Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXV

Robust Mixture Models for Modeling and Clustering Complex Time Series Data

Coordinatore del Corso: Prof. Nicola Sartori

Supervisore: Prof. Massimiliano Caporin

Co-supervisore: Prof. Alireza Nematollahi

Dottorando: Fariborz Setoudehtazangi

30/09/2023

Abstract

This innovative thesis explores the potential of mixture models for effectively modeling and clustering complex time series data characterized by skewed distributions, heavy tails, and conditional heteroskedasticity. Traditional models often struggle to capture the inherent characteristics of financial returns, such as asymmetry and non-normality, causing the development of more sophisticated techniques. This thesis presents two novel families of mixture models that offer flexible and powerful frameworks for capturing the complexities of financial data. First, a unified approach to enhance the robustness of GARCH models is proposed by incorporating the class of finite mixture of scale mixture of skew normal (SMSN) distributions for component errors. Second, a mixture of ARCH (MoARCH) models is introduced, using normal mean-variance mixture (NMVM) distributions. These innovative models provide effective solutions for modeling and analyzing financial data, allowing for more accurate representation and a better understanding of its intricacies.

The primary contribution of this thesis is a novel approach by augmenting the Generalized Autoregressive Conditional Heteroskedastic (GARCH) process pioneered by Bollerslev (1986) with a scale mixture of skew normal distribution to model financial time-series returns. This novel GARCH model surpasses the limitations of traditional approaches and significantly enhances financial modeling and forecasting capabilities. The effectiveness of this methodology is demonstrated through simulations and a real data example using log-returns of Apple stocks from February 1st, 2019, to March 8th, 2022.

Additionally, this research introduces the groundbreaking mixture of ARCH model based on NMVM distributions for effective modeling and clustering of time series data. The proposed approach is evaluated using simulated and real-world financial time series

data, showcasing its superior accuracy and robustness compared to conventional models when handling skewed and heavy-tailed data. By tackling the challenge of modeling and clustering time series data with skewness and heavy tails, this study addresses prevalent issues encountered in finance, economics, and related fields, where outliers and extreme events greatly impact analysis. The MoARCH models with NMVM distributions present a novel and promising approach for time-series classification and clustering tasks. The proposed model is rigorously evaluated using simulated and real-world financial time series data, demonstrating their superior accuracy and robustness, particularly when dealing with skewness, heavy tails, and conditional heteroskedasticity. These models offer improved decision-making capabilities in diverse domains, including finance, economics, and beyond.

Overall, this thesis represents a significant advance in time series analysis, providing powerful tools for modeling and clustering complex time series data. The novel mixture models overcome the limitations of traditional approaches and offer more accurate and reliable analyses. The findings and insights derived from this research contribute to the development of more robust statistical methods in various fields, opening up new avenues for research and facilitating better decision-making based on a comprehensive understanding of time series data.

Sommario

Questa tesi all'avanguardia esplora il potenziale dei modelli di miscele per la modellazione e il clustering efficaci di complessi dati di serie storiche caratterizzati da distribuzioni asimmetriche, code pesanti ed eteroschedasticità condizionale. I modelli tradizionali spesso faticano a catturare le caratteristiche intrinseche dei rendimenti finanziari, come l'asimmetria e la non-normalità, rendendo necessario lo sviluppo di tecniche più sofisticate. Questa tesi presenta due nuove famiglie di modelli di miscele che offrono strutture flessibili e potenti per catturare le complessità dei dati finanziari. In primo luogo, viene proposto un approccio unificato per migliorare la robustezza dei modelli GARCH incorporando la classe di miscele finite di distribuzioni di miscele di normali asimmetriche (SMSN) per gli errori dei componenti. In secondo luogo, viene introdotto un modello di miscele di ARCH (MoARCH), utilizzando distribuzioni di miscele di media-varianza normali (NMVM). Questi modelli innovativi forniscono soluzioni efficaci per la modellazione e l'analisi dei dati finanziari, consentendo una rappresentazione più accurata e una migliore comprensione delle sue complessità.

Il contributo principale di questa tesi è un nuovo approccio che potenzia il processo di Generalized Autoregressive Conditional Heteroskedastic (GARCH) sviluppato da Bollerslev (1986) con una distribuzione di miscele di normali asimmetriche per modellare i rendimenti delle serie storiche finanziarie. Questo nuovo modello GARCH supera i limiti degli approcci tradizionali e migliora significativamente le capacità di modellazione e previsione finanziaria. L'efficacia di questa metodologia viene dimostrata attraverso simulazioni e un esempio di dati reali utilizzando i log-rendimenti delle azioni di Apple dal 1° febbraio 2019 al 8 marzo 2022.

Inoltre, questa ricerca introduce il rivoluzionario modello di miscele di ARCH basato sulle distribuzioni NMVM per una modellazione e un clustering efficaci dei dati di

serie storiche. L'approccio proposto viene valutato utilizzando dati simulati e reali di serie storiche finanziarie, dimostrando la sua precisione e robustezza superiori rispetto ai modelli convenzionali nel trattare dati asimmetrici e a code pesanti. Affrontando la sfida di modellare e raggruppare dati di serie storiche con asimmetria e code pesanti, questo studio affronta problemi diffusi nell'ambito finanziario, economico e di settori correlati, in cui gli outliers e gli eventi estremi incidono notevolmente sull'analisi. I modelli MoARCH con distribuzioni NMVM presentano un approccio nuovo e promettente per le attività di classificazione e clustering delle serie storiche. Il modello proposto viene rigorosamente valutato utilizzando dati simulati e reali di serie storiche finanziarie, dimostrando la sua precisione e robustezza superiori, soprattutto quando si tratta di asimmetria, code pesanti ed eteroschedasticità condizionale. Questi modelli offrono miglioramenti nelle capacità di prendere decisioni in diversi settori, inclusi finanza, economia e oltre.

Nel complesso, questa tesi rappresenta un significativo avanzamento nell'analisi delle serie storiche, fornendo strumenti potenti per la modellazione e il clustering di dati di serie storiche complessi. I nuovi modelli di miscele superano i limiti degli approcci tradizionali e offrono analisi più accurate e affidabili. Le scoperte e le intuizioni derivate da questa ricerca contribuiscono allo sviluppo di metodi statistici più robusti in vari campi, aprendo nuove opportunità di ricerca e facilitando una migliore presa di decisioni basata su una comprensione completa dei dati di serie storiche.

To my family

Acknowledgements

As I reflect on the completion of my PhD thesis, I'm overwhelmed with gratitude. This journey, filled with challenges, personal growth, and endless inspiration, became possible due to a constellation of remarkable individuals.

Foremost among these luminaries are my exceptional mentors, Professor Massimiliano Caporin and Professor Alireza Nematollahi. Their unwavering support, wise guidance, and unwavering encouragement formed the basis of my research journey. Their innovative ideas, profound insights, and scholarly expertise were instrumental in shaping the course and outcomes of my work. Their mentorship has been a perpetual wellspring of inspiration, and I'm deeply grateful for the invaluable opportunities they provided.

Furthermore, when it comes to Professor Geoffrey McLachlan, I find myself at a loss for words to adequately express my admiration. His vast reservoir of knowledge and his remarkable personality have left such an indelible mark on me that I could write volumes in tribute. Being granted the privilege of serving as a visiting PhD scholar under Professor McLachlan's guidance at the esteemed Department of Physics and Mathematics of the University of Queensland (UQ) has filled me with profound gratitude. Professor McLachlan's presence is akin to a wellspring of inspiration, an embodiment of exceptional wisdom and insight. This unique opportunity to expand my intellectual horizons under his mentorship has not only fostered personal growth but has also enriched my intellectual journey in ways that words can scarcely capture.

To my extraordinary parents, whose unwavering love, support, and boundless motivation have laid the bedrock of my success, I owe a debt of gratitude beyond measure. Your steadfast commitment, sacrifices, unwavering encouragement, and tireless efforts have not only enabled me to chase my dreams but also brought me to this remarkable milestone. I am truly blessed to have parents as selfless and supportive as you, and your unwavering belief in me has been an unending source of strength.

To my companions in the XXXV cycle —Beppe, Caizhu, Cristian, Erika, Marco, Nicolas, Pietro, and Touqeer—as well as my dear friends Fereshteh, Nasrin, and Parvaneh, your collaboration, insights, and friendship have been incredibly important during my research journey. Being in our stimulating academic environment helped me learn and grow as a researcher. Your help meant a lot, and having you around made this journey truly amazing. All of you have made a lasting impact on my academic journey, and I'm very thankful for that.

Andrea Sottosanti, I want to express my heartfelt gratitude for the treasured memories and deep emotions we've shared. Your kindness and unwavering friendship have not only touched my heart but have left an enduring impression that time cannot fade. Your boundless generosity and genuine warmth are a wellspring of inspiration that continues to enrich my life. Once more, thank you from the bottom of my heart, Andrea.

Deepest appreciation is reserved for Dr. Tahere Manouchehri, whose unwavering assistance and steadfast support illuminated every step of my research journey. Her feedback, insights, and unmatched expertise played a pivotal role in shaping the final outcome of my work. My sincerest gratitude extends to her for the exceptional opportunities she provided me to expand my knowledge and skills as a researcher. Her guidance was instrumental, and I'm profoundly thankful for her contributions to my academic growth and success.

With heartfelt gratitude,

Fariborz
12 April 2023

Contents

List of Figures **xv**

List of Tables **xviii**

Introduction **3**
- Overview . . . 3
- Main contributions of the thesis . . . 5

1 Financial Time Series Analysis **9**
- 1.1 Characteristics of Financial Time Series . . . 9
- 1.2 Financial time series . . . 10
  - 1.2.1 (G)ARCH models . . . 13
- 1.3 Basic properties . . . 14
  - 1.3.1 Uniqueness and stationarity . . . 14
  - 1.3.2 Mean zero . . . 15
  - 1.3.3 Lack of serial correlation . . . 15
- 1.4 Unconditional variance . . . 16
- 1.5 Heavy tails of y_t . . . 16

2 Mixturing, heavy tailedness, asymmetry and conditional heteroskedasticity in financial returns modelling **19**
- 2.1 Abstract . . . 19
- 2.2 Introduction . . . 20
- 2.3 Preliminaries . . . 22
  - 2.3.1 The SMSN Distribution . . . 22
  - 2.3.2 The FM_m-SMSN Distributions . . . 25
- 2.4 The FM_m-SMSN-GARCH(1,1) Model . . . 27
  - 2.4.1 Model specification . . . 27
  - 2.4.2 ML parameter estimation via the ECME algorithm . . . 29
  - 2.4.3 The observed information matrix . . . 37
  - 2.4.4 Model selection criteria and performance assessment . . . 40
- 2.5 Simulation study and a real data example . . . 41
  - 2.5.1 Simulations . . . 41
    - 2.5.1.1 Model specifications . . . 41
    - 2.5.1.2 Part 1: Parameter Retrieval . . . 42

2.5.1.3 Part 2: Robustness, misspecification and model identification . . . 45
2.5.2 Real data analysis: Apple Inc. (AAPL) . . . 50
2.6 Conclusions . . . 57

3 Robust mixture ARCH modeling based on the normal mean-variance mixture distributions 59
3.1 Abstract . . . 59
3.2 Introduction . . . 60
3.3 Normal mean-variance mixture distributions . . . 61
3.4 Designing and Executing the NMVM-MoARCH . . . 62
3.5 Maximum penalized estimation of the model parameters . . . 65
3.5.1 Selection of tuning parameter and model selection . . . 69
3.6 Computational aspects . . . 70
3.6.1 Estimation of standard errors . . . 70
3.6.2 Initial values . . . 71
3.7 Numerical Study . . . 72
3.7.1 Finite sample properties . . . 72
3.7.2 classification accuracy . . . 75
3.8 Final Thoughts and Concluding Remarks . . . 76

4 Key Findings and Recommendations for Future Studies 79

Appendix Appendix A 83
A.1 Distinctive Examples within the SMSN Family . . . 83
A.2 The CML-step of the FM_m-SMSN-GARCH(1,1) model . . . 86
A.3 Proof of Proposition 3.1 . . . 88
A.4 Proof of proposition 1.4. . . . 89
A.5 . . . 92
A.6 . . . 94

Appendix Appendix B 105
B.1 Exploring the Formulation of the Generalized Hyperbolic Distribution (GH) . . . 105
B.2 Exploring the Various NMVM Distributions: An Overview of Key Examples . . . 106

Bibliography 113

List of Figures

1.1 The top left plot displays the actual price series, S_t, while the top right plot represents the corresponding log-returns, y_t. The middle and bottom plots, show the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the squared log-return series, y_t^2, respectively. . . . 12

1.2 The top plot displays histograms of the marginal residuals overlaid with estimated marginal densities, while the bottom plot presents the Q–Q plot of residuals based on the Gaussian distribution. . . . 13

2.1 Some typical graphs of the light-tailed SN densities with various shape, scale and degrees of freedom parameters. . . . 24

2.2 Some typical graphs of the heavy-tailed ST densities with various shape, scale and degrees of freedom parameters. . . . 25

2.3 AIC, BIC, EDC of 250 simulated FM_2-ST-GARCH(1,1) model against other competitors of the family. . . . 46

2.4 AIC, BIC, EDC of 250 simulated FM_2-ST-GARCH(1,1) model against FM_1-ST-GARCH(1,1) and FM_3-ST-GARCH(1,1) models. . . . 47

2.5 AIC, BIC, EDC of 250 simulated FM_2-ST-GARCH(1,1) model against FM_3-ST-GARCH(1,1) model. . . . 48

2.6 Time series plot of the Apple Inc. from 2/01/2019 to 03/08/2022. . . . 50

2.7 Daily Return of Apple Inc. . . . 51

2.8 Histograms of the marginal residuals of the FM_2-ST-GARCH(1,1) mixture model to Apple inc. data overlaid with estimated marginal densities. . . . 53

2.9 The Q–Q Plot of Residuals of Apple Data Based on the Gaussian (left) and FM_2-ST (right) Models. . . . 54

2.10 Real data (black) and their prediction (red) of closing price of Apple Inc. based on the FM_2-ST-GARCH Model. . . . 57

3.1 Absolute relative bias (ARB) and the root mean squared error (RMSE) for the estimates of parameters across various sample sizes. . . . 74

3.2 Miss Classification Error Rate (MCR) for different sample sizes . . . 76

A.1 Histogram and estimated density of marginal residuals of simulated data in a medium-FM_2-SMSN-GARCH (1,1) process. . . . 93

A.2 Histogram and estimated density of marginal residuals of simulated data in a two-Weak-FM_2-SMSN-GARCH (1,1) process. . . . 93

A.3 Bias of ML estimates based on the medium-FM_2-SN-GACH, medium-FM_2-ST-GACH, medium-FM_2-SSL-GACH and medium-FM_2-SCN-GACH models. . . . 97

A.4 Bias of ML estimates based on the weak-FM_2-SN-GACH, weak-FM_2-ST-GACH, weak-FM_2-SSL-GACH and weak-FM_2-SCN-GACH models. 98
A.5 MSE of ML estimates based on the medium-FM_2-SN-GACH, medium-FM_2-ST-GACH, medium-FM_2-SSL-GACH and medium-FM_2-SCN-GACH models. 99
A.6 MSE of ML estimates based on the weak-FM_2-SN-GACH, weak-FM_2-ST-GACH, weak-FM_2-SSL-GACH and weak-FM_2-SCN-GACH models. 100
A.7 Mean of ML estimates based on the medium-FM_2-SN-GACH, medium-FM_2-ST-GACH, medium-FM_2-SSL-GACH and medium-FM_2-SCN-GACH models. 101
A.8 Mean of ML estimates based on the weak-FM_2-SN-GACH, weak-FM_2-ST-GACH, weak-FM_2-SSL-GACH and weak-FM_2-SCN-GACH models. 102
A.9 SD of ML estimates based on the medium-FM_2-SN-GACH, medium-FM_2-ST-GACH, medium-FM_2-SSL-GACH and medium-FM_2-SCN-GACH models. 103
A.10 SD of ML estimates based on the weak-FM_2-SN-GACH, weak-FM_2-ST-GACH, weak-FM_2-SSL-GACH and weak-FM_2-SCN-GACH models. 104

List of Tables

2.1 Bias and MSE of ML estimates of medium-FM_2-SN–GARCH(1,1), medium-FM_2-ST–GARCH(1,1), medium-FM_2-SSL–GARCH(1,1) and medium-FM_2-SCN–GARCH(1,1) models with various sample sizes n = 250, 500, and 1000. 43

2.2 Bias and MSE of ML estimates of weak-FM_2-SN–GARCH(1,1), weak-FM_2-ST–GARCH(1,1), weak-FM_2-SSL–GARCH(1,1) and weak-FM_2-SCN–GARCH(1,1) models with various sample sizes n = 250, 500, and 1000. 44

2.3 ANOVA results 48

2.4 Tukey HSD test 49

2.5 Model selection criteria for the proposed FM_m-SMSN-GARCH(1,1) model with various number of components m = 1, 2, 3 52

2.6 ML estimation results with their standard errors for fitting the FM_2-ST-GARCH(1,1) model on the Apple Inc. data. 53

2.7 Performance Metrics for the Proposed FM_m-SMSN–GARCH Model . . . 56

3.1 Bias and MSE of ML estimates of the NMVBS − MoARCH with g = 2 across different sample sizes, n = 250, 500, and 1000 74

A.1 Initial values for simulation scenarios (two moderate and two weak components) . . . 92

A.2 Moderately components 92

A.3 Weakly components 92

A.4 Mean and SD of ML estimates of medium-FM_2-SN–GARCH(1,1), medium-FM_2-ST–GARCH(1,1), medium-FM_2-SSL–GARCH(1,1) and medium-FM_2-SCN–GARCH(1,1) models with various sample sizes n = 250, 500, and 1000 95

A.5 Mean and SD of ML estimates of weak-FM_2-SN–GARCH(1,1), weak-FM_2-ST–GARCH(1,1), weak-FM_2-SSL–GARCH(1,1) and weak-FM_2-SCN–GARCH(1,1) models with various sample sizes n = 250, 500, and 1000 96

Introduction

Overview

Modeling financial data sets poses intricate challenges due to the intricate nature of financial returns and their relationship to previous observations. The task becomes particularly complex when attempting to model data with intricate connections to past data points, as well as when encountering residuals that exhibit a lack of symmetry and heavy-tails even after fitting appropriate models.

In the realm of financial time series analysis, capturing the nuances of financial returns requires flexible and robust modeling techniques. Traditional models, such as the Generalized Autoregressive Conditional Heteroskedastic (GARCH) process and its variants, often assume that the residuals follow a Gaussian distribution. The assumption of Gaussian and/or symmetric innovations in conditional heteroskedastic models has been widely explored in the financial literature. While this assumption provides certain conveniences for statistical and probabilistic modeling, it inherently restricts the model's ability to capture important properties such as skewness and heavy tails, which are pervasive characteristics often observed in financial time series returns. Consequently, researchers have recognized the pressing need for more flexible and versatile distribution families that can accommodate the peculiarities of financial data.

In order to better address these features, alternative families of distributions have been proposed in the literature. For instance, (Azzalini, 1985) introduced the Skew-Normal distribution, which permits skewness in univariate distributions. This distribution was subsequently extended to the multivariate case by (Azzalini and Valle, 1996). While the Skew-Normal distribution is effective in capturing skewness, it does not fully account for heavy tails. On the other hand, (Lange and Sinsheimer, 1993) introduced the Normal/independent distributions, which focus on capturing heavy tails but lack the capability to model skewness. Among the suitable options capable of accommodating both skewness and heavy tails, the Skew-t distribution introduced by (Jones and Faddy, 2003) and the Skew-Slash distribution proposed by (Wang and Genton, 2006) stand out. These distributions exhibit unique characteristics that make them suitable for modeling financial data with asymmetry and heavy tails.

Furthermore, researchers have aimed to overcome it by combining the strengths of existing distributions and developing more comprehensive families. For instance, (Lachos and Labra, 2014) introduced the Skew-Normal/independent family of distributions, which merges the advantageous properties of the Skew-Normal distribution and the Normal/independent distributions. This family enables the simultaneous modeling of skewness and heavy tails, offering a more flexible framework for analyzing financial data. Another versatile family is the scale mixture of skew normal (SMSN) distribution proposed by (Branco and Dey, 2001), which represents a highly versatile family of distributions capable of effectively capturing lightly, heavily-tailed, symmetric and asymmetric simultaneously. This distribution provides researchers with a robust and flexible framework for financial data analysis, accommodating the complex characteristics observed in real-world financial datasets.

In line with the SMSN distributions, the NMVM distributions introduced by (McNeil *et al.*, 2015) offer a suitable framework for handling data with both pronounced skewness and heavy tails. This class of distributions encompasses various well-known (symmetric or asymmetrical) members, including the generalized hyperbolic (GH) distribution, the normal mean-variance mixture of Birnbaum-Saunders (NMVBS) distribution, and the normal mean-variance mixture of Lindley (NMVL) distribution, which are considered as special cases.

An essential advantage of both SMSN and NMVM distributions is their convenient stochastic representation, which facilitates the implementation of the expectation conditional maximization (ECM) algorithm for efficient maximum likelihood (ML) estimation. This methodological advancement ensures accurate estimation and reliable inference within the models framework, offering an effective and practical solution for analyzing data with complex characteristics.

Conditional heteroskedasticity models have long been recognized as effective tools for modeling and forecasting volatility in time series data. However, the application of these kinds of models in clustering time series data remains relatively unexplored yet holds promising potential.

Mixture models are powerful tools for clustering tasks, enabling the representation of complex data distributions and the discovery of hidden patterns. Traditionally, mixture models for classification tasks are based on the mixture of regression (MoR) model, assuming a regression structure for the response variable. However, when clustering financial time series data, the mixture of conditional heteroskedasticity models is more suitable. Financial time series often exhibit time-dependent volatility patterns, better captured by conditional heteroskedasticity models than regression models. Models like autoregressive conditional heteroskedasticity (ARCH) and its variants explicitly account for heteroskedasticity and time-varying variances, making them well-suited for capturing the volatility dynamics within each cluster.

By employing a mixture of ARCH models for clustering time series data, we can leverage the strengths of conditional heteroskedastic models in modeling volatility patterns while simultaneously benefiting from the flexibility of the mixture modeling framework. This combination allows us to identify distinct clusters based on both the mean behavior and the volatility dynamics, enabling a more comprehensive understanding of the underlying structure of the time series data. By leveraging the volatility modeling capabilities of ARCH models, we can capture the time-dependent characteristics inherent in time series data. This approach enhances our ability to uncover hidden patterns and provides a more accurate representation of the underlying data structure in time series clustering tasks.

Inspired by the aforementioned advantages and promising results, this thesis takes a significant step forward by proposing the use of the SMSN and NMVM distributions as an alternative model for innovations in GARCH and mixture of ARCH(MoARCH) models, respectively. Depending on the specific analytical framework employed, the SMSN and NMVM distributions offer a flexible and powerful approach to capturing the complex characteristics of financial data. By incorporating the SMSN and NMVM distributions into GARCH and MoARCH models, we aim to enhance the modeling, forecasting and clustering capabilities of these models, particularly in capturing the asymmetry, heavy tails, and other important features observed in financial time series.

Main contributions of the thesis

This thesis aims to advance the field of time series analysis by exploring the potential of mixture models for modeling and clustering time series data with complex distributions, including skewed, heavy-tailed distributions and conditional heteroskedasticity. To achieve this, we first conducted a thorough review of the relevant literature on time series modeling, mixture models, and other related statistical techniques. We then presented a unique approach for modeling and clustering time series data that can effectively handle skewness and heavy tails simultaneously.

One noteworthy accomplishment of this thesis is the introduction of two novel family of mixture models that can be used for modeling and clustering time series data. The first novel family of mixture models is based on a conditional heteroskedasticity model, which utilizes a finite mixture of scale mixture of skew normal distributions. This innovative family of mixture models provides a strong alternative to conventional ones. Financial returns exhibit several characteristics that make them challenging to model accurately. These characteristics include time-varying volatility, fat-tailed distributions, and the presence of skewness and asymmetry. Traditional GARCH models have been widely used to capture volatility clustering and persistence in financial returns, but they often assume a symmetric distribution for the innovations, such as the normal distribution. This assumption fails to capture the skewness and asymmetry commonly observed in financial data. To address this limitation, we propose developing a new GARCH model based on a scale mixture of skew normal distribution. The scale mixture of skew normal distribution is a flexible and powerful framework for modeling skewed and fat-tailed distributions. By incorporating this distribution within the GARCH framework, we aim to capture the asymmetry and non-normality in financial returns more effectively. The motivation for this research stems from the need for improved modeling techniques that can better capture the empirical features of financial data. By explicitly considering the skewness in the distribution of innovations, our proposed GARCH model can provide more accurate estimates of volatility and risk. This, in turn, can have important implications for various financial applications, such as portfolio optimization, risk management, option pricing, and value-at-risk estimation. Therefore, developing a GARCH model based on a scale mixture of skew normal distribution is a promising avenue for research in financial modeling. It addresses the limitations of traditional GARCH models and has the potential to improve risk assessment and decision-making in various financial applications.

Subsequently, based on the outcomes from the first set, the author introduces the second set of mixture models, which is a mixture of ARCH models utilizing normal mean-variance mixture distributions. This model presents an effective approach to modeling and clustering time series data. Using a mixture of ARCH models can be considered a new approach in the context of classification and clustering, as it combines the concepts of ARCH models with the mixture modeling framework. While there has been extensive research on the applications of ARCH models in financial econometrics, their application in classification and clustering is relatively less explored.

The motivation for using a mixture of ARCH models in classification and clustering tasks lies in capturing the heteroskedasticity (varying volatility) present in the data. ARCH models are well-suited for modeling time-varying volatility, as they allow the conditional variance of a variable to depend on its past values and other relevant information. By incorporating mixture modeling, the approach can handle complex data distributions that may exhibit multiple latent clusters. The introduction of these innovative mixture models opens up new opportunities for research, contributing to the ongoing development of more accurate and robust statistical methods for the analysis of time series data, driving progress in multiple fields.

The structure of the remaining chapters in the thesis is outlined as follows. In Chapter 1, we provide an overview of financial time series data, highlighting the challenges associated with their modeling. We also discuss various suitable models for analyzing such data.

Moving on to Chapter 2, we present a novel approach for modeling financial returns. This approach introduces a robust conditional heteroskedasticity model based on the Finite Mixture of Scale Mixture of Skew Normal distribution. Our proposed model offers several desirable features and contributes to the development of a convenient framework. To estimate the maximum likelihood parameters, we develop an EM-type algorithm. The effectiveness of the model is extensively demonstrated through comprehensive simulation studies and a real data example.

In Chapter 3, we introduce a robust mixture ARCH modeling approach based on the normal mean-variance mixture distributions. Similar to the previous chapter, this model provides various advantages and contributes to the development of a convenient framework for financial modeling. The estimation of maximum likelihood parameters is performed using an EM-type algorithm. We present comprehensive simulation studies and a real data example to illustrate the effectiveness of the proposed model.

Chapter 4 serves as the conclusion of the thesis, summarizing the key findings and insights derived from the research. Additionally, we outline potential future directions for further exploration and investigation in this field. Finally, the appendices provide additional supporting material and relevant details to enhance the understanding of the thesis. These appendices offer supplementary resources, including documentation, comprehensive data tables, technical derivations, and other pertinent content that further expand upon the topics discussed in the main chapters.

Chapter 1

Financial Time Series Analysis

1.1 Characteristics of Financial Time Series

Financial time series analysis plays a crucial role in understanding and predicting the behavior of financial markets. The sequential data points collected over time represent various financial indicators, such as stock prices, exchange rates, interest rates, commodity prices, and economic indices. Analyzing financial time series data provides valuable insights into market dynamics, trends, and volatility, which are essential for making informed investment decisions, managing risks, and economic forecasting.

Financial markets are complex and influenced by a multitude of factors, including economic indicators, political events, investor sentiment, and market psychology. As a result, financial time series data often exhibits distinct characteristics that set it apart from other types of time series data. These characteristics include trend, seasonality, volatility clustering, autocorrelation, and heteroskedasticity.

One key characteristic of financial time series is trend, which refers to the long-term movement in prices or values. Trends can be either upward (indicating a bullish market) or downward (indicating a bearish market) and provide insights into the overall direction of the market. Seasonality is another important characteristic found in financial time series. It refers to regular patterns or fluctuations that occur at fixed intervals of time, such as daily, weekly, monthly, or yearly cycles. Seasonality can arise due to various factors, such as calendar events, economic cycles, or investor behavior, and can impact the prices or values of financial assets. Volatility clustering is a phenomenon commonly observed in financial time series, where periods of high volatility tend to be followed by periods of high volatility, and periods of low volatility tend to be followed by periods of low volatility. This clustering effect suggests that volatility is not randomly distributed over time but rather exhibits persistence. Autocorrelation is the correlation between a time series and its lagged values. In financial time series, autocorrelation is often present, indicating that the past values of the series influence its future values. Understanding the autocorrelation structure is crucial for developing appropriate time series models. Heteroskedasticity is the presence of changing variance in a time series. Financial time series often exhibit heteroskedasticity, where the variance of the series changes over time. This changing volatility can have significant implications for risk management and forecasting.

To capture the dynamics and characteristics of financial time series, traditional time series models like the Autoregressive Integrated Moving Average (ARIMA) model have been widely used. ARIMA models incorporate autoregressive, differencing, and moving average components to capture temporal dependencies and trends. However, ARIMA models are not specifically designed to model changing volatility observed in financial time series.

The limitations of traditional models in capturing changing volatility have led to the development of specialized models, such as the Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. ARCH and GARCH models explicitly model the change in variance over time, allowing for more accurate modeling of financial time series exhibiting increasing or decreasing volatility.

In this chapter, we will delve into the intricacies of financial time series analysis and explore the application of ARCH and GARCH models. We will discuss the characteristics of financial time series, the limitations of traditional models, and the need for specialized models such as ARCH and GARCH.

1.2 Financial time series

In the realm of financial analysis, time series data plays a crucial role in understanding the behavior of market assets and making informed decisions. Consider a time series of prices for a financial asset, denoted as S_t, where t represents the time index ranging from 0 to T. However, analyzing the raw price series, S_t, can be challenging since it often exhibits non-stationary behavior due to the presence of a unit root. To address this, it is common practice to analyze log-returns on S_t, denoted as y_t. The log-returns are computed by taking the logarithmic difference between consecutive price observations:

$$y_t = \log(S_t) - \log(S_{t-1}) = \log\left(1 + \frac{S_t - S_{t-1}}{S_{t-1}}\right).$$

By Taylor-expanding the above expression, it becomes apparent that y_t is approximately equivalent to the relative return, $(S_t - S_{t-1})/S_{t-1}$. However, log-returns are preferred over relative returns due to their additive nature, which is not shared by relative returns.

To illustrate this concept, let's consider a concrete example of the daily closing values of the Apple index, which serves as the the closing price of Apple stocks. The data for this series spans from 2/01/2019 to 03/08/2022, and can be obtained from finance.yahoo.com.

Figure 1.1, displayed in the top left plot, visualizes the actual price series, S_t, while the top right plot represents the corresponding log-returns, y_t.

The log-return series, y_t, exhibits several stylized facts commonly observed in financial data. As depicted in the middle right plot of 1.1, the log-return series demonstrates a lack of correlation (typically, log-return series exhibit minimal correlation, particularly at higher lags). However, when we examine the squared log-return series, y_t^2, as shown in the middle left plot, a strong autocorrelation is evident, even for significant lag values.

Another characteristic of financial log-return series is their tendency to display heavy-tailed distributions, indicating the presence of outliers or extreme events. The bottom plot of Figure 1.2 presents a Q-Q plot, comparing the marginal distribution of y_t against the standard normal distribution, further illustrating this heavy-tailed behavior. The plot also demonstrates the existence of the so-called leverage effect. The series y_t exhibits different responses to positive and negative movements, or, in other words, the conditional distribution of $|y_t|\big|\{y_{t-1} > 0\}$ differs from that of $|y_t|\big|\{y_{t-1} < 0\}$. This phenomenon is illustrated by plotting the sample quantiles of the two conditional distributions against each other. The underlying explanation lies in the fact that the market reacts differently to positive and negative news, which aligns with intuitive expectations.

While statisticians typically prefer stationary data as it facilitates global parameter estimation using the entire dataset, proposing a stationary model for y_t that captures the aforementioned stylized facts is challenging. The log-return series does not exhibit a visually apparent stationary pattern due to the presence of clustered local variances (volatility) characterized by periods of low and high values. Consequently, fitting a linear time series model, such as an autoregressive moving average (ARMA) model, to y_t would yield estimated parameters close to zero due to the lack of serial correlation, which is not desirable in this context.

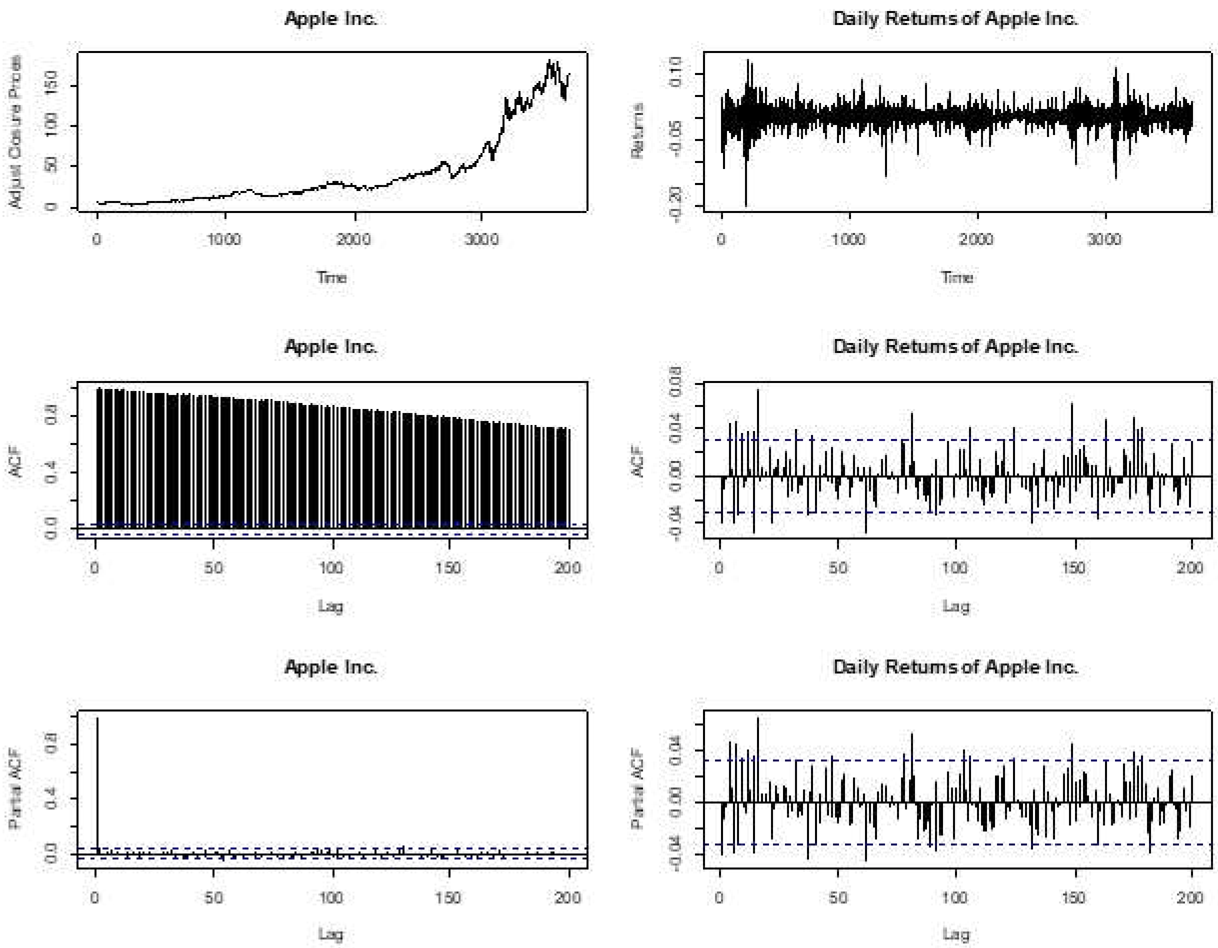

FIGURE 1.1: The top left plot displays the actual price series, S_t, while the top right plot represents the corresponding log-returns, y_t. The middle and bottom plots, show the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the squared log-return series, y_t^2, respectively.

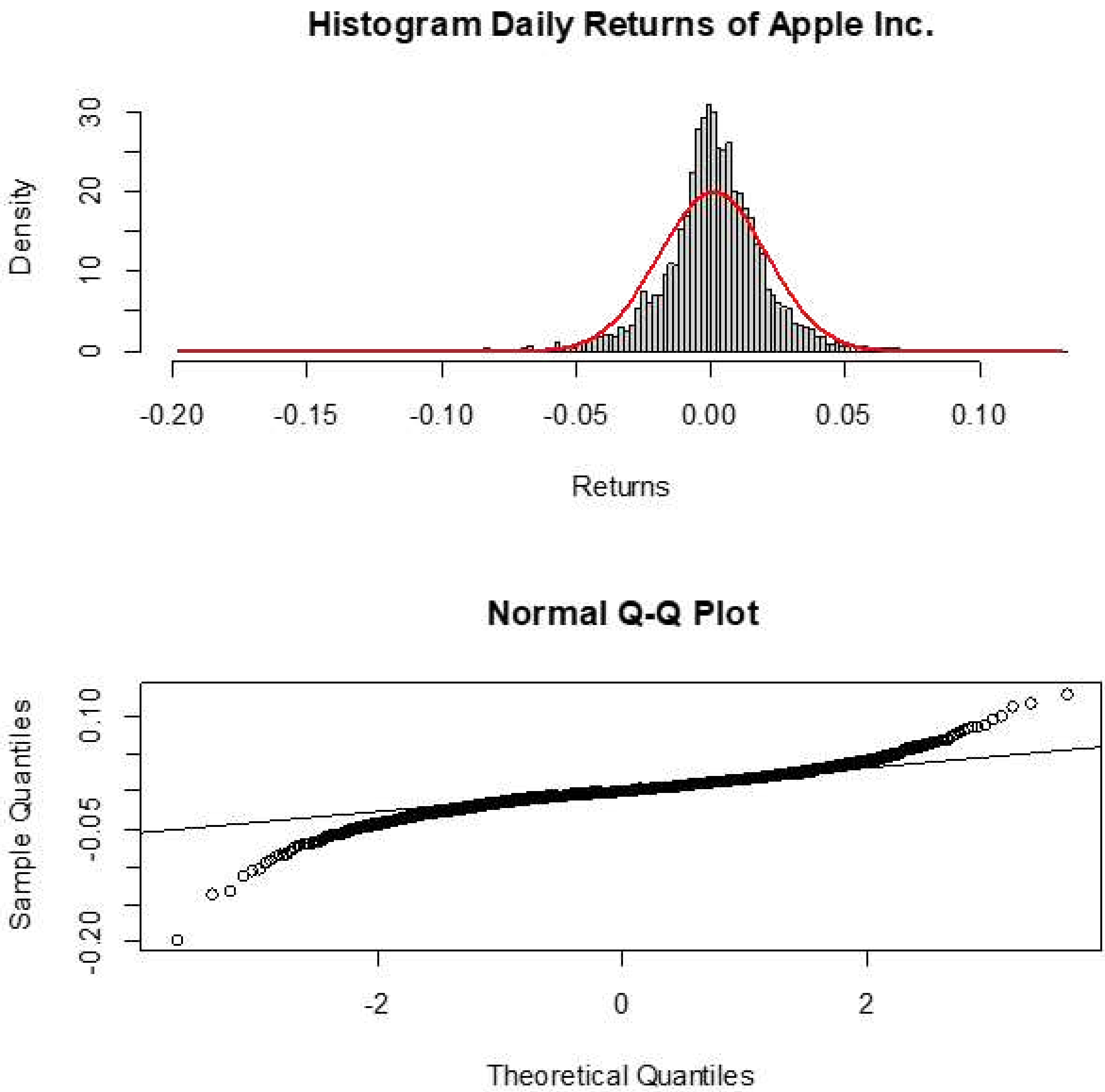

FIGURE 1.2: The top plot displays histograms of the marginal residuals overlaid with estimated marginal densities, while the bottom plot presents the Q–Q plot of residuals based on the Gaussian distribution.

1.2.1 (G)ARCH models

The challenges described earlier led to the development of a non-linear stationary model by (Engle, 1982), known as ARCH (Autoregressive Conditionally Heteroskedastic). The ARCH model incorporates an autoregressive-type process to capture the conditional variance evolution of the log-return series, Y_t. Subsequently, (Bollerslev, 1986) and (Taylor, 2008) independently generalized Engle's model, leading to the creation of the more realistic and widely used "GARCH" model. GARCH (Generalized

Autoregressive Conditionally Heteroscedastic) has become one of the most commonly employed models for financial time series analysis. Its success has inspired the development of numerous sophisticated models within the field. There is a extensive literature on GARCH modeling, with notable references including (Giraitis *et al.*, 2007) and the (Straumann, 2006) book. The GARCH(p, q) model is defined as follows:

$$y_t = \log\left(\frac{S_t}{S_{t-1}}\right) = \omega + \sqrt{h_t}\epsilon_t,$$

$$h_t = \mu_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{1.1}$$

where $\mu_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and the innovation sequence $\{\varepsilon_i\}, -\infty < i < \infty$ is independent and identically distributed with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$, incorporates the concept that the conditional variance (h_t) of y_t, based on information up to time $t-1$, exhibits an autoregressive structure and is positively correlated with its recent past and the squared returns (y_t^2). This framework captures the notion of volatility (or conditional variance) being persistent, whereby large (small) values of y_t^2 are likely to be followed by large (small) values. By incorporating these dynamics, the GARCH(p, q) model allows for a more accurate representation of the changing volatility patterns observed in financial time series.

1.3 Basic properties

1.3.1 Uniqueness and stationarity

Determining the existence of a unique and stationary solution for the system of equations (1.1) involves a complex analysis of the top Lyapunov exponent associated with a specific sequence of random matrices. (Bougerol and Picard, 1992) have presented a comprehensive theorem that provides a necessary and sufficient condition for the existence of such a second-order stationary solution. Furthermore, Proposition 3.3.3 (b) in Straumann's influential work (Straumann, 2006) suggests that a unique and stationary solution can be achieved if the sum of the coefficients α_i (where i ranges from 1 to p) and β_j (where j ranges from 1 to q) is strictly less than 1, expressed mathematically as:

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{1.2}$$

This condition serves as a crucial criterion to ensure the stability and convergence of the second-order stationary solution within the context of the system (1.1).

1.3.2 Mean zero

Introducing the information set, denoted as $\mathcal{F}_{t-1}$ and defined as the sigma-algebra generated by the random variables ε_i, with $-\infty < i \leq t-1$, i.e., $\mathcal{F}t - 1 = \sigma\varepsilon_i, -\infty < i \leq t-1$, we can delve into the intriguing properties of the GARCH model (1.1). Specifically, in the context where h_t is measurable with respect to $\mathcal{F}_{t-1}$, an interesting observation emerges regarding the mean of yt, denoted as $E(y_t)$:

$$\begin{aligned}
E(y_t) &= E(\sqrt{h_t}\varepsilon_t) \\
&= E[E(\sqrt{h_t}\varepsilon_t|\mathcal{F}_{t-1})] \\
&= E[\sqrt{h_t}E(\varepsilon_t|\mathcal{F}_{t-1})] \\
&= E[\sqrt{h_t} \cdot E(\varepsilon_t)] \\
&= E(\sqrt{h_t} \cdot 0) \\
&= 0.
\end{aligned}$$

This result demonstrates that the mean of y_t, within the context of the GARCH model, gracefully vanishes—a fascinating property intricately connected to the underlying dynamics governed by the information set $\mathcal{F}_{t-1}$.

1.3.3 Lack of serial correlation

Similarly, we can demonstrate that y_t is uncorrelated with y_{t+h} for $h > 0$:

$$\begin{aligned}
E(y_t y_{t+h}) &= E(y_t\sqrt{h_{t+h}}\varepsilon_{t+h}) \\
&= E[E(y_t\sqrt{h_{t+h}}\varepsilon_{t+h}|\mathcal{F}_{t+h-1})] \\
&= E[y_t\sqrt{h_{t+h}}E(\varepsilon_{t+h}|\mathcal{F}_{t+h-1})] \\
&= 0.
\end{aligned}$$

This result implies that the cross-covariance between y_t and y_{t+h} is zero, indicating no correlation between them. Through meticulous analysis utilizing the information set $\mathcal{F}_{t+h-1}$, we unveil the intriguing property of uncorrelated dynamics between y_t and y_{t+h}, showcasing the intricate relationships embedded within the GARCH model.

1.4 Unconditional variance

To compute the expected value $E(y_t^2)$, it is beneficial to explore an alternative representation of y_t^2. Let us define the sequence $Z_t = y_t^2 - h_t = h_t(\epsilon_t^2 - 1)$. By following the methodology outlined in Sections 1.3.2 and 1.3.3, it is possible to demonstrate that Z_t constitutes a martingale difference sequence with zero mean. Establishing the "lack of serial correlation" property (which necessitates $E(y_t^4) < \infty$, not always guaranteed) is more intricate. However, this definition offers the advantage of treating Z_t as a "white noise" sequence, enabling the utilization of extensive results on martingale difference sequences, as demonstrated in (Davidson, 1994).

Next, we proceed with the alternative representation:

$$y_t^2 = h_t + Z_t = \mu_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2 - \sum_{j=1}^{q} \beta_j Z_{t-j} + Z_t. \qquad (3)$$

Here, we denote $R = \max(p, q)$, $\alpha_i = 0$ for $i > p$, and $\beta_j = 0$ for $j > q$. Consequently, the above equation can be rewritten as:

$$y_t^2 = \mu_0 + \sum_{i=1}^{R} (\alpha_i + \beta_i) y_{t-i}^2 - \sum_{j=1}^{q} \beta_j Z_{t-j} + Z_t.$$

In essence, y_t^2 follows an autoregressive moving average (ARMA) process with martingale difference innovations.

By utilizing stationarity (implying $E(y_t^2) = E(y_{t+h}^2)$), we can readily obtain the unconditional variance:

$$E(y_t^2) = \mu_0 + \sum_{i=1}^{R} (\alpha_i + \beta_i) E(y_{t-i}^2) - \sum_{j=1}^{q} \beta_j E(Z_{t-j}) + E(Z_t) = \mu_0 + E(y_t^2) \sum_{i=1}^{R} (\alpha_i + \beta_i),$$

leading to the expression:

$$E(y_t^2) = \frac{\omega}{1 - \sum_{i=1}^{R} (\alpha_i + \beta_i)}.$$

This result once again underscores the significance of condition 1.2.

1.5 Heavy tails of y_t

In this section, we present an argument for the potential heavy-tailed nature of the GARCH model (1.1). For the sake of simplicity, we focus on demonstrating this for the

GARCH(1,1) model. We begin by assuming the following condition:

$$E(\alpha_1 \epsilon_t^2 + \beta_1)^{\frac{q}{2}} > 1 \tag{1.3}$$

for some $q > 0$. It is important to note that this condition is satisfied, for example, when $\epsilon_t \sim N(0,1)$, although it holds true in other cases as well.

We express the GARCH(1,1) model as follows:

$$h_{t+1} = \mu_0 + \alpha_1 y_t^2 + \beta_1 h_t = \mu_0 + (\alpha_1 \epsilon_t^2 + \beta_1) h_t$$

By utilizing the independence of ϵ_k from $\mathcal{F}_{k-1}$, we obtain:

$$E(\sqrt{h}_{t+1}^q) = E\left[\mu_0 + (\alpha_1 \varepsilon_t^2 + \beta_1) h_t\right]^{\frac{q}{2}} \geq E[(\alpha_1 \varepsilon_t^2 + \beta_1) h_t]^{\frac{q}{2}} = E(\alpha_1 \epsilon_t^2 + \beta_1)^{\frac{q}{2}} E(\sqrt{h}_t^q)$$

If $E(\sqrt{h}_t^q)$ were finite, then, due to stationarity, it would be equal to $E(\sqrt{h}_{t+1}^q)$. Consequently, simplifying the expression, we would obtain:

$$1 \geq E(\alpha_1 \epsilon_t^2 + \beta_1)^{\frac{q}{2}}$$

This contradicts the assumption stated in Equation (1.3). Thus, we conclude that $E(\sqrt{h}_t^q)$ is infinite, which implies that $E(y_t^2)$ is also infinite. Consequently, y_t does not possess finite moments, indicating a heavy-tailed distribution.

Chapter 2

Mixturing, heavy tailedness, asymmetry and conditional heteroskedasticity in financial returns modelling

2.1 Abstract

conditional heteroskedasticity models are commonly used for modelling financial time series data which are characterized by extreme and/or skewed observations. These data features might not be properly captured by the most commonly adopted distribution. In this chapter, a mixture model for financial time series characterized by conditional heteroscedasticity model is developed, introducing the Finite Mixture of Scale Mixture of Skew Normal of Generalized Autoregressive Conditional Heteroskedastic (FM_m-SMSN-GARCH) model. The SMSN distributions allow for the lightly/heavily-tailed symmetric and asymmetric distributions providing for flexibility to handle outliers and complex data. The proposed model has several desirable features, such as the development of a convenient hierarchical representation of the FM_m-SMSN family that makes it possible to construct a likelihood function to derive the maximum likelihood estimates via an EM–type algorithm. A comprehensive simulation study and real data example allow evaluating the better performance of the proposed method.

Keywords: GARCH models; SMSN distributions; finite mixture models; ECME algorithms; maximum likelihood estimates; stock market, classification.

2.2 Introduction

Statistical modeling and analysis based on finite mixtures of symmetric distributions, especially normal mixtures, have been applied in many fields; see, among many others, (Hunter *et al.*, 2007), (Holzmann *et al.*, 2006), (Salas-Gonzalez *et al.*, 2010) and (Kottas and Fellingham, 2012). In recent years, finite mixtures of asymmetric distributions have been developed as a powerful substitute to the normal mixtures in a wide range of applications. The benefit of asymmetric finite mixture models is to accommodate different characteristics, such as heavy tail, skewness, kurtosis, multimodality, and the presence of outliers; these features can be observed in many fields, such as finance, biostatistics, bioinformatics, image analysis, and medicine. Comprehensive surveys are available in (Lindsay, 1995), (Böhning, 1999), (Peel and McLachlan, 2000), (Frühwirth-Schnatter, 2006), (Mengersen *et al.*, 2011), (McLachlan *et al.*, 2019). (Pyne *et al.*, 2009), (McCulloch and Tsay, 1994), (Lee and McLachlan, 2013b) and (Manouchehri and Nematollahi, 2019).

We focus here on financial data modeling which is considered as a complex task, given that these time series exhibit skewness, kurtosis, heavy tail and multimodality. Financial returns usually reverberate a structure which can be logically illustrated with conditional heteroskedastic models, such as the Autoregressive Conditional Heteroskedastic (ARCH) process proposed by (Engle, 1982), as well as some variants and extensions of ARCH models, including the Generalized Autoregressive Conditional Heteroskedastic (GARCH) process of (Bollerslev, 1986), the Glosten - Jagannathan - Runkle GARCH (GJR-GARCH) model of (Glosten *et al.*, 1993) [1]. In many empirical applications, the distribution of model innovations follow simple specifications, the Guassian or the Student-T. Only part of the literature has focused on the use of flexible distributional assumptions.

In this chapter, we contribute to this field by introducing a GARCH model whose innovations follow a mixture of asymmetric distribution. Interestingly, in our proposal the components of the mixture are scale-mixtures of skew-normal distributions (SMSN), proposed by (Branco and Dey, 2001), an absorbing and acutely flexible family of probabilistic distributions. The SMSN class contains the whole family of scale mixtures of normal (SMN) distributions proposed by (Andrews and Mallows, 1974) as well as the normal and skew-normal (SN) densities. Furthermore, the Cauchy, the skew-Cauchy,

[1] We refer the reader to (Bollerslev *et al.*, 2010) for a more complete listing of the GARCH models develped within the financial econometrics literature in the last 40 years.

the skew-normal (SN), the skew-t (ST), the skew-slash (SSL) and the skew contaminated normal (SCN) distributions are particular symmetric and asymmetric members of this family and are all characterized by heavier tails than the SN (and the normal) one. This class of distributions leads to a more flexible approach to fit asymmetric and heavy-tailed empirical distributions and uses fewer components in the fitting of mixture models; see, e.g (Lin, 2010), (Lee and McLachlan, 2013a) and (Lee and McLachlan, 2013b). Finite mixtures with SMSN component densities is a rich class of distribution to simultaneously accommodate kurtosis, skewness and multimodality. Therefore, a GARCH model, together with the SMSN family for the innovations can be a suitable choice for modelling financial data. We refer to this novel family of mixture GARCH model as FM_m-SMSN-GARCH. This new mixture GARCH family has a useful stochastic representation and a convenient hierarchical representation that allows easier inferences and parameters estimation. Notably, thank to a representation of SMSN parameter estimation of our new GARCH family can be obtained within an Expectation–Maximization (EM) framework.

Among the several contributions focusing on the SMSN we mention (Azzalini and Capitanio, 1999) who raise various issues related to the skew normal family, (Basso *et al.*, 2010) who develop a robust approach to finite mixture modeling based on scale mixtures of skew-normal distributions, (Arellano-Valle *et al.*, 2006) who propose a unified framework for selection distributions and describe some of their theoretical properties and application, Azlini's recent book in collaboration with Capitanio (Azzalini, 2013) , and (Manouchehri and Nematollahi, 2022b) who consider a Bayesian and non-Bayesian approaches for the Periodic AR Models Based on the SMSN Innovations.

In this chapter, we develop two extensions of the EM-algorithm (Dempster *et al.*, 1977), including the ECM algorithm (Meng and Rubin, 1993) and the ECME algorithm (Liu and Rubin, 1994) using the stochastic representation of the proposed model.

The outline of the chapter is as follows: In Section 3.3, the useful properties of the SMSN family are reviewed for the aim of this work. In Section 3.4, the proposed GARCH model with FM_m-SMSN innovations is presented, where the estimation method are proposed to estimate the model parameters, by using the ECME algorithms. The Numerical studies and real practical examples are reported in Section 3.5. Finally, a brief discussion is given in Section 3.6.

2.3 Preliminaries

In this section, we provide a brief overview of the scale mixtures of skew-normal distributions (SMSN) introduced by (Branco and Dey, 2001). Additionally, we review the key properties and characteristics of the finite mixture of scale-mixtures of skew-normal (FM_m-SMSN) distributions.

2.3.1 The SMSN Distribution

The skew-normal (SN) distribution, initially proposed by (Azzalini, 1985), describes a random variable Z with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma^2 > 0$, and skewness/shape parameter λ. The probability density function (PDF) of the SN distribution is defined as follows:

$$\phi_{SN}(z; \mu, \sigma^2, \lambda) = 2\phi(z; \mu, \sigma^2)\Phi\left(\frac{\lambda(z-\mu)}{\sigma}\right), \quad z \in \mathbb{R}, \tag{2.1}$$

where $\phi(z; \mu, \sigma^2)$ represents the PDF of the univariate normal distribution with mean μ and variance σ^2, while $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) of the univariate standardized normal distribution. We denote the SN distribution as $Z \sim SN(\mu, \sigma^2, \lambda)$ to indicate that a random variable Z follows the SN distribution defined by Equation (2.1).

A random variable Y belongs to the SMSN family with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$ if it can be expressed as:

$$Y = \mu + k^{\frac{1}{2}}(U)\sigma Z_0, \tag{2.2}$$

where $Z_0 \sim SN(0, 1, \lambda) = SN(\lambda)$ represents a standard SN random variable, and U is a random variable with cumulative distribution function (CDF) $H(\cdot; \nu)$ and probability density function (PDF) $h(\cdot; \nu)$, where ν serves as the scale factor parameter. The distribution of U is typically indexed by ν, which can be a (potentially multivariate) parameter. Importantly, U is independent of Z_0. The function $k(\cdot)$ is a positive function, and μ denotes the location parameter.

The SMSN family gets its name from the fact that the conditional distribution of Y given $U = u$ follows a skew-normal distribution. Specifically, the density function of Y

is given by:

$$
\begin{aligned}
g(y;\mu,\sigma^2,\lambda,\nu) &= \int_0^\infty f(Y=y|U=u)dH(u;\nu) \\
&= \int_0^\infty 2\phi(y;\mu,k(u)\sigma^2)\Phi\left(k^{\frac{1}{2}}(u)\frac{\lambda(y-\mu)}{\sigma}\right)dH(u;\nu), \quad y\in\mathbb{R},
\end{aligned}
\tag{2.3}
$$

which means, $g(y;\mu,\sigma^2,\lambda,\nu)$ represents an infinite mixture of skew-normal densities, where U serves as the scale factor and $H(\cdot;\nu)$ is the mixing distribution. The random variable Y follows an SMSN distribution, denoted as $Y \sim SMSN(\mu;\sigma^2,\lambda,\nu)$ or $Y \sim SMSN(\mu;\sigma^2,\lambda,H)$, highlighting its dependency on the location parameter μ, scale parameter σ^2, skewness parameter λ, and the distribution H.

For further insights and statistical properties of the SMSN family, (Andrews and Mallows, 1974) offers a comprehensive reference. The following proposition summarizes some fundamental properties of the SMSN family.

Proposition 2.1. *(Andrews and Mallows, 1974) Suppose $Y \sim SMSN(\mu;\sigma^2,\lambda;H)$, where $U \sim H$ is the mixing random scale factor. Then the following properties hold:*

1. *If $E(k^{\frac{1}{2}}(U))<\infty$, then $\mu = E(Y) = \mu - b\Delta$.*

2. *If $E(k(U))<\infty$, then $\sigma_Y^2 = Var(Y) = \sigma^2 k_2 - \sigma^2 b^2\delta^2$.*

3. *If $E(k(U))<\infty$, the skewness coefficient is given by:*

$$
\gamma_1 = \frac{E(Y-\mu_y)^3}{[Var(Y)]^{\frac{3}{2}}} = \frac{\delta b_1 + \delta^3 b_2}{[k_2 - b^2\delta^2]^{\frac{3}{2}}},
$$

4. *If $E(k^2(U))<\infty$, the kurtosis coefficient is given by:*

$$
\gamma_2 = \frac{E(Y-\mu_y)^4}{[Var(Y)]^2} - 3 = \frac{3k_4 + \sqrt{\frac{2}{\pi}}\delta^2 c(b_4\delta^2 - b_3)}{[k_2 - b^2\delta^2]^2},
$$

5. *A random variable Y can be stochastically represented as:*

$$
Y = \mu + \Delta W + \varrho k^{\frac{1}{2}}(U)W_1, \tag{2.4}
$$

where $b = -\sqrt{\frac{2}{\pi}}k_1$, $k_n = E[k^{\frac{n}{2}}(U)]$, $b_1 = 3\sqrt{\frac{2}{\pi}}(k_3 - 3k_1k_2)$, $b_2 = 2\left(\frac{2}{\pi}\right)^{\left(\frac{3}{2}\right)}{k_1}^3 - \sqrt{\frac{2}{\pi}}k_3$, $b_3 = 6(2k_3 - k_1k_2)$, $b_4 = 4k_3 - \frac{6}{\pi}k_1^3$, $\Delta = \sigma\delta$, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, $\varrho^2 = \sigma^2 - \Delta^2$, $W = k^{\frac{1}{2}}(U)|W_0|$. It is worth noting that the relationships between (Δ,ϱ^2) and (σ,λ) are one-to-one, where $\sigma^2 = \varrho^2 + \Delta^2$ and $\varrho^2 = \frac{\Delta}{\lambda}$.

The SMSN family encompasses various special cases that arise from utilizing different distributions for the scale mixing random variable U. Here are a few notable examples:

- Skew-normal (SN): In this case, U is a constant with a probability of one.
- Skew-t (ST): Here, U follows a gamma distribution with shape and rate parameters both equal to $\frac{\nu}{2}$, denoted as $U \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$. The parameter ν represents the degrees of freedom.
- Skew-slash (SSL): In the SSL distribution, U is distributed according to a beta distribution with parameters ν and 1, denoted as $U \sim \text{Beta}(\nu, 1)$. The parameter ν is constrained to be greater than 0, i.e., $\nu > 0$.
- Skew-contaminated-normal (SCN): The SCN distribution introduces a mixture of two point masses at γ and 1, represented by the probability density function $h(u; \nu) = \nu \mathbb{I}(u = \gamma) + (1 - \nu)\mathbb{I}(u = 1)$, where $\mathbb{I}$ denotes the indicator function. The parameter vector is given by $\boldsymbol{\nu} = (\nu, \gamma)'$, with $0 < \nu < 1$ and $0 < \gamma \leq 1$.

  Please note that the Skew Normal (SN) density is a light-tailed distribution, whereas the Skew Student's t (ST) and Skew Slash (SSL) densities are heavy-tailed distributions. In Figure 2.1 and Figure 2.2, we provide graphs of both the light-tailed SN densities and the heavy-tailed ST densities with different shape parameters (1 and 0.9 for asymmetry and 0 for symmetry) and various degrees of freedom (30, 3, and 2.1).

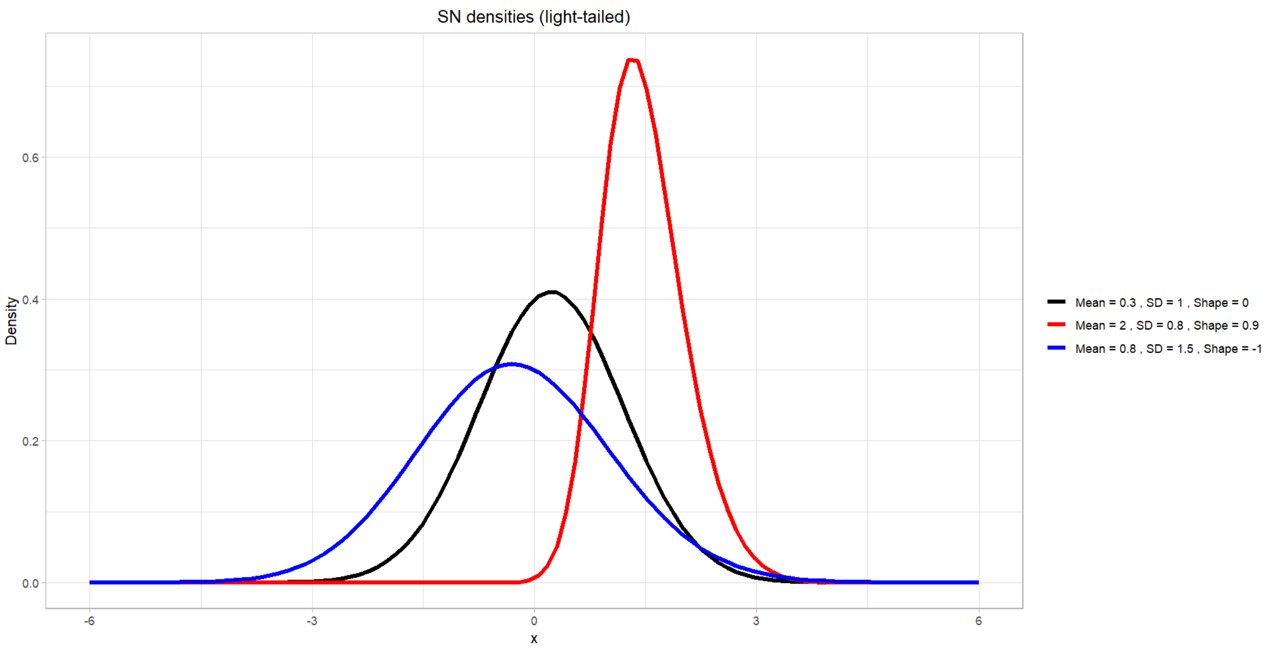

FIGURE 2.1: Some typical graphs of the light-tailed SN densities with various shape, scale and degrees of freedom parameters.

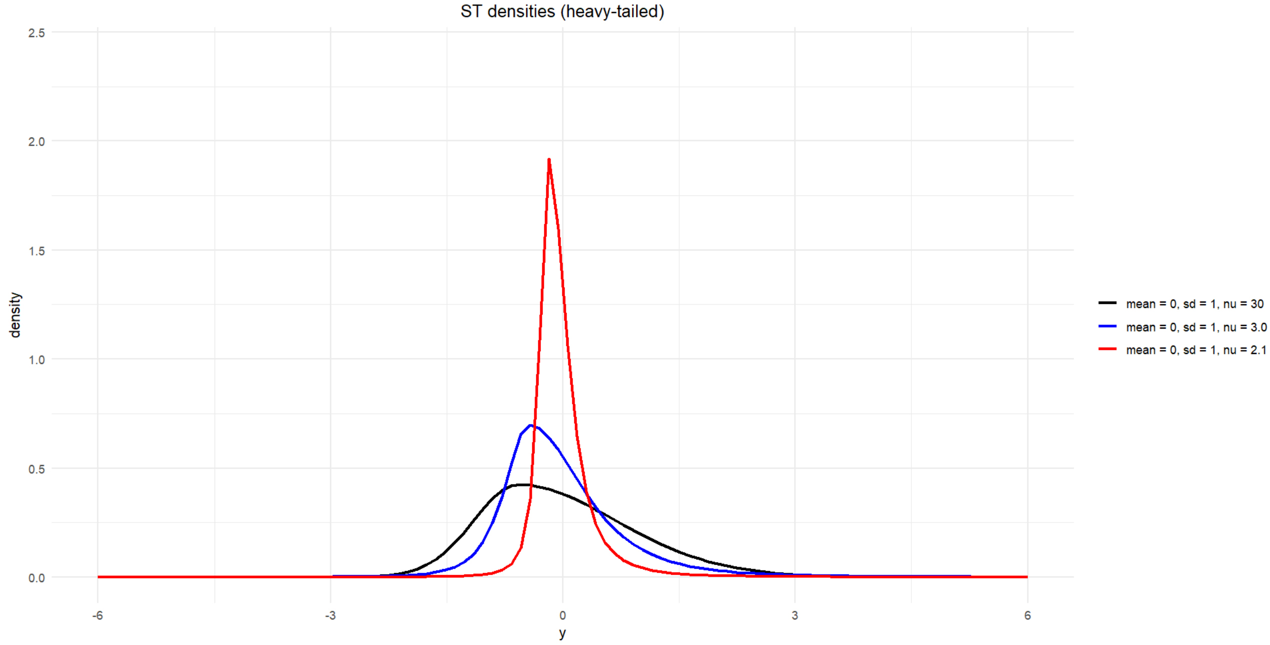

FIGURE 2.2: Some typical graphs of the heavy-tailed ST densities with various shape, scale and degrees of freedom parameters.

For additional technical details and information on SMSN distributions, as well as a proposition for computing the conditional moments necessary for the calculation of certain conditional expectations in the proposed EM-algorithm, please refer to Appendix A.1.

2.3.2 The FM$_m$-SMSN Distributions

The finite mixture of scale mixture of skew normal (FM$_m$-SMSN) distribution is a m-component mixture of SMSN distributions. This mixture distribution is characterized as follows:

$$f(y; \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\nu}) = \sum_{i=1}^{m} \pi_i g_i(y; \boldsymbol{\theta}_i, \boldsymbol{\nu}_i), \quad y \in \mathbb{R}, \tag{2.5}$$

where $\boldsymbol{\pi} = (\pi_1, ..., \pi_m)'$, with $\pi_i > 0$, $i = 1, ..., m$, are the mixing probabilities, such that $\sum_{i=1}^{m} \pi_i = 1$, $\boldsymbol{\theta} = (\boldsymbol{\theta_1}', ..., \boldsymbol{\theta_m}')'$, $\boldsymbol{\theta_i} = (\mu_i, \sigma_i{}^2, \lambda_i)'$, $\boldsymbol{\nu} = (\boldsymbol{\nu}'_1, ...\boldsymbol{\nu}'_m)'$, and as we mentioned before in equation (2.3), $g_i(.; \boldsymbol{\theta}_i, \boldsymbol{\nu}_i)$ is an $SMSN(\boldsymbol{\theta}_i, \boldsymbol{\nu}_i)$-component density for $i = 1, ..., m$.

To denote that a random variable Y has a m-component FM$_m$-SMSN density which showed in equation (2.5), we write $Y \sim FM_m - SMSN(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\nu})$. In terms of the components of the mixture, Equation (2.5) can be equivalently obtained by

$$Y|Z_i = 1 \sim SMSN(\boldsymbol{\theta}_i, \boldsymbol{\nu}_i), \quad i = 1, ..., m, \tag{2.6}$$

where $Z_1, ..., Z_m$ are independent random vectors, each one having a multinomial distribution, $\boldsymbol{Z} = (Z_1, ..., Z_m)' \sim Multinomial(1, \pi_1, \pi_2, ..., \pi_m)$ and with probability mass function is given by

$$p(Z_1 = z_1, ..., Z_m = z_m) = {\pi_1}^{z_1}{\pi_2}^{z_2}...{\pi_m}^{z_m}, \quad \{z_i = 0, 1\}, \quad i = 1, ..., m, \quad \sum_{i=1}^{m} z_i = 1.$$

We notice that integrating out $\boldsymbol{z} = (z_1, ..., z_m)'$, the marginal density in (2.5) can be obtained. The distribution of Y corresponds to the *ith* component of the mixture because just one component of Z can be equal to one and those that remain are zero $(Z_i = 1, \quad Z_j = 0 \quad \forall i \neq j)$; In order to see more detail, see (McLachlan *et al.*, 2019). The following proposition summarizes some properties of the FM$_m$-SMSN family.

Proposition 2.2. *(McLachlan* et al.*, 2019) Assuming, for $i = 1, ..., m$, that $k_{ri} = E({U_i}^{-r/2}), r = 1, 2$, where $U_i \sim H(; \nu_i)$, then the mean and variance of the random variable Y which is a FM_m-SMSN distribution are, respectively, given by*

$$E(Y) = \sum_{i=1}^{m} \pi_i(\mu_i - b_i\Delta_i), \tag{2.7}$$

and

$$Var(Y) = \sum_{i=1}^{m} \pi_i\Big(\big(k_{2i}\sigma_i^2 - b_i^2\Delta_i^2\big) + \big((\mu_i - \bar{\mu}) - (b_i\Delta_i - \bar{b\Delta})\big)^2\Big), \tag{2.8}$$

where $\bar{\mu} = \sum_{i=1}^{m} \pi_i\mu_i$, $\bar{b\Delta} = \sum_{i=1}^{m} \pi_i b_i \Delta_i$ such that $b_i = -\sqrt{\frac{2}{\pi}}k_{1i}$, and $\Delta_i = \sigma_i\delta_i$, such that $\delta_i = \frac{\lambda_i}{\sqrt{1+\lambda_i^2}}, \quad i = 1, ..., m$.

Therefore, by substituting (2.3) in (2.5), the density of the FM$_m$-SMSN family is given by

$$\begin{aligned} f(y; \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\nu}) &= \sum_{i=1}^{m} \pi_i g_i(y; \boldsymbol{\theta}_i, \boldsymbol{\nu}_i) \\ &= \sum_{i=1}^{m} \pi_i \int_0^{\infty} 2\phi(y; \mu_i, u_i^2\sigma_i^2)\Phi(\sqrt{u_i}\lambda_i(y - \mu_i)/\sigma_i)dH(u_i; \nu_i). \end{aligned} \tag{2.9}$$

Also, we note that when $\lambda_i = 0, \quad i = 1, ..., m$, the finite mixture of scale mixture normal (FM$_m$-SMN) model is obtained as a symmetric case of this family. If $H(1; \boldsymbol{\nu_i}) = 1$, the finite mixture of skew normal (FM_m-SN) model is generated from equation (2.9), and the other heavy tails distributions of this family are the finite mixture of skew-t (FM_m-ST), finite mixture of skew-slash (FM_m-SSL) and finite mixture of skew-contaminated-normal (FM_m-SCN), with ST, SCN and SSL component densities, respectively.

2.4 The FM_m-SMSN-GARCH(1,1) Model

In this section, we first introduce the proposed model, and we then develop a EM-type algorithm to estimate unknown parameters. A method for selecting the best model is also introduced. Finally, the observed information matrix of the proposed model is obtained.

2.4.1 Model specification

To account for the empirical evidence suggesting a time variation in volatility, (Engle, 1982) and (Bollerslev, 1986) proposed the Autoregressive Conditional Heteroscedasticity (ARCH) and the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models, respectively. The dynamics of a stock price process $(S_t)_{(t\in 0,1,\dots,T)}$ are characterized by the stationary process $(Y_t)_{(t\in 0,1,\dots,T)}$ defined by:

$$Y_t = log\Big(\frac{S_t}{S_{t-1}}\Big) = \omega + \sqrt{h_t}\epsilon_t, \tag{2.10}$$

where $(\epsilon_t)_{(t\in 0,1,\dots,T)}$, are independent identically distributed random variables defined on the sample space $(\Omega, \mathcal{F}, P)$ and $\Big(\mathcal{F}_t = \sigma\big(\epsilon_u, 1 \leq u \leq t\big)\Big)_{t\in(0,1,\dots,T)}$ is the associated information filtration. The volatility process $\sqrt{h_t}$ is driven by the set of parameters $\boldsymbol{\theta}_V$ and the distribution of $\epsilon_t \sim WN(0,1)$ by a set of parameters $\boldsymbol{\theta}_D$. In the following, we will refer to ϵ_t as the model innovation.

In an ARCH (1) model, the conditional volatility is a linear function of the square of the previous innovations:

$$Var(y_t|\mathcal{F}_{t-1}) = h_t = \mu_0 + \alpha_1(y_{t-1} - \omega)^2, \tag{2.11}$$

where $\mu_0 > 0$, is a constant, and α_1 is a non-negative parameter.

In a GARCH (1, 1) model, the conditional volatility is a linear function of both the square of the previous observed innovations and of the previous conditional volatilities:

$$Var(y_t|\mathcal{F}_{t-1}) = h_t = \mu_0 + \alpha_1(y_{t-1} - \omega)^2 + \beta_1 h_{t-1}, \tag{2.12}$$

where $\mu_0 > 0$, is a constant, α_1 and β_1 are non-negative parameters, to ensure non-negativity of conditional variance and $\alpha_1 + \beta_1 < 1$, to ensure covariance stationarity. For $t = 1, ..., T$, h_t is the conditional variance of y_t given $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$, the information set available until time $t - 1$.

There are many papers using non-Gaussian innovations and most of them are related to symmetric distribution. Our purpose is to use a flexible characterization of the innovations in the same spirit of (Haas *et al.*, 2004) that are using a mixed normal distribution coupled with a GARCH-type structure (termed MN-GARCH). Their model allows for conditional variance in each of the components as well as dynamic feedback between the components However, our proposal is different as we point at introducing a mixture on the innovation process. Our approach will allow for a more flexible parametrization of the innovations and, for simplicity, is introduced with a GARCH model for volatility. However, more flexible parametrizations for the conditional variance can be used.

In this paper, the innovation of the proposed model, ϵ_t, (see equation 2.10) follows a mixture:

$$f(\epsilon_t; \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\nu}) = \sum_{i=1}^{m} \pi_i g_i(\epsilon_t; \boldsymbol{\theta}_i, \boldsymbol{\nu}_i). \tag{2.13}$$

The following conditional density is obtained using equations, (2.10), (2.12), and (2.13):

$$f_{y_t|\mathcal{F}_{t-1}}(y_t, \boldsymbol{\Theta}) = \sum_{i=1}^{m} \pi_i h_t^{-\frac{1}{2}} \int_0^\infty \phi_{SN}\Big(\frac{y_t - \omega}{\sqrt{h_t}}; \mu_i, u_{ti}^{-1}\sigma_i^2, \lambda_i\Big) dH(u_{ti}; \nu_i); \quad t = 1, ..., T, \tag{2.14}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\varphi}', \boldsymbol{\pi}', \boldsymbol{\theta}', \boldsymbol{\nu}')'$, with $\boldsymbol{\varphi}' = (\omega, \mu_0, \alpha_1, \beta_1)$ being the parameter vector of the GARCH model.

We also mention that the number of components of the model is denoted by m, which is a known positive integer and that the weight of the i-th component is denoted by $\pi_i > 0$, and $\sum_{i=1}^{m} \pi_i = 1$, which follows a distribution $\boldsymbol{\pi} = (\pi_1, ..., \pi_m)'$. In addition, the innovation parameters of the GARCH model are represented by $\boldsymbol{\theta} = ({\boldsymbol{\theta}_1}', ...{\boldsymbol{\theta}_m}')'$ with $\theta_i = (\mu_i, {\sigma_i}^2, \lambda_i)'$, $\boldsymbol{\nu} = ({\boldsymbol{\nu}_1}', ...{\boldsymbol{\nu}_m}')'$ and the density of the innovation of the i-th component is an $SMSN(\boldsymbol{\theta}_i, \boldsymbol{\nu}_i) - distribution.$

For computational convenience, we assume that the m-component densities are obtained from the same distribution, meaning that $g_i(y; \boldsymbol{\theta}_i, \boldsymbol{\nu}_i) = g(y; \boldsymbol{\theta}_i, \boldsymbol{\nu}_i)$, and for ν_i

we also assume that $\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_m = \boldsymbol{\nu}$. This strategy works very well in the experimental studies we have done and greatly simplifies the optimization problem.

We remind that for $t = 1, ..., T$, the random variables ϵ_t are independent and identically distributed, so that $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = 1$. Therepfore, we must introduce restrictions on the parameters of the FM$_m$-SMSN to ensure that the GARCH innovations have unit variance and a null mean. These are provided in the following proposition.

Proposition 2.3. *Let $\epsilon_t \sim FM_m - SMSN(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\nu})$, then the zero mean and unit variance restrictions in 2.7 amd 2.8, respectively, imply that*

$$\mu_1 = b_1\Delta_1 - \frac{\sum_{i=1}^{m-1} \pi_i(\mu_i - b_i\Delta_i)}{\pi_1}, \tag{2.15}$$

and

$$\begin{aligned} \sigma_1^2 = &\frac{1 - \sum_{i=1}^{m-1} \pi_i\Big(\big(k_{2i}\sigma_i^2 - b_i^2\Delta_i^2\big) + \big((\mu_i - \bar{\mu}) - (b_i\Delta_i - \bar{b\Delta})\big)^2\Big)}{k_{21}} \\ &+ \frac{-\Big(\big((\mu_1 - \bar{\mu}) - (b_1\Delta_1 - \bar{b\Delta})\big)^2\Big) + b_1^2\Delta_1^2}{k_{21}}, \end{aligned} \tag{2.16}$$

where $\bar{\mu} = \sum_{i=1}^{m} \pi_i\mu_i$, $\bar{b\Delta} = \sum_{i=1}^{m} \pi_i b_i\Delta_i$ such that $b_i = -\sqrt{\frac{2}{\pi}}k_{1i}$, and $\Delta_i = \sigma_i\delta_i$, and $\delta_i = \frac{\lambda_i}{1+\lambda_i^2}$, $\quad i = 1, ..., m$.

Proof. The proof is provided in Appendix A.3. □

We note that to impose zero mean and unit variance, some constraints are imposed on the parameters of the first component of the mixture for convenience only.

2.4.2 ML parameter estimation via the ECME algorithm

We now proceed to the ML estimation of the unknown parameters of an FM$_m$-SMSN-GARCH(1,1) using the Expectation/Conditional Maximisation Either (ECME) algorithm, proposed by (Liu and Rubin, 1994), as an extension of the EM algorithm of (Dempster *et al.*, 1977). Without loss of generality, in this study, we suppose that the initial volatility, h_0, and initial value of the return series , y_0, are known.

By using the Markovian property of the GARCH(1,1) process, the likelihood function of the GARCH model with FM$_m$-SMSN innovations is obtained by

$$L(\boldsymbol{\Theta}, \boldsymbol{y}) = f(y_T|\mathcal{F}_{T-1})f(y_{T-1}|\mathcal{F}_{T-2}), ...f(y_1|\mathcal{F}_0) = \prod_{t=1}^{T} f(y_t|\mathcal{F}_{T-1}), \quad t = 1.....T.$$

Therefore, the respective log-likelihood function is provided by

$$\begin{aligned}
\ell(\boldsymbol{\Theta}, \boldsymbol{y}) &= \sum_{t=1}^{T} \ell_t(\boldsymbol{\Theta}, \boldsymbol{y}) \\
&= \sum_{t=1}^{T} \log \left(\sum_{i=1}^{m} \pi_i h_t^{-\frac{1}{2}} \int_0^{\infty} \phi_{SN}\Big(\frac{y_t - \omega}{\sqrt{h_t}}; \mu_i, u_{ti}^{-1}\sigma_i^2, \lambda_i\Big) dH(u_{ti}; \nu_i) \right) \\
&= \sum_{t=1}^{T} \log \left(\sum_{i=1}^{m} 2\pi_i h_t^{-\frac{1}{2}} \int_0^{\infty} \phi\Big(\frac{y_t - \omega}{\sqrt{h_t}}; \mu_i, u_{ti}^{-1}\sigma_i^2\Big) \Phi\Big(\sqrt{u_{ti}}\lambda_i\big(y_t - \omega - \mu_i\sqrt{h_t}\big)/\sqrt{h_t}\sigma_i\Big) dH(u_{ti}; \nu_i) \right).
\end{aligned} \tag{2.17}$$

Due to the complexity of the above log-likelihood function, there is no analytical solution to achieve the ML estimate of unknown parameter and maximization of the proposed log-likelihood function with respect to $\boldsymbol{\Theta}$ requires a high dimensional nonlinear optimization procedure. Therefore, we need to develop a numerical search algorithm. To do this, an innovative EM-type algorithm has been developed to obtain the ML estimate for the FM$_m$-SMSN-GARCH(1,1) model using the hierarchical representation of the proposed model.

To overcome computational complexity, we proceed with the derivation of the hierarchical representation of the FM$_m$-SMSN-GARCH(1,1) model. This will allow to estimate the parameters by resorting to an efficient expectation conditional maximization either (ECME) algorithm. We first define an auxiliary indicator for the component distributions in the GARCH mixture model, which can be expressed in terms of a multinomial random vector $\boldsymbol{Z}_t = (Z_{t1}, ..., Z_{tm})'$, which is denoted by

$$\boldsymbol{Z}_t = (Z_{t1}, ..., Z_{tm})' \sim Multinomial(1, \pi_1, \ldots, \pi_m),$$

where

$$Z_{ti} = \begin{cases} 1 & \text{if t-th observation belongs the i-th component;} \\ 0 & ow. \end{cases}, t = 1, ..., T; \quad i = 1, ..., m.$$

Only one element of $\boldsymbol{Z}_t$ is one (while the rest of elements are zero), and its label/-position determines the component-distribution for the tth innovation; see (Peel and McLachlan, 2000) for additional details. Also, according to this assumption y_t is independent of ϵ_t, and therefore also of $\boldsymbol{Z}_t$. Therefore, the proposed GARCH model can be equivalently represented by

$$\frac{y_t - \omega}{\sqrt{h_t}} | \mathcal{F}_{t-1}, Z_{ti} = 1 \sim SMSN(\mu_i, \sigma_i^2, \lambda_i, \boldsymbol{\nu}), \tag{2.18}$$

$$\boldsymbol{Z}_t \sim Multinomial(1, \boldsymbol{\pi}),$$

for $i = 1, ..., m, \quad t = 1, ..., T$. In addition, in order to obtain a likelihood function based on complete data, the proposed mixture-GARCH model can be easily displayed using the stochastic representation of the SMSN family given in 2.4:

$$y_t|\mathcal{F}_{t-1}, W_{ti} = w_{ti}, U_{ti} = u_{ti}, Z_{ti} = 1 \sim N\big(\omega + \sqrt{h_t}(\mu_i + \Delta_i w_{ti}), u_{ti}^{-1}\varrho_i^2 h_t\big), \tag{2.19}$$

$$W_t|U_{ti} = u_{ti}, Z_{ti} = 1 \sim TN_1\big(0, u_{ti}^{-1}\big)I_{(0,\infty)},$$

$$U_{ti}|Z_{ti} = 1 \sim H\big(u_{ti}; \boldsymbol{\nu}_i\big),$$

$$\boldsymbol{Z}_t \sim Multinomial(1, \boldsymbol{\pi}),$$

for $i = 1, ..., m, \quad t = 1, ..., T$, where $\varrho_i^2 = \sigma_i^2 - \Delta_i^2$, and the sequences $\big\{\mathcal{F}_{t-1}, ; t = 1, ..., T\big\}$, $\big\{Z_{ti}; i = 1, ..., T\big\}$, and $\big\{(U_{ti}, W_{ti}); i = 1, ..., m, t = 1, ..., T\big\}$ are independent. Also note that, the parameter vector, $\boldsymbol{\theta_i} = (\mu_i, \sigma_i^2, \lambda_i)'$ can be easily redefined by $\boldsymbol{\theta_i} = (\mu_i, \varrho_i^2, \Delta_i^2)'$, $i = 1, ..., m$, because of the objective relation between the two sets.

Let $\boldsymbol{\Omega} = \big\{\boldsymbol{y}, \boldsymbol{P}\big\}$, denote the complete data, where $\boldsymbol{y} = (y_1, ..., y_T)'$, and $\boldsymbol{P} = \big\{\boldsymbol{P}_{ti} = (W_{ti}, U_{ti}, Z_{ti})_{t=1}^T\big\}_{i=1}^m$ are the observable part (observed data) and missing part (hidden variables), respectively. Therefore, the likelihood-function based on the complete data is given by

$$L_{\boldsymbol{\Omega}}(\boldsymbol{\Theta}) = \prod_{t=1}^{T}\prod_{i=1}^{m}\Big(\pi_i f_{y_t|\mathcal{F}_{t-1}, W_{ti}, U_{ti}, Z_{ti}}\big(y_t; \boldsymbol{\varphi}, \Delta_i, \varrho_i^2\big) f_{W_{ti}|U_{ti}, Z_{ti}}\big(w_{ti}, \boldsymbol{\nu_i}\big) f_{U_{ti}|z_{ti}}\big(u_{ti}, \boldsymbol{\nu_i}\big)\Big)^{Z_{ti}}, \tag{2.20}$$

where
$f_{y_t|\mathcal{F}_{t-1}, W_{ti}, U_{ti}, Z_{ti}}\big(y_t; \boldsymbol{\varphi}, \Delta_i, \varrho_i^2\big) = \phi\Big(y_t; \omega + \sqrt{h_t}(\mu_i + \Delta_i w_{ti}), u_{ti}^{-1}\varrho_i^2 h_t\Big)$, $f_{W_{ti}|U_{ti}, Z_{ti}}\big(w_{ti}, \boldsymbol{\nu_i}\big) = TN_1\big(0, u_{ti}^{-1}\big)I_{(0,\infty)} = 2\phi\Big(w_{ti}; 0, u_{ti}^{-1}\Big)$, with $\phi(.)$ as the normal probability density function, and $f_{U_{ti}|z_{ti}}\big(u_{ti}, \boldsymbol{\nu_i}\big) = h\big(u_{ti}; \boldsymbol{\nu}_i\big)$, is a density function which is induced by the mixing distribution $H\big(u_{ti}; \boldsymbol{\nu}_i\big)$. Thus, the respective log likelihood-function based on the complete data using equation (2.20) is obtained by

$$\begin{aligned}\ell_{\boldsymbol{\Omega}}(\boldsymbol{\Theta}) = &\sum_{i=1}^{m}\sum_{t=1}^{T} Z_{ti}\log h\big(u_{ti};\boldsymbol{\nu}_i\big) + \sum_{i=1}^{m}\sum_{t=1}^{T} Z_{ti}\log \pi_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T} Z_{ti}\log \varrho_i^2 \\ &- \frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T} Z_{ti}h_t - \frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T} \frac{Z_{ti}U_{ti}}{h_t\varrho_i^2}\big(y_t-\omega-\sqrt{h_t}(\mu_i+\Delta_i w_{ti})\big)^2.\end{aligned} \tag{2.21}$$

In order to obtain ML estimates in models with missing variables, the EM algorithm proposed by (Dempster *et al.*, 1977) can be used. In addition, the EM has several suitable properties including ease of implementation as well as monotone convergence. However, if the M-step of the EM algorithm does not have a closed form expression, the algorithm loses some of its attraction. The expectation conditional maximization (ECM) and expectation conditional maximization either (ECME) algorithms proposed by (Meng and Rubin, 1993), (Liu and Rubin, 1994), respectively, are simple modification of EM in which the maximization-step is replaced by a sequence of computationally convenient conditional maximization (CM)-steps.

Here, to obtain ML estimates of unknown parameters of the FM_m-SMSN-GARCH(1,1) model, the ECME algorithm is used as an extention of ECM algorithm. The propose ECME algorithm has some desirable advantages that not only inherits the mentioned features of the ECM algorithm but it might also be faster than ECM. In fact, some CM-steps of the ECM algorithm are replaced by the CML-steps in the ECME algorithm. which maximize the corresponding constrained log-likelihood function.

Conditional expectation used in the ECME algorithm is $Q\big(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(k)}\big) = E_{\boldsymbol{\Theta}^{(k)}}\big(\ell_{\boldsymbol{\Omega}}(\boldsymbol{\Theta})|\boldsymbol{y}\big)$, where $\widehat{\boldsymbol{\Theta}}^{(k)}$ is the estimated value of $\boldsymbol{\Theta}$ in the $k-th$ step of the algorithm as follows:

In order to construct this function, we must first calculate conditional expectations, $\hat{Z}_{ti}^{(k)}$, $\hat{h}_t^{(k)}$, and $\widehat{h_t Z_{ti}}^{(k)}$, so given the standard properties of the conditional expectations, we have

$$\begin{aligned}\hat{Z}_{ti}^{(k)} &= E_{\boldsymbol{\Theta}^{(k)}}\big(Z_{ti}|y\big) = \frac{\pi_i^{(k)} g\big(y_t|\mathcal{F}_{t-1};\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{(\boldsymbol{k})},\hat{\boldsymbol{\nu}}_{\boldsymbol{i}}^{(\boldsymbol{k})}\big)}{\sum_{i=1}^{m}\pi_i^{(k)} g\big(y_t|\mathcal{F}_{t-1};\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{(\boldsymbol{k})},\hat{\boldsymbol{\nu}}_{\boldsymbol{i}}^{(\boldsymbol{k})}\big)}, \\ \hat{h}_t^{(k)} &= E_{\boldsymbol{\Theta}^{(k)}}\big(h_t|y\big) = \mu_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1},\end{aligned} \tag{2.22}$$

and

$$\widehat{h_t Z_{ti}}^{(k)} = E_{\boldsymbol{\Theta}^{(k)}}\big(h_t Z_{ti}|y\big) = \mu_0\hat{Z}_{ti}^{(k)} + \alpha_1 y_{t-1}^2\hat{Z}_{ti}^{(k)} + \beta_1 h_{t-1}\hat{Z}_{ti}^{(k)},$$

where $g\big(y_t|\mathcal{F}_{t-1};\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{(\boldsymbol{k})},\hat{\boldsymbol{\nu}}_{\boldsymbol{i}}^{(\boldsymbol{k})}\big) = \int_0^\infty \phi_{SN}\Big(\frac{y_t-\omega}{\sqrt{h_t}};\mu_i,u_{ti}^{-1}\sigma_i^2,\lambda_i\Big)dH(u_{ti};\nu_i)$, represents the $i-th$ component of the $SMSN$ distribution in the mixture-$GARCH$ model. In addition, considering the posterior moments in proposition 2.5 in Appendix B.1, we obtain the following conditional expectations

$$E_{\boldsymbol{\Theta}^{(k)}}\left(U_{ti}|y\right) = \left(\hat{U}\right)_{ti}^{(k)} = \hat{\omega}_{1ti}^{(k)}, \tag{2.23}$$

$$E_{\boldsymbol{\Theta}^{(k)}}\left(W_{ti}U_{ti}|y\right) = \left(\widehat{UW}\right)_{ti}^{(k)} = \hat{U}_{ti}^{(k)}\hat{m}_{ti}^{(k)} + \hat{M}_i^{(k)}\hat{\eta}_{ti}^{(k)}, \tag{2.24}$$

$$E_{\boldsymbol{\Theta}^{(k)}}\left(W_{ti}^2U_{ti}|y\right) = \left(\widehat{UW^2}\right)_{ti}^{(k)} = \hat{M}_i^{2(k)} + \hat{U}_{ti}^{(k)}\hat{m}_{ti}^{2(k)} + \hat{M}_i^{(k)}\hat{\eta}_{ti}^{(k)}\hat{m}_{ti}^{(k)}, \tag{2.25}$$

for $i = 1, ..., m$ and $t = 1, ..., T$, where $\omega_{1ti}^{(k)}$ and $\hat{\eta}_{ti}^{(k)}$, are obtained using steps (i) and (ii) of proposition 2.5 in Appendix B.1, with $\hat{M}_i^{2(k)} = \frac{\hat{\varrho}_i{}^{2(k)}}{\hat{\varrho}_i{}^{2(k)}+\hat{\Delta}_i{}^{2(k)}}$, and $\hat{m}_i^{(k)} = \frac{\hat{\Delta}_i{}^{(k)}}{\sqrt{\hat{h}_t^{(k)}\left(\hat{\varrho}_i{}^{2(k)}+\hat{\Delta}_i{}^{2(k)}\right)}}\left(y_t - \omega - \sqrt{h_t}\hat{\mu}_i^{(k)}\right) = \frac{\hat{\Delta}_i{}^{(k)}}{\hat{\varrho}_i{}^{2(k)}+\hat{\Delta}_i{}^{2(k)}}\left(\frac{y_t-\omega}{\sqrt{h_t}} - \hat{\mu}_i^{(k)}\right)$.

Note that all of these quantities should be evaluated at $\hat{\boldsymbol{\Theta}}^k = \boldsymbol{\Theta}$, where the estimated value of $\boldsymbol{\Theta}$ in the k-th step of the algorithm is denoted by $\hat{\boldsymbol{\Theta}}^k$. Conditional expectations $\hat{\omega}_{1ti}^{(k)}$ and $\hat{\eta}_{ti}^{(k)}$, and all of the conditional expectations in equations (2.23)-(2.25), have a closed-form expression, for the Skew-t (ST) and Skew contaminated-normal (SCN) of the SMSN family; see more details in (Basso *et al.*, 2010), (Lachos *et al.*, 2010b), and (Lachos *et al.*, 2010a). However, when working with the skew-slash (SSL) distribution, there is not a closed-form expression, so in this case, a Monte Carlo integration, together with the ECME algorithm (MC-ECME) can be applied to approximate the integrals; see (Wei and Tanner, 1990) and (McLachlan and Krishnan, 2007). In addition, since Z_{ti} and (U_{it}, W_{ti}) are independent, it follows that $\left(\widehat{UZ}\right)_{ti}^{(k)} = \left(\hat{Z}\right)_{ti}^{(k)}\left(\widehat{U}\right)_{ti}^{(k)}$, $\left(\widehat{UZW}\right)_{ti}^{(k)} = \left(\hat{Z}\right)_{ti}^{(k)}\left(\widehat{UW}\right)_{ti}^{(k)}$, and $\left(\widehat{UZW^2}\right)_{ti}^{(k)} = \left(\hat{Z}\right)_{ti}^{(k)}\left(\widehat{UW^2}\right)_{ti}^{(k)}$.

The proposed ECME algorithm for ML estimation of the FM_m-SMSN-GARCH(1,1) model proceed as follows:

$\mathbf{E-step}$: In E-step of the algorithm, we first calculate $E_{\boldsymbol{\Theta}^{(k)}}\left(\ell_{\boldsymbol{\Omega}}(\boldsymbol{\Theta})|\boldsymbol{y}\right)$ according to the conditional expectations of equations (2.22)-(2.25), so in the E-step of the algorithm, we have

$$
\begin{aligned}
Q\big(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)}\big) &= \sum_{i=1}^{m}\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}E\big(\log h(u_{ti};\boldsymbol{\nu}_i)|y\big)+\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}\log\pi_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}\log\varrho_i^2\\
&-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\big(\widehat{Z_{ti}h_t}\big)^{(k)}-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZU\big)}_{ti}^{(k)}}{h_t\varrho_i^2}\big(y_t-\omega-\sqrt{h_t}\mu_i\big)^2\\
&+\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZUW\big)}_{ti}^{(k)}}{h_t\varrho_i^2}\big(y_t-\omega-\sqrt{h_t}\mu_i\big)\sqrt{h_t}\Delta_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZUW^2\big)}_{ti}^{(k)}}{h_t\varrho_i^2}h_t{\Delta_i}^2\\
&=\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}E\big(\log h(u_{ti};\boldsymbol{\nu}_i)|y\big)+\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}\log\pi_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}\log\varrho_i^2\\
&-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\big(\widehat{Z_{ti}h_t}\big)^{(k)}-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZU\big)}_{ti}^{(k)}}{h_t\varrho_i^2}\big(y_t-\omega\big)^2+\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZU\big)}_{ti}^{(k)}}{\sqrt{h_t}\varrho_i^2}\big(y_t-\omega\big)\mu_i\\
&-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZU\big)}_{ti}^{(k)}}{\varrho_i^2}{\mu_i}^2+\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZUW\big)}_{ti}^{(k)}}{\sqrt{h_t}\varrho_i^2}\big(y_t-\omega\big)\Delta_i-\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZUW\big)}_{ti}^{(k)}}{\varrho_i^2}\big(\mu_i\Delta_i\big)\\
&-\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{\big(ZUW^2\big)}_{ti}^{(k)}}{\varrho_i^2}{\Delta_i}^2.
\end{aligned}
\tag{2.26}
$$

CM – steps

Updating the parameters in the CM-steps is done in the following sub-steps:

- i) Update π_i by maximizing (2.26) with respect to π_i which gives

$$\hat{\pi}_i^{(k+1)}=\frac{\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}}{n},$$

- ii) Update μ_i by maximizing (2.26) with respect to μ_i which gives

$$\hat{\mu}_i^{(k+1)}=\frac{\sum_{t=1}^{T}\frac{\widehat{\big(ZU\big)}_{ti}^{(k)}}{{\hat{\varrho}_i}^{2(k)}}\Big(\big(y_t-\omega\big)/\sqrt{\hat{h}_t}\Big)-\sum_{t=1}^{T}\frac{\widehat{\big(ZUW\big)}_{ti}^{(k)}\hat{\Delta_i}^{(k)}}{{\hat{\varrho}_i}^{2(k)}}}{\sum_{t=1}^{T}\frac{\widehat{\big(ZU\big)}_{ti}^{(k)}}{{\hat{\varrho}_i}^{2(k)}}},$$

- iii) Update ${\varrho_i}^2$ by maximizing (2.26) with respect to ϱ_i which gives

$$\hat{\varrho}_i^{2(k+1)} = \frac{\frac{1}{2}\sum_{t=1}^{T}\left(\widehat{(ZU)}_{ti}^{(k)}\left(\frac{(y_t-\omega)}{\sqrt{\hat{h}_t}} - \hat{\mu}_i^{(k)}\right)^2\right) - \sum_{t=1}^{T}\left(\widehat{(ZUW)}_{ti}^{(k)}\left(\frac{(y_t-\omega)}{\sqrt{\hat{h}_t}} - \hat{\mu}_i^{(k)}\right)\hat{\Delta}_i^{(k)}\right)}{\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}},$$

$$+\frac{\frac{1}{2}\sum_{t=1}^{T}\left(\widehat{(ZUW^2)}_{ti}^{(k)}\hat{\Delta}_i^{2(k)}\right)}{\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}},$$

- iv) Update Δ_i by maximizing (2.26) with respect to Δ_i which gives

$$\hat{\Delta}_i^{(k+1)} = \frac{\sum_{t=1}^{T}\frac{\widehat{(ZUW^2)}_{ti}^{(k)}}{\hat{\varrho_i}^{2(k)}}\left(\frac{(y_t-\omega)}{\sqrt{\hat{h}_t}} - \hat{\mu}_i^{(k)}\right)}{\frac{1}{2}\sum_{t=1}^{T}\frac{\widehat{(ZUW^2)}_{ti}^{(k)}}{\hat{\varrho_i}^{2(k)}}},$$

- v) Update ω by maximizing (2.26) with respect to ω which gives

$$\hat{\omega}^{(k+1)} = \frac{\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{(ZU)}_{ti}^{(k)}}{\sqrt{\hat{h}_t}\hat{\varrho_i}^{2(k)}}\left(\frac{y_t}{\sqrt{\hat{h}_t}} - \hat{\mu}_i^{(k)}\right) - \sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{(ZUW)}_{ti}^{(k)}}{\sqrt{\hat{h}_t}\hat{\varrho_i}^{2(k)}}\hat{\Delta}_i^{(k)}}{\sum_{i=1}^{m}\sum_{t=1}^{T}\frac{\widehat{(ZU)}_{ti}^{(k)}}{\hat{h}_t\hat{\varrho_i}^{2(k)}}},$$

- vi) In the last step (CML-step), update $\boldsymbol{\nu_i}$ by maximizing (2.26) with respect to $\boldsymbol{\nu_i}$ which gives

$$\hat{\boldsymbol{\nu}}_i^{(k+1)} = \underset{\boldsymbol{\nu}_i}{\operatorname{argmax}}\left(\sum_{t=1}^{T}\log\left(\hat{\pi}_i^{(k+1)}\mathbf{g}\left(y_t|\mathcal{F}_{t-1};\hat{\boldsymbol{\theta}}_i^{(k)},\boldsymbol{\nu}_i\right)\right)\right).$$

  More details on the CML-step are sketched in Appendix A.2. According to (Lin *et al.*, 2014) and (Zeller *et al.*, 2019), assuming the identical mixing component, i.e. $\nu_1 = \nu_2 = ... = \nu_m = \nu$, a more parsimonious model can be obtained. This setting changes the problem of nontrivial high-dimension optimization into the more simple one/two dimensional search.

- For ease of estimation, we re-parametrized the parameters $(\mu_0, \alpha_1, \beta_1)$ of the model. Since $(\mu_0, \alpha_1, \beta_1)$ must be greater than zero and $(\alpha_1 + \beta_1 < 1)$, they are parametrized as follows [2]:

$$\mu_0 = \exp(-\eta_1),$$
$$\alpha_1 = \frac{1}{1+\exp(-\eta_2)},$$

and

$$\beta_1 = \frac{1-\alpha_1}{1+\exp(-\eta_3)}.$$

The E-, CM- and CML-steps of the ECME algorithm are repeatedly alternated until a suitable convergence rule is satisfied, e.g. $\left|\ell\big(\widehat{\mathbf{\Theta}}^{(k+1)}, \boldsymbol{y}\big)/\ell\big(\widehat{\mathbf{\Theta}}^{(k)}, \boldsymbol{y}\big) - 1\right| \leq \textit{tolerance}.$ We also note that different value of tolerance can be selected. In the present paper we set the tolerance to 10^{-5}. To provide a comprehensive implementation guide, Algorithm 2 presents a summarized pseudocode of the aforementioned ECME algorithm.

[2]The reparametrization here reported is specific to the GARCH(1,1) model. Alternative designs must be chosen when specifying different conditional variance models.

Algorithm 1 Implementation procedure of the ECME algorithm for fitting the FM_m-SMSN-GARCH(1,1)

procedure *FM_m-SMSN-GARCH(1,1)*

  inputs: $\{y_t\}_{t=1}^{T}$- the set of input data; **m** - the number of components;
  $\epsilon = 10^{-5}$ - the prespecified tolerance.
  initialize: Obtain the starting value $\widehat{\boldsymbol{\Theta}}^{(0)}$.
  Compute the initial log-likelihood as $\ell\big(\widehat{\boldsymbol{\Theta}}^{(0)}, \boldsymbol{y}\big)$.
  Set "Convergence = False".
  for $k = 0$ **do**
    for ***t=1*** *to* ***T*** **do**
      for ***i=1*** *to* ***m*** **do**
        Obtain latent and missing information $\hat{Z}_{ti}^{(k)}$, $\widehat{h_t Z_{ti}}^{(k)}$, $\big(\widehat{UZ}\big)_{ti}^{(k)}$, $\big(\widehat{UZW}\big)_{ti}^{(k)}$, and $\big(\widehat{UZW^2}\big)_{ti}^{(k)}$.
      end
    end
    for ***i=1*** *to* ***m*** **do**
      Update the parameters $\hat{\pi}_i^{(k+1)}, \hat{\mu}_i^{(k+1)}, \hat{\varrho}_i^{2(k+1)}, \hat{\Delta}_i^{(k+1)}, \hat{\omega}^{(k+1)}$, and $\hat{\boldsymbol{\nu}}_{\boldsymbol{i}}^{(k+1)}$.
      Set $\widehat{\boldsymbol{\Theta}}^{(k+1)} = \{\widehat{\boldsymbol{\varphi}}^{(k+1)}, \widehat{\boldsymbol{\pi}}^{(k+1)}, \widehat{\boldsymbol{\theta}}^{(k+1)}, \widehat{\boldsymbol{\nu}}^{(k+1)}\}$.
    end
    Compute $\ell\big(\widehat{\boldsymbol{\Theta}}^{(k+1)}, \boldsymbol{y}\big)$ and set "Convergence = True" if $\Big|\ell\big(\widehat{\boldsymbol{\Theta}}^{(k+1)}, \boldsymbol{y}\big)/\ell\big(\widehat{\boldsymbol{\Theta}}^{(k)}, \boldsymbol{y}\big) - 1\Big| \leq \epsilon$.
    if *Convergence = True* **then**
      break
    end
  end
  return $\ell\big(\widehat{\boldsymbol{\Theta}}^{(k+1)}, \boldsymbol{y}\big)$ and $\Theta^{(k+1)}$.

end procedure

2.4.3 The observed information matrix

To compute the standard errors of maximum likelihood (ML) parameter estimates for the FM_m-GARCH model, we provide a systematic procedure. Our objective is to approximate the asymptotic covariance matrix of the ML estimates within the proposed mixture-GARCH model. To achieve this, we derive the observed information matrix denoted as $I(\boldsymbol{\Theta}|\boldsymbol{y}) = -\partial^2 \ell(\boldsymbol{\Theta}|\boldsymbol{y})/\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}'$, which captures the second-order partial derivatives of the log-likelihood function (2.17) with respect to the model parameters.

The information matrix is comprised of individual blocks, where each block is represented as $I_{(\boldsymbol{\gamma},\boldsymbol{\delta})} = \sum_{t=1}^{T} -\frac{\partial^2 \ell_t(\boldsymbol{\Theta}|\boldsymbol{y})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\delta}'}$. Here, $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ represents the parameter set $\boldsymbol{\varphi}, \pi_j, \mu_j, \sigma_j^2, \lambda_j$, and $\boldsymbol{\nu_j}$, with $j = 1, ..., m$, and $\boldsymbol{\varphi} = (\alpha_1, \beta_1)$.

Considering the maximum likelihood function of the FM$_m$-SMSN-GARCH(1,1) model in equation (2.17), we can express the elements of the information matrix as follows:

$$I_{(\gamma,\boldsymbol{\delta})} = \sum_{t=1}^{T}\left[\sum_{i=1}^{m}\frac{\partial^2 \log \pi_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\delta}'} - \frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2 \log h_t\sigma_i^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\delta}'} - \sum_{i=1}^{m}\frac{1}{M_{ti}^2}\frac{\partial M_{ti}}{\partial\boldsymbol{\gamma}}\frac{\partial M_{ti}}{\partial\boldsymbol{\delta}'} + \sum_{i=1}^{m}\frac{1}{M_{ti}}\frac{\partial^2 M_{ti}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\delta}'}\right],$$

Where

$$\frac{\partial M_{ti}}{\partial\boldsymbol{\gamma}} = I_{ti}^{\phi}(1)\frac{\partial B_{ti}}{\partial\boldsymbol{\gamma}} - \frac{1}{2}I_{ti}^{\Phi}\left(\frac{5}{2}\right)\frac{\partial S_{ti}}{\partial\boldsymbol{\gamma}},$$

and

$$\begin{aligned}\frac{\partial^2 M_{ti}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\delta}'} = {} & \frac{1}{4}I_{ti}^{\Phi}\left(\frac{5}{2}\right)\frac{\partial S_{ti}}{\partial\boldsymbol{\gamma}}\frac{\partial S_{ti}}{\partial\boldsymbol{\delta}'} - \frac{1}{2}I_{ti}^{\Phi}\left(\frac{3}{2}\right)\frac{\partial^2 S_{ti}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\delta}'} \\ & - \frac{1}{2}I_{ti}^{\phi}(2)\left(\frac{\partial B_{ti}}{\partial\boldsymbol{\gamma}}\frac{\partial S_{ti}}{\partial\boldsymbol{\delta}'} + \frac{\partial S_{ti}}{\partial\boldsymbol{\gamma}}\frac{\partial B_{ti}}{\partial\boldsymbol{\delta}'}\right) \\ & - I_{ti}^{\phi}(2)\frac{\partial B_{ti}}{\partial\boldsymbol{\gamma}}\frac{\partial S_{ti}}{\partial\boldsymbol{\delta}'} + I_{ti}^{\phi}(1)\frac{\partial^2 B_{ti}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\delta}'},\end{aligned}$$

with

$$I_{ti}^{\Phi}(\rho) = E_U\left(U^{\rho}\exp\left(-US_{ti}/2\right)\Phi(\sqrt{U}B_{ti})\right),$$

$$I_{ti}^{\phi}(\rho) = E_U\left(U^{\rho}\exp\left(-U(S_{ti}+B_{ti}^2)/2\right)\right),$$

$$M_{ti} = E_U\left(\sqrt{U}\exp\left(-US_{ti}/2\right)\Phi(\sqrt{U}B_{ti})\right),$$

$$S_{ti} = \frac{\left(y_t - \omega - \mu_i\sqrt{h_t}\right)^2}{h_t\sigma_i^2},$$

and

$$B_{ti} = \frac{\lambda_i\left(y_t - \omega - \mu_i\sqrt{h_t}\right)}{\sqrt{h_t\sigma_i^2}}.$$

In the following proposition, we provide the explicit expressions for $I_{ti}^{\Phi}(\rho)$ and $I_{ti}^{\phi}(\rho)$ for specific members of the FM$_m$-SMSN-GARCH(1,1) family. These expressions are derived to facilitate the computation of the information matrix and play a crucial role in the estimation process. For the detailed proofs of these expressions, we refer the interested readers to Appendix B.2.

Proposition 2.4. *(a) In the case of the FM_m-ST-GARCH(1,1) distribution (ST-GARCH):*

$$I_{ti}^{\Phi}(\rho) = \frac{2^{\rho+\frac{3}{2}}\Gamma\big(\frac{\nu+1}{2}\big)\big(\nu/2\big)^{\frac{\nu}{2}}\Gamma\big(\rho+\frac{\nu}{2}\big)(\nu+S_{ti})^{-(\rho-\frac{1}{2})}}{\Gamma\big(\frac{\nu}{2}\big)(\nu+S_{ti})^{\frac{\nu+1}{2}}\Gamma\big(\frac{\nu}{2}\big)} T\left(\frac{B_{ti}\sqrt{\nu+2\rho}}{\big(\nu+S_{ti}\big)^{\frac{1}{2}}};\nu+2\rho\right),$$

$$I_{ti}^{\phi}(\rho) = \frac{2^{(\rho+\frac{\nu}{2})}\frac{\nu}{2}^{\frac{\nu}{2}}\Gamma\big(\rho+\frac{\nu}{2}\big)}{\Gamma\big(\frac{\nu}{2}\big)}\left(\frac{1}{\nu+B_{ti}^2+S_{ti}}\right)^{(\rho+\frac{\nu}{2})},$$

and

$$\frac{\partial \ell_t(\boldsymbol{\Theta};\boldsymbol{y})}{\partial \nu} = \frac{1}{\sqrt{2\pi}\sigma_i}\Bigg(1+\log(\frac{\nu}{2})+DIG(\nu/2)I_{ti}^{\Phi}(1/2)-I_{ti}^{\Phi}(3/2)$$

$$+\int_0^{\infty} u^{\frac{1}{2}}\log(u)exp\big(-uS_{ti}/2\big)\Phi(\sqrt{u}B_{ti})dH(u|\nu)\Bigg),$$

where DIG denotes the digamma function.

(b) In the case of the FM_m-SSL-GARCH(1,1) distribution (SSL-GARCH):

$$I_{ti}^{\Phi}(\rho) = \frac{2^{2+\nu}\Gamma(\rho+\nu)}{S_{ti}^{\rho+\nu}}P\big(1;\rho+\nu,(\frac{S_{ti}}{2})\big),E\big(\Phi\big(\sqrt{Ga_{ti}}\big)B_{ti}\big),$$

$$I_{ti}^{\phi}(\rho) = \frac{\nu 2^{\rho+\nu}\Gamma(\rho+\nu)}{\sqrt{2\pi}\big(B_{ti}^2+S_{ti}\big)^{\rho+\nu}}P\big(1;\rho+\nu,(\frac{B_{ti}^2+S_{ti}}{2})\big),$$

where $Ga \sim Gamma\big(\rho+\nu,\frac{S_{ti}}{2}\big)I_{(0,1)}$, and $P\big(y;c,d\big)$ shows the Gamma(c,d)-distribution function evaluated at y.

and

$$\frac{\partial \ell_t(\boldsymbol{\Theta};\boldsymbol{y})}{\partial \nu} = 2\int_0^1 u^{\nu-1}(1+\nu\log u)\phi\Big(\frac{y_t-\omega}{\sqrt{h_t}};0,u^{-1}\sigma_i^2\Big)\Phi(\sqrt{u}B_{ti})du,$$

(c) In the case of the FM_m-SCN-GARCH(1,1) distribution (SCN-GARCH):

$$I_{ti}^{\Phi}(\rho) = \sqrt{2\pi}\Big(\nu\gamma^{\rho-/2}\phi\big(\sqrt{S_{ti}};0,\gamma^{-1}\big)\Phi(\sqrt{\gamma}B_{ti})+(1-\nu)\phi(\sqrt{S_{ti}};0,1)\Phi(B_{ti})\Big),$$

$$I_{ti}^{\phi}(\rho) = \nu\gamma^{\rho-1/2}\phi\big(\sqrt{B_{ti}^2 + S_{ti}}; 0, \gamma^{-1}\big) + (1-\nu)\phi\big(\sqrt{B_{ti}^2 + S_{ti}}; 0, 1\big),$$

and

$$\frac{\partial \ell_t(\boldsymbol{\Theta}; \boldsymbol{y})}{\partial \nu} = 2\Bigg(\phi\Big(\frac{y_t - \omega}{\sqrt{h_t}}; 0, \gamma^{-1}\sigma_i^2\Big)\Phi(\sqrt{\gamma}B_{ti}) - \phi\Big(\frac{y_t - \omega}{\sqrt{h_t}}; 0, \sigma_i^2\Big)\Phi(B_{ti})\Bigg),$$

$$\frac{\partial \ell_t(\boldsymbol{\Theta}; \boldsymbol{y})}{\partial \gamma} = \frac{\nu}{\sqrt{2\pi\sigma_i^2}}\sqrt{\gamma}\exp\left(-\gamma S_{ti}/2\right)\Big(\gamma^{-1}\Phi(\sqrt{\gamma}B_{ti}) + \phi(B_{ti}/\sqrt{\gamma})B_{ti}/\sqrt{\gamma} - \Phi(\sqrt{_{ti}})S_{ti}\Big).$$

As a result, the standard errors for a specific parameter estimate, denoted as $\hat{\theta}_r$, can be approximated by

$$S.E(\hat{\theta}_r) \approx \sqrt{[\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\Theta}}|\boldsymbol{y})]_{rr}}$$

where $[\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\Theta}}|\boldsymbol{y})]_{rr}$ is the (r,r)-th entry of the inverse of 2.4.3.

2.4.4 Model selection criteria and performance assessment

One of the key considerations in model selection is determining the most suitable model among a set of competing specifications. In our study, we have taken into account three well-established criteria: the Akaike Information Criterion (AIC) (Akaike, 1974), the Bayesian Information Criterion (BIC) (Schwarz, 1978) and the Efficient Determination Criterion (EDC) (Bai *et al.*, 1989). Using multiple model selection criteria, such as AIC, BIC, and EDC, provides a more comprehensive evaluation of competing models. Each criterion offers unique insights and focuses on different aspects of model selection, ensuring a robust and well-informed choice. The criteria are defined as follows:

$$AIC = -2\ell\big(\widehat{\boldsymbol{\Theta}}, \boldsymbol{y}\big) + 2D, \quad BIC = -2\ell\big(\widehat{\boldsymbol{\Theta}}, \boldsymbol{y}\big) + \log(n)D, \quad EDC = -2\ell\big(\widehat{\boldsymbol{\Theta}}, \boldsymbol{y}\big) + R_n D,$$

where $\ell\big(\widehat{\boldsymbol{\Theta}}, \boldsymbol{y}\big)$ is the maximized log-likelihood, D is the number of estimated parameters in the proposed mixture GARCH model and n is the sample size of the GARCH sample. For these three model selection criteria we state that $\pi_m = 1 - \sum_{i=1}^{m-1}\pi_i$, so it is obvious that the number of components is $m-1$. Furthermore, note that in the EDC criteria, R_n must be chosen so that $\frac{R_n}{n}$ and $\frac{R_n}{(loglogn)}$ go to zero while n goes to infinity. Therefore, we chose $R_n = 0.2\sqrt{n}$ coherently with (Bai *et al.*, 1989). The lower values of the AIC, BIC and EDC criteria indicate the best selection of the model.

2.5 Simulation study and a real data example

In this section, the performance and flexibility of the proposed FM_m-SMSN-GARCH(1,1) are evaluated using simulated and real data example. The implementations of the algorithms are based on the R software (Team *et al.*, 2013) version 4.2.0 with a core i7 760 processor 2.8 GHz, and a relative tolerance of 10^{-5} is used for convergence of the ECME-algorithms.

2.5.1 Simulations

In this part, we provide two parts of simulations. In the first part, we show the consistency properties of the proposed model and estimation methods considering different sample sizes. Moreover, to show the satisfaction of the proposed estimations, some simulations for FM_m-SMSN-GARCH(1,1) parameters recovery by simulating from them and estimating the proposed ML estimates are provided. Finally, in the second part of simulations, to show the performances (robustness, misspecification and model identification) of our models to model the data with unknown structure we generate asymmetry and heavy-tailed distributions with various components that belong to the class of scale mixtures of skew-normal (SMSN) distributions.

2.5.1.1 Model specifications

In our simulations, we investigate two distinct scenarios to analyze the performance of our model. These scenarios are as follows:

- Moderately Separated Model:

  In the first scenario, we consider a data generating process that comprises two components exhibiting clear separation. Appendix A.2 presents Table A.2, which reports the true parameters satisfying the constraints outlined in Proposition 2.3. Additionally, Figure A.1 provides an illustrative example of a simulated series that demonstrates the evident distinction between the two components.

- Weakly Separated Model:

  Moving to a more intricate setting, the second scenario involves a data generating process characterized by two components that are closely situated. Table A.3 in Appendix A.3 presents the corresponding parameter values, while Figure A.2 showcases a representative example. It is worth noting that, upon a cursory visual inspection, the two components are not easily distinguishable. However, discernible attributes such as asymmetry and skewness are noticeably present.

These contrasting scenarios enable us to evaluate the effectiveness of our model under varying degrees of component separation, thereby providing valuable insights into its robustness and performance characteristics.

2.5.1.2 Part 1: Parameter Retrieval

The first simulation study aims to assess the performance of the ML estimates obtained using the Expectation Conditional Maximization (ECME) algorithm, as well as their corresponding standard errors through the information-based method outlined in Section 3.4.

To evaluate the accuracy of the estimations, we compute the mean bias and mean-squared error (MSE) using the following formulas:

$$Bias = \frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_i - \theta), \quad MSE = \frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_i - \theta)^2,$$

where $\hat{\theta}_i$ represents the estimate of θ for the i-th sample. In each scenario, we generate samples of varying sizes: 250, 500, and 1000 observations, and repeat the experiment 500 times. The Bias and MSE values for the first scenario (moderate separated components) are presented in Table 2.1, while the results for the second scenario (weakly separated components) can be found in Table 2.2. Notably, in both scenarios, the numerical results consistently indicate that both Bias and MSE tend to converge to zero as the sample size increases. Moreover, Figures A.3 to A.10 in Appendix A.6 visually demonstrate the diminishing magnitudes of Bias, MSE, Mean, and SD of ML estimators based on the different members of the proposed model as the sample size, denoted by n, increases. This empirical evidence highlights the consistency of the ML estimators. These findings affirm that the proposed ML estimates of FM_m-SMSN-GARCH(1,1), obtained using the ECME algorithm, exhibit strong consistency properties. Furthermore, Tables A.4 and A.5 in Appendix A.6 illustrate the convergence of the mean and standard deviations (SD) of the estimated parameters toward their true values, with the standard deviations decreasing accordingly. These results collectively demonstrate the high accuracy of the point estimates in the considered scenarios.

Table 2.1: Bias and MSE of ML estimates of medium-FM$_2$-SN–GARCH(1,1), medium-FM$_2$-ST–GARCH(1,1), medium-FM$_2$-SSL–GARCH(1,1) and medium-FM$_2$-SCN–GARCH(1,1) models with various sample sizes n = 250, 500, and 1000.

Models		FM$_2$-SN–GARCH(1,1)			FM$_2$-ST–GARCH(1,1)		
Measure	Parameters	250	500	1000	250	500	1000
	μ_1	-0.089	-0.071	-0.044	-0.199	-0.172	-0.109
	μ_2	-0.243	-0.171	-0.111	-0.038	-0.029	0.017
	σ_1^2	0.472	0.471	0.458	-0.047	-0.028	-0.020
	σ_2^2	0.070	0.070	0.060	-0.412	-0.222	-0.192
	λ_1	0.321	0.209	0.205	-0.214	-0.043	0.064
	λ_2	-0.474	-0.118	-0.117	-0.108	-0.075	0.038
Bias	ν	–	–	–	-0.328	0.216	0.150
	γ	–	–	–	–	–	–
	π	2.9E-03	5.0E-04	2.0E-04	9.0E-04	4.0E-04	2.0E-04
	ω	0.125	0.008	0.001	-0.053	0.007	-0.004
	μ_0	7.7E-05	4.9E-05	5.0E-05	4.9E-05	5.0E-05	4.8E-05
	α_1	0.011	0.006	0.005	0.018	0.015	0.014
	β_1	-0.021	-0.009	-0.006	-0.055	-0.029	-0.015
	μ_1	0.165	0.051	0.030	0.098	0.042	0.017
	μ_2	0.181	0.098	0.039	0.113	0.032	0.013
	σ_1^2	0.732	0.437	0.338	0.019	0.005	0.003
	σ_2^2	0.084	0.024	0.018	0.518	0.255	0.137
	λ_1	0.595	0.219	0.140	0.590	0.184	0.094
	λ_2	5.527	1.785	0.681	2.0006	0.778	0.286
MSE	ν	–	–	–	2.7200	1.7158	0.9782
	γ	–	–	–	–	–	–
	π	2.9E-03	5.0E-04	2.0E-04	9.0E-04	4.0E-04	2.0E-04
	ω	0.096	0.069	0.061	0.079	0.065	0.062
	μ_0	0.000	0.000	0.000	0.000	0.000	0.000
	α_1	2.279E-03	6.9E-05	5.1E-05	4.06E-04	2.73E-04	2.08E-04
	β_1	9.757E-03	7.55E-04	3.70E-04	5.201E-03	1.788E-03	6.54E-04

Models		FM$_2$-SSL–GARCH(1,1)			FM$_2$-SCN–GARCH(1,1)		
Measure	Parameters	250	500	1000	250	500	1000
	μ_1	0.099	0.044	0.017	-0.048	-0.012	0.002
	μ_2	-0.256	-0.140	-0.100	-0.127	-0.109	-0.079
	σ_1^2	0.387	0.298	0.237	0.302	0.252	0.224
	σ_2^2	0.176	0.152	0.060	0.093	0.068	0.062
	λ_1	0.100	0.043	0.032	0.306	0.238	0.179
	λ_2	0.377	-0.117	-0.054	-0.295	-0.149	-0.116
Bias	ν	0.434	0.019	0.012	0.051	0.023	0.0001
	γ	–	–	–	0.102	0.042	0.031
	π	5.7E-03	-4.0E-04	4.0E-04	5.1E-03	1.1E-03	-1.0E-04
	ω	0.149	0.071	-0.001	0.149	0.006	-0.0001
	μ_0	6.6E-05	5.0E-05	5.4E-05	5.8E-05	5.2E-05	5.1E-05
	α_1	0.007	0.005	0.005	0.010	0.007	0.007
	β_1	-0.016	-0.007	-0.007	-0.022	-0.009	-0.007
	μ_1	0.117	0.054	0.027	0.108	0.049	0.018
	μ_2	0.128	0.031	0.015	0.061	0.026	0.012
	σ_1^2	0.374	0.237	0.136	0.381	0.211	0.121
	σ_2^2	0.086	0.049	0.030	0.068	0.032	0.018
	λ_1	0.379	0.219	0.095	0.778	0.290	0.131
	λ_2	0.222	0.631	0.173	0.341	0.761	0.356
MSE	ν	0.910	0.304	0.023	0.030	0.009	0.0003
	γ	–	–	–	0.079	0.029	0.004
	π	1.0E-03	4.0E-04	2.0E-04	9.0E-04	6.0E-04	3.0E-04
	ω	0.094	0.056	0.026	0.107	0.075	0.052
	μ_0	0.000	0.000	0.000	0.000	0.000	0.000
	α_1	1.64E-04	7.0E-05	4.9E-05	2.30E-04	1.01E-04	7.8E-05
	β_1	2.067E-03	8.33E-04	3.91E-04	2.301E-03	8.64E-04	4.64E-04

TABLE 2.2: Bias and MSE of ML estimates of weak-FM_2-SN–GARCH(1,1), weak-FM_2-ST–GARCH(1,1), weak-FM_2-SSL–GARCH(1,1) and weak-FM_2-SCN–GARCH(1,1) models with various sample sizes $n = 250$, 500, and 1000.

Models		FM_2-SN-GARCH(1,1)			FM_2-ST-GARCH(1,1)		
Measure	Parameters	250	500	1000	250	500	1000
	μ_1	-0.226	-0.205	-0.130	-0.61	-0.496	-0.348
	μ_2	0.283	0.245	0.136	-0.638	-0.544	-0.391
	σ_1^2	0.281	0.026	0.132	-0.210	-0.147	-0.079
	σ_2^2	1.166	0.796	0.625	-1.022	-0.807	-0.473
	λ_1	1.346	0.787	0.573	0.655	0.397	-0.144
	λ_2	-1.931	-1.660	-1.247	-0.615	-0.339	0.191
Bias	ν	–	–	–	1.326	0.422	0.181
		–	–	–	1.326	0.422	0.181
	π	0.072	0.056	0.044	0.060	0.047	0.026
	ω	1.94E-03	4.63E-04	3.03E-04	1.995E-03	3.84E-04	2.75E-04
	μ_0	9.19E-04	6.98E-05	4.22E-05	4.35E-04	8.71E-05	5.16E-05
	α_1	-0.015	-0.011	-0.007	-0.018	-0.012	0.003
	β_1	0.001	0.002	0.002	0.026	0.024	-0.004
	μ_1	0.079	0.068	0.042	0.523	0.386	0.151
	μ_2	0.158	0.122	0.062	0.806	0.423	0.250
	σ_1^2	1.021	0.252	0.258	0.093	0.082	0.043
	σ_2^2	4.763	2.600	1.328	1.584	1.289	0.689
	λ_1	2.743	0.980	0.535	2.002	0.813	0.280
	λ_2	5.371	3.931	1.966	2.715	1.636	0.692
MSE	ν	−−	−−	−−	3.505	4.1731	1.274
		–	–	–	3.505	4.173	1.274
	π	0.018	0.012	0.009	0.015	0.008	0.007
	ω	9.63E-05	5.13E-07	1.6E-07	1.43E-04	2.39E-07	1.32E-07
	μ_0	4.46E-05	9.59E-08	2.39E-09	1.6E-06	4.07E-08	4.52E-09
	α_1	3.58E-04	3.23E-04	1.77E-04	8.16E-04	3.01E-04	9.64E-05
	β_1	1.654E-03	1.205E-03	1.095E-03	7.88E-04	1.061E-03	1.85E-04

Models		FM_2-SSL-GARCH(1,1)			FM_2-SCN-GARCH(1,1)		
Measure	Parameters	250	500	1000	250	500	1000
	μ_1	0.060	0.024	0.009	0.208	0.175	0.092
	μ_2	0.079	-0.028	-0.018	0.189	0.150	0.092
	σ_1^2	0.099	0.044	0.019	0.120	0.086	0.067
	σ_2^2	-0.481	-0.312	-0.191	1.332	0.981	0.688
	λ_1	-0.293	-0.246	-0.208	1.399	0.946	0.752
	λ_2	0.306	0.264	0.223	-1.356	-0.985	-0.708
Bias	ν	1.251	-0.552	-0.379	0.153	0.115	0.090
		1.251	-0.552	-0.379	0.161	0.110	0.080
	π	0.100	0.077	0.039	-0.058	-0.041	0.002
	ω	0.001	0.0007	0.0003	0.151	0.039	0.004
	μ_0	1.45E-04	9.33E-05	5.08E-05	1.01E-04	8.34E-05	6.93E-05
	α_1	-0.013	-0.011	-0.008	-0.018	-0.009	-0.005
	β_1	0.006	0.011	0.008	0.011	0.005	-0.002
	μ_1	0.024	0.018	0.010	0.135	0.101	0.044
	μ_2	0.147	0.086	0.072	0.112	0.064	0.046
	σ_1^2	0.275	0.075	0.026	0.094	0.074	0.054
	σ_2^2	1.134	1.006	0.438	4.457	1.959	1.044
	λ_1	0.367	0.219	0.110	3.73	1.973	1.291
	λ_2	2.173	0.931	0.544	3.864	1.639	0.930
MSE	ν	3.201	2.042	0.779	0.032	0.014	0.010
		3.201	2.042	0.779	0.043	0.022	0.015
	π	0.017	0.014	0.007	0.006	0.003	0.004
	ω	1.58E-05	2.27E-06	2.38E-07	7.7247E-02	4.9E-03	3.14E-05
	μ_0	4.22E-08	2.84E-08	3.36E-09	3.63E-08	1.73E-08	9.12E-09
	α_1	4.68E-04	4.62E-04	1.69E-04	7.6E-04	3.41E-04	1.84E-04
	β_1	3.97E-04	1.03E-03	7.07E-04	1.78E-03	1.12E-03	1.12E-03

2.5.1.3 Part 2: Robustness, misspecification and model identification

This section focuses on examining the robustness, misspecification, and model identification aspects of the proposed methodology. We begin by simulating a sample of size 250 with 500 replications from a mixture GARCH model with FM_2-ST-component (moderate separated components), denoted as FM_2-ST-GARCH(1,1). Subsequently, we fit different members of this model family in three scenarios, each involving a varying number of components. The performance of these models is evaluated using various model selection criteria as discussed in Subsection 2.4.4.

The log-likelihood value, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Efficient Determination Criterion (EDC) for the fitted FM_m-SMSN-GARCH(1,1) models, with $m = 1, 2, 3$, are reported in Figures 2.3-2.5. Remarkably, as observed from these figures, all considered model selection criteria consistently favor the true model. They effectively identify the most appropriate model amidst the competition. For instance, in Figure 2.3, when considering the BIC criterion, a comparison of the BIC values for the FM_2-ST-GARCH(1,1) model with its counterparts in the family, namely FM_2-SN-GARCH(1,1), FM_2-SSL-GARCH(1,1), and FM_2-SCN-GARCH(1,1), clearly demonstrates that the true model, FM_2-ST-GARCH(1,1), is correctly selected as anticipated.

To reinforce the findings from the information criteria, we employ more robust statistical procedures, namely Analysis of Variance (ANOVA) and Tukey's Honestly Significant Difference (HSD) test. ANOVA yields a p-value significantly below 0.05 when comparing a sample of size 250 generated from FM_2-ST-GARCH(1,1) with its competitors, indicating significant differences in means between the groups. Table 2.3 presents the results of ANOVA for the different model selection criteria, namely log-likelihood, AIC, BIC, and EDC. When ANOVA yields a significant result, it signifies that at least one group differs from the others. However, it does not indicate which specific groups exhibit the differences. To identify these differences, we employ the Tukey's HSD test, which conducts pairwise comparisons. The results of the Tukey's HSD test are displayed in Table 2.4. Notably, the pairwise comparisons between FM_2-ST-GARCH(1,1) vs. FM_2-SN-GARCH(1,1), FM_2-SSL-GARCH(1,1) vs. FM_2-ST-GARCH(1,1), and FM_2-SCN-GARCH(1,1) vs. FM_2-ST-GARCH(1,1) all yield significant differences. Hence, the considerable power of all model selection criteria in determining the most appropriate model is convincingly established.

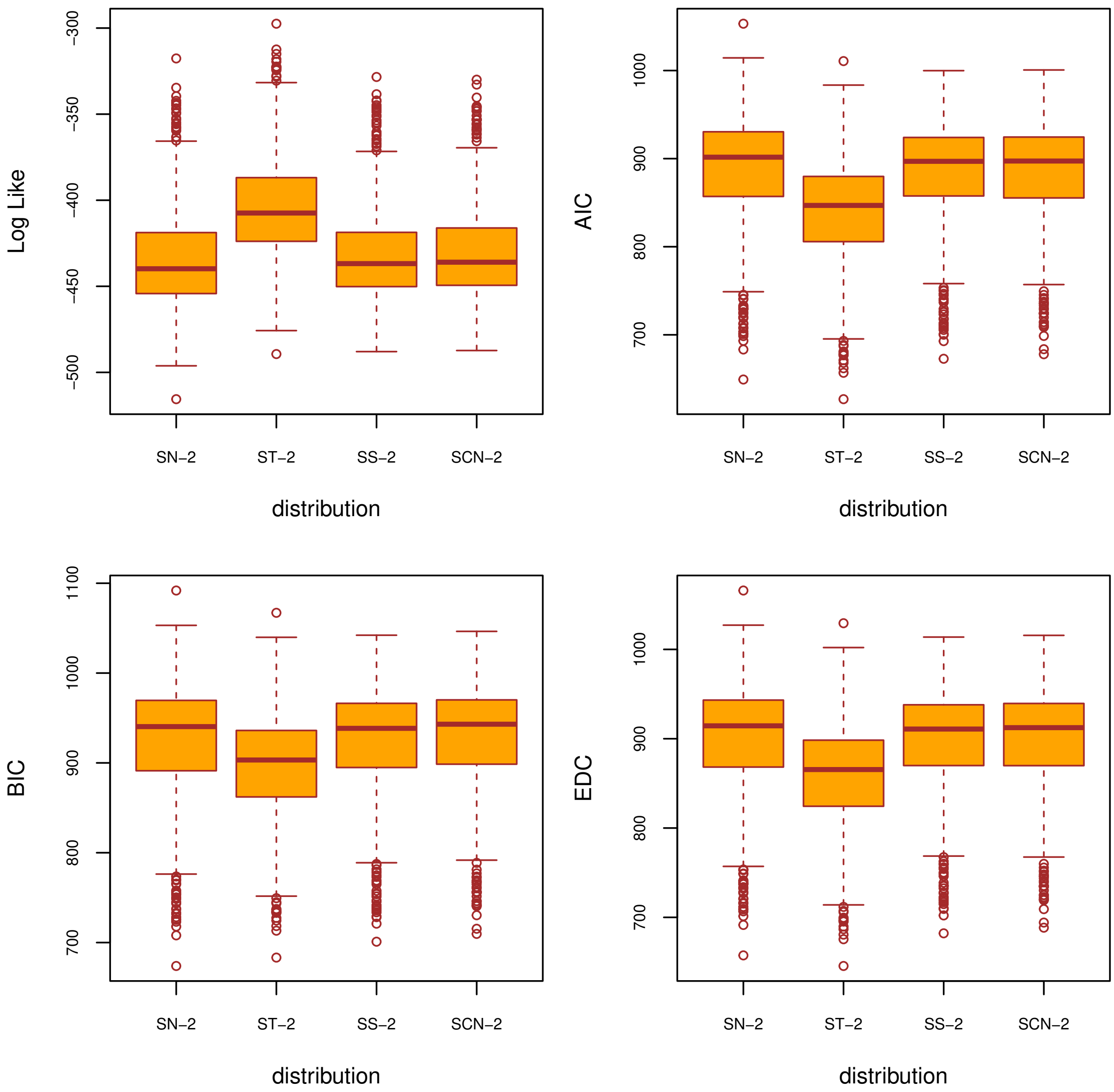

FIGURE 2.3: AIC, BIC, EDC of 250 simulated FM_2-ST-GARCH(1,1) model against other competitors of the family.

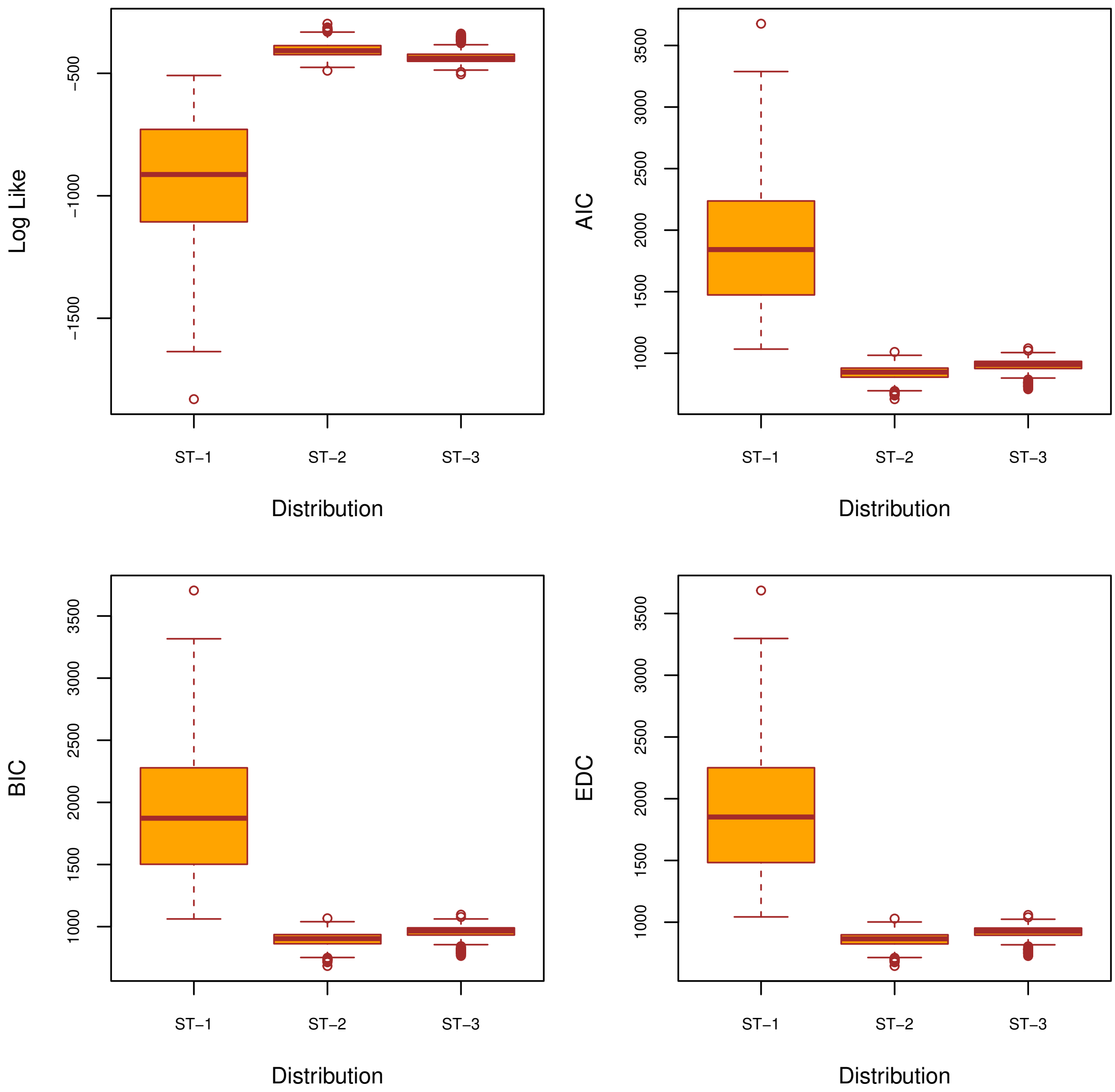

FIGURE 2.4: AIC, BIC, EDC of 250 simulated FM_2-ST-GARCH(1,1) model against FM_1-ST-GARCH(1,1) and FM_3-ST-GARCH(1,1) models.

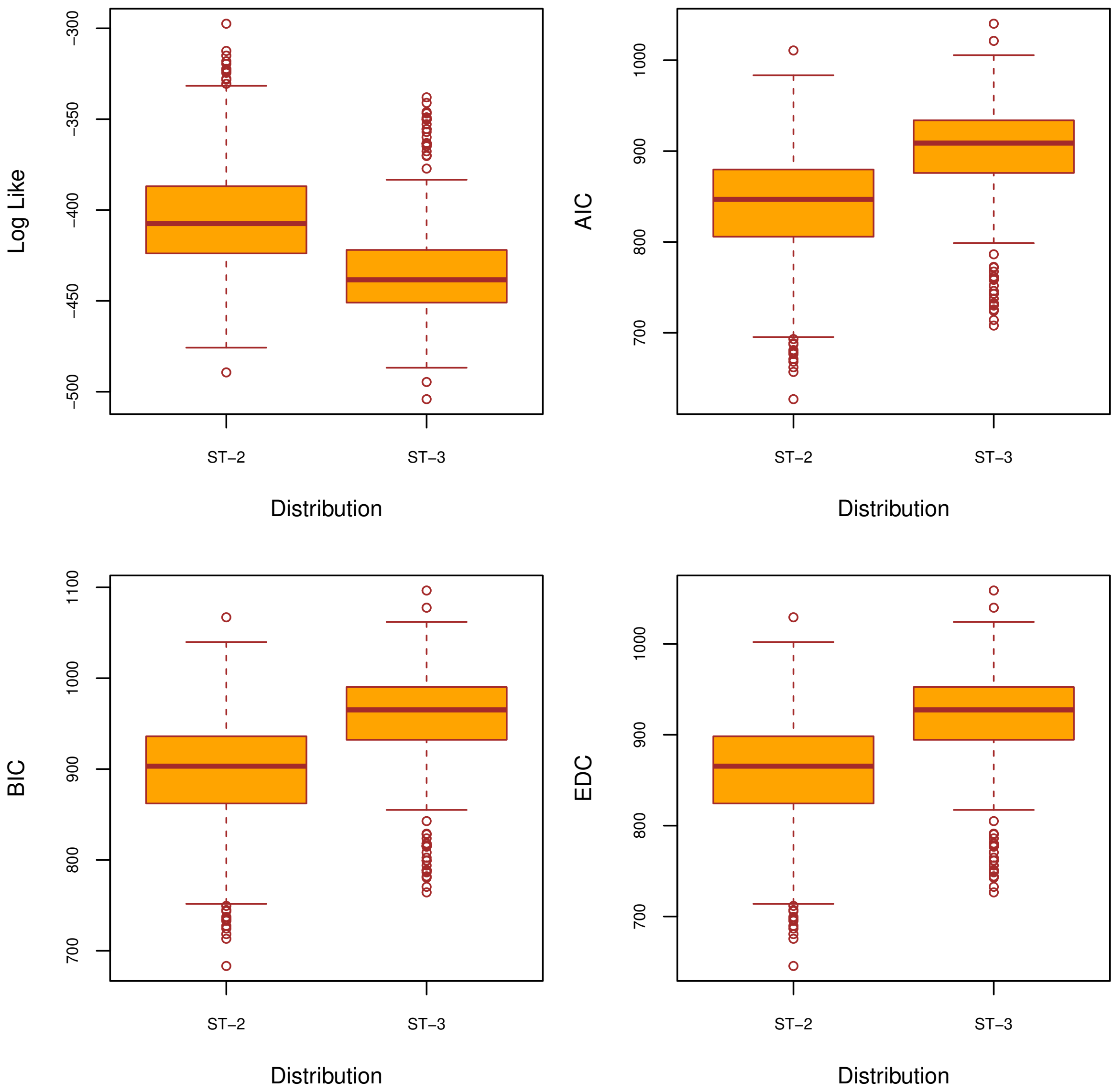

FIGURE 2.5: AIC, BIC, EDC of 250 simulated FM_2-ST-GARCH(1,1) model against FM_3-ST-GARCH(1,1) model.

TABLE 2.3: ANOVA results

Criteria	Sum sq	Mean sq	F-value	$Pr(> F)$
Log Like	303,296	101,099	117.37	0
AIC	815,630	271,877	73.814	0
BIC	317,972	105,991	25.552	3.22E-16
EDC	622,843	207,614	54.205	0

Table 2.4: Tukey HSD test

Criteria	Comparison	Difference	Lower Bound	Upper Bound	p-value
	FM_2-ST-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	30.283	25.510	35.055	0
	FM_2-SSL-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	2.591	-2.181	7.363	0.502
Log Like	FM_2-SCN-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	3.381	-1.390	8.154	0.263
	FM_2-SSL-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	-27.692	-32.464	-22.919	0
	FM_2-SCN-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	-26.901	-31.673	-22.128	0
	FM_2-SCN-GARCH(1,1) vs FM_2-SSL-GARCH(1,1)	0.790	-3.981	5.563	0.974
	FM_2-ST-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	-48.582	-58.451	-38.712	0
	FM_2-SSL-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	-3.262	-13.131	6.607	0.830
AIC	FM_2-SCN-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	-2.843	-12.712	7.025	0.880
	FM_2-SSL-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	45.319	35.450	55.189	0
	FM_2-SCN-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	45.738	35.869	55.607	0
	FM_2-SCN-GARCH(1,1) vs FM_2-SSL-GARCH(1,1)	0.418	-9.450	10.287	0.999536
	FM_2-ST-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	-27.481	-37.955	-17.008	0
	FM_2-SSL-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	0.118	-10.355	10.591	0.999
BIC	FM_2-SCN-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	4.058	-6.414	14.532	0.751
	FM_2-SSL-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	27.599	17.126	38.073	0
	FM_2-SCN-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	31.540	21.066	42.013	0
	FM_2-SCN-GARCH(1,1) vs FM_2-SSL-GARCH(1,1)	3.940	-6.533	14.413	0.768
	FM_2-ST-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	-41.617	-51.681	-31.553	0
	FM_2-SSL-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	-2.146	-12.210	7.917	0.947
EDC	FM_2-SCN-GARCH(1,1) vs FM_2-SN-GARCH(1,1)	-0.565	-10.629	9.498	0.998
	FM_2-SSL-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	39.471	29.407	49.535	0
	FM_2-SCN-GARCH(1,1) vs FM_2-ST-GARCH(1,1)	41.052	30.988	51.116	0
	FM_2-SCN-GARCH(1,1) vs FM_2-SSL-GARCH(1,1)	1.580	-8.483	11.645	0.977

2.5.2 Real data analysis: Apple Inc. (AAPL)

In this section, as an illustrative case study of the proposed models and methods, we consider an analysis of the 904 daily observations of the closing price of Apple stocks from 2/01/2019 to 03/08/2022; data have been recovered from finance.yahoo.com. The original time series plot of the price, is given in Figure 2.6. As the price series are known to be non-stationary, we analyze the returns, computed as

$$x_t = log(\frac{y_t}{y_{t-1}})$$

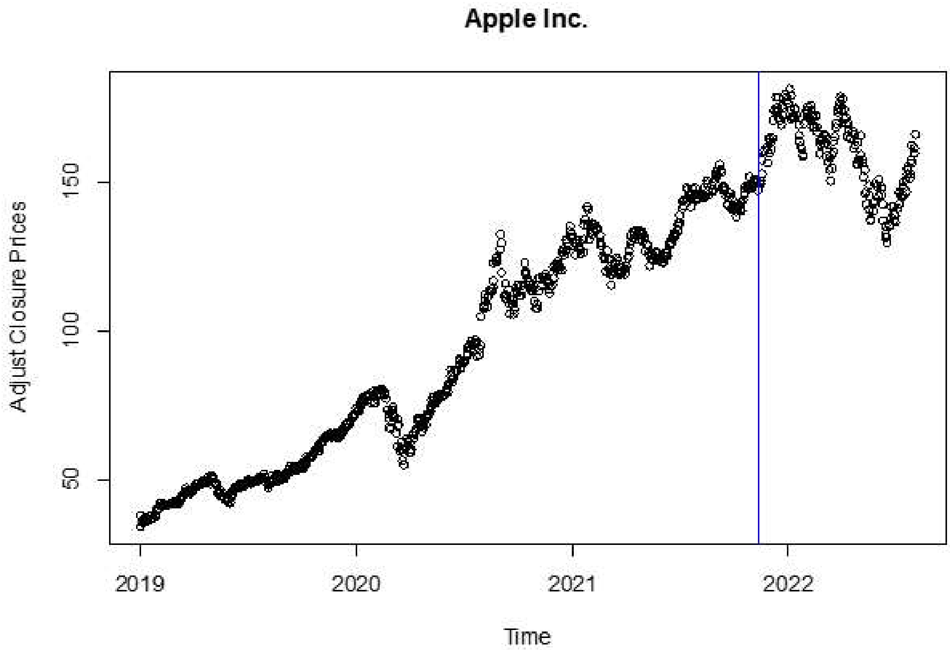

FIGURE 2.6: Time series plot of the Apple Inc. from 2/01/2019 to 03/08/2022.

The time series x_t has been plotted in Figure 2.7 where the volatility clustering effect is clearly visible, supporting the presence of conditional heteroskedasticity.

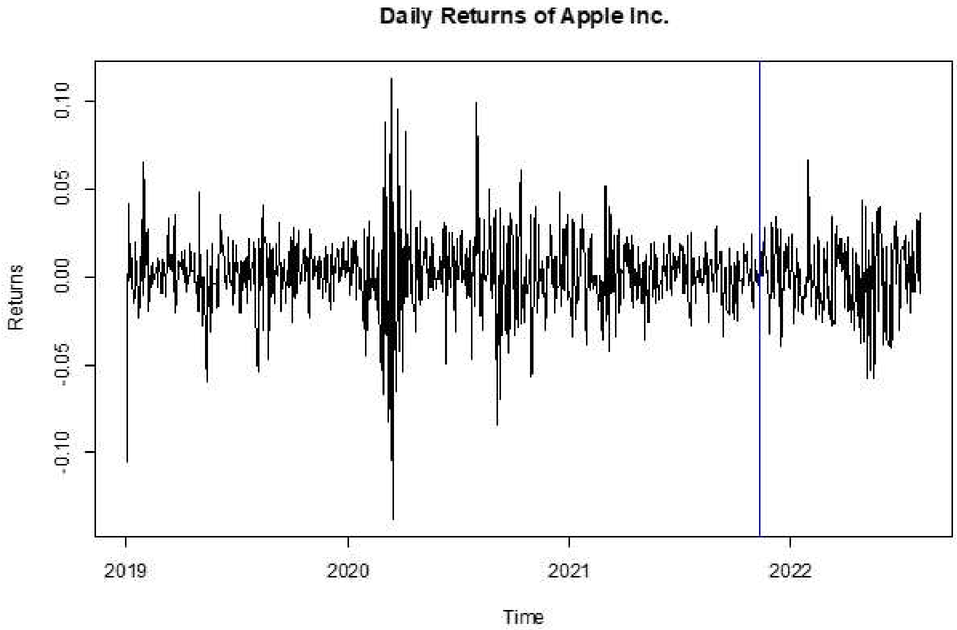

FIGURE 2.7: Daily Return of Apple Inc.

In the following analysis, we compare various competing models, including FM$_m$-N-GARCH(1,1), FM$_m$-T-GARCH(1,1), FM$_m$-SN-GARCH(1,1), FM$_m$-ST-GARCH(1,1), FM$_m$-SSL-GARCH(1,1), and FM$_m$-SCN-GARCH(1,1), by examining several information selection criteria. The selection process involves evaluating the values of these criteria, as presented in Table 2.5, to determine the models that perform optimally. Upon careful analysis, it becomes evident that the FM$_m$-SMSN–GARCH(1,1) models, specifically those with two or three components, consistently outperform the other models. This conclusion is supported by the Log-likelihood, AIC, BIC, and EDC criteria, all of which demonstrate the superiority of the FM$_m$-SMSN–GARCH(1,1) models with multiple components. Furthermore, we observe that both the FM$_m$-ST-GARCH(1,1) and FM$_m$-SSL-GARCH(1,1) models, when implemented with two or three components, outperform the FM$_m$-N-GARCH(1,1), FM$_m$-T-GARCH(1,1), FM$_m$-SN-GARCH(1,1), and FM$_m$-SCN-GARCH(1,1) models. This finding indicates that incorporating asymmetric distributions characterized by heavier tails and utilizing a finite mixture of components result in a superior fit compared to the simpler models FM$_m$-N-GARCH(1,1), FM$_m$-ST-GARCH(1,1), FM$_m$-SN-GARCH(1,1), and FM$_m$-SCN-GARCH(1,1) with $m = 1, 2, 3$. Additionally, when considering the Log-likelihood, AIC, BIC, and EDC criteria collectively, it becomes evident that the FM$_2$-ST–GARCH(1,1) model consistently outperforms all other competing models. This model exhibits the highest level of goodness-of-fit across the evaluated criteria, thus establishing its superiority in capturing the underlying characteristics of the dataset. Based on the comprehensive assessment of these selection criteria, it is clear that the FM_2-ST–GARCH(1,1) model stands out as the most suitable choice and is therefore designated as the best model for the given dataset.

TABLE 2.5: Model selection criteria for the proposed FM_m-SMSN-GARCH(1,1) model with various number of components m = 1, 2, 3

	m	Log-likelihood	AIC	BIC	EDC
FM_m-SN-GARCH(1,1)	1	1120.897	-2227.79	-2195.71	-2204.15
	2	1250.211	-2478.42	-2428	-2441.27
	3	1254.471	-2478.94	-2410.19	-2428.28
FM_m-ST-GARCH(1,1)	1	1251.497	-2486.99	-2450.33	-2459.97
	2	**1628.441**	**−3224.88**	**−3151.55**	**−3170.84**
	3	1253.450	-2482.9	-2427.9	-2442.37
FM_m-SSL-GARCH(1,1)	1	1249.611	-2483.22	-2446.55	-2456.2
	2	**1583.88**	**−3135.76**	**−3062.43**	**−3081.72**
	3	1252.447	-2480.89	-2425.89	-2440.36
FM_m-SCN-GARCH(1,1)	1	1248.366	-2478.73	-2437.48	-2448.33
	2	1251.980	-2477.96	-2418.38	-2434.05
	3	1253.102	-2472.2	-2394.29	-2414.78
FM_m-N-GARCH(1,1)	1	1098.814	-2183.63	-2151.54	-2159.98
	2	1246.188	-2470.38	-2419.96	-2433.22
	3	1251.907	-2473.81	-2405.06	-2423.15
FM_m-T-GARCH(1,1)	1	1247.908	-2479.82	-2443.15	-2452.79
	2	1247.340	-2470.68	-2415.68	-2430.15
	3	1245.659	-2459.32	-2385.98	-2405.27

Table 2.6 presents the maximum likelihood estimates and their estimates of the standard errors of the parameters (provided by the observed information matrix given in section 2.4.3) for the FM_2-ST–GARCH(1,1) model, which has been identified as the best fit for the return series of Apple Inc. The estimated values provide valuable insights into the underlying dynamics of the dataset. Figure 2.8 showcases the histograms of the marginal residuals obtained from the FM_2-ST–GARCH(1,1) model, with the estimated marginal densities overlaid for visual comparison. This graphical representation aids in understanding the distributional characteristics of the residuals and highlights the effectiveness of the chosen model in capturing the observed data patterns. Furthermore, Figure 2.9 displays Q–Q plots of the residuals based on the FM_2-N-GARCH(1,1) model. These plots provide a visual assessment of the goodness-of-fit between the theoretical

distribution assumed by the model and the empirical distribution of the residuals. Notably, the Q–Q plots reinforce the superiority of the FM_2-ST–GARCH(1,1) model, as they align more closely with the expected theoretical distribution. Taken together, both Figure 2.8 and Figure 2.9 substantiate the conclusion that the FM_2-ST–GARCH(1,1) model is the most appropriate and reliable choice for modeling the return series of Apple Inc. This model accurately captures the essential characteristics and exhibits superior performance when compared to other competing models.

TABLE 2.6: ML estimation results with their standard errors for fitting the FM_2-ST-GARCH(1,1) model on the Apple Inc. data.

Component	Parameters	ML Estimates	standard error
First Component	π_1	0.409	0.024
	μ_1	-0.467	0.184
	σ_1^2	0.672	0.375
	λ_1	1.302	0.837
	ν	3.747	1.860
Second Component	π_2	0.590	0.024
	μ_2	0.817	0.097
	σ_2^2	0.975	0.401
	λ_2	-1.296	0.940
	ν	3.747	1.860
GARCH Parameters	ω	-1.063	0.362
	μ_0	0.002	0.040
	α_1	0.017	0.130
	β_1	0.982	0.204

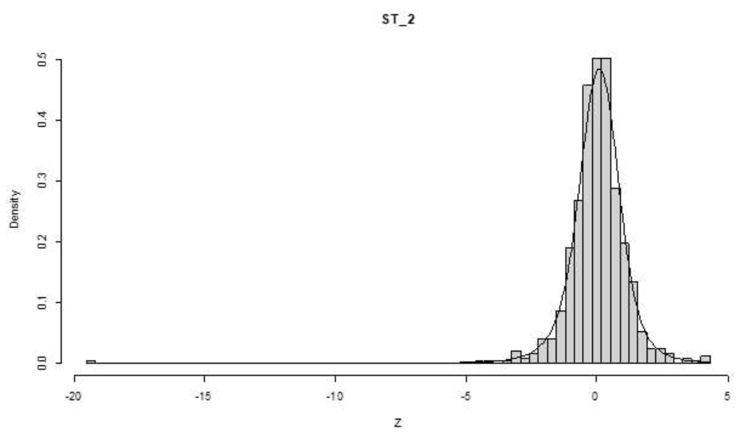

FIGURE 2.8: Histograms of the marginal residuals of the FM_2-ST-GARCH(1,1) mixture model to Apple inc. data overlaid with estimated marginal densities.

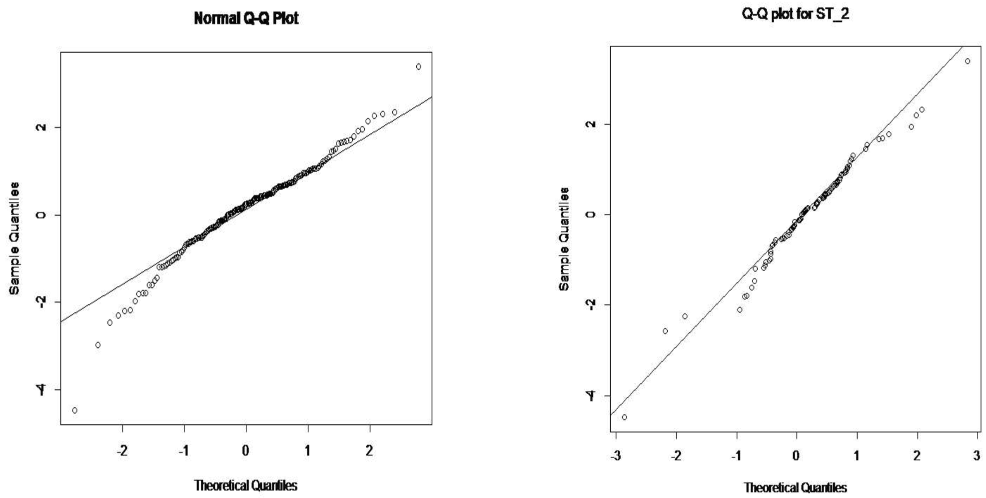

FIGURE 2.9: The Q–Q Plot of Residuals of Apple Data Based on the Gaussian (left) and FM_2-ST (right) Models.

The evaluation of the proposed models' forecasting accuracy is conducted using the AAE (Average Absolute Error) and RMSE (Root Mean Squared Error) criteria, defined as follows:

$$AAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$

These metrics provide insight into how well the models perform in predicting the data. The results of AAE and RMSE for both the training and test datasets are summarized in Table 2.7.

Upon comparing these criteria, several key observations emerge:

1. Model Complexity Matters: Models with higher complexity (m=2 and m=3) consistently outperform their simpler counterparts (m=1) in both training and test datasets. This highlights the importance of capturing the intricacies present in the financial data.

2. Asymmetric Distributions and Mixture Components: FM_2-ST-GARCH and FM_2-SSL-GARCH models, which incorporate skewness and utilize mixture components, exhibit superior predictive capabilities across all metrics. This suggests that financial data often follows asymmetric distributions with heavier tails and comprises a mixture of underlying components. These factors must be considered for accurate forecasting.

3. FM_2-ST-GARCH Shines: Among all the models and configurations, FM_2-ST-GARCH stands out as the top-performing model. It consistently achieves the lowest

AAE and RMSE values in both the training and test datasets. This exceptional performance underscores its suitability for modeling the complex dynamics inherent in financial time series data.

4. Versatile Models: Several models, especially those with m=2 and m=3, demonstrate strong predictive performance, showcasing their robustness and applicability across various scenarios. This versatility is valuable for financial analysts and decision-makers seeking reliable forecasts.

Therefore, the findings suggest that the proposed FM_m-SMSN-GARCH models, with their ability to capture the complexity of financial time series data, including skewness and mixture components, offer promising avenues for improving decision-making, risk management, and investment strategies in the realm of financial modeling and forecasting.

Figure 2.10 offers a visual depiction of the model's predictive prowess. It showcases the actual closing prices of Apple Inc. (depicted in black) alongside their corresponding predictions (highlighted in red), generated by the FM_2-ST-GARCH Model. This graphical representation serves as a concrete demonstration of the model's capability to forecast financial time series data accurately. It allows for a direct comparison between observed and predicted values, providing a vivid illustration of the model's effectiveness in capturing real-world financial trends and fluctuations

TABLE 2.7: Performance Metrics for the Proposed FM_m-SMSN–GARCH Model

Model	(m)	Training Data AAE	Training Data RMSE	Test Data AAE	Test Data RMSE
SN	1	53.390	65.368	88.421	112.377
	2	23.123	26.643	35.740	48.191
	3	22.272	31.421	26.146	37.015
ST	1	15.870	21.769	20.662	27.955
	2	**9.173**	**11.447**	**12.346**	**15.257**
	3	15.339	20.242	17.742	22.434
SSL	1	16.358	21.752	20.866	28.644
	2	12.964	15.860	13.100	16.027
	3	15.668	20.607	17.981	21.006
SCN	1	25.262	29.675	36.587	52.597
	2	16.752	21.025	22.175	28.993
	3	16.762	20.659	23.633	30.801
N	1	49.575	62.240	85.924	107.073
	2	48.203	64.847	81.274	103.493
	3	26.352	33.533	43.374	57.669
T	1	30.515	38.857	77.275	101.191
	2	28.138	35.553	63.529	75.809
	3	29.562	39.494	67.781	85.409

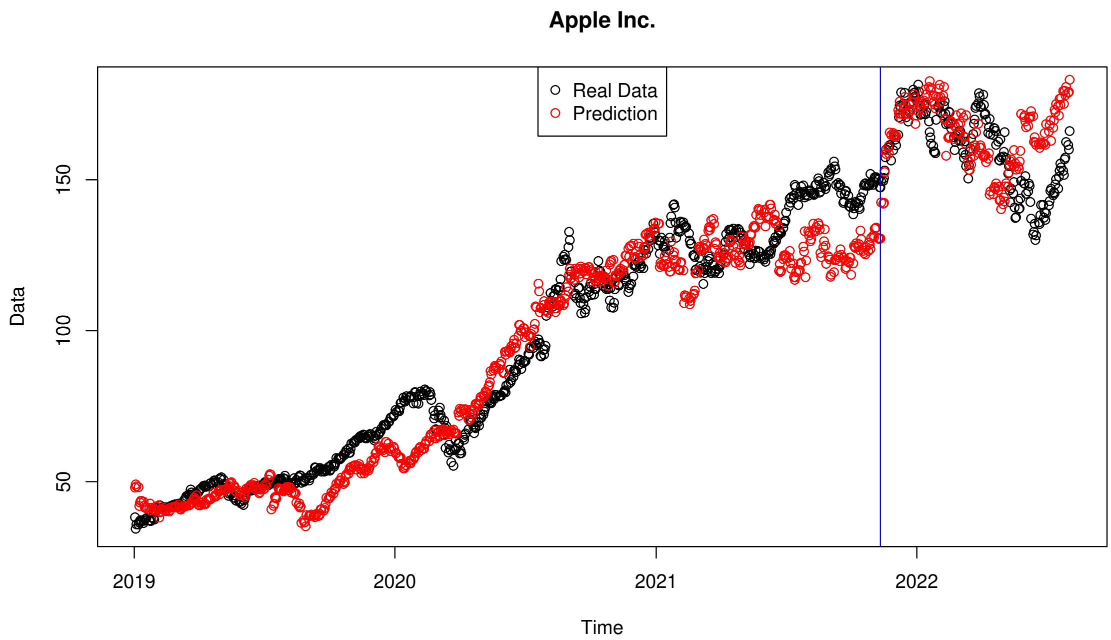

FIGURE 2.10: Real data (black) and their prediction (red) of closing price of Apple Inc. based on the FM_2-ST-GARCH Model.

2.6 Conclusions

A very flexible family of conditional heteroscedasticity named the FM_m-SMSN-GARCH(1,1) distributions has been proposed in this chapter, which is a novel tool alternative to existing methods for investigating features and modelling time series data. In fact, in the present study, a mixture model is proposed based on a flexible class of symmetric/asymmetric and lightly/heavy tailed distribution. The robust proposed model can efficiently and simultaneously deal with skewness, heavy-tailed-ness and mixture of components in the conditional heteroscedasticity process.

The novel model allows the researchers in various fields to analyze data in a very flexible methodology. A computationally efficient ECME algorithm is developed to compute the ML estimates of parameters under a convenient hierarchical representation of the model, and the performance of our proposed algorithm is confirmed by some extensive simulation studies. After obtaining the ML estimates through the proposed ECME algorithm, they are performed and coded with available by R package, where all codes are available upon request.

In order to evaluate the performance of the proposed model, a real data analysis has been conducted. The results demonstrate the model's capability to capture key characteristics of the data and provide accurate inference. Additionally, simulation experiments have been carried out to illustrate the desirable asymptotic properties of the ML estimates obtained using our proposed computational techniques.

Future research endeavors will focus on several projects. Firstly, we aim to develop a web-based shiny application (Manouchehri and Nematollahi, 2022a) that enables modeling, estimation, and robust inference for the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model with scale mixtures of skew-normal innovations. This application will provide a general and highly flexible class of error distributions, with a fully Bayesian inference implemented through the Markov chain Monte Carlo method. Another promising avenue for future work is the extension of our proposed methodology to incorporate the class of normal mean-variance mixture (NMVM) distributions (McNeil *et al.*, 2015), which can offer increased flexibility in modeling data exhibiting strong skewness and heavy-tailed features. Additionally, exploring multivariate variants of the current approach, inspired by existing research in this domain (Billio *et al.*, 2006; Caporin and McAleer, 2012; Billio and Caporin, 2005), presents another interesting extension of the current work.

Finally, we intend to extend the proposed model to the realm of Model-Based Clustering of Multiple Time Series, which holds significant promise. Prior works by (Fraley and Raftery, 2002) and (Aielli and Caporin, 2014) have explored similar topics. [3]

[3] We refer the reader to (Bouveyron *et al.*, 2019) for a more complete listing of Model-Based Clustering and Classification approaches.

Chapter 3

Robust mixture ARCH modeling based on the normal mean-variance mixture distributions

3.1 Abstract

The proposed chapter presents a novel and advanced solution to tackle the complexities of analyzing time series data with conditional heteroscedasticity. By introducing the Mixture of ARCH (MoARCH) model, the authors address the limitations of conventional ARCH models that assume a normal distribution for errors. The MoARCH model incorporates the Normal Mean Variance Mixture (NMVM) distribution, allowing it to effectively handle non-normal characteristics, such as skewness and heavy tails, in the data. This is achieved through a stochastic representation, which facilitates efficient parameter estimation using a combination of an EM-type algorithm and a penalized likelihood function. The study finds that the NMVM-MoARCH model outperforms traditional normal-based approaches in terms of accuracy and efficiency in analyzing time series data with skewness and heavy tails. This contribution to the field of model-based clustering has the potential to revolutionize the analysis of time series data and provide deeper insights and more valuable results.

Keywords: EM algorithm, MoARCH models, Normal mean-variance mixture distributions, Penalized likelihood.

3.2 Introduction

Clustering data into meaningful groups is a long-standing challenge for statisticians and data scientists. Model-based clustering is a statistical technique that seeks to identify a set of parameters for a statistical model that effectively explains the data and produces insightful groupings. One popular model used in this approach is the mixture model, which blends a set of probability distributions to characterize the data. The process of model-based clustering involves fitting the data to the mixture model and determining the parameters using maximum likelihood estimation, resulting in clusters based on the estimated parameters. This method is versatile and efficient, as it can handle complex distributions and missing information. In this field, there is a significant amount of literature available, covering a range of applications such as gene expression (Yeung *et al.*, 2001), medicine (Fraley and Raftery, 2002), social networks (Handcock *et al.*, 2007), high-dimensional data (Bouveyron and Brunet-Saumard, 2014), and time series (Fröhwirth-Schnatter and Kaufmann, 2008).

The Mixture of ARCH (AutoRegressive Conditional Heteroscedasticity proposed by (Engle, 1982) model is a statistical technique used to model and cluster time series data. It is designed to handle the problem of non-stationarity and conditional heteroscedasticity in time series data, which is a common issue in finance and economics. The mixture component allows for multiple ARCH models to be combined in order to more accurately capture the dynamic relationships between variables. The ARCH models are used to model the conditional variance of the time series, and the mixture component allows for multiple ARCH models to be combined to better capture the heterogeneity in the data. The application of the Mixture of ARCH model to time series data provides valuable insights into the relationships between variables and can help to identify patterns and clusters in the data. In the literature, most of the mixture models are based on the normal distribution due to its mathematical ease and simplicity. The commonly used approach for estimating the parameters of a mixture model is typically based on maximum likelihood (ML) with a normality assumption. This approach can produce accurate results when the error distribution follows a normal distribution. However, in many real-world scenarios, this assumption is not met, particularly when the data contains outliers or is highly skewed. This can lead to a normal distribution being insufficient in these cases.

In this study, a novel and robust Mixture ARCH model is introduced that utilizes the Normal Mean Variance Mixture (NMVM) distribution. The NMVM distribution, first introduced by (McNeil *et al.*, 2015), offers several key features such as the ability to handle skewness and heavy tails. The class of NMVM distributions encompasses several significant variants, including the generalized hyperbolic (GH) distribution (Barndorff-Nielsen and Halgreen, 1977), the Birnbaum-Saunders normal mean-variance mixture (NMVBS) distribution (Pourmousa *et al.*, 2015), and the Lindley normal mean-variance mixture (NMVL) distribution (Naderi *et al.*, 2018) as special cases. As a result, the novel NMVM-MoARCH model we present is both flexible and sturdy, and it is capable of addressing skewness and heavy tails in the MoARCH scenario. The research also incorporates a penalized likelihood function to find the most suitable number of components. Additionally, the stochastic representation of the model leads to the creation of two adaptations of the expectation-maximization algorithm (Dempster *et al.*, 1977), the ECM algorithm (Meng and Rubin, 1993) and the ECME algorithm (Liu and Rubin, 1994).

The main objective of this study is three-pronged: Firstly, to develop a MoARCH model using NMVM distributions; Secondly, to assess the efficacy of the proposed model through computational analysis; and Thirdly, to apply the findings to actual data. The paper is structured as follows: In Section 3.3, we provide a concise overview of the NMVM distribution and its salient features. Section 3.4 presents the proposed NMVM-MoARCH(q) model. Section 3.5 details the maximum penalized estimates (MPL) of the NMVM-ARCH(q) model using an EM-type algorithm, as well as the methods for determining the tuning parameter and the number of components. Section 3.6 deals with the practical aspects of estimating standard errors and selecting appropriate initial values. In Section 3.7, extensive simulation studies are carried out to evaluate the efficiency of the proposed approach. The results of these studies are then applied to actual data examples to demonstrate the practicality of the methodology. Finally, the work concludes with a summary of the key findings in Section 3.8.

3.3 Normal mean-variance mixture distributions

The NMVM distribution belongs to a family of flexible distributions formed by scaling the mean and variance of a normal random variable with a positive scale-valued mixing random variable. In particular, a random variable Y is said to follow the normal mean-variance mixture distribution class described by (McNeil *et al.*, 2015), indicated by $X \sim NMVM(\mu, \lambda, \sigma^2, \boldsymbol{\nu})$, if it is represented by a hierarchical mixing structure as follows:

$$X|(W = w) \sim N(\mu + w\lambda, w\sigma^2), \tag{3.1}$$

$$W \sim h(w; \boldsymbol{\nu}),$$

where $\mu \in \mathbb{R}$, $\lambda \in \mathbb{R}$, and W is a positive mixing random variable with a cumulative distribution function (CDF) $H(.;\boldsymbol{\nu})$, and a probability density function (PDF) $h(.;\boldsymbol{\nu})$, both of which are contingent on a possible parameter vector $\boldsymbol{\nu}$.

Furthermore, the random variable X is said to obey the NMVM distribution $Y \sim NMVM(\mu, \lambda, \sigma^2, \boldsymbol{\nu})$, with the following stochastic representation:

$$X = \mu + \lambda W + \sqrt{W} Z, \tag{3.2}$$

where $\mu \in \mathbb{R}$, $\lambda \in \mathbb{R}$, Z is a univariate normal random variable with mean 0 and variance σ^2. The random variable W is a positive variable with a PDF of $h(.;\boldsymbol{\nu})$, which is statistically independent of Z.

The NMVM distribution has a PDF given by:

$$f_{NMVM}(x;\mu,\lambda,\sigma^2,\boldsymbol{\nu}) = \int_0^\infty \phi(x;\mu + w\lambda, w\sigma^2) h(w;\boldsymbol{\nu}) dw. \tag{3.3}$$

In this paper, we make use of the NMVM distribution family, which provides a suitable tool for modeling data with asymmetric and heavier tails compared to Gaussian distributions. By considering the value of W, we can see that the mean and variance of Y are not constant. Researchers can utilize this favorable attribute to obtain appropriate distributions that can effectively model data with asymmetry and tails that are heavier than those of the normal distribution.

One of the special cases of the NMVM family is the GH distribution, as specified in (McNeil *et al.*, 2015). This subclass of NMVM is obtained by using the generalized inverse Gaussian (GIG) distribution (Good, 1953) as the PDF of W. In this paper, we mainly use the GH distribution to model our data. For more details on the GH distribution, please refer to Appendix B.1.

3.4 Designing and Executing the NMVM-MoARCH

The purpose of the new family of time series, NMVM-MoARCH(q), is to study the efficient and robust clustering of N multiple time series with respect to asset pricing and volatility. The ultimate goal is to categorize the time series into an optimal number, (g), of clusters, represented by $(C_1, .., C_g)$, as determined by the mixture model components.

Consider a panel of multiple time series, denoted as $\{y_{jt}\}$, $t = 1, \ldots, n_j$, and is observed for N units $j = 1, \ldots, N$. In fact, y_{jt} is the random variable representing the asset price at time t $(t = 1, ..., n_j)$. Additionally, let us define the component-indicator random variable Z_j, that assigns time series j to one of the g components or clusters

C_i, with probability $P(Z_j^{(i)} = 1) = \pi^{(i)}$, where $\sum_{i=1}^{g} \pi^{(i)} = 1$ and $\pi^{(i)} > 0$ for $i = 1, ..., g$ and $j = 1, ..., N$.

Let $\boldsymbol{y_j^2} = (1, y_{jt-1}^2, ..., y_{jt-q}^2)$, $j = 1, ..., N$, be a $(q+1) \times 1$ vector of variables and $\boldsymbol{\alpha_i} = (\alpha_{i0}, \alpha_{i1}, ..., \alpha_{iq})'$ be the coefficients for the conditional variance. Given $Z_j^{(i)} = 1$, in each single time series, we model the observations as follows:
Let $\left(y_{jt}^{(i)} | Z_j^{(i)} = 1\right) \equiv y_{jt}^{(i)}$. Consequently, we have the following relationship:

$$y_{jt} = \sqrt{h_{jt}^{(i)}} \varepsilon_{jt}^{(i)}; \quad \varepsilon_{jt}^{(i)} \sim \text{NMVM}(0, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu^{(i)}}), \quad \text{with probability} \quad \pi_i; \quad i = 1, \ldots, g. \tag{3.4}$$

where the conditional variance $h_{jt}^{(i)}$ is represented by the expression $\alpha_{i0} + \alpha_{i1} y_{j,t-1}^2 + \ldots + \alpha_{iq} y_{j,t-q}^2 = \boldsymbol{\alpha_i}' \boldsymbol{y_{jt}^2}$. Here, $\boldsymbol{y_j^2}$ is a $(q+1) \times 1$ vector of variables for time series j, and $\boldsymbol{\alpha_i}$ denotes the coefficients for the conditional variance. To ensure proper identification, it is essential to set the expected value of the conditional variance, $E[h_{jt}^{(i)}]$, equal to 1 when $\sigma^{2(i)}$ represents the unconditional variance of the innovations. Achieving this condition can be accomplished by setting α_{i0} to be equal to 1 minus the sum of all other α_{iq} coefficients. This reparameterization guarantees that the unconditional value of h is 1, thereby ensuring stability and accuracy in the modeling process.

Let $\boldsymbol{y_j} = (y_{j0}, .., y_{jn_j})'$ come from a i-component mixture ARCH model, if the conditional density function is given by:

$$f(y_{jt} | \mathcal{F}_{j,t-1}, \boldsymbol{\theta_i}) = f_{NMVM}\left(\frac{y_{jt}}{\sqrt{h_{jt}^{(i)}}}; 0, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu^{(i)}}\right), \text{ with prob. } \pi_i, \ i = 1, \ldots, g; \ j = 1, \ldots, N; \ t = 1, \ldots, n_j, \tag{3.5}$$

where $\boldsymbol{\theta_i} = (\boldsymbol{\alpha_i'}, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu^{(i)}})$.

The joint density function for the i-component mixture ARCH model can be defined through equation (3.5) as:

$$f_{MAR}(\boldsymbol{y}_j | \boldsymbol{\mathcal{F}}_j, \boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i \prod_{t=1}^{n_j} f(y_{jt} | \mathcal{F}_{j,t-1}, \boldsymbol{\theta}_i), \tag{3.6}$$

where $\boldsymbol{\Theta} = (\pi^{(1)}, ..., \pi^{(g-1)}, \boldsymbol{\theta_i}', ..., \boldsymbol{\theta_g}')'$ represents the complete set of parameters for the NMVM-MoARCH(q) and $\boldsymbol{\theta^{(i)}} = (\boldsymbol{\alpha_i'}, \lambda^{(i)}, \sigma^{2(i)}, , \boldsymbol{\nu^{(i)}})$. The value of $\boldsymbol{\nu^{(i)}}$ is different for different models: it is equal to $\nu^{(i)}$ for NMVL, NIG, NMVBS, and GHST, equal to $(\kappa^{(i)}, \psi^{(i)})'$ for VG, and equal to 0 for SL.

The log-likelihood of parameters for an iid sample $(\boldsymbol{y}_1, ..., \boldsymbol{y}_N)$ can be calculated using the characterization (3.6) as:

$$ell(\boldsymbol{\Theta}) = \sum_{j=1}^{N} \log f_{MAR}(\boldsymbol{y}_j|\boldsymbol{\mathcal{F}}_j, \boldsymbol{\Theta}) = \sum_{j=1}^{N} \log \left(\sum_{i=1}^{g} \pi^{(i)} \prod_{t=1}^{n_j} f_{NMVM}(y_{jt}; 0, \lambda^{(i)}, \sigma^{2(i)} h_{jt}^{(i)}, \boldsymbol{\nu^{(i)}}) \right). \tag{3.7}$$

The NMVM-MoARCH(q) model can be expressed as a two-level hierarchical representation:

$$\frac{y_{jt}}{\sqrt{h_{jt}^{(i)}}}|\mathcal{F}_{jt-1}, Z_j^{(i)} = 1 \sim \text{NMVM}\left(0, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu}^{(i)}\right),$$

$$\boldsymbol{Z_j} \sim Multinomial(1, \pi^{(1)}, ..., \pi^{(g)}).$$

By utilizing the stochastic representation of the NMVM, the hierarchical representation of the MoARCH(q) model that has been proposed is expressed as follows:

$$\frac{y_{jt}}{\sqrt{h_{jt}^{(i)}}}|\mathcal{F}_{jt-1}, Z_j^{(i)} = 1 \sim N\left(w_{jt}\lambda^{(i)}, w_{jt}\sigma^{2(i)}\right),$$

$$W_{jt}|Z_j^{(i)} = 1 \sim h(w_{jt}, \boldsymbol{\nu^{(i)}}),$$

$$\boldsymbol{Z_j} \sim Multinomial(1, \boldsymbol{\pi}^{(i)}).$$

The alternative form of the above hierarchical representation is:

$$y_{jt}|\mathcal{F}_{jt-1}, Wjt = w_{jt}, Z_j^{(i)} = 1 \sim N\left(\sqrt{h_{jt}^{(i)}}(w_j^{(i)}\lambda^{(i)}), w_{jt}\sigma^{2(i)}h_{jt-1}^{(i)}\right), \tag{3.8}$$

$$W_{jt}|Z_j^{(i)} = 1 \sim h(w_{jt}, \boldsymbol{\nu^{(i)}}),$$

$$p(Z_j^{(i)} = 1) = \pi^{(i)},$$

for $i = 1, ..., g,\ j = 1, ..., N,\ t = 1, ..., n_j$.

To address potential problems in applications, we tackle one such issue which is label switching. To resolve this, the maximum likelihood inferential paradigm is employed, making label switching a theoretical identifiability issue that can usually be resolved by ordering the mixing proportions in the form of $\pi^{(1)} >, ..., > \pi^{(g)}$. In cases where the mixing proportions are equal, other model parameters can be considered for total ordering.

Another potential issue is the presence of a large number of components in mixture models, which can lead to overfitting of the data and poor interpretations. On the other hand, having too few components can result in inflexibility in approximating the actual data structure. Hence, estimating the accurate number of components is crucial in mixture models. To address potential underestimation or overestimation, we use a penalized log-likelihood function which is a complex function that requires high-dimensional nonlinear optimization. To solve this computational complexity, we consider using an EM-type algorithm.

The penalized log-likelihood function is defined as follows:

$$\ell_{\text{pen}}(\boldsymbol{\Theta}) = \ell(\boldsymbol{\Theta}) - N\gamma D_{f.\text{MAR}} \sum_{i=1}^{g} \left(\log(\varepsilon + \pi^{(i)}) - \log(\varepsilon)\right), \tag{3.9}$$

where $\ell(\Theta)$ represents the log-likelihood function in equation, γ is a tuning parameter , ε is a small positive constant with a value of 10^{-6}, and $\mathrm{D_{f.MAR}}$ stands for the number of free parameters for each component. The free parameters in $\mathrm{GHST - MoARCH}$, $\mathrm{NIG - MoARCH}$, $\mathrm{NMVBS - MoARCH}$, and $\mathrm{NMVL - MoARCH}$ is $\mathrm{D_{f.MAR}} = (\mathrm{q}+1)+4$, while in $\mathrm{VG - MoARCH}$, it is $\mathrm{D_{f.MAR}} = (\mathrm{q}+1)+5$, and in $\mathrm{SL - MoARCH}$, it is $\mathrm{D_{f.MAR}} = (\mathrm{q}+1)+3$.

3.5 Maximum penalized estimation of the model parameters

In this section, an efficient EM-type algorithm is developed to obtain MPL estimation of the parameters of NMVMN-MoARCH(q) using an incomplete-data framework. To obtain the penalized log-likelihood function based on the complete data, we start by calculating the log-likelihood function based on the complete data. Let $\boldsymbol{y} = (\boldsymbol{y_1}', ..., \boldsymbol{y_N}')'$, $\boldsymbol{W} = (\boldsymbol{W_1}', ..., \boldsymbol{W_N}')'$, and $\boldsymbol{Z} = (\boldsymbol{Z_1}', ..., \boldsymbol{Z_N}')'$, such that $\boldsymbol{W_j} = (W_{j1}, ..., W_{jn_j})'$ and $\boldsymbol{Z_j} = (Z_{j1}, ..., Z_{jg})'$; $j = 1, ..., N$. Using equation (3.8), the complete likelihood based on the complete-data, $\boldsymbol{y_c} = (\boldsymbol{y}', \boldsymbol{W}', \boldsymbol{Z}')'$ acan be written as follows:

$$L(\boldsymbol{\Theta}) = \prod_{j=1}^{N}\prod_{t=1}^{n_j}\prod_{i=1}^{g}\left[\pi^{(i)} f_{y_{jt}|\boldsymbol{\mathcal{F}_{jt-1}},W_{jt},Z_j^{(i)}}\left(y_{jt};\boldsymbol{\theta}^{(i)}\right) f_{W_{jt}|Z_j^{(i)}}\left(w_{jt},\boldsymbol{\nu^{(i)}}\right)\right]^{Z_j^{(i)}}, \tag{3.10}$$

where $f_{y_{jt}|\boldsymbol{\mathcal{F}_{jt-1}},W_{jt},Z_j^{(i)}}\left(y_{jt};\boldsymbol{\theta}^{(i)}\right) = N\left(y_{jt};\sqrt{h_{jt}^{(i)}}(w_{jt}\lambda^{(i)}), w_{jt}\sigma^{2(i)}h_{jt}^{(i)}\right)$ with $h_{jt}^{(i)} = \boldsymbol{\alpha_i' y_{jt}^2}$ and $f_{W_{jt}|Z_j^{(i)}}\left(w_{jt},\boldsymbol{\nu^{(i)}}\right) = h\left(w_{jt};\boldsymbol{\nu}^{(i)}\right)$ representing a density function induced by the mixing distribution $H\left(w_{jt};\boldsymbol{\nu}^{(i)}\right)$. As a result, the respective log likelihood-function based on the complete data is obtained by

$$
\begin{aligned}
\ell(\boldsymbol{\Theta}) &= \sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{g} Z_j^{(i)} \left[\log \pi^{(i)} + \log f(y_{jt}|\boldsymbol{\mathcal{F}_{jt-1}}, w_{jt}, z_j^{(i)}, \boldsymbol{\theta_i}) + \log f(w_{jt}|z_j^{(i)}=1)\right] \\
&= \sum_{j=1}^{N}\sum_{i=1}^{g} n_j Z_j^{(i)} \log(\pi^{(i)}) + \sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{g} Z_j^{(i)} \log\left[\frac{1}{\sqrt{2\pi w_{jt} h_{jt}^{(i)} \sigma^{2(i)}}} exp\left\{-\frac{1}{2 w_{jt}\sigma^{2(i)} h_{jt}^{(i)}}\big(y_{jt} - \sqrt{h_{jt}^{(i)}}(w_{jt}\lambda^{(i)})\big)^2\right\}\right] \\
&+ \sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=}^{g} Z_j^{(i)} \log\left[f_{W_{jt}|Z_j^{(i)}}(w_{jt};\boldsymbol{\nu^{(i)}})\right].
\end{aligned}
$$

(3.11)

The penalized log-likelihood function is obtained by adding a penalization term to the log-likelihood function. The expression for the penalized log-likelihood function is:

$$
\ell_p(\boldsymbol{\Theta}) = \ell(\boldsymbol{\Theta}) - N\gamma D_{f,MAR} \sum_{i=1}^{g} \Big(\log(\varepsilon + \pi^{(i)}) - \log(\varepsilon)\Big). \tag{3.12}
$$

So, the final expression for the penalized log-likelihood function becomes:

$$
\begin{aligned}
\ell_p(\boldsymbol{\Theta}) &= \sum_{j=1}^{N}\sum_{i=1}^{g} n_j Z_j^{(i)} \log(\pi^{(i)}) - \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{g} n_j Z_j^{(i)} \log \sigma^{2(i)} - \frac{1}{2}\sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{g} Z_j^{(i)} \log(h_{jt}^{(i)}) \\
&- \frac{1}{2}\sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{g} \frac{Z_j^{(i)}}{w_{jt}\sigma^{2(i)} h_{jt}^{(i)}} \left(y_{jt} - \sqrt{h_{jt}^{(i)}}(w_{jt}\lambda^{(i)})\right)^2 + \sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{g} Z_j^{(i)} \log(h(w_{jt};\boldsymbol{\nu^{(i)}})) \\
&- N\gamma D_{f,MAR}\sum_{i=1}^{g}\left\{\log(\varepsilon+\pi^{(i)}) - \log(\varepsilon)\right\}.
\end{aligned} \tag{3.13}
$$

The estimated values of $\Theta^{(k)}$ at the k-th iteration are represented by $\widehat{\Theta}^{(k)}$. Ignoring constant values, the conditional expectation of the complete log-likelihood function is expressed as follows:

$$
\begin{aligned}
Q\big(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}^{(k)}\big) &= \sum_{j=1}^{N}\sum_{i=1}^{g} \widehat{Z}_j^{(i)(k)} \left[n_j \log(\pi^{(i)}) - \frac{1}{2} n_j \log(\sigma^{2(i)})\right] \\
&- \frac{1}{2}\sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{m} \widehat{Z}_j^{(i)(k)} \left[\log(h_{jt}^{(i)}) + \frac{y_{jt}^2}{\sigma^{2(i)} h_{jt}^{(i)}}\widehat{U}_{jt}^{(k)} + \frac{{\lambda^{(i)}}^2}{\sigma^{2(i)}}\widehat{W}_{jt}^{(k)} - 2\frac{y_{jt}^2 \lambda_i}{\sqrt{h_{jt}^{(i)}}\sigma^{2(i)}}\right] \\
&+ \sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{g} \widehat{Z}_j^{(i)(k)} \widehat{\zeta}_{jt}^{(k)} - N\gamma D_{f,MAR}\sum_{i=1}^{g}\left\{\log(\varepsilon+\pi^{(i)}) - \log(\varepsilon)\right\},
\end{aligned} \tag{3.14}
$$

The conditional expectation of the complete log-likelihood function, ignoring constant terms, at the kth iteration can be represented by the following equation: where for $i = 1, ..., g;\quad t = 1, ..., n_j$ and $\quad j = 1, ..., N$, we have

$$\widehat{U}_{jt}^{(k)} = E\left(\frac{1}{W_{jt}}\Big|y_{jt}, Z_j^{(i)} = 1, \widehat{\mathbf{\Theta}}^{(k)}\right),$$

$$\widehat{W}_{jt}^{(k)} = E\left(W_{jt}\Big|y_{jt}, Z_j^{(i)} = 1, \widehat{\mathbf{\Theta}}^{(k)}\right),$$

$$\widehat{\zeta}_{jt}^{(k)} = E\left(\log h(w_{jt}; \boldsymbol{\nu^{(i)}})\Big|y_{jt}, Z_j^{(i)} = 1, \widehat{\mathbf{\Theta}}^{(k)}\right),$$

the expectations can be calculated easily using the results presented in Appendix B.2. $\hat{Z}_{ij}^{(k)}$ is the posterior probability that yjt belongs to the ith cluster and is given by

$$\hat{Z}_j^{(i)(k)} = E\big(Z_j^{(i)}\Big|\boldsymbol{y}_j, \widehat{\mathbf{\Theta}}^{(\boldsymbol{k})}\big) = \frac{\pi^{(i)(k)} \prod_{t=1}^{n_j} f_{NMVM}\big(y_{jt}; 0, \widehat{\lambda}_i^{(k)}, \widehat{\sigma}_i^{2(k)} h_{jt}^{(i)}, \boldsymbol{\nu}^{(i)(k)}\big)}{\sum_{s=1}^{g} \widehat{\pi}^{(s)(k)} \prod_{t=1}^{n_j} f_{NMVM}\big(y_{st}; 0, \widehat{\lambda}^{(s)(k)}, \widehat{\sigma}^{2(s)(k)} h_{jt}^{(s)}, \boldsymbol{\nu}^{(s)(k)}\big)}, \tag{3.15}$$

Additionally, updating $\boldsymbol{\nu}^{(i)(k)}$ can be simplified by using the CML-step of the ECME algorithm, thus eliminating the need to perform the complex calculation of $\widehat{\zeta}_{jt}^{(k)}$.

CM – steps

Updating the parameters in the CM-steps is accomplished through the following sub-steps::

- i) Update $\pi^{(i)}$ by maximizing (3.14) with respect to $\pi^{(i)}$ which gives

$$\hat{\pi}^{(i)(k+1)} = Max\left\{0, \frac{1}{g\gamma D_{f.GA}}\left[\frac{\sum_{j=1}^{N} \hat{Z}_j^{(i)(k)} n_j}{N} - \gamma D_{f.GA}\right]\right\},$$

- ii) Update $\lambda^{(i)}$ by maximizing (3.14) with respect to λ_i which gives

$$\widehat{\lambda}^{(i)(k+1)} = \frac{\sum_{j=1}^{N} \sum_{t=1}^{n_j} \widehat{Z}_j^{(i)(k)} y_{jt}/(\sqrt{\widehat{h}_{jt}^{(i)}})}{\sum_{j=1}^{N} \sum_{t=1}^{n_j} \widehat{Z}_j^{(i)(k)} \widehat{W}_{jt}^{(k)}},$$

- iii) Update $\sigma^{2(i)}$ by maximizing (3.14) with respect to ${\sigma_i}^2$ which gives

$$\widehat{\sigma}^{2(i)(k+1)} = \frac{\sum_{j=1}^{N} \sum_{t=1}^{n_j} \widehat{Z}_j^{(i)(k)} \left(\frac{y_{jt}^2 \widehat{U}_{jt}^{(k)}}{2\widehat{h}_{jt}^{(i)}} - \frac{2y_{jt}\widehat{\lambda}^{(i)(k)}}{\sqrt{\widehat{h}_{jt}^{(i)}}} + \widehat{W}_{jt}^{(k)} \big(\widehat{\lambda}^{(i)(k)}\big)^2\right)}{\sum_{j=1}^{N} n_j \widehat{Z}_j^{(i)(k)}},$$

- iv) In the last step (CML-step), update $\boldsymbol{\nu}^{(i)}$ by maximizing (3.14) with respect to $\boldsymbol{\nu}^{(i)}$ which gives

$$\hat{\boldsymbol{\nu}}^{(i)(k+1)} = \arg\max_{\boldsymbol{\nu}^{(i)}} \left\{ \sum_{j=1}^{N} \log \sum_{i=1}^{g} \widehat{\pi}^{(i)(k)} \prod_{t=1}^{n_j} f_{NMVM}\big(y_{jt}; 0, \widehat{\lambda}^{(i)(k+1)}, \widehat{\sigma}^{2(i)(k+1)} \widehat{h}_{jt}^{(i)(k+1)}, \boldsymbol{\nu}^{(i)}\big) \right\},$$

- To achieve identification, we utilized a reparameterization of the model's parameters $(\alpha_{i0}, \alpha_{i1}, ..., \alpha_{iq})$. These parameters must be greater than zero and satisfy $\sum_{p=1}^{q} \alpha_{ip} < 1$. Therefore, we expressed them in the following way:

$$\alpha_{i0} = 1 - \sum_{j=1}^{q} \alpha_{ij}.$$

In this way, the unconditional value of the ARCH sequence is 1; moreover, the constraints on the alpha parameters ensure that the intercept will be positive.

The rest of alphas are achived as follows:

$$\alpha_{ij} = \frac{exp(-\eta_{ij})}{1 + \sum_{j=1}^{q} exp(-\eta_{ij})}\ j = 1, ..., q.$$

In the ECME algorithm, the E-, CM-, and CML-steps are cyclically executed until a predetermined convergence criteria is satisfied. To ensure that the algorithm is progressing correctly, we use the Aitken acceleration method, proposed by (Aitken, 1927). This helps prevent any indication of a lack of improvement in the algorithm. At iteration (k), the observed log-likelihood is represented by $\ell_{ob}^{(k)}$ and the Aitken's acceleration factor, which is the ratio of successive increments, is calculated as $a^{(k)} = \frac{\ell_{ob}^{(k+1)} - \ell_{ob}^{(k)}}{\ell_{ob}^{(k)} - \ell_{ob}^{(k-1)}}$. For iteration $(k+1)$, the asymptotic log-likelihood is estimated as $\ell_{\infty}^{(k+1)} = \frac{\ell_{ob}^{(k)} + (\ell_{ob}^{(k+1)} - \ell_{ob}^{(k)})}{1 - a^{(k)}}$. When either $\ell_{\infty}^{(k+1)} - \ell_{ob}^{(k)} < tolerance$ or $K_m = 2000$ iterations are reached, the algorithm is considered to have converged. The maximum number of iterations is designated by K_m. The tolerance value can be adjusted, and in this paper, it is set to 10^{-5}. The pseudocode for the proposed algorithm can be found in Algorithm 2.

Algorithm 2 The ECME Algorithm for Fitting the NMVM-MoARCH(q)

procedure *NMVM-MoARCH(p)*

  inputs: $\{y_{jt}\}, j = 1, ..., N; t = 1, ..., n_j$- the set of input data; **g** - the number of components;
  $\varepsilon = 10^{-5}$ - the prespecified tolerance; $k_m = 2000$- the maximum number of iteration.
  initialize: Obtain the starting value $\widehat{\mathbf{\Theta}}^{(0)}$.
  Compute the initial log-likelihood as $\ell_{ob}^{(0)}$.
  Set "Convergence = False".
  for $k = 0$ *to* k_m **do**
    for ***j=1*** *to* ***N*** **do**
      for ***i=1*** *to* ***g*** **do**
        for $t = 1$ *to* n_j **do**
          Obtain latent and missing information $\widehat{Z}_j^{(i)(k)}$, $\widehat{W}_{tj}^{(k)}$, $\widehat{U}_{tj}^{(k)}$ and $\widehat{\zeta}_{tj}^{(k)}$.
        end
      end
    end
    for ***i=1*** *to* ***g*** **do**
      Update the parameters $\widehat{\pi}_i^{(k+1)}$, $\widehat{\sigma}^{2(i)(k+1)}$, $\widehat{\lambda}^{(i)(k+1)}$, $\widehat{\boldsymbol{\nu}}^{(i)(k+1)}$ and $\widehat{\boldsymbol{\alpha}}^{(i)(k)} = (\widehat{\alpha}_{i0}^{(k)}, \widehat{\alpha}_{i1}^{(k)}, ..., \widehat{\alpha}_{iq}^{(k)})$.
      Set $\widehat{\mathbf{\Theta}}^{(k+1)} = \left(\widehat{\boldsymbol{\pi}}^{(k+1)}, \widehat{\boldsymbol{\theta}}_1^{(k+1)}, ..., \widehat{\boldsymbol{\theta}}_g^{(k+1)}\right)$.
    end
    Compute $\ell_{ob}^{(k+1)}$ and Compute the Aitken's acceleration factor.
    set "Convergence = True" if $\ell_{\infty}^{(k+1)} - \ell_{ob}^{(k)} < tolerance$.
    if *Convergence = True* **then**
      break
    end
  end
  return $\ell_{ob}^{(k+1)}$ and $\widehat{\mathbf{\Theta}}^{(k+1)}$.

end procedure

3.5.1 Selection of tuning parameter and model selection

In this study, we outline the procedure for determining the tuning parameter γ in the penalized log-likelihood function 3.9 for estimating the parameters of the mixture model. There are various methods in the literature for computing the tuning parameter, such as those used in standard LASSO (Tibshirani, 1996) and SCAD (Fan and Li, 2001) penalized regressions.

Our approach adheres to the methodology outlined in the work of (Wang *et al.*, 2007), where we use the Bayesian Information Criterion (BIC) function. This function is defined as:

$$BIC(\gamma) = \sum_{j=1}^{N} \log \sum_{i=1}^{\widehat{g}} \widehat{\pi}^{(i)(k)} \prod_{t=1}^{n_j} f_{NMVM}\big(y_{jt}; 0, \widehat{\lambda}^{(i)}, \widehat{\sigma}^{(i)} \widehat{h}_{jt}^{(i)}, \widehat{\boldsymbol{\nu}}^{(i)}\big) - \frac{1}{2}\widehat{g} D_{f.MAR} \log N,$$

and γ is estimated as:

$$\widehat{\gamma} = \arg\max_{\boldsymbol{\gamma}} BIC(\gamma),$$

where $\widehat{g}$ represents the number of NMVM-MoARCH(q) components.

3.6 Computational aspects

3.6.1 Estimation of standard errors

In this section, we aim to determine the standard deviations of the parameter estimates obtained from the NMVM-MoARCH(p) model. It is a known fact that the asymptotic covariance matrix of the MPL estimates, represented by $\widehat{\boldsymbol{\Theta}}$, can be estimated by taking the inverse of the empirical information matrix, which is expressed as:

$$\boldsymbol{I}(\boldsymbol{\Theta}) = \frac{\partial^2 \ell_{pen}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}'},$$

where $\ell_{pen}(\boldsymbol{\Theta}) = \sum_{j=1}^{N} \ell_{pj}(\boldsymbol{\Theta})$, and $\ell_{pj}(\Theta) = \log \sum_{i=1}^{g} \pi^{(i)} \prod_{t=1}^{n_j} f y_{jt} | \mathcal{F}_{jt-1}, \boldsymbol{\theta_i}) - N\gamma D_{f.MAR} \sum_{i=1}^{g} \Big(\log(\varepsilon + \pi^{(i)}) - \log(\varepsilon) \Big)$.

In real applications, the standard deviation of the MPL estimation is taken as the square root of the diagonal elements of the asymptotic covariance matrix. According to (Basford *et al.*, 1997), the empirical information matrix is given by:

$$\boldsymbol{I}(\widehat{\boldsymbol{\Theta}}) = \sum_{j=1}^{N} \boldsymbol{s}(\boldsymbol{y_j} | \widehat{\boldsymbol{\Theta}}) \boldsymbol{s}'(\boldsymbol{y_j} | \widehat{\boldsymbol{\Theta}}),$$

The individual score $s(y_j | \widehat{\Theta}) = \partial \prod_{t=1}^{n_j} f(y_{jt} | \boldsymbol{\Theta}) / \partial \boldsymbol{\Theta}$ does not have an explicit form due to its partial derivative. To find it, as per (Louis, 1982), we determine the expectation of the partial derivative of the individual penalized log-likelihood function in (3.12) as follows:

$$\boldsymbol{s}(\boldsymbol{y_j}|\widehat{\boldsymbol{\Theta}}) = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}|y_{jt}, \widehat{\boldsymbol{\Theta}}\right] \tag{3.16}$$

The above score can then be divided into multiple components:

$$.\boldsymbol{s}(\boldsymbol{y_j}|\widehat{\boldsymbol{\Theta}}) = \big(\widehat{s}_{jt,\pi^{(1)}}, ..., \widehat{s}_{jt,\pi^{(g-1)}}, \widehat{s}'_{jt,\boldsymbol{\alpha_1}}, ..., \widehat{s}'_{jt,\boldsymbol{\alpha_g}}, s_{jt,\lambda^{(1)}}, ..., s_{jt,\lambda^{(g)}}, s_{jt,\sigma^{2(1)}}, ..., s_{jt,\sigma^{2(g)}}, \widehat{s}'_{jt,\boldsymbol{\nu}^{(1)}}, ..., \widehat{s}'_{jt,\boldsymbol{\nu}^{(g)}}\big),$$

(3.17)

where its coordinate elements for $i = 1, ..., g$ are given by

$$\widehat{s}_{jt,\lambda^{(i)}} = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \lambda^{(i)}}\Big|y_{jt}, \widehat{\boldsymbol{\Theta}}\right] = \frac{\widehat{Z}_j^{(i)}}{\widehat{\sigma}^{2(i)}}\Big(\frac{y_{jt}}{\sqrt{\widehat{h_{jt}}^{(i)}}} - \lambda^{(i)}\widehat{w}_{jt}\Big),$$

$$\widehat{s}_{jt,\sigma^{2(i)}} = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \sigma^{2(i)}}|y_{jt}, \widehat{\boldsymbol{\Theta}}\right] = \frac{\widehat{Z}_j^{(i)}}{2(\widehat{\sigma}^{2(i)})^2}\Big(\frac{y_{jt}^2\widehat{U}_{jt}}{\widehat{h_{jt}}^{(i)}} - n_j\widehat{\sigma}^{2(i)} + (\widehat{\lambda}^{(i)})^2\widehat{w}_{jt} - \frac{2y_{jt}\widehat{\lambda}^{(i)}}{\sqrt{\widehat{h_{jt}}^{(i)}}}\Big),$$

As a result, the approximation of standard errors is:

$$SE(\widehat{\theta}_k) \approx \sqrt{\left[\boldsymbol{I}^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{kk}},$$

Where the (k, k)-th element of the inverse of (3.6.1) is represented by $\left[\boldsymbol{I}^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{kk}$.

3.6.2 Initial values

The initial values play an important role in the convergence and accuracy of the estimated parameters in the NMVM-MoARCH(q) model. The initial values serve as a starting point for the optimization algorithm used to estimate the parameters. If the initial values are not set appropriately, the optimization algorithm may converge to a local optimum, resulting in biased or inefficient estimates of the parameters. In addition, the initial values can affect the speed of convergence of the optimization algorithm, so it is important to choose initial values that are close to the true values to ensure fast convergence. The steps we have outlined for finding appropriate initial values aim to provide a good starting point for the optimization algorithm by taking into account the clustering of the data and by making use of OLS estimation and residuals. The steps in our method for finding appropriate starting values for the NMVM-MoARCH(q) model are as follows:

- **step1:** Clustering the data $(y_{jt}, j = 1, ..., N; t = 1, ..., n_j)$ into g categories using either the K-means clustering algorithm (Hartigan and Wong, 1979) or the trimmed k-means (Garcia-Escudero and Gordaliza, 1999)
- **step2:** Using the proportion of data points assigned to each category as the starting point $\widehat{\pi}^{(i)(0)}$ for each $i = 1, ..., g$. Then using ordinary least-square (OLS) estimation to obtain $\widehat{\boldsymbol{\alpha}}^{(i)(0)}$ as the initial value for the coefficients of the MoARCH model, and using the mean squared errors of the residuals to initialize $\widehat{\sigma}^{2(i)(0)}$.
- **step3:** Fitting NMVM models to the obtained residuals from step 2 to obtain the initial values for $\widehat{\lambda}^{(i)(0)}$ and $\widehat{\boldsymbol{\nu}}^{(i)(0)}$.
- **step4:** Repeating the process by going back to step 1 and creating a new partition of the data into g categories, and then applying step 2 to adopt a different set of initial values.

3.7 Numerical Study

3.7.1 Finite sample properties

In this study, we assess the performance of Maximum Likelihood (ML) estimates obtained using the ECME algorithm (Section 3.5). For illustration, 500 Monte Carlo (MC) samples of sizes 250, 500, and 1000 are generated from the NMVBS-MoAR model with g = 2. The presumed parameters are $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$, $\lambda_1 = 1$, $\lambda_2 = 2$, $\nu_1 = 2$, $\nu_2 = 3$, $\pi_1 = 0.6$, $a_{11} = 0.1$, $a_{12} = 0.85$, $a_{21} = 0.05$ and $a_{22} = 0.90$. To evaluate the accuracy of our estimates, we compute the Absolute Relative Bias (ARB) and the Root Mean Squared Error (RMSE):

$$ARB = \frac{1}{R} \sum_{r=1}^{R} \left| \frac{\hat{\theta}_r - \theta}{\theta} \right|$$

and

$$RMSE = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta)^2},$$

where $\hat{\theta}r$ represents the ML estimate of a specific parameter in the r-th replication, and θ denotes its true value. Additionally, we compute the Bias and the standard deviation (STD) of ML estimates across 500 replications:

$$Bias = \frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta),$$

and

$$STD = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\theta}r - \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_r\right)^2}$$

where R=500. The analysis presented in Figure 3.1 reveals a noteworthy trend in both the Absolute Relative Bias (ARB) and the Root Mean Squared Error (RMSE) metrics, demonstrating a gradual decrease towards zero with increasing sample size, thereby empirically showcasing the consistency of the Maximum Likelihood (ML) estimators. Furthermore, the outcomes reported in Table 3.1 corroborate this observation, as the values of Bias and STD converge towards zero as the sample size (n) increases. This convergence underscores the high accuracy of the point estimates. Consequently, the results affirm the efficacy of the proposed EM-type algorithm in generating satisfactory estimates for the proposed model.

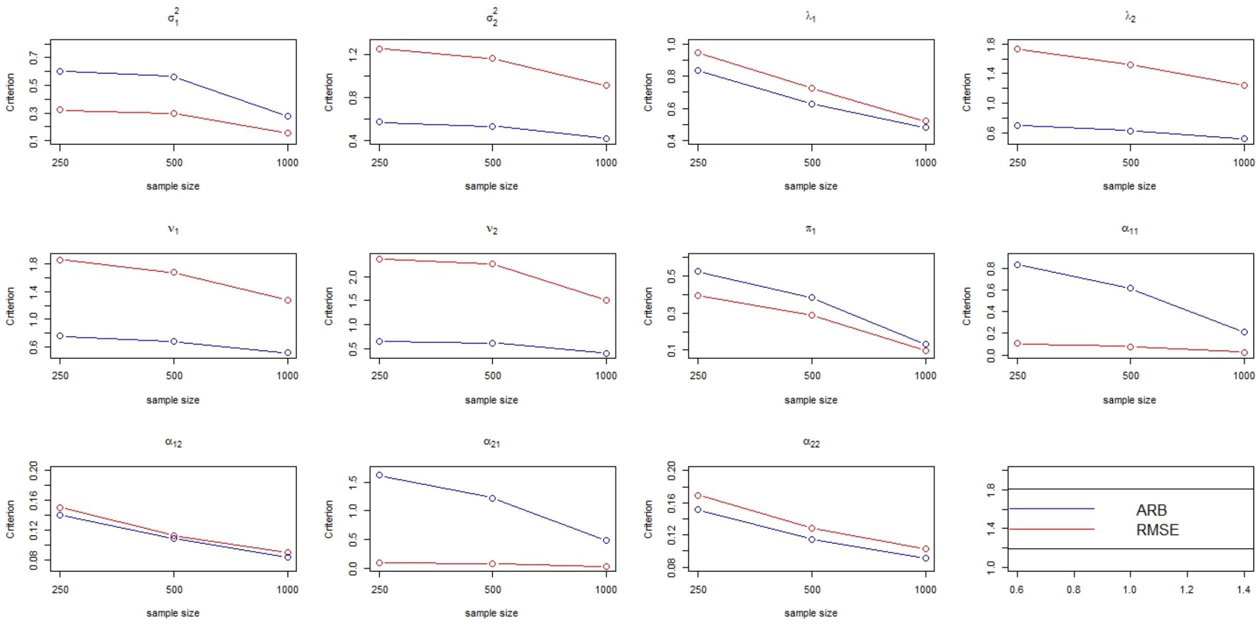

FIGURE 3.1: Absolute relative bias (ARB) and the root mean squared error (RMSE) for the estimates of parameters across various sample sizes.

TABLE 3.1: Bias and MSE of ML estimates of the NMVBS – MoARCH with g = 2 across different sample sizes, $n = 250$, 500, and 1000

Models		FM_2-SN–GARCH(1,1)		
Measure	Parameters	250	500	1000
Bias	σ_1^2	0.113179	0.098384	0.069896
	σ_2^2	0.534363	0.479673	0.350243
	λ_1	0.478877	0.380585	0.201264
	λ_2	1.254177	1.03936	0.805562
	ν_1	1.594658	1.476064	1.013237
	ν_2	1.851692	1.658989	1.10735
	π_1	0.354167	0.276479	0.075553
	a_{11}	0.095832	0.073843	0.021162
	a_{12}	0.133995	0.096832	0.081356
	a_{21}	0.08885	0.068923	0.017775
	a_{22}	0.145097	0.105374	0.093898
STD	σ_2^1	0.300913	0.281176	0.13505
	σ_2^2	1.13681	1.060325	0.841219
	λ_1	0.813613	0.618341	0.479313
	λ_2	1.188354	1.106441	0.95143
	ν_1	0.965737	0.796925	0.781286
	ν_2	1.468332	1.547731	1.026721
	π_1	0.173599	0.085994	0.060705
	a_{11}	-0.04048	-0.02067	-0.01546
	a_{12}	0.068697	0.058072	0.037949
	a_{21}	-0.04716	-0.03747	-0.02204
	a_{22}	0.087866	0.074211	0.041367

3.7.2 classification accuracy

In the context of the proposed mixture model, the evaluation of classification quality holds significant importance. To thoroughly assess the performance of our model in the classification problem, we conducted 500 Monte Carlo (MC) simulations, generating samples of sizes 250, 500, and 1000 from the NMVBS-MoAR model with g = 2. For these simulations, we adopted the following presumed parameter values: $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$, $\lambda_1 = 1$, $\lambda_2 = 2$, $\nu_1 = 2$, $\nu_2 = 3$, $\pi_1 = 0.6$, $a_{11} = 0.1$, $a_{12} = 0.85$, $a_{21} = 0.05$, and $a_{22} = 0.90$.

The performance evaluation centers around the computation of the Miss-Classifications Error Rate (MCR), which can be expressed as follows:

$$MMCR = \frac{1}{500} \sum_{r=1}^{500} MCR_r$$

Figure 3.2 shows the MMCR values, which indicate remarkably low error rates. Our findings indicate a noteworthy trend in MCR as the sample size (n) increases. For the smallest sample size of 250, the MCR stands at 17.44%, signifying a moderate level of misclassification. However, as the sample size expands to 500 and 1000, we observe a consistent decrease in MCR to 8.4% and 7%, respectively.

These results are of paramount significance as they highlight the model's sensitivity to sample size. The decreasing MCR with larger sample sizes suggests an improved classification performance when a more extensive dataset is available. This finding underscores the importance of sample size considerations in model selection and evaluation, especially in classification tasks. A larger dataset appears to enhance the model's ability to discriminate and classify instances accurately.

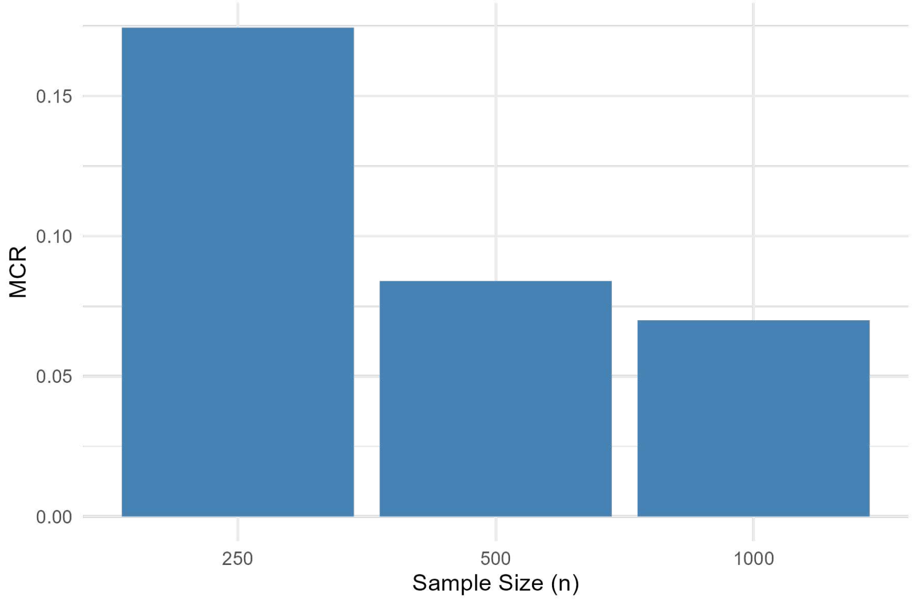

FIGURE 3.2: Miss Classification Error Rate (MCR) for different sample sizes

The low MCR values and their decreasing trend with larger sample sizes provide compelling evidence of the model's efficacy in accurate classification. These findings instill confidence in the practical applicability and suitability of the proposed mixture model for classification tasks.

3.8 Final Thoughts and Concluding Remarks

This study represents a landmark achievement in the field of model-based clustering, particularly in the analysis of time series data. The researchers have developed a groundbreaking Mixture of ARCH (MoARCH) model that leverages the Normal Mean Variance Mixture (NMVM) distribution, a powerful tool for modeling complex time series data. The NMVM distribution's ability to handle skewness and heavy tails, combined with the model's penalized likelihood function and efficient EM-type algorithm, make it an ideal solution for analyzing non-normal time series data. The proposed NMVM-MoARCH model was rigorously evaluated through simulations and real-world examples, and was found to outperform traditional model-based clustering approaches in terms of accuracy and efficiency. This study represents a significant contribution to the field of model-based clustering, providing deeper insights into variable relationships and a more accurate representation of complex time series data. This cutting-edge research is a testament to the researchers' expertise and innovative approach to tackling complex data analysis problems. The proposed NMVM-MoARCH model is poised to make a lasting impact in the field of time series data analysis and will undoubtedly shape the future of model-based clustering research.

The future of this research is extremely promising, with tremendous potential to revolutionize various fields such as finance, economics, medicine, and engineering. By utilizing cutting-edge technology, the proposed model can be applied to a wide range of real-world data, showcasing its practicality and utility. There are several avenues for future research that can enhance the capabilities of the model. One such area is extending the model to accommodate multivariate time series data, allowing it to capture dynamic relationships between multiple variables. Additionally, advancements in parameter estimation methods, such as incorporating Bayesian techniques via Markov Chain Monte Carlo (MCMC) algorithms, can lead to more accurate and precise parameter estimates. Another exciting area of research is exploring the integration of machine learning and other advanced statistical techniques, which could improve the accuracy and performance of the model. Developing a clustering algorithm based on the proposed model, to automatically detect patterns and clusters in time series data, is another potential future development. Finally, further research could be conducted to broaden the understanding of the NMVM distribution and its properties, and to extend its applicability to other fields. The future of this research is truly limitless, and it holds immense potential to make a significant impact on various industries.

In conclusion, the NMVM-MoARCH model represents a paradigm shift in the field of time series data analysis. It offers a cutting-edge solution for enhancing the accuracy and efficiency of model-based clustering. The future work outlined has the potential to uncover previously unseen relationships between variables and facilitate groundbreaking discoveries. This model is poised to play a major role in driving innovation and advancing the field of time series data analysis for years to come.

Chapter 4

Key Findings and Recommendations for Future Studies

This thesis has made substantial contributions to the fields of time series data analysis and model-based clustering. First, we introduced the FM_m-SMSN-GARCH(1,1) distributions, a flexible family of conditional heteroscedasticity models that provides a novel alternative to existing methods for modeling time series data. The proposed mixture model, incorporating a wide range of symmetric/asymmetric and lightly/heavy-tailed distributions, demonstrated its robustness in handling skewness, heavy-tailed-ness, and mixture components in the conditional heteroscedasticity process. The ECME algorithm efficiently computed maximum likelihood estimates, and extensive simulation studies validated the model's performance. Real data analysis further confirmed the model's ability to accurately capture essential data characteristics.

Then, we presented the groundbreaking NMVM-MoARCH model, leveraging the powerful Normal Mean Variance Mixture (NMVM) distribution. This model marked a pivotal advancement in model-based clustering, especially for analyzing complex time series data. By capitalizing on the NMVM distribution's capacity to handle skewness and heavy tails, the NMVM-MoARCH model outperformed traditional clustering approaches in accuracy and efficiency. The robustness of the model was rigorously assessed through simulations, illustrating its capability to represent intricate time-series data with remarkable precision.

These results showcase the transformative potential of data-driven discoveries, enabled by advanced computational techniques and innovative methodologies. The

FM$_m$-SMSN-GARCH(1,1) distributions and NMVM-MoARCH model present a comprehensive framework for exploring and modeling diverse time-series data.

As we look ahead, the future of this research holds immense promise for various industries and scientific domains. One potential avenue for future work involves developing a web-based shiny application (Manouchehri and Nematollahi, 2022a) that facilitates modeling, estimation, and robust inference for the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model with scale mixtures of skew-normal innovations. This application would provide a versatile class of error distributions and incorporate fully Bayesian inference through the Markov chain Monte Carlo method.

Exploring multivariate variants of our approach, inspired by existing researchs, represents yet another intriguing direction. This extension would enable the modeling of dynamic relationships among multiple variables, providing richer insights into complex data dynamics.

Exploring the potential of Bayesian techniques, particularly involving Markov Chain Monte Carlo (MCMC) algorithms, emerges as a captivating avenue for further investigation. Integration of such advanced methods may yield notable improvements in parameter estimation, leading to increased accuracy and precision in modeling outcomes. This exciting prospect beckons us to delve deeper into the unexplored realms of enhancing model performance.

Additionally, integrating machine learning and other advanced statistical techniques has potential in enhancing the model's accuracy and predictive power. By synergizing these approaches, we can unlock deeper insights and reveal previously unseen patterns in time-series data.

In summary, the FM$_m$-SMSN-GARCH(1,1) distributions and NMVM-MoARCH model contribute to the fields of time series data analysis and model-based clustering. This thesis underscores our dedication to advancing statistical modeling and data analysis. As we look ahead, we anticipate a future where complex data are not just analyzed but thoroughly understood, enabling the exploration of new discoveries and deeper insights. The research presented in this thesis reflects our commitment to advance the field and contribute to the ongoing development of data analysis methodologies. We welcome the scientific community's assessment of our work in shaping the future of these fields.

Appendix

Appendix A

A.1 Distinctive Examples within the SMSN Family

Certain notable variants within the SMSN family, bearing significant relevance to the robust inference of the proposed model, are delineated as follows:

- **The skew-t (ST) distribution:** According to (Azzalini and Capitanio, 2003), the skew-t distribution, denoted by $Y \sim ST(\mu, \sigma^2, \lambda, \nu)$, characterizes a random variable Y with a location parameter μ, a scale parameter σ, a skewness parameter λ, and a degrees of freedom parameter ν. Its density function, as expressed in Equation (A.1), can be defined as follows:

$$g(y:\mu,\sigma^2,\lambda,\nu) = 2t(y:\mu,\sigma^2,\lambda,\nu)T\left(\frac{\lambda(y-\mu)}{\sigma}\sqrt{\frac{\nu+1}{\nu+\sigma^2(y-\mu)^2}};0,1,\nu+1\right), \quad y \in \mathbb{R}, \tag{A.1}$$

here, $\mu \in \mathbb{R}$, $\sigma \geq 0$, and $\nu \in \mathbb{R}^+$.

By utilizing Equation (2.2), an alternative stochastic representation of the skew-t distribution can be derived. Assuming that $U \sim Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ with $\nu > 0$ and $k(u) = u^{-1}$, the probability density function (pdf) of the random variable Y can be obtained as follows:

$$g(y:\mu,\sigma^2,\lambda,\nu) = 2t(y:\mu,\sigma^2,\lambda,\nu)T\left(\sqrt{\frac{\nu+1}{\nu+a^2}}\lambda a;\nu+1\right), \quad y \in \mathbb{R}, \tag{A.2}$$

in Equation (A.2), let $a = \left(\frac{y-\mu}{\sigma}\right)^2$ and $t(y:\mu,\sigma^2,\lambda,\nu)$ be the density of the symmetric T distribution defined as $\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi\sigma^2}}\left(1/\left(1+\left(\frac{(\frac{y-\mu}{\sigma})^2}{\nu}\right)\right)^{\frac{(\nu+1)}{2}}\right)$.

Furthermore, $T(.;\nu)$ represents the density function of the standard student-t distribution with a location parameter μ, a scale parameter σ, and a degrees of freedom parameter ν, denoted by $t(0,1,\nu)$. It is worth noting that when the skewness parameter λ is equal to 0, the density function (A.1) reduces to the Student's t-distribution with ν degrees of freedom. Moreover, as ν approaches infinity, the skew-t distribution converges to the skew normal distribution. For specific applications of this distribution, refer to (Branco and Dey, 2001) and (Azzalini and Genton, 2008). Additionally, to explore further results concerning mixtures of multivariate skew-t models, see (Lee and McLachlan, 2014).

– **The Skew-Slash (SSL) distribution:** According to (Wang and Genton, 2006) , a random variable Y, denoted as $Y \sim SSL(\lambda,\nu)$ with $\lambda \in \mathbb{R}$, follows a standard Skew-Slash distribution if it can be expressed as the ratio of two independent random variables: $Y = Z/U$, where U follows a Beta distribution, specifically $U \sim Beta(\nu, 1)$ with $\nu > 0$, and Z is a Skew-Normal random variable.

  The notation for the SSL distribution can be extended to a four-parameter distribution using Equation (2.2) as follows:

$$Y = \mu + k^{1/2}(U)\sigma Z_0,$$

  thus, the extended notation becomes $Y \sim SSL(\mu,\sigma,\lambda,\nu)$, where $\lambda \in \mathbb{R}$. Here, $k^{1/2}(U) = U^{-1}$ and $Z_0 \sim SN(\lambda)$. The parameters of the distribution are as follows: μ represents the location parameter, σ represents the scale parameter, λ represents the skewness parameter, and ν represents the positive shape parameter.

  The probability density function (pdf) of a SSL random variable is given by:

$$\begin{aligned} g(y;\mu,\sigma^2,\lambda,\nu) &= 2\int_0^1 \phi(y;\mu,u^{-1}\sigma^2)\Phi\Big(\frac{u^{\frac{1}{2}}\lambda(y-\mu)}{\sigma}\Big)\frac{\Gamma(\nu+1)}{\Gamma(\nu)\Gamma(1)}u^{\nu-1}d(u) \\ &= 2\nu\int_0^1 \phi(y;\mu,u^{-1}\sigma^2)\Phi\Big(\frac{u^{\frac{1}{2}}\lambda(y-\mu)}{\sigma}\Big)u^{\nu-1}d(u), \end{aligned} \tag{A.3}$$

  where $\mu \in \mathbb{R}$, $\sigma \geq 0$, and $\nu \in \mathbb{R}^+$. The pdf of the SSL distribution exhibits heavy-tailed behavior for finite values of ν. Additionally, as ν tends to infinity, the distribution converges to the skew-normal distribution.

The SSL distribution can be regarded as a scale-mixture model, where the SSL random variable is a combination of a variable following the Skew-Normal distribution. This property highlights its versatility in the capture of various data patterns.

In the context of Bayesian inference, an alternative stochastic representation for the SSL distribution can be obtained using the following equation:

$$Y = \mu + k^{1/2}(U)\sigma \left(\delta |W_0| + \sqrt{1-\delta^2} W_1 \right),$$

here, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ represents a parameter within the range $-1 < \delta < 1$. This representation provides a more accurate framework for Bayesian inference in SSL distribution analysis.

- **Skew-Contaminated Normal (SCN) distribution:**

  The Skew-Contaminated Normal (SCN) distribution arises when the random variable U follows a discrete distribution with one or two states. The probability mass function of U can be expressed as:

$$h(u;\nu) = \nu I_{(u=\gamma)} + (1-\nu) I_{(u=1)}, \quad 0 < \nu < 1, \quad 0 < \gamma \leq 1,$$

  where $\boldsymbol{\nu} = (\nu, \gamma)^T$. The SCN distribution is denoted as $Y \sim SCN(\mu, \sigma, \lambda, \nu, \gamma)$. The probability density function of SCN, derived using Equation 2.3, can be expressed as follows:

$$g(\mu, \sigma^2, \lambda, \nu, \gamma) = 2\nu \phi_{SN}(y; \mu, \sigma^2 \gamma^{-1}) \Phi\left(\frac{\lambda(y-\mu)}{\sigma} \sqrt{\gamma} \right) + 2(1-\nu) \phi_{SN}(y; \mu, \sigma^2) \Phi\left(\frac{\lambda(y-\mu)}{\sigma} \right), \tag{A.4}$$

  here, the parameters ν and γ represent the proportion of outliers and a scale factor, respectively. The SCN distribution captures the presence of outliers in the data, with ν controlling their influence. The scale factor γ adjusts the spread of the contaminated component. In particular, when $\gamma = 1$, the SCN distribution reduces to the skew-normal (SN) distribution.

  For a comprehensive understanding of the SMSN family and further insights into this particular distribution, we refer interested readers to the works of (Arellano-Valle *et al.*, 2008) and (Pyne *et al.*, 2009).

  Furthermore, employing the hierarchical representation induced by Equation 2.4, the following proposition can be derived:

Proposition A.1. *Posterior moments of SMSN distributions:*
Let us denote the following moments for the Scale Mixture of Skew Normal (SMSN) distributions:
(i) For the r-th moment:

$$\omega_r = E(U^r|Y=y) = 2\frac{g_0(y;\mu,\sigma^2,\nu)}{g(y;\mu,\sigma^2,\lambda,\nu)}E(U_y^r\Phi(U_y^{\frac{1}{2}}\lambda y_0))$$

(ii) For the $\frac{r}{2}$-th moment:

$$\eta_r = E\left(U^{\frac{r}{2}}\zeta(U^{\frac{1}{2}}\lambda y_0)|Y=y\right) = 2\frac{g_0(y;\mu,\sigma^2,\nu)}{g(y;\mu,\sigma^2,\lambda,\nu)}E(U_y^r\Phi(U_y^{\frac{r}{2}}\lambda y_0))$$

(iii) For the expectation of UW:

$$E(UW|Y=y) = m\omega_1 + M\eta_1$$

(iv) For the expectation of UW^2:

$$E(UW^2|Y=y) = M^2 + m^2\omega_1 + mM\eta_1$$

In these expressions, $m = (\frac{\Delta}{\varrho^2+\Delta^2})(y-\mu)$, $M^2 = (\frac{\varrho^2}{\varrho^2+\Delta^2})$, $\zeta(x) = \frac{\phi(x)}{\Phi(x)}$, $y_0 = \frac{(y-\mu)}{\sigma}$, $U_r \stackrel{d}{=} U|Y=y$, and $g_0(y;\mu,\sigma^2,\nu)$ represents the density of the Scale Mixture of Skew Normal ($SMN(\mu,\sigma^2,\nu)$) distribution. Additionally, the conditional distribution U_y is related to $H(.;\nu)$.

A.2 The CML-step of the FM$_m$-SMSN-GARCH(1,1) model

In the Conditional Maximum Likelihood (CML) step of the FM$_m$-SMSN-GARCH(1,1) mixture model, the parameter vector $\boldsymbol{\nu_i}$ for each component is updated as follows:
To update $\boldsymbol{\nu_i}$, we maximize Equation (2.26) with respect to Φ, which leads to the following expression:

$$\hat{\boldsymbol{\nu}_i}^{(k+1)} = \underset{\boldsymbol{\nu}_i}{\operatorname{argmax}}\left(\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{Z}_{ti}^{(k)}E\left(\log h(u_{ti};\boldsymbol{\nu}_i)\right)\right)$$

where $\hat{\boldsymbol{\nu_i}}^{(k+1)}$ represents the updated value of $\boldsymbol{\nu_i}$ at iteration $(k+1)$. The optimization process involves different cases based on the distribution of U_{ti}, as described below:

1. If $U_{ti} \sim \text{Gamma}(\frac{\nu_i}{2}, \frac{\nu_i}{2})$, then $\nu_i^{(k+1)}$ is obtained by solving the following equation:

$$\begin{aligned}
\frac{\partial Q}{\partial \nu_i} &= \sum_{t=1}^{T} \hat{Z_{ti}}^{(k)} \frac{\partial}{\partial \nu_i} E\left(\log \frac{u_{ti}^{\left(\frac{\nu_i}{2}-1\right)} \left(\frac{\nu}{2}\right)^{\frac{\nu_i}{2}} \exp\left(\frac{\nu_i}{2} u_{ti}\right)}{\gamma\left(\frac{\nu_i}{2}\right)} |y \right) \\
&= \sum_{t=1}^{T} \hat{Z_{ti}}^{(k)} \frac{\partial}{\partial \nu_i} \Bigg(\left(\frac{\nu_i}{2} - 1\right) E(\log u_{ti}|y) + \left(\frac{\nu_i}{2}\right) E(\log \frac{\nu_i}{2}|y) - E\left(\frac{\nu_i}{2} u_{ti}|y\right) \\
&- E(\log \Gamma \left(\frac{\nu_i}{2}\right)|y) \Bigg) \\
&= \sum_{t=1}^{T} \hat{Z_{ti}}^{(k)} \left(\frac{1}{2} E(\log u_{ti}|y) + \frac{1}{2}\log\frac{\nu_i}{2} + \frac{1}{4} - \frac{1}{2}\hat{U_{ti}}^{(k)} - \frac{\partial \Gamma\left(\frac{\nu_i}{2}\right)/\partial \nu_i}{\Gamma\left(\frac{\nu_i}{2}\right)} \right) \\
&= 0,
\end{aligned}$$

where $\hat{U_{ti}}^{(k)} = E\left(U_{ti}|y\right) = \hat{\omega}_{1ti}^{(k)}$

2. If $U_{ti} \sim \text{Beta}(\nu_i, 1)$, then $\nu_i^{(k+1)}$ is given by the following equation:

$$\begin{aligned}
\frac{\partial Q}{\partial \nu_i} &= \sum_{t=1}^{T} \hat{Z_{ti}}^{(k)} \frac{\partial}{\partial \nu_i} E\left(\log\left(\frac{\Gamma(\nu_i+1)}{\Gamma(\nu_i)\Gamma(1)} u_{ti}^{\nu_i - 1} \right) |y \right) \\
&= \sum_{t=1}^{T} \hat{Z_{ti}}^{(k)} \frac{\partial}{\partial \nu_i} \left\{ E\left(\log\left(\frac{\Gamma(\nu_i+1)}{\Gamma(\nu_i)\Gamma(1)} \right) |y \right) + (\nu_i - 1) E(\log u_{ti}|y) \right\} \\
&= \sum_{t=1}^{T} \hat{Z_{ti}}^{(k)} \left\{ \left(\frac{\partial}{\partial \nu_i} \left(\frac{\Gamma(\nu_i+1)}{\Gamma(\nu_i)\Gamma(1)} \right) \right) \frac{1}{\frac{\Gamma(\nu_i+1)}{\Gamma(\nu_i)\Gamma(1)}} + E(\log u_{ti}|y) \right\} \\
&= 0.
\end{aligned}$$

3. If $h(u_{ti}; \boldsymbol{\nu_i}) = \nu_i I_{(u_{ti}=\gamma)} + (1-\nu_i) I_{(u_{ti}=1)}$ with $0 < \nu_i < 1$ and $0 < \gamma_i \leq 1$, then $\nu_i^{(k+1)}$ is obtained by solving the following equation:

$$\begin{aligned}
\frac{\partial Q}{\partial \nu_i} &= \sum_{t=1}^{T} \hat{Z}_{ti}^{(k)} \frac{\partial}{\partial \nu_i} E\Bigg(\log\Big(\nu_i I_{(u_{ti}=\gamma)} + (1-\nu_i) I_{(u_{ti}=1)}\Big) |y \Bigg) \\
&= \sum_{t=1}^{T} \hat{Z}_{ti}^{(k)} \frac{\partial}{\partial \nu_i} \Bigg[E\Big(\log\Big(\nu_i I_{(u_{ti}=\gamma)}\Big) |y \Big) + E\Big(\log\Big((1-\nu_i) I_{(u_{ti}=1)}\Big) |y \Big) \Bigg] \\
&= \sum_{t=1}^{T} \hat{Z}_{ti}^{(k)} \left(\frac{1}{\nu_i} I_{(u_{ti}=\gamma)} - \frac{1}{\nu_i - 1} I_{(u_{ti}=1)} \right) \\
&= 0.
\end{aligned}$$

These equations capture the iterative optimization process for updating the parameter vector $\boldsymbol{\nu_i}$ within the CML step of the FM$_m$-SMSN-GARCH(1,1) mixture model.

A.3 Proof of Proposition 3.1

- Proof of μ_1:
  Since $\epsilon_t \sim FM_m - SMSN(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\nu})$, according to Proposition 2.2, we have $E(\epsilon) = \sum_{i=1}^{m} \pi_i(\mu_i - b_i\Delta_i)$. Under the zero mean condition, we can deduce:

$$\begin{aligned}
\sum_{i=1}^{m} \pi_i(\mu_i - b_i\Delta_i) &= 0 \\
\Rightarrow \sum_{i=1}^{m-1} \pi_i(\mu_i - b_i\Delta_i) &= -\pi_1(\mu_1 - b_1\Delta_1) \\
\Rightarrow \sum_{i=1}^{m-1} \pi_i(\mu_i - b_i\Delta_i) &= -\pi_1\mu_1 + \pi_1 b_1\Delta_1 \\
\Rightarrow -\sum_{i=1}^{m-1} \frac{\pi_i}{\pi_1}(\mu_i - b_i\Delta_i) + b_1\Delta_1 &= \mu_1
\end{aligned}$$

- Proof of σ_1^2:
  Since $\epsilon_t \sim FM_m - SMSN(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\nu})$, according to Proposition 2.2, we have $Var(\epsilon_t) = \sum_{i=1}^{m} \pi_i \left((k_{2i}\sigma_i^2 - b_i^2\Delta_i^2) + ((\mu_i - \bar{\mu}) - (b_i\Delta_i - b\bar{\Delta}))^2 \right)$. Under the unit variance condition, we can deduce:

$$\begin{aligned}
&\sum_{i=1}^{m-1} \pi_i \left((k_{2i}\sigma_i^2 - b_i^2\Delta_i^2) + ((\mu_i - \bar{\mu}) - (b_i\Delta_i - b\bar{\Delta}))^2 \right) \\
&= 1 - \pi_1 k_{2i}\sigma_1^2 + \pi_1 b_1^2\Delta_1^2 - \pi_1 \left((\mu_1 - \bar{\mu}) - (b_1\Delta_1 - b\bar{\Delta}) \right)^2
\end{aligned}$$

Combining the above equations, we obtain:

$$\sum_{i=1}^{m-1} \pi_i \left((k_{2i}\sigma_i^2 - b_i^2\Delta_i^2) + ((\mu_i - \bar{\mu}) - (b_i\Delta_i - b\bar{\Delta}))^2 \right)$$
$$- 1 - \pi_1 b_1^2 \Delta_1^2 + \pi_1 \left((\mu_1 - \bar{\mu}) - (b_1\Delta_1 - b\bar{\Delta}) \right)^2 = -\pi_1(k_{2i}\sigma_1^2)$$

Therefore, we can express σ_1^2 as:

$$\sigma_1^2 = \frac{1}{k_{21}\pi_1} \Bigg\{ 1 - \sum_{i=1}^{m-1} \pi_i \left((k_{2i}\sigma_i^2 - b_i^2\Delta_i^2) + ((\mu_i - \bar{\mu}) - (b_i\Delta_i - \bar{b}\Delta))^2 \right)$$
$$- \left((\mu_1 - \bar{\mu}) - (b_1\Delta_1 - \bar{b}\Delta) \right)^2 + b_1^2\Delta_1^2 \Bigg\}$$

A.4 Proof of proposition 1.4.

Proof. (a)
Since $U \sim Gamma\big(\frac{\nu}{2}, \frac{\nu}{2}\big)$, therefore we have

$$\begin{aligned}
I_{ti}^{\Phi}(\rho) &= E_U\Big(U^\rho \exp{(-US_{ti}/2)} \Phi(\sqrt{U}B_{ti}) \Big) \\
&= 2\int_0^\infty u^\rho \exp{(-uS_{ti}/2)} \Phi(\sqrt{u}B_{ti}) \frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma\big(\frac{\nu}{2}\big)} u^{\frac{\nu}{2}-1} \exp\big(-\nu u/2\big) du \\
&= 2\frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma\big(\frac{\nu}{2}\big)} \int_0^\infty u^{\rho+\frac{\nu}{2}-1} \exp\big(-u(S_{ti}+\nu)/2\big) \Phi(\sqrt{u}B_{ti}) du \\
&= \frac{2^\rho \nu^{\frac{\nu}{2}} \Gamma\big(\rho + \frac{\nu}{2}\big)}{\sqrt{2\pi}\Gamma\big(\frac{\nu}{2}\big)(\nu + S_{ti})^{\rho+\frac{\nu}{2}}} T\left(\frac{B_{ti}\sqrt{\nu+2\rho}}{(\nu+S_{ti})^{\frac{1}{2}}}; \nu + 2\rho \right).
\end{aligned}$$

To be more clear, we notice that the last equality obtained by the following equation which was proved by (Lachos *et al.*, 2010b)

$$\begin{aligned}
&\int_0^\infty u^r \Phi(\sqrt{u}B) \frac{u^{\frac{\nu-1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} \exp\left(-\frac{1}{2}u(S+\nu)\right) \left(\frac{\nu+S}{2}\right)^{\frac{\nu+1}{2}} du \\
&= 2^{r+1} \frac{\Gamma\left(\frac{\nu+1+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} (\nu+S)^{-r} T\left(\sqrt{\frac{\nu+1+2r}{(\nu+S)^{\frac{1}{2}}}} B; \nu+1+2r \right).
\end{aligned}$$

Thus

$$\int_0^\infty u^\rho \exp\left(-\frac{uS_{ti}}{2}\right)\Phi\left(\sqrt{u}B_{ti}\right)\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}u^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu u}{2}\right)du$$
$$=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\left(\frac{S_{ti}+\nu}{2}\right)^{\frac{\nu+1}{2}}}\int_0^\infty u^{\rho-\frac{1}{2}}\exp\left(-\frac{u(S_{ti}+\nu)}{2}\right)\Phi\left(\sqrt{u}B_{ti}\right)\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\frac{u^{\frac{\nu-1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)}\left(\frac{S_{ti}+\nu}{2}\right)^{\frac{\nu+1}{2}}du$$
$$=\frac{2^{\rho+\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}\Gamma\left(\rho+\frac{\nu}{2}\right)(\nu+S_{ti})^{-(\rho-\frac{1}{2})}}{\Gamma\left(\frac{\nu}{2}\right)(\nu+S_{ti})^{\frac{\nu+1}{2}}\Gamma\left(\frac{\nu}{2}\right)}T\left(\frac{B_{ti}\sqrt{\nu+2\rho}}{(\nu+S_{ti})^{\frac{1}{2}}};\nu+2\rho\right).$$

So, we obtain

$$2\int_0^\infty u^\rho \exp(-uS_{ti}/2)\Phi(\sqrt{u}B_{ti})\frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}u^{\frac{\nu}{2}-1}\exp\big(-\nu u/2\big)du$$
$$=\frac{2^{\rho+\frac{3}{2}}\Gamma\left(\frac{\nu+1}{2}\right)\left(\nu/2\right)^{\frac{\nu}{2}}\Gamma\left(\rho+\frac{\nu}{2}\right)(\nu+S_{ti})^{-(\rho-\frac{1}{2})}}{\Gamma\left(\frac{\nu}{2}\right)(\nu+S_{ti})^{\frac{\nu+1}{2}}\Gamma\left(\frac{\nu}{2}\right)}T\left(\frac{B_{ti}\sqrt{\nu+2\rho}}{\left(\nu+S_{ti}\right)^{\frac{1}{2}}};\nu+2\rho\right)$$
$$=I_{ti}^{\Phi}(\rho).$$

And

$$I_{ti}^{\phi}(\rho)$$
$$=\mathbb{E}_U\left[U^\rho\exp\left(-\frac{U(S_{ti}+B_{ti}^2)}{2}\right)\right]$$
$$=\int_{-\infty}^\infty u^\rho\exp\left(-\frac{u(S_{ti}+B_{ti}^2)}{2}\right)\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}u^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu u}{2}\right)du$$
$$=\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\int_{-\infty}^\infty u^{\rho+\frac{\nu}{2}-1}\exp\left(-\frac{u(S_{ti}+B_{ti}^2+\nu)}{2}\right)du$$
$$=\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}\Gamma\left(\rho+\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left(\frac{1}{2}(S_{ti}+B_{ti}^2+\nu)\right)^{\rho+\frac{\nu}{2}}}\underbrace{\int_{-\infty}^\infty\frac{\left(\frac{1}{2}(S_{ti}+B_{ti}^2+\nu)\right)^{\rho+\frac{\nu}{2}}}{\Gamma(\rho+\frac{\nu}{2})}u^{\rho+\frac{\nu}{2}-1}\exp\left(-\frac{u(S_{ti}+B_{ti}^2+\nu)}{2}\right)du}_{=1\quad\text{since it follows a Gamma}\sim(\rho+\frac{\nu}{2},\frac{1}{2}(S_{ti}+B_{ti}^2+\nu))}$$
$$=\frac{2^{\rho+\frac{\nu}{2}}\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}\Gamma\left(\rho+\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{1}{\nu+B_{ti}^2+S_{ti}}\right)^{\rho+\frac{\nu}{2}}.$$

□

Proof. (b)
Since $U \sim Beta(\nu, 1)$, thus we have

$$\begin{aligned}
I_{ti}^{\Phi}(\rho) &= \mathbb{E}_U\left[U^{\rho}\exp\left(-\frac{US_{ti}}{2}\right)\Phi(\sqrt{U}B_{ti})\right] \\
&= 2\int_0^{\infty} u^{\rho}\exp\left(-\frac{uS_{ti}}{2}\right)\Phi(\sqrt{u}B_{ti})\frac{\Gamma(\nu+1)}{\Gamma(\nu)}u^{\nu-1}du \\
&= 2\nu\int_0^{\infty} u^{\rho+\nu-1}\exp\left(-\frac{uS_{ti}}{2}\right)\Phi(\sqrt{u}B_{ti})du \\
&= \frac{2^{2+\nu}\Gamma(\rho+\nu)}{S_{ti}^{\rho+\nu}}P\left(1;\rho+\nu,\frac{S_{ti}}{2}\right)E\left(\Phi\left(\sqrt{Ga_{ti}}\right)B_{ti}\right),
\end{aligned}$$

And

$$\begin{aligned}
I_{ti}^{\phi}(\rho) &= \mathbb{E}_U\left[U^{\rho}\exp\left(-\frac{U(S_{ti}+B_{ti}^2)}{2}\right)\right] \\
&= 2\int_0^{\infty} u^{\rho}\exp\left(-\frac{u(S_{ti}+B_{ti}^2)}{2}\right)\Phi(\sqrt{u}B_{ti})\frac{\Gamma(\nu+1)}{\Gamma(\nu)}u^{\nu-1}du \\
&= 2\nu\int_0^{\infty} u^{\rho+\nu-1}\exp\left(-\frac{u(S_{ti}+B_{ti}^2)}{2}\right)du \\
&= \frac{\nu\cdot 2^{\rho+\nu}\Gamma(\rho+\nu)}{\sqrt{2\pi}(B_{ti}^2+S_{ti})^{\rho+\nu}}P\left(1;\rho+\nu,\frac{B_{ti}^2+S_{ti}}{2}\right).
\end{aligned}$$

□

Proof. (c)
Since $h(u;\boldsymbol{\nu}) = \nu I_{(u=\gamma)} + (1-\nu)I_{(u=1)}, 0<\nu<1, 0<\gamma\leq 1$, so we obtain

$$\begin{aligned}
I_{ti}^{\Phi}(\rho) &= \mathbb{E}_U\left[U^{\rho}\exp\left(-\frac{US_{ti}}{2}\right)\Phi(\sqrt{U}B_{ti})\right] \\
&= \nu\gamma^{\rho}\exp\left(-\frac{\gamma S_{ti}}{2}\right)\Phi(\sqrt{\gamma}B_{ti}) + (1-\nu)\exp\left(-\frac{S_{ti}}{2}\right)\Phi(B_{ti}) \\
&= \sqrt{2\pi}\left(\nu\gamma^{\rho--/2}\phi\left(\sqrt{S_{ti}};0,\gamma^{-1}\right)\Phi(\sqrt{\gamma}B_{ti}) + (1-\nu)\phi\left(\sqrt{S_{ti}};0,1\right)\Phi(B_{ti})\right),
\end{aligned}$$

And

$$\begin{aligned}
I_{ti}^{\phi}(\rho) &= \mathbb{E}_U\left[U^{\rho}\exp\left(-\frac{U(S_{ti}+B_{ti}^2)}{2}\right)\right] \\
&= \nu\gamma^{\rho}\exp\left(-\frac{\gamma(S_{ti}+B_{ti}^2)}{2}\right) + (1-\nu)\exp\left(-\frac{S_{ti}+B_{ti}^2}{2}\right) \\
&= \nu\gamma^{\rho-1/2}\phi\left(\sqrt{B_{ti}^2+S_{ti}};0,\gamma^{-1}\right) + (1-\nu)\phi\left(\sqrt{B_{ti}^2+S_{ti}};0,1\right).
\end{aligned}$$

□

A.5

Table A.1: Initial values for simulation scenarios (two moderate and two weak components)

Table A.2: Moderately components

Parameters	SN	ST	SSL	SCN
μ_1	0.831	2.966	0.988	1.049
μ_2	-2	-1	-2	-2
σ_1^2	2.204	0.333	1.177	1.162
σ_2^2	0.5	2	0.5	0.5
λ_1	1	1	1	1
λ_2	-2	-2	-2	-2
ν	–	5	5	0.1
γ	–	–	–	0.2
π	0.6	0.6	0.6	0.6
ω	1.0E-04	1.0E-04	1.0E-04	1.0E-04
μ_0	1.0E-06	1.0E-06	1.0E-06	1.0E-06
α_1	0.05	0.05	0.05	0.05
β_1	0.9	0.9	0.9	0.9

Table A.3: Weakly components

Parameters	SN	ST	SSL	SCN
μ_1	-0.348	1.299	0.110	0.170
μ_2	0.5	1.5	0.4	0.2
σ_1^2	1.485	0.333	0.432	0.226
σ_2^2	2	2	2	2
λ_1	1	1	1	1
λ_2	-2	-2	-2	-2
ν	–	5	5	0.1
γ	–	5	5	0.2
π	0.6	0.6	0.6	0.6
ω	1.0E-04	1.0E-04	1.0E-04	1.0E-04
μ_0	1.0E-06	1.0E-06	1.0E-06	1.0E-06
α_1	0.05	0.05	0.05	0.05
β_1	0.9	0.9	0.9	0.9

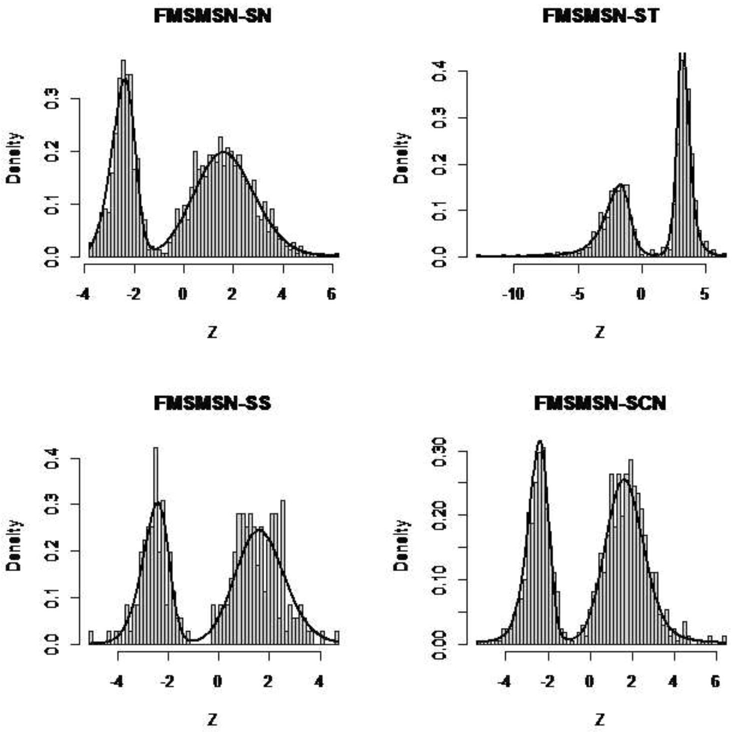

FIGURE A.1: Histogram and estimated density of marginal residuals of simulated data in a medium-FM_2-SMSN-GARCH (1,1) process.

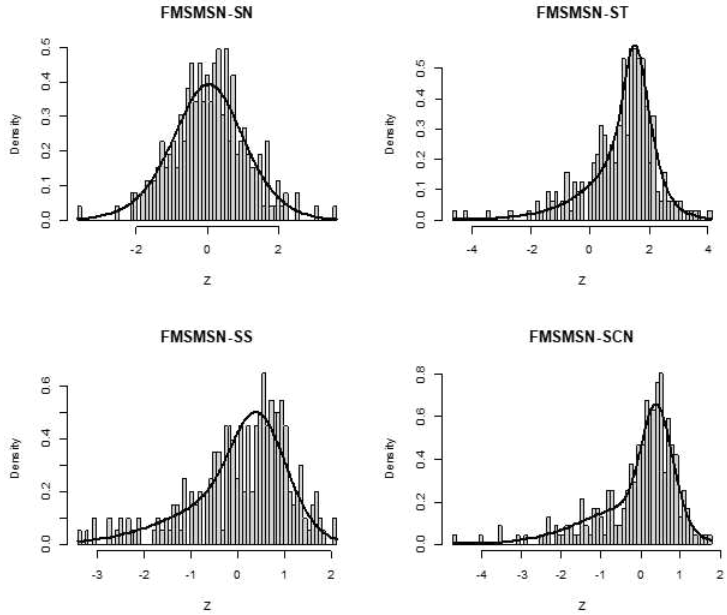

FIGURE A.2: Histogram and estimated density of marginal residuals of simulated data in a two-Weak-FM_2-SMSN-GARCH (1,1) process.

A.6

We present the mean and standard deviations (SD) of the ML estimates obtained through the Expectation Conditional Maximization Either (ECME) algorithm. The comprehensive results can be found in Tables A.4 and A.5. Additionally, to provide a visual representation of the ECME algorithm's performance, we have included Figures A.3 to A.10 related to the ML estimates.

Table A.4: Mean and SD of ML estimates of medium-FM$_2$-SN–GARCH(1,1), medium-FM$_2$-ST–GARCH(1,1), medium-FM$_2$-SSL–GARCH(1,1) and medium-FM$_2$-SCN–GARCH(1,1) models with various sample sizes $n = 250$, 500, and 1000

Models		FM$_2$-SN–GARCH(1,1)			FM$_2$-ST–GARCH(1,1)		
Measure	Parameters	250	500	1000	250	500	1000
Mean	μ_1	0.742	0.760	0.787	2.767	2.794	2.856
	μ_2	-2.243	-2.171	-2.111	-1.038	-1.029	-0.982
	σ_1^2	2.677	2.676	2.663	0.285	0.304	0.312
	σ_2^2	0.570	0.570	0.560	1.587	1.777	1.807
	λ_1	1.321	1.209	1.205	0.785	0.956	1.064
	λ_2	-2.474	-2.118	-2.117	-2.108	-2.075	-1.961
	ν	–	–	–	4.671	5.216	5.150
	γ	–	–	–	–	–	–
	π	0.608	0.603	0.600	0.595	0.598	0.600
	ω	0.125	0.008	0.001	-0.053	0.007	-0.004
	μ_0	7.8E-05	5.0E-05	5.1E-05	5.0E-05	5.1E-05	4.9E-05
	α_1	0.061	0.056	0.055	0.068	0.065	0.064
	β_1	0.878	0.890	0.893	0.844	0.871	0.885
SD	μ_1	0.396	0.216	0.167	0.242	0.113	0.075
	μ_2	0.350	0.262	0.163	0.334	0.178	0.116
	σ_1^2	0.713	0.463	0.357	0.129	0.071	0.053
	σ_2^2	0.281	0.141	0.123	0.590	0.453	0.317
	λ_1	0.701	0.419	0.312	0.737	0.427	0.300
	λ_2	2.302	1.330	0.817	1.410	0.879	0.534
	ν	–	–	–	1.616	1.291	0.977
	γ	–	–	–	–	–	–
	π	0.053	0.021	0.015	0.030	0.020	0.015
	ω	0.284	0.262	0.247	0.276	0.255	0.249
	μ_0	4.46E-04	2.7E-05	2.6E-05	2.7E-05	2.71E-05	2.5E-05
	α_1	0.046	0.005	0.004	0.006	0.004	0.002
	β_1	0.096	0.025	0.018	0.045	0.030	0.020

Models		FM$_2$-SSL–GARCH(1,1)			FM$_2$-SCN–GARCH(1,1)		
Measure	Parameters	250	500	1000	250	500	1000
Mean	μ_1	1.088	1.032	1.006	1.0005	1.036	1.052
	μ_2	-2.256	-2.140	-2.100	-2.127	-2.109	-2.079
	σ_1^2	1.564	1.475	1.414	1.465	1.414	1.387
	σ_2^2	0.676	0.652	0.560	0.593	0.568	0.562
	λ_1	1.100	1.043	1.032	1.306	1.238	1.179
	λ_2	-1.622	-2.117	-2.054	-2.295	-2.149	-2.116
	ν	5.434	5.019	5.012	0.151	0.123	0.100
	γ	–	–	–	0.302	0.242	0.231
	π	0.605	0.599	0.600	0.605	0.601	0.599
	ω	0.149	0.071	-0.001	0.149	0.006	0.000
	μ_0	6.7E-05	5.1E-05	5.5E-05	5.9E-05	5.3E-05	5.2E-05
	α_1	0.057	0.055	0.055	0.060	0.057	0.057
	β_1	0.883	0.892	0.892	0.877	0.890	0.892
SD	μ_1	0.327	0.230	0.164	0.326	0.221	0.137
	μ_2	0.249	0.109	0.075	0.212	0.121	0.078
	σ_1^2	0.473	0.385	0.282	0.538	0.384	0.267
	σ_2^2	0.234	0.161	0.163	0.243	0.165	0.122
	λ_1	0.607	0.466	0.307	0.827	0.483	0.314
	λ_2	1.039	0.785	0.413	2.062	0.859	0.585
	ν	0.849	0.551	0.153	0.166	0.096	0.016
	γ	–	–	–	0.262	0.166	0.054
	π	0.031	0.020	0.015	0.029	0.024	0.016
	ω	0.268	0.225	0.161	0.291	0.275	0.230
	μ_0	3.64E-04	2.7E-05	2.9E-05	2.48E-04	2.7E-05	2.6E-05
	α_1	0.010	0.006	0.004	0.010	0.006	0.005
	β_1	0.042	0.027	0.018	0.042	0.027	0.020

TABLE A.5: Mean and SD of ML estimates of weak-FM_2-SN–GARCH(1,1), weak-FM_2-ST–GARCH(1,1), weak-FM_2-SSL–GARCH(1,1) and weak-FM_2-SCN–GARCH(1,1) models with various sample sizes n = 250, 500, and 1000

Models		FM_2-SN–GARCH(1,1)			FM_2-ST–GARCH(1,1)		
Measure	Parameters	250	500	1000	250	500	1000
	μ_1	-0.574	-0.553	-0.478	0.686	0.802	0.951
	μ_2	0.783	0.745	0.636	0.861	0.955	1.108
	σ_1^2	1.767	1.511	1.617	0.122	0.185	0.253
	σ_2^2	3.166	2.796	2.625	0.977	1.192	1.526
	λ_1	2.346	1.787	1.573	1.655	1.397	0.855
	λ_2	-3.931	-3.660	-3.247	-2.615	-2.339	-1.808
Mean	ν	–	–	–	6.326	5.422	5.181
		–	–	–	6.326	5.422	5.181
	π	0.672	0.656	0.644	0.660	0.647	0.626
	ω	2.047E-03	5.63E-04	4.03E-04	2.095E-03	4.84E-04	3.75E-04
	μ_0	9.2E-04	7.08E-05	4.32E-05	4.36E-04	8.81E-05	5.26E-05
	α_1	0.034	0.038	0.042	0.031	0.037	0.053
	β_1	0.901	0.902	0.902	0.926	0.924	0.895
	μ_1	0.167	0.161	0.160	0.384	0.374	0.173
	μ_2	0.279	0.249	0.210	0.632	0.356	0.313
	σ_1^2	0.971	0.502	0.490	0.222	0.246	0.194
	σ_2^2	1.846	1.403	0.968	0.734	0.799	0.682
	λ_1	0.965	0.600	0.454	1.255	0.810	0.510
	λ_2	1.281	1.084	0.640	1.530	1.234	0.810
SD	ν	–	–	–	3.430	2.0006	1.115
		–	–	–	3.430	2.0006	1.115
	π	0.115	0.096	0.087	0.107	0.080	0.080
	ω	9.626E-03	5.47E-04	2.62E-04	1.1818E-02	3.03E-04	2.37E-04
	μ_0	6.622E-03	3.02E-04	2.47E-05	1.191E-03	1.82E-04	4.32E-05
	α_1	0.010	0.013	0.010	0.021	0.011	0.009
	β_1	0.040	0.034	0.033	0.008	0.021	0.012

Models		FM_2-SSL–GARCH(1,1)			FM_2-SCN–GARCH(1,1)		
	Measure	250	500	1000	250	500	1000
	μ_1	0.171	0.135	0.119	0.379	0.346	0.263
	μ_2	0.479	0.371	0.381	0.389	0.350	0.292
	σ_1^2	0.531	0.476	0.452	0.347	0.313	0.294
	σ_2^2	1.518	1.687	1.808	3.332	2.981	2.688
	λ_1	0.706	0.753	0.791	2.399	1.946	1.752
	λ_2	-1.693	-1.735	-1.776	-3.356	-2.985	-2.708
Mean	ν	6.251	4.447	4.620	0.253	0.215	0.190
		6.251	4.447	4.620	0.361	0.310	0.280
	π	0.700	0.677	0.639	0.541	0.558	0.602
	ω	0.001	0.0008	0.0004	0.152	0.0395	0.004
	μ_0	1.46E-04	9.43E-05	5.18E-05	1.02E-04	8.44E-05	7.03E-05
	α_1	0.0364	0.038	0.041	0.031	0.040	0.044
	β_1	0.906	0.911	0.908	0.911	0.905	0.897
	μ_1	0.143	0.132	0.102	0.303	0.266	0.191
	μ_2	0.376	0.292	0.268	0.277	0.204	0.193
	σ_1^2	0.516	0.271	0.162	0.283	0.259	0.223
	σ_2^2	0.951	0.954	0.634	1.639	0.999	0.756
	λ_1	0.53	0.398	0.259	1.332	1.039	0.852
	λ_2	1.444	0.928	0.704	1.424	0.818	0.655
SD	ν	1.280	1.319	0.797	0.096	0.039	0.050
		1.280	1.319	0.797	0.132	0.103	0.096
	π	0.085	0.092	0.080	0.080	0.045	0.034
	ω	0.003	0.001	0.0002	0.232	0.057	0.002
	μ_0	1.46E-04	1.41E-04	2.81E-05	1.62E-04	1.02E-04	6.58E-05
	α_1	0.016	0.018	0.009	0.020	0.015	0.012
	β_1	0.018	0.030	0.025	0.040	0.033	0.033

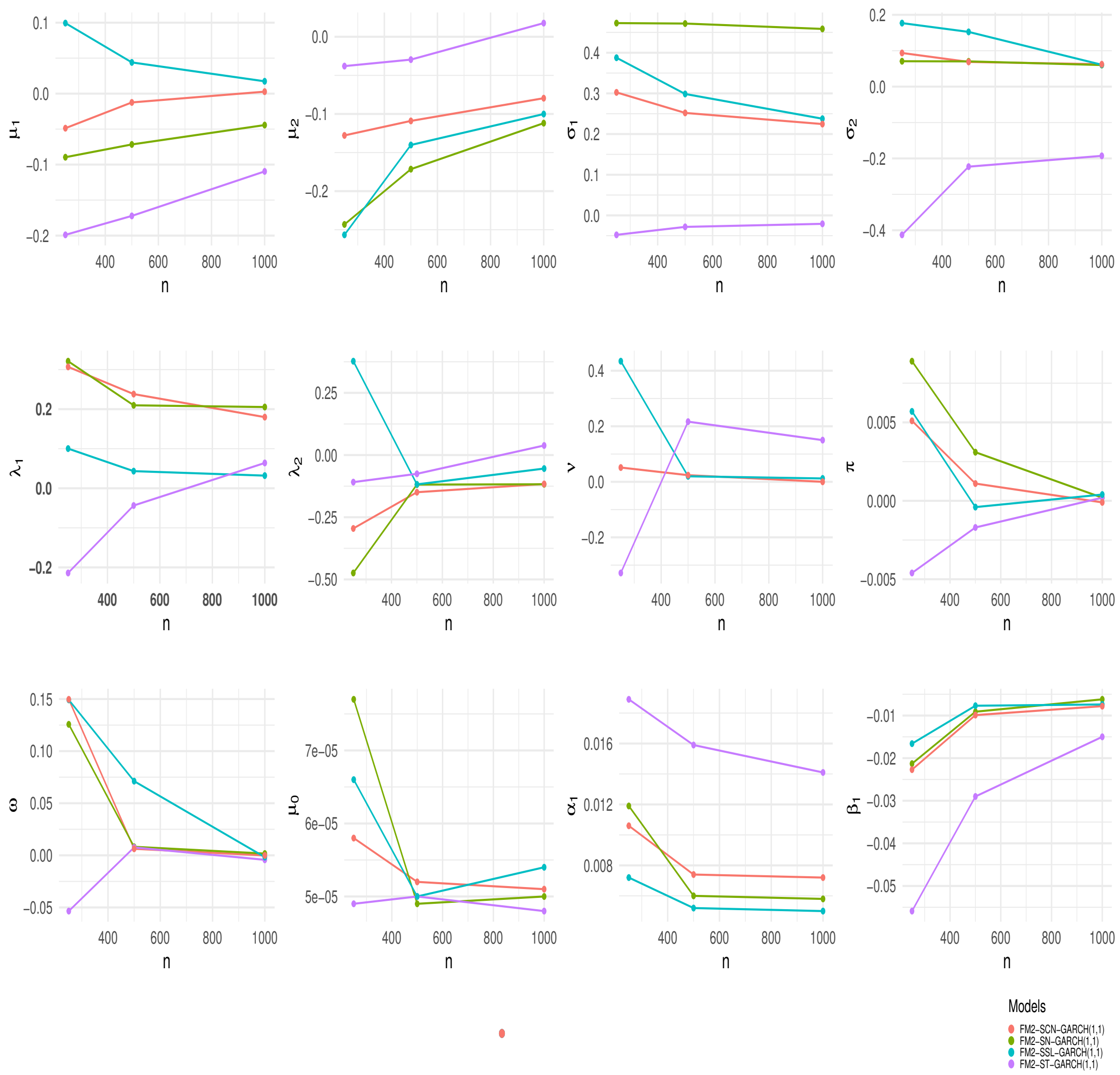

FIGURE A.3: Bias of ML estimates based on the medium-FM$_2$-SN-GACH, medium-FM$_2$-ST-GACH, medium-FM$_2$-SSL-GACH and medium-FM$_2$-SCN-GACH models.

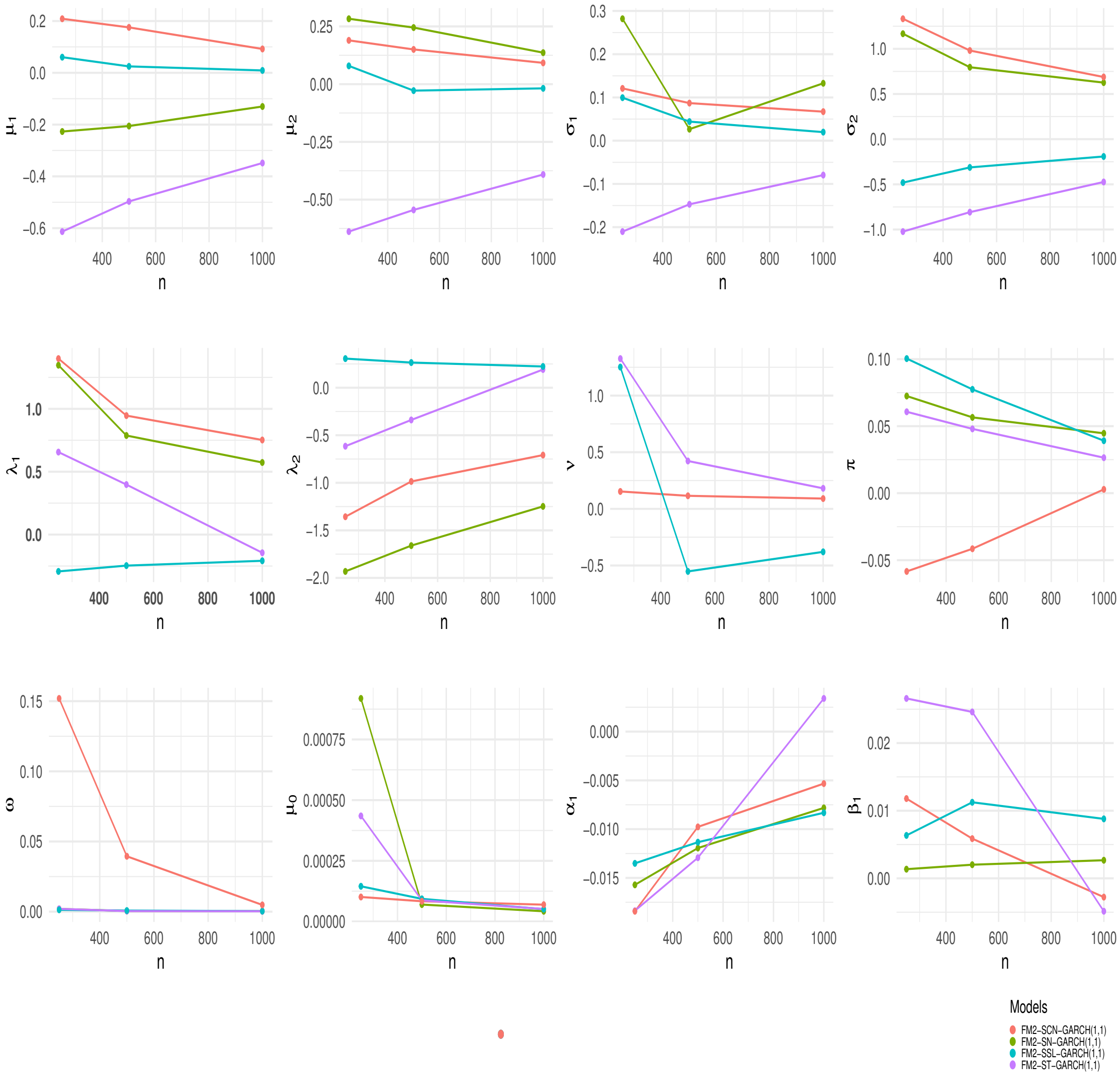

FIGURE A.4: Bias of ML estimates based on the weak-FM$_2$-SN-GACH, weak-FM$_2$-ST-GACH, weak-FM$_2$-SSL-GACH and weak-FM$_2$-SCN-GACH models.

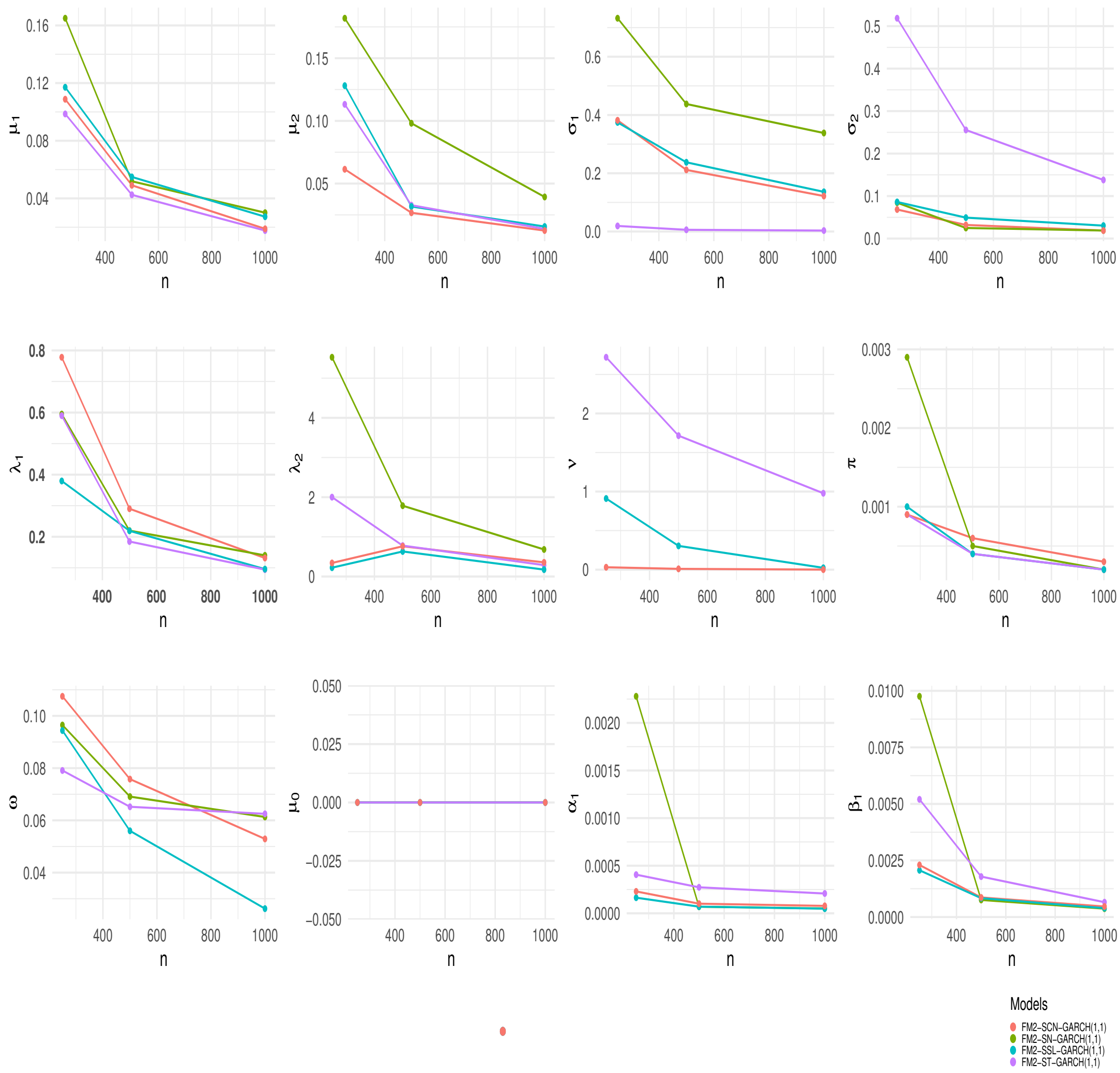

FIGURE A.5: MSE of ML estimates based on the medium-FM$_2$-SN-GACH, medium-FM$_2$-ST-GACH, medium-FM$_2$-SSL-GACH and medium-FM$_2$-SCN-GACH models.

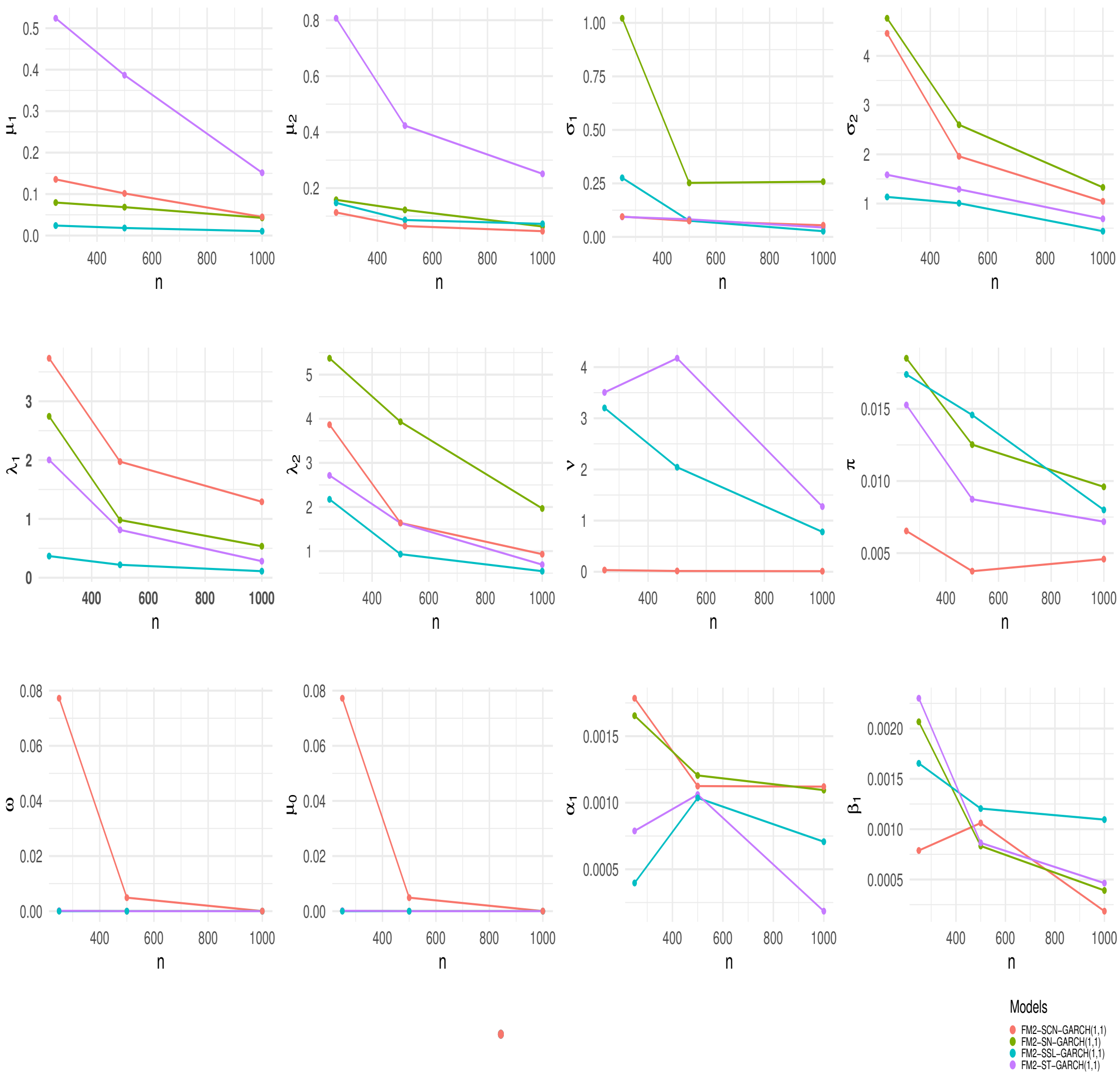

FIGURE A.6: MSE of ML estimates based on the weak-FM$_2$-SN-GACH, weak-FM$_2$-ST-GACH, weak-FM$_2$-SSL-GACH and weak-FM$_2$-SCN-GACH models.

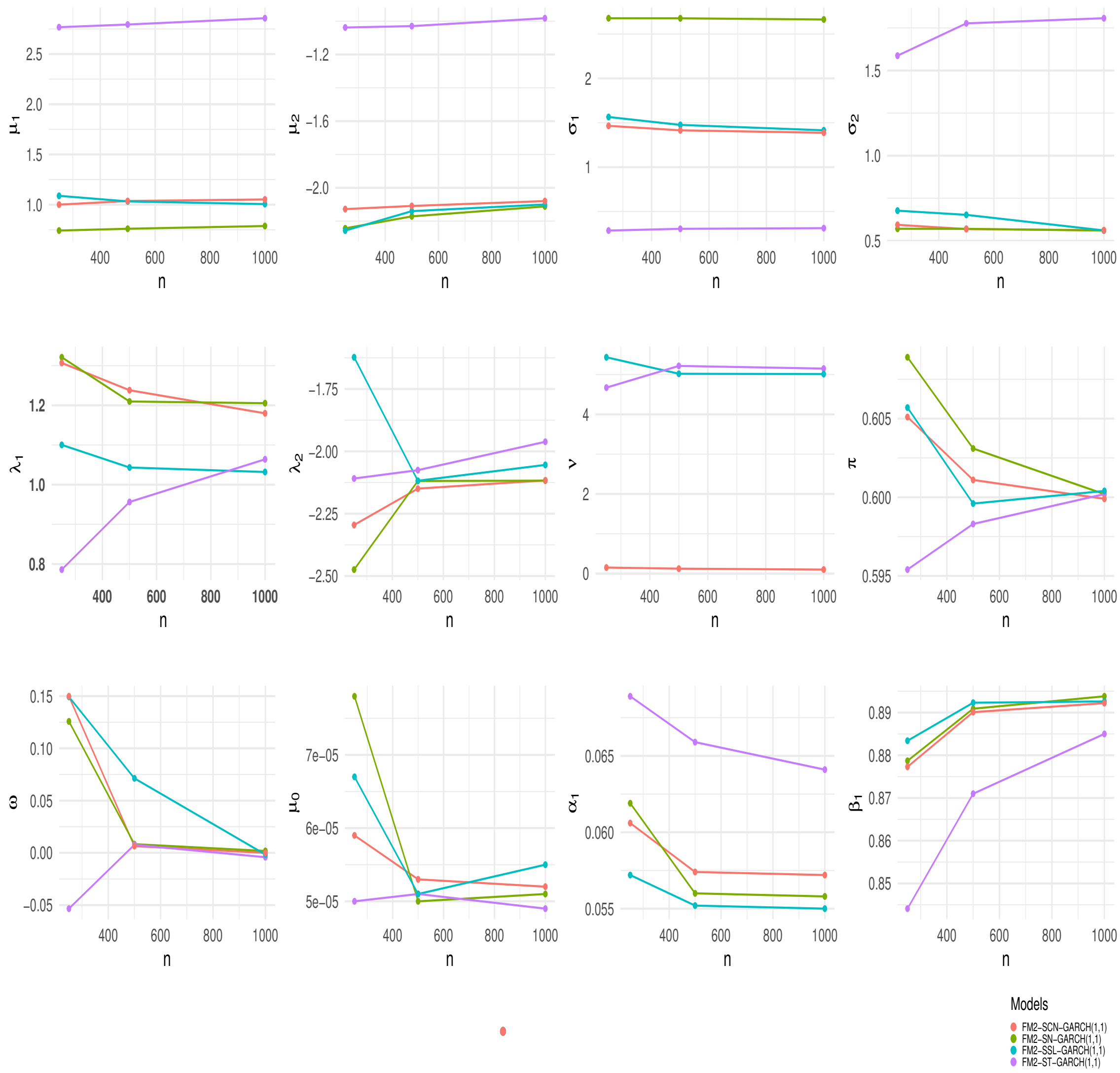

FIGURE A.7: Mean of ML estimates based on the medium-FM$_2$-SN-GACH, medium-FM$_2$-ST-GACH, medium-FM$_2$-SSL-GACH and medium-FM$_2$-SCN-GACH models.

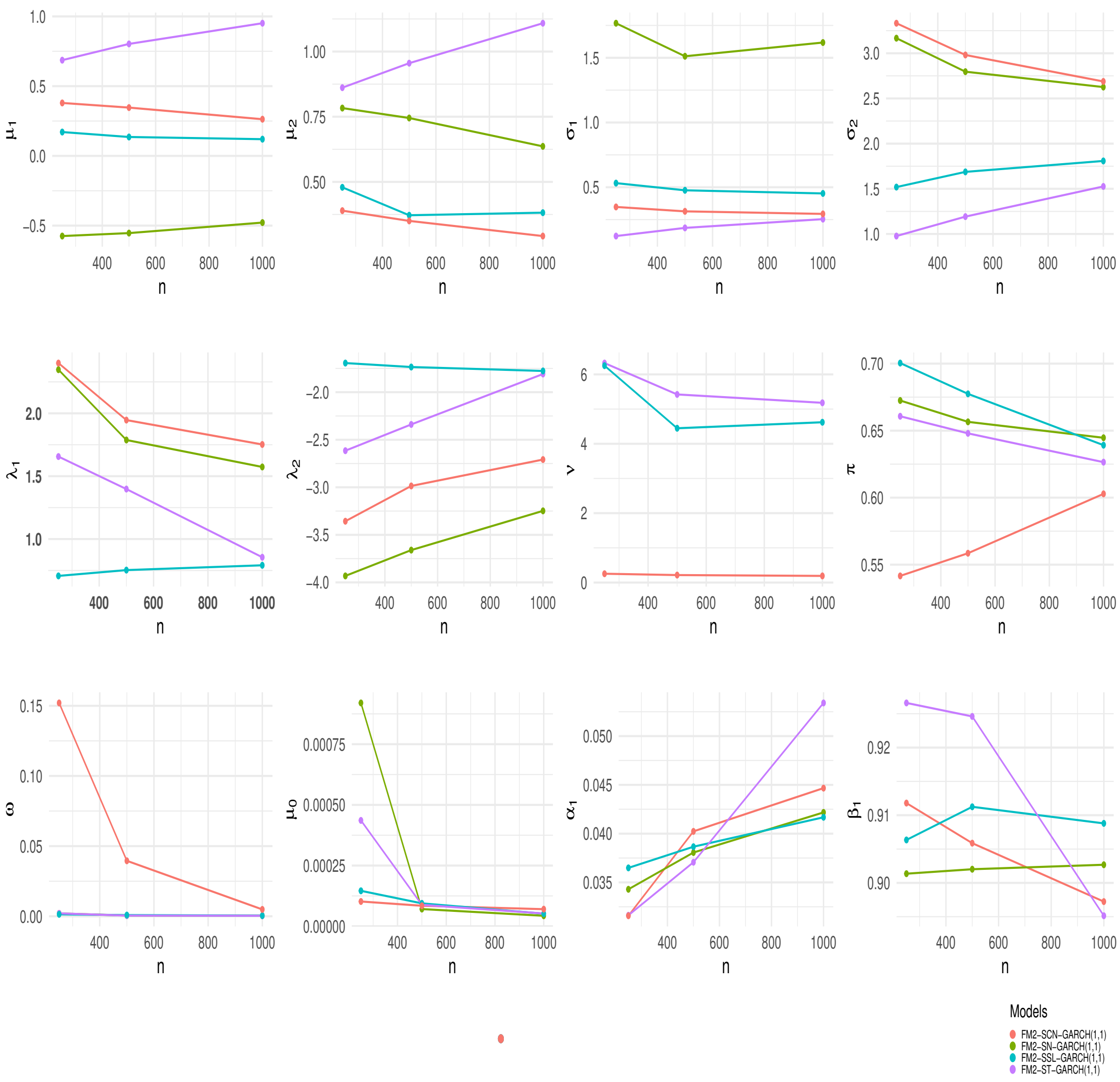

FIGURE A.8: Mean of ML estimates based on the weak-FM_2-SN-GACH, weak-FM_2-ST-GACH, weak-FM_2-SSL-GACH and weak-FM_2-SCN-GACH models.

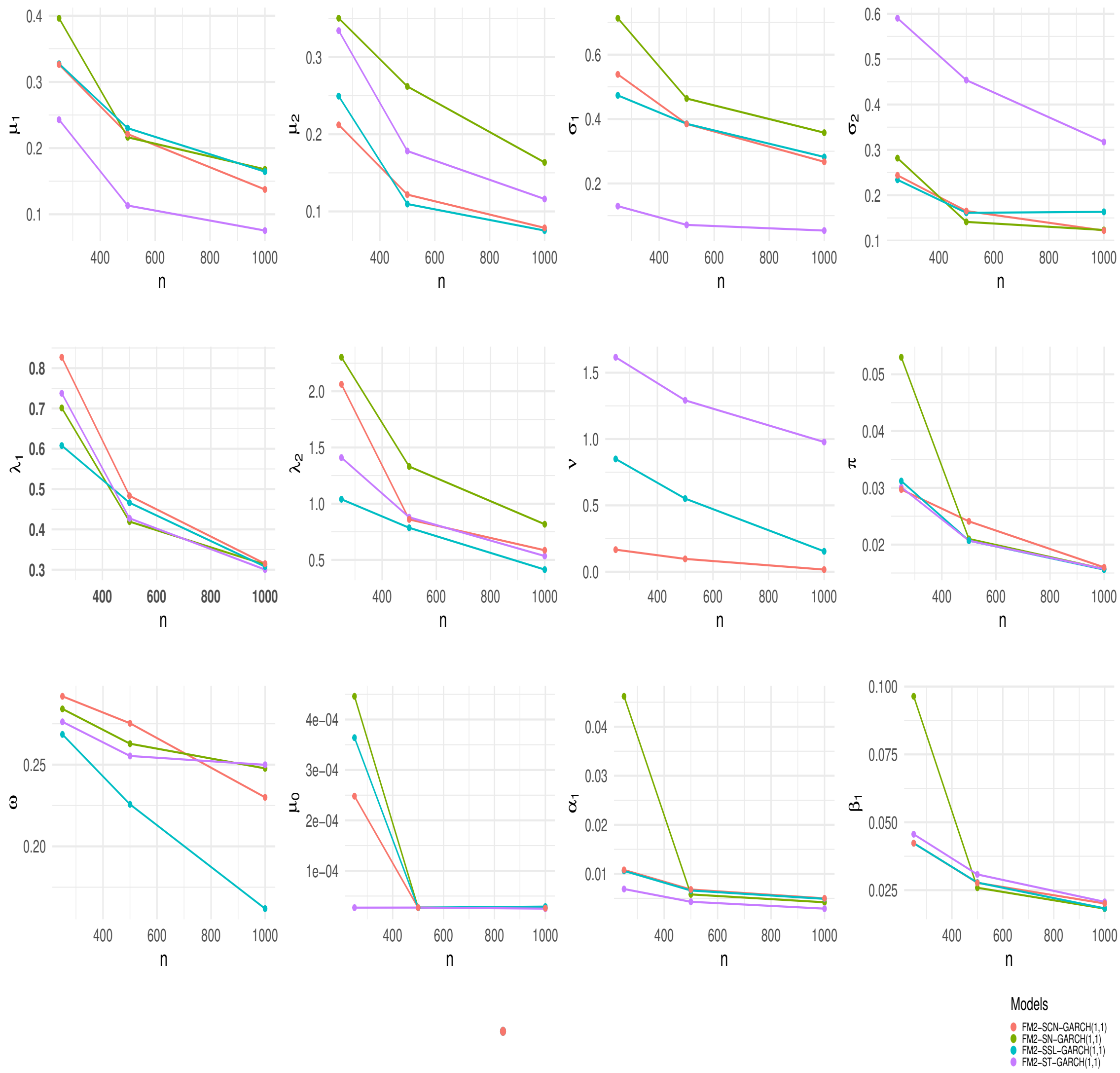

FIGURE A.9: SD of ML estimates based on the medium-FM_2-SN-GACH, medium-FM_2-ST-GACH, medium-FM_2-SSL-GACH and medium-FM_2-SCN-GACH models.

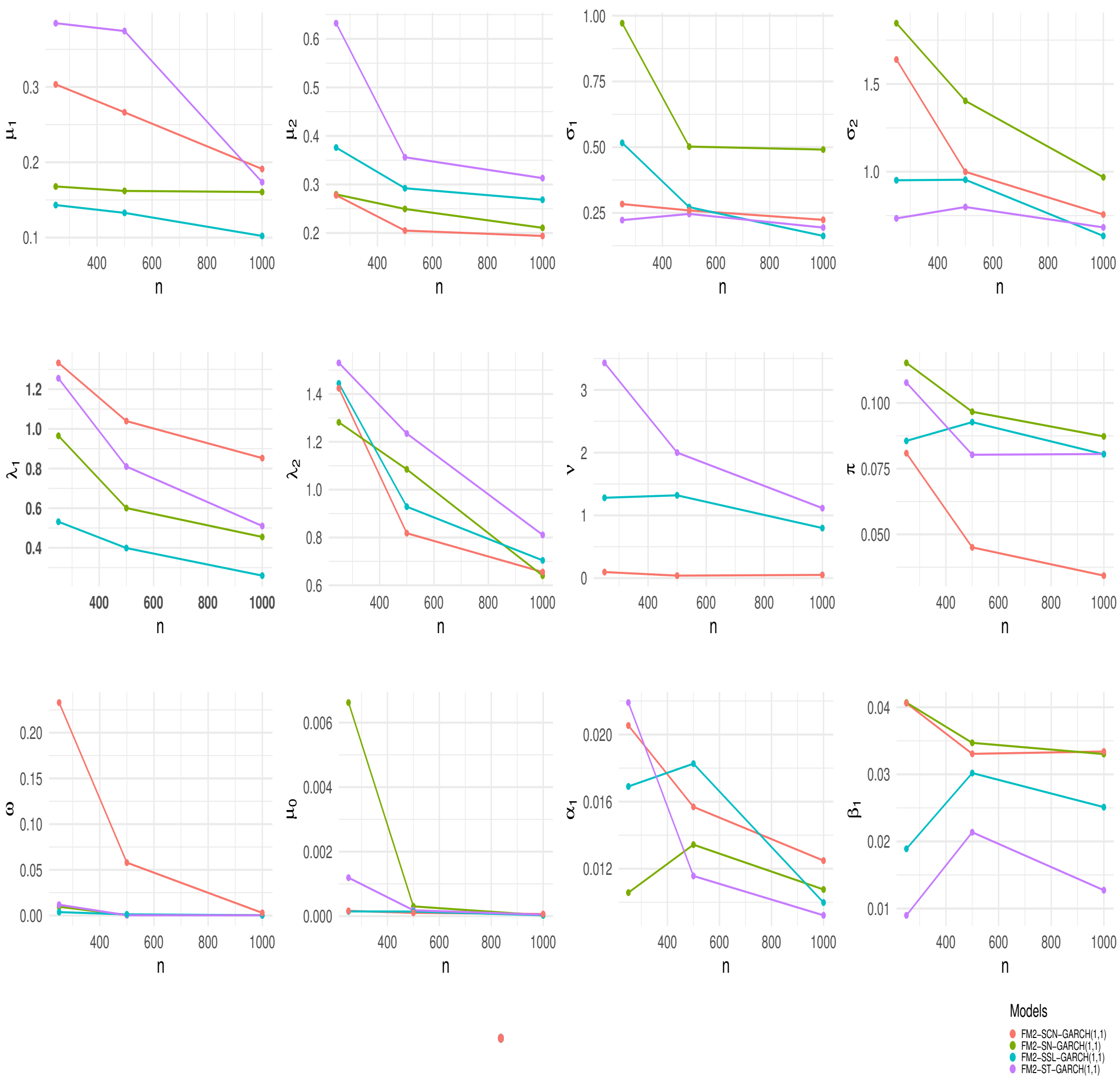

FIGURE A.10: SD of ML estimates based on the weak-FM$_2$-SN-GACH, weak-FM$_2$-ST-GACH, weak-FM$_2$-SSL-GACH and weak-FM$_2$-SCN-GACH models.

Appendix

Appendix B

B.1 Exploring the Formulation of the Generalized Hyperbolic Distribution (GH)

The GH distribution belongs to the NMVM family and is specified when the random mixing variable follows the GIG distribution ($W \sim GIG(\kappa, \chi, \psi)$). The probability density function (PDF) of the Generalized Inverse Gaussian (GIG) distribution for $w > 0$ is given as:

$$f_{GIG}(w; \kappa, \chi, \psi) = \frac{1}{2}\left(\frac{\psi}{\chi}\right)^{\frac{\kappa}{2}} \frac{w^{\kappa-1}}{K_\kappa(\sqrt{\psi\chi})} \exp\left\{-\frac{1}{2}\left(\frac{1}{w\chi} + w\psi\right)\right\}. \tag{B.1}$$

The modified Bessel function of the third kind, denoted as $K_\kappa(x)$, is defined for $x > 0$ and $\kappa \in \mathbb{R}$ as follows:

$$K_\kappa(x) = \frac{1}{2}\int_0^\infty y^{\kappa-1} \exp\left(-\frac{x}{2}\left(y + \frac{1}{y}\right)\right) dy$$

The parameters χ and ψ must satisfy the following conditions:
(i) If $\kappa > 0$, then $\chi \geq 0$ and $\psi > 0$.
(ii) If $\kappa < 0$, then $\chi > 0$ and $\psi \geq 0$.
(iii) If $\kappa = 0$, then $\chi > 0$ and $\psi > 0$.
These conditions ensure the validity of the GIG distribution and its associated modified Bessel function for different values of κ. This function is an important special function in mathematics and physics, widely used in various applications. The expectations $E(W^r)$ and $E(\log W)$ for $r = \pm 1$ are computed using the following expressions:
For $r = \pm 1$:

$$E(W^r) = \left(\frac{\chi}{\psi}\right)^{\frac{r}{2}} R_{(\kappa,r)}\left(\sqrt{\chi\psi}\right) \tag{B.2}$$

and

$$E(\log W) = \log\left(\sqrt{\frac{\chi}{\psi}}\right) + \frac{1}{K_\kappa(\chi\psi)}\frac{\partial K_\kappa(\chi\psi)}{\partial\kappa}$$

where $R_{(\kappa,a)}(c) = \frac{K_{\kappa+a}(c)}{K_\kappa(c)}$.
The GH (Generalized Hyperbolic) distribution describes a random variable X and is defined by the probability density function (PDF):

$$f_{GH}(x;\mu,\lambda,\sigma^2,\kappa,\chi,\psi) = \frac{\left(\frac{\psi}{\chi}\right)^{\kappa/2}\left(\psi+\frac{\lambda^2}{\sigma^2}\right)^{0.5-\kappa}}{\sqrt{2\pi\sigma^2}K_\kappa(\sqrt{\chi\psi})}\frac{K_{\kappa-0.5}\left(\sqrt{\left(\psi+\frac{\lambda^2}{\sigma^2}\right)\left(\chi+\delta(x;\mu,\sigma^2)\right)}\right)}{\left(\sqrt{\left(\psi+\frac{\lambda^2}{\sigma^2}\right)\left(\chi+\delta(x;\mu,\sigma^2)\right)}\right)^{0.5-\kappa}}$$
$$\times\exp\left\{\frac{(x-\mu)\lambda}{\sigma^2}\right\},$$

where $\delta(x;\mu,\sigma^2) = \frac{(x-\mu)^2}{\sigma^2}$.
These expressions define the expectations and the PDF of the GH distribution and are used in various mathematical and statistical applications.

B.2 Exploring the Various NMVM Distributions: An Overview of Key Examples

We present the distinct members of the NMVM distributions and the necessary conditional expectations for our suggested ECME algorithm.

- **The generalized hyperbolic skew-t distribution:** we introduce the Generalized Hyperbolic Skew-t (GHST) distribution, which is a specific case within the broader framework of the GH distribution. The GHST distribution characterizes the random variable (X) and is defined by the properties of its mixing random variable, (W), which follows an inverse gamma distribution denoted as $W \sim \text{Invgamma}(\frac{\nu}{2},\frac{\nu}{2})$.

  The GHST distribution exhibits different statistical moments, with its mean $E(W)$ being $\frac{\nu}{\nu-2}$ and the variance $Var(W)$ given by $\frac{2\nu^2}{(\nu-2)^2(\nu-4)}$. This distribution emerges as a special limiting case, characterized by the conditions $\kappa = -\frac{\nu}{2}$, $\chi = \nu$, and $\psi = 0$.

The probability density function (PDF) that governs the GHST distribution is expressed as:

$$f_{GHST}(x;\mu,\lambda,\sigma^2,\nu) = \left(\frac{\lambda^2/\sigma^2}{\nu+\delta(x,\mu,\sigma^2)}\right)^{(\nu+1)/4} \frac{\nu^{\nu/2}K_{\frac{\nu+1}{2}}\left(\sqrt{\lambda^2(\nu+\delta(x,\mu,\sigma^2))/\sigma^2}\right)}{\sqrt{2\pi\sigma^2}\Gamma(\nu/2)2^{\nu/2-1}exp\left\{(\mu-x)\lambda/\sigma^2\right\}},$$

Here, the function $\delta(x,\mu,\sigma^2)$ is defined as $\frac{(x-\mu)^2}{\sigma^2}$. Furthermore, applying Bayes' rule reveals that the conditional distribution of W given $X = x$ is a Generalized Inverse Gaussian (GIG) distribution, parameterized as $\kappa_{ST} = -\frac{(\nu+1)}{2}$, $\chi_{ST} = \delta(x,\mu,\sigma^2)+\nu$, and $\psi_{ST} = \frac{\lambda^2}{\sigma^2}$. Consequently, the k-th moment of W given $X = x$ can be easily computed using equation (B.2).

- **The normal inverse Gaussian distribution:** The Normal Inverse Gaussian (NIG) distribution is a specialized case derived from the Generalized Hyperbolic (GH) distribution by setting specific parameter values. In particular, this distribution is obtained by setting κ to -0.5, χ to 1, and ψ to ν^2 within the GH framework.

  The probability density function (PDF) characterizing the NIG distribution is expressed as follows:

$$f_{NIG}(x;\mu,\lambda,\sigma^2,\nu) = \sqrt{\frac{\psi_{NIG}}{\pi^2\sigma^2\chi_{NIG}}}K_1(\sqrt{\chi_{NIG}\psi_{NIG}})\exp\left\{\frac{(x-\mu)\lambda}{\sigma^2}+\nu\right\},$$

  Here, the parameters χ_{NIG} and ψ_{NIG} are defined as $\chi_{NIG} = 1+\delta(x,\mu,\sigma^2)$ and $\psi_{NIG} = \nu^2 + (\lambda/\sigma)^2$, respectively. Additionally, it's worth noting that $K_{0.5}(x) = K_{-0.5}(x) = \sqrt{\pi/2}x^{-0.5}\exp\{-x\}$.

  By applying Bayes' rule, it can be deduced that the conditional distribution of the mixing random variable W given $X = x$ follows a Generalized Inverse Gaussian (GIG) distribution with parameters $\kappa_{NIG} = -1$, $\chi_{NIG} = \chi_{NIG}$, and $\psi_{NIG} = \psi_{NIG}$. Consequently, the k-th moment of W given $X = x$ can be readily calculated using equation (B.2).

- **The variance-gamma distribution:** The Variance-Gamma (VG) distribution is derived from the Generalized Hyperbolic (GH) distribution by specific parameter settings, namely, setting $\chi = 0$ and $\kappa > 0$. This results in a distinct distribution that has important applications in various fields. The probability density function (PDF) for the VG distribution is expressed as follows:

$$f_{VG}(x;\mu,\lambda,\sigma^2,\kappa,\psi) = \frac{\psi^{\kappa}K_{\kappa-0.5}\big(\sqrt{\chi_{VG}\psi_{VG}}\big)}{2^{\kappa-1}\sqrt{2\pi\sigma^2}\Gamma(\kappa)\big(\sqrt{\chi_{VG}/\psi_{VG}}\big)^{0.5-\kappa}}\exp\left\{\frac{(x-\mu)\lambda}{\sigma^2}\right\},$$

Here, $\chi_{VG} = \delta(x,\mu,\sigma^2)$ and $\psi_{VG} = \psi^2 + (\lambda/\sigma)^2$. Notably, this distribution possesses distinctive properties that make it valuable in modeling various financial and statistical phenomena.

Furthermore, by applying Bayes' rule, it can be established that the conditional distribution of the mixing random variable W given $X = x$ follows a Generalized Inverse Gaussian (GIG) distribution with parameters $(\kappa - 0.5, \chi_{VG}, \psi_{VG})$. Additionally, leveraging equation (B.2), it is possible to determine the k-th moment of W given $X = x$ for the VG distribution. This mathematical relationship is vital for understanding the behavior of the VG distribution in different contexts.

- **The skew Laplace distribution:** The Skew Laplace (SL) distribution is a probability distribution that emerges when the mixing random variable, denoted as W in equation 3.1, follows an exponential distribution with a mean of 0.5, represented as $W \sim Exp(0.5)$. The Skew Laplace distribution is valuable in various statistical applications, especially when modeling asymmetric data. Its probability density function (PDF) is defined as:

$$f_{SL}(x;\mu,\sigma^2,\lambda) = \frac{1}{2\tau\sigma}\exp\left\{-\tau\frac{|x-\mu|}{\sigma} + \frac{\lambda(x-\mu)}{\sigma^2}\right\},$$

Here, $\tau = \sqrt{1 + \left(\frac{\lambda}{\sigma}\right)^2}$ is a crucial parameter that characterizes the shape and skewness of the distribution.

By employing Bayes' rule, it can be deduced that the conditional distribution of W given $X = x$ follows a Generalized Inverse Gaussian (GIG) distribution with parameters $(0.5, \chi_{ST}, \psi_{ST})$, where $\chi_{ST} = \delta(x,\mu,\sigma^2)$ and $\psi_{ST} = k^2$. Therefore, it is possible to compute the k-th moment of W given $X = x$ by utilizing equation (B.2). This expression provides essential insights into the behavior and statistical properties of the Skew Laplace distribution.

- **The normal mean-variance Birnbaum-Saunders distribution:** The Normal Mean-Variance Birnbaum-Saunders (NMVBS) distribution is a probabilistic model obtained by assuming that the mixing random variable W

in equation (3.1) follows a Birnbaum-Saunders (BS) distribution (Birnbaum and Saunders, 1969) with shape and scale parameters α and β, denoted as $W \sim BS(\alpha, \beta)$.

The probability density function (PDF) of the BS distribution can be expressed as a mixture of two Generalized Inverse Gaussian (GIG) distributions, as given in (Desmond, 1986):

$$f_{BS}(w; \alpha, \beta) = \frac{1}{2} f_{GIG}\left(w; \frac{1}{2}, \frac{\beta}{\alpha^2}, \frac{1}{\alpha\beta^2}\right) + \frac{1}{2} f_{GIG}\left(w; -\frac{1}{2}, \frac{\beta}{\alpha^2}, \frac{1}{\alpha\beta^2}\right).$$

If we assume that $W \sim BS(\alpha, 1)$, we obtain the Normal Mean-Variance Birnbaum-Saunders distribution, denoted as $f_{NMVBS}(x; \mu, \lambda, \sigma^2, \alpha)$, with the following PDF:

$$f_{NMVBS}(x; \mu, \lambda, \sigma^2, \alpha) = \frac{1}{2} f_{GH}\left(x; \mu, \lambda, \sigma^2, \frac{1}{2}, \frac{1}{\alpha^2}, \frac{1}{\alpha^2}\right) + f_{GH}\left(x; \mu, \lambda, \sigma^2, -\frac{1}{2}, \frac{1}{\alpha^2}, \frac{1}{\alpha^2}\right).$$

For any integer $r = \pm 1, \pm 2, ...$, the rth moment of W given $X = x$ can be calculated as follows:

$$E(W^r|X = x) = \left(\frac{\chi_{BS}}{\psi_{BS}}\right)^{k/2} \left\{p(x)R(0, r)\left(\sqrt{\chi_{BS}\psi_{BS}}\right) + (1 - p(x))R(-1, r)\left(\sqrt{\chi_{BS}\psi^*_{BS}}\right)\right\},$$

where $\chi_{BS} = \delta(x, \mu, \sigma^2) + \alpha^{-2}$, $\psi_{BS} = (\lambda/\sigma)^2 + \alpha^{-2}$, and

$$p(x) = \frac{f_{GH}\left(x; \mu, \lambda, \sigma^2, \frac{1}{2}, \frac{1}{\alpha^2}, \frac{1}{\alpha^2}\right)}{f_{GH}\left(x; \mu, \lambda, \sigma^2, \frac{1}{2}, \frac{1}{\alpha^2}, \frac{1}{\alpha^2}\right) + f_{GH}\left(x; \mu, \lambda, \sigma^2, -\frac{1}{2}, \frac{1}{\alpha^2}, \frac{1}{\alpha^2}\right)}.$$

This expression enables the calculation of moments for the NMVBS distribution, providing valuable insights into its statistical characteristics and applications.

– **The normal mean-variance Lindley distribution:** The Normal Mean-Variance Lindley (NMVL) distribution, introduced in (Naderi *et al.*, 2018), describes a random variable Y where the mixing random variable W follows a Lindley distribution (Lindley, 1958). The probability density function (PDF) of the Lindley distribution is proportional to $\alpha^2(1 + w^2)e^{-\alpha w}$, where α is a positive parameter and w is positive. The PDF of X is expressed as the mixture of two Generalized Hyperbolic (GH) distributions:

$$f_{NMVL}(x;\mu,\sigma^2,\lambda,\alpha)=\frac{\alpha}{\alpha+1}f_{GH}\big(x;\mu,\lambda,\sigma^2,1,0,2\alpha\big)+\frac{1}{\alpha+1}f_{GH}\big(x;\mu,\lambda,\sigma^2,2,0,2\alpha\big)$$

The expected value of W^r conditioned on $X=x$ can be calculated for $r=\pm1,\pm2,...$ as follows:

$$E(W^r|X=x)=\Big(\frac{\chi_{LI}}{\psi_{LI}}\Big)^{k/2}\Big\{p(x)R(0.5,r)\Big(\sqrt{\chi_{LI}\psi_{LI}}\Big)+(1-p(x))R(1.5,r)\Big(\sqrt{\chi_{LI}\zeta_{LI}}\Big)\Big\},$$

where $\chi_{LI}=\delta(x,\mu,\sigma^2)$, $\psi_{LI}=(\lambda/\sigma)^2+2\alpha$, and

$$p(x)=\frac{\alpha f_{GH}\big(x;\mu,\lambda,\sigma^2,1,0,2\alpha\big)}{\alpha f_{GH}\big(x;\mu,\lambda,\sigma^2,1,0,2\alpha\big)+f_{GH}\big(x;\mu,\lambda,\sigma^2,1,0,2\alpha\big)}.$$

In the paper by (Browne and McNicholas, 2015), a new parameterization was proposed for the Generalized Inverse Gaussian (GIG) distribution due to an identifiability issue. The issue is that $f_{GH}\big(x;\mu,\lambda,\sigma^2,\kappa,\chi,\psi\big)$ is equal to $f_{GH}\big(x;\mu,a\lambda,a\sigma^2,\kappa,\chi/a,a\psi\big)$ for any positive constant a, which leads to an identifiability problem. To resolve this, (Browne and McNicholas, 2015) set $\chi=\omega\eta$ and $\psi=\omega/\eta$ and showed that the finite mixture of GH distributions is identifiable with this new parameterization. This reparameterization is also consistent with the notation used in the paper by (McNeil *et al.*, 2015).

Bibliography

Aielli, G. P. and Caporin, M. (2014) Variance clustering improved dynamic conditional correlation mgarch estimators. *Computational Statistics & Data Analysis* **76**, 556–576.

Aitken, A. C. (1927) Xxv.—on bernoulli's numerical solution of algebraic equations. *Proceedings of the Royal Society of Edinburgh* **46**, 289–305.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Andrews, D. F. and Mallows, C. L. (1974) Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(1), 99–102.

Arellano-Valle, R. B., Branco, M. D. and Genton, M. G. (2006) A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics* **34**(4), 581–601.

Arellano-Valle, R. B., Castro, L. M., Genton, M. G. and Gómez, H. W. (2008) Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis* **3**(3), 513–539.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* pp. 171–178.

Azzalini, A. (2013) *The skew-normal and related families*. Volume 3. Cambridge University Press.

Azzalini, A. and Capitanio, A. (1999) Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**(3), 579–602.

Azzalini, A. and Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 367–389.

Azzalini, A. and Genton, M. G. (2008) Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review* **76**(1), 106–129.

Azzalini, A. and Valle, A. D. (1996) The multivariate skew-normal distribution. *Biometrika* **83**(4), 715–726.

Bai, Z.-D., Krishnaiah, P. R. and Zhao, L.-C. (1989) On rates of convergence of efficient detection criteria in signal processing with white noise. *IEEE Transactions on Information Theory* **35**(2), 380–388.

Barndorff-Nielsen, O. and Halgreen, C. (1977) Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **38**(4), 309–311.

Basford, K., Greenway, D., McLachlan, G. and Peel, D. (1997) Standard errors of fitted component means of normal mixtures. *Computational Statistics* **12**(1), 1–18.

Basso, R. M., Lachos, V. H., Cabral, C. R. B. and Ghosh, P. (2010) Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis* **54**(12), 2926–2941.

Billio, M. and Caporin, M. (2005) Multivariate markov switching dynamic conditional correlation garch representations for contagion analysis. *Statistical Methods and Applications* **14**(2), 145–161.

Billio, M., Caporin, M. and Gobbo, M. (2006) Flexible dynamic conditional correlation multivariate garch models for asset allocation. *Applied Financial Economics Letters* **2**(02), 123–130.

Birnbaum, Z. W. and Saunders, S. C. (1969) A new family of life distributions. *Journal of applied probability* **6**(2), 319–327.

Böhning, D. (1999) *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others*. Volume 81. CRC Press.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3), 307–327.

Bollerslev, T., Russell, J. and Watson, M. (2010) *Volatility and time series econometrics: essays in honor of Robert Engle*. Oxford University Press.

Bougerol, P. and Picard, N. (1992) Stationarity of garch processes and of some nonnegative time series. *Journal of econometrics* **52**(1-2), 115–127.

Bouveyron, C. and Brunet-Saumard, C. (2014) Model-based clustering of high-dimensional data: A review. *Computational Statistics & Data Analysis* **71**, 52–78.

Bouveyron, C., Celeux, G., Murphy, T. B. and Raftery, A. E. (2019) *Model-based clustering and classification for data science: with applications in R*. Volume 50. Cambridge University Press.

Branco, M. D. and Dey, D. K. (2001) A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**(1), 99–113.

Browne, R. P. and McNicholas, P. D. (2015) A mixture of generalized hyperbolic distributions. *Canadian Journal of Statistics* **43**(2), 176–198.

Caporin, M. and McAleer, M. (2012) Do we really need both bekk and dcc? a tale of two multivariate garch models. *Journal of Economic Surveys* **26**(4), 736–751.

Davidson, J. (1994) *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22.

Desmond, A. F. (1986) On the relationship between two fatigue-life models. *IEEE Transactions on Reliability* **35**(2), 167–169.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society* pp. 987–1007.

Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* **96**(456), 1348–1360.

Fraley, C. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association* **97**(458), 611–631.

Fröhwirth-Schnatter, S. and Kaufmann, S. (2008) Model-based clustering of multiple time series. *Journal of Business & Economic Statistics* **26**(1), 78–89.

Frühwirth-Schnatter, S. (2006) *Finite mixture and Markov switching models*. Springer Science & Business Media.

Garcia-Escudero, L. A. and Gordaliza, A. (1999) Robustness properties of k means and trimmed k means. *Journal of the American Statistical Association* **94**(447), 956–969.

Giraitis, L., Leipus, R. and Surgailis, D. (2007) Recent advances in arch modelling. *Long memory in economics* pp. 3–38.

Glosten, L. R., Jagannathan, R. and 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.

Good, I. J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* **40**(3-4), 237–264.

Haas, M., Mittnik, S. and Paolella, M. S. (2004) Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics* **2**(2), 211–250.

Handcock, M. S., Raftery, A. E. and Tantrum, J. M. (2007) Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **170**(2), 301–354.

Hartigan, J. A. and Wong, M. A. (1979) Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)* **28**(1), 100–108.

Holzmann, H., Munk, A. and Gneiting, T. (2006) Identifiability of finite mixtures of elliptical distributions. *Scandinavian Journal of Statistics* **33**(4), 753–763.

Hunter, D. R., Wang, S. and Hettmansperger, T. P. (2007) Inference for mixtures of symmetric distributions. *The Annals of Statistics* pp. 224–251.

Jones, M. C. and Faddy, M. (2003) A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(1), 159–174.

Kottas, A. and Fellingham, G. W. (2012) Bayesian semiparametric modeling and inference with mixtures of symmetric distributions. *Statistics and Computing* **22**(1), 93–106.

Lachos, V., Labra, F., Bolfarine, H. and Ghosh, P. (2010a) Multivariate measurement error models based on scale mixtures of the skew–normal distribution. *Statistics* **44**(6), 541–556.

Lachos, V. H., Ghosh, P. and Arellano-Valle, R. B. (2010b) Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* pp. 303–322.

Lachos, V. H. and Labra, F. V. (2014) Multivariate skew-normal/independent distributions: properties and inference. *Pro Mathematica* **28**(56), 11–53.

Lange, K. and Sinsheimer, J. S. (1993) Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* **2**(2), 175–198.

Lee, S. and McLachlan, G. J. (2014) Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics and Computing* **24**(2), 181–202.

Lee, S. X. and McLachlan, G. J. (2013a) On mixtures of skew normal and skew t -distributions. *Advances in Data Analysis and Classification* **7**(3), 241–266.

Lee, S. X. and McLachlan, G. J. (2013b) Model-based clustering and classification with non-normal mixture distributions. *Statistical Methods & Applications* **22**(4), 427–454.

Lin, T.-I. (2010) Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing* **20**(3), 343–356.

Lin, T.-I., Ho, H. J. and Lee, C.-R. (2014) Flexible mixture modelling using the multivariate skew-t-normal distribution. *Statistics and Computing* **24**(4), 531–546.

Lindley, D. V. (1958) Fiducial distributions and bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 102–107.

Lindsay, B. G. (1995) Mixture models: theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, pp. i–163.

Liu, C. and Rubin, D. B. (1994) The ecme algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika* **81**(4), 633–648.

Louis, T. A. (1982) Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **44**(2), 226–233.

Manouchehri, T. and Nematollahi, A. (2019) On the estimation problem of periodic autoregressive time series: symmetric and asymmetric innovations. *Journal of Statistical Computation and Simulation* **89**(1), 71–97.

Manouchehri, T. and Nematollahi, A. (2022a) Par (1) model analysis: a web-based shiny application for analysing periodic autoregressive models. *Journal of Statistical Computation and Simulation* pp. 1–22.

Manouchehri, T. and Nematollahi, A. (2022b) A comparison of the bayesian and non-bayesian approaches for the periodic ar models based on the smsn innovations. *Iranian Journal of Science and Technology, Transactions A: Science* **46**(2), 615–630.

McCulloch, R. E. and Tsay, R. S. (1994) Statistical analysis of economic time series via markov switching models. *Journal of Time Series Analysis* **15**(5), 523–539.

McLachlan, G. J. and Krishnan, T. (2007) *The EM algorithm and extensions.* John Wiley & Sons.

McLachlan, G. J., Lee, S. X. and Rathnayake, S. I. (2019) Finite mixture models. *Annual Review of Statistics and its Application* **6**, 355–378.

McNeil, A. J., Frey, R. and Embrechts, P. (2015) *Quantitative risk management: concepts, techniques and tools-revised edition.* Princeton University Press.

Meng, X.-L. and Rubin, D. B. (1993) Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika* **80**(2), 267–278.

Mengersen, K. L., Robert, C. and Titterington, M. (2011) *Mixtures: estimation and applications*. Volume 896. John Wiley & Sons.

Naderi, M., Arabpour, A. and Jamalizadeh, A. (2018) Multivariate normal mean-variance mixture distribution based on lindley distribution. *Communications in Statistics-Simulation and Computation* **47**(4), 1179–1192.

Peel, D. and McLachlan, G. J. (2000) Robust mixture modelling using the t distribution. *Statistics and Computing* **10**(4), 339–348.

Pourmousa, R., Jamalizadeh, A. and Rezapour, M. (2015) Multivariate normal mean–variance mixture distribution based on birnbaum–saunders distribution. *Journal of Statistical Computation and Simulation* **85**(13), 2736–2749.

Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T.-I., Maier, L. M., Baecher-Allan, C., McLachlan, G. J., Tamayo, P., Hafler, D. A. *et al.* (2009) Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences* **106**(21), 8519–8524.

Salas-Gonzalez, D., Kuruoglu, E. E. and Ruiz, D. P. (2010) Modelling with mixture of symmetric stable distributions using gibbs sampling. *Signal Processing* **90**(3), 774–783.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics* pp. 461–464.

Straumann, D. (2006) *Estimation in conditionally heteroscedastic time series models*. Volume 181. Springer Science & Business Media.

Taylor, S. J. (2008) *Modelling financial time series*. world scientific.

Team, R. C. *et al.* (2013) R: A language and environment for statistical computing .

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Wang, H., Li, R. and Tsai, C.-L. (2007) Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94**(3), 553–568.

Wang, J. and Genton, M. G. (2006) The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference* **136**(1), 209–220.

Wei, G. C. and Tanner, M. A. (1990) A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association* **85**(411), 699–704.

Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. and Ruzzo, W. L. (2001) Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17**(10), 977–987.

Zeller, C. B., Cabral, C. R. B., Lachos, V. H. and Benites, L. (2019) Finite mixture of regression models for censored data based on scale mixtures of normal distributions. *Advances in Data Analysis and Classification* **13**(1), 89–116.

Fariborz Setoudehtazangi

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 3483808129
e-mail: fariborz.setoudehtazangi@studenti.unipd.it

Current Position

- **PhD Student in Statistical Sciences**
  University of Padova
  Since February 2020 (expected completion: February 2023)
  Thesis title: "Time Series Clustering Based on a Mixture Model for Latent Volatility with Applications in Finance"
  Supervisor: Prof. Massimiliano Caporin

Research interests

- Statistical Machine Learning
- Computational Statistics
- Time Series
- Bayesian Statistics
- Multivariate Statistical Analysis
- Mixture Models and Model based clustering
- Longitudinal data
- Biostatistics
- Linear models
- Big Data Analysis, Data Mining, and Machine Learning

Education

- **Master of Science, Mathematical Statistics**
  Shiraz University, Faculty of Science, Department of Statistics
  Sep 2015 - Sep 2017
  Title of dissertation: "An Estimation Procedure in Autoregressive Models Based on the Finite Mixtures of Non-Gaussian Innovations"
  Supervisor: Prof. Alireza Nematollahi
  Final Mark: Excellent
- **Bachelor of Science, Applied Statistics**
  Shiraz University, Faculty of Science, Department of Statistics
  Sep 2011 - Sep 2015
  Title of dissertation: "Multivariate Statistical Analysis"
  Supervisor: Prof. Alireza Nematollahi
  Final Mark: Excellent (20/20)

Visiting periods

- **Visiting PhD Scholar**
  April 2023 - July 2023
  School of Mathematics and Physics, University of Queensland, Queensland, Australia
  Supervisor: Professor Geoff McLachlan

Awards and Scholarship

- **2023 ARC postgraduate research stipend** (Australian Research Council)
  Year: 2023
- **Postdoctoral Researcher**
  July 2023
  School of Mathematics and Physics, University of Queensland, Queensland, Australia
  Application Title: A Novel Approach to Semi-Supervised Statistical Machine Learning
  Supervisor: Professor Geoff McLachlan
- **PhD scholarship**
  Feb 2020 - July 2023
  Department of Statistical Sciences, University of Padova, Padova, Italy
- **Top student prize**
  Sep 2017
  Awarded as the top student in the Statistics Department (First Rank in M.Sc)
  The School of Sciences, Shiraz University, Shiraz, Iran
- **Top student prize**
  Sep 2015
  Awarded as the top student in the Statistics Department (First Rank in B.Sc)
  The School of Sciences, Shiraz University, Shiraz, Iran

Computer skills

- R/Rstudio
- Python
- LaTeX
- Minitab
- SPSS
- MS Office: Excel, Word, MathType, PowerPoint.

Languages

- Persian (Farsi): native
- English: fluent
- Italian: basic.

Research

Articles in journals

- Setoudeh, F. , Nematollahi. A. R., Manouchehri, T. , Caporin, M.(2023). Title (Time series clustering based on a mixture model for latent volatility with applications in finance). *Computational Statistics and Data Analysis (under review)*

- Setoudeh, F. , Nematollahi. A. R., Manouchehri, T. , Caporin, M.(2023). Title (Time series clustering based on a mixture model for latent volatility with applications in finance). *advances in data analysis and classification (under review)*
- Setoudeh, F. Spoto, F. Tezza, C. Fokianos, K. (2023). Title (Analysis of Neuronal Activity in Mice through Modern Time Series Methods). *Data Research Camp 2022, Venice, Italycontribute to the Springer Volume related to the Data Research Camp 2022. Springer Proceedings in Mathematics and Statistics (under review)*

Ongoing research

- Mixture of Linear Experts Model for Censored Data: A Novel Approach with Normal Mean-Variance Mixture Distribution (Setoudeh, F).

- A Novel Approach to Semi-Supervised Statistical Machine Learning (McLachlan, G and Setoudeh, F)

Conference presentations

Setoudeh, F. , Caporin, M. (2022) *Robust mixture GARCH modeling based on the scale mixture of skew normal distributions.* Statistical Data Science Conference (SDS), Bologna, Italy, , 26-28 August 2022.

Teaching Experience

- **Course: Time Series Analysis** *Dec 2018 - Jun 2019*
  Degree: Bachelor of Statistics
  Institution: Department of Statistics, College of Science, Fasa University, Fasa, Iran
  Course Description: Introduction to Time Series and Forecasting by Peter J. Brockwell and Richard A. Davis, using R programming. One semester, 4 hours per week
  Instructor: Prof. Fariborz Setoudehtazangi
- **Course: Engineering Statistics** *Dec 2018 - Jan 2020*
  Degree: Computer Engineering, Electronic Engineering, and Water Engineering
  Institution: Faculty of Engineering, Fasa University, Fasa, Iran
  Course Description: 4 semesters, 4 hours per week
  Instructor: Prof. Fariborz Setoudehtazangi
- **Course: Design of Experiment** *Dec 2018 - Jan 2020*
  Degree: Food Industry Engineering
  Institution: School of Agriculture, Fasa University, Fasa, Iran
  Course Description: 4 semesters, 4 hours per week
  Instructor: Prof. Fariborz Setoudehtazangi
- **Course: Statistical Methods** *Dec 2018 - Jan 2020*
  Degree: Bachelor of Statistics
  Institution: Department of Statistics, College of Science, Fasa University, Fasa, Iran
  Course Description: 4 semesters, 4 hours per week
  Instructor: Prof. Fariborz Setoudehtazangi
- **Course: Mathematical Statistics (1) & Multivariate Statistical Analysis (1) using SPSS and R** *Sept 2015 - Feb 2016*
  Degree: Bachelor of Statistics
  Institution: Department of Statistics, College of Science, Shiraz University
  Course Description: 2 hours per week per course
  Instructors: Prof. Mina Touhidi, Prof. Alireza Nematollahi
- **Course: Probability (2) & Multivariate Statistical Analysis (2) using SPSS and R** *Feb 2016 - June 2016*

Degree: Bachelor of Statistics
Institution: Department of Statistics, College of Science, Shiraz University
Course Description: 2 hours per week per course
Instructors: Prof. Zohre Shishebor and Prof. Alireza Nematollahi

- **Course: Mathematical Statistics (1) & Regression using SPSS and R** *Sep 2016 - Feb 2017*
  Degree: Bachelor of Statistics
  Institution: Department of Statistics, College of Science, Shiraz University
  Course Description: 2 hours per week per course
  Instructors: Prof. Rasool Borhani Haghighi and Prof. Mahmoud Kharati Kooapei
- **Course: Time Series using ITSM and R & Mathematical Statistics 2** *Feb 2017 - June 2017*
  Degree: Bachelor of Statistics
  Institution: Department of Statistics, College of Science, Shiraz University
  Course Description: 2 hours per week per course
  Instructors: rof. Zohreh Shishebor and Prof. Rasool Borhani Haghigh

References

Prof. Geoff McLachlan
Institution: School of Mathematics and Physics, The University of Queensland
Address: Brisbane, 4072, Australia
e-mail: g.mclachlan@uq.edu.au

Prof. Massimiliano Caporin
Institution: Department of Statistical Sciences, University of Padova
Address: Padova, 35121, Italy
e-mail: massimiliano.caporin@unipd.it

Prof. Alireza Nematollahi
Institution: Department of Statistics, Shiraz University
Address: Shiraz, Iran
e-mail: ar.nematollahi@shirazu.ac.ir