Deep Learning and High-Frequency Data Enhanced Bayesian Financial Risk Forecasting

by

Hewathanthrige Rangika Iroshani Peiris

A thesis submitted to fulfil requirements for the degree of

Doctor of Philosophy

in

Discipline of Business Analytics, Business School
The University of Sydney

March 2025

This research reported in this thesis was supported by the award of a Research Training Program scholarship to the PhD Candidate.

Supervisors:

Lead Supervisor:
Associate Professor Minh-Ngoc Tran,
Discipline of Business Analytics,
The University of Sydney Business School

Associate Supervisor:
Dr. Chao Wang,
Discipline of Business Analytics,
The University of Sydney Business School

Dedication

"To my late dad and mum, whose love, wisdom, and encouragement have been the foundation of my journey. Your memory continues to inspire and guide me every day."

Contents

Abstract

This thesis, titled "Deep Learning and High-Frequency Data Enhanced Bayesian Financial Risk Forecasting," presents an innovative approach in forecasting financial tail risks, particularly focusing on Value-at-Risk (VaR) and Expected Shortfall (ES) through a Bayesian framework. Recognizing the limitations of traditional risk models in capturing the complex, non-linear dynamics of financial markets, this work proposes novel methodologies that integrate realized volatility measures with advanced Bayesian and machine learning techniques, aiming to enhance predictive accuracy and robustness in risk forecasting.

The first part of this thesis introduces a semi-parametric framework for joint VaR and ES forecasting, incorporating multiple realized measures such as realized variance, bi-power variation, and realized kernel. Extending the realized exponential Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, this framework employs a quasi-likelihood function based on the asymmetric Laplace distribution, enabling Bayesian estimation through adaptive Markov Chain Monte Carlo methods. Empirical analyses demonstrate that the proposed model outperforms both parametric and semi-parametric counterparts, particularly in periods of high market volatility, underscoring its effectiveness in capturing the nuanced risk dynamics across multiple realized measures.

The second part of the thesis advances risk forecasting further by proposing a long-memory and non-linear model, termed RNN-HAR, for direct VaR prediction. This model integrates a Recurrent Neural

Network (RNN) with the heterogeneous autoregressive (HAR) model to capture long-memory effects and non-linear dependencies within realized volatility measures. Using loss-based Bayesian Sequential Monte Carlo (SMC) for inference, the RNN-HAR model demonstrates superior predictive performance across 31 global markets, consistently outperforming conventional HAR models in VaR forecasting.

Building on these contributions, the final part of the thesis proposes a novel RNN-based framework, RNN-ReESCAViaR, for the joint estimation of VaR and ES. By embedding an RNN within the Realized-ES-CAViaR (Conditional Autoregressive Value-at-Risk) model and employing an asymmetric Laplace quasi-likelihood function, this framework captures the interdependence between VaR and ES in a highly flexible manner. Bayesian SMC is again utilized, enhancing computational efficiency and allowing for real-time sequential prediction. Empirical results across 21 markets show that RNN-ReESCAViaR achieves significant improvements over both parametric and semi-parametric models, illustrating its capacity to adapt to complex volatility patterns in financial data.

The findings of this thesis also have important regulatory implications. The proposed models could play a significant role in regulatory stress testing frameworks, which are crucial for ensuring that financial institutions hold adequate capital during periods of market stress. By more accurately forecasting tail risks, including VaR and ES, the models can provide regulators with better tools to assess financial institutions' ability to withstand extreme market conditions. This aligns with the Basel III framework, which emphasizes the need for more accurate measures of tail risk and capital adequacy to ensure financial stability.

Collectively, these contributions represent a significant advance in Bayesian tail risk forecasting by integrating realized measures with machine learning and Bayesian techniques. The models developed in this thesis not only provide more accurate and resilient tail risk forecasts but also offer a flexible framework for capturing non-linear dependencies and long-memory effects, which are crucial for modern financial risk management. This research has substantial implications for risk management practices, particularly in volatile or crisis-prone periods, where robust, precise tail risk predictions are essential for maintaining financial stability.

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

I give permission for the digital version of my dissertation to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University of Sydney to restrict access for a period of time.

H. Rangika Iroshani Peiris
28th December 2024

Acknowledgements

”Foremost, I would like to express my heartfelt gratitude to my lead supervisor, Associate Professor Minh-Ngoc Tran, for accepting me as a supervisee partway through my candidature. His invaluable guidance, encouragement, and consistent support have been instrumental in bringing this thesis to completion. I am deeply grateful for our weekly meetings that kept me aligned with my research goals, his prompt responses to my queries, and the time he set aside to discuss both academic and personal matters. His understanding of my efforts to balance the roles of researcher, academic tutor, mother, and wife has been especially appreciated.

I also wish to thank my associate supervisor, Dr. Chao Wang, who has supported me since the beginning of my PhD journey. His guidance was vital in helping me pass my proposal defense. I am grateful for his ongoing assistance with model development, empirical analysis, technical support, paper submissions, and the publication process.

My sincere thanks go to Professor Richard Gerlach, who initially accepted me as a research student, recommended me for a PhD scholarship, and supported me with journal publications. His guidance has been foundational to my academic progress.

I am also deeply grateful to Professor Junbin Gao, who, as the postgraduate coordinator, recommended me for a research scholarship in 2020 and provided immense personal support during my mother’s passing and the COVID-19 lockdowns. His understanding and compassion encouraged me to persevere, and I am forever grateful for his support through the transition in supervision. I would also like to thank my annual progress review committee members, Professor Artem Prokhorov and Professor Jie Yin, for their valuable feedback and guidance, which significantly contributed to my progress, as well as Professor Andrey Vasnev for his encouragement throughout my research.

Further, I would like to thank the Business Analytics administration staff, unit coordinators, head tutors, and tutorial team members for their support in balancing my research and teaching roles.

I am grateful to the Australian government for awarding me the prestigious Research Training Program (RTP) scholarship. I still vividly recall receiving the award notification while traveling to my mother's funeral; the financial support from this scholarship was essential during the pandemic, enabling me and my family to stay focused and stable in Sydney.

The challenges of pursuing a PhD away from my family have been immense. While I have faced many adversities in life, including civil war and natural disasters, being away from family support in Sydney was particularly challenging. I am thankful for my fellow PhD students, especially Vica, Mehrnoosh, and Niles, for their camaraderie. I also extend my appreciation to Dr. Rasika, Dr. Sanjaye, and Dr. Nirmala, whose advice, shared experiences, and support provided much-needed comfort and encouragement along the way. I am especially grateful to my close friends who provided emotional support and motivation throughout this journey. Their understanding, kindness, and encouragement have meant so much to me.

I am also grateful to my former supervisors in Sri Lanka, Professor Prabhath Jayasinghe, who guided my MBA thesis, and Professor Sarath Banneheka, who supervised my undergraduate research. Their encouragement during my early academic career laid the groundwork for my academic journey.

My heartfelt thanks go to my family. To my elder brother, Praveen Peiris, and sister-in-law, Moksha Peiris, for their financial support with visa applications and ongoing help, and to my elder sister, Rasika Peiris, and brother-in-law, Himal De Silva, for their emotional support. My sister's willingness to take on family responsibilities allowed me to focus fully on my studies, and her daily prayers were a blessing during the most challenging times.

Finally, to my husband, Udesh Amarasinghe, I am incredibly grateful for standing by me throughout this PhD journey, taking on the role of caregiver for our son so I could pursue my research. And to my son, Bovindu Amarasinghe, thank you for your patience and love. Your warmth and affection have often been a source of strength and motivation, inspiring me to persevere. Thank you for choosing me as your mother in this life."

Use of AI and AI-Assisted Technologies in Thesis Writing

I would like to express my acknowledgement for utilizing AI and AI-assisted technologies during the creation of my thesis. In this context, I have utilized ChatGPT, Grammarly, and QuillBot exclusively for the purpose of text editing, with the intention of refining the clarity and language of my thesis. It is important to note that I did not employ these technologies in the generation of research data, analysis, result discussions, or drawing conclusions throughout the writing process.

Dissertation Conventions

The following conventions have been adopted in this dissertation:

Typesetting

This document was compiled using LaTeX2e. Texstudio 2.12.22 was used as a text editor interfaced to LaTeX2e. Inkscape 0.92.3 was used to produce schematic diagrams and other drawings.

Spelling

Australian English spelling conventions have been used, as defined in the Macquarie English Dictionary—A. Delbridge (Ed.), Macquarie Library, North Ryde, NSW, Australia, 2001.

Referencing

The Harvard reference style is used for referencing and citation in this dissertation.

System of Units

The units comply with the international system of units recommended in an Australian Standard: AS ISO 1000-1998 (Standards Australia Committee ME/71, Quantities, Units and Conversions 1998).

Acronyms

ALD	Asymmetric Laplace Distribution
BV	Bi-power Variation
DQ	Dynamic Quantile
ES	Expected Shortfall
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
HAR	Heterogeneous Autoregressive model
HS	Historical Simulation
MCMC	Markov Chain Monte Carlo
MCS	Model Confidence Set
MLE	Maximum Likelihood Estimation
QML	Quasi-Maximum Likelihood
REGARCH	Realized Exponential GARCH
RK	Realized Kernel
RM	Realized Measures
RV	Realized Variance
RNN	Recurrent Neural Network
SMC	Sequential Monte Carlo
VRate	VaR Violation Rate
VaR	Value-at-Risk

Authorship attribution statement

Chapter 4 of this thesis is published as
Peiris, R., Wang, C., Gerlach, R. and Tran, M.N., 2024. Semi-parametric financial risk forecasting incorporating multiple realized measures. Quantitative Finance (ABDC rating: A).

Chapter 5 of this thesis is published as:
Peiris, R., Tran, M.N., Wang, C. and Gerlach, R., 2024. Loss-based Bayesian Sequential Prediction of Value at Risk with a Long-Memory and Non-linear Realized Volatility Model. arXiv preprint arXiv:2408.13588. Under revision for Journal of Financial Econometrics (ABDC rating: A*)

I co-designed the research ideas and developed the research methodology with the co-authors (who are my supervisors), implemented the algorithms, analyzed the results, conducted the empirical studies, and wrote the manuscripts draft.

H. Rangika Iroshani Peiris
28th December 2024

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Minh-Ngoc Tran
26th December 2024

Publications and Conferences

Journal Articles

Published

- Peiris, R., Wang, C., Gerlach, R. and Tran, M.N., 2024. Semi-parametric financial risk forecasting incorporating multiple realized measures. Quantitative Finance (Rank A).

Under revision

- Peiris, R., Tran, M. N., Wang, C., & Gerlach, R. (2024). Loss-based Bayesian Sequential Prediction of Value at Risk with a Long-Memory and Non-linear Realized Volatility Model. Journal of Financial Econometrics. (Rank A*)

Conference Presentations

- Peiris, R., Wang, C., Gerlach, R., & Tran, M. N. (2024). Semi-parametric financial risk forecasting incorporating multiple realized measures. EAC-ISBA Conference 2024, The Education University of Hong Kong, 26th June 2024.
- Peiris, R., Wang, C., Gerlach, R., & Tran, M. N. (2024). Semi-parametric financial risk forecasting incorporating multiple realized measures. 26th International Conference on Computational Statistics (COMPSTAT 2024), University of Giessen, Germany, 29th August 2024.
- Peiris, R., Tran, M. N., Wang, C., & Gerlach, R.,(2024). Loss-based Bayesian Sequential Prediction of Value at Risk with a Long-Memory and Non-linear Realized Volatility Model. Interim Conference of the Asian Regional Section of the International Association for Statistical Computing (IASC-ARS), National Taiwan University, Taipei City, Taiwan, 13th December 2024.
- Peiris, R., Wang, C., Gerlach, R., & Tran, M. N. (2024). Semi-parametric financial risk forecasting incorporating multiple realized measures. 18th International Joint

Conference CFE-CMStatistics (CFE - 2024), King's College, London, UK, 15th December 2024.

- Peiris, R., Tran, M. N., Wang, C., & Gerlach, R., (2025). Loss-based Bayesian Sequential Prediction of Value at Risk with a Long-Memory and Non-linear Model. Contributed Poster Presentation, Joint Statistical Meetings (JSM 2025), Nashville, Tennessee, USA, 4th August 2025.

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the research focus on Bayesian tail risk forecasting with realized measures of volatility. It provides background to the topic, highlighting the importance of accurate risk forecasting in financial markets, particularly using realized measures and advanced modelling techniques. Additionally, the chapter presents a summary of the original contributions of this dissertation, which include novel modelling approaches that integrate Bayesian inference and deep learning. Finally, an outline of the dissertation structure is provided to guide the reader through the purpose and content of each chapter.

1.1 Introduction

Tail risk forecasting has become a central focus in financial risk management due to the severe impact of extreme market movements on investments. Traditional risk forecasting models often rely on return series alone, which may not capture the full extent of market dynamics. This dissertation addresses this limitation by using realized measures, high-frequency data-based measures of volatility that provide more precise insights into market behavior. Realized measures allow us to develop models that more accurately reflect the complexities of financial data, offering enhanced predictive capabilities for Value-at-Risk (VaR) and Expected Shortfall (ES).

In addition, the work explores the integration of Bayesian inference and deep learning techniques. Bayesian methods provide a principled and convenient way for capturing uncertainty, support robust model estimation and sequential prediction. Deep learning, specifically through Recurrent Neural Networks (RNNs), allows the models to capture long-term dependencies and non-linear patterns within the data, which are often critical for forecasting accuracy. This combination aims to advance tail risk forecasting, addressing gaps in both theoretical development and practical applications.

1.2 Summary of original contributions

The original contributions of this research are structured around the development and application of advanced statistical and machine learning models for tail risk forecasting. Key contributions include:

- The development of a semi-parametric framework that incorporates multiple realized measures to improve the accuracy of joint VaR and ES forecasting. This model leverages the asymmetric Laplace quasi-likelihood and adaptive Markov Chain Monte Carlo (MCMC) techniques, establishing a robust and flexible structure for financial risk forecasting.

- The extension of the heterogeneous autoregressive (HAR) model to include non-linear and long-memory characteristics by embedding an RNN. This extension, termed RNN-HAR, provides improved VaR forecasting performance by capturing complex market dynamics through Bayesian Sequential Monte Carlo (SMC) inference.

- A novel model, RNN-ReESCAViaR (Realized-ES-Conditional Autoregressive Value-at-Risk), for jointly forecasting VaR and ES that integrates realized measures with RNN-based non-linear modelling. This approach offers competitive forecast accuracy by jointly estimating VaR and ES, using SMC for efficient Bayesian inference.

These contributions not only enhance the theoretical landscape of risk modelling, but also provide practical value to stakeholders in the financial industry by improving the reliability of the forecast.

1.3 Dissertation structure

The structure of this dissertation is organized as follows:

- **Chapter 2** presents a review of risk modelling and forecasting methodologies, introducing core tail risk estimators such as VaR and ES, as well as methods for evaluating forecasting accuracy using metrics like quantile loss, joint loss, and the Dynamic Quantile test. Further it provides a comprehensive discussion on realized measures in tail risk forecasting.
- **Chapter 3** lays the foundation for Bayesian inference and deep learning methodologies relevant to the thesis. It details adaptive MCMC for Bayesian inference, explores loss-based Bayesian inference with SMC, and introduces RNNs as a deep learning approach capable of capturing non-linear time series patterns.
- **Chapter 4** presents the first main contribution of the dissertation, a semi-parametric framework for forecasting tail risk that leverages multiple realized measures. Traditional models often rely on a single realized measure or even only the return series, which can overlook key information embedded in volatility patterns. By integrating multiple realized measures, this framework aims to improve the accuracy of joint VaR and ES predictions. The chapter also outlines the use of Bayesian inference through adaptive MCMC techniques, providing a flexible yet robust estimation process that accommodates complex model dynamics. Empirical results demonstrate the advantages of this framework over traditional single-measure approaches, highlighting its potential impact on risk management practices. This work has been published by the Quantitative Finance journal (Rank A) under title of "Semi-parametric financial risk forecasting incorporating multiple realized measures".
- **Chapter 5** Building on the need for models that capture long-memory effects and non-linear relationships in financial data, Chapter 5 introduces the RNN-HAR model. This model extends the HAR structure, which is widely used for modelling long-memory characteristics, by embedding a RNN to account for non-linear dependencies in the data. By using a loss-based Bayesian inference approach, specifically Bayesian SMC, the model bypasses the need for a traditional likelihood function and instead uses a quantile-based loss function. This method not only enhances the efficiency of estimation but also

provides more accurate forecasts by allowing for flexibility in the modelling process. By employing Bayesian SMC for estimation, this model balances the computational demands of deep learning with the robustness of Bayesian inference. The RNN-HAR model's ability to handle both long-term dependencies and complex non-linear patterns results in improved one-step-ahead VaR forecasts, as shown in an extensive empirical analysis across multiple markets. This chapter illustrates how the model outperforms both the basic HAR and other extensions, demonstrating the value of integrating deep learning within established forecasting frameworks. This work is under revision with Journal of Financial Econometrics (Rank A*).

- **Chapter 6** presents the final main contribution, the RNN-ReESCAViaR model, which further advances tail risk forecasting by jointly estimating VaR and ES within a semi-parametric framework. This model uniquely leverages RNNs to capture non-linear dynamics and incorporates realized measures, separately influencing VaR and ES. Empirical analysis across 21 markets highlights the RNN-ReESCAViaR model's performance in jointly forecasting VaR and ES, outperforming both parametric and semi-parametric models. The chapter underscores how joint modelling of these two measures, combined with RNN and Bayesian techniques, offers a more comprehensive approach to tail risk forecasting that meets industry demands for accuracy and computational efficiency.

- **Chapter 7** concludes the dissertation, summarizes, the key findings and suggests potential directions for future research in tail risk forecasting.

Chapter 2

Realized measures in tail risk forecasting

This chapter presents an overview of tail risk forecasting and the role of realized measures in risk forecasting. Also, it explains the two types of risk estimators, VaR, ES and the risk forecasts evaluation methods which will be used in this thesis. Later, this chapter explains the realized measures that are used in this thesis.

2.1 Risk modelling and forecasting

Risk modelling and forecasting are essential components of financial risk management, aiming to quantify and anticipate potential losses in investment portfolios. In the field of finance, the risk modelling and forecasting are essential tools for understanding and managing uncertainties. The traditional risk measures often underestimate tail risk, leading to insufficient risk management strategies. To address this issue, tail risk forecasting models utilize more advance tail risk measures to capture the tail behaviour more accurately. This section describes the tail risk measures and the evaluation techniques of tail risk forecasts used in this thesis.

2.1.1 Tail risk estimators

The global stock market crash on October 19, 1987, known as "Black Monday," exposed vulnerabilities in existing financial risk management practices, especially in handling extreme market downturns. In response, the Basel Committee on Banking Supervision (BCBS) introduced the Basel I Accord in 1988 to establish standardized guidelines for managing financial risks and protecting institutions from unexpected losses. However, subsequent financial crises in the 1990s' such as the Orange County bankruptcy, the collapse of Barings Bank, and the Long-Term Capital Management hedge fund crisis, highlighted the need for improved methods to assess market risks effectively.

In 1989, Dennis Weatherstone, then CEO of J.P. Morgan & Co., introduced the "4-15 report", which quantified daily risks across the bank's operations. This practice evolved into the VaR framework, a measure that estimates the maximum potential loss of a portfolio within a specific confidence level (e.g., 95%) over a set time horizon. The Basel Committee endorsed VaR in 1996 as a benchmark for market risk assessment, incorporating it into the Basel II framework. This recommendation marked VaR's rise as a standard tool in financial risk management (Jorion, 2007).

However, VaR has significant shortcomings. While it identifies the threshold for potential losses, it does not account for the magnitude of losses exceeding that threshold. To address this gap, ES, also known as conditional VaR, was introduced by Artzner et al. (1999). ES provides the average loss when the VaR threshold is breached, making it a more comprehensive measure of tail risk. Moreover, ES satisfies the mathematical property of subadditivity, ensuring it captures diversification benefits more accurately than VaR.

Despite its advantages, ES was not initially adopted in Basel II, which relied solely on VaR. However, the global financial crisis of 2007–2009 exposed weaknesses in traditional risk measures, particularly their reliance on parametric assumptions like normality. In response, the

Basel III framework introduced ES as the primary tail risk measure for market risk, recognising its ability to account for extreme losses and better align with risk management goals.

This thesis employs both VaR and ES as key tail risk estimators due to their importance in regulatory frameworks and their widespread use in financial risk management. In the next section, detailed approaches to calculating VaR and ES using Gaussian and Student's t distributions are discussed, providing the foundational tools for tail risk modelling in this study.

Value at risk (VaR)

Numerous definitions for VaR are available in the literature. In this thesis, the formal definition of VaR proposed by McNeil et al. (2005) is adopted. VaR for an asset portfolio at a specified confidence level $\alpha \in (0, 1)$ is defined as the minimum loss, denoted as l such that the probability of the loss L exceeding l is not greater than $(1 - \alpha)$. According to the Basel III Accord, the common confidence level of $\alpha = 2.5\%$ is employed in this thesis. However, for the purposes of the empirical study, we also consider $\alpha = 1\%$ and $\alpha = 5\%$ to evaluate risk at different confidence levels. Gaussian and Student-t distributions are considered as loss distributions in the empirical studies to assess the robustness of VaR calculations under different assumptions about the underlying return distribution.

McNeil et al. (2005) shows that when the loss distribution F_L follows a Gaussian distribution with mean μ and variance σ^2 the the α level VaR can be given as:

$$VaR_\alpha = \mu + \sigma\Phi^{-1}(\alpha), \tag{2.1}$$

where $\Phi^{-1}(\alpha)$ is the inverse of the cumulative distribution function (CDF) of the standard normal distribution, i.e., the quantile function at the confidence level α.

The Student-t distribution is particularly useful for modelling financial returns when there is excess kurtosis (fat tails). Unlike the Gaussian distribution, which assumes that the tails of the distribution decay exponentially, the Student-t distribution provides a better fit for financial data that exhibits more extreme values or larger shocks. If the loss function F_L following a Student-t distribution with degrees of freedom $v > 2$, the VaR is computed using the formula:

$$VaR_\alpha = \mu + \sigma t_v^{-1}(\alpha)\sqrt{\frac{v-2}{v}}, \tag{2.2}$$

where $t_v^{-1}(\alpha)$ is the inverse of the CDF of the Student-t distribution with v degrees of freedom.

Expected shortfall (ES)

ES can be viewed as the average loss that exceeds the VaR at a given confidence level, α. If the loss distribution F_L follows a Gaussian distribution with mean μ and standard deviation σ then the ES at α can be cacluated as (McNeil et al., 2005):

$$ES_\alpha = \mu + \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}, \tag{2.3}$$

where ϕ is the probability density function (PDF) of the standard normal distribution, and $\Phi^{-1}(\alpha)$ is the inverse of the CDF of the standard normal distribution. This formula essentially computes the conditional expectation of the losses that are worse than the VaR level, incorporating the tail of the loss distribution.

For a loss distribution that follows a Student-t distribution with v degrees of freedom, the calculation of ES becomes more complex. In this case, the ES at confidence level α is given by:

$$ES_\alpha = \mu + \sigma\frac{g_v(t_v^{-1}(\alpha))}{1-\alpha}\left(\frac{v+(t_v^{-1}(\alpha))^2}{v-1}\right)\sqrt{\frac{v-2}{v}}, \tag{2.4}$$

where g_v is the PDF of the standard t-distribution, and $t_v^{-1}(\alpha)$ is the inverse CDF of the t-distribution. This formula adjusts for the heavier tails of the t-distribution compared to the normal distribution, providing a more accurate estimation of ES for fat-tailed risks.

While VaR and ES are essential measures for assessing the potential risk of a portfolio, their reliability and accuracy must be rigorously evaluated. In line with Basel II's guidelines, which recommend back-testing VaR models to assess forecast accuracy (Chen et al., 2012), it becomes critical to validate the predictive power of these models through various evaluation techniques. This ensures that the risk forecasts not only capture the underlying risk characteristics but also provide reliable predictions for financial decision-making. The next section discusses various evaluation techniques commonly used in the literature to assess VaR and ES forecasts. These methods enable us to determine the effectiveness of the proposed models in forecasting tail risk and help guide model improvements.

2.1.2 Evaluation of risk forecasts

The risk forecast evaluation is crucial because it allows us to assess the reliability and accuracy of our tail risk forecasts, which are essential for financial decision-making and risk management. A robust evaluation framework ensures that the models used to forecast risk are both effective in capturing market dynamics and consistent in their predictive capabilities.

To ensure the validity of our risk forecast evaluation, the in-sample and out-of-sample periods are kept independent. In Chapters 4 to 6, no information from the out-of-sample period is used during in-sample training and vice versa. Additionally, no hyperparameter tuning is performed in this thesis, which eliminates the risk of look-ahead bias during model training and evaluation.

In the context of tail risk forecasting, various evaluation methods are employed in the literature to measure the accuracy and adequacy of VaR and ES predictions. These methods aim to assess not just the point forecasts of risk, but also how well the models capture the tail behavior of the loss distributions. Some of the common evaluation criteria include the VaR violation Rate (VRate), the Dynamic Quantile (DQ) test, the Quantile loss Function, Joint loss Function, and Model Confidence Set (MCS). Each of these methods offers unique insights into the performance of risk models, focusing on different aspects of forecast quality such as calibration, coverage, and the ability to account for extreme events.

VaR Violation Rate (VRate)

The VRate is a straightforward and effective tool for evaluating the accuracy of VaR forecasts. It measures the proportion of instances where the actual return exceeds the predicted VaR level—referred to as violations. This is expressed mathematically as:

$$\text{VRate} = \frac{1}{m} \sum_{t=n+1}^{n+m} I\left(r_t < \text{VaR}_t\right), \tag{2.5}$$

where n is the in-sample size, m is the forecast sample size, I is an indicator function that equals 1 when the return exceeds the VaR at time t, and 0 otherwise, r_t is the actual return at time t. In an ideal scenario, if the model correctly specifies the return quantiles, the true VRate should match the confidence level α. A model with a VRate close to the nominal level indicates that the distribution assumption of the return is appropriate. This metric can be further refined for model comparison by considering the ratio VRate/α which acts as a standardised measure for comparing and ranking competing models. Models with this ratio closer to 1 are generally preferred, as they indicate a better alignment between the observed and expected violation frequencies (Chen et al., 2012).

While the VRate provides an essential measure of a model's ability to meet its specified quantile level, it focuses solely on the frequency of violations. This simplicity, while useful, does not capture the independence or temporal behavior of these violations, which are critical for assessing the robustness of VaR models.

To evaluate these additional aspects, methods such as the DQ test are employed (Engle and Manganelli, 2004). The DQ test extends the analysis by statistically examining whether violations occur randomly over time, ensuring a more thorough assessment of model performance. The next section will explore the DQ test in detail and its role in evaluating VaR models.

Dynamic Quantile (DQ) test

The DQ test, introduced by Engle and Manganelli (2004), evaluates the conditional coverage of VaR forecasts. This test is notable for its ability to assess whether the violations (instances where returns exceed the VaR threshold) occur both independently and at the expected rate, making it a robust tool for evaluating risk models. While their original paper proposes both in-sample and out-of-sample versions of the test, we focus exclusively on the out-of-sample approach due to its practical relevance for forecasting accuracy in dynamic financial environments. The test begins with the construction of a "hit" sequence:

$$H_t = I\left(r_t < \mathrm{VaR}_t\right) - \alpha,$$

where, $I\left(r_t < \mathrm{VaR}_t\right)$ is an indicator function equal to 1 if the observed return r_t below the forecast VaR, and α represents the confidence level. Under the null hypothesis, H_t should satisfy the following conditions:

- Unconditional Coverage: $\mathbb{E}(H_t) = 0$, implying that the proportion of violations matches the confidence level.
- Conditional Coverage: $\mathbb{E}(H_t W_{it}) = 0$, where W represents lagged hits and other explanatory variables, ensuring no systematic patterns in the violations.

These hypotheses validate that the model produces accurate and independent forecasts of risk. To test the null hypothesis, an artificial regression is constructed, regressing H_t on its lagged values and the VaR forecast:

$$H_t = \beta_0 + \beta_1 H_{t-1} + \beta_2 H_{t-2} + ... + \beta_k H_{t-k} + \beta_{k+1}\mathrm{VaR}_t + \epsilon_t, \tag{2.6}$$

where k is the number of lags. This formulation checks whether H_t is independent of past hits and the current VaR forecast. In this context, using 4 lagged hits (H_{t-1} to H_{t-4}) ensures a thorough examination of autocorrelation in the violations. This approach is particularly effective in financial applications, where dependencies in returns and risk forecasts may span multiple time periods. For instance:

$$\begin{aligned}
W_t^T &= (1, H_{t-1}, \mathrm{VaR}_t) \quad \text{denoted as } DQ_1,\\
W_t^T &= (1, H_{t-1}, H_{t-2}, \mathrm{VaR}_t) \quad \text{denoted as } DQ_2,\\
W_t^T &= (1, H_{t-1}, H_{t-2}, H_{t-3}, \mathrm{VaR}_t) \quad \text{denoted as } DQ_3,\\
W_t^T &= (1, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-4}, \mathrm{VaR}_t) \quad \text{denoted as } DQ_4.
\end{aligned}$$

The test statistic is derived as:

$$DQ\left(q\right)=\frac{H^{'}W\left(W^{'}W\right)^{-1}W^{'}H}{\alpha\left(1-\alpha\right)}\sim\chi_{q}^{2},\tag{2.7}$$

where W is a matrix of explanatory variables (including lagged hits and VaR forecasts), and q is the number of explanatory variables. The null hypothesis is that all regression coefficients, except the constant term, are equal to zero. This would imply that H_t is independent of past hits and the current VaR forecast, meaning that the violations (hits) are unbiased.

The use of lagged hits ensures the test captures potential dependencies in the violations over time, a crucial feature for evaluating dynamic financial models. Empirical studies, such as those by Chen et al. (2012), confirm that the test is relatively insensitive to the number of lags, though increasing lags enhances its ability to detect subtle dependencies. The DQ test complements simpler measures like Violation Rate (VRate) by providing a rigorous framework for evaluating the joint adequacy of unconditional and conditional coverage in risk forecasts. By incorporating lagged dependencies, it ensures the reliability of VaR models in capturing the dynamics of financial risk.

Building on the evaluation metrics discussed thus far, we now turn to Quantile loss, another important tool for assessing the accuracy of VaR models. In the following section, we explore how this method provides a more direct measure of quantile estimation, thereby enhancing our model evaluation framework.

Quantile loss function

In evaluating the performance of VaR models, traditional backtests based on violations often lack the sensitivity needed to distinguish between models with similar violation rates. To address this limitation, measuring the magnitude of uncovered losses becomes essential. One prominent method for this purpose is the Quantile loss Function, which directly assesses the accuracy of VaR forecasts by comparing observed returns to the predicted quantiles.

The Quantile loss function, as proposed by Lopez (1999) and further developed by Gneiting and Ranjan (2011), captures the magnitude of errors in VaR predictions by comparing the forecasted quantiles to the actual returns. Since this loss function is strictly consistent, it ensures that the forecasted quantile matches the true quantile, minimizing the expected loss when the model is correct. The Quantile loss function is given by:

$$S_{\alpha}\left(Q_t,r_t\right)=\sum_{t=n+1}^{n+m}\left(\alpha-I(r_t\leq\widehat{Q}_t)\right)\left(r_t-\widehat{Q}_t\right),\tag{2.8}$$

where $\widehat{Q}_t$ represents the forecasted quantile, n is the in-sample size, m is the forecast sample size, and r_t is the observed return. The model that minimizes this function is considered the most accurate, as it closely matches the true quantile, providing a precise measure of the risks that VaR forecasts intend to capture.

While the Quantile loss function provides a robust measure for evaluating the accuracy of VaR forecasts, it focuses solely on one aspect of tail risk. Effective risk management often requires a comprehensive assessment that accounts for both the frequency and magnitude of losses beyond the VaR threshold. To address this limitation, we now present the joint loss function, which evaluates VaR and ES together, offering a unified framework for assessing these complementary risk measures.

Joint loss function

The ES has traditionally been regarded as a non-elicitable risk measure, meaning that no standalone loss function is minimized by the true ES (Gneiting, 2011). However, the groundbreaking work of Fissler and Ziegel (2016) established that while ES is not elicitable alone, it is jointly elicitable alongside VaR. This joint elicitation enables the development of strictly consistent scoring functions that evaluate both risk measures simultaneously.

One widely used Jfunction, introduced by Fissler and Ziegel (2016), is given by the following scoring function:

$$\begin{aligned} S_t\left(r_t, \mathrm{VaR}_t, \mathrm{ES}_t\right) = (I_t - \alpha)G_1(\mathrm{VaR}_t) - I_t G_1(r_t) + G_2(\mathrm{ES}_t)\left(\mathrm{ES}_t - \mathrm{VaR}_t + \frac{I_t}{\alpha}(\mathrm{VaR}_t - r_t)\right) \ldots \\ - H(\mathrm{ES}_t) + a(r_t), \end{aligned} \tag{2.9}$$

where $I_t = 1$ if $r_t < \mathrm{VaR}_t$ and 0 otherwise, $G_1(\cdot)$ and $G_2(\cdot)$ are increasing functions, $H'(\cdot) = G_2(\cdot)$, and $a(\cdot)$ is an integrable function.

Building on this framework, Taylor (2019) proposed an ES-CAViaR model that jointly estimates VaR and ES using a quasi-likelihood derived from the asymmetric Laplace (AL) distribution. The AL-based quasi-log-likelihood, also a strictly consistent Joint loss function, is expressed as:

$$S_t(r_t, \widehat{Q}_t, \widehat{\mathrm{ES}}_t) = -\log\left(\frac{\alpha - 1}{\widehat{\mathrm{ES}}_t}\right) - \frac{(r_t - \widehat{Q}_t)(\alpha - I(r_t \leq \widehat{Q}_t))}{\alpha \widehat{\mathrm{ES}}_t}. \tag{2.10}$$

To formally assess model performance, the average joint loss is calculated as:

$$S = \frac{1}{m}\sum_{t=n+1}^{n+m} S_t \tag{2.11}$$

where S_t represents the individual loss for each observation. In this context, a lower value of S indicates that the model provides better predictions of VaR and ES, reflecting higher forecast accuracy.

In this approach, the AL quasi-likelihood serves both as a scoring function for evaluation and joint VaR and ES estimation. The resulting framework allows for the integration of Bayesian methods and provides a rigorous basis for comparing models that forecast VaR and ES simultaneously. These developments underscore the importance of Joint loss functions in modern risk forecasting. By addressing the interdependence of VaR and ES, these functions not only improve forecast accuracy but also provide a unified approach to model evaluation.

While the Joss function provides a comprehensive evaluation of forecast accuracy by assessing the precision of VaR and ES forecasts, both of which are key risk metrics, the Model Confidence Set (MCS) is a statistical tool used to compare and rank models based on their performance. The MCS helps determine whether differences in model performance are statistically significant, focusing on identifying the most reliable models. In the next section, we will introduce the MCS method and explore its application to assess model performance in this thesis.

Model Confidence Set (MCS)

The MCS test, introduced by Hansen et al. (2011), is a robust statistical method used for comparing multiple forecasting models. The goal of the MCS is to identify a set of models that are statistically indistinguishable from each other in terms of forecast performance, given a predefined confidence level. This approach is particularly useful when selecting models for applications like financial forecasting, where forecasting accuracy and reliability are paramount.

The procedure starts with a collection of competing models $\mathcal{M}_0 = \{\mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_m\}$, each evaluated by a loss function $L_{i,t}$ that measures the forecasting error at time t for model i (Hansen et al., 2011). The models are then compared through a series of pairwise tests based on the loss differences:

$$d_{ij,t} = L_{i,t} - L_{j,t} \text{ for all } i, j \in \mathcal{M}_0$$

The expected difference between the loss functions is defined as:

$$\mu_{ij} = E(d_{ij,t})$$

where μ_{ij} represents the expected loss difference between models i and j. If $\mu_{ij} \leq 0$, model i is considered superior to model j. The null hypothesis $H_{0,\mathcal{M}}$ assumes that the models in a subset $\mathcal{M}$ have equal predictive ability, i.e., $\mu_{ij} = 0$, for all pairs of models in $\mathcal{M}$. An alternative hypothesis $H_{A,\mathcal{M}}$ asserts that there are significant differences between some models.

To determine the superior models, the MCS employs equivalence tests and an elimination rule. The equivalence test checks whether the null hypothesis can be rejected for a set of models, while the elimination rule removes the models that do not meet the required statistical criteria. The test statistic for the equivalence test is computed as:

$$t_{ij} = \frac{\overline{d_{ij}}}{\sqrt{\mathrm{VaR}(\overline{d_{ij}})}},$$

where $\overline{d_{ij}}$ is the average loss difference over the sample period, and $\mathrm{VaR}(\overline{d_{ij}})$ is the estimated variance of the difference. The maximum of these test statistics across all pairs is used to determine the significance of the differences:

$$T_{\mathcal{M}} = \max_{(i,j \in \mathcal{M})} \left| t_{ij} \right|.$$

The MCS algorithm proceeds by iterating through these tests and progressively eliminating models from the set. At each iteration, the null hypothesis is tested, and if rejected, the weakest model is removed. The surviving models form the final model confidence set $\mathcal{M}^*_{1-\alpha}$, which contains the models that are not statistically dominated by others at a given confidence level $1 - \alpha$. This final set of models represents those with comparable predictive accuracy, offering a robust selection for further analysis (Hansen et al., 2011).

The MCS is particularly effective when comparing a large number of competing forecasting models, such as those used in financial risk management. For instance, Hansen et al. (2011) applied the MCS to compare different models for risk forecasting, showing its ability to identify superior models in terms of Quantile loss. By considering loss functions like Quantile loss, the MCS allows us to statistically evaluate which models provide the most reliable forecasts under different confidence levels. The results are particularly useful for model selection in risk forecasting, ensuring that the final chosen models are statistically robust and suitable for real-world applications.

In this section, various evaluation methods for assessing risk forecasts were discussed, with a particular focus on VaR and ES measures. We reviewed key VaR evaluation techniques, including the VRate measure, DQ test and Quantile loss. These methods help quantify the accuracy of risk forecasts, particularly in assessing the tail of the distribution. Additionally, the Joint loss Function for ES was considered as an important tool for evaluating the joint behavior of these risk measures. To ensure robustness, we also explored the MCS, a statistical procedure for identifying superior forecasting models based on their predictive accuracy.

While risk evaluation methods provide crucial tools for assessing the performance of risk forecasts, the accuracy of these forecasts depends heavily on the quality of the inputs used in the modelling process. In financial markets, realized measures have emerged as indispensable tools for capturing intraday price movements and improving tail risk forecasting. By incorporating high-frequency data, realized measures offer a refined understanding of market volatility that cannot be achieved using daily returns alone. The next section explores the concept of realized measures, focusing on their role in enhancing the precision of VaR and ES estimates.

2.2 Realized measures

In the context of tail risk forecasting, accurate volatility and risk measures are essential for predicting extreme market events such as large losses or financial crashes. Traditional models, which rely on daily returns, have proven to be insufficient in capturing the high-frequency dynamics that influence financial markets. These models often provide weak signals regarding current volatility levels and fail to adapt quickly to sudden market changes (Hansen et al., 2012). As a result, the need for more precise and timely measures of volatility has led to the emergence of realized measures, which leverage high-frequency data to offer a more informative and efficient view of volatility.

Realized measures (RMs) are derived from high-frequency intraday data and are considered superior to traditional volatility estimators. They offer a more accurate and unbiased representation of volatility by capturing the nuances of price dynamics that are often overlooked in daily returns (Andersen et al., 2001). These measures are particularly valuable in the forecasting of risk metrics such as VaR and ES, which are used to assess potential losses in extreme market scenarios (Barndorff-Nielsen et al., 2009).

A variety of realized measures have been proposed in the literature to estimate volatility and its components more accurately. Some of the most commonly used realized measures include Realized Variance (Andersen and Bollerslev (1998); Andersen et al. (2003)), Realized Range, (Christensen and Podolskij, 2007), Realized Kernel, (Barndorff-Nielsen et al. (2008); Barndorff-Nielsen and Shephard (2002)) and Bi-power Variation (Barndorff-Nielsen and Shephard, 2004). These measures have been applied in various models to improve the precision of volatility forecasts.

However, given the focus of this thesis on tail risk forecasting and the scope of the analysis, we concentrate on three realized measures that are particularly relevant for our study: RV, RK, and BV. Each of these measures was chosen for its distinct characteristics, reliability in providing estimates within the model, and widespread use in the literature. Moreover, each measure has distinct advantages in capturing different aspects of volatility dynamics and plays a crucial role in the accuracy of VaR and ES forecasts (Hansen et al., 2012).

Realized Variance (RV)

RV is one of the most widely used realized measures and is calculated as the sum of the squared intraday returns over a given period. Formally, it is expressed as:

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2,$$

where $r_{t,i}$ represents the intraday returns calculated over equally spaced intervals, and M is the number of intervals within the trading day. RV has become a standard measure for volatility estimation because it is unbiased and highly efficient (Andersen et al., 2001).

However, RV is not without its limitations. It can be a highly noisy estimate of volatility and is sensitive to overnight changes (Barndorff-Nielsen et al., 2008). Despite these challenges, RV remains a foundational measure in volatility forecasting, particularly when used in conjunction with other realized measures to improve robustness and accuracy (Andersen et al., 2003).

Realized Kernel (RK)

RK is a more advanced realized measure that was introduced to address the noise and market frictions often encountered in high-frequency data. RK combines intraday volatility estimation with kernel smoothing to provide more robust and efficient volatility estimates. The key advantage of RK is its ability to handle noisy data, making it particularly useful when working with high-frequency data prone to measurement errors (Barndorff-Nielsen et al., 2008). Formally, RK can be calculated by applying kernel smoothing to the high-frequency return data, allowing for a smoother and more stable estimate of volatility over time. RK is particularly valuable when working with noisy data, as it reduces the impact of outliers and small market frictions, leading to more reliable volatility forecasts (Barndorff-Nielsen et al., 2009). By incorporating RK into volatility models, researchers and practitioners can obtain more precise estimates of volatility, which can then be used to improve the accuracy of VaR and ES forecasts (Hansen et al., 2012).

The non-negative estimator as defined by Barndorff-Nielsen et al. (2008);Barndorff-Nielsen et al. (2009) as follows:

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \tag{2.12}$$

where $\gamma_h = \sum_{j=|h|+1}^{n} x_{j,t}x_{j-|h|,t}$.. The kernel weight function, $k(x)$ in the above equation is selected as the Parzen weight function for its compliance with the smoothness conditions, $k'(0) = k'(1) = 0$. This ensures a non-negative estimate. The Parzen kernel function is defined as:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2, \\ 2(1-x)^3 & 1/2 \leq x \leq 1, \\ 0 & x > 1, \end{cases} \tag{2.13}$$

where x_j is the high-frequency return in equations 2.12 and 2.13. H is the standard bandwith spelt out in Barndorff-Nielsen et al. (2009), which is given by;

$$H^* = c^* \zeta^{4/5} n^{3/5}, c^* = (12^2/0.269)^{1/5} = 3.5134, \zeta^2 = \frac{\omega^2}{T \int_0^T \sigma_u^4 du}. \tag{2.14}$$

Further details are available in Barndorff-Nielsen et al. (2009).

Bi-power Variation (BV)

BV is another realized measure used to estimate volatility. Unlike RV, which sums the squared intraday returns, BV is based on the absolute returns and is particularly effective in capturing jump components in volatility. It is defined as:

$$BV_t = \sum_{i=1}^{M} |r_{t,i}| \, |r_{t,i-1}| \, ,$$

where $r_{t,i}$ represents the intraday return at each time step. BV is designed to capture both the continuous variation in prices as well as the discrete jumps that may occur due to market events or shocks (Barndorff-Nielsen and Shephard, 2004).

BV is especially useful in scenarios where volatility is driven by large, abrupt changes in price, such as market jumps or shocks, which are often not captured by traditional volatility measures. By incorporating BV into risk models, practitioners can obtain a more comprehensive view of market dynamics and improve the accuracy of tail risk forecasts, particularly in extreme market conditions (Christensen and Podolskij, 2007).

The integration of realized measures into tail risk forecasting models enhances the precision of risk predictions, particularly in estimating extreme losses. Tail risk metrics such as VaR and ES are highly sensitive to the accuracy of the volatility estimates used in their calculation. Realized measures, by incorporating high-frequency data, allow for more timely adjustments to volatility estimates, which is essential in forecasting potential large losses in volatile market conditions (Andersen et al., 2003).

By capturing the finer details of intraday price movements, realized measures provide a richer and more reliable source of information for risk modelling. The use of multiple realized measures—such as RV, RK, and BV—can help mitigate the limitations of any single measure, offering a more robust and accurate framework for estimating volatility and forecasting risk.

As a result, the use of realized measures is essential for improving the effectiveness of tail risk forecasting, ensuring that models are better equipped to handle extreme market events and provide more reliable estimates of potential losses (Hansen and Huang, 2016).

2.3 Chapter summary

This chapter focused on tail risk forecasting and the role of realized measures in improving risk estimates. We discusse the key tail risk measures, VaR and ES, and explored various evaluation methods including VRate, the DQ test, Quantile loss, Joint loss Functions, and the MCS test. These techniques assess the accuracy and robustness of VaR and ES forecasts.

Additionally, the chapter highlighted the importance of realized measures, such as RV, RK, and BV, in forecasting volatility. By utilizing high-frequency data, these measures offer more precise and timely volatility estimates, significantly enhancing the forecasting of tail risks like VaR and ES.

Chapter 3

Foundations on Bayesian Inference and Deep Learning

This chapter outlines the methodological frameworks underpinning this thesis. It begins by introducing Bayesian inference, with a focus on the application of adaptive MCMC methods. The chapter then explores loss-based Bayesian inference, providing a detailed discussion of the Sequential Monte Carlo (SMC) algorithm. Finally, an overview of deep learning is presented, emphasising the role of Recurrent Neural Networks (RNNs) in modelling sequential data.

3.1 Introduction to Bayesian Inference

Bayesian inference is a powerful statistical framework that allows for the updating of probabilities based on prior knowledge and observed data. Central to this approach is Bayes' Theorem, which describes how to revise the probability of a hypothesis or model after observing new evidence. Specifically, Bayesian inference involves two main probability statements: the posterior distribution, $P(\theta|y)$, which updates our belief about the parameter θ after observing the data y, and the posterior predictive distribution, $P(y^{'}|y)$, which forecasts future or unobserved data y' based on the current data y (Gelman et al., 2013). These distributions provide a full characterization of uncertainty regarding model parameters and future observations. Mathematically, the posterior distribution is given by:

$$P(\theta|y) \propto P(\theta) \times P(y|\theta), \tag{3.1}$$

where, $P(\theta|y)$ is the posterior distribution of the model parameters θ given the data y, $P(\theta)$ is the prior distribution representing our belief about θ before observing any data, $P(y|\theta)$ is the likelihood function, which models the probability of the data y given the parameters θ.

The posterior distribution $P(\theta|y)$ updates our prior belief by incorporating information from the observed data. As new data becomes available, this posterior distribution can be updated in a sequential manner, where the posterior from one step serves as the prior for the next, allowing for continuous learning.

While the conceptual framework of Bayesian inference is straightforward, implementing it in practice often requires sophisticated computational methods, particularly for complex models with high-dimensional parameter spaces. Markov Chain Monte Carlo (MCMC) methods are widely used to sample from the posterior distribution when it is analytically intractable. MCMC generates a sequence of samples that approximate the target posterior distribution, with each sample being dependent on the previous one, thereby forming a Markov chain. The iterative process of MCMC allows for efficient exploration of complex parameter spaces, making it particularly useful for fitting high-dimensional models (Gelman et al., 2013).

In this thesis, we apply Bayesian inference using advanced MCMC techniques, including adaptive MCMC, which improves sampling efficiency by adjusting the sampling strategy based on the current state of the chain (Cowles, 2013). Additionally, we employ Sequential Monte Carlo (SMC) methods, which adaptively propagate a set of particles (samples) over time to estimate posterior distributions in models with sequential data (Chen et al., 2012). SMC is especially useful in scenarios where traditional MCMC may be computationally inefficient, such as when working with time-varying data. It is particularly efficient for expanding-window sequential forecasting, which is essential in volatility modeling and forecasting, especially in the context of financial risk.

Bayesian methods offer several advantages in fields such as financial modelling and machine learning, where they not only provide robust parameter estimation but also allow for the quantification of uncertainty, which is critical for risk forecasting and decision-making (Li et al. (2020); Nguyen et al. (2022a)). By integrating prior information with observed data, Bayesian approaches allow for continuous learning and better decision-making, making them particularly suitable for applications like VaR and ES forecasting.

3.1.1 Adaptive MCMC Algorithm

MCMC methods are a cornerstone of Bayesian inference, enabling sampling from complex posterior distributions. The efficiency of MCMC relies heavily on the proposal distribution, which generates candidate samples during the algorithm. However, since the true shape of the target posterior is unknown, selecting an appropriate proposal distribution is challenging. Adaptive MCMC methods address this problem by tuning the proposal distribution dynamically during the sampling process, thereby improving convergence and mixing efficiency (Chen and So (2006); Chen et al. (2012)). These algorithms aim to improve computational efficiency while maintaining the theoretical properties of MCMC, such as stationarity and ergodicity (Haario et al., 1999).

In high-dimensional spaces, where the parameter space can be large and highly correlated, standard MCMC methods, such as the Random Walk Metropolis-Hastings (RWMH) algorithm, often struggle with poor mixing and slow convergence. In such settings, the posterior distribution may exhibit strong correlations between parameters, leading to inefficient exploration and long mixing times. As Roberts and Rosenthal (2009) point out, the scaling of the proposal distribution in the random walk Metropolis algorithm plays a crucial role in the efficiency of sampling. In high-dimensional settings, improper scaling can lead to poor performance, especially when the posterior distribution is not well-explored by the random walk.

This issue becomes especially relevant in Bayesian tail risk forecasting, where models often involve high-dimensional posterior distributions with complex dependencies and correlations between parameters. Standard MCMC methods may struggle to sample effectively in such cases, making adaptive MCMC a more suitable approach. Adaptive MCMC methods address these challenges by dynamically adjusting the proposal distribution to better reflect the geometry of the posterior, allowing for more efficient exploration and faster convergence. This approach not only improves mixing but also ensures that the chain avoids local optima, helping it explore the entire posterior space more effectively.

These adaptive algorithms are particularly effective in high-dimensional spaces, where standard MCMC methods often struggle due to the correlations and complexity of the posterior distributions. The adaptive framework incorporates feedback from the sampling process, iteratively refining the proposal distribution, which improves convergence and mixing efficiency over time.

Gaussian Mixture Proposal and Covariance Adaptation

One of the key features of the adaptive MCMC algorithm is the use of a Gaussian mixture proposal during the burn-in phase. This mixture consists of two components: a smaller variance component for local refinement and a broader component for global exploration. This allows the algorithm to balance exploration and exploitation during the early stages of the sampling process. The covariance matrix adaptation follows the approach proposed by Haario et al. (2001), where the covariance matrix is updated dynamically based on the evolving sample distribution. In each iteration, the proposal distribution is refined based on the sample mean and covariance matrix from the burn-in phase, which is updated as:

$$C_{n+1} = \tfrac{n-1}{n} C_n + \tfrac{1}{n} \left(\theta_n - \overline{\theta}\right) \left(\theta_n - \overline{\theta}\right)^T,$$

where $\overline{\theta}$ is the running mean of the parameters. This covariance adaptation ensures that the proposal distribution increasingly reflects the posterior's geometry, facilitating faster convergence to the stationary distribution.

Additionally, when high correlations exist between parameters, a block update approach may be used, where correlated parameters are updated simultaneously to better reflect their dependencies. This strategy can significantly improve the efficiency of sampling in models with strong inter-parameter correlations.

Burn-In Phase and Acceptance Rate Tuning

The burn-in phase is the first stage of the adaptive MCMC algorithm. During this phase, a random walk Metropolis-Hastings (RWM) algorithm is used to generate candidate parameters θ^* as follows:

$$\theta^* = \theta^{j-1} + \epsilon,$$

where ϵ is a random step drawn from a Gaussian mixture proposal:

$$\epsilon = \rho N(0, \mathrm{diag}(c_1)) + (1-\rho) N(0, \omega \mathrm{diag}(c_1)),$$

where ρ is the mixing weight for the smaller variance component, typically set between 0.8 and 1, with $\rho = 0.95$ being a common choice (Chen et al., 2012). ω is a scaling factor for the broader component, often set to a large value, such as 100. The vector c_1 serves as a tuning parameter that adjusts the step size to achieve an optimal acceptance rate.

To ensure computational efficiency and adequate exploration, the acceptance rate is tuned to lie within a range of 0.15–0.5, as recommended by Gilks et al. (1995), with 0.234 being the theoretical optimum for high-dimensional parameter spaces (Roberts et al., 1997). These adjustments enhance the algorithm's ability to sample effectively from complex posterior distributions.

Updating the Proposal Distribution

Following the burn-in phase, the proposal distribution is updated iteratively to align with the posterior distribution. A common approach uses the sample mean $\overline{\theta}$ and covariance matrix C to defines a Gaussian proposal:

$$q(\theta^*|\theta^{(n)}) = (1-\beta)N(\theta^{(n)}, \sigma_{\text{opt}}C) + \beta(\theta^{(n)}, \sigma_{\text{ini}}I_d),$$

where $\sigma_{\text{opt}}C$ represents the tuned covariance matrix estimated from the burn-in phase. $\sigma_{\text{ini}}I_d$ introduces equal variance across all dimensions, allowing for a broader exploration of the parameter space during the burn-in phase. β is a small weight for large exploratory jumps. This adaptive formulation follows the methods proposed by Haario et al. (1999) and Chen et al. (2012), ensuring a balance between local refinement and global exploration.

MCMC Convergence and Efficiency Testing

To ensure the reliability and robustness of the adaptive MCMC algorithm, both convergence diagnostics and efficiency assessments are conducted. Specifically, the Gelman-Rubin diagnostic (Gelman et al., 2013) is utilized to evaluate the convergence of the MCMC chains, while the effective sample size (ESS) and autocorrelation time are employed to assess the sampling efficiency.

The convergence of the MCMC chains is assessed using the Gelman-Rubin diagnostic (Gelman et al., 2013), which compares the variance within each chain to the variance between chains. To apply this diagnostic, multiple independent chains are run, each initialised with wide ranging starting values, and only the post-burn-in iterations are considered for analysis. The process proceeds as follows for each parameter:

1. Chain Partitioning:

   Let m represent the number of chains, each of length n, following the burn-in period. For each chain j, the mean $\overline{\psi}_{\cdot j}$ and the overall mean $\overline{\psi}_{\cdot\cdot}$ are calculated:

   $$\overline{\psi}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{i,j}, \text{ and } \overline{\psi}_{\cdot\cdot} = \frac{1}{m}\sum_{j=1}^{m}\overline{\psi}_{\cdot j\cdot},$$

2. Variance Calculation:

   The between-chain variance (B) and within-chain variance (W) are computed as:

   $$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\overline{\psi}_{\cdot j} - \overline{\psi}_{\cdot\cdot}\right)^2, \text{ and } W = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(\psi_{i,j} - \overline{\psi}_{\cdot j}\right)^2\right)$$

3. Posterior Variance Estimate:

   The marginal posterior variance for the parameter ψ, denoted $\widehat{Var}(\psi)$, is a weighted average of the between-chain and within-chain variances:

   $$\widehat{Var}(\psi) = \frac{n-1}{n}W + \frac{1}{n}B.$$

4. Potential Scale Reduction Factor:

   The potential scale reduction factor, $\widehat{R}$, is calculated as:

   $$\widehat{R} = \sqrt{\frac{\widehat{Var}(\psi)}{W}}$$

   A value of $\widehat{R}$ close to 1 suggests satisfactory convergence of the chains to the stationary distribution, while values above 1.1 indicate that further iterations are required to improve convergence (Gelman et al., 2013).

Once convergence is confirmed, the efficiency of the MCMC chains is assessed using the effective sample size (ESS) and autocorrelation time.

1. Effective Sample Size (ESS):

   The ESS is an important metric that quantifies the number of independent samples obtained from the MCMC chains. To calculate the ESS, the following steps are followed:

   - Variogram Calculation:

     For each lag t, the variogram V_t is computed as;

     $$V_t = \frac{1}{m(n-t)}\sum_{j=1}^{m}\sum_{i=t+1}^{n}\left(\psi_{i,j} - \psi_{i-t,j}\right)^2$$

- Autocorrelation Estimate:
  The autocorrelation at lag t, $\widehat{\rho}_t$, is given by:

$$\widehat{\rho}_t = 1 - \frac{V_t}{2\widehat{Var}(\psi)}.$$

  The sum of autocorrelation estimates is truncated once the sum for two consecutive lags becomes negative, and the ESS is then estimated as:

$$\widehat{n}_{eff} = \frac{mn}{1+2\sum_{t=1}^{T}\widehat{\rho}_t},$$

  where T is the first odd positive integer such that $\widehat{\rho}_T + \widehat{\rho}_{T-1}$ becomes negative.

2. Autocorrelation Time:
   Autocorrelation time is another measure of sampling efficiency, quantifying the dependence between consecutive samples. It is computed as:

$$\widehat{\tau} = 1 + 2\sum_{t=1}^{T}\widehat{\rho}_t$$

   Shorter autocorrelation times indicate higher sampling efficiency, as it implies that consecutive samples are more independent (Chib and Jeliazkov, 2001).

Together, the Gelman-Rubin diagnostic, ESS, and autocorrelation time provide a comprehensive framework for verifying the convergence and efficiency of the adaptive MCMC algorithm. This multi-step approach ensures that the resulting parameter estimates are both reliable and efficient.

Adaptive MCMC methods have been widely adopted in financial econometrics. For example, Chen et al. (2012) used this approach to estimate dynamic financial models, while Gerlach et al. (2016) applied adaptive MCMC to Realized GARCH models. These studies highlight the effectiveness of adaptive MCMC in handling high-dimensional parameter spaces and improving computational performance (Wang et al., 2023).

While adaptive MCMC methods adjust the proposal distribution based on the sampling history to improve convergence, the next section explores loss-based Bayesian inference using Sequential Monte Carlo (SMC) methods. Unlike adaptive MCMC, which primarily focuses on optimizing sampling efficiency, loss-based SMC emphasizes minimizing the prediction error through sequential weighting and resampling techniques.

3.1.2 Sequential Monte Carlo (SMC) Algorithm

SMC methods are versatile Bayesian computation tools that approximate target distributions using a weighted set of particles. These particles evolve iteratively through reweighting, resampling, and mutation steps, enabling SMC to handle complex or high-dimensional models efficiently (Del Moral et al., 2006). This adaptability makes SMC particularly appealing for dynamic models and datasets with sequential structures.

SMC methods approximates the posterior distribution in the Bayesian framework given in equation (3.1) through a sequence of intermediate distributions, gradually transitioning from the prior to the posterior as new information is incorporated. Two key approaches to constructing the sequence of intermediate distributions are likelihood annealing and data annealing. Both share a common algorithmic framework, which is outlined below. Following algorithms are based on Nguyen et al. (2022b).

Likelihood Annealing: In likelihood annealing, intermediate distributions are defined as:

$$P_t(\theta) \propto P(\theta) \times P(y|\theta)^{\phi_t}, \tag{3.2}$$

where ϕ_t is a temperature parameter increasing from 0 to 1. This approach controls the influence of the likelihood function over the prior, facilitating a smooth transition to the posterior distribution (Del Moral et al., 2006). At iteration t, the likelihood annealing SMC algorithm has three steps:

- **Initialization:** Sample $\left\{\theta_j^0, W_j^0 = \frac{1}{M}\right\}_{j=1}^{M}$ from the prior distribution.
- **Reweighting:** Update particle weights for each iteration t:

  $$w_j^t = W_j^{t-1} \cdot \left[\frac{P(y|\theta_j^{t-1})^{\phi_t}}{P(y|\theta_j^{t-1})^{\phi_{t-1}}}\right],$$

  followed by normalization.

  Resampling: Resample particles when the effective sample size (ESS) drops below a threshold:

  $$ESS = \frac{1}{\sum_{j=1}^{M}(W_j^t)^2}$$

- **Mutation:** Apply MCMC steps to diversify particles. The resulting particles approximate the distribution $P_t(\theta)$.

Data Annealing: Data annealing constructs intermediate distributions using subsets of data $y_{1:t}$.

$$P_t(\theta) \propto P(\theta) \times P(y_{1:t}|\theta), \tag{3.3}$$

where t iterates over observations. This approach is effective for streaming data and models where processing the full dataset simultaneously is computationally prohibitive.

- **Initialization:** Sample $\left\{\theta_j^0, W_j^0 = \frac{1}{M}\right\}_{j=1}^{M}$ from the prior distribution.
- **Reweighting:** Update particle weights incrementally:

$$w_j^t = W_j^{t-1} \cdot p(y_t|\theta_j^{t-1}), \text{ followed by normalization.}$$

- **Resampling:** Resample particles when the ESS drops below a threshold:
- **Mutation:** Apply MCMC steps to improve particle diversity. Return the final particle set approximating the posterior distribution.

Likelihood-Based Bayesian SMC

Likelihood-based SMC methods form the foundation of traditional Bayesian inference by relying on the likelihood function to update beliefs about model parameters. This approach assumes a well-defined probabilistic model that explains how the observed data is generated. Using the likelihood function, the method evaluates the agreement between the model and the data, updating prior distributions to obtain posterior distributions.

A central feature of likelihood-based SMC is its reliance on annealing strategies to manage the transition from prior beliefs to the posterior. Annealing involves incrementally increasing the influence of the likelihood function over iterations, ensuring that the transition is smooth and computationally stable.

This framework is particularly effective when the underlying assumptions of the model, such as the distribution of the data, are well-specified and accurately reflect reality. However, as highlighted in the works of Bissiri et al. (2016) and Frazier et al. (2024), the reliance on strict distributional assumptions can limit its robustness, especially in situations where the data is complex or deviates from standard probabilistic structures. Despite these limitations, likelihood-based SMC remains a powerful and widely adopted tool for Bayesian inference and prediction in financial applications, as demonstrated in numerous empirical studies across risk forecasting and volatility modelling (Nguyen et al., 2022b).

Loss-Based Bayesian SMC

Loss-based SMC provides an alternative framework that moves beyond the constraints of likelihood-based methods by focusing on the use of loss functions rather than explicit probabilistic assumptions. This approach is rooted in generalised Bayesian inference, where the posterior distribution is constructed based on how well the parameters minimise a predefined loss function. Unlike the likelihood-based approach, which assumes a specific data-generating process, the loss-based method emphasises flexibility and robustness, allowing it to handle scenarios where the true distribution of the data is unknown or highly complex.

The development of loss-based inference has been discussed extensively in the literature, including the foundational work of Knoblauch et al. (2019) and Matsubara et al. (2021), who highlight its ability to address model misspecification. Similarly, Bissiri et al. (2016) and Frazier et al. (2024) argue that loss-based Bayesian methods provide a coherent framework for updating beliefs about parameters without requiring strict probabilistic assumptions. In financial risk forecasting, Li et al. (2023) advocate for loss-based Bayesian quantile regression as a more robust alternative to maximum likelihood estimation, particularly for modelling joint VaR and ES.

A key advantage of the loss-based approach is its adaptability. By focusing on minimising a loss function, researchers can tailor the method to different objectives, such as quantile scoring rules or other domain-specific measures. This flexibility makes it particularly suitable for financial applications, where the data often exhibit non-linear dynamics and irregularities that are difficult to capture using traditional likelihood-based methods.

Although loss-based SMC shares several procedural elements with likelihood-based SMC, such as the use of annealing and resampling to update particle weights, its reliance on loss functions enables it to operate effectively even when conventional model assumptions fail. As a result, this approach provides a robust alternative for sequential Bayesian inference, particularly in settings where data complexity or model uncertainty poses significant challenges.

3.2 Introduction to Deep Learning

Machine learning (ML) is a subfield of artificial intelligence that focuses on the development of algorithms that allow computers to learn from data and make decisions or predictions without explicit programming. Unlike traditional statistical methods, which often rely on predefined assumptions about the data, machine learning models automatically identify patterns and relationships within the data. This flexibility makes ML particularly effective for tasks involving high-dimensional and unstructured data such as images, text, and audio (Goodfellow, 2016).

Traditional methods, such as linear regression or analysis of variance, typically rely on assumptions like linearity or normality to model relationships between variables. These models aim to explain or interpret the data based on these assumptions. In contrast, machine learning focuses on learning from the data itself, which allows for greater flexibility and is often better suited for more complex or nonlinear data structures. ML techniques have proven especially effective in areas where traditional methods struggle, such as natural language processing, computer vision, and time series forecasting (James, 2013).

Machine learning algorithms are typically categorized into three main types based on the learning process. First, supervised learning, where models are trained using labeled data to predict outcomes or classify data points. Second, unsupervised learning, where algorithms identify hidden patterns or structures in data without any labels. And finally, Reinforcement learning, where models learn by interacting with an environment and receiving feedback in the form of rewards or penalties (Goodfellow, 2016).

The advantage of machine learning models is their ability to automatically learn patterns and improve over time, making them highly adaptable to new data. These models can continuously refine their performance as more data becomes available, which contrasts with traditional methods that may require re-specification or re-estimation when new data is introduced (Schmidhuber, 2015).

The development of deep learning, a subset of machine learning that employs neural networks with many layers, has significantly expanded the scope of what machine learning can achieve. This technique allows models to learn highly abstract features from large datasets, enabling breakthroughs in fields like image recognition, speech processing, and even time-series analysis. These advances will be explored in the next section, which introduces deep learning as an advanced form of machine learning.

Introduction to Deep Learning

Deep learning (DL) is a specialized subset of machine learning that involves neural networks with many layers, often referred to as deep neural networks (DNNs). These networks are designed to learn hierarchical representations of data, enabling models to automatically extract abstract features from raw input data, such as images, text, or speech, with minimal feature engineering. Deep learning techniques have revolutionized many fields, including computer vision, speech recognition, and natural language processing, by significantly improving performance over traditional machine learning models (Goodfellow, 2016; Schmidhuber, 2015).

At the core of deep learning is the artificial neural network (ANN), which is a computational model inspired by the structure and functioning of the human brain. A typical neural network consists of an input layer, one or more hidden layers, and an output layer. The hidden layers, which contain multiple neurons, are where deep learning models excel. They learn to represent data through non-linear transformations of the input data, allowing them to capture complex patterns and relationships.

One of the key features of deep learning is its ability to automatically learn features from data, making it highly effective for tasks where traditional machine learning techniques struggle. For instance, in image recognition, traditional methods required hand-crafted features (e.g., edges, corners), while deep learning models such as Convolutional Neural Networks (CNNs) automatically learn spatial hierarchies of features directly from raw image data. Similarly, deep learning models have enabled major advancements in natural language processing and time-series analysis by capturing sequential and temporal dependencies in data (Goodfellow, 2016), (LeCun et al., 2015).

However, deep learning models also come with challenges, such as the need for large amounts of labeled data and high computational resources. These models can also be prone to overfitting if not properly regularized. Despite these challenges, deep learning continues to set the benchmark for state-of-the-art performance in a wide range of applications.

3.2.1 Recurrent Neural Networks (RNNs)

RNNs are a class of deep learning models specifically designed to handle sequential data, where the order of the inputs carries important information. Unlike traditional feedforward neural networks (FNNs), which assume that all inputs are independent, RNNs are able to model time-dependent structures within the data by maintaining a hidden state that is updated recursively over time. This ability to "remember" previous inputs makes RNNs particularly effective for tasks where the order of the data matters, such as time-series forecasting, natural language processing, and speech recognition.

Let $D_t = (x_t, y_t), t = 1, 2...$ represent a sequence of data points, where x_t is the input at time t and y_t is the corresponding output. The goal of an RNN is to model the conditional mean $\hat{y}_t = E(y_t|x_t, D_{1:t-1})$, i.e., the expected output $\hat{y}_t$ given the current input x_t and all previous data $D_{1:t-1}$.

The basic RNN architecture can be represented with the following recurrence relations:

$$h_t = g_h\left(W^h\left[h_{t-1}, x_t\right] + b^h\right) \tag{3.4}$$

$$\hat{y}_t = g_y\left(W^y h_t + b^y\right), \tag{3.5}$$

where h_t is the hidden state at time t, W^h and W^y are weight matrices for the hidden state-to-input and hidden state-to-output connections, respectively. b^h and b^y are bias terms for the hidden and output layers, g_h and g_y are activation functions (commonly sigmoid or tanh) (Liu et al., 2024).

At each time step, the RNN updates its hidden state by combining the current input x_t and the previous hidden state h_{t-1}. This recursive structure allows RNNs to capture temporal dependencies and sequential patterns in the data. The output $\hat{y}_t$ is then computed using the current hidden state h_t.

This basic structure, while powerful, can face challenges when modeling long-term dependencies in time-series data. The hidden state h_t is updated using the weight matrix W^h, which could lead to issues such as exploding or vanishing gradients when dealing with long sequences. If the values in the weight matrix W^h are large, the gradients may explode, and if the values are small, the gradients may vanish, making training difficult for long time-series data (Goodfellow, 2016).

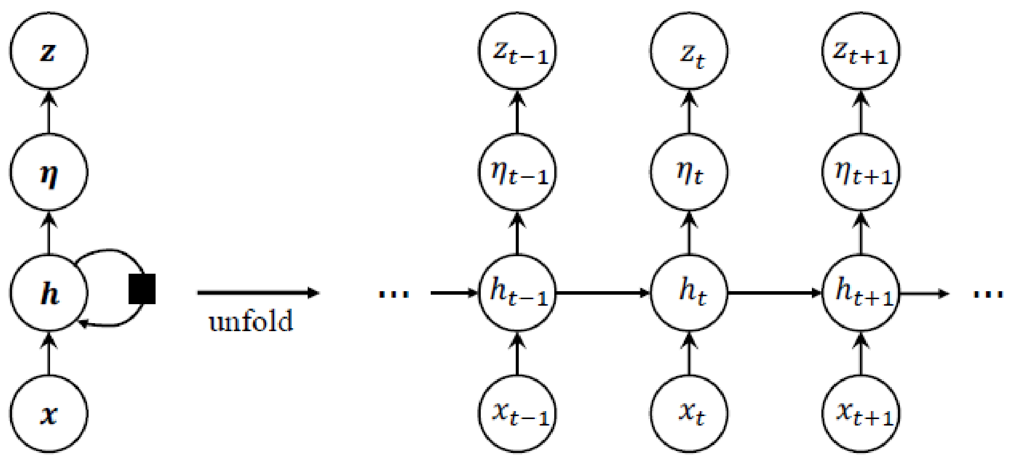

Figure 3.1. Graphical representation of the RNN model. Adapted from Nguyen et al. (2019).

A graphical representation of this RNN architecture is shown in Figure 3.1. On the left side, the diagram illustrates the RNN model where the hidden state h_t at each time step depends on both the previous hidden state h_{t-1} and the current input x_t. The right side shows an unfolded computational graph, where each node corresponds to a time step, clearly depicting the flow of information across time.

While the basic RNN structure is well-suited for capturing temporal dependencies in sequential data, it does face challenges in learning long-term dependencies due to issues like vanishing and exploding gradients. Despite these limitations, RNNs remain a powerful tool in time-series forecasting and have been successfully applied in various domains, including finance and risk modeling. In particular, RNNs can be leveraged to model complex, non-linear relationships and temporal dynamics in financial data, making them valuable for applications like tail risk forecasting, which will be discussed further in Chapter 5 and Chapter 6 of this thesis.

3.3 Chapter summary

This chapter introduced key concepts in Bayesian inference and deep learning. We discussed adaptive MCMC and SMC algorithms for efficient sampling and dynamic modeling in Bayesian inference. In the second part, we explored machine learning and deep learning, focusing on how neural networks can learn from data and make predictions. We then specifically introduced RNNs, highlighting their ability to model sequential data by capturing temporal dependencies. These foundational concepts provide a basis for the advanced techniques discussed in later chapters.

Chapter 4

Realized-ES-CAViaR-M model

In this chapter a semi-parametric VaR and ES forecasting framework employing multiple realized measures is developed. The proposed framework extends the realized exponential GARCH model to be semi-parametrically estimated, via a joint loss function, whilst extending existing quantile time series models to incorporate multiple realized measures. A quasi-likelihood is built, employing the asymmetric Laplace distribution that is directly linked to a joint loss function, which enables Bayesian inference for the proposed model. An adaptive Markov Chain Monte Carlo method is used for the model estimation. The empirical section evaluates the performance of the proposed framework with six stock markets from January 2000 to June 2022, covering the period of COVID-19. Three realized measures, including 5-minute realized variance (RV5), bi-power variation (BV), and realized kernel (RK), are incorporated and evaluated in the proposed framework. One-step-ahead 1% and 2.5% VaR and ES forecasting results of the proposed model are compared to a range of parametric and semi-parametric models, lending support to the effectiveness of the proposed framework.

4.1 Introduction

Financial risk management is an integral task for financial institutions. VaR is a standard tool for measuring and controlling financial market risks. Let $\mathcal{L}_t$ denote the information available at time t and

$$F_t(r) = Pr(r_t \leq r \mid \mathcal{L}_{t-1})$$

be the Cumulative Distribution Function (CDF) of the return r_t conditional on $\mathcal{L}_{t-1}$. Assuming that $F_t(\cdot)$ is strictly increasing and continuous on the real line $\mathbb{R}$, the one-step-ahead α-level VaR at time t can be defined as:

$$Q_t = F_t^{-1}(\alpha), \quad 0 < \alpha < 1.$$

However, as explained in chapter 2, VaR cannot measure the magnitude of the loss for violations and is not mathematically coherent, meaning that it is not a sub-additive measure and can favour non-diversification. Artzner et al. (1999) propose an alternative called ES, also called conditional VaR or tail VaR. ES calculates the expected loss conditional on exceeding a VaR threshold and is a coherent risk measure. The one-step-ahead α-level ES is the tail conditional expectation of r_t, i.e.:

$$\text{ES}_t = E(r_t \mid r_t \leq Q_t, \mathcal{L}_{t-1}).$$

The recent Basel III Accord (Basel Committee on Banking Supervision, 2023) places new emphasis on ES. This work focuses on the daily forecasting of VaR and ES on the lower/left tail. Following the Basel III Accord, the common $\alpha = 2.5\%$ probability level is studied in the paper. The more extreme 1% probability level is also investigated.

Volatility plays a crucial role in parametric tail risk forecasting. The GARCH model of Engle (1982a) and Bollerslev (1986) is widely used for modelling and forecasting volatility in the finance industry. Numerous extensions, such as EGARCH by Nelson (1991) and GJR-GARCH by Glosten et al. (1993), have been introduced to capture the well-known leverage effect. However, the volatility dynamics in these conventional GARCH models is driven by (daily) returns, which are considered potentially noisy signals for the volatility series.

The availability of high-frequency intra-day data has allowed the construction of many informative, efficient realized measures (RMs) of volatility. The most commonly used RMs include Realized variance (RV) (Andersen and Bollerslev (1998), Andersen et al. (2003)), Realized Range (RR) (Christensen and Podolskij (2007), Martens and Van Dijk (2007)), Realized

Kernel (RK) (Barndorff-Nielsen et al. (2009), and Bi-power variation (BV) (Barndorff-Nielsen and Shephard (2004)), etc.

Hansen et al. (2012) include a RM in their volatility equation via their realized GARCH framework, enabling joint modelling of returns and RMs using a measurement equation. Hansen and Huang (2016) extend the realized GARCH framework to include multiple RMs via the Realized Exponential GARCH (REGARCH) model. The REGARCH shows improved volatility forecasting performance compared to realized GARCH, and GARCH, demonstrating the usefulness of incorporating multiple realized measures in volatility modelling.

The tail risk forecasting accuracy of parametric models depends heavily on the choice of the distribution of the returns. Semi-parametric models, which do not rely on a specific return distribution, are also developed in the literature. Engle and Manganelli (2004) introduce the conditional auto-regressive VaR (CAViaR) model, which directly estimates VaR as the quantile of the conditional return distribution via a quantile regression framework. The model is optimized by minimising the quantile loss function. However, CAViaR models do not estimate ES.

Fissler and Ziegel (2016) show that VaR and ES are jointly elicitable for a class of joint loss functions, although ES is not elicitable by itself. This finding carries significant implications for the risk forecasting literature, particularly for researchers in the field of semi-parametric risk forecasting, who have new avenues to explore in joint VaR and ES modelling. Taylor (2019) proposes an ES-CAViaR framework to jointly and semi-parametrically estimate VaR and ES. A quasi-likelihood, built on the asymmetric Laplace (AL) distribution, allows the joint estimation of conditional VaR and ES. Taylor (2019) shows that the AL quasi-likelihood function falls into the class of strictly consistent loss functions developed by Fissler and Ziegel (2016). In the ES-CAViaR model, a CaViaR-type quantile equation models the VaR component. Then, ES is modelled via two proposed versions of a VaR to ES relationship: additive and multiplicative.

Gerlach and Wang (2020) incorporate a realized measure as an exogenous variable, extending ES-CAViaR models to the semi-parametric ES-X-CAViaR-X model class, finding improved VaR and ES forecast performance. Wang et al. (2023) further extend the work of Gerlach and Wang (2020) by introducing the semi-parametric Realized-ES-CAViaR framework, including a measurement equation to model the relationship between the tail risk measure and a RM; a leverage effect is also considered in the framework.

Three key facts motivate the development of our proposed framework. First, the REGARCH, using multiple RMs to model volatility, demonstrates improved performance compared to the realized GARCH, using only one RM. Second, the REGARCH is a parametric model requiring the specification of a return error distribution, while semi-parametric models, such as ES-CAViaR, do not, which is advantageous in many real return series. Third, incorporating a single RM into the semi-parametric modelling process, such as the ES-X-CAViaR-X or Realized-ES-CAViaR, can improve risk forecasting accuracy. Therefore, there is a gap in the literature regarding semi-parametric joint VaR and ES forecasting models with multiple realized measures; filling that gap is the primary aim of this paper.

The main contributions of this chapter are as follows. First, a new semi-parametric joint VaR and ES forecasting framework incorporating multiple RMs is proposed. This extends the quantile regression framework using multiple RMs as exogenous variables. The relationship between VaR and ES is modelled as time-varying and driven by the information from RMs. Further, a measurement equation is included in the framework to model the joint contemporaneous dependencies between the quantile series and multiple RMs. Second, an adaptive Bayesian MCMC algorithm is used to estimate the proposed model, including the parameters in the measurement equation variance-covariance matrix. Lastly, the effectiveness of the proposed framework is evaluated via a comprehensive empirical study, including 33 competing models and covering the period from January 2000 to June 2022.

This chapter is organized as follows. Section 4.2 reviews the relevant existing literature on tail risk forecasting models. Section 4.3 presents the proposed framework. The likelihood function and the adaptive Bayesian MCMC algorithm are presented in Section 4.4. Section 4.5 presents the empirical results. Section 4.6 summarizes the findings.

4.2 Background models

This section describes the relevant models used to forecast VaR and ES in the literature, while the properties of each model are described in the context of motivating the proposed framework. Fundamental concepts used in the model development process are also discussed.

4.2.1 Parametric GARCH-type models

Let $\mathbf{r} = \{r_t, t = 1, \ldots, T\}$ be a time series of daily returns. The key interest in parametric volatility modelling is the conditional variance, $\sigma_t^2 = \text{var}\,(r_t \mid \mathcal{L}_{t-1})$. σ_t is called the volatility. Here, $\mathbb{E}(r_t|\mathcal{L}_{t-1}) = 0$ is assumed, equivalent to working with demeaned returns in practice. The GARCH(1,1) model is:

$$r_t = \sigma_t z_t,$$
$$\sigma_t^2 = \omega_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where z_t is i.i.d. with zero mean and unit variance. Parametric approaches require a parametric distribution for z_t to be chosen, e.g., Gaussian or Student's t, to compute VaR and ES forecasts.

Francq and Zakoïan (2015) and Gao and Song (2008) consider a semi-parametric approach using historical simulation (HS), by modelling VaR and ES as constant multiples of the latent volatility σ_t which is assumed to follow a GARCH-type volatility model. Assuming a constant conditional return distribution with zero mean, the VaR and ES are modelled as:

$$Q_t = a_\alpha \sigma_t; \;\; \text{ES}_t = b_\alpha \sigma_t; \;\; \frac{\text{ES}_t}{Q_t} = \frac{b_\alpha}{a_\alpha} > 1, \tag{4.1}$$

where a_α and b_α are constant depending on the return distribution and can be estimated via HS on the standardized residuals $\frac{r_t}{\hat{\sigma}_t}$. The series $\hat{\sigma}_t$ is estimated first using quasi-maximum likelihood (QML) (Gao and Song, 2008).

4.2.2 Realized (E)GARCH model

A parametric realized GARCH framework, incorporating a RM into the volatility modelling process via a measurement equation, is developed in Hansen et al. (2012). Hansen and Huang (2016) further extend the realized GARCH by incorporating multiple RMs and propose the parametric realized EGARCH (REGARCH) model. A log-REGARCH specification is defined as:

$$\begin{aligned}
r_t &= \sigma_t z_t, \qquad (4.2)\\
\log(\sigma_t) &= \omega + \beta \log(\sigma_{t-1}) + \tau_1 z_{t-1} + \tau_2 \left(z_{t-1}^2 - 1\right) + \boldsymbol{\gamma}^T \mathbf{u}_{t-1},\\
\log(x_{j,t}) &= \xi_j + \varphi_j \log(\sigma_t) + \delta_{j,1} z_t + \delta_{j,2}(z_t^2 - 1) + u_{j,t}, \;\; j = 1, 2, ..., K.
\end{aligned}$$

The three log-REGARCH equations, in order, are the return equation, the GARCH or volatility equation, and the measurement equation, respectively. The measurement equation defines the contemporaneous relationship between the (ex-post) RMs of volatility and the (ex-ante) volatility. Here, K denotes the number of RMs, and $K = 1$ defines the original realized GARCH model. $\mathbf{x}_t = (x_{1,t}, ..., x_{K,t})^T$ is the vector of RMs, at time t, on the same scale as σ_t, e.g., $\sqrt{RV}$. z_t are i.i.d. with zero mean and unit variance and $\mathbf{u}_t = (u_{t,1}, ...)^T \sim N(0, \Sigma)$. z_t and $\mathbf{u}_t$ are mutually and serially independent. Further, the coefficient $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_K)^T$ of $\mathbf{u}_{t-1}$ represents how informative the RMs of day $t-1$ are about volatility on day t. The model uses two sets of leverage functions, both following the usual quadratic form, to model the leverage effect.

4.2.3 Semi-parametric ES-X-CAViaR-X model

Taylor (2019) proposes a semi-parametric class of models (called ES-CAViaR) to model the dynamics of VaR and ES jointly. Gerlach and Wang (2020) extend this model by adding various different ES to VaR relationships, allowing a single RM to influence both VaR and ES separately.

One of their proposed semi-parametric ES-X-CAViaR-X models is defined as follows:

$$\begin{aligned} Q_t &= \beta_0 + \beta_1 x_{t-1} + \beta_2 Q_{t-1}, \\ \omega_t &= \gamma_0 + \gamma_1 x_{t-1} + \gamma_2 \omega_{t-1}, \\ \mathrm{ES}_t &= Q_t - \omega_t. \end{aligned} \tag{4.3}$$

Here x_t is the RM. The dynamics of ES_t and Q_t have an additive, time-varying relationship, defined by ω_t, which is driven separately to Q_t, by the RM. The ω_t specification is directly generalized from a GARCH-type model. This specification allows the unknown conditional return distribution to change over time. The restriction $\gamma_0 \geq 0, \gamma_1 \geq 0, \gamma_2 \geq 0$ is employed to ensure that the VaR and ES series do not cross.

4.2.4 Semi-parametric Realized-ES-CAViaR models

The semi-parametric Realized-ES-CAViaR models (Wang et al., 2023) extend the ES-X-CAViaR-X model by incorporating a measurement equation:

Non-crossing of VaR and ES is enforced via the condition $\gamma_0 \geq 0, \gamma_1 \geq 0, \gamma_2 \geq 0$. The multiplicative error term $\epsilon_t = \frac{r_t}{Q_t}$ in the measurement equation facilitates the incorporation of the leverage effect into the model. Furthermore, the influence of the RM on VaR and ES is individually captured through the difference ω_t. As the relationship between VaR and ES varies over time, driven by the RM, the conditional return distribution also evolves. Compared to the ES-X-CAViaR-X model (Gerlach and Wang, 2020), the added measurement equation "completes" the model by regressing the RM on the quantile (can also be replaced with ES).

4.3 Proposed model

This section proposes a new Realized-ES-CAViaR-M model, employing multiple RMs to jointly and semi-parametrically model VaR and ES. The model extends the Realized-ES-CAViaR, via incorporating multiple RMs and adding a log specification, as well as the REGARCH, by virtue of being semi-parametric, i.e., the return distribution assumption is not required and the relationship between VaR and ES varies over time.

To motivate these proposals, now we present the process on developing the Realized-ES-CAViaR-M model. Given a REGARCH (equation (4.2)) with a parametric return distribution z_t in the return equation $r_t = \sigma_t z_t$, we have $\sigma_t = \frac{Q_t}{a_\alpha}$, meaning the quantile Q_t is proportional to the volatility σ_t. The constant a_α is equal to the CDF inverse of the selected parametric return distribution.

Since in the proposed model we do not assume the value of a_α, no return distribution is assumed and we have a semi-parametric approach, as defined in equation (4.1). Further, we define the multiplicative error $\epsilon_t = \frac{r_t}{Q_t}$ as that in the Realized-ES-CAViaR, then we have $\epsilon_t = \frac{\sigma_t z_t}{Q_t} = \frac{z_t}{a_\alpha}$. Substituting $\sigma_t = \frac{Q_t}{a_\alpha}$ and $z_t = \epsilon_t a_\alpha$ into the REGARCH framework and removing the return equation produce:

$$\log(\frac{Q_t}{a_\alpha}) = \omega + \beta \log(\frac{Q_{t-1}}{a_\alpha}) + \tau_1 \epsilon_{t-1} a_\alpha + \tau_2 \left(\epsilon_{t-1}^2 a_\alpha^2 - 1\right) + \boldsymbol{\gamma}^T \mathbf{u}_{t-1}, \tag{4.4}$$

$$\log(x_{j,t}) = \xi_j + \varphi_j \log(\frac{Q_t}{a_\alpha}) + \delta_{j,1} \epsilon_t a_\alpha + \delta_{j,2} (\epsilon_t^2 a_\alpha^2 - 1) + u_{j,t},\ j = 1, 2, ..., K. \tag{4.5}$$

Here, both Q_t and a_α have negative values, as the left tail 1% and 2.5% probability levels are considered. Multiplying both side of (4.4) with a constant $c = \log(-a_\alpha)$ and rearranging equation (4.5), we have:

$$\begin{aligned} \log(-Q_t) &= c\omega + \beta \log(-Q_{t-1}) + c\tau_1 a_\alpha \epsilon_{t-1} + c\tau_2 a_\alpha^2 \epsilon_{t-1}^2 - c\tau_2 + c\boldsymbol{\gamma}^T \mathbf{u}_{t-1}, \\ \log(x_{j,t}) &= \xi_j - \varphi_j c \log(-Q_t) + \delta_{j,1} a_\alpha \epsilon_t + \delta_{j,2} a_\alpha^2 \epsilon_t^2 - \delta_{j,2} + u_{j,t},\ j = 1, 2, ..., K. \end{aligned} \tag{4.6}$$

Setting $w^* = c\omega - c\tau_2, \tau_1^* = c\tau_1 a_\alpha, \tau_2^* = c\tau_2 a_\alpha^2, \boldsymbol{\gamma}^{*T} = c\boldsymbol{\gamma}^T, \xi_j^* = \xi_j - \delta_{j,2}, \varphi_j^* = -\varphi_j c, \delta_{j,1}^* = \delta_{j,1} a_\alpha, \delta_{j,2}^* = \delta_{j,2} a_\alpha^2$, the resulting model is:

$$\begin{aligned} \log(-Q_t) &= \omega^* + \beta \log(-Q_{t-1}) + \tau_1^* \epsilon_{t-1} + \tau_2^* \epsilon_{t-1}^2 + \boldsymbol{\gamma}^{*T} \mathbf{u}_{t-1}, \\ \log(x_{j,t}) &= \xi_j^* + \varphi_j^* \log(-Q_t) + \delta_{j,1}^* \epsilon_t + \delta_{j,2}^* \epsilon_t^2 + u_{j,t};\ j = 1, 2, ..., K. \end{aligned} \tag{4.7}$$

The current framework does not contain the ES component. Wang et al. (2023) develop an additive VaR to ES time varying w_t component which is directly driven by the realized measure. We extend the approach by developing a ω_t component that is also separately driven by multiple realized measures. Therefore, the proposed model is specified as:

Realized-ES-CAViaR-M

$$\log(-Q_t) = \omega + \beta \log(-Q_{t-1}) + \tau_1 \epsilon_{t-1} + \tau_2 \epsilon_{t-1}^2 + \boldsymbol{\gamma}^T \mathbf{u}_{t-1}, \tag{4.8}$$

$$\omega_t = \nu_0 + \nu_1 \omega_{t-1} + \boldsymbol{\psi}^T |\mathbf{u}|_{t-1}, \tag{4.9}$$

$$\mathrm{ES}_t = Q_t - \omega_t, \tag{4.10}$$

$$\log(x_{j,t}) = \xi_j + \varphi_j \log(-Q_t) + \delta_{j,1} \epsilon_t + \delta_{j,2} \epsilon_t^2 + u_{j,t};\ j = 1, 2, ..., K. \tag{4.11}$$

We omit the * in the derived model (4.7) for cleanness of the presentation. The model contains four equations: the *quantile* equation (4.8), the VaR-ES difference ω_t equation (4.9), the *ES* equation (4.10) and the *measurement* equations (4.11). As in REGARCH, K is the number of realized measures and $K = 1$ gives a log specification of the Realized-ES-CAViaR. Here $x_{j,t}$ is the square root of the RM, i.e., on the same scale as volatility. The measurement error vector

$\mathbf{u}_t \overset{\text{i.i.d.}}{\sim} N(0, \Sigma)$, as standard. Σ is the variance-covariance matrix of $\mathbf{u}_t$, with dimension $K \times K$. The key developments of the model are now discussed.

The quantile equation (4.8) extends the existing quantile regression (CAViaR) by introducing a log specification, including the leverage effect term and incorporating the information from multiple RMs. This makes the model analogous to REGARCH. This work studies the left tail quantile, for example, $\alpha = 1\%; 2.5\%$, where each quantile Q_t is less than 0, leading to the utilization of $(-Q_t)$ in the log operator. Incorporating a log specification also guarantees that $(-Q_t)$ is always positive, thus Q_t is guaranteed to be negative for the studied left tail. The quadratic leverage effect specification as in REGARCH is followed. The regression coefficients γ capture how influential the K lagged RMs are on next period (log-) quantile.

The ω_t equation (4.9) and ES equation (4.10) capture the time-varying and additive relationship between VaR and ES, all driven separately by the lagged RM vector, whose individual effects on the VaR to ES difference are given by $\boldsymbol{\psi}^T$, where $\boldsymbol{\psi} = (\psi_1, ..., \psi_K)^T$. Again, we constrain $\nu_0 \geq 0, \nu_1 \geq 0, \boldsymbol{\psi} \geq 0$ to ensure that the VaR and ES series do not cross. Since the ES_t is modelled as Q_t minus a non-negative time varying ω_t component, the leverage effect included in the quantile equation is implicitly applied to ES as well. Other relationships between VaR and ES, such as the multiplicative one in Gerlach and Wang (2020), could also be explored.

The measurement equations (4.11) complete the model by providing a way to model the joint contemporaneous dependence between the risk level and multiple RMs $x_{j,t}; j = 1, 2, ..., K$. The leverage function follows the quadratic form as in REGARCH.

Comparing to the Realized-ES-CAViaR model, the Realized-ES-CAViaR-M extends it via incorporating the information of multiple realized measures during the quantile and ES forecasting process. Further, a log specification is used in Realized-ES-CAViaR-M. As below, we also introduce the Log-Realized-ES-CAViaR which aims to investigate the impact of adding the log specification into the quantile regression and measurement equations. The model will be compared with other models in the empirical study.

Log-Realized-ES-CAViaR

$$\begin{aligned}
\log(-Q_t) &= \beta_0 + \beta_1 \log(x_{t-1}) + \beta_2 \log(-Q_{t-1}), \\
\omega_t &= \gamma_0 + \gamma_1 x_{t-1} + \gamma_2 \omega_{t-1}, \\
\text{ES}_t &= Q_t - \omega_t, \\
\log(x_t) &= \xi + \phi \log(-Q_t) + \tau_1 \epsilon_t + \tau_2 \left(\epsilon_t^2 - E(\epsilon^2)\right) + u_t.
\end{aligned} \tag{4.12}$$

4.4 Likelihood and model estimation

CAViaR-type models are typically estimated via minimising the quantile loss function, for which the latent quantile series is strictly consistent. Engle and Manganelli (2004) suggest that solutions to the optimization of the CAViaR type models can be heavily dependent on the chosen initial values. Li et al. (2023) demonstrate that maximum likelihood based quantile regression models are sensitive to initial conditions and advocate for using Bayesian approach as a more reliable alternative. With an intensive simulation study, Wang et al. (2023) have shown that Bayesian estimator is more accurate than the AL based QML estimator regarding parameter estimation and risk forecasting. Further, to conduct statistical inference Bayesian method can conveniently help us capture the parameter estimation uncertainty, by calculating the 95% credible intervals using the MCMC chains. Therefore, this work also employs the Bayesian methods.

4.4.1 Likelihood function for the proposed model

As discussed in Sections 4.2.3 and 4.2.4, the ES-CAViaR, ES-X-CAViaR-X and Realized-ES-CAViaR models are semi-parametric. However, Bayesian methods typically employ a parametric distributional assumption to form a likelihood.

Koenker and Machado (1999) note that the conventional quantile regression estimator is equivalent to an MLE based on the AL density, with a mode at the quantile. Discovering a specific link between ES_t and a dynamic σ_t, for the AL distribution, Taylor (2019) further extends this result to produce the conditional density function:

$$f_t(r) = \frac{(\alpha - 1)}{\text{ES}_t} \exp\left(\frac{(r_t - Q_t)(\alpha - I(r_t \leq Q_t))}{\alpha \text{ES}_t}\right), \tag{4.13}$$

As shown in Taylor (2019), the negative logarithm of this AL-based density function is strictly consistent for Q_t and ES_t jointly, meaning that it fits into the class of strictly consistent loss functions developed by Fissler and Ziegel (2016).

This density then allows a quasi-likelihood function to be built, given models for Q_t and ES_t, assuming a zero mean return, thus allowing Bayesian methods to be employed. Since r_t cannot follow an AL distribution with a mode at Q_t, the AL-based likelihood built on equation (4.13) is a quasi-likelihood function, whose mode coincides with the minimum of the joint loss function. The quasi-log-likelihood is then:

$$\ell(\mathbf{r}; \boldsymbol{\theta}) = \sum_{t=1}^{T} \left(\log \frac{(\alpha - 1)}{\text{ES}_t} + \frac{(r_t - Q_t)(\alpha - I(r_t \leq Q_t))}{\alpha \text{ES}_t}\right), \tag{4.14}$$

where $\mathbf{r} = \{r_1, r_2, ...r_T\}$ and the parameter vector is $\boldsymbol{\theta}$.

The full model likelihood also includes parts from the measurement equations. The AL-based return quasi-log-likelihood (4.14) combines with the likelihood for the RMs to produce the full quasi-log-likelihood for the proposed Realized-ES-CAViaR-M model:

$$\ell(\mathbf{r},\mathbf{X};\boldsymbol{\theta},\Sigma)=\ell(\mathbf{r};\boldsymbol{\theta})+\ell(\mathbf{X}|\mathbf{r};\boldsymbol{\theta},\Sigma),$$

where $\mathbf{X}$ is the set of multiple RMs: $\{x_{1,t},x_{2,t},\ldots,x_{K,t}\}$ and Σ is the covariance matrix of the measurement errors $\mathbf{u}_t$.

Thus, the full quasi-log-likelihood of the proposed model can be written as:

$$\begin{aligned}\ell(\mathbf{r},\mathbf{X};\boldsymbol{\theta},\Sigma)=&\sum_{t=1}^{n}\left(\log\frac{(\alpha-1)}{\mathrm{ES}_t}+\frac{(r_t-Q_t)(\alpha-I(r_t\le Q_t))}{\alpha\mathrm{ES}_t}\right)\\&-\frac{1}{2}\sum_{t=1}^{n}\left(k\log(2\pi)+\log(|\Sigma|)+\mathbf{u}_t^{'}(\boldsymbol{\theta})\Sigma^{-1}\mathbf{u}_t(\boldsymbol{\theta})\right),\end{aligned}\tag{4.15}$$

For any given value of $\boldsymbol{\theta}$, Hansen and Huang (2016) show that the RM based Gaussian likelihood yields the partial maximization concerning Σ as:

$$\hat{\Sigma}(\boldsymbol{\theta})=\frac{1}{n}\sum_{t=1}^{n}\mathbf{u}_t(\boldsymbol{\theta})\mathbf{u}_t(\boldsymbol{\theta})^{'},$$

where they point out that $\mathbf{u}_t$ in the above equation depends on θ, but does not depend on the covariance matrix Σ. Therefore, the maximization problem is simplified to finding $\operatorname{argmax}_{\boldsymbol{\theta}}\ell(\mathbf{r},\mathbf{X};\boldsymbol{\theta},\hat{\Sigma}(\boldsymbol{\theta}))$ since

$$\sum_{t-1}^{n}\mathbf{u}_t^{'}(\boldsymbol{\theta})\Sigma(\theta)^{-1}\mathbf{u}_t(\theta)=tr\left\{\sum_{t=1}^{n}\hat{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{u}_t(\boldsymbol{\theta})\mathbf{u}_t^{'}(\boldsymbol{\theta})\right\}=nK$$

which does not depend on $\boldsymbol{\theta}$.

From a Bayesian standpoint, the likelihood is an inverse Wishart distribution in Σ, which can then be integrated out from the likelihood, as discussed in the next section.

4.4.2 Bayesian estimation

The quasi-log-likelihood in (4.15) includes the logarithm of the multivariate Gaussian density for each measurement error vector $\mathbf{u}_t$. Denoting $\boldsymbol{\theta}_{-\Sigma}$ as the vector of all model parameters excluding Σ, the integrated measurement likelihood is:

$$\begin{aligned}p(\mathbf{X}|\mathbf{r},\boldsymbol{\theta}_{-\Sigma})&=\int p(\mathbf{X}|\mathbf{r},\boldsymbol{\theta})p(\Sigma)d\Sigma\\&=\int(2\pi)^{-0.5KT}|\Sigma|^{-0.5T}\exp\left[-0.5\sum_{t=1}^{T}\mathbf{u}_t^{'}\Sigma^{-1}\mathbf{u}_t\right]p(\Sigma)d\Sigma\,.\end{aligned}$$

Under a standard Jeffreys prior $p(\Sigma) \propto |\Sigma|^{-0.5(K+1)}$, the integration is proportional to an inverse Wishart density function in Σ, so the integral can be shown to be:

$$p(\mathbf{X}|\mathbf{r}, \boldsymbol{\theta}_{-\Sigma}) \propto |\hat{\Sigma}|^{-0.5(T-K-1)} \ , \tag{4.16}$$

where the terms not in Σ are ignored, and $\hat{\Sigma} = \frac{1}{T-K-1} \sum_{t=1}^{T} \mathbf{u}_t \mathbf{u}_t^{'}$ is the usual sample variance covariance estimator of Σ. Thus, the full integrated quasi-likelihood, replacing the log full measurement likelihood in (4.15) with its integrated version in (4.16), logged, is:

$$\begin{aligned} \ell(\mathbf{r}, \mathbf{X}; \boldsymbol{\theta}_{-\Sigma}) = \sum_{t=1}^{n} & \left(\log \frac{(\alpha - 1)}{\mathrm{ES}_t} + \frac{(r_t - Q_t)(\alpha - I(r_t \le Q_t))}{\alpha \mathrm{ES}_t} \right) \\ & - \frac{1}{2}(T - K - 1) \log(|\hat{\Sigma}|) \, . \end{aligned} \tag{4.17}$$

Priors are chosen to be flat over the regions sufficient for non-negativity of ω_t in equation (4.9), combined with the others, quite liberal and wide, limits to ensure finite parameter ranges and a proper prior. Thus, we choose $\pi(\boldsymbol{\theta}_{-\Sigma}) \propto I(A)$, being a flat prior for $\boldsymbol{\theta}_{-\Sigma}$ over the region A, and 0 elsewhere. To ensure finite parameter ranges, A restricts each element of $\boldsymbol{\theta}_{-\Sigma}$ to be inside $(-D_0, D_0)$. For example, stationarity requires $|\beta_1| < 1$, i.e., $D_0 = 1$ for β_1. In the empirical study, we choose $D_0 = 3$ for the other parameters, which is sufficiently large based on our analyses. To ensure non-negativity of ω_t, region A further restricts $\nu_0 \ge 0, \nu_1 \ge 0, \boldsymbol{\psi} \ge 0$. These priors take a log transformation and are combined with the integrated quasi-log-likelihood in equation (4.17) to construct the posterior distribution.

Following Chen et al. (2022), in the MCMC algorithm, to assist with the speed of mixing, the parameter vector is simulated in blocks, i.e., each block of parameters is simulated from its conditional posterior. Blocks are chosen so that parameters within each block tend to be more correlated in the posterior, whilst parameters not in the same block are less correlated; this aids in faster mixing and convergence. Table 4.1 details the blocking structure, based on the number of RMs in the model. The block-wise proposals are generated and accepted with the usual Metropolis algorithm, e.g., see Chen et al. (2022).

Table 4.1. Block structure of the employed MCMC. (Parameters in the variance-covariance matrix Σ have been integrated out. B_i represents the i^{th} parameter block)

Block number	**k=1** (12 parameters)	**k=2** (18 parameters)	**k=3** (24 parameters)
B_1	$\{\omega, \beta, \tau_1, \tau_2\}$	$\{\omega, \beta, \tau_1, \tau_2\}$	$\{\omega, \beta, \tau_1, \tau_2\}$
B_2	$\{\gamma, \delta_{11}, \delta_{12}\}$	$\{\gamma_1, \gamma_2, \xi_1, \xi_2\}$	$\{\gamma_1, \gamma_2, \gamma_3\}$
B_3	$\{\nu_0, \nu_1\}$	$\{\varphi_1, \varphi_2\}$	$\{\xi_1, \xi_2, \xi_3\}$
B_4	$\{\xi, \varphi, \psi\}$	$\{\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}\}$	$\{\varphi_1, \varphi_2, \varphi_3\}$
B_5		$\{\nu_0, \nu_1\}$	$\{\delta_{11}, \delta_{21}, \delta_{31}\}$
B_6		$\{\psi_1, \psi_2\}$	$\{\delta_{12}, \delta_{22}, \delta_{32}\}$
B_7			$\{\nu_0, \nu_1\}$
B_8			$\{\psi_1, \psi_2, \psi_3\}$

The proposal density is a mixture of three multivariate Gaussian proposal distributions, with a random walk mean vector for each block. The proposal variance-covariance matrix of each block in each mixture element is $C_i\Sigma$, where $C_1 = 1; C_2 = 100; C_3 = 0.01$, with Σ initially set to $\frac{2.38}{\sqrt{(d_i)}} I_{d_i}$, where d_i is the dimension of the i^{th} block and I_{d_i} is the identity matrix of dimension d_i. The vector of mixing weights (w) is (0.7, 0.15, 0.15), allowing both small and large proposal jumps to be considered. The covariance matrix for each block is tuned as in Chen et al. (2022), with target acceptance rates as in Roberts et al. (1997); i.e., acceptance rates: 0.44 for $d_i = 1$, 0.35 when $2 \leq d_i \leq 4$ and 0.234 for $d_i > 4$. The algorithm is run in epochs, where each epoch is $N = 20,000$ iterations, until the mean total absolute percentage difference in the sample variances of epoch iterates, over all parameters, is less than 10% (see Chen et al. (2022) for details); typically this takes 3 or 4 epochs. The last 10,000 iterates of the final epoch are used for estimation and inference.

4.5 Data and empirical study

This section presents the empirical analysis conducted to evaluate the performance of the proposed framework in forecasting VaR and ES. A comprehensive overview of the data and model comparisons is provided, followed by the estimation of model parameters and the evaluation of forecasting accuracy. This approach enables a clear comparison with benchmark models, highlighting the strengths and limitations of the proposed methodology.

4.5.1 Data description

Daily closing prices and RM data from January 2000 to June 2022 were downloaded from Oxford-man Institute's realized library (Heber et al., 2009). Three common RMs, including RV5, RK, and BV are considered. Six market indices, including S&P500 and NASDAQ in the US, FTSE 100 (UK), DAX (Germany), SMI (Swiss), and HSI (Hong Kong), are included in the study. Each data set is split into an initial in-sample period, from January 2000 to December 2011, and an out-of-sample forecasting period from January 2012 to June 2022. Our out-of-sample period includes the COVID-19 period. Figure 4.5.1 displays a time series plot of the absolute value of daily return, RV5, RK and BV of S&P500 for exposition.

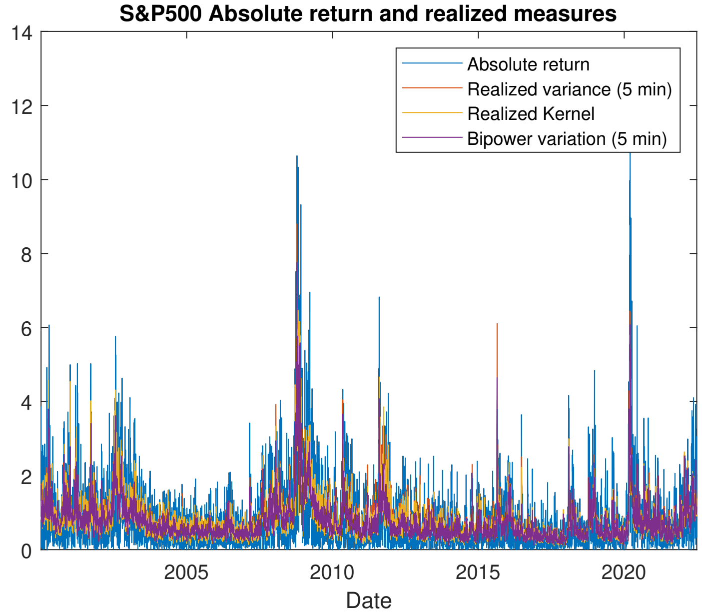

Figure 4.1. S&P500 absolute return series and three realized measures from January 2000 to June 2022.

Daily one-step-ahead forecasts of VaR and ES are calculated for the six return series at the 1% and 2.5% probability level in the forecast sample. A rolling window, with fixed in-sample size T, is used to estimate each of m one-step-ahead forecasts of VaR and ES in the forecast period for each series. Table 4.2 shows the total sample sizes, plus T and m in each market. T and m differ due to different non-trading days in each market.

Table 4.2. Summary of the selected data sets and their in-sample and out-of-sample split.

Index	Sample size	In-sample size (T)	Out-of-sample size (m)
S&P500	5634	3008	2626
FTSE	5667	3020	2647
NASDAQ	5636	3006	2630
HSI	5504	2937	2567
DAX	5697	3050	2647
SMI	5634	3013	2621

4.5.2 Models in comparison

Table 4.3 lists the 33 models considered in the tail risk forecasting study. As in Section 4.2, four groups of models are included, for comparison, as now discussed.

Table 4.3. A summary of the competing models in the empirical section, based on the model type, realized measures used, and the number of realized measures (K).

Model	Type	Realized measures	K
GARCH Models			
GARCH-t	Parametric	NA	0
EGARCH-t	Parametric	NA	0
GJR-GARCH-t	Parametric	NA	0
GARCH-QML-HS	Parametric	NA	0
EGARCH-QML-HS	Parametric	NA	0
GJR-GARCH-QML-HS	Parametric	NA	0
REGARCH-t Models			
RV5	Parametric	RV5	1
RK	Parametric	RK	1
BV	Parametric	BV	1
RV5-RK	Parametric	RV5, RK	2
RV5-BV	Parametric	RV5, BV	2
RK-BV	Parametric	RK, BV	2
RV5-RK-BV	Parametric	RV5, RK, BV	3
ES-CAViaR Models			
ES-CAViaR-Add	Semi-Parametric	NA	0
ES-CAViaR-X Models			
RV5	Semi-parametric	RV5	1
RK	Semi-parametric	RK	1
BV	Semi-parametric	BV	1
ES-X-CAViaR-X Models			
RV5	Semi-parametric	RV5	1
RK	Semi-parametric	RK	1
BV	Semi-parametric	BV	1
Realized-ES-CAViaR Models			
RV5	Semi-parametric	RV5	1
RK	Semi-parametric	RK	1
BV	Semi-parametric	BV	1
Log-Realized-ES-CAViaR Models			
RV5	Semi-parametric	RV5	1
RK	Semi-parametric	RK	1
BV	Semi-parametric	BV	1
Realized-ES-CAViaR-M Models			
RV5	Semi-parametric	RV5	1
RK	Semi-parametric	RK	1
BV	Semi-parametric	BV	1
RV5-RK	Semi-parametric	RV5, RK	2
RV5-BV	Semi-parametric	RV5, BV	2
RK-BV	Semi-parametric	RK, BV	2
RV5-RK-BV	Semi-parametric	RV5, RK, BV	3

Note: "0" represents that the model does not use realized measures. Grey shading highlights the proposed models.

Conventional GARCH (Bollerslev, 1986), EGARCH (Nelson, 1991), and GJR-GARCH models (Glosten et al., 1993), all with Student's t return error are included. The two-step QML-HS approach as described in Section 4.2.1 is also considered, with GARCH, EGARCH and GJR-GARCH employed as the volatility models.

Next, the parametric REGARCH model with RV5, RK, and BV is included. Student's t return error (REGARCH-t) with Gaussian measurement error is considered, as in Watanabe (2012). A similar MCMC algorithm is employed for the estimation of this model. There are seven different versions, which are models with one, two, or three RMs. We have also tested models with Gaussian errors for the return equations, however, these are outperformed by the Student's t return error models and hence not included to save space.

From the semi-parametric models, ES-CAViaR-X (additive), ES-X-CAViaR-X (Gerlach and Wang, 2020) and (Log-)Realized-ES-CAViaR ($K = 1$) (Wang et al., 2023) are included and estimated via similar adaptive MCMC algorithms. Finally, seven versions of the proposed Realized-ES-CAViaR-M framework are included, again being models with one, two or three RMs included.

4.5.3 Parameter estimates

In this section, we study the parameter estimates from the proposed model on 1% & 2.5% probability levels and their comparison to the REGARCH-t, based on parameter inference results for one forecasting step and some discussions for γ parameters of the the full out-of-sample period. Only the common parameters between Realized-ES-CAViaR-M and REGARCH-t are compared, since the REGARCH-t does not have the ES component related parameters, e.g., the ones in equation (4.9).

One forecasting step results

For the proposed Realized-ES-CAViaR-M on 1% & 2.5% probability levels and REGARCH-t with 3 RMs, Table 4.4 shows the parameter posterior means and the lower and upper quantiles (LQ and UQ) of the 95% credible intervals (CI), using the first moving window of S&P 500 data. Insignificant parameter estimates with CIs include 0 are highlighted in red. We have the following observations.

First, we find our REGARCH-t parameter estimates are in general consistent with the ones in Tables 2 and 3 of Hansen and Huang (2016). The volatility autoregressive parameter β estimate is close 1. The τ_1 and τ_2 estimates are both significant with the value of τ_1 as negative, so that negative returns have more impact on the future volatility. Regarding the coefficients γ_1 (coefficient of RV5), γ_2 (RK) and γ_3 (BV) that are used to model the information from RMs, interestingly we see that the γ_1 estimate is insignificant. For the parameters in the measurement equation, φ_1, φ_2 and φ_3 (regression coefficients between RMs and volatility) are all close to unity. Negative ξ_1, ξ_2 and ξ_3 estimates are produced meaning negative bias corrections are needed when regressing RMs versus volatility. This is to be expected, as RMs are only measured when the market is open, while the returns employed in this work are close-to-close and include overnight price movements, thus downward biased correction is needed. The leverage effect captured in the measurement equation is also significant, with the values of δ_{11}, δ_{21} and δ_{31} as negative.

Second, regarding the parameter estimates of the proposed Realized-ES-CAViaR-M model, we observe that the ranges of CIs are in general consistent with the ones from the REGARCH-t. When checking the values of the parameter estimates, we observe some distinctive while explainable behaviours. The β estimate is also close unity. However, we can see that the τ_1 estimates from the proposed Realized-ES-CAViaR-M are positive. This is because the left tail quantile Q_t has negative values, thus the defined multiplicative error $\epsilon_t = \frac{r_t}{Q_t}$ used in the leverage term have an opposite sign to the return r_t and z_t in REGARCH-t. With respect to the γ_1 (coefficient of RV5), γ_2 (RK) and γ_3 (BV) parameter estimates, it is very interesting to see that on 1% and 2.5% different RMs are significant, meaning for different probability levels and potentially for different forecasting steps different RMs could play more important role in risk forecasting. In the full-of-sample study to be shown in the following section, we will have more discussion on this.

Third, for the measurement equation of the proposed Realized-ES-CAViaR-M, although we use Q_t as regressor, we still see that the φ_1, φ_2 and φ_3 estimates are close to unity, which is consistent with the REGARCH-t. Meanwhile, when comparing to REGARCH-t, more negative ξ_1, ξ_2 and ξ_3 estimates are produced. As discussed in Section 4.3 when developing the proposed model, we have $\sigma_t = \frac{Q_t}{a_\alpha}$ with a_α as a number that is negative with absolute value that is greater than 1 for the considered 1% and 2.5% probability levels. For example with a standard Gaussian return distribution, the a_α values on 1% and 2.5% are equal to the Gaussian CDF inverse values of -2.3263 and -1.96, which is also why the 1% estimated Realized-ES-CAViaR-M has $\boldsymbol{\xi}$ estimates that are more negative than the ones of 2.5%, e.g., -1.2498 vs -1.0468 for ξ_1. Therefore, with more negative bias correction observed, the proposed model is still able to produce close to unity $\boldsymbol{\varphi}$ estimates as REGARCH-t does. Lastly, the leverage term related coefficients are also all significant, while the signs δ_{11}, δ_{21} and δ_{31} are again opposite to the ones from REGARCH-t.

Full out-of-sample results

In the REGARCH-t and proposed Realized-ES-CAViaR-M, the regression coefficients γ capture how influential the K lagged RMs are on next period volatility or quantile forecast. Meanwhile, a key contribution of the work is to incorporate the multiple RMs into the quantile (and ES) forecasting. In the previous section we have observed some interesting behaviours for the γ parameters. We now further explore how the γ estimates behave for the full out-of-sample

Table 4.4. Parameter posterior means and lower and upper quantiles of the 95% credible intervals of Realized-ES-CAViaR-M (1% & 2.5%) and REGARCH-t, with the 1st set of in-sample data of S&P500.

	Realized-ES-CAViaR-M (1%)			Realized-ES-CAViaR-M (2.5%)			REGARCH-t		
	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB
ω	0.0102	0.0040	0.0161	0.0046	-0.0007	0.0100	0.0010	-0.0080	0.0089
β	0.9717	0.9664	0.9772	0.9705	0.9650	0.9759	0.9660	0.9588	0.9729
τ_1	0.1812	0.1649	0.1980	0.1526	0.1377	0.1696	-0.1517	-0.1682	-0.1361
τ_2	0.1160	0.0904	0.1452	0.0850	0.0671	0.1029	0.0464	0.0372	0.0555
γ_1	0.0108	-0.0318	0.0509	-0.0046	-0.0472	0.0386	-0.0448	-0.0904	0.0042
γ_2	0.0015	-0.0257	0.0290	0.0300	0.0031	0.0572	0.0598	0.0303	0.0901
γ_3	0.2278	0.1927	0.2649	0.2081	0.1693	0.2486	0.2507	0.2023	0.2978
ξ_1	-1.2498	-1.2899	-1.2097	-1.0468	-1.0882	-1.0046	-0.4296	-0.4889	-0.3706
ξ_2	-1.3719	-1.4152	-1.3308	-1.1675	-1.2093	-1.1258	-0.5665	-0.6228	-0.5115
ξ_3	-1.3419	-1.3823	-1.2999	-1.1392	-1.1820	-1.0962	-0.6371	-0.6979	-0.5730
φ_1	1.0428	1.0088	1.0721	1.0400	1.0048	1.0794	1.0322	0.9876	1.0785
φ_2	1.0478	1.0120	1.0812	1.0442	1.0072	1.0825	1.0345	0.9899	1.0802
φ_3	1.0536	1.0167	1.0863	1.0530	1.0159	1.0929	1.0456	0.9982	1.0930
δ_{11}	0.1123	0.0901	0.1333	0.0939	0.0757	0.1132	-0.0861	-0.1039	-0.0683
δ_{21}	0.0636	0.0345	0.0924	0.0533	0.0297	0.0786	-0.0443	-0.0686	-0.0174
δ_{31}	0.1754	0.1537	0.1955	0.1457	0.1280	0.1640	-0.1362	-0.1537	-0.1192
δ_{12}	0.3635	0.3278	0.4010	0.2464	0.2199	0.2734	0.1111	0.0987	0.123
δ_{22}	0.6784	0.6280	0.7291	0.4619	0.4250	0.5011	0.2052	0.1864	0.2236
δ_{32}	0.2262	0.1947	0.2591	0.1540	0.1301	0.1778	0.0698	0.0591	0.0807
ν							10.5214	7.5990	15.2822

Note: Insignificant parameter estimates are highlighted in red.

period. Figure 4.2 displays the full out-of-sample γ_1 (RV5), γ_2 (RK) and γ_3 (BV) plots for Realized-ES-CAViaR-M (1% & 2.5%) and REGARCH-t with the S&P500 data. First, we observe that the general pattern of the γ estimates from the 1% & 2.5% Realized-ES-CAViaR-M and the REGARCH-t is consistent, with distinctive behaviours observed. This means the information from RMs is used differently in semi-parametric and parametric risk forecasting. Second, for both 1% & 2.5% risk forecasting and volatility forecasting, the BV seems to be the most influential variable. The implication of this observation on the empirical performance will be discussed in the following section. Third, as shown in Table 4.4, for the 1st forecasting step of 1% Realized-ES-CAViaR-M, both the γ_1 and γ_2 parameter estimates are insignificant, while in the latter forecasting steps the insignificant parameters could become significant, e.g., RV5 during the 2019 period. Further, we observe that for different markets and forecasting steps, the significance of the RMs could vary. Such observations demonstrate that REGARCH-t and the proposed Realized-ES-CAViaR-M are capable of selecting RMs for volatility and risk forecasting tasks using a data driven approach. Meanwhile, this naturally brings up the direction for the next step research. Via incorporating a much larger set of RMs and further developed modelling framework, we could investigate which (types) RMs are more important in volatility and risk forecasting and select such RMs using automatic variable selection technique, such as LASSO (Tibshirani, 1996).

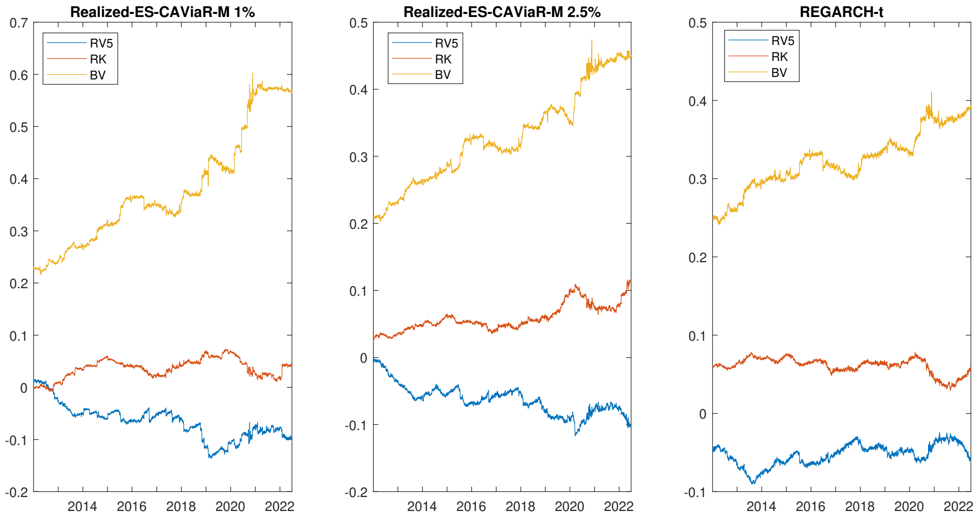

Figure 4.2. The full out-of-sample γ_1 (RV5), γ_2 (RK) and γ_3 (BV) plot for Realized-ES-CAViaR-M (1% & 2.5%) and REGARCH-t with the S&P500 data.

4.5.4 Assessing Value-at-Risk forecasts

This section discusses the evaluation of one-step-ahead VaR forecasting accuracy via using the strictly consistent quantile loss function:

$$\frac{1}{m}\sum_{t=T+1}^{T+m}\left(\alpha - I(r_t \le \widehat{Q}_t)\right)\left(r_t - \widehat{Q}_t\right), \tag{4.18}$$

where $\widehat{Q}_{T+1}, ..., \widehat{Q}_{T+m}$ are the quantile forecasts at level α. Since the quantile loss function is strictly consistent, the model with minimum sample quantile loss is preferred. Tables 4.5 and 4.6 present the VaR quantile loss function results on the tested 1% and 2.5% probability levels. The average rank based on the ranks of the quantile loss across six markets is also included in the "Avg Rank" column. The box indicates the favoured model, and the blue text indicates the second-ranked model in each column.

In general, for both probability levels the proposed Realized-ES-CAViaR-M models produce lower quantile loss values and are better ranked than the other models considered. For the

1% study, overall the best ranked model is the Realized-ES-CAViaR-M incorporating all three realized measures, followed by the Realized-ES-CAViaR-M with RV5 and BV. For the 2.5% study, the top two performing models for each market are all from the Realized-ES-CAViaR-M model class. The overall best ranked model is again Realized-ES-CAViaR with three realized measures, followed by Realized-ES-CAViaR-M with BV. The preferred performance of Realized-ES-CAViaR-M with three RMs demonstrate the effectiveness of incorporating multiple RMs in semi-parametric quantile forecasting. Further, based on the parameter estimates in Section 4.5.3, BV seems to be the most influential RM in the Realized-ES-CAViaR-M forecasting process. Such findings are supported by the results that the proposed models with BV are generally preferred and better ranked than the ones without BV. This means that although RK&RV are potentially less effective than BV in semi-parametric quantile forecasting, the Realized-ES-CAViaR-M with three RMs can still effective incorporate the useful information from RK&RV and combine it with BV (see Figure 4.2) in generating improved quantile forecasts.

Compared to the Realized-ES-CAViaR model of Wang et al. (2023), the Realized-ES-CAViaR-M extends it via incorporating the information of multiple realized measures during the VaR and ES forecasting process. Further, a log specification is used in Realized-ES-CAViaR-M, with an additional leverage employed in the quantile equation (4.8). Based on the empirical results, now we investigate the improved performance of the Realized-ES-CAViaR-M models is due to the log specification, the leverage term, or the inclusion of multiple RMs.

Regarding the log-specification, comparing the Realized-ES-CAViaR and Log-Realized-ES-CAViaR, the overall consistent performance of the two classes of models shows that the log specification does not significantly affect the risk forecasting performance.

Regarding the leverage effect, actually the (Log-)Realized-ES-CAViaR model of Wang et al. (2023) has already considered it in its quantile regression process. For example, in the Log-Realized-ES-CAViaR model (4.12) by substituting the $\log(x_{t-1})$ in the quantile equation with its expression in the measurement equation we get:

$$\log(-Q_t) = \beta_0 + \beta_1 \left(\xi + \phi\log(-Q_{t-1}) + \tau_1\epsilon_{t-1} + \tau_2\left(\epsilon_{t-1}^2 - E(\epsilon^2)\right) + u_{t-1}\right) + \beta_2\log(-Q_{t-1}),$$

thus the leverage effect is considered in its VaR forecasting. Consequently, as ES is equal to VaR subtracting ω_t, the leverage effect is also implicitly considered in the ES forecasting process.

Therefore, the overall favoured performance of the Realized-ES-CAViaR-M with three RMs confirms the key driver of its improved risk forecasting performance is the information contained in the multiple RMs, which is captured by the proposed framework.

The performance of the REGARCH-t on the 2.5% level is slightly better than the more extreme 1% level, demonstrating that fixing the return distribution for different probability levels could limit the performance of the parametric model. Meanwhile, the superior performance of Realized-ES-CAViaR-M shows the advantages of semi-parametric risk forecasting, as the selection of return error distribution is not required.

To statistically test whether the quantile loss differences between different models are significant, the model confidence set (MCS), introduced by Hansen et al. (2011), is employed. The MCS produces a group of models that is constructed such that it will contain the "superior" forecasting models, given a level of confidence. The MCS is used to assess the statistical significance for quantile loss (per equation (4.18)) under the 75% confidence level. We adopt the Matlab code downloaded from Kevin Sheppard's web page (Sheppard, 2009). Two methods, R and SQ, are calculated to test the competing models based on different rules of calculating the test statistic in the downloaded MCS code. In Tables 4.5 and 4.6, for each market we use grey shading to highlight the models included in MCS, based on the R method, for visual comparison of the models. As can be seen, the proposed Realized-ES-CAViaR-M models are more or equally likely to be included in the MCS, compared to other models. The GARCH-type models and ES-CAViaR-Add (without using realized data) are generally less likely to be included in the MCS, demonstrating the usefulness of incorporating realized data in either parametric or semi-parametric risk forecasting. In addition, Table 4.9 presents a summary of MCS results based on the quantile loss for both R and SQ methods. The R and SQ columns show the total number of times that each model is included in the 75% MCS across the six return series (the higher, the better). The results from the SQ method are consistent with the ones from the R method, with the Realized-ES-CAViaR-M type models being more or equally likely to be included in MCS compared to other models. The Realized-ES-CAViaR-M with three RMs is the only model that is always in the MCS across six markets for both tests and probability levels.

4.5.5 Assessing Expected Shortfall forecasts

The same 33 models are employed to generate one-step-ahead ES forecasts at the 1% and 2.5% probability levels for the same six series in the forecasting period.

As discussed in Section 4.4.1, Taylor (2019) shows that the negative of the quasi-log-likelihood function (4.13) is strictly consistent for Q_t and ES_t considered jointly, and fits into the class of strictly consistent joint loss functions for VaR and ES developed by Fissler and Ziegel (2016). We use the average joint loss $S = \frac{1}{m}\sum_{t=n+1}^{n+m} S_t$ to formally and jointly assess the VaR and ES forecasts from all models.

$$S_t(r_t, \widehat{Q}_t, \widehat{\text{ES}}_t) = -\log\left(\frac{\alpha - 1}{\widehat{\text{ES}}_t}\right) - \frac{(r_t - \widehat{Q}_t)(\alpha - I(r_t \leq \widehat{Q}_t))}{\alpha \widehat{\text{ES}}_t}. \tag{4.19}$$

Tables 4.7 and 4.8 show the 1% and 2.5% VaR and ES joint loss function values for each model and each market. The "Avg Rank" column calculated under the same way as Tables 4.5 and 4.6 is included. Grey shading is again used to highlight the models included in the 75% MCS, based on the R method with joint loss. The results are generally consistent with the ones from the quantile loss study. For each market, on the 1% level the top ranked models are more likely from the proposed model class, and on the 2.5% level all the top 2 ranked models are from the proposed Realized-ES-CAViaR-M models. Again, the overall best ranked model is the Realized-ES-CAViaR-M with all three realized measures. Although the parametric REGARCH-t with three RMs could generate competitive forecasting results, it is still consistently outperformed by its semi-parametric counter-part, the proposed Realized-ES-CAViaR-M, again demonstrating the usefulness of forecasting the risk without assuming the return error distribution. The performance of the Realized-ES-CAViaR and Log-Realized-ES-CAViaR are similar to each other, and they are outperformed by the Realized-ES-CAViaR-M with three realized measures. Such observations demonstrate the effectiveness of employing the information from multiple RMs in semi-parametric ES forecasting. The MCS results in Tables 4.7 and 4.8 support the Realized-ES-CAViaR-M model as being more or equally likely to be included in the MCS. A summary of the joint loss -based MCS results based on R and SQ methods are presented in Table 4.9. The results from the joint loss SQ method align with the ones from the R method. Across six markets, the Realized-ES-CAViaR-M with three RMs is the only model that is always in the MCS for both tests and probability levels.

Another interesting observation is that when the same realized measure(s) are included in the Realized-ES-CAViaR-M models, their VaR forecasting performance based on quantile loss could differ to the one from VaR&ES joint loss. For example, via comparing the rank results in Tables 4.6 and 4.8 on the 2.5% probability level, when only one realized measure is included in the Realized-ES-CAViaR-M models, they performed more competitively in terms of quantile loss than joint loss. Such observation is potentially related to the additive VaR to ES time varying relationship as described in equations (4.9) and (4.10), driven *separately* by the lagged RMs comparing to the ones in the quantile equation (4.8). There is potentially interesting future work based on such observation: which (types of) realized measures could be more useful in VaR forecasting? Is it the case that other (types of) realized measures more useful in ES forecasting? These questions could be investigated via studying a large set of RMs in VaR and ES forecasting via a further developed modelling framework. The LASSO approach can be used to conduct automatic variable selection, as discussed at the end of Section 4.5.3.

To summarize, with the 1% and 2.5% quantile and joint loss evaluations and MCS backtests, for VaR and ES forecasting accuracy comparison over six markets, the proposed Realized-ES-CAViaR-M framework has generally favourable performance compared to a range of competing models. The performance is most favourable for the proposed model using all three RMs.

Table 4.5. 1% VaR forecasting quantile loss on six indices.

Model	S&P500	FTSE	NASDAQ	HSI	DAX	SMI	Avg Rank
GARCH							
GARCH-t	89.8	91.4	105.8	100.4	111.1	88.0	32.7
EGARCH-t	87.5	88.1	102.6	96.2	107.5	85.5	23.8
GJR-GARCH-t	89.0	86.8	103.2	96.4	108.6	84.8	25.0
GARCH-QML-HS	88.3	88.1	101.0	98.0	110.3	86.5	29.5
EGARCH-QML-HS	86.5	86.1	99.0	95.4	106.4	84.4	17.5
GJR-GARCH-QML-HS	87.9	84.3	100.9	95.4	107.5	84.1	17.7
REGARCH-t							
RV5	85.2	88.7	97.1	97.9	110.1	84.5	26.3
RK	80.9	88.1	95.1	97.2	114.8	86.1	25.5
BV	82.7	88.9	95.5	100.4	108.9	86.1	25.0
RV5-RK	83.5	87.6	94.1	95.1	109.9	82.6	18.7
RV5-BV	80.6	87.9	93.8	100.3	108.8	82.6	19.2
RK-BV	80.1	87.9	93.8	95.5	109.1	83.3	16.8
RV5-RK-BV	79.7	87.3	93.9	95.5	108.9	82.6	15.5
ES-CAViaR							
ES-CAViaR-Add	87.2	89.3	100.4	97.7	112.4	87.1	29.5
ES-CAViaR-X							
RV5	87.8	85.3	93.0	95.7	109.9	82.2	17.2
RK	87.3	83.2	93.2	96.0	109.6	84.5	18.3
BV	88.2	84.4	91.9	95.9	107.7	80.6	12.2
ES-X-CAViaR-X							
RV5	82.8	84.6	93.0	95.5	110.0	82.3	15.7
RK	82.4	83.4	93.9	95.9	110.5	84.6	19.2
BV	79.2	84.4	92.1	96.0	108.3	80.8	9.8
Realized-ES-CAViaR							
RV5	83.1	85.1	93.2	95.0	110.1	82.1	14.8
RK	84.0	85.5	94.2	95.5	111.0	83.9	21.3
BV	78.5	85.6	92.1	95.8	108.9	82.1	12.2
Log Realized-ES-CAViaR							
RV5	82.2	85.2	92.5	95.0	109.9	82.2	13.0
RK	83.1	84.4	93.6	96.2	110.3	84.0	19.8
BV	78.6	84.9	90.8	95.8	109.1	81.6	10.5
Realized-ES-CAViaR-M							
RV5	80.9	85.4	92.7	94.4	108.2	81.4	9.2
RK	81.7	83.0	93.7	97.3	107.9	82.9	14.5
BV	78.0	83.6	92.0	95.2	107.1	82.3	7.2
RV5-RK	82.6	83.5	92.8	94.3	110.0	81.6	10.8
RV5-BV	77.9	83.6	92.1	95.0	106.5	81.5	4.8
RK-BV	78.0	82.9	91.7	95.7	106.6	81.4	5.2
RV5-RK-BV	77.8	82.8	91.9	94.8	107.0	81.3	2.7

Note: The box indicates the favoured models, and the blue text indicates the second-ranked model in each column. Grey shades the models that are included in the 75% MCS using the R method.

Table 4.6. 2.5% VaR forecasting quantile loss on six indices.

Model	S&P500	FTSE	NASDAQ	HSI	DAX	SMI	Avg Rank
GARCH							
GARCH-t	181.2	183.0	218.5	210.9	226.0	175.0	31.8
EGARCH-t	175.3	176.9	213.4	203.7	217.2	168.6	20.5
GJR-GARCH-t	174.6	176.8	211.1	204.7	219.9	168.0	21.2
GARCH-QML-HS	182.0	184.0	219.8	211.9	227.3	175.4	32.8
EGARCH-QML-HS	176.7	177.9	214.8	204.8	217.6	169.0	23.0
GJR-GARCH-QML-HS	175.5	177.6	212.3	205.4	220.3	168.4	24.3
REGARCH-t							
RV5	183.7	177.9	203.3	204.2	223.1	166.9	24.7
RK	169.6	177.9	201.1	202.7	223.9	168.7	21.8
BV	172.8	178.3	201.4	205.3	221.9	170.6	25.2
RV5-RK	171.4	176.5	197.8	199.9	220.6	164.7	13.7
RV5-BV	168.0	177.8	197.9	205.7	220.7	164.7	18.5
RK-BV	168.0	176.9	197.0	199.5	220.0	165.5	11.8
RV5-RK-BV	167.8	176.4	197.5	199.9	219.8	164.7	11.0
ES-CAViaR							
ES-CAViaR-Add	179.2	179.7	212.0	205.4	222.9	174.6	29.0
ES-CAViaR-X							
RV5	179.0	179.6	209.9	203.1	221.6	172.5	26.3
RK	177.6	182.0	207.3	203.6	221.5	173.3	26.5
BV	180.5	178.9	209.3	203.1	221.3	169.9	25.3
ES-X-CAViaR-X							
RV5	169.7	177.2	198.0	201.0	218.8	164.8	14.3
RK	168.2	176.3	197.3	202.4	218.8	166.4	13.3
BV	168.1	175.0	196.2	200.6	219.3	166.2	11.2
Realized-ES-CAViaR							
RV5	170.7	178.3	198.0	200.6	219.0	166.6	17.2
RK	171.3	179.0	197.7	201.8	218.6	168.0	18.3
BV	165.8	176.8	196.1	200.9	218.2	167.9	11.0
Log Realized-ES-CAViaR							
RV5	169.9	177.1	197.3	199.9	218.3	165.7	13.0
RK	169.8	176.5	196.8	205.7	218.5	167.1	13.8
BV	165.6	174.6	195.8	199.5	218.5	166.6	9.5
Realized-ES-CAViaR-M							
RV5	166.1	173.7	194.8	197.9	216.7	162.9	3.8
RK	166.3	173.4	195.5	197.7	215.6	164.2	4.2
BV	163.2	174.1	194.1	198.6	214.9	164.0	3.2
RV5-RK	166.4	174.8	198.4	198.3	224.3	168.6	15.7
RV5-BV	165.3	172.9	204.3	205.2	216.0	165.6	12.0
RK-BV	165.1	180.2	197.4	202.2	211.7	163.5	11.0
RV5-RK-BV	164.6	172.5	195.1	198.0	214.7	162.6	2.0

Note: The box indicates the favoured models, and the blue text indicates the second-ranked model in each column. Grey shades the models that are included in the 75% MCS using the R method.

Table 4.7. 1% VaR & ES joint loss function values across six indices.

Model	S&P500	FTSE	NASDAQ	HSI	DAX	SMI	Avg Rank
GARCH							
GARCH-t	5879.8	5881.2	6349.3	6050.6	6365.9	5681.8	30.0
EGARCH-t	5885.0	5928.9	6444.7	5973.9	6366.0	5686.3	30.5
GJR-GARCH-t	5934.1	5823.2	6361.8	5970.7	6390.5	5652.1	30.0
GARCH-QML-HS	5726.0	5734.9	6121.3	5969.0	6329.6	5588.3	25.7
EGARCH-QML-HS	5717.9	5769.3	6179.2	5932.7	6282.6	5573.6	23.8
GJR-GARCH-QML-HS	5749.1	5686.1	6167.0	5920.7	6311.0	5570.4	22.5
REGARCH-t							
RV5	5623.4	5842.3	6059.8	6003.9	6422.3	5587.5	28.3
RK	5364.9	5794.7	6053.8	5937.2	6598.9	5645.5	24.8
BV	5539.6	5797.1	5942.7	6052.4	6380.5	5595.8	26.2
RV5-RK	5511.5	5780.0	5970.8	5917.0	6449.2	5504.4	23.5
RV5-BV	5418.2	5775.0	5984.3	6070.6	6393.9	5484.3	23.3
RK-BV	5370.1	5768.2	5966.7	5938.2	6400.5	5514.2	22.0
RV5-RK-BV	5350.7	5734.7	5960.9	5931.8	6385.2	5493.9	19.2
ES-CAViaR							
ES-CAViaR-Add	5692.0	5758.5	6069.2	5977.1	6404.2	5620.4	27.5
ES-CAViaR-X							
RV5	5547.1	5597.8	5802.4	5901.2	6318.1	5416.8	15.8
RK	5568.5	5568.7	5793.3	5896.2	6311.2	5522.1	15.2
BV	5466.3	5605.9	5780.6	5922.5	6257.9	5370.4	11.8
ES-X-CAViaR-X							
RV5	5433.5	5582.8	5800.6	5896.1	6315.5	5407.5	11.8
RK	5453.1	5566.9	5816.2	5888.4	6315.0	5509.2	14.3
BV	5309.8	5594.0	5783.4	5926.7	6255.8	$\boxed{5363.8}$	9.2
Realized-ES-CAViaR							
RV5	5438.4	5583.8	5792.4	5874.3	6279.7	5432.4	10.7
RK	5479.9	5587.8	5814.6	5876.2	6295.7	5504.1	14.2
BV	5293.7	5593.9	5779.5	5912.5	6239.6	5414.4	8.0
Log-Realized-ES-CAViaR							
RV5	5394.5	5595.6	5773.8	5881.8	6327.9	5422.8	11
RK	5433.8	5567.7	5803.9	5897.1	6314.1	5497.8	13.7
BV	5299.3	5591.5	$\boxed{5752.0}$	5920.4	6272.3	5399.3	8.8
Realized-ES-CAViaR-M							
RV5	5414.6	5708.7	5806.8	5890.6	6270.7	5406.6	11.8
RK	5445.3	5564.2	5834.3	5913.6	6258.2	5459.4	13.2
BV	5298.2	5560.7	5792.2	5914.8	6204.8	5425.8	8.3
RV5-RK	5420.6	5567.8	5800.9	$\boxed{5873.2}$	6327.3	5421.6	11.2
RV5-BV	5291.4	5535.5	5794.6	5908.9	6184.6	5401.2	6.2
RK-BV	5289.0	5522.3	5784.8	5914.5	$\boxed{6182.6}$	5392.8	5.0
RV5-RK-BV	$\boxed{5285.6}$	$\boxed{5514.5}$	5787.0	5889.5	6191.4	5390.6	$\boxed{3.5}$

Note: The box indicates the favoured models, and the blue text indicates the second-ranked model in each column. Grey shades the models that are included in the 75% MCS using the R method.

Table 4.8. 2.5% VaR & ES joint loss function values across six indices.

Model	S&P500	FTSE	NASDAQ	HSI	DAX	SMI	Avg Rank
GARCH							
GARCH-t	5274.1	5325.9	5810.0	5641.2	5871.3	5138.3	31.5
EGARCH-t	5187.0	5265.3	5810.0	5553.3	5789.0	5070.5	26.0
GJR-GARCH-t	5207.1	5234.5	5762.1	5565.3	5819.2	5053.7	26.8
GARCH-QML-HS	5373.5	5422.0	5900.4	5701.6	5962.6	5206.3	33.0
EGARCH-QML-HS	5246.3	5314.7	5841.6	5590.7	5824.0	5110.3	29.7
GJR-GARCH-QML-HS	5273.4	5287.6	5811.0	5600.9	5867.4	5097.3	30.5
REGARCH-t							
RV5	5183.7	5234.1	5523.0	5559.8	5866.3	4994.9	25.3
RK	4868.2	5220.0	5516.3	5513.0	5915.2	5031.9	26.2
BV	5004.1	5219.9	5468.0	5566.4	5837.1	5027.7	23.2
RV5-RK	4941.3	5192.0	5443.2	5493.3	5849.0	4941.0	18.8
RV5-BV	4890.2	5212.4	5453.4	5586.2	5837.4	4932.3	20.5
RK-BV	4870.3	5204.2	5434.6	5490.1	5827.9	4948.9	16.7
RV5-RK-BV	4862.4	5180.0	5435.3	5549.5	5819.4	4936.4	17.0
ES-CAViaR							
ES-CAViaR-Add	5193.4	5198.2	5649.1	5558.2	5800.0	5094.8	25.2
ES-CAViaR-X							
RV5	4995.6	5166.6	5455.1	5509.1	5762.9	4963.6	19.7
RK	5000.8	5169.6	5453.7	5509.2	5762.4	4990.6	20.2
BV	4947.5	5146.9	5454.4	5507.8	5747.9	4951.2	17.3
ES-X-CAViaR-X							
RV5	4905.8	5146.3	5382.0	5494.9	5748.8	4905.0	12.8
RK	4900.7	5141.0	5380.2	5496.0	5742.2	4951.3	13.5
BV	4900.5	5117.4	5366.6	5491.2	5745.2	4923.0	10.7
Realized-ES-CAViaR							
RV5	4908.1	5154.8	5381.6	5482.7	5731.0	4945.1	12.5
RK	4949.7	5159.4	5389.6	5488.7	5724.3	4981.7	14.8
BV	4840.6	5130.4	5366.8	5489.8	5710.1	4959.1	9.0
Log Realized-ES-CAViaR							
RV5	4898.7	5151.0	5376.4	5487.3	5740.4	4910.1	10.5
RK	4904.6	5141.5	5377.1	5490.1	5739.0	4965.7	13.2
BV	4836.6	5113.6	5364.1	5500.7	5727.0	4924.4	8.3
Realized-ES-CAViaR-M							
RV5	4855.3	5102.8	5352.0	5465.3	5720.4	4895.0	4.5
RK	4873.3	5105.2	5360.0	5446.8	5709.3	4908.9	5.5
BV	4817.3	5091.5	5358.7	5462.9	5680.4	4910.0	3.5
RV5-RK	4868.9	5110.0	5398.8	5459.4	5817.2	5002.9	13.2
RV5-BV	4853.6	5077.2	5479.9	5543.9	5699.7	4898.2	10.3
RK-BV	4837.9	5191.8	5382.2	5506.9	5608.0	4892.6	9.5
RV5-RK-BV	4822.8	5066.5	5354.6	5455.7	5670.8	4877.7	1.7

Note: The box indicates the favoured models, and the blue text indicates the second-ranked model in each column. Grey shades the models that are included in the 75% MCS using the R method.

Table 4.9. 75% MCS with R and SQ methods on 1% and 2.5 % probability levels.

Model	1% Quantile		1% Joint		2.5% Quantile		2.5% Joint		
	R	**SQ**	**R**	**SQ**	**R**	**SQ**	**R**	**SQ**	**Total**
GARCH									
GARCH-t	2	2	2	2	0	0	0	0	8
EGARCH-t	2	3	2	2	3	3	0	1	16
GJR-GARCH-t	4	4	3	1	3	4	0	1	20
GARCH-QML-HS	2	2	3	2	0	0	0	0	9
EGARCH-QML-HS	5	5	3	3	3	1	0	1	21
GJR-GARCH-QML-HS	4	4	3	3	3	1	0	0	18
REGARCH-t									
RV5	2	3	0	1	1	1	0	1	9
RK	3	4	3	2	3	3	1	2	21
BV	2	3	4	2	3	1	1	2	18
RV5-RK	4	4	3	3	5	5	3	4	31
RV5-BV	3	5	2	3	3	3	2	4	25
RK-BV	4	5	3	4	5	5	3	4	33
RV5-RK-BV	4	6	4	4	5	5	3	5	36
ES-CAViaR									
ES-CAViaR-Add	2	2	2	2	0	0	0	1	9
ES-CAViaR-X									
RV5	5	5	4	5	1	1	1	2	24
RK	4	5	4	4	2	0	2	2	23
BV	5	5	4	5	2	2	3	4	30
ES-X-CAViaR-X									
RV5	5	5	4	5	6	4	4	5	38
RK	4	4	5	4	5	5	3	5	35
BV	5	6	5	6	5	5	4	5	41
Realized-ES-CAViaR									
RV5	5	5	4	5	5	3	4	5	36
RK	5	5	4	4	5	4	4	3	34
BV	6	6	5	6	5	4	5	5	42
Log-Realized-ES-CAViaR									
RV5	5	6	4	6	6	5	4	5	41
RK	5	5	5	5	5	4	3	4	36
BV	6	6	5	6	6	5	5	5	44
Realized-ES-CAViaR-M									
RV5	5	5	4	5	6	5	4	5	39
RK	5	5	5	6	6	6	5	5	43
BV	6	6	5	6	6	6	6	5	46
RV5-RK	5	6	4	5	4	4	3	4	35
RV5-BV	6	6	6	6	6	5	3	4	42
RK-BV	6	6	6	6	5	5	3	6	43
RV5-RK-BV	6	6	6	6	6	6	6	6	$\boxed{48}$

Note: The box indicates the favoured models, and the blue text indicates the second-ranked model based on the Total column.

4.6 Computational feasibility

This section provides an analysis of the computational efficiency of the models used in this study, with a particular focus on the Realized-ES-CAViaR-M model, which is the proposed model in this thesis. The efficiency of this model is emphasized in comparison to other semi-parametric models and parametric models that use realized measures, such as REGARCH-t models, and the existing models in the semi-parametric approach: ES-CAViaR-X and ES-X-CAViaR-X.

Hardware Specifications

All computations are performed using the Artemis High Performance Computing facility at the University of Sydney, leveraging high-performance computing resources to handle the computational demands of models involving large datasets and complex computations.

- Processor: Multi-core Intel Xeon processors
- Number of Cores: 16-24 cores, depending on the model
- Memory: Ranges from 18 GB to 64 GB per model
- Wall Time: Computational times vary based on model complexity, with Realized-ES-CAViaR-M being faster than other complex models

Computational efficiency

The proposed Realized-ES-CAViaR-M model is designed to efficiently handle the complexity of incorporating realized measures while maintaining competitive computational performance. Despite its added complexity, Realized-ES-CAViaR-M demonstrates superior computational efficiency compared to other semi-parametric models, such as ES-CAViaR-X and ES-X-CAViaR-X, as well as parametric models like REGARCH-t that also use realized measures.

The Realized-ES-CAViaR-M model (with 3 realized measures) is processed using 6 blocks across 24 cores, with each block taking 18 hours to complete. The ability to run these blocks in parallel allows for efficient handling of the large dataset.

Table 4.10. Computational time for S&P (m=2626) for $\alpha = 1\%$

Model	K	No of parallel blocks	No of cores	Memory	Hours for each block
Realized-ES-CAViar_M	**3**	6	24	64 GP	18
Realized-ES-CAViar_M models	**2**	6	24	64 GP	13
Realized-ES-CAViar_M models	**1**	6	24	64 GP	11
Log Realized-ES-CAViaR	**1**	1	1	18GP	90
Realzied-ES-CAViaR	**1**	1	1	18HP	66
ES-X-CAViaR-X	**1**	1	1	18GP	66
ES-CAViaR-X	**1**	1	1	18GP	35
ES-CAViaR	**0**	1	1	18GP	35
REGARCH-t	**3**	6	20	64GP	30
REGARCH-t	**2**	6	18	64GP	20
REGARCH-t	**1**	6	24	64GP	18

To manage the computational complexity of the Realized-ES-CAViaR-M and REGARCH-t models, the dataset (with 2626 observations) is divided into 6 blocks: 1-400, 401-800, 801-1200, 1201- 1600, 1601-2000, 2001-2626. Each block is processed in parallel using MATLAB's parfor loop, utilizing up to 24 cores for the Realized-ES-CAViaR-M models. This parallel execution ensures that the models can scale efficiently across multiple cores, minimizing the total processing time.

While ES-CAViaR-X and ES-X-CAViaR-X models do not benefit from the same level of parallelization (using only 1 core), the Realized-ES-CAViaR-M model leverages the parallel processing capabilities of the HPC Artemis facility, leading to faster wall time for the complex realized measure models. Despite being a more complex model that uses multiple realized measures, Realized-ES-CAViaR-M achieves faster computation times. In comparison, ES-X-CAViaR-X (1K) takes 66 hours per block, ES-CAViaR-X (1K) takes 35 hours per block, and Realized-ES-CAViaR-M (with 3 realized measures) finishes in 18 hours per block, which is significantly faster than ES-X-CAViaR-X and ES-CAViaR-X, despite Realized-ES-CAViaR-M being more complex and using a greater number of realized measures. This enhanced efficiency is achieved through the use of parallel computing and the optimized implementation of the Realized-ES-CAViaR-M model, allowing it to handle larger datasets more quickly without sacrificing model quality.

In conclusion, the Realized-ES-CAViaR-M model is not only computationally efficient despite its complexity but also outperforms existing semi-parametric models like ES-CAViaR-X and ES-X-CAViaR-X in terms of computational time. Its design allows for efficient use of parallel computing, which significantly reduces the time required to run complex models. The use of blocking methods and high-performance computing (HPC) infrastructure ensures that even complex datasets can be processed in a reasonable time frame.

Thus, Realized-ES-CAViaR-M is both a powerful and efficient tool for forecasting and risk management, making it a valuable contribution to the existing literature on semi-parametric models that use realized measures.

4.7 Chapter summary

This chapter proposes a new semi-parametric joint VaR and ES forecasting framework incorporating multiple realized measures. The proposed Realized-ES-CAViaR-M models generate highly competitive risk forecasting results regarding quantile loss, VaR and ES joint loss, and MCS backtest. In particular, the proposed Realized-ES-CAViaR-M models produce favourable results compared to their parametric counterpart, e.g., the REGARCH model, and the semi-parametric counterparts, e.g., ES-X-CAViaR-X and Realized-ES-CAViaR. Additionally, the Realized-ES-CAViaR-M models are computationally efficient, offering significant advantages in terms of processing time compared to existing models.

This work can be improved by considering more realized measures, including their sub-sampled versions and different frequencies. Moreover, the proposed framework includes single lags only, which can be extended to multiple lags. Finally, alternative versions of the model, such as a multiplicative time-varying relationship between VaR and ES via incorporating the information from multiple realized measures, could be considered in future work.

Chapter 5

RNN-HAR model

This chapter presents a long-memory and non-linear realized volatility model class is proposed for direct VaR forecasting. This model, referred to as RNN-HAR, extends the heterogeneous autoregressive (HAR) model, a framework known for efficiently capturing long memory in realized measures, by integrating a RNN to handle the non-linear dynamics. Quantile loss-based generalized Bayesian inference with Sequential Monte Carlo is employed for model estimation and sequential prediction in RNN-HAR. The empirical analysis is conducted using daily closing prices and realized measures with around 12 years of data till 2022, covering 31 market indices. The proposed model's one-step-ahead VaR forecasting performance is compared against a basic HAR model and its extensions. The results demonstrate that the proposed RNN-HAR model consistently outperforms all other models considered in the study.

5.1 Introduction

Volatility forecasting is a fundamental aspect of financial markets, playing a crucial role for both regulators and market practitioners involved in risk management and asset pricing. Accurate predictions of market volatility are vital for numerous applications, including setting capital reserves, pricing derivatives, and managing investment portfolios. The ability to anticipate market fluctuations can significantly enhance decision-making processes and mitigate financial risks.

Traditionally, parametric models are employed to forecast financial market volatility due to their ease of implementation and interpretability. Among these, the GARCH model and stochastic volatility models are widely used. The GARCH-type models, first introduced by Engle (1982b) and Bollerslev (1986), and further developed by others, provide a robust framework for capturing time-varying volatility by modeling the conditional variance as a linear function of past variances and squared returns. These models have proven effective in various financial contexts and remain a staple in the volatility forecasting literature (Taylor, 2008b).

However, traditional models often face challenges when applied to high-frequency intraday data, which is now widely available for many financial assets. Intraday data provides a granular view of market movements, offering richer and more detailed information compared to daily or lower-frequency data. The availability of high-frequency intra-day data has allowed the construction of many informative, efficient realized measures of volatility, such as the commonly used realized variance (RV, Andersen and Bollerslev (1998), Andersen et al. (2003)). Meanwhile, despite their popularity, short-memory models like the standard GARCH and stochastic volatility models struggle to accurately capture certain stylized features observed in high-frequency financial data, such as the long-range dependencies (Cont, 2001). Furthermore, these models typically do not fully leverage the richness of high-frequency data, potentially missing valuable information.

To address these limitations, researchers have explored long-memory volatility models. Among these, the FIGARCH (Fractionally Integrated GARCH) model introduced by Baillie et al. (1996) has been particularly influential. It extends the GARCH framework by incorporating fractional differencing to model long memory in volatility. While effective, these long-memory models often present estimation challenges and lack parsimony, making them less practical for widespread use (Tsay, 2010). Additionally, despite their advances, they still do not fully utilize high-frequency data. This gap has led to the development of models that explicitly incorporate high-frequency data to improve volatility forecasting. One such model is the Heterogeneous Autoregressive (HAR) model, introduced by Corsi (2009), designed to capture the long memory in realized volatility by leveraging information from high-frequency data.

The HAR model represents a significant advancement in financial econometrics by addressing the limitations of traditional volatility models. It captures long memory in realized volatility measures, making it suitable for analyzing high-frequency financial data. The HAR model operates as an additive cascade model, decomposing volatility into components influenced by different market participants' actions. Although not being formally a long-memory model, the HAR model effectively captures volatility persistence and other stylized facts observed in financial data streams. Its original formulation, using realized variance (RV) and ordinary least squares for estimation, can be extended to address patterns such as non-Gaussianity, spikes/outliers, and conditional heteroskedasticity (Clements and Preve, 2021). Section 5.2.1 provides a review of HAR and its extensions.

Despite its strengths, a limitation of the HAR model and its extensions is their reliance solely on realized measures to forecast volatility, ignoring the additional information contained in the return series. Returns, which encapsulate the cumulative effect of market dynamics, provide valuable insights into investor behavior, market trends, and external factors influencing asset prices. By incorporating returns into volatility modeling, our approach aims to enrich the predictive power of the HAR model. This integration not only reduces the impact of micro-structure noise contained in the realized measures but also leverages the stability and information inherent in the returns. Ultimately, one of our primary contributions is to extend the HAR model to include returns, thereby enhancing the accuracy and robustness of risk forecasts, improving risk management practices and contributing to a more comprehensive understanding of financial market dynamics.

While HAR models are effective in predicting realized volatility, applications in financial risk management often necessitate forecasting VaR. VaR is indispensable for assessing and managing financial risk across various contexts, crucial for regulatory compliance and portfolio management strategies, as highlighted in Frey and Embrechts (2010) and Christoffersen (2011). Recognizing the practical importance of VaR, our research focuses directly on modeling and forecasting VaR via extending the HAR framework. This perspective helps enhance the relevance of HAR models in real-world financial applications. To rigorously evaluate the accuracy of our VaR forecasts, we employ quantile scores, which provide a consistent and robust assessment of predictive performance. This methodological choice facilitates a straightforward comparison and validation of our forecasting models, ensuring that our approach meets the stringent demands of risk management practices without making restrictive assumptions about return distributions. Therefore, the second contribution of our research is advancing the utility of HAR models by empowering them to directly and semi-parametrically forecast VaR and hence avoiding marking assumption on the return distribution, thereby robustifying the risk modeling and forecasting practice.

The integration of machine learning techniques into financial econometrics has gained considerable attention in recent years, offering substantial benefits in terms of predictive accuracy and the ability to capture complex, non-linear relationships within financial data (Kim and Won, 2018; Nguyen et al., 2022a,b). Traditional models such as the HAR model, while are effective at modeling volatility dynamics, face inherent limitations due to their reliance on linear regression frameworks. These models struggle to capture the non-linear dependencies commonly present in financial time series. To address these limitations, recent studies have turned to machine learning techniques to enhance the HAR framework's forecasting capabilities. For example, Arnerić et al. (2018) employ Feedforward Neural Networks (FNNs) into the HAR type models to better capture the non-linear behavior of RV. Further, the studies by Christensen et al. (2023); Branco et al. (2024); Pourrezaee and Hajizadeh (2024); Zhang et al. (2022); Patton and Zhang (2022) demonstrate how machine learning models—such as deep learning neural networks and stacked machine learning models—can improve volatility and VaR forecasting accuracy. In particular, the Recurrent Neural Networks (RNNs) are frequently used in these work. RNNs are designed for processing sequential data, making them superior to FNNs in the time series forecasting task (Lipton et al., 2015). These advancements underscore the critical role of machine learning and neural network methods in overcoming the traditional limitations of econometric models, also making RNN a natural choice for our work.

Recognizing these advancements, the third contribution of our research makes a significant stride by integrating RNNs into the HAR framework. More precisely, we derive daily, weekly, and monthly effects of realized volatility using three RNN structures, thereby capturing the non-linear and long-term effects of these variances on VaR. We refer to our approach as RNN-HAR. By embedding RNNs within the HAR model, our methodology aims to enhance the accuracy and robustness of VaR predictions. This model forms a hybrid framework, combining machine learning and econometrics, where the non-linear RNN technique captures the complex relationships in the data, while the HAR process preserves the meaningful financial and economic interpretability essential for forecasting.

Comparing to the existing machine learning and neural network based approaches in the literature, our approach has distinctive features. Christensen et al. (2023) investigate how various ML techniques work for the RV forecasting. Branco et al. (2024) focus on forecasting RV directly, by including all inputs including lagged RV and return in a NN, whilst we embed the RNN in a model that separates out daily, weekly, monthly RV inputs in a top layer linear framework, allowing us to assess/interpret the significance and effects of these different time-framed inputs, but still allowing them to influence VaR via RNN. Further, we estimate and forecast VaR directly, not RV, using the quantile loss function (where the information of return is utilized) for parameter estimation, whilst Branco et al. (2024) estimate RV directly and then employ filtered historical simulation to indirectly estimate/forecast VaR. Pourrezaee and Hajizadeh (2024) propose using stacking machine learning methods for the Bitcoin volatility and VaR forecasting. The stacking models include the standard GARCH, HAR and other machine learning models such as neural network, support vector regression, Long Short-Term Memory, and random forest. The VaRs are produced by each individual methods and used in the subsequent stacking modelling process, while our approach embed the RNN component into the HAR framework for direct VaR forecasting. Zhang et al. (2022) utilize temporal convolutional networks to forecast stock volatility which is subsequently used to generate VaR forecasts parametrically, and the realized data are not considered. Patton and Zhang (2022) use a neural network for estimating the optimal "bespoke" RVs with the high-frequency data, to tailor the estimate of volatility to the application in which it will be used, while the work focuses on forecasting the realized volatility.

Recent advancements in financial econometrics have leveraged sophisticated statistical inference techniques to enhance the accuracy of risk forecasting models. Our work utilizes loss-based generalized Bayesian inference, in conjunction with Sequential Monte Carlo (SMC) methods, for model estimation and prediction in RNN-HAR. Loss-based Bayesian inference is invaluable in scenarios where the likelihood function may be challenging to specify or is not readily available; see, e.g., Bissiri et al. (2016) and Knoblauch et al. (2019). This approach does not require assumption on the distribution of the returns, avoiding the possible issue of model misspecification. Given the complexity of our proposed RNN-HAR model structure, using SMC for Bayesian inference and sequential prediction is pivotal, allowing us to handle the inherent challenges of Bayesian computation in sophisticated models such as RNN-HAR.

In summary, the novelty of this research is fourfold. First, we enrich the HAR framework with the information from the return series. Second, we model and predict VaR directly. Third, we extend the HAR model by incorporating RNNs. Fourth, we consider loss-based generalized Bayesian inference with SMC for model estimation and prediction. Lastly, we evaluate the performance of our proposed model against basic HAR and three other extended HAR models using empirical data spanning nearly two decades (2000–2022) including 31 market indices, demonstrating its superior forecasting capabilities.

This chapter is organized as follows. Section 5.2 reviews the relevant background models. Section 5.3 proposes the RNN-HAR model. Bayesian inference and prediction using SMC is presented in section 5.4. Section 5.5 presents the empirical results. Section 5.6 summarizes the findings.

5.2 Background

This section briefly provides some background on the existing HAR models, the Asymmetric Laplace density based quantile loss and RNN.

5.2.1 HAR type models

This section reviews the HAR model and its extensions; we focus on a selection of widely recognized models. These models offer diverse methodologies for capturing volatility dynamics and have been extensively studied in the literature for their efficacy in risk management and forecasting applications. We provide a detailed exposition of each model and its respective formulations.

The HAR model of Corsi (2009) forecasts future RV based on the past daily, weekly and monthly RVs:

$$\mathrm{RV}_t = \beta_0 + \beta_d \mathrm{RV}_{t-1} + \beta_w \mathrm{RV}_{t-1|t-5} + \beta_m \mathrm{RV}_{t-1|t-22} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \tag{5.1}$$

where RV_{t-1} is the daily RV input. The weekly RV input $\mathrm{RV}_{t-1|t-5} = \frac{1}{5}\sum_{j=t-1}^{t-5} \mathrm{RV}_j$ is the average RV of the stock market index from time $t-1$ to time $t-5$. The monthly RV input $\mathrm{RV}_{t-1|t-22} = \frac{1}{22}\sum_{j=t-1}^{t-22} \mathrm{RV}_j$ is the average RV of the stock market index from time $t-1$ to time $t-22$. The coefficients β_d, β_w and β_m measure how the past daily, weekly and monthly volatility patterns influence future RVs. Despite its simple architecture, the HAR model empirically achieves highly accurate RV forecasts, which makes it popular among researchers and practitioners.

The HAR model's simplicity allows for various extensions, which can enhance its performance in different ways. One approach is applying transformations to realized volatility, which impacts the model's structure and properties. Corsi et al. (2008) describe the square root transformation of the HAR model (SqrtHAR) as follows:

$$\sqrt{RV_t} = \beta_0 + \beta_d \sqrt{RV_{t-1}} + \beta_w \sqrt{\mathrm{RV}_{t-1|t-5}} + \beta_m \sqrt{\mathrm{RV}_{t-1|t-22}} + \epsilon_t. \tag{5.2}$$

They show that this SqrtHAR model helps stabilize variance and improve the robustness of volatility forecasts compared to the original HAR model.

Another approach involves modifying the structure of the model to incorporate additional factors. For instance, Corsi and Renò (2012) introduce the Leverage HAR (LevHAR) model, which integrates past aggregated negative returns into the HAR model to capture the leverage effect. Following Asai et al. (2012), we only include the negative part of heterogeneous return since the positive part is usually insignificant. Therefore, we use the following definition of

LevHAR model as in Asai et al. (2012):

$$\begin{aligned}\mathrm{RV}_t = \beta_0 + \beta_d RV_{t-1} + \beta_w \mathrm{RV}_{t-1|t-5} + \beta_m \mathrm{RV}_{t-1|t-22} + \gamma_1 y_{t-1} I\left[y_{t-1} < 0\right] \ldots \\ +\gamma_2 y_{t-1|t-5} I\left[y_{t-1|t-5} < 0\right] + \gamma_3 y_{t-1|t-22} I\left[y_{t-1|t-22} < 0\right] + \epsilon_t,\end{aligned} \tag{5.3}$$

where $y_{t-1|t-n} = \frac{1}{n}\sum_{j=t-1}^{t-n} \mathrm{y}_j$ is the average return of the stock market index from time $t-1$ to time $t-n$, defined under the same manner as weekly and monthly RV. $I[x < 0]$ is the indicator function, which takes 1 if x is negative and 0 otherwise. γ_1, γ_2 and γ_3 are the parameters to capture the leverage effect. Unlike the HAR and SqrtHAR models, the LevHAR considers the information from return during its modeling process, however, such information only contains the signs of returns.

While the HAR, SqrtHAR, LevHAR and other variants of HAR models (e.g., Patton and Sheppard (2015), Bollerslev et al. (2016)) represent significant advancements in volatility modeling, they share common limitations. These models primarily focus on capturing volatility dynamics through realized volatilities and other factors but do not explicitly incorporate returns into their frameworks, except the LevHAR model where only the signs of returns are considered. This omission is critical as returns play a fundamental role in determining asset price movements and directly influence risk measures such as VaR. By neglecting to integrate returns and forecast VaR, these models might overlook crucial aspects of financial risk assessment, limiting their applicability in practical risk management contexts. Addressing these gaps is essential for developing more comprehensive and accurate models that better reflect the complexities of financial markets.

Another potential avenue for enhancing modeling accuracy involves integrating the HAR model into the GARCH type models. Huang et al. (2016) introduce the Realized-HAR-GARCH model, which expands the volatility dynamic equation by integrating the HAR structure of realized variance into the GARCH equation. This model incorporates multiple lags of a realized measure and employs a measurement equation to make necessary adjustments. It focuses on capturing latent volatility associated with inter-day returns rather than the intra-day returns captured by realized measures. The authors argue that their Realized-HAR-GARCH model offers more nuanced dynamics for realized measures than the original HAR model. The Realized-HAR-GARCH model is written as:

$$\begin{aligned} y_t &= \mu + \sqrt{v_t} z_t, \\ v_t &= \omega + \gamma v_{t-1} + \beta_d RV_{t-1} + \beta_w RV_{t-1|t-5} + \beta_m RV_{t-1|t-22}, \\ RV_t &= \xi + \phi v_t + \tau_1 z_t + \tau_2 (z_t^2 - 1) + u_t, \end{aligned} \tag{5.4}$$

where v_t is the conditional variance, $z_t \sim N(0,1)$, and $u_t \sim N(0, \sigma_u^2)$ with z_t and u_t being independent. While the Realized-HAR-GARCH model improves the capture of long memory in underlying volatility and offers more accurate multi-period out-of-sample volatility forecasts over various forecast horizons (Huang et al., 2016), it does not forecast VaR directly. This limitation highlights a gap that our proposed model aims to fill by incorporating VaR forecasting directly into the framework.

Clements and Preve (2021) explore various aspects to improve the HAR model forecast performance. They consider different estimators, data transformation and combination schemes. There, they consider two weighted least squares schemes and robust regression as estimators, log and square root as transformations, and six combinations of the different estimators and transformations in the empirical study. They conclude that their simple remedies outperform the standard HAR forecasts. Further, they suggest estimating model parameters under a loss function coherent with the final application of the forecasts, such as VaR forecasting. This suggestion lays the foundation for our proposed model.

To summarize, despite the advancements made by the HAR model and its aforementioned extensions in capturing volatility dynamics, there remain significant gaps. These models do not explicitly incorporate returns into their frameworks and often neglect the direct forecasting of VaR.

5.2.2 The quantile loss and asymmetric Laplace density

Koenker and Machado (1999) note that the quantile regression estimator is equivalent to the maximum likelihood estimator based on the Asymmetric Laplace (AL) density with the mode being the quantile. Let y_t is the return on day t. Its quantile loss-based AL density is

$$f(y_t|Q_t,\sigma) = \frac{\alpha(1-\alpha)}{\sigma} \exp\left(\frac{-(y_t - Q_t)(\alpha - I(y_t \leqslant Q_t))}{\sigma}\right), \tag{5.5}$$

where Q_t is the α-level quantile, $\alpha \in (0,1)$ and σ is the scale parameter. See Taylor (2019) for more details. Although (5.5) is a well-defined density function, we do not assume that y_t follows the AL distribution. Instead, we view (5.5) as a loss-based density that allows us to define a likelihood-alike function and then perform the subsequent loss-based generalized Bayesian inference. Via incorporating this quantile loss-based AL density into the inference process, we avoid the need of specifying the probabilistic assumption about the return distribution; meaning that the proposed RNN-HAR is a semi-parametric risk forecasting framework. Employing the quantile loss also enables the loss-based generalized Bayesian inference which represents a paradigm shift in statistical modeling, diverging from traditional methods that rely on likelihood functions and strict distributional assumptions. Details of the inference method and technical implementation details are shown in Section 5.4.

5.2.3 RNN

RNN and its variants excel at learning and representing non-linear patterns and dependencies of the sequential data. In this paper, we use the basic RNN of **?**, while more sophisticated architectures, such as the Long Short-Term Memory (LSTM) model of **?**, could also be employed.

Let $\{D_t = (x_t, y_t), t = 1, 2, ...\}$ be the data with x_t the input and y_t the output. We use (x_t, y_t) in this section as a generic notation, not necessarily applicable to the return data in the other sections. The task is to model the conditional mean $\widehat{y_t} = \mathbb{E}(y_t|x_t, D_{1:t-1})$. The basic RNN model is defined as:

$$\begin{aligned} h_t &= \phi\left(Vx_t + Wh_{t-1} + b\right), \\ y_t &= \beta_0 + \beta_1 h_t + \epsilon_t. \end{aligned} \tag{5.6}$$

Here, ϕ is an activation function, such as Tanh or Sigmoid, ϵ_t is an random error, and h_t is the hidden state. The important feature of this RNN structure is the hidden state h_t, which feeds itself with its lagged value h_{t-1} and the current information from the input x_t, leading to a mechanism for storing and updating the memory in the data. Our research aims to integrate RNNs within the HAR framework to improve the VaR forecast of the latter.

5.3 The proposed RNN-HAR model

The HAR model is well-known for capturing the long memory effects in financial volatility, while RNNs excel in learning intricate patterns and non-linear dynamics from sequential data. In this study, we aim to integrate RNNs into the HAR model, leveraging the strengths of both frameworks. Our RNN-HAR model is specified as follows:

$$\text{VaR}_t = \beta_0 + \beta_d h_t^d + \beta_w h_t^w + \beta_m h_t^m, \tag{5.7a}$$

$$h_t^d = \text{RNN}(RV_{t-1}, h_{t-1}^d) = \phi(\alpha_0^d + \alpha_1^d RV_{t-1} + \alpha_2^d h_{t-1}^d), \tag{5.7b}$$

$$h_t^w = \text{RNN}(RV_{t-1|t-5}, h_{t-1}^w) = \phi(\alpha_0^w + \alpha_1^w RV_{t-1|t-5} + \alpha_2^w h_{t-1}^w), \tag{5.7c}$$

$$h_t^m = \text{RNN}(RV_{t-1|t-22}, h_{t-1}^m) = \phi(\alpha_0^m + \alpha_1^m RV_{t-1|t-22} + \alpha_2^m h_{t-1}^m). \tag{5.7d}$$

Figure 5.1 provides a visualization of the architecture of the proposed RNN-HAR model. The RNN-HAR model produces the VaR forecast at time t, given the information up to time $t-1$. It contains three RNNs that produce the daily, weekly and monthly hidden states, h_t^d, h_t^w and h_t^m. The model first computes these states using their lagged values, h_{t-1}^d, h_{t-1}^w, h_{t-1}^m, and the

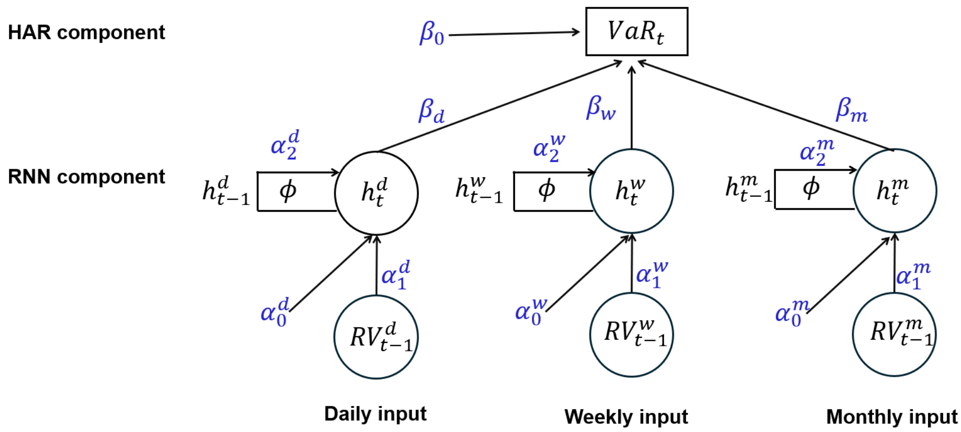

Figure 5.1. Visualization of the model architecture of the proposed RNN-HAR model. Model parameters are highlighted in blue.

daily RV_{t-1}, weekly $RV_{t-1|t-5}$ and monthly $RV_{t-1|t-22}$ as inputs. The activation function ϕ is selected as the tanh function. These calculations are specified in Equations (5.7b)-(5.7d). Then, the states h_t^d, h_t^w and h_t^m are aggregated, as in Equation (5.7a), to produce the forecast of VaR at time t. As in the standard HAR model, the parameters β_d, β_w and β_m weigh the influence of the past daily, weekly and monthly volatilities on the future VaR.[1]

The RNN-HAR model expresses VaR_t as a function comprising an intercept and three distinct RNN structures: one for daily data, one for weekly data, and another for monthly data. This approach explicitly captures the non-linear and long-term memory effects that daily, weekly, and monthly realized measures exert on VaR. By integrating RNN and HAR, our proposed model aims to improve the accuracy and robustness of VaR forecasting in the financial markets. Compared to the HAR model in (5.1), our new model introduces two key extensions. First, it employs three RNNs to model the non-linear and long-term dependence relationships between daily, weekly, and monthly RV and VaR. Second, it directly estimates VaR as the α-level quantile of the return distribution, leveraging the AL-based quantile loss function (5.5) for inference. For parametric models such as GARCH, VaR estimation typically requires an assumption about the return distribution. Avoiding such assumptions enhances statistical robustness. Our approach estimates VaR directly, using the quantile score as the loss function, thereby eliminating the need for a predefined return distribution. Unlike the HAR model, which relies on realized measures and disregards the return series (except the LevHAR model that uses the signs of returns), the RNN-HAR model uses the information contained in returns. By using the

[1]With only one hidden layer and one hidden unit for the daily, weekly and monthly RNN components, the proposed RNN-HAR already produces significantly improved VaR forecasts compared to existing models, as shown in Section 5.5. Therefore we choose to use only one hidden layer and one hidden unit in the paper. However, more complex RNN architectures with larger hidden units, such as LSTM, could be also explored as interesting future work.

AL-based quantile loss, the RNN-HAR model incorporates returns directly and aligns more closely with the ultimate goal of computing VaR.

Comparing to the existing HAR type models, the RNN-HAR extends them via incorporating the non-linear neural networks. To more explicitly evaluate the impact of incorporating RNNs, we include below the linear VaR-HAR framework for comparison, via modifying the models developed in **?**. The VaR-HAR directly forecasts VaR via employing the daily, weekly and monthly RV as below:

$$\mathrm{VaR}_t = \beta_0 + \beta_d \mathrm{RV}_{t-1} + \beta_w \mathrm{RV}_{t-1|t-5} + \beta_m \mathrm{RV}_{t-1|t-22}. \tag{5.8}$$

In Section 5.5, we compare our RNN-HAR with the linear VaR-HAR model (5.8), both with the quantile loss, to evaluate the impact of the three RNN components.

Furthermore, the proposed RNN-HAR model can be easily extended to other HAR specifications, such as the SqrtHAR (5.2) and LevHAR (5.3). For instance, to explicitly capture the leverage effect, the RNN-LevHAR model can be formulated as follows

$$\begin{aligned}
\mathrm{VaR}_t &= \beta_0 + \beta_d h_t^d + \beta_w h_t^w + \beta_m h_t^m + \gamma_1 y_{t-1} I\left[y_{t-1} < 0\right] ... \\
&+ \gamma_2 y_{t-1|t-5} I\left[y_{t-1|t-5} < 0\right] + \gamma_3 y_{t-1|t-22} I\left[y_{t-1|t-22} < 0\right], \\
h_t^d &= \mathrm{RNN}(RV_{t-1}, h_{t-1}^d) = \phi(\alpha_0^d + \alpha_1^d RV_{t-1} + \alpha_2^d h_{t-1}^d), \\
h_t^w &= \mathrm{RNN}(RV_{t-1|t-5}, h_{t-1}^w) = \phi(\alpha_0^w + \alpha_1^w RV_{t-1|t-5} + \alpha_2^w h_{t-1}^w), \\
h_t^m &= \mathrm{RNN}(RV_{t-1|t-22}, h_{t-1}^m) = \phi(\alpha_0^m + \alpha_1^m RV_{t-1|t-22} + \alpha_2^m h_{t-1}^m).
\end{aligned}$$

However, to maintain a clear focus, this paper primarily investigates the RNN-HAR model.

In summary, our proposed RNN-HAR model addresses several key limitations in existing volatility models. First, it integrates RNNs into the HAR framework, enabling the capture of complex, non-linear dependencies and long-range memory in volatility forecasting. Second, by employing the quantile loss function, the model eliminates the need to assume a specific return distribution, overcoming a common limitation of parametric models like GARCH. Third, instead of forecasting RV as in the HAR model, we directly estimate VaR, offering a more comprehensive and convenient framework for risk assessment. These contributions highlight the model's potential to enhance volatility forecasting accuracy and improve risk management strategies in financial markets. The next section details the model estimation process and the sequential forecasting of VaR using the loss-based Bayesian SMC method.

5.4 Loss-based Bayesian inference and forecasting

Recent advances in Bayesian computation have made it easy to use Bayesian approach. In addition to the standard MCMC method, alternative Bayesian inference techniques, such as

Sequential Monte Carlo, have significantly enhanced the accessibility of Bayesian models, even for large-scale applications like Bayesian deep learning. Bayesian approach also provides a convenient way for expanding-window sequential prediction as the posterior from the previous time step becomes the prior of the next; see, e.g., Nguyen et al. (2022a); Duan and Fulop (2015).

In addition, loss-based generalized Bayesian inference represents a paradigm shift in statistical modeling, diverging from traditional methods that rely on likelihood functions and strict distributional assumptions. See, e.g., Bissiri et al. (2016); Knoblauch et al. (2019); Matsubara et al. (2021); Frazier et al. (2024). Unlike classical frameworks where likelihood functions necessitate specific probabilistic assumptions about the data distribution, loss-based Bayesian inference emphasizes the use of loss functions to guide Bayesian inference and prediction. This approach is particularly advantageous in scenarios where underlying data distributions are complex or unknown, mitigating potential model misspecification by offering flexibility and robustness. By focusing on a loss function rather than a likelihood, researchers can tailor models to better reflect real-world uncertainties and variations by updating beliefs about model parameters in a robust manner, particularly in situations where justifying strict assumptions about data distributions is challenging. Recent developments, as discussed in Bissiri et al. (2016) and Knoblauch et al. (2019), highlight the use of loss functions in updating beliefs and parameter estimation without restrictive distributional assumptions, thereby enhancing the applicability and reliability of Bayesian models in complex data environments. Li et al. (2023) observe that MLE-based quantile regression models are sensitive to initial conditions and advocate for a loss-based Bayesian quantile regression approach for estimating joint VaR and ES models.

Building on these principles and given the AL based quantile loss in Equation (5.5), we adopt a quantile loss-based Bayesian approach to address the challenge of modeling financial time series and forecasting VaR without presuming a specific return distribution. This method effectively integrates prior knowledge with observed data, facilitating robust parameter estimation without imposing strong distributional assumptions.

Dealing with the scaling parameters, such as σ in (5.5), can be a not trivial problem in generalized Bayesian inference (Knoblauch et al., 2019). Fortunately, for the AL density in (5.5), by adopting an inverse Gamma prior on σ, its full conditional distribution is inverse Gamma. This allows straightforward integration and simplifying the likelihood function in subsequent Bayesian analysis steps. As a result, the posterior distribution of the parameter of interest is obtained without explicit consideration of σ (Gerlach et al., 2011).

In Bayesian inference, priors serve as our initial assumptions regarding model parameters before observing any data.

Following an exploration of different prior combinations, we opt for a Normal prior with a mean of zero and a variance of 0.01 for the recurrent parameters in our RNN-HAR model. This choice reflects our initial expectation that these parameters are centered around zero with small variability. For the model parameters $\beta_0, \beta_d, \beta_w, \beta_m$, we assume a Normal distribution with a mean of 0 and a standard deviation of 1, indicating our initial uncertainty and allowing for exploration across a spectrum of parameter values.

Integrating loss-based generalized Bayesian inference into the RNN-HAR framework yields several advantages: providing a coherent methodology for updating beliefs based on observed data, robustifing inference for the model parameters without relying on data assumptions of the returns, allowing convenient and efficient sequence prediction based on Sequence Monte Carlo. In Section 5.5.3, we evaluate the effectiveness of the employed loss-based SMC method via comparing its performance with the standard optimization routine in Matlab, to empirically demonstrate the advantages of the SMC method.

5.4.1 Sequential Monte Carlo

The SMC method is a powerful tool for Bayesian inference and sequential prediction; see, e.g., Del Moral et al. (2006) and Gunawan et al. (2022). This approach is particularly effective and convenient for volatility modeling and forecasting, where generating sequential expanding-window forecasts is essential.

SMC uses a set of samples, often called particles, to approximate a sequence of probability distributions. This sequential updating enables efficient estimation of posterior distributions in complex and non-linear models where conventional Monte Carlo methods are computationally demanding or impractical. SMC allows straightforward expanding-window one-step-ahead forecast calculations, which makes it particularly useful for volatility forecasting (Nguyen et al., 2022b). The SMC method also provides an accurate estimate of the marginal likelihood, which is an important quality often used for model selection. These attributes make SMC an attractive approach for Bayesian inference and sequential forecasting in our RNN-HAR model. To further demonstrate the effectiveness of the loss-based SMC method, in Section 5.5.3 we have compared the SMC method with the standard optimization routine.

There are two common SMC approaches in the literature: likelihood annealing and data annealing (Nguyen et al., 2022b). The first approach is designed for sampling from the posterior while the second is for sequential prediction, thus SMC with likelihood annealing is used for in-sample analysis and SMC with data annealing is used for out-of-sample forecasting in this paper. We present these two approaches in the following sections.

Likelihood annealing

The loss-based generalized posterior distribution in our RNN-HAR model is

$$\pi(\theta) = p(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)p(\theta), \tag{5.9}$$

where $p(\theta)$ is the prior and $p(y_{1:T}|\theta)$ is the loss-based likelihood-alike function derived from (5.5), and $\theta = \{\beta_0, \beta_d, \beta_w, \beta_m, \alpha_0^d, \alpha_1^d, \alpha_2^d, \alpha_0^w, \alpha_1^w, \alpha_2^w, \alpha_0^m, \alpha_1^m, \alpha_2^m, \sigma\}$ is the parameter vector set which includes all the parameters in RNN-HAR. For sampling from the generalized posterior $\pi(\theta)$, SMC first samples a set of M weighted particles $\left\{W_0^j, \theta_0^j\right\}_{j=1}^M$ from an easy-to-sample distribution $\pi_0(\theta)$, such as the prior $p(\theta)$, and then traverses these particles through intermediate distributions $\pi_t(\theta), t = 1, ..., K$ with $\pi_K(\theta) = \pi(\theta)$. In this paper, we set $\pi_0(\theta) = p(\theta)$ as it is possible to sample from the prior $p(\theta)$. The likelihood annealing SMC sampler uses the following intermediate distributions

$$\pi_t(\theta) := \pi_t(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)^{\gamma_t} p(\theta), \tag{5.10}$$

where the γ_t are referred to as the temperature levels satisfying $0 = \gamma_0 < \gamma_1 < \gamma_2 < ... < \gamma_k = 1$. Note that the sequence of distributions π_t requires the full training data $y_{1:T}$ to be available. SMC with a likelihood annealing sampler is suitable for in-sample analysis.

Several methods exist to implement SMC in practice; here, we consider the method used by Nguyen et al. (2022b), which uses three main steps: reweighting, resampling, and a Markov move.

Reweighting: At the beginning of iteration t, the set of weighed particles $\left\{W_{t-1}^j, \theta_{t-1}^j\right\}_{j=1}^M$ that approximate the intermediate distribution $\pi_{t-1}(\theta)$ is reweighted to approximate the target $\pi_t(\theta)$. The efficiency of these weighted particles as a representation of $\pi_t(\theta)$ is often measured by the effective sample size (ESS) defined in (5.13).

Resampling: The particles are resampled if the ESS is below a pre-specified threshold.

Markov Move: The resulting equally weighted samples are then refreshed by a Markov kernel whose invariant distribution is $\pi_t(\theta)$.

Following Nguyen et al. (2022b), we choose the tempering sequence γ_t adaptively to ensure a sufficient level of particle efficiency by selecting the next value of γ_t such that ESS stays above a threshold. We now present the likelihood annealing SMC sampler, which is adapted from Nguyen et al. (2022b). Note that we do not compute the marginal likelihood estimate as its meaning is not well justified in the generalized Bayesian inference setting.

1. Sample $\theta_0^j \sim p(\theta)$ and set $W_0^j = 1/M$ for $j = 1...M$
2. For $t = 1, ..., K$
   (a) Resampling: Compute the unnormalized weights
   $$w_t^j = W_{t-1}^j \frac{p(y_{1:T}|\theta_{t-1}^j)^{\gamma_t} p(\theta_{t-1}^j)}{p(y_{1:T}|\theta_{t-1}^j)^{\gamma_{t-1}} p(\theta_{t-1}^j)} = W_{t-1}^j p(y_{1:T}|\theta_{t-1}^j)^{\gamma_t - \gamma_{t-1}}, j = 1, ..., M \tag{5.11}$$
   and set the new normalized weights
   $$w_t^j = \frac{w_t^j}{\sum_{s=1}^M w_t^s}, j = 1, ..., M \tag{5.12}$$
   (b) Compute the effective sample size (ESS)
   $$ESS = \frac{1}{\sum_{j=1}^M \left(w_t^j\right)^2}, j = 1, ..., M \tag{5.13}$$
   if ESS $< cM$ for some $0 < c < 1$, **then**
   - Resampling: Resample from $\left\{\theta_{t-1}^j\right\}_{j=1}^M$ using the weights $\left\{W_t^j\right\}_{j=1}^M$, and then set $W_t^j = 1/M$ for $j = 1, ..., M$, to obtain the new equally-weighted particles $\left\{\theta_t^j, W_t^j\right\}_{j=1}^M$.
   - Markov move: For each $j = 1, ..., M$ move the sample θ_t^j according to N_{lik} random walk Metropolis-Hasting steps:
     – Generate a proposal $\theta_t^{j'}$ from a from a multivariate Normal distribution $N(\theta_t^j, \Sigma_t)$ with Σ_t the covariance matrix.
     – Set $\theta_t^j = \theta_t^{j'}$ with the probability
     $$\min\left(1, \frac{p(y_{1:T}|\theta_t^{j'})^{\gamma_t} p(\theta_t^{j'})}{p(y_{1:T}|\theta_t^j)^{\gamma_t} p(\theta_t^j)}\right); \tag{5.14}$$
     otherwise keep θ_t^j unchanged.

   end

Data annealing

For out-of-sample expanding-window forecasts where the posterior of the model parameters θ is updated once new data arrive, it is necessary to use SMC with the data annealing (Nguyen et al., 2022b). The following sequence of distributions is used to generate weighted particles in this SMC sampler.

$$\pi_t(\theta) := \pi_t(\theta|y_{1:t}) \propto p(y_{1:t}|\theta)p(\theta) \propto \pi_{t-1}(\theta)p(y_t|\theta, y_{1:t-1}), \quad t = T+1, ... \tag{5.15}$$

with $y_{1:t}$ the data available up to time t, and $y_{1:T_{in}}$ the in-sample data. The SMC procedure for sampling from the sequence $\pi_t(\theta)$ in (5.15) is the same as before, except that the unnormalized weights become

$$w_t^j = W_{t-1}^j \frac{p(y_{1:t}|\theta_{t-1}^j)p(\theta_{t-1}^j)}{p(y_{1:t-1}|\theta_{t-1}^j)p(\theta_{t-1}^j)} = W_{t-1}^j p(y_t|y_{1:t-1},\theta_{t-1}^j), j = 1, ..., M \tag{5.16}$$

In line with Nguyen et al. (2022b), we employ SMC with likelihood annealing for in-sample Bayesian inference and SMC with data annealing for generating one-step-ahead forecasts with expanding window. The specific implementation settings for the SMC samplers are outlined below.

Table 5.1. SMC settings

Variable	Description	Value
K	Number of annealing levels	10,000
M	Number of particles	2,000
c	Constant of the ESS threshold	0.8
N_{lik}	Number of Markov moves in the SMC with likelihood annealing	10
N_{data}	Number of Markov moves in the SMC with data annealing	20

5.5 Empirical study

5.5.1 Data description

The initial daily closing prices and realized variance calculated with 5-minute high-frequency data were sourced from the Oxford-man Institute's realized library (Heber et al., 2009), covering 31 global market indices for the period from January 2000 to June 2022. Daily return values were computed based on the daily price data. Our analysis encompasses significant events such as the COVID-19 pandemic. Due to varying non-trading days across different markets throughout the period under study, sample sizes and forecasting periods vary across each series. To ensure consistency, we standardized the length of all return series to the last $T = 3000$ observations (around 12 years), except for the BVLG market which has only 2398 values. These series were then divided into an in-sample period comprising the first $T_{in} = 2000$ observations and an out-of-sample period comprising the last $T_{out} = 1000$ observations.

Table 5.2 provides descriptive statistics for each market in this study. Among these markets, the emerging market BVSP exhibits the highest standard deviation, indicating greater variability than the more established markets. On the other hand, IXIC stands out as having the highest mean return.

Table 5.2. Summary statistics of the return series.

Market	N	Mean	Std	Skewness	Kurtosis	Min	Max
AEX	3000	0.0234	1.1183	-0.5723	9.7439	-10.6384	8.6462
AORD	3000	0.0005	0.9052	-0.8246	8.8584	-7.0728	4.4291
BFX	3000	0.0103	1.1379	-0.9672	15.2145	-14.2235	6.8949
BSESN	3000	-0.0081	1.1011	-0.9482	16.5074	-13.8222	8.2043
BVLG	2398	0.0000	1.1353	-0.8747	10.1403	-10.9619	6.5051
BVSP	3000	-0.0197	1.5750	-0.7933	14.5011	-16.0260	12.9906
DJI	3000	0.0209	1.0794	-0.9663	24.5143	-13.8247	10.7360
FCHI	3000	0.0156	1.2663	-0.5953	9.9544	-11.9977	7.7889
FTMIB	3000	0.0011	1.5565	-1.0556	14.1095	-18.5434	8.5472
FTSE	3000	0.0089	1.0202	-0.5866	10.4931	-10.1382	7.7806
GDAXI	3000	0.0136	1.2752	-0.4607	9.4086	-11.8749	9.7516
GSPTSE	3000	-0.0004	0.9093	-1.8510	33.5737	-13.1944	9.1019
HSI	3000	-0.0038	1.2255	-0.1899	6.0910	-5.9839	8.7032
IBEX	3000	-0.0026	1.3829	-0.7636	11.5616	-12.7119	8.1168
IXIC	3000	0.0358	1.2680	-0.7839	11.9963	-13.1586	8.9088
KS11	3000	-0.0036	1.0185	-0.5060	10.6059	-10.1935	7.0957
KSE	3000	-0.0148	1.0454	-0.5878	6.8834	-7.3188	4.6228
MXX	3000	-0.0203	0.9770	-0.5030	7.5941	-6.9709	5.0541
N255	3000	0.0266	1.3326	-0.4221	8.0216	-11.1593	7.7255
NSEI	3000	-0.0014	1.1074	-0.9426	15.5131	-13.6741	8.0057
OMXC20	3000	0.0109	1.1455	-0.3443	5.7130	-7.8569	5.1068
OMXHPI	3000	0.0102	1.1816	-0.6401	8.8448	-10.7945	6.1853
OMXSPI	3000	0.0056	1.1507	-0.7847	10.3512	-11.8285	6.9907
OSEAX	3000	-0.0011	1.1294	-0.6811	8.9272	-9.8730	5.8014
RUT	3000	0.0120	1.4329	-0.8890	13.8693	-15.2513	8.8834
SMSI	3000	-0.0041	1.3624	-0.7651	12.1109	-14.0552	8.1712
S&P500	3000	0.0250	1.1006	-0.8630	18.0908	-12.6874	8.9440
SSEC	3000	-0.0113	1.3264	-0.9804	9.5522	-8.8919	5.6243
SSMI	3000	0.0117	0.9885	-0.8894	12.5799	-10.1410	6.7734
STI	3000	0.0223	0.9610	-1.5480	26.6530	-14.7222	5.9946
STOXX50E	3000	0.0135	1.2777	-0.5419	9.6881	-11.9999	8.6605

5.5.2 Forecasting set-up

We perform an empirical analysis to compare the proposed RNN-HAR model with the conventional HAR model as in Equation (5.1) and its extensions. We consider the extensions that can be executed using the available dataset for this comparative analysis, including SqrtHAR as detailed in Equation (5.2), LevHAR as presented in Equation (5.3), RHARGARCH as described in Equation (5.4) and the semi-parametric VaR-HAR as specified in Equation (5.8).

It is worth noting that our proposed RNN-HAR and the linear VaR-HAR models directly output VaR forecasts, while for the parametric HAR type and RHARGARCH models the Realized Variance RV_{t+1} and the conditional variance v_{t+1} forecasts are produced, given information up to t. We first calculate $\sqrt{RV_{t+1}}$ and $\sqrt{v_{t+1}}$ as the volatility forecast estimator, then the inverse of the standard Normal cdf is used as the scaling factor to transform the volatility forecast to VaR forecast $\widehat{\text{VaR}}_{t+1}$ (Clements and Preve, 2021).

We implement the following procedure for a daily expanding window one-step-ahead forecasting with parameters re-estimated at each forecasting step, to keep the estimated models current and utilize historical data efficiently. To forecast VaR in the RNN-HAR model at the time t within the test data ($T_{in} + 1 \leqslant t \leqslant T_{in} + T_{out}$), let $\{\theta^{(i)}\}_{i=1}^{M}$ be the particles approximating the posterior distribution $\pi_t(\theta) = p(\theta|y_{1:t})$. For each particle $\theta^{(i)}$, we compute $Q_{t+1}^{(i)}$ according to the RNN-HAR equation (5.7), which represents an estimate of VaR forecast $\widehat{\text{VaR}}_{t+1}$. The resulting M values $\{Q_{t+1}^{(i)}\}_{i=1}^{M}$ represent the posterior predictive distribution of VaR_{t+1}, given the information up to time t. The arithmetic mean, $\widehat{Q}_{t+1}$, of these realizations serves as the point forecast for VaR_{t+1}. The similar forecasting procedure is also applied to the linear VaR-HAR model with quantile loss.

Since quantiles are elicitable, as defined in Gneiting (2011), and the standard quantile score function is strictly consistent: the expected quantile score is minimized with the true quantile series, a key VaR forecasting accuracy evaluation metric employed is the predictive quantile score. Evaluated on the out-of-sample data of size T_{out}, the quantile score is computed as:

$$\text{QS} = \frac{1}{T_{out}} \sum_{t=T_{in}+1}^{T_{in}+T_{out}} (y_t - \widehat{\text{VaR}}_t)(\alpha - I(y_t \leqslant \widehat{\text{VaR}}_t)) \;, \tag{5.17}$$

where $\widehat{\text{VaR}}_{T_{in}+1}, \ldots, \widehat{\text{VaR}}_{T_{in}+T_{out}}$ is a series of quantile forecasts at probability level α for out-of-sample returns $y_{T_{in}+1}, \ldots, y_{T_{in}+T_{out}}$.

5.5.3 Comparison on the inference methods

Before investigating the forecasting results from various competing models, we first evaluate the effectiveness of the employed loss-based SMC method via comparing its performance with the standard optimization routine. To start, for the linear VaR-HAR model, we estimate the model via both the SMC method and the standard constrained optimization routine 'fmincon' with the interior point algorithm in Matlab: https://au.mathworks.com/help/optim/ug/fmincon.html. For the linear VaR-HAR model (5.8) estimated with SMC and fmincon, Table 5.3 presents the out-of-sample quantile score values as in Equation (5.17) across 31 indices on the 1% probability level. In addition, for each market the parameter estimates using the first in-sample data produced by both inference hods are also included for comparison. In general, although both methods returned comparable parameter estimates across 31 markets, the out-of-sample quantile scores from the loss-based Bayesian SMC method are favored for 29 out of 31 indices, demonstrating the effectiveness of employing the SMC approach.

In addition, we also tested the estimation of the RNN-HAR model using the fmincon optimization routine. However, we observed convergence issues, and the fmincon outputs were significantly less stable and robust compared to the SMC method. Consequently, SMC was chosen as the inference method for both the proposed RNN-HAR model and the linear VaR-HAR model.

Table 5.3. For the linear VaR-HAR model (5.8) estimated with SMC and fmincon, the out-of-sample quantile score values across 31 markets on the 1% probability level. The parameter estimates are produced by the first in-sample data for each market. The last rows show the average values of 31 markets. Blue highlighting indicates the favored quantile score via comparing the SMC and fmincon.

	Linear VaR-HAR estimated with SMC					Linear VaR-HAR estimated with fmincon				
Market	QS	β_0	β_d	β_w	β_m	QS	β_0	β_d	β_w	β_m
AEX	0.0405	-1.4536	-1.0757	0.2006	-0.1523	0.0778	-1.3462	-1.1491	0.2278	-0.1538
AORD	0.0414	-1.4128	-0.7802	0.3342	-0.5898	0.0462	-1.4933	-0.9931	0.5431	-0.5560
BFX	0.0421	-1.3053	-1.2055	0.0448	-0.0106	0.0473	-1.3250	-1.2396	0.0463	0.0219
BSESN	0.0475	-1.3533	-0.0944	0.0151	-0.9877	0.0628	-1.3747	-0.0502	0.0305	-1.0130
BVLG	0.0383	-1.3246	-1.0081	-0.4194	0.0618	0.0512	-1.3182	-0.9281	-0.4582	0.0308
BVSP	0.0560	-2.3298	-0.4106	-0.1340	0.0389	0.0625	-2.3645	-0.3961	-0.1601	0.0605
DJI	0.0415	-1.356	-0.6504	-0.2506	-0.0663	0.0427	-1.3753	-0.6623	-0.1589	-0.0935
FCHI	0.0518	-1.7900	-0.6449	0.1471	-0.2668	0.0628	-1.7798	-0.6318	0.1691	-0.2984
FTMIB	0.0643	-2.5848	-0.3137	-0.0463	-0.0709	0.0708	-2.7106	-0.2880	-0.0845	-0.0025
FTSE	0.0447	-1.6049	-0.2086	-0.2299	-0.2969	0.0474	-1.6267	-0.2165	-0.1887	-0.3087
GDAXI	0.053	-1.7795	-0.5354	0.0798	-0.4612	0.0564	-1.7679	-0.5607	0.0910	-0.4455
GSPTSE	0.0439	-1.4800	-0.9687	0.3469	-0.4159	0.0449	-1.4664	-0.9977	0.3740	-0.4170
HSI	0.0450	-1.3654	-0.6090	-0.248	-0.3308	0.0488	-1.2939	-0.5661	-0.3290	-0.3352
IBEX	0.0488	-2.196	-0.3747	-0.0309	-0.1050	0.0487	-2.2244	-0.3964	-0.0025	-0.0990
IXIC	0.0491	-1.8906	-0.6612	0.0324	-0.0413	0.0550	-1.8853	-0.7065	0.0343	0.0343
KS11	0.0372	-1.5016	-0.7149	0.4231	-0.9325	0.0480	-1.5061	-0.7930	0.5912	-1.0056
KSE	0.0440	-1.8672	0.0164	-1.8388	0.8581	0.0605	-1.8423	-0.0351	-1.7903	0.8462
MXX	0.0424	-1.6464	-0.5226	-0.0120	-0.0608	0.0473	-1.6232	-0.5125	-0.0388	-0.0634
N225	0.0401	-2.2739	-0.3738	-0.0088	-0.3090	0.0406	-2.2913	-0.3678	-0.0011	-0.3211
NSEI	0.0471	-1.3097	-0.1116	-0.0492	-0.8404	0.0608	-1.2973	-0.1189	-0.0268	-0.8734
OMXC20	0.0418	-2.0063	-0.1989	-0.3810	-0.1118	0.0437	-1.9774	-0.2359	-0.3464	-0.0991
OMXHPI	0.0481	-1.7042	-0.6576	0.3479	-0.6638	0.0705	-1.7094	-0.6352	0.3705	-0.6952
OMXSPI	0.0469	-1.7509	-0.285	-0.2583	-0.3195	0.0598	-1.7488	-0.2785	-0.2507	-0.3293
OSEAX	0.0430	-1.6700	-0.3910	-0.1290	-0.4080	0.0523	-1.6759	-0.4298	-0.1329	-0.3692
RUT	0.0562	-2.1142	-0.2916	0.0275	-0.2212	0.0661	-2.1214	-0.2824	0.0221	-0.2197
SMSI	0.0507	-2.4481	-0.1652	-0.0867	-0.1267	0.0545	-2.4827	-0.1546	-0.0835	-0.1272
SPX	0.0433	-1.4931	-0.7093	-0.0621	-0.2226	0.0441	-1.4972	-0.7259	-0.0017	-0.2782
SSEC	0.0495	-2.3985	0.2147	-0.4492	-0.5282	0.0485	-2.5148	0.2495	-0.5301	-0.4024
SSMI	0.0399	-1.4275	-1.0967	0.0814	-0.0141	0.0404	-1.4161	-1.0996	0.0473	0.0391
STI	0.0331	-1.1957	-0.2484	-0.8835	-0.0254	0.0388	-1.1797	-0.1532	-0.9375	-0.0657
STOXX50E	0.0498	-2.1253	-0.2740	-0.0250	-0.2731	0.0517	-2.2088	-0.2801	-0.0183	-0.2288
Average	0.0458	-1.7471	-0.4952	-0.1117	-0.2546	0.0533	-1.7563	-0.5044	-0.0965	-0.2506

5.5.4 Parameter estimates

An additional advantage of the Bayesian method is its capability of generating the full posterior distribution for parameter inference. Figures 5.2 and 5.3 present the SMC in-sample likelihood annealing posterior distributions for the coefficients β from the RNN-HAR model on the probability level $\alpha = 1\%$, estimated with the S&P500 dataset. As shown, all the parameters β_0, β_d, β_w and β_m are significant with the 95% credible intervals not including 0. In addition, we also investigate the posterior distributions for all 9 α parameters to identify whether non-linear relationship between the RV input and the future hidden states are significant. It can be seen that all the parameters are significant, except α_2^w in the given example. In particular, the coefficients α_1^d, α_1^w and α_1^m of the daily, weekly and monthly RV input are significant.

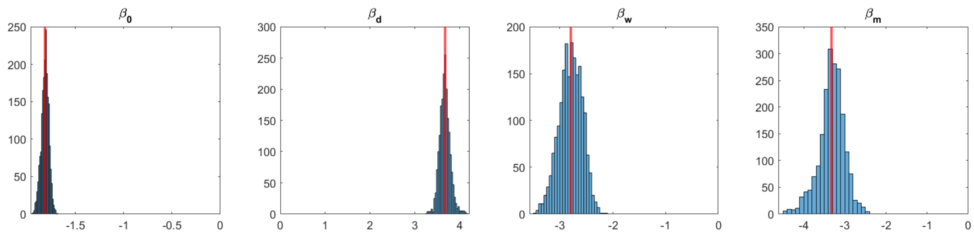

Figure 5.2. RNN-HAR β_0, β_d, β_w and β_m parameters' in-sample likelihood annealing posterior distribution plots with the S&P500 data on the 1% probability level. The vertical lines in red represent the posterior means for each parameter.

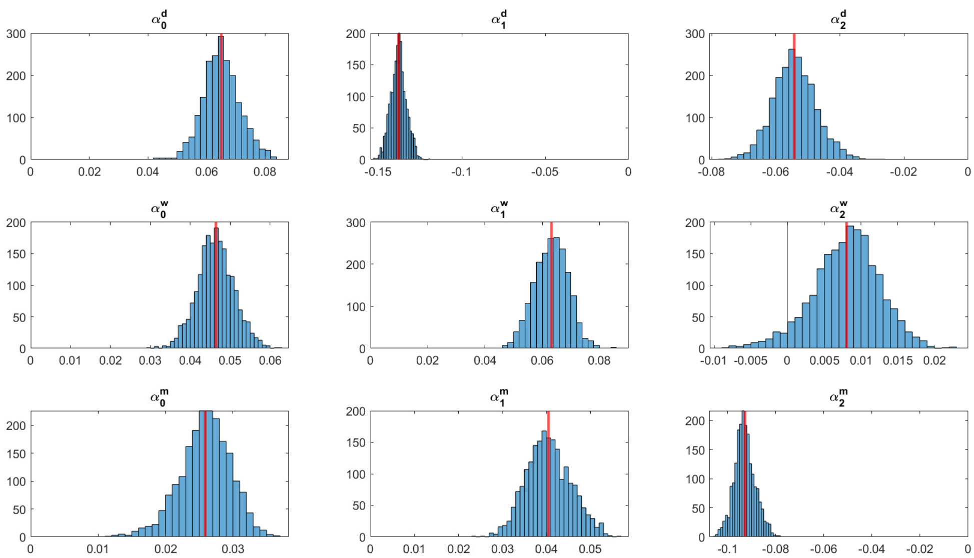

Figure 5.3. RNN-HAR α parameters' SMC in-sample likelihood annealing posterior distribution plots with the S&P500 data on the 1% probability level. The vertical lines in red represent the posterior means for each parameter.

To provide further comparison on the behavior of the models, Table 5.4 presents the parameter estimates averaged across 31 markets using the first in-sample data of each market, for the RNN-HAR estimated with SMC, linear VaR-HAR estimated with SMC and fmincon, HAR, SqrtHAR and Lev-HAR models. Fist, it can be seen that the β_d, β_w and β_m parameter estimates are relatively close across the HAR, SqrtHAR and LevHAR models, as all of these models aim to produce RV forecast. The estimates of the γ_1, γ_2 and γ_3 parameters in the LevHAR are all negative, confirming the existence of the leverage effect: negative returns have higher impact on the RV forecasts. Meanwhile, since the non-linear RNN-HAR and linear VaR-HAR directly produce VaR forecasts, the parameters estimated from these two models behave differently to the ones from the HAR type models.

Table 5.4. Parameter estimates average across 31 markets using the first in-sample data of each market, for the non-linear RNN-HAR estimated with SMC, linear VaR-HAR estimated with SMC and fmincon, HAR, SqrtHAR and Lev-HAR models.

Parameter	RNN-HAR-SMC	VaR-HAR-SMC	VaR-HAR-fmincon	HAR	SqrtHAR	LevHAR
β_0	-2.2139	-1.7471	-1.7563	0.2584	-0.0053	0.0857
β_d	0.2265	-0.4952	-0.5044	0.2768	0.3338	0.1650
β_w	0.3272	-0.1117	-0.0965	0.2633	0.3410	0.2091
β_m	1.2640	-0.2546	-0.2506	0.2496	0.2183	0.2604
α_0^d (γ_1)	0.0160					-0.3882
α_1^d (γ_2)	-0.0118					-0.8270
α_2^d (γ_3)	-0.0175					-0.9346
α_0^w	0.0102					
α_1^w	-0.0106					
α_2^w	-0.0112					
α_0^m	0.0142					
α_1^m	-0.0222					
α_2^m	-0.0037					

5.5.5 Evaluating VaR forecast performance

Following the guidelines outlined in the Basel III Capital Accord (BCBS, 2019), our analysis focuses on daily one-step-ahead VaR forecasts at the probability level of $\alpha = 2.5\%$. Additionally, we also explore probability levels of $\alpha = 1\%$ and $\alpha = 5\%$ for further empirical analysis purposes.

To evaluate the performance of VaR forecasts, we consider several criteria. First, we begin with an intuitive comparison by visualizing the VaR forecasts produced by the RNN-HAR model alongside the violated returns (i.e., instances where returns exceed VaR) and comparing them with those from the HAR model. Second, we assess the violation rate, which measures the proportion of violated returns over the forecasting period. Third, we compare VaR forecasting accuracy by analyzing the quantile scores (5.17). The statistical significance of the quantile score results is assessed using the Model Confidence Set (MCS) of Hansen et al. (2011) and the Diebold–Mariano (DM) test of Diebold and Mariano (1995). Finally, we employ the Dynamic Quantile (DQ) test of Engle and Manganelli (2004) to backtest the VaR forecasts from various models.

To start, the top panel of Figure 5.4 presents the S&P500 out-of-sample returns of from June 2020 to June 2022 and the corresponding 2.5% VaR forecasts produced by the RNN-HAR and HAR models. The middle and bottom panels show the violated returns with the RNN-HAR and HAR. The VaR forecasts from two models have distinctive behaviors, in particular during the more volatile period. For example, as pointed by the arrow in the bottom panel, from Jan 2022 to July 2022 it can seen that the VaR forecasts of the HAR model are more frequently violated by returns than the ones from RNN-HAR, meaning the VaR forecasts produced by the HAR do not provide sufficient loss coverage.

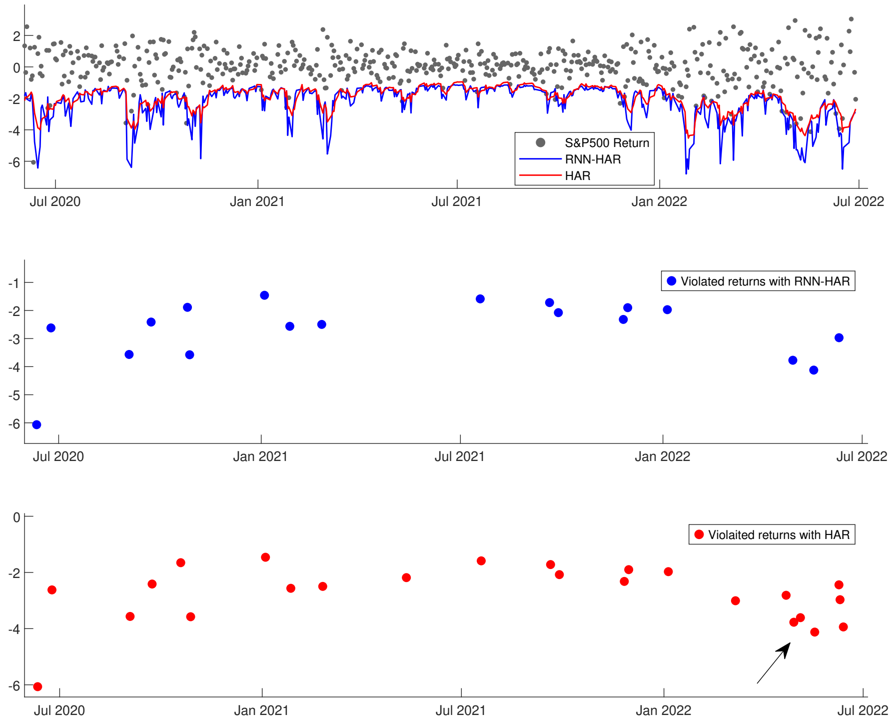

Figure 5.4. The top panel shows the S&P500 out-of-sample returns from June 2020 to June 2022 and the corresponding VaR forecasts from RNN-HAR and HAR models on the 2.5% probability level. The middle and bottom panels show the violated returns (when return exceeds the VaR) for the HAR-RNN and HAR models respectively.

To have a more systematic view on the VaR violations, we now access the violation rate

$$\text{VRate} = \frac{1}{T_{out}} \sum_{t=T_{in}+1}^{T_{in}+T_{out}} I(y_t \leqslant \widehat{\text{VaR}}_t), \tag{5.18}$$

where T_{in} is the in-sample size and T_{out} is the test sample size. Models with $\frac{\text{VRate}}{\alpha}$ close to 1 are preferred. Table 5.5 shows the $\frac{\text{VRate}}{\alpha}$ values for various models across the 31 markets and three different α values. The results indicate that the RNN-HAR model outperforms the other models. Specifically, for $\alpha = 1\%$, the RNN-HAR model achieves the preferred violation rates in 27 out of the 31 markets, with its performance followed by the linear semi-parametric VaR-HAR model. For the HAR-type models, the Normal distribution-based VaR forecasting clearly fails to provide adequate risk coverage at this extreme probability level, as evidenced by the generally high violation rates. On the 2.5% probability level, RNN-HAR is still the most preferred model with favored violation rate for 23 markets, again followed by the linear VaR-HAR model. Comparing to the 1% study, the performance of the HAR type models is generally improved with the $\frac{\text{VRate}}{\alpha}$ values that are much close to 1, as 2.5% is a less extreme probability level. On 5% level which is the least challenging task for the Normal distribution, the performance of the HAR type models is clearly improved with violation rates that are much closer to 5% comparing to the 1% and 2.5% study. Meanwhile, the RNN-HAR model remains favored, achieving preferred violation rates in 23 out of 31 markets.

Overall, when comparing the RNN-HAR and the linear VaR-RNN models—both estimated using quantile loss-based SMC—the superior performance of RNN-HAR underscores the benefit of incorporating non-linearity via RNNs in the VaR forecasting process. In addition, the comparison between the semi-parametric, loss-based linear VaR-HAR and the parametric HAR-type models demonstrates the advantage of adapting the HAR framework for direct risk forecasting using the quantile loss-based SMC method.

Table 5.5. $\frac{\text{VRate}}{\alpha}$ values for the 6 models in comparison across the 31 markets with $\alpha = 1\%, 2.5\%$ and 5%. Blue highlighting indicates the model with the $\frac{\text{VRate}}{\alpha}$ cloest to 1 for each mark and α.

	$\alpha = 1\%$						$\alpha = 2.5\%$						$\alpha = 5\%$					
Market	RNN-HAR	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH	RNN-HAR	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH	RNN-HAR	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH
AEX	1.4	1.4	2.9	3.3	2.5	3.1	1.40	1.52	1.72	1.80	1.68	1.88	1.26	1.34	1.32	1.66	1.26	1.38
AORD	1.7	2.0	3.3	5.6	3.1	3.4	1.48	1.40	1.68	1.68	1.64	1.92	1.10	1.20	1.18	2.02	1.16	1.36
BFX	1.0	1.1	2.3	2.8	2.1	2.7	1.08	1.32	1.68	1.64	1.56	2.04	1.18	1.28	1.24	1.22	1.22	1.36
BSESN	1.8	2.0	2.5	6.4	2.3	2.5	1.36	1.48	1.76	1.80	1.60	1.68	1.16	1.18	1.20	2.10	1.16	1.22
BVLG	0.8	0.6	1.5	1.9	2.9	1.9	0.81	0.76	1.77	1.40	1.72	1.46	0.91	1.06	1.15	1.13	1.19	1.21
BVSP	1.9	1.7	2.1	2.5	1.8	2.1	1.32	1.36	1.52	1.44	1.52	1.36	1.12	1.06	1.16	1.22	1.12	1.08
DJI	2.3	2.4	2.8	5.3	2.5	3.1	1.48	1.36	1.96	1.92	1.64	1.96	1.24	1.40	1.46	1.92	1.44	1.46
FCHI	1.8	2.2	3.4	4.0	3.0	3.6	1.32	1.40	2.00	1.96	1.88	2.08	1.20	1.32	1.44	1.54	1.40	1.44
FTMIB	1.5	1.6	2.9	3.3	2.9	3.0	1.16	1.08	1.64	1.72	1.60	1.72	0.96	0.98	1.18	1.22	1.16	1.16
FTSE	2.5	2.5	3.3	3.6	3.0	3.9	1.48	1.44	1.76	1.76	1.76	2.24	1.18	1.22	1.32	1.60	1.20	1.48
GDAXI	2.2	2.3	4.0	4.6	3.6	3.2	1.64	1.72	2.52	2.56	2.44	2.08	1.12	1.26	1.86	1.74	1.74	1.64
GSPTSE	1.4	1.4	3.3	6.8	3.1	3.4	1.24	1.36	2.00	1.96	2.00	1.96	1.10	1.12	1.38	2.50	1.34	1.40
HIS	1.3	1.6	2.6	2.4	2.6	2.8	1.28	1.36	1.68	1.64	1.72	1.76	1.26	1.26	1.32	1.56	1.26	1.28
IBEX	1.5	1.5	1.9	2.0	2.0	2.3	1.24	1.16	1.16	1.24	1.16	1.48	0.88	0.92	0.86	0.98	0.88	0.96
IXIC	1.4	1.5	2.1	2.4	2.0	2.4	1.24	1.16	1.36	1.40	1.36	1.68	1.12	1.28	1.32	1.44	1.28	1.42
KS11	0.9	0.9	2.2	2.1	1.9	2.3	0.80	1.08	1.84	1.84	1.60	1.68	1.14	1.20	1.54	1.52	1.38	1.44
KSE	1.3	1.7	2.9	2.6	2.9	2.7	1.20	1.28	1.96	2.12	1.92	1.92	1.40	1.40	1.52	1.50	1.50	1.40
MXX	1.5	1.7	1.1	1.5	1.2	2.8	1.24	1.36	1.00	1.04	0.88	1.84	1.36	1.42	0.86	0.90	0.84	1.38
N225	1.0	1.2	2.2	2.8	2.0	2.0	1.00	0.92	1.60	1.72	1.48	1.32	1.06	1.08	1.34	1.34	1.26	1.12
NSEI	2.0	2.3	2.6	8.0	2.5	2.5	1.32	1.40	1.48	1.56	1.48	1.52	1.12	1.16	1.14	2.34	1.12	1.24
OMXC20	1.3	1.1	1.9	2.2	1.9	2.0	1.12	1.16	1.28	1.32	1.28	1.32	1.02	1.02	1.08	1.32	1.06	1.10
OMXHPI	1.5	1.7	3.7	4.0	3.6	1.8	1.32	1.36	2.24	2.32	2.16	1.56	1.14	1.10	1.64	1.72	1.58	1.20
OMXSPI	2.0	1.8	3.8	4.4	3.5	2.5	1.32	1.48	2.28	2.32	2.28	1.92	1.16	1.20	1.60	1.70	1.58	1.36
OSEAX	1.4	1.2	2.2	5.2	2.0	2.6	1.12	1.12	1.24	1.16	1.36	1.48	1.12	1.22	1.04	1.82	1.08	1.08
RUT	1.3	1.4	1.6	2.1	1.2	2.5	1.24	1.32	1.16	1.20	1.00	1.76	1.02	1.20	0.94	1.10	0.86	1.46
SMSI	1.3	1.4	1.7	2.2	1.7	2.7	1.00	1.12	1.32	1.32	1.08	1.52	0.98	0.96	0.92	1.18	0.96	1.02
S&P500	2.1	1.9	3.2	4.8	2.7	3.5	1.36	1.64	1.96	2.08	1.80	2.00	1.36	1.46	1.40	2.06	1.32	1.48
SSEC	0.8	1.0	2.3	4.3	2.1	2.4	1.00	1.12	1.40	1.40	1.32	1.48	1.12	1.16	1.16	1.88	1.04	1.16
SSMI	1.3	1.8	2.6	2.8	2.0	3.3	1.04	1.12	1.60	1.68	1.44	1.56	1.14	1.12	1.24	1.26	1.20	1.20
STI	1.2	1.2	1.6	1.7	1.5	2.2	1.12	1.32	1.12	1.12	1.16	1.36	1.28	1.40	0.90	0.90	0.94	1.12
STOXX50E	1.8	2.0	3.2	3.9	2.8	3.2	1.44	1.44	1.92	1.92	1.80	1.92	1.16	1.26	1.32	1.42	1.26	1.34

Next, we employ the strictly consistent quantile score from (5.17) to formally evaluate the VaR forecasting accuracy of the considered models. To provide further intuition on the quantile score function, we rewrite (5.17) as follows:

$$\mathrm{QS}_t = \begin{cases} (y_t - \widehat{\mathrm{VaR}}_t)(\alpha - 1), & \text{if } y_t \leqslant \widehat{\mathrm{VaR}}_t, \\ (y_t - \widehat{\mathrm{VaR}}_t)\alpha, & \text{otherwise.} \end{cases} \tag{5.19}$$

With $\alpha = 1\%$, for instance, the quantile score penalizes under-estimated VaR forecasts (when the return violates the VaR) 99 times more than over-estimated (conservative) VaR forecasts. To statistically test whether the quantile loss differences between different models are significant, the MCS, introduced by Hansen et al. (2011), is employed. MCS produces a set of models that contain the "superior" forecasting models, given a level of confidence. The MCS is used to evaluate the statistical significance for quantile scores under the 75% confidence level. We adopt the Matlab code downloaded from Kevin Sheppard's web page (Sheppard, 2009). In addition, we use the one-sided DM test (Diebold and Mariano, 2002) to compare the proposed RNN-HAR with each of the other models in comparison and check whether the RNN-HAR significantly outperforms based on the quantile score.

On the considered 1%, 2.5% and 5% probability levels, the quantile scores, MCS and DM testing results are presented in Tables 5.6, 5.7 and 5.8 respectively. In general, across all three probability levels the favored model is the proposed RNN-HAR model which most frequently generates the smallest quantile scores. The RNN-HAR is the only model that is included in the MCS for all 31 markets on 1% and 2.5% probability levels, and is included in the MCS for 29 markets on the 5% probability level. These findings underscore the effectiveness of the proposed model across varying probability levels. The performance of the RNN-HAR is followed by the linear VaR-HAR model. The VaR-HAR in general more frequently produces the preferred quantile scores and is more likely to be included in the MCS than the other HAR type models, especially for the most extreme 1% probability level. The DM testing results also show that the RNN-HAR statistically outperforms the semi-parametric VaR-HAR model estimated with SMC for 11, 14 and 19 times, for the 1%, 2.5% and 5% probability levels respectively. Comparing to HAR type models, the RNN-HAR significantly outperforms the HAR type models for at least 10 out of the 31 markets for the 1% and 2.5% studies. On the 5% probability level which is a less challenging task for the parametric methods, comparing to the HAR type models the RNN-HAR still produces significantly improved loss results for 5 to 16 markets, and the SqrtHAR is least likely to be outperformed by the RNN-HAR.

These observations are generally consistent with those from the violation rate study. For the non-linear RNN-HAR and the linear VaR-RNN models—both estimated using the quantile loss-based SMC—the superior performance of RNN-HAR highlights the importance of capturing the non-linear relationship between the daily, weekly, and monthly RV inputs and

VaR forecasts. The statistically significant parameter estimates in the RNN equations (5.7b) to (5.7d), as discussed in Section 5.5.4, further confirm the presence of such a non-linear relationship. Additionally, regarding the effectiveness of using the quantile loss function to adapt conventional RNNs for semi-parametric VaR forecasting, the generally superior performance of the linear VaR-HAR model compared to parametric HAR-type models provides supporting evidence.

Table 5.6. The left panel shows the $\alpha = 1\%$ quantile scores and the 75% MCS results across 31 markets. The model with the lowest quantile score is highlighted in blue, and models in the MCS are highlighted in grey shading. The right panel shows the one-sided DM test p-values via comparing the proposed RNN-HAR with other competing models. Grey shading indicates the RNN-HAR significantly outperforms the competing model on the 10% significance level.

	Quantile scores and 75% MCS						DM test comparing RNN-HAR vs				
Market	RNN-HAR	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH
AEX	0.0397	0.0405	0.0439	0.0504	0.0431	0.0523	0.33	0.06	0	0.10	0.02
AORD	0.0380	0.0414	0.0426	0.0516	0.0420	0.0519	0	0.01	0	0.03	0.01
BFX	0.0452	0.0421	0.0482	0.0511	0.0480	0.0551	0.74	0.14	0.04	0.18	0.04
BSESN	0.0460	0.0475	0.0505	0.0711	0.0492	0.0484	0.24	0.02	0	0.05	0.09
BVLG	0.0404	0.0383	0.0422	0.0436	0.0663	0.0432	0.68	0.01	0.01	0.01	0.22
BVSP	0.0682	0.0560	0.0644	0.0640	0.0631	0.0634	0.91	0.75	0.76	0.81	0.80
DJI	0.0436	0.0415	0.0508	0.0565	0.0490	0.0515	0.75	0.02	0	0.03	0
FCHI	0.0477	0.0518	0.0552	0.0578	0.0545	0.0574	0.02	0.01	0	0.01	0.01
FTMIB	0.0643	0.0643	0.0667	0.0665	0.0664	0.0678	0.50	0.11	0.26	0.10	0.04
FTSE	0.0440	0.0447	0.0501	0.0517	0.0486	0.0593	0.28	0	0.01	0	0.01
GDAXI	0.0528	0.0530	0.0649	0.0641	0.0634	0.0600	0.45	0	0	0	0
GSPTSE	0.0426	0.0439	0.0507	0.0569	0.0483	0.0563	0.06	0	0	0.01	0
HIS	0.0426	0.0450	0.0471	0.0427	0.0462	0.0481	0	0.03	0.47	0.04	0.02
IBEX	0.0517	0.0488	0.0506	0.0499	0.0500	0.0569	0.88	0.66	0.74	0.75	0.06
IXIC	0.0464	0.0491	0.0484	0.0504	0.0472	0.0484	0.03	0.22	0.11	0.38	0.25
KS11	0.0367	0.0372	0.0398	0.0360	0.0382	0.0393	0.36	0.14	0.61	0.28	0.20
KSE	0.0399	0.0440	0.0489	0.0464	0.0488	0.0460	0.04	0	0.03	0	0.01
MXX	0.0427	0.0424	0.0438	0.0392	0.0423	0.0465	0.62	0.32	0.89	0.62	0.07
N225	0.0391	0.0401	0.0412	0.0439	0.0405	0.0446	0.12	0.18	0.05	0.24	0.05
NSEI	0.0485	0.0471	0.0509	0.0862	0.0500	0.0496	0.65	0.14	0	0.20	0.37
OMXC20	0.0422	0.0418	0.0440	0.0432	0.0426	0.0435	0.60	0.12	0.32	0.38	0.21
OMXHPI	0.0454	0.0481	0.0532	0.0529	0.0516	0.0460	0.02	0	0.01	0.01	0.32
OMXSPI	0.0451	0.0469	0.0551	0.0564	0.0532	0.0483	0.08	0	0	0	0.01
OSEAX	0.0421	0.0430	0.0437	0.0735	0.0418	0.0565	0.37	0.29	0	0.54	0.01
RUT	0.0559	0.0562	0.0567	0.0574	0.0541	0.0717	0.48	0.45	0.38	0.71	0.01
SMSI	0.0497	0.0507	0.0512	0.0496	0.0503	0.0541	0.05	0.12	0.53	0.30	0.03
SPX	0.0449	0.0433	0.0487	0.0545	0.0477	0.0486	0.72	0.16	0.01	0.20	0.15
SSEC	0.0474	0.0495	0.0500	0.0595	0.0490	0.0498	0.02	0.17	0.01	0.26	0.18
SSMI	0.0355	0.0399	0.0368	0.0361	0.0355	0.0496	0.07	0.33	0.42	0.51	0.03
STI	0.0332	0.0331	0.0355	0.0350	0.0357	0.0369	0.56	0.14	0.22	0.15	0.07
STOXX50E	0.0484	0.0498	0.0547	0.0590	0.0528	0.0660	0.28	0	0	0.02	0

Table 5.7. The left panel shows the $\alpha = 2.5\%$ quantile scores and the 75% MCS results for the 6 models in comparison across 31 markets. The model with the lowest quantile score is highlighted in blue, and models in the MCS are highlighted in grey shading. The right panel shows the one-sided DM test p-values via comparing the proposed RNN-HAR with other competing models. Grey shading indicates the proposed RNN-HAR significantly outperforms the competing model on the 10% significant level.

	Quantile scores and 75% MCS						DM test comparing RNN-HAR vs				
Market	RNN-HAR	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH
AEX	0.0840	0.0869	0.0882	0.0887	0.0867	0.0947	0.24	0.02	0.01	0.07	0.01
AORD	0.0756	0.0813	0.0790	0.0792	0.0782	0.0868	0	0.03	0.03	0.10	0.03
BFX	0.0873	0.0903	0.0925	0.0924	0.0913	0.0982	0.27	0.02	0.02	0.06	0.01
BSESN	0.0871	0.0890	0.0900	0.0903	0.0880	0.0880	0.30	0.13	0.11	0.35	0.33
BVLG	0.0826	0.0842	0.1079	0.0831	0.1067	0.0839	0.34	0.01	0.02	0.02	0.37
BVSP	0.1169	0.1103	0.1167	0.1165	0.1175	0.1172	0.82	0.51	0.54	0.40	0.46
DJI	0.0834	0.0835	0.0912	0.0919	0.0887	0.0920	0.49	0.03	0.02	0.07	0.02
FCHI	0.0953	0.0989	0.1013	0.1014	0.0996	0.1029	0.07	0	0	0.01	0
FTMIB	0.1145	0.1148	0.1145	0.1144	0.1142	0.1158	0.36	0.51	0.52	0.55	0.31
FTSE	0.0861	0.0856	0.0900	0.0912	0.0886	0.0987	0.56	0.03	0.02	0.12	0.02
GDAXI	0.0990	0.1025	0.1101	0.1101	0.1082	0.1051	0	0	0	0	0
GSPTSE	0.0774	0.0772	0.0803	0.0800	0.0789	0.0872	0.54	0.14	0.16	0.24	0.02
HIS	0.0919	0.0946	0.0934	0.0937	0.0926	0.0948	0	0.14	0.10	0.26	0.05
IBEX	0.0961	0.0946	0.0938	0.0941	0.0939	0.1015	0.70	0.78	0.74	0.84	0.05
IXIC	0.0993	0.1029	0.0996	0.0996	0.0985	0.0993	0.18	0.47	0.46	0.61	0.49
KS11	0.0741	0.0750	0.0780	0.0779	0.0758	0.0777	0.37	0.08	0.08	0.23	0.11
KSE	0.0836	0.0865	0.0901	0.0901	0.0897	0.0872	0	0.02	0.02	0.03	0.05
MXX	0.0809	0.0826	0.0806	0.0803	0.0795	0.0827	0.04	0.59	0.68	0.86	0.13
N225	0.0810	0.0827	0.0836	0.0838	0.0827	0.0877	0.05	0.11	0.10	0.18	0.02
NSEI	0.0882	0.0903	0.0912	0.0914	0.0896	0.0892	0.30	0.12	0.12	0.30	0.32
OMXC20	0.0834	0.0847	0.0840	0.0842	0.0829	0.0838	0.03	0.39	0.35	0.66	0.42
OMXHPI	0.0848	0.0874	0.0944	0.0943	0.0926	0.0870	0.03	0	0	0.01	0.07
OMXSPI	0.0885	0.0929	0.0986	0.0981	0.0960	0.0921	0	0	0	0.01	0.02
OSEAX	0.0845	0.0917	0.0856	0.0851	0.0843	0.0989	0	0.28	0.37	0.56	0.02
RUT	0.1095	0.1085	0.1117	0.1118	0.1085	0.1244	0.58	0.36	0.35	0.72	0.03
SMSI	0.0911	0.0925	0.0936	0.0929	0.0913	0.0976	0.08	0.07	0.13	0.44	0.02
SPX	0.0867	0.0894	0.0923	0.0924	0.0898	0.0925	0.31	0.10	0.10	0.19	0.08
SSEC	0.0887	0.0910	0.0889	0.0903	0.0880	0.0890	0.09	0.45	0.21	0.71	0.41
SSMI	0.0705	0.0756	0.0730	0.0747	0.0705	0.0837	0	0.18	0.07	0.52	0.05
STI	0.0643	0.0631	0.0650	0.0649	0.0651	0.0657	0.78	0.37	0.38	0.36	0.27
STOXX50E	0.0954	0.0973	0.1019	0.1023	0.0987	0.1107	0.13	0.01	0.01	0.06	0.01

Table 5.8. The left panel shows the $\alpha = 5\%$ quantile scores and the 75% MCS results across 31 markets. The model with the lowest quantile score is highlighted in blue, and models in the MCS are highlighted in grey shading. The right panel shows the two sided DM testing p-values via comparing the proposed RNN-HAR with other models. Grey shading indicates the p-values that are less than 10%.

	Quantile scores and 75% MCS						DM test comparing RNN-HAR vs				
Market	RNN-HAR	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH	VaR-HAR	HAR	LevHAR	SqrtHAR	RHARGARCH
AEX	0.1408	0.1502	0.1448	0.1473	0.1431	0.1499	0.01	0.10	0.05	0.22	0.04
AORD	0.1254	0.1329	0.1243	0.1336	0.1232	0.1314	0	0.74	0.02	0.90	0.13
BFX	0.1468	0.1535	0.1503	0.1504	0.1486	0.1554	0.02	0.12	0.16	0.29	0.05
BSESN	0.1424	0.1435	0.1415	0.1563	0.1391	0.1397	0.42	0.66	0.01	0.94	0.93
BVLG	0.1377	0.1436	0.1386	0.1375	0.1600	0.1395	0.02	0.04	0.02	0.01	0.29
BVSP	0.1924	0.1869	0.1876	0.1901	0.1892	0.1893	0.74	0.82	0.69	0.85	0.83
DJI	0.1389	0.1418	0.1461	0.1504	0.1420	0.1446	0.35	0.02	0.01	0.14	0.08
FCHI	0.1544	0.1620	0.1610	0.1616	0.1587	0.1617	0.01	0.01	0.02	0.04	0.01
FTMIB	0.1746	0.1773	0.1757	0.1753	0.1753	0.1770	0	0.25	0.41	0.33	0.12
FTSE	0.1383	0.1395	0.1396	0.1407	0.1384	0.1471	0.35	0.27	0.19	0.49	0.06
GDAXI	0.1601	0.1632	0.1679	0.1668	0.1660	0.1650	0.01	0	0.03	0.01	0.02
GSPTSE	0.1134	0.1178	0.1195	0.1270	0.1188	0.1266	0	0.02	0	0.05	0.02
HIS	0.1550	0.1572	0.1563	0.1552	0.1556	0.1568	0	0.23	0.47	0.36	0.19
IBEX	0.1509	0.1519	0.1512	0.1509	0.1506	0.1564	0.39	0.46	0.51	0.56	0.08
IXIC	0.1675	0.175	0.1695	0.1721	0.1669	0.1684	0.07	0.28	0.12	0.57	0.41
KS11	0.1274	0.1313	0.1292	0.1251	0.1266	0.1283	0	0.20	0.83	0.69	0.35
KSE	0.1422	0.1442	0.1445	0.1422	0.1438	0.1419	0.16	0.17	0.50	0.29	0.55
MXX	0.1324	0.1333	0.1297	0.1276	0.1292	0.1307	0.14	0.88	0.91	0.88	0.92
N225	0.1398	0.1420	0.1415	0.1434	0.1404	0.1458	0.14	0.20	0.09	0.38	0.04
NSEI	0.1413	0.1443	0.1430	0.1685	0.1405	0.1407	0.27	0.25	0	0.63	0.66
OMXC20	0.1388	0.1390	0.1376	0.1353	0.1367	0.1367	0.47	0.71	0.93	0.92	0.85
OMXHPI	0.1373	0.1449	0.1467	0.1451	0.1447	0.1417	0	0.01	0.02	0.02	0.01
OMXSPI	0.1475	0.1536	0.1537	0.1501	0.1504	0.1493	0	0.02	0.24	0.16	0.17
OSEAX	0.1423	0.1508	0.1406	0.1619	0.1413	0.1534	0	0.74	0	0.72	0.05
RUT	0.1818	0.1805	0.1842	0.1866	0.1817	0.1936	0.58	0.34	0.18	0.51	0.08
SMSI	0.1466	0.1494	0.1501	0.1488	0.1468	0.1521	0.06	0.03	0.19	0.46	0.03
SPX	0.1432	0.1495	0.1472	0.1510	0.1446	0.1482	0.07	0.11	0.03	0.31	0.07
SSEC	0.1403	0.1439	0.1392	0.1527	0.1389	0.1399	0.05	0.89	0	0.95	0.65
SSMI	0.1156	0.1255	0.1204	0.1175	0.1179	0.1273	0	0.05	0.30	0.14	0.06
STI	0.1045	0.1055	0.1037	0.1033	0.1046	0.1043	0.20	0.74	0.78	0.46	0.56
STOXX50E	0.1537	0.1607	0.1602	0.1613	0.1562	0.1692	0	0.02	0.02	0.16	0.01

Lastly, we use the DQ test (Engle and Manganelli, 2004) to backtest and validate the VaR forecasts. As pointed out by Engle and Manganelli (2004), the out-of-sample DQ test does not depend on the estimation procedure, but only the sequences of the VaR forecasts and the corresponding return values are needed. In this test, a series of "Hits" are calculated by $H_t = I(y_t < \text{VaR}_t) - \alpha$ for the null hypothesis and iid series with rate α. When the null hypothesis is true, it can be shown that $\mathbb{E}(H_t) = 0$ and $\mathbb{E}(H_t W_{it}) = 0$ (Gerlach et al., 2016) where W has q explanatory variables that are in the information set at time $t-1$ when the forecast VaR_t is made. To check whether all parameters in a regression of H on W equal zero, the following DQ test statistic has been derived as:

$$DQ(q) = \frac{H'W\left(W'W\right)^{-1}W'H}{\alpha(1-\alpha)} \sim \chi^2_q, \tag{5.20}$$

Following Engle and Manganelli (2004) and Gerlach et al. (2016), we employ DQ test with lags from 1 to 4, respectively, as follows:

$$\begin{aligned}
W_t^T &= (1, H_{t-1}, \text{VaR}_t) \quad \text{denoted as } DQ_1,\\
W_t^T &= (1, H_{t-1}, H_{t-2}, \text{VaR}_t) \quad \text{denoted as } DQ_2,\\
W_t^T &= (1, H_{t-1}, H_{t-2}, H_{t-3}, \text{VaR}_t) \quad \text{denoted as } DQ_3,\\
W_t^T &= (1, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-4}, \text{VaR}_t) \quad \text{denoted as } DQ_4.
\end{aligned}$$

Table 5.9 shows the total number of markets for which each model is rejected by the DQ tests, where the significance level is selected as the corresponding α. Less rejections are better. The proposed RNN-HAR model has the lowest number of rejections out of 31 markets considered for each probability level. Again, its performance is followed by the semi-parametric VaR-HAR which in general outperforms the conventional HAR models.

Table 5.9. Summary of the DQ backtesting results across 31 markets, with $\alpha = 1\%$, $\alpha = 2.5\%$ and $\alpha = 5\%$. Blue highlighting indicates the model with the least number of rejections for each test on each probability level.

	$\alpha = 1\%$				$\alpha = 2.5\%$				$\alpha = 5\%$			
	DQ1	DQ2	DQ3	DQ4	DQ1	DQ2	DQ3	DQ4	DQ1	DQ2	DQ3	DQ4
RNN-HAR	8	8	14	14	7	5	8	7	4	3	4	4
VaR-HAR	10	10	14	12	7	6	8	8	12	12	13	11
HAR	25	25	26	25	23	27	25	24	16	15	16	16
LevHAR	28	28	29	29	21	26	22	21	19	20	20	20
SqrtHAR	21	20	22	23	21	21	20	22	14	15	15	15
RHARGARCH	28	27	28	28	25	23	25	26	21	20	20	19

In addition to the primary evaluation techniques discussed in this chapter, the MCS test, as introduced in Chapter 2, was applied to assess the relative performance of the proposed models. However, the results of the MCS test were consistent across all models and markets, showing no significant differences in predictive performance. As such, the MCS results are not presented here, but the application of the test supports the robustness of the models in this study.

5.5.6 Computational efficiency

In this study, all models, including RNN-HAR, VaR-HAR, HAR, LevHAR, SqrtHAR, and RHARGARCH, were run for all 31 markets included in the analysis. The training times for each model varied depending on their complexity. While the traditional models (VaR-HAR, HAR, LevHAR, SqrtHAR, and RHARGARCH) typically took 10-15 minutes to run per market, the RNN-HAR model, owing to the inclusion of the RNN component, required approximately 40-45 minutes per market.

The computational demands for the models reflect the complexity of the algorithms. The RNN-HAR model, due to its more intricate structure, naturally requires more time to process. This increase in computational time is consistent with the trade-off between the added complexity of capturing non-linear dependencies and the resulting performance improvements.

All computations were performed on a personal laptop with 12 GB of RAM and a single-core processor. While the models were run sequentially, the use of relatively moderate computational resources did not prevent the models from being run for all 31 markets in the study. The RNN-HAR model's increased computational time is justified by its superior ability to capture complex, non-linear patterns in the data, leading to enhanced forecasting performance.

In conclusion, despite the longer computational time for the RNN-HAR model, the trade-off is justified by its improved performance. The model's ability to capture intricate temporal dependencies makes it a valuable forecasting tool, offering substantial benefits over traditional models, even though it requires more computational resources.

5.6 Chapter summary

This chapter extends the HAR model by integrating RNN structures to capture daily, weekly, and monthly non-linear and long-term effects of realized variances on VaR. The proposed RNN-HAR model directly estimates VaR and leverages quantile scores, eliminating the need for distributional assumptions. This approach incorporates return series alongside realized measures, reducing potential inaccuracies associated with relying solely on the latter. We use SMC with likelihood annealing for in-sample analysis and SMC with data annealing for out-of-sample sequential forecasting. Our extensive empirical study covers 31 major stock markets, demonstrating that the proposed RNN-HAR model outperforms other HAR extensions across all three probability levels we considered.

The chapter focuses mainly on one-step-ahead VaR forecast. Our approach can be extended to multi-step-ahead forecasts. For a given horizon h, equation (5.7a) can be modified as

$$\mathrm{VaR}_{t+h} = \beta_0 + \beta_d h_t^d + \beta_w h_t^w + \beta_m h_t^m.$$

After training, this model can be used for h-step-ahead VaR forecast together with its associated uncertainty.

Another promising avenue for future research involves further enriching the model with LSTM structures and incorporating multiple realized measures. These advancements hold potential for achieving even higher levels of predictive accuracy and robustness in financial risk forecasting.

Chapter 6

RNN-ReESCAViaR model

In this chapter a non-linear model framework is proposed for joint VaR and ES forecasting. The framework, referred to as RNN-ReESCAViaR, extends the Realized-ES-CAViaR Model, a framework where the realized measure influences VaR and ES separately by integrating an RNN to handle non-linear dynamics. The link between the quantile score function and the AL density has established a flexible, likelihood-based framework for the joint modelling of VaR and ES. Recognizing the importance of accurately capturing the dynamic interdependence between these risk measures in financial applications, our proposed framework integrates asymmetric Laplace quasi-likelihood with RNN-based time series techniques. Bayesian inference is conducted using the SMC algorithm, enhancing model estimation and forecasting performance. Daily closing prices and realized variance from 2000 to 2022 across 21 markets is selected for the empirical study. One-step-ahead VaR and ES forecasting performances is compared against four semi-parametric and parametric models. Compared to all other models, our proposed RNN-ReESCAViaR model demonstrates better results.

6.1 Introduction

Recent global events highlight the need for strong risk management practices. For example, the COVID-19 pandemic caused major disruptions and volatility in financial markets. Many institutions faced significant losses because they could not accurately assess risks during this crisis. This situation revealed the shortcomings of traditional risk measures that often fail to account for extreme events. As a result, many firms have re-evaluated their risk management strategies to better handle sudden market shocks (Baker et al., 2020). Moreover, ongoing geopolitical tensions, such as the conflict between Russia and Ukraine, have shown how interconnected global markets can be. These events have led to dramatic changes in commodity prices and financial stability, reinforcing the need for adaptable risk management practices. Companies now realise that a proactive approach protects assets and helps them seize opportunities in uncertain times.

In light of these experiences, there is a strong need for risk management practices that are both robust and flexible enough to adapt to rapid changes. VaR and ES are key measures to assess financial risk. VaR is one of the most commonly used measures for evaluating tail risk but has significant limitations. VaR estimates the potential loss at a specified confidence level, providing a quantile of the profit and loss distribution. Its simplicity has made it a popular choice for measuring market risk. However, VaR cannot capture the severity of losses that occur beyond the threshold, failing to consider the shape and structure of the tail of the return distribution (Artzner et al., 1999). This lack of insight means that VaR can lead to an incomplete understanding of risk exposure. Moreover, VaR is not a coherent risk measure, as it does not satisfy the property of sub-additivity. This implies that the risk of a combined portfolio could be assessed as lower than the risk of its individual components, which can promote risky behaviour through non-diversification (Acerbi and Tasche, 2002). These shortcomings highlight the necessity of incorporating additional risk measures. In contrast, ES has gained popularity for its favourable properties. ES measures the average loss expected in scenarios where losses exceed the VaR threshold, thus providing a more comprehensive assessment of tail risk (Roccioletti, 2015). ES accounts for the magnitude of extreme losses as a coherent risk measure, making it a vital complement to VaR.

The Basel Committee on Banking Supervision (2023) has advocated for a transition from relying solely on VaR at the 1% level to adopting ES at the 2.5% level, emphasizing the importance of a more robust risk assessment framework. However, it is crucial to note that ES inherently depends on the corresponding VaR estimate; thus, it cannot be elicited independently. The practical evaluation of risk necessitates considering the joint (VaR, ES) pair, as highlighted by Fissler and Ziegel (2016). This combined approach allows for a more thorough understanding of tail risks, aligning with current regulatory recommendations and enhancing overall risk management strategies.

Traditional parametric models, particularly GARCH models, have been extensively used to estimate VaR and ES in financial risk management. GARCH models, introduced by Bollerslev (1986), allow for the modelling of time-varying volatility, which is crucial for capturing the dynamic nature of financial returns. However, these models typically operate under specific distributional assumptions, often assuming that returns follow a normal or t-distribution. This reliance on parametric assumptions can lead to significant inaccuracies in risk estimation, particularly during periods of market stress when the true distribution may exhibit heavy tails or skewness (Engle, 2001). Extensions of the GARCH framework, such as EGARCH (Nelson, 1991) and GJR-GARCH (Glosten et al., 1993), aim to address some of these limitations by allowing for asymmetric effects in volatility. Despite these advancements, the inherent assumptions of these models can still lead to misestimation during extreme market events, necessitating the exploration of more flexible modelling frameworks.

Semi-parametric time series quantile models, such as the Conditional Autoregressive Value at Risk (CAViaR) framework introduced by Engle and Manganelli (2004), focus on directly estimating the quantile (VaR) by minimizing a quantile loss function, akin to quantile regression techniques. However, it is important to note that these models do not explicitly estimate ES. The challenge in estimating ES arises from its non-elicitable nature; as Gneiting (2011) discusses, no unique loss function exists that can be minimized by ES in expectation. To address this limitation, Fissler and Ziegel (2016) propose a family of joint loss functions that are strictly consistent for concurrently estimating the true VaR and ES series. These loss functions are uniquely minimized in expectation by the true VaR and ES, facilitating the development and estimation of joint models for both risk measures.

Building upon these advancements, Taylor (2019) introduces a joint semi-parametric model for VaR and ES, referred to as ES-CAViaR. This model employs an asymmetric Laplace (AL) distribution with a time-varying scale, which allows for constructing a quasi-likelihood function. Taylor (2019) demonstrates that this quasi-likelihood is directly related to the class of strictly consistent loss functions identified by Fissler and Ziegel (2016), thereby enabling the joint estimation of conditional VaR and ES. This approach enhances the modelling of tail risks and provides a more comprehensive framework for assessing financial risk.

Integrating realized measures in VaR and ES forecasting marks a significant advancement in risk modelling. Realized measures, such as realized volatility, provide valuable insights into the underlying dynamics of asset returns. Parametric models that incorporate these measures, like RGARCH Hansen et al. (2012) and REGARCH Hansen and Huang (2016), enhance traditional GARCH specifications by utilizing realized volatility data, thereby improving forecasting accuracy. In a semi-parametric context, Gerlach and Wang (2020) extend the ES-CAViaR framework (Taylor, 2019) by incorporating realized measures as exogenous variables, demonstrating enhanced forecasting accuracy for both VaR and ES. Building on this foundation, Wang et al. (2023) introduce the Realized-ES-CAViaR framework, suggesting that it can be further extended into other nonlinear models as proposed by Gerlach et al. (2011). While this model (Gerlach et al., 2011) improves flexibility for dynamic quantile estimation, it fails to fully leverage the advanced techniques' capacity to capture non-linear dynamics and temporal dependencies present in financial data.

The involvement of machine learning techniques in financial forecasting, particularly for VaR and ES estimation, has gained significant attention. As highlighted by Sako et al. (2022), several machine learning models have been proposed for financial forecasting. Artificial neural networks (ANNs), also called neural networks (NN), constitute a subset of machine learning, more precisely deep learning (DL), which is based on numerical algorithms that simulate the behaviour of biological systems composed of neurons.

Xu et al. (2016) introduce a nonparametric Conditional Autoregressive Expectile (NCARE) model that utilizes Artificial Neural Networks (ANNs) to offer flexibility in capturing nonlinear relationships in risk estimation. While this approach extends the CARE model by Taylor (2008a) avoiding predetermined functional forms, it does not fully address the sequential nature of financial data. Compared to ANN, RNN has shown superior results due to the inherently sequential nature of time series data. RNNs are designed to capture temporal dependencies by feeding the output of a previous time step as an input to the current step, enabling them to remember past information. This ability to model time-dependent patterns makes RNNs more suitable for financial forecasting tasks, where capturing dynamic relationships is crucial (Sako et al., 2022). Further, RNNs are particularly well-suited for financial time series forecasting because they can process sequential data and retain the memory of recent events through interconnected hidden neurons (Wang et al., 2022). This capacity to learn from past data makes RNNs effective in modelling the autocorrelation characteristics of stock prices, which often do not adhere to random walk assumptions. Therefore, our proposed model integrates RNNs to forecast VaR and ES, offering a more comprehensive and accurate framework for risk forecasting.

Despite the advancements in estimating VaR and ES, a significant gap remains in the joint estimation of these two measures using advanced techniques. While scholars have explored the combination of quantile regression with machine learning methods such as quantile regression neural networks (QRNN) and LASSO-QR—these models often struggle to capture long memory effects and nonlinear dependencies, leading to inaccuracies, particularly at higher confidence levels (Wang et al., 2024). Moreover, existing machine learning approaches have not effectively addressed the challenge of severe VaR violations, where actual losses exceed the predicted VaR values.

To address these challenges, our proposed RNN-ReESCAViaR framework aims to fill the gap by integrating RNNs for the joint estimation of VaR and ES, where the realized measure influences both quantities separately. Our framework retains a semi-parametric framework, thus eliminating reliance on specific distributional assumptions regarding returns. By effectively capturing the non-linear dependencies in the joint dynamics of VaR and ES, our approach provides a more robust and flexible method for risk assessment. Leveraging Bayesian SMC for parameter estimation, the RNN-ReESCAViaR framework enhances predictive accuracy in volatile market conditions. This research advances the joint estimation of VaR and ES while contributing to the broader literature on financial risk assessment, highlighting the effectiveness of non-linear models.

The remainder of this paper is organized as follows: Section 6.2 details the proposed model, section 6.3 describes the methodology, Section 6.4 presents the empirical results, and Section 6.5 summarizes the findings and their implications.

6.2 Proposed models

The proposed RNN-ReESCAViaR framework is an extension of the Realized-ES-CAViaR model (Wang et al., 2023), which focuses on the joint estimation of VaR and ES using realized measure. By incorporating a RNN, the proposed model further enables effective capture of the non-linear relationships present in financial data, providing a comprehensive approach to sequential forecasting. Wang et al. (2023) introduce the additive version of the Realized-ES-CAViaR model as follows:

Realized-ES-CAViaR

$$Q_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 Q_{t-1}, \tag{6.1}$$

$$\omega_t = \gamma_0 + \gamma_1 x_{t-1} + \gamma_2 \omega_{t-1}, \tag{6.2}$$

$$\mathrm{ES}_t = Q_t - \omega_t, \tag{6.3}$$

$$x_t = \xi + \phi|\mathrm{ES}_t| + \tau_1 \epsilon_t + \tau_2 \left(\epsilon_t^2 - E(\epsilon^2)\right) + u_t, \tag{6.4}$$

where $\gamma_0 \geqslant 0, \gamma_1 \geqslant 0, \gamma_2 \geqslant 0$, to ensure that the VaR and ES estimates do not cross. Here the equation (6.1) is the quantile equation, equation (6.2) is the ω_t equation, equation (6.3) is the ES equation, equation (6.4) is the measurement equation where x_t denotes the realized measures. Here we use realized variance (RV) as the realized measure.

In our proposed framework, we extend the Realized-ES-CAViaR model by using RNN. To identify the most effective structure for our proposed RNN-ReESCAViaR framework, we initially tested nine variations of the model, each differing in the parameters influenced by the latent state h_t and the specific inputs fed into the RNN. These variations included models where either one, two, or all three key parameters were driven by h_t, and different combinations of $x_{t-1}, Q_{t-1}, ES_{t-1}$ were used as RNN inputs. After evaluating each version on a subset of six diverse markets, we selected two final versions, RNN-ReESCAViaR-1h, with only one parameter dependent on h_t, and RNN-ReESCAViaR-2h, with two parameters dependent on h_t with all three RNN inputs $x_{t-1}, Q_{t-1}, ES_{t-1}$ in the models. These models were chosen based on their forecasting accuracy, stability, and computational efficiency, providing a balance of complexity and performance suited to our empirical analysis.

Proposed RNN-ReESCAViaR-1h model:

$$
\begin{aligned}
Q_t &= \beta_t + \beta_1 x_{t-1} + \beta_2 Q_{t-1} && (6.5)\\
\omega_t &= \nu_0 + \nu_1 x_{t-1} + \nu_2 \omega_{t-1}, && (6.6)\\
\mathrm{ES}_t &= Q_t - \omega_t && (6.7)\\
x_t &= \xi_0 + \phi_1(|\mathrm{ES}_t|) + \delta_1 \epsilon_t + \delta_2(\epsilon_t^2 - E(\epsilon^2)) + \mathbf{u}_t && (6.8)\\
\beta_t &= \beta_0 + \beta_3 h_t && (6.9)\\
h_t &= \phi(\alpha_0 + \alpha_1 x_{t-1} + \alpha_2 Q_{t-1} + \alpha_3 \mathrm{ES}_{t-1} + \alpha_4 h_{t-1}) && (6.10)
\end{aligned}
$$

Proposed RNN-ReESCAViaR-2h model:

$$
\begin{aligned}
Q_t &= \beta_t + \beta_1 x_{t-1} + \beta_2 Q_{t-1} && \\
\omega_t &= \nu_t + \nu_1 x_{t-1} + \nu_2 \omega_{t-1}, && (6.11)\\
\mathrm{ES}_t &= Q_t - \omega_t && \\
x_t &= \xi_0 + \phi_1(|\mathrm{ES}_t|) + \delta_1 \epsilon_t + \delta_2(\epsilon_t^2 - E(\epsilon^2)) + \mathbf{u}_t && \\
\beta_t &= \beta_0 + \beta_3 h_t && \\
\nu_t &= \nu_0 + \nu_3 h_t && (6.12)\\
h_t &= \phi(\alpha_0 + \alpha_1 x_{t-1} + \alpha_2 Q_{t-1} + \alpha_3 \mathrm{ES}_{t-1} + \alpha_4 h_{t-1}) &&
\end{aligned}
$$

In the above models, the equation (6.5) is the quantile equation, where the quantile Q_t is influenced by the lag value of the realized measure (x) and the quantile. Here β_t is driven by h_t and this equation is common to both versions. The measurement equation (6.8) is common to both versions and captures the relationship between the realized measure and the ES. Further it includes the leverage term to capture the leverage effect and there is no parameters driven by h_t here. In Wang et al. (2023) they consider both additive and multiplicative versions in jointly modelling VaR and ES, but in this proposed framework, we only consider the additive version of the model as in ES equation (6.7), and it is common to both versions of the model we consider in this paper.

The recurrent update equation (6.10), h_t, serves as a fundamental component in both versions of the proposed RNN-ReESCAViaR framework. This equation updates the hidden state by incorporating lagged values of the x_{t-1},Q_{t-1} and ES_{t-1}, along with the previous hidden state h_{t-1} and a constant term. This recursive structure allows the model to capture temporal dependencies and nonlinear dynamics inherent in financial time series data. By integrating these lagged variables, the hidden state h_t effectively encapsulates historical information influencing the current state, enhancing the model's ability to adapt and respond to evolving market conditions.

The two versions of the RNN-ReESCAViaR framework differ primarily in treating the ω_t equation and the associated parameter ν. In the RNN-ReESCAViaR-1h model, the ω_t equation incorporates x_{t-1} and ω_{t-1} but the parameter ν_0 is a constant and does not depend on the hidden state h_t. This limitation restricts the model's capacity to adapt to changes in market dynamics based on the hidden state information. In contrast, the RNN-ReESCAViaR- 2h model introduces a more flexible approach by allowing ν_t to be a function of h_t, as outlined in equation (6.11). This modification enables the model to leverage past information more effectively, capturing the influence of the realized measure on the ω_t through a recurrent structure. By integrating h_t into the dynamics of ν_t, the RNN-ReESCAViaR-2h model enhances its ability to respond to fluctuations in market conditions, thus providing a more robust framework for estimating VaR and ES. Following section details the likelihood and model estimation process.

6.3 Model estimation

Our model estimation approach utilizes Bayesian techniques to ensure robust and consistent forecasting of VaR and ES through the RNN-ReESCAViaR framework. CAViaR-type models, as highlighted by Engle and Manganelli (2004), are often highly sensitive to initial conditions, which can impact the stability of model outputs. Li et al. (2023) address this by recommending Bayesian methods as a more dependable alternative, particularly through the use of Bayesian Sequential Monte Carlo (SMC), which proves more resilient to initial-value dependencies than maximum likelihood methods.

6.3.1 Likelihood function

Following Wang et al. (2023), our proposed model's likelihood function is grounded in the Asymmetric Laplace (AL) distribution. Koenker and Machado (1999), interpret the conventional quantile regression estimator as a maximum likelihood estimator (MLE) under the AL density, which features a mode at the quantile. Taylor (2019) establishes a conditional density function linking ES_t to a dynamic measure of volatility σ_t:

$$\lambda(r,\theta) = \frac{(\alpha-1)}{\mathrm{ES}_t} \exp\left(\frac{(r_t - Q_t)(\alpha - I(r_t \leq Q_t))}{\alpha \mathrm{ES}_t}\right), \tag{6.13}$$

This density enables the formulation of a quasi-likelihood function that captures the relationship between ES_t and quantiles. The negative logarithm of this AL-based density is consistent with the joint loss function framework presented by Fissler and Ziegel (2016), allowing us to employ Bayesian methodologies effectively.

To construct the full quasi-log-likelihood function for our model, we incorporate measurement equations reflecting the realized measures. Following Wang et al. (2023), we assume that the measurement error u_t (iid) follows a normal distribution $u_t \sim N(0, \sigma_u^2)$. Therefore, the overall quasi-log-likelihood can be expressed as:

$$\begin{aligned}\lambda(r, \mathbf{X}; \theta) &= \lambda(r;\theta) + \lambda(\mathbf{X}|r;\theta) \\ &= \sum_{t=1}^{n} \left(\log \frac{(\alpha-1)}{\mathrm{ES}_t} + \frac{(r_t - Q_t)(\alpha - I(r_t \leq Q_t))}{\alpha \mathrm{ES}_t}\right) \\ &- \frac{1}{2}\sum_{t=1}^{n}\left(\log(2\pi) + \log(\sigma_u^2) + u_t^2/\sigma_u^2\right),\end{aligned} \tag{6.14}$$

where $\mathbf{X} = X_t, X_2, ..., X_n$ and $u_t = \mathbf{X}_t - \xi_0 - \phi_1(|\mathrm{ES}_t|) - \delta_1\epsilon_t - \delta_2(\epsilon_t^2 - E(\epsilon^2))$.

6.3.2 Bayesian estimation

Bayesian methods have become increasingly popular in statistical modelling, particularly for their strength in parameter estimation and their ability to incorporate prior knowledge with observed data. These methods provide a systematic approach that enhances accuracy in various applications, especially in financial contexts. Wang et al. (2023) show through extensive simulation studies that Bayesian estimators outperform traditional AL-based quasi-maximum likelihood (QML) estimators regarding parameter estimation and risk forecasting. This emphasizes the benefits of using Bayesian approaches when facing challenges like parameter sensitivity and model uncertainty. Li et al. (2023) also highlight the reliability of Bayesian methods, pointing out that they can effectively manage the sensitivity of initial values in modeling. They argue that by treating initial values as unknown parameters within a Bayesian framework, researchers can improve estimation robustness, leading to better forecasts for risk measures such as VaR and ES.

6.3.3 Sequential Monte Carlo (SMC)

SMC methods enhance the Bayesian framework by providing efficient estimation and forecasting. SMC is particularly valuable for generating a complete predictive density for risk measures. It reuses intermediate samples obtained through a data-annealing process, which improves accuracy while keeping computations manageable (Nguyen et al. (2022b); Gunawan et al. (2022)). As Del Moral et al. (2006) note, SMC methods excel in handling complex state spaces and high-dimensional problems, making them ideal for financial modelling. SMC methods are particularly beneficial within the Bayesian framework. They enable efficient sampling from complex posterior distributions, facilitating the computation of predictive densities for VaR and ES forecasts (Li et al., 2023). The algorithm implemented in this study is adapted from Nguyen et al. (2022b), ensuring that the estimation process is both reliable and aligned with current best practices in the field.

6.3.4 Prior distributions

Specifying prior distributions for model parameters in the Bayesian framework is essential, as these priors reflect our initial beliefs before observing the data. We carefully evaluated several candidate distributions for each parameter before selecting the final priors. Table 6.1 summarizes the prior distributions for each parameter in the RNN-ReESCAViaR framework.

Table 6.1. Prior distributions

	RNN-ReESCAViaR-1h	RNN-ReESCAViaR-2h
Parameter	Prior	Prior
β_1	N(-0.9, 0.1)	N(-0.9, 0.1)
β_2	N(0.3, 0.1)	N(0.3, 0.1)
ν_1	N(0.1, 0.1)	N(0.1, 0.1)
ν_2	N(0.7, 0.1)	N(0.7, 0.1)
ϕ_1	N(0.5, 0.1)	N(0.5, 0.1)
δ_1	N(0.3, 0.1)	N(0.3, 0.1)
δ_2	N(0.5, 0.1)	N(0.5, 0.1)
β_0	N(-0.3, 0.1)	N(-0.3, 0.1)
β_3	N(-0.2, 0.1)	N(-0.2, 0.1)
ν_0	N(0.01, 0.1)	N(0.01, 0.1)
ν_3		N(0.01, 0.1)
ξ_0	N(-0.1, 0.1)	N(-0.1, 0.1)
α_0	$N(0.1, 0.1^2)$	$N(0.1, 0.1^2)$
α_1	$N(0.1, 0.1^2)$	$N(0.1, 0.1^2)$
α_2	$N(0.1, 0.1^2)$	$N(0.1, 0.1^2)$
α_3	$N(0.1, 0.1^2)$	$N(0.1, 0.1^2)$
α_4	$N(0.1, 0.1^2)$	$N(0.1, 0.1^2)$
σ	IG(0.5, 1)	IG(0.5, 1)

We employ SMC samplers for Bayesian inference and sequential forecasting, specifically leveraging likelihood annealing for in-sample analysis and data annealing for out-of-sample, expanding-window forecasts, following the approach of Nguyen et al. (2022b).

6.3.5 Likelihood Annealing SMC for in-sample Analysis

More details about SMC algorithum is explained in chapter 3 and chapter 5. Here we revisit the main steps of SMC algorithum. In the likelihood annealing approach, the goal is to sample from a generalized posterior distribution of model parameters, $\pi(\theta)$, based on all training data, $y_{1:T}$, The posterior is defined as:

$$\pi(\theta) = p(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)p(\theta), \tag{6.15}$$

where $p(\theta)$ is the prior and $p(y_{1:T}|\theta)$ is a likelihood-like function as in equation (6.14).

The SMC sampler begins by initializing M particles $\left\{W_0^j, \theta_0^j\right\}_{j=1}^{M}$ drawn from the prior distribution, where each particle weight W_0^j is initially set to $1/M$. Particles are then sequentially traversed through a set of tempered distributions,

$$\pi_t(\theta) := \pi_t(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)^{\gamma_t} p(\theta), \tag{6.16}$$

where the temperature levels $\{\gamma_t\}_{t=0}^{K}$ range from 0 to 1.

SMC Algorithm steps:

1. Reweighting: At each step t compute the unnormalized particle weights,

$$w_t^j = W_{t-1}^j p(y_{1:T}|\theta_{t-1}^j)^{\gamma_t - \gamma_{t-1}}, j = 1, ..., M \tag{6.17}$$

and normalize them by dividing each weight by the sum of all w_t^j, yielding the updated weights W_t^j.

2. Effective Sample Size (ESS): Compute ESS to assess particle efficiency:

$$\text{ESS} = \frac{1}{\sum_{j=1}^{M} \left(W_t^j\right)^2}, j = 1, ..., M \tag{6.18}$$

If ESS falls below a threshold, resample the particles.

3. Resampling and Markov Move: After resampling, the equally weighted particles are perturbed through a Metropolis-Hastings step. For each particle, propose a new state from $N(\theta_t^j, \Sigma_t)$ with Σ_t the covariance matrix and accept the proposal with probability:

$$\min\left(1, \frac{p(y_{1:T}|\theta_t^{j'})^{\gamma_t} p(\theta_t^{j'})}{p(y_{1:T}|\theta_t^{j})^{\gamma_t} p(\theta_t^{j})}\right); \tag{6.19}$$

6.3.6 Data Annealing SMC for out-of-sample forecasting

To generate one-step-ahead forecasts, we use data annealing, where the posterior distribution is updated as new data become available.

SMC Data Annealing steps:

1. Reweighting: Update the weights of particles $\left\{W_{t-1}^j\right\}_{j=1}^{M}$ according to the predictive density:

$$w_t^j = W_{t-1}^j p(y_t|y_{1:t-1}, \theta_{t-1}^j) \tag{6.20}$$

2. Effective Sample Size (ESS): Calculate ESS as before. If ESS falls below a threshold, resample the particles to maintain diversity.

3. Markov Move: Perturb each particle through a Metropolis-Hastings step using the updated posterior $\pi_t(\theta)$.

These steps ensure that our SMC procedure is well-suited for both in-sample and out-of-sample analyses, with likelihood annealing handling posterior sampling from the full dataset and data annealing managing sequential updates for forecasting. Table 6.2 represents the SMC settings used in this research.

Table 6.2. SMC settings

Variable	Description	Value
K	Number of annealing levels	10,000
M	Number of particles	2,000
c	Constant of the ESS threshold	0.8
N_{lik}	Number of Markov moves in the SMC with likelihood annealing	10
N_{data}	Number of Markov moves in the SMC with data annealing	15

6.4 Data and empirical study

6.4.1 Data description

Daily closing prices and RV data from January 2000 to June 2022 were downloaded from Oxford-man Institute's realized library (Heber et al., 2009). Data from 21 markets are included in this study. Each data set is split into an initial in-sample period, from January 2000 to December 2011, and an out-of-sample forecasting period from January 2012 to June 2022 including the COVID-19 period. Daily one-step-ahead forecasts of VaR and ES are calculated for 21 at the 2.5% and 1% probability levels in the forecast sample. A rolling window, with fixed in-sample size T, is used to estimate each of m one-step-ahead forecasts of VaR and ES in the forecast period for each series. Table 6.3 shows the total sample sizes, plus T and m in each market. T and m differ due to different non-trading days in each market.

Table 6.3. Summary of the selected data sets and their in-sample and out-of-sample split.

Index	Sample size	In-sample size (T)	Out-of-sample size (m)
AEX	5734	3056	2678
AORD	5675	3025	2650
BFX	5732	3054	2678
BVSP	5530	2954	2576
DJI	5629	3007	2622
FCHI	5736	3055	2681
FTSE	5667	3020	2647
GDAXI	5697	3049	2648
GSPTSE	5036	2413	2623
HIS	5504	2937	2567
IBEX	5699	3024	2675
IXIC	5636	3006	2630
KS11	5531	2955	2576
KSE	5477	2885	2592
MXX	5636	3009	2627
N225	5459	2909	2550
OSEAX	5186	2581	2605
RUT	5633	3006	2627
S&P500	5634	3008	2626
SSEC	5426	2886	2540
SSMI	5634	3012	2622

Figure 6.1 displays the time series plot of the absolute value of daily return and RV of FTSE as an example.

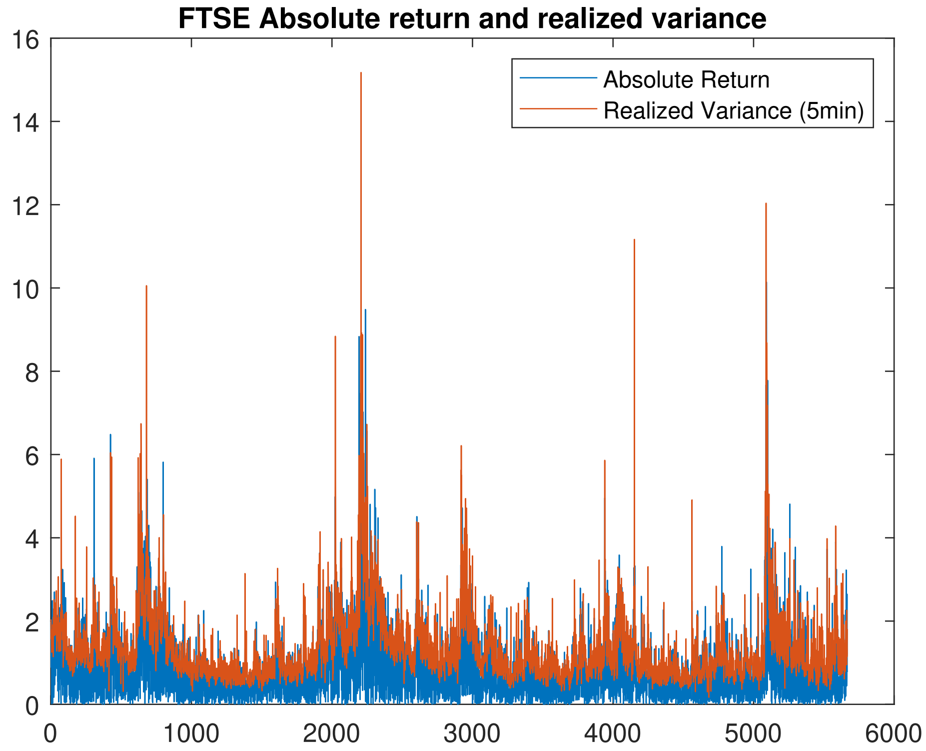

Figure 6.1. FTSE absolute return series and realized variance

6.4.2 Models in comparison

In our empirical study, we benchmark the performance of our proposed framework against prominent models in the literature that incorporate realized measures. Models that rely solely on return series do not capture the complete effect of realized measures, which are essential for accurately modeling financial risk. For this reason, we exclude models without realized measures from the comparison. We include the parametric REGARCH model (Hansen and Huang, 2016) with t distributions and three semi-parametric models, ES-CAViaR-X, ES-X-CAViaR-X (Gerlach and Wang, 2020), and Realized-ES-CAViaR (Wang et al., 2022) as they each provide valuable insights by integrating a realized measure into the risk forecasting process. All 4 models use Bayesian MCMC for parameter estimation.

Figure 6.2 displays the 2.5% VaR and ES forecasts of FTSE for the proposed RNN-ReESCAViaR-2h as an example.

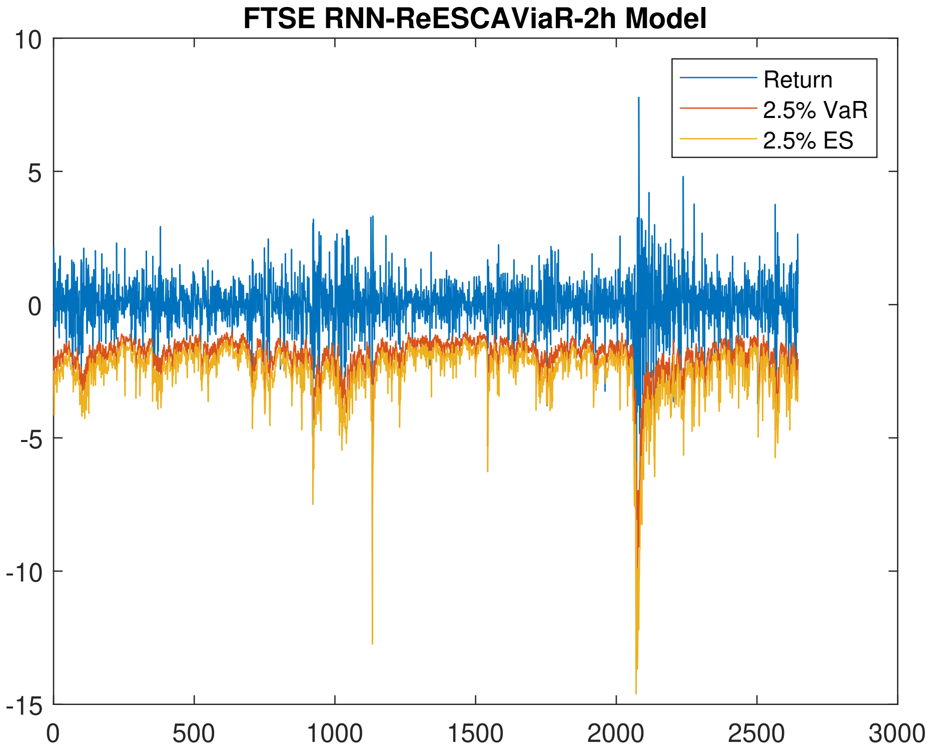

Figure 6.2. Proposed RNN-ReESCAViaR-2h model Return, VaR and ES for FTSE

Assessing VaR forecasts

This section discusses the evaluation of VaR forecasts. First, the VaR violation rate (VRate) is used as an informal criterion to assess VaR forecasting accuracy. VRate is the proportion of returns in the forecast period that violate or exceed the VaR forecasts $\widehat{Q}_t$.

$$\text{VRate } = \frac{1}{m} \sum_{t=T+1}^{T+m} I(r_t < \widehat{Q}_t) \tag{6.21}$$

where T is the in-sample size and m is the forecasting sample size. Models with VRate close to nominal, that is, $\alpha = 2.5\%$, or equivalently $\frac{\text{VRate}}{\alpha}$ close to 1, are preferred. Table 6.4 summarizes the VRates for each model over the return series. The results show that the proposed RNN-ReESCAViaR models give overall better results for 43% of the markets when $\alpha = 2.5\%$ and for $\alpha = 1\%$, it gives better performance for 38% of the total markets considered. However, compared the RNN-ReESCAViaR-1h model, RNN-ReESCAViaR-2h gives better results. When considering individual models for both α values, Realized-ES-CaViaR models perform better.

Table 6.4. VRate

	$\alpha = 2.5\%$						$\alpha = 1\%$					
Market	RNN-ReESCAViaR-1h	RNN-ReESCAViaR-2h	REGARCHt	ES-CaViaR-X	ES-X-CaViaR-X	Realized-ES-CaViaR	RNN-ReESCAViaR-1h	RNN-ReESCAViaR-2h	REGARCHt	ES-CaViaR-X	ES-X-CaViaR-X	Realized-ES-CaViaR
AEX	0.60	1.03	1.30	**1.02**	1.08	1.14	1.49	**1.27**	1.90	1.31	1.34	1.34
AORD	1.30	0.94	1.15	0.95	**0.98**	0.95	2.23	2.19	1.85	**1.06**	**1.06**	1.09
BFX	0.82	**0.96**	1.28	0.90	0.88	0.90	0.86	**1.01**	1.31	1.01	0.97	0.97
BVSP	0.39	0.47	0.92	**0.98**	0.95	0.93	0.93	0.58	**1.01**	**1.01**	0.89	0.97
DJI	**0.98**	**0.98**	1.50	1.16	1.14	1.07	1.64	**1.11**	1.87	1.68	1.60	1.53
FCHI	0.91	**1.01**	1.39	1.22	1.21	1.21	1.72	1.64	1.79	**1.45**	**1.45**	**1.45**
FTSE	**1.10**	1.15	1.57	1.25	1.25	1.21	2.08	**1.62**	2.27	1.70	**1.62**	**1.62**
GDAXI	**1.10**	1.33	1.59	1.42	1.36	1.39	2.19	**1.66**	1.85	1.89	1.77	1.85
GSPTSE	1.08	0.85	1.40	1.07	1.05	**0.99**	1.49	1.45	1.60	1.33	1.33	**1.26**
HIS	0.87	**0.97**	1.18	1.14	1.12	1.09	1.60	1.40	1.60	1.32	1.29	**1.21**
IBEX	0.73	0.72	1.18	**0.99**	**0.99**	1.03	**1.01**	**1.01**	1.23	1.20	1.20	1.16
IXIC	0.53	0.62	1.51	**1.08**	1.16	1.16	0.57	1.48	1.75	**1.22**	1.25	1.25
KS11	0.51	0.59	1.23	0.89	0.87	**0.90**	1.28	0.78	1.28	**0.89**	0.85	**0.89**
KSE	**1.02**	0.93	1.22	1.06	1.05	0.94	1.66	1.77	1.66	1.16	1.43	**1.00**
MXX	0.55	0.44	**0.91**	0.65	0.65	0.64	0.53	0.53	**0.72**	0.57	0.57	0.57
N225	1.24	1.27	1.24	1.19	1.21	**1.15**	2.04	1.84	1.61	1.45	1.45	**1.29**
OSEAX	0.64	0.66	1.29	0.91	0.92	**0.94**	1.15	0.77	1.42	0.81	0.88	**1.04**
RUT	0.38	0.38	**0.90**	0.65	0.69	0.61	0.49	0.42	**0.95**	0.80	0.80	0.76
S&P500	1.10	1.19	1.40	1.08	1.08	**1.07**	1.41	**1.22**	1.90	1.37	1.37	1.29
SSEC	0.71	0.71	1.04	1.20	1.13	**1.01**	1.57	1.18	1.26	1.26	1.30	**0.94**
SSMI	0.92	**0.98**	1.43	1.08	1.07	1.10	1.56	1.22	1.64	**1.11**	**1.11**	1.14

Note: Bold numbers indicate the favoured model for each market.

As a more formal test, the quantile loss as in equation (6.22) is also used to assess VaR forecasting accuracy.

$$\frac{1}{m}\sum_{t=T+1}^{T+m}\left(\alpha - I(r_t \leq \widehat{Q}_t)\right)\left(r_t - \widehat{Q}_t\right), \tag{6.22}$$

where $\widehat{Q}_{T+1}, ..., \widehat{Q}_{T+m}$ are the quantile forecasts at level α. Since the quantile loss function is strictly consistent, the model with minimum sample quantile loss is preferred. Table 6.5 shows the 2.5% and 1% VaR quantile loss function results. The average rank based on the ranks of the quantile loss across the 21 markets is also included in the "Avg Rank". Results show that our proposed model, RNN-ReESCAViaR-2h, performs better for both α values. When $\alpha = 2.5\%$, and $\alpha = 1\%$, the RNN-ReESCAViaR framework gives better performance for 43% and 48%, respectively, for all the markets considered.

Table 6.5. Quantile loss values

	$\alpha = 2.5\%$						$\alpha = 1\%$					
Market	RNN-ReESCAViaR-1h	RNN-ReESCAViaR-2h	REGARCHt	ES-CaViaR-X	ES-X-CaViaR-X	Realized-ES-CaViaR	RNN-ReESCAViaR-1h	RNN-ReESCAViaR-2h	REGARCHt	ES-CaViaR-X	ES-X-CaViaR-X	Realized-ES-CaViaR
AEX	191.17	**188.27**	201.09	191.82	191.68	192.52	92.04	**90.49**	106.55	91.85	92.08	92.10
AORD	159.47	158.56	**157.55**	161.22	161.03	161.77	83.88	83.57	81.83	78.26	**78.20**	78.51
BFX	189.94	**189.30**	202.57	190.95	190.81	191.74	93.91	**93.87**	98.74	93.89	93.92	94.04
BVSP	268.63	265.29	252.45	**250.74**	251.00	254.50	**123.82**	126.20	125.17	127.67	127.47	127.52
DJI	163.86	**163.79**	175.42	164.34	164.70	168.30	84.48	83.11	83.27	81.51	**81.32**	82.70
FCHI	216.66	**216.56**	225.65	219.69	219.72	220.17	108.82	**108.49**	111.74	108.91	108.77	109.02
FTSE	177.14	**176.43**	179.15	177.04	177.27	178.07	92.19	87.95	89.35	85.31	**84.86**	85.08
GSPTSE	**147.06**	147.54	149.34	149.76	150.09	153.05	79.54	78.98	**75.46**	78.65	78.72	78.03
HIS	203.47	202.06	204.13	200.98	**200.90**	201.16	97.11	95.46	98.67	95.70	95.55	**94.92**
IBEX	237.66	238.01	240.47	236.43	**236.32**	239.13	**117.90**	118.10	120.60	119.77	119.52	121.03
IXIC	199.65	199.23	207.93	**197.30**	198.04	198.38	**90.92**	91.19	97.36	93.12	93.15	93.42
KS11	164.01	159.93	160.34	**156.37**	156.38	156.56	75.54	74.78	78.58	73.91	73.98	**73.91**
KSE	198.04	197.73	201.35	**196.83**	197.19	202.10	103.00	104.09	102.56	**97.73**	103.80	100.95
MXX	178.87	181.44	**170.44**	173.71	173.65	174.33	85.70	85.42	**82.13**	85.93	86.05	85.94
N225	226.12	227.23	227.47	225.50	225.07	**224.35**	118.63	114.51	114.89	112.63	112.68	**110.91**
OSEAX	187.24	186.96	183.24	**182.94**	182.97	187.84	86.58	**85.87**	88.36	87.29	87.74	88.08
RUT	236.12	234.34	230.92	221.81	**221.59**	221.84	**106.53**	108.98	110.49	108.16	108.31	107.95
S&P500	166.54	**166.12**	172.76	169.67	169.64	170.27	**82.30**	82.55	83.25	82.64	82.88	83.04
SSEC	240.86	241.50	245.67	**239.54**	241.78	242.49	124.67	121.19	127.05	**119.82**	120.26	122.86
SSMI	164.22	**164.21**	168.44	164.94	164.64	166.27	83.62	82.74	83.61	82.20	82.39	**82.11**
Avg Quantile loss	196.79	196.30	199.45	194.81	194.93	196.39	97.30	96.44	98.64	95.95	96.28	96.08
Avg Rank	3.38	2.86	4.90	2.62	2.71	4.52	3.76	2.90	4.76	2.76	3.19	3.29

Note: Bold numbers indicate the favoured model for each market

To assess the statistical significance of quantile loss differences among models, we initially applied the Model Confidence Set (MCS) method, introduced by Hansen et al. (2011). The MCS identifies a set of models expected to contain the best-performing models within a chosen confidence level, thus allowing for an objective comparison of forecasting models. MCS tests under the 90% confidence level and employs test statistics (R and SQ) to compare quantile loss among models. However, these results are omitted from the final analysis, as they do not contribute additional insight into model performance.

Assessing ES forecasts

As discussed in Section likelihood function, Taylor (2019) shows that the negative of the quasi-log-likelihood function (6.13) is strictly consistent for Q_t and ES_t considered jointly, and fits into the class of strictly consistent joint loss functions for VaR and ES developed by Fissler and Ziegel (2016). We use the average joint loss $S = \frac{1}{m}\sum_{t=n+1}^{n+m} S_t$ to formally and jointly assess the VaR and ES forecasts from all models.

$$S_t(r_t, \widehat{Q}_t, \widehat{\text{ES}}_t) = -\log\left(\frac{\alpha - 1}{\widehat{\text{ES}}_t}\right) - \frac{(r_t - \widehat{Q}_t)(\alpha - I(r_t \leq \widehat{Q}_t))}{\alpha\widehat{\text{ES}}_t}. \tag{6.23}$$

Table 6.6 shows the 2.5% and 1% VaR and ES joint loss function values for each model and market. The "Avg Rank" calculated in the same way as Table 6.5 is included. The results are generally consistent with the ones from the quantile loss study. From all markets considered, our proposed RNN-ReESCAViaR framework performs 52% better than the other models for $\alpha = 2.5\%$, and it is 47.62% for $\alpha = 1\%$. Comparing individual models, proposed RNN-ReESCAViaR-1h performs as the best model for 6 markets while RNN-ReESCAViaR-2h model performs the best model for 5 markets when $\alpha = 2.5\%$ considering all 21 markets. When $\alpha = 1\%$, the proposed RNN-ReESCAViaR-2h is the best model for six markets, the highest from all models considered.

6.4.2 Models in comparison

Table 6.6. Joint loss values

	$\alpha = 2.5\%$						$\alpha = 1\%$					
Market	RNN-ReESCAViaR-1h	RNN-ReESCAViaR-2h	REGARCHt	ES-CaViaR-X	ES-X-CaViaR-X	Realized-ES-CaViaR	RNN-ReESCAViaR-1h	RNN-ReESCAViaR-2h	REGARCHt	ES-CaViaR-X	ES-X-CaViaR-X	Realized-ES-CaViaR
AEX	5343.50	**5306.37**	5570.70	5351.90	5339.70	5357.60	5750.40	**5719.63**	6326.00	5808.20	5787.50	5780.80
AORD	4943.11	**4890.43**	4899.90	4938.50	4937.10	4939.60	5732.41	5653.74	5587.40	5427.50	**5422.20**	5424.80
BFX	5298.69	**5287.19**	5468.50	5314.10	5305.30	5327.90	**5753.70**	5756.81	5988.30	5778.70	5763.80	5783.70
BVSP	6239.11	6199.93	6033.20	**6031.50**	6033.60	6037.60	6551.81	6593.15	**6546.00**	6704.40	6626.80	6614.90
DJI	**4739.53**	4756.26	4950.80	4802.10	4795.70	4833.60	5325.75	**5287.01**	5380.60	5351.00	5327.40	5338.10
FCHI	**5699.86**	5706.52	5948.70	5781.50	5772.10	5769.40	6292.64	**6247.77**	6538.00	6317.20	6291.20	6281.70
FTSE	5116.01	**5103.21**	5254.30	5146.00	5147.90	5151.60	5842.72	5646.57	5822.60	5600.90	5586.70	**5583.80**
GDAXI	**5676.93**	5702.25	6039.60	5753.50	5752.00	5738.70	6383.98	**6225.09**	6539.90	6325.20	6322.10	6289.70
GSPTSE	**4508.24**	4512.19	4575.60	4561.80	4558.80	4624.40	5253.85	5155.11	**5067.90**	5170.80	5145.20	5139.10
HIS	5528.73	5508.38	5548.40	5495.60	5493.90	**5487.70**	5945.17	5895.92	6017.70	5902.80	5897.60	**5872.70**
IBEX	5953.18	5941.34	6002.90	5948.70	**5936.70**	5955.80	**6416.75**	6426.27	6550.10	6533.00	6498.00	6532.90
IXIC	5478.72	5417.17	5600.60	**5381.90**	5382.70	5384.90	**5728.80**	5756.75	6095.50	5806.30	5804.10	5794.80
KS11	4953.46	4893.38	4895.20	4803.90	**4796.20**	4797.00	5228.12	5232.10	5390.50	5228.20	5228.70	**5222.00**
KSE	5500.74	5480.37	5564.40	5486.40	**5479.40**	5551.20	6338.21	6217.04	6190.20	**6049.80**	6160.10	6112.50
MXX	5216.54	5281.13	**5100.70**	5144.60	5143.70	5152.50	5616.00	5605.75	**5556.60**	5640.80	5641.90	5642.10
N225	5743.71	5753.75	5809.60	5751.80	5752.10	**5731.80**	6554.42	6380.24	6398.90	6300.90	6308.20	**6249.30**
OSEAX	5233.93	5245.54	5187.10	5177.10	**5169.90**	5213.30	5582.64	**5552.68**	5598.70	5564.10	5564.60	5580.30
RUT	5917.43	5856.20	5787.60	5668.10	5663.50	**5660.60**	6106.26	6191.64	6227.50	6117.80	6111.50	**6097.70**
S&P500	4824.20	**4821.04**	4986.30	4914.90	4905.50	4905.00	**5346.66**	5374.89	5505.00	5428.90	5437.60	5436.80
SSEC	**5710.62**	5734.71	5809.30	5738.60	5726.50	5712.70	6340.37	**6239.14**	6405.30	6355.80	6305.00	6294.30
SSMI	**4878.91**	4888.26	5032.50	4916.80	4904.70	4942.60	5473.63	5429.21	5559.80	5419.10	**5410.20**	5431.80
Avg Joint loss	5357.39	5346.93	5431.71	5338.54	5333.19	5346.45	5884.01	5837.45	5966.31	5849.11	5840.02	5819.57
Avg Rank	3.29	2.95	5.05	3.43	2.62	3.67	3.52	2.62	4.95	3.81	3.14	2.80

Note: Bold numbers indicate the favoured model for each market.

In summary, based on 1% and 2.5% VRate, quantile and joint loss evaluations for VaR and ES forecasting accuracy across 21 markets, the proposed RNN-ReESCAViaR framework demonstrates consistently strong performance compared to competing models, underscoring its effectiveness and robustness in risk forecasts. The results indicate moderate gains over existing approaches, with the potential for further improvement by refining the RNN settings within the model.

In addition to the primary evaluation techniques discussed in this chapter, the MCS test, as introduced in Chapter 2, was applied to assess the relative performance of the proposed models. However, the results of the MCS test were consistent across all models and markets, showing no significant differences in predictive performance. As such, the MCS results are not presented here, but the application of the test supports the robustness of the models in this study.

While the proposed RNN-ReESCAViaR framework has demonstrated strong performance in joint forecasting of VaR and ES across various evaluation metrics, it is important to highlight that the backtesting and evaluation of ES separately were not included in this study. Although the Dynamic ES test by Patton et al. (2019), the bootstrap-based method by McNeil and Frey (2000), and the method by Du and Escanciano (2017) are well-established techniques for assessing ES forecasting, they were not employed here due to the scope and focus of the current work. These tests could be considered in future research to further enhance the robustness of the model's ES predictions and to validate its performance in more granular detail. Future extensions of this model may also explore the incorporation of multiple realized measures and alternative model configurations to further improve both VaR and ES forecasting accuracy.

6.4.3 Computational efficiency

This section compares the computational efficiency of the proposed RNN-ReESCAViaR models with other models used for joint VaR and ES forecasting. While the RNN-ReESCAViaR models utilize the SMC method for Bayesian inference, the other models (Realized-ES-CAViaR, ES-X-CAViaR-X, ES-CAViaR-X, and REGARCH-t) employ the Bayesian MCMC method.

For the proposed RNN-ReESCAViaR models, we used two configurations: RNN-ReESCAViaR-1h and RNN-ReESCAViaR-2h, which involve the use of RNNs and Bayesian SMC. Given the complexity of these models, they required substantial computational resources to run across 21 markets in the study.

- The RNN-ReESCAViaR-1h model took approximately 6404.45 seconds (1.77 hours) to run for all 21 markets.
- The RNN-ReESCAViaR-2h model took approximately 37801.48 seconds (10.5 hours) to run for all 21 markets.

These models demonstrated efficient performance even when scaled across multiple markets, with processing times that are relatively fast for models involving RNNs and Bayesian SMC methods. For comparison, the other models—such as Realized-ES-CAViaR, ES-X-CAViaR-X, and REGARCH-t—were run for just one market (e.g., S&P market with m = 2626 and $\alpha = 1\%$) with the following computational times:

Realized-ES-CAViaR: 66 hours per block (1 core, 18GB memory).

ES-X-CAViaR-X: 66 hours per block (1 core, 18GB memory).

ES-CAViaR-X: 35 hours per block (1 core, 18GB memory).

REGARCH-t: 18 hours per block (6 blocks, 24 cores, 64GB memory).

The proposed RNN-ReESCAViaR models were able to handle the computational load for all 21 markets within significantly less time compared to the models running for a single market, demonstrating the scalability and efficiency of the framework.

In conclusion, the RNN-ReESCAViaR models, despite their complexity, provide a highly efficient solution for forecasting joint VaR and ES across a large number of markets. The computational times for all 21 markets show that the proposed models offer an efficient approach to risk forecasting, particularly given the additional complexities introduced by the RNN and Bayesian SMC methods.

6.5 Chapter summary

This study introduces an RNN-based, semi-parametric framework for jointly forecasting VaR and ES, incorporating realized measures to enhance risk prediction. The proposed RNN-ReESCAViaR models demonstrate competitive performance across VaR rates, quantile loss, and joint loss for VaR and ES, showing favourable results relative to both parametric models (e.g., REGARCH) and other semi-parametric approaches (e.g., ES-CaViaR-X, ES-X-CaViaR-X, and Realized-ES-CaViaR). By utilizing RNN, the framework better captures complex non-linear patterns in financial data that traditional models may miss, enhancing forecasting accuracy. Furthermore, adopting Bayesian SMC for model estimation improves computational efficiency, significantly reducing run time compared to Bayesian MCMC methods.

Future work may further enhance model performance by optimizing RNN settings and refining SMC parameters. Additional extensions could include incorporating multiple realized measures at various frequencies and their subsampled versions. While the current model relies on single lags, it can be extended to multiple lags for a more comprehensive temporal structure. Lastly, alternative model configurations, such as a multiplicative, time-varying relationship between VaR and ES driven by multiple realized measures, present promising directions for future research.

Chapter 7

Conclusion and future directions

This chapter concludes the dissertation by summarizing the key contributions of the research and highlighting directions for future work in tail risk forecasting. It emphasizes the integration of realized measures, Bayesian techniques, and machine learning methods, particularly RNNs, in improving forecasting performance for VaR and ES. Furthermore, the chapter discusses the regulatory implications of these models, particularly how deep learning-based risk forecasting could support regulatory stress testing and align with the Basel III requirements on risk management.

In this thesis, we have proposed innovative approaches to forecasting financial tail risk by utilizing multiple realized measures, Bayesian inference, and advanced machine learning techniques, primarily RNNs. Each chapter has built upon previous work, gradually advancing the precision and robustness of VaR and ES predictions, which are critical for effective financial risk management.

Chapter 4 introduces a novel semi-parametric framework for joint forecasting of VaR and ES, the Realized-ES-CAViaR-M model, which incorporates multiple RMs. The model demonstrates significant improvements over both parametric models, such as the REGARCH model, and existing semi-parametric models, including ES-X-CAViaR-X and Realized-ES-CAViaR. The Realized-ES-CAViaR-M model shows competitive performance across several key metrics, such as quantile loss, VaR and ES joint loss, and MCS backtesting. These results confirm the potential of integrating multiple realized measures for more accurate and robust risk forecasting.

This research also has important regulatory implications. The proposed model could play a significant role in regulatory stress testing frameworks, which are crucial for ensuring that financial institutions hold adequate capital during periods of market stress. By more accurately forecasting tail risks, including VaR and ES, the model can provide regulators with better tools to assess financial institutions' ability to withstand extreme market conditions. This aligns with the Basel III framework, which emphasizes the need for more accurate measures of tail risk and capital adequacy to ensure financial stability. The flexibility and accuracy of the Realized-ES-CAViaR-M model make it a promising candidate for stress testing scenarios, where accurate forecasts of potential losses are critical for determining appropriate capital buffers.

Despite these promising findings, there are several avenues for further enhancement. Future research could explore the inclusion of additional realized measures, including their subsampled versions and variations at different frequencies, to further improve forecasting accuracy. The current framework uses single lags, which could be extended to incorporate multiple lags for more comprehensive dynamic relationships between VaR and ES. Furthermore, alternative model structures, such as introducing a multiplicative time-varying relationship between VaR and ES through the integration of information from multiple realized measures, offer a promising direction for refinement. These extensions could further enhance the applicability of this framework in the context of real-time risk management and regulatory compliance.

A key area for future research involves identifying the most influential realized measures in forecasting volatility and risk. By expanding the set of RMs and enhancing the modelling framework, it will be possible to apply automatic variable selection techniques, such as LASSO (Tibshirani, 1996), to better capture the most relevant RMs for improved forecasting performance. This approach could lead to more efficient models that focus on the most impactful risk indicators, contributing to enhanced risk management and decision-making processes.

Chapter 5 introduces the RNN-HAR model, a long-memory, non-linear realized volatility framework tailored for VaR forecasting. This model extends the traditional HAR structure, which is renowned for capturing long memory in realized volatility, by integrating RNNs to address non-linear market dynamics. The RNN-HAR model captures daily, weekly, and monthly realized variance effects, providing a richer temporal perspective for estimating risk measures. This is particularly beneficial for risk management because it allows for a more granular and dynamic understanding of market volatility across different time horizons.

A key innovation lies in employing Bayesian SMC with a quantile loss function, enabling loss-based Bayesian inference directly aligned with risk forecasting objectives. This approach ensures that the model effectively captures long-term dependencies while adapting to the non-linearities inherent in financial data. Empirical analysis across 31 global markets demonstrates the model's superior performance over traditional HAR models and their extensions, affirming its efficacy in forecasting risk measures across various significance levels. These findings have clear practical implications, as the model can be directly applied in financial institutions' risk management practices, enhancing their ability to forecast risk more accurately and manage capital reserves more effectively, in line with Basel III requirements.

Future research could further enhance the RNN-HAR model by incorporating global optimization techniques for parameter estimation, (Goffe et al., 1994; Jerrell and Campione, 2001; Sun et al., 2024), potentially improving forecasting precision. Moreover, expanding the range of realized measures or exploring their subsampled and multi-frequency variants offers an opportunity to enrich the model's input features. The integration of advanced architectures, such as Long Short-Term Memory (LSTM) networks, may provide additional flexibility in modeling complex temporal dependencies and improve its predictive robustness. These enhancements could further solidify the RNN-HAR framework as a leading tool for non-linear and long-memory risk forecasting in financial markets, which is crucial for stress testing and regulatory compliance.

Chapter 6 introduces an innovative semi-parametric framework, RNN-ReESCAViaR, for jointly forecasting VaR and ES by incorporating realized measures to enhance risk prediction by incorporating RNN. Leveraging the capabilities of RNNs, the model effectively captures complex non-linear patterns in financial data, outperforming traditional parametric models (e.g., REGARCH) and other semi-parametric approaches (e.g., ES-CAViaR-X, ES-X-CAViaR-X, and Realized-ES-CAViaR). The model demonstrates better forecasting performance across multiple metrics, including VaR rates, quantile loss, and joint loss for VaR and ES.

The adoption of Bayesian SMC for model estimation significantly enhances computational efficiency, offering a robust alternative to the Bayesian MCMC approach. This combination of advanced machine learning techniques and Bayesian estimation underscores the effectiveness of the RNN-ReESCAViaR framework in risk forecasting. The model's ability to provide real-time, dynamic risk forecasts makes it particularly suitable for financial institutions that need to continuously monitor and manage their risk exposure in a rapidly changing market environment.

Future research could focus on several promising directions to further enhance this framework. First, optimizing RNN configurations and exploring advanced structures such as LSTM networks may improve the model's ability to capture long-term dependencies and complex temporal relationships. Incorporating multiple realized measures at various frequencies, including their subsampled versions, could enrich the model's inputs, while employing techniques like LASSO for variable selection could identify the most impactful realized measures for risk forecasting. These improvements would make the RNN-ReESCAViaR framework even more adaptable to real-time risk management and better aligned with Basel III's objectives of improving the accuracy and reliability of risk assessments in stress scenarios.

Extending the model to include multiple lags in its structure could provide a more comprehensive temporal representation, potentially improving predictive accuracy. Alternative configurations, such as incorporating a multiplicative and time-varying relationship between VaR and ES driven by multiple realized measures, also hold promise. Finally, exploring global optimization techniques for parameter estimation and testing log-transformed versions of the model could yield additional improvements in efficiency and performance. These enhancements and extensions present valuable opportunities for advancing the RNN-ReESCAViaR framework, solidifying its role as a leading tool for joint risk forecasting in financial markets, and contributing to stress testing in regulatory settings.

Collectively, the contributions from Chapters 4, 5, and 6 provide a significant advancement in the field of tail risk forecasting. By combining realized measures with machine learning and Bayesian inference, this thesis demonstrates a practical path forward for enhancing the predictive accuracy of VaR and ES measures, especially under extreme market conditions. The models proposed in this work not only deepen the theoretical understanding of financial risk but also provide practical solutions that align with the fast-evolving demands of the finance industry. These contributions underscore the effectiveness of machine learning methods, such as RNNs, in capturing the complex dependencies inherent in financial data, while the adoption of Bayesian techniques enhances estimation precision and forecasting performance. Future work could build upon this foundation by further optimizing RNN settings, experimenting with alternative neural network architectures like LSTM models, and incorporating additional realized measures or subsampled data across various time frames to capture even more detailed volatility patterns. Furthermore, these models can contribute significantly to regulatory stress testing, providing regulators and financial institutions with more accurate tools to assess capital adequacy and manage financial stability under stressed market conditions.

Before concluding, it is worth noting that an area for potential future research lies in extending the models developed here to bi-variate and multi-variate settings. In particular, copula models could be explored to model the joint distributions of multiple asset classes or risk factors, enabling the analysis of systemic risk. The application of CoVaR (Conditional Value-at-Risk) could be a key extension, assessing the risk contribution of individual institutions to the overall systemic risk. These extensions would provide a more comprehensive view of financial stability, offering valuable insights for regulators and financial institutions in understanding interdependencies and systemic risk.

In conclusion, this thesis sets forth a compelling case for using realized measures, Bayesian methods, and machine learning approaches to significantly enhance risk forecasting. This research advances both the theoretical understanding and practical implementation of financial tail risk models, offering a robust, adaptable framework that meets the increasingly complex challenges of risk management in the modern financial landscape, and helping to ensure compliance with Basel III regulations.

Bibliography

Acerbi, C. and Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value at risk. *Economic notes*, 31(2):379–388. (Cited on page 96.)

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905. (Cited on pages 15, 34, and 63.)

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55. (Cited on pages 15 and 16.)

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625. (Cited on pages 15, 16, 17, 34, and 63.)

Arnerić, J., Poklepović, T., and Teai, J. W. (2018). Neural network approach in forecasting realized variance using high-frequency data. *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy*, 9(2):18–34. (Cited on page 65.)

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228. (Cited on pages 6, 34, and 96.)

Asai, M., McAleer, M., and Medeiros, M. C. (2012). Asymmetry and long memory in volatility modeling. *Journal of Financial Econometrics*, 10(3):495–512. (Cited on pages 67 and 68.)

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1):3–30. (Cited on page 63.)

Baker, S. R., Bloom, N., Davis, S. J., and Terry, S. J. (2020). Covid-induced economic uncertainty. Technical report, National Bureau of Economic Research. (Cited on page 96.)

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536. (Cited on pages 15 and 16.)

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. (Cited on pages 15, 16, 17, and 35.)

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(2):253–280. (Cited on page 15.)

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37. (Cited on pages 15, 17, and 35.)

Basel Committee on Banking Supervision (2023). MAR *Calculation of* RWA *for market risk*. Bank for International Settlements. (Cited on pages 34 and 96.)

BCBS (2019). Minimum capital requirements for market risk. *Suiza: BCBS*. (Cited on page 83.)

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):1103–1130. (Cited on pages 27, 28, 66, and 73.)

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327. (Cited on pages 34, 46, 63, and 97.)

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18. (Cited on page 68.)

Branco, R. R., Rubesam, A., and Zevallos, M. (2024). Forecasting realized volatility: Does anything beat linear models? *Journal of Empirical Finance*, page 101524. (Cited on pages 65 and 66.)

Chen, C. W. and So, M. K. (2006). On a threshold heteroscedastic model. *International Journal of Forecasting*, 22(1):73–89. (Cited on page 21.)

Chen, Q., Gerlach, R., and Lu, Z. (2012). Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics & Data Analysis*, 56(11):3498–3516. (Cited on pages 8, 9, 11, 20, 21, 23, and 25.)

Chen, W. Y., Peters, G. W., Gerlach, R. H., and Sisson, S. A. (2022). Dynamic quantile function models. *Quantitative Finance*, 22(9):1665–1691. (Cited on pages 43 and 44.)

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American statistical association*, 96(453):270–281. (Cited on page 25.)

Christensen, K. and Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141(2):323–349. (Cited on pages 15, 17, and 34.)

Christensen, K., Siggaard, M., and Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5):1680–1727. (Cited on page 65.)

Christoffersen, P. (2011). *Elements of financial risk management*. Academic press. (Cited on page 64.)

Clements, A. and Preve, D. P. (2021). A practical guide to harnessing the HAR volatility model. *Journal of Banking & Finance*, 133:106285. (Cited on pages 64, 69, and 79.)

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2):223. (Cited on page 63.)

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196. (Cited on pages 63 and 67.)

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27(1-3):46–78. (Cited on page 67.)

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30(3):368–380. (Cited on page 67.)

Cowles, M. K. (2013). *Applied Bayesian statistics: with R and OpenBUGS examples*, volume 98. Springer Science & Business Media. (Cited on page 20.)

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436. (Cited on pages 26, 74, and 103.)

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263. (Cited on page 84.)

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144. (Cited on page 88.)

Du, Z. and Escanciano, J. C. (2017). Backtesting expected shortfall: accounting for tail risk. *Management Science*, 63(4):940–958. (Cited on page 115.)

Duan, J.-C. and Fulop, A. (2015). Density-tempered marginalized sequential Monte Carlo samplers. *Journal of Business & Economic Statistics*, 33(2):192–202. (Cited on page 73.)

Engle, R. (2001). Garch 101: The use of ARCH/GARCH models in applied econometrics. *Journal of economic perspectives*, 15(4):157–168. (Cited on page 97.)

Engle, R. F. (1982a). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007. (Cited on page 34.)

Engle, R. F. (1982b). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007. (Cited on page 63.)

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381. (Cited on pages 9, 10, 35, 41, 84, 91, 92, 97, and 101.)

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44(4):1680–1707. (Cited on pages 12, 35, 41, 52, 96, 97, 102, and 113.)

Francq, C. and Zakoïan, J.-M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics*, 184(1):158–173. (Cited on page 36.)

Frazier, D. T., Loaiza-Maya, R., Martin, G. M., and Koo, B. (2024). Loss-based variational Bayes prediction. *Journal of Computational and Graphical Statistics*, (just-accepted):1–31. (Cited on pages 27, 28, and 73.)

Frey, R. and Embrechts, P. (2010). *Quantitative risk management*. Princeton University Press. (Cited on page 64.)

Gao, F. and Song, F. (2008). Estimation risk in GARCH VaR and ES estimates. *Econometric Theory*, 24(5):1404–1424. (Cited on page 37.)

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, third edition. edition. (Cited on pages 20, 23, and 24.)

Gerlach, R., Chen, C. W., and Lin, E. M. (2016). Bayesian assessment of dynamic quantile forecasts. *Journal of Forecasting*, 35(8):751–764. (Cited on pages 25, 91, and 92.)

Gerlach, R. and Wang, C. (2020). Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. *International Journal of Forecasting*, 36(2):489–506. (Cited on pages 35, 37, 38, 40, 47, 98, and 108.)

Gerlach, R. H., Chen, C. W., and Chan, N. Y. (2011). Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *Journal of business & economic statistics*, 29(4):481–492. (Cited on pages 73 and 98.)

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press. (Cited on page 23.)

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801. (Cited on pages 34, 46, and 97.)

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762. (Cited on pages 12, 79, and 97.)

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422. (Cited on page 11.)

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of econometrics*, 60(1-2):65–99. (Cited on page 119.)

Goodfellow, I. (2016). Deep learning. (Cited on pages 28, 29, 30, and 31.)

Gunawan, D., Kohn, R., and Tran, M. N. (2022). Flexible and robust particle tempering for state space models. *Econometrics and Statistics*. (Cited on pages 74 and 103.)

Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational statistics*, 14:375–395. (Cited on pages 21 and 23.)

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. (Cited on page 22.)

Hansen, P. R. and Huang, Z. (2016). Exponential GARCH modelling with realized measures of volatility. *Journal of Business & Economic Statistics*, 34(2):269–287. (Cited on pages 18, 35, 37, 42, 47, 97, and 108.)

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906. (Cited on pages 15, 16, 35, 37, and 97.)

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497. (Cited on pages 13, 14, 52, 84, 88, and 112.)

Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). Oxford-man institute's realized library, version 0.3. *Oxford-Man Institute, University of Oxford*. (Cited on pages 44, 77, and 106.)

Huang, Z., Liu, H., and Wang, T. (2016). Modeling long memory volatility using realized measures of volatility: A realized HAR GARCH model. *Economic Modelling*, 52:812–821. (Cited on pages 68 and 69.)

James, G. (2013). An introduction to statistical learning. (Cited on page 29.)

Jerrell, M. E. and Campione, W. A. (2001). Global optimization of econometric functions. *Journal of Global Optimization*, 20:273–295. (Cited on page 119.)

Jorion, P. (2007). Value at risk: The new benchmark for managing financial risk. (Cited on page 6.)

Kim, H. Y. and Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, 103:25 – 37. (Cited on page 65.)

Knoblauch, J., Jewson, J., and Damoulas, T. (2019). Generalized variational inference: Three arguments for deriving new posteriors. *arXiv preprint arXiv:1904.02063*. (Cited on pages 28, 66, and 73.)

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448):1296–1310. (Cited on pages 41, 69, and 101.)

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444. (Cited on page 30.)

Li, D., Clements, A., and Drovandi, C. (2023). A bayesian approach for more reliable tail risk forecasts. *Journal of Financial Stability*, 64:101098. (Cited on pages 28, 41, 73, 101, 102, and 103.)

Li, Z., Tran, M.-N., Wang, C., Gerlach, R., and Gao, J. (2020). A bayesian long short-term memory model for value at risk and expected shortfall joint forecasting. *arXiv preprint arXiv:2001.08374*. (Cited on page 21.)

Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *ProQuest One Academic*. (Cited on page 65.)

Liu, C., Wang, C., Tran, M.-N., and Kohn, R. (2024). A long short-term memory enhanced realized conditional heteroskedasticity model. *Economic Modelling*, page 106922. (Cited on page 30.)

Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates. *Economic Review-Federal Reserve Bank of San Francisco*, (2):3. (Cited on page 11.)

Martens, M. and Van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, 138(1):181–207. (Cited on page 34.)

Matsubara, T., Knoblauch, J., Briol, F.-X., Oates, C., et al. (2021). Robust generalised bayesian inference for intractable likelihoods. *arXiv preprint arXiv:2104.07359*. (Cited on pages 28 and 73.)

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3-4):271–300. (Cited on page 115.)

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press. (Cited on pages 7 and 8.)

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370. (Cited on pages 34, 46, and 97.)

Nguyen, N., Tran, M.-N., Gunawan, D., and Kohn, R. (2019). A long short-term memory stochastic volatility model. *arXiv preprint arXiv:1906.02884*, page 120. (Cited on pages xxiv and 31.)

Nguyen, T.-N., Tran, M.-N., Gunawan, D., and Kohn, R. (2022a). A statistical recurrent stochastic volatility model for stock markets. *Journal of Business & Economic Statistics*, 41(2):414–428. (Cited on pages 21, 65, and 73.)

Nguyen, T.-N., Tran, M.-N., and Kohn, R. (2022b). Recurrent conditional heteroskedasticity. *Journal of Applied Econometrics*, 37(5):1031–1054. (Cited on pages 26, 27, 65, 74, 75, 76, 77, 103, and 104.)

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697. (Cited on page 68.)

Patton, A. J. and Zhang, H. (2022). Bespoke realized volatility: Tailored measures of risk for volatility prediction. *Available at SSRN 4315106*. (Cited on pages 65 and 66.)

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of econometrics*, 211(2):388–413. (Cited on page 115.)

Pourrezaee, A. and Hajizadeh, E. (2024). Forecasting bitcoin volatility and value-at-risk using stacking machine learning models with intraday data. *Computational Economics*, pages 1–31. (Cited on pages 65 and 66.)

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied probability*, 7(1):110–120. (Cited on pages 23 and 44.)

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of computational and graphical statistics*, 18(2):349–367. (Cited on page 21.)

Roccioletti, S. (2015). *Backtesting value at risk and expected shortfall*. Springer. (Cited on page 96.)

Sako, K., Mpinda, B. N., and Rodrigues, P. C. (2022). Neural networks for financial time series forecasting. *Entropy*, 24(5):657. (Cited on page 98.)

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–117. (Cited on page 29.)

Sheppard, K. (2009). MFE MATLAB function reference financial econometrics. *Unpublished paper, Oxford University, Oxford. Available at: http://www. kevinsheppard. com/images/9/95/MFE_Toolbox_Documentation. pdf*. (Cited on pages 52 and 88.)

Sun, Y., Mutalib, S., Omar, N., and Tian, L. (2024). A novel integrated approach for stock prediction based on modal decomposition technology and machine learning. *IEEE Access*. (Cited on page 119.)

Taylor, J. W. (2008a). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252. (Cited on page 98.)

Taylor, J. W. (2019). 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. (Cited on pages 12, 35, 37, 41, 52, 69, 97, 98, 102, and 113.)

Taylor, S. (2008b). *Modelling financial time series*. World Scientific. (Cited on page 63.)

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288. (Cited on pages 50 and 118.)

Tsay, R. S. (2010). *Analysis of financial time series*. John Wiley & sons. (Cited on page 63.)

Wang, C., Gerlach, R., and Chen, Q. (2023). A semi-parametric conditional autoregressive joint value-at-risk and expected shortfall modeling framework incorporating realized measures. *Quantitative Finance*, 23(2):309–334. (Cited on pages 25, 35, 38, 39, 41, 47, 51, 98, 99, 100, 101, and 102.)

Wang, J., Cui, Q., Sun, X., and He, M. (2022). Asian stock markets closing index forecast based on secondary decomposition, multi-factor analysis and attention-based lstm model. *Engineering Applications of Artificial Intelligence*, 113:104908. (Cited on pages 98 and 108.)

Wang, J., Wang, S., Lv, M., and Jiang, H. (2024). Forecasting var and es by using deep quantile regression, gans-based scenario generation, and heterogeneous market hypothesis. *Financial Innovation*, 10(1):36. (Cited on page 98.)

Watanabe, T. (2012). Quantile forecasts of financial returns using realized GARCH models. *The Japanese Economic Review*, 63(1):68–80. (Cited on page 47.)

Xu, Q., Liu, X., Jiang, C., and Yu, K. (2016). Nonparametric conditional autoregressive expectile model via neural network with applications to estimating financial risk. *Applied stochastic models in business and industry*, 32(6):882–908. (Cited on page 98.)

Zhang, C.-X., Li, J., Huang, X.-F., Zhang, J.-S., and Huang, H.-C. (2022). Forecasting stock volatility and value-at-risk based on temporal convolutional networks. *Expert Systems with Applications*, 207:117951. (Cited on pages 65 and 66.)