UNIVERSITY OF SOUTHAMPTON

Faculty of Social Sciences
School of Mathematical sciences

Essays on Risk Management and Portfolio Allocation Using Tail Measures

by
Faridah Alruwaili
ORCID: 0009-0001-3881-7724

A thesis for the degree of
Doctor of Philosophy

July 2024

University of Southampton

Abstract

Faculty of Social Sciences
School of Mathematical sciences

Doctor of Philosophy

Essays on Risk Management and Portfolio Allocation Using Tail Measures

by
Faridah Alruwaili
ORCID: 0009-0001-3881-7724

This thesis delves into effective methods for managing risks and decision-making processes in finance. The research comprises three main chapters, each addressing critical challenges in time series analysis, portfolio optimization, and risk assessment.

Chapter 2 introduces the Robust Model Averaging Marginal Regressions (RMAMAR) procedure, a novel approach that combines one-dimensional marginal regression functions to approximate conditional regression functions robustly. By employing local linear estimation and robust M-estimators, RMAMAR addresses the curse of dimensionality and enhances parameter estimation accuracy, particularly in high-dimensional datasets.

Chapter 3 extends dynamic portfolio choice methodologies under Expected Utility (EU) frameworks to incorporate investors' quantile preferences, focusing on specific quantiles of the returns distribution. Through empirical applications and simulations, this chapter demonstrates the effectiveness of constructing optimal portfolios under quantile preferences with multiple conditioning variables, showcasing superior performance during market crises.

Chapter 4 proposes a new approach to backtesting risk measures by introducing a univariate score function that combines the marginal/conditional score functions for forecasting Value-at-Risk (VaR) and Systemic Risk (SR). This method ensures an equitable and comprehensive assessment of both risk measures, overcoming the limitations of existing methods that prioritize one measure over the other based on the equality of VaR across models. Furthermore, Chapter 4 conducts a comparative analysis of different identification functions for backtesting, including the one-dimensional function proposed by Banulescu-Radu et al. (2021) and the two-dimensional function introduced byFissler and Hoga (2023), to evaluate the potential risk associated with employing identification functions that are not strictly defined for backtesting purposes.

Overall, this thesis contributes to advancing risk management and decision-making methodologies by providing robust and practical strategies.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Signed:..... Date:.....

Acknowledgements

First and foremost, I thank Allah, the Most Merciful and the Most Compassionate, who makes all things possible for His blessings given to me during this journey and in completing my thesis.

I would like to express special thanks to my supervisor, Prof. Jose Olmo, for his constant support, valuable feedback, and discussion. Prof. Olmo exemplified for me the scientific spirit of a true scholar. It was a privilege to be supervised by Prof. Olmo with the incredible human side that helped to overcome challenges throughout this journey. Many thanks to Prof. Zudi Lu and Prof. Huifu Xu for their guidance and support.

I extend sincere love and thanks to my Mother for her daily prayers, support, and unconditional love, who taught me to see the light at the end of the tunnel. With her endless love, I am blessed.

I would also like to express my special thanks, deepest gratitude, and love to my husband, Zaid, and my children, Sara, Reema, and Azzam, for their encouragement, patience, unconditional love, and belief in me when I doubt myself. My gratitude also extends to my siblings and friends for their prayers, love and support.

Lastly, I dedicate this thesis to the spirit of my father, whom I dearly wish were alive to share in my success.

Chapter 1

Introduction

1.1 Motivation

Modern statistical analysis and financial modelling fundamentally require robust approaches that can manage high-dimensional data, capture dependencies among variables, and effectively evaluate forecasting models. This thesis addresses these crucial requirements by developing and applying advanced statistical techniques in diverse areas of risk management and financial modelling. Specifically, the thesis focuses on three main topics: (1) the development of Robust Model Averaging MArginal Regressions (RMAMAR) for nonparametric time series analysis with high-dimensional data, (2) the construction of dynamic optimal portfolios using Quantile Preferences (QP) with multiple conditioning variables, and (3) the testing and comparison of forecasting models considering both Value at Risk (VaR) and systemic risk (SR) components. Each topic contributes unique insights and methodologies to advance the field of statistical analysis and financial modelling, with the goal of providing robust and effective solutions to complex challenges in risk management and financial decision-making. Detailed motivations, with references to relevant literature, are available in the introductory sections of each chapter, detailing the rationale behind the research conducted within that specific chapter. The motivation of each chapter are as follows:

Chapter 2: Robust Model Averaging MArginal Regressions (RMAMAR) for Nonparametric Time Series Analysis with high-dimensional data:
In the Second chapter of this thesis, we investigate nonparametric time series analysis, where dealing with the curse of dimensionality is a major challenge for achieving accurate estimation and forecasting. In the presence of high-dimensional predictors, traditional nonparametric methods often provide poor estimation due to the exponential increase in required sample sizes Fan and Yao (2003). Recently, Li et al. (2015) proposed a semiparametric model, namely Model Averaging MArginal

Regressions (MAMAR), to address this challenge of nonlinear time series analysis, particularly in cases where the conditioning information involves a large number of lags. MAMAR provides a flexible prediction method by averaging a collection of marginal nonparametric models, utilizing suitable weights derived through least squares minimisation. However, real-world datasets often exhibit outliers, which are observations that significantly deviate from the majority of the data. Additionally, these datasets may contain heavy-tailed distributions or deviate from assumed distributional forms. The presence of outliers, heavy tails, or distributional deviations can significantly affect the accuracy of model estimates if the estimator is sensitive to such phenomena.
Although least squares estimators are the most popular classical regression techniques for estimating a model's parameters, these traditional estimators are extremely sensitive and thus easily affected by outliers Tukey (1960), Huber (1964) and Hampel (1968). This sensitivity can lead to poor performance, particularly in scenarios like financial data where deviations are common occurrences. This chapter introduces the Robust Model Averaging Marginal Regressions (RMAMAR) procedure, motivated by the necessity for robust estimation techniques. RMAMAR aims to forecast the future using conditional time series regression, incorporating a high-dimensional vector of lagged predictors. By robustly combining one-dimensional marginal regression functions, RMAMAR minimises the impact of extreme events and enhances the accuracy and robustness of parameter estimation within high-dimensional settings. The approach employs local linear estimation and robust estimation techniques, such as M-estimators, to achieve these objectives.

Chapter 3: Large-Dimensional Dynamic Optimal Portfolio Construction with Quantile Preferences:
In Chapter 3, we address the demand for robust methodologies by focusing on dynamic portfolio selection, a crucial aspect of risk management and investment strategy, particularly in scenarios with multiple conditioning variables. In practice, choosing a dynamic portfolio often involves considering numerous conditioning or forecasting variables that capture changes in the investment opportunity set over time. Various methods have been proposed to characterise the dependency of the portfolio decision on a set of conditioning variables such as Brandt (1999). Nevertheless, dealing with data that have a high number of dimensions frequently presents difficulties in optimizing portfolios Fan and Yao (2003), and the need to capture dependencies among variables is crucial. Chen et al. (2016) suggested a novel data-driven approach under the Expected Utility (EU) framework to estimate the optimal portfolio considering multiple conditioning variables. Chen et al. (2016) 's work is partly motivated by the Model Averaging MArginal Regression (MAMAR) method introduced by Li et al. (2015). Motivated by Chen et al. (2016), Chapter 3 departs from the EU framework and investigates investors' quantile preferences, by focusing on specific quantiles of the returns distribution rather than just the mean.

The chapter aims to develop methodologies that not only overcome the challenges posed by high-dimensional data in portfolio choice under quantile preferences but also offer practical and robust frameworks for optimizing portfolio allocations. By extending model averaging techniques to dynamic portfolio selection and utilizing QP along with multiple conditioning variables, our proposed methodology offers a robust approach to constructing dynamic optimal portfolios.

Chapter 4: Testing and Comparing Forecasting Models for VaR and Systemic Risk: Given the significance of systemic risk measures, developing accurate statistical assessment tools for evaluating different models' predictive performance becomes crucial. Introducing such tools, known as backtesting in finance, is the core objective of Chapter 4. Due to the existence of multiple alternative risk forecasting models, the backtesting tries to make forecast comparisons and rank models, and this is similar to the model selection procedure in statistics, known as comparative backtesting. In the light of the comparative backtesting, Fissler and Hoga (2023) present a valuable framework for comparing risk forecasts by introducing the concept of multi-objective score function equipped with the lexicographic order. Based on that, Fissler and Hoga (2023) proposed Diebold–Mariano One and a Half-Sided test with bivariate scores to compare risk measures forecasting.
However Fissler and Hoga (2023) methodology has a significant limitation. The lexicographic order approach employed in their tests imposes a constraint on the evaluation and comparison of risk measures. Specifically, the assessment is restricted to either VaR or SR conditioning on the equality of VaR across the models under consideration. This constraint implies that the evaluation and comparison of the risk measures cannot be conducted on an equal footing.
Motivated by the work of Fissler and Hoga (2023), in Chapter 4, we propose an alternative univariate score function to test and compare models' forecasting by considering the forecasting of both VaR and SR components without implementing the lexicographic order. This proposed score function is designed to evaluate both risk measures, VaR and SR, on an equitable basis unlike the method of Fissler and Hoga (2023). Our univariate score function is defined as the sum of the marginal/conditional score functions for forecasting VaR and SR. By defining our score function as the sum of these individual score functions, we ensure that both risk measures are assessed simultaneously and on an equal footing, without the need to prioritize one over the other based on the equality of VaR across models. This alternative score function avoids the restrictive assumption of equal expected values for the first risk measure, which is a prerequisite for the lexicographic order approach.
Following the introduction of the motivations behind each chapter in the thesis, we outline the chapters in detail, along with the contributions, in the subsequent section.

1.2 Outline and Contributions of the Thesis

This section offers an overview of the thesis structure and highlight the contributions to the literature. The thesis consists of five chapters.
In Chapter 2, we propose the Robsut Model Averaging MArginal Regressions (RMAMAR) procedure, which robustly approximates the conditional regression function by combining one-dimensional marginal regression functions in an affine manner. Initially, we employ the local linear estimation technique to estimate the conditional marginal regressions. Then, we combine the fitted conditional marginal regressions using weight parameters determined through M-estimators to enhance the robustness of the parameter estimation. Hence, our contributions in Chapter 2 are as follows; First, in this proposing approach, we enhance the robustness of the MAMAR approach by employing robust M-estimation techniques, which are less sensitive to outliers compared to least squares methods used in Li et al. (2015). Moreover, our RMAMAR addresses the curse of dimensionality, a common challenge in statistical modelling, particularly in high-dimensional datasets. By incorporating one-dimensional marginal nonparametric regressions, similar to MAMAR approach proposed by Li et al. (2015), we overcome the challenges associated with high-dimensional conditional regression functions. Furthermore, we establish the theoretical foundation for our robust estimator by demonstrating its uniform convergence under some mild conditions. In addition, we investigate the performance of our RMAMAR approach through Monte Carlo simulations and empirical application; in the simulation section, we conduct three simulation examples across various scenarios and distributional assumptions. Our findings show that the proposed RMAMAR approach performs accurately in the presence of high-dimensional predictors, and it consistently outperformed other competitive models, such as MAMAR approach and nonlinear additive models, demonstrating its robustness in various settings. In the empirical application, our approach performs well under the presence of heavy tails and the occurrence of abnormal observations. Moreover, the results highlight RMAMAR ability to uncover nonlinear lag effects compared to the linear Autoregressive model. In addition, RMAMAR provides the lowest Mean Absolute Error (MAE) in both in-sample and out-of-sample analysis, underscoring its superior predictive performance compared to other models conducted in the study.
The Chapter is structured as follows: Section 2.2 describes our proposed Robust Model Averaging MArginal Regression (RMAMAR). In Section 2.3, we establish the theoretical foundation for our robust estimator by demonstrating its uniform convergence under some mild conditions and deriving the estimator consistency. Section 2.4 introduces Monte Carlo simulation to examine our proposed RMAMAR method in understanding time series lag effects in applications, where three simulation examples following various scenarios are introduced. As an empirical

application of the proposed robust method, we analyse rainfall in North London data in Section 2.5. Section 2.6 gives the conclusion of the Chapter.
In Chapter 3, we extend the semiparametric dynamic portfolio choice under EU framework proposed by Chen et al. (2016) by incorporating investors' quantile preferences, focusing on specific quantiles of the returns distribution rather than just the mean return. To the best of our knowledge, no prior research has been conducted to avoid the curse of dimensionality issue on the portfolio choice problem under the QP setting when the number of conditioning variables is large. Thus, our contributions are threefold; first, we construct dynamic optimal portfolios using quantile preferences with multiple conditioning variables. This is accomplished in two stages. In the first stage, we derive the optimal portfolio weights based on single conditioning variable X_j in the information set, where $j = 1, ..., p,$. This involves, for each j-th conditioning variable $X_{j,t-1} = x_j$ and a given $\tau \in (0, 1)$, we obtain the optimal portfolio weights using quantile regressions. In the second stage, the optimal portfolios constructed from the individual conditioning variables are combined by a model averaging approach to obtain an optimal portfolio based on multiple conditioning variables. This approach is similar in spirit to the MAMAR method proposed in Chen et al. (2016) but adapted to a quantile setting. The second contribution is to show that the proposed quantile method based on conditioning variables effectively captures investor's downside preferences during crises such as the 2007 financial crisis and the COVID-19 pandemic. This is achieved by means of an empirical application to a diversified portfolio comprising six major financial assets such as bonds, gold, and stock indexes. The optimal portfolios obtained from the conditional quantile regression approach outperform the unconditional counterpart portfolio strategies under different metrics in out-of-sample settings. The third contribution is to show the ability of this investment strategy to construct optimal portfolios under quantile preferences in large dimensions.
The Chapter is structured as follows. The methodology for estimating the dynamic portfolio choice under the QP setting is described in Section 2. This estimation can be accomplished in two steps. In the first step, we select the marginal optimal portfolio weights $w_j(x_j)$ under QP by maximising the conditional quantile portfolio problem. We combine the marginal optimal portfolio weights in the second step through the model averaging approach. A description of the out-of-sample performance methodology, along with a review of the approaches considered in the evaluation, is provided in Section 3. Section 4 implements our approach to construct optimal portfolios under QP with multiple covariates in two empirical exercises. The first exercise focuses on a small portfolio allocation problem that comprises six major financial assets such as bonds, gold, and stock indexes and six conditioning variables. In contrast, the second portfolio allocation problem considers all the assets traded in the FTSE100 over the last 16 years. The Chapter is concluded in Section 5.
In Chapter 4, the lexicographic order approach employed in the One and a Half-Sided

tests proposed by Fissler and Hoga (2023) imposes a constraint on the evaluation and comparison of risk measures. Specifically, the assessment is restricted to either VaR or SR , conditioning on the equality of VaR across the models under consideration. To address this limitation, our contributions in this Chapter are the following: First, we propose an alternative univariate score function to test and compare models' forecasting by considering the forecasting of both VaR and SR components without implementing the lexicographic order in our score function, which goes beyond the backtesting implemented by Fissler and Hoga (2023). Our proposed approach introduces a univariate score function that combines the marginal/conditional score functions for forecasting VaR and SR measure. By defining our score function as the sum of these individual score functions, we ensure that both risk measures are assessed concurrently and on an equal footing, without the need to prioritize one over the other based on the equality of VaR across models. Consequently, our proposed univariate score function addresses the limitation of the existing methodology by providing a more equitable framework for risk forecast comparison.
Moreover, to assess the risk of employing non strict identification function in backtesting, we conduct a comparative analysis of our identification function along with the one-dimensional identification function introduced by Banulescu-Radu et al. (2021) in addition to the two-dimensional identification function introduced by Fissler and Hoga (2023). Specifically, we evaluate the power of the three identification functions, in identifying misspecified systemic risk forecasts under various sample sizes and distributional scenarios. A misspecified risk measure is defined as a measure that fails to capture the true tail risk dynamics over time accurately. Through our analysis, we demonstrate the superiority of our proposed identification function compared to Banulescu-Radu et al. (2021) identification function. In Particular, the backtest proposed by Banulescu-Radu et al. (2021) exhibits a complete loss of power in distinguishing between correct and misspecified forecasts. In contrast, our identification function, aligning with Fissler and Hoga (2023) identification function, successfully identifies misspecified forecasts almost with certainty with almost 100% across different distributional assumptions and sample sizes. These results underscore the power and consistency of our proposed identification function in detecting misspecified systemic risk forecasts, outperforming Banulescu-Radu et al. (2021) method. Furthermore, through a comprehensive simulation analysis, we show that the risk measures VaR and SR are elicitable under our joint score function and our identification function is strict and has the ability to detect the misspecified forecasting models with almost certainty.
Chapter 4 is constructed as follows: Section (4.2) introduces the systemic risk measures CoVaR and CoES. Section 4.3 discusses backtesting systemic risk measures, reviewing the multi-objective score function by Fissler and Hoga (2023) and proposing an alternative univariate score function, which includes weighted extensions. Furthermore, we present in (4.3.3.3) a comparative analysis of our identification

function with the one-dimensional function by Banulescu-Radu et al. (2021) and the two-dimensional function by Fissler and Hoga (2023). In Section (4.3.3.4), we investigate the elicitability property of our proposed score function, $S^{(\text{VaR+SR})}$, where SR represents systemic risk, through comprehensive simulations. Monte Carlo simulations in Subsection (4.4.1) examine the finite-sample performance of our test $\mathcal{T}_n^{OS.uni}$ when testing the null hypothesis $H_0 : \mathrm{E}[\bar{d}_n^{uni}] \leq 0$ under different scenarios, where d_n^{uni} represents the score differences. We apply our test to the forecasts of (VaR, CoVaR) and (VaR, CoVaR, CoES) obtained from a bivariate GARCH(1,1) model with a t-copula for innovations, and the time-varying correlation ρ_t follows the Generalized Autoregressive Score (GAS) model. One-step-ahead forecasts are the focus throughout our analysis. The simulation results indicate the following: Firstly, the power of our tests increases significantly as the sample size (n) increases. Secondly, in general, it is easier to detect differences in predictive ability of two models when considering scenarios where both VaR and SR forecasts are accurate for one model while being misspecified for the other. Thirdly, even when introducing a small difference in the predictive accuracy of the VaR forecasts, our tests exhibit the power to identify this differentiation, even in cases where the systemic risk forecasts demonstrate comparable forecasting performance. This finding highlights the sensitivity of our tests in detecting marginal differences in model forecasting. Furthermore, across the various scenarios examined in this study, our tests generally demonstrate superior performance compared to those proposed by Fissler and Hoga (2023) in evaluations incorporating both (VaR + CoVaR) and (VaR + CoVaR + CoES). Finally, comparisons involving CoVaR and CoES exhibit higher power compared to those depending on CoVaR only, potentially attributed to the richer informational content offered by the CoES component, increasing the overall power of the analysis.
In (4.4.2), an empirical application focuses on the daily log-returns of the DAX 30 index, using the daily log-returns of the S&P 500 index as a reference variable. The study compares systemic risk forecasts obtained from two copula models: a benchmark Gaussian copula model and a t-copula model. In both models, the correlation parameter, capturing the dependence structure between the two indices, is modeled using the dynamic GAS framework proposed byCreal et al. (2013). Our findings indicate that the t-copula exhibits superior predictive performance, supported by p-values of 0.005 for (VaR,CoVaR) and 0.0204 for (VaR,CoVaR,CoES), aligning with its popularity in empirical studies. The Chapter concluded in Section (4.6).
In Chapter 5 we conclude the thesis.

Chapter 2

Semiparametric Robust Averaging

2.1 Introduction

In recent years, for forecasting the future, econometricians and statisticians have been challenged by the large number of variables as well as the ambiguity surrounding the functional forms in regression and time series analysis. When conducting regression analysis, selecting from a large number of potential explanatory (covariates) variables presents a significant challenge. The challenge becomes even worse in time series analysis because we consider all possible lags of all possible predictor variables as candidate variables during the estimation and forecasting steps. This complexity underscores the need to carefully investigate the process of constructing our forecasting models, particularly in determining ”what” variables we should include and ”how” they are related to the dependent variable. Various methods have been introduced in the literature to address the challenge of determining ”what” variables to include in time series models to identify the optimal fitting without overfitting, especially when faced with numerous lagged observations. Model selection procedures were introduced first in the literature to address this challenge. In regression and time series analysis, the model selection technique is defined as the process of identifying the optimal model that includes only necessary variables based on specific criteria from a set of candidates. Model selection (variable selection) procedures recently gained popularity with the rise of econometric and statistical models known as high-dimensional models that include a large number of time series lagged observations. For instance, in fields such as labour economics and financial econometrics, researchers can have situations where regression models may include hundreds or even thousands of regressors Belloni and Chernozhukov (2011). In this case, including a large number of lagged observations will lead to increased complexities in modelling and inference as selecting the most relevant variables becomes crucial. A considerable amount of literature introduces model selection procedures and criteria proposed according to some optimality considerations, such

as the Akaike Information Criterion method (AIC) Akaike (1973), the Mallows's Cp method (Mallows (1973)), the Bayesian Information Criterion method (BIC) Schwarz (1978), the generalized AIC method Nishii (1984), the cross-validation(CV) method Stone (1974), the generalized CV (GCV) method (Craven and Wahba (1978)) and others, see. Shao (1997) and Ding et al. (2018) for references and a thorough overview of various model selection techniques.
Although model selection techniques have been widely employed in various fields, in many cases, when conducting statistical analysis, it is uncommon to find one single solution, and this, due to factors such as the limited amount of data available, leads to different combinations of predictors that could be all reasonable explanations for our model. In addition, the existence of various estimation and forecasting approaches in the literature might give slightly different results and conclusions; for more details about that, the reader is referred to Hastie et al. (2009) and Kuhn et al. (2013).
Furthermore, Selection from several candidate models relies on various criteria, such as AIC or BIC, and this variability in the selection procedure may yield a different selected model, introducing uncertainty into the selection process. For more discussion on model uncertainty see Chatfield (1995), Draper (1995) and Yuan and Yang (2005). Pötscher (1991), demonstrated that using AIC for model selection can result in biased inference. Subsequent studies by Kabaila (1995) investigated the negative impacts of model selection on confidence intervals and prediction intervals. References Leeb and Pötscher (2003) and Leeb and Pötscher (2006) discuss the limitations of post-model-selection estimators, which are obtained by first selecting a model (e.g., using AIC) and then estimating the parameters in the selected model using the same data set. They show that it is impossible to estimate the unconditional distribution of these estimators with reasonable accuracy, even asymptotically. In particular, they prove that no estimator for this distribution can be uniformly consistent, not even locally.. Traditional model selection methods can become time-consuming when dealing with a large candidate models and lagged predictor variables, as they involve exploring every possible combination to identify the best-fitting model.
Considering the limitations of model selection discussed above, model averaging was introduced in the literature as an alternative to overcome the difficulties in choosing between a large number of covariates (lagged predictor variables) and the problem of dimensionality in both regression and time series settings. Model averaging combines information from multiple candidate models and weighs them based on some criterion to obtain more accurate and robust estimates of parameters or predictions. By including multiple models and weighting their contributions based on their predictive performance or other criteria, model averaging offers a flexible approach to help reduce the uncertainty and variability that appear in model selection with possibly reducing modelling biases Peng and Yang (2022). Rather than depending mainly on a single model, model averaging considers the insights derived from multiple models,

giving each model a certain level of weight based on its performance. The idea of combining all candidate models instead of depending on one final model to obtain improvements in forecasting was introduced first by Bates and Granger (1969). There exists an extensive body of literature on Bayesian model averaging, including Hoeting et al. (1999), Raftery et al. (1997) and Hjort and Claeskens (2003) for an overview of Bayesian model averaging. However, more recent attention has focused on constructing optimal model averaging weights for frequentist models. For instance, the Mallows model averaging introduced by Hansen (2007) selects the model weights by minimising Mallows' criterion over a set of discrete weight values. Wan et al. (2010) provide more robust theoretical results for using Mallow's criterion, which concentrated on two of Hansen (2007) 's assumptions. To allow for heteroskedasticity models, Hansen and Racine (2012) introduced a jackknife model averaging approach where weights are selected to minimize a leave-one-out cross-validation criterion function. They established that the proposed estimator achieves the minimum asymptotic squared error. Zhang et al. (2013) extend the jackknife model averaging framework to models with dependent data. Zhang et al. (2016) investigate optimal model averaging estimators for generalised linear models. For recent surveys on variable selection and model averaging, one can refer to the surveys by Claeskens and Hjort (2008) and Fan and Lv (2008, 2010) and references therein.
Almost all mentioned studies are averaging under the parametric dynamic structure. Although parametric models have gained substantial attention due to their simplicity, nonparametric procedures based on data have been widely utilised to address the challenge of "how" to determine the unknown functional forms of econometric models. Compared to parametric models, nonparametric procedures, such as kernel smoothing methods, local polynomial regression and smoothing splines, offer less structural restriction and more flexibility in capturing the complex relationship between dependent and explanatory variables. See, for instance, (Li and Racine (2023),Ullah and Pagan (1999)) for discussions on kernel smoothing procedures, and (Fan (2018), Fan et al. (1995)) for local polynomial regression, Eubank (1999)) for coverage spline methods.
Let (Y_t, X_t^T) be stationary time series data, where Y_t represents the response of the time series data and $X_t = (X_{t1}, \cdots, X_{td})^T$ represents a d-dimensional random vector containing the available information up to time t-1. The components of X_t may consist of lagged values of the response Y_t itself, as well as various time series predictor variables. Consequently, the dimensionality d of X_t can become substantially large, as observed in Li et al. (2015) and in practical scenarios. In many applications, we are interested in estimating/forecasting the regression function
$E(Y_t|X_t = x), x = (x_1, \cdots, x_d)^T$. The nonparametric methods can effectively estimate regression functions $E(Y_t|X_t = x), x = (x_1, \cdots, x_d)^T$ when the dimension "d" is small. However, it is widely recognized that the performance of nonparametric estimation diminishes when applied directly in high-dimensional settings, especially as the

dimension "d" increases, particularly $d > 3$. This deterioration in performance can be attributed to the exponential increase in the required sample size to attain estimation accuracy comparable to that of a one-dimensional function. This reduction in the performance as the dimensionality escalates is known as the "curse of dimensionality" see (Stone (1980, 1982)), Ibragimov and Hasminskii (1981) and Silverman (2018). Therefore, direct utilization of nonparametric methods may not be recommended in practice as it could lead to the potential curse of dimensionality, leading to poor estimation and forecasting. As a result, several nonlinear and nonparametric approaches have been introduced in the literature to avoid the curse of dimensionality, such as varying coefficient models, additive models and partially linear models, see Fan and Yao (2003), Teräsvirta et al. (2010), Gao (2007) and Li and Racine (2023) for comprehensive reviews. As we mentioned earlier, in the context of time series analysis, the problem becomes more pronounced where the conditioning information $E(Y_t|Y_{t1}, \cdots, Y_{td})$ may include an infinite number of lags, i.e., $d = \infty$ Li et al. (2015). To address this challenge of nonlinear time series analysis, especially when dealing with a large number of lagged conditioning variables, several semiparametric methods have been proposed, such as Linton and Mammen (2005, 2008), Linton and Sancetta (2009) and Chen and Ghysels (2011). However, these proposed approaches are computationally intensive (refer to aforementioned references). The reader is referred to Li et al. (2015) for further discussions and to explore the limitations of these proposed models.

Recently, Li et al. (2015) addressed the limitations of the mentioned earlier methods and proposed a semiparametric model, namely Model Averaging MArginal Regressions (MAMAR), for forecasting the regression function $E(Y_t|X_t = x), x = (x_1, \cdots, x_d)^T$ using conditional time series regression by utilizing a high-dimensional vector of lagged predictors. MAMAR provides a flexible prediction approach by averaging a set of marginal nonparametric models with appropriate weights assigned by minimising the least squares criterion. In the MAMAR method, the individual marginal regression functions are estimated by a nonparametric kernel estimation technique. Motivated by Li et al. (2015), Chen et al. (2016) extended MAMAR to dynamic portfolio choice problem. Huang and Li (2018) and Chen et al. (2018) extended the MAMAR method to panel data by providing the asymptotic results of the proposed procedure and to ultra-high dimensional time series data, respectively. Li et al. (2018) extended MAMAR by constructing a varying coefficient regression model and Peng and Lu (2021) explored the idea of model averaging marginal logistic regressions on a binary time series classification.

Traditional statistical approaches often rely on the assumption that observations follow a normal distribution. However, this assumption is frequently violated in real-world datasets, which exhibit deviations from normality, due to several reasons such as the presence of heavy-tailed distributions and the occurrence of abnormal observations, commonly referred to as outliers. Heavy-tailed distributions,

characterized by a higher probability of extreme values compared to the normal distribution, are a common occurrence in various fields such as finance, insurance, and risk management. Additionally, real-world datasets often contain abnormal observations, or outliers, observation deviate significantly from the rest of the data. These outliers can appear due to various factors, such as measurement errors, data entry mistakes, or genuine but infrequent events. Ignoring or improperly handling these events can lead to biased estimates and the accuracy of the model estimates can be significantly affected as highlighted by Tukey (1960). For more discussion on the impact of normality assumption violation, the reader is referred to Atkinson (1985), Belsley et al. (2005) and Gujarati et al. (2003), among others.

Least squares estimators are the most popular classical regression techniques for estimating the parameters of a model. However, least square estimators are extremely sensitive and thus easily affected by outliers and heavy-tailed distributions Tukey (1960), Huber (1964) and Hampel (1968). This sensitivity often leads to poor performance, particularly in scenarios where deviations are common, such as in financial data. Consequently, it underscores the necessity for robust estimation techniques like M-estimators (generalizations of a Maximum Likelihood estimator), which aim to minimise the impact of these extreme events. For comprehensive reviews of M-estimators, see De Menezes et al. (2021), Peracchi (1990) and van der Vaart et al. (1996).

Although the MAMAR technique performs well in out-of-sample prediction, it is expected to be highly influenced by outliers or/and heavy-tailed distributions, which will reduce its efficiency when dealing with commonly used non-normal errors. To address the limitation of the MAMAR approach and its sensitivity to outlying observations and extreme cases, De Gooijer and Zerom (2019) introduced a computational approach for obtaining quantile predictions based on penalised high dimensional quantile averaging. Tu et al. (2021) proposed a semiparametric model averaging prediction under independent and identically distributed (i.i.d) data. Zhan et al. (2023) developed a model averaging approach to estimate the conditional quantiles utilising a collection of semiparametric varying coefficient models by adopting a uniform structure across all models, and each sub-model contains only a single varying coefficient.

Therefore, motivated by Li et al. (2015), De Gooijer and Zerom (2019), Tu et al. (2021) and Zhan et al. (2023), we propose a procedure called Robsut Model Averaging MArginal Regressions (RMAMAR) that aims to approximate the conditional time series regression by robustly combining one-dimensional marginal regression functions in an affine manner. Initially, we employ the local linear estimation technique to estimate the conditional marginal regressions. Then, we combine the fitted conditional marginal regressions using weight parameters determined through M-estimators to enhance the robustness of the parameter estimation.

Thus, our contributions in this Chapter are as follows: we enhance the robustness of

the MAMAR approach by employing robust M-estimation techniques, which are less sensitive to outliers compared to least squares methods used in Li et al. (2015). Moreover, our RMAMAR addresses the 'curse of dimensionality', a common challenge in statistical modelling, particularly in high-dimensional datasets. By incorporating one-dimensional marginal nonparametric regressions, similar to the MAMAR approach proposed by Li et al. (2015), we overcome issues associated with high-dimensional conditional regression functions. Furthermore, we establish the theoretical foundation for our robust estimator by demonstrating its uniform convergence under some mild conditions and deriving the estimator consistency. In addition, through extensive Monte Carlo simulations and real data analysis across various scenarios and distributional assumptions, our findings show that the proposed RMAMAR approach performs accurately in the presence of high-dimensional predictors, and it consistently outperformed alternative methods such as the MAMAR approach and nonlinear additive models. In the empirical application, our approach performs well under the presence of heavy tails and the occurrence of abnormal observations. Moreover, the results highlight RMAMAR ability to uncover nonlinear lag effects compared to the linear Autoregressive model. In addition, RMAMAR provides the lowest Mean Absolute Error (MAE) in both in-sample and out-of-sample analysis, underscoring its superior predictive performance compared to other models conducted in the study.

The Chapter is structured as follows: Section 2.2 describes our proposed Robust Model Averaging MArginal Regression (RMAMAR). In Section 2.3, we establish the theoretical foundation for our robust estimator by demonstrating its uniform convergence under some mild conditions and deriving the estimator consistency. Section 2.4 introduces Monte Carlo simulation to examine our proposed RMAMAR method in understanding time series lag effects in applications, where three simulation examples following various scenarios are introduced. As an empirical application of the proposed robust method, we analyse rainfall in North London data in Section 2.5. Section 2.6 gives the conclusion of the Chapter.

2.2 RMAMAR Approach

The model averaging is useful for prediction by fitting multiple candidate models and assigning higher weights to the most optimal candidate models. This strategy enhances prediction accuracy and decreases the risk of model misspecification (Yang (2001),Hansen (2007) and Hjort and Claeskens (2003)). Let $(Y_t, X_t^T), 1 \leq t \leq n$ be stationary time series data, where Y_t represents the response of the time series data and $X_t = (X_{t1}, \cdots, X_{td})^T$ represents a d-dimensional random vector containing the available information up to time t-1. The components of X_t may consist of various predictor variables, including lagged values of the response variable Y_t itself.

Consequently, the dimensionality d of X_t can become substantially large, as observed in Li et al. (2015) and in practical scenarios.
To avoid the curse of dimensionality, mainly when the dimension of the covariates exceeds ($d > 3$), Li et al. (2015) introduced a semiparametric model for forecasting the multivariate regression function $E(Y_t|X_t = x)$, where$x = (x_1, \cdots, x_d)^T$, by employing conditional time series regression and incorporating a high dimensional vector of lagged predictors. Li et al. (2015) proposed a flexible time series prediction approach called MAMAR, wherein they approximate a multivariate regression function by averaging a collection of nonparametric models, each representing a one-dimensional marginal regression function. The weights for this approximation are estimated by minimising the least squares. This method estimates the individual marginal regression functions through a nonparametric kernel estimation technique. In the following subsection, we first describe the MAMAR approach, including the estimation procedures and employing model averaging by combining these estimates. Then, we introduce our proposed RMAMAR.

2.2.1 MAMAR Approach

In the MAMAR approach, we approximate the conditional mean function $m(x) = E(Y|X = x)$ by a linear combination of one-dimensional regression functions. Such that we approximate

$$\begin{aligned} &m(x) = E(Y|X = x) \text{ by} \\ &m_\omega(x) = \sum_{j=1}^{d} \omega_j m_j(x_j) \end{aligned} \tag{2.1}$$

Here, $m_j(x_j) = E(Y|X_j = x_j)$ represents the marginal conditional regression functions for each individual predictor X_j, each of these marginal regression functions can be viewed as a nonlinear candidate model, ω_j represents the weights assigned to each regression function $m_j(x_j)$ and d is the dimensionality of the predictor space. Therefore, our objective is to obtain $\omega_{op} = (\omega_{op,1}, \cdots, \omega_{op,d})^T$ that minimizes

$$E[Y - \sum_{j=1}^{d} \omega_j E(Y|X_j)]^2, \tag{2.2}$$

According to equation (2.1), in order to obtain the estimator of $m(x)$, we should estimate the marginal conditional regression functions and obtain the weights. In the MAMAR approach, this process begins with estimating the marginal conditional regressions $m_j(x_j)$ by utilising the Nadaraya-Watson kernel method, which is proposed independently by Nadaraya (1964) and Watson (1964). This method provides a flexible way to estimate the conditional mean by weighting each

observation based on its distance from the point at which the mean is being computed.

$$\widehat{m}_j(x_j) = \frac{\sum_{t=1}^{n} Y_t K(\frac{X_{tj}-x_j}{h_j})}{\sum_{t=1}^{n} K(\frac{X_{tj}-x_j}{h_j})}, \tag{2.3}$$

Where K represents a kernel function, which is a non-negative real-valued smooth integrable function satisfying the normalisation property: $\int K(u)du = 1$, and symmetry $K(-u) = K(u) \quad$ for all $u \in \mathbb{R}$. Table 2.1 provides examples of common symmetric kernel functions, such as the Gaussian kernel, Uniform kernel and the Epanechnikov kernel. The parameter $h_j > 0$ is the smoothing bandwidth, which needs to be appropriately selected to control the width of the kernel function. The reader is referred to Ruppert et al. (1995) and Heidenreich et al. (2013) for a comprehensive review of optimal bandwidth selection with further references.

Kernel	$K(x)$	$\int x^2 K(x)\,dx$	$\int K(x)^2\,dx$
Uniform	$\frac{1}{2}\mathbb{I}(\lvert x\rvert \leq 1)$	$\frac{1}{3}$	$\frac{1}{2}$
Epanechnikov	$\frac{3}{4}(1-x^2)\mathbb{I}(\lvert x\rvert \leq 1)$	$\frac{1}{5}$	$\frac{3}{5}$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$	1	$\frac{1}{2\sqrt{\pi}}$

TABLE 2.1: Comparison of Standard Kernel Functions

Next, after estimating the marginal regression functions $\widehat{m}_j(x_j), j = 1, \cdots, d$ at the sample points, our objective is to combine them by assigning weights $\omega_{op,j}$ to each conditional mean component. These weights represent the contribution of each $\widehat{m}_j(x_j)$ to the overall estimation of the conditional mean function. The weight vector is estimated by minimising the least squares sample objective function given by:

$$\min_{\omega_1,\ldots,\omega_d} \sum_{t=1}^{n} \left(Y_t - \sum_{j=1}^{d} \omega_j \widehat{m}_j(x_{tj}) \right)^2, \tag{2.4}$$

where Y_t is the observed response and $\widehat{m}_j(x_{tj})$ represents the estimated one-dimensional regression functions. Here, ω_j is the weight associated with the j-th predictor variable.
Once the weight vector $\widehat{\omega}_{op} = (\widehat{\omega}_{op,1}, \cdots \widehat{\omega}_{op,d})^T$ is estimated, where each weight component $\widehat{\omega}_{op,j}$ represents the weight assigned to each conditional mean function, we compute the conditional regression function $\widehat{m}(x)$ by combining the estimated marginal conditional regression functions $\widehat{m}_j(x_j)$ with the corresponding weights :

$$\widehat{m}(x) := \widehat{m}_{\widehat{\omega}}(x) = \sum_{j=1}^{d} \widehat{\omega}_{op,j} \widehat{m}_j(x_j) \tag{2.5}$$

where $x = (x_1, ..., x_d)^T$.

Least squares estimators are the most popular classical regression techniques for estimating the parameters of a model. However, least square estimators are extremely sensitive and thus easily affected by outliers, and heavy-tailed distributions see Tukey (1960), Huber (1964) and Hampel (1968). For instance, Tukey (1960) showed that while least square estimators perform well when applying to data sampled from a normal distribution, they tend to perform poorly when the distribution is changed slightly, such as by introducing contamination. This sensitivity often leads to poor performance, particularly in scenarios where deviations are common, such as in financial data. Consequently, it underscores the need for robust estimation techniques like M-estimators (generalisations of a Maximum Likelihood estimator), which aim to minimise the influence of these extreme events. For comprehensive reviews of M-estimators, see Huber (2004), De Menezes et al. (2021), Peracchi (1990) and van der Vaart et al. (1996).
While the MAMAR technique employs least squares and performs well in out-of-sample prediction, its sensitivity to outliers or heavy-tailed distributions may reduce its efficiency when handling commonly used non-normal errors. To address this limitation of the MAMAR approach, we propose RMAMAR, which aims to approximate a multivariate regression function more robustly while combining one-dimensional marginal regression functions in an affine manner. We utilise robust M-estimation techniques, such as the Huber function proposed by Huber (1964), to improve parameter estimation's robustness in such situations. By minimising objective functions designed to handle outliers and deviations more effectively, RMAMAR provides more robust parameter estimates than MAMAR. This robustness ensures that the resulting model accurately captures the underlying relationships in the data, making RMAMAR particularly valuable in scenarios where the data exhibits characteristics such as asymmetry or heavy tails. In summary, adopting RMAMAR as a statistical modelling technique is suitable for various applications across diverse fields where data quality and robustness are critical issues. Before introducing the RMAMAR approach, we present some literature reviews about robust M-estimators.

2.2.1.1 Overview of Robust M-estimators

As mentioned before, least square estimators were conducted widely in the literature; however, Tukey (1960) illustrated that those estimators are extremely sensitive to outliers and heavy-tailed distributions. A single outlier can have a significant impact on the estimates. To overcome the drawbacks of nonrobust estimators, the field of robust statistics emerged in the 1960s with seminal work of Tukey (1960), the article by Huber (1964), and the thesis by Hampel (1968). These contributions and others aim to propose robust regression techniques capable of resisting the effects of outliers and

nonnormality. As part of this effort, Peter Bickel, Peter Huber, and Frank Hampel were invited to collaborate with John Tukey during the academic year 1970-1971 at Princeton University, aiming to advance robust statistics Hampel (1997). In the Princeton robustness study, see Andrews et al. (1972) for more details; the primary objective was to develop compromise estimators capable of performing well across the range between the normal and heavy-tailed distributions. In the field of robust estimation, various classes of estimators such as L-estimators, R-estimators and M-estimators (generalisations of a Maximum Likelihood estimator introduced by Huber (1964)) have been developed to address the challenges of outliers and deviations from standard assumptions. Among these classes, M-estimators offer computational efficiency despite being computationally more expensive Huber (2004),De Menezes et al. (2021) and Prata et al. (2009) among others. In addition, M-estimators offer a simpler and more straightforward approach to robust estimation than L- and R-estimators, as a fixed function determines their shape, simplifying the estimation process. Defining the robust properties upfront can be challenging in L- and R-estimators, as their behaviour may vary depending on the data and estimation procedure. The mathematical structure of M-estimators, including their robust objective functions and influence functions, enables them to effectively handle outliers and deviations in the data, making them robust tools for statistical estimation (Rey (2012), Huber and Ronchetti (2009)). These estimators prioritise the majority of the data clustered around the mean while minimising the impact of outliers, which are typically located far from the mean, through the use of robust objective functions that downweight the impact of outliers relative to the majority of the data.
Instead of estimating parameter β by minimising the sum of squared error as it is done in traditional least squares, the M-estimator replaces that with a robust criterion as follows:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho(x_i, \beta) \tag{2.6}$$

Here, ρ is a robust loss function, and the type of ρ function has been investigated in extensive literature. Table (2.2) presents two popular types of robust loss functions and corresponding derivatives. Furthermore, Figure (2.1) illustrates various loss functions, Least Absolute Deviation (LAD), OLS, and Huber Loss, with different values of k.

Type	$\rho(\varepsilon_i)$	$\psi(\varepsilon_i)$
Huber	$\begin{cases} \frac{1}{2}\varepsilon_i^2 & \text{if } \lvert\varepsilon_i\rvert \le k \\ k\left(\lvert\varepsilon_i\rvert - \frac{1}{2}k\right) & \text{if } \lvert\varepsilon_i\rvert > k \end{cases}$	$\begin{cases} \varepsilon_i & \text{if } \lvert\varepsilon_i\rvert \le k \\ k\,\text{sign}(\varepsilon_i) & \text{if } \lvert\varepsilon_i\rvert > k \end{cases}$
Tukey	$\begin{cases} \frac{k^2}{6}\left(1 - \left[1 - \left(\frac{\varepsilon}{k}\right)^2\right]^3\right) & \text{if } \lvert\varepsilon\rvert \le k \\ \frac{k^2}{6} & \text{otherwise} \end{cases}$	$\begin{cases} \varepsilon\left[1 - \left(\frac{\varepsilon}{k}\right)^2\right]^2 & \text{if } \lvert\varepsilon\rvert \le k \\ 0 & \text{otherwise} \end{cases}$

TABLE 2.2: Comparison of Robust Loss Functions: Different ρ Functions and Corresponding Derivatives ψ for Huber and Tukey Losses, where ε_i denotes the residuals from a regression model, $\varepsilon_i = (x_i - x_i^T\beta)$ where x_i represents the explanatory variables and β represents the regression coefficients.

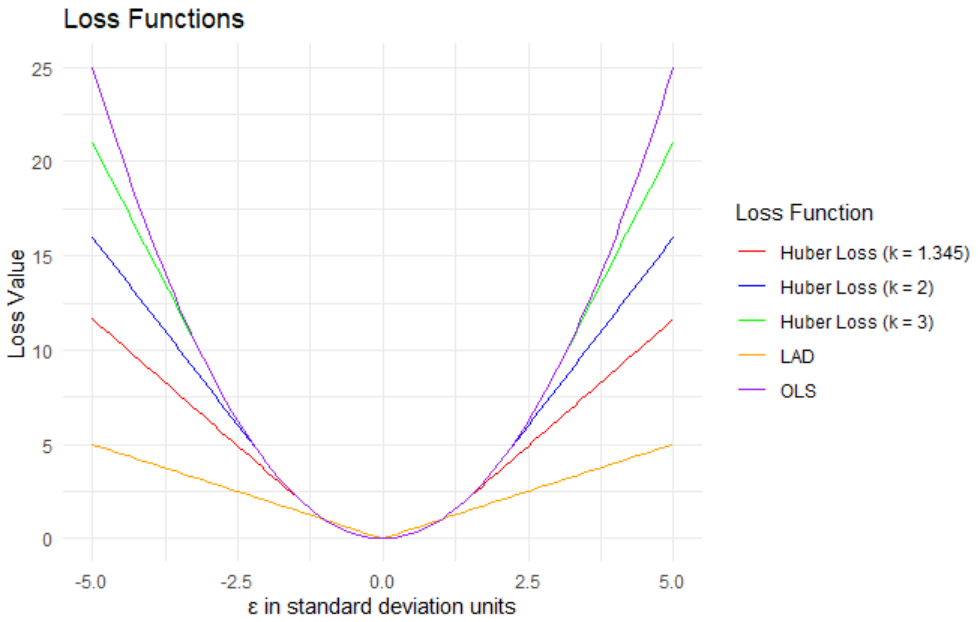


FIGURE 2.1: Plot of various loss functions: Least Absolute Deviation (LAD), Ordinary Least Squares (OLS), and Huber Loss with different values of k.

In the literature, several general properties are considered for ρ to be satisfied: $\rho(\varepsilon \ge 0)$; $\rho(\varepsilon) = \rho(-\varepsilon)$; $\rho(0) = 0$, $\rho(\varepsilon_i) \ge \rho(\varepsilon_j)$ for $|\varepsilon_i| \ge |\varepsilon_j|$ and continuous derivative with respect to the coefficients to enable numerical stability and find the minima.

The M-estimation method assigns smaller weights to observations with outliers, making it more robust to outliers compared to least squares estimation. To find the estimator for β in the M-estimation method, we differentiate (2.6) with respect to β and set the partial derivative to zero:

$$\sum_{i=1}^{n} \psi(\varepsilon_i) x_i = 0 \tag{2.7}$$

where $\varepsilon_i = y_i - x_i^T\beta$ represents the residual for the i-th observation. Here, ψ represents the derivative of ρ. To solve (2.7), we define the weight function $W(\varepsilon) = \frac{\psi(\varepsilon)}{\varepsilon}$ for $\varepsilon \neq 0$, and $W(\varepsilon) = \psi'(0)$ for $\varepsilon = 0$. Substituting the expression for $W(\varepsilon)$ into Eq (2.7), we obtain:

$$\begin{aligned} \sum_{i=1}^{n} W_i \varepsilon_i &= 0 \\ \sum_{i=1}^{n} W_i \varepsilon_i x_i &= 0 \end{aligned} \tag{2.8}$$

where $W_i = W(\varepsilon_i)$. The solution to this estimating equation (2.8), denoted as $\hat{\beta}_n$, can be found by iteratively adjusting the weights W_i and the parameter β until convergence is achieved. This method, known as iteratively reweighted least squares, is employed to estimate the parameters in M-estimation.

A substantial amount of literature has investigated the asymptotic properties of the solution $\hat{\beta}_n$ with different forms of ρ. Several authors consider the assumption e_i is i.i.d and discuss particular choices of ρ or ψ, we start with the seminal work of Huber (1964, 1973), Bai et al. (1990) , Bai et al. (1992) ,Niemiro (1992), He and Shao (1996), Chen et al. (1990) and Arcones (1998) among others. For literature relaxing the i.i.d assumption of the error and allowing different dependent errors see, for example, Gastwirth and Rubin (1975), Lin et al. (2009), Phillips (1991), Cui et al. (2004), Berlinet et al. (2000), Gao et al. (2009) and Babu (1989), among others.

Now, after introducing the M-estimators class, we present our RMAMAR approach. At RMAMAR, we aim to develop a robust M-estimation framework to approximate conditional regression. Similar to the MAMAR approach, we approximate the conditional regression function by a robust linear combination of one-dimensional regression functions, which is described by the following:

$$m_{\rho w}(x) = \sum_{j=1}^{d} \omega_{\rho j} m_{\rho j}\left(x_j\right) \tag{2.9}$$

Here, d is the dimensionality of the covariate space and is assumed to have a fixed value, $m_{\rho j}(x_j)$ represents a regression function for the j-th predictor variable X_j, the subscript ρ represents the robust fitting technique we employ it compared to MAMAR, and $\omega_{\rho j}$ denotes the weight associated with each regression function $m_{\rho j}(x_j)$ in the RMAMAR framework. These weights are based on robust estimation methods, such as M-estimation, with a robust objective function, like the Huber loss function (as defined in Table (2.2)). Therefore, we aim to obtain $\omega_{op} = (\omega_{op,1}, \cdots, \omega_{op,d})^T$ that minimizes

$$E\ \rho\left(Y - \sum_{j=1}^{d} \omega_{\rho j}\ m_{\rho j}(X_j)\right). \tag{2.10}$$

Here, ρ is a robust loss function, such as the Huber loss, penalising the difference between the observed response variable Y. The predicted values, which are calculated based on the regression functions $m_{\rho j}(X_j), j = 1, \cdots, d$ and weighted by the robust weights $\omega_{\rho j}, j = 1, \cdots, d$. This penalty aims to mitigate the impact of outliers and deviations from the assumed distributional form, ensuring the robustness of the estimation process.

The regression functions $m_{\rho j}(X_j), j = 1, \cdots, d$ are unknown but one dimensional, and they can be estimated using nonparametric approaches. A substantial amount of literature on modelling the relationship between the response (dependent) variable and covariate (independent) variables with various linear smoothing techniques to estimate nonparametric regression functions, such as kernel, spline, and local polynomial methods. For a comprehensive review of available methods and findings on both theory and applications, the reader is referred to the books authored by Eubank (1988), Härdle (1990), Wahba (1990), Hastie and Tibshirani (1990), Fan and Gijbels (1996) and Bosq (2012) among others. In recent years, the local linear method has gained popularity among the mentioned linear smoothing techniques due to the method's numerical and theoretical attractive properties, including bias reduction and the method's robustness by its ability to adapt the edge effects compared to the Nadaraya–Watson type estimator utilised in the MAMAR approach, for further details on the methods and results, the reader is referred to Fan (1993), and the book by Fan and Gijbels (1996).

In RMAMAR, for x_j close to X_{tj} we aim to estimate $m_{\rho j}\left(x_j\right)$ in a more robust way by conducting the local linear estimation method. Using Taylor expansion gives,

$$m_{\rho j}(X_{tj}) \approx m_{\rho j}(x_j) + \grave{m}_{\rho j}(x_j)(X_{tj} - x_j)$$

$$\equiv a_j + b_j(X_{tj} - x_j) \quad ; t = 1, ..., n \quad , j = 1, ..., d \text{ and d is fixed,}$$

where $\grave{m}_{\rho j}(x_j)$ is the first order derivative of $m_{\rho j}(x_j)$.

The estimator is defined as $\widehat{m}_{\rho j}(x) = \widehat{a}$ where $\widehat{a}$ and $\widehat{b}$ minimizes

$$\min_{a_j, b_j} \sum_{t=1}^{n} \left(Y_t - \sum_{j=1}^{d} a_j - b_j(X_{tj} - x_j) \right)^2 K(\frac{X_{tj} - x_j}{h_j}), \tag{2.11}$$

where K represents a kernel function and h_j is a chosen bandwidth that was explained earlier.

2.2.1.2 Practical Limitations of Local Linear Estimation Procedures

While local linear estimation is a powerful technique for estimating marginal regression functions in the RMAMAR methodology, offering advantages such as reduced sensitivity to boundary effects compared to the Nadaraya-Watson estimator used in the original MAMAR approach, it is crucial to acknowledge and discuss the practical limitations one may encounter during implementation. One of the key challenges lies in the optimal choice of the bandwidth parameter, which determines the size of the local neighborhood used for fitting the local linear regression at each point. A smaller bandwidth leads to a more flexible fit but may introduce higher variance, while a larger bandwidth produces a smoother fit but may introduce bias. Finding the optimal balance between bias and variance is crucial for accurate estimation. Several bandwidth selection methods, including cross-validation, plug-in methods, and rule-of-thumb approaches, have been proposed, each with its own advantages and limitations. The choice of bandwidth selection method can significantly impact the estimation results, and it is often recommended to try multiple methods and compare the results to assess the sensitivity of the estimates to the bandwidth choice.
Another practical limitation is the potential for boundary bias, which occurs when estimating the regression function near the boundaries of the predictor space, where the local neighborhood used for fitting the local linear regression is asymmetric. This asymmetry can lead to biased estimates at the boundaries. Various techniques have been proposed to mitigate boundary bias, such as boundary kernel methods, local polynomial fitting with higher-order polynomials, and reflection methods, but these techniques may introduce additional complexity and computational burden to the estimation procedure. Addressing the practical limitations of local linear estimation in the context of the RMAMAR methodology is an important area for future research. Further investigations could focus on developing robust and data-driven bandwidth selection methods specifically tailored for the RMAMAR approach, as well as exploring advanced techniques for handling boundary issues, such as adaptive bandwidth selection or boundary correction methods. These efforts will improve the accuracy and reliability of the marginal regression function estimates, enhancing the applicability and performance of the RMAMAR approach in real-world scenarios.

In the next step and after estimating the marginal regression functions at the sample points, we aim to assign weights $\omega_{\rho j}^{op}$ to each conditional mean component $m_{\rho j}(x_j)$. These weights $\omega_{\rho j}$ indicate the influence or importance of each $m_{\rho j}(x_j)$ to the overall estimation of the conditional mean function. The weight vector is estimated by minimising

$$\min_{\omega_1,\cdots,\omega_d} \sum_{t=1}^{n} \rho \left(Y_t - \sum_{j=1}^{d} \omega_{\rho j} \widehat{m}_{\rho j}(X_{tj}) \right), \tag{2.12}$$

where ρ represents the general convex loss function used in M-estimation, Y_t denotes the observed response variable, $\omega_{\rho j}$ represents the theoretical or true weights associated with each predictor variable in the robust linear regression model, and $\widehat{m}_{\rho j}(x_{tj})$ represents the estimated conditional mean function for the j-th predictor variable, obtained through a robust local M-estimation method in the previous step.

2.2.2 Clarification on the Dimensionality of the Auxiliary Regression

The auxiliary regression step in RMAMAR, as shown in Equation (2.12), can be viewed as a regression problem where the observed response Y_t is regressed on the estimated marginal regressions $\hat{m}_j(X_{tj})$ using a robust loss function ρ. It is important to note that while this auxiliary regression involves multiple predictors (the estimated marginal regressions), these predictors are one-dimensional marginal regressions, which can be estimated efficiently using nonparametric techniques without suffering from the curse of dimensionality.
The key advantage of the RMAMAR approach is that it breaks down the high-dimensional conditional regression problem into a series of one-dimensional marginal regressions, which can be estimated reliably using nonparametric methods. The auxiliary regression step is then used to combine these one-dimensional marginal regressions in a robust manner, effectively approximating the high-dimensional conditional regression function. Furthermore, the auxiliary regression step in RMAMAR does not require estimating a high-dimensional regression model directly. Instead, it relies on the estimated marginal regressions, which are low-dimensional and can be obtained efficiently using local linear estimation or other nonparametric techniques.
It is worth noting that the dimensionality of the auxiliary regression is equal to the number of marginal regressions considered, which can be chosen based on domain knowledge or variable selection techniques. In many practical applications, including only the most relevant marginal regressions may be sufficient, further reducing the dimensionality of the auxiliary regression step.
In summary, while the RMAMAR approach involves an auxiliary regression step to estimate the weights for combining the marginal regressions, this step does not require estimating a high-dimensional regression model directly. Instead, it leverages the low-dimensional marginal regressions, which can be estimated efficiently using nonparametric techniques, and combines them in a robust manner through the auxiliary regression. This approach effectively mitigates the curse of dimensionality while providing a flexible and robust approximation to the high-dimensional conditional regression function.
Once the weight vector $\widehat{\omega}_{\rho}^{op}$ is estimated through equation (2.12), where each weight component $\widehat{\omega}_{\rho j}$ in the vector represents the weight assigned to each conditional mean function $\widehat{m}_{\rho j}(x_{tj})$, we combine the estimated one-dimensional regression functions

$\widehat{m}_{\rho j}(x_{tj})$ with their corresponding weights to form the estimated robust conditional regression function, defined as following

$$\widehat{m}(x) := \widehat{m}_{\rho\widehat{\omega}}(x) = \sum_{j=1}^{d} \widehat{\omega}_{\rho j}^{\text{op}} \, \widehat{m}_{\rho j}(x_j) \tag{2.13}$$

where $x = (x_1, ..., x_d)^T$.

2.2.2.1 Considerations for Omitted Variables and Endogeneity in RMAMAR

The RMAMAR approach, introduced as an extension of the MAMAR framework by Li et al. (2015), employs marginal regressions to approximate the conditional regression function. While this approach is motivated by the desire to avoid the curse of dimensionality in high-dimensional time series analysis, it naturally raises concerns regarding the potential impact of omitted variables and endogeneity on the model's performance and the validity of the averaging process.
Omitted variable bias is a common issue in regression analysis, occurring when important predictors are excluded from the model, and these excluded predictors are correlated with the included predictors. In the context of marginal regressions, if the excluded predictors have a significant influence on the response variable and are correlated with the included predictors, it can lead to biased estimates of the marginal effects. The included predictors may partially capture the effects of the omitted variables, resulting in inaccurate estimates of the true relationships.
The presence of omitted variables and high correlation among predictors may have implications for the averaging process in the RMAMAR methodology. If the excluded predictors are strongly correlated with the included predictors, the averaging process may not effectively capture the true underlying relationships, potentially leading to suboptimal performance. The assigned averaging weights to the marginal regressions may not accurately reflect the importance of each predictor, as the effects of omitted variables may be inadvertently incorporated into the weights of the included predictors.
However, it is important to recognize that the RMAMAR approach does not assume that the approximating model using marginal regressions is the true data-generating process. As stated in Li et al. (2015), the marginal regression averaging is viewed as a flexible approximation or model averaging device, particularly useful when dealing with high-dimensional datasets with many potential predictors of unknown functional forms. The key assumption required for the approximation to be exact is a weaker condition that allows for some dependence between covariates, rather than requiring full independence.
Furthermore, the RMAMAR method involves a trade-off between bias and variance. By focusing on marginal regressions, it may indeed introduce some bias due to

omitted variables or endogeneity concerns. However, this potential increase in bias is traded off against a reduction in variance, especially when dealing with high-dimensional data or complex nonlinear relationships. The marginal regression approximation can be considered a pragmatic use of lower-dimensional relationships to build a more complex predictor, without imposing restrictive parametric assumptions. This property arguably provides some robustness compared to misspecified parametric models in the presence of omitted variables.
To mitigate the potential limitations of omitted variables and endogeneity, the RMAMAR approach incorporates robust estimation techniques, such as M-estimation with the Huber loss function. By downweighting the influence of outliers and heavy-tailed distributions, the robust approach aims to mitigate the impact of potential deviations from model assumptions, including the potential effects of omitted variables or endogeneity.
Additionally, to address the issue of omitted variable bias, the concept of "limited correlatedness" among the predictors can be considered as an underlying assumption when using marginal regressions. This assumption suggests that the correlations between the included and excluded predictors should be relatively low. By ensuring that the included predictors are not strongly correlated with the excluded predictors, the effects of omitted variables can be minimized, and the marginal regressions can provide more reliable estimates of the individual predictor effects.
Moreover, variable selection techniques or regularization methods can be employed to identify and include the most relevant predictors or lags in the analysis, minimizing the impact of omitted variables. For instance, Lasso (Least Absolute Shrinkage and Selection Operator) is a popular regularization technique that introduces an L1 penalty term to the objective function, encouraging sparsity and effectively performing variable selection. By selecting a subset of relevant predictors, Lasso can help reduce the impact of omitted variables and improve the interpretability of the model.
It is crucial to acknowledge that the impact of omitted variables and endogeneity on the RMAMAR methodology is an important area for future research. Further investigations could focus on developing robust methods specifically designed to handle these issues and improve the performance of the averaging process in the presence of correlated predictors. This may involve exploring advanced variable selection techniques, incorporating regularization methods to handle high-dimensional data, or developing novel approaches to account for endogeneity in the context of marginal regressions.
In conclusion, while the use of marginal regressions in the RMAMAR methodology raises important considerations regarding the potential impact of omitted variables and endogeneity, the robust nature of the approach, combined with variable selection strategies and the trade-off between bias and variance reduction, aims to mitigate these limitations. By acknowledging these concerns and continuing to explore robust methods and techniques, the RMAMAR methodology can be further enhanced to

provide more reliable and accurate results, ensuring its robustness and effectiveness in various practical applications.
In the next Section, we investigate the properties of our estimator $\widehat{\omega}_{n\rho}^{\text{op}}$ in case the robust marginal regression functions $m_{\rho j}(x_j)$ are known. In the rest of this chapter, we omit ρ from $m_\rho(X_t)$, $\widehat{\omega}_{n\rho}^{\text{op}}$ and drop the suffix n and represent $\widehat{\omega}_{n\rho}^{\text{op}}$ by $\widehat{\omega}$ for simplicity. Before introducing the main Theorem and the model assumptions, we define the α-mixing process.

Definition (α-mixing)

For a stochastic process $X := (X_t, t \in Z)$ on given probability space$(\Omega, \mathcal{F}, P)$, we define $\mathcal{F}_n^l = \sigma\{X_t : n \leq t \leq l\}$ as the σ-algebra generated by the random variables X_t. We define the measure of dependence for any two σ-algebra $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$ as

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

For the stochastic process$\{X_t\}$, and a positive integer m, the coefficient $\alpha(m)$, serves as the dependence coefficient and expressed as

$$\alpha(m) = \alpha(X, m) := \sup_{n \in Z} \alpha(\mathcal{F}_{-\infty}^n, \mathcal{F}_{n+m}^\infty)$$

The stochastic process$\{X_t\}$ is said to be α-mixing if is satisfies $\alpha(m) \to 0$ as $m \to \infty$, this condition proposed by Rosenblatt (1956).
The α-mixing condition is crucial in various limit theorems for stochastic processes, including central limit theorems as employed by Rosenblatt (1956). Further details on limit theorems for α-mixing processes can be found in Bradley (1986) and Lin and Lu (1996). Several literature proposed regression estimation methods under various mixing conditions, such as α-mixing, ρ-mixing, β-mixing, uniform strong mixing (ϕ-mixing) and $\psi-$mixing. These conditions are related, with α-mixing being a weaker and less restrictive assumption compared to others. The following diagram illustrates the relationship between these mixing conditions as discussed in the Bradley (2005) survey.

$$\begin{array}{ccccccc} \psi\text{-mixing} & \Rightarrow & \phi\text{-mixing} & \Rightarrow & \begin{array}{c}\rho\text{-mixing} \\ \beta\text{-mixing}\end{array} & \Rightarrow & \alpha\text{-mixing} \end{array}$$

Doukhan (2012) investigated the relationships between those mixing conditions with examples of models. Moreover, for a survey on various mixing conditions, the reader is referred to Bradley (2005) and Bobbia et al. (2022).

Next, we establish the uniform convergence of our estimator, which assures the estimator's consistency.

2.3 Analyzing Estimator Behavior: Establishing Uniform Convergence

In this section, we introduce Theorem 1, which establishes the uniform convergence properties of our M-estimator $\hat{\omega}$, indicating that as the sample size increases, the estimator converges uniformly to the true parameter value across the entire parameter space.

Before stating the main results, we need to assume the following regularity assumptions on $\rho, m(X_t)$ and e_t. The assumptions are standard assumptions and are often employed for statistical inference to establish the theoretical framework for M-estimation in regression models under α-mixing dependence. see, for instance, Bai et al. (1992), Wu (2007), Gao et al. (2009).
A1. ρ takes the form of a convex function, with important examples such as the Huber loss function. We assume a series of strictly stationary random errors for the error,$\{e_t\}$.
A2. Let ψ represents the derivative of ρ and ψ satisfies the following properties:
(i) As $|s| \to 0$, the expected value of the derivative function $\psi(e_1 + s)$ behave approximately as a linear function of s, with a small error term such as $k_1 s + o(|s|)$ where k_1 is a positive constant.

$$\mathrm{E}(\psi(e_1 + s)) = k_1 s + o(|s|).$$

(ii) For any s approaching zero, the expected square difference between $\psi(e_1 + s)$ and $\psi(e_1)$ is bounded by a function $k_2(|s|)$, where k_2 is continuous at $s = 0$ and satisfies $k_2(|s|) = O(|s|)$. This condition bounds the fluctuation of $\psi(e_1 + s)$ around $\psi(e_1)$ as s approaches zero.

$$\mathrm{E}(\psi(e_1 + s) - \psi(e_1))^2 \leq k_2(|s|),$$

Moreover, for any sequence of random variables $\{\eta_n\}$ such that η converges in probability to zero ($\eta_n = o_P(1)$), the expected square difference between $\psi(e_1 + \eta_n)$ and $\psi(e_1)$ tends to zero.

$$\mathrm{E}(\psi(e_1 + \eta_n) - \psi(e_1))^2 = o(1).$$

(iii) The expectation of the square of the derivative function ψ evaluated at e_1 denoted by $\xi_0 = \mathrm{E}\left(\psi^2\left(e_1\right)\right)$ is finite, which indicates that the second moment of $\psi(e_1)$ is bounded.
(iv) As $|s|$ and $|l|$ tends to zero, and for sufficiently large t, the expected value of the product $\left(\psi\left(e_1+s\right)\psi\left(e_t+l\right)\right)$ can be approximated as the sum of the expected products $(\psi(e_1)\psi(e_t)$ and $k_1^2 sl$ with a small error term $o(|sl|)$

$$\mathrm{E}\left(\psi\left(e_1+s\right)\psi\left(e_t+l\right)\right) = \mathrm{E}\left(\psi\left(e_1\right)\psi\left(e_t\right)\right) + k_1^2 sl + o(|sl|).$$

(v) Both sequences $\{e_t\}$ and $\{m(X_t)\}$ are strictly stationary. Additionally, for all $h \leq t$, $m(X_h)$ and e_t are mutually independent for all $h \leq t$.
A3. Let $\{m(X_t)\}$ represents a sequence of stationary α-mixing values. The fourth moment of $\|m(X_1)\|$ denoted as $\mathrm{E}\left(\|m(X_1)\|^{4+\gamma}\right)$ is finite, indicating limited variability in the distribution of $m(X_1)$. In addition, $\sum_{n=1}^{\infty} \alpha^{\gamma/(4+\gamma)}(n)$ converges, ensuring certain regularity properties of the sequence $\{m(X_t)\}$ in terms of its mixing behaviour. Here, $\|\cdot\|$ denotes the L_2-distance and $\gamma > 0$.

We introduce some notation for simplicity. Let

$$\Theta_n(\omega) = \sum_{t=1}^{n} \rho\left(Y_t - \sum_{j=1}^{d} m_j(X_{tj})^\top \omega_j\right), \Psi_n(\omega) = \sum_{t=1}^{n}\sum_{j=1}^{d} m_j(X_{tj})\psi\left(Y_t - \sum_{j=1}^{d} m_j(X_{tj})^\top \omega_j\right) \tag{2.14}$$

Theorem. Suppose that assumptions A1 through A3 hold true, then for a given constant $c > 0$ and sufficiently large n, we obtain the following

$$\sup_{|T_n^{1/2}(\omega-\omega_0)\|\leq c} \left|\Theta_n(\omega) - \Theta_n\left(\omega_0\right) + \left(\omega-\omega_0\right)^\top \Psi_n\left(\omega_0\right) - \frac{1}{2}\left(\omega-\omega_0\right)^\top R_n\left(\omega-\omega_0\right)\right| = o_P(1), \tag{2.15}$$

Here, T_n is a matrix obtained by scaling the sum of two terms by the sample size n. The first term is a combination of the autocovariance matrix $\boldsymbol{\Omega}_0 = \mathrm{E}\left(m(X)_0 m(X)_0^\top\right)$ multiplied by $\zeta_i = \mathrm{E}\left(\psi\left(e_1\right)\psi\left(e_{i+1}\right)\right)$, and the cross-covariance between $m(X_0)$ and its lagged versions $m(X_i)$ for $(i \geq 1)$ which is denoted as $2\sum_{i=1}^{\infty}\zeta_i \mathrm{E}\left(X_0 X_i^\top\right)$.
Furthermore, $\boldsymbol{R}_n$ denoted as $\boldsymbol{R}_n = nk_1\boldsymbol{\Omega}_0$.

To simplify the expression and notation, the proof concentrates solely on the scenario where $d = 1$, given that the basic ideas of our methodology remain applicable for $d \geq 2$.
Subsequently, we present two basic lemmas. The first lemma establishes two covariance inequalities applicable to α-mixing processes. The proofs of these inequalities lemma can be found in the literature on α-mixing processes, such as Lin

and Lu (1996). The second lemma is referenced as Lemma 1 in Bai et al. (1992), and its proof is provided therein.

Lemma 1. If $\mathrm{E}\left(|X|^p + |Y|^q\right) < \infty$ for some $p, q \geq 1$ and $1/p + 1/q < 1$, then

$$|\operatorname{cov}(X, Y)| \leq 8\alpha^{1/r} \left[\mathrm{E}\left(|X|^p\right)\right]^{1/p} \left[\mathrm{E}\left(|Y|^q\right)\right]^{1/q},$$

where $r = (1 - 1/p - 1/q)^{-1}$ and $\alpha = \sup_{A \in \sigma(X), B \in \sigma(Y)} |\Pr(AB) - \Pr(A)\Pr(B)|$. If $\Pr\left(|X| \leq C_1\right) = 1$ and $\Pr\left(|Y| \leq C_2\right) = 1$ for some $C_1 > 0, C_2 > 0$, then

$$|\operatorname{cov}(X, Y)| \leq 4\alpha C_1 C_2.$$

lemma 2. Under the assumptions A1-A2(i), we have

$$E\left[\rho\left(e_1 + s\right) - \rho\left(e_1\right)\right] = \frac{1}{2} s' A s + o\left(\|s\|^2\right) \quad \text{as} \quad s \to 0.$$

Proof of Theorem Define,

$$G_{n,t} = \rho\left(e_t - m(X_t)\left(\omega - \omega_0\right)\right) - \rho\left(e_t\right) - m(X_t)\left(\omega - \omega_0\right)\psi\left(e_t\right) \text{ where } e_t = Y_t - m(X_t)^\top \omega_0.$$

Given the definition of $G_{n,t}$, we rewrite $\Theta_n(\omega)$ and $\Theta_n(\omega_0)$ in (2.14) as follows:

$$\Theta_n(\omega) = \sum_{t=1}^{n} \rho\left(Y_t - m(X_t)^\top \omega\right) = \sum_{t=1}^{n} \rho\left(e_t - m(X_t)^\top (\omega - \omega_0)\right)$$

$$\Theta_n(\omega_0) = \sum_{t=1}^{n} \rho\left(Y_t - m(X_t)^\top \omega_0\right) = \sum_{t=1}^{n} \rho\left(e_t\right)$$

Note that,

$$\begin{aligned}
&\Theta_n(\omega) - \Theta_n\left(\omega_0\right) + \left(\omega - \omega_0\right)\Psi_n\left(\omega_0\right) - \mathrm{E}\left(\Theta_n(\omega) - \Theta_n\left(\omega_0\right) + \left(\omega - \omega_0\right)\Psi_n\left(\omega_0\right)\right) \\
&= \sum_{t=1}^{n} \left(G_{n,t} - \mathrm{E}\left(G_{n,t}\right)\right)
\end{aligned}$$

First, we want to demonstrate

$$\operatorname{var}\left(\sum_{t=1}^{n} G_{n,t}\right) = o(1) \tag{2.16}$$

We know that

$$\operatorname{var}\left(\sum_{t=1}^{n} G_{n,t}\right) = \sum_{t=1}^{n} \operatorname{var}\left(G_{n,t}\right) + \sum_{h \neq t} \operatorname{cov}\left(G_{n,h}, G_{n,t}\right). \tag{2.17}$$

We aim to show that both components in the right-hand side (RHS) of (2.17) tend to zero as n increases, which indicates that var $(\sum_{t=1}^{n} G_{n,t})$ tends to zero as well.
Note that,

$$G_{n,t} = \int_0^{m(X_t)(\omega-\omega_0)} (\psi(e_t + v) - \psi(e_t))\, dv \tag{2.18}$$

Given assumption A2(ii): for a constant C we have, for $\left|T_n^{1/2}(\omega - \omega_0)\right| \leq C$,

$$\sum_{t=1}^{n} E(G_{n,t}^2) = \sum_{t=1}^{n} E\left(\int_0^{m(X_t)(\omega-\omega_0)} (\psi(e_t + v) - \psi(e_t))\, dv\right)^2$$

To bound the expectation, we utilize the Cauchy-Schwarz Inequality, which states that for any two integrable functions f and g, the following inequality holds:

$$\left(\int_a^b f(x)g(x)dx\right)^2 \leq \int_a^b f(x)^2 dx \cdot \int_a^b g(x)^2 dx$$

Hence,

$$\begin{aligned}
\sum_{t=1}^{n} \mathrm{E}\left(G_{n,t}^2\right) &= \sum_{t=1}^{n} \mathrm{E}\left[\left(\int_0^{m(X_t)(\omega-\omega_0)} (\psi(e_t + v) - \psi(e_t))\, dv\right)^2\right] \\
&\leq \sum_{t=1}^{n} \mathrm{E}\left[|m(X_t)(\omega - \omega_0)| \int_0^{m(X_t)(\omega-\omega_0)} (\psi(e_t + v) - \psi(e_t))^2\, dv\right] \\
&\quad \text{We implement the Law of Iterated Expectations: } E[X] = E[E[X|Y]] \\
&= \sum_{t=1}^{n} \mathrm{E}\left[|m(X_t)(\omega - \omega_0)| \cdot \mathrm{E}\left(\int_0^{m(X_t)(\omega-\omega_0)} (\psi(e_t + v) - \psi(e_t))^2\, dv | m(X_t)\right)\right] \\
&\quad \text{After integrating out the randomness of } m(X_t), \text{ we no longer condition on} \\
&\quad m(X_t) \text{ in the outer expectation, which simplifies the expression to:} \\
&= \sum_{t=1}^{n} \mathrm{E}\left[|m(X_t)(\omega - \omega_0)| \cdot \int_0^{m(X_t)(\omega-\omega_0)} \mathrm{E}(\psi(e_t + v) - \psi(e_t))^2\, dv\right] \\
&\quad \text{Since both } |m(X_t)(\omega - \omega_0)| \text{ and the integral are bounded by constants, we} \\
&\quad \text{bound the entire expression by a constant } C \text{ times the expectation of} \\
&\quad |m(X_t)(\omega - \omega_0)| \text{ multiplied by the integral.} \\
&\leq \sum_{t=1}^{n} C\mathrm{E}\left(|m(X_t)(\omega - \omega_0)| \cdot \int_0^{m(X_t)(\omega-\omega_0)} |v| dv\right) \\
&\quad \text{Because the expectation of } |m(X_t)(\omega - \omega_0)| \text{ is of order } O(n^{-1/2}) \text{ given the} \\
&\quad \text{condition } |T_n^{\frac{1}{2}}(\omega - \omega_0)| \leq C, \text{ and we are multiplying it by a constant } C, \\
&\quad \text{where the product remains of order } O(n^{-1/2}). \text{ Hence, the entire expression is} \\
&\quad \text{of order } O(n^{-1/2}), \text{ which is equivalent to } o(1). \\
&\leq O\left(n^{-1/2}\right) = o(1).
\end{aligned}$$

Therefore, we have

$$\sum_{t=1}^{n} \operatorname{var}\left(G_{n,t}\right) \leq \sum_{t=1}^{n} \mathrm{E}\left(G_{n,t}^{2}\right) = o(1) \tag{2.19}$$

We split the summation $\sum_{h\neq t} \operatorname{cov}\left(G_{n,h}, G_{n,t}\right)$ into two parts as follows,

$$\sum_{h\neq t} \operatorname{cov}\left(G_{n,h}, G_{n,t}\right) = 2\sum_{t=2}^{q_n}(n-t+1)\operatorname{cov}\left(G_{n,1}, G_{n,t}\right) + 2\sum_{t=q_n+1}^{n}(n-t+1)\operatorname{cov}\left(G_{n,1}, G_{n,t}\right), \tag{2.20}$$

Here, q_n represents a sequence of values such that $q_n \to \infty$ as n increases, but at a slower rate than $\sqrt{n}$ such that $q_n = o(\sqrt{n})$. Since $q_n = o(\sqrt{n})$, as n increases, the value of q_n also increases. Consequently, the upper limit of the sum q_n becomes larger and larger, but the contribution of each term diminishes as n grows larger. Therefore, the overall sum $\sum_{t=2}^{q_n}(n-t+1)\operatorname{cov}\left(G_{n,1}, G_{n,t}\right)$ is of the order $O\left(\frac{1}{\sqrt{n}}\right)$, indicating that it is bounded by a quantity that diminishes to zero as $n \to \infty$. Thus,

$$\sum_{t=2}^{q_n}(n-t+1)\operatorname{cov}\left(G_{n,1}, G_{n,t}\right) = o(1). \tag{2.21}$$

In addition, assumption A2(iv) states that as $|v_1|, |v_2|$ approaches zero, for sufficient large n where $|t-h| > q_n$, we have

$$\mathrm{E}\left(\left(\psi\left(e_t+v_1\right)-\psi\left(e_t\right)\right)\left(\psi\left(e_h+v_2\right)-\psi\left(e_h\right)\right)\right) = k_1^2 v_1 v_2 + o\left(|v_1|\,|v_2|\right), \tag{2.22}$$

We start by computing $\mathrm{E}\left(G_{n,h}G_{n,t}\right)$,

$$\begin{aligned}
&\mathrm{E}\left(G_{n,h}G_{n,t}\right)\\
&= \mathrm{E}\left(\int_0^{m(X_t)(\omega-\omega_0)}\int_0^{m(X_h)(\omega-\omega_0)}\left(\left(\psi\left(e_t+v_1\right)-\psi\left(e_t\right)\right)\left(\psi\left(e_h+v_2\right)-\psi\left(e_h\right)\right)\right)dv_1dv_2\right)\\
&\quad \text{From (2.22) we obtain.}\\
&= k_1^2\mathrm{E}\left(\int_0^{m(X_t)(\omega-\omega_0)}\int_0^{m(X_h)(\omega-\omega_0)} v_1v_2dv_1dv_2\right) + o\left(\mathrm{E}\left(\int_0^{m(X_t)(\omega-\omega_0)}\int_0^{m(X_h)(\omega-\omega_0)} v_1v_2dv_1dv_2\right)\right).
\end{aligned}$$

Note that the product of the expectations of $G_{n,h}$ and $G_{n,t}$ is given by,

$$\mathrm{E}\left(G_{n,h}\right)\mathrm{E}\left(G_{n,t}\right) = k_1^2\mathrm{E}\left(\int_0^{m(X_t)(\omega-\omega_0)} v_1dv_1\right)\mathrm{E}\left(\int_0^{m(X_h)(\omega-\omega_0)} v_2dv_2\right).$$

Thus, by invoking Lemma 1 to $G_{n,h}$ and $G_{n,t}$ and simplify the computation, we obtain for $\left|T_n^{1/2}\left(\omega-\omega_0\right)\right| \leq C$,

$$\left|\operatorname{cov}\left(G_{n,h}, G_{n,t}\right)\right| \leq Cn^{-2}\alpha^{\gamma/(4+\gamma)}(|h-t|) + o\left(n^{-2}\right) \tag{2.23}$$

By utilizing the result of (2.23) and assumption A3 we have

$$\begin{aligned}\sum_{t=q_n+1}^{n}(n-t+1)\left|\operatorname{cov}\left(G_{n,1},G_{n,t}\right)\right| &= Cn^{-1}\\ \sum_{t=q_n+1}^{n}\alpha^{\gamma/(4+\gamma)}(t)+o\left(n\left(n-q_n\right)n^{-2}\right) &= o(1).\end{aligned} \tag{2.24}$$

Hence, Equation (2.16) is concluded from (2.19),(2.21), and (2.24). Therefore, Equation (2.16) implies that $G_{n,t}$ becomes more stable as n increases. Consequently,

$$\Theta_n(\omega)-\Theta_n\left(\omega_0\right)+\left(\omega-\omega_0\right)\Psi_n\left(\omega_0\right)-\mathrm{E}\left(\Theta_n(\omega)-\Theta_n\left(\omega_0\right)+\left(\omega-\omega_0\right)\Psi_n\left(\omega_0\right)\right)=o_P(1) \tag{2.25}$$

By implementing Lemma 2, we have

$$\mathrm{E}\left(\Theta_n(\beta)-\Theta_n\left(\beta_0\right)\right)=\frac{1}{2}R_n\left(\beta-\beta_0\right)^2 \text{ and } \mathrm{E}\left(\Psi_n\left(\beta_0\right)\right)=0. \tag{2.26}$$

Therefore,

$$\mathrm{E}\left(\Theta_n(\omega)-\Theta_n\left(\omega_0\right)+\left(\omega-\omega_0\right)\Psi_n\left(\omega_0\right)\right)=\frac{1}{2}R_n\left(\beta-\beta_0\right)^2 \tag{2.27}$$

Substituting (2.27) into (2.25) we have,

$$\Theta_n(\omega)-\Theta_n\left(\omega_0\right)-\left(\omega-\omega_0\right)\Psi_n\left(\omega_0\right)-\frac{1}{2}R_n\left(\beta-\beta_0\right)^2=o_P(1) \tag{2.28}$$

Since $\Theta_n(\omega)-\Theta_n\left(\omega_0\right)-\left(\omega-\omega_0\right)\Psi_n\left(\omega_0\right)$ is convex in ω and $R_n\left(\beta-\beta_0\right)^2$ is continuous and convex in ω, by Theorem 10.8 of (Rockafellar (1969), p.90), we prove(2.15) by (2.28).

Remark. If we assumed $\mathbf{\Omega}_0$ is positive definite, then based on equation (2.15) and employing same argument as presented in (Bai et al. (1992) ,Theorem 2.2), the consistency of the estimator will be derived easily.

$$\widehat{\omega}-\omega_0=o_p(1). \tag{2.29}$$

Generality of the Proof for Higher Dimensions

It is important to discuss whether the proof based on d=1 is without loss of generality. While the presented proof focuses on the case where d=1 for simplicity and clarity, the main ideas and techniques used in the proof can be extended to higher dimensions ($d>1$) with some modifications.

The proof for the case of d=1 holds significant value, even though it may not be directly applicable to higher dimensions. By presenting the proof for the one-dimensional case, we gain important insights into the fundamental ideas and

techniques that form the foundation for extending the results to the multi-dimensional setting. The proof for d=1 serves as a stepping stone, showcasing the key concepts and mathematical tools that are essential for establishing the theoretical properties of the RMAMAR estimator. It highlights the crucial steps, such as the decomposition of the objective function, the use of mixing conditions, and the application of limit theorems, which can be adapted and generalized to accommodate higher dimensions. By understanding the proof for d=1, researchers can grasp the underlying logic and structure that govern the behavior of the RMAMAR estimator. This understanding is invaluable when considering the extension to higher dimensions, as it provides a roadmap for navigating the complexities and challenges that arise in the multi-dimensional setting.

In the case of $d > 1$, the key steps of the proof remain similar. The main difference lies in the notation and the dimensionality of the vectors and matrices involved. For example, the weight vector ω and the marginal regression functions $m_j(x_j)$ would be replaced by their multi-dimensional counterparts, and the summations would be performed over all dimensions.

The assumptions and conditions used in the proof, such as the mixing conditions and moment bounds, would need to be extended to accommodate the higher-dimensional setting. However, the overall structure and logic of the proof would remain intact.

It is worth noting that the extension of the proof to higher dimensions introduce additional technical complexities and require more elaborate notation. Nevertheless, the fundamental ideas and techniques used in the proof, such as the decomposition of the objective function, the use of mixing conditions, and the application of limit theorems, can be adapted to handle the multi-dimensional case.

Future research could focus on rigorously establishing the theoretical properties of the RMAMAR estimator in the multi-dimensional setting, building upon the ideas and techniques presented in this proof.

In summary, while the proof based on d=1 is presented for simplicity and clarity, the main ideas and techniques used in the proof can be extended to higher dimensions with appropriate modifications. The presented proof provides a solid foundation for future research on the theoretical properties of the RMAMAR estimator in the general multi-dimensional setting.

2.4 Monte Carlo Simulation

In this Monte Carlo simulation study, We demonstrate the practical application of our RMAMAR model through simulated data, aiming to examine how lagged information influences the forecasting of time series data. For comparative purposes, our simulation models are similar to that in Li et al. (2015).

The selection of the smoothing bandwidth h_j, as indicated in Equation (2.11), is crucial for effective smoothing. In our simulations, we employ the Cross-Validation (CV) method using the h.select function provided in the R package sm. However, it's important to note that we are aware of its sensitivity to outliers when CV is applied. We leave this aspect for future investigation.

In Example 1, we consider the linear Autoregressive (AR) model in our simulation. We conduct a comparative study involving three different model predictions: the linear AR model of order 10, the nonlinear MAMAR of lag 10 proposed in Li et al. (2015) and the nonlinear RMAMAR of lag 10 proposed in this Chapter.
The estimation process for the linear AR model relies on the ARIMA with the maximum likelihood (ML) method, implemented by the "ML" function within the R package STATS. In nonlinear MAMAR, we implement a nonparametric model averaging approach, specifically using a nonrobust local constant method. This is achieved by fitting local constant regressions to the predictor variables using the sm.regression function from the sm package. Then, we estimate the weights that combine the fitted regressions by lm function in R, which utilizes least square estimation. The estimation procedure for our proposed approach, RMAMAR, is provided in this Chapter's section (2.2). Across all three simulation examples, we employ the robust Huber loss function introduced earlier in Table (2.2) to calculate the weights. The next Empirical Application section provides more details about the robust Huber loss function.
In Example 2 and 3, we introduce a nonlinear additive AR model in our simulation. We compare three prediction models: the nonlinear additive autoregressive model of order 10, the nonlinear MAMAR of lag 10 and RMAMAR of lag 10. The estimation technique in the nonlinear additive AR model is based on the Generalized Additive Models (GAM) with smoothing splines in the GAM R package. GAM, introduced first by Hastie and Tibshirani (1986), is a class of statistical models that extend the framework of a linear model to include nonlinear relationships between predictors and the response (dependent) variable. This extension is represented by modelling the nonlinear relationships through smooth functions alongside the linear relationship between the response (dependent) and predictor (covariate) variables. These smooth functions are represented through nonparametric regression methods like smoothing splines.
To gain a deep understanding and evaluate the forecasting performance of our proposed approach compared to other forecasting approaches, we generate different simulation scenarios that start with a simple structure of linear models to complex nonlinear structures with interactions. First, we introduce our general framework model, the foundation for generating the three simulation scenarios we consider in this study.

$$
\begin{aligned}
Y_t &= \sum_{j=1}^{10} \vartheta_{j^*}(Y_{t-j}) + \varepsilon_t \\
\vartheta_{j^*}(Y_{t-j}) &= a_j Y_{t-j} + \delta \frac{\exp(-jY_{t-j})}{1+\exp(-jY_{t-j})} + \gamma \cos(Y_{t-j}Y_{t-1})
\end{aligned}
\tag{2.30}
$$

Here, Y_t represents the response (dependent) variable at time t, and j is the lag variable. In assessing the robustness of our proposed method, we consider the error to follow two types of distribution: a t-distribution with 3 degrees of freedom $\varepsilon_t \sim (t_3)$ and a normal distribution $\varepsilon_t \sim i.i.d.N(0,\sigma^2)$. The $\varepsilon_t \sim (t_3)$ distribution exhibits heavier tails than the normal distribution, enabling us to evaluate method performance under conditions of increased variability or outliers. The values of σ^2 and the coefficients a_j in (2.30) for $j = 1, 2, \cdots, 10$ are presented in Table (2.3). These values are the estimated values of a linear Autoregressive model of order 10, AR(10), using the whole data set in our time series daily rainfall in North London in the Empirical Application Section (2.5). The parameters δ and γ are constants that take different values in our simulations, such as $\delta = 0, 0.1, 5$ and $\gamma = 0, 0.1, .5$. From (2.30), we observe that the parameter δ introduces nonlinear transformations of lagged values, and γ captures the interaction between Y_{t-j} and Y_{t-1}. By adjusting the values of δ and γ, we construct three model structures to simulate our time series data. Example 1 considers the case where δ and $\gamma = 0$. In this case, model (2.30) simplifies to a pure linear AR(10). In the second example, a simulation derived from a nonlinear additive autoregressive model of order 10, the model is constructed by setting $\delta \neq 0$ and $\gamma = 0$. To improve the complexity of the model structure, in example 3, we include the interactions between Y_{t-j} and Y_{t-1} by considering $\gamma \neq 0$.
To generate the stationary time series data of size n from (2.30), as a general framework of our three examples, we follow the process of iterating the model starting from initial values $Y_1 = \cdots = Y_{10}$ are set equal to zero. To ensure the stationarity of the data, we delete the first 100 observations among the $(100 + n)$ observations obtained through this iteration. We consider the performance under different sample sizes $n = 140, 200$ and 250. We split the generated time series data n into two subsets: an estimation sample and a prediction sample denoted as n_{est}, n_{pred}, respectively. The estimation sample is implemented for estimating the model parameters and determined by $n_{est} = n - n_{pred}$. To evaluate the model performance, we introduce the prediction sample to forecast the dependent variable values utilizing the estimation models. The size of the n_{pred} is fixed at 50, therefore the n_{est} will take the values $90, 150$ and 200 for $n = 140, 200$ and 250, respectively. Based on estimated model we examine one-step-ahead prediction of $Y_{n_{est}+t}$, say $\widehat{Y}_{n_{est}+t}$, for $t = 1, 2, \ldots, n_{pred}$. The prediction methods' performance are evaluated by employing the mean squared

prediction error (MSPE) on the prediction sample, which is defined by

$$\text{MSPE} = \frac{1}{n_{\text{pred}}} \sum_{t=1}^{n_{\text{pred}}} \left(Y_{n_{est}+i} - \widehat{Y}_{n_{\text{est}}+i} \right)^2. \tag{2.31}$$

Following the Li et al. (2015) simulation examples setting, the simulation process of all three examples is repeated 100 times. Next, we present the simulation results for each example by providing boxplots of the MSPE from 100 repetitions for different prediction methods under two types of error distribution.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	σ^2
0.2255	0.0174	0.0544	-0.0082	0.0264	0.0046	0.023	0.0347	0.0108	-0.0513	13.11

TABLE 2.3: The variance σ^2 and the coefficients a_d in the model (2.30).

2.4.1 Example 1. Linear Autoregressive Model

We generate the random samples of length $n = 140, 200$ and 250 from the following model

$$Y_t = \textstyle\sum_{j=1}^{10} a_j Y_{t-j} + \varepsilon_t \tag{2.32}$$

As mentioned earlier, we investigate the models' performance across different sample sizes and distributions. We compare linear AR(10), nonlinear nonrobust MAMAR proposed by Li et al. (2015) and nonlinear RMAMAR proposed in this Chapter. The comparison of prediction methods using mean squared prediction error (MSPE) evaluation provides valuable insights into their relative performance. We expect a superior linear AR(10) model performance in this example since the underlying data generation process follows a linear AR(10) model. Therefore, employing a linear AR(10) model for prediction aligns closely with the true data-generating process, resulting in superior predictive accuracy compared to other approaches as confirmed in Figures (2.2 and 2.3). In the case of $\varepsilon_t \sim (t_3)$, the linear AR(10) method consistently outperforms all prediction approaches, especially with smaller sample sizes as provided in Figure (2.2). However, our proposed nonlinear RMAMAR of lag 10's prediction accuracy is better than the nonrobust nonlinear MAMAR of lag 10 across all sample sizes. Specifically, with an increase in sample size, the performance of our RMAMAR approach significantly improves, approaching the performance of the linear AR(10) method. For instance, in a sample size of 200, the median MSPE of the AR(10) method is 2.87, slightly better than our RMAMAR approach at 2.98. In contrast, in the sample size of 250, the performance gap decreases further, indicating that our method's predictive accuracy improves with larger datasets.
However, when the underlying distribution is normal, Figure (2.3) illustrates that the AR(10) method consistently outperforms all approaches, and the behaviour becomes

more evident with small sample sizes. At the same time, the RMAMAR performs better compared to MAMAR. Overall, the results in case of $\varepsilon_t \sim (t_3)$ show the robustness of our proposed RMAMAR to extreme cases, such as heavy-tailed data distributions, where RMAMAR offers a competitive alternative approach, and the prediction accuracy can be significantly improved, especially with larger datasets.

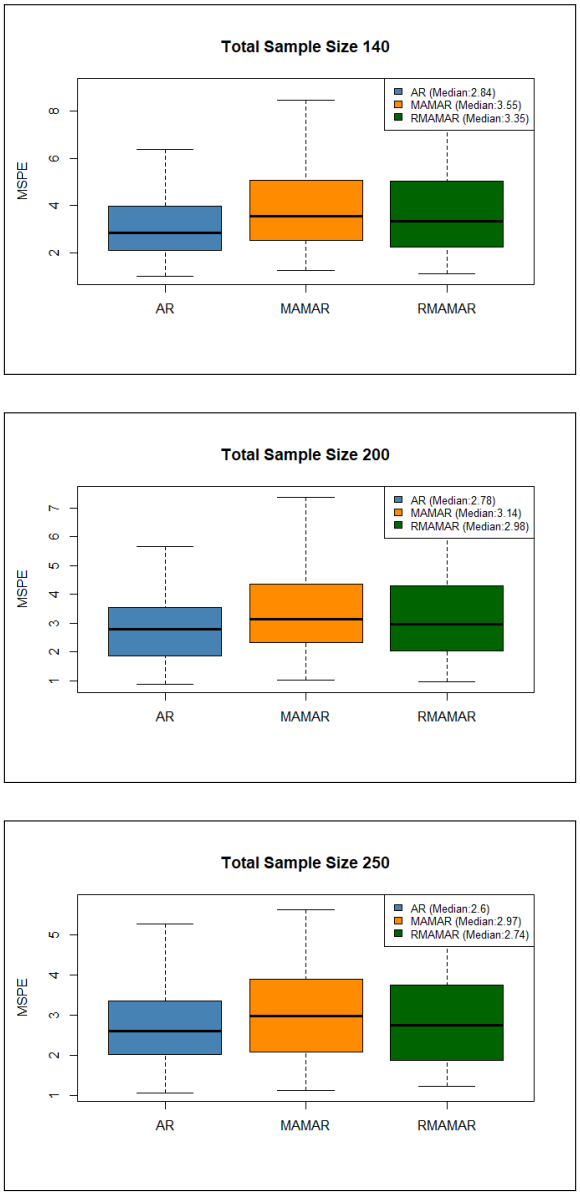


FIGURE 2.2: Ex1. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for linear Autoregressive (AR) models with lag 10, nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10, and nonlinear Robust MAMAR (RMAMAR) of lag 10. The errors ε_t are generated from a Student's t-distribution with 3 degrees of freedom $\varepsilon_t \sim (t_3)$. The sample sizes n are 140, 200, and 250 for each respective model. The median MSPE values for AR, MAMAR, and RMAMAR are provided in the legend.

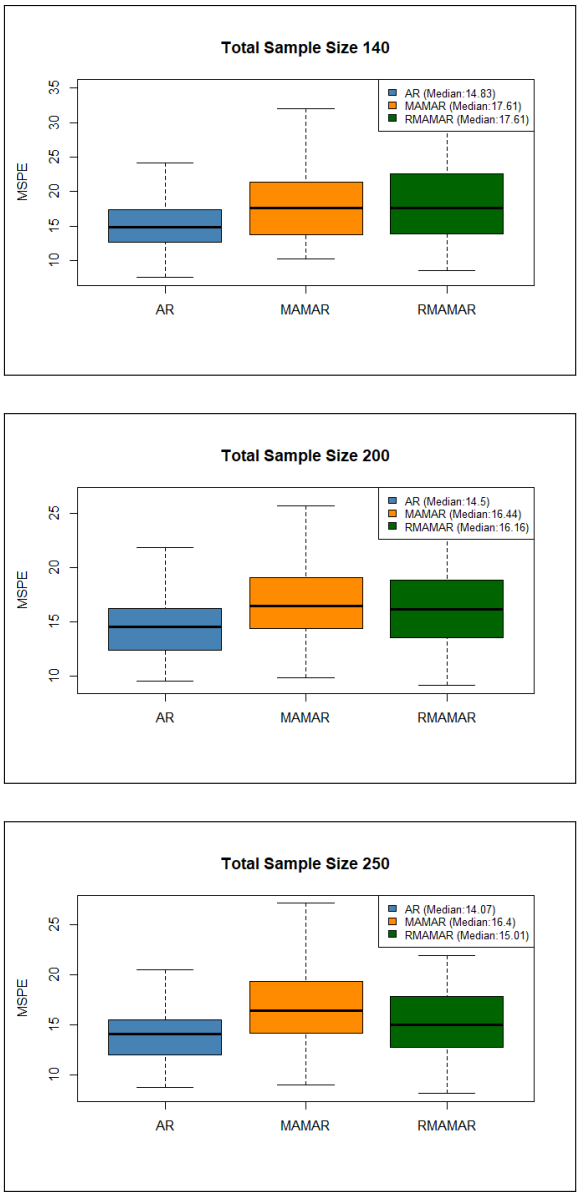


FIGURE 2.3: Ex1. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for linear Autoregressive (AR) models with lag 10, nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10 and nonlinear Robust MAMAR (RMAMAR) of lag 10. The errors ε_t are generated from Normal distribution $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. The sample sizes n are 140, 200, and 250 for each respective model. The median MSPE values for AR, MAMAR, and RMAMAR are provided in the legend.

2.4.2 Example 2. Nonlinear Additive Autoregressive Model

The underlying data generation process is characterized by a purely nonlinear additive AR model of order 10, as defined by the following:

$$Y_t = \sum_{j=1}^{10} a_j Y_{t-j} + \delta \frac{\exp\left(-jY_{t-j}\right)}{1+\exp\left(-jY_{t-j}\right)}) + \varepsilon_t \tag{2.33}$$

This study evaluates the prediction performance of three methods: nonlinear MAMAR of lag 10, nonlinear RMAMAR of lag 10, and nonlinear additive AR model of order 10, abbreviated as (GAM) for simplicity. Figure (2.4) displays MSEP evaluation under $\varepsilon_t \sim (t_3)$ across variations in sample size and the δ values: 0.1 and 0.5. We observe that from equation (2.33), by increasing the δ values, we effectively increase the nonlinearity in the model, which can help to assess the three approaches' ability to capture the higher degree of nonlinearity in the models. Across all sample sizes and nonlinearity levels, the RMAMAR approach consistently demonstrates the best performance, as indicated by its lower median MSEP values and smaller interquartile ranges (boxes) compared to MAMAR and GAM. This outperformance of RMAMAR is clear when the sample size is small, indicating the ability of our approach to capture nonlinear relationships and perform well even with limited data and greater nonlinearity degree represented by $\delta = 0.5$. Moreover, the performance of the GAM and MAMAR methods improves with larger sample sizes, although GAM is slightly better than MAMAR when the sample size increases. Moreover, Figure (2.5) shows the performance of the three prediction methods when the underlying distribution is normal. In this scenario, our RMAMAR maintains its advantages compared to the other approaches, highlighting its effectiveness in handling various data conditions across variations in sample size and the level of nonlinearity represented by δ values.

In summary, the consistent outperformance of our proposed method, the nonlinear RMAMAR of lag 10 across different distributional scenarios, underscores its effectiveness in capturing and predicting outcomes in real-world data settings, where data may exhibit nonlinearities and uncertainties.

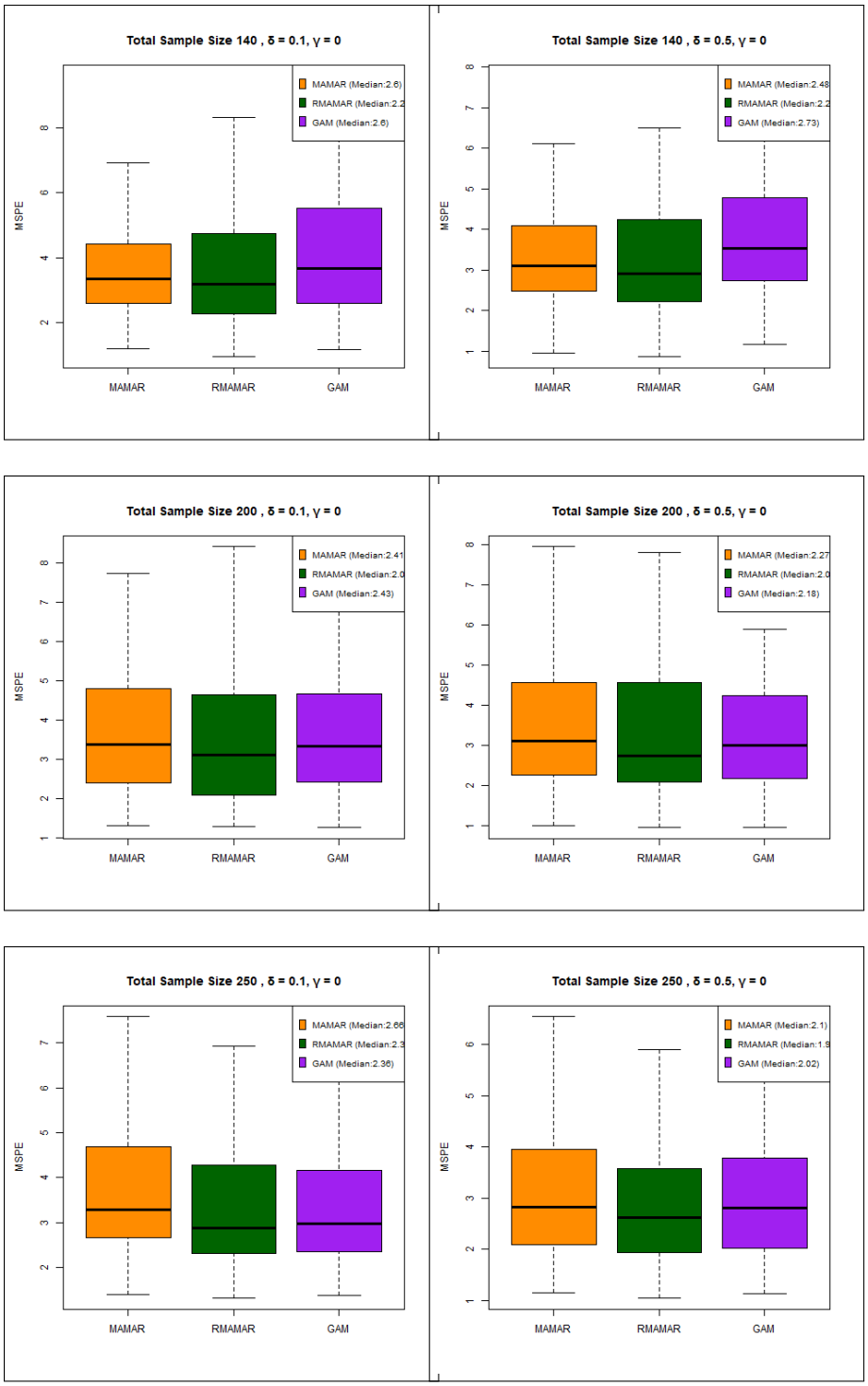


FIGURE 2.4: Ex2. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10, nonlinear Robust MAMAR (RMAMAR) of lag 10 and nonlinear additive AR model of order 10, abbreviated as (GAM). The errors are $\varepsilon_t \sim (t_3)$. The sample sizes n are 140, 200, and 250 with two different values of δ : 0.1 and 0.5. The median MSPE values for MAMAR, RMAMAR and GAM are provided in the legend.

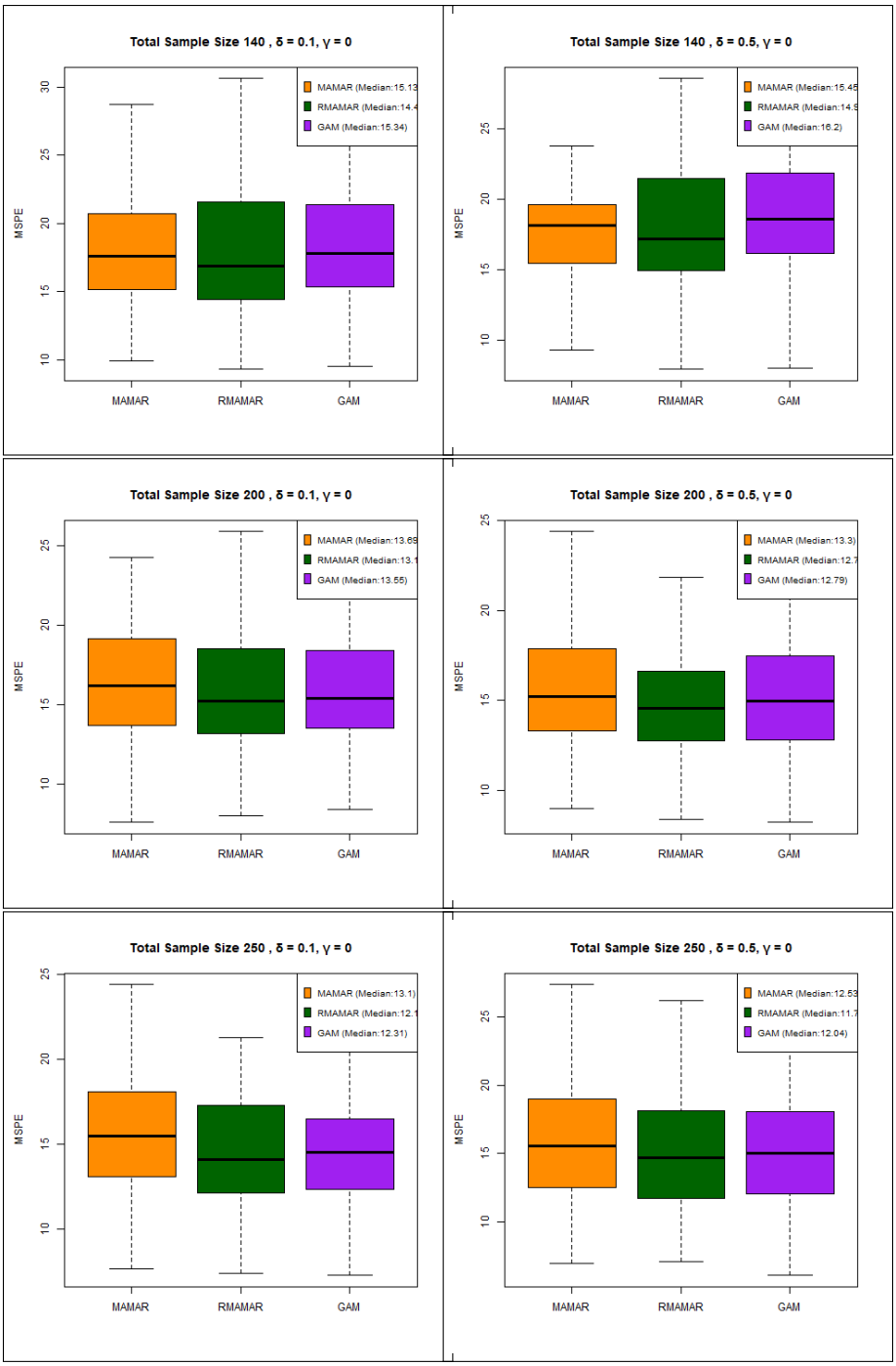


FIGURE 2.5: Ex2. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10, nonlinear Robust MAMAR (RMAMAR) of lag 10 and nonlinear additive AR model of order 10, abbreviated as (GAM). The errors ε_t are generated from Normal distribution $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. The sample sizes n are 140, 200, and 250 with two different values of δ : 0.1 and 0.5. The median MSPE values for MAMAR, RMAMAR and GAM are provided in the legend.

2.4.3 Example 3. Nonlinear Autoregressive Model with Interactions

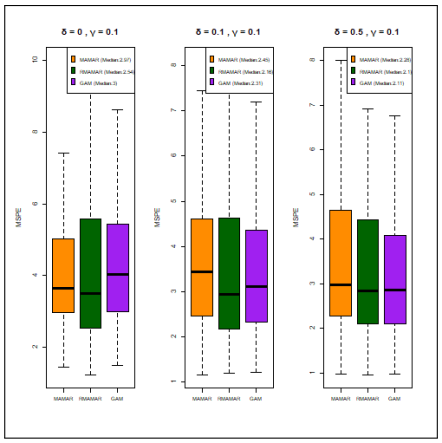
In this example, γ in Eq (2.30) takes non-zero values while δ could be either 0 or non-zero. The model is defined as:

$$Y_t = \sum_{j=1}^{10} a_j Y_{t-j} + \delta \frac{\exp(-jY_{t-j})}{1+\exp(-jY_{t-j})} + \gamma \cos(Y_{t-j}Y_{t-1}) + \varepsilon_t \tag{2.34}$$

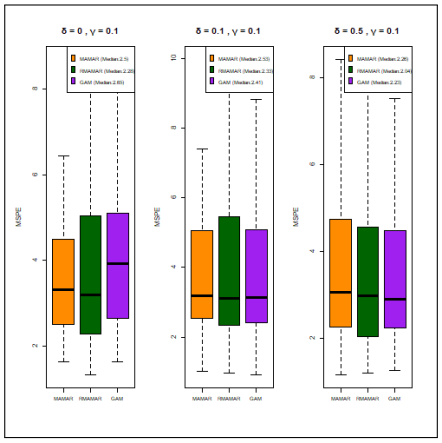
Under this scenario, for $\gamma \neq 0$, model (2.34) is a nonlinear additive AR(10) with interaction between Y_{t-j} and Y_{t-1}. In this simulation example, the predictive performance of three methods, nonlinear MAMAR, nonlinear RMAMAR and GAM, is evaluated under t-distribution and normal distribution scenarios. Figure (2.6) represents the scenario where γ is set to a low value of 0.1. In this case, the model in (2.34) closely resembles a purely to a purely additive AR model under t distribution since the impact of the cosine term on the overall model behaviour is relatively small. As presented, our nonlinear RMAMAR approach consistently outperforms the other methods across all sample sizes. It introduces greater nonlinearity into the model by increasing the δ values, which can capture more complex relationships in the data. Moreover, the MAMAR approach slightly improves over GAM when $\delta = 0$ in all the sample sizes. However, GAM becomes slightly better as the δ value increases while the behaviour of our outperforming approach is maintained across sample sizes. By setting $\gamma = 0.5$, the model (2.34) significantly deviates from the purely additive AR structure due to the stronger influence of the cosine term that introduces interactions between lags in the model. As illustrated in Figure (2.7), RMAMAR performs better than GAM when increasing the sample size and setting $\delta = 0$ but worse than the MAMAR method. All three approaches perform similarly by increasing the $\delta = 0.5$ and sample size. Figures (2.8) and (2.9) show the simulation under the normal distribution and considering two scenarios $\gamma = 0.1$ and $\gamma = 0.5$, respectively. In both scenarios, our RMAMAR approach consistently outperforms all other approaches across different sample sizes, and with increasing the δ values all over the sample sizes, the GAM approach shows improvement while MAMAR is in contrast, especially with small γ value where the model is close to being pure nonlinear additive AR model.

In summary, our simulation investigation shows that compared to other existing prediction methods we consider in the three analysis models, our proposed RMAMAR overall illustrates consistent performance compared to nonrobust MAMAR and GAM across different distributional scenarios. When the actual model is purely additive, as shown in Example 2, it outperforms the nonrobust version across a variation of sample sizes and δ values. However, when a purely additive modelling framework is violated, as presented in Example 3, RMAMAR provides better results with a low value of γ where the model structure closely approximates a purely

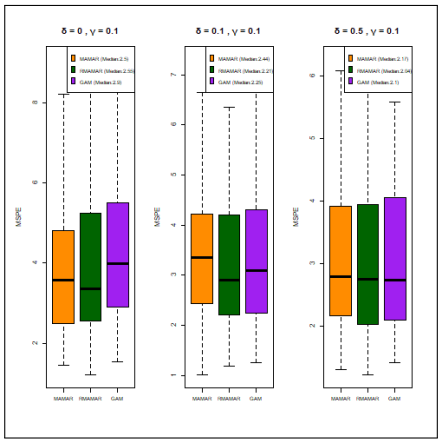
additive model, and the three approaches start to act similarly when the sample size increases, and δ value increases as well. This demonstrates that our proposed procedure can precisely lead to satisfactory prediction performances within purely nonlinear additive AR model structures, even with limited sample sizes.



(A) Sample size $n = 140$

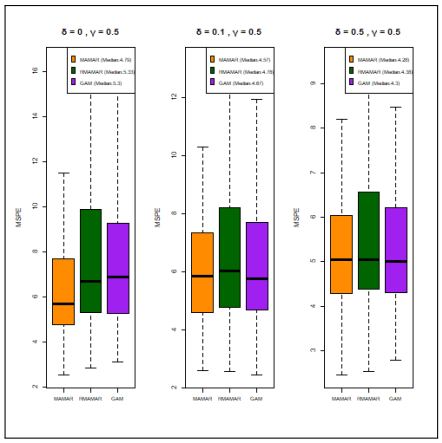


(B) Sample size $n = 200$

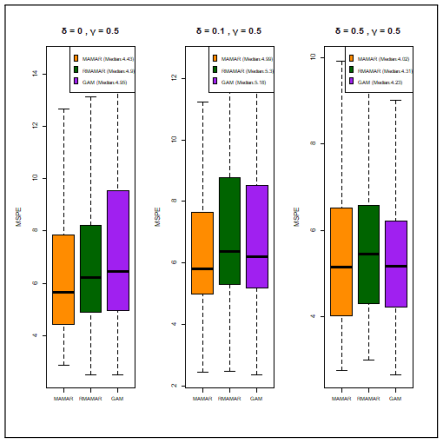


(C) Sample size $n = 250$

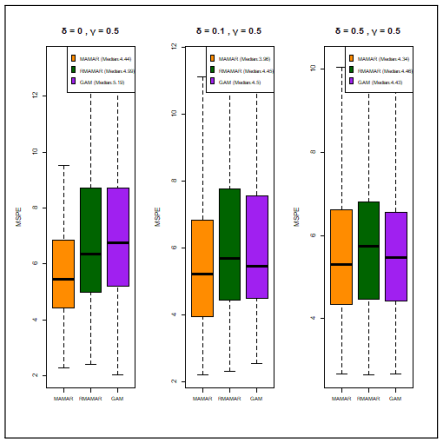
FIGURE 2.6: Ex3. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10, nonlinear Robust MAMAR (RMAMAR) of lag 10 and nonlinear additive AR model of order 10, abbreviated as (GAM). The errors $\varepsilon_t \sim (t_3)$. The sample sizes n are 140, 200, and 250 with $\gamma = 0.1$ and different values of δ. The median MSPE values for MAMAR, RMAMAR and GAM are provided in the legend.



(A) Sample size $n = 140$

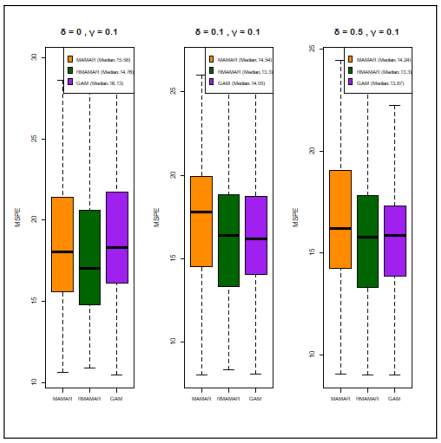


(B) Sample size $n = 200$

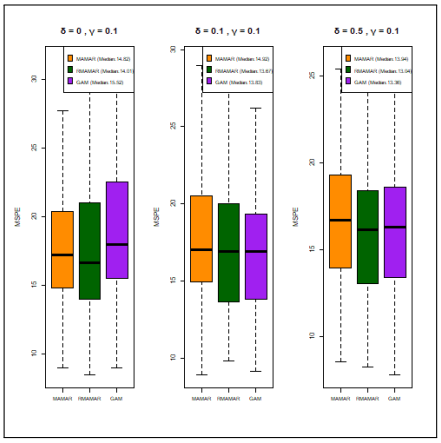


(C) Sample size $n = 250$

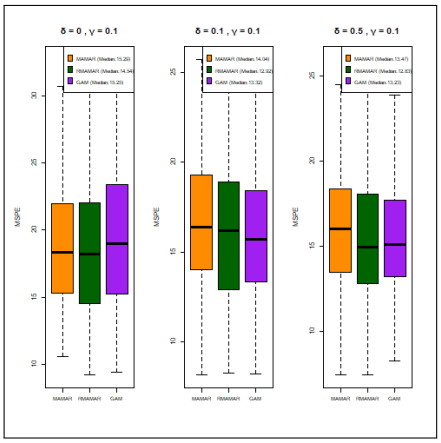
FIGURE 2.7: Ex3. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10, nonlinear Robust MAMAR (RMAMAR) of lag 10 and nonlinear additive AR model of order 10, abbreviated as (GAM). The errors $\varepsilon_t \sim (t_3)$. The sample sizes n are 140, 200, and 250 with $\gamma = 0.5$ and different values of δ. The median MSPE values for MAMAR, RMAMAR and GAM are provided in the legend.



(A) Sample size $n = 140$

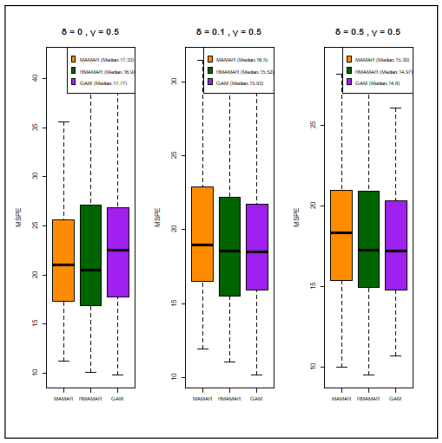


(B) Sample size $n = 200$

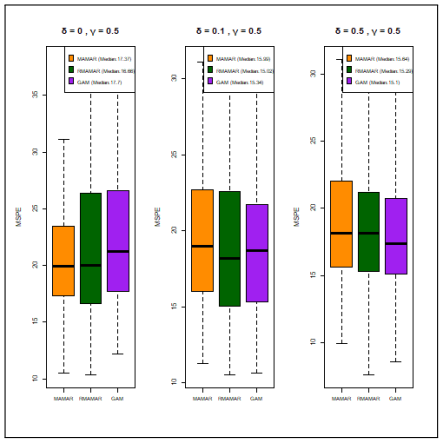


(C) Sample size $n = 250$

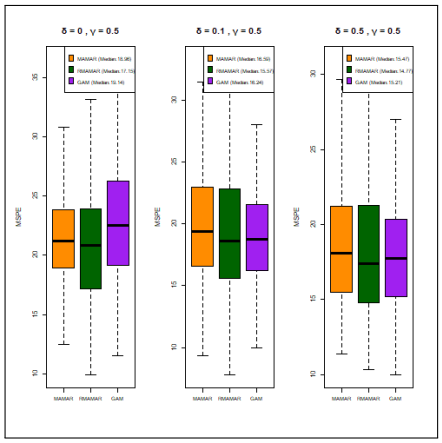
FIGURE 2.8: Ex3. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10, nonlinear Robust MAMAR (RMAMAR) of lag 10 and nonlinear additive AR model of order 10, abbreviated as (GAM). The errors ε_t are generated from Normal distribution $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. The sample sizes n are 140, 200, and 250 with $\gamma = 0.1$ and different values of δ. The median MSPE values for MAMAR, RMAMAR and GAM are provided in the legend.



(A) Sample size $n = 140$



(B) Sample size $n = 200$



(C) Sample size $n = 250$

FIGURE 2.9: Ex3. Boxplot illustrating 100 repetitions of the Mean Squared Prediction Error (MSPE) for nonlinear Model Averaging MArginal Regression (MAMAR) of lag 10, nonlinear Robust MAMAR (RMAMAR) of lag 10 and nonlinear additive AR model of order 10, abbreviated as (GAM). The errors ε_t are generated from Normal distribution $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. The sample sizes n are 140, 200, and 250 with $\gamma = 0.5$ and different values of δ. The median MSPE values for MAMAR, RMAMAR and GAM are provided in the legend.

2.5 Empirical application: Analysis of North London Daily Rainfall Data

Weather forecasting is vital for daily planning, with agriculture and various industries relying heavily on accurate predictions. Forecasting weather encompasses multiple aspects, including snowfall, air pressure, rainfall, and temperature variations, which significantly impact human activities. Among these, rainfall intensity is a critical parameter due to its profound influence on various natural processes. Rainfall affects vegetation distribution, crop planting and harvest, water resource management, and more.

Rainfall also plays a crucial role in managing reservoir water levels, but its unpredictability can lead to natural disasters such as floods. Studying changes in the intensity and frequency of heavy rainfall events is crucial for understanding climate change. However, forecasting rainfall is a formidable challenge due to its multidimensional and highly nonlinear nature, requiring thorough investigation and analysis for accurate predictions. Time series analysis has emerged as a significant tool in this endeavour, with various classical methods explored for forecasting rainfall intensity.

2.5.1 Data Analysis

This study intends to utilize various time series methodologies to investigate, compare, and evaluate the forecast accuracy of future areal rainfall in North London. The data involved in this study is Northern London rainfall amount, and it is daily data, which is available at London DATASTORE https://data.london.gov.uk/dataset/daily-areal-rainfall. Areal rainfall is calculated using data collected at one or more rain gauges within each unit. This is then averaged and weighted across each areal unit. The data covers 1-October 2007 to 30-September- 2010, giving 1096 observations. Time series data may exhibit seasonality, trends, and various patterns, such as exponential or linear patterns. In response to these scenarios and before modelling the rainfall forecast, preliminary statistical tests were applied to monitor these fluctuating features.

The daily rainfall data plot illustrated in Figure(2.10) appears stationary since the series' mean and variance do not seem to vary significantly over time. Furthermore, the autocorrelation function (ACF) and partial autocorrelation function (PACF), as illustrated in Figure (2.11) and (2.12), respectively, provide evidence supporting the stationarity of the daily rainfall data. The rapid decay and lack of significant autocorrelations and partial autocorrelations at higher lags suggest that the series does not have strong long-term dependence. The absence of significant spikes at regular

seasonal lags in both plots also indicates that the seasonal pattern in the data is not captured by the ACF and PACF. To formally assess the stationarity of the series, we conduct the Dickey-Fuller test, a widely used statistical test to evaluate the presence of unit roots, which are indicative of non-stationarity. The p-value of 0.01 indicates that the null hypothesis of a unit root (non-stationarity) can be rejected at the 5% significance level, providing further evidence that the time series is stationary.

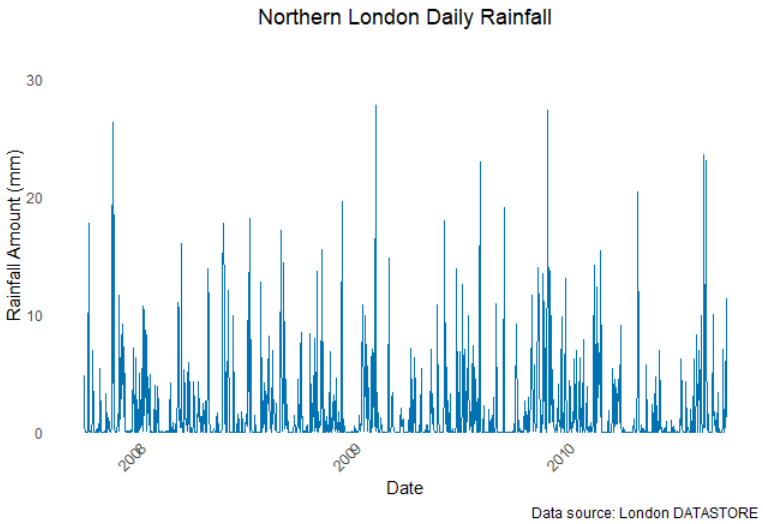


FIGURE 2.10: Daily rainfall data

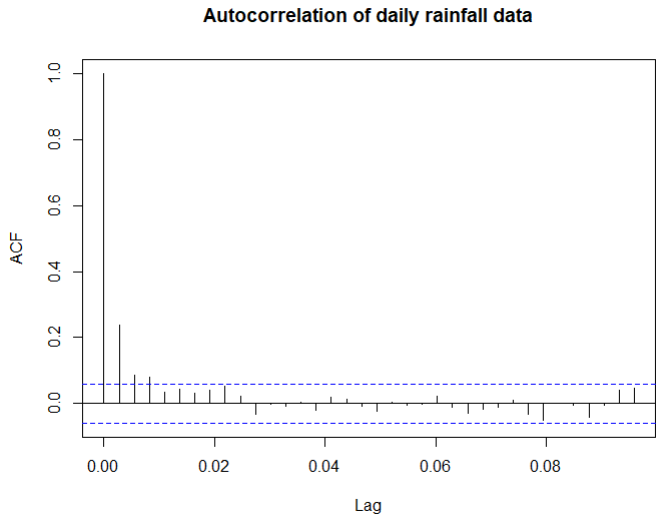


FIGURE 2.11: ACF plot for daily rainfall data

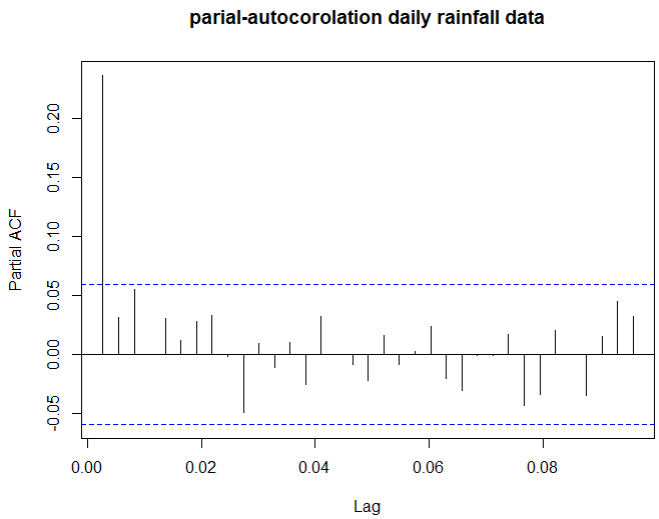


FIGURE 2.12: PACF plot for daily rainfall data

We divide the total sample size of $n = 1096$ into two subsets: one for model estimation and the other for prediction evaluation. Specifically, we set $n_{pred} = 30$, resulting in an estimation sample size of $n_{est} = n - n_{pred} = 1066$. In this study, our focus lies on the one-step-ahead prediction of the areal rainfall (mm) in North London using the information of lagged rainfall observations. We aim to determine whether utilizing lagged rainfall observations is useful in improving the explanation and prediction of the future. Our time series forecasting model will treat each lagged rainfall observation as an individual predictor.
To do that, we start by applying the linear Autoregressive (AR(p)) model to the estimation sub-sample, where we test p model values ranging from 1 to 30. The Akaike Information Criterion (AIC) is commonly utilized for model selection. As suggested by Li et al. (2015), we use AIC values to identify the best-fit model as an initial step. Table (2.4) presents the AIC values for p =1 to 10, the complete AIC values for p=1 to 30 are provided in Table (2.8) in the Chapter Appendix.
Our main objective in this chapter is to consider models with a large number of lag variables. Although the AR(3) model demonstrated the lowest AIC among the tested models, we select AR(10) to introduce additional complexities to our model and examine how effectively our RMAMAR approach handles the inherent complexities of such datasets. This choice aligns with our focus on addressing the curse of dimensionality that arises when the dimension exceeds 3 (Fan and Yao (2003)). It is worth noting that in such a high-dimensional case, selecting optimal lags is an interesting issue that can be considered in future work.

AR Order	AIC
p = 1	5795.5
p = 2	5796.4
p = 3	5795.2
p = 4	5797.2
p = 5	5798.3
p = 6	5800.2
p = 7	5801.0
p = 8	5801.8
p = 9	5803.8
p = 10	**5803.4**

TABLE 2.4: AIC Values for AR Models

The estimated coefficients of AR(10) are provided in Table (2.5), indicating the influence of lags up to 10 days on the daily areal rainfall data for North London. In this linear analysis, the p-values indicate that the coefficients of $ar_2 - ar_{10}$ at the 5% significance level are insignificant from zero.

Coefficient	Estimate	Standard Error	p-value
intercept	1.830	0.170	$<$ 2e-16 ***
ar_1	0.228	0.030	$<$ 2e-16 ***
ar_2	0.018	0.031	0.5583
ar_3	0.052	0.031	0.0919 .
ar_4	-0.005	0.031	0.8564
ar_5	0.025	0.031	0.4168
ar_6	0.001	0.031	0.9718
ar_7	0.027	0.031	0.3943
ar_8	0.036	0.031	0.2583
ar_9	0.012	0.032	0.6941
ar_{10}	-0.049	0.031	0.1195
AIC	5803.43		

TABLE 2.5: Estimated Coefficients , Standard Errors and p-values for AR(10) Model

Upon examining the kernel density estimate of the daily rainfall series in Figure (2.13), we observe that it does not follow a Gaussian (normal) distribution. The distribution is heavily skewed to the right, with a long right tail, indicating the presence of a significant number of high rainfall events. The peak of the distribution is close to zero, suggesting that most of the rainfall amounts are relatively small, while the occurrence of large rainfall amounts is less frequent but still notable. The asymmetry in the distribution suggests that the rainfall data is not well-described by a symmetric distribution like the Gaussian. The non-Gaussian nature of the rainfall distribution

suggest that linear models based on the assumption of normality may not fully capture the characteristics of the data. Hence, we apply the semiparametric RMAMAR approximation of lag 10 proposed in this chapter along with the semiparametric MAMAR of lag 10 introduced by Li et al. (2015) and the nonlinear additive AR model of order 10 (for short, we use GAM) in addition to the linear AR model of order 10. The estimation procedure of AR(10) and GAM is similar to the methods explained earlier in Section (2.4). The MAMAR approximation method is expressed as,

$$E(y_t|y_{t-1},\cdots,y_{t-10}) = w_0 + \sum_{j=1}^{10} w_j \cdot E(y_t|y_{t-j}). \tag{2.35}$$

Here, we estimate the marginal conditional expectation $E(y_t|y_{t-j})$ using a local constant estimator (Nadaraya–Watson). Then we estimate the weight coefficients $\omega_0,\cdots,w_{10}$ through least squares estimation as follows

$$\min_{w_0,\cdots,w_{10}} \sum_{t=11}^{n_{\text{est}}} \left[y_t - w_0 - \sum_{j=1}^{10} w_j \cdot \widehat{E}(y_t|y_{t-j}) \right]^2, \tag{2.36}$$

Here, n_{est} represents the number of observations in the estimation sample, the term $\widehat{E}(y_t|y_{t-j})$ denotes the local constant estimator of the conditional expectation of y_t given y_{t-j}.

In contrast, in the RMAMAR approximation, we first estimate the marginal conditional regression functions $m_{\rho j}(y_t|y_{t-j})$ by local linear estimator. Then we obtain the weight coefficients $\omega_0,\cdots,w_{10}$ by the following

$$\min_{w_0,\cdots,w_{10}} \sum_{t=11}^{n_{est}} \rho \left[y_t - w_0 - \sum_{j=1}^{10} w_j \widehat{m}_{\rho j}(y_t|y_{t-j}) \right]. \tag{2.37}$$

Here, ρ represents a robust check function. Among various robust loss functions available, two popular loss functions adopted widely in the literature are the Huber loss function and the Tukey loss function, introduced earlier in Table (2.2). We conduct the Huber loss function in this analysis for its desirable optimization properties. Specifically, the Huber loss function corresponds to a convex optimization problem, guaranteeing a unique solution. In contrast, the Tukey loss function may lead to multiple local minima and complicate the optimization process. Recall the definition of the Huber loss function:

$$\rho(\varepsilon_i) = \begin{cases} \frac{1}{2}\varepsilon_i^2 & \text{if } |\varepsilon_i| \le k \\ k\left(|\varepsilon_i| - \frac{1}{2}k\right) & \text{if } |\varepsilon_i| > k \end{cases} \tag{2.38}$$

As we observe from the definition of the Huber loss function in (2.38), it employs a quadratic piecewise function for small errors similar to the least squares loss, which is sensitive to small errors that provide accurate estimates of the data central tendency. On the other hand, for large errors exceeding a threshold determined by the parameter k, it adopts a linear behaviour similar to the absolute deviation loss function, which enhances robustness against outliers. This combination of features allows the Huber loss function to find a good balance between sensitivity to small errors and robustness to outliers, making it appropriate for diverse applications. Furthermore, $\widehat{m}_\rho(y_t|y_{t-j})$ is the local linear estimator of $m_\rho(y_t|y_{t-j})$. We use the R function "lprobust" from the package "nprobust" to perform local linear regression where confidence intervals produced by lprobust are robust and bias-corrected; see Calonico et al. (2018) for more details. We fixed the bandwidth parameter at h = 0.5 for the RMAMAR and MAMAR for simplicity and illustrative purposes. Additional bandwidth selection methods and their results are provided in the Chapter Appendix.

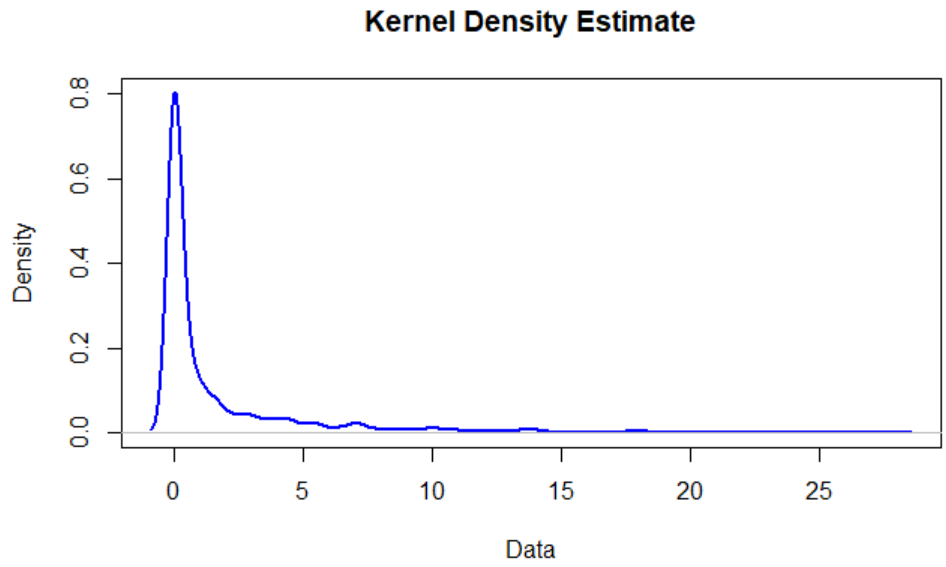


FIGURE 2.13: Kernel density

Table (2.6) illustrates the estimated coefficients of the lag effects in RMAMAR. Notably, we observe that almost all the coefficients are significant at the 5% level. In particular, the lag effects ranging from 2 to 10, which were considered insignificant in the linear AR model (Table (2.5)) demonstrate significance in RMAMAR framework (Table (2.6)). This shift suggests that these lag effects may exhibit nonlinear behaviour in our data, a phenomenon our approach can capture effectively. In addition, AIC is a widely used criterion for model selection; comparing the AIC values of the RMAMAR of lag 10 (2645.8) and the AR(10) model (5803.43) suggests that the RMAMAR of lag 10 is preferred over the AR(10) model. Lower AIC values imply a more optimal fit of the model to the data, and in this case, the RMAMAR model demonstrates substantially

lower AIC compared to the AR(10) model, indicating its superiority in capturing the data.

Moreover, since we are dealing with data that has heavy tails and the occurrence of abnormal observations, as presented in Figure (2.13), this can consequently significantly influence the mean squared error (MSE) therefore implementing the mean absolute deviation (MAD) can be a more robust measure in such cases. Table (2.7) shows MAE values for various forecasting approaches using both in-sample and out-of-sample data. In the estimation sample, our robust semiparametric approach RMAMAR method has MAE of 1.755, smaller than that of the linear AR model (2.257), the nonrobust semiparametric approach MAMAR (1.979) and the GAM (2.153). This suggests that our robust approach, RMAMAR, provides more accurate performance in the estimation sample. Furthermore, within the evaluation sample for one-step-ahead prediction with a size of $n_{pred} = 30$, our RMAMAR demonstrates the lowest MAE of 1.632, followed by the GAM with MAE of 2.017, the nonrobust approach MAMAR with MAE of 2.19 and the linear AR model with MAE of 2.323. These findings underscore the superior predictive performance of our robust semiparametric approach compared to linear AR, nonlinear additive AR and nonrobust semiparametric methods.
Overall, this analysis demonstrates the effectiveness of our robust semiparametric approach in uncovering nonlinear lag effects compared to the linear AR. Further, RMAMAR demonstrates the lowest MAE using both in-sample and out-of-sample data with heavy tails, underscoring our approach's superior robustness in handling challenging data and providing accurate predictions. In this study, we set the bandwidth parameter h = 0.5 for both RMAMAR and MAMAR models for simplicity and illustrative purposes. Further investigation of bandwidth selection methods and their results are provided in the Chapter Appendix. The results of additional bandwidth selection methods consistently demonstrate the superiority of our RMAMAR approach over MAMAR. It is important to highlight that the performance of kernel-based models, such as local linear regression and local constant regression, can be impacted by the choice of bandwidth, and it is essential to note that determining the optimal bandwidth selection remains an area for future investigation.

Coefficient	Estimated Value	Standard Error	p-value
ω_0	-3.1732	0.2992	<2e-16 ***
ω_1	0.3592	0.0316	<2e-16 ***
ω_2	0.2185	0.0618	0.0004 ***
ω_3	0.2119	0.0645	0.0010 ***
ω_4	0.2397	0.0940	0.0108 *
ω_5	0.1090	0.0650	0.0936 .
ω_6	0.2876	0.0771	0.0002 ***
ω_7	0.2068	0.0709	0.0035 **
ω_8	0.1387	0.0761	0.0683 .
ω_9	0.1705	0.0814	0.0362 *
ω_{10}	0.2926	0.0742	0.0001 ***
AIC	2645.814		

TABLE 2.6: Estimated coefficients and their standard errors in RMAMAR with lags from 1 to 10 for the estimation sub-sample of the series

MAE	Model			
	AR	MAMAR	GAM	RMAMAR
In Sample	2.257	1.979	2.153	1.755
Out of Sample	2.323	2.199	2.017	1.632

TABLE 2.7: Comparison of Mean Absolute Error (MAE) for the AR(10), MAMAR, GAM and RMAMAR models, considering both in-sample estimation and out-of-sample performance.

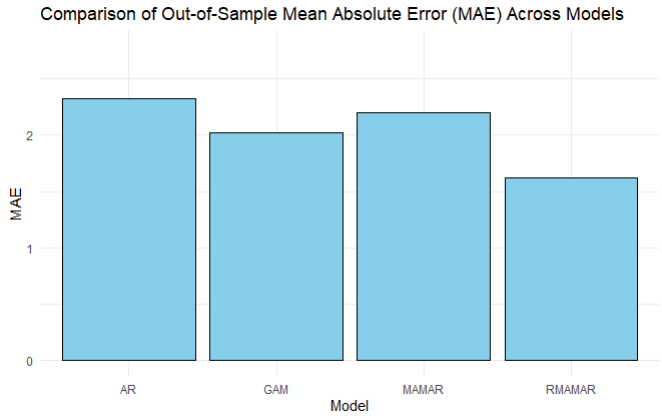


FIGURE 2.14: Mean Absolute Error (MAE) out-of-sample performance for the linear AR(10), nonlinear MAMAR of lag 10, nonlinear additive AR of order 10 (GAM) and nonlinear Robust MAMAR of lag 10.

2.6 Conclusion

In this chapter, we proposed the Robust Model Averaging MArginal Regression (RMAMAR) method as an extension of the Model Averaging MArginal Regression (MAMAR) methodology introduced by Li et al. (2015). RMAMAR integrates robust estimation techniques with nonparametric regression to provide a more robust approach for modeling conditional relationships in high-dimensional time series data. We investigated the behavior of our estimator by demonstrating uniform convergence, leading to the consistency of the estimator.
To evaluate the performance of RMAMAR approach, we conducted Monte Carlo simulations and real data analysis. The simulation study investigated our approach under three scenarios and various distributional assumptions, including t-distributions and normal distributions. We also implemented different parameter values in our simulated models to assess the sensitivity of our approach. The results showed that RMAMAR could lead to satisfactory prediction performances within purely nonlinear additive AR model structures, even with limited sample sizes, compared to alternative methods such as MAMAR and nonlinear additive models across different distributional scenarios. In the real data analysis, we applied RMAMAR to rainfall data characterized by heavy tails. Our analysis demonstrated RMAMAR's ability to uncover nonlinear lag effects compared to the linear Autoregressive (AR) model. Furthermore, RMAMAR achieved the lowest Mean Absolute Error (MAE) both in-sample and out-of-sample, indicating improved predictive ability compared to the other models considered in the analysis.
This chapter's key contributions are as follows: First, we extended the MAMAR approach by incorporating robust estimation techniques, enhancing its robustness to outliers and heavy-tailed distributions. Second, we established the theoretical properties of our estimator, including uniform convergence and consistency. Third, through extensive simulations and real data analysis, we demonstrated the superior performance of RMAMAR in capturing nonlinear relationships and providing accurate predictions, even in the presence of outliers and heavy-tailed distributions. In summary, our study highlights the advantages of the proposed RMAMAR method in understanding and modelling time series data containing outliers and nonlinear dependencies. By combining robust estimation techniques in model averaging with nonparametric regression, RMAMAR offers a flexible and robust approach for analyzing complex time series datasets in various applications.

Appendix

AR Order	AIC Value
1	5795.51
2	5796.41
3	5795.29
4	5797.28
5	5798.40
6	5800.29
7	5801.09
8	5801.85
9	5803.85
10	5803.43
11	5805.31
12	5806.74
13	5808.44
14	5809.80
15	5811.45
16	5813.34
17	5815.23
18	5816.85
19	5818.31
20	5820.07
21	5821.92
22	5823.08
23	5823.99
24	5825.44
25	5827.43
26	5829.40
27	5831.11
28	5830.73
29	5830.97
30	5831.98

TABLE 2.8: AIC Values for AR Models

MAE	Model	
	MAMAR	RMAMAR
In Sample	2.069	1.767
Out of Sample	2.008	1.621

TABLE 2.9: Comparison of Mean Absolute Error (MAE) for MAMAR and RMAMAR models, evaluating both in-sample estimation and out-of-sample performance. The optimal bandwidth, determined through simple cross-validation, is h=0.9314397 using the h.select function within R package sm.

MAE	Model	
	MAMAR	RMAMAR
In Sample	1.923	1.751
Out of Sample	2.203	1.644

TABLE 2.10: Comparison of Mean Absolute Error (MAE) for MAMAR and RMAMAR models, evaluating both in-sample estimation and out-of-sample performance. The optimal bandwidth, determined through the rule of thumb, is h=0.331656 using the bw.nrd function from the stats R package.

MAE	Model	
	MAMAR	RMAMAR
In Sample	2.1624	1.7951
Out of Sample	1.848	1.519

TABLE 2.11: Comparison of Mean Absolute Error (MAE) for the MAMAR and RMAMAR models, evaluating both in-sample estimation and out-of-sample performance. The optimal bandwidth, determined through biased cross-validation, is h= 1.0516 using bw.bcv function from stats package in R.

MAE	Model	
	MAMAR	RMAMAR
In Sample	2.1624	1.7951
Out of Sample	1.848	1.519

TABLE 2.12: Comparison of Mean Absolute Error (MAE) for the MAMAR and RMAMAR models, evaluating both in-sample estimation and out-of-sample performance. The optimal bandwidth, determined through unbiased cross-validation, is h= 0.1099 using bw.ucv function from stats package in R.

Chapter 3

Large Dimensional Optimal Portfolios Under Quantile Preferences

3.1 Introduction

A portfolio refers to a collection of securities or other assets. Constructing an optimal portfolio is one of the most significant objectives in financial research and represents a critical application of decision theory in uncertain environments. Throughout history, traditional models, grounded in the Expected Utility (EU) framework, have long served as a roadmap for investors' investment decisions. The EU hypothesis is a popular paradigm in economics; it was formalised in von Neumann and Morgenstern (1947) and has been proven to hold under general conditions of uncertainty (Savage (1972)). However, criticisms of the EU framework have been made over the past three decades, supported by experimental studies see for instance Simon (1979), Payne et al. (1992), Baltussen and Post (2011), Kahneman and Tversky (1981), Rabin (2000) and Tversky and Kahneman (1992). In the complex world of financial markets, investors often exhibit preferences beyond EU considerations.
This Chapter departs from the EU framework and sheds light on investors' quantile preferences, by focusing on specific quantiles of the returns distribution rather than just the mean. When faced with uncertain alternatives, an agent under the theory of expected utility selects the choice that maximises the expected value of his utility function. In the case of quantile utility, the agent maximises the utility function at a selected τ quantile instead of its mean. For example, in the case of the 0.25 quantile, the agent with quantile preferences focuses on maximising that quantile for a given target τ probability. Quantile preferences (QP) have several attractive features. The quantile, $\tau \in (0, 1)$, is a single-dimensional parameter that captures risk attitude. The

decision maker, a quantile-maximiser, is more risk averse when the τ value is low Castro et al. (2022). Therefore, the risk attitude measure is simple and intuitive. QP and properties of a quantile model for an individual's behaviour have been attracting increasing attention recently. Manski (1988) was the first to investigate QP properties, which Chambers (2009), Rostek (2010) and de Castro and Galvao (2022) later axiomatized. Chambers (2009) axiomatised the QP within a risk framework, demonstrating that these preferences satisfy the criteria of monotonicity, ordinal covariance, and continuity. On the other hand, Rostek (2010) axiomatised quantile maximisation within the subjective uncertainty framework proposed by Savage (1954). de Castro and Galvao (2022) provided an alternate axiomatisation for the static scenario incorporating an uncertain context and a finite state space. Recently, there has been a growing interest in studying dynamic quantile preferences. The study conducted by Giovannetti (2013) investigates a two-period standard economy featuring one risky and one risk-free asset, focusing on an asset pricing model under QP maximisation. In de Castro and Galvao (2019), a dynamic model of rational behaviour under uncertainty is proposed, wherein the agent's preferences are determined by a quantile of the returns distribution instead of the expected utility. Recently, in Castro et al. (2022), the portfolio selection problem is investigated within a model where individuals demonstrate QP. de Castro et al. (2022) utilise an experimental study to evaluate the importance of QP and the investors' choices that are consistent with the occurrence of QP.

In practical scenarios, the selection of a dynamic portfolio frequently relies on various conditioning or forecasting variables that reflect changes in the investment opportunity across time. In empirical studies, variables such as interest rates, inflation rates, dividend yield, default premium, lagged excess return, and others are commonly considered. For instance, in empirical studies of portfolio choice in single-period contexts, Avramov (2002) and Kandel and Stambaugh (1996) implemented the dividend yield as conditioning variables, Brandt (1999) examined various variables, including the dividend yield, default premium, term premium and lagged return in his portfolio choice problems while Chen et al. (2016) focused on lagged return. In multi-period settings, Campbell and Viceira (2001) used the Treasury bill yield as conditioning variables, while several studies such as Balduzzi and Lynch (1999), Campbell and Viceira (1999) and Barberis (2000) considered the dividend yield as conditioning variables.

Generally, the portfolio decision's dependency on the conditioning variables can be described in two ways. The first approach relies on a parametric statistical model in which returns from risky assets are correlated with the conditioning variables. This approach is commonly illustrated by linear regression-based models, which condition the expected return of assets on various factors using linear relationships.[1] The second strategy for characterising the dependency of the portfolio decision on a set of

[1]The basic Capital Asset Pricing Model (CAPM) and Fama-French models stand out as foundational frameworks (see Treynor(1961, 1962), Sharpe (1964), Mossin (1966)); Fama and French (1992, 1993) extend

conditioning variables is based on nonparametric techniques. Brandt (1999) introduces a nonparametric approach within the Expected Utility (EU) framework, in addition to analyze the asymptotic properties of the estimated optimal portfolio choice. The decision-making process for portfolio selection is conditioned on factors such as dividend yield, default premium, term premium, and lagged excess return. However, both techniques perform poorly in cases where the dimensionality of the conditioning variables is high due to the phenomenon known as the curse of dimensionality; see Fan and Yao (2003) for a discussion of the curse of dimensionality problem. This challenge suggests that when there are several conditioning variables, it may not be desirable to directly utilise the parametric technique or Brandt (1999) 's nonparametric approach. In order to avoid the challenges posed by the curse of dimensionality, Chen et al. (2016) propose a novel data-driven approach within the EU framework to estimate the optimal portfolio choice even when dealing with multiple conditioning variables of high dimensionality. In Chen et al. (2016) empirical applications, decisions to construct an optimal portfolio are made conditionally based on the lagged excess return. Chen et al. (2016) 's work is partly inspired by the Model Averaging MArginal Regression (MAMAR) approach proposed by Li et al. (2015), which has demonstrated a good performance in estimating the conditional mean regression function and in out-of-sample prediction. Model averaging in the presence of large dimensional predictors is a relatively new technique that has been mainly used to estimate and forecast the conditional mean (Ando and Li (2014); Li et al. (2015); Cheng and Hansen (2015); Chen et al. (2016); Chen et al. (2018)). Tu et al. (2021) explored the extension of high-dimensional forecasting into the conditional quantile context by developing a semiparametric model averaging prediction under i.i.d data. Conversely, De Gooijer and Zerom (2019) extended MAMAR approach by proposing a penalized high-dimensional quantile averaging in time series data.

To the best of our knowledge, no prior research has tackled the challenge of overcoming the curse of dimensionality in portfolio selection under the framework of QP when dealing with a large number of conditioning variables. Thus, our contributions are threefold; first, we construct dynamic optimal portfolios using quantile preferences with multiple conditioning variables. This is accomplished in two stages. In the first stage, we derive the optimal portfolio weights based on single conditioning variables. This involves, for each j-th conditioning variable $X_{j,t-1} = x_j$ and a given $\tau \in (0,1)$, we obtain the optimal portfolio weights using quantile regressions. In the second stage, the optimal portfolios constructed from the individual conditioning variables are combined by a model-averaging approach to obtain an optimal portfolio based on multiple conditioning variables. This approach is similar in spirit to the MAMAR method proposed in Chen et al. (2016) but adapted to

the basic CAPM by introducing additional factors such as size and book-to-market, as explanatory variables for describing the cross-section of stock returns. More recently, Fama and French (2015) include two additional risk factors in their three-factor model exploiting cross-sectional differences in profitability and investment across assets.

a quantile setting. The second contribution demonstrates that the proposed quantile method based on conditioning variables effectively aligns with investor's downside preferences during significant crises, such as the 2007 global financial crisis and the COVID-19 pandemic. This is achieved by employing an empirical application to a portfolio of six major financial indices. The optimal portfolios obtained from the conditional quantile regression approach outperform the unconditional counterpart portfolio strategies under different metrics in out-of-sample settings. The third contribution is to show the ability of this investment strategy to construct optimal portfolios under quantile preferences in large dimensions.

The Chapter is organized as follows: Section 2 presents the methodology for estimating dynamic portfolio choices within the framework of quantile preferences. This estimation can be accomplished in two steps. In the first step, we select the marginal optimal portfolio weights $w_j(x_j)$ under QP by maximising the conditional quantile portfolio problem. We combine the marginal optimal portfolio weights in the second step through the model averaging approach. A description of the out-of-sample performance methodology, along with a review of the approaches considered in the evaluation, is provided in Section 3. Section 4 implements our approach to construct optimal portfolios under QP with multiple covariates in two empirical exercises. The first exercise focuses on a small portfolio allocation problem that comprises six index assets and six conditioning variables. In contrast, the second portfolio allocation problem considers all the assets traded in the FTSE100 over the last 16 years. The Chapter is concluded in Section 5.

3.2 Methodology for Estimating the Dynamic Quantile Portfolio Choice

Within the domain of decision theory under uncertainty, it is customary to assume that individuals are endowed with a utility function denoted as u(·), which serves to characterize their preferences. Several candidate functions have dominated the literature on portfolio theory. Thus, the quadratic utility function defined as $u(W) = W - (\gamma/2)W^2$ and the Constant Absolute Risk Aversion utility (CARA) function, defined as $u(W) = 1 - exp(-\gamma W)$, closely correspond to the classical mean-variance framework developed by Markowitz (1952), with W denoting investor's wealth. More recently, the literature has focused on the Constant Relative Risk Aversion utility function (CRRA), given by the family of power utility functions

$$u(W) = \begin{cases} \frac{W^{1-\gamma}}{1-\gamma} & , \gamma \neq 1 \\ \ln(W) & , \gamma = 1, \end{cases} \tag{3.1}$$

where γ is risk aversion parameter.

The following subsection introduces the optimal portfolio choice problem in the expected utility framework for a large set of conditional variables. To do this, we review the nonparametric setting introduced by Brandt (1999) and Chen et al. (2016).

3.2.1 Expected Utility Framework

Suppose that there are n assets, whose return vector at time t is denoted by $\mathbf{R}_t = (R_{1t}, .., R_{nt})^T$. In the expected utility framework, we aim to solve the following general maximisation problem given by:

$$\begin{aligned} &\mathrm{E}[u(w^T \mathbf{R}_t) | X_{1,t-1}, ..., X_{p,t-1}], \\ &\text{s.t. } \mathbf{1}_n^\top w = \sum_{i=1}^{n} w_i = 1, \end{aligned} \tag{3.2}$$

Here, $\mathbf{X}_t = (X_{1t}, ..., X_{pt})^T$ represents a vector of p conditioning or forecasting variables X_{jt}. Under the expected utility framework, to solve the following general maximisation problem:

$$\begin{aligned} &\mathrm{E}[u(w^T \mathbf{R}_t) | X_{1,t-1}, ..., X_{p,t-1}], \\ &\text{s.t. } \mathbf{1}_n^\top w = \sum_{i=1}^{n} w_i = 1, \end{aligned} \tag{3.3}$$

Brandt (1999) develops a nonparametric approach for estimating the optimal portfolio weights without explicitly modelling returns or portfolio weights. Following Brandt (1999) approach, we differentiate $u(.)$ in (3.3) with respect to w_i to the first order, then we determine the optimal portfolio by solving the following for $w_1, ..., w_{n-1}$:

$$\mathrm{E}\left[(R_{it} - R_{nt})\, \dot{u}\left(w^\top \mathbf{R}_t \right) \mid X_{1,t-1}, \ldots, X_{p,t-1} \right] = 0 \quad \text{a.s. }, i = 1, \ldots, n-1, \tag{3.4}$$

Here, $\dot{u}(.)$ represents the derivative of the utility function $u(.)$. By utilizing the constraint $\sum_{i=1}^{n} w_i = 1$, we can compute the last element w_n in w. To derive the solution of (3.4), Brandt (1999) suggests employing a kernel-based smoothing technique. However, when p (the dimension of the conditioning variables is large), the proposed conditional approach often performs poorly due to the "curse of dimensionality". To address this challenge, Chen et al. (2016) introduces a technique to estimate dynamic portfolio choices with large number of conditioning variables through a model averaging approach. In Chen et al. (2016) approach, the marginal objective function is defined for each univariate conditioning variable X_j as,

$$\begin{aligned} &\mathrm{E}[u(w^T\mathbf{R}_t)|X_{j,t-1}=x_j], \\ &\text{s.t. } \mathbf{1}_n^\top w=\sum_{i=1}^{n} w_i=1, \end{aligned} \tag{3.5}$$

Here, $w^T=(w_1,\ldots,w_n)$, $\mathbf{1}_n$ represents the n-dimensional column vector of ones, and $u(.)$ denotes a concave utility function evaluating the investor's wealth at time t. The corresponding first-order conditions for the marginal optimal portfolio weights w_j assessed at the conditioning variable x_j are:

$$\mathrm{E}\left[(R_{it}-R_{nt})\,\dot{u}\left(w_j^T\mathbf{R}_t\right)\mid X_{j,t-1}=x_j\right]=0 \quad \text{a.s. }, i=1,\ldots,n-1 \tag{3.6}$$

Here, $w_j^T=[w_{1(j)},\ldots,w_{n(j)}]$ with $w_{n(j)}=1-\sum_{i=1}^{n-1} w_{i(j)}$.

By utilizing the sample data, the first-order conditions of the marginal optimal portfolio, as described in (3.6), can be represented as follows for each $j=1,\cdots,p$:

$$\frac{1}{Th}\sum_{t=1}^{T}\left[(R_{it}-R_{nt})\,\dot{u}\left(w^\top\mathbf{R}_t\right)K\left(\frac{X_{j,t-1}-x_j}{h}\right)\right]=0 \quad \text{a.s. }, i=1,\ldots,n-1. \tag{3.7}$$

Here, $K(.)$ represents the kernel function, and $h\to 0$ denotes a bandwidth parameter that converges to zero as $T\to\infty$. The solution of (3.7) is denoted by $\widehat{w}_j^T=[\widehat{w}_{1(j)},\ldots,\widehat{w}_{n(j)}]$ where $\widehat{w}_{n(j)}=1-\sum_{i=1}^{n-1}\widehat{w}_{ij}$.
As proposed by Chen et al. (2016), after estimating the optimal portfolios, we construct a joint portfolio choice by combining the marginal optimal portfolios selected across all the conditioning variables using the model averaging weights, which is expressed as follows:

$$\widehat{w}_a(\boldsymbol{X_t})=\sum_{j=1}^{p} a_j\widehat{w}_j(X_{jt}) \quad \text{with}\sum_{j=1}^{p} a_j=1, \tag{3.8}$$

Here, $\mathbf{a}=(a_1,...,a_p)^T$ represents the weights assigned to each marginal optimal portfolio selected for each conditioning variable X_j. Next, we discuss the selection of the weights $\mathbf{a}=(a_1,...,a_p)^T$ in (3.8).
The performance of the portfolio choice in (3.8) mainly depends on the determination of the model averaging weights $a_1,...,a_p$. Let $w_{\mathbf{a},t}\equiv w_a(\boldsymbol{X}_t)=\sum_{j=1}^{p} a_j w_j(X_{jt})$. The optimal model averaging weights $\mathbf{a}_o$ are obtained by maximising:

$$\mathbf{a}_o=\arg\max_{\mathbf{a}}\mathrm{E}\left\{u\left[\sum_{j=1}^{p} a_j w_j^T(X_{j,t-1})\mathbf{R}_t\right]\right\},\ \text{s.t. }\sum_{j=1}^{p} a_j=1. \tag{3.9}$$

Here we replace the weights $w_j^T(X_{j,t-1})$ by the estimated values $\widehat{w}_j^T(X_{j,t-1})$, that is the solution of (3.7). We estimate $\mathbf{a}_o$ by $\widehat{\mathbf{a}} = (\widehat{a}_1, ..., \widehat{a}_p)^T$, which satisfies the following equations:

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{R}_{wj,t} - \hat{R}_{wp,t}) \cdot \dot{u}\left(\sum_{j=1}^{p}\hat{a}_j\hat{R}_{wj,t}\right) = 0 \text{ for } j = 1, ..., p-1,$$

and $\hat{a}_p = 1 - \sum_{j=1}^{p-1}\hat{a}_j$, where $\hat{R}_{wj,t} = \hat{w}_j(X_{j,t-1})R$ and $\hat{R}_{wp,t}$ is the return obtained from the last conditioning variable p. For a comprehensive understanding of the estimation techniques for the model averaging weights $\mathbf{a} = (a_1, ..., a_p)^T$, the reader is referred to Chen et al. (2016), Section 4.
The following subsection adapts this approach to the recent quantile utility framework developed by de Castro and Galvao (2019) and de Castro et al. (2022).

3.2.2 Quantile Utility Framework

For a specified risk attitude $\tau \in (0, 1)$, in the presence of conditioning variables within the QP framework, the optimal portfolio weights are obtained as a solution to the following general maximisation problem:

$$\begin{aligned} &Q_\tau[u(w^T\mathbf{R}_t)|X_{1,t-1}, ..., X_{p,t-1}], \\ \text{s.t. } &\mathbf{1}_n^\top w = \sum_{i=1}^{n} w_i = 1, w_i \geq 0, \end{aligned} \tag{3.10}$$

Here, $w = (w_1, \ldots, w_n)^T$ represents the portfolio weights for n assets, $\mathbf{1}_n$ denotes a column vector of ones with n dimensions and $u(.)$ denotes a utility function that exhibits strict monotonicity with respect to wealth.

Remark 1. Quantiles exhibit several essential properties; a well-known property is its equivariance under monotonic transformations. Formally, if $f(\cdot) : \mathbb{R} \to \mathbb{R}$ is continuous and strictly in creasing function, then

$$Q_\tau[f(X)] = f(Q_\tau[X]). \tag{3.11}$$

In simpler terms, this means that the quantiles of the transformed random variable $f(X)$ correspond to the transformed quantiles of the original variable X. Due to this property, the selection of the utility function becomes irrelevant under QP if it is strictly monotonic, which is formalised in the following remark.

Remark 2. suppose u(.) is a continuous and monotonically increasing utility function defined over the domain of the random variables $w^T\mathbf{R}_t$, subject to the constraint $\mathbf{1}_n^\top w = \sum_{i=1}^n w_i = 1$, where $\mathbf{1}_n$ represents the n-dimensional column vector of ones. The optimal portfolio weight w^* is a solution to (3.10) if and only if it is a solution to the following:

$$\begin{aligned} &\max_w Q_\tau[w^T\mathbf{R}_t|X_{1,t-1},...,X_{p,t-1}], \\ &\text{s.t. } \mathbf{1}_n^\top w = \sum_{i=1}^n w_i = 1, w_i \geq 0. \end{aligned} \tag{3.12}$$

Proof. The objective function of (3.10) is to maximize

$$Q_\tau[u(w^T\mathbf{R}_t)|X_{1,t-1},...,X_{p,t-1}].$$

Noting that $u(.)$ is a continuously increasing utility function and since the quantile is invariant to monotone transformations, hence the former maximisation problem can be written as:

$$\begin{aligned} \arg\max_w Q_\tau[u(w^T\mathbf{R}_t)|X_{1,t-1},...,X_{p,t-1}] &= \arg\max_w u(Q_\tau[w^T\mathbf{R}_t|X_{1,t-1},...,X_{p,t-1}]) \\ &= \arg\max_w Q_\tau[w^T\mathbf{R}_t|X_{1,t-1},...,X_{p,t-1}]. \end{aligned}$$

Our primary focus in the remainder of the Chapter is to analyse the problem described in (3.12) with particular emphasis on scenarios involving a large number of conditioning variables.

3.2.2.1 Model Averaging in a Quantile Setting

To address the curse of dimensionality problem within the QP framework, we are motivated by the methodology presented in Chen et al. (2016) and propose a model averaging approach for portfolio selection under QP. Our strategy focuses on approximating the quantile function $Q_\tau[w^T\mathbf{R}_t|X_{1,t-1},...,X_{p,t-1}]$ by a weighted sum of quantile functions corresponding to each conditioning variable. This weighted sum is expressed as $\sum_{j=1}^p a_j Q_\tau[w_j^T(X_{j,t-1})\mathbf{R}t|X_{j,t-1}]$, where a_j represents the optimal model averaging weights.
By adopting this strategy, our objective shifts from directly maximising $Q_\tau[w^T\mathbf{R}_t|X_{1,t-1},...,X_{p,t-1}]$, to maximising the weighted sum $\sum_{j=1}^p a_j Q_\tau[w_j^T(X_{j,t-1})\mathbf{R}_t|X_{j,t-1}]$ which helps to overcome the challenge of the curse of dimensionality in the presence of large number of conditioning variables. We

decompose the problem into p separate portfolio problems, each involving an individual conditioning variable, and subsequently, we formulate the joint dynamic optimal portfolio by assigning weights to each individual portfolio, with the weights reflecting the importance or influence of each variable on the overall portfolio.
Our approach is accomplished in two stages. In the first stage, we solve the portfolio choice for each individual conditioning variable X_j where $j = 1, ..., p,$. This involves, for each j-th conditioning variable $X_{j,t-1} = x_j$ and a given $\tau \in (0, 1)$, we obtain the optimal portfolio weights under QP, this can be formulated as:

$$\begin{aligned} &\max_{w} \; Q_\tau[w_j^T(x_j)\mathbf{R}_t | X_{j,t-1} = x_j], \\ &\text{s.t. } \sum_{i=1}^{n} w_i = 1, w_i \geq 0. \end{aligned} \tag{3.13}$$

Here, $w_j = [w_{j,1}, \ldots, w_{j,n}]^T$ is the vector of portfolio weights for the j-th conditioning variable, $\mathbf{R}_t$ is the vector of asset returns at time t, and $Q_\tau[\cdot]$ denotes the conditional quantile function at the τ-th quantile level. The optimization problem in (3.13) aims to find the portfolio weights w_j that maximize the conditional quantile of the portfolio returns given the conditioning variable $X_{j,t-1} = x_j$.

By solving this optimization problem for each conditioning variable $j = 1, \ldots, p$, we obtain p vectors of portfolio weights, denoted as $w_{\text{op},1}^T, w_{\text{op},2}^T, \ldots, w_{\text{op,p}}^T$, where each vector corresponds to the optimal portfolio choice for a specific conditioning variable.

$$\mathbf{W}_{(p\times n)} = \begin{bmatrix} w_{\text{op},1}^T \\ w_{\text{op},2}^T \\ \vdots \\ w_{\text{op,p}}^T \end{bmatrix} = \begin{bmatrix} w_{1(1)} & \cdots & w_{n(1)} \\ w_{1(2)} & \cdots & w_{n(2)} \\ \vdots & \ddots & \vdots \\ w_{1(p)} & \cdots & w_{n(p)} \end{bmatrix} \tag{3.14}$$

Here, n represents the portfolio assets' number, and p is the number of conditioning variables.
In the second stage, our goal now is to combine these marginal portfolio choices (p vectors) to establish a joint portfolio choice that accounts for all the conditioning variables. The joint portfolio choice, denoted by $w_{a(\tau)}(\boldsymbol{x})$, is obtained as follows:

$$\begin{aligned} &w_{a(\tau)}(\boldsymbol{x}) = \sum_{j=1}^{p} a_j w_{\text{op,j}}^T(x_j) \; , \\ &\text{s.t. } \sum_{j=1}^{p} a_j = 1, \; a_j \geq 0, \quad \forall j = 1, \ldots, p. \end{aligned} \tag{3.15}$$

By formulating the joint portfolio choice as a weighted average of the marginal portfolio choices, we effectively combine the information from each conditioning variable, allowing for a more comprehensive portfolio selection strategy.[2]
Next, we discuss the selection of the marginal optimal portfolio weights under QP and their weighted average, a_j.

Stage 1. Estimation of the Marginal Optimal Portfolio Weights $w_j(x_j)$

In order to compute the quantile portfolio return for each specific conditioning variable in $\mathbf{X}_{t-1}$, we first replace the unknown portfolio weight vector w^T with a grid of random portfolio weights meeting our total investment constraint $\sum_{i=1}^{n} w_i = 1, w_i \geq 0$.
For $\tau \in (0,1)$ and each known vector $w^T = (w_1, ..., w_n)$ in the grid, where n is the number of assets, we estimate $Q_\tau[w^T \mathbf{R}_t | X_{j,t-1} = x_j]$ by applying linear quantile regression:[3]

$$w^T \mathbf{R}_t = \alpha(\tau) + \beta(\tau) X_{j,t-1} + \epsilon_t(\tau). \tag{3.16}$$

Here, t is the number of observations, $\alpha(\tau)$ and $\beta(\tau)$ denote the quantile regression coefficients at a specific quantile level $\tau \in (0,1)$, $X_{j,t-1}$ represents the j-th conditioning variable where $j = 1, \ldots, p$, and $\epsilon_t(\tau)$ is the tth error term associated with the quantile τ. The error term represents the difference between the observed and predicted values at τ- quantile level; for instance, at the τ-quantile, the median of the error term is zero when $\tau = 0.5$.
We obtain $(\widehat{\alpha}(\tau), \widehat{\beta}(\tau))$ such that,

$$\widehat{Q}_\tau[w^T \mathbf{R}_t | X_{j,t-1} = x_j] = \widehat{\alpha}(\tau) + \widehat{\beta}(\tau) X_{j,t-1}. \tag{3.17}$$

We iterate through each weight vector w^T in the defined grid and obtain a different $\widehat{Q}_\tau$ for each iteration. We next select the optimal quantile portfolio weights $w_j^T(x_j)$

[2]Our proposed methodology does not involve directly averaging quantiles. Instead, we employ a model averaging approach to combine the optimal portfolio weights obtained from separate quantile regressions for each conditioning variable. In the first stage, we solve a quantile optimization problem for each conditioning variable X_j to determine the optimal portfolio weights w_j that maximize the conditional quantile of the portfolio return, $Q_\tau[(w_j^\top R_t)|X_j]$. These optimal portfolio weights are not quantiles themselves but rather the solutions to optimization problems that utilize quantile regression techniques. In the second stage, we combine these optimal portfolio weights using a weighted average, effectively integrating the information from each conditioning variable to construct a joint portfolio that considers all available information. This model averaging approach allows us to address the curse of dimensionality while incorporating the information from multiple conditioning variables in a computationally efficient manner.

[3]Introduced by Koenker and Bassett (1978), quantile regression captures the relationship between specific quantiles of the returns distributions, and inherently adjusts for outliers and extreme events, providing a more efficient risk assessment, see, for instance, Hampel et al. (1986), Huber (2004) and Huber and Ronchetti (2009). Quantile regression has been widely employed in various financial applications throughout the last decade, see Engle and Manganelli (2004), for conditional Value at Risk models; Barnes and Hughes (2002) to study the CAPM for different quantiles of the cross-section of stock market returns; or Allen et al. (2011) that integrate quantile methods into the Fama-French three-factor model.

evaluated at the conditioning variable x_j that maximize $Q_\tau[w_j^T(x_j)\mathbf{R}_t|X_{j,t-1} = x_j]$. i.e. we optimize over w^T on the grid,

$$\begin{aligned} &\max_{w} Q_\tau[(w^T\mathbf{R}_t)|X_{j,t-1} = x_j], \\ &\text{s.t. } \mathbf{1}_n^\top w = \sum_{i=1}^{n} w_i = 1, w_i \geq 0. \end{aligned} \tag{3.18}$$

Let $\widehat{w}_{\text{op,j}}^T$ denote the optimal portfolio weight vector for a specific conditioning variable x_j in (3.18) by using quantile regression. We repeat stage 1 for all the conditioning variables $\mathbf{X}_{t-1} = X_{1,t-1}, ..., X_{p,t-1}$. Therefore, for $j = 1, ...p$ we will have

$$\mathbf{W}^*_{(p\times n)} = \begin{bmatrix} \widehat{w}_{\text{op,1}}^T \\ \widehat{w}_{\text{op,2}}^T \\ \vdots \\ \widehat{w}_{\text{op,p}}^T \end{bmatrix} = \begin{bmatrix} \widehat{w}_{1(1)} & \cdots & \widehat{w}_{n(1)} \\ \widehat{w}_{1(2)} & \cdots & \widehat{w}_{n(2)} \\ \vdots & \ddots & \vdots \\ \widehat{w}_{1(p)} & \cdots & \widehat{w}_{n(p)} \end{bmatrix}$$

We then combine all the estimated optimal portfolio weights by a model averaging approach, which will be explained in the following stage.

Grid-based Approach for Determining Optimal Weights

In our analysis, we employ a grid-based approach to generate a comprehensive set of portfolio weight combinations for determining the optimal weights that maximize the conditional quantile of the portfolio return. Specifically, we use Monte Carlo sampling from uniform distributions to generate a large number of random portfolio weight vectors (e.g., 1000) that satisfy the constraints of full investment ($\sum_{i=1}^{n} w_i = 1$) and non-negative weights ($w_i \geq 0$). This approach provides flexibility in terms of the number of weight combinations considered and the choice of distribution for generating the weights.

While the grid-based approach offers several advantages, it is essential to acknowledge its limitations. Firstly, generating a large grid and evaluating the performance of each weight combination can be computationally intensive, especially as the number of assets or the grid size increases. Secondly, the choice of distribution (currently uniform) for generating the weights may influence the results, and different distributions may be more suitable depending on the specific characteristics of the assets or the investor's preferences. Lastly, while the grid approach explores a wide range of weight combinations, it may not guarantee finding the globally optimal solution, especially if the grid is not sufficiently dense or if the optimization landscape is complex. Despite these limitations, the empirical results demonstrate the effectiveness of the grid approach, combined with quantile regression, in capturing

relevant information and providing useful insights for portfolio selection. The approach can be further refined or adapted based on specific requirements, such as increasing the grid size, using alternative distributions, or incorporating more advanced optimization techniques. Future research could explore potential avenues for enhancing the grid approach, such as using adaptive or iterative grid generation methods, incorporating domain knowledge to guide the weight distribution, or employing parallel computing to handle larger grid sizes efficiently.

Stage 2. Estimation of the weighted average of quantiles

Earlier in (3.15), we introduced the joint portfolio choice, $w_{a(\tau)}(\boldsymbol{x})$, as a weighted average of the marginal optimal portfolio choices, $w^T_{\text{op,j}}(x_j)$, that obtained from solving the individual optimization problems for each conditioning variable. The Performance of (3.15) mainly depends on the selection of the model averaging weights $\mathbf{a} = (a_1, ..., a_p)^T$. In this stage, we aim to combine all the marginal optimal portfolio choices $\widehat{w}^T_{\text{op,j}}(x_j)$ over $j = 1, ..., p$ that were previously selected in Stage 1 in order to establish the joint portfolio choice. The process for estimating the unknown model averaging weights **a** will be discussed next.
Let $\boldsymbol{w}_{a(\tau),t} \equiv w_{a(\tau)}(\boldsymbol{X}_t) = \sum_{j=1}^{p} a_j \, w^T_{\text{op,j}}(X_{jt})$. The optimal model averaging weights $\widehat{\boldsymbol{a}}$ are obtained by solving the following optimization problem:

$$\begin{aligned} \widehat{\boldsymbol{a}} &= \arg\max_{\boldsymbol{a}} E\left\{ u\left[\sum_{j=1}^{p} a_j \, w^T_{\text{op,j}}(X_{j,t-1})\mathbf{R}_t \right] \right\}, \\ \text{s.t.} \quad & \sum_{j=1}^{p} a_j = 1, \quad a_j \geq 0. \end{aligned} \tag{3.19}$$

Here in equation (3.19), we replace the $w^T_{\text{op,j}}(X_{j,t-1})$ with the estimated values $\widehat{w}^T_{\text{op,j}}(X_{j,t-1})$, which are obtained from the solution of Equation (3.18).

It is important to note that by introducing the notation $\boldsymbol{w}_{a(\tau),t}$ we are illustrating to the reader that the expression $a_j \, \widehat{w}^T_{\text{op,j}}(x_j)$ implemented in (3.19) depends on "j" unlike the returns $\mathbf{R}_t$.

Remark. The expression $\left[\sum_{j=1}^{p} a_j \, w^T_{\text{op,j}}(X_{j,t-1})\mathbf{R}_t\right]$ in (3.19) can be illustrated by using the matrix notation as:

$$\mathbf{a}_{(1\times p)}\hat{\mathbf{W}}^*_{(p\times n)}\mathbf{R}_{(n\times 1)} = \begin{bmatrix} a_1 & \cdots & a_p \end{bmatrix} \begin{bmatrix} \hat{w}_{1(1)} & \cdots & \hat{w}_{n(1)} \\ \hat{w}_{1(2)} & \cdots & \hat{w}_{n(2)} \\ \vdots & \ddots & \vdots \\ \hat{w}_{1(p)} & \cdots & \hat{w}_{n(p)} \end{bmatrix} \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix}$$

Motivation for Model Averaging in Stage 2

Our methodology comprises two distinct stages with specific objectives. In Stage 1, our primary goal is to incorporate the investor's quantile preferences and obtain the marginal optimal portfolio weights for each conditioning variable. This is achieved by maximizing the quantile utility function through quantile regression, allowing us to capture the investor's risk preferences and construct portfolios that align with their desired level of downside protection.

However, in Stage 2, our objective shifts to combining the marginal optimal portfolios obtained from Stage 1 to create a comprehensive portfolio that takes into account the information from all the conditioning variables. At this stage, we employ the model averaging approach, where the marginal optimal portfolios are aggregated using a set of weights determined by optimizing a specific objective function. The model averaging approach serves as a practical and effective method for combining the marginal optimal portfolios, allowing us to exploit the diversification benefits of multiple conditioning variables.

By optimizing the weights in the model averaging approach, we aim to find the optimal combination that balances the contributions of each conditioning variable to the overall portfolio performance. The model averaging approach provides a flexible and intuitive framework for aggregating different components or portfolios, making it a suitable choice for this stage of our methodology. It is important to note that the use of model averaging in Stage 2 does not diminish the significance of the quantile preferences established in Stage 1. The quantile preferences are inherently captured and incorporated into the final portfolio through the marginal optimal weights obtained in Stage 1, and the model averaging approach in Stage 2 serves as an effective means to combine these weights.

While alternative approaches, such as optimizing a quantile utility function in Stage 2, were explored, the results were not as satisfactory compared to the model averaging approach. Optimizing a quantile utility function in Stage 2 can be challenging due to the non-additive nature of quantiles and the potential for distortions when combining quantiles from different distributions. In contrast, the use of model averaging in Stage 2 has proven to be a practical and reliable choice for combining the marginal optimal portfolios and achieving the desired portfolio characteristics.

Our empirical results demonstrate the effectiveness of this approach in capturing the investor's preferences and providing satisfactory portfolio performance. The use of model averaging in Stage 2, combined with the quantile preferences in Stage 1, has shown to yield stable and robust portfolio optimization outcomes, as evidenced by the favorable out-of-sample performance and risk measures in our analysis.

3.3 Out-of-Sample Performance Evaluation

In this section, three criteria are used to evaluate our proposed method across various τ values: (i) out-of-sample portfolio Variance statistics,(ii) out-of-sample Sharpe ratio criteria and (iii) out-of-sample Sortino ratio criteria. To conduct a comparative analysis, we compare out-of-sample performance to various portfolio strategies from the existing literature. We arrange our analysis systematically to ensure a comprehensive evaluation of the portfolio strategies. Initially, we conduct an "unconditional" comparison of various portfolio strategies, particularly the unconditional quantile method (Uncond.Q_τ) that has different τ values, against a set of benchmark competitors. These benchmarks include the Equally Weighted Portfolio (EWP), the unconditional Markowitz's Mean-variance portfolio (referred to as Markowitz portfolio throughout the Chapter), and the unconditional Minimum-variance portfolio (Min.Var).

Our goal in studying these strategies is to produce insights that will assist investors in making the optimal decisions when choosing a portfolio strategy that meets specific risk-return goals. Following this comparison, we aim to further explore by comparing the unconditional quantile approach (Uncond.Q_τ) to the conditional quantile approach (Cond.Q_τ). By comparing these quantile approaches, we aim to understand the advantages of including conditioning information in the quantile strategy and evaluate its impact on improving portfolio performance and risk management. An EWP is an investment portfolio where all assets or securities are allocated an equal weight $\frac{1}{n}$ where n represents the number of assets in the portfolio. The unconditional Markowitz portfolio is a fundamental concept in modern portfolio theory. There are three approaches to interpret Markowitz's portfolio optimization problem. First, the objective function is maximising the utility function of the investor:

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \boldsymbol{\mu}^T\mathbf{w} - \gamma\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} \quad & \mathbf{1}_n^T\mathbf{w} = \sum_{i=1}^n w_i = 1, \mathbf{w} \geq 0. \end{aligned} \tag{3.20}$$

This problem is a convex quadratic with linear constraints, which has a closed-form solution:

$$\mathbf{w}_{\text{Markowitz}} = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} + \nu\mathbf{1})$$

where ν is the optimal dual variable, $\nu = \frac{2\gamma - \mathbf{1}^T\Sigma^{-1}\mu}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$. Here, γ is the risk aversion parameter, representing the investor's attitude towards risk. A higher value of γ implies a more risk-averse investor. This risk aversion parameter helps adjust the trade-off between portfolio return and risk. The second alternative reformulation for Markowitz portfolio is minimising the portfolio variance with a constraint that

imposes a lower bound on the expected return :

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \boldsymbol{\mu}^\top \mathbf{w} \geq R_{\min}, \\ & \mathbf{1}_n^T \mathbf{w} = \sum_{i=1}^n w_i = 1, \mathbf{w} \geq 0. \end{aligned} \tag{3.21}$$

Here $R_{\min}$ stands for the lower bound on expected return for the investment. Alternatively, we set the objective function by maximising the expected portfolio return for a target risk as a constraint:

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \boldsymbol{\mu}^T \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \leq \delta^2, \\ & \mathbf{1}_n^T \mathbf{w} = \sum_{i=1}^n w_i = 1, \mathbf{w} \geq 0, \end{aligned} \tag{3.22}$$

where δ denotes an upper bound on the portfolio risk.
The third portfolio strategy we construct in this study is the unconditional Minimum-variance portfolio (Min.Var). This technique is a risk-based portfolio where the optimisation explicitly does not focus on expected returns but mainly on risk reduction. It seeks to construct a portfolio with the lowest possible variance or standard deviation of returns.

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \mathbf{1}_n^T \mathbf{w} = \sum_{i=1}^n w_i = 1, \mathbf{w} \geq 0. \end{aligned} \tag{3.23}$$

Which has a closed-form analytical solution:

$$\mathbf{w}_{\text{Min.Var}} \quad = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{3.24}$$

The last portfolio strategy is the unconditional quantile portfolio (Uncond.Q_τ). For a given $\tau \in (0,1)$ the objective function is:

$$\begin{aligned} & \max_w \; Q_\tau[w^T \mathbf{R}_t], \\ & \text{s.t. } \mathbf{1}_n^T \mathbf{w} = \sum_{i=1}^n w_i = 1, \mathbf{w} \geq 0. \end{aligned} \tag{3.25}$$

The Uncond.Q_τ can be implemented by considering the following steps. First, we consider a grid of portfolio weights. For each known $w^T = (w_1, ..., w_n)$, we compute the portfolio return $w^T \mathbf{R}_t$. Then, we sort the data in ascending order to calculate the empirical quantile of the portfolio return. The quantile is given by the order statistic in the position $\tau\, T$, with the quantile $\tau \in (0,1)$ and T the number of observations. We

iterate through each weight vector w^T in the defined grid and calculate the empirical quantile of the portfolio for each iteration. Subsequently, we construct our portfolio by selecting the weight vector w^T from the grid that maximises the quantile of the portfolio return.
In all described portfolio strategies, the constraint $\mathbf{1}_n^T\mathbf{w} = \sum_{i=1}^n w_i = 1$ ensures that the total sum of portfolio weights equals one. In addition, the constraint $\mathbf{w} \geq 0$ represents the presence of a short-selling constraint. For both Markowitz and Min.Var portfolios, the expected return is calculated as the weighted sum of the unconditional individual asset returns, and the portfolio variance is computed based on the unconditional covariance matrix of asset returns.

3.3.1 Performance Evaluation Metrics

We employ the out-of-sample evaluation criteria, namely the portfolio Sharpe ratio, the Sortino ratio and the Variance, to compare the performance of our proposed approach to the previously discussed strategies. To conduct this comparison, we adopt the following rolling window strategy. We implement a window size M=12 months(1 year), including m = 11 months for the in-sample estimation window, and h represents the out-of-sample, which is one month (20 days),$m < T$ and T is the total number of months in the data set. The window will move forward at each step by one month (20 days). In the first rolling window, data from day 1 through m is included, and in the subsequent window, observations from day 21 through m + 20 are considered. This process continues iteratively. Diagram 3.1 illustrates the partitioning process.

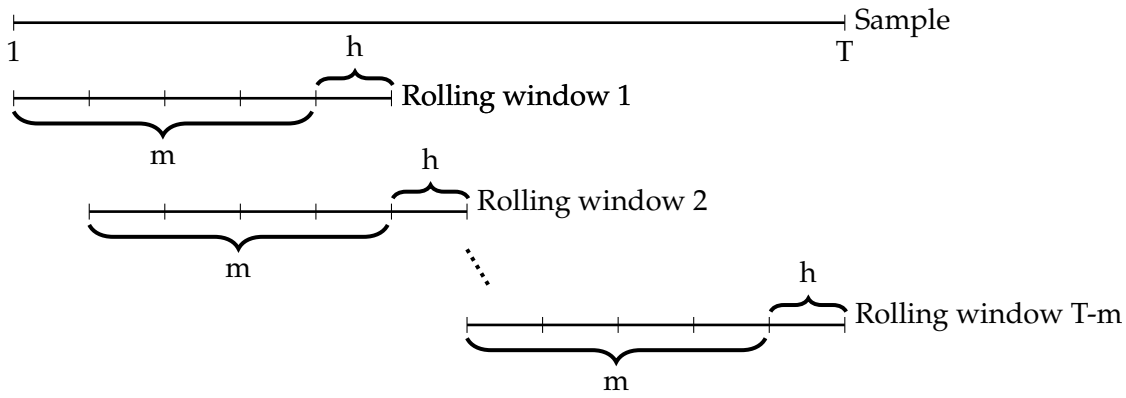


FIGURE 3.1: Rolling window process

Based on the in-sample estimations (m), we construct various portfolios. At the end of this partitioning process, for each strategy k, we have generated $T - m$ portfolio weight vectors; that is w_t^k for $t = m, ..., T - 1$. For each strategy k, holding the portfolio w_t^k for one month gives the out-of-sample portfolio return at time $t + 1$ denoted as $\widehat{R}_{(p),t+1} = w_t^T R_{t+1}$, where R_{t+1} denotes the asset returns vector.

The three criteria employed in this study assess the strategies' risk-adjusted performance. The Sharpe ratio calculates a portfolio's risk-adjusted return, measuring the excess return per unit of volatility (or standard deviation). A higher Sharpe ratio indicates a more favourable risk-adjusted performance. The out-of-sample Variance statistics and Sharpe ratio for strategy k are:

$$\left(\widehat{\sigma}^k\right)^2 = \frac{1}{T-m-1}\sum_{t=m}^{T-1}\left(\mathrm{w}_t^{k^\top} R_{t+1} - \widehat{\mu}^k\right)^2, \tag{3.26}$$

with,

$$\widehat{\mu}^k = \frac{1}{T-m}\sum_{t=m}^{T-1}\mathrm{w}_t^{k^\top} R_{t+1} \tag{3.27}$$

$$\widehat{\text{Sharpe}}^k = \frac{\widehat{\mu}^k}{\widehat{\sigma}^k} \tag{3.28}$$

While the Sharpe ratio evaluates the strategy's upside and downside volatility, the Sortino ratio focuses only on the downside volatility of the strategy. It is calculated by dividing the strategy's excess return by its downside deviation. This helps evaluate the strategies' downside risk and identify those with lower downside volatility, which captures investor preferences for minimising downside deviations during challenging times:

$$\widehat{\text{Sortino}}^k = \frac{\widehat{\mu}^k - MAR}{\widehat{\sigma}_d^k} \tag{3.29}$$

Here $\widehat{\sigma}_d^k$ is the standard deviation of the downside (negative) returns of a portfolio which can be defined as

$$\widehat{\sigma}_d^k = \sqrt{\frac{1}{T-m}\sum_{t=m}^{T-1}\left(\min\left(\mathrm{w}_t^{k^\top} R_{t+1} - MAR, 0\right)\right)^2},$$

Where MAR stands for Minimum Acceptable Return, a higher Sortino ratio suggests superior risk-adjusted performance regarding downside risk, the reader is referred to Sortino and Van Der Meer (1991), Mohan et al. (2016) and Basile et al. (2016) for further discussion of the Sortino ratio and MAR.

Bootstrapping is employed for statistical inference through resampling techniques to assess the significance of the mean Sharpe (Sortino) ratio differences for given pairs of portfolios. We compute the differences between the two strategies' mean Sharpe (Sortino) ratios for each observation, then resample these differences with replacement and compute the mean differences for each bootstrap sample. The hypothesis we are testing for each portfolio comparison is as follows: the Null Hypothesis(H_0): the mean Sharpe (Sortino) ratio of our strategy is less than or equal to the mean Sharpe (Sortino) ratio of the competitor's strategy. i.e. $H_0 : \mu_{\text{our strategy}} \leq \mu_{\text{competitor}}$. The Alternative

Hypothesis (H_a): Our strategy's mean Sharpe (Sortino) ratio is greater than the competitor's. Suppose the p-value is less than the predefined significance level (e.g., 0.05); then, we reject the null hypothesis, showing that our strategy has a statistically significant higher Sharpe (Sortino) ratio than the competitor's strategy.

3.4 Empirical application

This section presents an empirical application that demonstrates the methodologies developed in the Chapter. We investigate the dynamic quantile portfolio choice by considering multiple conditioning variables and allowing the dimension of the conditioning variables to be large. In both examples, we utilise our approach on a rolling-window basis as described in section 3.3.1. We also employ the out-of-sample evaluation criteria: the Sharpe ratio, the Sortino ratio and the portfolio out-of-sample Variance statistics. In addition, we implement bootstrapping methods to measure the significance of the mean Sharpe (Sortino) ratio differences for given pairs of portfolios. In the below empirical exercises, our objective is to allocate the optimal portfolio weights for the conditional quantile preferences model across the quantile levels $\tau \in (0,1)$, utilizing lag-one returns as the conditioning variables, i.e. $\mathbf{X}_{t-1} = \mathbf{R}_{t-1}^T$. Hence, the number of conditioning variables involved in the exercises equals the number of assets evaluated for investment purposes, denoted as $p = n$. The lag-one returns are chosen for convenience of computation. In both examples, we set the Minimum Acceptable Return (MAR) $= 3\%$. After analysing the historical performance of Portfolio 1 (6 assets) and Portfolio 2 (68 assets), we found that Portfolio 1 has mean returns that fall between 0% and 4% and standard deviations indicating different levels of risk. Portfolio 2, with the majority (51 out of 68) of assets, achieved returns above the 3% MAR. Therefore, the 3% MAR for both portfolios reflects the balance between achieving a reasonable return and managing risk. We also apply the CRRA utility function with $\gamma = 2$ for both examples. We consider additional γ values commonly used in the literature to indicate various risk aversion preferences as a robustness exercise. The findings are presented in the Appendix.
In both examples, our data covers the period from 2007 to 2022, capturing a range of distresses with varying effects on the global economy. For example, the global financial crisis occurred between 2007 and 2010 and greatly influenced the global economy. This event started in the US mortgage market and led to global financial instability. From 2011- 2015, there were global events, such as the European debt crisis and the Chinese stock market crash. Greece, Portugal, Ireland, Italy, and Spain, known as the "PIIGS" countries, were the most heavily affected by the European debt crisis. The dramatic drop in stock prices on the Shanghai and Shenzhen stock exchanges during this time is called the "Chinese stock market crash". Compared to the Global Financial Crisis, both events concentrated more on specific countries and had a more

indirect and comparatively lesser influence on the global economy. Between 2016 and 2019, the economy was steady compared to prior financial events. Finally, the COVID-19 Pandemic covers the period 2020-2022, and it is a global health crisis that has had significant impacts on social, economic, and public health worldwide. It originated in Wuhan, China, and quickly spread globally through international travel and person-to-person transmission, resulting in a sharp increase in cases in most countries.

After analysing the global economy impact of the four periods, it becomes clear that some periods can be categorised as crises, such as the Global Financial Crisis 2007-2010 and the COVID-19 Pandemic 2020-2022, while we consider 2011-2015 and 2016-2019 as non-crisis since it had a relatively more minor impact because it was more focused on specific countries not globally. This division and differentiating between crisis and non-crisis periods can provide valuable information about our portfolio analysis for several reasons. For instance, it allows for a comparative analysis of the portfolio's performance during different market conditions. This helps to understand the portfolio behaviour during times of stability vs times of crises, enabling us to evaluate their flexibility and ability to adapt to changing economic conditions.

As a result, to achieve a more comprehensive understanding of the portfolio's performance in both examples, we first analyse our optimal portfolio selection over the entire study period from 2007 to 2022. Furthermore, to better understand conditional information's role in influencing conditional portfolio strategies during challenging times, we divide the out-of-sample evaluation of the entire period into two categories: crises and non-crises, as presented in diagram 3.2.

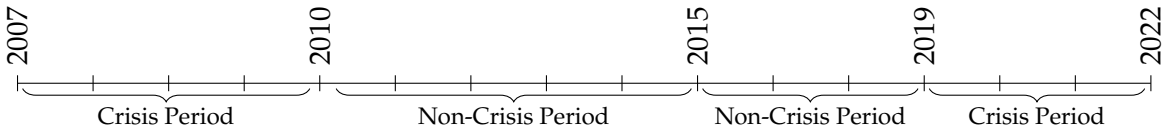


FIGURE 3.2: Timeline of Financial Crises and Global Events (2007 -2022)

3.4.1 Example 1. Small Portfolio Allocation Problem

3.4.1.1 Data Description

In this analysis, we examine the optimal portfolio selection of an investor with quantile preferences who can distribute wealth among six assets: the Bloomberg US Aggregate Bond Index, Treasury Inflation-Protected Securities (TIPS), SPDR Gold Shares, S&P 500, Dow Jones and NASDAQ. The Bloomberg U.S. Aggregate Bond is an exchange-traded fund (ETF) representing a broad U.S. fixed-income market index. The second asset is TIPS, a type of U.S. government bond that protects against

inflation. SPDR Gold Shares is an ETF that tracks the performance of the price of gold. It is one of the most popular and widely traded gold ETFs. The last three assets are stock indexes: S&P 500, Dow Jones and NASDAQ. We consider a sample of daily returns data that covers the period from January 2007 to December 2022. The lag-1 returns of the six assets serve as conditioning variables, i.e. $\mathbf{X}_{t-1} = \mathbf{R}_{t-1}^T$. Hence, we seek to choose an optimal quantile portfolio that depends on six variables.

3.4.1.2 Preliminary Analysis

In Figure 3.3, a plot of the daily log returns on each asset is displayed. Additionally, Figure 3.4 reports nonparametric kernel Gaussian estimates illustrating the unconditional density function of financial returns for the six assets. Visually examining the density plots reveals that all six density functions have similar mean returns but have notably different standard deviations. However, a formal statistical test rejects the null hypothesis that all pairwise combinations have equal mean. Furthermore, the Shapiro-Wilk normality test has been implemented for the six assets. The resulting p-values for all assets are extremely small, suggesting the rejection of the null hypothesis of normality and indicating significant deviations in the return distributions for all assets from a normal distribution.

Table 3.1 summarises the descriptive statistics of all asset returns, including information about each asset's mean return, standard deviation, and highest and lowest observed returns. Based on these summary statistics, it can be inferred that the stock indexes exhibit the highest expected return and variance and are followed by SPDR Gold Shares in terms of expected return and risk. Conversely, the bond indexes have the lowest mean and variance. In Figure 3.5, we select Cond.$Q_{0.05}$ as a representative example of our proposed approach, showing how our proposed portfolio dynamically allocates weights over time.

3.4.1.3 Discussion of the Out-of-Sample Performance analysis

In this example, we have calculated 181 out-of-sample portfolio Variance statistics, Sharpe ratios and Sortino ratios for each portfolio strategy. We constructed the optimal portfolio across different τ values in our proposed portfolio approach.

First, we make a comparison between unconditional portfolio strategies: EWP, Markowitz, Min.Var, Uncond.$Q_{0.01}$, Uncond.$Q_{0.05}$ and Uncond.$Q_{0.10}$. Table 3.2 provides summary statistics, including mean and standard deviation Sharpe ratios across unconditional portfolio strategies covering the entire study period 2007-2022. Among all the strategies, the Uncond.$Q_{0.01}$ and EWP strategies achieve the highest mean Sharpe ratio at 0.082 and 0.080, respectively, while the Markowitz strategy

shows the lowest mean Sharpe ratio at 0.035. Second, to study the role of the conditioning set, we conducted a comparison between conditional quantile strategies (Cond.Q_τ) and unconditional quantile strategies (Uncond.Q_τ). Table 3.3 illustrates the mean and standard deviation of Sharpe ratios for both conditional and unconditional quantile portfolio strategies throughout the study period from 2007 to 2022. In this Table, among all the strategies, the Cond.$Q_{0.05}$ strategy exhibits the highest mean Sharpe ratios at around 0.09. In contrast, the Uncond.$Q_{0.05}$ strategy shows the lowest mean Sharpe ratios at 0.063. This value suggests that, compared to other strategies, the Uncond.$Q_{0.05}$ strategy might not have performed as effectively on a risk-adjusted basis during this time frame. Regarding risk, the Uncond.$Q_{0.01}$ strategy shows slightly the highest variability with a standard deviation of 0.255.

To measure the statistical difference in means between the out-of-sample Sharpe ratios of a benchmark portfolio and portfolio k competitor, we first check the normality of the Sharpe ratios for all portfolio strategies by conducting various methods such as histograms, Q-Q plots, the Shapiro-Wilk Test and Kolmogorov-Smirnov test, Skewness and Excess Kurtosis. All strategies show positive skewness, suggesting they are all right-skewed to some extent. The excess kurtosis values are positive, indicating that all strategies are heavy-tailed compared to a normal distribution. Since bootstrapping does not require the assumption of normality and can provide robust results even when data does not have a normal distribution, we implement bootstrapping for our analyses.

Bootstrapping is a powerful resampling approach that involves regularly collecting replacement samples from the original data and recalculating the statistic for each sample to estimate the sampling distribution of a statistic. In our analysis, the basic bootstrapping method is utilised to compare the returns of a benchmark portfolio against the competitor. We calculate the difference in means between our benchmark strategy and the competitor for each portfolio of interest. The following process bootstrapped the pair of returns: We randomly drew pairs of returns with replacements from the original data. Then, for each resampled dataset, we calculate the mean difference. We create a distribution of the mean difference by repeating this process for several iterations (set to 1000 in this example).

For each portfolio comparison, we set Cond.Q_τ as the benchmark portfolio and compare it with Uncond.Q_τ with the same τ value. The following were calculated: Mean of the bootstrapped mean differences, 95% confidence interval of the mean difference and p-value to test the Null and Alternative Hypotheses. The Null hypothesis (H_0): the mean of the Sharpe ratio of Cond.Q_τ is less than or equal to the mean of the Sharpe ratio of Uncond.Q_τ with the same τ value of the competitor. The alternative hypothesis (H_a) specifies that the mean of Sharpe ratio of Cond.Q_τ is greater than the mean of the competitor. Table 3.4 reports the bootstrap analysis for the study period from 2007 to 2022. Based on bootstrap resampling and a significance

level of $\alpha = 0.05$, the results for Cond.$Q_{0.05}$ and Cond.$Q_{0.10}$ are positive but only statistically significant for Cond.$Q_{0.05}$. In contrast, the sample is very small for Cond.$Q_{0.01}$, and the results are unreliable.

To examine the influence of conditional information in improving portfolio strategy performance across varying market conditions and investor preferences, Table 3.5 presents a comprehensive overview of all quantile portfolio strategies and their mean Sharpe ratios and standard deviation during both crisis and non-crisis periods. Given the statistically significant evidence showing that the Cond.$Q_{0.05}$ strategy's mean Sharpe ratio outperforms the Uncond.$Q_{0.05}$ strategy, our analysis centres on examining the performance of these two strategies over both crisis and non-crisis periods. The Cond.$Q_{0.05}$ has consistently been superior to Uncond.$Q_{0.05}$ in both crisis periods. Specifically, Cond.$Q_{0.05}$ yields mean Sharpe ratios of 0.076 and 0.032, whereas the Uncond.$Q_{0.05}$ achieved only 0.065 and 0.008, respectively. However, the standard deviation differences between the two strategies are relatively small. This implies that even during unstable market conditions, the conditional strategy provides higher returns while maintaining a slightly smaller level of risk than the unconditional quantile portfolio. The benefits of conditional information are still seen even during the two non-crisis periods. The Cond.$Q_{0.05}$ portfolio again outperforms with a Sharpe ratio of 0.062 and 0.163 compared to 0.045 and 0.121 for the Uncond.$Q_{0.05}$.

The standard deviations are also closely matched between the two strategies, with Cond.$Q_{0.05}$ standard deviations at 0.258 and 0.241 and the Uncond.$Q_{0.05}$ at 0.259 and 0.253, respectively. Regardless of the market conditions, this shows that while the conditional strategy leads to better mean Sharpe ratio returns, it offers a slightly smaller risk level than the unconditional quantile strategy. Our analysis indicates that including conditional information may enhance portfolio strategies and improve performance under various market conditions.

The second out-of-sample evaluation criterion is the Sortino ratio, where risk-adjusted performance measures focus on the downside risk of portfolios. Table 3.6 covers the entire study period 2007-2022 and provides an overview of the performance of out-of-sample Sortino ratios across unconditional portfolios. The Min.Var and Markowitz strategies display the highest mean Sortino ratios with 0.522 and 0.486 values, respectively, and they also carry the highest variability in Sortino ratios at 3.24 and 1.77, respectively, over the 2007-2022 period. However, the EWP and Uncond.$Q_{0.05}$ report the lowest mean Sortino ratio with standard deviation at 0.449 and 0.477, respectively. Table 3.7 compares the conditional and unconditional quantile portfolio strategies from the 2007-2022 study period using Sortino ratios' mean and standard deviation. All Cond.Q_{τ} portfolios dominate in terms of risk-adjusted returns when considering negative volatility, while Uncond.$Q_{0.05}$ exhibits the lowest mean Sortino ratio with 0.10.

We employ bootstrapping methods to measure the statistical difference between the out-of-sample Sortino ratios for the return of Cond.Q_τ quantile portfolio compared to Uncond.Q_τ quantile strategy having the same τ value. Table 3.8 represents the bootstrap analysis covering the study period 2007-2022. We test the null hypothesis that Cond.Q_τ's Sortino ratios are less than or equal to that of Uncond.Q_τ portfolio with the same τ value. If the p-value is smaller than 5%, it provides evidence for rejecting the null hypothesis, suggesting that Cond.Q_τ exhibits a significantly higher Sortino ratio than Uncond.Q_τ. Based on bootstrap resampling, Table 3.8 and corresponding p-value provide evidence to support the rejection of the null hypothesis, indicating that the mean Sortino ratio of Cond.$Q_{0.05}$ strategy is significantly greater than Uncond.$Q_{0.05}$ portfolio. On the other hand, while the mean differences between the rest Cond.Q_τ strategies vs the Uncond.Q_τ with the same τ value are positive, suggesting a higher mean Sharpe ratio for those Cond.Q_τ strategies, the evidence is not enough to reject the null hypothesis.

To assess the impact of conditional information on portfolio outcomes across varying market conditions and investor preferences, Table 3.9 presents a comprehensive overview of all quantile portfolio strategies and their mean Sortino ratios with the standard deviation during both crisis and non-crisis periods. The primary focus of our analysis centres on comparing the performance of conditional and unconditional quantile portfolios with $\tau = 0.05$ across the crises and non-crises intervals. During the initial crisis phase, Uncond.$Q_{0.05}$ portfolios have a Sortino ratio of 0.16 to 0.13 for the Cond.$Q_{0.05}$, which indicates superior downside risk management for Uncond.$Q_{0.05}$ compared to the Cond.$Q_{0.05}$ portfolio. The conditional portfolio, however, performed better than the non-conditional portfolio in the subsequent crisis, with a value of 0.07 to 0.04. Despite these figures, the Cond.$Q_{0.05}$ portfolio reports a sample standard deviation of 0.40 across both crises. In contrast, the Uncond.$Q_{0.05}$ shows greater variability in the first period with a deviation of 0.53, stabilising to 0.40 in the next phase. The Cond.$Q_{0.05}$ portfolio consistently exceeded the Uncond.$Q_{0.05}$ portfolio in the non-crisis period, registering Sortino ratios of 0.12 and 0.24, respectively, compared to 0.08 and 0.12. Nevertheless, this superior performance was paired with increased variability, as evidenced by the standard deviations of 0.56 and 0.58 for the conditional strategy compared to 0.46 and 0.50 for the Uncond.$Q_{0.05}$ strategy. In conclusion, while the conditional strategy has often offered better mean Sortino ratios, especially in non-crisis times, it does so with increased variability in performance. Table 3.10 presents the average out-of-sample portfolio Variance statistics. The comparison of Cond.Q_τ and Uncond.Q_τ were relatively close to each other in the case of $\tau = 0.01, 0.05$. However, the Cond.$Q_{0.10}$ portfolio displays lower mean out-of-sample Variance statistics, at 0.541 compared to Uncond.$Q_{0.10}$ strategy with 0.627 out-of-sample Variance statistics.

Overall, our empirical analysis shows the superior performance of Cond.Q_τ strategy

during the crisis periods. Specifically across the Sharpe ratio metric, Cond.$Q_{0.01}$ and Cond.$Q_{0.05}$ strategies demonstrate their adaptability to manage risk during unstable market conditions. Furthermore, when considering the Sortino ratio, most of the time, the Cond.$Q_{0.01}$ and Cond.$Q_{0.05}$ consistently outperform the remaining competitors. These findings emphasise the tail risk sensitivity of Cond.$Q_{0.01}$ and Cond.$Q_{0.05}$, which is beneficial during crises when extreme events are more likely to occur. Moreover, the risk-averse nature of these lower quantiles, i.e., lower quantiles, reflects a higher level of risk aversion and thereby represents an advantage during unstable market conditions.

3.4.2 Example 2. Large Portfolio Allocation Problem

In this example, we aim to empirically explore the performance of our quantile portfolio strategy conditioning on a large number of assets. This way, we can assess the ability to use our weighted average approach in high-dimensional data. Therefore, we consider a sample of daily returns data on FTSE-100 collected from January 2007 to July 2022. The FTSE-100 index consists of 100 companies; however, our data covers the entire period for 68 of these companies. The remaining companies have entered and exited the index during the period under consideration. Therefore, our analysis primarily focuses on these 68 companies.

We utilize a dataset comprising $T = 3911$ observations on stock returns from 68 companies. The lag-1 returns on these 68 stocks serve as the conditioning variables. Our methodology will be applied using the same rolling-window framework, CRRA utility function, out-of-sample evaluation criteria, and bootstrapping techniques as demonstrated in Example 1. By utilising these approaches and metrics, we aim to comprehensively understand the performance and significance of the portfolios under consideration. In this example, we have calculated 185 out-of-sample Sharpe ratios, Sortino ratios and Variance statistics for each portfolio strategy.

Figures 3.8 and 3.9 illustrate the Sharpe and Sortino ratios of different portfolios covering 2007-2022. Initially, we compare the mean and standard deviation of portfolio Sharpe ratios between unconditional portfolio strategies, with the results provided in Table 3.11. In this Table, the EWP strategy shows the highest return at 0.064, while the Markowitz strategy has the lowest mean Sharpe ratio at 0.008. Additionally, Table 3.12 presents a comparison between Cond.Q_τ and Uncond.Q_τ in terms of a mean and standard deviation of Sharpe ratios. The Cond.$Q_{0.05}$ and Cond.$Q_{0.10}$ strategies provide higher mean returns compared to the Uncond.Q_τ strategies with the same τ value.

We employ bootstrapping methods to measure the statistical difference between the out-of-sample Sharpe ratios for the return of Cond.Q_τ compared to Uncond.Q_τ with

the same τ value. Table 3.13 reports the bootstrap analysis for the study period from 2007 to 2022. The null hypothesis tested is that the mean Sharpe ratio of Cond.Q_τ portfolio is less than or equal to that of Uncond.Q_τ strategy. If the p-value is smaller than 5%, there is strong evidence to support the rejection of the null hypothesis and conclude that Cond.Q_τ has a significantly higher Sharpe ratio than Uncond.Q_τ portfolio. Although Cond.Q_τ strategies with $\tau = .05$ and 0.10 have always had a higher out-of-sample mean Sharpe ratio than Uncond.Q_τ portfolios with the same τ values, the p-values do not report evidence to reject the null hypothesis, indicating that there is no statistically significant difference in the Sharpe ratios between Conditional Q_τ strategies and Unconditional Q_τ strategies.

Table 3.14 provides an overview of all quantile portfolio strategies and their mean Sharpe ratios and standard deviation during both intervals, crisis and non-crisis. The results present the impact of conditional information in enhancing portfolio strategy performance across varying market conditions and investor preferences. To maintain consistency with Example 1, we compare the performance of a conditional quantile portfolio with $\tau = 0.05$ to the unconditional quantile portfolio with the same τ value throughout both crisis and non-crisis periods. By including conditional information, the portfolio has generally been superior to the unconditional portfolio in terms of the Sharpe ratio over the two crisis periods. Specifically, Cond.$Q_{0.05}$ yields a mean Sharpe Ratio of 0.036 and 0.047, whereas the Uncond.$Q_{0.05}$ achieves only 0.028 and 0.039. However, the differences remained relatively close when assessing the portfolios' volatility, represented by the standard deviations. This implies that while the conditional strategy provides higher returns, even during unstable market conditions, it maintains a similar level of risk to the unconditional quantile portfolio. Even during the two non-crisis intervals, the benefits of conditional information persisted. The Cond.$Q_{0.05}$ portfolio again outperforms the Uncond.$Q_{0.05}$, achieving slightly higher Sharpe ratios of 0.086 and 0.067 than the unconditional portfolio of 0.085 and 0.064. The volatility levels remain closely matched between the two strategies during these periods, with the conditional strategy's standard deviations at 0.227 and 0.230 and the unconditional's at 0.232 and 0.234. This finding suggests that even during unstable market conditions, the Cond.$Q_{0.05}$ strategy still offers higher returns while maintaining a comparable level of risk to the Uncond.$Q_{0.05}$ portfolio. Overall, our findings indicate that portfolio strategies can be enhanced by including conditional information and offering potential improvement in performance under a range of market conditions. Moreover, as the conditional strategy yields better returns, it does not add any extra risk compared to strategies that do not consider such conditions.

In our assessment of the Sortino ratios, we introduce two tables providing the performance of various portfolio strategies over the entire study period. Table 3.15 compares unconditional portfolio strategies, presenting each strategy's mean and standard deviation of the Sortino ratios. It shows that, compared to all strategies,

Min.Var strategy has the highest mean returns at 0.132 and the lowest standard deviation at 0.445. By exploring the comparison of the quantile of conditional and unconditional portfolios, the analysis findings are captured in Table 3.16. The results show that Cond.$Q_{0.05}$ has a slightly higher mean return than Uncond.$Q_{0.05}$ while all other strategies show a comparable mean return. Additionally, to examine the statistical difference among the out-of-sample Sortino ratios for Uncond.Q_τ portfolios compared to Cond.Q_τ portfolios, we have conducted a bootstrapping analysis. Table (3.17) reports the bootstrapping results where the null hypothesis is that the mean Sortino ratios of portfolio Cond.Q_τ are less than or equal to that of Uncond.Q_τ with the same τ value. When the p-value is less than 5%, there is strong evidence supporting the rejection of the null hypothesis, suggesting that the Conditional Q_τ strategy exhibits a significantly higher Sharpe ratio compared to the competitor. Based on bootstrap resampling and a significance level of $\alpha = 0.05$, the positive mean differences suggest that Cond.$Q_{0.05}$ might have a higher mean Sharpe ratio than Uncond.$Q_{0.05}$. However, the corresponding p-values suggest no evidence to reject the null hypothesis and conclude that Cond.$Q_{0.05}$ has no difference in Sortino ratio than other strategies.

To assess the impact of conditional information on portfolio outcomes across varying market conditions and investor preferences, Table 3.18 presents a comprehensive overview of all quantile portfolio strategies and their mean Sortino ratios with the standard deviation during both crisis and non-crisis periods. To maintain consistency with Example 1, we focus mainly on comparing the performance of conditional and unconditional quantile portfolios with $\tau = 0.05$ over the crises and non-crises time intervals. With mean Sortino ratios of 0.084 and 0.090 compared to 0.077 and 0.070, the Cond.$Q_{0.05}$ portfolio outperforms the Uncond.$Q_{0.05}$ strategy during both crisis phases, demonstrating superior downside risk management. In the first phase, the standard deviation of the Uncond.$Q_{0.05}$ portfolio shows an increase at 0.38, then decreases at 31; the Cond.$Q_{0.05}$ portfolio consistently shows a standard deviation of 0.35 across both crises.

During non-crisis times, the Cond.$Q_{0.05}$ portfolio consistently outperforms the Uncond.$Q_{0.05}$ with mean Sortino ratios of 0.189 and 0.099 vs 0.177 and 0.093. This better performance is accompanied by high variability, as shown by the standard deviations of 0.641 in the first non-crisis period compared to the unconditional standard deviations of 0.582. In contrast, the differences remain relatively small in the second non-crisis period, with 0.371 and 0.380, respectively. In conclusion, the Cond.$Q_{0.05}$ technique frequently improves average Sortino ratios across crisis and crisis intervals. Furthermore, compared to non-crisis periods, it displays a more stable risk-adjusted profile during a crisis period, emphasising the desirable balance between performance potential and consistency. Table 3.19 presents the average out-of-sample portfolio Variance statistics. All the Uncond.Q_τ strategies display lower

mean out-of-sample Variance statistics than Cond.Q_τ strategies corresponding to the same τ value.

3.5 Conclusion

In this Chapter, we extended the semiparametric dynamic portfolio choice under EU framework proposed by Chen et al. (2016) by incorporating investors' quantile preferences, focusing on specific quantiles of the returns distribution rather than just the mean return. To the best of our knowledge, no prior research has been conducted to avoid the curse of dimensionality issue on the portfolio choice problem under the QP setting when the number of conditioning variables is large. We proposed a methodology for estimating the dynamic portfolio choice under the QP setting. This estimation can be accomplished in two steps. In the first step, we select the marginal optimal portfolio weights $w_j(x_j)$ under QP by maximising the conditional quantile portfolio problem. This involves, for each j-th conditioning variable $X_{j,t-1} = x_j$ and a given $\tau \in (0, 1)$, we obtain the optimal portfolio weights using quantile regressions. In the second stage, the optimal portfolios constructed from the individual conditioning variables are combined by a model averaging approach to obtain an optimal portfolio based on multiple conditioning variables. This approach is similar in spirit to the MAMAR method proposed in Chen et al. (2016) but adapted to a quantile setting. We implemented our approach to construct optimal portfolios under QP with multiple covariates in two empirical exercises. The first exercise focuses on a small portfolio allocation problem that comprises six major financial assets such as bonds, gold, and stock indexes and six conditioning variables. In contrast, the second portfolio allocation problem considers all the assets traded in the FTSE100 over the last 16 years. The empirical applications demonstrated that the proposed methodology performs well most of the time in the out-of-sample evaluation. It illustrated that our proposed quantile method effectively captures the investor's preference for avoiding large downside deviations during crises such as the 2007 financial crisis and the COVID-19 pandemic. It showed that optimal portfolios obtained from the conditional quantile regression approach outperform the unconditional counterpart portfolio strategies under different metrics in out-of-sample settings. This indicates that including conditional information may enhance portfolio strategies and improve performance under various market conditions.

Appendix

Variable	N	Mean	Std.Dev.	Median	Trimmed	MAD	Min	Max	Range	Skew	Kurtosis	SE
S&P 500	4027	0.02	1.30	0.07	0.06	0.76	-12.77	10.96	23.72	-0.51	11.80	0.02
NASDAQ	4027	0.04	1.44	0.10	0.08	0.94	-13.15	11.16	24.31	-0.44	7.53	0.02
Dow Jones	4027	0.02	1.23	0.06	0.06	0.73	-13.84	10.76	24.61	-0.47	14.67	0.02
SPDR Gold Shares	4027	0.02	1.12	0.05	0.04	0.83	-9.19	10.70	19.89	-0.25	6.55	0.02
U.S. Aggregate Bond	4027	0.00	0.34	0.01	0.01	0.22	-7.08	3.80	10.88	-2.62	65.21	0.01
Treasury Inflation-Protected Securities	4027	0.00	0.42	0.01	0.01	0.31	-3.00	4.36	7.35	0.04	9.99	0.01

TABLE 3.1: Summary Statistics

Portfolio	**Mean**	**Std.Dev.**
EWP	0.08	0.253
Markowitz	0.035	0.244
Min.Var	0.054	0.238
Uncond.$Q_{0.01}$	0.082	0.255
Uncond.$Q_{0.05}$	0.063	0.255
Uncond.$Q_{0.10}$	0.068	0.245

TABLE 3.2: Ex.1: Mean and Standard Deviation for the unconditional Portfolio out-of-sample Sharpe Ratios covering the period 2007-2022.

τ **value**	**Portfolio**	**Mean**	**Std.Dev.**
0.01	Cond.$Q_{0.01}$	0.071	0.244
	Uncond.$Q_{0.01}$	0.082	0.255
0.05	Cond.$Q_{0.05}$	0.086	0.249
	Uncond.$Q_{0.05}$	0.063	0.252
0.10	Cond.$Q_{0.10}$	0.08	0.253
	Uncond.$Q_{0.10}$	0.068	0.245

TABLE 3.3: Mean and Standard Deviation of the Portfolio out-of-sample Sharpe Ratios based on quantile preference, covering the period 2007-2022

Portfolio	Bootstrapping analysis		
	Mean Diff.(Bootstrap Estimation)	**95% conf. interval**	**p-value**
Cond.$Q_{0.01}$ **vs Uncond.**$Q_{0.01}$	-0.0327	[-0.1197 , 0.0496]	0.781
Cond.$Q_{0.05}$ **vs Uncond.**$Q_{0.05}$	0.0958	[0.0306 , 0.1706]	0.006
Cond.$Q_{0.10}$ **vs Uncond.**$Q_{0.10}$	0.0381	[-0.0485 , 0.1205]	0.181

TABLE 3.4: Ex.1:Bootstrap out of sample mean Sharpe ratio analysis for Cond.Q_τ vs Uncond.Q_τ strategy with the same τ value. The null hypothesis: the mean Sharpe ratios of portfolio Cond.Q_τ is less than or equal to that of Uncond.Q_τ. If the p-value is smaller than 5%, there is strong evidence to reject the null hypothesis and conclude that Cond.Q_τ has a significantly higher mean Sharpe ratio than Uncond.Q_τ.

$\tau = 0.01$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.01}$	0.057 (0.239)	0.044 (0.250)	0.125 (0.242)	0.058 (0.241)
Uncond.$Q_{0.01}$	0.050 (0.224)	0.061 (0.265)	0.166 (0.250)	0.040 (0.258)
$\tau = 0.05$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.05}$	0.076 (0.226)	0.062 (0.258)	0.163 (0.241)	0.032 (0.251)
Uncond.$Q_{0.05}$	0.068 (0.232)	0.045 (0.259)	0.121 (0.253)	0.008 (0.252)
$\tau = 0.10$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.10}$	0.054 (0.248)	0.049 (0.256)	0.173 (0.239)	0.034 (0.250)
Uncond.$Q_{0.10}$	0.065 (0.229)	0.062 (0.248)	0.132 (0.243)	-0.006 (0.246)

TABLE 3.5: Ex. 1: Mean and Standard Deviation of the Portfolio out-of-sample Sharpe Ratios, based on a quantile preference, over different periods.

Portfolio	Mean	Std.Dev.
EWP	0.108	0.449
Markowitz	0.486	1.77
Min.Var	0.522	3.24
Uncond.$Q_{0.01}$	0.13	0.474
Uncond.$Q_{0.05}$	0.10	0.477
Uncond.$Q_{0.10}$	0.124	0.521

TABLE 3.6: Ex.1:Mean and Standard Deviation for the unconditional Portfolio out-of-sample Sortino Ratios covers the period 2007-2022.

τ value	Portfolio	Mean	Std.Dev.
0.01	Cond.$Q_{0.01}$	0.13	0.527
	Uncond.$Q_{0.01}$	0.13	0.474
0.05	Cond.$Q_{0.05}$	0.144	0.508
	Uncond.$Q_{0.05}$	0.10	0.477
0.10	Cond.$Q_{0.10}$	0.175	0.834
	Uncond.$Q_{0.10}$	0.124	0.521

TABLE 3.7: Ex.1:Mean and Standard Deviation of the Portfolio out-of-sample Sortion Ratios, based on a quantile preference, covers the period 2007-2022.

Portfolio	Bootstrapping analysis		
	Mean Diff.(Bootstrap Estimation)	95% conf. interval	p-value
Cond.$Q_{0.01}$ vs Uncond.$Q_{0.01}$	0.0054	[-0.2181 , 0.251]	0.507
Cond.$Q_{0.05}$ vs Uncond.$Q_{0.05}$	0.2059	[0.0061 , 0.431]	0.023
Cond.$Q_{0.10}$ vs Uncond.$Q_{0.10}$	0.2230	[-0.0974 , 0.657]	0.118

TABLE 3.8: Ex.1:Bootstrap out of sample mean Sortino ratio analysis for Cond.Q_τ vs Uncond.Q_τ strategy with the same τ value. The null hypothesis: the mean Sortino ratios of portfolio Cond.Q_τ is less than or equal to that of Uncond.Q_τ. If the p-value is smaller than 5%, there is strong evidence to reject the null hypothesis and conclude that Cond.Q_τ has a significantly higher mean Sortino ratio than Uncond.Q_τ.

$\tau = 0.01$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.01}$	0.11 (0.42)	0.12 (0.60)	0.19 (0.59)	0.10 (0.38)
Uncond.$Q_{0.01}$	0.08 (0.40)	0.11 (0.50)	0.22 (0.50)	0.10 (0.46)
$\tau = 0.05$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.05}$	0.13 (0.40)	0.12 (0.56)	0.24 (0.58)	0.07 (0.40)
Uncond.$Q_{0.05}$	0.16 (0.53)	0.08 (0.46)	0.12 (0.50)	0.04 (0.40)
$\tau = 0.10$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.10}$	0.10 (0.43)	0.09 (0.54)	0.41 (1.40)	0.07 (0.39)
Uncond.$Q_{0.10}$	0.13 (0.49)	0.11 (0.47)	0.21 (0.64)	0.03 (0.46)

TABLE 3.9: Ex.1:Mean and Standard Deviation of the Portfolio out-of-sample Sortino Ratios, based on a quantile preference, over different periods.

τ **value**	**Portfolio**	**Mean of Portfolio Variance**
0.01	Cond.$Q_{0.01}$	0.569
	Uncond.$Q_{0.01}$	0.579
0.05	Cond.$Q_{0.05}$	0.598
	Uncond.$Q_{0.05}$	0.589
0.10	Cond.$Q_{0.10}$	0.541
	Uncond.$Q_{0.10}$	0.627

TABLE 3.10: Ex.1:Mean of the Portfolio out-of-sample Variance statistics, based on a quantile preference, covers the period 2007-2022.

Portfolio	Mean	Std.Dev.
EWP	0.064	0.244
Markowitz	0.008	0.224
Min.Var	0.061	0.213
Uncond.$Q_{0.01}$	0.062	0.226
Uncond.$Q_{0.05}$	0.06	0.225
Uncond.$Q_{0.10}$	0.062	0.222

TABLE 3.11: Ex.2: Mean and Standard Deviation for the unconditional Portfolio out-of-sample Sharpe Ratios covers the period 2007-2022.

τ value	Portfolio	Mean	Std.Dev.
0.01	Cond.$Q_{0.01}$	0.059	0.231
	Uncond.$Q_{0.01}$	0.062	0.226
0.05	Cond.$Q_{0.05}$	0.064	0.221
	Uncond.$Q_{0.05}$	0.06	0.225
0.10	Cond.$Q_{0.10}$	0.065	0.231
	Uncond.$Q_{0.10}$	0.062	0.222

TABLE 3.12: Ex.2:Mean and Standard Deviation of the Portfolio out-of-sample Sharpe Ratios, based on a quantile preference, covers the period 2007-2022.

Portfolio	Bootstrapping analysis		
	Mean Diff.(Bootstrap Estimation)	95% conf. interval	p-value
Cond.$Q_{0.01}$ vs Uncond.$Q_{0.01}$	-0.0207	[-0.0626 , 0.023]	0.826
Cond.$Q_{0.05}$ vs Uncond.$Q_{0.05}$	0.0250	[-0.0123 , 0.065]	0.103
Cond.$Q_{0.10}$ vs Uncond.$Q_{0.10}$	0.0044	[-0.033 , 0.0402]	0.405

TABLE 3.13: Ex.2:Bootstrap out of sample mean Sharpe ratio analysis for Cond.Q_τ vs Uncond.Q_τ strategy with the same τ value. The null hypothesis: the mean Sharpe ratios of portfolio Cond.Q_τ is less than or equal to that of Uncond.Q_τ. If the p-value is smaller than 5%, there is strong evidence to reject the null hypothesis and conclude that Cond.Q_τ has a significantly higher mean Sharpe ratio than Uncond.Q_τ.

$\tau = 0.01$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.01}$	0.033 (0.231)	0.083 (0.240)	0.046 (0.240)	0.058 (0.220)
Uncond.$Q_{0.01}$	0.039 (0.225)	0.080 (0.239)	0.066 (0.234)	0.045 (0.210)
$\tau = 0.05$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.05}$	0.036 (0.218)	0.086 (0.227)	0.067 (0.230)	0.047 (0.212)
Uncond.$Q_{0.05}$	0.028 (0.230)	0.085 (0.232)	0.064 (0.234)	0.039 (0.204)
$\tau = 0.10$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.10}$	0.037 (0.230)	0.087 (0.247)	0.069 (0.239)	0.049 (0.201)
Uncond.$Q_{0.10}$	0.037 (0.218)	0.090 (0.233)	0.055 (0.229)	0.040 (0.208)

TABLE 3.14: Ex.2:Mean and Standard Deviation of the Portfolio out-of-sample Sharpe Ratios, based on a quantile preference, over different periods

Portfolio	**Mean**	**Std.Dev.**
EWP	0.121	0.476
Markowitz	0.093	0.503
Min.Var	0.132	0.445
Uncond.$Q_{0.01}$	0.121	0.497
Uncond.$Q_{0.05}$	0.113	0.446
Uncond.$Q_{0.10}$	0.124	0.504

TABLE 3.15: Ex.2:Mean and Standard Deviation for the unconditional Portfolio out-of-sample Sortino Ratios covers the period 2007-2022

τ value	Portfolio	Mean	Std.Dev.
0.01	Cond.$Q_{0.01}$	0.122	0.473
	Uncond.$Q_{0.01}$	0.121	0.497
0.05	Cond.$Q_{0.05}$	0.125	0.47
	Uncond.$Q_{0.05}$	0.113	0.446
0.10	Cond.$Q_{0.10}$	0.126	0.469
	Uncond.$Q_{0.10}$	0.124	0.504

TABLE 3.16: Ex.2:Mean and Standard Deviation of the Portfolio out-of-sample Sortion Ratios, based on a quantile preference, covers the period 2007-2022

Portfolio	Bootstrapping analysis		
	Mean Diff.(Bootstrap Estimation)	95% conf. interval	p-value
Cond.$Q_{0.01}$ **vs Uncond.**$Q_{0.01}$	-0.0094	[-0.1209 , 0.0965]	0.573
Cond.$Q_{0.05}$ **vs Uncond.**$Q_{0.05}$	0.0447	[-0.1462 , 0.1306]	0.183
Cond.$Q_{0.10}$ **vs Uncond.**$Q_{0.10}$	-0.0084	[-0.1268 , 0.0904]	0.551

TABLE 3.17: Ex.1:Bootstrap out of sample mean Sortino ratio analysis for Cond.Q_τ vs Uncond.Q_τ strategy with the same τ value. The null hypothesis: the mean Sortino ratios of portfolio Cond.Q_τ is less than or equal to that of Uncond.Q_τ. If the p-value is smaller than 5%, there is strong evidence to reject the null hypothesis and conclude that Cond.Q_τ has a significantly higher mean Sortino ratio than Uncond.Q_τ.

$\tau = 0.01$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.01}$	0.086 (0.398)	0.186 (0.618)	0.074 (0.377)	0.117 (0.394)
Uncond.$Q_{0.01}$	0.086 (0.376)	0.183 (0.693)	0.097 (0.381)	0.086 (0.355)
$\tau = 0.05$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.05}$	0.084 (0.356)	0.189 (0.641)	0.099 (0.371)	0.090 (0.353)
Uncond.$Q_{0.05}$	0.077 (0.389)	0.177 (0.582)	0.093 (0.380)	0.070 (0.318)
$\tau = 0.10$				
Portfolio	**2007-2010**	**2011-2015**	**2016-2019**	**2020-2022**
Cond.$Q_{0.10}$	0.094 (0.413)	0.186 (0.617)	0.101 (0.385)	0.090 (0.345)
Uncond.$Q_{0.10}$	0.086 (0.380)	0.212 (0.721)	0.074 (0.348)	0.077 (0.338)

TABLE 3.18: Ex.2:Mean and Standard Deviation of the Portfolio out-of-sample Sortino Ratios, based on a quantile preference, over different time periods

τ **value**	**Portfolio**	**Mean of Portfolio Variance**
0.01	Cond.$Q_{0.01}$	2.87
	Uncond.$Q_{0.01}$	2.76
0.05	Cond.$Q_{0.05}$	3.85
	Uncond.$Q_{0.05}$	3.58
0.10	Cond.$Q_{0.10}$	2.28
	Uncond.$Q_{0.10}$	1.64

TABLE 3.19: Ex.2:Mean of the Portfolio out-of-sample Variance statistics, based on a quantile preference, covers the period 2007-2022

Comparison between the Sharpe and Sortino ratios of the conditional quantile method, considering various γ values that align with different τ values.

Portfolio	Mean and Standard Deviation		
	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
Cond.$Q_{0.01}$	0.071	0.067	0.076
	(0.244)	(0.231)	(0.242)
Cond.$Q_{0.05}$	0.086	0.082	0.083
	(0.249)	(0.251)	(0.247)
Cond.$Q_{0.10}$	0.080	0.075	0.079
	(0.253)	(0.249)	(0.253)

TABLE 3.20: Ex.1: Mean and (Standard Deviation) of the Portfolio out-of-sample Sharpe ratio, covering the period 2007-2022 for different γ values of $\tau = 0.01, 0.05, 0.10$ respectively.

Portfolio	Mean and Standard Deviation		
	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
Cond.$Q_{0.01}$	0.130	0.104	0.119
	(0.527)	(0.450)	(0.490)
Cond.$Q_{0.05}$	0.144	0.169	0.146
	(0.508)	(0.649)	(0.563)
Cond.$Q_{0.10}$	0.175	0.158	0.135
	(0.834)	(0.715)	(0.520)

TABLE 3.21: Ex.1: Mean and (Standard Deviation) of the Portfolio out-of-sample Sortino ratio, covering the period 2007-2022 for different γ values of $\tau = 0.01, 0.05, 0.10$ respectively.

Portfolio	Mean and Standard Deviation		
	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
Cond.$Q_{0.01}$	0.0587	0.0594	0.0631
	(0.231)	(0.222)	(0.229)
Cond.$Q_{0.05}$	0.0641	0.0667	0.0644
	(0.221)	(0.218)	(0.223)
Cond.$Q_{0.10}$	0.0649	0.0585	0.0615
	(0.231)	(0.229)	(0.224)

TABLE 3.22: Ex.2: Mean and (Standard Deviation) of the Portfolio out-of-sample Sharpe ratio, covering the period 2007-2022 for different γ values of $\tau = 0.01, 0.05, 0.10$ respectively.

Portfolio	Mean and Standard Deviation		
	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
Cond.$Q_{0.01}$	0.122	0.129	0.121
	(0.473)	(0.512)	(0.437)
Cond.$Q_{0.05}$	0.125	0.118	0.127
	(0.470)	(0.414)	(0.482)
Cond.$Q_{0.10}$	0.126	0.140	0.147
	(0.469)	(0.599)	(0.739)

TABLE 3.23: Ex.2: Mean and (Standard Deviation) of the Portfolio out-of-sample Sortino ratio, covering the period 2007-2022 for different γ values of $\tau = 0.01, 0.05, 0.10$ respectively.

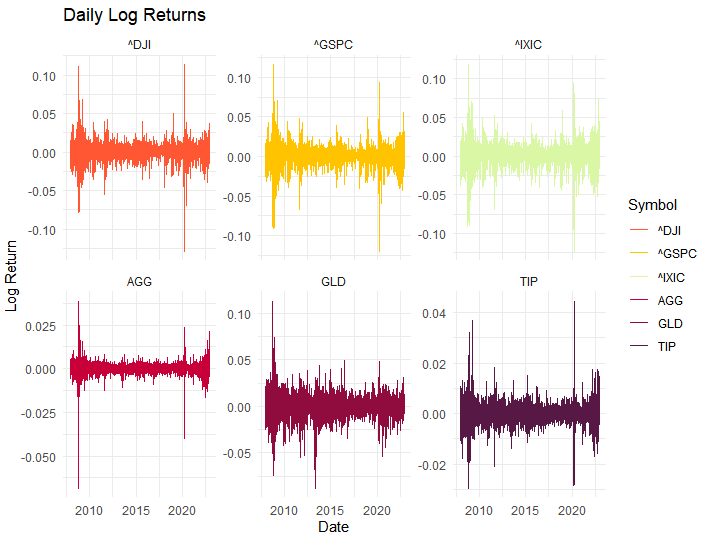


FIGURE 3.3: Ex.1:Daily return of the six assets: (S&P 500 = GSPC, NASDAQ = IXIC, Dow Jones = DJI, SPDR Gold Shares = GLD, U.S. Aggregate Bond = AGG, Treasury Inflation-Protected Securities = TIP)

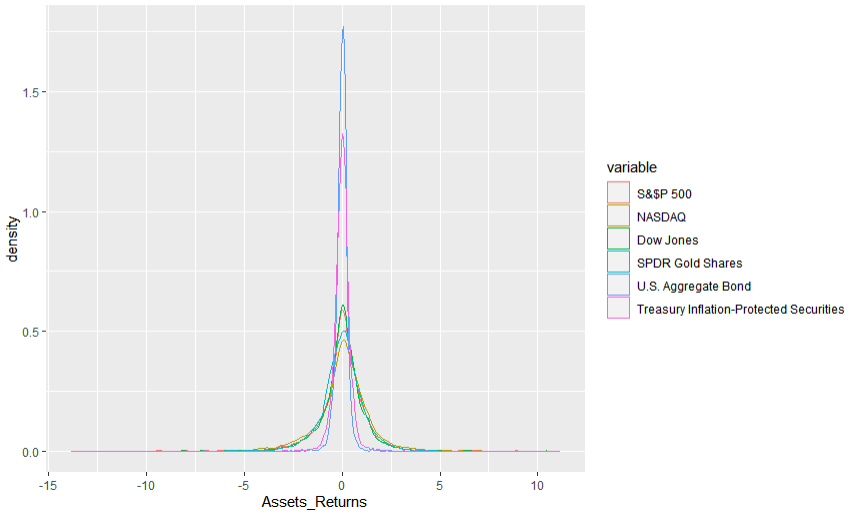
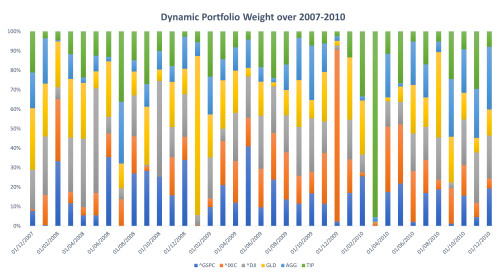
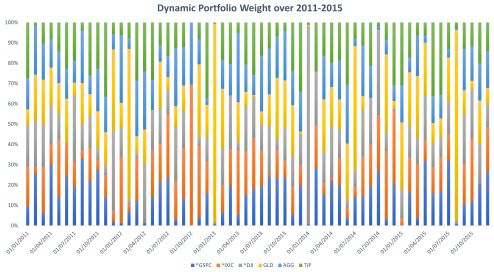


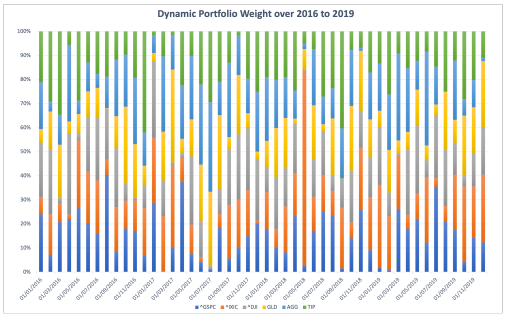
FIGURE 3.4: Ex.1:Nonparametric kernel estimates of the unconditional densities of daily log-returns on the six assets



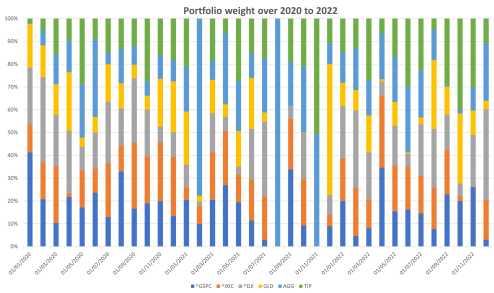
(A) 2007-2010



(B) 2011-2015



(C) 2016-2019



(D) 2020-2022

FIGURE 3.5: Ex1.The Dynamics of Portfolio Weights in Cond.$Q_{0.05}$ strategy over 2007 to 2022

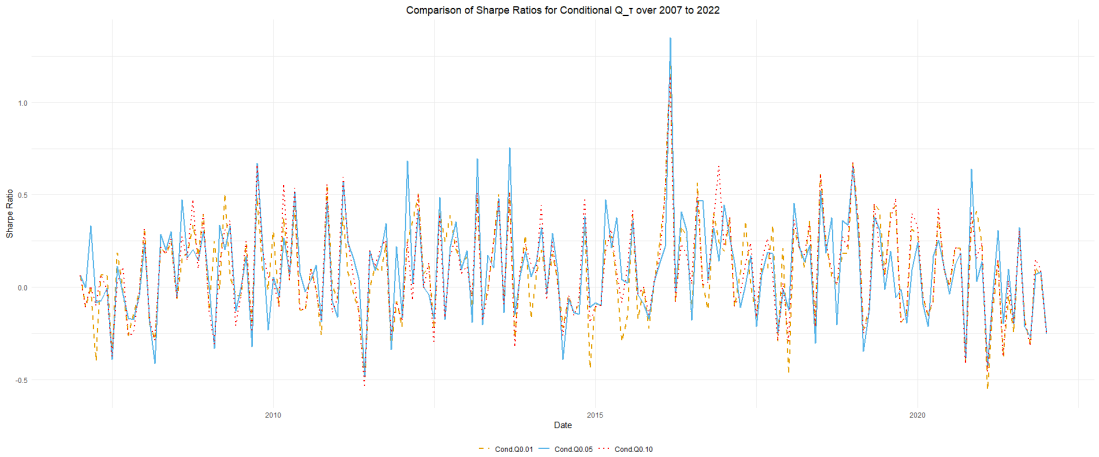


FIGURE 3.6: Ex1. Comparison of Sharpe Ratios for Conditional Quantile Portfolio Strategies with Different Risk Levels over 2007-2022

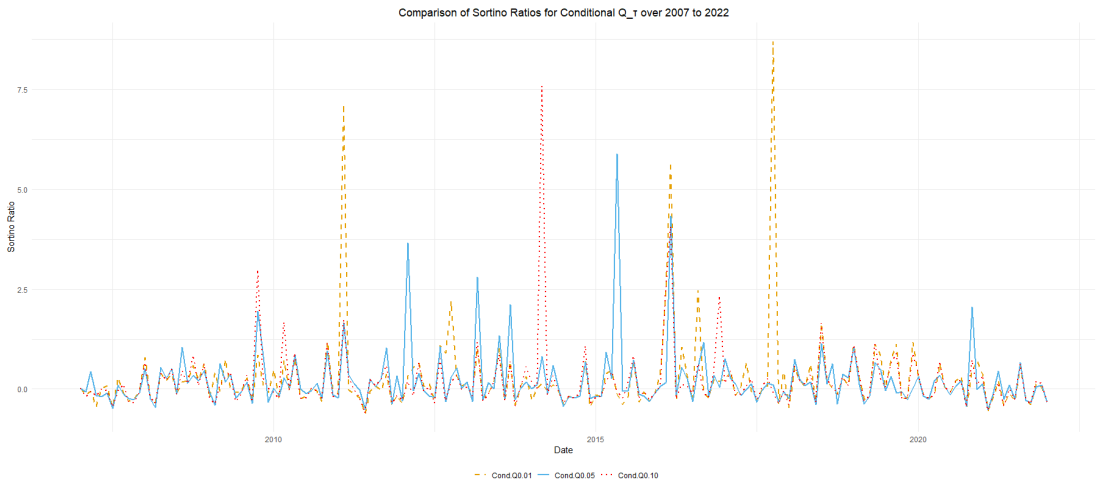


FIGURE 3.7: Ex1. Comparison of Sortion Ratios for Conditional Quantile Portfolio Strategies with Different Risk Levels over 2007-2022

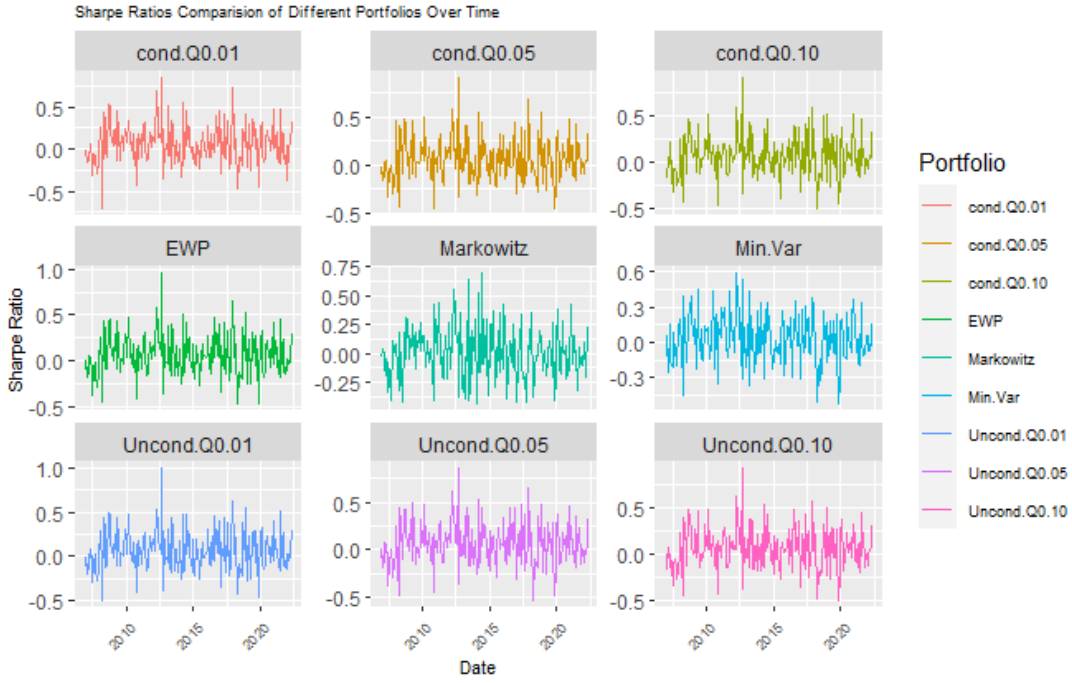


FIGURE 3.8: Ex2. Comparison of Sharpe Ratios across Portfolios2007-2022

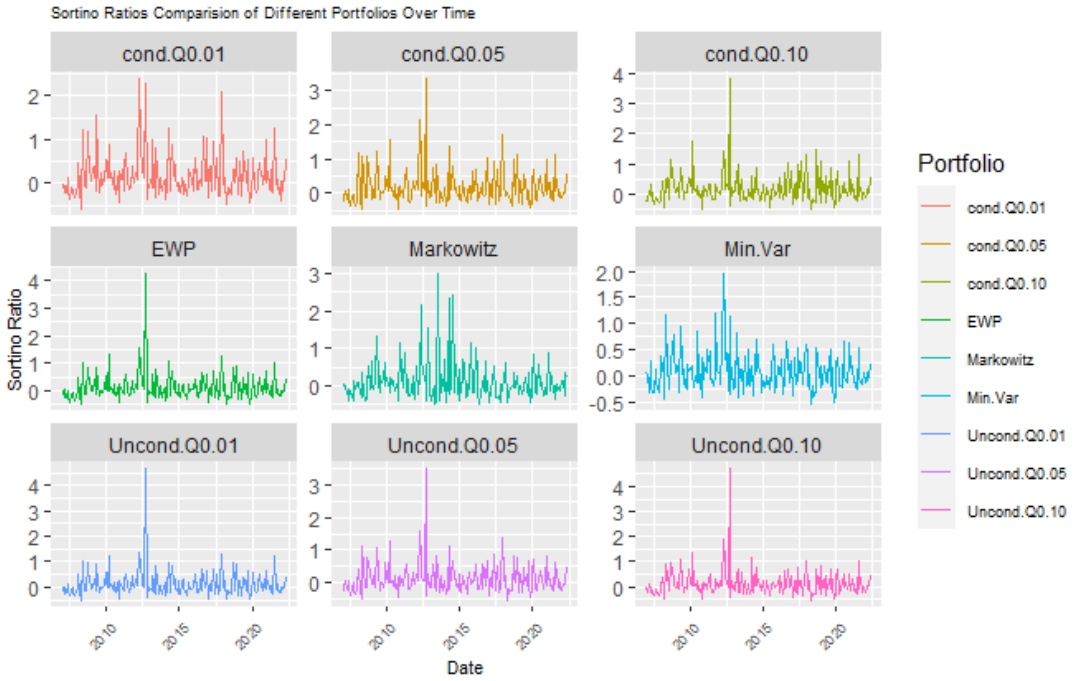


FIGURE 3.9: Ex2. Comparison of Sortino Ratios across Portfolios 2007-2022

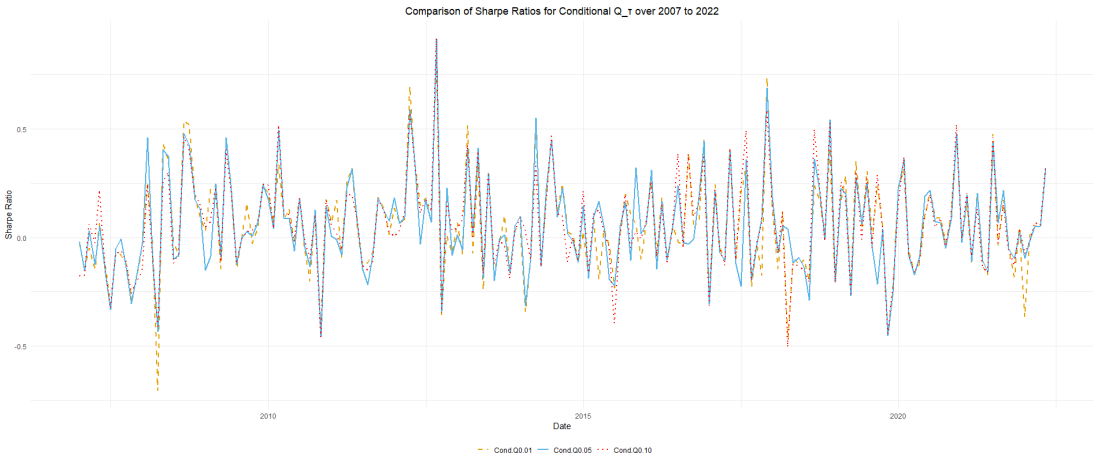


FIGURE 3.10: Ex2. Comparison of Sharpe Ratios for Conditional Quantile Portfolio Strategies with Different Risk Levels over 2007-2022

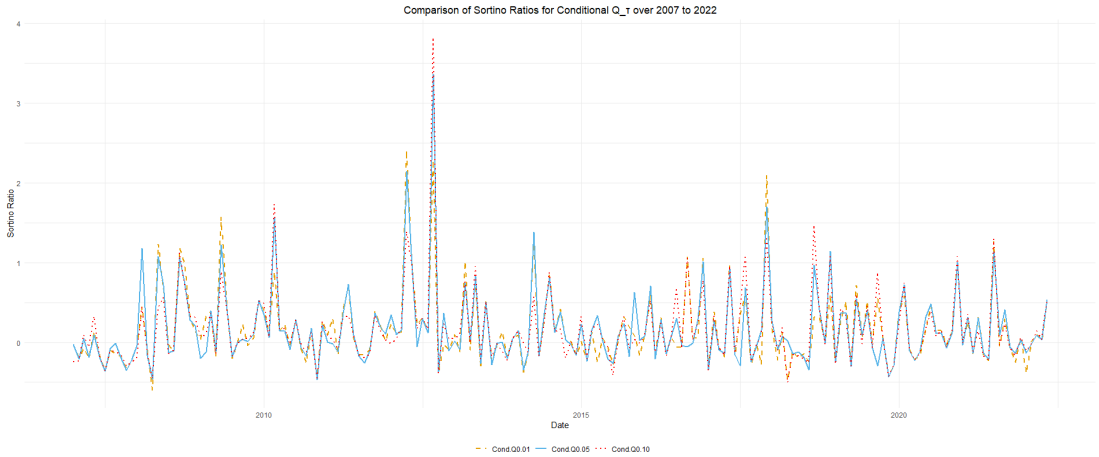


FIGURE 3.11: Ex2. Comparison of Sortion Ratios for Conditional Quantile Portfolio Strategies with Different Risk Levels over 2007-2022

Chapter 4

Simplifying the Complexity: Backtesting of Systemic Risk Measures

4.1 Introduction

Crises such as the US subprime mortgage in 2008–2009 and the COVID-19 pandemic in 2020 have brought attention to the interdependence within the financial system because losses among these institutions naturally spread in times of distress, exposing the entire financial system to vulnerability. Financial regulations, like the Basel capital requirements, aim to constrain the risk associated with each institution when considered in isolation. Consequently, while individual risks might be effectively managed under regular conditions, the system as a whole is sometimes induced to be fragile and vulnerable to significant macroeconomic shocks. Systemic risk is the potential for an extremely unfavourable event at an individual institution or a set of interconnected institutions to trigger a domino effect. It may lead to serious instability or the collapse of a whole financial system. The European Central Bank (ECB) defines systemic risk as ”a risk of financial instability so widespread that it impairs the functioning of a financial system to the point where economic growth and welfare suffer materially European Central Bank (ECB) (2010) ”. These crises have also underscored the significance of managing and measuring systemic risk in the financial sector rather than solely regulating individual financial institutions to protect overall financial stability. Prior to that, risk assessment mainly centred on the marginal distribution of losses, and the two standard metrics for assessing firm-level risk were the Value at Risk (VaR) and Expected Shortfall (ES). These measures aim to assess the potential loss a firm could experience during an extreme event. Regarding VaR, the seminal work by Artzner et al. (1999) establishes a robust axiomatic framework for

risk measures. VaR represents the maximum loss (negative log return) that a firm could experience with a confidence level of $1-\alpha$, meaning that the probability of the loss exceeding this VaR value is α. The parameter α is commonly set at 1% or 5%. For instance, when α is 5%, VaR indicates the maximum loss the firm could have with 95% confidence. The concept of VaR was soon extended to ES by Rockafellar et al. (2000) and Acerbi and Tasche (2002), among others. ES provides a more comprehensive risk measure by considering the expected value of losses beyond the VaR threshold. ES calculates the average loss (negative log return) given that the losses exceed the VaR at a specified confidence level α. Mathematically, it is defined as: $ES_\alpha = E[X|X > VaR_\alpha]$,where X represents the loss (negative log return).

Over the last decade and as a consequence of the crises, systemic risk has been an active area of research, with a substantial amount of literature focusing on measuring systemic risk. The primary objective of a systemic risk measure is to offer a comprehensive assessment and understanding of vulnerabilities present in the financial system. It aims to quantify these risks and contribute to creating policies and regulations that boost stability, thereby diminishing the probability of systemic crises. Two alternative perspectives are involved in addressing the measurement of systemic risk: quantifying systemic risk for the entire financial sector and determining the impact of each institution on the overall systemic risk to understand how the distress of a particular institution could spread through the financial system, contributing to broader risks. Researchers have developed and proposed a variety of methodologies and models to examine the two perspectives, the comprehensive level of risk within the financial system and the marginal systemic risk of individual institutions, see(Giesecke and Kim (2011), Acharya et al. (2017) Pedersen et al. (2010) Brownlees and Engle (2017), Huang et al. (2009), Tobias and Brunnermeier (2016), Feinstein et al. (2017) and others).

In this Chapter, we reexamine the forecasting of two significant systemic risk measures. We explore the risk measures introduced first by Tobias and Brunnermeier (2016) Conditional value at risk (CoVaR) and Conditional Expected Shortfall (CoES). These measures, designed to assess the interdependencies among individual firms and the overall market, are extensions of the traditional VaR and ES in the context of systemic risk. If the losses of interest are represented by Y and the losses of a reference position are denoted by X, then the VaR(ES) of Y at a certain confidence level α, given that a reference position X is in distress (undergoing an extreme adverse event), is represented by $CoVaR_{\alpha|\beta}(Y|X)$ ($CoES_{\alpha|\beta}(Y|X)$), where the distress event for X is defined as $X \geq VaR_\beta(X)$. Estimating and forecasting systemic risk measures such as CoVaR and CoES has become a critical problem as the need for systemic risk prevention in financial institutions becomes more widely recognised. Forecasting the above-mentioned systemic risk measures requires appropriate models for the marginal distribution of X and Y and their dependence structure. It involves developing models that accurately capture the individual behaviour of variables X

and Y and their dependence structure. Although correlation coefficients have been widely used to measure dependency, they fail to capture crucial tail behaviour of the joint probability distributions(see, e.g.,Embrechts (1999) and Embrechts et al. (2002)). Consequently, the literature has introduced a variety of modelling approaches to address these requirements (see Girardi and Ergün (2013)), Calabrese and Osmetti (2019), Bernardi and Catania (2019),Oh and Patton (2018),Reboredo and Ugolini (2015),Mainik and Schaanning (2014), Bernardi et al. (2017))
Given the significance of systemic risk measures, developing accurate statistical assessment tools for evaluating different models' predictive performance becomes crucial. Introducing such tools, known as backtesting in finance, is the core objective of this Chapter. Backtesting aims to address two main questions. The initial question is how well the forecasts of risk measures align with the observed losses in the sequence. Hence, backtesting is implemented with the aim of examining the accuracy of tail risk forecasts, which is similar to model validation in statistics. Fissler and Hoga (2023) and Fissler et al. (2015) call this backtesting approach as "traditional backtests". Escanciano and Olmo (2010) emphasises that the core of backtesting involves evaluating how actual outcomes align with risk measures generated by a model.
Due to the existence of multiple alternative risk forecasting models, the second primary question that backtesting tries to answer is which prediction model for risk measures demonstrates superior predictive accuracy relative to a set of competitors. Therefore, backtesting is utilised to make forecast comparisons and model ranking, and this is similar to the model selection procedure in statistics, known as comparative backtesting. In the light of the second question under comparative backtesting, Fissler and Hoga (2023) present a valuable framework for comparing risk forecasts by introducing the concept of multi-objective elicitability, which operates with bivariate scores $\boldsymbol{S}$ mapping to R^2 equipped with the lexicographic order. However, the lexicographic order approach employed in the One and a Half-Sided tests proposed by Fissler and Hoga (2023) imposes a constraint on the evaluation and comparison of risk measures. Specifically, the assessment is confined to either VaR or the systemic risk measure (SR) contingent upon the equality of VaR across the models under consideration. This indicates that Fissler and Hoga (2023) approach fails to assess both risk measures on an equal footing. To overcome this limitation, our contributions in this chapter are as follows: First, we propose an alternative approach by introducing a univariate score function that combines the marginal/conditional score functions for forecasting VaR and SR measure. By defining our score function as the summation of these individual score functions, we ensure an equitable assessment of both risk measures without the need to prioritize one over the other based on the equality of VaR across models, which goes beyond the backtesting implemented by Fissler and Hoga (2023). Moreover, we examine the risk of employing identification functions that are not strictly defined for backtesting by conducting a comparative analysis of our identification function along with the one-dimensional identification function

introduced by Banulescu-Radu et al. (2021) in addition to the two-dimensional identification function introduced by Fissler and Hoga (2023). Specifically, we evaluate the power of the three identification functions in identifying misspecified systemic risk forecasts under various sample sizes and distributional scenarios. A misspecified risk measure is defined as a measure that fails to capture the true tail risk dynamics over time accurately. Through our analysis, we demonstrate the superiority of our proposed identification function compared to Banulescu-Radu et al. (2021) identification function. In Particular, the backtest proposed by Banulescu-Radu et al. (2021) exhibits a complete loss of power in distinguishing between correct and misspecified forecasts. In contrast, our identification function, aligning with Fissler and Hoga (2023) identification function, successfully identifies misspecified forecasts almost with certainty (100%) across different distributional assumptions and sample sizes. These results underscore the power of our proposed identification function in detecting misspecified systemic risk forecasts, outperforming Banulescu-Radu et al. (2021) method. Furthermore, through a comprehensive simulation analysis, we illustrate the elicitability of our proposed score function. Moreover, we show through a Monte Carlo simulation the following: Firstly, the power of our tests increases significantly as the sample size (n) grows larger. Secondly, our tests show a good size control with a rejection frequency close to a 5% significance level, even for a small sample size ($n = 500$). Thirdly, detecting differences in the predictive ability of two models becomes easier when considering scenarios where both VaR and SR forecasts are accurate for one model while being misspecified for the other. Fourth, despite introducing only a small difference in the predictive accuracy of the VaR forecasts, our test has the power to detect this differentiation, even in cases where the systemic risk forecasts demonstrate a comparable performance across models. This finding highlighted the good sensitivity of our tests to detect marginal differences in model forecasting. Furthermore, across the various scenarios examined in this study, our tests generally demonstrate slightly higher power compared to those proposed by Fissler and Hoga (2023) in evaluations incorporating both (VaR + CoVaR) and (VaR + CoVaR + CoES). Finally, comparisons involving CoVaR and CoES exhibit higher power than those depending on CoVaR only, potentially attributed to the richer informational content offered by the CoES component, increasing the overall power of the analysis. The Chapter is constructed as follows: In Section (4.2), we introduce the systemic risk measures CoVaR and CoES. Section (4.3) introduces backtesting systemic risk measures with a review of Fissler and Hoga (2023) multi-objective score function and the alternative proposed univariate score function with also extending the univariate score functions to include weights. Furthermore, in (4.3.3.3), we focus on examining the risks of employing non-strict identification functions for backtesting systemic risk measures. Specifically, we conduct a comparative analysis of our identification function along with the one-dimensional identification function introduced by Banulescu-Radu et al. (2021) in addition to the two-dimensional identification

function introduced by Fissler and Hoga (2023). In addition, in (4.3.3.4), we investigate the elicitable property of our proposed score function, $S^{(\text{VaR+SR})}$ where SR represents the systemic risk, through a comprehensive simulation study. Monte Carlo simulations are illustrated in (4.4.1), where we evaluate the performance of the proposed backtesting procedure under a more realistic and dynamic setting. The generated data captures the potential heterogeneity and dynamic dependencies commonly observed in financial time series data. We apply our proposed test to the forecasts of (VaR, CoVaR) and (VaR, CoVaR, CoES) obtained from a bivariate GARCH(1,1) model with a t-copula for innovations with time varying correlation ρ_t follows the Generalized Autoregressive Score (GAS) model. We focus on one-step-ahead forecasts throughout our analysis. In (4.4.2), an empirical application focuses on the daily log losses of the DAX 30 index, using the daily log losses of the S&P 500 index as a reference variable. The study compares the systemic risk forecasts obtained from two different copula models: a benchmark Gaussian copula model and a t-copula model. In both copula models, the correlation parameter, which captures the dependence structure between the two indices, is modelled using a dynamic approach known as the generalized autoregressive score (GAS) framework, as proposed by Creal et al. (2013). Our findings indicate that the t-copula exhibits superior predictive performance, supported by p-values of 0.005 for (VaR,CoVaR) and 0.0204 for (VaR,CoVaR,CoES), aligning with its popularity in empirical studies. The Chapter concluded in Section (4.6).

4.2 Systemic Risk Measures CoVaR and CoES

The fragility of the financial system, highlighted by financial crises, underscores the importance of analyzing the interdependence among the assets or liabilities of various financial institutions in crisis analysis (see Embrechts et al. (2005), Kaas et al. (2009), Goovaerts et al. (2011), Asimit and Gerrard (2016), Gupta et al. (2021), Dhaene et al. (2022)). Classical risk measures, such as VaR and ES, were ineffective in assessing global contagion during several financial crises. Since VaR only focuses on a single institution, many scholars believe that this well-known classical risk metric cannot adequately represent the systemic nature of risk. According to Danielsson et al. (2011), VaR does not view an institution as a component of a complex system capable of generating new risks. Therefore, several researchers have proposed various empirical tools aimed at better measuring systemic risk, and numerous studies provide in-depth analysis of systemic risk and quantify it using measures such as conditional value-at-risk (CoVaR)[1], coexpected shortfall(CoES), and others. According to Tobias

[1]In some literature, conditional value-at-risk (CVaR), is defined as expected shortfall (ES) as discussed in this chapter (see Rockafellar and Uryasev (2002), Sarykalin et al. (2008), Noyan and Rudolf (2013)). However, in contrast to CVaR, conditional value at risk, or CoVaR for short, is presented in this chapter as a systemic risk measure following Tobias and Brunnermeier (2016) and Girardi and Ergün (2013).

and Brunnermeier (2016), $\text{CoVaR}_{i|j}$ is defined as the value at risk (VaR) of a financial institution i conditioning on another institution j being at its VaR exactly. Girardi and Ergün (2013) extend the definition of CoVaR introduced by Tobias and Brunnermeier (2016) by permitting the financial distress event for institution j to occur when its return falls at or below its VaR ($\text{R}_j \leq \text{VaR}_j$), instead of exactly at its VaR ($\text{R}_j = \text{VaR}_j$). While CoVaR is formulated based on conditional probability and the VaR of the system, CoES focuses on conditional expectations, emphasizing the tail of the loss distribution and the expected shortfall (see Tobias and Brunnermeier (2016) and Mainik and Schaanning (2014)). In particular, the tail of a distribution represents extreme events, and CoES measures the average of losses beyond a certain threshold when an extreme event happens. We now formally define the two systemic risk measures employed in the chapter.

Definition (CoVaR and CoES)

Given significance levels $\alpha \in (0,1), \beta \in [0,1)$,
(i) CoVaR is defined by following Girardi and Tolga Ergün (2013) and Banulescu-Radu et al. (2021) as:

$$\text{CoVaR}_{\alpha|\beta}(Y \mid X) := \text{CoVaR}_{\alpha|\beta}\left(F_{X,Y}\right) := \text{VaR}_{\alpha}\left(F_{Y|X\geq \text{VaR}_{\beta}(X)}\right), \tag{4.1}$$

here $F_{Y|X\geq \text{VaR}_{\beta}(X)} = \text{P}\left\{Y \leq \cdot \mid X \geq \text{VaR}_{\beta}(X)\right\}$. Throughout this chapter, we follow the sign convention used by Fissler and Hoga (2023), where positive values represent losses and negative values represent gains for both the position of interest (Y) and the reference position (X). For special cases, include $\beta = 0$, it simplifies to $\text{CoVaR}_{\alpha|0}(Y \mid X) := \text{VaR}_{\alpha}(Y)$, and if $\beta = \alpha$ it simplifies to $\text{CoVaR}_{\alpha}(Y \mid X) = \text{CoVaR}_{\alpha|\alpha}(Y \mid X)$. This equation denotes the VaR_{α} of the distribution of Y conditional on X being greater than or equal to the VaR at a specific level β.

(ii) CoES is defined as:

$$\text{CoES}_{\alpha|\beta}(Y \mid X) := \text{CoES}_{\alpha|\beta}\left(F_{X,Y}\right) := \frac{1}{1-\alpha}\int_{\alpha}^{1} \text{CoVaR}_{\gamma|\beta}(Y \mid X)\text{d}\gamma. \tag{4.2}$$

For continuous $F_{X,Y}$, CoES simplifies to the conditional expectation of Y given that Y is greater than or equal to the conditional CoVaR and X is greater than or equal to its β-quantile: $\text{CoES}_{\alpha|\beta}(Y \mid X) = \text{E}\left[Y \mid Y \geq \text{CoVaR}_{\alpha|\beta}(Y \mid X), X \geq \text{VaR}_{\beta}(X)\right]$. Special cases include $\text{CoES}_{\alpha|0}(Y \mid X) = \text{ES}_{\alpha}(Y)$, where $\beta = 0$, and $\text{CoES}\alpha(Y \mid X) = \text{CoES}\alpha \mid \alpha(Y \mid X)$ if $\beta = \alpha$.

Remark. The definition of CoVaR in Eq(4.1) is proposed by Girardi and Ergün (2013) and Banulescu-Radu et al. (2021). Tobias and Brunnermeier (2016) defines CoVaR as CoVaR_{β}, representing the β-quantile of the conditional distribution function

$F_Y\left(\cdot \mid X = \text{VaR}_\beta(X)\right) = \text{P}\left\{Y \leq \cdot \mid X = \text{VaR}_\beta(X)\right\}$. This definition captures the idea of measuring the risk of Y under the specific condition that X is precisely at its-$\beta-$quantile. However, Tobias and Brunnermeier (2016) definition has some limitations. First, if F_X is continuous, the probability associated with the event $\{X = VaR_\beta(X)\}$ can be zero because the probability of any specific point is zero in a continuous random variable. Conditioning on a specific point is challenging in such continuous distributions where the probability of a single point is zero. Secondly, it fails to completely capture the tail risk of F_X.

4.3 Backtesting Systemic Risk Measures

Considering the significance of systemic risk measures, it becomes crucial to develop accurate statistical evaluations for evaluating the accuracy of different models' forecasting, such tools known in finance as backtests. Two primary goals are considered for backtesting. First, we aim to measure/test if the models' forecasts are accurate, this process is similar to model validation in statistics. Fissler and Hoga (2023) and Fissler et al. (2015) call this backtesting approach as "traditional backtests". Jorion (2007) defined the backtesting approach aligns with this goal as a formal statistical tool that compares actual losses (realized outcomes) with the predicted loss values generated by risk prediction models. Given that the actual value of the risk measure cannot be directly observed, making these comparisons often relies on violations. For example, in the context of VaR, we consider a violation when the realized portfolio return falls below the forecasted VaR. Escanciano and Olmo (2010) emphasize that the core of backtesting involves evaluating how actual outcomes align with risk measures generated by a model. Hence, the natural question this backtesting deals with is deciding how well the forecasts of risk measures align with the observed losses in the sequence.
Due to the existence of multiple alternative risk prediction models, the second primary purpose of backtesting lies in the necessity of forecast comparison and ranking, and this similar to model selection procedure in statistics. This approach of backtesting is known as comparative backtesting. Here, we deal with the question of which prediction model for risk measures demonstrates superior predictive accuracy relative to the others. In the following subsection, we introduce basic definitions of properties that are important for ranking the forecasts and forecast validation.

4.3.1 Basic Definitions

We introduce foundational concepts that drawing from the decision-theoretic framework illustrated and discussed by Osband (1985), Lambert et al. (2008), Gneiting (2011), Emmer et al. (2015), as well as Fissler and Ziegel (2016, 2019) and Fissler and

Hoga (2023). The set $\mathcal{A}$, a subset of $\mathbb{R}^k$, represents the entire domain of potential forecasts or predictions. The observation domain is specified as $\mathbb{R}^d$. A functional T is considered, which takes a class of distributions $\mathcal{F}$ (a subset of the distributions on $\mathbb{R}^d$) as input and produces an output in $\mathcal{A}$. For simplicity, the discussion is restricted to single-valued functionals.

Definition of Integrability. A function $a : \mathbb{R}^d \to \mathbb{R}$ is said to be $\mathcal{F}$-integrable if the integral of the absolute value of a with respect to any distribution F in $\mathcal{F}$ is finite. For such a function a, an induced map $\bar{a} : \mathcal{F} \to \mathbb{R}$ can be defined as $\bar{a}(F) = \int a(y)\mathrm{d}F(y)$. Similarly, a function $g : \mathcal{A} \times \mathbb{R}^d \to \mathbb{R}$ is called $\mathcal{F}$-integrable if $g(r, \cdot)$ is $\mathcal{F}$-integrable for all r in $\mathcal{A}$.

Definition of Consistent Scoring Function. An $\mathcal{F}$-integrable function $S : \mathcal{A} \times \mathbb{R}^d \to \mathbb{R}$ is termed an $\mathcal{F}$-consistent scoring function for T if the following condition holds: For any distribution F in $\mathcal{F}$ and any r in $\mathcal{A}$, the expected value of $S(T(F), F)$ with respect to F is less than or equal to the expected value of $S(r, F)$ with respect to F. If, additionally, equality holds only when $T(F) = r$, then S is called a strictly $\mathcal{F}$-consistent scoring function for T.

Definition of Elicitability. The functional T is said to be elicitable on $\mathcal{F}$ if there exists a strictly $\mathcal{F}$-consistent scoring function for T.

Definition of Identification Function An $\mathcal{F}$-integrable function $V : \mathcal{A} \times \mathbb{R}^d \to \mathbb{R}^m$ is called an $\mathcal{F}$-identification function for T if the expected value of $V(T(F), F)$ with respect to F is equal to the zero vector for all F in $\mathcal{F}$. If, furthermore, for any F in $\mathcal{F}$ and any r in $\mathcal{A}$, the expected value of $V(r, F)$ with respect to F being equal to the zero vector implies that $T(F) = r$, then V is termed a strict $\mathcal{F}$-identification function for T.

Definition of Identifiability. The functional T is identifiable on $\mathcal{F}$ if there exists a strict $\mathcal{F}$-identification function for T.

Definition of Conditional Elicitability by Fissler and Hoga (2023). Let us consider two functionals, T_1 and T_2, where T_1 maps from a class of distributions $\mathcal{F}$ to an action domain $\mathcal{A}_1$, and T_2 maps from $\mathcal{F}$ to another action domain $\mathcal{A}_2$. The functional T_2 is said to be conditionally elicitable with respect to T_1 on $\mathcal{F}$ if the following conditions hold:

1- The functional T_1 is elicitable on $\mathcal{F}$.

2- For any specific value r_1 in $\mathcal{A}_1$, the functional T_2 is elicitable on the subset of distributions $\mathcal{F}_{r_1} \subset \mathcal{F}$, where $\mathcal{F}_{r_1}$ is defined as the set of distributions F in $\mathcal{F}$ such that $T_1(F) = r_1$.

Definition of Conditional Identifiabilit by Fissler and Hoga (2023). The functional T_2 is said to be conditionally identifiable with respect to T_1 on $\mathcal{F}$ if the following conditions hold:

1- The functional T_1 is identifiable on $\mathcal{F}$.

2- For any specific value r_1 in $\mathcal{A}_1$, the functional T_2 is identifiable on the subset of distributions $\mathcal{F}_{r_1} \subset \mathcal{F}$, where $\mathcal{F}_{r_1}$ is defined as the set of distributions F in $\mathcal{F}$ such that $T_1(F) = r_1$.

For the purpose of comparative backtesting of systemic risk measures, the Diebold and Mariano (1995) (DM) test is a widely used statistical test in econometrics and finance in forecast comparisons. This statistical test uses formal hypothesis tests to address sampling uncertainty in forecast comparisons. In particular, the DM test compares the expected loss differential between two competing forecasts. It relies on scalar scoring functions, which provide a single numerical value summarizing the forecast errors. A significant result reflecting the power of the DM test in assessing different models' accuracy comes from West (1996). This study shows the asymptotic validity of DM when applied to forecast errors based on estimated residuals. This allows for a more practical application of the DM test in situations where true residuals cannot be directly observed and must be estimated from the available data. In the following subsection, we will introduce Fissler and Hoga (2023) methodology that adjusts the DM test to be implemented in the multi-objective score function.

4.3.2 Review of Multi-objective Score Function Backtesting

The DM tests introduce formal hypothesis tests for addressing the challenges associated with sampling uncertainty in forecast comparisons. These tests are frequently employed in empirical forecast evaluations and continue to be explored in theoretical research. Following the principle of strict consistency, DM tests have traditionally relied on scalar scoring functions. Fissler and Hoga (2023) methodology adjusts the DM tests to be implemented for the proposed multi-objective elicitability scores equipped with the lexicographic order to compare systemic risk forecasts. Utilising the lexicographic order on R^2 enables the comparison of all forecasts, including those that may be misspecified; throughout the chapter, the term a "misspecified risk measure" is a measure that is not able to capture the true tail risk dynamically over time. Fissler and Hoga (2023) introduce and establish the concept of conditional identifiability and conditional elicitability of $\text{CoVaR}_{\alpha|\beta}(Y|X)$ and $(\text{CoVaR}_{\alpha|\beta}(Y|X), \text{CoES}_{\alpha|\beta}(Y|X))\text{withVaR}_{beta}(X)$, respectively.

The foundation of Fissler and Hoga (2023) methodology lies in the derivation of multi-objective scores, denoted as $\boldsymbol{S} = (S_1, S_2)'$, through Theorem 4.2 in their paper which will be represented in the below subsection.

4.3.2.1 Definition of Fissler and Hoga (2023) Multi-Objective Scores and Identification Functions

We introduce the multi-objective scores proposed by Fissler and Hoga (2023), which are based on the lexicographic order $\preceq_{\text{lex}}$ on $\mathbb{R}^2$ as outlined in their Theorem 4.2. These scores consist of three main components:

1- **VaR Score** Defined as:

$$\begin{aligned} S^{\text{VaR}}(v,(x,y)) =&(\mathbb{1}\{x \leq v\} - \beta)h(v) \\ &- \mathbb{1}\{x \leq v\}h(x) + a^{\text{VaR}}(x,y) \end{aligned} \tag{4.3}$$

where, $\mathbb{1}x \leq v$ is the indicator function, $h : \mathbb{R} \to \mathbb{R}$ is strictly increasing, and $a^{\text{VaR}}(x,y)$ is a term chosen based on integrability and sign considerations.
2- **(VaR, CoVaR) Score.** Represented as:

$$\begin{aligned} \boldsymbol{S}^{(\text{VaR,CoVaR})}((v,c),(x,y)) &= \begin{pmatrix} S^{\text{VaR}}(v,(x,y)) \\ S_v^{\text{CoVaR}}(c,(x,y)) \end{pmatrix}, \\ S_v^{\text{CoVaR}}(c,(x,y)) &= \mathbb{1}\{x > v\}[(\mathbb{1}\{y \leq c\} - \alpha)g(c) \\ &\quad -\mathbb{1}\{y \leq c\}g(y) + a(y)] \\ &\quad +a^{\text{CoVaR}}(x,y), \end{aligned} \tag{4.4}$$

where $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing and S_v^{CoVaR} is $\mathcal{F}$ integrable for all $v \in \mathbb{R}$.
3- **(VaR, CoVaR, CoES) Score.** Given by:

$$\begin{aligned} \boldsymbol{S}^{(\text{VaR,CoVaR,CoES})}((v,c,e),(x,y)) &= \begin{pmatrix} S^{\text{VaR}}(v,(x,y)) \\ S_v^{(\text{CoVaR,CoES})}((c,e),(x,y)) \end{pmatrix}, \\ S_v^{(\text{CoVaR,CoES})}((c,e),(x,y)) \quad &= \mathbb{1}\{x > v\}\Big[(\mathbb{1}\{y \leq c\} - \alpha)g(c) - \mathbb{1}\{y \leq c\}g(y) \\ &+ \phi'(e)\left(e - \frac{1}{1-\alpha}(y\mathbb{1}\{y > c\} + c(\mathbb{1}\{y \leq c\} - \alpha))\right) \\ &- \phi(e) + a(y)\Big] + a^{\text{CoES}}(x,y), \end{aligned} \tag{4.5}$$

Where $\phi : \mathbb{R} \to \mathbb{R}$ is strictly convex with a subgradient $\phi' < 0$.
The functions $a^{\text{VaR}}(x,y)$, $a^{\text{CoVaR}}(x,y)$, and $a^{\text{CoES}}(x,y)$ are selected based on the integrability of the scoring functions and their respective signs.

Moreover, we present the identification functions proposed by Fissler and Hoga (2023) as illustrated in Theorem S.3.1. in their supplementary materials.

1- **VaR Strict Identification Function.** Defined as:

$$V^{(\text{VaR})}(v,(x,y)) = \mathbb{1}\{x \leq v\} - \beta. \tag{4.6}$$

2- **(VaR,CoVaR) Strict Identification Function.** Given by:

$$\boldsymbol{V}^{(\text{VaR,CoVaR})}((v,c),(x,y)) = \begin{pmatrix} \mathbb{1}\{x \leq v\} - \beta \\ \mathbb{1}\{x > v\}[\mathbb{1}\{y \leq c\} - \alpha] \end{pmatrix}. \tag{4.7}$$

3- **(VaR, CoVaR, CoES) Strict Identification Function.** Represented as:

$$\boldsymbol{V}^{(\text{VaR,CoVaR,CoES})}((v,c,e),(x,y)) = \begin{pmatrix} \mathbb{1}\{x \leq v\} - \beta \\ \mathbb{1}\{x > v\}[\mathbb{1}\{y \leq c\} - \alpha] \\ \mathbb{1}\{x > v\}\left[e - \frac{1}{1-\alpha}(y\mathbb{1}\{y > c\} + c(\mathbb{1}\{y \leq c\} - \alpha))\right] \end{pmatrix}. \tag{4.8}$$

In the following subsection, we present the implementation of comparative backtests with the multi-objective scores as proposed by Fissler and Hoga (2023).

4.3.2.2 Diebold–Mariano tests for multi-objective scores

Two-Sided Tests

Consider two sequences of forecasts $\{\boldsymbol{r}_{1,t}\}_{t=1,\dots,n}$ and $\{\boldsymbol{r}_{2,t}\}_{t=1,\dots,n}$ as the series of generated by the models and $\boldsymbol{S} = (S_1, S_2)'$ denote one of the multi-objective scores provided earlier. For instance, if $\boldsymbol{S} = \boldsymbol{S}^{(\text{VaR,CoVaR,CoES})}$, then $\boldsymbol{r}_{i,t} = (\widehat{\text{VaR}}_{i,t}, \widehat{\text{CoVaR}}_{i,t}, \widehat{\text{CoES}}_{i,t})$ for $i = 1, 2$. These scores are used to define bivariate score differences $\boldsymbol{d}_t = (d_{1t}, d_{2t})' = \boldsymbol{S}(\boldsymbol{r}_{1,t}, (X_t, Y_t)) - \boldsymbol{S}(\boldsymbol{r}_{2,t}, (X_t, Y_t))$ at each time t, capturing the forecast differences between the competing models, where $(X_t, Y_t)_{t=1,\dots,n}$ represent the verifying observations. Fissler and Hoga (2023) propose a two-sided test to assess whether both forecast models predict equally well on average. The null hypothesis is given by $H_0^{=} : \mathrm{E}\left[\bar{\boldsymbol{d}}_n\right] = \boldsymbol{0}$ for all $n = 1, 2, \dots$, where $\bar{\boldsymbol{d}}_n := (\bar{d}_{1n}, \bar{d}_{2n})' := \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{d}_t$. The test statistic for the two-sided test is defined as:

$$\mathcal{T}_n = n\bar{\boldsymbol{d}}_n' \widehat{\boldsymbol{\Omega}}_n^{-1} \bar{\boldsymbol{d}}_n,$$

where $\widehat{\boldsymbol{\Omega}}_n$ is a consistent estimator of the variance-covariance matrix of $\sqrt{n}\bar{\boldsymbol{d}}_n$ under the null hypothesis. Under suitable regularity conditions, $\mathcal{T}_n$ converges in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$.

One and a Half-Sided Tests

Fissler and Hoga (2023) introduce the One and a Half-sided tests to assess the superiority of risk forecasts $\{r_{2,t}\}_{t=1,\dots,n}$ over benchmark forecasts $\{r_{1,t}\}_{t=1,\dots,n}$. The null hypothesis under consideration takes the form:

$$H_0^{\preceq_{\text{lex}}} : \mathrm{E}\left[\bar{d}_{1n}\right] = 0 \text{ and } \mathrm{E}\left[\bar{d}_{2n}\right] \leq 0 \text{ for all } n = 1, 2, \dots. \tag{4.9}$$

which states that the expected average score difference for the first component (VaR) is zero, and the expected average score difference for the second component (e.g., CoVaR or CoES) is less than or equal to zero.

To test the null hypothesis $H_0^{\preceq_{\text{lex}}}$, Fissler and Hoga (2023) proposed the the following test statistic:

$$\mathcal{T}_n^{\text{OS}} = n \left(\bar{d}_{1n}, \max\left\{\bar{d}_{2n}, \left(\widehat{\sigma}_{12,n}/\widehat{\sigma}_{11,n}\right)\bar{d}_{1n}\right\}\right) \widehat{\mathbf{\Omega}}_n^{-1} \begin{pmatrix} \bar{d}_{1n} \\ \max\left\{\bar{d}_{2n}, \left(\widehat{\sigma}_{12,n}/\widehat{\sigma}_{11,n}\right)\bar{d}_{1n}\right\} \end{pmatrix}.$$

Here, $\widehat{\mathbf{\Omega}}_n$ is an estimator of the variance-covariance matrix of the average score differences, defined as:

$$\begin{aligned} \widehat{\mathbf{\Omega}}_n = \begin{pmatrix} \widehat{\sigma}_{11,n} & \widehat{\sigma}_{12,n} \\ \hat{\sigma}_{12,n} & \widehat{\sigma}_{22,n} \end{pmatrix} &= \frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{d}_t - \overline{\boldsymbol{d}}_n\right)\left(\boldsymbol{d}_t - \overline{\boldsymbol{d}}_n\right)' \\ &+ \frac{1}{n}\sum_{h=1}^{m_n} w_{n,h} \sum_{t=h+1}^{n}\left[\left(\boldsymbol{d}_t - \overline{\boldsymbol{d}}_n\right)\left(\boldsymbol{d}_{t-h} - \overline{\boldsymbol{d}}_n\right)' + \left(\boldsymbol{d}_{t-h} - \overline{\boldsymbol{d}}_n\right)\left(\boldsymbol{d}_t - \overline{\boldsymbol{d}}_n\right)'\right], \end{aligned} \tag{4.10}$$

where, m_n represents a sequence of integers that tends to infinity as n increases and $w_{n,h}$ is a uniformly bounded scalar triangular array, with $w_{n,h}$ approaching 1 as n increases for all $h = 1, \ldots, m_n$.

Although the One and a Half-Sided tests provide a useful framework for comparing risk forecasts through multi-objective scores, their methodology suffers from a significant limitation. The use of a lexicographic order operates under the assumption that the expected value of the first risk measure, VaR, denoted as $\mathrm{E}[\bar{d}_{1n}]$, is consistent across the comparative forecasts. This constraint implies that the evaluation and comparison of the risk measures cannot be conducted on an equal footing. Instead, the assessment is restricted to either VaR or SR contingent upon the equality of VaR across the models under consideration.
To address this limitation, we propose an alternative approach that goes beyond the tests implemented by Fissler and Hoga (2023). Our approach aims to assess both risk measures, VaR and SR, on an equitable basis unlike the lexicographic order method, which evaluates either VaR or SR depending on the equality of VaR across models. In the following subsection, we introduce our proposed alternative approach.

4.3.3 Alternative Univariate Backtesting Systemic Risk Forecasts

In response to the limitation inherent in the Diebold-Mariano (DM) tests utilising two-dimensional multi-objective scores proposed by Fissler and Hoga (2023), in particular focusing on systemic risk forecasts with the assumption of the equality of VaRs forecasts, we propose an alternative score function that is designed to address a

more comprehensive evaluation of risk models by considering both risk measures forecasting, and it is one-dimensional score function. Our proposed score function for two risk measures $S(r_1, r_2, y)$ is defined as the sum of two individual score functions:

$$S(r_1, r_2, y) = S_1(r_1, y) + S_{2,r_1}(r_2, y), \tag{4.11}$$

Here, $S_1(r_1, y)$ represents the marginal score function for the first risk measure, and $S_{2,r_1}(r_2, y)$ represents the conditional score function for the second risk measure given the realisation of the first risk measure r_1. This score function is designed to provide a joint assessment of forecasts for both risk measures, considering both individual forecasts and their relationship with each other.
Most existing studies in the literature focus on evaluating either VaR or SR measures individually. In contrast, our approach introduces a novel univariate score function, $S^{(\text{VaR+SR})}$, that combines VaR and systemic risk measures for a comprehensive evaluation of risk forecasts. By incorporating both VaR and systemic risk measures into a single score function, we aimed to provide a more holistic view of risk assessment in the financial sector. While the existing literature primarily concentrates on individual risk measures, our work explores the potential benefits of jointly assessing VaR and systemic risk measures. We believe that this joint assessment framework offers a unique perspective on risk forecasting and highlights the importance of considering multiple risk criteria collectively in risk management practices. Our objectives and methodological approach differ from studies focusing solely on individual risk measures, and we believe that our work contributes to the literature by proposing this novel framework for joint risk assessment, which has received limited attention in the existing literature. The novelty of our approach suggests the potential for further research in this direction, as the integration of multiple risk criteria is an area that warrants further exploration and investigation.

After introducing the general score function $S(r_1, r_2, y)$, we now introduce our score function $S^{(\text{VaR+CoVaR})}((v, c), (x, y))$ for the joint assessment of forecasts for two risk measures: VaR and CoVaR, which will be defined as follows:

$$S^{(\text{VaR+CoVaR})}((v, c), (x, y)) = S^{\text{VaR}}(v, (x, y)) + S_v^{\text{CoVaR}}(c, (x, y)) \tag{4.12}$$

Here, v represents VaR forecasts, and c represents CoVaR forecasts. The first component $S^{\text{VaR}}(v, (x, y))$ represents the scoring function associated with the first risk measure VaR. This scoring function independently assesses the accuracy of the VaR forecast. The second component $S_v^{\text{CoVaR}}(c, (x, y))$ is associated with the second risk measure component CoVaR given the realisation of VaR, where the marginal scores $S^{\text{VaR}}(v, (x, y))$ and $S_v^{\text{CoVaR}}(c, (x, y))$ are defined in (4.3) and (4.4).
As illustrated from (4.12), our proposed univariate score function is formulated to ensure that both risk measures, VaR and CoVaR, are assessed concurrently and on an equal footing without the need to prioritize one over the other based on the equality of

VaR across models. By avoiding the restrictive assumption of equal expected values for the first risk measure, a prerequisite for the lexicographic order approach employed in the One and a Half-Sided tests by Fissler and Hoga (2023), our joint score function facilitates a more holistic and balanced evaluation of risk forecasts.
After introducing our consistent scoring function $S^{(\text{VaR+CoVaR})}$, we derive our identification function $V^{(\text{VaR+CoVaR})}((v,c),(x,y))$ as the summation of the identification functions for VaR and CoVaR, respectively. Specifically, the identification function is given by:

$$V^{(\text{VaR+CoVaR})}((v,c),(x,y)) = \left(\frac{\partial S^{\text{VaR}}(v,(x,y))}{\partial v}\right) + \left(\frac{\partial S_v^{\text{CoVaR}}(c,(x,y))}{\partial c}\right) \tag{4.13}$$

Where the first term, $\frac{\partial S^{\text{VaR}}(v,(x,y))}{\partial v}$, represents the identification function for VaR, and the second term, $\frac{\partial S_v^{\text{CoVaR}}(c,(x,y))}{\partial c}$, represents the identification function for CoVaR. Which leads to :

$$V^{(\text{VaR+CoVaR})}((v,c),(x,y)) = \Big(\mathbb{1}x \leq v - \beta + \mathbb{1}x > v\,[\mathbb{1}y \leq c - \alpha] \Big) \tag{4.14}$$

In (4.3.3.3), we evaluate the power of our proposed identification function, $V^{(\text{VaR+CoVaR})}$, in identifying misspecified systemic risk forecasts under various sample sizes and distributional scenarios. Specifically, we conduct a comparative analysis of our identification function $V^{(\text{VaR+CoVaR})}$ along with the one-dimensional identification function introduced by Banulescu-Radu et al. (2021) in addition to the two-dimensional identification function introduced by $\boldsymbol{V}^{(\text{VaR,CoVaR})}$. Moreover, in (4.3.3.4) we prove through a numerical simulation analysis that our score function $S^{(\text{VaR+CoVaR})}$ is elicitable under various simulation scenarios.

4.3.3.1 Comparing Competing Risk Models through Univariate Scoring Function

To compare the forecasts of two competing risk models, we analyse the sequences of forecasts $\{\boldsymbol{r}_{1,t}\}_{t=1,\dots,n}$ and $\{\boldsymbol{r}_{2,t}\}_{t=1,\dots,n}$. We consider the score function $S^{(\text{VaR+CoVaR})}$, which evaluates the performance of these forecasts, each forecast consists of $\boldsymbol{r}_{i,t} = (\widehat{\text{VaR}}_{i,t}, \widehat{\text{CoVaR}}_{i,t})$ for $i = 1, 2$. The verifying observations are denoted as $\{(\boldsymbol{X}_t, \boldsymbol{Y}_t)\}_{t=1,\dots,n}$. The score function $S^{(\text{VaR+CoVaR})}$ is defined as follows:

$$S^{(\text{VaR+CoVaR})}(\boldsymbol{r}_{i,t},(x,y)) = S^{\text{VaR}}(\widehat{\text{VaR}}_{i,t},(x,y)) + S^{\text{CoVaR}}(\widehat{\text{CoVaR}}_{i,t},(x,y)) \tag{4.15}$$

To compare the two forecasts at time t, we compute the univariate score difference d_t^{uni} between the two sequences:

$$d_t^{uni} = S^{(\text{VaR+CoVaR})}(\boldsymbol{r}_{1,t},(x,y)) - S^{(\text{VaR+CoVaR})}(\boldsymbol{r}_{2,t},(x,y)) \tag{4.16}$$

Subsequently, we calculate the test statistic for a one-sided test for univariate loss differences d_t^{uni} where the sample variance of a univariate time series is computed as:

$$\widehat{\sigma}_n = \frac{\sum_{t=1}^{n}(d_t^{uni} - \bar{d}^{uni})^2}{n}. \tag{4.17}$$

We defined the test Statistic $\mathcal{T}_n^{OS.uni}$ as follows:

$$\mathcal{T}_n^{OS.uni} = \frac{\sqrt{n}}{\sqrt{\widehat{\sigma}_n}} \cdot \bar{d}^{uni} \tag{4.18}$$

The hypothesis we are testing,

$$H_0 : \mathrm{E}\left[\bar{d}_n^{uni}\right] \leq 0 \quad \text{vs} \quad H_a : \mathrm{E}\left[\bar{d}_n^{uni}\right] > 0.$$

Under the null hypothesis $H_0 : \mathrm{E}[\bar{d}_n^{uni}] \leq 0$, to properly interpret the test statistic $\mathcal{T}_n^{OS.uni}$ and compute the p-value, we need to specify the asymptotic distribution of the test statistic under the equality condition, i.e., when $\mathrm{E}[\bar{d}_n^{uni}] = 0$. Specifically, under the null hypothesis of equality $H_0 : \mathrm{E}[\bar{d}_n^{uni}] = 0$, the test statistic $\mathcal{T}_n^{OS.uni}$ converges in distribution to a standard normal asymptotically:

$$\mathcal{T}_n^{OS.uni} \xrightarrow{d} N(0,1)$$

Therefore, to test the one-sided alternative $H_a : \mathrm{E}[\bar{d}_n^{uni}] > 0$, we can apply a standard one-sided z-test (or equivalently, a one-sided t-test) and reject the null hypothesis $H_0 : \mathrm{E}[\bar{d}_n^{uni}] \leq 0$ in the right tail when the observed value of $\mathcal{T}_n^{OS.uni}$ is sufficiently large and positive. A small p-value supports the conclusion that the first model is statistically significantly worse in terms of mean loss compared to the second model.

4.3.3.2 Extension of the Proposed Univariate Score function

We are extending our initial proposed score function to include three risk measures by introducing the CoES risk measure. Following a similar framework, we denote the score of three risk measures as $S^{(\mathrm{VaR,CoVaR,CoES})}$. Specifically, we define the score function for three risk measures as

$$S^{(\mathrm{VaR+CoVaR+CoES})}((v,c,e),(x,y)) = S^{\mathrm{VaR}}(v,(x,y)) + S_v^{(\mathrm{CoVaR,CoES})}((c,e),(x,y)) \tag{4.19}$$

Where $S_v^{(\mathrm{CoVaR,CoES})}((c,e),(x,y))$ defined in (4.5). The advantage of incorporating CoES into the score function lies in its ability to provide insights beyond the VaR and

CoVaR framework. CoES represents the expected shortfall conditioned on the joint occurrence of extreme events captured by VaR and CoVaR, and it will lead to a more comprehensive risk assessment. This extension makes the score function capable of distinguishing between models based not only on individual risk measures (VaR) and conditional risk (CoVaR) but also on their ability to capture joint dynamics in extreme events (CoES). This allows for a more detailed assessment of model performance and ranking.

We also could extend the score functions to include weights. For instance, the expression defined in (4.12) after including the weights will be as follows:

$$S^{(\text{VaR+CoVaR})}((v,c),(x,y)) = \gamma \cdot S^{\text{VaR}}(v,(x,y)) + \delta \cdot S_v^{\text{CoVaR}}(c,(x,y)) \tag{4.20}$$

Here, we introduce the weights for the VaR and CoVaR components as γ and $1-\gamma$, respectively. We can adjust these weights according to the desired emphasis on each risk measure. If we seek to prioritize VaR, we increase the value of γ relative to $1-\gamma$. Conversely, we increase the value of $1-\gamma$ if we want to emphasize the systemic risk component (CoVaR).

Our investigation continues by next examining the risk of employing identification functions that are not strictly defined for backtesting systemic risk measures by conducting a comparative analysis of our identification function $V^{(\text{VaR+CoVaR})}$ along with the one-dimensional identification function introduced by Banulescu-Radu et al. (2021) in addition to the two-dimensional identification function introduced by $\boldsymbol{V}^{(\text{VaR,CoVaR})}$.

4.3.3.3 Comparison of Identification Functions for Backtesting Systemic Risk Measures

The main objective of this analysis is to investigate the performance of identification functions in backtesting systemic risk measures, specifically VaR and CoVaR. To illustrate this, we evaluate the power of three identification functions in detecting misspecified forecasts. We consider our proposed identification function $V^{(\text{VaR+CoVaR})}$, defined in equation (4.14), the strict identification function $\boldsymbol{V}^{(\text{VaR,CoVaR})}$ proposed by Fissler and Hoga (2023), given in equation (4.7), and the one-dimensional identification function V^{BR} proposed by Banulescu-Radu et al. (2021), defined in equation (4.21) in the next Remark 4.3.3.3.1.

Remark 4.3.3.3.1. Banulescu-Radu et al. (2021) propose in their equation (4) a one dimensional identification function V^{BR} for the pair $\Big(\text{VaR}_\beta(X), \text{CoVaR}_{\alpha|\beta}(Y \mid X)\Big)$. The function V^{BR} maps pairs of real numbers (v,c) and (x,y) to a real number and is

defined as,

$$V^{BR}((v,c),(x,y)) = \mathbb{1}\{x > v\}\mathbb{1}\{y > c\} - (1-\alpha)(1-\beta). \tag{4.21}$$

A simulation study is conducted where we generate bivariate normal random variables (X_t, Y_t) with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$. Our objective is to forecast VaR_β of X_t and $\text{CoVaR}_{\alpha|\beta}$ of Y_t given X_t at confidence levels $\alpha = \beta = 0.95$. We consider two sets of forecasts: accurately specified forecasts for these risk measures computed and denoted as $(\widehat{\text{VaR}}_{\beta,t}, \widehat{\text{CoVaR}}_{\alpha|\beta,t})$, and intentionally misspecified forecasts $(\widehat{\text{VaR}}_{\beta',t}, \widehat{\text{CoVaR}}_{\alpha'|\beta',t})$ with different confidence levels $(\alpha', \beta') = (0.75, 0.99)$.

The correctly specified forecasts are $\widehat{\text{VaR}}_{\beta,t} = \text{VaR}_{\beta,t} \approx 1.64$ and $\widehat{\text{CoVaR}}_{\alpha|\beta,t} = \text{CoVaR}_{\alpha|\beta,t} \approx 3.23$, while the misspecified forecasts are $(\widehat{\text{VaR}}_{\beta',t}, \widehat{\text{CoVaR}}_{\alpha'|\beta',t}) \approx (2.33, 2.23)$. It is important to note that in this scenario, we satisfy the condition $(1-\alpha')(1-\beta') = (1-\alpha)(1-\beta)$. Under this condition, as illustrated in Fissler and Hoga (2023) that the identification function V^{BR} of Banulescu-Radu et al. (2021) cannot uniquely identify the true VaR and CoVaR values from the data. Specifically, V^{BR} satisfies:

$$\bar{V}\left(\left(\text{VaR}_\beta(X), \text{CoVaR}_{\alpha|\beta}(X \mid Y)\right), F_{X,Y}\right) = 0$$

However, it also satisfies:

$$\bar{V}\left(\left(\text{VaR}_{\beta'}(X), \text{CoVaR}_{\alpha'|\beta'}(X \mid Y)\right), F_{X,Y}\right) = 0$$

as long as $(1-\alpha')(1-\beta') = (1-\alpha)(1-\beta)$.

The analysis examines the null hypothesis of correct unconditional calibration for each identification function.

Definition 4.3.3.3.1. A sequence of predictions $\{r_t\}_{t\in\mathbb{N}}$ is considered calibrated for a parameter Θ on average if

$$E[V(r_t, (X_t, Y_t))] = 0 \quad \forall\, t \in \mathbb{N}. \tag{4.22}$$

For $\boldsymbol{V}^{(\text{VaR,CoVaR})}$, the null hypothesis of correct unconditional calibration is:

$$H_0 : \text{E}\left[\boldsymbol{V}^{(\text{VaR,CoVaR})}\left((\widehat{\text{VaR}}_t, \widehat{\text{CoVaR}}_t), (X_t, Y_t)\right)\right] = \mathbf{0} \quad \text{for all } t = 1, 2, \ldots$$

However, for the non-strict V^{BR}, the null hypothesis effectively becomes:

$$H_0 : \mathrm{E}\left[V^{BR}\left((\widehat{\mathrm{VaR}}_t, \widehat{\mathrm{CoVaR}}_t), (X_t, Y_t)\right)\right] = 0 \quad \text{for all } t = 1, 2, \ldots$$

While for our identification function $V^{(\mathrm{VaR+CoVaR})}$,
$H_0 : E[V^{(\mathrm{VaR+CoVaR})}((\widehat{\mathrm{VaR}}_t, \widehat{\mathrm{CoVaR}}_t), (X_t, Y_t))] = 0 \quad \forall\, t = 1, 2, \ldots$

To test for the null hypothesis of correct unconditional calibration of the three identification functions, we follow Nolde and Ziegel (2017) and Fissler and Hoga (2023) by introducing the following test statistic provided by Fissler and Hoga (2023) in the supplementary materials :

$$\mathcal{T}_n^V := n \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{V}(\boldsymbol{r}_t, (X_t, Y_t))\right)' \widehat{\boldsymbol{\Sigma}}_n^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{V}(\boldsymbol{r}_t, (X_t, Y_t))\right). \tag{4.23}$$

Based on this expression, the test statistic for each identification function is calculated. For instance, for $\mathcal{T}_n^{\boldsymbol{V}^{(VaR,CoVaR)}}$ proposed by Fissler and Hoga (2023) the $\widehat{\boldsymbol{\Sigma}}_n$ is defined similarly to $\widehat{\boldsymbol{\Omega}}_n$ in equation (4.10), but we replace $\boldsymbol{d}_t$ with $\boldsymbol{V}^{(\mathrm{VaR,CoVaR})}(\boldsymbol{r}_t, (X_t, Y_t))$ as defined in (4.7). The test statistic $\mathcal{T}_n^V$ in (4.23) follows an asymptotic χ^2 distribution with degrees of freedom equal to the dimension of $\boldsymbol{V}$ under the null hypothesis, see Fissler and Hoga (2023) for more details. However, in the case of our one dimensional identification function, the test statistic, denoted as $\mathcal{T}_n^{V^{(VaR+CoVaR)}}$, will define the $\widehat{\boldsymbol{\Sigma}}_n$ in (4.23) similarly to $\widehat{\sigma}_n$ in equation (4.17), but we replace d_t^{uni} with $V^{(\mathrm{VaR+CoVaR})}(\boldsymbol{r}_t, (X_t, Y_t))$ as defined in (4.14).

Let $z_{\alpha/2}$ denote the $(1 - \alpha/2)$-quantile of the standard normal distribution. Then, the null hypothesis of correct unconditional calibration for $V^{(\mathrm{VaR+CoVaR})}$ is rejected at the α significance level if:

$$|\mathcal{T}_n^{V^{(\mathrm{VaR+CoVaR})}}| > z_{\alpha/2}$$

Rejection of the null hypothesis implies that the forecasts of VaR and CoVaR are not correctly calibrated, as assessed by the identification function $V^{(\mathrm{VaR+CoVaR})}$. This suggests that the forecasts are misspecified and do not align with the true underlying distribution of the data.

It is important to note that the test statistic $\mathcal{T}_n^{V^{(\mathrm{VaR+CoVaR})}}$ represents a two-sided test and rejection of the null hypothesis occurs for both large positive and large negative values of the test statistic, indicating potential misspecification in either direction. By comparing the absolute value of $\mathcal{T}_n^{V^{(\mathrm{VaR+CoVaR})}}$ to the critical value $z_{\alpha/2}$, we can assess the statistical significance of the deviations from correct calibration and make informed decisions regarding the adequacy of the forecasts based on the proposed identification function.

The simulation results are based on

$$\overline{\boldsymbol{V}}^{(\widehat{\text{VaR}},\widehat{\text{CoVaR}})} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{V}^{(\text{VaR,CoVaR})}\left((\widehat{\text{VaR}}_t, \widehat{\text{CoVaR}}_t), (X_t, Y_t)\right),$$

$$\overline{V}^{BR} = \frac{1}{n}\sum_{t=1}^{n} V^{BR}\left((\widehat{\text{VaR}}_t, \widehat{\text{CoVaR}}_t), (X_t, Y_t)\right),$$

and

$$\overline{V}^{(\widehat{\text{VaR}}+\widehat{\text{CoVaR}})} = \frac{1}{n}\sum_{t=1}^{n} V^{(\text{VaR+CoVaR})}\left((\widehat{\text{VaR}}_t, \widehat{\text{CoVaR}}_t), (X_t, Y_t)\right).$$

Table (4.1) presents rejection frequencies, derived from 10,000 replications, for three tests conducted at a 5% significance level. The investigation is under different sample sizes ($n \in \{500, 1000\}$) and considers three identification functions: $\boldsymbol{V}^{(\text{VaR,CoVaR})}$, V^{BR}, and $V^{(\text{VaR+CoVaR})}$. When employing $\boldsymbol{V}^{(\text{VaR,CoVaR})}$, the test keeps size, with approximately 6% of correctly specified forecasts leading to rejection of the null hypothesis (close to the expected 5% level). Additionally, it demonstrates high power in detecting misspecified forecasts, nearly identifying them with certainty (close to 100% rejection) across both sample sizes. In contrast, the identification function V^{BR} fails to differentiate between correctly specified and misspecified forecasts effectively. At a sample size of 500 (1000), the null hypothesis is rejected at similar frequencies of around 29% (8%) for both types of forecasts. This limitation highlights the risk of employing non-strict identification functions for backtesting systemic risk measures. Notably, the performance of $V^{(\text{VaR+CoVaR})}$ is comparable to $\boldsymbol{V}^{(\text{VaR,CoVaR})}$. It also keeps size, with the rejection frequency for correctly specified forecasts being approximately 6% (5.5%) at a sample size of 500 (1000). Moreover, similar to $\boldsymbol{V}^{(\text{VaR,CoVaR})}$, misspecified forecasts are nearly identified with certainty (100% rejection) across both sample sizes. These findings underscore the ability of our proposed identification function to distinguish between correct and misspecified forecasts, a crucial aspect of effective backtesting procedures for systemic risk measures.

Sample Size	Identification Function	Correctly Specified Forecasts	Misspecified Forecasts
500	$\boldsymbol{V}^{(\text{VaR,CoVaR})}$	6.13	99.97
	V^{BR}	29.06	28.72
	$V^{(\text{VaR+CoVaR})}$	6.69	99.69
1000	$\boldsymbol{V}^{(\text{VaR,CoVaR})}$	6.68	100
	V^{BR}	8.57	7.98
	$V^{(\text{VaR+CoVaR})}$	5.86	100

TABLE 4.1: Rejection frequencies (%) of H_0 based on $\overline{\boldsymbol{V}}^{(\widehat{\text{VaR}},\widehat{\text{CoVaR}})}$, $\overline{V}^{BR}$, and $\overline{V}^{(\widehat{\text{VaR}}+\widehat{\text{CoVaR}})}$ for different sample sizes under Normal distribution. Results are displayed for correctly specified and misspecified forecasts.

We have also examined the three identification functions under a t-distribution with 5 degrees of freedom t_5. The covariance matrix for the t-distribution, denoted as $\mathbf{\Sigma}_t^{(5)}$, is obtained by scaling the covariance matrix of a corresponding normal distribution,$\mathbf{\Sigma}$, by a factor determined by the degrees of freedom parameter (ν) of the t-distribution. Specifically, $\mathbf{\Sigma}_t^{(5)} = \frac{\nu}{\nu-2} \times \mathbf{\Sigma} = \frac{5}{3} \times \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$. Given the assumed distribution and a specific significance level $\alpha = \beta = 0.95$, the correctly specified forecasts, denoted as $\widehat{\text{VaR}}_{\beta,t} \approx 2.02$ and $\widehat{\text{CoVaR}}_{\alpha|\beta,t} \approx 4.72$. Additionally, we introduce the misspecified $\widehat{\text{VaR}}_{\beta',t} \approx 3.36$ and $\widehat{\text{CoVaR}}_{\alpha'|\beta',t} \approx 4.76$, where $\alpha' = 0.75$ and $\beta' = 0.99$. Table (4.2) presents the rejection frequencies of the null hypothesis derived from 10,000 replications for three tests conducted at a 5% significance level and sample sizes$n \in \{500, 1000\}$ under t distribution. The simulation results corroborate the findings presented in Table (4.1). In particular, the identification function proposed by Banulescu-Radu et al. (2021) exhibits a clear limitation in distinguishing between correctly specified and misspecified forecasts. It rejects the null hypothesis at approximately similar frequencies for both types of forecasts, failing to differentiate between them effectively. In contrast, our proposed identification function, $V^{(\text{VaR+CoVaR})}$, demonstrates comparable performance to the identification function introduced by Fissler and Hoga (2023). Both of these identification functions exhibit a high degree of power in detecting misspecified forecasts, rejecting the null hypothesis 100% frequency across the different sample sizes examined.

Sample Size	**Identification Function**	**Correctly Specified Forecasts**	**Misspecified Forecasts**
500	$\boldsymbol{V}^{(\text{VaR,CoVaR})}$	9.08	100
	V^{BR}	93.67	98.73
	$V^{(\text{VaR+CoVaR})}$	9.78	100
1000	$\boldsymbol{V}^{(\text{VaR,CoVaR})}$	10.98	100
	V^{BR}	87.54	97.2
	$V^{(\text{VaR+CoVaR})}$	8.83	100

TABLE 4.2: Rejection frequencies (%) of H_0 based on $\overline{\boldsymbol{V}}^{(\widehat{\text{VaR}},\widehat{\text{CoVaR}})}$, $\overline{V}^{BR}$, and $\overline{V}^{(\widehat{\text{VaR}}+\widehat{\text{CoVaR}})}$ for different sample sizes under t_ν where $\nu = 5$. Results are displayed for correctly specified and misspecified forecasts.

Our investigation continues by next examining the elicitability of our proposed univariate score function $S^{(\text{VaR+SR})}$ through a simulation exercise. This exercise systematically tests the score function behaviour across various scenarios and evaluates its sensitivity and accuracy.

4.3.3.4 Exploring the Elicitability of Our Univariate Score Function Through Numerical Simulation

We conduct a comprehensive simulation study to numerically investigate the elicitability of our proposed score function $S^{(\text{VaR+CoVaR})}$. We introduce $S^{(\text{VaR+CoVaR+CoES})}$ later in this section.

In this study, we systematically implement various simulation scenarios to validate our score functions. We start the analysis within a univariate framework, where we generate data following a normal distribution with the objective to minimise $S^{(\text{VaR})}$, the reader is referred to equation (4.3) for $S^{(\text{VaR})}$ definition, we set the risk levels for the first scenario at $\alpha = 0.95$. The solution to this optimisation problem is the candidate for VaR value. This experimental setting evaluates the accuracy of the $S^{(\text{VaR})}$ in estimating the quantile associated with the specified coverage probability α by comparing the solution of the optimisation with the theoretical VaR value. The elicitability will be illustrated when we are able to retrieve the VaR theoretical value from the minimisation exercise. Then, we extend our evaluation within a bivariate framework and as a general setting and for simplicity, we set the risk levels for all scenarios at $\alpha = \beta = 0.95$. For a bivariate framework our variables follow a normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. We set $\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and the matrix Σ is considered as a symmetric matrix where the off-diagonal of the matrix denoted as ρ. We start the analysis under a bivariate framework by generating two independent random variables under $N(\mu, \Sigma)$. We introduce our joint score function $S^{(\text{VaR+VaR})}$ that is defined in the general proposed univariate score function in (4.11) where the marginal scores that we combine are $S^{(\text{VaR})}$ and $S^{(\text{VaR})}$. In this scenario 2, our objective is to minimise $S^{(\text{VaR+VaR})}$ and allocate the solutions of this optimisation problem that correspond to the candidate for VaR values of each marginal score function and evaluate the two VaR solutions with the theoretical VaR values of each independent random variable. The elicitability will be illustrated when we are able to retrieve the two VaR theoretical values from the minimisation exercise.

In Scenario 3, we repeat Scenario 2 but with replacing one of the $S^{(\text{VaR})}$ with $S^{(\text{CoVaR})}$. Therefore, our objective is to minimise the joint score $S^{(\text{VaR+CoVaR})}$ that is introduced in (4.12), where the marginal scores $S^{(\text{VaR})}$ and $S^{(\text{CoVaR})}$ are defined earlier in (4.3) and (4.4), respectively. Since we generate two independent random variables under $N(\mu, \Sigma)$, then the CoVaR value will be the VaR value of the first VaR. Again, the elicitability will be illustrated when we retrieve the two VaR theoretical values from the minimisation exercise.

To ensure that the proposed score function accurately estimates risk measures in the presence of correlations, in Scenarios 4 and 5, we minimise $S^{(\text{VaR+VaR})}$ following a bivariate distribution under $N(\mu, \Sigma)$, where the correlation coefficients ρ for Scenarios 4 and 5 are 0.4 and -0.4, respectively. The elicitability will be illustrated when we are

able to retrieve the two VaR theoretical values from the minimisation exercise under these scenarios.
In Scenarios 6 and 7, we replicate Scenarios 4 and 5 by replacing one of $S^{(\text{VaR})}$ components by $S^{(\text{CoVaR})}$. Therefore, our objective is to minimise the joint score $S^{(\text{VaR+CoVaR})}$ that is introduced in (4.12), where the marginal scores $S^{(\text{VaR})}$ and $S^{(\text{CoVaR})}$ are defined earlier in (4.3) and (4.4), respectively. This adjustment allows us to examine the performance of $S^{(\text{VaR+CoVaR})}$ in estimating risk measures in the presence of correlations. The elicitability will be illustrated when we are able to retrieve the VaR theoretical value and CoVaR theoretical value from the minimisation exercise under these scenarios. Table (4.3) summarises the scenarios conducted in our analysis.

Scenario	Score Function	Distribution	Correlation Coefficients (ρ)
1	$S^{(\text{VaR})}$	N(0,1)	N/A
2	$S^{(\text{VaR+VaR})}$	$N(\mu, \Sigma)$	0
3	$S^{(\text{VaR+CoVaR})}$	$N(\mu, \Sigma)$	0
4	$S^{(\text{VaR+VaR})}$	$N(\mu, \Sigma)$	0.4
5	$S^{(\text{VaR+VaR})}$	$N(\mu, \Sigma)$	-0.4
6	$S^{(\text{VaR+CoVaR})}$	$N(\mu, \Sigma)$	0.4
7	$S^{(\text{VaR+CoVaR})}$	$N(\mu, \Sigma)$	-0.4

TABLE 4.3: Summary of Simulation Scenarios of $S^{(\text{VaR+CoVaR})}$ score function.

To provide robust empirical evidence regarding our proposed score function, we iterate the simulation procedure 100 times, and as a reference point, we compute the theoretical value corresponding to each scenario. We consider computing the squared difference between the fixed theoretical value and the estimated optimal value across all iterations. Subsequently, we compute the mean of these squared differences over all iterations. The theoretical values of CoVaR obtained from the conditional normal distribution corresponding to the specific α and β, which is obtained as follows:
If we have a multivariate normal vector $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, then the conditional quantile of the multivariate normal vector $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ at level β, given that the realization of the first variable X as its VaR_α is given by:

$$\text{CoVaR}_\beta(\boldsymbol{Y} \mid X = \text{VaR}_\alpha) = \overline{\boldsymbol{\mu}} + z_\beta \sqrt{\overline{\Sigma}} \quad ,$$

here, $\overline{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\text{VaR}_\alpha - \boldsymbol{\mu}_2)$, z_β is the critical value from the standard normal distribution for quantile β, and $\sqrt{\overline{\Sigma}} = \sqrt{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}$.
Table (4.4) presents the simulation study results evaluating the performance of our proposed score function, with risk levels $\alpha = \beta = 0.95$, under various structures. We

include the mean of optimal values across all iterations, the theoretical values and the Mean Squared Error (MSE) of each Scenario where these results contribute to the discussion of the elicitable property of our proposed loss functions. Overall, the low MSE values in Table (4.4) suggest that the estimated optimal values obtained from our minimisation exercises closely align with the fixed theoretical values on average, which suggests that our score functions successfully are able to provide accurate estimates that closely match the theoretical values across multiple simulation runs.

Scenario	Theoretical Value	Mean Optimal Value	MSE
1	v: 1.6449	v^* : 1.6431	0.0005
2	v_1: 1.6449	v_1^*:1.6446	0.0004
	v_2: 1.6449	v_2^*: 1.6439	0.0004
3	v: 1.6449	v^*: 1.6760	0.0017
	c : 1.6449	c^* : 1.5835	0.0234
4	v_1 1.6449	v_1^*:1.6408	0.0004
	v_2 1.6449	v_2^* : 1.6405	0.0005
5	v_1: 1.6449	v_1^* : 1.6456	0.0004
	v_2: 1.6449	v_2^* : 1.6442	0.0004
6	v: 1.6449	v^*:1.6493	0.0006
	c: 2.3538	c^*: 2.3737	0.0144
7	v: 1.6449	v^*: 1.6473	0.0006
	c: 0.7005	c^* : 0.6945	0.0073

TABLE 4.4: Comparison of theoretical values, mean optimal values and MSE which represents the mean of squared differences between the fixed theoretical value and the estimated optimal value across 100 iterations. This simulation study evaluates the performance of our proposed score function, with risk levels $\alpha = \beta = 0.95$, under various structures. The findings presented in this section contribute to discussing the elicitable property of the proposed loss functions. The v^* and c^* represent the mean optimal values of VaR and CoVaR, respectively.

As a robustness check of the sensitivity of our score functions, in scenarios 6 and 7, we assess the effect of varying risk levels on the performance of our proposed score function $S^{(\text{VaR+CoVaR})}$ by examining a range of values for α and β. Tables (4.5) and (4.6) provide the theoretical CoVaR values at different risk levels of α and β and the optimal CoVaR values computed through the simulation in Scenarios 6 and 7. From the findings illustrated in Tables (4.5) and (4.6), we conclude that there is consistency between optimal and theoretical CoVaR values across various risk levels, suggesting that the optimisation process provides results that capture the conditional tail behaviour successfully.

Furthermore, in the Appendix of this Chapter, Figure (4.7) represents the MSE values of the CoVaR estimation component in the score function $S^{(\text{VaR+CoVaR})}$ across a range of risk levels specified by $\alpha = \beta \in [0.1, 0.99]$ for correlation coefficients ρ = 0.4 and ρ=

-0.4. Moreover, we analyze the impact of varying correlation coefficients, in addition to the risk levels specified by $\alpha = \beta \in [0.1, 0.99]$, on the performance of our score function $S^{(\text{VaR+CoVaR})}$ by examining a range of positive and negative correlation values. For the positive correlation case, we utilise $\rho = 0.9, 0.7, 0.5, 0.3$, and 0.1, while for the negative correlation case, we use $\rho = -0.9, -0.7, -0.5, -0.3$, and -0.1. Figures (4.8) and (4.9) in the Appendix of this Chapter illustrate the MSE values of the CoVaR estimation component in the score function $S^{(\text{VaR+CoVaR})}$ across positive and negative correlation values, respectively.

Risk Level	CoVaR Theoretical Value	Mean Optimal Value	MSE
$\alpha = \beta = 0.97$	2.6761	2.6529	0.0186
$\alpha = \beta = 0.90$	1.895	1.894	0.0071
$\alpha = \beta = 0.85$	1.589	1.577	0.0035
$\alpha = \beta = 0.80$	1.346	1.339	0.0013
$\alpha = \beta = 0.75$	1.139	1.1360	0.0006

TABLE 4.5: Summary of results for simulation study investigates the performance of our proposed risk measures $S^{(\text{VaR+SR})}$ in Scenario 6, correlation coefficient ρ = 0.4, with various risk levels $\alpha = \beta$. The analysis is based on CoVaR estimation, where the CoVaR theoretical value, mean optimal value, and Mean Squared Error (MSE) are examined. The mean optimal value represents the average CoVaR value obtained from the optimisation process across 100 runs.

Risk Level	CoVaR Theoretical Value	Mean Optimal Value	MSE
$\alpha = \beta = 0.97$	0.8345	0.8140	0.0134
$\alpha = \beta = 0.90$	0.490	0.4798	0.0024
$\alpha = \beta = 0.85$	0.345	0.3434	0.0014
$\alpha = \beta = 0.80$	0.227	0.2244	0.0009
$\alpha = \beta = 0.75$	0.124	0.1228	0.0006

TABLE 4.6: Summary of results for simulation study investigates the performance of our proposed risk measures $S^{(\text{VaR+SR})}$ in Scenario 7, correlation coefficient ρ = -0.4, with various risk levels $\alpha = \beta$. The analysis is based on CoVaR estimation, where the CoVaR theoretical value, mean optimal value, and Mean Squared Error (MSE) are examined. The mean optimal value represents the average CoVaR value obtained from the optimisation process across 100 runs.

Additionally, we extend the score function from $S^{(\text{VaR+CoVaR})}$ to $S^{(\text{VaR+CoVaR+CoES})}$, where the objective is to introduce the concept of CoES alongside VaR and CoVaR. Our objective is to minimise the joint score function $S^{(\text{VaR+CoVaR+CoES})}$ that is introduced in (4.19) where the marginal scores $S^{(\text{VaR})}$ and $S^{(\text{CoVaR+CoES})}$ are defined earlier in (4.3) and (4.5). Under the assumption that $\boldsymbol{Y}$ follows a bivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ, we compute the theoretical CoES to compare the theoretical values with our optimisation values that we allocate from minimising $S^{(\text{VaR+CoVaR+CoES})}$.

$$\text{CoES}_\beta(\boldsymbol{Y} \mid X = \text{VaR}_\alpha) = \mu_{Y|X} + \sigma_{Y|X} \frac{1}{\alpha} \int_{1-\alpha}^{1} \Phi^{-1}(u) du$$

Where, $\mu_{Y|X}$ and $\sigma_{Y|X}$ are the conditional mean and standard deviation of Y given $X = \text{VaR}_\alpha$, $\Phi^{-1}(u)$ is the inverse cumulative distribution function (quantile function) of the standard normal distribution evaluated at u. The term $\frac{1}{\alpha}\int_{1-\alpha}^{1}\Phi^{-1}(u)du$ represents the conditional expectation of the inverse standard normal distribution in the right tail region beyond the α-quantile.

For $S^{(\text{VaR+CoVaR+CoES})}$, we conduct a comprehensive simulation exercise by examining the impact of changing the correlation coefficient values on the our score function $S^{(\text{VaR+CoVaR+CoES})}$ across the risk levels interval $\alpha = \beta \in [0.1, 0.99]$. In the Appendix of this Chapter, Figures (4.10) and (4.11) illustrate the MSE values of the CoES estimation component based on the score function $S^{(\text{VaR+CoVaR+CoES})}$ across a range of $\alpha = \beta \in [0.1, 0.99]$ values for positive and negative correlation coefficients, respectively.
Overall, after conducting different simulation scenarios, we have drawn essential conclusions regarding our score function's elicitability by systematically testing its behaviour across various scenarios, including univariate and multivariate distributions and comparing the results with theoretical values. The successful retrieval of quantiles corresponding to specified coverage probabilities in different simulation scenarios and the consistent results obtained from the score functions underscore the consistency of our joint score function across different risk measures, demonstrating the joint score function's elicitability. Additionally, we assess the effect of varying risk levels on the performance of our joint score function by examining a range of values for α and β. The results of the optimal values and theoretical CoVaR and CoES values across various risk levels prove our score function's robustness and effectiveness in conducting comparative backtesting and risk assessment tasks.

4.4 Numerical Evidence: Monte Carlo Simulation and Empirical Application

In Subsection(4.4.1), the simulation study aims to evaluate the performance of the proposed backtesting procedure under a more realistic and dynamic setting unlike the simulation conducted in (4.3.3.3) where we focus mainly in investigate the performance of our identification function $V^{(\text{VaR+CoVaR})}$ and other existing identification functions and illustrate the risks of employing non-strict identification functions for backtesting systemic risk measures.
Consistent with the models and dataset outlined in Fissler and Hoga (2023), we employ a bivariate GARCH(1,1) model with a t-copula for innovations driven by GAS in the simulation subsection. In the empirical application subsection, we conduct a comparative analysis by implementing our proposed backtest, contrasting the bivariate GARCH(1,1) model with a Gaussian-copula for innovations driven by GAS

against the GJR-GARCH(1,1) model with a t-copula driven by GAS for the dependence model. This approach ensures consistency and smooths the way for a comprehensive analysis of the given models across simulated and real-world scenarios.

Additionally, it is important to note that in both the Monte Carlo simulations and the empirical application, we follow the same setting and data-generating process employed in the Fissler and Hoga (2023) paper. This ensures consistency and facilitates direct comparison with the results presented therein.

4.4.1 Monte Carlo simulations

In our simulation study, we examine how well our test $\mathcal{T}_n^{OS.uni}$ performs in finite samples when testing H_0. We apply this test to forecasts of (VaR, CoVaR) and (VaR, CoVaR, CoES) obtained from a bivariate GARCH(1,1) model with a t-copula for innovations. We focus on one-step-ahead forecasts throughout our analysis.

4.4.1.1 Data Generating Process

Following the same data generating process setting illustrated in the supplementary materials Fissler and Hoga (2023), this simulation study is built on a carefully designed process that starts with a bivariate GARCH model and copula theory. The behaviour of two variables, X_t and Y_t, is generated by a GARCH(1,1) structure, where their variances evolve over time according to specific formulas.

$$\begin{aligned} X_t &= \sigma_{x,t}\varepsilon_{x,t}, \quad \sigma_{x,t}^2 = \omega_x + \alpha_x X_{t-1}^2 + \beta_x \sigma_{x,t-1}^2, \\ Y_t &= \sigma_{y,t}\varepsilon_{y,t}, \quad \sigma_{y,t}^2 = \omega_y + \alpha_y Y_{t-1}^2 + \beta_y \sigma_{y,t-1}^2, \end{aligned} \tag{4.24}$$

Here, $\omega_z > 0, \alpha_z \geq 0, \beta_z \geq 0$ $(z \in \{x, y\})$. The innovations in these variables, $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$, have different distributions: $\varepsilon_{x,t}$ is standard normal, while $\varepsilon_{y,t}$ follows a standardized Student's t_5 distribution. The key part is the dependence between X_t and Y_t, which is modelled by t-copula with time-varying correlation parameter $\rho_t \in (-1, 1)$. This correlation follows the Generalized Autoregressive Score (GAS) model. In particular, following Creal et al. (2013) and Fissler and Hoga (2023) we set:

$$\begin{aligned} f_t &= \omega^\dagger + \alpha^\dagger s_{t-1} + \beta^\dagger f_{t-1}, \\ s_{t-1} &:= \frac{\mathrm{d}}{\mathrm{d}f_{t-1}} \log c\left(\boldsymbol{U}_{t-1}; \vartheta, \Delta\left(f_{t-1}\right)\right), \\ \rho_t &= \Delta(f_t), \\ \Delta(x) &= [1 - \exp(-x)]/[1 + \exp(x)] \in (-1, 1). \end{aligned} \tag{4.25}$$

Table (4.7) presents the empirically plausible values for the marginal GARCH processes and the t-copula dependence structure (GAS innovations) implemented in this study, following Fissler and Hoga (2023) simulation setting. Figure (4.1) summarises the Data Generating Process (DGP). For more details and discussion regarding the set-up of the DGP, the reader is referred to supplementary materials in Fissler and Hoga (2023).

Parameter	Value
Marginal Parameters (GARCH)	
ω_x, ω_y	0.001
α_x, α_y	0.2
β_x, β_y	0.79
Dependence Parameters (GAS)	
$\omega^\dagger$	0.001
$\alpha^\dagger$	0.1
$\beta^\dagger$	0.99
ϑ	5

TABLE 4.7: Parameter Values for Data Generating Process

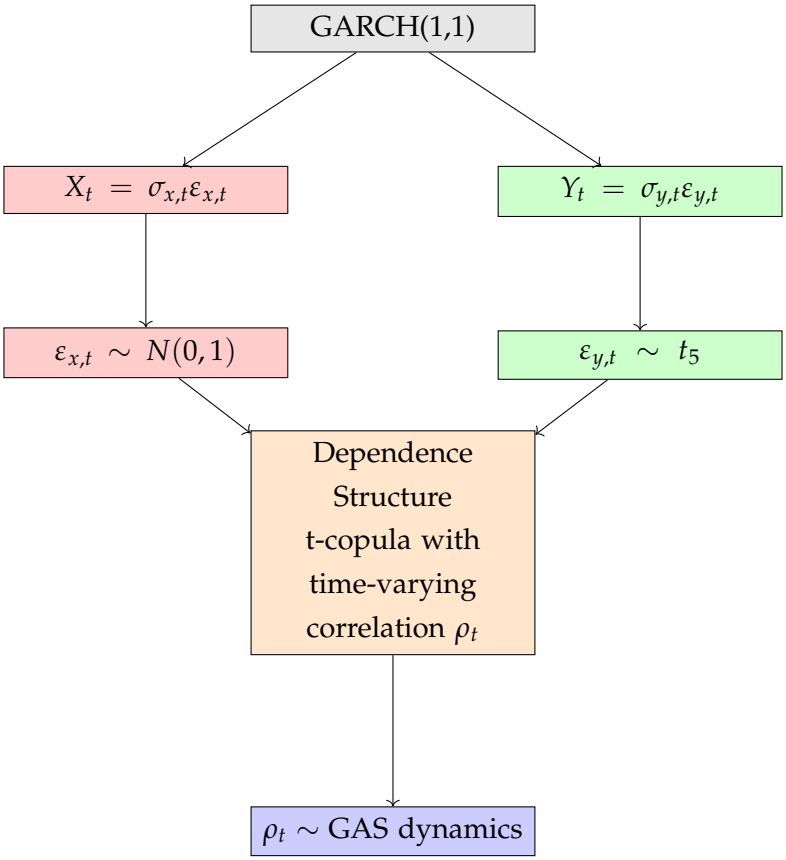


FIGURE 4.1: Flowchart illustrating the components of the data generating process

4.4.1.2 Risk Forecasts

In this subsection, we describe the procedure of forecasting systemic risk measures and a model specification utilised for the forecasting. A common practice in financial risk forecasting is a rolling window approach. We have introduced the rolling window approach in Chapter (3), and we will adjust this approach for the example displays in (4.4.2). In this study, instead of a rolling-window approach, a fixed-window approach is employed due to computational feasibility concerns, especially we conduct 10,000 replications for our analysis to ensure robust estimates with minimal standard errors. The generated data, denoted as$\{(X_t, Y_t)\}_{t=-r+1,\dots,n}$ from a bivariate GARCH model in (4.24) with GAS t-copula for the dependence structure to capture the joint dynamics of X_t and Y_t is split into an in-sample period$\{(X_t, Y_t)\}_{t=-r+1,\dots,0}$ and an out-of-sample period$\{(X_t, Y_t)\}_{t=1,\dots,n}$. The parameter estimation is performed using the in-sample data, then we utilize the estimated parameters to generate risk forecasts (VaR, SR) for the out-of-sample window. Next, we will describe the in-sample parameter estimation and the out-of-sample risk forecasting.
After generating the data, the marginal parameters, $(\omega_x, \alpha_x, \beta_x)$ and $(\omega_y, \alpha_y, \beta_y)$, are estimated through the in-sample data using Gaussian quasi-maximum likelihood techniques (QMLE) (Francq and Zakoian (2004),Fissler and Hoga (2023)). More specifically, we compute the conditional variances $\widehat{\sigma}^2_{x,t}$ and $\widehat{\sigma}^2_{y,t}$ for X_t and Y_t respectively, along with the standardised residuals $\widehat{\varepsilon}_{x,t}$ and $\widehat{\varepsilon}_{y,t}$. Then we obtain the estimated Probability Integral Transforms (PITs) $\widehat{\boldsymbol{U}}_t$ using the empirical cumulative distribution functions of the standardized residuals where $\widehat{\boldsymbol{U}}_t = (\widehat{F}_x(\widehat{\varepsilon}_{x,t}), \widehat{F}_y(\widehat{\varepsilon}_{y,t}))$. After that, we estimate the marginal parameters $(\omega^\dagger, \alpha^\dagger, \beta^\dagger, \vartheta)$ by employing the maximum likelihood estimation based on $\widehat{\boldsymbol{U}}_t$ as proposed by Creal et al. (2013). Then, we utilize the estimated parameters to generate risk forecasts, including VaR and SR for the out-of-sample window.
Next, we describe one-step-ahead forecasting of the conditional risk measures $\mathrm{VaR}_t(X_t)$, $\mathrm{CoVaR}_t(Y_t \mid X_t)$, and $\mathrm{CoES}\,t(Y_t \mid X_t)$ for out-of-sample observations. We repeat the definition of the risk measures forecasted including VaR, CoVaR and CoES as presented in the supplementary materials in Fissler and Hoga (2023). VaR_t for X_t, denoted as the β-quantile of the conditional distribution $F_{X_t|\mathfrak{F}_{t-1}}$, where α and β are risk levels set to 0.95 and $\mathfrak{F}_{t-1}$ represents the information set up to time $t-1$. i.e. $\mathfrak{F}_{t-1} = \sigma((X_{t-1}, Y_{t-1}), (X_{t-2}, Y_{t-2}), \dots)$.

$$\mathrm{VaR}_t(X_t) = \mathrm{VaR}_\beta\left(F_{X_t|\mathfrak{F}_{t-1}}\right)$$

Similarly, CoVaR_t for Y_t given X_t is denoted as

$$\text{CoVaR}_t\left(Y_t \mid X_t\right) = \text{CoVaR}_{\alpha|\beta}\left(F_{(X_t,Y_t)|\mathfrak{F}_{t-1}}\right)$$

In addition, (CoES_t) for Y_t given X_t is forecasted as $\text{CoES}_{\alpha|\beta}$, both defined based on the conditional distribution and past information.

$$\text{CoES}_t\left(Y_t \mid X_t\right) = \text{CoES}_{\alpha|\beta}\left(F_{(X_t,Y_t)|\mathfrak{F}_{t-1}}\right)$$

Where, $F_{(X_t,Y_t)|\mathfrak{F}_{t-1}}(x,y) = \text{P}\left\{X_t \leq x, Y_t \leq y \mid \mathfrak{F}_{t-1}\right\} =: \text{P}_{t-1}\left\{X_t \leq x, Y_t \leq y\right\}$ for $x, y \in \mathbb{R}$.

We obtain $\widehat{\text{VaR}}_t$ for X_t using the estimated parameters obtained from the in-sample window, by using conditional standard deviation $\widehat{\sigma}_{x,t}$ and the empirical β-quantile of standardized residuals $\{\widehat{\varepsilon}_{x,t=-r+1,\ldots,0}\}$ i.e.
$\widehat{\text{VaR}}_t := \widehat{\text{VaR}}_t\left(X_t\right) = \widehat{\sigma}_{x,t}\widehat{\text{VaR}}_{x,\varepsilon}, \quad t = 1,\ldots,n.$
However, CoVaR and CoES are obtained by numerically solving integral equations involving the estimated t-copula dependence structure. In particular, CoVaR is determined by solving

$$\text{P}_{t-1}\left\{Y_t > \text{CoVaR}_t\left(Y_t \mid X_t\right) \mid X_t \geq \text{VaR}_t\left(X_t\right)\right\} = 1-\alpha.$$

Additionally, considering the probability that X_t itself exceeding its VaR level , denoted as $1-\beta$, straightforward computations lead to the implicit definition of $\text{CoVaR}_t\left(Y_t \mid X_t\right)$. This definition involves a double integral over the joint distribution of Y_t and X_t.

$$(1-\alpha)(1-\beta) = \int_{F_y\left(\text{CoVaR}_t(Y_t|X_t)/\sigma_{y,t}\right)}^{1} \int_{\beta}^{1} c\left(\left(u_1,u_2\right),\vartheta,\rho_t\right) \mathrm{d}u_1\, \mathrm{d}u_2. \tag{4.26}$$

Substituting estimates, for example, replacing ϑ with the in-sample estimate $\widehat{\vartheta}$ and F_y with the empirical cdf $\widehat{F}_y$,etc. Then numerically solving the resulting equation (4.26) gives the forecasts $\widehat{\text{CoVaR}}_t$ for $t = 1,\ldots,n$. After obtaining the CoVaR forecasts $\widehat{\text{CoVaR}}_t$, we proceed to calculate the prediction of CoES. From (4.2) we derive CoES predictions $\widehat{\text{CoES}}_t$ for $t = 1,\ldots,n$. The methods used to obtain CoVaR and CoES in this study were further detailed in the supplementary materials in Fissler and Hoga (2023).
Following Fissler and Hoga (2023) to derive the contaminated forecasts, we multiply the correctly specified forecasts with independent and identically distributed Weibull distributed random variables. This noise is assumed to be independent and

Weibull-distributed, ensuring that the forecasts remain positive, and this is significant due to the requirements of the 0-homogeneous loss functions, see (Fissler and Hoga (2023)). The equations for this contamination process are as follows:

Let $\widehat{\boldsymbol{r}}_t = (\widehat{r}_t^{\text{VaR}}, \widehat{\boldsymbol{r}}_t^{\text{SR}}) = (\widehat{\text{VaR}}_t, \widehat{\text{CoVaR}}_t, \widehat{\text{CoES}}_t)$ be the vector of correctly specified forecasts for time t, where $\widehat{\text{VaR}}_t$ is the VaR forecast, $\widehat{\text{CoVaR}}_t$ is the CoVaR forecast, and $\widehat{\text{CoES}}_t$ is the CoES forecast.

In particular, to generate two sets of contaminated forecasts, $\widehat{\boldsymbol{r}}_{1,t,}$ and $\widehat{\boldsymbol{r}}_{2,t}$, we implement the following:

$$\begin{aligned}
\widehat{r}_{i,t,}^{\text{VaR}} &= \widehat{\text{VaR}}_t \cdot \epsilon_{i,t}^{\text{VaR}} \\
\widehat{r}_{i,t}^{\text{CoVaR}} &= \widehat{\text{CoVaR}}_t \cdot \epsilon_{i,t}^{\text{CoVaR}} \\
\widehat{r}_{i,t}^{\text{CoES}} &= \widehat{\text{CoES}}_t \cdot \epsilon_{i,t}^{\text{CoES}}
\end{aligned}$$

where $i = 1, 2$, and $\epsilon_{i,t}^{\text{VaR}}{}_{t=1,\ldots,n}$, $\epsilon_{i,t}^{\text{CoVaR}}{}_{t=1,\ldots,n}$, and $\epsilon_{i,t}^{\text{CoES}}{}_{t=1,\ldots,n}$ are mutually independent and identically distributed Weibull-distributed random variables with the following density function:

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, \quad x > 0$$

We choose the shape parameter $k = 10$ and the scale parameter $\lambda = 0.3$, which results in a mean of $\lambda\Gamma(1 + 1/k) \approx 0.285$ for the multiplicative noise; for further detail, see Fissler and Hoga (2023). This choice ensures that the contaminated forecasts $\widehat{\boldsymbol{r}}_{1,t}$ and $\widehat{\boldsymbol{r}}_{t,(2)}$ are seriously misspecified on average.

These equations and the chosen parameters for the Weibull distribution allow to generate two distinct sets of contaminated forecasts that are misspecified on average due to the multiplicative noise Fissler and Hoga (2023). The Flowchart (4.2) illustrates the steps involved in the risk forecasting process.

4.4.1.3 Test description

In the assessment of the risk forecasts, the score functions we choose have a 0-homogeneous property such that the score differences remain unchanged when a positive constant scales both forecasts and observations. Using 0-homogeneous scores ensures that the comparisons between different models remain constant regardless of the scale used for the forecasting and observations. This property is important in making the model comparison meaningful and robust. In addition, having a 0-homogeneous score often leads to higher power of DM tests where the test is more likely to distinguish between forecasting models (Nolde and Ziegel (2017), Patton

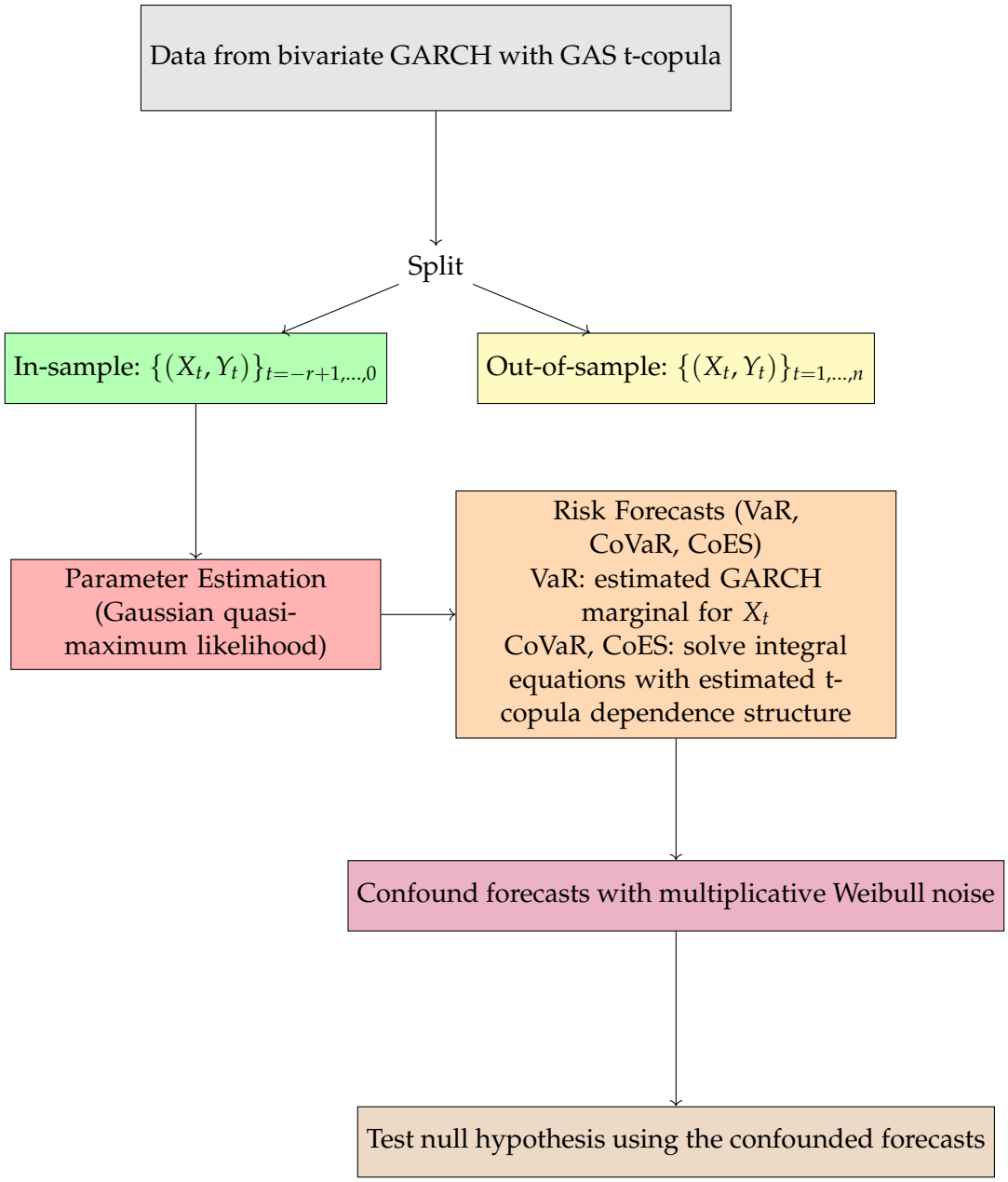


FIGURE 4.2: Flowchart depicting the steps involved in the risk forecasting process

et al. (2019), Taylor (2019) and Fissler and Hoga (2023)). To achieve 0-homogeneity in the VaR score function in (4.3) and following (Nolde and Ziegel (2017), example 4 and the supplementary materials in Fissler and Hoga (2023)) we set $h(z) = log(z)$ and $a^{\text{VaR}}(x, y) = log(x)$. Also, in $S_v^{\text{CoVaR}}(c, (x, y))$ component from (4.4) we choose $g(z) = log(z)$, $a(y) = log(y)$ and $a^{\text{CoVaR}}(x, y) = 0$, this yielding to

$$S_v^{\text{CoVaR}}(c, (x, y)) = \mathbb{1}\{x > v\}[(\mathbb{1}\{y \leq c\} - \alpha)\, \log(c) + \mathbb{1}\{y \geq c\}\log(y)] \tag{4.27}$$

In addition, for the component $S_v^{\text{CoVaR,CoES}}$ in (4.5) we achieve 0-homogeneity by selecting $g(z) = 0$, $a(y) = a^{\text{CoVaR}}(x, y) = 0$ and $\phi(z) = -log(z)(z > 0)$, leading to

$$S_v^{(\text{CoVaR,CoES})}((c,e),(x,y)) \quad = \frac{\mathbb{1}\{x>v\}}{1-\alpha}\left[(\mathbb{1}\{y>c\}\frac{y-c}{e}+(1-\alpha)\left(\frac{c}{e}-1+log(e)\right)\right]. \tag{4.28}$$

Now we investigate the behaviour of our tests after simulating $R = 10000$ trajectories $\{X_t, Y_t\}_{t=-r+1,\ldots,n}$ from bivariate GARCH(1,1) model with GAS-driven t-copula as presented in (4.24) and (4.25) respectively. Here, $r = 1000$ and $n \in \{500, 1000\}$. We fit the model, forecast VaR and systemic risk (SR) components, and compute score differences for each trajectory. Then we obtain $\mathcal{T}_n^{OS.uni}$ test under different scenarios in our analysis and investigate the behaviour of $S^{(\text{VaR+CoVaR})}$ and $S^{(\text{VaR+CoVaR+CoES})}$ components under each scenario. The first scenario under consideration is when VaR and SR forecasts both are comparable, here we compare $r_{1,t} = \widehat{r}_{1,t} = (\widehat{r}_{1,t}^{\text{VaR}}, \widehat{r}_{1,t}^{\text{SR}})$ and $r_{2,t} = \widehat{r}_{2,t} = (\widehat{r}_{2,t}^{\text{VaR}}, \widehat{r}_{2,t}^{\text{SR}})$. Therefore, both forecast sets are equally misspecified, which indicates that there is no difference between the two forecast models.

In the second scenario, we compare the forecast of the two models where we have comparable VaR forecasts and better SR in the second model, which is illustrated as
$r_{1,t} = \widehat{r}_{1,t} = (\widehat{r}_{1,t}^{\text{VaR}}, \widehat{r}_{1,t}^{\text{SR}})$
$r_{2,t} = (\widehat{r}_{2,t}^{\text{VaR}}, \widehat{r}_t^{\text{SR}}) = (\widehat{\text{VaR}_{2,\text{t}}} + \widehat{\text{CoVaR}_\text{t}} + \widehat{\text{CoES}_\text{t}})$

Here, the $r_{2,t}$ forecast combines the misspecified VaR forecast with the correctly specified systemic risk forecasts, implying that the systemic risk forecasts of $r_{1,t}$ are inferior to those of $r_{2,t}$, leading to an expectation of positive score differences under this scenario. For notation simplicity, in the results and table illustration, we will use (*) to indicate accurate forecasting in a model component. For example, $r_{2,t}$ in scenario 2 will be presented in Table (4.8) as $r_{2,t} = \widehat{r}_{2,t} = (\widehat{r}_{2,t}^{\text{VaR}}, \widehat{r}_t^{\text{SR}^*})$ where (*) indicates an accurate forecasts in terms of SR component in the second model.
Table (4.8) presents the results of both scenarios under consideration in test statistics of $S^{(\text{VaR+CoVaR})}$ and $S^{(\text{VaR+CoVaR+CoES})}$ and for different sample sizes.

In the first scenario, where both VaR and Systemic Risk (SR) forecasts are assumed comparable, the test accurately assesses the average difference between the models should be zero. For $S^{(\text{VaR+CoVaR})}$, the rejection frequencies are low, at 7% for a sample size of 500 and 5% for a sample size of 1000. Similarly, for $S^{(\text{VaR+CoVaR+CoES})}$, the rejection frequencies are 5.5% for both sample sizes of 500 and 1000. These low rejection frequencies suggest that the tests are able to maintain the correct size when the null hypothesis is true, even at a relatively small sample size of 500 observations. In other words, the tests do not reject the null hypothesis too frequently when there is no actual difference in forecasting performance between the models being compared. The fact that the rejection frequencies are close to the nominal significance level at 5%) indicates that the tests have good size control. On the other hand, the test demonstrates higher sensitivity in scenario 2, where $r_{2,t}$ outperforms $r_{1,t}$. The

increased rejection frequencies 80% (94%) of the test of $S^{(\text{VaR+CoVaR})}$ and 83% (96%) for $S^{(\text{VaR+CoVaR+CoES})}$ in the sample size of 500 (1000) illustrate the test's ability to distinguish between models with varying systemic risk forecasting accuracy.

Sample Size	Under the assumption, $r_{1,t}^{\text{VaR}} = \widehat{r}_{1,t}^{\text{VaR}}$ and $r_{2,t}^{\text{VaR}} = \widehat{r}_{2,t}^{\text{VaR}}$				
	$\text{S}^{(\text{VaR+SR})}$ **score results**			$\text{S}^{(\text{VaR,SR})}$ **score results**	
	Systemic Risk Forecasts	(VaR+ CoVaR)	(VaR+ CoVaR+ CoES)	(VaR, CoVaR)	(VaR, CoVaR, CoES)
500	$r_{1,t}^{\text{SR}} = \widehat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \widehat{r}_{2,t}^{\text{SR}}$	7	5.5	3.5	3
	$r_{1,t}^{\text{SR}} = \widehat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \widehat{r}_{t}^{\text{SR}*}$	80	83	80	80
1000	$r_{1,t}^{\text{SR}} = \widehat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \widehat{r}_{2,t}^{\text{SR}}$	5	5.5	5	4.5
	$r_{1,t}^{\text{SR}} = \widehat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \widehat{r}_{t}^{\text{SR}*}$	94	96	94	95

TABLE 4.8: Rejection Frequencies (%) of H_0 for $r_{1,t} = (r_{1,t}^{\text{VaR}}, r_{1,t}^{\text{SR}})$ and $r_{2,t} = (r_{2,t}^{\text{VaR}}, r_{2,t}^{\text{SR}})$, assuming VaR forecasts of both models are equally misspecified: $r_{1,t}^{\text{VaR}} = \widehat{r}_{1,t}^{\text{VaR}}$ and $r_{2,t}^{\text{VaR}} = \widehat{r}_{2,t}^{\text{VaR}}$ while we consider various systemic risk scenarios. The last two columns of the table present the results of Fissler and Hoga score function $\text{S}^{(\text{VaR,SR})}$. For $\text{S}^{(\text{VaR+SR})}$, we obtain $\mathcal{T}_n^{OS.uni}$ to test : $H_0 : \text{E}\left[\bar{d}_n^{uni}\right] \leq 0$. While for $\text{S}^{(\text{VaR,SR})}$, we obtain One and a Half-Sided test, $\mathcal{T}_n^{OS}$, to examine $H_0^{\preceq_{\text{lex}}} : \text{E}\left[\bar{d}_{1n}\right] = 0$ and $\text{E}\left[\bar{d}_{2n}\right] \leq 0$. The (*) indicates accurate forecasts in the second model.

In the initial two scenarios, we assume that VaR forecasts are comparable between the models, while the accuracy of systemic risk (SR) forecasts varies. Now, we shift to the scenarios where we have differences in the predictive ability of the VaR predictions. First, we compare the forecast of the two models, where we have better VaR forecasting in the second model and equal accuracy in the SR component. To have a better VaR forecast in the second model, we follow Fissler and Hoga (2023) by introducing the following:

$$\widehat{r}_{2,t}^{VaR^*} = \widehat{\text{VaR}}_t(1.03 \times \epsilon_{2,t}^{VaR})$$

Here,$1.03 \times 0.285 = 0.29355$, this will lead to a higher mean with 0.29355 of the multiplicative noise $1.03 \times \epsilon_{2,t}^{VaR}$ compared with the multiplicative noise of $\epsilon_{1,t}^{VaR}$. By having this and recalling that both $\epsilon_{1,t}^{VaR}$ and $\epsilon_{2,t}^{VaR}$ are Weibull-distributed with identical shape and scale parameters, this will lead to have the average VaR losses of forecasts $\widehat{r}_{1,t}^{VaR}$ are higher than those of the less biased $\widehat{r}_{2,t}^{VaR^*}$. Under this scenario, we have, $r_{1,t} = (\widehat{r}_{1,t}^{\text{VaR}}, \widehat{r}_{1,t}^{\text{SR}})$ and $r_{2,t} = (\widehat{r}_{2,t}^{\text{VaR}^*}, \widehat{r}_{2,t}^{\text{SR}})$ where VaR forecasts in the second model is better while we have equal accuracy in SR component in both models, such that $\text{E}\left[\bar{d}_n\right] > 0$. The final scenario under consideration when the forecasting of both VaR and SR in the second model is better compared to the first model such that $r_{1,t} = (\widehat{r}_{1,t}^{\text{VaR}}, \widehat{r}_{1,t}^{\text{SR}})$ and $r_{2,t} = (\widehat{r}_{2,t}^{\text{VaR}^*}, \widehat{r}_{t}^{\text{SR}^*})$. Here, we implement the same step as the previous scenario to have better VaR forecasting. Table (4.9) presents rejection

frequencies (%) of the null hypothesis ($H_0 : E[\bar{d}_n^{uni}] \leq 0$) for different scenarios and sample sizes, considering the assumptions that $r_{1,t}^{\text{VaR}} = \hat{r}_{1,t}^{\text{VaR}}$ and $r_{2,t}^{\text{VaR}} = \hat{r}_{2,t}^{\text{VaR}^*}$ where (*) indicates an accurate VaR forecasts in term of the second model. In scenario 1, Even though there is a marginal difference in the predictive accuracy of the VaR forecasts while the systemic risk forecasts are comparable, our tests are sensitive enough to detect this difference. Leading to rejection frequencies of 44% (72%) for $S^{(\text{VaR+CoVaR})}$ and 36.5% (64.5%)for $S^{(\text{VaR+CoVaR+CoES})}$ in the sample size of 500 (1000). Furthermore, our tests demonstrate higher power in detecting this slight variation between the forecasting models compared to the test conducted by Fissler and Hoga (2023), with rejection frequencies of 44 % compared to 39.5 % in Fissler and Hoga (2023). Additionally, increasing the sample size enhances the power of our test to identify this variability, with the rejection frequency rising to 72 % compared to Fissler and Hoga (2023) test with 61.5% when the sample size is increased.

Moving to Scenario 2, where both VaR and systemic risk forecasts are considered accurate in the second model, the tests demonstrate higher sensitivity, yielding rejection frequencies of 94%(99.5%) for $S^{(\text{VaR+CoVaR})}$ and 93.5 %(100%) for $S^{(\text{VaR+CoVaR+CoES})}$ in the sample size of 500 (1000). In this scenario, introducing an additional small difference in VaR forecasts results in a higher rejection frequency than Scenario 2 in Table (4.8). Specifically, when only SR forecasts are considered superior, the rejection frequency for a sample size of 500 (1000) is 80% (94%) in the $S^{(\text{VaR+CoVaR})}$. However, when both VaR and SR are considered accurate in the second model, the rejection frequency increases to 94% (99.5%) in $S^{(\text{VaR+CoVaR})}$. A similar pattern is observed in the $S^{(\text{VaR+CoVaR+CoES})}$ combination.

Moreover, comparing our test with the one conducted by Fissler and Hoga (2023), we observe that our test shows higher power, with a rejection frequency of 94% compared to 88.5% for a sample size of 500. However, in sample size 1000, the rejection frequency of our test is slightly higher, with 99.5% compared to 98.5% of Fissler and Hoga (2023) test.

Overall, although we introduce a small difference in the predictive accuracy of the VaR forecasts, our tests are capable of detecting this differentiation, even in cases where the systemic risk forecasts demonstrate comparable forecasting. This indicates that our tests have a good sensitivity to detect marginal differences in model forecasting. Additionally, our tests effectively capture differences in forecasting accuracy, showing increased rejection rates in situations where both VaR and systemic risk forecasts are considered accurate in the second model.

Sample Size	Under the assumption, $r_{1,t}^{\text{VaR}} = \hat{r}_{1,t}^{\text{VaR}}$ and $r_{2,t}^{\text{VaR}} = \hat{r}_{2,t}^{\text{VaR}*}$				
	$\text{S}^{(\text{VaR+SR})}$ **score results**			$\text{S}^{(\text{VaR,SR})}$ **score results**	
	Systemic Risk Forecasts	(VaR+ CoVaR)	(VaR+ CoVaR+ CoES)	(VaR, CoVaR)	(VaR, CoVaR, CoES)
500	$r_{1,t}^{\text{SR}} = \hat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \hat{r}_{2,t}^{\text{SR}}$	44	36.5	39.5	39
	$r_{1,t}^{\text{SR}} = \hat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \hat{r}_{2,t}^{\text{SR}*}$	94	93.5	88.5	91
1000	$r_{1,t}^{\text{SR}} = \hat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \hat{r}_{2,t}^{\text{SR}}$	72	64.5	61.5	61
	$r_{1,t}^{\text{SR}} = \hat{r}_{1,t}^{\text{SR}}$ $r_{2,t}^{\text{SR}} = \hat{r}_{2,t}^{\text{SR}*}$	99.5	100	98.5	98.5

TABLE 4.9: Rejection Frequencies (%) of H_0 for $r_{1,t} = (r_{1,t}^{\text{VaR}}, r_{1,t}^{\text{SR}})$ and $r_{2,t} = (r_{2,t}^{\text{VaR}}, r_{2,t}^{\text{SR}})$, assuming the VaR forecasts of the second model is better than the first model, $r_{1,t}^{\text{VaR}} = \hat{r}_{1,t}^{\text{VaR}}$ and $r_{2,t}^{\text{VaR}} = \hat{r}_{2,t}^{\text{VaR}*}$. The last two columns of the table present the results of Fissler and Hoga score function $\text{S}^{(\text{VaR,SR})}$.
For $\text{S}^{(\text{VaR+SR})}$, we obtain $\mathcal{T}_n^{OS.uni}$ to test : $H_0 : \text{E}\left[\bar{d}_n^{uni}\right] \leq 0$. While for $\text{S}^{(\text{VaR,SR})}$, we obtain One and a Half-Sided test, $\mathcal{T}_n^{OS}$, to examine
$H_0^{\preceq_{\text{lex}}} : \text{E}\left[\bar{d}_{1n}\right] = 0$ and $\text{E}\left[\bar{d}_{2n}\right] \leq 0$.
The (*) indicates accurate forecasts in the second model.

The simulation study yielded several main results as follows:
Firstly, the power of our tests increases significantly as the sample size (n) grows larger. Secondly, detecting differences in the predictive ability of two models becomes more easier when considering scenarios where both VaR and SR forecasts are accurate for one model while being misspecified for the other. Thirdly, despite introducing only a small difference in the predictive accuracy of the VaR forecasts, our test has the power to detect this differentiation, even in cases where the systemic risk forecasts demonstrate a comparable performance across models. This finding highlighted the good sensitivity of our tests to detect marginal differences in model forecasting. Furthermore, across the various scenarios examined in this study, our tests generally demonstrate superior performance compared to those proposed by Fissler and Hoga (2023) in evaluations incorporating both (VaR + CoVaR) and (VaR + CoVaR + CoES). Finally, comparisons involving CoVaR and CoES exhibit higher power compared to those depending on CoVaR only, potentially attributed to the richer informational content offered by the CoES component, increasing the overall power of the analysis.

4.4.2 Empirical Application

We analyze daily log-losses on the S&P 500 and DAX 30 from 2000 to 2020, denoted by $X_{-r+1}, \ldots, X_n$ and $Y_{-r+1}, \ldots, Y_n$ respectively. The data are sourced from `www.wsj.com/market-data/quotes` (using the ticker symbols SPX and DAX). We calculate the logarithmic return of the stock index
$Z_t = -\log\left(P_{Z,t}/P_{Z,t-1}\right) (Z \in \{X, Y\})$, where $P_{Z,t}$ is the stock index value at time t,

and $P_{Z,t-1}$ is the stock index value at the previous time step $(t-1)$. The number of observations where the data on both indexes are available is 5193 observations. Throughout the analysis, we fix the risk levels α and β at 0.95, and we set r to 1000, which denotes the length of the rolling window. We compare the (VaR, CoVaR, CoES) forecast of the rolling window for the series $\{(X_t, Y_t)\}_{t=1,\dots,n}$. By choosing X (S&P 500) and Y (DAX 30) for the analysis, we focus on assessing the risk of large losses in the DAX 30 under the scenario where S&P 500 is in distress.
In short-term risk management, conditional risk measures that take into account the available information up to the previous time period are considered more informative than unconditional risk measures. These conditional risk measures, denoted as $F(X_t, Y_t|\mathfrak{F}_{t-1}(x,y))$, rely on the conditional distribution of the random variables (X_t, Y_t) given the information set generated by past observations(X_{t-1}, Y_{t-1}), (X_{t-2}, Y_{t-2}), and so on, as well as any exogenous information available up to time t-1. The forecasted conditional risk measures include the $\text{VaR}_t(X_t) = \text{VaR}_\beta(F_{X_t|\mathfrak{F}_{t-1}})$, $\text{CoVaR}_t(Y_t|X_t) = \text{CoVaR}_{\alpha|\beta}(F_{(X_t,Y_t)|\mathfrak{F}_{t-1}})$, and $\text{CoES}_t(Y_t|X_t) = \text{CoES}_{\alpha|\beta}(F_{(X_t,Y_t)|\mathfrak{F}_{t-1}})$. The dependence on risk levels α and β is omitted for simplicity in notation.

For risk measure forecasting, we employ two models. The first approach involves fitting separate GARCH(1,1) specifications to the time series X_t and Y_t. For modelling the dependence between the innovations $(\varepsilon_{x,t}, \varepsilon_{y,t})$, we adopt a Gaussian copula density with time-varying correlation parameter $\rho_t \in (-1, 1)$, where the correlation dynamics follow the GAS process. Conditional on past information $\mathfrak{F}_{t-1}$, the innovations are assumed to follow a Gaussian copula with correlation ρ_t evolving according to GAS dynamics. The same models implemented in the simulation with ($\vartheta = \infty$ in (4.25)), models' details provided in Subsection (4.4.1.1).

In the second forecasting model, we utilise the GJR-GARCH(1,1) proposed by Glosten et al. (1993) with t-copula driven by GAS for the dependence model. The GJR-GARCH(1,1) specifications are:

$$\begin{aligned} X_t &= \sigma_{x,t}\varepsilon_{x,t}, \quad \sigma^2_{x,t} = \omega_x + \alpha_x X^2_{t-1} + \beta_x \sigma^2_{x,t-1} + \gamma_x X^2_{t-1} I_{(X_{t-1}<0)}, \\ Y_t &= \sigma_{y,t}\varepsilon_{y,t}, \quad \sigma^2_{y,t} = \omega_y + \alpha_y Y^2_{t-1} + \beta_y \sigma^2_{y,t-1} + \gamma_y Y^2_{t-1} I_{(Y_{t-1}<0)}. \end{aligned} \tag{4.29}$$

Here, γ_x and γ_y represent additional parameters that accommodate varying impacts of positive and negative shocks on volatility. The dependence between $(\varepsilon_{x,t}, \varepsilon_{y,t})$ is modelled using a t-copula density with time-varying correlation driven via GAS dynamics as in(4.25). Figure (4.3) illustrates the two models employed in the risk measure forecasting.

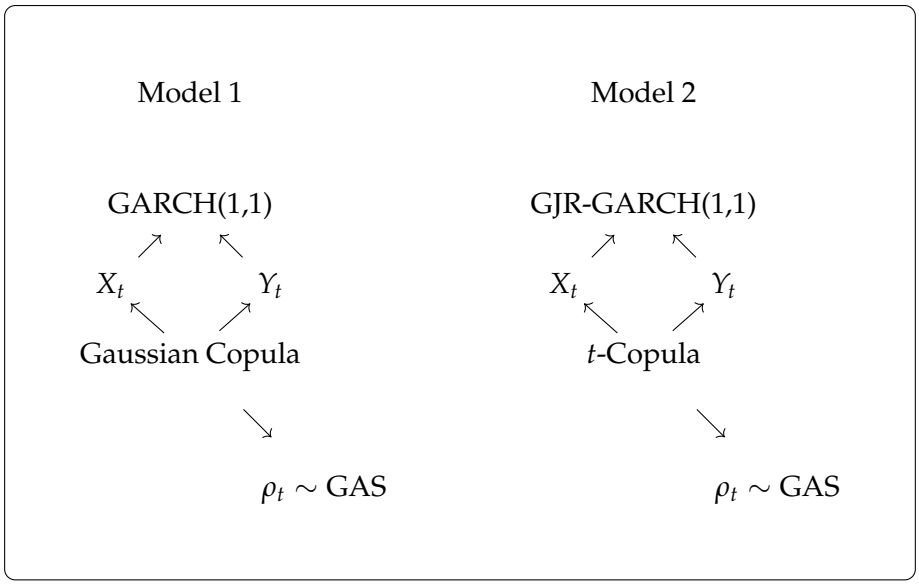


FIGURE 4.3: The Two Models Employed for Risk Measure Forecasting

For both models, following Fissler and Hoga (2023), we do not make assumptions about the probability distribution of the error terms $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$. Instead, in the estimation stage, we use the Gaussian quasi-maximum likelihood estimator (QMLE) for the GARCH-type marginal models, which is robust to other error distributions. Furthermore, when forecasting risk, we use the empirical cumulative distribution functions (cdfs) of the error terms (F_x and F_y) instead of assuming a specific theoretical distribution. The details regarding the risk prediction calculation are introduced in (4.4.1.2).

The forecasting comparisons are conducted between the GARCH-Gauss model with Gaussian copula and the GJR-GARCH with t-copula by evaluating their ability to predict extreme events in the DAX 30 index with the scenario that the S&P 500 index experiences VaR violations, and it investigates how well the models forecast extreme losses in the DAX index during these events. The comparison involves calculating CoVaR and CoES forecasts for both models, and it involves two series of rolling-window predictions $\boldsymbol{r}_{1,t}$ and $\boldsymbol{r}_{2,t}$ obtained by refitting the models daily with a rolling window of length $r = 1000$. This generates n=4193 forecasts from the GARCH model with Gaussian copula, denoted as $\{\boldsymbol{r}_{1,t}\}_{t=1,\ldots,n}$, and from the GJR-GARCH with t-copula, denoted as $\{\boldsymbol{r}_{2,t}\}_{t=1,\ldots,n}$. We consider $\boldsymbol{r}_{1,t}$ as a benchmark forecasts which are to be better by $\boldsymbol{r}_{2,t}$.

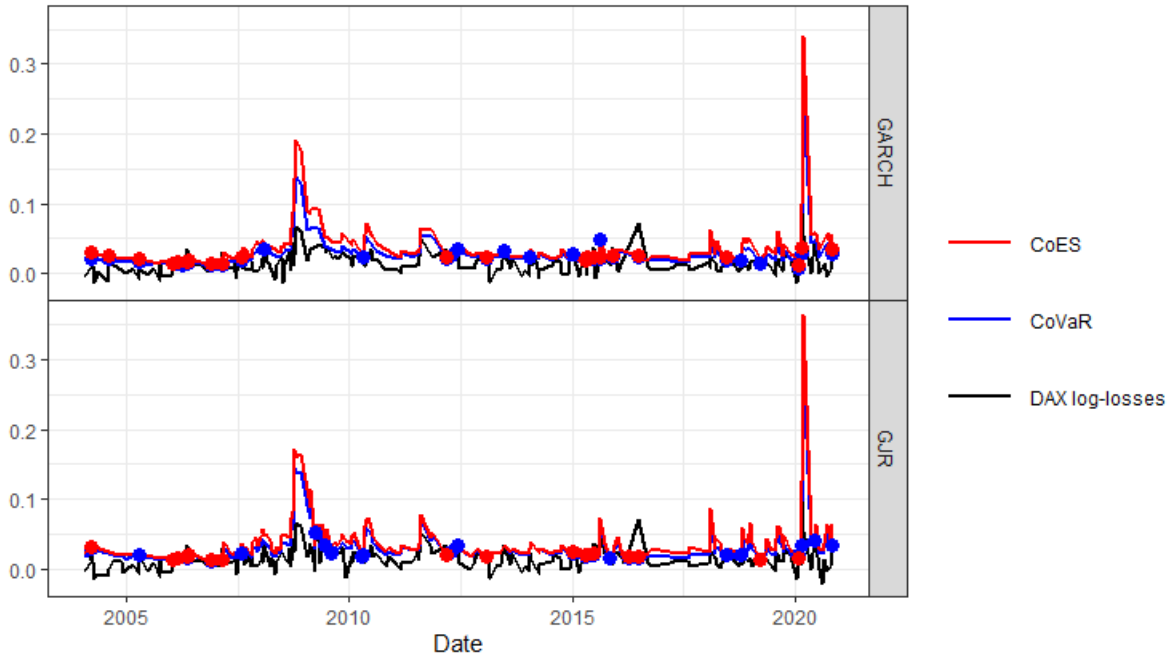


FIGURE 4.4: Top: GARCH model with Gaussian copula forecasting CoVaR(CoES) as the blue(red) line along with DAX log losses on days when the S&P 500 exceeds its VaR forecast. Bottom: similar to the top but with GJR-GARCH with t-copula.

In Figure (4.4), we visually represent the specific periods when VaR violations occurred in the S&P 500 index and then examine how well the two models forecast the corresponding losses(CoVaR and CoES) in the DAX index during those times. Here, the GARCH model with Gaussian copula is provided in the top panel, while the bottom panel corresponds to GJR-GARCH with t-copula. The black line represents the DAX log losses on days when the S&P 500 exceeds its VaR forecast. As we can observe, the black lines between the upper and lower panels vary slightly, and this is because of the distinct marginal models, which lead to varying VaR forecasts.

As previously mentioned, we consider the GARCH model with Gaussian copula as our benchmark and to test the Null Hypothesis H_0 for the following risk measures:(VaR, CoVaR) and (VaR, CoVaR, CoES). We anticipate that the GJR-GARCH with the t-copula model to have a better risk forecast by having lower scores than the GARCH model with the Gaussian copula, potentially leading to a rejection of the H_0. This anticipation is because of the informal comparison that has been conducted by in Fissler and Hoga (2023), which shows higher non-exceedance frequencies in both (VaR, CoVaR) and (VaR, CoVaR, CoES) GJR-GARCH with the t-copula compared to the GARCH model with Gaussian copula. In addition, substantial empirical evidence and studies favouring GJR-GARCH compared to simple GARCH(1,1), see for instance(Glosten et al. (1993),Brownlees et al. (2011), Liu and Hung (2010)) and GAS-t cupola models (Creal et al. (2013) and Bernardi and Catania (2019)). We perform the tests using $\mathcal{T}_n^{\text{OS.uni}}$ with $S^{(\text{VaR+CoVaR})}$, and $S^{(\text{VaR+CoVaR+CoES})}$, along with (4.28) and

(4.27), which feature 0-homogeneous score differences, and utilising $\widehat{\sigma}_n$ from (4.17); details are in Subsection 4.4.1.3 and the supplementary materials in Fissler and Hoga (2023). It is frequently advised to utilize scoring functions for VaR and ES forecasts that result in 0-homogeneous score differences, enabling unit-consistent and powerful comparisons (Nolde and Ziegel (2017),Patton et al. (2019)), where the powerful comparison has been illustrated in Fissler and Hoga (2023).
Let $\bar{d}_n^{uni}$ be the score differences that are implemented to compare the two forecasts as described in (4.16). Upon computation of $\bar{d}_n^{uni}$, we observe that it yields positive values, indicating that the GJR-GARCH with t-copula has lower scores for both the $S^{(\text{VaR+CoVaR})}$ and the $S^{(\text{VaR+CoVaR+CoES})}$ forecasts. These score differences are statistically significant at 5% level, with p-values of 0.0055 for the $\mathcal{T}_n^{OS.uni}$ -based Wald test for $S^{(\text{VaR+CoVaR})}$ and 0.0204 for $S^{(\text{VaR+CoVaR+CoES})}$. This indicates that the GJR–GARCH model with t-copula performs well compared to the GARCH model with Gaussian copula.
Furthermore, we extend the analysis to include weights in our joint score functions and investigate the impact of different weight combinations on the statistical significance of the Wald Test results for the score functions $S^{(\text{VaR+CoVaR})}$ and $S^{(\text{VaR+CoVaR+CoES})}$. We examine various weights ranging from 0 to 1 with a step size of 0.1, assigned to the VaR, CoVaR, and CoES components.
Tables (4.10) and (4.11) display the p-value results of the weighted score functions $S^{(\text{VaR+CoVaR})}$ and $S^{(\text{VaR+CoVaR+CoES})}$, respectively. Upon examining the weights and the corresponding p-values of the weighted score function $S^{(\text{VaR+CoVaR+CoES})}$, we observe a significant negative correlation (-0.69) between the weight assigned to VaR and the resulting p-values. This implies that a higher weight assigned to VaR tends to be associated with lower p-values, indicating higher statistical significance.
In contrast, the correlation coefficient between weights on CoVaR and p-values is -0.079, while the weight assigned to CoES exhibits a strong positive correlation (0.77) with the p-values. This suggests that higher weights assigned to CoES are associated with higher p-values, indicating lower statistical significance.
These findings emphasize the critical role of weight assignment in evaluating systemic risk. The strong negative correlation (-0.69) between VaR weight and resulting p-values aligns with the popularity of the t-copula model in empirical work.
Moreover, we provide visual representations of the p-values, and the weighted score functions $S^{(\text{VaR+CoVaR})}$ and $S^{(\text{VaR+CoVaR+CoES})}$ in Figure (4.5) and (4.6). Through additional consideration of weight assignment, the performance of our proposed score functions $S^{(\text{VaR+CoVaR})}$ and $S^{(\text{VaR+CoVaR+CoES})}$ could further improve, a task we leave for future investigation.

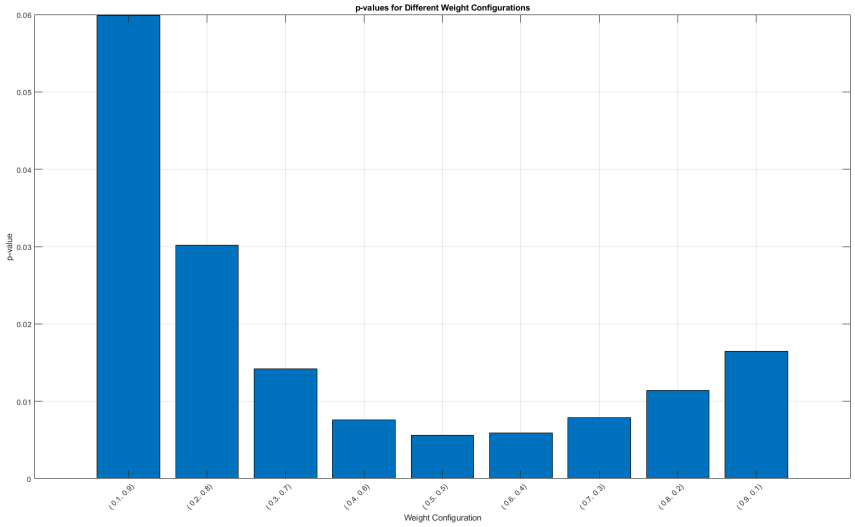


FIGURE 4.5: The bar plot displays the p-value results of the weighted score function $S^{(\text{VaR}+\text{CoVaR}}$. We examine various weights ranging from 0 to 1 with a step size of 0.1 on our score function to study the impact of these weights. Each bar represents a specific weight configuration denoted as (VaR, CoVaR). The x-axis labels indicate the weight configurations in the format (VaR, CoVaR), while the y-axis represents the corresponding p-values obtained from the statistical analysis. This plot provides insights into the significance of risk forecasts generated under various weight combinations, aiding in the evaluation of their predictive performance. Additionally, the statistical significance of the score differences at a 5% level is assessed using the $\mathcal{T}_n^{OS.uni}$ -based Wald test for $S^{(\text{VaR}+\text{CoVaR})}$. Low p-values indicate that the GJR–GARCH model with t-copula performs well compared to the GARCH model with Gaussian copula.

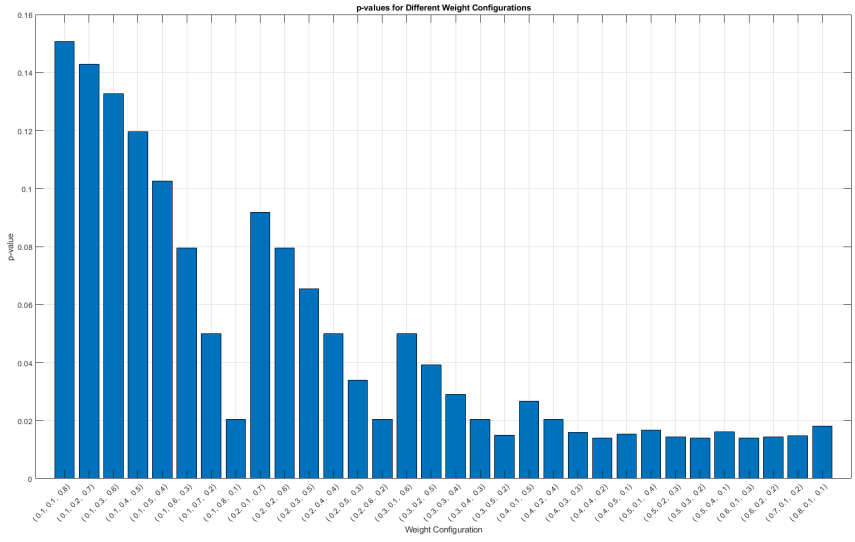


FIGURE 4.6: The bar plot displays the p-value results of the weighted score function $S^{(\text{VaR+CoVaR+CoES})}$. We examine various weights ranging from 0 to 1 with a step size of 0.1 on our score function to study the impact of these weights. Each bar represents a specific weight configuration denoted as (VaR, CoVaR, CoES). The x-axis labels indicate the weight configurations in the format (VaR, CoVaR, CoES), while the y-axis represents the corresponding p-values obtained from the statistical analysis. This plot provides insights into the significance of risk forecasts generated under various weight combinations, aiding in the evaluation of their predictive performance. Additionally, the statistical significance of the score differences at a 5% level is assessed using the $\mathcal{T}_n^{OS.uni}$ -based Wald test for $S^{(\text{VaR+CoVaR+CoES})}$. Low p-values indicate that the GJR–GARCH model with t-copula performs well compared to the GARCH model with Gaussian copula.

p-value	Weight Assignment
0.0599	$\gamma = 0.1$, $\delta = 0.9$
0.0302	$\gamma = 0.2$, $\delta = 0.8$
0.0142	$\gamma = 0.3$, $\delta = 0.7$
0.0076	$\gamma = 0.4$, $\delta = 0.6$
0.0056	$\gamma = 0.5$, $\delta = 0.5$
0.0059	$\gamma = 0.6$, $\delta = 0.4$
0.0079	$\gamma = 0.7$, $\delta = 0.3$
0.0114	$\gamma = 0.8$, $\delta = 0.2$
0.0165	$\gamma = 0.9$, $\delta = 0.1$

TABLE 4.10: The table presents the impact of different combinations of weights on the statistical significance of the Wald Test results for the score function $S^{(\text{VaR+CoVaR})}$. Each row represents a combination of weights for VaRVaR (γ) and CoVaR δ and the Weights range from 0 to 1 with a step size of 0.1. Additionally, the statistical significance of the score differences at a 5% level is assessed using the $\mathcal{T}_n^{OS.uni}$ -based Wald test for $S^{(\text{VaR+CoVaR})}$, with low p-values indicating superior performance of the GJR–GARCH model with t-copula compared to the GARCH model with Gaussian copula.

p-value	Weight Assignment
0.1508	$\gamma = 0.1$,$\delta = 0.1$, $\lambda = 0.8$
0.1428	$\gamma = 0.1$,$\delta = 0.2$, $\lambda = 0.7$
0.1327	$\gamma = 0.1$,$\delta = 0.3$, $\lambda = 0.6$
0.1197	$\gamma = 0.1$,$\delta = 0.4$, $\lambda = 0.5$
0.1025	$\gamma = 0.1$,$\delta = 0.5$, $\lambda = 0.4$
0.0795	$\gamma = 0.1$,$\delta = 0.6$, $\lambda = 0.3$
0.0499	$\gamma = 0.1$,$\delta = 0.7$, $\lambda = 0.2$
0.0205	$\gamma = 0.1$,$\delta = 0.8$, $\lambda = 0.1$
0.0919	$\gamma = 0.2$,$\delta = 0.1$, $\lambda = 0.7$
0.0795	$\gamma = 0.2$,$\delta = 0.2$, $\lambda = 0.6$
0.0655	$\gamma = 0.2$,$\delta = 0.3$, $\lambda = 0.5$
0.0499	$\gamma = 0.2$,$\delta = 0.4$, $\lambda = 0.4$
0.034	$\gamma = 0.2$,$\delta = 0.5$, $\lambda = 0.3$
0.0205	$\gamma = 0.2$,$\delta = 0.6$, $\lambda = 0.2$
0.0499	$\gamma = 0.3$,$\delta = 0.1$, $\lambda = 0.6$
0.0392	$\gamma = 0.3$,$\delta = 0.2$, $\lambda = 0.5$
0.029	$\gamma = 0.3$,$\delta = 0.3$, $\lambda = 0.4$
0.0205	$\gamma = 0.3$,$\delta = 0.4$, $\lambda = 0.3$
0.015	$\gamma = 0.3$,$\delta = 0.5$, $\lambda = 0.2$
0.0267	$\gamma = 0.4$,$\delta = 0.1$, $\lambda = 0.5$
0.0205	$\gamma = 0.4$,$\delta = 0.2$, $\lambda = 0.4$
0.016	$\gamma = 0.4$,$\delta = 0.3$, $\lambda = 0.3$
0.0139	$\gamma = 0.4$,$\delta = 0.4$, $\lambda = 0.2$
0.0153	$\gamma = 0.4$,$\delta = 0.5$, $\lambda = 0.1$
0.0167	$\gamma = 0.5$,$\delta = 0.1$, $\lambda = 0.4$
0.0144	$\gamma = 0.5$,$\delta = 0.2$, $\lambda = 0.3$
0.0139	$\gamma = 0.5$,$\delta = 0.3$, $\lambda = 0.2$
0.0162	$\gamma = 0.5$,$\delta = 0.4$, $\lambda = 0.1$
0.0139	$\gamma = 0.6$,$\delta = 0.1$, $\lambda = 0.3$
0.0143	$\gamma = 0.6$,$\delta = 0.2$, $\lambda = 0.2$
0.0148	$\gamma = 0.7$,$\delta = 0.1$, $\lambda = 0.2$
0.018	$\gamma = 0.8$,$\delta = 0.1$, $\lambda = 0.1$

TABLE 4.11: The table presents the impact of different weight combinations on the statistical significance of the Wald Test results for the score function $S^{(\text{VaR+CoVaR+CoES})}$. Each subtable pair represents fixed values of the weights assigned to VaR (γ) while varying the weights for CoVaR (δ) and CoES (λ). The weights range from 0 to 1 with a step size of 0.1. Additionally, the statistical significance of the score differences at a 5% level is assessed using the $\mathcal{T}_n^{OS.uni}$ -based Wald test for $S^{(\text{VaR+CoVaR})+\text{CoES})}$, with low p-values indicating superior performance of the GJR–GARCH model with t-copula compared to the GARCH model with Gaussian copula.

4.5 Discussion

4.5.1 Sensitivity to Risk Levels

Our study primarily focuses on risk levels of $\alpha = \beta = 0.95$ following Fissler and Hoga (2023) setting, which is a commonly used setting in practice. However, it is important to acknowledge the sensitivity of the results to these tuning parameters. To examine the sensitivity of our proposed score function $S^{(\text{VaR+SR})}$ to varying risk levels, we conducted additional simulation analyses across a range of α and β values in 4.3.3.4, where we investigate the elicitability of our proposed score function through numerical simulations. The results, presented in Tables 4.5 and 4.6, and Figures 4.7, 4.8, 4.9, 4.10 and 4.11, demonstrate that our score function exhibits consistent performance and robustness across different risk level combinations.
Furthermore, in the empirical application section 4.4.2, we explored the impact of different weight assignments on the statistical significance of the Wald test results for the score functions $S^{(\text{VaR+CoVaR})}$ and $S^{(\text{VaR+CoVaR+CoES})}$. This analysis, while not directly related to the sensitivity to risk levels, provided insights into the sensitivity of the results to the prioritization of individual risk measures within the combined score function. While the choice of risk levels can influence the specific numerical values obtained, our overall findings and conclusions regarding the effectiveness of our proposed score function in jointly assessing VaR and systemic risk measures remain valid across a reasonable range of risk level settings.

4.5.2 Limitations and Shortcomings

While our proposed univariate score function $S^{(\text{VaR+SR})}$ offers a novel approach to jointly assessing VaR and SR, it is important to acknowledge potential limitations and shortcomings associated with considering a combination statistic. One key limitation lies in the interpretability and the ability to discern the relative importance of individual risk components within the combined score. By combining multiple risk measures into a single score function, the specific contributions or relative importance of each individual risk measure may be obscured, potentially reducing the interpretability of the overall score.
To mitigate this issue and gain further insights into the individual performance of the models, our proposed score function allows for the inclusion of weights, enabling users to prioritize one risk measure over the other based on their specific requirements. However, this flexibility in weighting introduces another challenge – determining optimal weighting strategies. The choice of weights can significantly influence the overall score and its sensitivity to changes in individual risk measures, making the determination of appropriate weighting strategies context-dependent and potentially challenging.

It is important to note that the proposed test statistic aims to provide an overall assessment of a model's performance in capturing both individual and systemic risk simultaneously. While it is possible that one model outperforms the other in different risk measures, the joint evaluation of VaR and SR is crucial in many practical applications, such as risk management and regulatory oversight. A model that performs well in both measures may be preferred over a model that excels in one measure but performs poorly in the other. When the test indicates a significant difference between two models, it suggests that one model has a better overall performance in capturing both VaR and SR, although it does not necessarily imply superiority in both measures individually. Despite these potential limitations and shortcomings, our proposed approach provides a unique perspective on risk assessment by jointly considering VaR and SR measures. Future research could focus on addressing these limitations, such as developing robust weighting strategies, exploring alternative combination methods, or incorporating additional risk measures to enhance the comprehensiveness of the risk assessment framework.

4.6 Conclusion

In conclusion, our proposed score function, $S^{(\text{VaR}+\text{SR})}$, offers a simpler and straightforward solution that addresses the limitation observed in Fissler and Hoga (2023) approach. Specifically, Fissler and Hoga (2023) requires a constraint on the evaluation and comparison of risk measures by stating that the assessment is confined to either VaR or SR contingent upon the equality of VaR across the models under consideration. This indicates that Fissler and Hoga (2023) approach fails to assess both risk measures on an equal footing. However, defining our univariate score function as the sum of the marginal/conditional score functions for forecasting VaR and SR without the need to prioritize one over the other based on the equality of VaR across models goes beyond the backtesting implemented by Fissler and Hoga (2023). Moreover, the risk of employing identification functions that are not strictly defined for backtesting is examined by conducting a comparative analysis of our identification function along with the one-dimensional identification function introduced by Banulescu-Radu et al. (2021) in addition to the two-dimensional identification function introduced by Fissler and Hoga (2023). Specifically, we evaluated the power of the three identification functions in identifying misspecified systemic risk forecasts under various sample sizes and distributional scenarios. A misspecified risk measure is defined as a measure that fails to capture the true tail risk dynamics over time accurately. Through our analysis, we demonstrated the superiority of our proposed identification function compared to Banulescu-Radu et al. (2021) identification function. In Particular, the backtest proposed by Banulescu-Radu et al. (2021) exhibits a complete loss of power in distinguishing between correct and misspecified forecasts.

In contrast, our identification function, aligning with Fissler and Hoga (2023) identification function, successfully identifies misspecified forecasts almost with certainty with almost 100% across different distributional assumptions and sample sizes. These results underscore the power of our proposed identification function in detecting misspecified systemic risk forecasts, outperforming Banulescu-Radu et al. (2021) method. Furthermore, through a comprehensive simulation analysis, we illustrate the elicitability of our proposed score function. Moreover, we showed through a Monte Carlo simulation the following: Firstly, the power of our tests increases significantly as the sample size (n) grows larger. Secondly, our tests show a good size control with a rejection frequency close to a 5% significance level, even for a small sample size ($n = 500$). Thirdly, detecting differences in the predictive ability of two models becomes more easier when considering scenarios where both VaR and SR forecasts are accurate for one model while being misspecified for the other. Fourth, despite introducing only a small difference in the predictive accuracy of the VaR forecasts, our test has the power to detect this differentiation, even in cases where the systemic risk forecasts demonstrate a comparable performance across models. This finding highlighted the good sensitivity of our tests to detect marginal differences in model forecasting. Furthermore, across the various scenarios examined in this study, our tests generally demonstrate slightly higher power compared to those proposed by Fissler and Hoga (2023) in evaluations incorporating both (VaR + CoVaR) and (VaR + CoVaR + CoES). Finally, comparisons involving CoVaR and CoES exhibit higher power compared to those depending on CoVaR only, potentially attributed to the richer informational content offered by the CoES component, increasing the overall power of the analysis. Furthermore, our analysis extends to include weights into our proposed joint score functions, offering flexibility in prioritizing various risk measures such as VaR and CoVaR by adjusting the allocated weights for each component according to our desired emphasis. We observed a significant negative correlation (-0.69) between VaR weight and resulting p-values. This suggests that assigning a higher weight to VaR tends to be associated with lower p-values, indicating higher statistical significance. On the other hand, a strong positive correlation (0.77) was observed for CoES weight and resulting p-values. These findings highlight the critical role of weight assignment in evaluating systemic risk forecasting. Through additional consideration of weight assignment, the performance of our proposed score function could further improve, a task we leave for future investigation.

Appendix

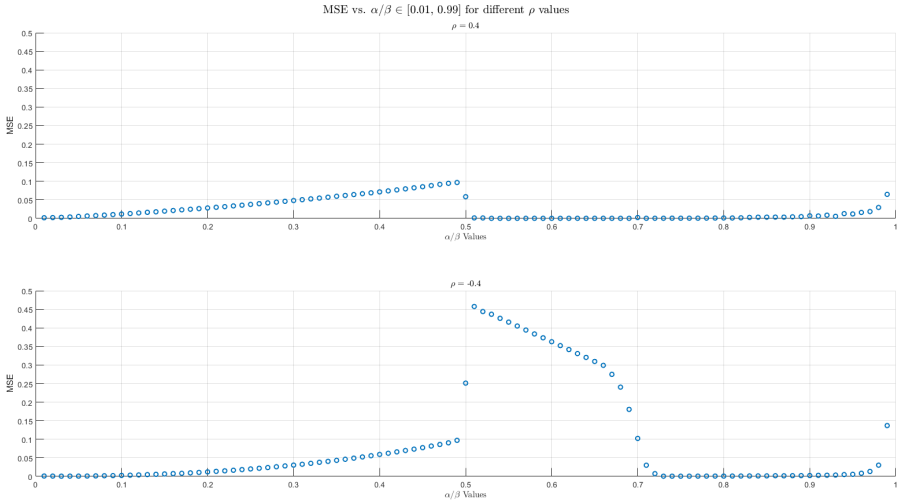


FIGURE 4.7: The plot illustrates the Mean Squared Error (MSE) values of the CoVaR estimation component in the score function $S^{(\text{VaR+CoVaR})}$ across a range of α/β values (0.01 to 0.99) for different correlation coefficients ρ (0.4 and -0.4). Each subplot represents a different ρ value: the first subplot corresponds to $\rho = 0.4$, while the second subplot corresponds to $\rho = -0.4$. The MSE is computed over 10000 simulations, with each simulation generating data and computing MSE over 100 iterations. This involves running the CoVaR estimation algorithm multiple times to assess its performance.

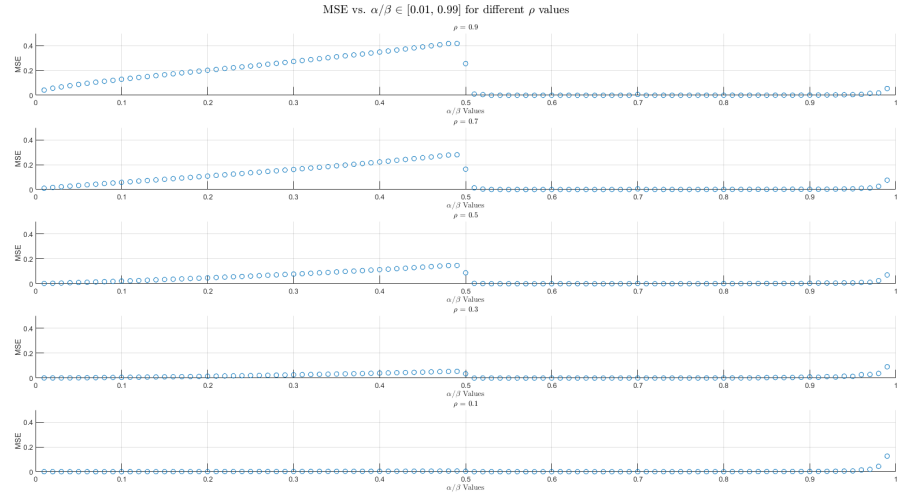


FIGURE 4.8: The plot illustrates the Mean Squared Error (MSE) values of the CoVaR estimation component in the score function $S^{(\text{VaR+CoVaR})}$ across a range of α/β values (0.01 to 0.99) for different correlation coefficients ρ (0.9, 0.7, 0.5, 0.3 and 0.1). Each subplot represents a different ρ value. The MSE is computed over 10000 simulations, with each simulation generating data and computing MSE over 100 iterations. This involves running the CoVaR estimation algorithm multiple times to assess its performance.

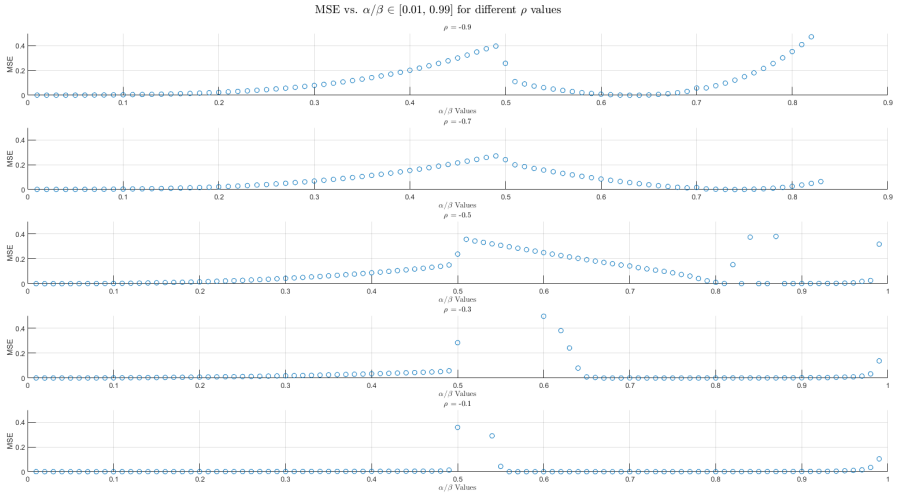


FIGURE 4.9: The plot illustrates the Mean Squared Error (MSE) values of the CoVaR estimation component in the score function $S^{(\text{VaR+CoVaR})}$ across a range of α/β values (0.01 to 0.99) for different correlation coefficients ρ (-0.9, -0.7, -0.5, -0.3 and -0.1). Each subplot represents a different ρ value. The MSE is computed over 10000 simulations, with each simulation generating data and computing MSE over 100 iterations. This involves running the CoVaR estimation algorithm multiple times to assess its performance.

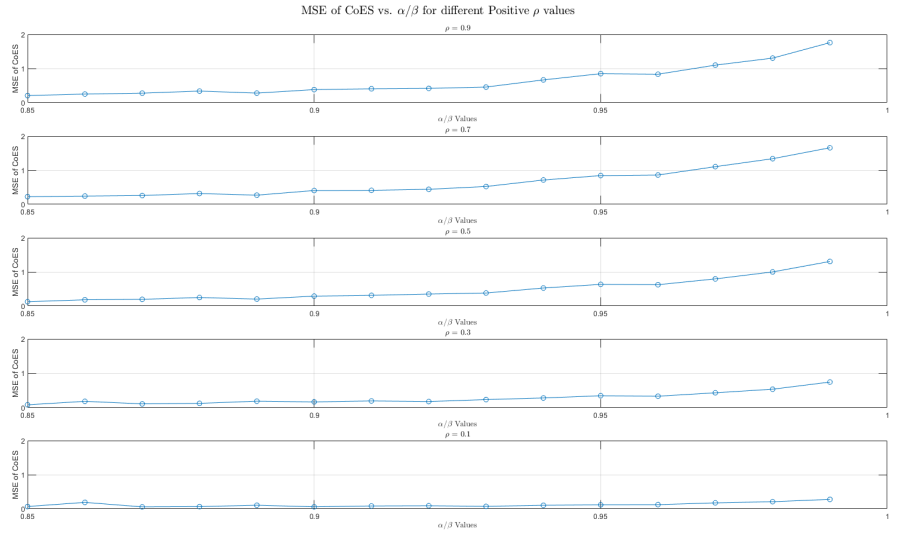


FIGURE 4.10: The plot illustrates the Mean Squared Error (MSE) values of the CoES estimation component in the score function $S^{(\text{VaR+CoVaR+CoES})}$ across a range of α/β values (0.01 to 0.99) for different correlation coefficients ρ (0.9, 0.7, 0.5, 0.3 and 0.1). Each subplot represents a different ρ value. The MSE is computed over 10000 simulations, with each simulation generating data and computing MSE over 100 iterations. This involves running the CoES estimation algorithm multiple times to assess its performance.

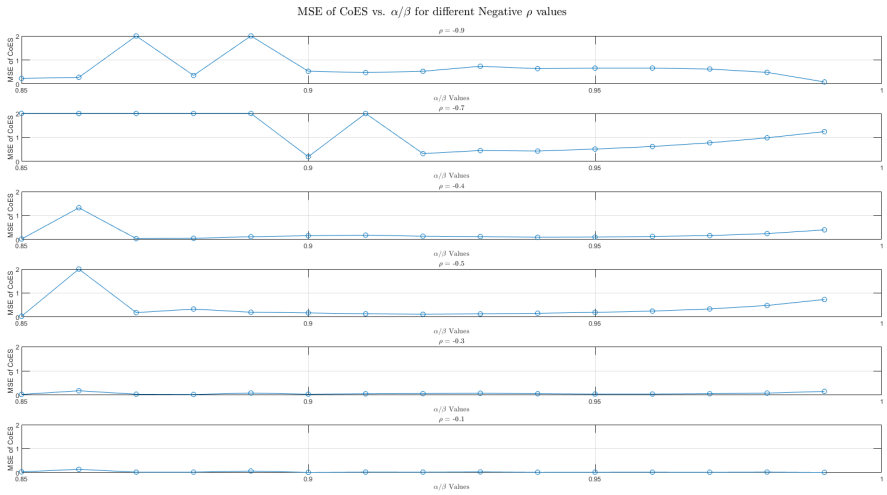


FIGURE 4.11: The plot illustrates the Mean Squared Error (MSE) values of the CoES estimation component in the score function $S^{(\text{VaR+CoVaR+CoES})}$ across a range of α/β values (0.01 to 0.99) for different correlation coefficients ρ (-0.9, -0.7, -0.5, -0.3 and -0.1). Each subplot represents a different ρ value. The MSE is computed over 10000 simulations, with each simulation generating data and computing MSE over 100 iterations. This involves running the CoES estimation algorithm multiple times to assess its performance.

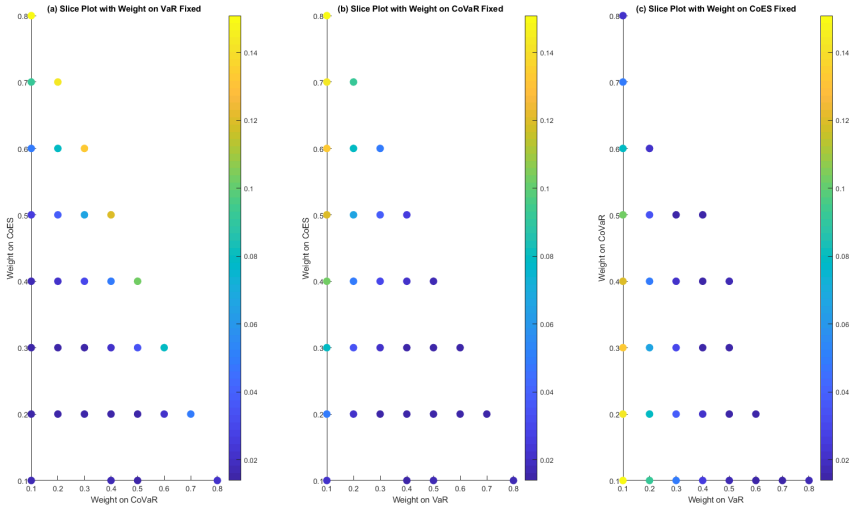


FIGURE 4.12: The slice plots illustrate the impact of different combinations of weights on the statistical significance of the Wald Test results for the score function $S^{(\text{VaR+CoVaR+CoES})}$. Each subplot represents a fixed value of one weight component, while the other two weight components vary. (a) Slice plot with weight on VaR fixed, (b) Slice plot with weight on CoVaR fixed, and (c) Slice plot with weight on CoES fixed. Each marker's position represents a combination of weight values. Marker colour indicates the Wald Test p-value associated with each combination, with darker colours representing lower p-values.

Chapter 5

Conclusions and Future Extension

Throughout this thesis, we have undertaken a journey to address crucial challenges in statistical analysis and financial modelling, spanning robust time series analysis, dynamic portfolio construction, and comparative backtesting of forecasting models for risk assessment. Each chapter has contributed unique methodologies and insights, together advancing the understanding and practice of statistical finance.

Chapter 2 introduced the Robust Model Averaging Marginal Regressions (RMAMAR) procedure as an extension of the Model Averaging MArginal Regression (MAMAR) methodology proposed by Li et al. (2015), RMAMAR aims to mitigate the curse of dimensionality and outlier sensitivity in nonparametric time series analysis. By combining one-dimensional marginal regression functions through robust M-estimation techniques, RMAMAR overcomes the limitations of least squares methods and effectively addresses the curse of dimensionality.
To evaluate the performance of the RMAMAR method, we conducted Monte Carlo simulations and real data analysis. In the simulation study, we investigated our approach under three scenarios and across various distributional assumptions, including t-distributions and normal distributions. To assess the sensitivity of our approach, we also implemented different parameter values in our simulated models. The simulation study results showed that, compared to alternative methods across different distributional scenarios such as MAMAR and nonlinear additive models, our RMAMAR approach can precisely lead to satisfactory prediction performances within purely nonlinear additive AR model structures, even with limited sample sizes.
In the real data analysis, we applied our RMAMAR approach to rainfall data characterized by a long tail to the right. Our analysis exhibited the ability of RMAMAR to uncover nonlinear lag effects compared to the linear Autoregressive (AR) model. Additionally, RMAMAR displayed the lowest Mean Absolute Error (MAE), both in-sample and out-of-sample, demonstrating an improved predictive ability compared to the different models conducted in the analysis.

Overall, our study highlights the advantages of the proposed RMAMAR method in understanding and modelling time series data that contain outliers and nonlinear dependencies. RMAMAR offers a flexible and robust approach for analyzing complex time series datasets in various applications by combining robust estimation techniques in model averaging with nonparametric regression.

In Chapter 3, we explored a dynamic portfolio with a series of conditioning variables under quantile preferences (QP), a departure from traditional Expected Utility frameworks. Our approach, which incorporates large dimensional conditioning variables and model averaging techniques, provides practical insights into constructing dynamic optimal portfolios that align with investors' downside preferences. In our proposed approach, we initially estimate the marginal optimal portfolio choice under QP by solving the conditional quantile function for each individual conditioning variable. Subsequently, through a model averaging process, we combine the optimal weights obtained from the previous step to form the overall dynamic optimal portfolio. The empirical evidence has underscored the effectiveness of this methodology in capturing downside risk preferences during unstable market periods, thereby highlighting its practical relevance for risk-aware investment strategies in both small and large dimensions.

As a future work, in Stage 2, where the estimation of the optimal model averaging weights of quantiles, $\boldsymbol{a}_o(\tau)$ are obtained by maximizing:

$$\begin{aligned} \boldsymbol{a}_o(\tau) &= \arg\max_{\boldsymbol{a}} E\left\{ u\left[\sum_{j=1}^{p} a_j(X_j)\; w_{\mathrm{op},j}^{T}(X_{j,t-1})\mathbf{R}_t \right] \right\}, \\ \text{s.t.} \quad & \sum_{j=1}^{p} a_j(X_j) = 1, \quad a_j \geq 0. \end{aligned} \tag{5.1}$$

one may consider replacing the expected utility maximization in Eq(5.1) with a quantile-based approach as follows

$$\begin{aligned} \boldsymbol{a}_o(\tau) &= \arg\max_{\boldsymbol{a}} Q_\tau\left\{ u\left[\sum_{j=1}^{p} a_j(X_j)\; w_{\mathrm{op},j}^{T}(X_{j,t-1})\mathbf{R}_t \right] \right\}, \\ \text{s.t.} \quad & \sum_{j=1}^{p} a_j(X_j) = 1, \quad a_j \geq 0. \end{aligned} \tag{5.2}$$

The proposed method in (5.2) aims to improve the robustness and accuracy of model averaging techniques by optimising quantile-based utility functions. Moreover, we can extend its applicability by relaxing the constraint $a_j \geq 0$, thereby allowing for the possibility of negative values for a_j.

Chapter 4 addressed the crucial task of comparing forecasting models for Value at Risk (VaR) and systemic risk (SR) components. our proposed score function, $S^{(\text{VaR+SR})}$, offers a robust solution to address the limitation observed in Fissler and Hoga (2023) backtesting approach. Specifically, Fissler and Hoga (2023) imposes a restrictive condition, by indicating that the two systemic risk forecasts only influence the ranking in the lexicographic order under the condition $E[\bar{d}_{1n}] = 0$. However, by building upon this existing framework, we introduced an alternative univariate score defined as the sum of the marginal/conditional score functions for forecasting VaR and SR without the utilization of the lexicographic order. Our score function provides a simple and straightforward univariate score function without any restrictive assumption, thus overcoming the limitations identified in Fissler and Hoga (2023) approach. Through extensive simulations and empirical analyses we have implemented in this Chapter, we have proven the power of our score function in identifying misspecified forecasts and distinguishing between different risk models, thereby directly contributing to the practical needs of risk management. We also compared our approach with the two-dimensional identification function introduced by Fissler and Hoga (2023) and the one-dimensional identification function introduced by Banulescu-Radu et al. (2021). Specifically, we test the power of the three identification functions under different sample sizes in detecting specific misspecified systemic risk forecasts. The backtesting conducted by Banulescu-Radu et al. (2021) indicates a complete inability to distinguish between correct and misspecified forecasts. Conversely, our identification function aligns with the conclusion drawn by Fissler and Hoga (2023), successfully identifying the misspecified forecasts with near certainty. Furthermore, our analysis extends to include weights in score functions, allowing flexibility in emphasizing different risk measures such as VaR and CoVaR by adjusting the weights assigned to each component according to our desired emphasis. We observed a significant negative correlation (-0.82) between VaR weight and resulting p-values, contrasting with a positive correlation (0.71) for CoES weight. These findings highlight the critical role of weight assignment in evaluating systemic risk. Through additional consideration of weight assignment, we anticipate that the performance of our proposed score function could further improve, a task we leave for future investigation.

In conclusion, the findings presented in this thesis contribute not only to the advancement of statistical finance but also offer practical implications for decision-makers and practitioners navigating the complexities of modern financial markets. By addressing critical challenges in time series analysis, portfolio optimization, and risk assessment, this research offers practical solutions and strategies that can be used to make more informed choices and enhance risk management in the dynamic financial landscape.

References

Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of banking & finance*, 26(7):1487–1503, 2002.

Viral V Acharya, Lasse H Pedersen, Thomas Philippon, and Matthew Richardson. Measuring systemic risk. *The review of financial studies*, 30(1):2–47, 2017.

Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. *Proceedings of the 2nd International Symposium on Information Theory*, pages 267–281, 1973.

DE Allen, RJ Powell, and AK Singh. Quantile regression as a tool for portfolio investment decisions during times of financial distress. *Annals of Financial Economics*, 6(01):1150003, 2011.

Tomohiro Ando and Ker-Chau Li. A model-averaging approach for high-dimensional regression. *Journal of the American Statistical Association*, 109(505):254–265, 2014.

D. F. Andrews, Peter J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey. *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, N.J., 1972.

Miguel A Arcones. Second order representations of the least absolute deviation regression estimator. *Annals of the Institute of Statistical Mathematics*, 50:87–117, 1998.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

Alexandru V Asimit and Russell Gerrard. On the worst and least possible asymptotic dependence. *Journal of Multivariate Analysis*, 144:218–234, 2016.

Anthony Curtis Atkinson. Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis. *(No Title)*, 1985.

Doron Avramov. Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458, 2002.

Gutti Jogesh Babu. Strong representations for lad estimators in linear models. *Probability theory and related fields*, 83(4):547–558, 1989.

ZD Bai, XR Chen, BQ Miao, and C Radhakrishna Rao. Asymptotic theory of least distances estimate in multivariate linear models. *Statistics*, 21(4):503–519, 1990.

ZD Bai, C Radhakrishna Rao, and Y Wu. M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statistica Sinica*, pages 237–254, 1992.

Pierluigi Balduzzi and Anthony W Lynch. Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, 52(1):47–78, 1999.

Guido Baltussen and Gerrit T Post. Irrational diversification: An examination of individual portfolio choice. *Journal of Financial and Quantitative Analysis*, 46(5): 1463–1491, 2011.

Denisa Banulescu-Radu, Christophe Hurlin, Jérémy Leymarie, and Olivier Scaillet. Backtesting marginal expected shortfall and related systemic risk measures. *Management science*, 67(9):5730–5754, 2021.

Nicholas Barberis. Investing for the long run when returns are predictable. *The Journal of Finance*, 55(1):225–264, 2000.

Michelle L Barnes and Anthony Tony W Hughes. A quantile regression analysis of the cross section of stock market returns. 2002.

Ignazio Basile, Pierpaolo Ferrari, et al. *Asset management and institutional investors*. Springer, 2016.

John M Bates and Clive WJ Granger. The combination of forecasts. *Journal of the operational research society*, 20(4):451–468, 1969.

Alexandre Belloni and Victor Chernozhukov. ℓ 1-penalized quantile regression in high-dimensional sparse models. 2011.

David A Belsley, Edwin Kuh, and Roy E Welsch. *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, 2005.

Alain Berlinet, Friedrich Liese, and Igor Vajda. Necessary and sufficient conditions for consistency of m-estimates in regression models with general errors. *Journal of statistical planning and inference*, 89(1-2):243–267, 2000.

Mauro Bernardi and Leopoldo Catania. Switching generalized autoregressive score copula models with application to systemic risk. *Journal of Applied Econometrics*, 34 (1):43–65, 2019.

Mauro Bernardi, Antonello Maruotti, and Lea Petrella. Multiple risk measures for multivariate dynamic heavy–tailed models. *Journal of Empirical Finance*, 43:1–32, 2017.

Benjamin Bobbia, Paul Doukhan, and Xiequan Fan. A review on some weak dependence conditions. 2022.

Denis Bosq. *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media, 2012.

Richard C Bradley. Basic properties of strong mixing conditions. *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 165–192, 1986.

Richard C Bradley. Basic properties of strong mixing conditions. a survey and some open questions. 2005.

Michael W Brandt. Estimating portfolio and consumption choice: A conditional euler equations approach. *The Journal of Finance*, 54(5):1609–1645, 1999.

Christian Brownlees and Robert F Engle. Srisk: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, 30(1):48–79, 2017.

Christian T Brownlees, Robert F Engle, and Bryan T Kelly. A practical guide to volatility forecasting through calm and storm. *Available at SSRN 1502915*, 2011.

Raffaella Calabrese and Silvia Angela Osmetti. A new approach to measure systemic risk: A bivariate copula model for dependent censored data. *European Journal of Operational Research*, 279(3):1053–1064, 2019.

Sebastian Calonico, Matias D Cattaneo, and Max H Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.

John Y Campbell and Luis M Viceira. Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics*, 114(2): 433–495, 1999.

John Y Campbell and Luis M Viceira. Who should buy long-term bonds? *American Economic Review*, 91(1):99–127, 2001.

Luciano de Castro, Antonio F Galvao, Gabriel Montes-Rojas, and Jose Olmo. Portfolio selection in quantile decision models. *Annals of finance*, 18(2):133–181, 2022.

Christopher Chambers. An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(2):335–342, 2009.

Chris Chatfield. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 158(3):419–444, 1995.

Jia Chen, Degui Li, Oliver Linton, and Zudi Lu. Semiparametric dynamic portfolio choice with multiple conditioning variables. *Journal of Econometrics*, 194(2):309–318, 2016.

Jia Chen, Degui Li, Oliver Linton, and Zudi Lu. Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. *Journal of the American Statistical Association*, 113(522):919–932, 2018.

X. R. Chen, Z. D. Bai, L. Zhao, and Y. Wu. Asymptotic normality of minimum l1-norm estimates in linear models. *Science in China. Series A*, 33:1311–1328, 1990.

Xilong Chen and Eric Ghysels. News—good or bad—and its impact on volatility predictions over multiple horizons. *The Review of Financial Studies*, 24(1):46–81, 2011.

Xu Cheng and Bruce E Hansen. Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, 186(2):280–293, 2015.

Gerda Claeskens and Nils Lid Hjort. Model selection and model averaging. *Cambridge books*, 2008.

Peter Craven and Grace Wahba. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische mathematik*, 31(4):377–403, 1978.

Drew Creal, Siem Jan Koopman, and André Lucas. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795, 2013.

Hengjian Cui, Xuming He, and Kai W Ng. M-estimation for linear models with spatially-correlated errors. *Statistics & probability letters*, 66(4):383–393, 2004.

Jon Danielsson, Kevin R James, Marcela Valenzuela, and Ilknur Zer. Model risk of systemic risk models. *London School of Economics (mimeo)*, 2011.

Luciano de Castro and Antonio F Galvao. Dynamic quantile models of rational behavior. *Econometrica*, 87(6):1893–1939, 2019.

Luciano de Castro and Antonio F Galvao. Static and dynamic quantile preferences. *Economic Theory*, 73(2-3):747–779, 2022.

Luciano de Castro, Antonio F Galvao, Charles N Noussair, and Liang Qiao. Do people maximize quantiles? *Games and economic behavior*, 132:22–40, 2022.

Jan G De Gooijer and Dawit Zerom. Semiparametric quantile averaging in the presence of high-dimensional predictors. *International Journal of Forecasting*, 35(3): 891–909, 2019.

DQF De Menezes, Diego Martinez Prata, Argimiro R Secchi, and José Carlos Pinto. A review on robust m-estimators for regression analysis. *Computers & Chemical Engineering*, 147:107254, 2021.

Jan Dhaene, Roger JA Laeven, and Yiying Zhang. Systemic risk: Conditional distortion risk measures. *Insurance: Mathematics and Economics*, 102:126–145, 2022.

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 1995.

Jie Ding, Vahid Tarokh, and Yuhong Yang. Model selection techniques: An overview. *IEEE Signal Processing Magazine*, 35(6):16–34, 2018.

Paul Doukhan. *Mixing: properties and examples*, volume 85. Springer Science & Business Media, 2012.

David Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1):45–70, 1995.

Paul Embrechts. Correlation: pitfalls and alternatives. *Risk Magazine*, pages 69–71, 1999.

Paul Embrechts, Alexander McNeil, and Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223, 2002.

Paul Embrechts, Andrea Höing, and Giovanni Puccetti. Worst var scenarios. *Insurance: Mathematics and Economics*, 37(1):115–134, 2005.

Susanne Emmer, Marie Kratz, and Dirk Tasche. What is the best risk measure in practice? a comparison of standard measures. *arXiv preprint arXiv:1312.1645*, 2015.

Robert F Engle and Simone Manganelli. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of business & economic statistics*, 22(4):367–381, 2004.

J Carlos Escanciano and Jose Olmo. Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics*, 28(1):36–51, 2010.

Randall L Eubank. Spline smoothing and nonparametric regression. *(No Title)*, 1988.

Randall L Eubank. *Nonparametric regression and spline smoothing*. CRC press, 1999.

European Central Bank (ECB). Financial networks and financial stability. *Financial Stability Reviews*, pages 138–146, June 2010.

Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

Eugene F Fama and Kenneth R French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

Jianqing Fan. Local linear regression smoothers and their minimax efficiencies. *The annals of Statistics*, pages 196–216, 1993.

Jianqing Fan. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Routledge, 2018.

Jianqing Fan and Irène Gijbels. *Local regression*. Chapman & Hall/CRC, Boca Raton, Fla., 1996. ISBN 0412983214, 9780412983214. URL http://www.worldcat.org/search?qt=worldcat_org_all&q=0412983214.

Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70 (5):849–911, 2008.

Jianqing Fan and Jinchi Lv. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101, 2010.

Jianqing Fan and Qiwei Yao. *Nonlinear time series: nonparametric and parametric methods*, volume 20. Springer, 2003.

Jianqing Fan, Nancy E Heckman, and Matt P Wand. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429):141–150, 1995.

Zachary Feinstein, Birgit Rudloff, and Stefan Weber. Measures of systemic risk. *SIAM Journal on Financial Mathematics*, 8(1):672–708, 2017.

Tobias Fissler and Yannick Hoga. Backtesting systemic risk forecasts using multi-objective elicitability. *Journal of Business & Economic Statistics*, (just-accepted): 1–30, 2023.

Tobias Fissler and Johanna F Ziegel. Higher order elicitability and osband's principle. 2016.

Tobias Fissler and Johanna F Ziegel. Order-sensitivity and equivariance of scoring functions. 2019.

Tobias Fissler, Johanna F Ziegel, and Tilmann Gneiting. Expected shortfall is jointly elicitable with value at risk-implications for backtesting. *arXiv preprint arXiv:1507.00244*, 2015.

Christian Francq and Jean-Michel Zakoian. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637, 2004.

Jiti Gao. *Nonlinear time series: semiparametric and nonparametric methods*. Chapman and Hall/CRC, 2007.

Jiti Gao, Degui Li, and Zhengyan Lin. Robust estimation in parametric time series models under long-and short-range-dependent structures. *Australian & New Zealand Journal of Statistics*, 51(2):161–181, 2009.

Joseph L Gastwirth and Herman Rubin. The behavior of robust estimators on dependent data. *The Annals of Statistics*, pages 1070–1100, 1975.

Kay Giesecke and Baeho Kim. Systemic risk: What defaults are telling us. *Management Science*, 57(8):1387–1405, 2011.

Bruno C Giovannetti. Asset pricing under quantile utility maximization. *Review of Financial Economics*, 22(4):169–179, 2013.

Giulio Girardi and A Tolga Ergün. Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance*, 37(8):3169–3180, 2013.

Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. 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, 1993.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

Marc J Goovaerts, Rob Kaas, and Roger JA Laeven. Worst case risk measurement: Back to the future? *Insurance: Mathematics and Economics*, 49(3):380–392, 2011.

Damodar N Gujarati et al. Basic econometrics" fourth edition mcgraw-hill. *New York*, 2003.

Aparna Gupta, Runzu Wang, and Yueliang Lu. Addressing systemic risk using contingent convertible debt–a network analysis. *European Journal of Operational Research*, 290(1):263–277, 2021.

F. R. Hampel, E. Ronchetti, P. J. Rousseeuw, and W. E. Stahel. *Robust statistics: The approach based on influence functions*. John Wiley and Sons, New York, 1986.

Frank Hampel. Some additional notes on the\princeton robustness year. *The practice of data analysis: Essays in honor of John W. Tukey*, pages 133–153, 1997.

Frank Rudolf Hampel. *Contributions to the theory of robust estimation*. University of California, Berkeley, 1968.

Bruce E Hansen. Least squares model averaging. *Econometrica*, 75(4):1175–1189, 2007.

Bruce E Hansen and Jeffrey S Racine. Jackknife model averaging. *Journal of Econometrics*, 167(1):38–46, 2012.

Wolfgang Härdle. *Applied nonparametric regression*. Number 19. Cambridge university press, 1990.

T. J. Hastie and R. J. Tibshirani. *Generalized additive models*. London: Chapman & Hall, 1990. ISBN 0412343908.

Trevor Hastie and Robert Tibshirani. Generalized Additive Models. *Statistical Science*, 1(3):297 – 310, 1986. . URL https://doi.org/10.1214/ss/1177013604.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

Xuming He and Qi-Man Shao. A general bahadur representation of m-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics*, 24(6):2608–2630, 1996.

Nils-Bastian Heidenreich, Anja Schindler, and Stefan Sperlich. Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA Advances in Statistical Analysis*, 97:403–433, 2013.

Nils Lid Hjort and Gerda Claeskens. Frequentist model average estimators. *Journal of the American Statistical Association*, 98(464):879–899, 2003.

Jennifer A Hoeting, David Madigan, Adrian E Raftery, and Chris T Volinsky. Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. *Statistical science*, 14(4):382–417, 1999.

Tao Huang and Jialiang Li. Semiparametric model average prediction in panel data analysis. *Journal of Nonparametric Statistics*, 30(1):125–144, 2018.

Xin Huang, Hao Zhou, and Haibin Zhu. A framework for assessing the systemic risk of major financial institutions. *Journal of Banking & Finance*, 33(11):2036–2049, 2009.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. ISSN 00034851. URL http://www.jstor.org/stable/2238020.

Peter J. Huber. Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*, 1(5):799 – 821, 1973. . URL https://doi.org/10.1214/aos/1176342503.

Peter J Huber. *Robust statistics*, volume 523. John Wiley & Sons, 2004.

Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. John Wiley & Sons, Inc., United States, 2009. ISBN 9780470129906. .

I. Ibragimov and R. Hasminskii. *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York, 1981.

Philippe Jorion. *Value at risk: the new benchmark for managing financial risk*. McGraw-Hill, 2007.

Rob Kaas, Roger JA Laeven, and Roger B Nelsen. Worst var scenarios with given marginals and measures of association. *Insurance: Mathematics and Economics*, 44(2): 146–158, 2009.

Paul Kabaila. The effect of model selection on confidence regions and prediction regions. *Econometric Theory*, 11(3):537–549, 1995.

Daniel Kahneman and Amos Tversky. *The simulation heuristic*. National Technical Information Service, 1981.

Shmuel Kandel and Robert F Stambaugh. On the predictability of stock returns: an asset-allocation perspective. *The Journal of Finance*, 51(2):385–424, 1996.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Max Kuhn, Kjell Johnson, et al. *Applied predictive modeling*, volume 26. Springer, 2013.

Nicolas S Lambert, David M Pennock, and Yoav Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138, 2008.

Hannes Leeb and Benedikt M Pötscher. The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory*, 19(1):100–142, 2003.

Hannes Leeb and Benedikt M Pötscher. Can one estimate the conditional distribution of post-model-selection estimators? 2006.

Degui Li, Oliver Linton, and Zudi Lu. A flexible semiparametric forecasting model for time series. *Journal of Econometrics*, 187(1):345–357, 2015.

Jialiang Li, Xiaochao Xia, Weng Kee Wong, and David Nott. Varying-coefficient semiparametric model averaging prediction. *Biometrics*, 74(4):1417–1426, 2018.

Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics: theory and practice*. Princeton University Press, 2023.

Z Lin and C Lu. Limit theorems of mixing dependent random variables, 1996.

Zhengyan Lin, Degui Li, and Jiti Gao. Local linear m-estimation in non-parametric spatial regression. *Journal of Time Series Analysis*, 30(3):286–314, 2009.

Oliver Linton and Enno Mammen. Estimating semiparametric arch (∞) models by kernel smoothing methods 1. *Econometrica*, 73(3):771–836, 2005.

Oliver Linton and Enno Mammen. Nonparametric transformation to white noise. *Journal of Econometrics*, 142(1):241–264, 2008.

Oliver Linton and Alessio Sancetta. Consistent estimation of a general nonparametric regression function in time series. *Journal of Econometrics*, 152(1):70–78, 2009.

Hung-Chun Liu and Jui-Cheng Hung. Forecasting s&p-100 stock index volatility: The role of volatility asymmetry and distributional assumption in garch models. *Expert Systems with Applications*, 37(7):4928–4934, 2010.

Georg Mainik and Eric Schaanning. On dependence consistency of covarand some other systemic risk measures. *Statistics & Risk Modeling*, 31(1):49–77, 2014.

C. L. Mallows. Some comments on cp. *Technometrics*, 15(4):661–675, November 1973.

Charles Manski. Ordinal utility models of decision making under uncertainty. *Theory and Decision*, 25:79–104, 1988.

Harry Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, 3 1952. . URL https://www.jstor.org/stable/2975974.

Vivek Mohan, Jai Govind Singh, and Weerakorn Ongsakul. Sortino ratio based portfolio optimization considering evs and renewable energy in microgrid power market. *IEEE Transactions on Sustainable Energy*, 8(1):219–229, 2016.

Jan Mossin. Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, pages 768–783, 1966.

Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

Wojciech Niemiro. Asymptotics for m-estimators defined by convex minimization. *The Annals of Statistics*, 20(3):1514–1533, 1992. ISSN 00905364. URL http://www.jstor.org/stable/2242025.

Ryuei Nishii. Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, pages 758–765, 1984.

Natalia Nolde and Johanna F Ziegel. Elicitability and backtesting: Perspectives for banking regulation. 2017.

Nilay Noyan and Gábor Rudolf. Optimization with multivariate conditional value-at-risk constraints. *Operations research*, 61(4):990–1013, 2013.

Dong Hwan Oh and Andrew J Patton. Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business & Economic Statistics*, 36(2): 181–195, 2018.

Kent Harold Osband. *Providing Incentives for Better Cost Forecasting (Prediction, Uncertainty Elicitation)*. University of California, Berkeley, 1985.

Andrew J Patton, Johanna F Ziegel, and Rui Chen. Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of econometrics*, 211(2):388–413, 2019.

John W Payne, James R Bettman, and Eric J Johnson. Behavioral decision research: A constructive processing perspective. *Annual review of psychology*, 43(1):87–131, 1992.

Lasse Heje Pedersen, Viral Acharya, Thomas Philippon, and Matt Richardson. Measuring systemic risk. *NYU Working Paper*, 2010.

Jingfu Peng and Yuhong Yang. On improvability of model selection by model averaging. *Journal of econometrics*, 229(2):246–262, 2022.

Rong Peng and Zudi Lu. Semiparametric averaging of nonlinear marginal logistic regressions and forecasting for time series classification. *Econometrics and Statistics*, 2021.

Franco Peracchi. Robust m-estimators. *Econometric Reviews*, 9(1):1–30, 1990.

Peter CB Phillips. A shortcut to lad estimator asymptotics. *Econometric theory*, 7(4): 450–463, 1991.

Benedikt M Pötscher. Effects of model selection on inference. *Econometric Theory*, 7(2): 163–185, 1991.

Diego Martinez Prata, Enrique Luis Lima, and José Carlos Pinto. Nonlinear dynamic data reconciliation in real time in actual processes. In *Computer Aided Chemical Engineering*, volume 27, pages 47–54. Elsevier, 2009.

Matthew Rabin. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5):1281–1292, 2000. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2999450.

Adrian E Raftery, David Madigan, and Jennifer A Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437): 179–191, 1997.

Juan C Reboredo and Andrea Ugolini. Systemic risk in european sovereign debt markets: A covar-copula approach. *Journal of International Money and Finance*, 51: 214–244, 2015.

William JJ Rey. *Introduction to robust and quasi-robust statistical methods*. Springer Science & Business Media, 2012.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1969.

R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Murray Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the national Academy of Sciences*, 42(1):43–47, 1956.

Marzena Rostek. Quantile maximization in decision theory. *The Review of Economic Studies*, 77(1):339–371, 2010.

David Ruppert, Simon J Sheather, and Matthew P Wand. An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270, 1995.

Sergey Sarykalin, Gaia Serraino, and Stan Uryasev. Value-at-risk vs. conditional value-at-risk in risk management and optimization. In *State-of-the-art decision-making tools in the information-intensive age*, pages 270–294. Informs, 2008.

Leonard Savage. J. the foundations of statistics. *John Wiley and Sons, New York, and Chapman and Hall, London*, 9:19–132, 1954.

Leonard J Savage. *The foundations of statistics*. Courier Corporation, 1972.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

Jun Shao. An asymptotic theory for linear model selection. *Statistica sinica*, pages 221–242, 1997.

William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.

Herbert A Simon. Rational decision making in business organizations. *The American economic review*, 69(4):493–513, 1979.

Frank A Sortino and Robert Van Der Meer. Downside risk. *Journal of portfolio Management*, 17(4):27, 1991.

Charles J Stone. Optimal rates of convergence for nonparametric estimators. *The annals of Statistics*, pages 1348–1360, 1980.

Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053, 1982.

Mervyn Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2):111–133, 1974.

James W Taylor. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133, 2019.

Timo Teräsvirta, Dag Tjøstheim, and Clive WJ Granger. Modelling nonlinear economic time series. 2010.

Adrian Tobias and Markus K Brunnermeier. Covar. *The American Economic Review*, 106 (7):1705, 2016.

Jack L Treynor. Market value, time, and risk. *Time, and Risk (August 8, 1961)*, 1961.

Jack L Treynor. Jack treynor's' toward a theory of market value of risky assets'. *Available at SSRN 628187*, 1962.

Jingwen Tu, Hu Yang, Chaohui Guo, and Jing Lv. Model averaging marginal regression for high dimensional conditional quantile prediction. *Statistical Papers*, 62:2661–2689, 2021.

John Wilder Tukey. A survey of sampling from contaminated distributions. *Contributions to probability and statistics*, pages 448–485, 1960.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5:297–323, 1992.

Aman Ullah and Adrian Pagan. *Nonparametric econometrics*. Cambridge university press Cambridge, 1999.

Aad W van der Vaart, Jon A Wellner, Aad W van der Vaart, and Jon A Wellner. M-estimators. *Weak Convergence and Empirical Processes: With Applications to Statistics*, pages 284–308, 1996.

John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 2nd edition, 1947.

Grace Wahba. *Spline models for observational data*. SIAM, 1990.

Alan TK Wan, Xinyu Zhang, and Guohua Zou. Least squares model averaging by mallows criterion. *Journal of Econometrics*, 156(2):277–283, 2010.

Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.

Kenneth D West. Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084, 1996.

Wei Biao Wu. M-estimation of linear models with dependent errors. 2007.

Yuhong Yang. Adaptive regression by mixing. *Journal of the American Statistical Association*, 96(454):574–588, 2001.

Zheng Yuan and Yuhong Yang. Combining linear regression models: When and how? *Journal of the American Statistical Association*, 100(472):1202–1214, 2005.

Zishu Zhan, Yang Li, Yuhong Yang, and Cunjie Lin. Model averaging for semiparametric varying coefficient quantile regression models. *Annals of the Institute of Statistical Mathematics*, 75(4):649–681, 2023.

Xinyu Zhang, Alan TK Wan, and Guohua Zou. Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics*, 174(2):82–94, 2013.

Xinyu Zhang, Dalei Yu, Guohua Zou, and Hua Liang. Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. *Journal of the American Statistical Association*, 111(516):1775–1790, 2016.