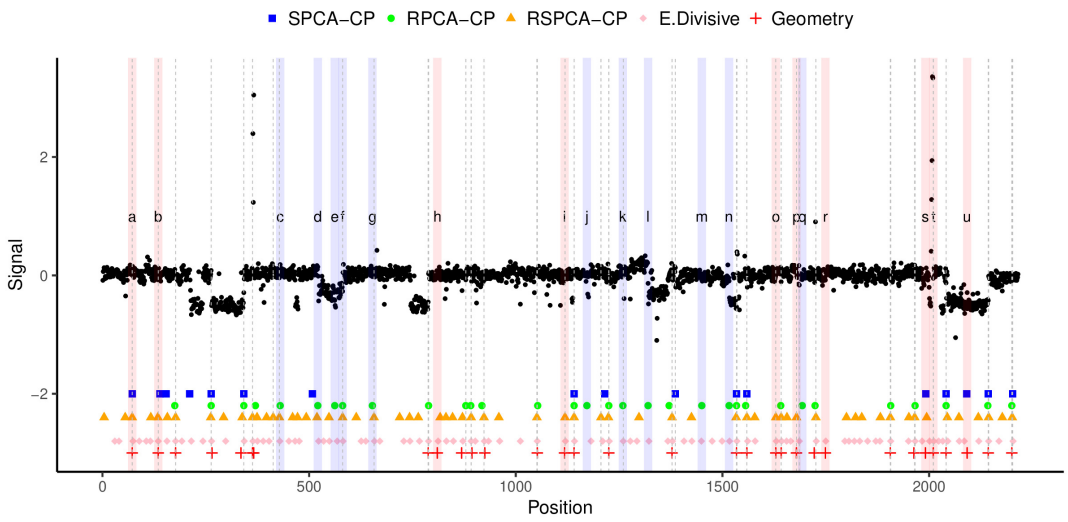

Figure 4.6: The estimated *cps* of mico-array data sequence for individual 15 with a bladder tumor by using SPCA-CP, RPCA-CP, RSPCA-CP, 'E.divisive', and 'Geometry', where dashed vertical lines indicate the *cps* detected by more than one methods ('E.divisive' excluded).

43 matrix (Nicholas and Matteson, 2014). The locations of *cps* across each micro-array set are assumed to be similar.

Table 4.7 presents the estimated number of *cps* and elapsed time by SPCA-CP, RPCA-CP, RSPCA-CP, 'E.divisive', and 'Geometry'. Among these methods, 'Geometry' and 'E.divisive' are respectively the most time-efficiency and time-consumption methods. Moreover, RSPCA-CP, and 'E.divisive' tend to overestimate distributional changes, by detecting 63, and 96 *cps*, respectively. However, SPCA-CP underestimates the number of *cps*, just detecting 17 *cps*. RPCA-CP and 'Geometry' identify the same total number of *cps*, but their locations differ between the two methods. Figures 4.5 and 4.6 visually show the locations of detected *cps* of the six methods for individuals 10 and 15, respectively, where the difference between RPCA-CP and 'Geometry' is shaded and labelled by letters from 'a' to 'u'. The red line denotes the *cps* detected by 'Geometry' but not by RPCA-CP, and vice versa for the blue lines. For individual 10 (see Figure 4.5), RPCA-CP detects some changes in mean at locations 'c-g', 'j-n' and 'q', while 'Geometry' dismisses these changes. The points 'c', 'g', and 'j-m' show changes in mean, while the points 'd-f' show changes in variance. Some *cps* are detected by both RPCA-CP and 'Geometry', but with slightly different locations, as exemplified by the locations 'n-q'. For individual 15 (see Figure 4.6), RPCA-CP captures the mean shifts at 'd-f', 'l' and 'n', while 'Geometry' does not. Moreover, the locations marked in red are mis-detected by 'Geometry'. Especially for positions 's-t', it is challenging to discern whether the observed variations are due to changes in distribution or simply outliers.

Table 4.8: The exact dates of *cps* estimated by SPCA-CP, RPCA-CP, RSPCA-CP, 'E.divisive', and 'Geometry'.

SPCA-CP	RPCA-CP	RSPCA-CP	'E.divisive'	'Geometry'
2020-02-21	2020-02-21	2020-02-25	2020-02-24	2020-03-12
2020-04-21			2020-04-22	2020-03-25
2020-06-29	2020-06-29	2020-06-15		
		2020-11-24		2020-11-24
		2021-01-22		
		2021-09-22		
	2022-02-09	2022-02-16		2022-02-09
2023-03-29	2023-03-22	2023-03-22		2023-03-22

4.6.2 Example 2: Dow Jones Industrial Average (DJIA)

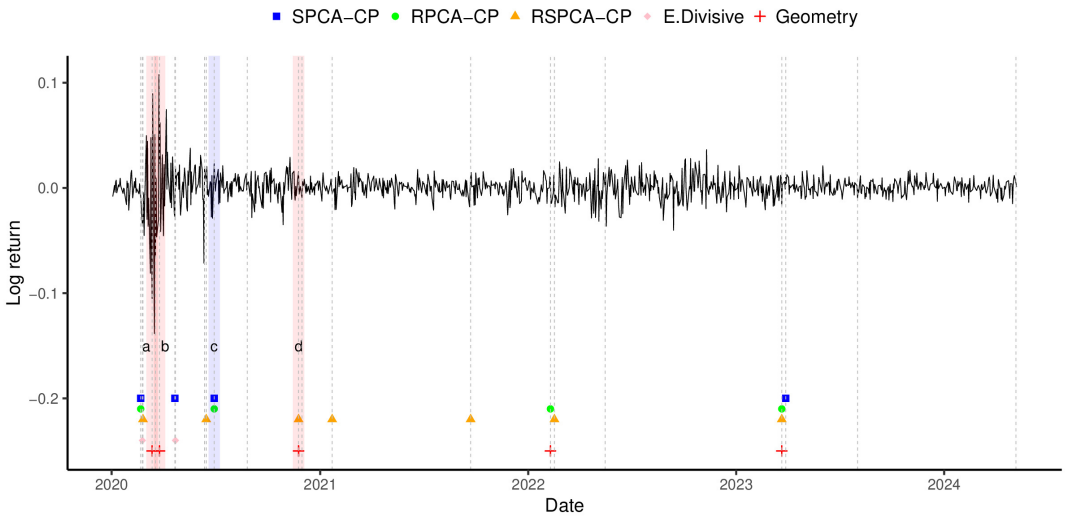

Figure 4.7: Daily log returns of the DJIA index from Jan. 1st, 2020 to May. 6th, 2024. The dashed vertical lines indicate the dates of *cps* estimated by SPCA-CP, RPCA-CP, RSPCA-CP, 'E.divisive', and 'Geometry'.

We consider the daily log returns of 30 companies in the Dow Jones Industrial Average (DJIA) from January 1, 2020, to May 6, 2024. The stock symbols include "MMM", "AXP", "AMGN", "AAPL", "BA", "CAT", "CVX", "CSCO", "KO", "DIS", "GS", "HD", "HON", "IBM", "INTC", "JNJ", "JPM", "MCD", "MRK", "MSFT", "NKE", "PG", "TRV", "UNH", "VZ", "WBA", "WMT", "DOW", "V", "CRM". This information composes the observed data matrix $\boldsymbol{X}$ with 1093 observations, each of which has dimension 30. The stock symbols of the companies are listed since the DJIA index changed its components over time. The original stock prices are collected from `https://ca.finance.yahoo.com/`. The DJIA index is downloaded from `https://ca.investing.com/indices/us-30-historical-data`.

Among this observation matrix $\boldsymbol{X}$, there are 30 missing values.

Table 4.8 lists the detected *cps* via SPCA-CP, RPCA-CP, RSPCA-CP, RSPCA-CP, 'E.divisive', and 'Geometry'. Figure 4.7 visually shows these *cps* of the DJIA index from January 1, 2020, to May 6, 2024. The different *cps* via RPCA-CP and 'Geometry' methods are shaded and labelled by letters 'a-d' in Figures 4.7. The red lines are the dates of *cps* that 'Geometry' identifies, but RPCA-CP does not. While the blue line is the *cp* detected by RPCA-CP but not 'Geometry'. Clearly, RSPCA-CP tends to overestimate the number of *cps*, while 'E.divisive' underestimates it, just having two. However, SPCA-CP, RPCA-CP, and 'Geometry' exhibit similar results on the number of *cps*. Some *cps* can be obvious, say, around February 21, 2020, which are detected timely by SPCA-CP, RPCA-CP, RSPCA-CP, RSPCA-CP, and 'E.divisive'. The change-point might be caused by the COVID-19 pandemic and the sudden crash of stock markets worldwide. However, 'Geometry' fails to capture this change, but provides deferred *cps* on March 12, and March 18, respectively. After that, 'Geometry' captures the date of March 25 as the end of the pandemic, which is much earlier than the dates detected by other methods. Both 'E.divisive' and SPCA-CP detect the date around April 21, 2020 as the end of the pandemic, coinciding with global stock markets re-entering a bull market. RPCA-CP and RSPCA-CP detect the *cps* at the end of the pandemic in June 2020, as shown by date 'c' in Figure 4.7. Clearly, from April 2020 to June 2020, the early phase of COVID-19, there was a significant correlation between the outbreak and volatility in stock markets, as noted by Pak et al. (2020). SPCA-CP detects both *cps*, marking the pandemic's shock weakening from strong in April 2020 to mild in June 2020. From February 9, 2022, to March 22, 2023, the influence of an interest rate hike was observed. All methods identify these two *cps* except 'E.divisive'. SPCA-CP does not capture the beginning of the interest rate hike. Overall, these *cps* detected by the RPCA-CP are more reasonable and interpretable.

4.7 Conclusion

In this chapter, we propose the multiple CPD method for complex data with moderately high dimensions, high correlations, outliers and/or heavy-tailed distribution. This method admits a flexible architecture by leveraging the strengths of both well-studied PCA and univariate CPD methods. We provide the theoretical justification of the proposed method to guarantee its feasibility. Numerically, we examine the performance of the PCA-uCPD method and compare them with some well-known methods, such as 'E.divisive', 'KCP', and 'changepoints' for low-dimensional data, and 'E.divisive' and 'Geometry' for high-dimensional data. By considering various data generation models and types of *cps*, it is

found that RPCA-CP achieves superior CPD performance with changes in mean and variance for multivariate data. As for moderately high-dimensional data, RPCA-CP achieves at least the performance of 'Geometry'. Furthermore, its performance can be strengthened when dealing with high-dimensional data with corrupted entries and heavy-tailed distribution. Finally, we demonstrate the outstanding performance of RPCA-CP on the ACGH data and the superior performance of both RPCA-CP and 'Geometry' on the DJIA data.

Chapter 5

High-Dimensional Multivariate Quantile Regression with Multi-Platform Data Integration

5.1 Introduction

In many modern applications, particularly in finance, data are increasingly collected from multiple platforms or sources, such as different financial markets or exchanges. Although these datasets are often generated for a common analytical objective and share a set of important predictors, they may exhibit substantial heterogeneity in scale, noise level, and dependence structure. Effectively integrating information across platforms while accounting for such heterogeneity has therefore become an important challenge in contemporary statistical modeling.

Existing data integration methods have primarily focused on conditional mean regression, which may be inadequate for financial data that frequently exhibit heavy-tailed distributions, outliers, and heterogeneous covariate effects. Quantile regression provides a robust alternative by modeling conditional quantiles of the response distribution, thereby offering a more comprehensive characterization of risk and tail behavior. While quantile regression has been extensively studied in univariate settings and has been successfully applied in a wide range of fields, its extension to multi-platform and multivariate response settings remains relatively underdeveloped.

In many financial applications, responses are inherently multivariate and often strongly correlated, such as returns from multiple assets or indices observed across different markets. Ignoring these dependencies may lead to inefficient estimation and an incomplete under-

standing of joint conditional behavior. Although recent work has advanced multivariate quantile regression methodologies, fundamental challenges remain due to the lack of a natural ordering for multivariate responses and the computational complexity associated with high-dimensional estimation.

Motivated by these challenges, this chapter develops a multivariate quantile regression framework for integrating high-dimensional data across multiple platforms. The proposed approach directly models vector-valued responses within a loss-based optimization framework, allowing for flexible dependence structures across response dimensions without imposing restrictive distributional assumptions. To exploit shared information across platforms while accommodating platform-specific heterogeneity, we incorporate an ℓ_1/ℓ_2 cross-platform group penalty that encourages the identification of predictors with common effects across platforms.

The resulting framework generalizes several existing models, including univariate quantile regression for multi-platform data integration and multivariate quantile regression in single-platform settings. An efficient smoothing-based optimization algorithm is developed to address the non-differentiability of the quantile loss function, and its convergence properties are established. Through simulation studies and real data applications, we demonstrate that the proposed method improves estimation accuracy and variable selection performance compared with single-platform and univariate approaches.

The remainder of this chapter is organized as follows. Section 5.2 introduces the proposed methodology and optimization algorithm. Section 5.3 presents simulation studies. Section 5.4 applies the proposed method to real data, and Section 5.5 concludes the chapter.

5.2 Methodology

5.2.1 Model setup

We consider data collected from K distinct platforms, where each platform provides n independent observations comprising a q-dimensional response vector $\boldsymbol{Y}_k \in \mathbb{R}^{n\times q}$ and a p-dimensional set of explanatory variables $\boldsymbol{X}_k \in \mathbb{R}^{n\times p}$, for $k = 1, \ldots, K$. Our goal is to jointly estimate the q marginal conditional quantiles of the response $\boldsymbol{Y}_k$ across all platforms. The high-dimensional setting allows the number of covariates p to increase with the sample size n. Furthermore, the response vectors $\boldsymbol{Y}_k$ may be dependent across platforms. We model the linear relationship between the responses and predictors via

$$\boldsymbol{Y}_k = \boldsymbol{X}_k \boldsymbol{B}_k(\boldsymbol{\tau}) + \boldsymbol{E}_k(\boldsymbol{\tau}), \tag{5.2.1}$$

where $\boldsymbol{B}_k(\boldsymbol{\tau}) \in \mathbb{R}^{p\times q}$ denotes the quantile regression coefficient matrix that varies by quantile $\boldsymbol{\tau} \in (0,1)^q$, providing flexibility beyond mean regression. $\boldsymbol{E}_k(\boldsymbol{\tau}) \in \mathbb{R}^{n\times q}$ is the error matrix whose rows are independently and identically distributed given $\boldsymbol{X}_k$. Denote the i-th row of $\boldsymbol{E}_k(\boldsymbol{\tau})$ as $\boldsymbol{e}_{ki} \in \mathbb{R}^q$ for $i = 1, \ldots, n$. We assume that the marginal quantiles of $\boldsymbol{e}_{ki}$ are equal to zero, $Q_{\boldsymbol{\tau}}(\boldsymbol{e}_{ki}) = \mathbf{0}$. The error assumption does not impose a parametric distribution, making the model robust to skewness, heavy tails, and outliers. For notational simplicity, the dependence of $\boldsymbol{B}_k$ and $\boldsymbol{E}_k$ on $\boldsymbol{\tau}$ is suppressed hereafter.

To be specific, we detail the structures of notations. For the platform k, the response matrix $\boldsymbol{Y}_k \in \mathbb{R}^{n\times q}$ is given by

$$\boldsymbol{Y}_k = [\boldsymbol{y}_{k1}^\top, \boldsymbol{y}_{k2}^\top, \ldots, \boldsymbol{y}_{kn}^\top]^\top = \begin{bmatrix} y_{k11} & y_{k12} & \cdots & y_{k1q} \\ y_{k21} & y_{k22} & \cdots & y_{k2q} \\ \vdots & \ddots & & \vdots \\ y_{kn1} & y_{kn2} & \cdots & y_{knq} \end{bmatrix},$$

where y_{kij} represents the i-th sample of response j at platform k, $i = 1, \ldots, n$, $j = 1, \ldots, q$. We remark that the matrix $\boldsymbol{Y}_k$ may have different scales across platforms. Rewrite the design matrix and quantile regression coefficient matrix respectively as $\boldsymbol{X}_k \in \mathbb{R}^{n\times p}$ and $\boldsymbol{B}_k \in \mathbb{R}^{p\times q}$, i.e.,

$$\boldsymbol{X}_k = \begin{bmatrix} x_{k11} & x_{k12} & \cdots & x_{k1p} \\ x_{k21} & x_{k22} & \cdots & x_{k2p} \\ \vdots & \ddots & & \vdots \\ x_{kn1} & x_{kn2} & \cdots & x_{knp} \end{bmatrix}, \qquad \boldsymbol{B}_k = \begin{bmatrix} \beta_{k11} & \beta_{k12} & \cdots & \beta_{k1q} \\ \beta_{k21} & \beta_{k22} & \cdots & \beta_{k2q} \\ \vdots & \ddots & & \vdots \\ \beta_{kp1} & \beta_{kp2} & \cdots & \beta_{kpq} \end{bmatrix}.$$

Denote the i-th row and m-th column vectors of $\boldsymbol{X}_k \in \mathbb{R}^{n\times p}$ respectively as $\boldsymbol{x}_{ki} = (x_{ki1}, \ldots, x_{kip})^\top \in \mathbb{R}^p$ and $\dot{\boldsymbol{x}}_{km} = (x_{k1m}, \ldots, x_{knm})^\top \in \mathbb{R}^n$ for $i = 1, \ldots, n; m = 1, \ldots, p$. Let $\dot{\boldsymbol{b}}_{kj} \in \mathbb{R}^p$ denote the j-th column of $\boldsymbol{B}_k$, which represents the regression vector for the j-th response and corresponds to the p covariates on platform k. Similarly, let $\boldsymbol{b}_{km} \in \mathbb{R}^q$ be the m-th row of $\boldsymbol{B}_k$, representing the regression vector that captures the effects of the m-th predictor across the q responses. As for the random error matrix $\boldsymbol{E}_k \in \mathbb{R}^{n\times q}$, it can be rewritten as,

$$\boldsymbol{E}_k = [\boldsymbol{e}_{k1}^\top, \boldsymbol{e}_{k2}^\top, \cdots, \boldsymbol{e}_{kn}^\top]^\top = \begin{bmatrix} \varepsilon_{k11} & \varepsilon_{k12} & \cdots & \varepsilon_{k1q} \\ \varepsilon_{k21} & \varepsilon_{k22} & \cdots & \varepsilon_{k2q} \\ \vdots & \ddots & & \\ \varepsilon_{kn1} & \varepsilon_{kn2} & \cdots & \varepsilon_{knq} \end{bmatrix}.$$

In this chapter, we aim to build a high-dimensional quantile regression model with multivariate responses across K platforms. Denote $\boldsymbol{B} = (\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_K) \in \mathbb{R}^{p\times Kq}$. The pooled

quantile loss function is defined as

$$Q_{\boldsymbol{\tau}}(\boldsymbol{B}) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{q}\rho_{\tau_j}(y_{kij} - \boldsymbol{x}_{ki}^{\top}\dot{\boldsymbol{b}}_{kj}), \tag{5.2.2}$$

where $\rho_\tau(u) = u(\tau - 1(u < 0))$ is the element-wise check loss function. The pooled quantile loss was also used by Dai et al. (2023) to construct the quantile regression model for multi-source data. Their work considered a univariate response per platform and simultaneously modeled multiple quantile levels in order to identify predictors that influence the response across all quantile levels. Furthermore, we focus on sparse estimation in the high-dimensional setting, where most entries of $\boldsymbol{B}_k$ are zero, implying that only a subset of predictors influence the responses. We therefore impose an appropriate regularization term to select the relevant predictors, a common technique in quantile regression modeling (Petrella and Raponi, 2019; Chen and Chen, 2024; Dai et al., 2023; Chen et al., 2020). The methodological challenge in our setting is to use a penalty that can identify predictors that influence the responses across multiple platforms. We assume that the set of important predictors is common across all platforms, that is, the nonzero rows of $\boldsymbol{B}_k$ are identical for each k. Denote $\boldsymbol{b}^{(m)} = [\boldsymbol{b}_{1m}^{\top}, \boldsymbol{b}_{2m}^{\top}, \ldots, \boldsymbol{b}_{Km}^{\top}]^{\top} \in \mathbb{R}^{Kq}$ the regression coefficients of the m-th predictor across all responses and all platforms. Thus, a sparse solution of $\hat{\boldsymbol{B}}$ can be obtained by minimizing the penalized objective function, i.e.,

$$\min_{\boldsymbol{B}\in\mathbb{R}^{p\times Kq}} L_\tau(\boldsymbol{B}) = Q_\tau(\boldsymbol{B}) + \lambda_n\|\boldsymbol{B}\|_{\ell_1/\ell_2}, \tag{5.2.3}$$

where $\|\boldsymbol{B}\|_{\ell_1/\ell_2} = \sum_{m=1}^{p}\|\boldsymbol{b}^{(m)}\|_2$, and $\boldsymbol{b}^{(m)} = [\boldsymbol{b}_{1m}^{\top}, \boldsymbol{b}_{2m}^{\top}, \ldots, \boldsymbol{b}_{Km}^{\top}]^{\top} \in \mathbb{R}^{Kq}$, denotes the regression coefficients of the m-th predictor across all responses and all platforms. Indeed, problem (5.2.3) provides a general framework for high-dimensional quantile regression. It encompasses several important special cases. When $q = 1$ and $K \geq 2$, it becomes a univariate quantile regression model for multi-source data integration. When $q \geq 2$ and $K = 1$, it simplifies to the multivariate quantile regression problem studied in Chen and Chen (2024). The most basic case, where $q = K = 1$, corresponds to high-dimensional quantile regression with Lasso penalization, as is in Belloni and Chernozhukov (2011) and Li and Zhu (2008).

5.2.2 Parameter estimation

The non-smooth nature of the quantile loss function presents significant computational difficulties. To overcome this, Chen and Chen (2024) developed efficient algorithms by leveraging smoothing approximations. Their approach utilizes Nesterov's smoothing technique of

Nesterov (2005), in conjunction with majorization-minimization (MM) algorithms to solve the optimization problem more efficiently.

In our multi-platform framework, we adopt a smoothing technique similar to that of Chen and Chen (2024) to facilitate data integration and parameter estimation. The lower and upper smoothing approximations of the quantile objective function $Q_\tau(\boldsymbol{B})$ are defined respectively as follows,

$$\underline{Q}_{\tau,\omega}(\boldsymbol{B}) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{q}\underline{\rho}_{\tau_j,\omega}(y_{kij} - \boldsymbol{x}_{ki}^\top \dot{\boldsymbol{b}}_{kj}), \tag{5.2.4}$$

$$\overline{Q}_{\tau,\omega}(\boldsymbol{B}) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{q}\overline{\rho}_{\tau_j,\omega}(y_{kij} - \boldsymbol{x}_{ki}^\top \dot{\boldsymbol{b}}_{kj}), \tag{5.2.5}$$

where $\underline{\rho}_{\tau,\omega}(u)$ and $\overline{\rho}_{\tau,\omega}(u)$ are smoothing approximations of the check loss function $\rho_\tau(u)$ (Aravkin et al., 2014), and are given by

$$\underline{\rho}_{\tau,\omega}(u) = \begin{cases} (\tau-1)u - \omega(1-\tau^2)/2, & u < (\tau-1)\omega, \\ u^2/(2\omega), & (\tau-1)\omega \le u \le \tau\omega, \\ \tau u - \omega\tau^2/2, & u > \tau\omega, \end{cases}$$

$$\overline{\rho}_{\tau,\omega}(u) = \begin{cases} (\tau-1)u, & u < -\omega, \\ (\tau-1)u - (u+\omega)^2/(4\omega), & -\omega \le u \le \omega, \\ \tau u, & u > \omega. \end{cases}$$

It is easy to verify that for any $\boldsymbol{B}$, $\underline{Q}_{\boldsymbol{\tau},\omega}(\boldsymbol{B}) \le Q_{\boldsymbol{\tau}}(\boldsymbol{B}) \le \overline{Q}_{\boldsymbol{\tau},\omega}(\boldsymbol{B})$, and $\lim_{\omega\to 0}\underline{Q}_{\boldsymbol{\tau},\omega}(\boldsymbol{B}) = \lim_{\omega\to 0}\overline{Q}_{\boldsymbol{\tau},\omega}(\boldsymbol{B}) = Q_{\boldsymbol{\tau}}(\boldsymbol{B})$.

Thus, the minimization problem (5.2.3) is equal to

$$\min_{\boldsymbol{B}\in\mathbb{R}^{p\times Kq}} \underline{L}_{\tau,\omega}(\boldsymbol{B}) = \underline{Q}_{\tau,\omega}(\boldsymbol{B}) + \lambda_n\|\boldsymbol{B}\|_{\ell_1/\ell_2}, \tag{5.2.6}$$

or

$$\min_{\boldsymbol{B}\in\mathbb{R}^{p\times Kq}} \overline{L}_{\tau,\omega}(\boldsymbol{B}) = \overline{Q}_{\tau,\omega}(\boldsymbol{B}) + \lambda_n\|\boldsymbol{B}\|_{\ell_1/\ell_2}. \tag{5.2.7}$$

We plan to use the MM technique to solve the minimization problems in Eqs.(5.2.6)-(5.2.7). Denote the first derivatives of $\underline{Q}_{\tau,\omega}(\boldsymbol{B})$ and $\overline{Q}_{\tau,\omega}(\boldsymbol{B})$ respectively as $\nabla\underline{Q}_{\tau,\omega}(\boldsymbol{B})$ and $\nabla\overline{Q}_{\tau,\omega}(\boldsymbol{B})$. Let $\lambda_{\max}(\boldsymbol{X}_k^\top\boldsymbol{X}_k)$ be the maximum eigenvalue of $\boldsymbol{X}_k^\top\boldsymbol{X}_k$) for $k = 1,\ldots,K$. Before proceeding, we present the following proposition.

Proposition 5.2.1 *$\nabla \underline{Q}_{\tau,\omega}(\boldsymbol{B})$ and $\nabla \overline{Q}_{\tau,\omega}(\boldsymbol{B})$ are Lipschitz continuous with Lipschitz constants*

$$\underline{L}_c = \frac{1}{\omega n^2} \max_k \lambda_{\max}(\boldsymbol{X}_k^\top \boldsymbol{X}_k), \tag{5.2.8}$$

$$\overline{L}_c = \frac{1}{2\omega n^2} \max_k \lambda_{\max}(\boldsymbol{X}_k^\top \boldsymbol{X}_k), \tag{5.2.9}$$

respectively.

Next, we revisit Lemma 2.1 in Chen and Chen (2024), which provides a sparse proxy solution to the regularized problem.

Lemma 5.2.1 *(Chen and Chen, 2024) Given $\lambda > 0$ and $\boldsymbol{\eta} \in \mathbb{R}^q$, the minimizer of the optimization problem*

$$\min_{\boldsymbol{b} \in \mathbb{R}^q} \ \lambda \|\boldsymbol{b}\|_2 + \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{\eta}\|_2^2 \tag{5.2.10}$$

can be expressed as shrink$(\boldsymbol{\eta}, \lambda) := \max\{1 - \lambda/\|\boldsymbol{\eta}\|_2, 0\} \cdot \boldsymbol{\eta}$.

Now, we will take the optimization problem (5.2.6) as an example to illustrate the MM algorithm in detail, which can be extended to the optimization problem (5.2.7) directly. Ortega and Rheinboldt (2000) first studied the MM algorithm, followed by many applications in penalized quantile regression models (Peng and Wang., 2020; Zhang et al., 2017; Chen et al., 2020).

Define a majorized function of $\underline{Q}_{\tau,\omega}(\boldsymbol{B})$ as $M_{\tau,\omega,\xi}(\boldsymbol{B}, \boldsymbol{Z})$ for any $\boldsymbol{Z} \in \mathbb{R}^{p \times Kq}$, i.e.,

$$M_{\tau,\omega,\xi}(\boldsymbol{B}, \boldsymbol{Z}) = \underline{Q}_{\tau,\omega}(\boldsymbol{Z}) + \langle \nabla \underline{Q}_{\tau,\omega}(\boldsymbol{Z}), \boldsymbol{B} - \boldsymbol{Z} \rangle + \frac{L_\omega + \xi}{2} \|\boldsymbol{B} - \boldsymbol{Z}\|_F^2$$

such that $M_{\tau,\omega,\xi}(\boldsymbol{B}, \boldsymbol{Z}) > \underline{Q}_{\tau,\omega}(\boldsymbol{Z})$ and $M_{\tau,\omega,\xi}(\boldsymbol{B}, \boldsymbol{B}) = \underline{Q}_{\tau,\omega}(\boldsymbol{B})$, with $\xi > 0$ and $L_\omega \geq \underline{L}_c$. Then we consider the minimization problem

$$\min_{\boldsymbol{B} \in \mathbb{R}^{p \times Kq}} M_{\tau,\omega,\xi}(\boldsymbol{B}, \boldsymbol{Z}) + \lambda_n \|\boldsymbol{B}\|_{\ell_1/\ell_2}. \tag{5.2.11}$$

Denote the unique minimizer of problem (5.2.11) as $P_{L_\xi}(\boldsymbol{Z})$. By Lemma (5.2.1), we have that

$$P_{L_\xi}(\boldsymbol{Z}) = \text{Shrink}\Big(\boldsymbol{G}, \tfrac{\lambda_n}{L_\omega + \xi}\Big) = (\text{Shrink}(\boldsymbol{g}_1^\top, \tfrac{\lambda_n}{L_\omega + \xi}), \ldots, \text{Shrink}(\boldsymbol{g}_p^\top, \tfrac{\lambda_n}{L_\omega + \xi}))^\top, \tag{5.2.12}$$

where

$$\boldsymbol{G} = \boldsymbol{Z} - \nabla \underline{Q}_{\tau,\omega}(\boldsymbol{Z})/(L_\omega + \xi) = (\boldsymbol{g}_1^\top, \ldots, \boldsymbol{g}_p^\top)^\top \in \mathbb{R}^{p \times Kq}, \tag{5.2.13}$$

and $\boldsymbol{g}_m \in \mathbb{R}^{Kq}$ represents the m-th row of $\boldsymbol{G}$ for $m = 1, \ldots, p$. The procedure for parameter estimation is summarized in Algorithm 2.

Remark 6 *For the linear-search step, $1/(L_\omega + \xi)$ is the step size, which needs to be as large as possible under constraint $M_{\boldsymbol{\tau},\omega,\xi}(\boldsymbol{B}, \boldsymbol{Z}) > Q_{\boldsymbol{\tau},\omega}(\boldsymbol{B})$ (by Lemma 2.1 in Toh and Yun (2010)). Denote $R^{(t+1)} = \nabla Q_{\boldsymbol{\tau},\omega}(\boldsymbol{B}^{(t+1)}) - \nabla Q_{\boldsymbol{\tau},\omega}(\boldsymbol{B}^{(t)}) - \bar{L}(\boldsymbol{B}^{(t+1)} - \boldsymbol{B}^{(t)})$. The termination conditions are defined as*

$$C_1 = \max_{1 \le k \le K} C_{1,k} \le \delta, \qquad C_2 = \max_{1 \le k \le K} C_{2,k} \le 5\delta,$$

and δ is a small positive value. We take $\delta = 10^{-5}$, and

$$C_{1,k} = \frac{\|R^{(t+1)}\|_F}{L_c^{(t)} \max(1, \|\boldsymbol{B}_k^{(t+1)}\|_F)}, \tag{5.2.14}$$

$$C_{2,k} = \frac{|\ \|\boldsymbol{Y}_k - \boldsymbol{X}_k \boldsymbol{B}_k^{(t+1)}\|_F - \|\boldsymbol{Y}_k - \boldsymbol{X}_k \boldsymbol{B}_k^{(t)}\|_F\ |}{\max(1, \|\boldsymbol{Y}_k\|_F)}. \tag{5.2.15}$$

Now, we establish the convergence of the MM algorithm.

Theorem 5.2.1 *Let $(\boldsymbol{B}^{(t)}{}_{t \ge 0})$ be the sequence generated by Algorithm (2). Assume that*

(A1) The objective function $\underline{L}_{\tau,\omega}(\boldsymbol{B})$ is continuous and bounded below.

(A2) The level set

$$\mathcal{L}_0 = \left\{ \boldsymbol{B} : \underline{L}_{\tau,\omega}(\boldsymbol{B}) \le \underline{L}_{\tau,\omega}(\boldsymbol{B}^{(0)}) \right\}$$

is compact.

(A3) For every $\boldsymbol{B}^{(t)}$, the MM subproblem

$$\min_{\boldsymbol{B}} \left\{ M_{\tau,\omega,\xi}(\boldsymbol{B}, \boldsymbol{B}^{(t)}) + \lambda_n \|\boldsymbol{B}\|_{\ell_1/\ell_2} \right\}$$

admits a unique minimizer.

(A4) The objective $\underline{L}_{\tau,\omega}(\boldsymbol{B})$ has a unique stationary point $\hat{\boldsymbol{B}}$ in $\mathcal{L}_0$.

Then

$$\boldsymbol{B}^{(t)} \to \hat{\boldsymbol{B}}, \quad t \to \infty.$$

Algorithm 2 Multi-platform Smoothing Quantile MM Algorithm

Input: Data $\{\boldsymbol{X}_k, \boldsymbol{Y}_k\}_{k=1}^K$; tuning parameters $\omega > 0$, $\xi > 0$, $\tau > 0$, $\lambda_n > 0$, $0 < \alpha < 1$; maximum iterations $T_{\max}$; tolerance δ.
Output: Estimated quantile regression coefficient matrix $\widehat{\boldsymbol{B}}$.

Initialization:

Set $\boldsymbol{B}_k^{(0)} = \mathbf{0}_{p_k \times q}$ for $k = 1, \ldots, K$, and define

$$\boldsymbol{B}^{(0)} = (\boldsymbol{B}_1^{(0)}, \boldsymbol{B}_2^{(0)}, \ldots, \boldsymbol{B}_K^{(0)}) \in \mathbb{R}^{p \times Kq}.$$

Compute the Lipschitz constant $\underline{L}_c$ according to (5.2.8) (lower smoothing). Set

$$L_0 = \underline{L}_c + \xi, \qquad L_{\min} = 0.01 L_0,$$

and initialize

$$L^{(0)} = \max(\alpha L_0,\, L_{\min}).$$

Main Iteration(for $t = 0$ to $T_{\max}$) do:

Linesearch step:
Initialize $L^{(\ell)} = L^{(0)}$. **For** $\ell = 1$ **to** 100 **do:**

- Compute the proximal MM update

$$\boldsymbol{B}^{\text{temp}} = \boldsymbol{P}_{L^{(\ell)}}\left(\boldsymbol{B}^{(t)}\right).$$

- Evaluate the smoothed loss $\underline{Q}_{\tau,\omega}(\boldsymbol{B}^{\text{temp}})$.
- Check the termination conditions:

$$\underline{Q}_{\tau,\omega}(\boldsymbol{B}^{\text{temp}}) \le M_{\tau,\omega,\xi}\left(\boldsymbol{B}^{\text{temp}},\, \boldsymbol{B}^{(t)}\right).$$

If the MM condition holds or $\ell = 100$, then break; otherwise update $L^{(\ell+1)} = L^{(\ell)}/\alpha$.

Stopping criterion: Compute $C_{1,k}$ and $C_{2,k}$ for $k = 1, \ldots, K$. If

$$\max_{1 \le k \le K} C_{1,k} \le \delta \quad \text{and} \quad \max_{1 \le k \le K} C_{2,k} \le 5\,\delta,$$

then stop,else update $\boldsymbol{B}^{(t+1)} = \boldsymbol{B}^{\text{temp}}$.

Return: $\widehat{\boldsymbol{B}} = \boldsymbol{B}^{(t+1)}$.

5.3 Simulation Study

In our study, we proposed the lower-smoothing (LS) and upper-smoothing (US) quantile regression methods via data integration (DI), respectively named as DI-LS and DI-US. To further evaluate the impact of platform heterogeneity, we benchmark our proposed methods against two baseline strategies for multi-platform data, i.e., the separated platform analysis (SPA) that analyzes each platform independently, and the naive pooled data analysis (nPD) that naively merges all platform data into a single dataset for analysis. After that, we apply our method to the separated ($q = 1$) and naive pooled datasets ($K = 1$), resulting in four benchmark methods, say SPA-Low, SPA-Up, nPD-Low and nPD-Up.

We remark that the regularization parameter λ is selected from a candidate set by minimizing an adjusted Akaike Information Criterion (A-AIC) (Chen and Chen, 2024), i.e.,

$$\text{A-AIC}(\lambda_n) = \log Q_{\boldsymbol{\tau}}(\hat{\boldsymbol{B}}) + \log(n)\frac{\text{supp}(\hat{\boldsymbol{B}})}{n}, \tag{5.3.1}$$

where $\text{supp}(\hat{\boldsymbol{B}})$ denotes the number of nonzero coefficients in $\hat{\boldsymbol{B}}$. The candidate values for λ are sampled at equally intervals between $\lambda_{\min}$ to $\lambda_{\max}$.

5.3.1 Evaluation metrics

To effectively evaluate methods performance, we employ five key metrics. Assessment focuses on two primary primary aspects: (i) estimation and prediction accuracy, measured by squared error (SE) and prediction error (PE), and (ii) variable selection accuracy, measured by false positive rate (FPR) and false negative rate (FNR). For these four accuracy metrics, lower values indicate better performance. Additionally, we report computational time (Time) to quantify algorithmic efficiency. Formal definitions are provided below.

(1) Root squared error (RSE):

$$\text{RSE}_B \;=\; \|\hat{\boldsymbol{B}} - \boldsymbol{B}^{\text{true}}\|_F,$$

where $\hat{\boldsymbol{B}}$ and $\boldsymbol{B}^{\text{true}}$ are the estimated and true regression coefficient matrices, respectively. The Frobenius norm $\|.\|_F$ measures the overall discrepancy between the estimated and true coefficients. The denominator normalizes the error by the total number of parameters across all platforms (K), predictors (p), and responses (q), making RMSE comparable across different dimensions.

(2) Prediction error (PE):

$$\text{PE} \;=\; \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{q}\rho_{\tau_j}\Big(y_{kij}-\boldsymbol{x}_{ki}^{\top}\hat{\boldsymbol{b}}_{kj}(\tau_j)\Big)\,, \tag{5.3.2}$$

where $y_{kij}, \boldsymbol{x}_{ki}$ are out-of samples, and $\hat{\boldsymbol{b}}_{kj}(\tau_j)$ is the estimated regression vector for response j at quantile level τ_j on platform k. This metric extends the approach of Park et al. (2023) to the K-platform setting. We adopt this definition because it directly incorporates the quantile loss function ρ, making it more suitable for evaluating quantile regression models than prediction error metrics based on mean squared error.

(3) False positive ratio (FPR):

$$\text{FPR} \;=\; \frac{\#\{\text{truly zero coefficients incorrectly estimated as nonzero}\}}{\#\{\text{truly zero coefficients}\}}.$$

(4) False negative ratio (FNR):

$$\text{FNR} \;=\; \frac{\#\{\text{truly nonzero coefficients incorrectly estimated as zero}\}}{\#\{\text{truly nonzero coefficients}\}}.$$

(5) Computational efficiency (Time): it is recorded as the CPU time (in seconds) required for a complete estimation of model parameters.

5.3.2 Multivariate quantile regression for multi-platform data

We first investigate the multivariate response case with $K = 2$, $q = 50$, $p = 500$. The simulated data $(\boldsymbol{X}, \boldsymbol{Y})$ are generated from model (5.2.1), where $\boldsymbol{x}_{ki}$ follows $\mathcal{MVN}_p(\mathbf{0},\, \boldsymbol{\Omega})$, with $\Omega_{i,j} = 0.5^{|i-j|}$ for $i, j = 1, \ldots, p$. Denote $\boldsymbol{e}_{\cdot i} = \big(\boldsymbol{e}_{1i}^{\top}, \ldots, \boldsymbol{e}_{Ki}^{\top}\big)^{\top} \in \mathbb{R}^{Kq}$. We consider the following three types of random errors, i.e.,

- Multivariate normal distribution: $\boldsymbol{e}_{\cdot i} \sim \mathcal{MVN}_{Kq}(\mathbf{0},\ \boldsymbol{\Sigma})$;
- Multivariate t distribution with degrees of freedom 3 ($\nu = 3$): $\boldsymbol{e}_{\cdot i} \sim \mathcal{MVT}_{\nu}\big(\mathbf{0},\ \frac{\nu-2}{\nu}\boldsymbol{\Sigma}\big)$;
- Multivariate asymmetric Laplace distribution (MAL): $\mathcal{MAL}(\mathbf{0},\ \boldsymbol{\Sigma})$.

Table 5.1: The averaged RSE, PE, FPR, FNR and Time (standard errors in parentheses) over 100 simulation replications given different quantile levels (τ), with $q = 5$, $p = 500$, and $K = 2$, where data between two platforms are independent ($d_2 = 0$).

τ	Errors	Methods	RSE_B	PE	FPR	FNR	Time
0.25	$\mathcal{MVN}$	SPA-LS	3.3521(.505)	3.8617(.201)	0.0016(.001)	0.0000(.000)	0.0118(.004)
		nPD-LS	10.115(1.51)	6.7521(.917)	0.0005(.001)	0.0000(.000)	0.0435(.016)
		DI-LS	3.2093(.414)	3.7702(.172)	0.0005(.001)	0.0000(.000)	0.0176(.006)
		SPA-US	3.4938(.432)	3.8259(.185)	0.0025(.002)	0.0000(.000)	0.0365(.009)
		nPD-US	6.5030(.616)	4.5075(.362)	0.0007(.001)	0.0000(.000)	0.0738(.020)
		DI-US	3.3209(.398)	3.7363(.169)	0.0008(.001)	0.0000(.000)	0.0467(.012)
	$\mathcal{MVT}$	SPA-LS	4.2630(.772)	5.4287(.354)	0.0014(.001)	0.0000(.000)	0.0147(.006)
		nPD-LS	11.062(1.75)	8.4240(.943)	0.0005(.001)	0.0000(.000)	0.0552(.021)
		DI-LS	4.0362(.603)	5.3068(.305)	0.0004(.001)	0.0000(.000)	0.0191(.007)
		SPA-US	4.5014(.641)	5.4301(.315)	0.0019(.002)	0.0000(.000)	0.0478(.011)
		nPD-US	7.3896(.851)	6.2262(.471)	0.0008(.001)	0.0000(.000)	0.0932(.030)
		DI-US	4.1899(.603)	5.2910(.303)	0.0005(.001)	0.0000(.000)	0.0546(.015)
	$\mathcal{MAL}$	SPA-LS	2.8289(.423)	3.5532(.212)	0.0014(.001)	0.0000(.000)	0.0106(.003)
		nPD-LS	9.6864(1.52)	6.4161(.941)	0.0005(.001)	0.0000(.000)	0.0409(.015)
		DI-LS	2.7252(.362)	3.4753(.196)	0.0004(.001)	0.0000(.000)	0.0153(.004)
		SPA-US	3.0023(.383)	3.5358(.206)	0.0019(.002)	0.0000(.000)	0.0328(.007)
		nPD-US	6.1464(.546)	4.1537(.358)	0.0009(.001)	0.0000(.000)	0.0612(.016)
		DI-US	2.8116(.347)	3.4409(.190)	0.0004(.001)	0.0000(.000)	0.0412(.012)
0.50	$\mathcal{MVN}$	SPA-LS	2.2668(.301)	4.5823(.176)	0.0039(.002)	0.0000(.000)	0.0098(.002)
		nPD-LS	7.4235(.976)	6.4624(.676)	0.0005(.001)	0.0000(.000)	0.0450(.015)
		DI-LS	2.0719(.289)	4.4898(.166)	0.0008(.001)	0.0000(.000)	0.0161(.004)
		SPA-US	2.4810(.344)	4.6753(.207)	0.0021(.001)	0.0000(.000)	0.0353(.006)
		nPD-US	5.6002(.295)	5.1259(.268)	0.0008(.001)	0.0000(.000)	0.0812(.018)
		DI-US	2.1625(.309)	4.5299(.176)	0.0006(.001)	0.0000(.000)	0.0490(.011)
	$\mathcal{MVT}$	SPA-LS	2.6276(.381)	6.1616(.307)	0.0035(.002)	0.0000(.000)	0.0112(.003)
		nPD-LS	8.1992(1.29)	8.3794(.876)	0.0004(.001)	0.0000(.000)	0.0554(.018)
		DI-LS	2.4270(.368)	6.0598(.307)	0.0006(.001)	0.0000(.000)	0.0190(.006)
		SPA-US	2.9415(.418)	6.2965(.319)	0.0017(.001)	0.0000(.000)	0.0429(.010)
		nPD-US	6.0387(.499)	6.8611(.435)	0.0008(.001)	0.0000(.000)	0.0978(.024)
		DI-US	2.5525(.366)	6.1171(.297)	0.0004(.001)	0.0000(.000)	0.0569(.015)
	$\mathcal{MAL}$	SPA-LS	1.6549(.234)	3.9747(.194)	0.0037(.002)	0.0000(.000)	0.0085(.002)
		nPD-LS	6.9985(.946)	5.9021(.728)	0.0003(.000)	0.0000(.000)	0.0422(.013)
		DI-LS	1.5395(.216)	3.9110(.187)	0.0006(.001)	0.0000(.000)	0.0148(.004)
		SPA-US	1.8553(.270)	4.0635(.198)	0.0017(.001)	0.0000(.000)	0.0291(.006)
		nPD-US	5.2166(.220)	4.4743(.265)	0.0007(.001)	0.0000(.000)	0.0665(.015)
		DI-US	1.5909(.237)	3.9392(.189)	0.0002(.001)	0.0000(.000)	0.0404(.011)
0.75	$\mathcal{MVN}$	SPA-LS	3.4125(.496)	3.8802(.162)	0.0014(.001)	0.0000(.000)	0.0122(.004)
		nPD-LS	10.202(1.47)	6.8048(.903)	0.0006(.001)	0.0020(.020)	0.0496(.017)
		DI-LS	3.2175(.387)	3.7687(.162)	0.0003(.001)	0.0000(.000)	0.0159(.005)
		SPA-US	3.5483(.409)	3.8465(.166)	0.0021(.001)	0.0000(.000)	0.0381(.009)
		nPD-US	6.5169(.555)	4.5170(.301)	0.0007(.001)	0.0000(.000)	0.0720(.019)
		DI-US	3.3165(.365)	3.7340(.157)	0.0006(.001)	0.0000(.000)	0.0450(.011)
	$\mathcal{MVT}$	SPA-LS	4.0718(.591)	5.4073(.330)	0.0016(.001)	0.0000(.000)	0.0134(.004)
		nPD-LS	11.047(1.81)	8.4234(1.06)	0.0005(.001)	0.0000(.000)	0.0578(.027)
		DI-LS	3.9139(.528)	5.3025(.315)	0.0004(.001)	0.0000(.000)	0.0177(.006)
		SPA-US	4.3772(.576)	5.4335(.317)	0.0018(.002)	0.0000(.000)	0.0472(.011)
		nPD-US	7.3427(.837)	6.2281(.507)	0.0006(.001)	0.0000(.000)	0.0935(.031)
		DI-US	4.0770(.538)	5.2888(.312)	0.0003(.001)	0.0000(.000)	0.0557(.015)
	$\mathcal{MAL}$	SPA-LS	2.8485(.473)	3.5594(.205)	0.0016(.001)	0.0000(.000)	0.0110(.004)
		nPD-LS	9.7848(1.51)	6.4625(.959)	0.0005(.001)	0.0000(.000)	0.0425(.016)
		DI-LS	2.7470(.383)	3.4895(.191)	0.0004(.001)	0.0000(.000)	0.0153(.004)
		SPA-US	3.0234(.419)	3.5439(.196)	0.0020(.001)	0.0000(.000)	0.0341(.008)
		nPD-US	6.1394(.535)	4.1497(.338)	0.0008(.001)	0.0000(.000)	0.0606(.016)
		DI-US	2.8089(.360)	3.4455(.184)	0.0006(.001)	0.0000(.000)	0.0407(.011)

Table 5.2: The averaged RSE, PE, FPR, FNR and Time (standard errors in parentheses) over 100 simulation replications given different quantile levels (τ), with $q = 5$, $p = 500$, and $K = 2$, where data between two platforms are weakly correlated ($d_2 = 0.3$).

τ	Errors	Methods	RSE_B	PE	FPR	FNR	Time
0.25	$\mathcal{MVN}$	SPA-LS	3.3582(.512)	3.8736(.210)	0.0016(.001)	0.0000(.000)	0.0119(.004)
		nPD-LS	10.1196(1.499)	6.7628(.907)	0.0005(.001)	0.0000(.000)	0.0454(.017)
		DI-LS	3.2167(.410)	3.7828(.179)	0.0006(.001)	0.0000(.000)	0.0164(.006)
		SPA-US	3.5047(.417)	3.8392(.183)	0.0024(.002)	0.0000(.000)	0.0364(.008)
		nPD-US	6.5074(.625)	4.5156(.370)	0.0007(.001)	0.0000(.000)	0.0736(.018)
		DI-US	3.3280(.412)	3.7472(.180)	0.0007(.001)	0.0000(.000)	0.0459(.013)
	$\mathcal{MVT}$	SPA-LS	4.2707(.780)	5.4463(.344)	0.0015(.001)	0.0010(.010)	0.0141(.005)
		nPD-LS	11.0897(1.781)	8.4617(.980)	0.0004(.001)	0.0000(.000)	0.0556(.023)
		DI-LS	4.0485(.608)	5.3261(.298)	0.0004(.001)	0.0000(.000)	0.0182(.006)
		SPA-US	4.5523(.678)	5.4655(.330)	0.0017(.002)	0.0000(.000)	0.0493(.012)
		nPD-US	7.4568(.918)	6.2861(.502)	0.0005(.001)	0.0000(.000)	0.1008(.036)
		DI-US	4.2082(.610)	5.3125(.302)	0.0005(.001)	0.0000(.000)	0.0555(.016)
	$\mathcal{MAL}$	SPA-LS	2.8456(.425)	3.5584(.220)	0.0015(.001)	0.0000(.000)	0.0105(.003)
		nPD-LS	9.6954(1.525)	6.4257(.944)	0.0005(.001)	0.0000(.000)	0.0416(.016)
		DI-LS	2.7438(.366)	3.4856(.201)	0.0002(.001)	0.0000(.000)	0.0150(.004)
		SPA-US	2.9826(.376)	3.5286(.214)	0.0022(.002)	0.0000(.000)	0.0329(.007)
		nPD-US	6.1749(.548)	4.1802(.359)	0.0007(.001)	0.0000(.000)	0.0643(.016)
		DI-US	2.8140(.358)	3.4433(.199)	0.0004(.001)	0.0000(.000)	0.0393(.011)
0.50	$\mathcal{MVN}$	SPA-LS	2.2549(.297)	4.5877(.180)	0.0039(.002)	0.0000(.000)	0.0098(.002)
		nPD-LS	7.4218(.971)	6.4724(.675)	0.0004(.001)	0.0000(.000)	0.0462(.015)
		DI-LS	2.0612(.292)	4.4962(.171)	0.0009(.001)	0.0000(.000)	0.0160(.004)
		SPA-US	2.4884(.333)	4.6879(.199)	0.0020(.002)	0.0000(.000)	0.0350(.007)
		nPD-US	5.6051(.289)	5.1416(.262)	0.0008(.001)	0.0000(.000)	0.0816(.018)
		DI-US	2.1475(.285)	4.5352(.176)	0.0007(.001)	0.0000(.000)	0.0482(.011)
	$\mathcal{MVT}$	SPA-LS	2.6362(.402)	6.1800(.309)	0.0036(.002)	0.0000(.000)	0.0110(.003)
		nPD-LS	8.2187(1.305)	8.4114(.888)	0.0003(.001)	0.0000(.000)	0.0546(.017)
		DI-LS	2.4358(.385)	6.0791(.310)	0.0007(.001)	0.0000(.000)	0.0184(.006)
		SPA-US	2.9540(.445)	6.3185(.326)	0.0017(.001)	0.0000(.000)	0.0438(.009)
		nPD-US	6.0775(.561)	6.9082(.475)	0.0007(.001)	0.0000(.000)	0.1014(.036)
		DI-US	2.5448(.386)	6.1286(.304)	0.0004(.001)	0.0000(.000)	0.0583(.016)
	$\mathcal{MAL}$	SPA-LS	1.6643(.241)	3.9813(.200)	0.0036(.002)	0.0000(.000)	0.0083(.002)
		nPD-LS	7.0030(.950)	5.9100(.735)	0.0003(.000)	0.0000(.000)	0.0415(.013)
		DI-LS	1.5467(.220)	3.9171(.193)	0.0006(.001)	0.0000(.000)	0.0148(.003)
		SPA-US	1.8728(.276)	4.0742(.207)	0.0016(.001)	0.0000(.000)	0.0292(.006)
		nPD-US	5.2379(.237)	4.5012(.273)	0.0006(.001)	0.0000(.000)	0.0686(.016)
		DI-US	1.5691(.226)	3.9316(.197)	0.0004(.001)	0.0000(.000)	0.0392(.010)
0.75	$\mathcal{MVN}$	SPA-LS	3.4029(.506)	3.8826(.175)	0.0014(.001)	0.0000(.000)	0.0124(.004)
		nPD-LS	10.2075(1.480)	6.8044(.915)	0.0006(.001)	0.0020(.020)	0.0486(.018)
		DI-LS	3.2153(.398)	3.7773(.176)	0.0004(.001)	0.0000(.000)	0.0159(.004)
		SPA-US	3.5440(.416)	3.8526(.167)	0.0023(.002)	0.0000(.000)	0.0380(.009)
		nPD-US	6.5056(.578)	4.5111(.316)	0.0009(.001)	0.0000(.000)	0.0735(.021)
		DI-US	3.3169(.390)	3.7421(.162)	0.0006(.001)	0.0000(.000)	0.0466(.012)
	$\mathcal{MVT}$	SPA-LS	4.0719(.625)	5.4147(.347)	0.0017(.001)	0.0000(.000)	0.0137(.005)
		nPD-LS	11.0485(1.828)	8.4314(1.065)	0.0005(.001)	0.0000(.000)	0.0557(.031)
		DI-LS	3.9182(.532)	5.3124(.328)	0.0003(.001)	0.0000(.000)	0.0178(.006)
		SPA-US	4.3694(.580)	5.4342(.333)	0.0019(.002)	0.0000(.000)	0.0461(.010)
		nPD-US	7.3534(.820)	6.2431(.493)	0.0006(.001)	0.0000(.000)	0.0954(.034)
		DI-US	4.0817(.548)	5.2977(.317)	0.0004(.001)	0.0000(.000)	0.0565(.017)
	$\mathcal{MAL}$	SPA-LS	2.8516 (.477)	3.5668 (.210)	0.0015 (.001)	0.0000 (.000)	0.0108 (.003)
		nPD-LS	9.7866 (1.505)	6.4598 (.945)	0.0006 (.001)	0.0000 (.000)	0.0411 (.014)
		DI-LS	2.7446 (.388)	3.4912 (.199)	0.0005 (.001)	0.0000 (.000)	0.0149 (.004)
		SPA-US	3.0274 (.412)	3.5513 (.195)	0.0021 (.001)	0.0000 (.000)	0.0342 (.008)
		nPD-US	6.1684 (.551)	4.1765 (.356)	0.0007 (.001)	0.0000 (.000)	0.0645 (.018)
		DI-US	2.8211 (.369)	3.4547 (.195)	0.0006 (.001)	0.0000 (.000)	0.0409 (.011)

In this way, all three distributions are standardized to share the same mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$, ensuring comparability, i.e.,

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & d_2 & \cdots & d_2 \\ d_2 & \mathbf{\Sigma}_2 & \cdots & d_2 \\ \vdots & \vdots & \ddots & \vdots \\ d_2 & d_2 & \cdots & \mathbf{\Sigma}_K \end{bmatrix}, \text{ and } \mathbf{\Sigma}_k = \begin{bmatrix} 1 & d_1 & \cdots & d_1 \\ d_1 & 1 & \cdots & d_1 \\ \vdots & \vdots & \ddots & \vdots \\ d_1 & d_1 & \cdots & 1 \end{bmatrix}$$

for $k = 1, \dots K$, where d_1 and d_2 represent the covariance within and between platforms, respectively. Herein, we consider the following two scenarios, i.e.,

- Data are strongly correlated within platforms, but independent between platforms: $d_1 = 0.8, d_2 = 0$;

- Data are strongly correlated within platforms, but weakly correlated between platforms: $d_1 = 0.8, d_2 = 0.3$.

Let $s = \lfloor 0.01\, p \rfloor$ be the number of nonzero rows (sparsity level), $\lfloor x \rfloor$ represents the largest integer less than or equal to x. For each platform $k = 1, \dots, K$, the regression coefficient matrix $\boldsymbol{B}_k \in \mathbb{R}^{p \times q}$ is given by (Chen and Chen, 2024)

$$\boldsymbol{B}_k = \begin{pmatrix} \mathbf{\Lambda}\, \boldsymbol{B}_{k1} \\ \mathbf{0}_{(p-s)\times q} \end{pmatrix},$$

where the entries of $\boldsymbol{B}_{k1} \in \mathbb{R}^{s \times q}$ are i.i.d. random variables drown from $\{-1, 1\}$ with probability 0.5 and $\mathbf{\Lambda} \in \mathbb{R}^{s \times s}$ is diagonal with linearly increasing entries, i.e.,

$$\mathbf{\Lambda} = \mathrm{diag}\Big(\tfrac{10}{s}, \tfrac{20}{s}, \cdots, \tfrac{10(s-1)}{s}, 10\Big) = \mathrm{diag}\Big(\tfrac{10}{s} \cdot 1, \tfrac{10}{s} \cdot 2, \cdots, \tfrac{10}{s} \cdot s\Big).$$

Following the specified settings, a dataset of $n = 400$ samples is generated per platform. Each dataset is subsequently divided into training (50%) and test (50%) subsets.

Table 5.1 reports the simulation results where data between two platforms are independent ($d_2 = 0$). Overall, the SPA (SPA-LS, SPA-US) and DI (DI-LS, DI-US) methods exhibit comparable performance, and both consistently outperform nPD (nPD-LS, nPD-US) by achieving smaller RSE and PE values. This result indicates that nPD, which naively merges data from all platforms into a single dataset, is inappropriate and does not fully utilize cross-platform information. From the perspective of variable selection, as compared with SPA, the DI method demonstrates clear advantages based on the FPR and FNR values. It yields a sparser model with higher selection accuracy while maintaining strong predictive

performance. These improvements arise because DI more effectively integrates information across multiple platforms, which leads to its superior overall performance.

In Table 5.2, we further report the simulation results for the case where data from two platforms exhibit weak correlation ($d_2 = 0.3$). Clearly, the overall conclusions remain consistent with those from the independent case (Table 5.1), indicating that the proposed data integration methods maintain stable performance even in the presence of mild correlations across platforms.

5.3.3 Unvariate quantile regression for multi-platform data

We now examine the reduced cases involving a univariate response from multi-platforms, by letting $K = 2$, $p = 500$, and $q = 1$. The data generation follows the procedure described in Section 3.2. The primary difference is the absence of within-platform correlation, as the response is univariate.

Table 5.3 presents the simulation results for the univariate response case with $q = 1$ when the two platforms are independent with $d_2 = 0$. The overall conclusions are consistent with those observed in the multivariate response settings, except that the choice of smoothing functions has a more pronounced impact. Specifically, when the same smoothing functions are employed, the proposed multi-platform algorithms consistently achieve smaller RSE, PE, and FPR values than the competing methods. For example, DI-LS outperforms SPA-LS and nPD-LS, while DI-US shows superior performance compared to DI-LS. In addition, methods based on the upper smoothing function generally exhibit better estimation accuracy. All methods perform well under the normal and AL distributions, whereas a modest degradation in performance is observed under the Student-t distribution. In particular, the heavy-tailed nature of the Student-t distribution has a more pronounced impact at the quantile levels $\tau = 0.25$ and $\tau = 0.75$, indicating that the estimation of more extreme quantiles is more challenging and more sensitive to heavy-tailed behavior. We further report the simulation results for the univariate response cases when platforms exhibit weak correlation ($d_2 = 0.3$) in Table 5.4. The results remain largely consistent with those in the independent scenario, indicating that the proposed multi-platform algorithms maintain stable and robust performance even in the presence of mild cross-platform dependence.

5.4 Real data applications

5.4.1 U.S. stock market indices

Zhang et al. (2024) regarded three financial indices, say S&P 500, Dow Jones, and VIX,

Table 5.3: The averaged RSE, PE, FPR, FNR and Time (standard errors in parentheses) over 100 simulation replications given different quantile levels (τ), with $q = 1$, $p = 500$, and $K = 2$, where data between two platforms are independent ($d_2 = 0$).

τ	Errors	Methods	RSE_B	PE	FPR	FNR	Time
0.25	$\mathcal{MVN}$	SPA-LS	7.4209(.024)	2.3843(.044)	0.0022(.002)	0.1290(.124)	0.0467(.027)
		nPD-LS	9.1466(.748)	2.8960(.854)	0.0013(.001)	0.0720(.093)	0.1021(.067)
		DI-LS	4.6404(.014)	1.5386(.510)	0.0004(.001)	0.0560(.090)	0.0540(.057)
		SPA-US	2.9649(.384)	1.0580(.255)	0.0048(.003)	0.0270(.049)	0.0311(.021)
		nPD-US	5.7458(.234)	1.7427(.614)	0.0019(.002)	0.0460(.086)	0.0795(.058)
		DI-US	2.3701(.086)	0.9316(.195)	0.0014(.002)	0.0060(.034)	0.0412(.050)
	$\mathcal{MVT}$	SPA-LS	7.7117(.741)	2.6721(.997)	0.0020(.002)	0.1420(.117)	0.0443(.028)
		nPD-LS	9.6240(.928)	3.2214(.817)	0.0011(.001)	0.0850(.111)	0.1042(.064)
		DI-LS	5.3121(.020)	1.8846(.480)	0.0006(.001)	0.0480(.086)	0.0598(.062)
		SPA-US	3.6806(.436)	1.4144(.268)	0.0042(.003)	0.0610(.067)	0.0360(.018)
		nPD-US	6.4626(.408)	2.1293(.594)	0.0018(.002)	0.0450(.086)	0.1049(.061)
		DI-US	3.0085(.253)	1.2805(.223)	0.0014(.002)	0.0200(.060)	0.0386(.034)
	$\mathcal{MAL}$	SPA-LS	7.8690(.816)	2.4883(.058)	0.0023(.002)	0.1470(.121)	0.0419(.029)
		nPD-LS	9.5436(.955)	3.0578(.913)	0.0013(.001)	0.0930(.125)	0.1103(.069)
		DI-LS	4.9141(.097)	1.6057(.546)	0.0005(.001)	0.0420(.082)	0.0634(.060)
		SPA-US	2.9801(.447)	1.0324(.282)	0.0044(.003)	0.0330(.055)	0.0285(.018)
		nPD-US	6.0297(.440)	1.8106(.661)	0.0017(.002)	0.0390(.083)	0.0854(.065)
		DI-US	2.3222(.158)	0.8987(.228)	0.0017(.002)	0.0100(.044)	0.0440(.050)
0.50	$\mathcal{MVN}$	SPA-LS	7.7430(.873)	2.9227(.086)	0.0021(.002)	0.1540(.128)	0.0484(.027)
		nPD-LS	7.4640(.677)	2.8683(.957)	0.0016(.001)	0.0170(.055)	0.0884(.064)
		DI-LS	2.9372(.486)	1.4317(.434)	0.0006(.001)	0.0000(.000)	0.0431(.051)
		SPA-US	1.5528(.529)	1.0413(.116)	0.0062(.003)	0.0000(.000)	0.0211(.008)
		nPD-US	4.0930(.707)	1.5838(.540)	0.0017(.001)	0.0000(.000)	0.0617(.049)
		DI-US	1.3215(.409)	0.9828(.083)	0.0021(.002)	0.0000(.000)	0.0276(.008)
	$\mathcal{MVT}$	SPA-LS	7.5652(.072)	3.0998(.086)	0.0019(.002)	0.1540(.133)	0.0465(.026)
		nPD-LS	7.9366(.814)	3.2593(.910)	0.0014(.001)	0.0160(.039)	0.0930(.065)
		DI-LS	3.5163(.651)	1.8470(.445)	0.0007(.001)	0.0000(.000)	0.0478(.059)
		SPA-US	1.9900(.702)	1.4204(.162)	0.0061(.004)	0.0000(.000)	0.0257(.008)
		nPD-US	4.7693(.961)	2.0566(.564)	0.0021(.002)	0.0000(.000)	0.0836(.061)
		DI-US	1.6364(.564)	1.3303(.127)	0.0019(.002)	0.0000(.000)	0.0304(.008)
	$\mathcal{MAL}$	SPA-LS	7.9248(.885)	2.8767(.065)	0.0019(.002)	0.1670(.134)	0.0454(.027)
		nPD-LS	7.7969(.690)	2.9732(.950)	0.0014(.001)	0.0130(.042)	0.0906(.069)
		DI-LS	3.0268(.578)	1.4225(.465)	0.0007(.001)	0.0000(.000)	0.0391(.051)
		SPA-US	1.3327(.577)	0.9310(.121)	0.0061(.004)	0.0000(.000)	0.0187(.009)
		nPD-US	4.1426(.701)	1.5508(.531)	0.0020(.002)	0.0000(.000)	0.0577(.052)
		DI-US	1.0873(.402)	0.8692(.081)	0.0017(.002)	0.0000(.000)	0.0241(.007)
0.75	$\mathcal{MVN}$	SPA-LS	7.5570(.883)	2.4571(.060)	0.0023(.002)	0.1370(.118)	0.0442(.027)
		nPD-LS	9.2062(.683)	2.9179(.795)	0.0011(.001)	0.0930(.112)	0.0933(.066)
		DI-LS	4.7356(.863)	1.5447(.471)	0.0004(.001)	0.0460(.085)	0.0669(.063)
		SPA-US	3.0476(.288)	1.0634(.235)	0.0045(.003)	0.0340(.050)	0.0310(.020)
		nPD-US	5.8351(.267)	1.7467(.594)	0.0019(.002)	0.0400(.078)	0.0772(.057)
		DI-US	2.4284(.999)	0.9372(.186)	0.0019(.002)	0.0040(.028)	0.0402(.046)
	$\mathcal{MVT}$	SPA-LS	7.3624(.014)	2.5629(.008)	0.0018(.002)	0.1250(.107)	0.0399(.028)
		nPD-LS	9.5220(.836)	3.1604(.775)	0.0014(.001)	0.0690(.085)	0.1004(.065)
		DI-LS	5.2130(.918)	1.8578(.456)	0.0006(.001)	0.0380(.079)	0.0710(.068)
		SPA-US	3.6118(.279)	1.4117(.239)	0.0043(.003)	0.0560(.066)	0.0343(.017)
		nPD-US	6.3585(.273)	2.0892(.540)	0.0019(.002)	0.0380(.078)	0.0923(.058)
		DI-US	2.9266(.075)	1.2724(.205)	0.0015(.002)	0.0140(.051)	0.0353(.033)
	$\mathcal{MAL}$	SPA-LS	7.5445(.975)	2.3432(.008)	0.0021(.002)	0.1440(.126)	0.0420(.028)
		nPD-LS	9.5959(.894)	3.0019(.889)	0.0012(.001)	0.0790(.113)	0.1140(.072)
		DI-LS	4.9035(.065)	1.5697(.522)	0.0003(.001)	0.0400(.080)	0.0610(.060)
		SPA-US	3.0860(.432)	1.0507(.266)	0.0039(.003)	0.0430(.052)	0.0304(.022)
		nPD-US	6.0773(.257)	1.7863(.596)	0.0020(.002)	0.0420(.082)	0.0850(.061)
		DI-US	2.4222(.157)	0.9190(.216)	0.0015(.002)	0.0040(.028)	0.0383(.043)

Table 5.4: The averaged RSE, PE, FPR, FNR and Time (standard errors in parentheses) over 100 simulation replications given different quantile levels (τ), with $q = 1$, $p = 500$, and $K = 2$, where data between two platforms are weakly correlated ($d_2 = 0.3$).

τ	Errors	Methods	RMSE_B	PE	FPR	FNR	Time
0.25	$\mathcal{MVN}$	SPA-LS	7.4722(.401)	2.4040(.105)	0.0021(.002)	0.1310(.126)	0.0479(.026)
		nPD-LS	9.1309(.271)	2.8908(.836)	0.0013(.001)	0.0710(.091)	0.0990(.065)
		DI-LS	4.6360(.201)	1.5353(.509)	0.0004(.001)	0.0560(.090)	0.0533(.056)
		SPA-US	2.9668(.385)	1.0582(.255)	0.0046(.003)	0.0270(.049)	0.0318(.023)
		nPD-US	5.7367(.225)	1.7368(.615)	0.0020(.002)	0.0460(.086)	0.0792(.058)
		DI-US	2.3730(.108)	0.9329(.194)	0.0014(.002)	0.0080(.039)	0.0447(.055)
	$\mathcal{MVT}$	SPA-LS	7.7603(.375)	2.6886(.996)	0.0020(.002)	0.1430(.117)	0.0429(.027)
		nPD-LS	9.6448(.294)	3.2288(.816)	0.0010(.001)	0.0890(.114)	0.1060(.067)
		DI-LS	5.3201(.203)	1.8843(.488)	0.0005(.001)	0.0480(.086)	0.0628(.061)
		SPA-US	3.6731(.435)	1.4124(.270)	0.0043(.003)	0.0610(.067)	0.0353(.018)
		nPD-US	6.4627(.434)	2.1262(.593)	0.0020(.002)	0.0470(.086)	0.0961(.061)
		DI-US	3.0177(.126)	1.2807(.229)	0.0017(.002)	0.0180(.058)	0.0390(.036)
	$\mathcal{MAL}$	SPA-LS	7.9436(.384)	2.5123(.106)	0.0024(.002)	0.1470(.123)	0.0430(.029)
		nPD-LS	9.5266(.293)	3.0503(.902)	0.0012(.001)	0.0930(.121)	0.1027(.066)
		DI-LS	4.9072(.209)	1.6014(.544)	0.0005(.001)	0.0420(.082)	0.0660(.061)
		SPA-US	2.9968(.448)	1.0365(.283)	0.0044(.003)	0.0320(.055)	0.0308(.019)
		nPD-US	5.9428(.233)	1.7806(.624)	0.0022(.002)	0.0360(.079)	0.0778(.056)
		DI-US	2.3226(.152)	0.8987(.227)	0.0019(.002)	0.0100(.044)	0.0385(.048)
0.50	$\mathcal{MVN}$	SPA-LS	7.7488(.847)	2.9256(.077)	0.0021(.002)	0.1560(.127)	0.0475(.026)
		nPD-LS	7.4676(.660)	2.8760(.945)	0.0013(.001)	0.0140(.047)	0.0885(.061)
		DI-LS	2.9343(.490)	1.4338(.437)	0.0006(.001)	0.0000(.000)	0.0415(.052)
		SPA-US	1.5554(.556)	1.0433(.119)	0.0063(.003)	0.0000(.000)	0.0213(.008)
		nPD-US	4.0985(.721)	1.5881(.545)	0.0016(.001)	0.0000(.000)	0.0631(.051)
		DI-US	1.3164(.412)	0.9830(.082)	0.0019(.002)	0.0000(.000)	0.0273(.008)
	$\mathcal{MVT}$	SPA-LS	7.5170(.119)	3.0863(.101)	0.0019(.002)	0.1540(.132)	0.0453(.027)
		nPD-LS	7.9223(.776)	3.2561(.893)	0.0013(.001)	0.0140(.038)	0.0947(.066)
		DI-LS	3.5222(.667)	1.8525(.452)	0.0007(.001)	0.0000(.000)	0.0458(.056)
		SPA-US	1.9742(.702)	1.4175(.158)	0.0066(.004)	0.0000(.000)	0.0262(.008)
		nPD-US	4.8095(.020)	2.0686(.585)	0.0021(.001)	0.0000(.000)	0.0851(.062)
		DI-US	1.6461(.572)	1.3346(.129)	0.0018(.002)	0.0000(.000)	0.0319(.010)
	$\mathcal{MAL}$	SPA-LS	7.9485(.874)	2.8834(.058)	0.0019(.002)	0.1640(.133)	0.0459(.027)
		nPD-LS	7.8045(.694)	2.9758(.949)	0.0014(.001)	0.0140(.043)	0.0910(.068)
		DI-LS	3.0252(.568)	1.4190(.462)	0.0007(.001)	0.0000(.000)	0.0338(.044)
		SPA-US	1.3354(.565)	0.9309(.112)	0.0061(.004)	0.0000(.000)	0.0188(.009)
		nPD-US	4.1498(.714)	1.5526(.533)	0.0019(.001)	0.0000(.000)	0.0597(.054)
		DI-US	1.0942(.398)	0.8711(.082)	0.0015(.002)	0.0000(.000)	0.0236(.007)
0.75	$\mathcal{MVN}$	SPA-LS	7.5511(.897)	2.4573(.065)	0.0023(.002)	0.1330(.116)	0.0423(.027)
		nPD-LS	9.2173(.709)	2.9253(.812)	0.0011(.001)	0.0960(.119)	0.0998(.064)
		DI-LS	4.7428(.858)	1.5469(.472)	0.0005(.001)	0.0460(.085)	0.0677(.063)
		SPA-US	3.0367(.284)	1.0642(.236)	0.0046(.003)	0.0330(.049)	0.0300(.021)
		nPD-US	5.8364(.289)	1.7489(.595)	0.0019(.002)	0.0420(.081)	0.0756(.056)
		DI-US	2.4322(.002)	0.9397(.185)	0.0017(.002)	0.0060(.034)	0.0368(.044)
	$\mathcal{MVT}$	SPA-LS	7.3183(.038)	2.5564(.016)	0.0018(.002)	0.1270(.107)	0.0404(.027)
		nPD-LS	9.5656(.892)	3.1833(.799)	0.0012(.001)	0.0770(.095)	0.0985(.063)
		DI-LS	5.2120(.922)	1.8554(.457)	0.0006(.001)	0.0360(.077)	0.0672(.065)
		SPA-US	3.6145(.283)	1.4131(.237)	0.0042(.003)	0.0560(.064)	0.0335(.017)
		nPD-US	6.3779(.314)	2.0940(.548)	0.0019(.002)	0.0380(.078)	0.0935(.062)
		DI-US	2.9288(.083)	1.2742(.203)	0.0012(.001)	0.0100(.044)	0.0416(.040)
	$\mathcal{MAL}$	SPA-LS	7.5524(.961)	2.3506(.008)	0.0021(.002)	0.1430(.126)	0.0435(.027)
		nPD-LS	9.5855(.892)	2.9972(.886)	0.0013(.001)	0.0770(.111)	0.1154(.071)
		DI-LS	4.9011(.074)	1.5702(.525)	0.0003(.001)	0.0400(.080)	0.0574(.059)
		SPA-US	3.0911(.435)	1.0502(.265)	0.0038(.003)	0.0450(.052)	0.0285(.021)
		nPD-US	6.0640(.224)	1.7796(.588)	0.0021(.002)	0.0420(.082)	0.0791(.057)
		DI-US	2.4231(.168)	0.9182(.219)	0.0015(.002)	0.0040(.028)	0.0411(.048)

Table 5.5: Methods comparison based on average prediction error (PE) (standard deviations in parentheses) over five random partitions of in-sample and out-of-sample data from multiple U.S. stock market indices.

Methods	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
SPA-LS	$28.9647_{(4.996)}$	$22.1664_{(2.617)}$	$22.1590_{(0.495)}$	$22.2746_{(4.359)}$	$31.3910_{(6.255)}$
nPD-LS	$30.5522_{(5.436)}$	$23.5004_{(3.084)}$	$22.3675_{(1.213)}$	$23.2787_{(4.904)}$	$33.1099_{(6.815)}$
DI-LS	$26.4879_{(4.697)}$	$21.3239_{(1.857)}$	$22.1754_{(1.060)}$	$21.9970_{(4.618)}$	$28.8022_{(5.535)}$
SPA-US	$36.4774_{(5.117)}$	$24.8355_{(2.705)}$	$21.6213_{(0.590)}$	$24.7946_{(4.625)}$	$39.9282_{(6.838)}$
nPD-US	$33.4391_{(5.672)}$	$24.1276_{(3.014)}$	$21.8445_{(1.323)}$	$23.8007_{(4.916)}$	$36.9738_{(7.530)}$
DI-US	$27.8361_{(4.909)}$	$21.4671_{(1.897)}$	$21.7320_{(1.157)}$	$22.0939_{(4.788)}$	$30.8690_{(6.201)}$

as distinct platforms, and performed data integration via a Bayesian model selection method. Similarly, we consider a two-platform data integration analysis of U.S. stock market data. The first platform comprised the S&P 500 index and its volatility index (VIX), and the second consisted of the NASDAQ-100 index and its volatility index (VXN). Given that the responses are derived from the same underlying entity, the U.S. stock market, it is reasonable to anticipate the presence of unobserved correlations among both the responses and the predictors. In a similar data analysis vein, we apply this multi-platform data to a multivariate quantile regression setting. Weekly data from October 2022 to September 2025 were collected, we take take log returns to obtain approximately independent observations. Candidate predictors were selected from the union of actively traded stocks in the S&P 500 or NASDAQ-100 across both platforms, excluding stocks with missing values. This process resulted in a final set of $n = 156, p = 516$.

To evaluate the predictive performance of the models, we repeatedly partitioned the data into a training set (80%) and a test set (20%), randomly. The prediction error (PE) defined in Eq.(5.3.2) is computed on the held-out test set. The mean PE is then reported as the final performance measure. The tuning parameter λ_n is selected from a candidate sequence of 20 values equally spaced ranging from 0.1 to 0.2. The optimal λ_n is chosen by minimizing the criterion A-AIC given in Eq.(5.3.1).

We report the average PE and its standard deviation in Table 5.5. Among the six methods, the DI-LS and DI-US methods generally achieve smaller prediction errors and lower variability compared with other methods, and this difference is more pronounced at $\tau = 0.05$ and $\tau = 0.95$. The differences in PE become less apparent at $\tau = 0.5$.

5.4.2 Multiple experiments on liver toxicity study

Dai et al. (2023) used data from a liver toxicity study (Bushel et al., 2007) to examine the global relationship between predictors and a univariate response by modeling multiple

Table 5.6: Methods comparison based on average prediction error (PE) (standard deviations in parentheses) over five random partitions of in-sample and out-of-sample data from liver toxicity study in multiple experiments.

Methods	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
SPA-LS	48.1203(7.978)	41.4775(4.652)	39.2979(3.576)	44.9945(7.376)	49.5768(10.582)
nPD-LS	45.2024(6.070)	39.9426(4.820)	40.5614(4.065)	45.0863(7.832)	44.6887(8.221)
DI-LS	43.7597(5.747)	38.6137(5.271)	39.1924(3.489)	43.7917(8.304)	42.8275(8.421)
SPA-US	51.7011(8.543)	42.5705(4.469)	39.3668(3.935)	45.7572(7.297)	54.4705(11.623)
nPD-US	45.2766(5.920)	39.5474(4.839)	40.3698(4.102)	45.0723(7.897)	46.5945(8.174)
DI-US	43.7601(5.435)	38.1623(5.237)	38.8291(3.484)	43.5213(8.340)	43.6081(8.678)

quantiles simultaneously, where they took the log-transformed cholesterol level as the response and gene expression profiles of 3,116 genes as predictors. The study included two experiments in which rats were exposed to either non-toxic (50 or 150 mg/kg) or toxic (1,500 or 2,000 mg/kg) doses of acetaminophen (paracetamol). These two experiments were treated as two distinct sources of data (i.e., $K = 2$) due to the differences in experimental conditions. The dataset is available in the R package `mixOmics` (Rohart et al., 2017).

In a similar vein, our study analyzed ten clinical measurements as multivariate response variables. Each measurement was log-transformed to reduce skewness and then standardized to ensure comparability across variables. From the 3,116 genes measured for the 64 subjects (32 per dose group), the top 200 genes for each group were selected as predictors, where the selection was based on the largest sums of squared correlations with the ten clinical responses. We carried out a comparison of methods based on average PE across multiple quantile levels. The comparison is based on five random partitions of in-sample and out-of-sample data, with standard deviations of PEs reported in parentheses (see Table 5.6). These results are consistent with those in Table 5.5, where the data integration methods (DI-LS and DI-US) demonstrate superior prediction accuracy, as evidenced by their lower PEs and standard deviations. However, this superiority varies across different quantile levels. The data integration technique notably outperforms separated platform analysis and naively pooled methods, particularly in the left and right tails.

5.5 Conclusion

In this chapter, we develop a robust framework for high-dimensional data integration (DI) based on a multivariate quantile regression model, with univariate quantile regression as a special case. By effectively integrating heterogeneous information across multiple platforms, the proposed approach overcomes the limitations of naive pooled (nPD) and separated (SPA)

analyses, enabling a more comprehensive characterization of complex dependence structures in high-dimensional data.

From a methodological perspective, the non-differentiability of the quantile loss function is addressed through upper and lower smoothing techniques, which allow efficient optimization via a differentiable objective function. The incorporation of an ℓ_2/ℓ_1 penalty facilitates interpretable variable selection by identifying predictors that exert consistent effects across platforms. Estimation is carried out using an MM algorithm, and theoretical results are established to ensure the consistency of the proposed estimator.

Extensive simulation studies across a range of error distributions, quantile levels, and response dimensions demonstrate that the data integration approach achieves a favorable balance between model sparsity, selection accuracy, and predictive performance. Empirical applications to financial market indices and liver toxicity data from multiple experiments further confirm the practical advantages of the proposed method, showing consistent improvements in predictive accuracy over both naive pooling (nPD) and separated (SPA) analysis. Overall, this chapter provides a scalable and robust tool for integrating multi-platform data, offering a reliable foundation for high-dimensional statistical analysis.

Chapter 6

Conclusion

This thesis develops robust statistical methodologies for analyzing complex financial time series index data, focusing on regime-switching models, change-point detection, and data integration. Motivated by challenges such as structural breaks, heavy-tailed distributions, outliers, dependence structures, and heterogeneous data sources, the proposed frameworks aim to improve modeling accuracy and interpretability in financial applications.

The first chapter reviews major regime-switching models, including threshold, hidden Markov switching (HMS), hidden semi-Markov switching (HSMS), and smooth transition models. By comparing their switching mechanisms, model structures, and estimation methods, the review highlights that threshold and smooth transition models provide parsimonious representations of nonlinear dynamics, while HMS and HSMS are better suited for capturing business cycle behavior. In particular, HSMS models offer flexible sojourn distributions that allow more realistic modeling of regime durations.

Building on this review, the thesis proposes a robust hidden semi-Markov switching (rHSMS) model that accommodates heavy-tailed data using a Hube ρ function. The Hidden semi Markov framework relaxed the assumption of the sojourn distribution and provides a likelihood based EM estimation algorithm.The rHSMS relaxes restrictive sojourn assumptions and employs a likelihood-based EM algorithm for estimation. Simulation and empirical studies using S&P 500 and Hang Seng Index returns demonstrate improved parameter estimation, more accurate state prediction, better fitting of extreme observations, and smaller bootstrap standard errors compared with GHMS and SC-HMM, highlighting the value of robustness in financial regime-switching analysis.

The second methodological contribution focuses on the multiple change point detection for complex data characterized by moderate to high dimensionality, strong correlations, and heavy tailed distribution. The proposed RPCA-CP framework integrates robust principal component analysis with univariate CPD techniques, leveraging dimensionality reduc-

tion while preserving structural changes. Theoretical justification ensures the feasibility of the approach, and numerical experiments show that RPCA-CP consistently outperforms or matches state-of-the-art methods across a wide range of data-generating scenarios. Its strong performance on real-world datasets, including ACGH and DJIA data, demonstrates its practical value for detecting structural changes in noisy and high-dimensional environments.

The third major contribution addresses high-dimensional data integration through a robust multivariate quantile regression framework. By extending univariate quantile regression to multivariate responses, the proposed DI method effectively fuses heterogeneous information across multiple platforms. The use of smoothed check loss functions enables efficient optimization, while the ℓ_1/ℓ_2 penalty facilitates interpretable variable selection across platforms. The MM algorithm and theoretical consistency results provide solid methodological grounding. Simulation studies and real-data applications confirm that the DI approach consistently outperforms naive pooled and separated analyses, achieving better predictive accuracy and more reliable inference by leveraging cross-platform dependence.

In summary, these contributions advance robust statistical modeling along three key directions. First, they enhance regime-switching models by incorporating robustness and flexible duration dynamics. Second, they improve change-point detection methods for heavy-tailed, moderately high-dimensional data. Third, they develop data-integration techniques for multivariate quantile regression models that are heterogeneous across platforms. Collectively, these methods emphasize robustness, high-dimensional applicability, flexibility, and interpretability, making them well suited to modern applications in which classical modeling assumptions are frequently violated.

Several promising directions remain for future research. While this thesis primarily focuses on univariate time series within the HSMS framework, extending the rHSMS model to multivariate settings is both natural and important. Such an extension would need to address complex dependence structures and incorporate emerging multivariate risk measures. In addition, further integration of regime-switching dynamics with high-dimensional and multi-platform modeling frameworks could lead to powerful tools for systemic risk analysis and biomedical data integration.

For the change-point detection framework, this thesis concentrates on offline detection. A natural next step is to develop so-called online change-point detection methods that enable real-time monitoring and rapid response to structural changes. For the PCA-uCP model, improving computational efficiency by developing principled approaches to reduce the number of principal components is another important direction.

Finally, future work will explore data integration in more complex settings. In particular, it would be valuable to consider large-volume data sources with heterogeneous predictor

structures and varying data quality across platforms.

Overall, this thesis provides a solid methodological foundation and opens multiple avenues for future theoretical development and empirical investigation in robust statistical modeling.

Appendix

The technical proofs are similar to that of Chen and Chen (2024).

Proof of Proposition 5.2.1. : We compute the first and second derivatives of the smoothed objective $\underline{Q}_{\tau,\omega}(\boldsymbol{B})$. Denote $\nabla \underline{Q}_{\tau,\omega}(\boldsymbol{B}) = (\nabla \underline{Q}_{\tau,\omega}(\boldsymbol{B}_1), \ldots, \nabla \underline{Q}_{\tau,\omega}(\boldsymbol{B}_K)) \in \mathbb{R}^{p\times Kq}$. Herein, $\nabla \underline{Q}_{\tau,\omega}(\boldsymbol{B}_k) \in \mathbb{R}^{p\times q}$ represents the gradient with respect to $\boldsymbol{B}_k$, $k = 1, \ldots, K$, i.e.,

$$\nabla \underline{Q}_{\tau,\omega}(\boldsymbol{B}_k) = \frac{\partial \underline{Q}_{\tau,\omega}(\boldsymbol{B})}{\partial \boldsymbol{B}_k} = -\frac{1}{n}\boldsymbol{X}_k^\top \boldsymbol{R}_{kj},$$

where

$$\boldsymbol{R}_{kj} = \nabla \underline{\rho}_{\tau_j,\omega}(\frac{y_{kij} - \boldsymbol{x}_{ki}^\top \dot{\boldsymbol{b}}_{kj}}{n}),$$

with $\nabla \underline{\rho}_{\tau_j,\omega}(u)$ being the derivative of the lower smoothing quantile Huber loss, i.e.,

$$\nabla \underline{\rho}_{\tau_j,\omega}(u) = \begin{cases} \tau_j - 1, & u < -(\tau_j - 1)\omega, \\ u/\omega, & -(\tau_j - 1)\omega \le u \le \tau_j \omega, \\ \tau_j, & u > \tau_j \omega. \end{cases}$$

Denote $\nabla^2 \underline{Q}_{\tau,\omega}(\boldsymbol{B}) = (\nabla^2 \underline{Q}_{\tau,\omega}(\boldsymbol{B}_1), \ldots, \nabla^2 \underline{Q}_{\tau,\omega}(\boldsymbol{B}_K))$, where

$$\nabla^2 \underline{Q}_{\tau,\omega}(\boldsymbol{B}_k) = \frac{\partial^2 \underline{Q}_{\tau,\omega}(\boldsymbol{B})}{\partial \boldsymbol{B}_k^2} = \frac{1}{n^2}\boldsymbol{X}_k^\top \boldsymbol{W}_{kj} \boldsymbol{X}_k \in \mathbb{R}^{p\times p},$$

where

$$\mathbf{W}_{kj} = \operatorname{diag}\left(\nabla^2 \underline{\rho}_{\tau_j,\omega}\left(\frac{y_{k1j} - \boldsymbol{x}_{k1}^\top \dot{\boldsymbol{b}}_{kj}}{n} \right), \ \ldots, \ \nabla^2 \underline{\rho}_{\tau_j,\omega}\left(\frac{y_{knj} - \boldsymbol{x}_{kn}^\top \dot{\boldsymbol{b}}_{kj}}{n} \right) \right) \in \mathbb{R}^{n\times n}.$$

Herein, $\nabla^2 \underline{\rho}_{\tau_j,\omega}(u)$ is the second derivative (piecewise constant) of the lower smoothing quan-

tile Huber loss, i.e.,

$$\nabla^2 \underline{\rho}_{\tau_j,\omega}(u) = \begin{cases} 1/\omega, & (\tau_j - 1)\omega \le u \le \tau_j \omega, \\ 0, & otherwise. \end{cases}$$

$\nabla \underline{Q}_{\tau,\omega}(\boldsymbol{B})$ is Lipschitz continuous with a Lipschitz constant

$$\underline{L}_c = \sup_{\boldsymbol{B}} \lambda_{\max}(\nabla \underline{Q}^2_{\tau,\omega}(\boldsymbol{B})).$$

Since $\lambda_{\max}(\frac{1}{n^2}\boldsymbol{X}_k^\top \boldsymbol{W}_{kj}\boldsymbol{X}_k) \le \frac{1}{n^2\omega}\lambda_{\max}(\boldsymbol{X}_k^\top \boldsymbol{X}_k)$,

$$\underline{L}_c = \frac{1}{n^2\omega} \max_k \lambda_{\max}(\boldsymbol{X}_k^\top \boldsymbol{X}_k).$$

Similarly, it is easy to prove that $\nabla \overline{Q}_{\tau,\omega}(\boldsymbol{B})$ is Lipschitz continuous with a Lipschitz constant of

$$\overline{L}_c = \underline{L}_c/2.$$

■

Proof of Theorem 5.2.1. : Recall that the objective function is

$$\underline{L}_{\tau,\omega}(\boldsymbol{B}) = \underline{Q}_{\tau,\omega}(\boldsymbol{B}) + \lambda_n \|\boldsymbol{B}\|_{\ell_1/\ell_2}.$$

By the MM construction, the surrogate function satisfies

$$\underline{Q}_{\tau,\omega}(\boldsymbol{B}) \le M_{\tau,\omega,\xi}(\boldsymbol{B}, \boldsymbol{B}^{(t)}), \quad \forall \boldsymbol{B},$$

with equality when $\boldsymbol{B} = \boldsymbol{B}^{(t)}$. Therefore,

$$\begin{aligned} \underline{L}_{\tau,\omega}(\boldsymbol{B}^{(t+1)}) &= \underline{Q}_{\tau,\omega}(\boldsymbol{B}^{(t+1)}) + \lambda_n \|\boldsymbol{B}^{(t+1)}\|_{\ell_1/\ell_2} \\ &\le M_{\tau,\omega,\xi}(\boldsymbol{B}^{(t+1)}, \boldsymbol{B}^{(t)}) + \lambda_n \|\boldsymbol{B}^{(t+1)}\|_{\ell_1/\ell_2}. \end{aligned} \tag{A.1}$$

Since $\boldsymbol{B}^{(t+1)}$ minimizes the MM surrogate, we have

$$M_{\tau,\omega,\xi}(\boldsymbol{B}^{(t+1)}, \boldsymbol{B}^{(t)}) + \lambda_n \|\boldsymbol{B}^{(t+1)}\|_{\ell_1/\ell_2} \le M_{\tau,\omega,\xi}(\boldsymbol{B}^{(t)}, \boldsymbol{B}^{(t)}) + \lambda_n \|\boldsymbol{B}^{(t)}\|_{\ell_1/\ell_2}. \tag{A.2}$$

Using the identity

$$M_{\tau,\omega,\xi}(\boldsymbol{B}^{(t)}, \boldsymbol{B}^{(t)}) = \underline{Q}_{\tau,\omega}(\boldsymbol{B}^{(t)}),$$

we combine (A.1) and (A.2) to obtain

$$\underline{L}_{\tau,\omega}(\boldsymbol{B}^{(t+1)}) \leq \underline{L}_{\tau,\omega}(\boldsymbol{B}^{(t)}).$$

Hence, $\{\underline{L}_{\tau,\omega}(\boldsymbol{B}^{(t)})\}$ is non-increasing. By Assumption (A1) and (A2), the level set

$$\mathcal{L}_0 = \left\{\boldsymbol{B} : \underline{L}_{\tau,\omega}(\boldsymbol{B}) \leq \underline{L}_{\tau,\omega}(\boldsymbol{B}^{(0)})\right\}$$

is compact. Since $\underline{L}_{\tau,\omega}(\boldsymbol{B}^{(t)})$ is non-increasing, all iterates satisfy

$$\boldsymbol{B}^{(t)} \in \mathcal{L}_0, \quad \forall t,$$

and thus the sequence $\{\boldsymbol{B}^{(t)}\}$ is bounded. Therefore, there exists a convergent subsequence

$$\boldsymbol{B}^{(t_k)} \to \bar{\boldsymbol{B}}.$$

We now show that $\bar{\boldsymbol{B}}$ is a stationary point. From the optimality of the MM step and continuity of the surrogate function, standard MM arguments imply that every limit point of $\{\boldsymbol{B}^{(t)}\}$ satisfies the first-order optimality condition of $\underline{L}_{\tau,\omega}$. Thus, every accumulation point is a stationary point. By Assumption (A4), the stationary point in $\mathcal{L}_0$ is unique and equal to $\hat{\boldsymbol{B}}$. Hence all subsequences converge to the same limit, implying that the full sequence converges:

$$\boldsymbol{B}^{(t)} \to \hat{\boldsymbol{B}}.$$

This completes the proof.

■

Bibliography

Akaikei, H. (1973). Information theory and an extension of maximum likelihood principle. In Proc. 2nd int. symp. on information theory (pp. 267-281).

Allen, D. E., Singh, A. K., Powell, R. J., McAleer, M., Taylor, J., & Thomas, L. (2013). Return-volatility relationship: Insights from linear and non-linear quantile regression (No. 13-020/III). Tinbergen Institute Discussion Paper.

Amini, M., Bayat, A., & Salehian, R. (2022). hhsmm: an R package for hidden hybrid Markov semi-Markov models. Computational Statistics, 1-53.

Aminikhanghahi, S., & Cook, D.J. (2017). A survey of methods for time series change point detection. Knowledge and information systems, 51(2), 339-367.

Amisano, G., & Fagan, G. (2013). Money growth and inflation: A regime switching approach. Journal of International Money and Finance, 33, 118-145.

Ando, R. K., Zhang, T., & Bartlett, P. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. Journal of machine learning research, 6(11).

Andrecut, M. (2009). Parallel GPU implementation of iterative PCA algorithms. Journal of Computational Biology, 16(11), 1593-1599.

Ang, A., & Bekaert, G. (2002). Regime switches in interest rates. Journal of Business & Economic Statistics, 20(2), 163-182.

Angelosante, D., & Giannakis, G.B. (2012). Group lassoing change-points in piecewise-constant AR processes. EURASIP Journal on Advances in Signal Processing, 2012, 1-16.

Aravkin, A. Y., Kambadur, A., Lozano, A. C., & Luss, R. (2014). Sparse quantile huber regression for efficient and robust estimation. arXiv preprint arXiv:1402.4624.

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2010). Contour detection and hierarchical image segmentation. IEEE transactions on pattern analysis and machine intelligence, 33(5), 898-916.

Arlot, S., Celisse, A., & Harchaoui, Z. (2019). A kernel multiple change-point algorithm via model selection. Journal of machine learning research, 20(162), 1-56.

Audrino, F., & Bühlmann, P. (2001). Tree-structured generalized autoregressive conditional heteroscedastic models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 63(4), 727-744.

Auger, I. E., & Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. Bulletin of mathematical biology, 51(1), 39-54.

Augustyniak, M. (2014). Maximum likelihood estimation of the Markov switching GARCH model. Computational Statistics & Data Analysis, 76, 61-75.

Avanesov, V., & Buzun, N. (2018). Change-point detection in high-dimensional covariance structure.

Ayodeji, I. O. (2016). A Three-State Markov-Modulated Switching Model for Exchange Rates. Journal of Applied mathematics, 2016(1), 5061749.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. Econometrica, 47-78.

Balcilar, M., Gupta, R., & Miller, S.M. (2015). Regime switching model of US crude oil and stock market prices: 1859 to 2013. Energy Economics, 49, 317-327.

Baum, L. E., & Eagon, J. A. (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology.

Baum, C. F., Barkoulas, J. T., & Caglayan, M. (2001). Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. Journal of International Money and Finance, 20(3), 379-399.

Bazzi, M., Blasques, F., Koopman, S. J., & Lucas, A. (2017). Time-varying transition probabilities for Markov regime switching models. Journal of Time Series Analysis, 38(3), 458-478.

Beder, T. S. (1995). VaR: Seductive but dangerous. Financial Analysts Journal, 51(5), 12-24.

Belloni, A., and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high dimensional sparse models. Annals of Statistics, 39(1), 82–130.

Bera, A. K., Galvao Jr, A. F., Montes-Rojas, G. V., & Park, S. Y. (2016). Asymmetric laplace regression: maximum likelihood, maximum entropy and quantile regression. Journal of Econometric Methods, 5(1), 79-101.

Bernardi, M., Maruotti, A., & Petrella, L. (2018). Multivariate Markov Switching models and tail risk interdependence. arXiv:1312.6407v3 [stat.ME].

Bleakley, K., & Vert, J.P. (2011). The group fused lasso for multiple change-point detection. arXiv preprint arXiv:1106.4199.

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical science, 17(3), 235-255.

Breunig, R., Najarian, S., & Pagan, A. (2003). Specification testing of Markov switching models. Oxford Bulletin of Economics and Statistics, 65, 703-72

Brockwell, P. J., Liu, J., & Tweedie, R. L. (1992). On the existence of stationary threshold autoregressive moving-average processes. Journal of Time Series Analysis, 13(2), 95-107.

Brooks, C. (2001). A double-threshold GARCH model for the French Franc/Deutschmark exchange rate. Journal of Forecasting, 20(2), 135-143.

Buffington, J., & Elliott, R.J. (2002). American options with regime switching. International Journal of Theoretical and Applied Finance, 5(05), 497-514.

Bulla, J., & Bulla, I. (2006). Stylized facts of financial time series and hidden semi-Markov models. Computational statistics & data analysis, 51(4), 2192-2209.

Bulla, J., Bulla, I., & Nenadić, O. (2010). hsmm—An R package for analyzing hidden semi-Markov models. Computational Statistics & Data Analysis, 54(3), 611-619.

Bulla, J. (2011). Hidden Markov models with t components. Increased persistence and other aspects. Quantitative Finance, 11(3), 459-475.

Bushel, P. R., Wolfinger, R. D., & Gibson, G. (2007). Simultaneous clustering of gene expression data with clinical chemistry and pathological evaluations reveals phenotypic prototypes. BMC Systems Biology, 1(1), 15.

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. Journal of the ACM (JACM), 58(3), 1-37.

Cardot, H., Lecuelle, G., Schlich, P., & Visalli, M. (2019). Estimating finite mixtures of semi-Markov chains: An application to the segmentation of temporal sensory data. Journal of the Royal Statistical Society Series C: Applied Statistics, 68(5), 1281-1303.

Carlier, G., Chernozhukov, V., De Bie, G., & Galichon, A. (2022). Vector quantile regression and optimal transport, from theory to numerics. Empirical Economics, 62(1), 35-62.

Chakraborty, B. (2003). On multivariate quantile regression. Journal of statistical planning and inference, 110(1-2), 109-132.

Chan, K. S., Petruccelli, J. D., Tong, H., & Woolford, S. W. (1985). A multiple-threshold AR (1) model. Journal of applied probability, 22(2), 267-279.

Chan, K. S., & Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. Advances in applied probability, 17(3), 666-678.

Chandler, R. E., & Bate, S. (2007). Inference for clustered data using the independence loglikelihood. Biometrika, 94(1), 167-183.

Chang, Y., Choi, Y., & Park, J. Y. (2017). A new approach to model regime switching. Journal of Econometrics, 196(1), 127-143.

Chauvet, M.(1998) An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching. International Economic Review, 39:969–996

Chen, B., & Chen, C. (2024). Fast optimization methods for high-dimensional row-sparse multivariate quantile linear regression. Journal of Statistical Computation and Simulation, 94(1), 69-102.

Chen, B. Z., Kong, L. C., & Xu, N. N. (2020). Smoothing quantile regression with elastic net penalty. Pac. J. Optim., 16(3), 369-393.

Chen, C. W. (1998). A Bayesian analysis of generalized threshold autoregressive models. Statistics & probability letters, 40(1), 15-22.

Chen, C. W., Chiang, T. C., & So, M. K. (2003). Asymmetrical reaction to US stock-return news: evidence from major stock markets based on a double-threshold model. Journal of economics and business, 55(5-6), 487-502.

Chen, C. W., So, M. K., & Gerlach, R. H. (2005). Assessing and testing for threshold nonlinearity in stock returns. Australian & New Zealand Journal of Statistics, 47(4), 473-488.

Chen, C. W., & So, M. K. (2006). On a threshold heteroscedastic model. International Journal of Forecasting, 22(1), 73-89.

Chen, C. W., Liu, F. C., & So, M. K. (2008). Heavy-tailed-distributed threshold stochastic volatility models in financial time series. Australian & New Zealand Journal of Statistics, 50(1), 29-51.

Chen, C. W., Gerlach, R., & Lin, E. M. (2008). Volatility forecasting using threshold heteroskedastic models of the intra-day range. Computational Statistics & Data Analysis, 52(6), 2990-3010.

Chen, C. W., Gerlach, R. H., & Lin, A. M. (2010). Falling and explosive, dormant, and rising markets via multiple-regime financial time series models. Applied Stochastic Models in Business and Industry, 26(1), 28-49.

Chen, C. W., Gerlach, R. H., & Lin, A. M. (2011). Multi-regime nonlinear capital asset pricing models. Quantitative Finance, 11(9), 1421-1438.

Chen, C. W., Liu, F. C., & So, M. K. (2011). A review of threshold time series models in finance. Statistics and its Interface, 4(2), 167-181.

Chen, H., & Chu, L. (2023). Graph-based change-point analysis. Annual Review of Statistics and Its Application, 10, 475-499.

Chen, J., Gupta, A. K., & Gupta, A. K. (2000). Parametric statistical change point analysis (Vol. 192). Boston: Birkhäuser.

Chen, R., & Tsay, R. S. (1991). On the ergodicity of TAR (1) processes. The Annals of Applied Probability, 613-634.

Chen, R. (1995). Threshold variable selection in open-loop threshold autoregressive models. Journal of time series analysis, 16(5), 461-481.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2022). Fast algorithms for the quantile regression process. Empirical economics, 62(1), 7-33.

Chkili, W., & Nguyen, D.K. (2014). Exchange rate movements and stock market returns in a regime-switching environment: Evidence for BRICS countries. Research in International Business and Finance, 31, 46-56.

Cho, H., & Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. Journal of the Royal Statistical Society Series B: Statistical Methodology, 77(2), 475-507.

Choi, K., & Hammoudeh, S. (2010). Volatility behavior of oil, industrial commodity and stock markets in a regime-switching environment. Energy Policy, 38(8), 4388–4399.

Contreras-Reyes, J. E. (2023). Information quantity evaluation of nonlinear time series processes and applications. Physica D: Nonlinear Phenomena, 445, 133620.

Corradin, R., Danese, L., & Ongaro, A. (2022). Bayesian nonparametric change point detection for multivariate time series with missing observations. International Journal of Approximate Reasoning, 143, 26-43.

Cosslett, S.R., & Lee, L.F. (1985). Serial correlation in latent discrete variable models. Journal of Econometrics, 27(1), 79-97.

Crosier, R.B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. Technometrics, 30(3), 291-303.

Croux, C., Filzmoser, P., & Fritz, H. (2013). Robust sparse principal component analysis. Technometrics, 55(2), 202-214.

Csörgö, M., & Horváth, L. (1997). Limit theorems in change-point analysis. (No Title).

Dai, G., Müller, U. U., & Carroll, R. J. (2023). Data integration in high dimension with multiple quantiles. Statistica Sinica, 33(1), 169-191.

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science, 369(6500), eabb9789.

Du, H., & Duan, Z. (2022). Finder: A novel approach of change point detection for multivariate time series. Applied Intelligence, 52(3), 2496-2509.

Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. Journal of Business & Economic Statistics, 15(1), 26-34.

Eitrheim, Ø., & Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. Journal of econometrics, 74(1), 59-75.

Elliott, R. J., & Siu, T. K. (2010). On risk minimizing portfolios under a Markovian regime-switching Black-Scholes economy. Annals of Operations Research, 176, 271-291.

Enikeeva, F., & Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives.

Fearnhead, P., & Rigaill, G. (2020). Relating and comparing methods for detecting changes in mean. Stat, 9(1), e291.

Freguson, J.D. (1980). Variable duration models for speech. In: Proceedings of the Symposium on the Application of Hidden Markov Models to Text and Speech, 1980. Princeton, New Jersey, pp. 143-179.

Fernandes, M., Guerre, E., & Horta, E. (2021). Smoothing quantile regressions. Journal of Business & Economic Statistics, 39(1), 338-357.

Fikri, N., Rida, M., Abghour, N., Moussaid, K., & El Omri, A. (2019). An adaptive and real-time based architecture for financial data integration. Journal of Big Data, 6(1), 97.

Francq, C., & Zakoı, J. M. (2008). Deriving the autocovariances of powers of Markov switching GARCH models, with applications to statistical inference. Computational Statistics & Data Analysis, 52(6), 3027-3046.

Frick, K., Munk, A., & Sieling, H. (2014). Multiscale change point inference. Journal of the Royal Statistical Society Series B: Statistical Methodology, 76(3), 495-580.

Gao, X., & Carroll, R. J. (2017). Data integration with high dimensionality. Biometrika, 104(2), 251-272.

Gao, X., & Zhong, Y. (2019). DI-LSMQRearn: a biomarker selection algorithm on cross-platform data. Bioinformatics, 35(21), 4465-4468.

Garcia, R.,& Perron, P. (1996). An analysis of the real interest rate under regime shifts. The review of economics and statistics, 111-125.

Gerlach, R., & Chen, C. W. (2008). Bayesian inference and model comparison for asymmetric smooth transition heteroskedastic models. Statistics and Computing, 18, 391-408.

Gerlach, R., Chen, C. W., Lin, D. S., & Huang, M. H. (2006). Asymmetric responses of international stock markets to trading volume. Physica A: Statistical Mechanics and its Applications, 360(2), 422-444.

Geweke, J. (1994). Bayesian comparison of econometric models. Working Paper, Federal Reserve Bank of Minneapolis, Minnesota.

Ghaddar, D. K., & Tong, H. (1981). Data transformation and self-exciting threshold autoregression. Journal of the Royal Statistical Society Series C: Applied Statistics, 30(3), 238-248.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. The journal of finance, 48(5), 1779-1801.

Goldfeld, S.M., & Quandt, R.E. (1973). A Markov model for switching regressions. Journal of econometrics, 1(1), 3-15.

Górecki, T., Horváth, L., & Kokoszka, P. (2018). Change point detection in heteroscedastic time series. Econometrics and statistics, 7, 63-88.

Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics, 42(1), 27–62.

Grundy, T., Killick, R., & Mihaylov, G. (2020). High-dimensional changepoint detection via a geometrically inspired mapping. Statistics and Computing, 30(4), 1155-1166.

Guédon, Y. (2003). Estimating Hidden Semi-Markov Chains From Discrete Sequences. Journal of Computational and Graphical Statistics, 12(3), 604-639.

Guo, B., Wu, Y., & Xie, H. (2011). A segmented regime-switching model with its application to stock market indices. Journal of Applied Statistics, 38(10), 2241-2252.

Guo, F., Chen, C.R., & Huang, Y.S. (2011). Markets contagion during financial crisis: A regime-switching approach. International Review of Economics and Finance, 20(1), 95–109.

Hagerud, G. E. (1997). A smooth transition ARCH model for asset returns. Stockholm School of Economics, the Economic Research Inst..

Hallin, M., Paindaveine, D., Šiman, M., Wei, Y., Serfling, R., Zuo, Y., ... & Mizera, I. (2010). MULTIVARIATE QUANTILES AND MULTIPLE-OUTPUT REGRESSION QUANTILES: FROM L_1 OPTIMIZATION TO HALFSPACE DEPTH [with Discussion and Rejoinder]. The Annals of Statistics, 635-703.

Hamilton, J.D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica, 57(2), 357–384.

Hamilton, J.D. (1994) Time Series Analysis. Princeton University Press, Princeton NJ

Hamilton, J.D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. Journal of econometrics, 64(1-2), 307-333.

Hamilton, J.D., & Lin, G. (1996). Stock market volatility and the business cycle. Journal of applied econometrics, 11(5), 573-593.

Hamilton, J. D., & Raj, B. (2002). New directions in business cycle research and financial analysis. Empirical Economics, 27, 149-162.

Hardy M.R. (2001). Regime-Switching Model of Long-Term Stock Returns. North American Actuarial Journal, 5(2), 41-53.

Haynes, K., Fearnhead, P., & Eckley, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. Statistics and computing, 27, 1293-1305.

He, X., Pan, X., Tan, K. M., & Zhou, W. X. (2023). Smoothed quantile regression with large-scale inference. Journal of Econometrics, 232(2), 367-388.

Hong, H. G., Christiani, D. C., & Li, Y. (2019). Quantile regression for survival data in modern cancer research: expanding statistical tools for precision medicine. Precision clinical medicine, 2(2), 90-99.

Hsu, D., Kakade, S.M., & Zhang, T. (2012). A spectral algorithm for learning hidden markov models. Journal of Computer and System Sciences, 78(5), 1460–1480.

Huang, W.H. (2015). The determinants of household electricity consumption in Taiwan: Evidence from quantile regression. Energy, 87, 120-133.

Inclan, C., & Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. Journal of the American Statistical Association, 89(427), 913-923.

Jackson, J.E. (1991) A user's guide to principal components (John Wiley and Sons, New York).

James, B., James, K. L., & Siegmund, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. Statistica Sinica, 69-90.

James, N. A., Zhang, W., Matteson, D. S., & Zhang, M. W. (2023). Package 'ecp'.

Jansen, E. S., & Teräsvirta, T. (1996). Testing parameter constancy and super exogeneity in econometric equations. Oxford Bulletin of Economics and Statistics, 58(4), 735-763.

Jegadeesh, N., Noh, J., Pukthuanthong, K., Roll, R., & Wang, J. (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. Journal of Financial Economics, 133(2), 273-298.

Jin, B., Shi, X., & Wu, Y. (2013). A novel and fast methodology for simultaneous multiple structural break estimation and variable selection for nonstationary time series models. Statistics and Computing, 23, 221-231.

Jolliffe, I.T. (2002) Principal Component Analysis (Springer, New York, 2002).

Kent, J. T. (1982). Robust properties of likelihood ratio tests. Biometrika, 69(1), 19-27.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association, 107(500), 1590-1598.

Kim, A. Y., Marzban, C., Percival, D. B., & Stuetzle, W. (2009). Using labeled data to evaluate change detectors in a multivariate streaming environment. Signal Processing, 89(12), 2529-2536.

Kim, C.J., & Nelson, C.R. (1999). State-space models with regime switching: classical and Gibbs-sampling approaches with applications. MIT Press Books, 1.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50.

Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. Journal of the american statistical association, 94(448), 1296-1310.

Koenker, R. (2017). Quantile regression: 40 years on. Annual review of economics, 9(1), 155-176.

Kong, L., & Mizera, I. (2012). Quantile tomography: using quantiles with multivariate data. Statistica Sinica, 1589-1610.

Kotz, S., Kozubowski, T., & Podgórski, K. (2001). The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance (No. 183). Springer Science & Business Media.

Kozubowski, T. J., Mazur, S., & Podgorski, K. (2022). Matrix variate generalized laplace distributions (No. 7/2022). Working Paper.

Kozumi, H., & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. Journal of statistical computation and simulation, 81(11), 1565-1578.

Krolzig, H. M. (1998). Econometric modelling of Markov switching vector auto-regressions using MSVAR for Ox.

Lai, T.L. (1995). Sequential changepoint detection in quality control and dynamical systems. Journal of the Royal Statistical Society: Series B (Methodological), 57(4), 613-644.

Langrock, R. & Zucchini, W.(2011). Hidden Markov models with arbitrary state dwell-time distributions. Computational Statistics & Data Analysis, 55(1), 715-724.

Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. Stochastic Processes and their Applications, 83(1), 79-102.

Lévy-Leduc, C., & Roueff, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. The Annals of Applied Statistics, 637-662.

Li, J. (2020). Asymptotic distribution-free change-point detection based on interpoint distances for high-dimensional data. Journal of Nonparametric Statistics, 32(1), 157-184.

Li, C. W., & Li, W. K. (1996). On a double-threshold autoregressive heteroscedastic time series model. Journal of applied econometrics, 11(3), 253-274.

Li, D., & Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. Journal of Econometrics, 167(1), 240-253.

Li, X., & Zhang, X. (2024). fastcpd: Fast Change Point Detection in R. arXiv preprint arXiv:2404.05933.

Li, Y., and Zhu, J.(2008). L_1-norm quantile regression. Journal of Computational and Graphical Statistics,17(1), 163–185.

Li, J., Xu, M., Zhong, P. S., & Li, L. (2019). Change point detection in the mean of high-dimensional time series data under dependence. arXiv preprint arXiv:1903.07006.

Li, J., Zhang, X., & Jeske, D. R. (2013). Nonparametric multivariate CUSUM control charts for location and scale changes. Journal of Nonparametric Statistics, 25(1), 1-20.

Li, Y., Nan, B., & Zhu, J. (2015). Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure. Biometrics, 71(2), 354-363.

Lin, Y., Wu, Y., Wang, X., & Ding, H. (2020). A segmented generalized Markov regime-switching model with its application in financial time series data. Journal of Statistical Computation and Simulation, 90(5), 839-853.

Lindsay, B. G. (1988, December). Composite likelihood methods. In Statistical Inference from Stochastic Processes: Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference Held August 9-15, 1987, with Support from the National Science Foundation and the Army Research Office (Vol. 80, p. 221). American Mathematical Soc..

Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. Journal of Applied probability, 36(3), 688-705.

Ling, S., & Tong, H. (2005). Testing for a linear MA model against threshold MA models.

Ling, S., Tong, H., & Li, D. (2007). Ergodicity and invertibility of threshold moving-average models.

Lipkova, J., Chen, R. J., Chen, B., Lu, M. Y., Barbieri, M., Shao, D., ... & Mahmood, F. (2022). Artificial intelligence for multimodal data integration in oncology. Cancer cell, 40(10), 1095-1110.

Liu, H., Gao, C., & Samworth, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection.

Liu, M., Li, Q., & Zhu, F. (2019). Threshold negative binomial autoregressive model. Statistics, 53(1), 1-25.

Liu, M., Li, Q., & Zhu, F. (2020). Self-excited hysteretic negative binomial autoregression. AStA Advances in Statistical Analysis, 104, 385-415.

Ma, F., Wahab, M. I. M., Huang, D., & Xu, W. (2017). Forecasting the realized volatility of the oil futures market: A regime switching approach. Energy Economics, 67, 136-145.

Maiti, M., Vuković, D., Mukherjee, A., Paikarao, P. D., & Yadav, J. K. (2022). Advanced data integration in banking, financial, and insurance software in the age of COVID-19. Software: Practice and Experience, 52(4), 887-903.

Maity, A. K., Carroll, R. J., & Mallick, B. K. (2019). Integration of survival and binary data for variable selection and prediction: a Bayesian approach. Journal of the Royal Statistical Society Series C: Applied Statistics, 68(5), 1577-1595.

Markovitz, H. M. (1959). Portfolio selection: Efficient diversification of investments. John Wiley.

Maruotti, A., Petrella, L., & Sposito, L. (2021). Hidden semi-Markov-switching quantile regression for time series. Computational Statistics & Data Analysis, 159, 107208.

Maruotti, A., Punzo, A., & Bagnato, L. (2019). Hidden Markov and semi-Markov models with multivariate leptokurtic-normal components for robust modeling of daily returns series. Journal of Financial Econometrics, 17(1), 91-117.

Maruotti, A., & Punzo, A. (2021). Initialization of hidden Markov and semi-Markov models: A critical evaluation of several strategies. International Statistical Review, 89(3), 447-480.

Matteson, D. S., & James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. Journal of the American Statistical Association, 109(505), 334-345.

Medhioub, I. (2015). A Markov switching three regime model of Tunisian business cycle. American Journal of Economics, 5(3), 394-403.

Meyers, R.A. (Ed.). (2011). Complex systems in finance and econometrics. Springer New York.

Michael, P., Nobay, A. R., & Peel, D. A. (1997). Transactions costs and nonlinear adjustment in real exchange rates; An empirical investigation. Journal of political economy, 105(4), 862-879.

Moradi, M., Cronie, O., Pérez-Goya, U., & Mateu, J. (2023). Hierarchical spatio-temporal change-point detection. The American Statistician, 77(4), 390-400.

Muggeo, V. M., & Adelfio, G. (2011). Efficient change point detection for genomic sequences of continuous measurements. Bioinformatics, 27(2), 161-166.

Nesterov, Y. (2005). Smooth minimization of non-smooth functions. Mathematical programming, 103(1), 127-152.

James, N. A., & Matteson, D. S. (2014). ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. Journal of Statistical Software, 62(7), 1–25. https://www.jstatsoft.org/v62/i07/.

Olivier, B., Guérin-Dugué, A., & Durand, J. B. (2022). Hidden semi-Markov models to segment reading phases from eye movements. Journal of Eye Movement Research, 15(4).

Ombao, H. C., Raz, J. A., von Sachs, R., & Malow, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. Journal of the American Statistical Association, 96(454), 543-560.

Ortega, J. M., & Rheinboldt, W. C. (2000). Iterative solution of nonlinear equations in several variables. Society for Industrial and Applied Mathematics.

Pak, A., Adegboye, O. A., Adekunle, A. I., Rahman, K. M., McBryde, E. S., & Eisen, D. P. (2020). Economic consequences of the COVID-19 outbreak: the need for epidemic preparedness. Frontiers in public health, 8, 241.

Pan, L., Gao, Q., Wei, K., Yu, Y., Qin, G., & Wang, T. (2025). A robust transfer learning approach for high-dimensional linear regression to support integration of multi-source gene expression data. PLoS computational biology, 21(1), e1012739.

Park, S., Kim, H., & Lee, E. R. (2023). Regional quantile regression for multiple responses. Computational Statistics & Data Analysis, 188, 107826.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin philosophical magazine and journal of science, 2(11), 559-572.

Peng, B., & Wang, L. (2015). An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression. Journal of Computational and Graphical Statistics, 24(3), 676-694.

Petrella, L., & Raponi, V. (2019). Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. Journal of Multivariate Analysis, 173, 70-84.

Petruccelli, J. D., & Woolford, S. W. (1984). A threshold AR (1) model. Journal of applied probability, 21(2), 270-286.

Pilliat, E., Carpentier, A., & Verzelen, N. (2023). Optimal multiple change-point detection for high-dimensional data. Electronic Journal of Statistics, 17(1), 1240-1315.

Pimentel, M.A., Santos, M.D., Springer, D.B., & Clifford, G.D. (2015). Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices. Physiological Measurement, 36(8), 1717–1727.

Qiu, P., & Hawkins, D. (2001). A rank-based multivariate CUSUM procedure. Technometrics, 43(2), 120-132.

Qiu, P., & Hawkins, D. (2003). A nonparametric multivariate cumulative sum procedure for detecting shifts in all directions. Journal of the Royal Statistical Society Series D: The Statistician, 52(2), 151-164.

Qin, S., & Wu, Y. (2020). General matching quantiles M-estimation. Computational Statistics & Data Analysis, 147, 106941.

Qin, S., Tan, Z., & Wu, Y. (2024). On robust estimation of hidden semi-Markov regime-switching models. Annals of Operations Research, 1-33.

Qin, S., Zhang, G., Gao, X., & Wu, Y. (2026). Multi-Platform Multivariate Regression with Group Sparsity for High-Dimensional Data Integration. Entropy, 28(2), 135.

Rohart F, Gautier B, Singh A, Lê Cao K-A (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 13(11): e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Romano, Y., Patterson, E., & Candes, E. (2019). Conformalized quantile regression. Advances in neural information processing systems, 32.

Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2), 461–464. http://www.jstor.org/stable/2958889

Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. Biometrics, 507-512.

Sen, A., & Srivastava, M. S. (1975). On tests for detecting change in mean. The Annals of statistics, 98-108.

Sheta, A. F., Ahmed, S. E. M., & Faris, H. (2015). A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. Soft Computing, 7(8), 2.

Shi, X., Wang, X., Wei, D., and Wu. Y. (2016), A sequential multiple change-point detection procedure via VIF regression, *Computational Statistics*, **31(2)**, 671-691.

Shi, Y., & Ho, K. Y. (2015). Long memory and regime switching: A simulation study on the Markov regime-switching ARFIMA model. Journal of Banking & Finance, 61, S189-S204.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika, 63(1), 117-126.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group lasso. Journal of computational and graphical statistics, 22(2), 231-245.

So, M. E. P., Lam, K., & Li, W. K. (1998). A stochastic volatility model with Markov switching. Journal of Business & Economic Statistics, 16(2), 244-253.

So, M. K., Li, W. K., & Lam, K. (2002). A threshold stochastic volatility model. Journal of Forecasting, 21(7), 473-500.

Srivastava, M. S., & Worsley, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. Journal of the American Statistical Association, 81(393), 199-204.

Su, A., Wang, A., Ye, C., Zhou, C., Zhang, G., Chen, G., ... & Xiao, Z. (2024). Tablegpt2: A large multimodal model with tabular data integration. arXiv preprint arXiv:2411.02059.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the american Statistical association, 89(425), 208-218.

Tian, Y.Z.S.A.I.G., Tang, M.L., Wong, C., & Tian, M.Z. (2024). Bayesian analysis of joint quantile regression for multi-response longitudinal data with application to primary biliary cirrhosis sequential cohort study. Statistical Methods in Medical Research, 33(7), 1163-1184.

Toh, K. C., & Yun, S. (2010). An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. Pacific Journal of optimization, 6(615-640), 15.

Tong, H. (1978). On a threshold model.

Tong,H (1983) Thresholdmodels in non-linear time series analysis, Lecture Notes in Statistics, No. 21. Springer, Heidelberg

Tong, H. (2012). Threshold models in non-linear time series analysis (Vol. 21). Springer Science & Business Media.

Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. Journal of the Royal Statistical Society: Series B (Methodological), 42(3), 245-268.

Van den Burg, G. J., & Williams, C. K. (2020). An evaluation of change point detection algorithms. arXiv preprint arXiv:2003.06222.

van Dijk, D. (1999). Smooth transition models: extensions and outlier robust inference (No. 200).

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. Statistica Sinica, 5-42.

Vert, J. P., & Bleakley, K. (2010). Fast detection of multiple change-points shared by many signals using group LARS. Advances in neural information processing systems, 23.

Wang, F., Liang, D., Li, Y., & Ma, S. (2023). Prior information-assisted integrative analysis of multiple datasets. Bioinformatics, 39(8), btad452.

Wang, T., & Samworth, R.J. (2018). High dimensional change point estimation via sparse projection. Journal of the Royal Statistical Society Series B: Statistical Methodology, 80(1), 57-83.

Wen, F., Xiao, J., Huang, C., & Xia, X. (2018). Interaction between oil and US dollar exchange rate: nonlinear causality, time-varying influence and structural breaks in volatility. Applied Economics, 50(3), 319-334.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.

Wright, J., Ganesh, A., Rao, S., Peng, Y., & Ma, Y. (2009). Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. Advances in neural information processing systems, 22.

Wu, S., & Chen, R. (2007). Threshold variable determination and threshold variable driven switching autoregressive models. Statistica Sinica, 17(1), 241-S38.

Xiao, S., & Dong, M. (2015). Hidden semi-Markov model-based reputation management system for online to offline (O2O) e-commerce markets. Decision Support Systems, 77, 87-99.

Xie, Y., Huang, J., & Willett, R. (2012). Change-point detection for high-dimensional time series with missing data. IEEE Journal of Selected Topics in Signal Processing, 7(1), 12-27.

Xu, H., Dubey, P., & Yu, Y. (2021). Online network change point detection with missing values and temporal dependence. arXiv preprint arXiv:2110.06450.

Yang, K., Yu, X., Zhang, Q., & Dong, X. (2022). On MCMC sampling in self-exciting integer-valued threshold time series models. Computational Statistics & Data Analysis, 169, 107410.

Yang, K., Xu, N., Li, H., Zhao, Y., & Dong, X. (2023). Multivariate threshold integer-valued autoregressive processes with explanatory variables. Applied Mathematical Modelling, 124, 142-166.

Yin, G. & Zhou, X.Y. (2004). Markowitz's mean-variance portfolio selection with regime switching: from discrete-time models to their continuous-time limits, in IEEE Transactions on Automatic Control, 3(49). 349-360.

Ypma, J., Borchers, H.W., Eddelbuettel, D., & Ypma, M.J. (2020). Package 'nloptr'.

Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. Statistics & Probability Letters, 54(4), 437-447.

Yu, S.Z. (2010). Hidden semi-Markov models. Artificial intelligence, 174(2), 215-243.

Yu, S.Z., & Kobayashi, H. (2003). An efficient forward-backward algorithm for an explicit-duration hidden Markov model. IEEE signal processing letters, 10(1), 11-14.

Yurchenko, Y. (2021). Matrix variate and tensor variate Laplace distributions. arXiv preprint arXiv:2104.05669.

Zakoian, J. M. (1994). Threshold heteroskedastic models. Journal of Economic Dynamics and control, 18(5), 931-955.

Zhang, G., Wu, Y., & Gao, X. (2024). Bayesian Model Selection via Composite Likelihood for High-dimensional Data Integration. Canadian Journal of Statistics, 52(3), 924-938.

Zhang, L., Kong, L. C., Li, Y., & Zhou, S. L. (2017). A smoothing iterative method for quantile regression with nonconvex lp penalty. J. Ind. Manag. Optim, 13(1), 93-112.

Zhang, L., & Drikvandi, R. (2023). High dimensional change points: challenges and some proposals.

Zhang, T., Wang, D., & Yang, K. (2021). Quasi-maximum exponential likelihood estimation for double-threshold GARCH models. Canadian Journal of Statistics, 49(4), 1152-1178.

Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. IEEE transactions on knowledge and data engineering, 34(12), 5586-5609.

Zheng, K., Li, Y. & Xu, W. (2021). Regime switching model: spectral clustering hidden Markov model. Annals of Operations Research, 303, 297–319.

Zhu, K. & Liu, T. (2018). Online Tool Wear Monitoring Via Hidden Semi-Markov Model With Dependent Durations. IEEE Transactions on Industrial Informatics, 14(1), 69–78.

Zou, C., Yin, G., Feng, L., & Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems.

Zou, C., Wang, G., & Li, R. (2020). Consistent selection of the number of change-points via sample-splitting. Annals of Statistics, 48(1), 413.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. Journal of Computational and Graphical Statistics, 15(2), 265–286.

Zou, H., & Xue, L. (2018). A selective overview of sparse principal component analysis. Proceedings of the IEEE, 106(8), 1311–1320.

Zucchini, W., MacDonald, I.L. & Langrock R. (2016). Hidden Markov models for time series: an introduction using R, Second Edition. Chapman and Hall/CRC.

The S&P 500 index in 2024: a market driven once again by the Mag 7. (n.d.). First Trust Economics. https://www.ftportfolios.com/Commentary/EconomicResearch/2025/1/8/the-sp-500-index-in-2024-a-market-driven-once-again-by-the-mag-7