THE UNIVERSITY OF SYDNEY

DOCTORAL THESIS

Bayesian Parametric Financial Risk Forecasting Employing Multiple High-Frequency Realized Measures

Author:
Vica TENDENAN

Supervisors:
Prof. Richard GERLACH
Assoc. Prof. Boris CHOY
Dr. Chao WANG

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

DISCIPLINE OF BUSINESS ANALYTICS
BUSINESS SCHOOL

2023

Declaration of Authorship

I, Vica Tendenan, declare that this thesis titled 'Bayesian Parametric Tail Risk Forecasting Using Realized EGARCH Models' and the work presented in it are my own. I confirm that:

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

Authorship Attribution Statement

I co-designed the research with the co-authors (who are my supervisors), developed the research methodology, conducted the simulation and empirical studies, analysed the results, and co-wrote the manuscript draft.

Vica Tendenan

12 July 2023

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

Prof. Richard Gerlach

14 July 2023

"It takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"

— Lewis Carroll, Alice in Wonderland

Dedicated to my son, Arsenio Amartya Vorca,
born 20 February 2014.

Abstract

Tail risk forecasting plays a crucial role in prudent risk management strategy, as required by the Basel accord, in preventing devastating losses and providing efficient capital allocation. Recent developments in volatility modeling to forecast market risks have shifted to modeling high frequency data of financial asset returns to improve the performance of volatility forecasting models. This volatility modeling development also attempts to capture extreme but rare events that may have a disastrous impact on a global scale. Various estimation techniques have also been explored to provide better parameter inference and prediction.

This thesis aims to develop parametric volatility modeling using multiple high-frequency realized volatility measures to forecast two types of tail risks, that is the Value at Risk (VaR) and the Expected Shortfall (ES), at 2.5% and 1% forecast levels. An extension of the realized exponential generalized autoregressive conditional heteroskedasticity model (realized EGARCH) is proposed by allowing the return equation errors to follow a standardized Student-t distribution and a standardized skewed Student-t distribution. It employs three robust types of robust realized volatility measures in the measurement equations of the realized EGARCH model, which channel the impact of the realized measures on the expectations of future volatility. These realized volatility measures are subsampled realized variance (RVSS), subsampled realized range (RRSS), and realized kernel (RK). A Bayesian technique is proposed to estimate the proposed model, and its performance is proven to be more favorable than that of a frequentist maximum likelihood (ML) approach in simulation studies. The proposed models are used in an empirical study to forecast the VaR and ES of seven market indices. The results of the empirical study suggest that the proposed extension of the realized EGARCH model employing the standardized skewed Student-t distribution improves the accuracy of tail risk prediction. Including the RRSS (either employed individually or jointly with RVSS and/or RK) in the realized EGARCH model is found to be important in improving the forecast performance of the model.

A variable selection method is further proposed based on the Least Absolute Shrinkage and Selection Operator (Lasso) approach to facilitate the selection of realized volatility measures in the realized EGARCH model. A cross validation (CV) method and a Bayesian Lasso (BLasso) approach are proposed and compared in both simulation and empirical studies using seven market indices. In the empirical study, RK, RVSS, RRSS, and Range variables are incorporated into the RE-SkN model as measures of realized volatility. Even though the estimated penalty parameters produced by the CV method are higher than that estimated by the BLasso approach for the simulation and empirical studies, both approaches tend to shrink the parameter coefficients relevant to the RK and Range variables, suggesting that the RRSS and RVSS provide a stronger signal about future volatility than the RK and Range variables. The results also suggest that the BLasso approach leads to more sparse realized EGARCH models than the CV approach.

A further extension of the realized EGARCH model is proposed by allowing the return error distribution to follow a standardized two-sided Weibull distribution. A Bayesian approach

is also proposed to estimate the model. It produces less biased and more precise parameter estimates than the ML estimation. The tail prediction accuracy of the model is compared with that of the realized EGARCH employing the standardized skewed Student-*t* distribution and other competing models in an empirical study employing RVSS, RRSS, and RK. The empirical study suggests that the forecast performance of the realized EGARCH employing the standardized two-sided Weibull distribution is relatively equal to that of the one using the standardized skewed Student-*t* distribution.

Acknowledgements

I would like to sincerely thank all of my supervisors who have guided me in my PhD study at the University of Sydney. Professor Richard Gerlach, with his extensive knowledge and profound experience in the field of Bayesian econometrics, has supervised me in navigating my academic life in Australia and in finding a new perspective in conducting forecasting analysis. I am also grateful for the opportunity to work as a tutor in some of your classes. I am indebted to Associate Professor Boris Choy for all the support in my research and life in Sydney, as well as a generous opportunity to be included in his teaching team. I learned abundantly from your excellence in teaching and your approach toward your teaching team and students. I would also like to thank Dr. Chao Wang, who also supervised and mentored me in the last two and half years of my study. Thank you for your patience in answering so many technical questions, your insightful comments, corrections, and encouragement.

I would like to express my sincere acknowledgment to the Australian Government for funding my study, both my Master's and PhD via the Australia Awards Scholarship. I am thankful for the opportunity to study a a world-class university and to have a wonderful and memorable life in Australia. My appreciation extends to the Business Analytics discipline staff members who have involved me in various teaching teams throughout my PhD study. This opportunity has given me abundant teaching experience. Also, thanks to Fiona Crawford from Editorial Collective for proofreading service for this thesis.

My profound gratitude goes to Dr. Bagus Santoso, the founder of DEFINIT, who has supervised and mentored me since I was still his undergraduate student. I am forever grateful for your unwavering support and for giving me space to grow and expand my horizon. I learn so much from your perfectionism and professionalism. Thank you for your trust and confidence in me and for making DEFINIT a second home for me in Yogyakarta. Thank you to my fellow DEFINITers for their kind supports and understanding while I am pursuing my PhD. I look forward to working with you to provide significant contributions to our office and our country, Indonesia.

My PhD journey is obviously one of the hardest times of my life, especially during the Covid 19 pandemic. I am thankful for the new friendships that have formed along the way. I found many angels on my winding path. Thank you Rangika, for our discussions over lunches in the research center lounge, and for always lending me your ears. Thank you Relyta Neilsen and her family for being my guardian angel in Sydney. Thank you Toraya Sikamalik New South Wales for all the family gatherings, chats, laughs, and traditional Toraja foods. For my fellow Indonesian students in Sydney, thank you so much for our time together, whether it was potluck parties, road trips, children's play dates, photo sessions, or simply chatting over coffee. You make this journey more memorable and colorful.

My special thanks go to the New South Wales government for its excellent health and education support for my son, who has special needs, during my study. Your excellent service

has made my PhD journey as a mother much more bearable.

I want to express my heartfelt gratitude to my parents and siblings for their continued prayers and unfailing encouragement. I hope I can make you proud.

Last but most importantly, thank you to my beloved husband "Papong", for being there all the way, for putting your career on hold while I pursue this PhD, and for being a very supportive husband and a great father to our son. You are the Yin of my Yang.

Praise the Lord, Jesus Christ. With His blessing, I could finish this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Financial Risk and Forecasting

The 2008 global financial crisis has underscored the importance of effective financial risk management and the need for precise risk measurement to prevent disastrous losses. The environment in the financial services industry is dynamic, changing rapidly and characterised by evolving potential risks that financial institutions need to manage. The nature of the risks faced by financial institutions also evolves as financial innovation progresses. Therefore, it is necessary for financial institutions to develop accurate risk measures that can effectively adjust to the fast-changing environment in which they operate.

Engle and Manganelli (2001) broadly define risk as the level of uncertainty about net returns in the future. Financial institutions face a variety of risks, with four types of risk mainly discussed in the finance literature as follows.

1. Credit risk: the potential loss suffered if a counterpart fails to fulfil its obligation.
2. Operational risk: the risk associated with errors in payments and transactions, as well as fraud and regulatory risks.
3. Liquidity risk: the loss caused by sudden unanticipated extreme negative cash flow, especially when a firm has a sudden need for liquid assets while its assets are mostly illiquid at that particular period of time.
4. Market risk: relates to the uncertainty of earning in the future due to market fluctuations. It is the most crucial risk since it measures potential losses caused by a decrease in the market value of an asset.

This study focuses on forecasting market risk. Currently, there are two main tools employed by financial institutions in measuring market risk: Value at Risk (VaR) and Expected Shortfall (ES) or conditional VaR. While VaR has been very popular among risk managers due to its simplicity, ES is recommended by the Basel Committee on Banking Supervision (2012) because it performs better in capturing fat-tailed events.

Volatility forecasting plays a crucial role in risk management and is a major input in investment and monetary policies. It is argued that asset price volatility is one of the main risk

factors in financial markets. Therefore, governments and monetary authorities take into account the volatility of key economic and financial indicators to assess the vulnerability of the economy to adverse shocks. The results of volatility analysis provide valuable input into the design of preemptive actions to mitigate the impact of potential worst case scenarios.

Because it is essential that real-time forecast of market risks are accurate, both financial market actors and researchers have developed models and methodologies to measure and forecast portfolio return volatility. Recent attempts at volatility modeling to forecast market risks have shifted from modeling quarterly and monthly data to shorter period or high frequency data, such as weekly, daily, and even intraday data of financial asset returns. Incorporating these high-frequency data into volatility forecasting modeling significantly improves the performance of volatility forecasting models.

This thesis has the following structure. Chapter 1 provides an introduction to the thesis with a brief discussion on market risk and volatility modeling and outlines the theoretical background of Bayesian inference. Chapter 2 presents a theoretical review of VaR and ES and various types of tail risk forecast evaluation methods, ranging from VaR and ES backtests, quantile loss functions, VaR and ES joint loss function, and model confidence set.

Chapter 3 presents an extension of the Realized EGARCH models from a Bayesian perspective as a new method of modeling volatility by introducing the Student-t distribution and the Skewed Student-t distribution in the return equation (RE-tN model and RE-SkN model, respectively). To improve the out-of-forecast performance of the proposed model, three types of realized measures are selected to be utilized, individually or jointly, in the measurement equation of the Realized EGARCH models. These are the subsampled Realized Variance (RVSS), subsampled Realized Range (RRSS), and Realized Kernel (RK). An adaptive Bayesian Markov Chain Monte Carlo (MCMC) procedure is adopted and applied to the proposed models to help account for the estimation of risk forecasts. The proposed models are estimated by using the Maximum Likelihood (ML) frequentist method and adaptive Bayesian MCMC method. The superior performance of the proposed adaptive MCMC method over the ML method is demonstrated via a simulation study. The proposed models are applied to seven market indices and used to produce VaR and ES forecasts. Various VaR and ES backtests are conducted to evaluate the the accuracy of 2.5% and 1% VaR and ES forecasts of the proposed models. The forecasting performance of the proposed model is compared to that of various competing models, such as the original realized GARCH and realized EGARCH models, the Conditional Autoregressive Expectile (CARE) model, and several extensions of the GARCH-type models.

Chapter 4 builds on the RE-SkN model with a focus on performing the selection of realized measures and improving the prediction accuracy of the RE-SkN model by using the least absolute shrinkage and selection operator (Lasso) approach to build a parsimonious RE-SkN model. Two variable selection methods are adapted and applied to the RE-SkN model: a cross validation (CV) method and a Bayesian Lasso (BLasso) approach. The penalty parameters of both methods are compared in both a simulation study and an empirical study. In

the empirical study, RK, RVSS, RRSS, and Range variables are incorporated into the RE-SkN model as measures of realized volatility. The prediction accuracy of the RE-SkN models in forecasting VaR and ES by using both variable selection methods is evaluated by comparing the quantile losses and the joint losses to the original RE-SkN model (without the Lasso approach) and to the GARCH and the EGARCH models. The latter two models also incorporate the Skewed Student-*t* distribution in the return equation.

Chapter 5 proposes a further extension of the Realized EGARCH model by exploring another alternative return distribution, that is, a standardized two-sided Weibull (STW) distribution. A Bayesian approach based on the Adaptive MCMC is also developed for the proposed Realized EGARCH model extension. A simulation study is conducted to compare the unbiasedness and precision of the MCMC estimation and the ML estimation of the RE-STW model. In an empirical study using seven market indices, the tail risk prediction accuracy of the RE-STW model is evaluated in comparison with that of the RE-SkN type models proposed in Chapter 3 and other competing models such as GARCH-family models and the CARE model.

Chapter 6.1 presents the conclusion of the thesis and suggests possible future research avenues.

1.2 Volatility Modeling

The available methods for estimating VaR and ES can be classified into three main categories: parametric, semi-parametric, and non-parametric (Engle & Manganelli, 2004; Taylor, 2008). The parametric approach includes the use of volatility models that estimate conditional quantiles using the resulting forecasts of conditional volatility. Statistically speaking, the volatility of a series of returns (r_t) is the square root of the conditional variance of the return given the available information available prior to time t (F_{t-1}).

$$\sigma_t = \sqrt{\text{Var}(r_t) \mid F_{t-1}}$$

The semi-parametric approach is based on Extreme Value Theory (EVT) and quantile regression. However, EVT cannot be directly applied to estimate VaR and ES, since financial returns usually suffer heteroskedasticity or do not have constant variance. Meanwhile, a variant of quantile regressions called the CAViaR model, proposed by Engle and Manganelli (2004), is useful for calculating the conditional quantile, but does not provide a clear mechanism for estimating ES. Taylor (2008) then introduced a new class of univariate models, called Conditional Autoregressive Expectile Models (CARE), which deliver estimates for both VaR and ES.

The non-parametric approach involves the use of rolling historical quantiles. VaR is calculated as the quantile of the distribution of the historical returns of the most recent moving window, while ES is measured as the mean of the returns that exceed the estimated VaR. The main difficulty in applying this approach is in determining the length of the previous

periods to be included in the moving window. On the one hand, including short past observations may result in a high level of sampling error. On the other hand, including long past observations may produce less sensitive estimates that require a lengthy timeframe in responding to the dynamics in the true distribution (Taylor, 2008).

In particular, the first approach, volatility models, have received significant attention in risk modeling. Early models developed to capture volatility dynamics are the Autoregressive Conditional Heteroskedasticity models in the seminal paper of Engle (1982) and the generalized ARCH model of Bollerslev (1986). The ARCH and GARCH models can capture the volatility clustering phenomenon first observed by Mandelbrot (1963, p. 418) as the periods where "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes."

ARCH models estimate the conditional mean and conditional variance of a series simultaneously and assume that the variance of the error in the present period (t) is influenced by the errors of previous periods (Enders, 2014). The ARCH of order p for a series of returns (r_t) is written in the following equations:

$$\begin{aligned} r_t &= \mu + \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2. \end{aligned} \tag{1.1}$$

where $a_t = \sigma_t z_t$, μ is the mean of the return, α_0, α_1, ..., α_p are the parameters, $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i > 0$, and z_t are independent and identically distributed error processes (iid) with zero mean and unit variance. The conditional variance in Equation 1.2 is often denoted as h_t. In this case, the shocks are written as $a_t = \sqrt{h_t} z_t$. The stationarity condition requires $\sum_{j=1}^{p} \alpha_i < 1$.

Bollerslev (1986) proposed a more general process to improve the flexibility of the lag structure, called Generalized Autoregressive Conditional Heteroskedasticity (GARCH). The general form of the GARCH model for a series of returns (r_t) is defined by the following equations:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2. \tag{1.2}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0; \beta_j \geq 0$, and the stationarity condition requires $\sum_{i=1}^{q} \beta_j + \sum_{j=1}^{p} \alpha_i < 1$.

GARCH models are essentially an extension of ARCH models, where conditional variances are modeled as an Autoregressive Moving Average (ARMA) process. The key element in the GARCH framework is the conditional variance specification, which is used to extract the information about the expected volatility. The standard GARCH models use squared returns as a proxy for the current volatility level, which then can be used to predict future volatility level (Bollerslev, 1986).

ARCH and GARCH models have undergone various further extensions. These include Exponential GARCH (EGARCH) models developed by Nelson (1991) and GJR-ARCH models by Glosten et al. (1993). EGARCH and GJR-ARCH models both aim to model the asymmetric volatility phenomenon, where negative and positive shocks tend to have asymmetric impacts on conditional volatility.

The specification of error distributions in volatility models is also an important aspect of the estimation of parametric tail risks. Early financial time series modeling based on ARCH and GARCH type models was largely based on the assumption of normal errors, while much evidence shows that financial time series are highly leptokurtic and often skewed. For example, Bollerslev (1987) allowed for conditional t-distributed errors in ARCH and GARCH models; Hansen (1994) and Fernández and Steel (1998) proposed a skewed Student-t distribution; Chen et al. (2012) combined an asymmetric Laplace distribution (ALD) with a GJR-GARCH model; Chen and Gerlach (2013) developed a more flexible extension of the ALD, a two-sided Weibull distribution, as the conditional return distribution in four GARCH-type models.

1.3 Joint Models for Returns and Realized Measures of Volatility

The traditional volatility models above use daily returns to model conditional variance, which provides weak signals about the current level of volatility (Hansen et al., 2012). In addition, these models have a slow adjustment to volatility movements, since they rely on long and slowly decaying weighted moving averages of past squared returns (Andersen et al., 2003).

The availability of high-frequency data and improved computing power has given rise to the introduction of realized volatility measures in place of returns in forecasting VaR and ES. The realized volatility can be measured by using the empirical variability of daily return computed from high-frequency intra-period returns. The realized volatility is considered unbiased, highly efficient, and more informative in modeling return volatility (Andersen et al., 2003).

Researchers have used various volatility measures, such as realized variance (Andersen & Bollerslev, 1998; Andersen et al., 2001, 2003), realized kernel (Barndorff–Nielsen et al., 2008; Barndorff–Nielsen & Shephard, 2002), and realized range (Christensen & Podolskij, 2007; Martens & van Dijk, 2007). Realized variance (RV), which is simply the sum of the intraday squared returns, is a very popular measure. However, RV is highly sensitive to market friction when estimating the volatility of very high frequency data, such as minute-to-minute or second-to-second data (Barndorff–Nielsen et al., 2008)).

Realized range (RR) was introduced by Martens and van Dijk (2007) and Christensen and Podolskij (2007) as a realized estimator. It is based on the price range observed during a certain time period. In theory, there is more information on volatility in RR compared to

that in RV, since it uses all the price dynamics in a time period, not just the price at the start and end of each subperiod (Gerlach et al., 2017).

The class of realized kernel (RK) estimators was introduced by Barndorff–Nielsen et al. (2008). RK can be used in the estimation of the quadratic variation of an underlying efficient price process from high-frequency noisy data. It combines intra-day volatility estimation and kernel smoothing in producing robust and efficient estimators in the presence of certain types of market friction.

Realized volatility estimation has encouraged the emergence of a new class of GARCH models that use various realized volatility measures, the first of which was called GARCH-X. GARCH-X model was estimated by Engle (2002) using realized variance. However, GARCH-X models could not capture change in the realized measures and could not be used to forecast return and volatility more than one period ahead (Hansen et al., 2012).

Engle and Gallo (2006) later improved the GARCH-X model and developed the complete model known as the Multiplicative Error Model (MEM), which takes into account the conditional dynamics of daily returns and adds a latent volatility process for each estimated realized measure. They found that MEM could produce an accurate forecast of market volatility one month ahead.

Recently, two new frameworks have been introduced with an objective to jointly model asset returns and realized measures of volatility. These models are the realized GARCH model developed by Hansen et al. (2012) and the realized exponential GARCH model introduced by Hansen and Huang (2016). The realized GARCH model is an extension of the GARCH-type models that take advantage of the availability of the high-frequency data by jointly modeling the returns and realized measures of volatility. The key feature of the realized GARCH model is an additional equation, called measurement equation, that contemporaneously relates the conditional variance with a realized measure. Hansen and Huang (2016) proposed the realized EGARCH model, which improves the realized GARCH model by allowing the use of multiple realized volatility measures and facilitating a more flexible modeling of the dependence between returns and volatility. Another extension of the realized GARCH model, called the threshold realized GARCH model, was recently proposed by Chen and Watanabe (2019a) by incorporating high-frequency realized volatility measures into a threshold framework to look at different volatility behaviors across regimes.

The original realized GARCH model Hansen et al. (2012) and the realized EGARCH model (Hansen & Huang, 2016) both assume Gausiannity of both observation and measurement equation errors. Although adopting the normal distribution for measurement equation errors is adequate, the study of Watanabe (2012) and Contino and Gerlach (2017) on the realized GARCH model suggests that the choice of observation error distribution is found to be important; in most cases the skewed Student-t distribution is favored. Wang et al. (2019) also extended the realized GARCH model by proposing the two-sided Weibull distribution developed by Chen and Gerlach (2013) for observation equation errors.

1.4 Bayesian Estimation for GARCH-type Models

GARCH model estimation in many statistical and econometric software programs is mainly conducted by using classical Maximum Likelihood (ML) techniques, which can be categorized as a frequentist approach. The main advantages of the ML approach are its simple procedure in producing the results and that it is easy to understand. Therefore, the ML estimation can be easily executed by nontechnical users (Ardia & Hoogerheide, 2010). From a theoretical point of view, ML estimators are asymptotically optimal under certain conditions, as discussed in Bollerslev et al. (1994) and Lee and Hansen (1994).

Despite these above favorable features, the ML estimation of GARCH-type models suffers from several drawbacks. First, since the optimization procedure is usually encumbered by positivity constraints and stationarity condition, the algorithm can be less stable and produces less consistent results, which will then create problems in the standard error calculation (Silvapulle & Sen, 2011). This problem might be more severe for complex models such as the realized EGARCH with multiple realized measures. Second, for realized EGARCH models, the asymptotic analysis and properties of the ML estimators are complicated and have not been adequately studied. For simplification, Hansen and Huang (2016) used a score function and the numerical Hessian matrix of the log-likelihood function to compute the standard errors for the estimators.

It is argued that the application of a Bayesian approach in GARCH-type modeling will lead to better inference and prediction compared to classical estimation approaches. First, the parameter constraints and the the covariance stationarity condition in the GARCH-type can be incorporated by specifying appropriate prior. Second, inference on the non-linear functions of the GARCH-type parameters is straightforward in the Bayesian setting. Appropriate MCMC procedures will allow us to obtain the posterior distribution of any non-linear function of the GARCH-type model parameters. The issue of local maxima convergence or convergence to incorrect values, encountered in ML estimation of complex GARCH-type models, can be avoided by using MCMC techniques. Third, Bayesian estimation could provide more reliable and consistent results even for small finite samples (Ardia & Hoogerheide, 2010; Virbickaite et al., 2015).

Fourth, the Bayesian approach presents a natural way to assess parameter uncertainty in volatility estimation that can be used in decision making, for instance, forecasting purposes. The predictive distributions of the future volatilities and the measure of precision for risk measures can be obtained via predictive interval (Ausín & Galeano, 2007; Paap, 2002). Relevant to this, the predictive distribution of risk measures, such as Value at Risk is of paramount importance in risk management (Jorion, 2001). Finally, the complexity of the GARCH-type model can be handled by using modern Bayesian computational methods, such as MCMC procedures (Ausín & Galeano, 2007). Although MCMC methods are computationally expensive, the standard errors can be easily calculated from the posterior distributions produced by the simulated chains.

In the last decade, Bayesian inference in empirical modeling has gained popularity in a wide range of fields of research because of its attractive properties and appealing performance. This approach has also been used to estimate GARCH-type models. Some examples of research that has developed a Bayesian estimator for the previous GARCH-type model are Bauwens and Lubrano (1998), Ardia et al. (2012), Jin and Maheu (2012), and Chen et al., 2014.

Bauwens and Lubrano (1998) employ a Gibbs sampler to perform Bayesian inference on GARCH models. Takahashi et al. (2009) put forward a method to model both realized volatility and daily return using a widely recognized stochastic volatility model. They approached it from a Bayesian perspective and developed an efficient MCMC sampling algorithm. Building upon this work, Takahashi et al. (2016) expanded the realized stochastic volatility model proposed by (Takahashi et al., 2009) by incorporating a broader range of distribution known as the generalized hyperbolic skew Student t distribution to model financial returns.

Ardia et al. (2012) use Monte Carlo simulation methods on a non-linear regression model for a very small data set and a mixture of GARCH model for a large data set. Jin and Maheu (2012) conduct Bayesian estimation with a range of multivariate GARCH models and realized covariance and returns. Chen et al. (2014) employ Bayesian computational method in estimating smoothly time-varying (TV) GARCH models.

Nevertheless, until recently, there have been very limited studies using Bayesian methods in Realized GARCH and Realized EGARCH models. Among the few that have developed a Bayesian estimator for the Realized GARCH model are Gerlach and Wang (2016), Contino and Gerlach (2017), Wang et al. (2019), and Chen and Watanabe (2019a) for the Threshold Realized GARCH model.

As for the Realized EGARCH model, there is even a limited study that estimates the Realized EGARCH Model in a Bayesian setting. Among the few is a recent paper of Chen et al. (2021). They apply several realized models, including the Realized GARCH of Hansen et al. (2012), the Realized EGARCH of Hansen and Huang (2016), the Realized Heterogeneous Autoregressive-GARCH or HAR-GARCH of Hansen and Huang (2016), and the realized Threshold GARCH model Chen and Watanabe (2019b).

Another variant of realized volatility model estimated using Bayesian MCMC is introduced by Chen et al. (2022), that is a realized hysteretic GARCH. The model is equivalent to a combination of a three regime nonlinear framework with daily return and realized volatility.

The latest study by Chen et al. (2023) introduces a new approach called RES-CAViaR, which incorporates daily realized volatility and expected shortfall (ES) into the conditional autoregressive value-at-risk (CAViaR) model. This model is used to forecast both VaR (value-at-risk) and ES simultaneously. Chen et al. (2023) utilize the Bayesian adaptive MCMC method to estimate all the unknown parameters and make joint predictions of daily VaR and ES. This estimation and prediction process covers a 4-year out-of-sample period, which includes the period of the COVID-19 pandemic.

1.5 Bayesian Inference

Bayesian computation is based on the Bayes theorem that aims to update the probability of a model by using a set of data in a systematic manner. The purpose of a Bayesian inference is to produce two main probability statements about a parameter, θ, or unobserved data, $\widetilde{y}$, which are conditional on the observed value of y. These probability statements are called posterior distribution, $p(\theta|y)$, and posterior predictive distribution, $p(\widetilde{y}|y)$. By using Bayesian inference, the probability model is fitted to the data set and the results are summarized in a probability distribution on the model parameters and on unobserved variables, such as the forecasts for new observations (Gelman et al., 2014).

The likelihood function and a prior distribution are two main components in a Bayesian model. According to the Bayes theorem, the posterior probability of a model is proportional to the product of the prior probability and the likelihood (Cowles, 2013), which can be simply written as

$$P(\theta \mid y) = \frac{P(\theta)P(y|\theta)}{p(y)} \propto P(\theta)\, x\, P(y \mid \theta). \tag{1.3}$$

The prior probability reflects previous knowledge of an outcome that is assessed before analysing the data. The data is then used to update this prior probability to obtain the posterior probability. In Bayesian sequential analysis, the posterior probability obtained from one step will become the prior probability for the next step (Cowles, 2013).

The estimation of a Bayesian approach requires advanced Bayesian computation methods, which mostly use the MCMC scheme. MCMC is the main computational method in Bayesian analysis that draws values of θ from approximate distribution and makes a correction on the draws to produce a better approximation of the target posterior density, $p(\theta|y)$. The sampling process is performed sequentially, and the distribution of the sampled draws is determined by the last value drawn. Therefore, the draw creates a Markov chain. MCMC is particularly useful when sampling θ directly from the posterior distribution, $p(\theta|y)$, is not possible or computationally inefficient. MCMC facilitates iterative sampling in such a way that in each step we will draw from a distribution that is closer to the target posterior density (Gelman et al., 2014).

The advantage of the MCMC method is that it can facilitate the fitting of complicated Bayesian models that have a high number of dimensions (Cowles, 2013). The key success of MCMC simulation is the improvement of the approximate distributions at each simulation step to achieve a convergence to the target distribution (Gelman et al., 2014).

1.5.1 The Metropolis algorithm

The Metropolis algorithm, introduced by Metropolis et al. (1953), adapts a random walk by using an acceptance or rejection rule to converge to the specified target distribution. Gelman et al. (2014) explain the Metropolis algorithm as follows.

1. A starting point θ^0, for which $p(\theta^0|y) > 0$, is drawn from a starting distribution $p_0(\theta)$.
2. For t=1,2,...:
   (a) Draw a proposal sample θ^* from a *jumping distribution* (or *proposal distribution*) at time t, $J_t(\theta^*|\theta^{t-1})$.
   (b) Perform the calculation of the densities ratio:
   $$r = \frac{p(\theta^*|y)}{p(\theta^{t-1}|y)} \tag{1.4}$$
   (c) Set
   $$\theta^t = \begin{cases} \theta^* & \text{with probability min(r,1)} \\ \theta^{t-1} & \text{otherwise} \end{cases}$$
   This step requires a uniform random number to be generated.
3. The acceptance probability is calculated as:
   $$A^t = min\{1, r\} \tag{1.5}$$
4. A threshold value C^t is drawn from a uniform distribution $[0, 1]$ to be compared with A^t. The proposed sample θ^* is accepted if $A^t > C^t$, that is, if the jump increases the posterior density. In this case, $\theta^t = \theta^*$. On the contrary, if the jump decreases the posterior density, $\theta^t = \theta^*$ with probability equal to the density ratio, r, otherwise set $\theta^t = \theta^{t-1}$. If the jump is rejected, when $\theta^t = \theta^{t-1}$, this is still counted as an iteration.
5. The parameter estimates are generated based on the N proposed and accepted samples $\theta^1, \theta^2, ..., \theta^N$ that converge to the target distribution $p(\theta|y)$.

1.5.2 Adaptive MCMC algorithm

The choice of an effective proposal distribution for MCMC methods is crucial to achieve convergence of the algorithm within a reasonable amount of time. This choice includes not only the size, but also the spatial orientation, of the proposal distribution. Since the target density is unknown, the choice of the effective proposal distribution is very difficult in practice. The possible solution to this problem is to use adaptive MCMC algorithms, where the proposal distribution is tuned using the history of the process (Gelman et al., 1996; Gilks et al., 1998; Haario et al., 2001).

The adaptive MCMC method adapted in this thesis is the method employed by Chen et al. (2012), who extended this procedure from Chen and So (2006). This method improves the speed of convergence and chain mixing since it enables occasional large jumps that will provide a higher opportunity for the chain to cover the posterior distribution and not stay too long in a particular position. This approach has also been used by Contino and Gerlach (2017), Gerlach and Wang (2016), and Wang et al. (2019) for the Realized GARCH model.

The adaptive MCMC method consists of two steps, the burn-in step and the MCMC sampling period. In the burn-in step, the estimation of parameter θ is performed using the random walk Metropolis algorithm. The proposed vector θ^* is simulated from a mixture Gaussian proposal distribution.

$$\theta^p = \theta^{(j-1)} + \varepsilon;\ \varepsilon \sim \rho N(0, \text{diag}\{c_1\}) + (1-\rho)N(0, \omega\text{diag}\{c_1\}) \tag{1.6}$$

where ρ is recommended to be between 0.8 and 1. Chen et al. (2012) propose setting $\rho = 0.95$. ω should be large, for example 100, c_1 is a vector of positive number to be tuned to achieve a certain observed acceptance rate. Chen et al. (2012) follow the recommendation of Gilks et al. (1996) that the observed acceptance rate lies $\in (0.15, 0.5)$. Another recommendation for the tuning proisof 0.234 under quite general conditions.

In the second step, the standard random walk Metropolis algorithm is run, employing a mixture of Gaussian proposal distributions. Proposed samples are drawn from a proposal distribution with the acceptance probability in Equation (3). The algorithm is run for a high number of iterations for each parameter that is used for the estimates for the posterior mean for each parameter.

The choice of algorithm employed in the burn-in period is particularly important, especially to achieve faster convergence of the algorithm. There are several alternatives that can be used, such as the Adaptive Proposal algorithm (Haario et al., 1999), the Adaptive Metropolis algorithm (Haario et al., 2001), and the Robust Adaptive Metropolis (RAM) algorithm (Vihola, 2012). Each of these algorithms is briefly discussed in the following sections.

1.5.3 Adaptive proposal algorithm

The Adaptive Proposal (AP) algorithm was introduced by Haario et al. (1999) and used by Contino and Gerlach (2017) in estimating the realized GARCH model of Hansen et al. (2012). The purpose of the Adaptive Proposal is to tune the proposal distribution along the search based on the covariance estimated from a fixed number of past states. Under this procedure, the MCMC process will adapt to the target distribution. Haario et al. (1999) argue that the adaptive burn-in will produce an efficient Gaussian proposal distribution for the standard Metropolis algorithm.

Suppose that at least H points $\{X_1, X_{t-H+1}, ..., X_{t-1}, X_t\}$ have been generated at time t. H here is a fixed integer that functions as a *memory* parameter, which is the number of parameter vectors previously sampled used to evaluate the covariance matrix. The AP algorithm will be used to simulate a proposal vector θ^* from the multivariate Gaussian proposal distribution as given below.

$$\theta^* \sim \mathcal{N}\Big(X_t, c_d^2 R_t + c_d \varepsilon\Big), \tag{1.7}$$

where c_d is the scaling factor that depends only on the dimension of the parameter vector d, R_t is the dxd-covariance matrix determined by the H points $H_{t-H+1}, X_{t-H+2}, ..., X_t$, and ε is

a small positive constant to prevent zero variances. The value of $c_d = 2.4/\sqrt{d}$ is based on the recommendation of Gelman et al. (1996). The acceptance rule follows the usual random walk Metropolis probability.

1.5.4 Adaptive Metropolis algorithm

The AP algorithm utilizes a Gaussian distribution in the proposal distribution, which is centered on the current state, and the calculation of the covariance is based on a fixed finite number of previous states. As shown in Haario et al. (1999), the averages produced from the samples using the AP algorithm are biased. Therefore, Haario et al. (2001) then propose the Adaptive Metropolis (AM) algorithm that incorporates all the information from previous states to calculate the covariance of the proposal distribution. Suppose that the states $X_0, X_1, ..., X_{t-1}$ have been sampled at time $t-1$, the AM algorithm then sets the covariance C_t to be:

$$C_t = \begin{cases} C_0 & t \leq t_0 \\ c_d^2 \operatorname{cov}(X_0, ..., X_{t-1}) + c_d^2 \varepsilon I_d & t > t_0 \end{cases} \tag{1.8}$$

where I_t denotes the d-dimensional identity matrix, $t_0 > 0$ is the length of the initial period and C_0 is the initial covariance, which is a positive definite matrix set arbitrarily to our best prior knowledge.

There are two main advantages of the AM algorithm. First, it uses the information accumulated since the start of the simulation. This immediate adaptation results in a more effective search and therefore can reduce the number of necessary function evaluations. Second, unlike the AP algorithm, the AM algorithm has correct ergodicity properties with the assumption that the target density is bounded from above and has bounded support. This is proven in Haario et al. (2001).

Another version of the AM algorithm is proposed by Roberts and Rosenthal (2009) to improve the convergence of the algorithm and satisfy the Weak Law of Large Numbers. Suppose we have a d-dimensional target distribution $\pi(\cdot)$, the Metropolis algorithm with proposal distribution given at iteration n is given by

$$\begin{aligned} Q_n(x, \cdot) &= N(x, (0.1)^2 I_d/d) && \text{for } n \leq 2d \\ Q_n(x, \cdot) &= (1-\beta) N(x, (2.38)^2 \Sigma_n/d) + \beta N(x, (0.1)^2 I_d/d) && \text{for } n > 2d \end{aligned} \tag{1.9}$$

where Σ_n denotes the current empirical estimate of the covariance of the target distribution based on the previous runs, and β is a small positive constant. Roberts and Rosenthal (2009) suggest $\beta = 0.05$. As suggested by Roberts et al. (1997), the optimal proposal in large-dimensional context is $N(x, (2.38)^2\Sigma/d)$. Therefore, Roberts and Rosenthal (2009) approximate this using $N(x, (2.38)^2\Sigma_n/d)$.

1.5.5 Robust adaptive Metropolis algorithm

The AM algorithm coerces a constant scaling and adapts the scale of the proposed distribution. However, it does not adapt the shape of the distribution. Vihola (2012) proposes a new approach called Robust Adaptive Metropolis (RAM) algorithm that is based on a single matrix update formula that is computationally equivalent to the covariance factor update in the AM algorithm. The RAM algorithm estimates the shape of the target distribution and simultaneously allows attainment of a given mean acceptance rate. This algorithm avoids the use of empirical covariance, which can be problematic in some settings, especially if the proposal distribution has no finite second moment.

The RAM is defined recursively through the following process:

1. Compute the proposal $Y_n := X_{n-1} + S_{n-1}U_n$, where S_{n-1} is a lower diagonal matrix with positive diagonal elements and U_n is a vector of random variable from d-dimensional Gaussian distribution.
2. The proposal is accepted with probability $\alpha_n := \min\left\{1, \frac{\pi(Y_n)}{\pi(X_{n-1})}\right\}$.
3. If the proposal Y_n is accepted, set $X_n := Y_n$. Otherwise, set $X_n := X_{n-1}$.
4. Compute the lower diagonal matrix, S_n, satisfying the equation:

$$S_n S_n^T = S_{n-1}\left(I + \eta_n(\alpha_n - \alpha^*)\frac{U_n U_n^T}{||U_n||^2}S_{n-1}^T\right), \tag{1.10}$$

   where $\{\eta_n\}_{n\geq 1} \subset (0,1]$ is a step size sequence decaying to zero, often defined as $\eta_n = n^{-\gamma}$ with an exponent $\gamma \in (1/2, 1)$, $\alpha^* \in (0,1)$ is the target mean acceptance probability, and $I \in \mathbb{R}^{dxd}$ is the identity matrix.

For theoretical details of the RAM algorithm, see Vihola (2012).

1.6 Convergence Diagnosis and Efficiency Testing of the MCMC Algorithm

The convergence of the adaptive MCMC method is assessed by adopting the Gelman-Rubin diagnostic (Gelman et al., 2014). Further, the efficiency of the MCMC method is evaluated using an effective sample size test (Gelman et al., 2014) and the autocorrelation time (Chib & Jeliazkov, 2001).

In the Gelman-Rubin diagnostic, the convergence diagnosis is conducted on multiple independently simulated chains. Each chain is split in half, and then we simultaneously test the mixing and stationarity of the resulting half chains. For the mixing test, we check whether the chains have mixed well, while for the stationarity we check whether the first and second half of each chain have traversed the same distribution.

For each parameter, the Gelman-Rubin statistics is calculated as follows:

1. Several simulated chains are run, consisting of burn-in and post burn-in periods. The Gelman-Rubin statistics are conducted only on the post burn-in iterations, since the calculation of the MCMC posterior mean and root mean square error (RMSE) are based on the post burn-in iterations. Each of these chains of the post-burn-in is split into the first and second half. Let m and n denote the number of chains after splitting and the length of each chain, respectively.

2. Let $\psi_{i,j}$ denotes the ith iteration in jth chain $(i = 1, ..., n; j = 1, ..., m)$, the between-chain variance (B) is calculated as:

$$B = \frac{n}{m-1} \sum_{j\equiv 1}^{m} (\bar{\psi}_{\cdot j} - \bar{\psi}_{\cdot\cdot})^2 \tag{1.11}$$

   where $\bar{\psi}_{\cdot j} = (1/n) \sum_{i=1}^{n} \psi_{ij}$, and $\bar{\psi}_{\cdot\cdot} = (1/m) \sum_{j=1}^{m} \bar{\psi}_{\cdot j}$. Meanwhile, the whithin-chain variance (W) is given by:

$$W = \frac{1}{m} \sum_{j=1}^{m} \left[\frac{1}{n-1} \sum_{i=1}^{n} (\psi_{ij} - \bar{\psi}_{\cdot j})^2 \right]. \tag{1.12}$$

3. The marginal posterior variance of the estimand, $\hat{Var}(\theta)$, is calculated as a weighted average of W and B, that is,

$$\hat{Var}(\theta) = \frac{n-1}{n} W + \frac{1}{n} B \tag{1.13}$$

4. The Gelman-Rubin statistics for each parameter are calculated to assess the potential scale reduction of the current distribution for ψ if the simulations were continued in the limit $n \to \infty$ as follows:

$$\widehat{R} = \sqrt{\frac{\hat{Var}(\theta)}{W}}. \tag{1.14}$$

   The value of $\widehat{R}$ approaching 1 generally suggests satisfactory chain convergence. Gelman et al. (2014) recommend 1.1 as a threshold. A higher $\widehat{R}$ indicates a higher potential scale reduction, thus suggesting that further simulations may be necessary to improve convergence to the stationary distribution.

After the simulated chains have good convergence, an approximate 'effective number of independent simulation draws' for parameter ψ can be computed in three steps as follows.

1. Using the information within and between chains, compute the variogram V_t at each lag t:

$$V_t = \frac{1}{m(n-t)} \sum_{j=1}^{n} \sum_{i=t+1}^{m} (\psi_{i,j} - \psi_{i-t,j})^2. \tag{1.15}$$

2. Inverting $E(\psi_i - \psi_{i-t})^2 = 2(1-\rho_t)\text{var}(\psi)$ to estimate the sum of correlations $hat\rho_t$:

$$\hat{\rho}_t = 1 - \frac{V_t}{2\hat{Var}(\theta)}, \tag{1.16}$$

   where $\hat{Var}(\theta)$ is the estimated marginal posterior variance in Equation (1.13).

3. Since the sample correlation of large values of t is too noisy, a partial sum of all $\hat{\rho}_t$ values is computed in lieu of summing all of $\hat{\rho}_t$ values. The partial sum stars from lag 0 and continues until the sum of autocorrelation estimates for two consecutive lags is negative. The effective sample size is then estimated as

$$\hat{n}_{\text{eff}} = \frac{mn}{1 + 2\sum_{t=1}^{T}\widehat{\rho}_t}, \tag{1.17}$$

   where the stopping point T is the first odd positive integer for which $\hat{\rho}_{T+1} + \hat{\rho}_{T+2}$ is negative.

   The suggested benchmark of the effective sample size is $5m$.

Further, the autocorrelation time is calculated as follows:

$$\hat{t} = \frac{\sum_{j-1}^{m}\left(1 + 2\sum_{l=1}^{T}\hat{\rho}_j(l)\right)}{m}, \tag{1.18}$$

where $\hat{\rho}_j(l)$ is the sample autocorrelation at lag l for each parameter in the jth chain. While Chib and Jeliazkov (2001) employ the Parzen window to determine T, this study follows an ad-hoc approach of Wang et al. (2019) where the chosen T satisfies $\hat{\rho}_j(T) + \hat{\rho}_j(T-1) < 0.1$.

1.7 Chapter Summary

This chapter gives a brief overview of the various key concepts that are covered in this thesis and a review of the available literature on volatility modeling. Relevant literature on Bayesian techniques is also briefly discussed. The following chapter will provide a more detailed overview of the tail risk estimators and various tail risk forecast evaluation methods used in this thesis.

Chapter 2

Asset Return Distribution, Tail Risk Estimators, and Forecast Evaluation

This chapter presents a brief review of various types of distributions used to model financial asset return and provides an overview of two types of tail risk estimators used in this thesis, that is, Value at Risk (VaR) and Expected Shortfall (ES) or conditional VaR. This chapter also discusses various tail risk forecast evaluation methods used in this thesis, ranging from VaR and ES backtests, quantile loss functions, VaR and ES joint loss functions, and model confidence set.

2.1 Review of the Distribution of Financial Asset Returns

The distributional form of financial asset returns plays an important theoretical and empirical role in risk management. Building on the theory of speculation of Bachelier (1900), earlier theories on price changes were based on the normal or Gaussian distribution (Brada et al., 1966; Laurent, 1959; Osborne, 1959). Such an assumption has been applied in building optimal portfolio selection, deriving the capital asset pricing model (Markowitz, 1952; Sharpe, 1964), and hedging strategy of an option by assuming a Brownian motion for the price of the underlying asset (Black & Scholes, 1973; Merton, 1973).

The shape of the conditional return is, without doubt, one of the important aspects in the estimation of tail risks. Including skewness is especially of paramount importance in financial and investment decision making (Aït-sahali & Brandt, 2001; Harvey & Siddique, 1999). As highlighted by the global financial crisis of 2008, a left-skewed asset portfolio can have a devastating impact on the overall portfolio: one extreme loss could wipe out all earlier gains.

There is substantial empirical evidence that daily asset returns are leptokurtic and rather left-skewed. The study of Mandelbrot (1963, 1967) on price variation and Fama (1965) on stock price behavior are among the first that sparked a wide interest in the statistical distribution of financial asset returns. Their works found that excess kurtosis and heavy tails exist in financial data, hence rejecting the normal assumption. Mandelbrot (1963) and Fama (1963, 1965) proposed a stable Paretian distribution, as an alternative for modeling asset

return due to its stability and invariance, that is, the shape parameter or characteristic exponent α and skewness index β are constant under addition (Fama, 1965; Mittnik & Rachev, 1989). However, later empirical studies on stock returns show the violation of the stability assumption, in particular the shape parameter does not remain constant as the sampling period increases (Akgiray & Booth, 1988; Hsu et al., 1974; Lau et al., 1990). In response to the empirical consistencies of the stable Paretian distribution in modeling asset returns, some alternative fat-tailed distributions have been proposed. A mixture of a normal distribution and a stable Paretian distribution was proposed by DuMouchel (1973). Praetz (1972) and Robert and Nicholas (1974) suggested that the Student-t model has greater descriptive validity than the stable Paretian distribution. Nevertheless, these alternative distributions have a drawback in that they are not based on a probabilistic scheme, describing them as limiting distributions. They are not stable, while stability is a desirable property for asset return models according to Mittnik and Rachev (1993).

To overcome this challenge, some research has adopted the asymmetric Laplace (AL) distribution for conditional return. The AL distribution was first introduced in the work of Hinkley and Revankar (1977) and has been used in financial modeling, for example, in Madan and Seneta (1990). Chen et al. (2012) combined an AL distribution and a GJR-GARCH (Glosten et al., 1993) model to estimate VaR and ES. Compared to normal and Student-t distributions, the asymmetric Laplace distribution was found to be more conservative, consistently overestimating tail risk levels during the global financial crisis period. Hansen (1994) proposed a standardized skewed Student-t distribution to overcome the weakness of the Student-t distribution, which could be potentially restrictive since the Student-t distribution only allows for a variation in location, scale, and tail fatness. The skewed Student-t distribution is more flexible and allows for richer behaviors, being able to model both leptokurtosis and asymmetry. This type of distribution is used to model the return distribution of the Realized EGARCH model in Chapters 3 and 4.

Another special case of an extreme-value distribution is the Weibull distribution, which is a more flexible extension of the Laplace distribution. In addition to fatter tails, the Weibull distribution also has exponential tails, which makes it very attractive. The exponential tails guarantee the existence of all moments (Mittnik & Rachev, 1989). A symmetric, two-sided 'modified' Weibull distribution was developed by Sornette et al. (2000).

A more flexible asymmetric two-sided Weibull distribution was then further proposed by Chen and Gerlach (2013) to use as a conditional return distribution. Since a conditional error in a GARCH-type model should have a mean 0 and variance 1, Chen and Gerlach (2013) developed the standardized two-sided Weibull distribution (STW). In Chapter 5, the STW distribution is adopted to model return distribution in the Realized EGARCH model.

2.2 Tail Risk Estimators

2.2.1 Value at risk

The need to improve financial risk management has led to a uniform measure of risk called value at risk (VaR). Jorion (2001) discusses a brief history of VaR, which can be traced back to the well-known financial crises of the early 1990s that affected Orange County (California), Barings (UK), Metallgesellschaft (Germany), Daiwa (Japan), and many others.

During the late 1980s, Till Guldimann, while serving as the head of global research at J.P. Morgan, can be credited as the originator of the term "Value at Risk." The bank's risk management group faced a decision regarding whether they should invest in long-maturity bonds, which would generate stable earnings but expose them to fluctuations in market values, or opt for cash investments, which would maintain a constant market value. The bank prioritized "Value at Risk" over "earnings risks," leading to the emergence of VaR.

During that period, there was considerable concern about properly managing derivative risk. The Group of Thirty (G-30), which included a representative from J.P. Morgan, served as a platform for discussing optimal risk management practices. In July 1993, the term "Value at Risk" gained widespread attention through the publication of the G-30 report, marking its initial notable appearance.

Value at Risk (VaR) is the worst expected loss of a certain portfolio within a given confidence level over a target horizon (Jorion, 1996). Nowadays, it is one of the most frequently used tools in assessing the market risk of an investment. The main reason for the popularity of VaR among risk managers is that it provides a simple measure of market risk in a single number synthesized from various complex unfavorable possibilities (Engle & Manganelli, 2004). This simplicity allows unsophisticated users to comprehend the result in managing risks in an efficient manner.

VaR is useful not only for financial institutions but also for regulatory committees. For financial institutions, the VaR can be used to assess the risk facing their business, since it is a number that indicates the maximum loss of a financial position for a certain probability in a given horizon. It is viewed as the measure of loss due to a rare event when the market operates in normal circumstances. For regulatory committees, the VaR is considered as the minimum loss due to extraordinary events in the market. Therefore, the regulatory committee can use VaR as an input in setting margin requirements (Tsay, 2005).

Tsay (2005) defines the VaR under a probabilistic framework. Suppose that we are measuring the risk of a financial position for the next ℓ periods at time t. The change in asset value in financial position from time t to $t+\ell$ is denoted by $\Delta V(\ell)$, which is a random variable in time t and measured in dollars. Let the cumulative distributed function (CDF) of $\Delta V(\ell)$ be denoted by $F_\ell(x)$. The VaR of a long position over the time period ℓ with probability p is defined as (Tsay, 2005):

$$p = Pr[\Delta V(\ell) \leq VaR] = F_\ell(VaR) \tag{2.1}$$

The holder of a long financial position experiences a loss when $\Delta V(\ell) < 0$, therefore if p is small, the VaR of a long position is usually negative, which indicates a loss. The probability that the holder suffers a loss greater than or equal to VaR over ℓ periods is p. Another way to interpret this is that there is a $(1-p)$ probability that the holder will encounter loss less than or equal to VaR over ℓ periods.

Meanwhile, the holder of a short position would encounter a loss if there is an increase in the asset value $\Delta V(\ell) > 0$. In this case, VaR is defined as (Tsay, 2005):

$$p = Pr[\Delta V(\ell) \geq VaR] = 1 - Pr[\Delta V(\ell) \leq VaR] = 1 - F_\ell(VaR) \tag{2.2}$$

The VaR of a short position is usually positive when p is small. Positive values indicate a loss. It can be concluded from the above definition that VaR explains the behavior of the tail of the CDF of $F_\ell(x)$. A long position will focus on the left tail of $F_\ell(x)$, while a short position will focus more on the right tail of $F_\ell(x)$. The discussion in the VaR analysis is usually based on a long position, therefore, focusing on the left tail.

The quantity $p-$th quantile of $F_\ell(x)$, x_p, for any univariate CDF $F_\ell(x)$ and probability p where $0 < p < 1$, is calculated as follows (Tsay, 2005):

$$x_p = \inf\{x | F_\ell(x) \geq p\} \tag{2.3}$$

where inf is the smallest real number that satisfies $F_\ell(x) \geq p$.

When discussed from this probabilistic framework, the VaR can be viewed as the quantile of the loss distribution. When the CDF of $F_\ell(x)$ is known, VaR can therefore be calculated as the pth quantile of $F_\ell(x)$, ie, VaR = x_p. In practice, the CDF is unknown, and therefore the VaR estimation focuses on estimating the CDF and its quantiles.

A formal definition of VaR is also given by McNeil et al. (2005). The VaR of an asset portfolio at a given confidence level $\alpha \in (0,1)$ is given by the smallest value l such that the probability that the loss L is greater than l is not greater than $(1-\alpha)$.

$$VaR_\alpha = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}. \tag{2.4}$$

The typical values for α are $\alpha = 0.95$, $\alpha = 0.975$, or $\alpha = 0.99$.

In this thesis, four types of loss distributions are analyzed, that is, the Gaussian, the Student-t, the skewed Student-t, and the STW distribution. Chapters 3 and 4 focus on the use of the standardized skewed Student-t in tail risk forecasting, while Chapter 5 focuses on STW distribution. The other two types of distribution, the Gaussian and the Student-t distributions, are included in those three chapters for comparison purposes.

If the loss distribution F_L follows a Gaussian distribution with mean μ and variance σ^2, McNeil et al. (2005) show that VaR at a given confidence level $\alpha \in (0,1)$ can be calculated

as:

$$VaR_\alpha = \mu + \sigma\Phi^{-1}(\alpha) \text{ and } VaR_\alpha^{\text{mean}} = \sigma\Phi^{-1}(\alpha), \tag{2.5}$$

where Φ is the standard normal df and Φ^{-1} is the α-quantile of Φ.

If the loss L is such that $(L-\mu)/\sigma$ follows a Student-t distribution with v degrees of freedom, this loss model can be written as $L \sim t(v, \mu, \sigma^2)$ and the moments are denoted as $E(L) = \mu$ and $var(L) = v\sigma^2/(v-2)$ if $v = 2$. In this case, σ is not the standard deviation of the distribution. VaR is calculated using the inverse CDF of the t-distribution t_v^{-1} as specified in McNeil et al. (2005):

$$VaR_\alpha = \mu + \sigma t_v^{-1}(\alpha)\sqrt{\frac{v-2}{v}}, \tag{2.6}$$

where t_v is the df of standard t.

Meanwhile, the VaR for the case where the loss follows the standardized skewed Student-t and the STW distributions are presented in Section 3.4 and Section 5.3, respectively.

Recently, the VaR has been subject to strong criticism because of three main problems. First, it only measures the short-term volatility of the market price and, therefore, is considered less suitable for sustainable prudent risk management (Basel Committee on Banking Supervision, 2012). Second, the VaR cannot be used to calculate the potential loss for fat-tailed events. VaR is essentially a quantile of the distribution of the potential loss, and information on the size of the losses beyond the particular quantile is not available. The VaR at the confidence level α does not provide any information on the severity of losses that occur with a probability less than $1-\alpha$ (Embrechts et al., 1999; McNeil et al., 2005). Third, the VaR is considered an incoherent measure of market risks. It is not 'subadditive' because individual risks can sum up to more than portfolio risks, which is not mathematically coherent (Artzner et al., 1997, 1999).

2.2.2 Expected shortfall

Expected Shortfall (ES) has been proposed as an alternative approach to overcome the shortcomings of VaR in a seminal paper by Artzner et al. (1997, 1999) since ES has the important property of coherence. The ES is also called the conditional VaR or mean excess, $E(X|X > \hat{x}_p)$, where $\hat{x}_p$ is the quantile estimate. In other words, the ES measures the conditional expectation of the return that is higher than the VaR.

The ES measures the average losses of the worst cases at a specified confidence interval for a certain period of time (n). For example, if (n) is 5 days and the confidence interval is 95%, ES is the average loss over those 5 days when the loss is higher than the 95th percentile of the loss distribution.

The ES is also called the Tail VaR or Superquantile of a loss variable L (Emmer et al., 2013). If a portfolio loss is defined as L and $E(|L|) < \infty$ and the loss distribution function F_L, the

ES at confidence level $\alpha \in (0,1)$ can be defined as (Emmer et al., 2013; McNeil et al., 2005):

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)du \tag{2.7}$$

where $q_u(F_L) = F_L^{\leftarrow}(u)$ denotes the quantile function of F_L.

McNeil et al. (2005) show that the relationship between VaR and ES can then be expressed as follows:

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du \tag{2.8}$$

Rather than setting a certain confidence level α, the average of VaR across all levels $u \geq \alpha$ is calculated, and hence the further tail of the loss distribution can be examined. It is obvious that ES_α depends only on the loss distribution and that $ES_\alpha \geq VaR_\alpha$.

ES can be viewed as an extension of VaR by exploring further details of the tailed distribution and therefore can provide a measure of the magnitude of loss of tailed events, which VaR cannot provide (Yamai & Yoshiba, 2005). Since May 2012, the Basel Committee on Banking Supervision (2012) has recommended the use of ES to better capture the risks of fat-tailed events and ES has been widely employed to capture fat-tailed risks ever since. The Basel III Accord, which was implemented in 2019, places new emphasis on ES: 'ES must be computed on a daily basis for the bank-wide internal models to determine market risk capital requirements. ES must also be computed on a daily basis for each trading desk that uses the internal models approach (IMA).' (Basel Committee on Banking Supervision, 2019, p. 89).

If the loss distribution F_L follows a Gaussian distribution with mean μ and variance σ^2, McNeil et al. (2005) shows that the ES can be calculated as:

$$ES_\alpha = \mu + \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha} \tag{2.9}$$

where ϕ denotes the density of the standard normal distribution and Φ^{-1} is the inverse of the normal CDF.

Meanwhile, when the loss distribution F_L follows a Student t distribution with v degrees of freedom, ES can be calculated as (McNeil et al., 2005):

$$ES_\alpha = \mu + \sigma\frac{g_v(t_v^{-1}(\alpha))}{1-\alpha}\left(\frac{v+(t_v^{-1}(\alpha))^2}{v-1}\right)\sqrt{\frac{v-2}{v}} \tag{2.10}$$

where t_v is the density function, g_v the density of standard t, t_v^{-1} is the inverse CDF, and v is the estimated degrees of freedom with a restriction $v > 2$.

Meanwhile, the ES for the case where the loss follows the standardized skewed Student-t and STW distributions is presented in Sections 3.4 and 5.3, respectively.

2.3 Tail Risk Forecast Evaluation

It is crucial that financial regulators assess the forecast accuracy of models employed in forecasting the VaR and ES. Since VaR and ES are not observable, defining a statistic to evaluate the validity of VaR estimates is not straightforward. The procedure for assessing forecast accuracy is known as 'backtesting', which is a statistical method to systematically compare actual losses with the corresponding projected VaR estimates. As an example, if one day VaR is calculated based on 99% confidence level, then under normal circumstances, it is expected that a violation or exception will take place every 100 days on average.

Accuracy evaluation of a set of VaR forecasts can be conducted by examining the forecasts as one-sided interval forecasts. A violation or exception of the VaR occurs when the *ex-post* return is lower than the VaR. The violation rate is then calculated as the proportion of observations where return r_t exceeds the VaR level. The computation of the VaR violation rate is given by Chen et al. (2012) as follows:

$$VaR_{VR} = \frac{1}{m} \sum_{t=n+1}^{n+m} I(r_t < VaR_t). \tag{2.11}$$

A model that has a correct specification for the return quantile will have a violation rate equal to the nominal level α. Therefore, the ratio of violation rate to α can be used to assess competing models.

There are various tests to evaluate the forecast performance of the proposed models. They can be categorized into VaR backtests, ES backtests, loss function evaluation, and model confidence set (MCS). There are three standard VaR backtest procedures to formally evaluate the accuracy of the VaR forecast. These are the unconditional coverage (UC) test of Kupiec (1995), the conditional coverage (CC) test of Christoffersen (1998), and the dynamic quantile (DQ) test of Engle and Manganelli (2004). These tests typically focus on coverage tests that are based on a comparison between the number of times losses exceed the VaR and the expected number using statistical tests.

Additionally, to assess whether the ES forecast is specified correctly, two recently introduced ES backtests are employed. The first ES backtest utilized in this study is the bivariate Expected Shortfall Regression (ESR) backtest developed by Bayer and Dimitriadis (2018). The second ES backtest employed is the test developed by Couperier and Leymarie (2019).

In addition to testing VaR violations, measuring the size of uncovered losses is also of paramount importance. To calculate the uncovered losses, Lopez (1999) proposes the use of a loss function. However, the ES is not an elicitable risk measure (Gneiting, 2011) and in general there is no loss function minimized by the true ES. Therefore, this study employs two special VaR and ES joint loss functions based on Fissler and Ziegel (2016). Finally, the Model Confidence Set (MCS) of Hansen et al. (2011) is employed to construct a set of 'superior' models (SSM). Each of these types of tests is explained in the following sections.

2.3.1 The Unconditional Coverage test

The Unconditional Coverage (UC) test, introduced by Kupiec (1995), is one of the first VaR backtesting approaches. The UC test is a nonparametric approach that uses the proportion of the VaR violations or exceptions, hence called the proportion of failures test. For a given number of observations T, we observe the number of VaR violations at the confidence level α, x. The UC test can be employed to test whether the ratio x/T, or denoted as π, is statistically different from α. For a given portfolio and model, binomial theory is applied to evaluate the difference between the number of expected and observed VaR violations. The probability of observing $x = \sum_{t=1}^{T} H_t$ violations given T observations is $(1-\alpha)^{T-x}\alpha^x$.

The UC test is performed as a Likelihood-Ratio test that takes the following form:

$$LR_{UC} = -2\ln[(1-\alpha)^{T-x}\alpha^x] + 2\ln[(1-\pi)^{T-x}\pi^x] \tag{2.12}$$

Under the null hypothesis, the UC test has a χ^2 distribution with one degree of freedom. The 5% critical value of $\chi^2(1)$ is 3.841. Therefore, if $LR_{UC} > 3.841$, the null hypothesis can be rejected at $\alpha = 5\%$. Nevertheless, the statistical power of this test is rather poor, as discussed by Kupiec (1995), therefore, the test result should be interpreted with caution.

TABLE 2.1: Critical values for LR_{UC} statistics

	Significance Level		
	1%	5%	10%
Asymptotic $\chi^2(1)$	6.632	3.842	2.706
Finite sample	5.497	5.025	3.555

Lopez (1999) developed critical values for LR_{UC} that can be referred to. Critical values are presented in Table 2.1.

2.3.2 The Conditional Coverage test

The conditional coverage (CC) test developed by Christoffersen (1998) improves the UC test not only by testing the joint assumption of unconditional coverage but also by taking into account the independence of the violations. There are two main reasons for testing the independence of violations. These are: (1) in the presence of the dependence in violations, the asymptotic results on the distribution of the UC test will be invalid, and (2) the VaR violation clustering is an important alarm signal for any financial institution.

The Independent test examines the independence property of the VaR violation. It tests whether the probability of the VaR violation is independent over time (Christoffersen, 1998). An accurate model will suggest that a VaR violation in period t will not depend on the occurrence of a violation in the previous period (Jorion, 2001).

The sequence of I_t consists of 0s and 1s. For any two consecutive days, we will have four outcomes; 00, 01, 10, and 11. We then define $T_{i,j}(i = 0,1; j = 0,1)$ as the number of days

in which state j occurs in one day while it was at state i the day before. $T_{0,0}$ is the number of days that the previous day's indicator is 0 and the subsequent day's indicator is 0, and $T_{1,1}$ is the number of days that the previous day's indicator is 1 and the subsequent day's indicator is 1 (Jorion, 2001). The outcome of the contingency table of the deviation indicator in Christoffersen (1998) is presented in Table 2.2.

TABLE 2.2: Deviation indicator outcome

	$I_{t-1}=0$	$I_{t-1}=1$	
$I_t=0$	T_{00}	T_{10}	$T_{00}+T_{10}$
$I_t=1$	T_{01}	T_{11}	$T_{01}+T_{11}$
	$T_{00}+T_{01}$	$T_{10}+T_{11}$	**T**

We can then define π_0 as the conditional probability that 01 occurs when the previous day is 0 and π_1 as the conditional probability that 11 occurs when the previous day is 1:

$$\pi_0 = \frac{T_{01}}{T_{00}+T_{01}}, \pi_1 = \frac{T_{11}}{T_{10}+T_{11}}, \pi = \frac{T_{01}+T_{11}}{T_{00}+T_{01}+T_{10}+T_{11}} = \frac{T_{01}+T_{11}}{T}. \tag{2.13}$$

Under the null hypothesis of violation independence across days, $\pi = \pi_0 = \pi_1$

(Jorion, 2001) shows that the Likelihood-Ratio test for the Independent test can be written as:

$$\begin{aligned} LR_{Ind} = &-2\ln(1-\pi)^{(T_{00}+T_{10})}\pi^{(T_{01}+T_{11})} \\ &+2\ln[(1-\pi_0)^{T_{00}}\pi_0^{T_{01}}(1-\pi_1)^{T_{10}}\pi_1^{T_{11}}] \end{aligned} \tag{2.14}$$

The first term of Equation 2.14 is the maximized likelihood under the null hypothesis that violations are independent across days. The second term is the maximized likelihood for the observed data.

The LR statistic in Equation 2.14 also tends in distribution to the $\chi^2(1)$. If the value of LR_{Ind} is lower than 3.841, the model passes the backtest, otherwise rejected.

The conditional coverage test provides a more accurate VaR measure, since it considers not only unconditional coverage, but also independence properties (group exception) as discussed in Christoffersen (1998). The joint test statistics are given by:

$$LR_{CC} = LR_{UC} + LR_{Ind} \tag{2.15}$$

The joint LR follows the χ^2 distribution with two degrees of freedom under the null hypothesis. If the value of LR_{CC} is lower than 5.991, the model passes the backtest, otherwise rejected.

The critical values of LR_{CC} can also refer to those developed by Lopez (1999) as presented in Table 2.3.

TABLE 2.3: Critical values for LR_{CC} statistics

	Significance Level		
	1%	5%	10%
Asymptotic $\chi^2(2)$	9.210	5.992	4.605
Finite sample	6.007	5.015	5.005

2.3.3 The Dynamic Quantile test

The Dynamic Quantile (DQ) test was proposed by Engle and Manganelli (2004). It is a regression-based model of hits variable on a constant, the lagged values of the hit variable, and any function of past information. Dumitrescu et al. (2012) formally define $\text{Hit}_t(\alpha) = I_t(\alpha) - \alpha$ as the demeaned violation. It will take the value of $1-\alpha$ if $r_t < VaR$ and $-\alpha$ otherwise. Given the information known at $t-1$, the conditional expectation of $\text{Hit}_t(\alpha)$ is zero.

$$E(Hit_t(\alpha)|\mathcal{I}_t - 1) = E(Hit_t(\alpha)) = 0 \tag{2.16}$$

Under the assumption of the CC test above, $\text{Hit}_t(\alpha)$ must be uncorrelated with its past values and with other lagged variables, and its expected value is zero. Dumitrescu et al. (2012) suggest that the DQ test can be used to test the CC assumption using the following linear regression model.

$$\begin{aligned} Hit_t(\alpha) = \delta + \sum_{k=1}^{K} \beta_k Hit_{t-k}(\alpha) \\ + \sum_{k=1}^{K} \gamma_k g[Hit_{t-k}(\alpha), Hit_{t-k-1}(\alpha), ..., z_{t-k}, z_{t-k-1,...}] + \varepsilon_t, \end{aligned} \tag{2.17}$$

where ε is a discrete *i.i.d* process, $g(.)$ is a function of past violations and of variables z_{t-k} belonging to the entire set of information $\mathcal{F}_{t-1}$.

The choices for z_{t-k} comprise the lagged returns r_{t-k}, past squared returns r^2_{t-k}, past predicted VaR, $VaR_{t-k|t-k-1}(\alpha)$, or implicit volatility data.

The test for the conditional efficiency corresponds to the test for the joint nullity of all coefficients:

$$H_0 : \delta = \beta_1 = ... = \beta_k = \gamma_1 = ... = \gamma_k = 0, \ \ \forall k = 1, .., K$$

Meanwhile, the hypothesis of the independence of VaR violations is:

$$H_0 : \beta_1 = ... = \beta_k = \gamma_1 = ... = \gamma_k = 0, \ \ \forall k = 1, .., K$$

and when $\delta = 0$, the hypothesis of unconditional coverage will be fulfilled. Therefore, the null hypothesis, $E[Hit_t(\alpha)] = E(\varepsilon) = 0$ indicates that $\Pr[I_t(\alpha) = 1] = E[I_t(\alpha)] = \alpha$ (Dumitrescu et al., 2012).

The DQ test for conditional coverage takes the following form:

$$DQ_{CC} = \frac{\hat{\Psi}' Z' Z \hat{\Psi}}{\alpha(1-\alpha)} \xrightarrow[T\to\infty]{L} \chi^2(2k+1) \tag{2.18}$$

where $\Psi = (\delta, \beta_1, .., \beta_k, \gamma_1, \gamma_k)'$ is the vector of the $2K+1$ parameters and Z is the matrix of the explanatory variables in Equation 2.17.

While Engle and Manganelli (2004) recommend the use of 1 lag and 4 lags for the DQ test, Chen et al. (2012) find that the DQ test is less sensitive to the choice of the number of lags. Furthermore, Berkowitz et al. (2011) suggest that the DQ test is superior in evaluating VaR at the 1% level compared to other VaR backtests that have lower power against VaR misspecified models.

2.3.4 Regression-Based ES backtesting

The ES backtesting framework was recently introduced by Bayer and Dimitriadis (2018). It is based on a joint regression model for the quantile and the ES. Unlike other backtests that require forecasts of various quantities such as VaR and volatility, this test only requires ES forecasts as input parameters.

Bayer and Dimitriadis (2018) define the conditional quantile of a series of returns r_t given the information set $\mathcal{F}_{t-1}$ at level $\tau \in (0,1)$, or the VaR at level τ, as

$$Q_\tau(r_t|\mathcal{F}_{t-1}) = F_T^{-1}(\tau) = \inf r \in \mathbb{R} : F_t(r) \geq \tau \tag{2.19}$$

Meanwhile, the functional ES at level τ given $\mathcal{F}_{t-1}$ is defined as

$$ES_\tau(r_t|\mathcal{F}_{t-1}) = \frac{1}{\tau}\int_0^\tau F_t^{-1}(s)\mathrm{d}s. \tag{2.20}$$

In the case that the distributional function F_t is continuous in its τ- quantile, the above definition is simplified to the truncated tail mean of r_t

$$ES_\tau(r_t|\mathcal{F}_{t-1}) = E_t[r_t|r_t \leq Q_t(r_t|\mathcal{F}_{t-1})]. \tag{2.21}$$

This backtest aims to test whether a series of ES forecasts $\{\hat{e}_t, t = 1, ..., T\}$, is correctly specified relative to a series of realized returns $\{r_t, t = 1, .., T\}$. To do this, Bayer and Dimitriadis (2018) propose two types of ES regression (ESR) backtest, that is (1) the bivariate ESR backtest and (2) the intercept ESR backtest.

Under the bivariate ESR backtest method, the returns r_t are regressed on the forecast $\hat{e}_t$ and an intercept.

$$r_t = \alpha + \beta\hat{e}_t + u_t^e, \tag{2.22}$$

where $ES_\tau(u_t^e|\mathcal{F}_{t-1}) = 0$. Given that $\hat{e}_t$ is generated using the information set $\mathcal{F}_{t-1}$, the condition on the error term is equivalent to

$$ES_\tau(r_t|\mathbb{F}_{t-1}) = \alpha + \beta\hat{e}_t. \tag{2.23}$$

The null and alternative hypotheses are written as

$$H_0 : (\alpha, \beta) = (0, 1)$$
$$H_1 : (\alpha, \beta) \neq (0, 1).$$

Under the null hypothesis, the ES forecast is correctly specified, since this implies $\hat{e}_t = ES_\tau(r_t|\mathcal{F}_{t-1})$.

It is not possible to estimate the parameters (α, β) in Equation 2.22 by M- or Z/GMM-estimation stand-alone by using a semiparametric method without specifying the full conditional distribution of u_t^e due to the non-elicitability of ES (Gneiting, 2011). However, a regression technique proposed by Dimitriadis and Bayer (2017) that models the quantile and the ES jointly can be used to estimate these parameters as follows:

$$r_t = \gamma + \delta\hat{e}_t + u_t^q, \tag{2.24}$$
$$r_t = \alpha + \beta\hat{e}_t + u_t^e, \tag{2.25}$$

where $Q_\tau(u_t^q|\mathcal{F}_{t-1}) = 0$ and $ES_\tau(u_t^e|\mathcal{F}_{t-1}) = 0$.

The test is conducted by testing the null hypothesis only based on the parameters (α, β) in Equation 2.25 and a Wald test statistic is used for this test as follows:

$$T_{ESR} = \big((\hat{\alpha}, \hat{\beta})' - (0, 1)'\big)' \hat{\Sigma}_{ES}^{-1} \big((\hat{\alpha}, \hat{\beta})' - (0, 1)'\big)', \tag{2.26}$$

where $\hat{\Sigma}_{ES}$ is an estimator for the asymptotic covariance matrix of the M-estimator of the parameters (α, β). The test statistics is asymptotically χ^2-distributed with two degrees of freedom.

Meanwhile, under the framework of the intercept ESR backtest, the forecast errors $r_t - \hat{e}_t$ are regressed on an intercept

$$r_t - \hat{e}_t = \alpha + u_t^e, \tag{2.27}$$

where $ES_\tau(u_t^e|\mathcal{F}_{t-1}) = 0$.

Unlike the bivariate ESR backtest, which can only accommodate two-sided hypotheses testing, the intercept ESR backtesting procedure allows the specification of both one-sided and two-sided tests. The null and alternative hypotheses are written as follows.

$$H_0^{2s} : \alpha = 0 \text{ versus } H_1^{2s} : \alpha \neq 0$$
$$H_0^{1s} : \alpha \geq 0 \text{ versus } H_1^{1s} : \alpha < 0.$$

This procedure is a t-test based on asymptotic covariance and on the bootstrap procedure discussed in Bayer and Dimitriadis (2018). The ESR backtest is conducted based on the 'R' code downloaded from https://rdrr.io/github/BayerSe/esback/.

2.3.5 ES backtesting using multi-quantile regression

Couperier and Leymarie (2019) introduced a method to test the ES based on the theory of multi-quantile regression theory. This test exploits the relationship between VaR and ES, where ES can be approximated by several VaRs. This is based on the definition of a Riemann sum, where the τ-level ES is approximated as a finite sum of several VaR series, as given by:

$$ES_t(\tau) \approx \frac{1}{p}\sum_{j=1}^{p} VaR_t(u_j), \tag{2.28}$$

where the risk level u_j is defined by $u_j = \tau + (j-1)\frac{1-\tau}{p}$ for $j = 1, 2, ..., p$, and p represents the number of quantiles applied in the approximation.

Let L_t denote the portfolio loss observed at time t, for $t = 1, 2, ..., T$, and let $\mathcal{F}_{t-1}$ denote the information set available at time $t-1$ with $\left(L_{t-1}, L_{t-2}, ...\right) \subseteq \mathcal{F}_{t-1}$. The u_j-th VaR level of the L_t distribution is then defined as the quantity $VaR_t(u_j)$ such that

$$Pr(L_t) \geq VaR_t(u_j)|\mathcal{F}_{t-1} = u_j. \tag{2.29}$$

Couperier and Leymarie (2019) generalize the idea of Gaglianone et al. (2011), who introduced VaR as a regressor in a quantile regression model into the assessment of multiple VaRs. This is carried out by estimating a multi-quantile regression model where the *ex-post* losses $\{L_t, t = 1, ..., T\}$ are regressed on the p VaR forecasts $\{VaR_t(u_j), t = 1, .., T\}_{j=1,2,..,p}$

$$L_t = \beta_0(u_j) + \beta_1(u_j)VaR(u_j) + \epsilon_{j,t} \quad \forall j = 1, 2, ..., p, \tag{2.30}$$

where $\beta_0(u_j)$ and $\beta_1(u_j)$ denote the intercept and slope parameters at level u_j, respectively, and $\epsilon_{j,t}$ denotes the error term at risk level u_j and time t such that the u_j-th conditional quantile of $\epsilon_{j,t}$ satisfies $Q_{\epsilon j,t}(u_j; \mathcal{F}_{t-1}) = 0$.

Given the multi-quantile regression model in Equation (2.30), the u_j-th conditional quantile of L_t is defined as

$$Q_{L_t}(u_j; \mathcal{F}_{t-1}) = \beta_0(u_j) + \beta_1(u_j)VaR(u_j) \quad \forall j = 1, 2, ..., p. \tag{2.31}$$

The procedure verifies whether $VaR(u_j)$ perfectly matches $Q_{L_t}(u_j; \mathcal{F}_{t-1})$. Under the null hypothesis that a sequence of VaR is valid, the coefficients satisfy $\beta_0(u_j) = 0$ and $\beta_1(u_j) = 1$, for $j = 1, 2, ..., p$. These parameter values will produce a valid tail distribution of the ES model and the ES forecasts.

There are four null hypotheses for the ES backtests based on the method by Couperier and Leymarie (2019) to assess different settings that the regression coefficients should meet for valid ES forecasts. These hypotheses $H_{0,J_1}, H_{0,J_2}, H_{0,I}, H_{0,S}$ are given by

$$H_{0,J_1} : \sum_{j=1}^{p} \big(\beta_0(u_j) + \beta_1(u_j)\big) = p, \tag{2.32}$$

$$H_{0,J_2} : \sum_{j=1}^{p} \beta_0(u_j) = 0, \text{ and } \sum_{j=1}^{p} \beta_1(u_j) = p \tag{2.33}$$

$$H_{0,I} : \sum_{j=1}^{p} \beta_0(u_j) = 0, \tag{2.34}$$

$$H_{0,S} : \sum_{j=1}^{p} \beta_1(u_j) = p, \tag{2.35}$$

where J_1 and J_2 denote joint backtests, I refers to the intercept backtest, and S indicates the slope backtest.

The null hypotheses of the joint backtests, H_{0,J_1} and H_{0,J_2}, focus on the expected value of the intercept $\beta_0(u_j)$ and the slope parameter $\beta_1(u_j)$ for $j = 1, 2, .., p$. While H_{0,J_1} sums both $\beta_0(u_j)$ and $\beta_1(u_j)$, H_{0,J_2} takes separate sums of the coefficients. The intercept backtest and slope backtests hypotheses, $H_{0,I}$ and $H_{0,S}$ respectively, complement the joint backtests to make easier the identification of misspecification. Rejection of $H_{0,I}$ will suggest constant forecasting errors across time, while when $H_{0,S}$ is rejected, this indicates the presence of time-varying errors, since they fluctuate along with the VaR prediction.

In terms of test statistics, suppose that $W, W \in J_1, J_2, I, S$ is the generic notation for test statistics and consider the classical formulation of a Wald-type test such as $H_{0,W} : R_W\beta = q_W$, then Couperier and Leymarie (2019) shows that the test statistics can be generally expressed as follows:

$$W = T\big(R_W\hat{\beta} - q_W\big)'\big(R_W\hat{\Sigma}R_W'\big)^{-1}\big(R_W\hat{\beta} - q_W\big). \tag{2.36}$$

where T denotes the sample size and $\hat{\Sigma}$ represents a consistent estimator of the asymptotic covariance matrix. The test statistics asymptotically follow a χ^2 distribution with one degree of freedom for J_1, I, S and two degrees of freedom for J_2. Since the coefficients are aggregated, the asymptotic distributions are based on a small and fixed number of degrees of freedom at any value of p. Therefore, Couperier and Leymarie (2019) argue that the critical values of the four backtests are unchanged regardless of the degree of accuracy used to approximate the ES.

The ES-MQR tests in this study adapt the companion Matlab code, which is available to download from http://www.runmycode.org/companion/view/3358.

2.3.6 Quantile loss function

The main weakness of VaR backtests is that they show very little power to distinguish between different, yet close, alternatives. Therefore, in addition to testing VaR violations, measuring the size of uncovered losses is also of paramount importance. To calculate the uncovered losses, Lopez (1999) proposes the use of a loss function, which is based on the distance between the observed returns and the VaR forecasts in lieu of a hypothesis testing approach based on statistical tests. The distance represents the losses that have not been covered.

One of the most commonly used loss functions is the tick loss function of Giacomini and Komunjer (2005), often referred to as quantile score. This can be done through a criterion function that is minimized in a quantile regression estimation (Koenker & Bassett Jr, 1978). As explained by Gneiting (2011), quantiles are elicitable, since the scoring function is strictly consistent. Therefore, the accuracy of the VaR forecast can be compared using the following quantile loss function.

$$S_\alpha(Q_t, r_t) = \sum_{t=n+1}^{n+m} (r_t - Q_t)(\alpha - I(r_t < Q_t)) \tag{2.37}$$

where $Q_{t+1}, ..., Q_{n+m}$ is a series of quantile forecasts. True $\text{VaR}_{\alpha,t}$ minimizes $E\left[S_\alpha(\text{VaR}_{\alpha,t}, r_t) \mid \mathcal{F}_{t-1}\right]$. Therefore, the most accurate VaR forecasting model should minimize the scoring function in Equation (2.37).

2.3.7 VaR and ES joint loss function

The ES is not an elicitable risk measure (Gneiting, 2011). Therefore, in general, there is no loss function minimized by the true ES. Nevertheless, Fissler and Ziegel (2016) show that ES is elicitable of higher order in the sense that VaR and ES are jointly elicitable. They develop a strictly consistent function that is minimized by the true VaR and ES:

$$\begin{aligned} S_t(r_t, VaR_t, ES_t) = {} & (I_t - \alpha)G_1(VaR_t) - I_t G_1(r_t) \\ & + G_2(ES_t)\left(ES_t - VaR_t + \frac{I_t}{\alpha}(VaR_t - r_t)\right) \\ & - H(ES_t) + a(r_t), \end{aligned} \tag{2.38}$$

where $I_t = 1$ if $r_t < VaR_t$ and 0 otherwise for $t = 1, ..., T$, $G_1()$ is increasing, $G_2()$ is strictly increasing and strictly convex, $G_2 = H'$ and $\lim_{x \to -\infty} G_2(x) = 0$ and a$(\cdot)$ is a real valued integrable function.

This study employs two special joint loss functions. The first is based on the recommendation of Fissler and Ziegel (2016) which has been used in the study of Wang et al. (2019), while the second is based on Taylor (2019). The first joint loss function considers $G_1(x) = x, G_2(x) = \exp(x), H(x) = \exp(x)$ and $a(r_t) = 1 - \log(1 - \alpha)$ which gives the following

scoring function:

$$\begin{aligned} S_t(r_t, VaR_t, ES_t) = (I_t - \alpha)(VaR_t) - I_t(r_t) \\ + \exp(ES_t)\Big(ES_t - VaR_t + \frac{I_t}{\alpha}(VaR_t - r_t)\Big) \\ - \exp(ES_t) + 1 - \log(1-\alpha), \end{aligned} \tag{2.39}$$

where the loss function $S = \sum_{t-1}^{T} S_t$ is strictly consistent and is jointly minimized by the true VaR and ES series.

The second joint loss function is a method for predicting ES corresponding to VaR forecasts using the equivalence between quantile regression and maximum likelihood based on an asymmetric Laplace (AL) density (Taylor, 2019). In this framework, a joint model of conditional VaR and ES is estimated by maximizing the AL log-likelihood. The scoring function proposed by Taylor (2019) is based on the scoring function of Fissler and Ziegel (2016) in Equation 2.38 by considering $G_1 = 0, G_2 = -1/x, H(x) = -\log(-x)$ and $a(r_t) = 1 - \log(1-\alpha)$. The function is the negative of the AL log-likelihood, referred to as the AL log score, which takes the following form:

$$S_t(r_t, VaR_t, ES_t) = -\log\left(\frac{\alpha - 1}{ES_t}\right) - \frac{(r_t - VaR_t)(\alpha - I(r_t \leq VaR_t))}{\alpha ES_t}. \tag{2.40}$$

A strictly consistent scoring function can be used as the loss function in model comparison (Gneiting & Raftery, 2007). The scoring function in Equation 2.40 is strictly consistent and the average score across a sample indicates a joint measure of the accuracy of the VaR and ES forecast.

2.3.8 Model confidence set

The AL log score is then used as the loss function in the calculation of the Model Confidence Set (MCS). The MCS, developed by Hansen et al. (2011), consists of a sequence of tests used to construct a set of 'superior' models (SSM) for which the null hypothesis of equal predictive ability (EPA) is not rejected. The MCS procedure will produce a set of models with a given confidence level.

The MCS test procedure described in Hansen et al. (2011) starts from a collection of a finite number of competing objects denoted $\mathcal{M}^0$ with a given probability and indexed by $i = 1, ..., m_0$. The evaluation of the object is based on the loss $L_{i,t}$, associated with the object i in the period t, where $t = 1, ..., n$. The relative performance variable is defined as

$$d_{ij,t} = L_{i,t} - L_{j,t}, \quad \text{for all } i, j \in \mathcal{M}^0. \tag{2.41}$$

Then, a sequence of significance tests finds the best set of models $\mathcal{M}^*$ identified as:

$$\mathcal{M}^* \equiv \{i \in \mathcal{M}^0 : \mu_{i,j} \leq 0 \quad \text{for all} \quad j \in \mathcal{M}^0\}. \tag{2.42}$$

where $\mu_{i,j} = E(d_{ij,t})$ is the expected loss difference between each pair of models. It is assumed to be finite and does not depend on t for all $i, j \in \mathcal{M}^0$. The objects are ranked in terms of expected loss. Object i is more superior to object j if $\mu_{ij} < 0$.

The hypothesis of the MCS test is as follows.

$$H_{0,\mathcal{M}} : \mu_{ij} = 0, \quad \text{for all } i, j \in \mathcal{M}, \tag{2.43}$$

where $\mathcal{M} \subset \mathcal{M}^0$. The alternative hypothesis $H_{\mathcal{A},\mathcal{M}}$ is $\mu_{ij} \neq 0$ for some $i, j \in \mathcal{M}$.

An equivalence test ($\delta_{\mathcal{M}}$) and an elimination rule ($e_{\mathcal{M}}$) are the basis of the MCS procedure. The purpose of the equivalence test is to test the null hypothesis $H_{0,\mathcal{M}}$ for any $\mathcal{M} \subset \mathcal{M}^0$. When $H_{0,\mathcal{M}}$ is rejected, the elimination rule is used to identify the object of $\mathcal{M}$ that should be eliminated from $\mathcal{M}$.

To test the hypothesis, the following $t-$statistics are constructed by Hansen et al. (2011).

$$t_{ij} = \frac{\bar{d_{ij}}}{\sqrt{\widehat{\text{var}}(\bar{d_{ij}})}} \tag{2.44}$$

$$\mathcal{T}_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} \mid t_{ij} \mid, \tag{2.45}$$

where $\bar{d_{ij}} \equiv n^{-1}\Sigma_{t-1}^{n} d_{ij,t}$ and $\widehat{\text{var}}(\bar{d_{ij}})$ denotes the estimate of $\text{var}(\bar{d_{ij}})$.

These test statistics do not have standard asymptotic distributions because those distributions depend on nuisance parameters. To address this problem, the MCS procedure employs bootstrap methods to estimate relevant distributions.

The construction of the set of superior models based on the MCS algorithm is as follows.

Step 1: Set $\mathcal{M} = \mathcal{M}^0$ that includes all objects.

Step 2: Test the null hypothesis $H_{0,\mathcal{M}}$ using the equivalence test at a given significance level α.

Step 3: If $H_{0,\mathcal{M}}$ is not rejected, then $\widehat{\mathcal{M}^*_{1-\alpha}} = \mathcal{M}$. If $H_{0,\mathcal{M}}$ is rejected, use the elimination rule to remove an object from $\mathcal{M}$.

Step 4: Repeat the procedure from *Step 1*.

The remaining surviving objects are denoted $\widehat{\mathcal{M}^*_{1-\alpha}}$, which Hansen et al. (2011) refers to as the *model confidence set*.

The MCS test is performed using two methods, that is, the R method and the SQ method. The first uses the sums of absolute values in the calculation of the MCS test statistic, while the latter uses the summed squares. For details, see Hansen et al. (2011, p. 465). The MCS

approach is available in the Matlab MFE toolbox of Kevin Sheppard, and the associated Matlab code can be downloaded from https://www.kevinsheppard.com/code/matlab/mfe-toolbox/.

2.4 Chapter Summary

This chapter presents a review of the distribution of financial asset returns relevant to tail-risk forecasting methods used in the following chapters of this thesis. The theoretical review on VaR and ES is further discussed in relation to the financial asset return distributions. A wide range of tail risk forecasting methods used in this thesis is also presented in a detailed manner. The following three chapters present novel approaches to risk forecasting by using various extensions of the Realized EGARCH models using high-frequency data estimated in a Bayesian setting.

Chapter 3

Tail Risk Forecasting Using Bayesian Realized EGARCH Model

3.1 Introduction

Forecasting the volatility of financial assets plays a crucial role in risk management and has been a focus of attention for both researchers and market practitioners. It is evident that asset returns exhibit time-varying volatility, usually have clustered volatility, and tend to respond asymmetrically towards positive and negative shocks. Therefore, modeling such dynamics is of paramount importance for risk forecasting purposes. The seminal works of Engle (1982) on Autoregressive Conditional Heteroskedasticity (ARCH) models and Bollerslev (1986) on the Generalized ARCH (GARCH) model have pioneered the development of various extensions of the models. GARCH-type models have been applied in a wide variety of fields requiring volatility forecasts, including in forecasting the risk of financial assets.

Most of the models utilize daily returns in modeling the conditional variance since the squared returns can provide an unbiased estimate of the current level of volatility that can be used to form expectations for future periods. However, daily returns provide weak signals about the current level of volatility (Hansen et al., 2012). As a consequence, GARCH models tend to be inefficient in reflecting true volatility during highly volatile periods, since they over-react in responding to large shocks and take a long time to revert back to the long-term mean (Andersen et al., 2003).

Following the availability of high-frequency data, more efficient realized measures of volatility have been proposed, such as realized variance (Andersen & Bollerslev, 1998; Andersen et al., 2003), realized range (Christensen & Podolskij, 2007; Martens & van Dijk, 2007), and realized kernel (Barndorff–Nielsen et al., 2008; Barndorff–Nielsen & Shephard, 2002). These realized measures provide more information than the daily return and accurate proxies of volatility compared to squared returns, since they incorporate the tick-by-tick intra-day process. Therefore, the realized measures have been increasingly used in recent studies in volatility modeling.

Hansen et al. (2012) took advantage of the realized measures by introducing a new model that provides joint modeling for the returns and realized volatility measures, namely the

Realized GARCH model. This model adds a measurement equation to the GARCH model, allowing joint dependence between volatility and a realized volatility measure. The Realized GARCH model is extended by Hansen and Huang (2016) via the Realized EGARCH model that allows the model to incorporate multiple realized volatility measures, hence the model has more than one measurement equation.

This chapter builds on the original Realized EGARCH model by Hansen and Huang (2016) and extensions of the Realized GARCH models by Watanabe (2012) and Contino and Gerlach (2017). The focus of this chapter is to forecast tail risks (VaR and ES) based on the extension of the Realized EGARCH model. The extension of the realized EGARCH is conducted in several ways. First, two types of return distributions, that is, a standardized Student-*t* and a standardized skewed Student-*t* distribution, are proposed, to take into account the non-Gaussianity in the error distribution. The measurement equation errors are kept to follow a normal distribution. The two frameworks are called RE-SkN and RE-tN. Second, three types of realized measures are utilized in the realized EGARCH model, either individually or jointly. Those are the subsampled realized variance (RVSS), subsampled realized range (RRSS), and realized kernel (RK). The subsampling process is based on Zhang et al. (2005) to deal with micro-structure effects. The choice of RVSS and RRSS is motivated by the recent work of Gerlach and Wang (2016) and Wang et al. (2019) who find that RVSS and RRSS could improve the out-of-sample forecasting of their proposed models. Meanwhile, the RK is chosen because it has some robustness to market micro-structure effects (Barndorff–Nielsen et al., 2008) and has been used consistently in the original realized GARCH and realized EGARCH models. Third, an adaptive MCMC procedure is adopted and applied to the proposed models. To evaluate the performance of the proposed RE-SkN and RE-tN models, their 2.5% and 1% (following the recommendation of the Basel Committee 2012, 2016, 2019) VaR and ES forecasts are compared to those of various competing models, such as the original realized GARCH and realized EGARCH models, the Conditional Autoregressive Expectile (CARE) model by Taylor (2008), and several extensions of GARCH-type models.

After this chapter was completed and the whole thesis was near completion, Chen et al. (2021) publish a study that employs the Bayesian method to forecast VaR and ES through the realized EGARCH and realized HAR (heterogeneous autoregressive) GARCH models. They also employ the skewed Student *t* distribution and also apply adaptive Bayesian MCMC sampling. The research however only employ one realized measure, that is realized kernel of the S&P500. Their research findings indicate that the HAR-GARCH model using the skewed Student *t* distribution performs better than other realized GARCH models in terms of violation rate and expected shortfall. However, in terms of forecasting volatility, the realized EGARCH model is found to have the best performance.

This chapter is structured as follows. Section 3.2 briefly discusses three types of realized volatility measures used in this study. Section 3.3 presents the specification of the proposed extension of the Realized EGARCH model. The Bayesian estimation of the proposed models is discussed in Section 3.5. Section 3.6 presents the result of the simulation study of the

Realized EGARCH model using a standardized skewed Student-t distribution for the observation equation and a Gaussian distribution for the measurement equation based on the maximum likelihood frequentist estimation and the proposed Bayesian estimation. In Section 3.7, the empirical study is presented, including the description of the data, the parameter estimates of the proposed models, and the evaluation of the tail risk forecasts. Finally, Section 3.8 concludes this capture and discusses possible future studies.

3.2 Realized Volatility Measures

This section discusses three types of realized measures employed in this study, that is, realized variance (RV), realized range (RR), and realized kernel (RK). For day t, let H_t, L_t, and C_t denote the intra-day high, low, and closing prices, respectively. The daily close-to-close log return is calculated as

$$r_t = \log(C_t) - \log(C_{t-1}),$$

where r_t^2 is the volatility estimate.

In a high-frequency intraday framework, each day t from open to close can be divided into N intervals of equal size with length $\triangle$, where each intraday time is subscripted as $i = 0, 1, 2, ..., N$. Let $P_{t-1+i\triangle}$ denote the closing price in the i-th interval of day t. RV is calculated as the sum of the N intra-day squared returns, at frequency $\triangle$, for day t:

$$RV_t^{\triangle} = \sum_{i=1}^{N} \left[\log(P_{t-1+i\triangle}) - \log(P_{t-1+(i-1)\triangle}) \right]^2. \tag{3.1}$$

The RR, proposed by Martens and van Dijk (2007) and Christensen and Podolskij (2007), may contain more information about volatility than the RV, in the same way that the intra-day range contains more information than squared returns, that is, it uses all the price movements in a time period, not just the price at the start and end of each sub-period. Let $H_{t,i} = \sup_{(i-1)\triangle<j<i\triangle} P_{t-1+j}$, and $L_{t,i} = \inf_{(i-1)\triangle<j<i\triangle} P_{t-1+j}$ denote the high and low prices during the i-th interval of day t, respectively; the RR is simply the sum of the squared intraperiod ranges:

$$RR_t^{\triangle} = \frac{\sum_{i=1}^{N} (\log H_{t,i} - \log L_{t,i})^2}{4 \log 2}. \tag{3.2}$$

In an ideal situation where prices are continuously observed and without measurement error, both the RV and the RR will generate a perfect estimate of volatility. Unfortunately, the RV suffers from a well-known bias problem due to market micro-structure noise. The bias tends to worsen as the sampling frequency of intra-day returns increases (Hansen & Lunde, 2006) and is evident in volatility signature plots (Andersen et al., 2000; Fang, 1996).

However, Martens and van Dijk (2007) show that the RR is more competitive and sometimes more efficient than the RV, under some micro-structure conditions and levels. As $N \to \infty$, the scaling factor of $4\log 2$ makes the RR an approximately unbiased estimator. Even though N is finite for empirical data and Equation 3.2 is biased, when employed in the realized GARCH and realized EGARCH models, the measurement equation parameters can adjust for such bias. Therefore, the RR is not required to be unbiased in the realized GARCH and the realized EGARCH models. Gerlach and Wang (2016) proved that employing the RR as the measurement equation variable in a realized GARCH model can improve predictive likelihoods, accuracy, and efficiency in forecasting empirical tail risks.

Different approaches can be adopted to reduce the market micro-structure noise in the RV and RR, such as scaling and subsampling approaches. The scaling process was proposed by Martens and van Dijk (2007) as follows:

$$RV_{S,t}^{\triangle} = \frac{\sum_{l=1}^{q} RV_{t-l}}{\sum_{l=1}^{q} RV_{t-l}^{\triangle}} RV_t^{\triangle}, \tag{3.3}$$

$$RR_{S,t}^{\triangle} = \frac{\sum_{l=1}^{q} RR_{t-l}}{\sum_{l=1}^{q} RR_{t-l}^{\triangle}} RR_t^{\triangle}, \tag{3.4}$$

where RV_t and RR_t represent the daily return square and the range square on day t, respectively, and q is the number of days employed to estimate the scaling factors.

The motivation behind the scaling process is that micro-structure noise has less effect on the daily return and range than on their high frequency counterparts. Therefore, the scaling factor can help reduce the bias in the RV and RR via the scaling and smoothing processes.

Meanwhile the subsampling approach was proposed by Zhang et al. (2005), by which a number of sub grids is selected from the original grid of observation times and then the average of estimators derived from the sub grids is calculated. For day t, N equally sized samples are grouped into M non-overlapping subsets $\Theta^{(m)}$ with size $N/M = n_k$, which means:

$$\Theta = \bigcup_{m=1}^{M} \Theta^{(m)}, \text{ where } \Theta^{(k)} \cap \Theta^{(l)} = \emptyset, \text{ when } k \neq l.$$

The subsampling process is then implemented on the subsets $\Theta^{(i)}$ with n_k interval:

$$\Theta^{(i)} = i, i + n_k, ..., i + n_k(M-2), i + n_k(M-1), \text{ where } i = 0, 1, 2..., n_k - 1.$$

Wang et al. (2019) apply the subsampling procedures on the RV and RR. Let $C_{t,i} = P_{t-1+i\triangle}$

denote the log closing price at the i-th interval of day t, the RV with the subsets Θ^i is calculated as:

$$RV_i = \sum_{m=1}^{M} (C_{t,i+n_k m} - C_{t,i+n_k(m-1)})^2; \text{ where } i = 0, 1, 2..., n_k - 1.$$

Suppose there are T minutes per trading day, the T/M RV with T/N subsampling is given by:

$$RV_{T/M,T/N} = \frac{\sum_{i=0}^{n_k-1} RV_i}{n_k}. \tag{3.5}$$

Let $H_{t,i} = \sup_{(i+n_k(m-1))\triangle<j<(i+n_k m)\triangle} P_{t-1+j}$ and $L_{t,i} = \inf_{(i+n_k(m-1))\triangle<j<(i+n_k m)\triangle} P_{t-1+j}$ denote the high and low prices during the interval $i + n_k(m-1)$ and $i + n_k m$, respectively. Wang et al. (2019) propose the T/M RR with T/N subsampling that takes the following form:

$$RR_i = \sum_{m=1}^{M} (H_{t,i} - L_{t,i})^2; \text{ where } i = 0, 1, 2..., n_k - 1. \tag{3.6}$$

$$RR_{T/M,T/N} = \frac{\sum_{i=0}^{n_k-1} RR_i}{4 \log 2 n_k}. \tag{3.7}$$

For details regarding the properties of the subsampled RR, please refer to Wang et al. (2019).

The selection of sampling frequency needs to take into account a trade-off between bias and variance, for example, see Bandi and Russell (2008) and Zhang et al. (2005). The RV and RR are usually computed from intra-day returns sampled at a moderate frequency, such as 5-minute frequency, to balance the bias and variance. This study employs 5-minute RV and RR with 1 min subsampling proposed by Wang et al. (2019), as follows:

$$\begin{aligned}
&RV_{5,1,0} = (\log C_{t5} - \log C_{t0})^2 + (\log C_{t10} - \log C_{t5})^2 + ... \\
&RV_{5,1,1} = (\log C_{t6} - \log C_{t1})^2 + (\log C_{t11} - \log C_{t6})^2 + ... \\
&\vdots \\
&RV_{5,1} = \frac{\sum_{i=0}^{4} RV_{5,1,i}}{5}, \\
&RR_{5,1,0} = (\log H_{t0\le t\le t5} - \log L_{t0\le t\le t5})^2 + (\log H_{t5\le t\le t10} - \log L_{t5\le t\le t10})^2 + ... \\
&RR_{5,1,1} = (\log H_{t1\le t\le t6} - \log L_{t1\le t\le t6})^2 + (\log H_{t6\le t\le t11} - \log L_{t6\le t\le t11})^2 + ... \\
&\vdots RR_{5,1} = \frac{\sum_{i=0}^{4} RR_{5,1,i}}{4 \log(2) 5}.
\end{aligned} \tag{3.8}$$

The third realized measure proposed for this study is the realized kernel (RK), which is

somewhat robust to noise. The RK was developed by Barndorff–Nielsen et al. (2008) Barndorff–Nielsen et al. (2009) and was also employed by the original realized GARCH and realized EGARCH models. Their non-negative estimators take the following form:

$$K(X) = \sum_{h=-H}^{H} k\Big(\frac{h}{H+1}\Big)\gamma_h, \text{ where } \gamma_h = \sum_{j=|h|+1}^{n} x_{j,t}x_{j-|h|,t}. \tag{3.9}$$

The kernel weight function, $k(x)$, in Equation 3.9, is chosen to be the Parzen weight function, since it satisfies the smoothness conditions, $k'(0) = k'(1) = 0$. This will ensure a non-negative estimate. The parzen kernel function is defined as follows:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2, \\ 2(1-x)^3 & 1/2 \leq x \leq 1, \\ 0 & x > 1, \end{cases} \tag{3.10}$$

where x_j is the high-frequency return in Equations (3.9) and (3.10), H is a standard bandwidth spelt out in Barndorff–Nielsen et al. (2009), given by:

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

For details, please see Barndorff–Nielsen et al. (2009).

3.3 The Specification of the Proposed Realized EGARCH Models with the Skewed Student-*t* Distribution

This section describes the Realized EGARCH model and the proposed variants of the model.

3.3.1 Model description

The realized EGARCH model of Hansen and Huang (2016) with K realized measures takes the following form:

$$\begin{aligned} r_t &= \mu_t + \sqrt{h_t}\epsilon_t \\ \log h_t &= \omega + \beta \log h_{t-1} + \tau(\epsilon_{t-1}) + \gamma' u_{t-1} \\ \log x_{k,t} &= \xi_k + \varphi_k \log h_t + \delta_{(k)}(\epsilon_t) + u_{k,t}, \quad k = 1, ..., K. \end{aligned} \tag{3.11}$$

where r_t is the daily return, $x_t = (x_{1,t}, ..., x_{K,t})$ is a vector of realized measures, the conditional mean $\mu_t = E(r_t|\mathcal{F}_{t-1})$, conditional variance $h_t = \text{var}(r_t|\mathcal{F}_{t-1})$, $\{\mathcal{F}_t\}$ is a filtration so that (r_t, x_t) is adapted to $\mathcal{F}_t$, $u_t = (u_{1,t}, ..., u_{K,t})'$, $\epsilon_t \overset{iid}{\sim} D_1(0,1)$, $u_t \overset{iid}{\sim} D_2(0,\Sigma)$, ϵ_t and $u_t = (u_{1,t}, ..., u_{K,t})'$ are mutually and serially independent.

The leverage functions $\tau(\epsilon)$ and $\delta_k(\epsilon)$, $k = 1, ...K$, allow for a more realistic independence between ϵ_t and u_t in practice. The leverage functions can be expressed as $\tau(\epsilon_t) = \tau' a(\epsilon_t)$ and $\delta_{(k)}(\epsilon_t) = \delta'_k b(\epsilon_t)$, $k = 1, ..., K$, where $a(\epsilon_t)$ and $b(\epsilon_t)$ are known functions of ϵ_t. Modeling the dependence between return shocks and volatility shocks, which is empirically important, is facilitated by these leverage functions. The volatility shock, $v_t = E(\log h_{t+1}|\mathcal{F}_t) - E(\log h_{t+1}|\mathcal{F}_{t-1})$, is given by $v_t = \tau(\epsilon_t) + \gamma' u_t$.

Following Hansen and Huang (2016), Hermite polynomials are adopted in a simple quadratic form for leverage functions.

$$\tau(\epsilon) = \tau_1\epsilon + \tau_2(\epsilon^2 - 1) \tag{3.12}$$

$$\delta_{(k)}(\epsilon) = \delta_{k,1}\epsilon + \delta_{k,2}(\epsilon^2 - 1), \quad k = 1..., K. \tag{3.13}$$

The realized EGARCH model in Equation 3.11 consists of three equations. The first equation is the return equation, which is standard in GARCH models. The second equation is the GARCH equation and the third equation is the measurement equation(s). There are two main features of the GARCH equation in the realized EGARCH model. The first feature is the presence of a leverage function $\tau(\epsilon_{t-1})$, which distinguishes the realized EGARCH model from the realized GARCH model, which only includes a leverage function in the measurement equation. The second key feature is the last term $\gamma' u_{t-1}$, which facilitates channeling of the impact of the realized measure on the expectations of future volatility. Since u_t is K-dimensional, it allows us to incorporate multiple realized measures of volatility. The parameter β represents the persistence of volatility, while γ reflects the information the realized measures contain about future volatility. The measurement equation shows the relationship between the (*ex-post*) realized measures of volatility and the (*ex-ante*) conditional variance. Hansen and Huang (2016) use RK, the daily range, and six types of realized variances and adopt a Gaussian specification, $D_1(0, 1) = N(0, 1$ and $D_2(0, \Sigma) = N(0, \Sigma)$.

This study extends the realized EGARCH model by allowing $D_1(0, 1)$ to be a standardized Student-*t*, and a standardized skewed Student-*t* of Hansen (1994), following Watanabe (2012) and Contino and Gerlach (2017) for the case of the realized GARCH. These proposed models are denoted hereafter as RE-tN and RE-SkN, respectively, while the original realized EGARCH model employing normal distribution for both return and measurement equations is denoted as RE-NN.

3.3.2 The skewed Student-*t* distribution

The Student-*t* distribution could be potentially restrictive, since it only allows for variation in location, scale, and tail fatness. Meanwhile, the skewed Student-*t* distribution is more flexible and allows for richer behaviors. It is able to model not only leptokurtosis, but also

asymmetry (Hansen, 1994). The PDF of a standardized skewed Student-t distribution is:

$$(\varepsilon_t|\nu,\lambda) = \begin{cases} bc\left(1+\frac{1}{\nu-2}\left(\frac{b\varepsilon_t+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} & if \quad \varepsilon_t < -\frac{a}{b} \\ bc\left(1+\frac{1}{\nu-2}\left(\frac{b\varepsilon_t+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} & if \quad \varepsilon_t \geq -\frac{a}{b} \end{cases} \tag{3.14}$$

where ν is the degrees of freedom parameter that controls the kurtosis of the distribution with range $2 < \nu < \infty$, λ is the skewness parameter with range $-1 < \lambda < 1$, while a, b, and c are some constants given by:

$$a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right), \quad b = \sqrt{1+3\lambda^2-a^2}, \quad c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}.$$

Hansen (1994) show that this is a proper density function that has zero mean and a unit variance and specializes to a Student-t density by setting $\lambda = 0$. When $\lambda > 0$, the mode of the density is to the left of zero and the variable is right-skewed, and *vice versa* when $\lambda < 0$.

3.3.3 Stationarity condition

It is important to find the model parameters to ensure that h_t has a finite expected value or higher moments. The stationarity condition of the realized EGARCH model in Equation 3.11 can be derived by subtituting the measurement equation into the GARCH equation:

$$\log h_t = \omega - \gamma'\xi_k + (\beta - \gamma'\varphi_k)\log h_{t-1} + a_t, \tag{3.15}$$

where $a_t = \tau(\epsilon_{t-1}) + \gamma'(\log x_{k,t-1} - \delta_{(k)}(\epsilon_t))$.

Taking the expectation of both sides of Equation 3.15, the long-run value of $\log h$ is derived as:

$$E(\log h) = \frac{\omega - \gamma'\xi_k}{1-(\beta-\gamma'\varphi_k)}. \tag{3.16}$$

The stationarity of the realized EGARCH model in Equation 3.11 is then achieved by imposing the following restriction:

$$\beta - \gamma'\varphi_k < 1 \tag{3.17}$$

3.4 VaR and ES for the Standardized Skewed Student-*t* Distribution

For the skewed Student-*t* distribution, the VaR estimation is conducted using the inverse CDF, $skt^{-1}_{\nu,\lambda}$ as specified in Contino and Gerlach (2017) as follows:

$$VaR_\alpha = \mu + \sigma skt^{-1}_{v,\lambda}, \tag{3.18}$$

$$skt^{-1}(\alpha \mid v, \lambda) = \begin{cases} \frac{1-\lambda}{b}\sqrt{\frac{v-2}{v}}t_v^{-1}\left(\frac{\alpha}{1-\lambda}, v\right) - \frac{a}{b}, & \text{if } \alpha < \frac{1-\lambda}{2} \\ \frac{1+\lambda}{b}\sqrt{\frac{v-2}{v}}t_v^{-1}\left(0.5 + \frac{\alpha}{1+\lambda}\left(\alpha - \frac{1-\lambda}{2}\right), v\right) - \frac{a}{b}, & \text{if } \alpha \geq \frac{1-\lambda}{2}. \end{cases}$$

where ν is the degrees of freedom parameter that controls the kurtosis of the distribution with range $2 < \nu < \infty$, λ is the skewness parameter with range $-1 < \lambda < 1$, while a, b, and c are some constants given by:

$$a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right), \quad b = \sqrt{1+3\lambda^2-a^2}, \quad c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}.$$

For the standardized skewed Student-*t* distribution, Contino and Gerlach (2017) derived the ES, given the standardized VaR value Θ, as follows:

$$ES_\alpha = \mu + \sigma E[\epsilon \mid \epsilon < \Theta], \tag{3.19}$$

where

$$\begin{aligned} E[\epsilon \mid < \Theta] &= \frac{1}{skt_{v,\lambda}(\Theta)}\int_{-\infty}^{\Theta} bc\left(1 + \frac{1}{v-2}\left(\frac{b\epsilon+a}{1-\lambda}\right)^2\right)^{-\frac{v+1}{2}} \epsilon d\epsilon \\ &= \frac{c(1-\lambda)^2}{b \cdot skt_{v,\lambda}(\Theta)}\frac{v-2}{1-v}\left(1 + \frac{1}{v-2}\left(\frac{b\Theta+a}{1-\lambda}\right)^2\right)^{\frac{1-v}{2}} - \frac{a}{b}, \end{aligned} \tag{3.20}$$

where $skt_{v,\lambda}$ is the CDF of the Skewed Student-*t* distribution:

$$skt(\epsilon \mid v, \lambda) = \begin{cases} (1-\lambda)t_v\left(\sqrt{\frac{v}{v-2}}\left(\frac{b\epsilon+a}{1-\lambda}\right), v\right) & \text{if } \epsilon < -\frac{a}{b} \\ \frac{1-\lambda}{2}(1+\lambda)\left[t_v\left(\sqrt{\frac{v}{v-2}}\left(\frac{b\epsilon+a}{1-\lambda}\right), v\right) - 0.5\right] & \text{if } \epsilon \geq -\frac{a}{b}. \end{cases}$$

3.5 Bayesian Estimation of the Realized EGARCH Model

This section discusses the Bayesian estimation approach and MCMC sampling procedures employed in estimating the proposed model and conducting the tail risk forecasting. Basically, the Bayesian model is a probability model consisting of a likelihood function $p(y|\theta)$ and a prior distribution $\pi(\theta)$. The product of the likelihood and the prior probability is equivalent to the posterior distribution $p(\theta|y)$.

3.5.1 Likelihood and model parameters

Following Hansen and Huang (2016), where $\epsilon_t \overset{iid}{\sim} D_1(0,1)$ and $u_t \overset{iid}{\sim} D_2(0,\Sigma)$, the log-likelihood function for the Realized EGARCH model in Equation 3.12 is given by:

$$\begin{aligned}\ell(r,x;\theta,\Sigma) = -\frac{1}{2}\sum_{t=1}^{n}\Big[&\log(2\pi) + \log h_t + \epsilon_t^2 + K\log(2\pi) \\ &+ \log(|\Sigma|) + u_t'\Sigma^{-1}u_t\Big]\end{aligned} \tag{3.21}$$

where $\epsilon_t = \epsilon_t(\theta) = (r_t - \mu)/\sqrt{h_t}$ and $u_{k,t}(\theta) = \tilde{x}_{k,t} - \xi_k - \varphi_k\tilde{h}_t(\theta) - \delta_k' b(\epsilon_t(\theta))$.

Hansen and Huang (2016) used the score function and the numerical Hessian matrix of the log likelihood function to compute the standard errors for the estimators. This model, hereafter, is denoted RE-NN (Realized EGARCH with Normal-Normal errors) model.

Under the proposed choice $D_1 \sim Skt_{\nu,\lambda}(0,1)$ and $D_2 = N(0,\Sigma)$ as in Equation 3.14, hereafter referred to as RE-SkN model (Realized EGARCH with Skewed Student-t-Normal errors), the log-likelihood function for model (3.12) takes the following form:

$$\ell(r;\theta) = \begin{cases} T[\ell(b)+\ell(c)] - \sum_{t=1}^{T}\left[\frac{v+1}{2}\log\left(1+\frac{1}{v-2}\left(\frac{b\epsilon_t+a}{1-\lambda}\right)^2\right) + 0.5\log(h_t)\right] & \text{if } \epsilon_t < -\frac{a}{b}, \\ T[\ell(b)+\ell(c)] - \sum_{t=1}^{T}\left[\frac{v+1}{2}\log\left(1+\frac{1}{v-2}\left(\frac{b\epsilon_t+a}{1+\lambda}\right)^2\right) + 0.5\log(h_t)\right] & \text{if } \epsilon_t \geq -\frac{a}{b}. \end{cases}$$

$$\ell(x;\theta,\Sigma) = -\frac{1}{2}\sum_{t=1}^{n}\Big[K\log(2\pi) + \log(|\Sigma|) + u_t'\Sigma^{-1}u_t\Big]$$

$$\ell(r,x;\theta,\Sigma) = \ell(r;\theta) + \ell(x;\theta,\Sigma). \tag{3.22}$$

Meanwhile, the log-likelihood function for model (3.11) with distribution choice $D_1 \sim t_\nu(0,1)$ and $D_2 = N(0,\Sigma)$ is:

$$\begin{aligned}\ell(r,x;\theta,\Sigma) = T\Big[&\log\Gamma\Big(\frac{v+1}{2}\Big) - \log\Gamma\Big(\frac{v}{2}\Big) - 0.5\log[(v-2)\pi]\Big] \\ &- \sum_{t=1}^{T}\left[\frac{v+1}{2}\log\left(1+\frac{(r_t-\mu)^2}{(v-2)h_t}\right) + 0.5\log h_t\right] \\ &- \frac{1}{2}\sum_{t=1}^{n}\Big[K\log(2\pi) + \log(|\Sigma|) + u_t'\Sigma^{-1}u_t\Big].\end{aligned} \tag{3.23}$$

This model is subsequently denoted as RE-tN model (Realized EGARCH with Student-t-Normal errors).

The parameters for those realized EGARCH models are presented in Table 3.1.

TABLE 3.1: Realized EGARCH models and parameters

Model	Parameters
RE-NN	$\mu, \omega, \beta, \tau_1, \tau_2, \gamma^{'}, \xi_k, \varphi_k, \delta_{k,1}, \delta_{k,2}, \Sigma$
RE-SkN	$\mu, \omega, \beta, \tau_1, \tau_2, \gamma^{'}, \xi_k, \varphi_k, \delta_{k,1}, \delta_{k,2}, \Sigma, \nu_{skw}, \lambda$
RE-tN	$\mu, \omega, \beta, \tau_1, \tau_2, \gamma^{'}, \xi_k, \varphi_k, \delta_{k,1}, \delta_{k,2}, \Sigma, \nu$

Joint Conditional Likelihood is given by:

$$\mathcal{L}(\theta, \Sigma, r, x) = \prod_{t=1}^{T} p(r_t, x_t | \mathcal{F}_{t-1}, \theta, \Sigma) \tag{3.24}$$

$$= \prod_{t=1}^{T} p(r_t | \mathcal{F}_{t-1}, \theta) p(x_t | r_t, \mathcal{F}_{t-1}, \theta, \Sigma) \tag{3.25}$$

where $\mathcal{F}_t = (r_1, ..., r_t; x_1, ..., x_t)'$; the predictive density $p(r_t \mid \mathcal{F}_{t-1}, \theta)$ is determined by the distributional choice of ϵ_t; and the conditional density $p(x_t | r_t, \mathcal{F}_{t-1}, \theta, \Sigma)$ is determined by the distribution of u_t.

3.5.2 Prior specification

A combination of mostly uninformative and Jeffreys-type priors is chosen over the possible region for the parameters.

$$\pi(\theta) \propto \frac{1}{\sigma^2_{u,k=1}} \frac{1}{\sigma^2_{u,k=2}} .. \frac{1}{\sigma^2_{u,k=K}} \frac{1}{\nu^2} I(A) \tag{3.26}$$

where A is the region described by the parameter restriction in Table 3.2. The indicator function $I(A) = 1$ over the region A and zero otherwise.

TABLE 3.2: Parameter restriction

Parameter	Bounds
$\mu, \omega, \tau_1, \tau_2, \xi, \delta_1, \delta_2$	$\pm\infty$
σ^2_u	> 0
ν	$(4, 200)$
λ	$(-1, 1)$
$\beta - \gamma^{'} \varphi$	< 1

A standard Jeffreys prior is adopted for the variance of u_t in Equation 3.11. The prior of the degrees of freedom parameter (ν) is ν^{-2}, which is an improper prior obtained by being flat on ν^{-1} (Bauwens et al., 2006). This is equivalent to $U(0, \frac{1}{4})$ so that $\nu > 4$. This ensures that the first four moments of the distribution of u_t are finite (Chen et al., 2006). Similar research has adopted this type of prior on σ^2_u and ν on Bayesian estimation of the previous model, the

realized GARCH models; for example, see Gerlach and Wang (2016), Contino and Gerlach (2017), and Wang et al. (2019).

3.5.3 Adaptive MCMC method

In order to numerically perform an inference on model parameters, an MCMC procedure is performed to draw samples from the posterior distribution of the model parameters. An adaptive MCMC method of Contino and Gerlach (2017) is adopted and extended, considering the balance between the target acceptance rate and convergence, the scaling factor for the tuning process, and the selection of the block of parameters in the updating process. The adaptive MCMC approach consists of two steps, a burn-in period and a post-burn-in period.

In the burn-in period, a Robust Adaptive Metropolis (RAM) algorithm of Vihola (2012) is employed. The RAM algorithm estimates the shape of the target distribution and simultaneously allows us to attain a given mean acceptance rate. RAM is defined recursively through the following process:

1. Compute the proposal $Y_n := X_{n-1} + S_{n-1}U_n$, where S_{n-1} is a lower diagonal matrix with positive diagonal elements and U_n is a vector of random variable from d-dimensional Gaussian distribution.
2. The proposal is accepted with probability $\alpha_n := \min\left\{1, \frac{\pi(Y_n)}{\pi(X_{n-1})}\right\}$.
3. If the proposal Y_n is accepted, set $X_n := Y_n$. Otherwise, set $X_n := X_{n-1}$.
4. Compute the lower diagonal matrix, S_n, satisfying the equation:

$$S_n S_n^T = S_{n-1}\left(I + \eta_n(\alpha_n - \alpha^*)\frac{U_n U_n^T}{||U_n||^2}S_{n-1}^T\right), \tag{3.27}$$

   where $\{\eta_n\}_{n\geq 1} \subset (0,1]$ is a step size sequence decaying to zero, often defined as $\eta_n = n^{-\gamma}$ with an exponent $\gamma \in (1/2, 1)$, $\alpha^* \in (0,1)$ is the target mean acceptance probability, and $I \in \mathbb{R}^{dxd}$ is the identity matrix.

The initial proposal covariance is set to $(2.3^2/d_i)I_{d_i}$ (Roberts & Rosenthal, 2009), where d_i denotes the dimension of the parameter block i and I_{d_i} is the identity matrix of dimension d_i. To obtain S_n in step (4), rank-one Cholesky update of S_{n-1} is conducted when $(\alpha_n - \alpha^*) > 0$, or downdate of S_{n-1} when $(\alpha_n - \alpha^*) < 0$. This approach is computationally more efficient than computing the right hand side of Equation 3.27 and performing new Cholesky decomposition. Following Vihola (2012), $\eta_n = \min\{1, d \cdot n^{-2/3}\}$, where d is the dimension of the block of parameters. The target mean acceptance rate $\alpha^* = 0.234$ (if $d > 4$), or 0.35 (if $2 \leq d \leq 4$), or 0.44 (if $d = 1$), as recommended by Roberts et al. (1997). For theoretical details of the RAM algorithm, see Vihola (2012).

The algorithm is run for 20,000 iterations for each model parameter to achieve convergence. The first half of burn-in iterates (10,000 iterates) is discarded, then the remaining second half is used to calculate the sample mean (X_n) and covariance matrix (Σ) for each parameter as the proposal distribution in the post burn-in period. In the post burn-in period, a standard

RWM algorithm is employed and a mixture of three Gaussian proposal distribution is used as follows:

$$Y_n \sim \mathcal{N}\left(X_{n-1}, C_i\Sigma\right), \tag{3.28}$$

where $C_1 = 1; C_2 = 100, C_3 = 0.01$, following Wang et al. (2019).

The algorithm is run for 10,000 iterations for each model parameter.

The convergence of the adaptive MCMC method is assessed via the Gelman-Rubin diagnostic, effective sample size testing (Gelman et al., 2014) and the autocorrelation time (Chib & Jeliazkov, 2001) as has been discussed in Section 1.6. To determine the setting of the parameter blocks, two settings of the parameter blocks are compared: 1 block (as in Contino and Gerlach (2017)) and 4 blocks. The latter setting is adapted from Wang et al. (2019) and adjusted to the models proposed in this chapter as follows: $\theta_1 = (\mu)$, $\theta_2 = (\omega, \beta, \tau', \gamma', \varphi_k)$, $\theta_3 = (\xi_k, \delta_k', \Sigma)$, and $\theta_4 = (\nu, \lambda)$. The parameters within the same parameter block are more likely to be correlated in the posterior (or likelihood) than between blocks. For example, the stationarity condition causes a correlation between iterates of $\beta, \gamma', \varphi_k$. Therefore, those parameters are grouped in the same block.

3.5.4 Tail risk forecasting

The proposed models are estimated using the proposed MCMC procedures with fixed size in-sample data combined with a rolling window approach to produce 1000 one-step ahead forecasts. The post burn-in iterates in the above MCMC draws are used to generate the variance (denoted σ^2 in Section 2.2 or h in Section 3.3.1). This will, in turn, form MCMC iterates for VaR and ES forecasts at 2.5% and 1%, respectively, based on Equations 3.18 and 3.19. The posterior mean estimates of the VaR and ES are then used as the final forecast. The accuracy of each model can then be measured by its relative forecast performance.

3.6 Simulation Study

A simulation study is conducted to compare the performance of MCMC estimators and the maximum likelihood (ML) estimators. The purpose is to demonstrate the superior performance of the MCMC estimators in terms of unbiasedness and precision properties.

3.6.1 Simulation set up and model

A total of 1000 replicated datasets are simulated from the proposed RE-SkN model in Equation 3.29 using one realized measure, with a sample size of 2000.

$$\begin{aligned} r_t &= 0.0 + \sqrt{h_t}\epsilon_t \\ \log h_t &= -0.12 + 0.98 \log h_{t-1} - 0.12\epsilon_{t-1} + 0.04(\epsilon_{t-1}^2 - 1) + 0.47u_{t-1} \\ \log x_t &= -0.17 + 0.94 \log h_t - 0.09\epsilon_t + 0.06(\epsilon_t^2 - 1) + u_t, \end{aligned} \tag{3.29}$$

where $\epsilon_t \sim Skt_{\nu=4.4,\lambda=0.5}$, $u_t \sim N(0, 0.15)$, and $h_0 = 0.0025$.

All initial parameter values were arbitrarily set equal to 0.1, except for $\nu = 5$, for both estimation methods. Tail risks are considered for the quantile level $\alpha = 2.5\%$ and $\alpha = 1\%$, following the recommendation of the Basel Committee (2012, 2016, 2019).

At the early stage, the simulation was run on multiple computers. To gain time efficiency, the final simulation is run on the University of Sydney's Ártemis High-Performance Computing (HPC). The simulation job is divided into 40 chunks to reduce processing time, each of which estimates the model in Equation 3.29 and forecasts the tail risk for time $t + 1$ for 25 datasets. All 40 processing jobs are run simultaneously in Artemis, each of which takes approximately 9.5 hours to complete.

3.6.2 MCMC convergence and efficiency

To assess the convergence of the MCMC method, $m = 10$ independent MCMC chains with different starting values are run for two different parameter block settings. Then, the variance between and within the chains is analyzed according to the Gelman-Rubin diagnostic (Gelman et al., 2014). Table 3.3 presents the summary of the Gelman-Rubin diagnostic $\hat{R}$, the number of effective sample size n_{eff} (Gelman et al., 2014), and the autocorrelation time $\hat{t}$ (Chib & Jeliazkov, 2001) for the settings of 1 parameter block and 4 parameter blocks. Favorable results are indicated by the values printed in bold. The $\hat{R}$ for each parameter is very close to 1 and all $\hat{R}$ are < 1.1, the recommended threshold by Gelman et al. (2014), except for ν in the 1 block setting. The $\hat{R}$ under the 4 block settings is closer to 1 for 11 of 13 parameters, suggesting an improved convergence property under this parameter block setting.

The result also suggests that the number of effective sample sizes for most parameters, n_{eff}, is higher than the suggested benchmark ($5m$) by Gelman et al. (2014). The number of effective sample size for 10 parameters under the 4 block setting is higher than that under the 1 block setting. Regarding the correlation time $\hat{t}$, the parameters under the 4 block setting also have a more favorable property, that is, they have a smaller autocorrelation time. Overall, the MCMC method under the 4 block setting provides a much improved convergence property, is more efficient, and has a smaller autocorrelation time than the 1 block setting.

TABLE 3.3: Summary statistics of the Gelman-Rubin diagnostic, effective sample size and autocorrelation time of the RE-SkN model estimated using Bayesian method for the simulated dataset

Parameters	**1 Block**			**4 Blocks**		
	$\hat{R}$	n_{eff}	$\hat{t}$	$\hat{R}$	n_{eff}	$\hat{t}$
μ	**1.0103**	1289.4	**34.3**	1.0104	**1324.3**	36.2
ω	1.0194	1224.9	38.6	**1.0053**	**1404.7**	**34.1**
β	1.0115	1415.3	33.8	**1.0043**	**1456.3**	**33.0**
γ	1.0124	1014.8	38.0	**1.0052**	**1567.7**	**30.9**
τ_1	**1.0080**	**1468.4**	**31.3**	1.0117	1424.6	33.5
τ_2	1.0399	275.9	70.2	**1.0074**	**1579.2**	**30.6**
ξ	1.0316	565.9	83.1	**1.0062**	**1380.2**	**35.2**
φ	1.0299	509.4	88.2	**1.0075**	**1373.3**	**35.3**
δ_1	1.0082	**1609.9**	**29.6**	**1.0052**	1213.2	33.4
δ_2	1.0761	111.0	436.2	**1.0080**	**1450.8**	**33.2**
σ_u^2	1.0094	**1551.6**	**27.9**	**1.0047**	1366.5	35.5
ν	1.3026	37.2	1341.3	**1.0097**	**1297.1**	**37.6**
λ	1.0165	901.9	33.2	**1.0095**	**1461.1**	**32.5**

Convergence can also be visually assessed by using the trace plots of the MCMC iterations. Trace plots of $\mu, \beta, \gamma, \tau_1$, and λ during the burn-in in Figure 3.1 show a very rapid convergence of the algorithm, in less than 5000 iterations. The stationarity of those parameters is also shown during the post burn-in periods, as indicated by the trace plots in Figure 3.2. More details on the trace plots of 10 independent chains for each parameter can be found in Appendix A.1. The trace plots of the MCMC chains show rapid convergence and chain mixing during the burn-in period. The acceptance rates for the 4 blocks during the burn-in period are 0.4303, 2120, 0.3426, and 0.3477, very close to the target acceptance rate.

FIGURE 3.1: RAM iterations with simulated data from the RE-SkN model

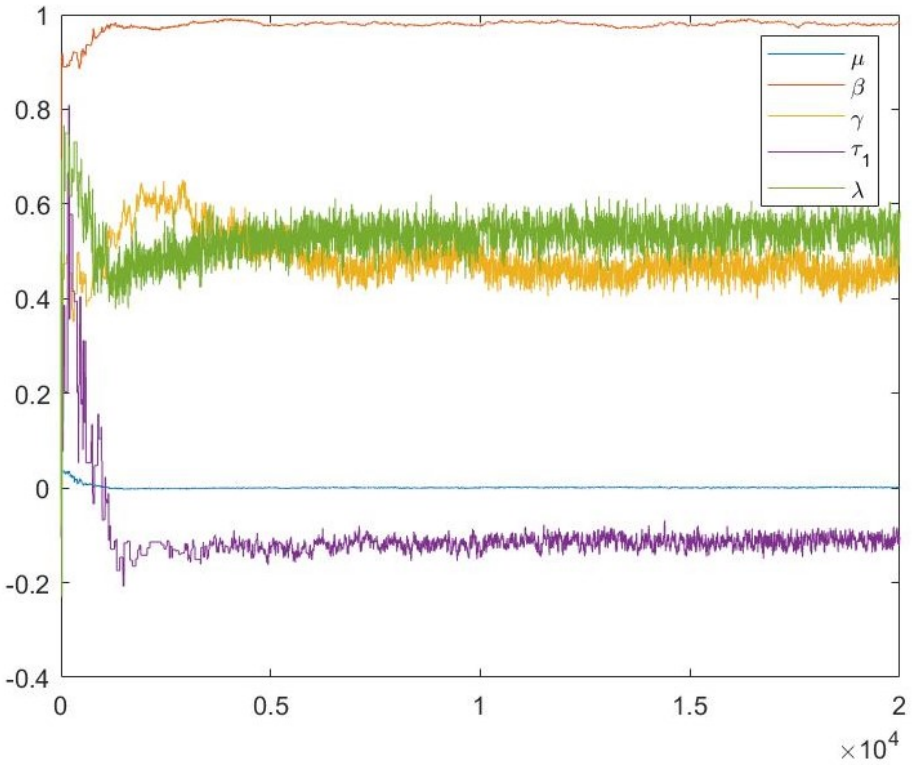
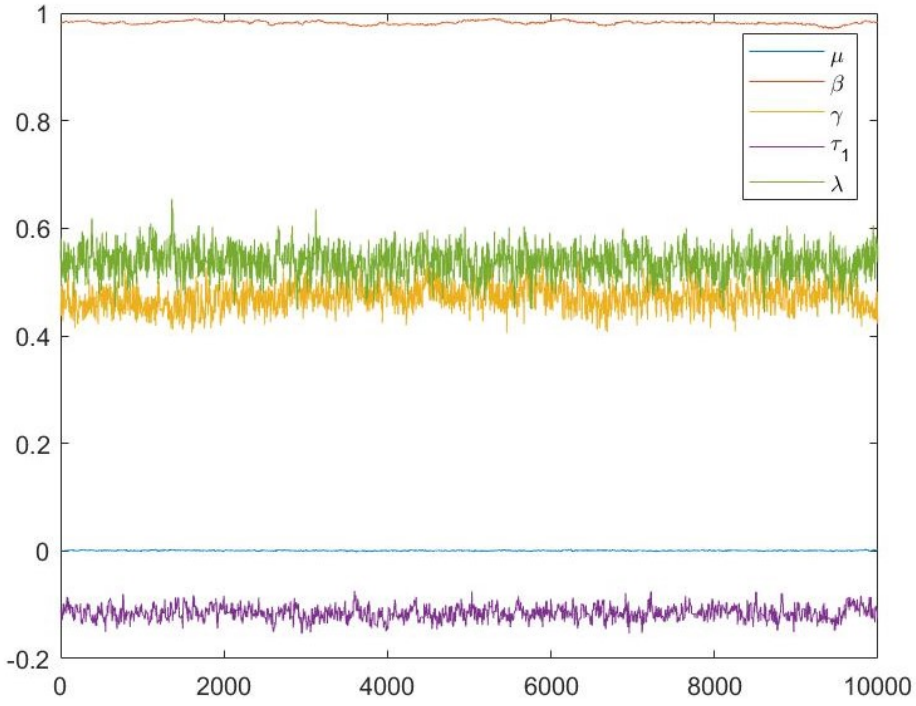


FIGURE 3.2: RWM iterations with simulated data from the RE-SkN model



3.6.3 Simulation results

The estimation results of the proposed RE-SkN model based on the ML method and the MCMC method using the 4 parameter block setting are summarized in Table 3.4. The posterior means of the parameters and the root mean square errors (RMSE) are also presented. The optimal parameter coefficients and the RMSE are printed in bold.

The RMSE for parameter i is calculated as:

$$RMSE_i = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\tilde{\theta}_{n,i} - \theta_i)^2} \tag{3.30}$$

where N is the number of datasets, $\tilde{\theta}$ is the estimated parameter, and θ is the true parameter.

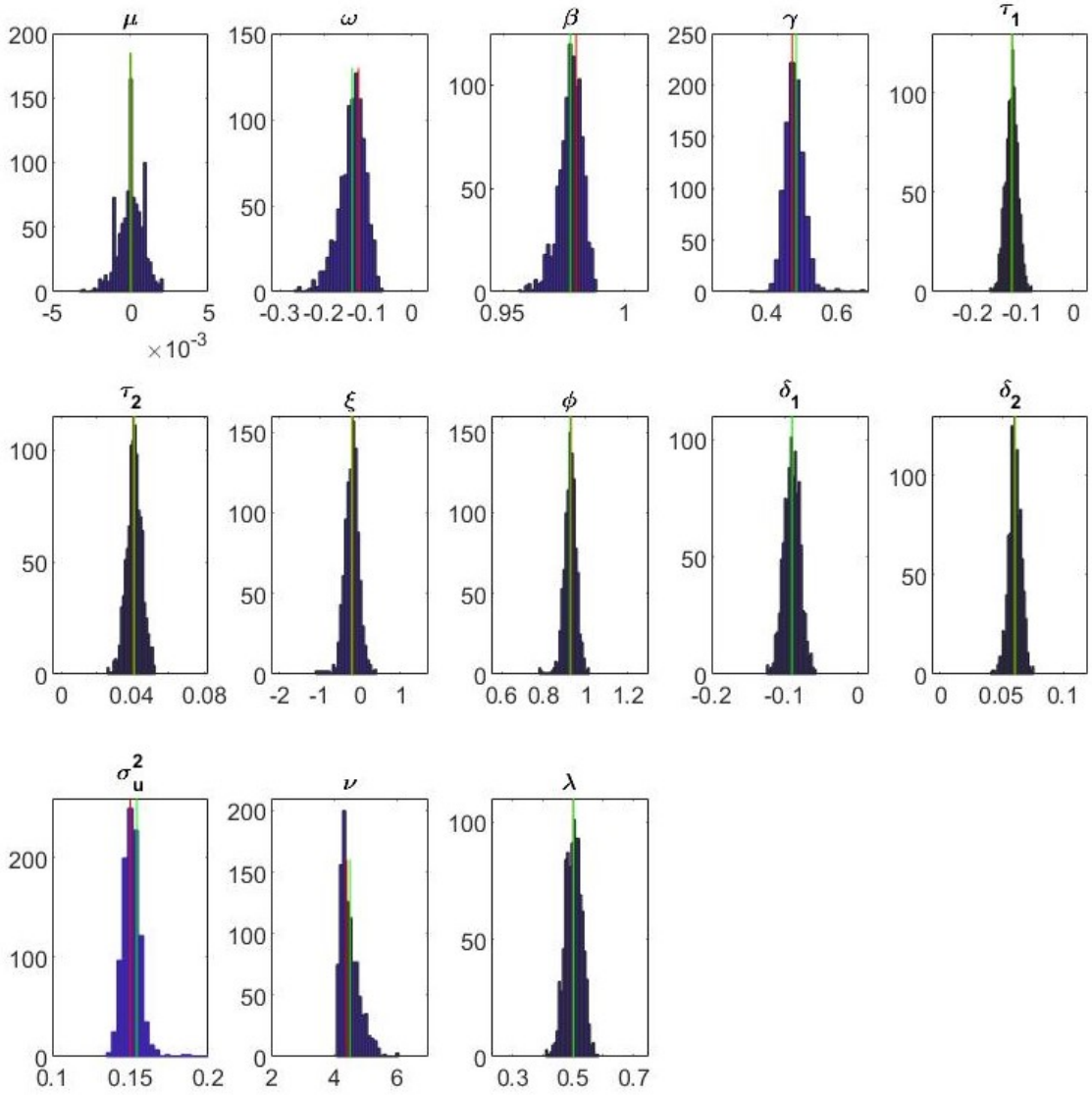
TABLE 3.4: Summary statistics of the RE-SkN Model parameters based on ML and MCMC Methods

n=2000	**True**	**ML**		**MCMC**	
Parameter		**Mean**	**RMSE**	**Mean**	**RMSE**
μ	0.00	0.0028	0.0176	**0.0001**	**0.0002**
ω	-0.12	-0.1407	0.1208	**-0.1342**	**0.0329**
β	0.98	0.9579	0.1114	**0.9777**	**0.0053**
γ	0.47	0.4951	0.1658	**0.4815**	**0.0531**
τ_1	-0.12	-0.1133	0.0665	**-0.1210**	**0.0125**
τ_2	0.04	0.0512	0.0606	**0.0407**	**0.0042**
ξ	-0.17	-0.0896	0.6982	**-0.2002**	**0.1884**
φ	0.93	0.9452	0.1742	**0.9263**	**0.0267**
δ_1	-0.09	-0.0826	0.0587	**-0.0901**	**0.0115**
δ_2	0.06	0.0786	0.1266	**0.0608**	**0.0050**
σ_u^2	0.15	0.2154	0.4433	**0.1542**	**0.0386**
ν	4.4	**4.3858**	0.4132	4.5075	**0.3685**
λ	0.5	0.4895	0.0962	**0.5014**	**0.0256**
2.5% VaR_{t+1}	-0.0763	-0.0857	0.1169	**-0.0768**	**0.0037**
2.5% ES_{t+1}	-0.0948	-0.1062	0.1459	**-0.0952**	**0.0055**
1% VaR_{t+1}	-0.0916	-0.1047	0.1472	**-0.0921**	**0.0049**
1% ES_{t+1}	-0.1130	-0.1291	0.1817	**-0.1133**	**0.0073**

For each dataset, the true one-step-ahead VaR and ES forecasts are calculated and averaged over the 1000 data sets. The average is given in the 'True' column. Both the ML estimators and MCMC estimators generate accurate estimates for the parameters and the VaR and ES forecasts, being very close to the true values. However, the MCMC method produces more favorable estimators, both in terms of unbiasedness and precision. All parameters and tail risk forecasts under the MCMC framework, except the parameter ν, have a smaller bias than those under the ML method. Regarding estimation precision, all MCMC estimators have smaller root mean square errors (RMSE) than their respective ML counterparts. The

histogram of the posterior mean of each parameter of the proposed RE-SkN model over the 1000 datasets is shown in Figure 3.3.

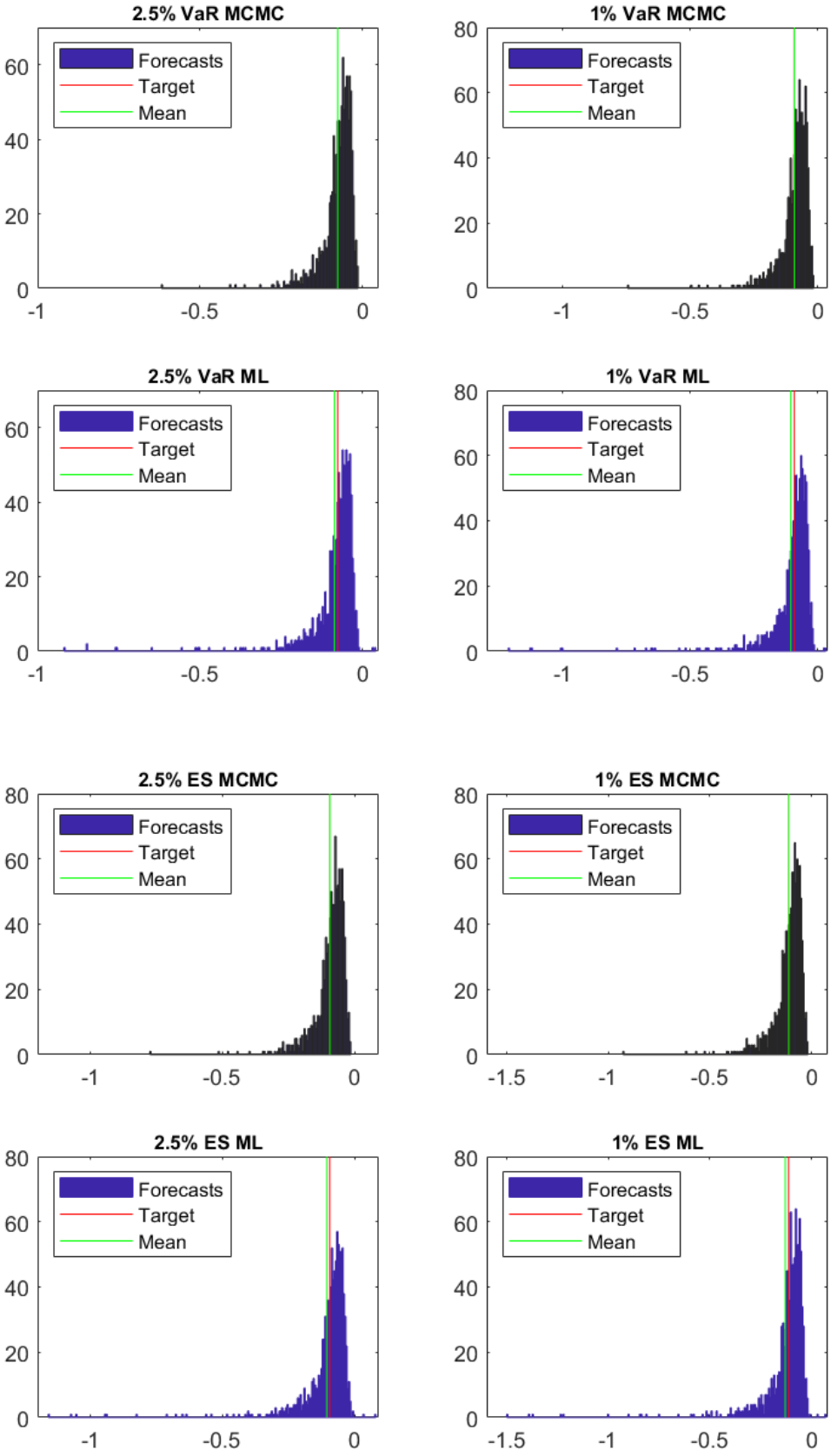
FIGURE 3.3: Histogram of 1000 Posterior means for the RE-SkN model



Note: Red line indicates true parameter and the green line indicates the mean estimate.

The performance of the MCMC and ML estimators in terms of unbiasedness and precision of the risk forecasts is also illustrated using the distribution of 1000 2.5% and 1% VaR and ES forecasts in Figure 3.4. Although both the MCMC and ML methods generate VaR and ES forecasts relatively close to the true value of VaR and ES (represented by the red horizontal lines), it can be seen that the ML method produces more outlying forecasts than the MCMC method. The mean values of both VaR and ES forecasts generated using the MCMC method are also very close to their respective true values, compared to those of the ML method. This might be due to the convergence issue in the ML method, which results in a higher RMSE of the risk forecasts. This suggests that the MCMC method is preferable compared to the ML method, as it produces more accurate risk forecasts.

FIGURE 3.4: Histogram of 1000 2.5% and 1% VaR and ES true and forecasts: MCMC and ML estimation



3.7 Empirical Study

This section presents the empirical results of the proposed models, the RE-tN and RE-SkN types models, estimated using the MCMC approach using returns and three types of realized measures (individual or combined) specified in Section 3.2. The risk forecast performance of the models is evaluated based on various tests discussed in Section 2.3 and compared with the original realized EGARCH model estimated also within a Bayesian setting (the RE-NN type models) .

Several GARCH family type models, all with Student-t distribution, are also included for comparison purposes, such as the GARCH, EGARCH, and GJR-GARCH models. A GARCH-EVT-t model is also included to produce conditional quantile and ES estimates by applying the peaks over threshold of the extreme value theory (EVT) method to the standardized residuals of the GARCH model with Student-t distribution. Following Taylor (2008), the threshold is 10% unconditional quantile for the lower tail. For details on GARCH-EVT-t models, see McNeil and Frey (2000) and Section 7.2.3 in McNeil et al. (2005). Comparison models also include a filtered GARCH approach (GARCH-HS-t) used in Wang et al. (2019), where a standardized VaR and ES are estimated based on a GARCH with Student-t distribution (GARCH-t). Returns $(r_1, .., r_n)$ are standardized by using their estimated conditional standard deviation, $rt/\sqrt{\hat{h}_t}$. The standardized VaR and ES are then multiplied by the forecast $\sqrt{\hat{h}_{n+1}}$ from the GARCH-t model.

The VaR and ES forecasts are also generated based on the symmetric absolute value conditional autoregressive expectiles (CARE-SAV) model of Taylor (2008) and the realized GARCH models of Hansen et al. (2011) with a Student-t distribution for the observation equation errors for the return (RG-tN). The GARCH-type models are estimated using the Econometrics toolbox in Matlab, while the GARCH-EVT-t, CARE-SAV, realized GARCH, and realized EGARCH models are estimated using the Matlab code developed for this thesis. The parameters of the models are used to generate 1000 one-step-ahead VaR and ES forecasts.

3.7.1 Data

This study utilizes the data of daily, including open, high, low, and closing prices, as well as high frequency data, observed at 5-minute and 1-minute intervals within trading hours. The daily return is calculated based on the close-to-close prices, including overnight price movements. The subsampled RV (RVSS) and subsampled RR (RRSS) used in this study are based on the dataset used in Wang et al. (2019), in which the daily RV and RR are calculated based on the 5-minute data, and the 1-minute data are used to generate daily RVSS and RRSS. Meanwhile, the data for the realized kernel using a Parzen weight function are obtained from the Oxford-Man Institute's realized library, Library Version: 0.3, which also originates from Thomson Reuters Tick History.

The data are collected for seven market indices: ASX 200 (Australia), DAX (Germany), FTSE 100 (UK), Hang Seng (Hong Kong), NASDAQ (US), SMI (Swiss), and S&P 500 (US), covering

the period April 2000 to June 2016. The summary statistics for each data series are shown in Table 3.5. All return series are leptokurtic, with kurtosis higher than 3, and they all exhibit negative skewness. The result of the standard test for normality (Jarque & Bera, 1987) shows that none of the return series is normally distributed, as JB statistics are all higher than the 5.99, the 5% critical JB test value.

TABLE 3.5: Descriptive statistics for each return series

Indices	Begin	End	Obs	Mean	Std	Skewness	Kurtosis	Min	Max	JB
ASX 200	13/07/2000	16/06/2016	3966	0.0103	1.011	-0.412	7.654	-7.713	5.628	3690.5
DAX	5/04/2000	17/06/2016	4066	0.0085	1.530	-0.044	6.544	-7.434	10.685	2129.1
FTSE 100	6/04/2000	17/06/2016	4081	-0.0005	1.226	-0.161	9.107	-9.266	9.384	6358.9
HangSeng	11/04/2000	17/06/2016	3937	0.0034	1.531	-0.079	11.375	-13.582	13.407	11511.0
NASDAQ	14/04/2000	17/06/2016	4001	0.0036	1.619	-0.064	9.597	-12.452	13.255	7258.0
SMI	4/07/2005	17/06/2016	2730	0.0076	1.185	-0.563	14.804	-12.735	10.788	15992.0
SP500	11/04/2000	17/06/2016	4018	0.0093	1.262	-0.316	11.897	-10.003	10.957	13320.0

Note: JB is the Jarque-Bera statistic to test the null hypothesis of normality.

3.7.2 Estimation of the Realized EGARCH type models

The estimation of the proposed Realized EGARCH model and the forecast for VaR and ES are run using the Artemis HPC. Given that the measurement error distribution is Normal (N) and there are three types of realized measures (subsampled RV (RVSS), subsampled RR (RRSS), and Realized Kernel (RK)) and three types of error distributions (Normal (N), standardized student-t (t), and standardized skewed student-t distributions (Sk)), the combination resulted in nine types of RE models to run for each market index. These types of distribution codes are used in several tables reporting the empirical results in the next parts of this thesis. For seven market indices, this gives us a total of 63 RE-type models to estimate. In the forecasting process, a fixed size in-sample data is combined with a rolling window approach to produce 1000 one-step ahead forecasts. Due to the high need for computation time, each model is re-estimated after moving forward 10 observations. Suppose that we have 2000 in-sample and 1000 forecast periods. Observations $t = 1$ up to $t = 2000$ are estimated to obtain $\tilde{\theta_1}$, which is then used to forecast the tail risk for t_{2001} to t_{2010}, observations $t = 11$ up to $t = 2010$ are estimated to obtain $\tilde{\theta_2}$, which is then used to forecast tail risk for t_{2011} to t_{2020}, etc.

The running time of each RE-type model differs according to the number of realized volatility measures included in the model and the type of error distributions. The RE-type model that includes only one realized volatility combined with a Gaussian distribution has the shortest running time. More complex models with more parameters to estimate take longer to complete. Please see Table 3.6 for the running time of various RE-type models for the S&P 500 index.

TABLE 3.6: The running time of the Realized EGARCH type models for S&P 500

Realized measures	**Error Distributions***	**Running time (hh:mm:ss)**
RRSS	NN	64:51:41
RRSS	tN	87:36:35
RRSS	SkN	85:59:00
RRSS-RK	NN	74:25:11
RRSS-RK	tN	106:43:30
RRSS-RK	SkN	102:12:49
RVSS-RRSS-RK	NN	77:13:37
RVSS-RRSS-RK	tN	100:34:56
RVSS-RRSS-RK	SkN	111:59:31

3.7.3 Model parameter estimates

Before discussing tail risk forecast evaluations, the posterior means of the parameters for all forecasting steps of the proposed RE-SkN and RE-tN type models are presented and discussed. The RE-NN type models are also presented for comparison. There are several tables presented in the following sections containing abbreviations of various model names. In each model name, RE refers to the realized EGARCH model, RV refers to the RVSS, RR refers to the RRSS, and RK refers to the Realized Kernel. The last term of the model name refers to the type of error distributions: N for the Normal distribution, t for the standardized Student-t distribution, and Sk for the standardized Skewed Student-t distribution.

The discussion in this section is based on the posterior mean of the parameters for the S&P 500 dataset presented in Table 3.7 and Table 3.8. In general, the parameter estimates are consistent with those in Table 3 of Hansen and Huang (2016). First, the γ parameter, which provides the channel for the impact of the realized measures on future volatility, ranges from 0.141 to 0.357 in the models with a single realized measure. Furthermore, the estimated γ in the models using the RVSS and RRSS are much greater (0.237 and 0.222 on average, respectively) than those using the RK (0.046 on average), indicating that RVSS and RRSS provide a stronger signal than the RK. The estimated value γ in models that employ RVSS and RRSS is several times larger than the typical value of α (about 0.05) in a conventional GARCH model, which measures the coefficient associated with squared returns. This suggests that the realized measures offer a stronger signal about future volatility than does the squared return.

Second, the estimates of β are close to 1 in all cases and $\beta - \gamma'\varphi_k$ are also relatively high, although they are less than unity, suggesting volatility persistence. The volatility is less persistent in the models employing the RVSS and RRSS, compared to those employing RK. For example, the average estimates of $\beta - \gamma'\varphi_k$ of the RE-RV-SkN and RE-RR-SkN models are 0.648 and 0.622, respectively, while that of the RE-RK-SkN model is 0.829.

TABLE 3.7: Estimated parameter mean for all forecasting steps for each parameter: S&P 500-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.0001	-0.322	0.965	0.319			-0.179	0.037	-0.934			0.967		
RE-RV-SkN	0.0000	-0.324	0.965	0.336			-0.186	0.039	-1.165			0.943		
RE-RV-tN	0.0003	-0.305	0.968	0.335			-0.183	0.038	-1.158			0.944		
RE-RR-NN	0.0000	-0.326	0.965		0.344		-0.177	0.039		-1.066			0.977	
RE-RR-SkN	0.0000	-0.327	0.965		0.357		-0.183	0.041		-1.227			0.961	
RE-RR-tN	0.0003	-0.313	0.967		0.356		-0.179	0.040		-1.210			0.963	
RE-RK-NN	0.0001	-0.264	0.972			0.141	-0.174	0.031			-0.538			1.003
RE-RK-SkN	0.0001	-0.272	0.971			0.144	-0.181	0.032			-0.694			0.987
RE-RK-tN	0.0003	-0.254	0.973			0.144	-0.178	0.032			-0.662			0.990
RE-RV-RR-NN	0.0001	-0.355	0.962	0.185	0.125		-0.162	0.042	-0.239	-0.354		1.041	1.054	
RE-RV-RR-SkN	0.0001	-0.348	0.963	0.198	0.133		-0.172	0.044	-0.707	-0.817		0.991	1.004	
RE-RV-RR-tN	0.0003	-0.334	0.965	0.199	0.128		-0.168	0.043	-0.587	-0.702		1.005	1.018	
RE-RV-RK-NN	0.0002	-0.330	0.965	0.256		0.011	-0.173	0.033	-0.220		-0.150	1.043		1.043
RE-RV-RK-SkN	0.0001	-0.339	0.964	0.271		0.013	-0.182	0.034	-0.539		-0.422	1.009		1.015
RE-RV-RK-tN	0.0004	-0.318	0.967	0.266		0.013	-0.177	0.034	-0.425		-0.337	1.022		1.024
RE-RR-RK-NN	0.0001	-0.339	0.964		0.282	0.011	-0.174	0.034		-0.408	-0.205		1.048	1.038
RE-RR-RK-SkN	0.0001	-0.347	0.963		0.295	0.012	-0.182	0.036		-0.700	-0.457		1.016	1.011
RE-RR-RK-tN	0.0003	-0.333	0.965		0.291	0.013	-0.178	0.036		-0.592	-0.360		1.028	1.021
RE-RV-RR-RK-NN	0.0002	-0.349	0.963	0.158	0.120	0.015	-0.167	0.039	-0.244	-0.360	-0.117	1.041	1.053	1.047
RE-RV-RR-RK-SkN	0.0002	-0.353	0.962	0.163	0.118	0.015	-0.169	0.039	-0.275	-0.407	-0.162	1.037	1.048	1.042
RE-RV-RR-RK-tN	0.0004	-0.337	0.965	0.161	0.117	0.016	-0.165	0.040	-0.246	-0.368	-0.138	1.042	1.053	1.045
Average	**0.0002**	**-0.323**	**0.966**	**0.237**	**0.222**	**0.046**	**-0.176**	**0.037**	**-0.562**	**-0.684**	**-0.353**	**1.007**	**1.019**	**1.022**

TABLE 3.8: Estimated parameter mean for all forecasting steps for each parameter: S&P 500-Part 2

Model	$\delta_{1,1}$	$\delta_{1,2}$	$\delta_{1,3}$	$\delta_{2,1}$	$\delta_{2,2}$	$\delta_{2,3}$	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RV-NN	-0.159			0.064			0.194					
RE-RV-SkN	-0.160			0.065			0.194				10.266	-0.167
RE-RV-tN	-0.156			0.064			0.194			9.594		
RE-RR-NN		-0.153			0.065			0.175				
RE-RR-SkN		-0.154			0.066			0.175			10.688	-0.165
RE-RR-tN		-0.150			0.066			0.175		10.187		
RE-RK-NN			-0.053			0.205			0.413			
RE-RK-SkN			-0.056			0.204			0.413		10.599	-0.166
RE-RK-tN			-0.042			0.205			0.413	9.750		
RE-RV-RR-NN	-0.158	-0.151		0.063	0.066		0.194	0.176				
RE-RV-RR-SkN	-0.159	-0.153		0.064	0.067		0.195	0.176			10.568	-0.162
RE-RV-RR-tN	-0.155	-0.147		0.064	0.067		0.195	0.176		9.766		
RE-RV-RK-NN	-0.158		-0.050	0.062		0.202	0.194		0.402			
RE-RV-RK-SkN	-0.159		-0.053	0.063		0.202	0.194		0.402		10.797	-0.161
RE-RV-RK-tN	-0.156		-0.041	0.063		0.203	0.194		0.401	10.093		
RE-RR-RK-NN		-0.152	-0.054		0.064	0.204		0.176	0.402			
RE-RR-RK-SkN		-0.153	-0.056		0.065	0.203		0.176	0.402		11.078	-0.160
RE-RR-RK-tN		-0.150	-0.043		0.065	0.204		0.176	0.401	10.370		
RE-RV-RR-RK-NN	-0.157	-0.150	-0.050	0.063	0.066	0.203	0.195	0.176	0.403			
RE-RV-RR-RK-SkN	-0.158	-0.151	-0.051	0.063	0.065	0.203	0.195	0.176	0.403		11.047	-0.157
RE-RV-RR-RK-tN	-0.154	-0.147	-0.038	0.063	0.066	0.204	0.195	0.175	0.403	9.762		
Average	**-0.157**	**-0.151**	**-0.049**	**0.064**	**0.066**	**0.203**	**0.194**	**0.176**	**0.405**	**9.932**	**10.720**	**-0.162**

Third, the leverage functions $(\tau_1, \tau_2, \delta_1, \delta_2)$ are of the expected sign: τ_1 and δ_1 are negative and τ_2 and δ_2 positive. The degree of asymmetry in the leverage effect is consistent across models, where the average estimate of τ_1 is -0.176 and τ_2 0.037. However, the estimated δ_1 and δ_2 tend to be smaller (more negative for δ_1) when using RVSS and RRSS than when using RK. The fact that $\delta_1 < 0$ indicates that x_t will be larger when $\epsilon_t < 0$ than when $\epsilon_t > 0$, which will make h_{t+1} larger when $\epsilon_t < 0$ through the variance equation if $\gamma > 0$. This result of leverage effects is consistent with the well-known stock market phenomenon that there is a negative correlation between today's return and tomorrow's volatility.

Fourth, the realized measures are not perfectly unbiased estimators of true volatility. For the realized measures to be unbiased, ξ and φ would be 0 and 1, respectively. The realized measures are measured from market open to close, so they may slightly underestimate volatility on average. Nevertheless, the parameters in the realized EGARCH models can adjust for the bias. The estimated φ is, in fact, close to 1 in all cases, suggesting that the realized measures are roughly proportional to the conditional variance minus a negative correction given by the estimate of ξ.

Fifth, the parameter σ_u^2 partially indicates how much error is contained in the realized measures, compared with $\log(h)$, but the log-scale makes this difficult to interpret. The models using the RRSS generate a smaller σ_u compared to those using the RVSS and RK, meaning

that the RRSS can potentially provide extra efficiency. This is consistent with the findings of Martens and van Dijk (2007), Christensen and Podolskij (2007), and Wang et al. (2019), which conclude that the realized range has lower mean squared errors than the realized variance, which may allow models employing the realized range to produce higher accuracy and efficiency in volatility estimation and forecasting.

Sixth, the estimates of ν are around 9 to 12, indicating that the choices of the Student-t and skewed Student-t return equation errors are justified, compared with the usual Gaussian distribution. Lastly, the estimated λ is negative, suggesting that the return is skewed to the left.

The posterior means of the parameters for every market index are presented in Appendix A.2. The discussion on the parameter result tables for other market indices is omitted to save space, but the findings are very similar. As can also be seen in Table 3.9, the average of the posterior mean of each parameter over all forecasting steps across models does not differ much across market indices, in terms of magnitude and sign. However, there are several exceptions. First, the average of the posterior mean of $\delta_{2,3}$ for ASX 200 is negative while those of other market indices are positive. In this case, $\delta_{2,3}$ refers to δ_2 for the measurement equation employing RK. Second, the average of the posterior mean of ξ_3 for ASX 200 and FTSE 100 is positive, while those of the other five market indices are negative. ξ_3 also belongs to the measurement equation employing RK. Lastly, the average of the posterior mean of ν and ν_{skw} for ASX 200 is comparatively higher than those of other market indices.

TABLE 3.9: The average posterior mean of all forecasting steps across models and market indices

Parameter	ASX 200	DAX	FTSE 100	Hang Seng	NASDAQ	SMI	SP500
μ	0.0003	0.0004	0.0000	0.0001	0.0003	0.0000	0.0002
ω	-0.198	-0.192	-0.227	-0.194	-0.338	-0.307	-0.323
β	0.979	0.978	0.975	0.978	0.962	0.967	0.966
γ_1	0.019	0.118	0.177	0.122	0.220	0.156	0.237
γ_2	0.212	0.135	0.193	0.208	0.231	0.296	0.222
γ_3	0.094	0.067	0.036	0.055	0.089	0.056	0.046
τ_1	-0.102	-0.124	-0.125	-0.047	-0.149	-0.114	-0.176
τ_2	0.028	0.039	0.043	0.048	0.043	0.030	0.037
ξ_1	-1.209	-0.079	-0.066	-0.939	-0.945	0.092	-0.562
ξ_2	-1.864	-0.092	-0.530	-1.554	-1.186	-0.123	-0.684
ξ_3	0.341	-0.053	0.011	-0.930	-0.983	-0.929	-0.353
φ_1	0.964	1.048	1.054	1.004	0.983	1.074	1.007
φ_2	0.944	1.051	1.023	0.956	0.987	1.043	1.019
ϕ_3	1.057	1.054	1.036	1.013	0.969	0.895	1.022
$\delta_{1,1}$	-0.087	-0.154	-0.141	-0.106	-0.157	-0.147	-0.157
$\delta_{1,2}$	-0.083	-0.140	-0.123	-0.091	-0.155	-0.123	-0.151
$\delta_{1,3}$	-0.161	-0.121	-0.073	-0.098	-0.135	-0.118	-0.049
$\delta_{2,1}$	0.125	0.084	0.073	0.058	0.042	0.051	0.064
$\delta_{2,2}$	0.107	0.100	0.072	0.041	0.039	0.044	0.066
$\delta_{2,3}$	-0.017	0.131	0.186	0.111	0.098	0.044	0.203
$\sigma^2_{u_1}$	0.342	0.209	0.158	0.203	0.194	0.222	0.194
$\sigma^2_{u_2}$	0.257	0.241	0.131	0.151	0.176	0.096	0.176
$\sigma^2_{u_3}$	0.631	0.315	0.291	0.475	0.351	0.817	0.405
ν	26.336	15.224	21.085	8.643	12.874	8.840	9.932
ν_{skw}	57.076	18.125	28.327	8.720	14.062	9.194	10.720
λ	-0.152	-0.135	-0.152	-0.061	-0.152	-0.141	-0.162

3.7.4 VaR and ES backtests

Prudent risk management requires accurate risk forecasts. On the one hand, financial regulators are usually concerned about how many times losses exceed the VaR (VaR violations) and the uncovered loss size. On the other hand, risk managers must maintain the balance between the safety goal and the profit maximization goal. An excessively high VaR will result in inefficiency, since financial institutions face large opportunity costs of capital due to reserving too much capital to buffer against market risks.

The VaR violation rates (VRates) for each model for all market indices are presented in Table 3.10 for the 2.5% forecasting and Table 3.11 for the 1% forecasting. Overall, the proposed RE-tN and RE-SkN type models generate more optimal VaR forecast series compared to other competing models, except for the CARE-SAV model for the 2.5% forecasting. The proposed models are consistently conservative, in the sense that the VRates are closer to the nominal VRates of 2.5% and 1%. Other competing models tend to be anti-conservative, generating

VaR forecasts much higher than 2.5% and 1%, except for the CARE-SAV and GARCH-HS-t models. The GARCH-EVT-t in particular produces too low VaR forecasts for the 2.5% level.

For the 2.5% VaR forecasting, the model with the average VRate closest to the nominal VRate of 2.5% is the CARE-SAV model (2.54%), followed by the RE-RR-tN model (2.59%), RE-RV-RR-tN model (2.60%), and the GARCH-HS-t model (2.60%). Individually, the CARE-SAV model produces the most accurate VRates for three of seven indices, while the RE-RV-SkN, RE-RR-tN, and RE-RK-SkN models produce the most accurate VRates for two of seven indices.

The favorable performance of the proposed RE-SkN and RE-tN type models is more distinctive for the 1% VaR forecast level. The best models generating average VRate closest to the nominal VRate of 1% are the RE-RR-tN and RE-RK-tN models (both 1%), followed by the RE-RR-RK-tN model (0.99%) and RE-RV-RR-RK-tN model (1.04%). Individually, the model that generates the most accurate VRates is the RE-RV-RR-RK-SkN model (four of seven indices), followed by the RE-RR-SkN and RE-RV-RR-RK-tN models (three of seven indices).

With respect to the number of realized measure employed, there is no clear pattern whether using multiple realized measures improves the 2.5% VaR forecast accuracy. However, a clear pattern can be observed for the 1% forecasting level, where models using multiple realized measures produce a VRate closer to 1% more frequently than those using a single realized measure. The result also suggests that the models utilizing the RRSS (individually or jointly with other realized measures) with the Student-*t* and Skewed Student-*t* distributions for observation equation errors tend to be more conservative and produce relatively accurate VaR violation rates.

TABLE 3.10: 2.5% VaR forecasting violation rate

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Average
GARCH-t	***3.9***	***4.0***	***4.2***	3.3	***3.8***	***3.5***	***3.6***	3.76
EGARCH-t	3.3	3.5	***3.7***	2.8	***3.7***	3.5	2.9	3.34
GJR-t	***3.8***	***3.6***	***3.6***	2.8	***3.8***	3.0	3.3	3.41
GARCH-HS-t	**2.6**	**2.4**	3.1	2.7	**2.4**	2.8	2.2	2.60
GARCH-EVT-t	***0.3***	1.7	1.7	***1.3***	1.8	***1.4***	1.8	1.43
CARE-SAV	**2.6**	**2.6**	**2.6**	2.8	**2.5**	**2.4**	**2.3**	2.54
RG-RV-tN	2.8	***4.4***	***4.2***	3.4	***3.6***	***3.9***	***4.5***	3.83
RG-RR-tN	**2.5**	***4.0***	***3.8***	**2.4**	2.9	***4.2***	3.5	3.33
RG-RK-tN	***3.6***	***4.4***	***4.1***	3.2	***4.1***	***1.2***	3.3	3.41
RE-RV-NN	3.1	***4.0***	***4.3***	2.9	***3.7***	3.2	***3.6***	3.54
RE-RV-SkN	**2.5**	3.4	3.2	**2.5**	2.9	2.9	3.3	2.96
RE-RV-tN	2.7	3.4	3.4	2.3	**2.5**	**2.4**	3.0	2.81
RE-RR-NN	2.9	***3.9***	***3.8***	**2.4**	3.2	3.3	3.5	3.29
RE-RR-SkN	2.1	3.2	3.0	1.8	**2.5**	3.0	2.9	2.64
RE-RR-tN	**2.5**	3.1	3.1	***1.4***	**2.5**	2.8	**2.7**	2.59
RE-RK-NN	3.1	***4.1***	***3.9***	2.8	***3.6***	**2.5**	3.2	3.31
RE-RK-SkN	2.8	3.3	3.1	**2.5**	3.0	1.7	**2.6**	2.71
RE-RK-tN	1.7	3.2	2.9	1.9	3.1	***1.1***	**2.4**	2.33
RE-RV-RR-NN	2.7	***4.0***	***4.0***	**2.5**	3.3	3.3	***3.7***	3.36
RE-RV-RR-SkN	1.8	3.3	3.0	2.1	**2.6**	2.9	3.1	2.69
RE-RV-RR-tN	2.2	3.2	3.2	***1.5***	**2.6**	2.7	2.8	2.60
RE-RV-RK-NN	3.2	***4.1***	***4.3***	2.9	***3.7***	3.1	***3.9***	3.60
RE-RV-RK-SkN	2.7	4.1	3.1	**2.5**	3.0	2.8	3.2	3.06
RE-RV-RK-tN	3.0	3.4	3.2	2.1	2.8	**2.4**	3.0	2.84
RE-RR-RK-NN	3.2	***4.0***	***3.8***	2.3	3.4	3.2	***3.6***	3.36
RE-RR-RK-SkN	2.3	3.4	**2.8**	1.9	**2.6**	2.9	2.8	2.67
RE-RR-RK-tN	**2.6**	3.2	2.9	***1.5***	**2.5**	1.2	**2.7**	2.37
RE-RV-RR-RK-NN	3.0	***4.1***	***4.1***	**2.4**	3.4	3.2	***3.8***	3.43
RE-RV-RR-RK-SkN	2.1	4.1	3.1	2.0	**2.6**	2.8	3.1	2.83
RE-RV-RR-RK-tN	2.3	3.4	3.1	1.6	**2.5**	2.7	**2.7**	2.61

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates that the violation rate is significantly different to 2.5% by the UC test.

TABLE 3.11: 1% VaR forecasting violation rate

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Average
GARCH-t	***2.1***	1.5	***1.8***	***1.7***	***1.9***	1.5	1.6	1.73
EGARCH-t	1.5	1.6	1.4	1.4	***1.7***	***1.9***	1.4	1.56
GJR-t	***1.8***	1.4	***2.0***	1.3	1.6	1.5	1.3	1.56
GARCH-HS-t	1.3	**1.1**	1.3	**1.1**	**0.9**	**1.1**	**0.9**	1.10
GARCH-EVT-t	***0.3***	***1.7***	***1.7***	1.3	***1.8***	1.4	1.6	1.40
CARE-SAV	0.8	**0.9**	1.2	0.8	**0.9**	0.8	**1.1**	0.93
RG-RV-tN	**1.1**	***2.5***	***2.3***	***1.9***	***2.1***	***2.3***	***2.6***	2.11
RG-RR-tN	**1.0**	***2.2***	1.5	1.3	1.4	***2.5***	***2.0***	1.70
RG-RK-tN	1.2	***2.6***	***1.8***	1.5	***2.6***	0.8	***1.7***	1.74
RE-RV-NN	1.2	***2.7***	***2.3***	***1.8***	***2.1***	***2.3***	***2.6***	2.14
RE-RV-SkN	0.7	1.6	1.5	1.3	1.3	1.5	1.3	1.31
RE-RV-tN	**1.0**	1.5	1.5	**0.9**	1.3	1.3	1.3	1.26
RE-RR-NN	**1.1**	***2.7***	***1.7***	1.4	***1.9***	***2.6***	2.2	1.94
RE-RR-SkN	0.7	1.3	**1.1**	**0.9**	**0.9**	1.5	**1.0**	1.06
RE-RR-tN	0.8	1.4	1.2	0.5	0.8	1.3	**1.0**	1.00
RE-RK-NN	1.6	***2.7***	***1.9***	1.6	***2.3***	***1.7***	***2.1***	1.99
RE-RK-SkN	**1.1**	1.4	**1.1**	**1.1**	2.3	0.7	0.6	1.19
RE-RK-tN	0.7	1.6	1.2	**1.1**	1.3	0.5	0.6	1.00
RE-RV-RR-NN	**1.1**	***2.8***	***1.8***	1.5	***1.9***	***2.5***	***2.2***	1.97
RE-RV-RR-SkN	0.6	1.4	1.2	**1.1**	1.2	1.5	**0.9**	1.13
RE-RV-RR-tN	0.7	1.6	1.4	0.5	1.2	1.2	0.6	1.03
RE-RV-RK-NN	1.6	***2.8***	***2.1***	***1.7***	***2.2***	***2.2***	***2.4***	2.14
RE-RV-RK-SkN	1.2	2.8	1.4	1.3	1.3	1.4	1.2	1.51
RE-RV-RK-tN	1.2	1.5	1.4	**1.0**	1.3	1.2	**1.0**	1.23
RE-RR-RK-NN	1.4	***2.8***	***1.7***	1.4	***2.1***	***2.5***	***1.9***	1.97
RE-RR-RK-SkN	**0.9**	1.4	**1.1**	**1.0**	1.2	1.4	**1.1**	1.16
RE-RR-RK-tN	**1.1**	1.4	**0.9**	0.5	**1.1**	1.2	0.7	0.99
RE-RV-RR-RK-NN	1.3	***2.8***	***1.7***	1.5	***2.0***	***2.4***	***2.3***	2.00
RE-RV-RR-RK-SkN	**1.0**	***1.8***	**1.1**	**1.0**	1.2	1.3	**1.0**	1.20
RE-RV-RR-RK-tN	**1.0**	1.4	1.4	0.5	**1.1**	**1.0**	**0.9**	1.04

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates that the violation rate is significantly different to 1% by the UC test.

The result of three types of VaR backtests (UC, CC, and DQ tests) conducted at 5% significance level is presented in Table 3.12. The proposed RE-SkN and RE-tN type models have a lower number of rejections than the competing models, as highlighted in green. The standard GARCH-family models have relatively high rejection rates, as expected. The RE-NN type models tend to also have a relatively high number of rejections. Meanwhile, most of the RE-SkN and RE-tN type models are not rejected in the backtests for all indices. Specifically, the 2.5% VaR forecasts generated by the RE-RK-SkN, RE-RV-RK-tN, and RE-RV-RR-RK-SkN models are not rejected for all three types of tests. For the 1% VaR forecasts, the RE-RK-SkN and RE-RR-SkN models have no rejection for all three types of tests. This finding is consistent across all types of realized measures, either using a single or multiple realized measures.

TABLE 3.12: VaR Backtest: UC, CC, and DQ tests

Model	$\alpha = 2.5\%$			$\alpha = 1\%$		
	UC	CC	DQ4	UC	CC	DQ4
GARCH-t	7	4	6	4	4	5
EGARCH-t	3	1	3	2	1	4
GJR-t	4	3	2	2	1	3
GARCH-HS-t	**0**	7	7	**0**	7	7
GARCH-EVT-t	3	3	6	4	3	6
CARE-SAV	**0**	1	3	**0**	2	2
RG-RV-tN	5	4	5	6	6	6
RG-RR-tN	3	3	3	3	3	5
RG-RK-tN	5	4	5	4	5	5
RE-RV-NN	4	2	4	6	5	5
RE-RV-SkN	**0**	**0**	1	**0**	**0**	2
RE-RV-tN	**0**	**0**	1	**0**	**0**	2
RE-RR-NN	2	2	2	4	3	4
RE-RR-SkN	**0**	**0**	2	**0**	**0**	**0**
RE-RR-tN	1	1	1	**0**	**0**	1
RE-RK-NN	3	2	3	5	4	5
RE-RK-SkN	**0**	**0**	**0**	**0**	**0**	**0**
RE-RK-tN	1	1	**0**	**0**	**0**	1
RE-RV-RR-NN	3	2	2	5	5	5
RE-RV-RR-SkN	**0**	**0**	1	**0**	**0**	2
RE-RV-RR-tN	1	**0**	1	**0**	**0**	2
RE-RV-RK-NN	4	3	4	6	5	6
RE-RV-RK-SkN	**0**	**0**	1	**0**	**0**	2
RE-RV-RK-tN	**0**	**0**	**0**	**0**	**0**	2
RE-RR-RK-NN	3	2	3	5	4	5
RE-RR-RK-SkN	**0**	**0**	2	**0**	**0**	1
RE-RR-RK-tN	1	**0**	1	**0**	**0**	2
RE-RV-RR-RK-NN	3	3	4	5	4	5
RE-RV-RR-RK-SkN	**0**	**0**	**0**	1	1	2
RE-RV-RR-RK-tN	**0**	**0**	1	**0**	**0**	1

Note: For each test, a green highlight indicates the model with the least number of rejections

Table 3.13 shows the result of the bivariate ESR backtests (Bayer & Dimitriadis, 2018) in Equation 2.22 across all market indices at the 5% significance level. There is a small number of rejections for the null hypothesis of correct ES forecasts by almost all ESR backtests, even for the traditional GARCH-family models employing the Student-*t* distribution, except for the GARCH-EVT-t model. This finding is consistent with the empirical studies conducted by Bayer and Dimitriadis (2018) that GARCH and GJR-GARCH models employing Student-*t* and skewed Student-*t* are not rejected by the ESR backtests. From Table 3.13, we can see the difference in performance across models at the 1% forecasting level, where the proposed

RE-tN and RE-SkN class models could generate more accurate ES forecasts compared to the competing models and the RE-NN class models. All proposed RE-tN and RE-SkN models have 0 ESR backtest rejection, except for the RE-RR-SkN and RE-RK-tN models. The best performing models with only one rejection for the 2.5% ES forecasts and no rejection for the 1% ES forecasts are the RE-RV-SkN, RE-RV-RK-SkN, RE-RR-RK-SkN, RE-RV-RR-RK-SkN, and RE-RV-RR-RK-tN models.

TABLE 3.13: ES regression-based backtest (ESR)

Model	**Total Rejections**	
	α **=2.5%**	α **=1%**
GARCH-t	1	1
EGARCH-t	2	1
GJR-t	1	1
GARCH-HS-t	1	1
GARCH-EVT-t	7	5
CARE-SAV	4	3
RG-RV-tN	3	1
RG-RR-tN	1	1
RG-RK-tN	2	1
RE-RV-NN	3	1
RE-RV-SkN	1	0
RE-RV-tN	2	0
RE-RR-NN	2	0
RE-RR-SkN	1	1
RE-RR-tN	2	0
RE-RK-NN	2	2
RE-RK-SkN	2	0
RE-RK-tN	2	1
RE-RV-RR-NN	2	0
RE-RV-RR-SkN	2	0
RE-RV-RR-tN	2	0
RE-RV-RK-NN	3	0
RE-RV-RK-SkN	1	0
RE-RV-RK-tN	2	0
RE-RR-RK-NN	2	0
RE-RR-RK-SkN	1	0
RE-RR-RK-tN	2	0
RE-RV-RR-RK-NN	3	0
RE-RV-RR-RK-SkN	1	0
RE-RV-RR-RK-tN	1	0

Note: Under the null hypothesis, the ES forecast are correctly specified. The tests are conducted at 5% level of significance. Less rejection is favored.

Table 3.14 presents the results of the ES backtests based on the multi-quantile regression (ES-MQR) approach of Couperier and Leymarie (2019) in Equation 2.30 across models and all market indices. The ES-MQR test is conducted by using the number of quantile $p = 6$ at the coverage level $\tau = 2.5\%$, as recommended by Couperier and Leymarie (2019). As an example, the six quantile VaR forecasts based on the RE-RV-RR-RK-SkN model for the S&P 500 used for the ES-MQR backtests are presented in Figure 3.5.

TABLE 3.14: ES backtests based on the multi-quantile regression ES-MQR) method

Model	**Hypothesis**				**Total**
	J_1	J_2	I	S	**Rejection**
GARCH-t	4	4	4	4	16
EGARCH-t	3	5	3	4	15
GJR-t	4	3	4	5	16
GARCH-HS-t	4	6	4	4	*18*
GARCH-EVT-t	7	7	7	7	***28***
CARE-SAV	2	7	**0**	4	13
RG-RV-tN	3	3	2	3	11
RG-RR-tN	2	3	1	3	9
RG-RK-tN	1	4	2	1	8
RE-RV-NN	1	2	1	2	6
RE-RV-SkN	2	5	**0**	4	11
RE-RV-tN	2	6	2	4	14
RE-RR-NN	1	3	1	1	6
RE-RR-SkN	2	4	1	3	10
RE-RR-tN	3	6	**0**	3	12
RE-RK-NN	1	**1**	1	1	**4**
RE-RK-SkN	1	6	1	3	11
RE-RK-tN	2	7	1	3	13
RE-RV-RR-NN	1	2	1	1	5
RE-RV-RR-SkN	1	4	1	3	9
RE-RV-RR-tN	2	6	1	4	13
RE-RV-RK-NN	**0**	1	**0**	1	**2**
RE-RV-RK-SkN	**0**	3	**0**	2	5
RE-RV-RK-tN	2	5	1	2	10
RE-RR-RK-NN	**0**	2	**0**	**0**	**2**
RE-RR-RK-SkN	**0**	4	**0**	1	5
RE-RR-RK-tN	**0**	5	**0**	2	7
RE-RV-RR-RK-NN	**0**	**1**	**0**	1	**2**
RE-RV-RR-RK-SkN	**0**	4	**0**	3	7
RE-RV-RR-RK-tN	1	6	**0**	1	8

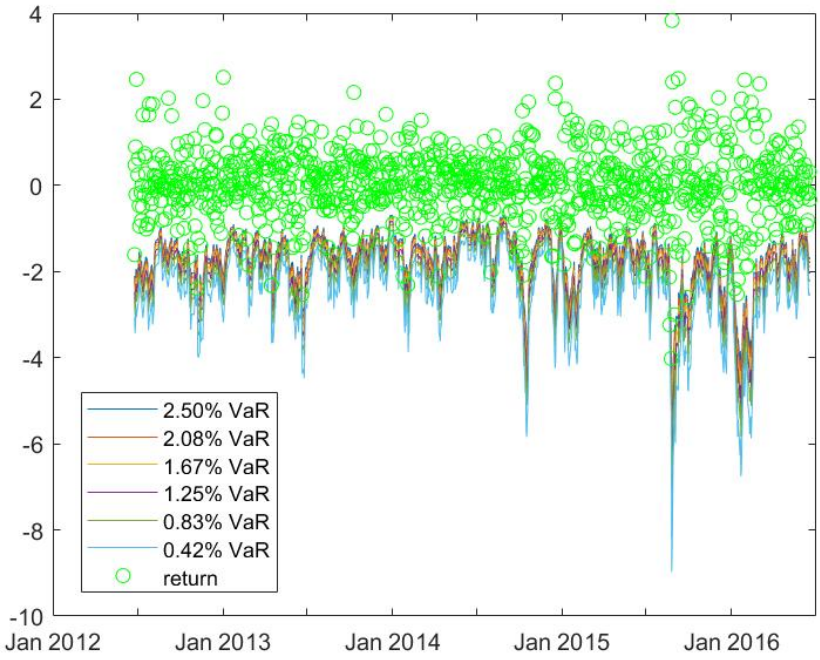
Note: a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates the least favored model, italics indicates the 2nd lowest ranked model in each column.

All four hypothesis testings of the ES-MQR backtests are conducted at the 5% significance level based on bootstrap critical values. With respect to the total rejections of four hypotheses of valid ES forecasts for seven market indices, the ES-MQR backtests of the realized EGARCH models estimated using the adaptive MCMC approach (the RE-NN,RE-tN, RE-SkN class models) are less likely to be rejected than the standard GARCH-type models using the Student-t distribution. However, some of those proposed models only perform slightly better than the CARE-SAV and RG-tN models, or even slightly worse than the RG-tN models. For instance, the RE-RV-tN model has a higher total number of rejections compared to the CARE-SAV and RG-tN models.

In terms of the individual hypotheses testings, the first joint backtests that sum the intercept and slope coefficients (H_{0,J_1}) and the intercept backtests ($H_{0,I}$) for the realized EGARCH models are less likely to reject the validity of ES forecasts. The non-rejection of the $H_{0,I}$ implies that the forecasting errors are not constant across time. The other joint backtests that separately sum the intercept and slope coefficients (H_{0,J_2}) tend to reject the validity of ES forecasts in most models. This is due to the rejection of the slope backtests ($H_{0,S}$), which indicates that the errors are time-varying since they fluctuate with respect to the VaR forecasts.

Although the RE-SkN type models are relatively preferable than the RE-tN type models, the ES-MQR backtests surprisingly are less likely to reject the tail risk validity produced by the RE-NN class models compared to the RE-SkN and RE-tN class models. This is rather contradictory to the result of the ESR backtests, which provide more supports for the RE-SkN and RE-tN type models. However, the rejection of the ES-MQR backtests for the RE-SkN and RE-tN models is mainly due to the rejection of H_{0,J_2} and $H_{0,S}$ hypotheses. For the other two hypothesis tests of ES backtests (H_{0,J_1} and $H_{0,I}$), the RE-SkN and RE-tN type models perform as well as the RE-NN type models. Lastly, the realized EGARCH models with multiple realized measures have lower ES-MQR backtest rejections than those with a single realized measure.

FIGURE 3.5: Six quantiles VaR forecasts of the RE-RV-RR-RK-SkN model for S&P 500



3.7.5 Quantile loss functions

Tables 3.15 and 3.16 present the quantile loss function values at the 2.5% and 1% forecasting levels, respectively. The proposed RE-SkN and RE-tN type models in general generate smaller quantile losses and rank better compared to other models and the RE-NN class models.

For the 2.5% VaR forecast, the best model generating the lowest average quantile loss value is the RE-RV-RK-SkN model (63.215), followed by the RE-RV-RR-RK-SkN model (63.220). In terms of average rank, the RE-RV-RK-SkN model also ranks the best, followed by the RE-RV-RK-tN model. In contrast, the GARCH-EVT-t and GARCH-t models produce the highest average quantile loss values, which is consistent with the VRate of these models in Table 3.10. On the one hand, the GARCH-EVT-t model tends to underestimate risks with an average VRate of 1.43%, which is much lower than the nominal VRate of 2.5%. On the other hand, the GARCH-t model overestimates risks, producing a VRate much higher than 2.5%, that is, 3.76%.

For the 1% forecast, the model that has the lowest average quantile loss value (29.541) and ranks the best (5.4 on average) is the RE-RV-RR-RK-SkN model. As discussed, the VRate of this model is also the best for four of the seven market indices in this study, although its average VRate across seven market indices is not the lowest. The second-best model is the RE-RV-RK-SkN model with an average quantile loss value of 29.643, while the RE-RV-RR-SkN model has the second-best average rank (7.3). It is evident for both the 2.5% and 1% forecasting levels that the proposed RE-SkN-type models are preferred in VaR forecasting.

TABLE 3.15: Quantile loss function values; $\alpha = 2.5\%$

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Mean Loss	Mean Rank
GARCH-t	*57.801*	80.213	*62.499*	76.265	66.537	69.750	56.700	*67.103*	*27.0*
EGARCH-t	54.925	78.204	59.180	73.726	63.206	68.774	53.840	64.550	19.3
GJR-t	56.707	79.343	59.779	73.114	63.810	68.708	54.000	65.066	22.3
GARCH-HS-t	57.756	80.133	62.454	75.875	66.424	69.844	56.700	67.027	26.4
GARCH-EVT-t	***90.850***	***108.287***	***86.737***	***91.598***	***72.950***	***90.156***	***67.292***	***86.839***	***30.0***
CARE-SAV	55.334	78.659	59.965	*76.349*	65.408	72.118	*57.845*	66.526	25.0
RG-RV-tN	54.166	*82.657*	60.951	72.225	64.966	69.571	55.521	65.723	22.7
RG-RR-tN	54.156	81.807	59.683	71.781	64.224	70.438	54.264	65.193	20.9
RG-RK-tN	55.098	81.962	61.489	73.008	*66.660*	*75.288*	55.097	66.943	26.0
RE-RV-NN	53.699	79.982	59.452	71.985	62.160	67.764	53.254	64.042	16.9
RE-RV-SkN	53.332	78.604	58.022	71.824	62.090	67.034	52.091	63.285	8.7
RE-RV-tN	53.519	78.514	57.899	72.241	61.884	67.207	52.038	63.329	9.0
RE-RR-NN	53.613	79.792	58.678	**71.569**	62.005	68.185	52.879	63.817	14.3
RE-RR-SkN	53.739	78.936	**57.547**	71.975	62.152	67.599	52.127	63.439	11.0
RE-RR-tN	53.390	78.577	57.752	74.610	62.013	**66.993**	52.320	63.665	11.1
RE-RK-NN	54.447	79.329	59.070	72.537	62.742	68.530	52.383	64.148	18.4
RE-RK-SkN	54.373	**78.117**	58.089	72.280	62.317	70.062	52.056	63.899	14.7
RE-RK-tN	54.746	**77.736**	58.008	73.024	62.089	73.830	52.649	64.583	16.1
RE-RV-RR-NN	53.446	79.815	58.846	72.016	61.962	68.169	52.821	63.868	14.6
RE-RV-RR-SkN	54.085	79.033	57.728	72.343	61.917	67.620	51.832	63.509	11.6
RE-RV-RR-tN	53.558	78.577	57.610	73.943	61.867	67.276	**51.379**	63.459	9.4
RE-RV-RK-NN	53.509	79.804	59.171	72.176	62.366	67.595	52.870	63.927	15.9
RE-RV-RK-SkN	53.241	78.329	57.592	72.277	61.785	67.165	52.118	**63.215**	**6.4**
RE-RV-RK-tN	53.345	78.397	57.780	72.922	61.642	**67.030**	**51.490**	63.229	**6.6**
RE-RR-RK-NN	53.396	80.086	58.403	72.002	61.636	67.816	52.544	63.698	12.6
RE-RR-RK-SkN	**53.161**	78.558	57.783	**71.747**	62.122	67.688	52.353	63.345	9.0
RE-RR-RK-tN	53.355	78.536	**57.328**	74.310	**61.632**	67.338	51.556	63.437	7.7
RE-RV-RR-RK-NN	53.165	80.139	58.833	72.020	62.253	67.841	52.839	63.870	15.1
RE-RV-RR-RK-SkN	53.251	78.859	57.684	72.228	**61.261**	67.420	51.838	**63.220**	7.3
RE-RV-RR-RK-tN	**53.145**	78.552	57.801	74.028	61.701	67.201	52.083	63.502	8.7

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates the least favored model, italics indicates the 2nd lowest ranked model, in each column.

TABLE 3.16: Quantile loss function values; $\alpha = 1\%$

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Mean Loss	Mean Rank
GARCH-t	*26.731*	36.925	29.449	36.873	29.736	35.759	26.410	31.698	23.7
EGARCH-t	25.572	38.023	27.965	35.068	28.153	35.976	24.696	30.779	17.7
GJR-t	25.772	38.542	27.581	34.493	28.474	36.222	24.720	30.829	18.7
GARCH-HS-t	*26.731*	36.526	*29.561*	36.510	29.717	35.802	26.410	31.608	23.6
GARCH-EVT-t	***37.392***	***49.205***	***40.302***	***39.483***	***33.299***	***47.263***	***30.225***	***39.596***	***30.0***
CARE-SAV	25.362	**35.768**	29.088	*37.863*	30.125	37.272	*27.199*	31.811	22.3
RG-RV-tN	25.007	*39.541*	29.222	34.610	30.344	37.100	26.679	31.786	23.4
RG-RR-tN	25.086	38.481	28.128	**33.116**	29.329	37.810	25.252	31.029	19.0
RG-RK-tN	26.214	38.794	29.262	34.777	*31.920*	*40.033*	26.400	*32.486*	*26.4*
RE-RV-NN	24.943	38.711	28.378	35.116	29.267	36.626	25.437	31.211	21.6
RE-RV-SkN	24.876	35.926	26.976	33.231	28.452	34.398	23.735	29.656	8.0
RE-RV-tN	**24.828**	36.122	26.941	33.584	28.308	34.580	23.622	29.712	8.6
RE-RR-NN	24.943	38.154	27.932	33.827	28.659	37.001	24.584	30.728	16.6
RE-RR-SkN	25.060	35.888	**26.705**	33.276	28.427	34.554	23.855	29.681	8.3
RE-RR-tN	24.923	**35.808**	26.832	35.024	28.385	**34.128**	23.911	29.858	9.3
RE-RK-NN	26.082	38.280	28.103	35.040	29.244	36.500	24.074	31.046	20.9
RE-RK-SkN	25.505	36.129	27.066	34.351	28.172	36.665	23.611	30.214	13.9
RE-RK-tN	25.855	36.242	27.082	34.080	28.081	38.816	23.922	30.582	15.9
RE-RV-RR-NN	24.834	38.541	27.985	34.482	28.514	36.915	24.493	30.824	17.0
RE-RV-RR-SkN	25.267	36.011	26.812	33.555	28.036	34.817	**23.205**	29.672	**7.3**
RE-RV-RR-tN	24.917	36.237	26.842	34.658	28.164	34.382	23.396	29.799	8.4
RE-RV-RK-NN	25.873	38.653	28.166	34.859	29.185	36.423	24.772	31.133	22.1
RE-RV-RK-SkN	25.104	36.011	26.711	33.571	28.292	34.405	23.404	**29.643**	7.7
RE-RV-RK-tN	25.314	35.883	26.922	33.719	28.203	34.449	23.397	29.698	8.4
RE-RR-RK-NN	25.559	38.588	27.622	34.300	28.382	36.587	24.215	30.751	17.6
RE-RR-RK-SkN	25.012	36.003	26.934	**33.012**	28.330	34.793	23.701	29.683	9.0
RE-RR-RK-tN	25.270	36.119	**26.494**	34.614	28.137	**34.343**	23.563	29.791	8.7
RE-RV-RR-RK-NN	**24.821**	38.954	27.975	34.606	28.704	36.549	24.487	30.871	17.3
RE-RV-RR-RK-SkN	25.025	35.869	26.817	33.443	**27.534**	34.725	**23.372**	**29.541**	**5.4**
RE-RV-RR-RK-tN	24.977	35.986	26.896	33.935	**28.010**	34.664	23.630	29.728	7.9

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates the least favored model, italics indicates the 2nd lowest ranked model, in each column.

3.7.6 Joint loss function of VaR and ES

The values of the special case of the VaR and ES joint loss functions of Fissler and Ziegel (2016), based on Equation 2.39, are presented in Table 3.17 for the 2.5% forecasting level and Table 3.18 for the 1% forecasting level. The proposed RE-SkN type models show a more desirable property compared to other models, generating lower VaR and ES joint loss values for five of seven market indices for either the 2.5% or 1% forecast level. The RE-RR-SkN model consistently produces the lowest average of the VaR and ES joint losses for both forecasting levels, followed by the RE-RR-RK-SkN model for the 2.5% forecasting level and the RE-RV-RR-SkN model for the 1% forecasting level. On average, the RE-RR-SkN model also ranks first for the 1% forecasting level and second for the 2.5% forecasting level (the RE-RR-RK-SkN model ranks first for the 2.5% forecasting level). As regards the type of realized measures, the proposed RE-SkN type models employing RRSS, both individually or jointly with RVSS and/or RK, are more likely to have lower VaR and ES joint losses. It

can be concluded that the choice of the RRSS variable and the skewed Student-*t* distribution as the return equation errors can improve the tail risk forecasting efficiency.

TABLE 3.17: VaR and ES joint loss function values; $\alpha = 2.5\%$

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Mean Loss	Mean Rank
GARCH-t	*986.5*	1068.8	*1004.8*	1049.8	1015.6	1019.9	970.8	1016.6	*25.1*
EGARCH-t	970.7	1073.3	995.5	1042.6	1002.8	1036.1	948.2	1009.9	24.0
GJR-t	980.2	1076.4	986.3	1037.2	1003.6	1030.2	951.1	1009.3	23.7
GARCH-HS-t	965.2	1058.7	988.8	1045.8	1007.1	1012.1	962.9	1005.8	20.4
GARCH-EVT-t	***1085.0***	***1128.6***	***1092.2***	***1091.6***	***1035.5***	***1157.3***	***1018.4***	***1086.9***	***30.0***
CARE-SAV	974.2	**1057.6**	990.9	*1054.7*	1018.0	1037.5	*983.7*	*1016.7*	23.9
RG-RV-tN	941.3	*1090.1*	1001.7	1037.5	1009.4	1026.3	966.1	1010.3	24.0
RG-RR-tN	941.2	1077.6	984.5	1028.7	996.4	*1041.1*	945.4	1002.1	18.4
RG-RK-tN	957.6	1083.1	1002.2	1031.8	*1026.7*	1033.2	954.2	1012.7	24.9
RE-RV-NN	943.7	1078.1	986.2	1035.4	998.4	1019.9	949.7	1001.6	20.7
RE-RV-SkN	939.1	1060.4	969.4	1031.7	989.6	**997.6**	928.5	988.0	8.1
RE-RV-tN	941.1	1066.6	972.7	1033.5	989.1	1002.4	928.7	990.6	12.6
RE-RR-NN	941.7	1072.3	978.0	**1028.5**	991.4	1027.5	926.6	995.1	13.9
RE-RR-SkN	940.6	1058.9	**963.9**	1030.5	985.8	1001.9	926.5	**986.9**	**5.6**
RE-RR-tN	938.7	1061.1	968.6	1037.7	985.9	1003.4	927.9	989.0	9.6
RE-RK-NN	957.2	1077.0	988.2	1033.1	1004.0	1016.4	931.2	1001.0	19.9
RE-RK-SkN	951.6	1061.9	974.0	1029.8	992.5	1012.9	**923.2**	992.3	11.3
RE-RK-tN	952.2	1063.5	976.1	1031.8	994.3	1028.4	926.6	996.1	16.0
RE-RV-RR-NN	940.0	1076.0	979.0	1031.2	991.4	1026.5	937.7	997.4	15.7
RE-RV-RR-SkN	943.4	1062.8	967.0	1031.1	986.5	1002.2	924.0	988.1	8.4
RE-RV-RR-tN	940.8	1065.2	969.0	1036.5	987.1	1004.3	924.1	989.6	10.6
RE-RV-RK-NN	947.4	1078.0	983.9	1034.0	998.9	1018.2	940.9	1000.2	19.6
RE-RV-RK-SkN	940.2	1059.0	966.8	1029.6	990.5	**998.2**	925.9	987.2	6.1
RE-RV-RK-tN	942.8	1065.4	972.2	1034.5	991.0	1000.6	925.4	990.3	11.7
RE-RR-RK-NN	943.3	1076.7	974.5	1031.0	988.6	1022.0	933.8	995.7	15.1
RE-RR-RK-SkN	**936.2**	**1057.6**	966.9	**1028.6**	988.7	1003.9	927.3	**987.0**	5.7
RE-RR-RK-tN	940.5	1065.4	**964.5**	1037.0	**984.8**	1006.1	**921.3**	988.5	8.9
RE-RV-RR-RK-NN	939.2	1079.4	980.1	1029.9	994.1	1022.8	941.8	998.2	16.1
RE-RV-RR-RK-SkN	939.5	1064.7	966.5	1030.8	**983.6**	1001.3	924.0	987.2	**5.1**
RE-RV-RR-RK-tN	**937.4**	1065.3	969.7	1035.5	986.2	1004.8	925.1	989.1	9.6

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates the least favored model, italics indicates the 2nd lowest ranked model, in each column.

TABLE 3.18: VaR and ES joint loss function values; $\alpha = 1\%$

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Mean Loss	Mean Rank
GARCH-t	981.2	1042.2	982.0	1015.4	983.0	999.3	957.1	994.3	22.29
EGARCH-t	*987.3*	1064.7	*1006.9*	1015.8	971.9	1043.6	938.6	*1004.1*	24.57
GJR-t	980.9	1059.5	974.5	1007.7	975.6	1034.1	938.2	995.8	20.86
GARCH-HS-t	956.3	1032.8	977.1	1010.6	973.9	995.6	955.7	986.0	18.43
GARCH-EVT-t	***1033.8***	***1120.7***	***1116.0***	***1040.2***	***1050.1***	***1247.3***	***998.2***	***1086.6***	***30.00***
CARE-SAV	963.5	1026.1	982.0	*1026.1*	993.9	1012.9	*975.8*	997.2	21.57
RG-RV-tN	**930.0**	1062.5	996.2	1016.1	991.2	1032.5	961.4	998.6	21.57
RG-RR-tN	**931.6**	1043.9	969.5	1001.0	970.5	1049.2	930.4	985.1	13.86
RG-RK-tN	951.9	1052.4	992.6	1003.3	*1010.5*	1014.7	944.9	995.7	21.29
RE-RV-NN	944.3	1066.1	983.8	1021.3	997.2	1041.8	964.7	1002.8	*25.14*
RE-RV-SkN	940.4	1026.9	955.3	1005.3	969.3	985.5	916.8	971.4	9.86
RE-RV-tN	941.5	1036.5	959.4	1004.8	967.5	993.9	911.8	973.6	12.14
RE-RR-NN	940.7	1052.3	973.7	1006.3	979.4	*1050.5*	912.1	987.9	18.00
RE-RR-SkN	938.0	**1023.1**	**949.5**	1001.7	962.1	**984.5**	915.5	**967.8**	**5.14**
RE-RR-tN	938.1	1025.3	954.2	1004.5	962.0	989.0	913.6	969.5	7.71
RE-RK-NN	965.4	*1069.3*	995.7	1013.0	996.9	1023.5	923.2	998.1	23.86
RE-RK-SkN	946.3	1035.4	970.3	1003.6	967.7	999.3	912.2	976.4	13.43
RE-RK-tN	946.5	1036.1	973.7	**999.3**	970.0	1013.1	912.1	978.7	13.14
RE-RV-RR-NN	936.7	1062.4	973.7	1009.5	979.0	1047.7	934.6	992.0	18.71
RE-RV-RR-SkN	940.6	1028.6	953.7	1002.0	961.5	987.4	**906.0**	**968.5**	**5.57**
RE-RV-RR-tN	939.1	1034.1	955.6	1004.5	964.1	992.2	910.4	971.4	9.14
RE-RV-RK-NN	960.4	1067.9	981.3	1015.4	995.4	1036.3	941.8	999.8	24.14
RE-RV-RK-SkN	940.6	1027.6	954.7	1000.5	970.6	**984.2**	910.7	969.8	6.86
RE-RV-RK-tN	945.3	1033.6	961.2	1004.2	970.7	990.4	911.1	973.8	11.71
RE-RR-RK-NN	949.3	1060.7	967.3	1008.7	974.3	1039.4	927.7	989.6	19.14
RE-RR-RK-SkN	934.7	**1023.7**	954.5	**999.7**	965.4	988.4	913.6	968.6	**5.57**
RE-RR-RK-tN	942.2	1033.6	**950.0**	1003.2	**961.1**	992.4	**909.6**	970.3	7.29
RE-RV-RR-RK-NN	938.7	1068.2	976.6	1007.5	983.3	1042.0	941.1	993.9	20.71
RE-RV-RR-RK-SkN	938.3	1032.2	953.3	1001.9	**957.3**	988.5	910.9	968.9	**5.57**
RE-RV-RR-RK-tN	934.2	1033.7	956.8	1001.0	961.7	996.9	910.8	970.7	7.43

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates the least favored model, italics indicates the 2nd lowest ranked model, in each column.

The other special case of the joint loss function of VaR and ES forecasts based on the asymmetric Laplace (AL) log score by Taylor (2019) in Equation 2.40 across models and market indices is presented in Table 3.19 for the 2.5% forecast level and Table 3.20 for the 1% forecast level, respectively. The proposed RE-SkN type models also produce the smallest average of the AL log scores (highlighted in green), that is, four of seven market indices for the 2.5% forecast level and five of seven market indices for the 1% forecast level. Some models in this class type also have the second lowest AL log score average.

The RE-RV-RR-RK-SkN model consistently generates the lowest joint loss and ranks first, on average, for both forecast levels. For the 2.5% forecasting level, the second-best models with the smallest joint loss values are the RE-RR-SkN, RE-RR-RK-SkN, and RE-RV-RK-SkN models, while for the 1% forecasting level, the second-best model is the RE-RR-SkN model.

The comparison of VaR and ES joint loss functions using two different score functions, one according to Equation 2.39 and another according to Equation 2.40, shows the superiority of

the RE-SkN-type models. It is also evident that the choice of the RRSS variable, either individually or jointly with the other two realized measures, can improve the tail risk forecast performance of the RE-SkN type models.

TABLE 3.19: Joint loss for VaR and ES forecasts based on AL log score; $\alpha = 2.5\%$

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Mean loss	Mean Rank
GARCH-t	*1.901*	2.209	*1.934*	2.130	2.013	2.033	1.848	*2.010*	*26.00*
EGARCH-t	1.837	2.207	1.898	2.099	1.961	2.057	1.775	1.976	24.00
GJR-t	1.873	2.222	1.882	2.081	1.972	2.045	1.785	1.980	24.43
GARCH-HS-t	1.823	2.172	1.892	2.119	1.974	2.009	1.817	1.972	21.57
GARCH-EVT-t	***2.317***	***2.499***	***2.280***	***2.340***	***2.093***	***2.481***	***2.030***	***2.291***	***30.00***
CARE-SAV	1.831	2.170	1.892	*2.141*	1.996	2.079	*1.870*	1.997	24.00
RG-RV-tN	1.762	*2.273*	1.919	2.080	1.984	2.043	1.829	1.984	24.43
RG-RR-tN	1.759	2.236	1.874	2.052	1.943	2.079	1.770	1.959	19.29
RG-RK-tN	1.808	2.255	1.925	2.074	*2.035*	*2.080*	1.796	1.996	*26.00*
RE-RV-NN	1.762	2.216	1.870	2.067	1.931	2.020	1.769	1.948	19.14
RE-RV-SkN	1.749	2.169	1.830	2.057	1.911	**1.965**	1.716	1.914	7.43
RE-RV-tN	1.757	2.183	1.835	2.064	1.914	1.976	1.717	1.921	12.14
RE-RR-NN	1.756	2.202	1.852	**2.050**	1.914	2.033	1.716	1.932	13.43
RE-RR-SkN	1.753	2.168	**1.816**	2.057	1.903	1.977	1.711	**1.912**	5.86
RE-RR-tN	1.749	2.170	1.826	2.091	1.902	1.976	1.716	1.918	9.14
RE-RK-NN	1.791	2.210	1.876	2.067	1.945	2.009	1.725	1.946	19.00
RE-RK-SkN	1.779	2.171	1.840	2.057	1.919	2.012	1.705	1.926	12.29
RE-RK-tN	1.776	2.172	1.844	2.067	1.928	2.067	1.716	1.939	16.57
RE-RV-RR-NN	1.753	2.210	1.855	2.058	1.914	2.030	1.741	1.937	15.43
RE-RV-RR-SkN	1.761	2.181	1.823	2.060	1.904	1.977	1.705	1.916	9.00
RE-RV-RR-tN	1.754	2.185	1.826	2.082	1.904	1.979	**1.703**	1.919	10.71
RE-RV-RK-NN	1.775	2.215	1.867	2.065	1.933	2.010	1.750	1.945	18.71
RE-RV-RK-SkN	1.751	**2.166**	1.822	2.057	1.911	**1.966**	1.711	**1.912**	**5.43**
RE-RV-RK-tN	1.762	2.182	1.834	2.071	1.915	1.969	1.709	1.920	12.00
RE-RR-RK-NN	1.762	2.213	1.842	2.058	1.910	2.018	1.731	1.933	15.00
RE-RR-RK-SkN	**1.741**	**2.165**	1.823	**2.052**	1.909	1.981	1.714	**1.912**	**5.43**
RE-RR-RK-tN	1.755	2.182	**1.815**	2.087	**1.901**	1.982	**1.698**	1.917	9.14
RE-RV-RR-RK-NN	1.748	2.219	1.857	2.056	1.926	2.021	1.750	1.940	15.14
RE-RV-RR-RK-SkN	1.748	2.180	1.821	2.059	**1.894**	1.971	1.704	**1.911**	**5.14**
RE-RV-RR-RK-tN	**1.743**	2.179	1.828	2.081	1.901	1.979	1.710	1.918	8.86

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates the least favored model, italics indicates the 2nd lowest ranked model, in each column.

TABLE 3.20: Joint loss for VaR and ES forecasts based on AL log score; $\alpha = 1\%$

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P 500	Mean loss	Mean Rank
GARCH-t	*2.064*	2.365	2.072	2.293	2.110	2.246	1.980	2.161	*23.86*
EGARCH-t	2.027	2.441	2.094	2.261	2.051	2.347	1.903	2.161	*23.86*
GJR-t	2.029	2.439	2.025	2.231	2.063	2.335	1.900	2.146	21.71
GARCH-HS-t	1.960	2.347	2.056	2.266	2.049	2.230	1.968	2.125	20.29
GARCH-EVT-t	***2.364***	***2.797***	***2.591***	***2.434***	***2.358***	***3.046***	***2.190***	***2.540***	***30.00***
CARE-SAV	1.962	2.321	2.070	*2.340*	2.128	2.296	*2.025*	2.163	23.00
RG-RV-tN	**1.885**	*2.472*	*2.102*	2.262	2.139	2.354	1.998	2.173	23.57
RG-RR-tN	**1.888**	2.404	2.021	2.195	2.060	*2.405*	1.894	2.124	15.57
RG-RK-tN	1.956	2.437	2.094	2.232	*2.225*	2.334	1.950	*2.175*	23.29
RE-RV-NN	1.911	2.441	2.046	2.267	2.106	2.372	1.960	2.158	*23.86*
RE-RV-SkN	1.900	2.303	1.965	2.202	2.032	**2.189**	1.828	2.060	8.57
RE-RV-tN	1.903	2.327	1.972	2.205	2.038	2.211	1.816	2.067	11.29
RE-RR-NN	1.902	2.401	2.021	2.212	2.057	2.381	1.821	2.114	16.71
RE-RR-SkN	1.898	**2.294**	**1.948**	2.197	2.016	2.189	1.827	**2.053**	6.00
RE-RR-tN	1.895	**2.295**	1.960	2.233	2.013	2.192	1.824	2.059	8.14
RE-RK-NN	1.973	2.439	2.069	2.249	2.105	2.311	1.849	2.142	21.71
RE-RK-SkN	1.924	2.325	1.996	2.215	2.023	2.250	1.816	2.078	12.71
RE-RK-tN	1.926	2.328	2.003	2.201	2.036	2.323	1.821	2.091	13.71
RE-RV-RR-NN	1.894	2.430	2.022	2.230	2.054	2.372	1.880	2.126	17.71
RE-RV-RR-SkN	1.908	2.312	1.958	2.199	2.008	2.199	**1.795**	2.054	6.00
RE-RV-RR-tN	1.898	2.336	1.963	2.225	2.015	2.203	**1.808**	2.064	9.14
RE-RV-RK-NN	1.976	2.443	2.041	2.249	2.101	2.340	1.905	2.151	23.29
RE-RV-RK-SkN	1.909	2.307	1.958	**2.195**	2.031	**2.185**	1.809	2.056	6.00
RE-RV-RK-tN	1.927	2.320	1.975	2.207	2.036	2.199	1.811	2.068	11.14
RE-RR-RK-NN	1.939	2.427	2.002	2.227	2.051	2.345	1.860	2.122	18.00
RE-RR-RK-SkN	1.891	2.297	1.962	**2.187**	2.021	2.202	1.820	2.054	**5.86**
RE-RR-RK-tN	1.918	2.320	**1.944**	2.222	2.012	2.203	1.809	2.061	8.29
RE-RV-RR-RK-NN	1.896	2.450	2.027	2.228	2.076	2.354	1.893	2.132	19.57
RE-RV-RR-RK-SkN	1.898	2.312	1.957	2.196	**1.991**	2.195	1.808	**2.051**	**4.43**
RE-RV-RR-RK-tN	1.891	2.316	1.966	2.205	**2.007**	2.219	1.813	2.059	7.43

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked model, bold italics indicates the least favored model, italics indicates the 2nd lowest ranked model, in each column.

3.7.7 Model confidence set

The joint loss function based on the AL log score function (Equation 2.40) for the 2.5% and 1% forecast levels are also incorporated as the loss function in the calculation of model confidence set (MCS) (Hansen et al., 2011). Tests are carried out using both R and SQ methods at 90% and 75% confidence levels. The results of the tests are presented in Table 3.21. The models that consistently enter the MCS for both MCS methods and across market indices are highligted in green. Overall, the proposed extensions of the realized EGARCH models (the RE-SkN and RE-tN type models) demonstrate superior performance compared to the competing models. The RE-SkN and RE-tN type models are more likely to be included in the R and SQ MCS methods, both at the 90% and 75% confidence levels. They are also more favored than the original realized EGARCH models that are estimated using the proposed adaptive MCMC in this study (the RE-NN type models). TThe superiority of the RE-SkN

and RE-tN model types is confirmed by the MCS p-values of those models, which consistently demonstrate high values in most cases. This empirical evidence suggests that those models are highly likely to be members of the set of superior models. For details regarding the MCS *p*-values for each model at 90% and 75% confidence levels using both R and SQ methods, please see Appendix A.3.

TABLE 3.21: 90% and 75% Model confidence set with R and SQ methods

Model	**α=2.5%**				**α=1%**			
	90% MCS		**75% MCS**		**90% MCS**		**75% MCS**	
	R	**SQ**	**R**	**SQ**	**R**	**SQ**	**R**	**SQ**
GARCH-t	4	2	1	2	5	5	3	3
EGARCH-t	5	5	5	3	6	5	5	4
GJR-t	4	3	4	2	6	5	6	5
GARCH-HS	4	4	3	2	6	5	5	4
GARCH-t-EVT	2	2	1	1	3	2	2	2
CARE-SAV	4	4	4	2	4	6	5	3
RG-RV-tN	5	2	3	2	5	4	4	2
RG-RR-tN	5	4	6	3	6	5	6	4
RG-RK-tN	6	3	3	3	5	4	3	2
RE-RV-NN	6	4	6	3	5	4	3	2
RE-RV-SkN	7	7	7	7	7	6	7	6
RE-RV-tN	6	5	6	4	7	5	7	4
RE-RR-NN	7	6	7	5	7	6	7	5
RE-RR-SkN	6	7	7	7	6	6	6	6
RE-RR-tN	6	6	6	6	6	6	5	6
RE-RK-NN	6	4	6	3	5	5	3	2
RE-RK-SkN	7	7	7	7	7	6	7	6
RE-RK-tN	7	5	7	5	7	6	7	6
RE-RV-RR-NN	6	5	6	5	6	5	5	5
RE-RV-RR-SkN	6	6	7	6	6	6	6	6
RE-RV-RR-tN	7	6	6	6	7	6	7	6
RE-RV-RK-NN	6	5	6	3	5	5	3	3
RE-RV-RK-SkN	7	7	7	7	7	7	7	7
RE-RV-RK-tN	7	7	7	7	7	7	7	7
RE-RR-RK-NN	6	5	7	4	6	5	6	4
RE-RR-RK-SkN	7	6	7	6	7	7	7	7
RE-RR-RK-tN	7	6	7	6	7	6	6	6
RE-RV-RR-RK-NN	6	5	6	4	6	5	5	4
RE-RV-RR-RK-SkN	7	7	7	7	7	7	7	7
RE-RV-RR-RK-tN	7	7	7	7	7	7	7	7

Note: For each MCS method, a green highlight indicates favored models for all market indices.

The results in Table 3.21 also indicate that there is a tendency that realized EGARCH models employing multiple realized measures have a more favorable performance than those using

a single realized measure. In particular, there are 3 RE-SkN type models and 2 RE-tN type models using multiple realized measures that are consistently included in the MCS using both the R and SQ methods at both confidence levels, that is, the RE-RV-RR-SkN, RE-RV-RK-SkN, RE-RV-RR-RK-SkN, RE-RV-RK-tN, and RE-RV-RR-RK-tN models. The RE-RV-RR-RK-SkN is also the most preferred model based on the VaR and ES joint loss function values using the AL log score function (Equation 2.40). Meanwhile, the RE-RR-SkN model, which is the second best model based on the joint loss function values using the loss functions in Equation 2.39 and Equation 2.40, is all included in the MCS for seven market indices when the MCS incorporates the 2.5% VaR and ES levels. Although the RE-RR-SkN model is included in the MCS for only six market indices when using the 1% tail risk forecasting, the performance of this model overall is still highly favorable.

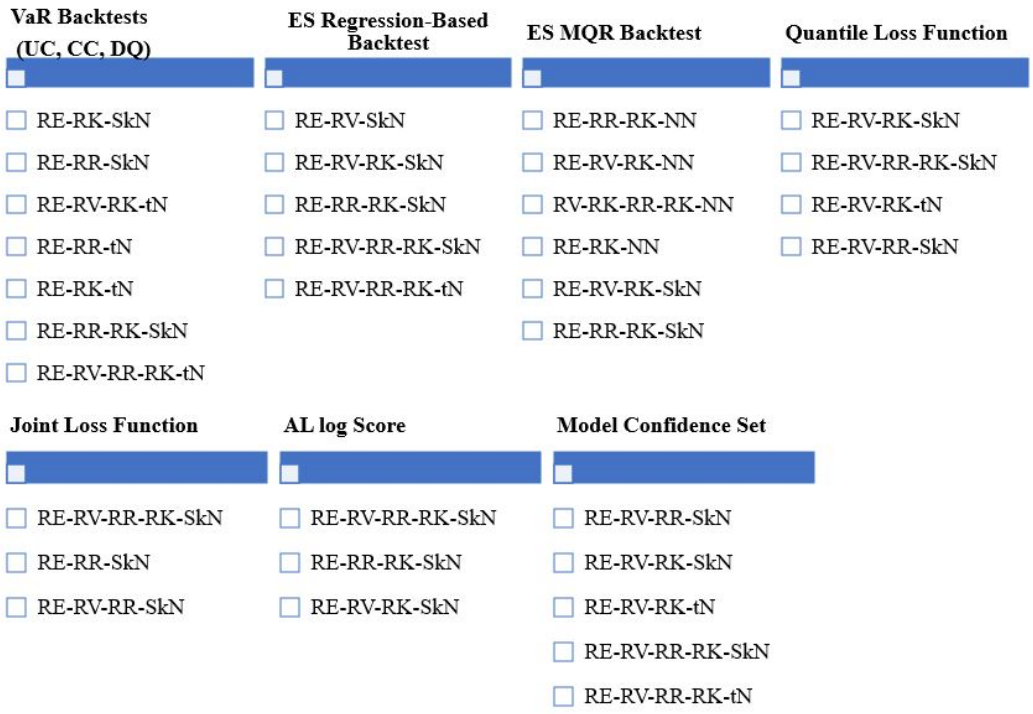
3.8 Chapter Summary

In this chapter, a Bayesian approach is proposed to estimate the realized EGARCH model for the purpose of VaR and ES forecasting. The original realized EGARCH model is extended by employing Student-*t* and skewed Student-*t* distributions for the observation equation errors (RE-tN and RE-SkN type models). The result of the simulation study demonstrates the advantage of the MCMC estimators over the maximum likelihood estimators in terms of unbiasedness and efficiency.

In the empirical study using seven market indices, the subsampled realized variance, the subsampled realized range, as well as the realized kernel are chosen to improve the out-of-sample forecasting performance of the models. The results of the empirical study show that the choice of the return equation error distribution is found to be consequential. The proposed RE-SkN class models have a superior tail risk forecasting performance in most cases compared to the standard GARCH-type models, CARE models, and realized GARCH models employing the Student-*t* distribution for the observation equation errors. The proposed RE-SkN class models consistently produce conservative and accurate violation rates and sufficient risk coverage, are generally less likely to be rejected in both VaR and ES backtests, produce the lowest joint loss function values, and most of the time are included in the model confidence set.

This study also finds that both the RE-SkN and RE-tN type models are more favorable than the original realized EGARCH model using the Gaussian distribution estimated using the proposed adaptive MCMC method (RE-NN type models). However, the RE-SkN type models have a better performance in many cases than the RE-tN types model. As can be seen in Figure **??**, most of the best models suggested by various VaR and ES forecast evaluation tests are RE-SkN type models. With respect to the choice of realized measures, the subsampled realized range (either employed individually or jointly with subsampled realized variance and/or realized kernel) is found to be crucial in improving the accuracy of the tail risk forecasts.

FIGURE 3.6: Summary of Best Models based on VaR and ES Forecast Evaluation



This study concludes that the RE-SkN and RE-tN type models incorporating a subsampled realized range should be considered in tail risk forecasting. This will help financial institutions to improve capital allocation efficiency and allow them to have larger profit maximization opportunities while at the same time maintaining safe goals in accordance with the Basel Capital Accords.

This study can be extended by: allowing other distributions for both the return and measurement equation errors, for example, by using asymmetric Laplace or two-sided Weibull distributions; incorporating a higher number of different realized measures; using other data frequencies and subsampling frequency for the realized measures; and employing other MCMC algorithms. Another area to explore is to develop a variable selection and shrinkage method for selecting realized measures in the realized EGARCH models under a Bayesian framework.

Chapter 4

The Realized EGARCH Model with Lasso Approach

4.1 Introduction

There is a rapidly growing literature on realized volatility measures. One of the most important areas using realized volatility measures is the forecasting of the volatility of financial assets. A comprehensive review on the subject can be found at the Realized Library of the Oxford-Man Institute of Quantitative Finance. The realized EGARCH model is one of a few models that benefits from research into realized volatility measures. One of the advantages of the realized EGARCH model is that u_t is K-dimensional, which means the realized EGARCH model can incorporate multiple realized volatility measures. Although this is a very attractive property, as the dimension K grows, the number of the realized EGARCH model parameters would increase $4K$ with innovations Σ. Incorporating many realized volatility measures could then be computationally inefficient.

It is crucial to select a sparse and parsimonious model for forecasting. For the realized EGARCH model, few studies have investigated model selection with respect to the selection of realized volatility measures to include in the measurement equations. The original study of the realized EGARCH by Hansen and Huang (2016) provided model selection by comparing the partial log likelihoods of several realized EGARCH models using multiple realized measures. Further research on this model has mainly incorporated a single realized volatility measures: for example, Wu et al. (2019) used a realized variance, whereas Wu and Xie (2021) used a realized kernel. This chapter aims to fill the gap in the literature by developing methods to select realized volatility measures in the realized EGARCH model for the purpose of obtaining a sparse model for tail risk forecasting.

Variable selection techniques have been rapidly evolving over the past few decades. classical model selection goes back to the work of Akaike (1973, 1974), who proposed using the Kullback-Leibner (KL) divergence of a fitted model from the true model in selecting a model. This work then led to the Akaike Information Criterion (AIC). A Bayesian approach was later proposed by Schwarz (1978) and led to the Schwarz Information Criterion (SIC).

Both the AIC and the BIC suggest a unified approach to choosing a parameter vector θ that maximizes the penalized likelihood.

Another seminal contribution to the model selection method was proposed by Tibshirani (1996), which is called the least absolute shrinkage and selection operator (Lasso). The Lasso is based on a penalized ℓ_1 regression and is essentially a shrunk regression that performs shrinkage and selection. Although shrinkage will result in an increase in bias, penalized regression prevents overfitting (the bias-variance trade-off). Penalized regression is extensively discussed in Hastie et al. (2015).

The Lasso approach is now one of the most popular machine learning techniques in econometrics for high-dimensional estimation problems for several reasons. First, it possess statistical accuracy for prediction and variable selection. Second, the Lasso computation is highly feasible. Third, it offers a general method that can be applied in a wide range of models due to the nature of the Lasso as a penalized likelihood approach (Bühlmann & Van De Geer, 2011).

The estimation of the Lasso estimator requires the penalty or the tuning parameter λ, which governs the intensity of the penalization. The Lasso estimator may not be consistent and have a slow convergence rate if λ is not chosen appropriately (Chatterjee, 2014; Chetverikov et al., 2021). While selection of the penalty parameter is of paramount importance, this task is, however, one the main challenges of the Lasso approach. Many studies have discussed the choice of the penalty parameter, but in practice researchers mainly rely on a frequentist framework using the cross validation (CV) method (see Bühlmann and Van De Geer (2011), Hastie et al. (2009), Hastie et al. (2015) for textbook-level discussions). Another method for obtaining the solution to the penalized regression is based on the Bayesian framework by employing a specific prior. This approach has been proven to perform similarly to, or better than, its frequentist counterparts (Hans, 2009; Kyung et al., 2010; Li & Lin, 2010).

The purpose of this chapter is to perform the selection of the realized volatility measures in the realized EGARCH model incorporating a standardized skewed Student-*t* distribution proposed in Chapter 3 Section 3.3.1 (RE-SkN model). The RE-SkN model with the Lasso approach is hereafter referred to as the **RE-SkN-Lasso model**. This chapter considers both the frequentist and Bayesian approaches in choosing the tuning parameter. A CV approach is developed by adapting Hastie et al. (2009) and Hyndman and Athanasopoulos (2018) to suit the time series data employed in this study, while the Bayesian Lasso (BLasso) approach developed in this chapter refers to Park and Casella (2008). The resulting sparse model is expected to improve tail risk forecasts of financial asset returns. This chapter also aims to compare the performance of the CV and BLasso approaches when choosing the tuning parameter in terms of the forecast performance of the resulting model of each approach.

This chapter is structured as follows. Section 4.2 provides a review of the literature and an overview of penalized regression methods. Sections 4.4 discuss the proposed model selection of the RE-SkN models using the Lasso approach, consisting of the CV and BLasso approaches. The tail forecasting methodology in this chapter is presented in Section 4.5. The

simulation study is presented in Section 4.6, while the empirical study is discussed in Section 4.7. Lastly, the conclusion and possible relevant future studies are discussed in Section 4.8.

4.2 Overview of Penalized Regression

This section provides a brief summary of penalized estimation methods with an emphasis on the Lasso approach. This section also includes a short review on the Bayesian perspective of the Lasso.

The following is a normal linear regression model.

$$y = \mu 1_n + X\beta + \varepsilon, \tag{4.1}$$

where y is the $n \times 1$ vector of response variable, μ is the overall mean, X is the $n \times p$ matrix of explanatory variables, β is the $p \times 1$ vector of regression parameters, and ε is the $n \times 1$ vector of independent and identically distributed normal errors with mean 0 and variance σ^2.

The Ordinary Least Squares (OLS) estimator $\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'y$ is obtained by minimizing the residual sum of squares (RSS) $= (y - X\beta)'(y - X\beta)$. The OLS estimator is well known to be highly unstable in the presence of multicollinearity and produce a non-unique estimator when the number of variables p exceeds the number of observations n. Specifically, Tibshirani (1996) discussed two main reasons why OLS estimates are less satisfactory. First, the OLS estimates have low prediction accuracy since they often have low bias but large variance. The coefficients are sometimes shrunk or set to 0 to improve the prediction accuracy, sacrificing a little bias to reduce the variance of the predicted values. Second, we usually prefer a smaller subset of predictors that demonstrate the strongest effects to make it easier to interpret.

Penalized regression methods have gained increasing attention, especially in estimating coefficients in the high-dimensional setting $p > n$ to improve the accuracy of the OLS prediction. Penalized regression adds a penalty term to the OLS minimization problem that aims to shrink small coefficients towards zero while keeping large coefficients large.

In the early stage, Hoerl and Kennard (1970) developed ridge regression, which is based on ℓ_2 norm regularization that minimizes

$$RSS + \lambda \sum_{j=1}^{p} |\beta_j|^2, \tag{4.2}$$

where $\lambda \geq 0$ is the penalty parameter controlling the amount of shrinkage.

As λ increases, the greater the amount of shrinkage and the coefficients are shrunk toward zero. By this, the ridge regression reduces the complexity of the model and could alleviate multicollinearity (Hastie et al., 2009). However, it does not serve the purpose of variable selection since it keeps all predictors.

The Lasso is another regularization technique that is currently one of the most useful and well studied (Tibshirani, 1996). It is a more recent shrinkage method that can overcome the disadvantage of ridge regression. Lasso estimates are obtained by minimizing RSS with the constraint on the ℓ_1 norm, instead of the ℓ_2 norm, of β.

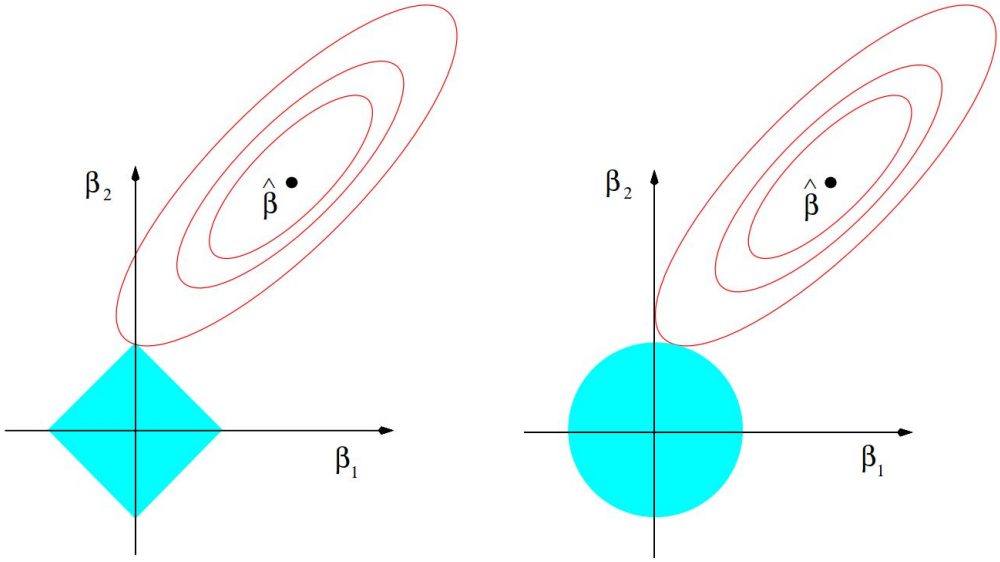
$$RSS + \lambda \sum_{j=1}^{p} |\beta_j|. \tag{4.3}$$

The reason for using the ℓ_1 norm instead of the ℓ_2 norm or any ℓ_q norm is the uniqueness of the ℓ_1 norm. On the one hand, the solutions are not sparse when $q > 1$. On the other hand, $q > 1$ yields sparse solutions, but the problem is non-convex, resulting in a very complex minimization computation. The smallest value that produces a convex problem is $q = 1$. This convexity yields sparse solution vectors and allows for high computational efficiency (Hastie et al., 2015).

Similar to the ridge regression, the Lasso also shrinks the coefficient estimates toward zero. However, the Lasso has several advantages over traditional estimation methods. When the tuning parameter λ is sufficiently large, the ℓ_1 penalty could force some of the coefficient estimates to be equal to zero. This allows us to perform both variable selection and parameter estimation simultaneously (Fan & Li, 2001; Tibshirani, 1996). The Lasso could then produce *sparse* models that only include a subset of the variables (Hastie et al., 2009). The estimated regression coefficients can also have the same efficiency as the oracle procedure because they allow an adaptive amount of shrinkage for each regression coefficient (Wang et al., 2007; Zou, 2006). In addition, the Lasso has excellent computational properties. The Lasso solution path in the OLS setting is piece-wise linear and has a very predictable movement (Efron et al., 2004; Osborne et al., 2000).

Figure 4.1 shows the comparison of Lasso (left panel) and Ridge (right panel) with only two parameters β_1 and β_2. The red elliptical contours are the RSS, while the blue areas are the constraint regions, taking the shape of a diamond for the Lasso and a disk for the ridge. The solutions correspond to the point where the elliptical contours meet with the constraint region. Since the diamond has corners, unlike the disk, then if the Lasso solution occurs at a corner, one parameter β_j is equal to zero. When the number of variables $p > 2$, the diamond has many corners, flat edges and faces; there are chances that the estimated parameters are equal to zero. For details, see Hastie et al. (2009).

FIGURE 4.1: The estimation for Lasso and Ridge



Notes: Left: Contour lines of RSS for the Lasso estimation, with $\hat{\beta}$ being the OLS estimator and the constraint region $|\beta_1| + |\beta_2| \leq t$. Right: Analogous to the left panel but for the Ridge estimation with the constraint region $\beta_1^2 + \beta_2^2 \leq t^2$.

Source: Hastie et al. (2009).

4.3 Methods for tuning parameter selection

This section provides a brief overview of the CV and BLasso methods approaches.

4.3.1 Cross validation

Cross validation (CV) is the simplest and one of the most popular strategies to estimate prediction error of statistical predictor functions. It was first clearly stated by Stone (1974) and is now one of the most popular strategies for algorithm selection widely used in applied machine learning. The main idea behind CV is to split data, once or several times, into the training sample and the validation sample. The training sample is used to train each algorithm, while the validation sample is used to estimate the risk of the algorithm. Then, CV selects the algorithm with the smallest estimated risk (Arlot & Celisse, 2010; Friedl et al., 2002). In this context, the CV method is used to select the tuning parameter that gives the smallest loss.

The common approach to choosing the tuning parameter λ using the CV approach is to choose a grid of λ values. The cross-validation errors for each value of λ are then calculated. The selected λ is the one that gives the smallest cross-validation errors. The final model is refitted using the selected value of the tuning parameter (Friedl et al., 2002; Hastie et al., 2009).

The K-fold cross validation and leave one out cross validation (LOOCV) are the most common forms of CV. In the K-fold CV, the set of observations is randomly divided into K groups or folds of approximately equal size. The first fold is kept as a validation set, and the remaining $K - 1$ folds are treated as a training set to fit a model. The prediction error of the fitted

model is then computed on the validation set. This process is repeated for $k = 1, 2, ..., K$, each time a different fold is treated as a validation set (Alpaydin, 2020; Hastie et al., 2009).

Suppose an indexing function is denoted by $\kappa : \{1, ..., N\} \mapsto \{1, ..., K\}$. This function indicates the partition to which the allocation of observation i is conducted randomly. The fitted function is denoted by $\hat{f}^{-k}(x)$ after removing the set k of the data. The CV estimate of the prediction error is computed as:

$$CV(\hat{f}) = \frac{1}{N} \sum_{i=1}^{N} L(y_i, \hat{f}^{-\kappa(i)}(x_i)). \tag{4.4}$$

Let $f(x, \lambda)$ be a given set of models and $\hat{f}^{-k}(x, \lambda)$ be the λ -th model fit without set k of the data. For this set of models, the CV estimate of the prediction error, which gives the estimate of the test error curve, is given by:

$$CV(\hat{f}, \lambda) = \frac{1}{N} \sum_{i=1}^{N} L(y_i, \hat{f}^{-\kappa(i)}(x_i, \lambda)). \tag{4.5}$$

The tuning parameter $\hat{\lambda} = \arg\min_\lambda CV(\hat{f}, \lambda)$ for the final model is $f(x, \hat{\lambda})$. A detailed description of the K-fold cross-validation can be found in Hastie et al. (2009).

The LOOCV (Geisser, 1975; Stone, 1974)) is a special case of K-fold CV, when the number of folds (K) equals the number of instances (N). This method involves splitting the set of observations into two parts. However, instead of creating two subsets of comparable size, a single observation is successively 'left out' and used for the validation, while the remaining observations are used as the training set. Under the LOOCV, $\kappa(i) = i$ and the fit estimation for the i-th observation is carried out using all the data after excluding the i-th (Arlot & Celisse, 2010; Hastie et al., 2009).

The computational complexity of the CV method increases as the number of validation sets increases (Shao, 1993). When the data size is large and the model takes a substantial time to fit, using the LOOCV is then computationally expensive. In this respect, the k-fold CV has a computational advantage over LOOCV, especially when $k << N$. The test error rate produced by the k-fold CV is also more accurate than that of LOOCV. Another aspect to consider is the bias-variance trade-off. The estimators produced by LOOCV are approximately unbiased, while those of the k-fold CV have some bias, especially when using small k. However, the LOOCV can have a high variance while the k-fold CV can compromise the bias and variance with the recommended $k = 5$ or $k = 10$ (Hastie et al., 2009).

4.3.2 Bayesian Lasso

Rapid development in Bayesian computation has motivated the implementation of a Bayesian framework in penalized variable selection methods. There are three main advantages to adopting a Bayesian approach to penalized regression, as discussed in Van de Schoot and Miocević (2020) and Van Erp et al. (2019). The first advantage is the natural penalization

in a Bayesian framework by choosing a specific parametric form for the prior that could shrink less significant coefficients towards zero and at the same time retain the substantial coefficient as large. Shrinkage priors peak at zero and usually have heavy tails, allowing large coefficients to escape the shrinkage (see Van de Schoot and Miocević (2020) for discussion of different shrinkage priors). In this case, the prior has the same function as the penalty term in the frequentist penalization methods. When combined with a specific posterior estimate, some types of prior distributions could result in identical solutions as the classical penalized regression, for instance, the double exponential prior distribution discussed in Park and Casella (2008). Kyung et al. (2010) and Li and Lin (2010) even argued that this Bayesian Lasso (BLasso) version could have a more favorable performance than the classical penalization methods. The second advantage is the simultaneous estimation of the penalty parameter with other model parameters. The third advantage is the flexibility of the Bayesian penalization with respect to the type of penalties. The frequentist penalization methods usually consider convex penalty functions and rely on optimization methods to obtain the minimum value of the penalized regression function, whereas the BLasso relies on MCMC sampling, which facilitates easier implementation of non-convex penalties.

Under the Bayesian Lasso framework, the Lasso estimates are equivalent to posterior mode estimates when the regression parameters θ are assigned independent and identical Laplace (double exponential) priors (Tibshirani, 1996). The Laplace distribution has a sharper peak and heavier tails than the normal one. The most interesting feature of the advantages of the Laplace distribution is that it can be represented as a scale mixture of normal distributions with independent exponentially distributed variances, as shown in Andrews and Mallows (1974):

$$\frac{\lambda}{2}e^{-\lambda|z|} = \int_o^\infty \frac{1}{\sqrt{2\pi s}}e^{-z^2/(2s)}\frac{\lambda^2}{2}e^{-\lambda^2 s/2}ds. \tag{4.6}$$

The tuning parameter λ can be selected using either an empirical Bayesian (EB) approach or a hierarchical Bayes approach. In the EB setting, the tuning parameter λ is estimated via marginal maximum likelihood, whereas the hierarchical Bayes approach assigns appropriate hyperpriors on the λ, which makes it possible to draw posterior inference on the tuning parameter (Leng et al., 2014; Park & Casella, 2008). When the marginal likelihood for the tuning parameter is not available in a closed form, a multistep method by using expectation-maximization (EM) algorithm can be employed to solve this problem (Casella, 2001). However, this involves running many Gibbs samplers, hence may be computationally inefficient. The alternative method, hierarchical Bayes, treats λ as a random variable by assigning an appropriate prior on λ^2 (Leng et al., 2014).

The hierarchical Bayes method offers several advantages over the EB method (Berger, 2013; Ntzoufras, 2011; Robert, 2007). First, unlike the EB approach, hierarchical Bayes automatically incorporates hyperparameter estimation errors. Second, the hierarchical approach also allows the use of both structural prior information at the first stage of the stage structure and subjective prior information at the second stage. Third, the hierarchical structure improves

the robustness of the resulting estimators, since it can reduce the arbitrariness of the hyperparameter choice and the averaging of the posterior results across different prior choices of parameters of interest. Fourth, while the solution of likelihood equations is required in the EB theory, numerical integration such as MCMC algorithms is required in the hierarchical Bayes approach and this yields conditional distributions. Lastly, the hierarchical structure provides a simpler interpretation and model computation due to simplified posterior distribution.

The attractive feature of the Laplace distribution combined with the advantage of hierarchical Bayes has motivated a number of studies to use the Laplace prior in hierarchical Bayesian approach. The Laplace prior was incorporated by Figueiredo (2003) in an expectation maximization (EM) training algorithm that yields a maximum a posteriori (MAP) estimate. The Laplace prior was also used in the sparse logistic regression model in the gene selection problem, for example, in sparse logistic regression (SLogReg) by Shevade and Keerthi (2003) and a competing sparse logistic regression with Bayesian regularization (BLogReg) by Cawley and Talbot (2006). Bae and Mallick (2004) constructed two levels of a hierarchical Bayesian model involving the Laplace prior and used an MCMC technique in variable selection applied to a leukemia and breast cancer dataset. Yuan and Lin (2005) applied the Laplace approximation in developing an empirical Bayes method to compute Bayesian estimates with a quick Lasso algorithm.

Park and Casella (2008) proposed a fully hierarchical Bayesian Lasso (BLasso) model by exploiting the representation of the Laplace distribution as a scale mixture of normal (SMN) distributions with an exponential mixing density (Andrews & Mallows, 1974). The proposed Gibbs sampling uses a conditional Laplace prior that takes the following specification:

$$\pi(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} \exp\left\{-\frac{\lambda|\beta_j|}{\sqrt{\sigma^2}}\right\} \tag{4.7}$$

and putting a non-informative scale-invariant marginal prior on σ^2, i.e. $\pi(\sigma^2) \propto 1/\sigma^2$.

The full BLasso model proposed by Park and Casella (2008) consists of the following hierarchical representation:

$$\begin{aligned} y|\mu, X, \beta, \sigma^2 &\sim N_n(\mu 1_n + X\beta, \sigma^2 I_n), \\ \beta|\sigma^2, \tau_1^2, \ldots, \tau_p^2 &\sim N_p(0_p, \sigma^2 D_\tau), \\ D_\tau &= \text{diag}(\tau_1^2, \ldots, \tau_p^2), \\ \sigma^2, \tau_1^2, \ldots, \tau_p^2 &\sim \pi(\sigma^2) \prod_{j=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2/2}, \\ \sigma^2, \tau_1^2, \ldots, \tau_p^2 &> 0. \end{aligned} \tag{4.8}$$

Park and Casella (2008) suggested using the improper prior density $\pi(\sigma^2) \propto 1/\sigma^2$ to model the error variance, while the paramater μ could be assigned an independent, flat prior. As $\lambda^2 \to \infty$, the prior distribution for λ^2 approaches 0 sufficiently fast but relatively flat.

4.4 The Selection of the Realized Volatility Measures in the RE-SkN-Lasso Model

This research aims to perform the selection of realized volatility measures in the realized EGARCH models employing a standardized skewed Student-t distribution in the return equation (the RE-SkN model) described in Section 3.3.1. The selection of variables is carried out using the Lasso approach (RE-SkN-Lasso model). A crucial step in the Lasso approach is choosing the optimal tuning parameter that will lead to the best-performing model to use in tail risk forecasting. Hereafter, the tuning parameter for the RE-SkN-Lasso model is denoted by λ_l to distinguish it from the skewness parameter, λ. Two methods are compared in choosing the optimal tuning parameter: the CV and the hierarchical Bayesian Lasso (BLasso) approach.

4.4.1 Penalized log likelihood

Naturally, penalized least squares can be extended to likelihood-type models (see Fan and Li (2001) for a discussion on variable selection via penalized likelihood). The model parameters are obtained by minimizing the negative penalized likelihood function with respect to β.

For the case of the RE-SkN model with Lasso, suppose that the parameter vector of the RE-SkN model is $\theta = (\eta, \gamma,)$, where $\eta = (\mu, \omega, \beta, \tau_1, \tau_2, \xi_1, .., \xi_K, \varphi_1, .., \varphi_K, \delta_{1,1}, .., \delta_{1,K}, \delta_{2,1}, .., \sigma_{u,1}, .., \sigma_{u,K})^T$ and $\gamma = (\gamma_1, .., \gamma_K)^T$. The parameters of the RE-SkN model are estimated by minimizing the following loss function, which is based on the penalized log-likelihood function:

$$\mathcal{L}(\theta) + \sum \lambda_l |\gamma_k|, \tag{4.9}$$

where $\mathcal{L}(\theta)$ is the negative log-likelihood function of the RE-SkN model in Equation 3.22, $\lambda_l > 0$ is the tuning parameter, and $k = 1, ..., K$ is the realized volatility measures. Since the focus of this study is to select realized measures, the penalties are then only imposed on γ_k parameters, which are the reflection of the strength of information provided by the realized measures about future volatility. When a certain γ_k coefficient is shrunk to 0, the particular realized measure is not sufficiently informative to be included in the model and can then be eliminated.

4.4.2 Cross validation method for Lasso RE-SkN model

The proposed frequentist procedure to select the RE-SkN-Lasso model combines the time series CV approach and the rolling window approach. This, hereafter, is called the **RE-SkN-Lasso-CV model**.

This study proposes two main steps to select the realized measures for the RE-SkN-Lasso model. The first step is to calculate the predictive log-likelihood for each proposed tuning parameter $\lambda_l \in \mathcal{S}_\tau$. The second step is to select the λ_l parameter from a range of possible values of the λ_l parameter by adapting the k-fold CV technique of Hastie et al. (2009) and

the time series CV of Hyndman and Athanasopoulos (2018) to suit the time series data employed in this study. The chronological order of the time series data is maintained so as not to lose the autoregressive dynamic of the RE-SkN model.

The setting of the training and validation samples in this study is adjusted from the usual CV method, which requires that the data be i.i.d. (see Arlot and Celisse (2010)). The purpose is to preserve the temporal dependence between observations in the high-frequency time-series data employed in this study.

Suppose that we have a total sample of $T + M + N$ observations. The dataset is divided into three: training sample (T observations); validation sample (M observations); and forecast sample (N observations). With the rolling approach, the rolling sample is constant, T, for every estimation window. Observations $T + 1$ to $T + M$ are used as a validation sample, while observations numbers $T + M + 1$ to $T + M + N$ are used as a forecast sample.

The selection of the tuning parameter can be conducted by searching over a given finite candidate set that will produce the smallest predictive risk (Lei, 2020). In this study, the proposed candidate set covers a wide range of possible tuning parameter values, $\lambda_l \in \mathcal{S}_\tau$: $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 1.5, .., 20, 21, .., 40, 50, .., 100, 200, .., 3000, 4000, .., 20000\}$.

The selection of the best RE-SkN model across different tuning parameters is performed using the predictive log likelihood, $\ell(r;\theta)$, since it measures the fit for the return series, as in Hansen et al. (2012) and Hansen and Huang (2016). The RE-SkN model that minimizes the loss is chosen as the final model, which is equivalent to the model with the highest predictive log-likelihood.

A rolling approach with a fixed T in-sample observation is employed to calculate predictive log-likelihoods for M validation samples as follows:

1. Pick $\lambda_l = 0.001$.
2. Based on sample data $1, ..., T$, the Realized EGARCH model is estimated by minimizing the negative log-likelihood plus the penalty in Equation 4.9 by plugging λ_l in Step 1.
3. Obtain the parameter of the Realized EGARCH model, $\hat{\theta}_T$, using the in-sample information $\mathcal{F}_T$, then generate $u_{k,T+1}$, z_{T+1}, and h_{T+1}.
4. Calculate the predictive log-likelihood for the return equation for period T+1, ℓ_{T+1}

$$\ell_{T+1}(r;\theta_T) = \begin{cases} T[\ell(b)+\ell(c)] - \sum_{t=1}^{T}\left[\frac{v+1}{2}\log\left(1+\frac{1}{v-2}\left(\frac{b\epsilon_t+a}{1-\lambda}\right)^2\right)+0.5\log(h_t)\right] & \text{if } \epsilon_t < -\frac{a}{b}, \\ T[\ell(b)+\ell(c)] - \sum_{t=1}^{T}\left[\frac{v+1}{2}\log\left(1+\frac{1}{v-2}\left(\frac{b\epsilon_t+a}{1+\lambda}\right)^2\right)+0.5\log(h_t)\right] & \text{if } \epsilon_t \geq -\frac{a}{b}. \end{cases}$$

where ν is the degree-of-freedom parameter that controls the kurtosis of the distribution with range $2 < \nu < \infty$, λ is the skewness parameter with range $-1 < \lambda < 1$, while a, b, and c are some constants given by:

$$a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right), \quad b = \sqrt{1+3\lambda^2-a^2}, \quad c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}.$$

5. Repeat Steps 1 to 4 with the rolling approach using a fixed in-sample of T observation to calculate ℓ_{T+2} until ℓ_{T+M}.
6. These log-density estimates are summed to estimate predictive log-likelihood:

$$\sum_{i=T+1}^{T+M} \ell_i(r;\theta_{i-1}). \tag{4.10}$$

7. For each other λ_l parameter, repeat Steps 2 to 6.
8. Select an optimum tuning parameter based on the above CV method, denoted as λ_l^{CV}, which is the λ_l parameter that produces the highest predictive log-likelihood.

The selected RE-SkN model is then used for tail risk forecasting (Section 4.5).

4.4.3 Bayesian Lasso approach for the RE-SkN model

In developing the Bayesian Lasso approach for the RE-SkN model, this study adopts the hierarchical Bayes approach (hereafter referred to as the **RE-SkN-BLasso model**) instead of the EB approach for the RE-SkN Lasso model due to the advantage of the hierarchical Bayes approach over the EB approach discussed in Section 4.3.2.

Unlike the CV approach, the BLasso approach does not use the validation sample, only the estimation sample and the forecast sample. To obtain a consistent dataset with the one used for the RE-SkN-Lasso-CV model, the BLasso approach uses a constant rolling sample, T, for every estimation window, while observations numbers $T+M+1$ to $T+M+N$ are used as the forecast sample.

Prior and posterior distributions

Consider the penalized log-likelihood function in Equation 4.9. Inspired by Park and Casella (2008), the Laplace prior specification applied on γ is rewritten as a scale mixture of normal:

$$\pi(\gamma) = \prod_{j=1}^{p} \frac{\lambda_l}{2} \exp\left\{-\lambda_l|\gamma_j|\right\}, \tag{4.11}$$

$$\gamma \mid \tau_1^2,..,\tau_k^2, \sim N_K(0_K, D_\tau),\ D_\tau = \text{diag}(\tau_1^2,..,\tau_K^2), \tag{4.12}$$

$$\tau_{l,1}^2,..,\tau_{l,k}^2 \mid \lambda_l^2 \sim \prod_{k=1}^{K} \frac{(\lambda_l^2)}{2} e^{-(\lambda_l^2)\frac{(\tau_k^2)}{2}}. \tag{4.13}$$

In order to choose λ_l, a class of gamma priors on λ_l^2 (not λ_l) is considered in Equation 4.14 for convenience, due to the way λ_l enters the posterior.

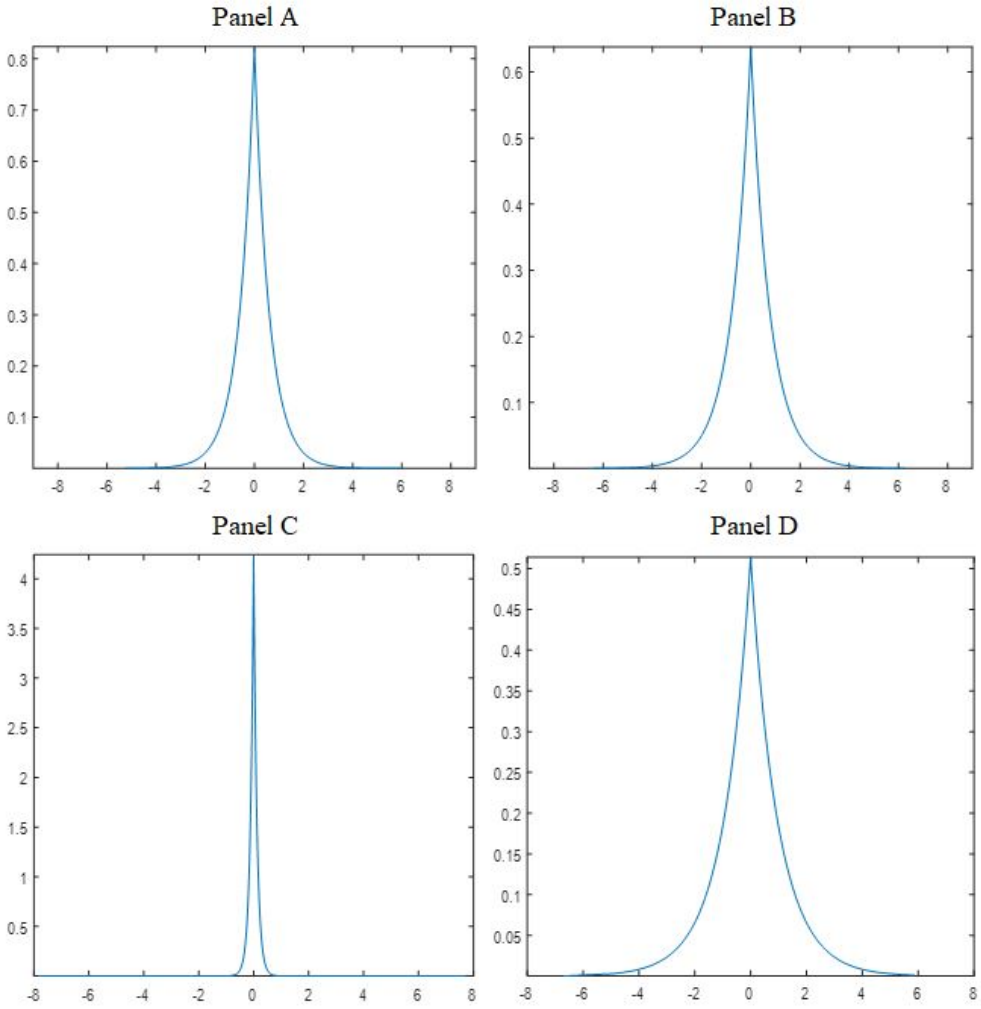
$$\pi(\lambda_l^2) = \frac{\delta^r}{\Gamma(r)}(\lambda_l^2)^{r-1}e^{-\delta\lambda_l^2}, \qquad \lambda_l^2 > 0 (r > 0, \delta > 0), \tag{4.14}$$

where r and δ are the hyperparameters: r is the shape parameter and δ is the rate parameter. Please note that the hypermarameter δ and the variance $\tau_{l,k}^2$ in Equation 4.13 are different notation from $\delta_{(k)}(\epsilon_t)$ and τ_1 and τ_2, respectively, in the leverage functions of the RE-SkN model in Equation 3.11.

While the hyperparameters r and δ could be fixed to some small values to obtain a flat prior, this study examines various combinations of the hyperparameters to obtain more appropriate prior distributions of λ_l^2 and γ. The combination of hyperparameter considered ranges of $r = \{0.1, 0.5, 1, 2, 3, 4, 5, 10\}$ and $\delta = \{0.1, 0.5, 1, 1.78, 2, 3, 4, 5, 10\}$ gives 72 possible combinations of r and δ. These candidate sets include the combination of $r = 1$ and $\delta = 1.78$ (Park & Casella, 2008) and $r = 1$ and $\delta = 10$ (Kyung et al., 2010).

Figure 4.2 presents several plots of prior densities for γ^2 (due to limited space, only a few plots of prior densities for γ are presented in this thesis). Panels A ($r = 0.1$ and $\delta = 0.1$) and B ($r = 1$ and $\delta = 1.78$) provide a reasonable and expected range of γ parameter, that is, mainly between -2 and 2. Panel C ($r = 10$ and $\delta = 0.1$) gives a too narrow range of γ, while Panel D ($r = 1$ and $\delta = 10$) gives a too wide range of γ, between -6 and 6. The chosen hyperparameters are $r = 1$ and $\delta = 1.78$ (Panel B), since these combinations are consistent with the hyperpriors used in the study of Park and Casella (2008). This study also finds that the width of the γ parameter is rather sensitive to the change of δ and less sensitive to the change of r. The higher is δ, the wider the interval of the γ values.

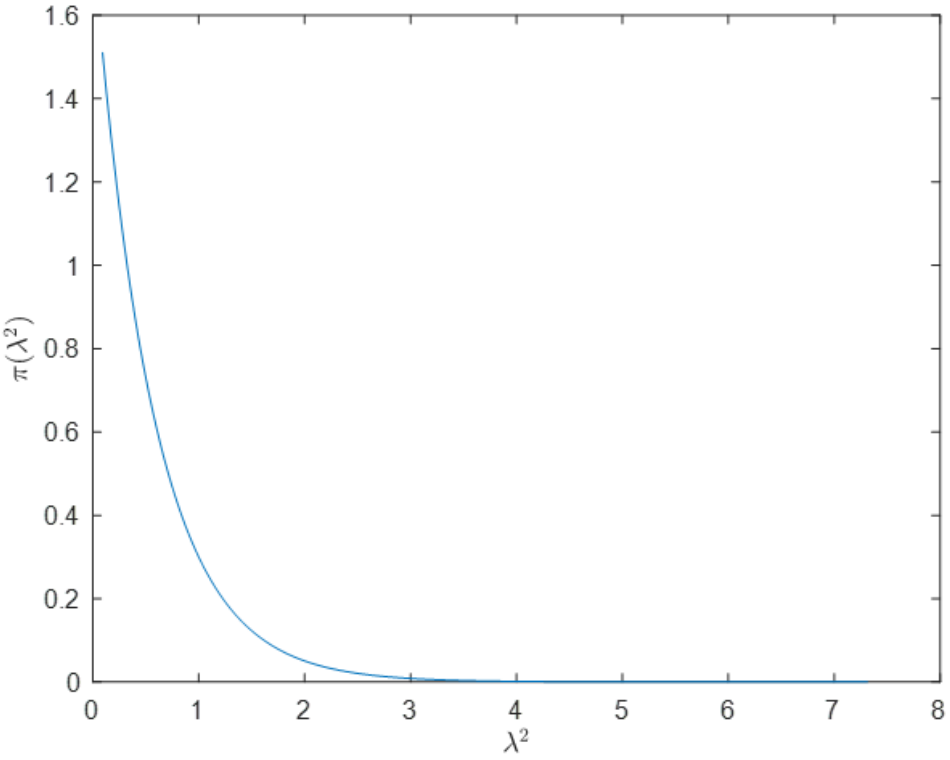
FIGURE 4.2: The prior distribution of γ



Notes: Panel A: $r = 0.1$ and $\delta = 0.1$, Panel B: $r = 1$ and $\delta = 1.78$, Panel C: $r = 10$ and $\delta = 0.1$, and Panel D: $r = 10$ and $\delta = 10$.

The plots of the prior densities for λ_l^2 based on the selected hyperparameters $r = 1$ and $\delta = 1.78$ are depicted in Figure 4.3. It can be seen that the prior densities for λ_l^2 die out approaching 0 very fast as $\lambda_l^2 \to \infty$.

FIGURE 4.3: The prior on λ_l^2 with $r = 1$ and $\delta = 1.78$



For other RE-SkN-Lasso model parameters η, a combination of mostly uninformative and Jeffreys-type priors, is chosen over the possible region for the parameters.

$$\pi(\eta) \propto \frac{1}{\sigma^2_{u,k=1}} \frac{1}{\sigma^2_{u,k=2}} .. \frac{1}{\sigma^2_{u,k=K}} \frac{1}{\nu^2} I(A) \tag{4.15}$$

where A is the region described by the parameter restriction in Table 4.1. The indicator function $I(B) = 1$ over region A and zero otherwise.

TABLE 4.1: Parameter Restriction of BLasso RE-SkN model

Parameter	Bounds
$\mu, \omega, \tau_1, \tau_2, \xi, \delta_1, \delta_2$	$\pm\infty$
σ_u^2	> 0
ν	$(4, 200)$
λ	$(-1, 1)$
$\beta - \gamma^{'}\varphi$	< 1
τ_l^2	> 0
λ_l^2	> 0

The joint prior distribution is:

$$\pi(\gamma, \eta, \tau_l^2, \lambda_l^2) \propto \pi(\gamma|\tau_l^2)\pi(\eta|r, x; \gamma)\pi(\tau_l^2|\lambda_l^2)\pi(\lambda_l^2), \tag{4.16}$$

and the joint posterior distribution for all parameters is obtained by multiplying the penalized likelihood, $L(r, x|\gamma, \eta)$ and the joint prior in Equation 4.16

$$\begin{aligned}\pi(\gamma, \eta, \tau_l^2, \lambda_l^2|r, x) &\propto L(r, x|\gamma, \eta) \cdot \pi(\gamma, \eta, \tau_l^2, \lambda_l^2) \\ &\propto L(r, x|\gamma, \eta) \cdot \pi(\gamma|\tau_l^2)\pi(\eta|r, x; \gamma)\pi(\tau_l^2|\lambda_l^2)\pi(\lambda_l^2).\end{aligned} \tag{4.17}$$

Adaptive MCMC method

Analytically deriving the posterior distribution for the above BLasso hierarchy for the RE-SkN model requires calculating intractable integrals. Since the posterior distribution in Equation 4.17 does not have an explicit form, this study proceeds with sampling techniques based on the proposed Adaptive MCMC in Section 3.5.3 to draw parameters from the posterior distribution.

In the burn-in period, a Robust Adaptive Metropolis (RAM) algorithm of Vihola (2012) is employed. In the post-burn-in period, a standard RWM algorithm is employed and a mixture of three Gaussian proposal distributions is used. The parameter block is adjusted to take into account the additional Lasso-related parameters. The parameter block consists of five blocks: $\theta_1 = \mu$, $\theta_2 = (\omega, \tau', \xi_k, \delta_k', \Sigma)$, $\theta_3 = (\beta, \gamma', \varphi)$, $\theta_4 = (\nu, \lambda)$, and $\theta_5 = ((\tau_k^2)', \lambda_l)$.

The posterior means of the parameters are calculated. The optimal tuning parameter based on the BLasso method is denoted by λ_l^{Bayes}, which is the posterior mean of the post-burn-in iterates, given a γ prior with the chosen hyperparameters $r = 1$ and $\delta = 1.78$.

4.5 Tail Risk Forecasting

After obtaining the tuning parameter, we move on to the next step, that is, forecasting tail risks for the last N observations, that is, observations $T + M + 1$ to $T + M + N$. The details of the forecast step are as follows:

1. For $i = 1$, the RE-SkN model is estimated for observation $M + i$ to $T + M + i - 1$, by maximizing

$$\mathcal{L}(\theta) + \sum \lambda_l|\gamma_j|. \tag{4.18}$$

2. Obtain the parameter of the RE-SkN model, $\hat{\theta}_{T+M}$, using the in-sample information $\mathcal{F}_{T+M}$, then generate $u_{k,T+M+1}$, z_{T+M+1}, and h_{T+M+1}.

3. Estimate tail risks, VaR_{T+M+1} based on Equation 3.18 and ES_{T+M+1} based on Equation 3.19 from the α-percentile of r_{T+M+1}, $\alpha = 2.5$ % and $\alpha = 1$ %.

4. Repeat Steps 1 to 3 above for $i = 2$ up to $i = M$ with the rolling approach using a fixed in-sample of T observation to estimate VaR_{T+M+2} up to VaR_{T+M+N} and ES_{T+M+2} up to ES_{T+M+N}.

For the CV method, the same value of the tuning parameter λ_l^{CV} is used in Equation 4.18 to perform the forecast for N observations. Meanwhile, the resulting tuning parameter by the BLasso approach λ_l^{Bayes} changes for different sets of data estimated during the rolling window forecasting.

4.6 Simulation Study

A simulation study is conducted to compare the performance of the proposed CV and the BLasso approaches in choosing the tuning parameters λ_l as well as the selection of the realized volatility measures in the RE-SkN model.

4.6.1 Simulation set up and model

A total sample of 3000 observations is simulated from the proposed RE-SkN model in Equation 4.19 using four realized measures. In this model, $\gamma_2 = 0$ and $\gamma_4 = 0$.

$$\begin{aligned} r_t &= -0.11 + \sqrt{h_t}\epsilon_t \\ \log h_t &= 0.05 + 0.98 \log h_{t-1} - 0.12\epsilon_{t-1} + 0.04(\epsilon_{t-1}^2 - 1) + 0.47u_{1,t-1} + 0.25u_{3,t-1} \\ \log x_{1,t} &= -0.17 + 0.93 \log h_t - 0.09\epsilon_t + 0.06(\epsilon_t^2 - 1) + u_{1,t} \\ \log x_{2,t} &= -0.75 + 0.94 \log h_t - 0.11\epsilon_t + 0.08(\epsilon_t^2 - 1) + u_{2,t} \\ \log x_{3,t} &= -0.13 + 0.95 \log h_t - 0.08\epsilon_t + 0.10(\epsilon_t^2 - 1) + u_{3,t} \\ \log x_{4,t} &= -0.80 + 0.96 \log h_t - 0.13\epsilon_t + 0.12(\epsilon_t^2 - 1) + u_{4,t}, \end{aligned} \tag{4.19}$$

where $\epsilon_t \sim Skt_{\nu=5,\lambda=0.5}$, $u_t \sim N(0, \sigma_{u,k}^2)$, $\sigma_{u,1}^2 = 0.15, \sigma_{u,2}^2 = 0.18, \sigma_{u,3}^2 = 0.12, \sigma_{u,4}^2 = 0.20$ and $h_0 = 0.0025$.

A rolling window approach with a fix T=2000 in-sample observations is employed to estimate the RE-SkN-Lasso model. For the CV method, observations t_{2001} to t_{2500} are used as M=500 validation samples, while the remaining observations N=500, t_{2501} to t_{3000}, are used as the forecast sample. For the BLasso method, a fixed T=2000 in-sample observations starting from $t_5$01 is used with a rolling window approach. The forecast sample is N=500, which are observations t_{2501} to t_{3000}.

The choice of the optimum tuning parameter λ_l is performed by using both the CV approach proposed in Section 4.4.2 and the BLasso approach proposed in Section 4.4.3. Under the CV approach, the optimal λ_l is chosen from a range of $\lambda_l \in \mathcal{S}_\tau$: $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 1.5, .., 20, 21, .., 40, 50, .., 100, 200, .., 3000, 4000, .., 20000\}$. Meanwhile, the optimal λ_l under the proposed BLasso approach is the posterior mean of the λ_l computed over the post-burn iterates.

4.6.2 Simulation procedures

The estimation of the RE-SkN model using CV and BLasso approaches requires high computing power and substantial computing time. To manage this challenge, the University of Sydney's Artemis High-Performance Computing (HPC) is used. It allows the use of several computers simultaneously through the network. The estimation of the RE-SkN model using the CV approach is broken down into 117 separate estimations of the RE-SkN models, each of which has a different value of the tuning parameter $\lambda_l \in \mathcal{S}_\tau$: $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 1.5, .., 20, 21, .., 40, 50, .., 100, 200, .., 3000, 4000, .., 20000\}$.
The estimation of the RE-SkN model by using the BLasso approach is divided into 50 separate estimations, each of which consists of 10 rolling window estimations and tail risk forecasts.

The summary of the estimation time of both methods is presented in Table 4.2. The total estimation time of the CV method over all range of tuning parameter values is much higher than the total estimation time of the BLasso method.

TABLE 4.2: Estimation time of the simulated data

	CV	**BLasso**
Number of estimation	17	50
The shortest estimation time	2:02:19	3:01:04
The longest estimation time	3:59:24	4:42:42
Average estimation time	2:53:02	3:46:09
Total estimation time	337:25:44	188:27:34

4.6.3 Simulation results: MCMC convergence

To assess the convergence of the MCMC algorithm, the simulated dataset generated using the model in Section 4.6.1 is estimated by employing the Robust Adaptive Metropolis (RAM) algorithm in the burn-in period and the standard Random Walk Metropolis (RWM) in the post-burn-in period. The MCMC is run with the following set-up: $m = 7$ independent chains with different starting values were run.

Initially, the MCMC algorithm was run for 20,000 iterations in the burn-in period and 10,000 iterations in the post-burn-in period. Parameters ξ and φ are found to take quite a long time to converge, so the iteration in the burn-in period is extended to 40,000 to improve the convergence. The number of iterations in the post-burn-in is observed until 20,000 iterations, but it is kept at 10,000 iterations in the final simulation, since the parameters have converged and mixed well after more than 20,000 iterations in the burn-in period.

In Table 4.3, the Gelman-Rubin statistics ($\hat{R}$) are very close to 1, suggesting convergence. The effective number of independent simulation draws, n_{eff}, are higher than 5m (the reference benchmark suggested by Gelman et al. (2014), confirming that the algorithm is efficient. The autocorrelation time, $\hat{t}$, for most parameters is quite small. Parameters ξ tend to have the longest autocorrelation time; they take longer to converge compared to other parameters.

TABLE 4.3: Summary statistics of the Gelman-Rubin diagnostic, effective sample size and autocorrelation time of the RE-SkN-BLasso model for the simulated dataset

Parameters	$\hat{R}$	$n_{\hat{eff}}$	$\hat{t}$
μ	1.0092	708.89	28.43
ω	1.0270	156.63	316.60
β	1.0325	163.65	302.97
γ_1	1.0020	2201.41	17.39
γ_2	1.0050	1104.48	42.51
γ_3	1.0018	1646.57	26.12
γ_4	1.0015	1970.95	22.97
τ_1	1.0219	473.33	99.55
τ_2	1.0534	103.76	459.00
ξ_1	1.1254	56.86	870.55
ξ_2	1.0862	55.18	901.90
ξ_3	1.0948	58.30	844.88
ξ_4	1.0959	69.14	709.58
φ_1	1.0428	171.52	250.14
φ_2	1.0228	161.72	279.60
φ_3	1.0448	162.55	282.54
φ_4	1.0248	183.62	236.99
$\delta_{1,1}$	1.0213	652.02	59.34
$\delta_{1,2}$	1.0046	630.15	58.92
$\delta_{1,3}$	1.0085	731.04	64.93
$\delta_{1,4}$	1.0052	571.39	73.20
$\delta_{2,1}$	1.0628	98.03	488.27
$\delta_{2,2}$	1.0529	83.08	588.10
$\delta_{2,3}$	1.0857	65.57	753.44
$\delta_{2,4}$	1.0634	77.69	636.66
$\sigma^2_{u_1}$	1.0033	1639.05	28.92
$\sigma^2_{u_2}$	1.0046	1213.94	35.57
$\sigma^2_{u_3}$	1.0019	1658.33	23.60
$\sigma^2_{u_4}$	1.0038	967.73	34.93
ν	1.0027	1529.16	23.82
λ	1.0018	8413.37	4.01
τ_{l_1}	1.0066	845.94	53.86
τ_{l_2}	1.0099	691.88	66.48
τ_{l_3}	1.0113	622.04	75.91
τ_{l_4}	1.0141	350.26	139.10
λ_l^{Bayes}	1.0084	1074.07	39.29

Trace plots of all parameters are also examined. However, due to space concerns, only the trace plots of the main parameters are presented in this chapter. These are γ_1 (Figure 4.4), γ_2 (Figure (4.5), γ_3 (Figure 4.6), γ_4 (Figure 4.7), and the tuning parameter λ_l (Figure 4.8. The trace plot on the left shows the burn-in iterations using the RAM algorithm, while the plot on the right shows the post-burn-in iterations using the RWM algorithm. All trace plots show that the MCMC chains rapidly converge to the stationary distribution and have a good chain mixing property.

FIGURE 4.4: Trace plots of of γ_1

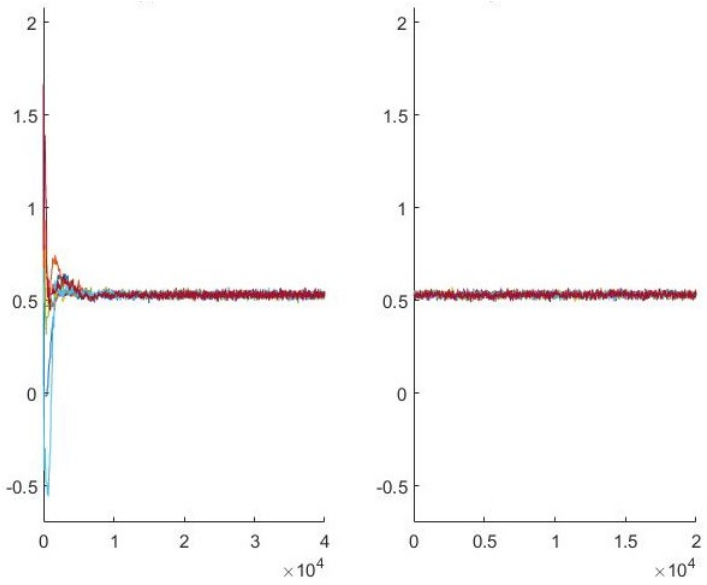


FIGURE 4.5: Trace plots of of γ_2

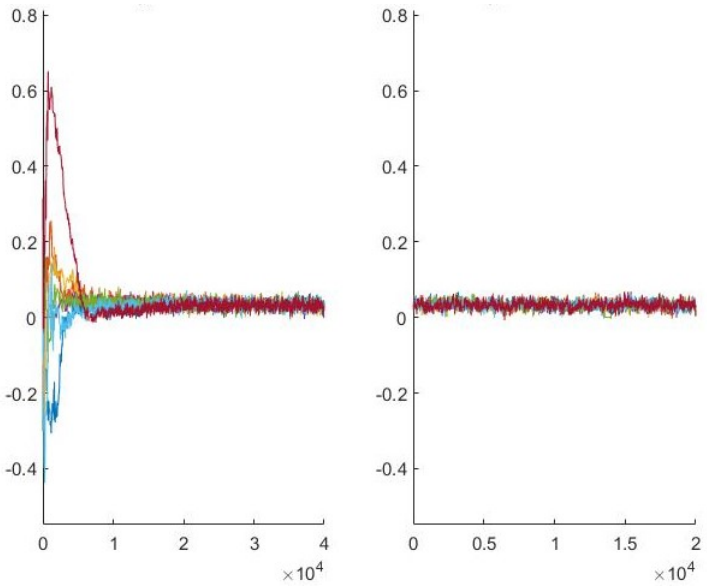


FIGURE 4.6: Trace plots of of γ_3

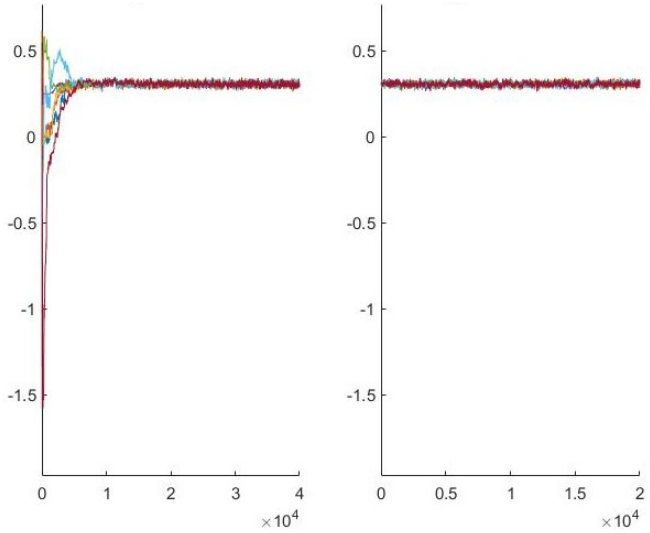


FIGURE 4.7: Trace plots of of γ_4

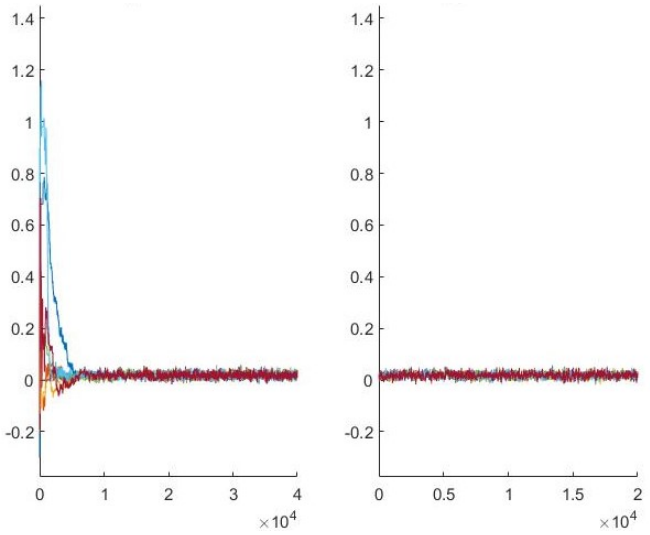
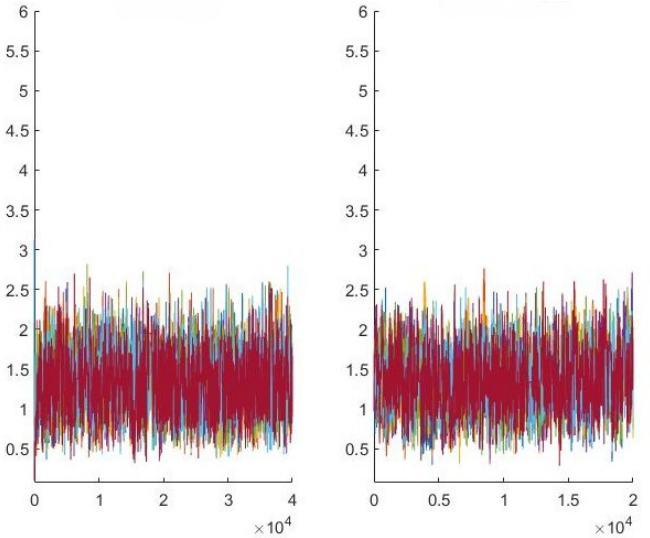


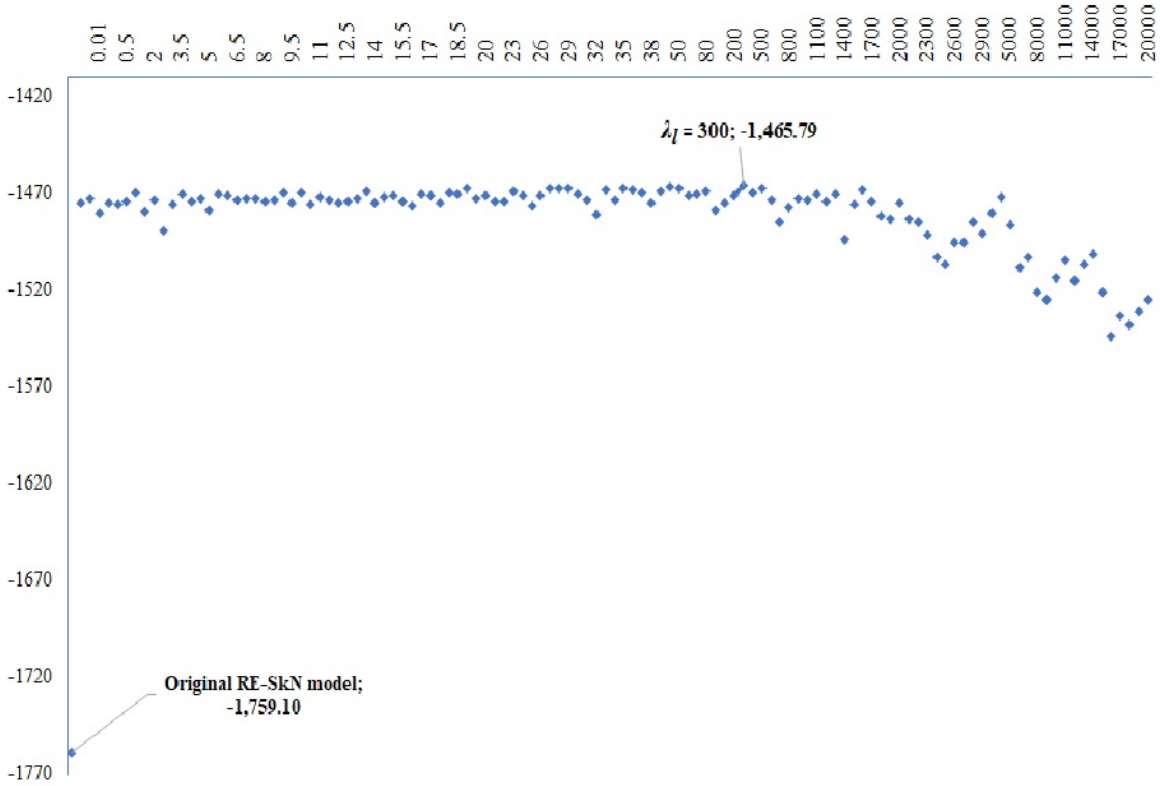
FIGURE 4.8: Trace plots of of λ_l



4.6.4 Simulation results: model parameters

This section presents the resulting parameters of the RE-SkN Lasso model both using the CV and the BLasso approaches. Under the CV approach, the plot of the predictive log-likelihood of the return, $\ell(r;\theta)$, over the validation set is shown in Figure 4.9. In general, the RE-SkN model with a penalty parameter produces a higher $\ell(r;\theta)$ than the RE-SkN model without a penalty, suggesting that predictive likelihood analysis favors the RE-SkN model with the Lasso approach compared to the original RE-SkN model. The optimum tuning parameter, which yields the highest $\ell(r;\theta)$ is $\lambda_l = 300$. Nevertheless, the choice of λ_l values up to $\lambda_l = 2000$ gives relatively equally high values of $\ell(r;\theta)$. The $\ell(r;\theta)$ starts to show a decreasing trend for $\lambda_l > 1700$. It could be argued that any value of the tuning parameter in the proposed set to $\lambda_l < 1700$ yields relatively high $\ell(r;\theta)$, much greater than that of the RE-SkN model without penalty.

FIGURE 4.9: Predictive log-likelihood using the CV approach



The estimated parameters of the RE-SkN-Lasso model using both CV and BLasso approaches are presented in Table 4.4. The estimated parameters of the original RE-SkN model estimated using Maximum Likelihood are also presented for comparison purposes in column 'No Lasso.' The resulting parameters of the RE-SkN-Lasso model using the CV approach is based on the optimal $\lambda_l^{CV} = 300$, are presented in column 'With Lasso (CV)', while those using the BLasso approach are presented in column 'BLasso'. The optimal tuning parameter based on the BLasso approach is $\lambda_l^{Bayes} = 1.4120$, which is the posterior mean of the post-burn-in iterates, given a γ prior with the chosen hyperparameters $r = 1$ and $\delta = 1.78$.

Although the tuning parameters are largely different, the predictive log-likelihood when using small values of λ_l under the CV approach is still high, as shown in Figure 4.9.

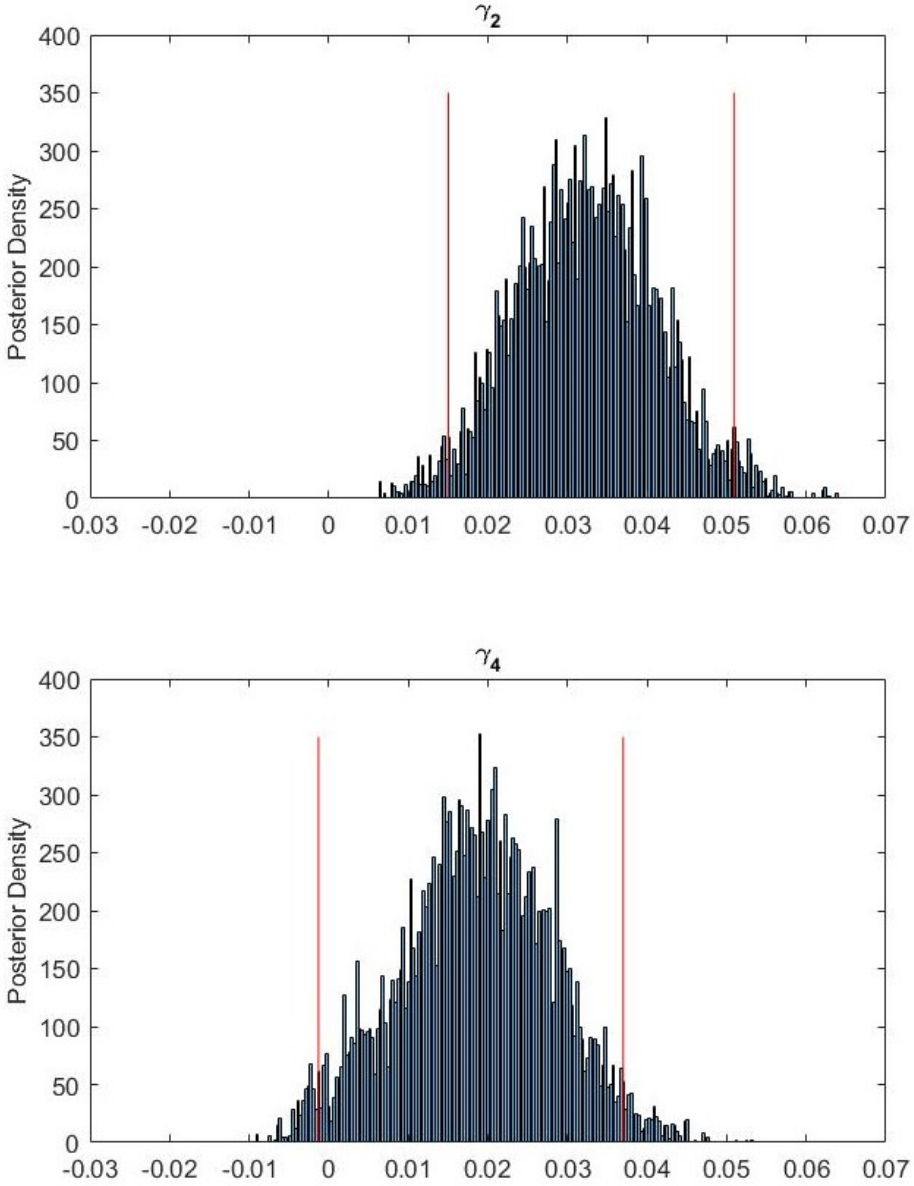
TABLE 4.4: Parameters of RE-SkN-Lasso model using simulated data

Parm	True Value	Maximum Likelihood		BLasso
		No Lasso	With Lasso (CV)	
μ	-0.11	-0.0179	-0.1284	-0.0955
ω	0.05	0.2673	0.0719	0.0678
β	0.98	0.9412	0.9803	0.9789
γ_1	0.47	0.6327	0.4664	0.5295
γ_2	**0**	**0.1200**	**0.0013**	**0.0350**
γ_3	0.25	0.3490	0.2338	0.3063
γ_4	**0**	**0.0618**	**0.0005**	**0.0188**
τ_1	-0.12	0.0370	-0.1413	-0.1243
τ_2	0.04	0.0801	0.0518	0.0519
ξ_1	-0.17	-0.0723	-0.3604	-0.3928
ξ_2	-0.75	-0.5566	-0.8608	-0.9656
ξ_3	-0.13	-0.0280	-0.2961	-0.3517
ξ_4	-0.8	-0.6503	-0.9326	-0.9770
φ_1	0.93	0.7965	0.9390	0.9284
φ_2	0.94	0.7910	0.9236	0.9352
φ_3	0.95	0.8095	0.9491	0.9482
φ_4	0.96	0.8153	0.9565	0.9488
$\delta_{1,1}$	-0.09	-0.0103	-0.0845	-0.0927
$\delta_{1,2}$	-0.11	-0.0627	-0.1107	-0.1145
$\delta_{1,3}$	-0.08	-0.0109	-0.0800	-0.0796
$\delta_{1,4}$	-0.13	-0.0527	-0.1492	-0.1445
$\delta_{2,1}$	0.06	0.0453	0.0750	0.0787
$\delta_{2,2}$	0.08	0.0958	0.0956	0.1015
$\delta_{2,3}$	0.1	0.0887	0.1263	0.1276
$\delta_{2,4}$	0.12	0.0978	0.1552	0.1567
σ^2_{u1}	0.15	0.4278	0.1611	0.1820
σ^2_{u2}	0.18	0.4345	0.1957	0.2143
σ^2_{u3}	0.12	0.4118	0.1289	0.1498
σ^2_{u4}	0.2	0.4780	0.2092	0.2262
ν	5	4.7630	4.5355	4.0344
λ	0.5	0.4288	0.4975	0.5376
λ_l			**300**	**1.4120**
$\tau_{l,1}$				0.7404
$\tau_{l,2}$				0.2956
$\tau_{l,3}$				0.5933
$\tau_{l,4}$				0.2469

The simulation study confirms that the Lasso approach could result in a sparser realized EGARH model, allowing the selection of realized measures of volatility to include in the

measurement equations. In Table 4.4, the RE-SkN-Lasso model estimated using the CV and BLasso approaches shrinks estimates of γ_2 and γ_4 parameters much closer to 0 (true value) compared to the RE-SkN model without penalty parameter, although the resulting coefficients are not exactly 0. The estimate of γ_4 is particularly very close to 0 for both methods.

FIGURE 4.10: Posterior density of γ_2 and γ_4



Note: Red lines are the lower and upper bounds of 95% credible interval.

The CV technique performs slightly better than the BLasso approach in shrinking the coefficient of γ_2 and γ_4 close to 0. The posterior realization of γ_2 and γ_4 based on the BLasso approach is presented in Figure 4.10. The 95% credible interval of γ_4 includes 0, however that of γ_2 does not. It is important to note that the tuning parameter chosen by the CV approach is much higher that that chosen by the BLasso approach. When λ_l is sufficiently large, the ℓ_1 penalty has a stronger shrinkage effect. However, the bias increases as λ_l increases due

to larger variance (James et al., 2013). As for the BLasso approach, the recommended tuning parameter values is relatively low, resulting in the estimates of γ_2 and γ_4 not shrinking very close to 0. With the lower value of tuning parameter, the BLasso method has less bias although with an increase in variance.

To evaluate the prediction error of both methods, 500 one-step-ahead forecasts of VaR and ES by using a fix 2000 in-sample with a rolling window approach are produced. Under the CV apporoach, the RE-SkN-Lasso model is estimated with a fixed $\lambda_l^{CV} = 300$ using the Maximum Likelihood method. Under the BLasso approach, λ_l^{Bayes} differs for every set of forecast estimation, based on the posterior means of the MCMC iterates.

The averages of 500 sets parameters of the RE-SkN-Lasso model estimated using both the CV and BLasso approaches over the forecast period are presented in Table 4.5. On average, the CV approach produces estimates of γ_2 and γ_4 closer to 0 than does the BLasso approach. However, the estimates of γ_2 and γ_4 under the BLasso approach on average improves, much closer than 0 than those estimated in Table 4.4. The tuning parameters based on the BLasso approach are consistently lower than that of the CV approach, that is on average 1.3698.

4.6.5 Tail risk forecasts accuracy

TABLE 4.5: The average of parameters of RE-SkN-Lasso model using simulated data over the forecast period

Parm	True Value	Maximum Likelihood		BLasso
		No Lasso	With Lasso (CV)	
μ	-0.11	0.0313	0.0068	-0.1459
ω	0.05	0.0641	0.0445	0.0495
β	0.98	0.9853	0.9863	0.9796
γ_1	0.47	0.4949	0.467	0.5049
γ_2	**0**	**0.0376**	**0.0003**	**0.0180**
γ_3	0.25	0.2056	0.2359	0.2746
γ_4	**0**	**0.0187**	**0.0003**	**-0.0076**
τ_1	-0.12	-0.1013	-0.1199	-0.102
τ_2	0.04	0.1387	0.0496	0.0361
ξ_1	-0.17	0.2443	-0.3492	-0.0626
ξ_2	-0.75	-0.3297	-0.9079	-0.6113
ξ_3	-0.13	0.2671	-0.3196	-0.0374
ξ_4	-0.8	-0.3442	-0.9468	-0.6338
φ_1	0.93	0.6975	0.9297	0.9213
φ_2	0.94	0.6984	0.9291	0.9176
φ_3	0.95	0.7154	0.9495	0.9422
φ_4	0.96	0.7171	0.9524	0.9324
$\delta_{1,1}$	-0.09	-0.0509	-0.077	-0.0745
$\delta_{1,2}$	-0.11	-0.0683	-0.1117	-0.0992
$\delta_{1,3}$	-0.08	-0.0431	-0.0739	-0.0698
$\delta_{1,4}$	-0.13	-0.0974	-0.1337	-0.1225
$\delta_{2,1}$	0.06	0.0325	0.0749	0.0548
$\delta_{2,2}$	0.08	0.064	0.0982	0.0708
$\delta_{2,3}$	0.1	0.0721	0.1228	0.089
$\delta_{2,4}$	0.12	0.0837	0.151	0.1093
σ^2_{u1}	0.15	0.4263	0.1587	0.1579
σ^2_{u2}	0.18	0.457	0.1925	0.1936
σ^2_{u3}	0.12	0.417	0.1204	0.1239
σ^2_{u4}	0.2	0.4912	0.1998	0.2015
ν	5	4.5597	4.1347	5.6156
λ	0.5	0.3973	0.5304	0.5017
λ_l			**300**	**1.3698**
$\tau_{l,1}$				0.906
$\tau_{l,2}$				0.2174
$\tau_{l,3}$				0.6515
$\tau_{l,4}$				0.2119

Figures 4.11 and 4.12 show the forecast sample returns of the simulated data and the associated forecast of the VaR and ES series at the forecast levels of 2.5% and 1%. The CV approach

generates relatively more extreme tail risk forecasts than the BLasso approach. This is particularly very clear visually at times where there is a persistence of extreme return.

FIGURE 4.11: Tail risk forecasts of simulated data; $\alpha = 2.5\%$

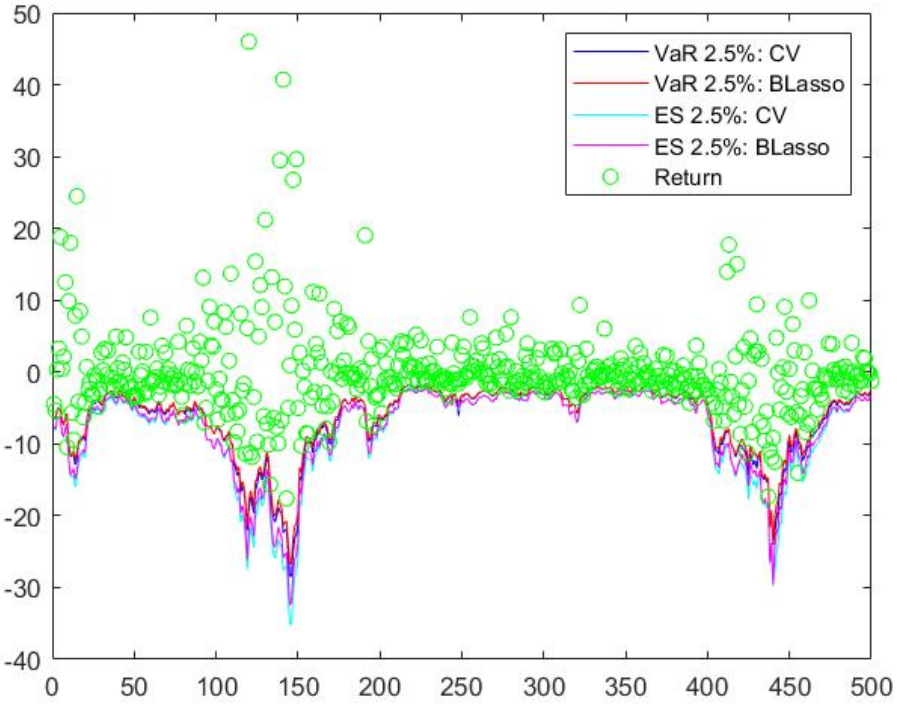
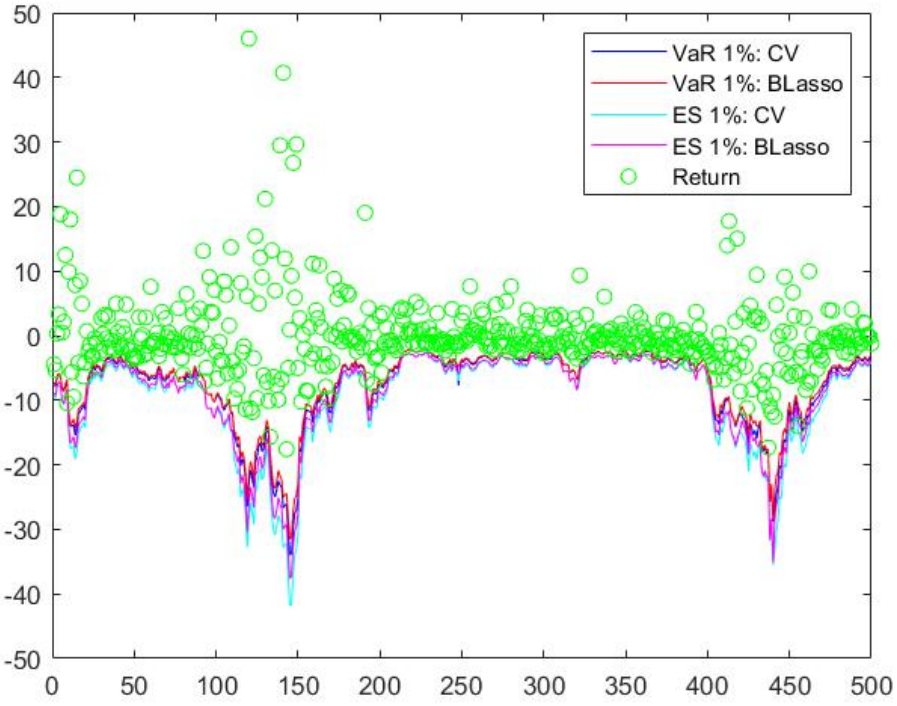


FIGURE 4.12: Tail risk forecasts of simulated data; $\alpha = 1\%$



The accuracy of the VaR and ES forecast during the forecast period is measured using the root mean square errors (RMSE). The RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\tilde{\theta}_{t,i} - \theta_{t,i})^2}, \tag{4.20}$$

where T is the number of forecasts, i is the type of tail forecast, $\tilde{\theta}$ is the estimated tail forecast, and θ is the true tail forecast.

The RSME of the tail risk forecasts of the RE-SkN-Lasso is presented in Table 4.6. For comparison purposes, the RMSE of the VaR and the ES resulting from the original RE-SkN model, both estimated using the ML method and the Bayesian method, are also presented. The RE-SkN-BLasso produces the most accurate VaR and ES forecasts (have the smallest RMSEs), outperforming all other methods at all forecast levels. Interestingly, the original RE-SkN model without the Lasso approach estimated using the proposed adaptive MCMC in this thesis has a better prediction accuracy than the RE-SkN model estimated using the CV method. It could then be argued that the proposed Bayesian approach in this thesis could improve tail risk forecast prediction. However, the CV approach has a better forecast performance than the original RE-SkN model (without the Lasso approach) estimated using the ML technique. In the following sections, a more thorough analysis on tail risk forecast performance is conducted for historical data in the following sections.

TABLE 4.6: Tail risk forecasts and RMSE of simulated data

RE-SkN Model	Tail Risk Forecasts	Mean	RMSE
True Value	$VaR_{2.5\%}$	-6.4921	
	$VaR_{1\%}$	-7.7013	
	$ES_{2.5\%}$	-8.0426	
	$ES_{1\%}$	-9.4288	
Original RE-SkN-ML	$VaR_{2.5\%}$	-6.6371	0.3472
	$VaR_{1\%}$	-7.9316	0.4570
	$ES_{2.5\%}$	-8.1905	0.4758
	$ES_{1\%}$	-9.7631	0.7024
RE-SkN-Lasso-CV	$VaR_{2.5\%}$	-6.6412	0.3154
	$VaR_{1\%}$	-7.9400	0.4304
	$ES_{2.5\%}$	-8.2265	0.4399
	$ES_{1\%}$	-9.8059	0.6772
Original RE-SkN-MCMC	$VaR_{2.5\%}$	-6.3078	0.2454
	$VaR_{1\%}$	-7.4138	0.3715
	$ES_{2.5\%}$	-7.7720	0.3798
	$ES_{1\%}$	-8.9915	0.5867
RE-SkN-BLasso	$VaR_{2.5\%}$	-6.3637	**0.2321**
	$VaR_{1\%}$	-7.4985	**0.3494**
	$ES_{2.5\%}$	-7.8452	**0.3644**
	$ES_{1\%}$	-9.1215	**0.5636**

4.7 Empirical Study

This section presents the empirical results of the proposed RE-SkN-Lasso models, both estimated using the proposed CV and BLasso approaches in forecasting VaR and ES. The selected tuning parameter over the forecast period and the shrinkage of γ parameters are discussed.

In addition to the three types of realized volatility measures employed in Chapter 3 (RVSS, RRSS, and RK), this chapter includes another realized measure, which is a range-based proxy (Parkinson, 1980) in Equation 4.21.

$$Range_t = (\log H_{t,i} - \log L_{t,i})^2. \tag{4.21}$$

The range of high and low prices has been proven to be a more superior volatility estimator than the squared return, providing a more efficient and less noisy estimator (Alizadeh et al., 2002; Parkinson, 1980). However, the range data are expected to be outperformed by the RVSS and RRSS.

The order of realized volatility measures in the measurement equations of the RE-SkN-Lasso models is the following: RK, RVSS, RRSS, and Range. In Chapter 3, the forecast performance of RK is found to be lower than that of RVSS and RRSS. Therefore, this historical study aims to investigate whether the RE-SkN-Lasso model can confirm that the variables RK and Range are less informative about future volatility than RVSS and RRSS and how far the coefficients of γ_1 (for RK) and γ_4 (for Range) are shrunk toward 0 by the RE-SkN-Lasso models.

The prediction accuracy of the RE-SkN-Lasso-CV and RE-SkN-BLasso models are compared. The forecast performance of several competing models is also presented for comparison purposes. The competing models include the proposed RE-SkN model in Chapter 3, both estimated using the ML approach (refered to as RE-SkN-ML in this chapter) and estimated using the Adaptive MCMC approach in Section 3.5.3 (RE-SkN-MCMC in this chapter). GARCH and EGARCH models are also included, this time using skewed Student-t distribution; these are called GARCH-Sk and EGARCH-Sk in this chapter, respectively. All models are estimated using the Matlab code developed for this thesis.

4.7.1 Data

This chapter uses the same historical dataset used in Chapter 3, which is sourced from the dataset used in Wang et al. (2019) and the Oxford-Man Institute's Realized Library, Library Version: 0.3. The dataset includes the daily return data for seven market indices: ASX 200 (Australia), DAX (Germany), FTSE 100 (UK), Hang Seng (Hong Kong), NASDAQ (US), SMI (Swiss), and S&P500 (US). The additional realized volatility measure used in this chapter, the range variable, is also sourced from the dataset used in Wang et al. (2019).

Suppose that the complete data is $t = T + N + M$, where T is the fixed in-sample observations. Under the CV approach, T is the training set, $N = 500$ is the validation sample, and $M = 500$ is the forecast sample. As for the BLasso approach, a rolling window approach with fixed in-sample observations T, starting from $t - N + 1$, is used as the estimation sample. The remaining observations $N = 500$ are used as the forecast sample. This is to ensure that both methods use a consistent dataset in producing parameter estimates. The rolling window approach is employed to produce 500 one-step-ahead forecasts.

The forecast period includes the 2015-2016 stock market selloff, when the world stock market declined substantially, particularly on 24 August 2015 (the August 24 flash card or the 2015 stock market crash). At this time a large sell-off in Asia, which was partly due to fears about the economic slowdown in China, triggered a significant drop in US and European markets.

4.7.2 Data estimation procedures

Given a very high computing load, the estimation of the RE-SkN-Lasso model using the CV and BLasso approaches is conducted by using the University of Sydney's Artemis High-Performance Computing (HPC). The estimation is divided into separate tasks to run on different computers. For the CV method, the data estimation for each market index is divided into 117 tasks, each running a RE-SkN-Lasso model with a different value of the tuning parameter. The optimal tuning parameter is then used to estimate the model to produce 500 one-step-ahead tail risk forecasts. For the BLasso method, the data estimation for each market index is broken down into 50 tasks; each task re-estimates the model 10 times to produce 10 sets of parameters to forecast 10 one-step-ahead tail risks, thus totaling 500 one-step-ahead forecasts.

The estimation time of the RE-SkN-Lasso model based on the CV method for each value of the tuning parameter is presented in Table 4.7. The estimation time for each task ranges from roughly 2-6 hours, depending on the length of the in-sample period for each market index. The column 'Total' is the sum of the estimation time over 117 tasks for each market index. The estimation time for SMI is the shortest while that for DAX is the longest.

The estimation time of the RE-SkN-Lasso model based on the BLasso approach for each task is presented in Table 4.8. The column 'Total' is the sum of the estimation time over 50 tasks for each market index. The estimation time for each task ranges from 2.5-6 hours, with a fairly consistent duration between tasks, that is, roughly around 4.5 hours. The estimation time for the SMI is also the shortest, while that for the Hang Seng is the longest.

Overall, the estimation of RE-SkN-Lasso CV models takes longer to complete, almost twice the length of time as the RE-SkN-BLasso models. This is due to the grid search process adopted to select the value of the tuning parameter that is used in the forecasting of tail risk under the CV approach, λ_l^{CV}.

TABLE 4.7: Estimation time: RE-SkN-Lasso-CV

Market Index	Min	Max	Mean	Total
ASX 200	2:28:04	4:39:37	3:18:10	386:24:49
DAX	3:26:04	6:02:36	4:24:22	515:30:40
FTSE 100	2:54:17	5:23:25	4:01:05	470:05:56
Hang Seng	2:10:23	3:47:46	3:01:46	351:25:18
NASDAQ	2:48:07	4:40:47	3:37:55	424:56:42
SMI	1:51:38	3:05:32	2:25:04	282:53:09
S&P 500	2:24:18	4:42:45	3:29:13	407:57:56

TABLE 4.8: Estimation time: RE-SkN-BLasso

Market Index	**Min**	**Max**	**Mean**	**Total**
ASX 200	3:22:54	5:12:46	4:17:58	214:58:26
DAX	3:22:10	5:21:55	4:15:40	213:03:20
FTSE 100	3:22:08	5:35:00	4:16:50	214:01:39
Hang Seng	4:05:39	6:06:07	4:39:08	227:57:14
NASDAQ	3:56:48	5:48:49	4:42:14	235:11:54
SMI	2:37:44	3:49:01	3:13:34	161:18:18
S&P 500	3:29:41	5:14:38	4:22:14	218:31:24

4.7.3 Predictive log-likelihood

The plot of the predictive log-likelihood $\sum \ell(r;\theta)$ over the validation period $M = 500$ resulting from the RE-SkN-Lasso model is presented for each market index. Figure 4.13 presents $\sum \ell(r;\theta)$ for the ASX 200, Figure 4.14 for the DAX, Figure 4.15 for the FTSE 100, Figure 4.16 for the Hang Seng, Figure 4.17 forthe NASDAQ, Figure 4.18 for the SMI, and Figure 4.19 for the S&P 500. For comparison purposes, the $\sum \ell(r;\theta)$ of several competing models, such as GARCH-Sk, EGARCH-Sk, and the original RE-SkN model estimated using the ML method, are also included in the plots.

Visual analysis of predictive log-likelihood across market indices and across models suggests that the RE-SkN-Lasso CV model with the optimum λ_l^{CV} consistently produces the highest predictive log-likelihood compared to other competing models across market indices. A closer inspection also suggests several findings. First, it is very obvious that the RE-SkN type models, both with Lasso approach and without Lasso approach, produce much higher predictive log-likelihood than GARCH-Sk and EGARCH-Sk models, which is as expected. Second, the use of the Lasso approach in the RE-SkN model has improved the predictive log-likelihood in most cases (such as the case of the ASX 200, FTSE 100, Hang Seng, and SMI. However, the Lasso method in the case of the DAX does not consistently produce the highest predictive log-likelihood, although the optimum tuning parameter does generate a higher predictive log-likelihood than the original RE-SkN model. For the S&P 500, the Lasso approach could only produce a marginal improvement in the predictive log-likelihood. Third, a higher predictive log-likelihood is produced by the RE-SkN-Lasso-CV model given that the tuning parameter value is approximately less than 500. The use of a tuning parameter higher than 500 results in a decrease in the predictive log-likelihood.

FIGURE 4.13: Predictive log likelihood: ASX 200

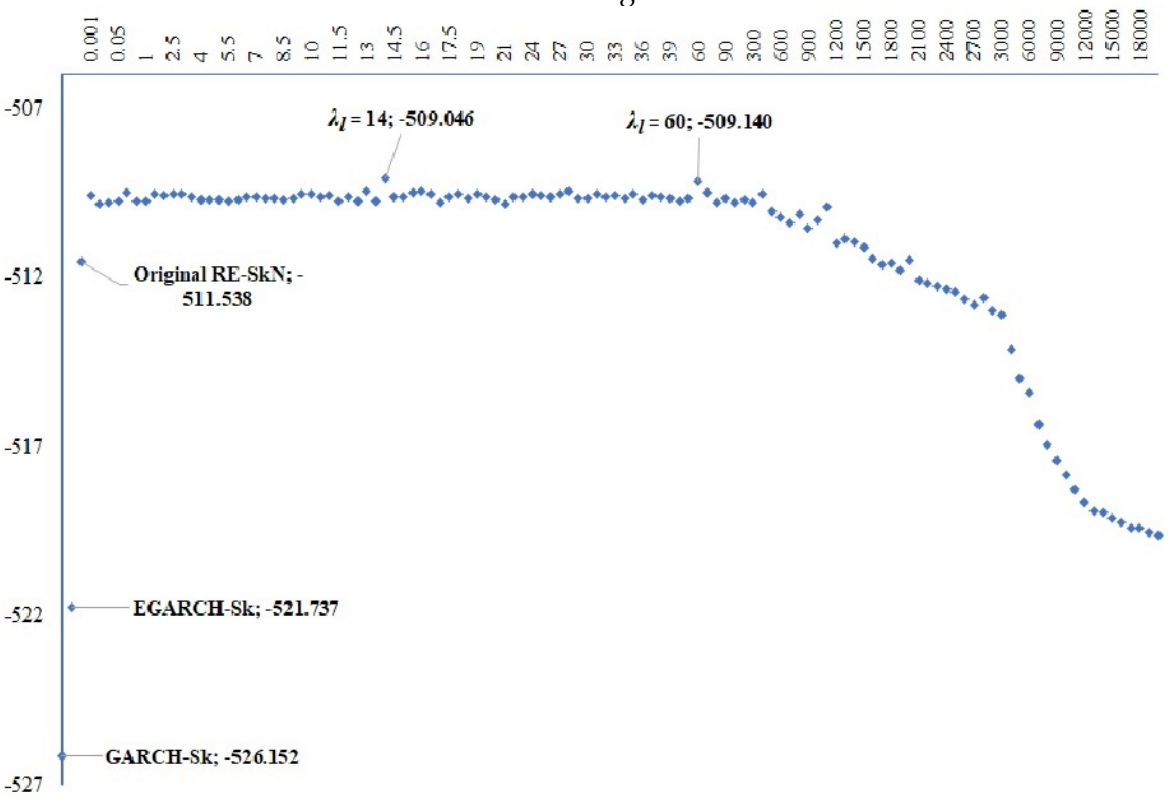


FIGURE 4.14: Predictive log likelihood: DAX

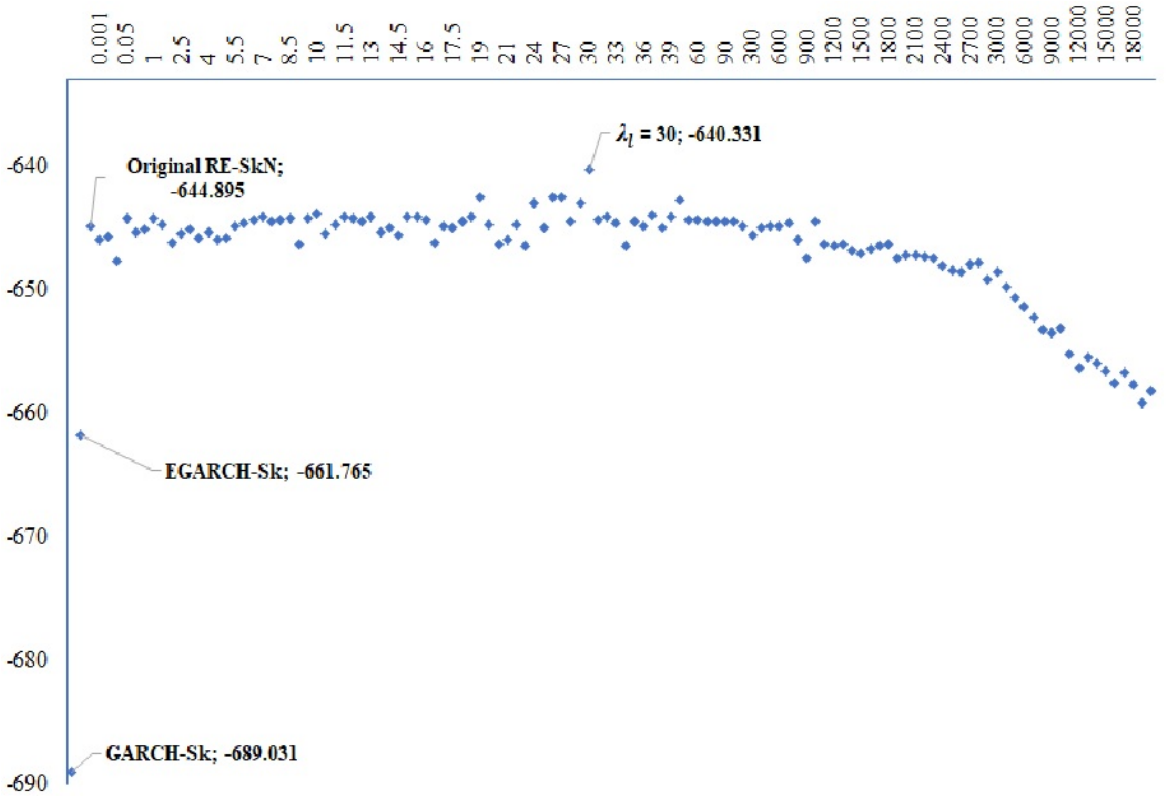


FIGURE 4.15: Predictive log likelihood: FTSE 100

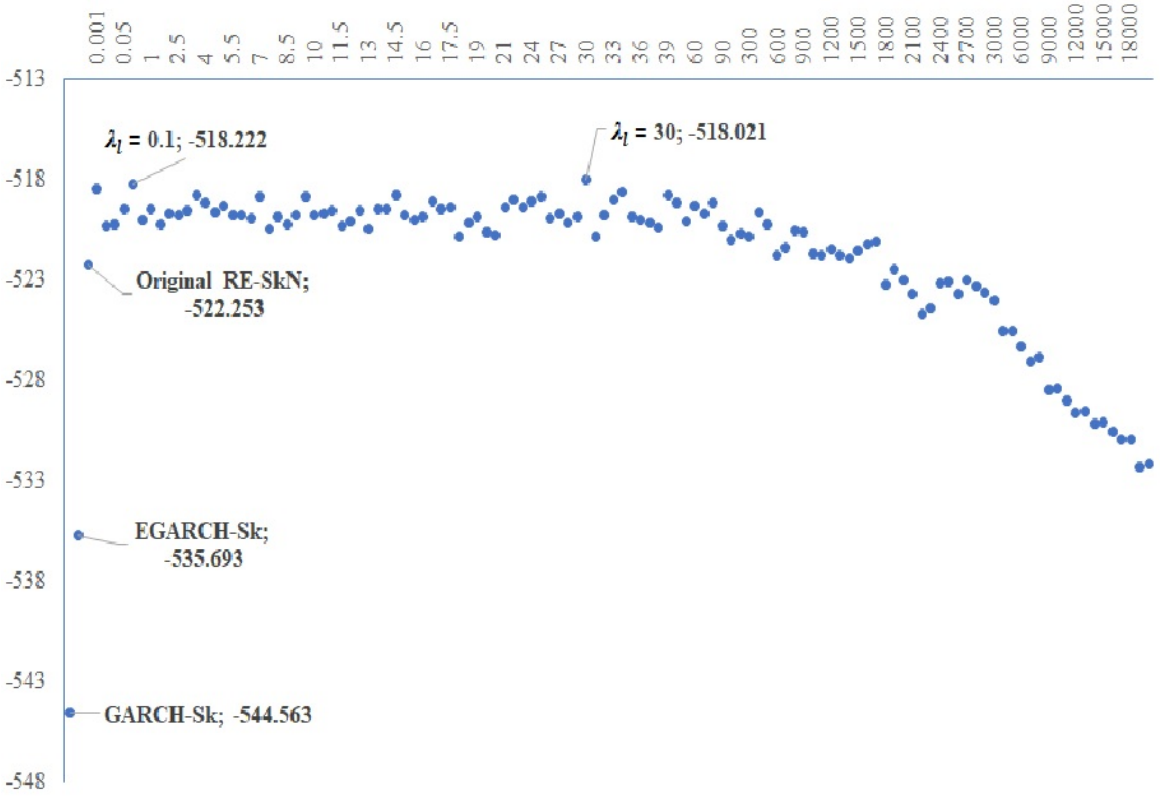


FIGURE 4.16: Predictive log likelihood: Hang Seng

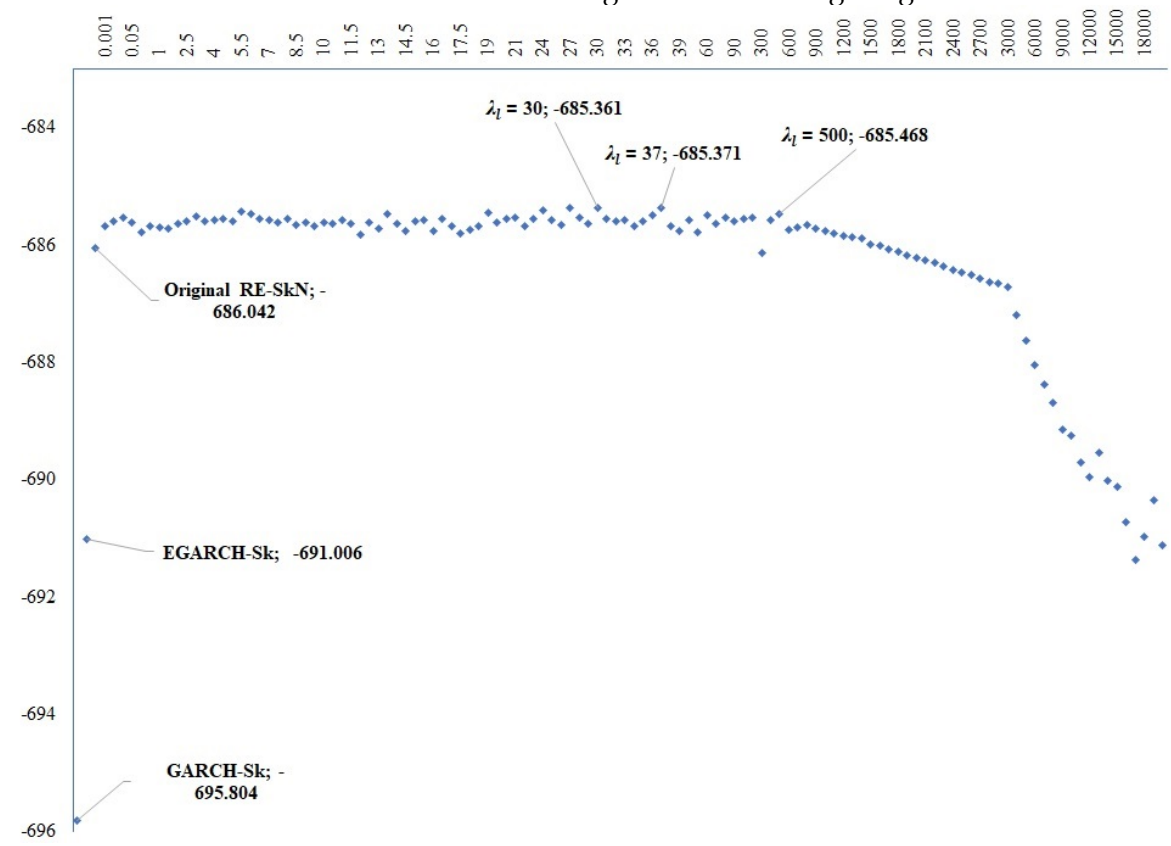


FIGURE 4.17: Predictive log likelihood: NASDAQ

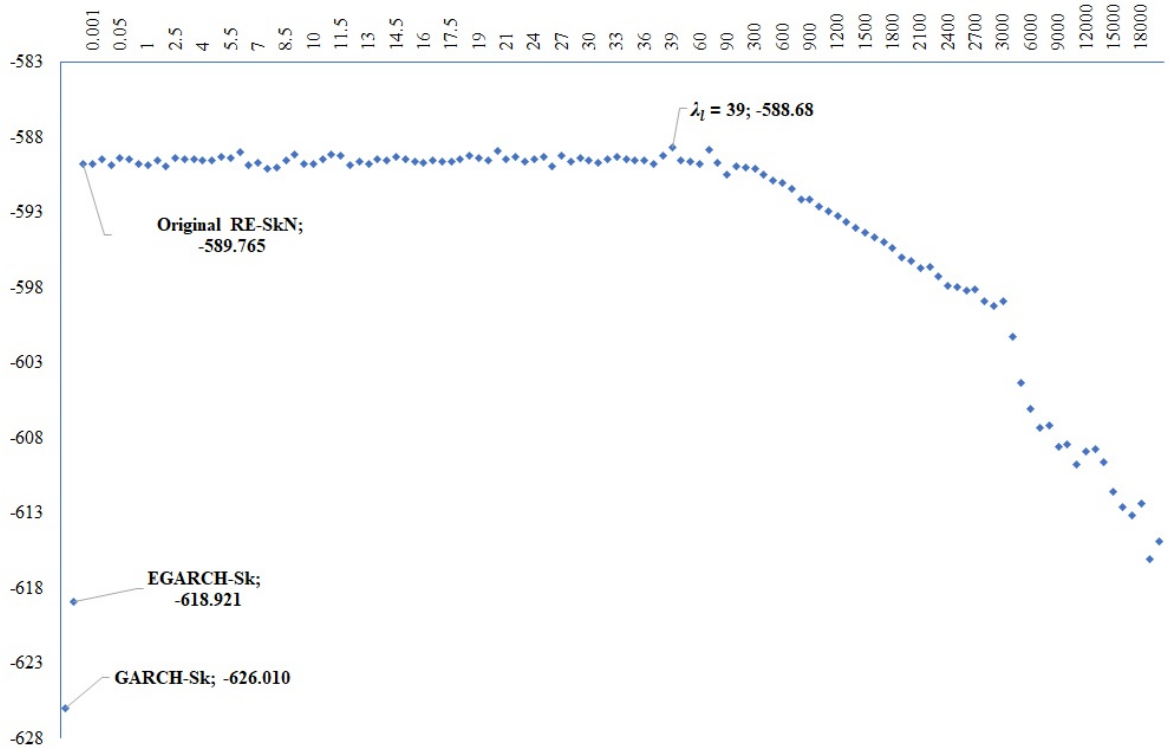


FIGURE 4.18: Predictive log likelihood: SMI

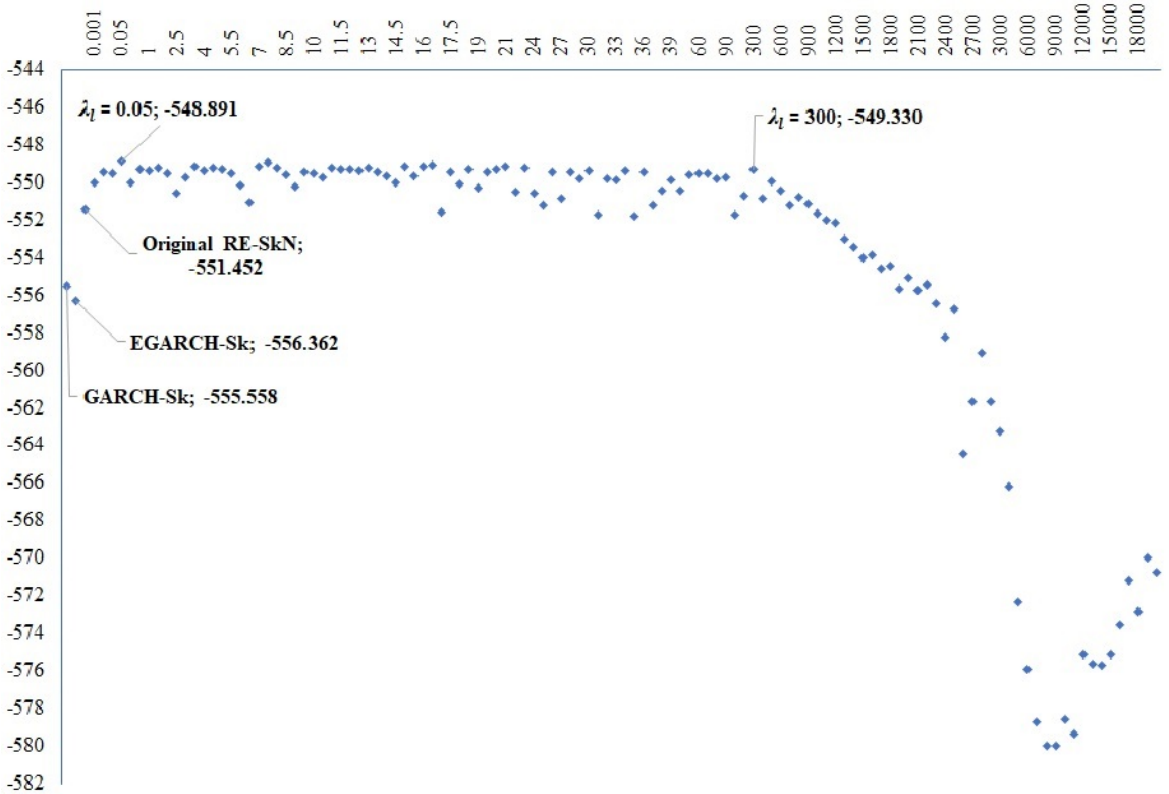
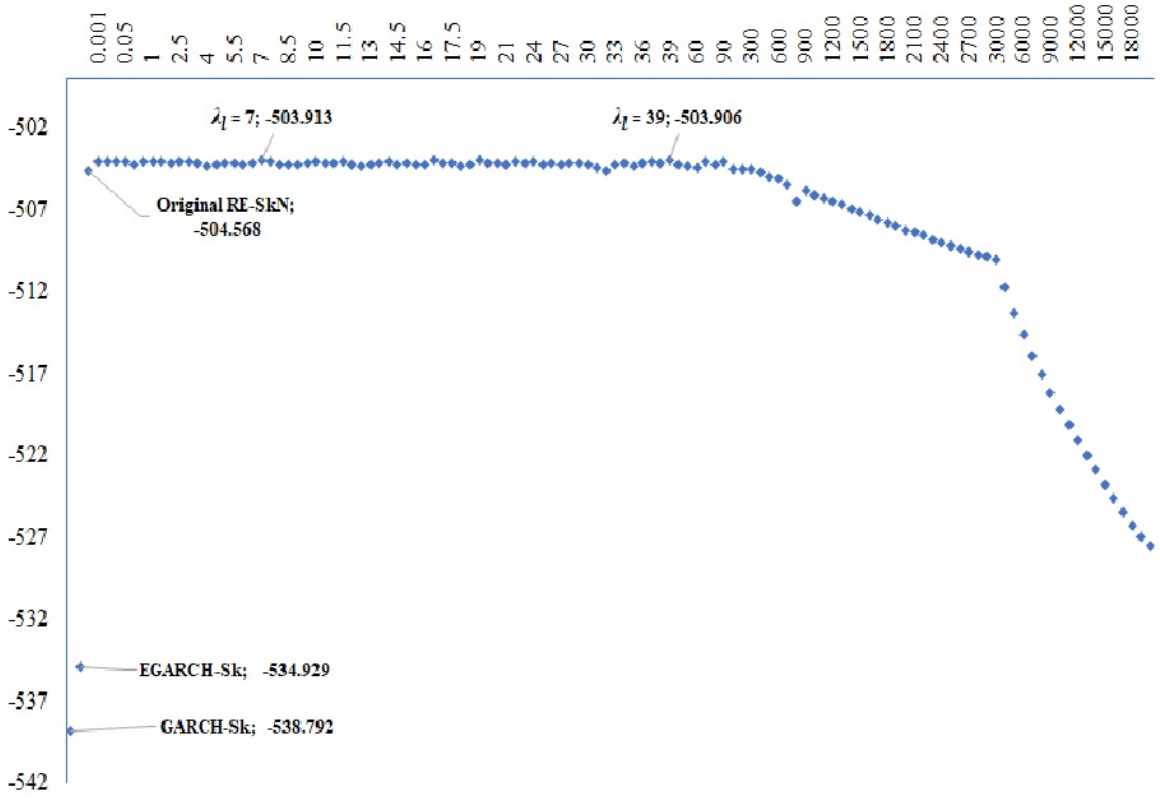


FIGURE 4.19: The predictive log likelihood for S&P500 return series



4.7.4 Model parameter estimates

The selected tuning parameter value based on the CV approach λ_l^{CV} for each market index is the one that gives the highest predictive log-likelihood $\ell(r|\theta)$ in the validation period. The selected tuning parameter is fixed and used in the RE-SkN-Lasso-CV to forecast out-of-sample tail risks. The λ_l^{CV} for each market return index is presented in Table 4.9. The tuning parameter values are diverse and fairly moderate, $\lambda_l^{CV} = 30$ for three market indices and $\lambda_l^{CV} = 39$ for two market indices. The selected tuning parameter in the RE-SkN-Lasso model for SMI is the smallest, $\lambda_l^{CV} = 0.05$.

Meanwhile, the value of the tuning parameter used in the RE-SkN-BLasso model varies across 500 sets of parameters resulting from the posterior predictive means of the MCMC iterations. The average values of the tuning parameters λ_l^{Bayes} for each market index are also presented in Table 4.9. The average of λ_l^{Bayes} is highly consistent across all market indices, ranging from 1.362 (ASX 200) to 1.428 (Hang Seng). The tuning parameter values of the RE-SkN-BLasso models are comparatively much lower than the selected tuning parameters in the RE-SkN-Lasso-CV models, except for the RE-SkN-Lasso-CV model for SMI.

TABLE 4.9: The tuning parameter value over the forecast period

	λ_l^{CV} (fixed)	**Average of** λ_l^{Bayes}
ASX 200	14	1.362
DAX	30	1.420
FTSE 100	30	1.412
Hang Seng	30	1.428
NASDAQ	39	1.407
SMI	0.05	1.418
S&P 500	39	1.419

One of the main interests in this empirical study is to investigate whether the proposed Lasso approach in this study could shrink some of the γ coefficients (close) to 0, to give a justification for eliminating the corresponding realized volatility measures from the RE-SkN model. The averages of the γ parameters of the RE-SkN-Lasso-CV model and the posterior means of MCMC iterations of γ parameters over the forecast period are presented in Table 4.10. Information about the future volatility contained by the RK variable is represented by γ_1, RVSS by γ_2, RRSS by γ_3, and Range by γ_4. The variables RK and Range are of particular interest; therefore, the coefficients of γ_1 and γ_4 are printed in bold to make them easy to see. The distributions of posterior means of parameters γ_1 and γ_4 over the forecast period for each market index can be found in Appendix B.

Both models confirm that the RK and Range variables are less informative about the future volatility of asset returns compared to the RVSS and RRSS. This can be seen in the estimated coefficients of γ_1, which are lower compared to the coefficients of γ_2 and γ_3 in all seven market indices. The proposed Lasso approach shrinks the coefficients of γ_1 to close to 0 for all market indices. The coefficients of γ_1 are especially very small, very close to 0, for the case of DAX, SMI, and S&P 500.

As for the Range variable, the estimated coefficients of γ_4 are lower compared to the coefficients of γ_2 and γ_3 in four of the seven market indices. Nevertheless, the coefficients of γ_4 are slightly higher than γ_1 in most indices. It is concluded that although the Range variable contains weak information to predict future volatility compared to the RVSS and RVSS, it can provide a better signal about future volatility than the RK. A special case is for DAX, where the Range variable is more informative than the other robust measures, RVSS and RRSS, as indicated by a higher coefficient of γ_4 for DAX than for γ_2 and γ_3.

This finding is consistent for both the RE-SkN-Lasso-CV model and the RE-SkN-BLasso model, suggesting that future estimation and extension of the realized EGARCH model might consider not including RK and Range in the measurement equations to aim for a sparse yet informative forecasting model.

An impressive finding is that the coefficient γ_3 is consistently higher than other realized volatility measures. In Chapter 3, the RE-SkN model including this variable is shown to have a favorable forecasting performance. The Lasso approach applied to the RE-SkN model

in most cases does not shrink γ_3 close to 0. This finding suggests that the RRSS is most informative in predicting the future volatility of stock market return compared to other realized volatility measures used in this study.

TABLE 4.10: The average of γ parameters over the forecast period

Index	γ_1	γ_2	γ_3	γ_4
Average of parameters of RE-SkN-Lasso-CV				
ASX 200	**0.0294**	-0.0953	0.2359	**0.1346**
DAX	**-0.0036**	0.0641	0.0766	**0.1405**
FTSE 100	**-0.0183**	0.1364	0.0805	**0.0654**
Hang Seng	**0.0273**	0.0411	0.1344	**0.0398**
NASDAQ	**0.0500**	0.1893	0.0819	**0.0505**
SMI	**0.0090**	0.0217	0.2694	**0.0791**
S&P	**0.0018**	0.2006	0.0791	**0.0484**
Average of posterior means of RE-SkN-BLasso				
ASX 200	**0.0294**	-0.0953	0.2359	**0.1346**
DAX	**-0.0036**	0.0641	0.0766	**0.1405**
FTSE 100	**-0.0183**	0.1364	0.0805	**0.0654**
Hang Seng	**0.0273**	0.0411	0.1344	**0.0398**
NASDAQ	**0.0500**	0.1893	0.0819	**0.0505**
SMI	**0.0090**	0.0217	0.2694	**0.0791**
S&P	**0.0018**	0.2006	0.0791	**0.0484**

4.7.5 Tail risk forecasts accuracy

This section discusses the tail risk prediction accuracy of the RE-SkN-Lasso and RE-SkN-BLasso models for seven market indices compared to several competing models. The estimation of the models in this chapter is computationally much more challenging than in other chapters. Therefore, in this chapter, only selected tail risk accuracy tests are performed. First, the VaR backtest is conducted based on the Unconditional Coverage (UC) test (Kupiec, 1995) outlined in Section 2.3.1, the Conditional Coverage (CC) test outlined in Section 2.3.2, and the Dynamic Quantile (DQ) test (Dumitrescu et al., 2012; Engle & Manganelli, 2004) using four lags outlined in Section 2.3.3. The quantile loss function based on Section 2.3.6 Equation 2.37 and the VaR and ES Joint Loss function based on the asymmetric Laplace (AL) log score in Section 2.3.7 Equation 2.40 are also estimated and analysed.

Figure 4.20 visualizes 1% ES forecasts for the S&P 500 based on the RE-SkN-Lasso-CV and RE-SkN-BLasso models compared to the original RE-SkN models without penalty parameter, that is the RE-SkN-ML and RE-SkN-MCMC models. The ES forecasts resulting from four models are very close. In the case of the S&P500 index, the RE-SkN-ML model seems to produce slightly more extreme ES forecasts than other RE-SkN type models. As for the RE-SkN-Lasso type models, the ES forecasts resulting from the RE-SkN-BLasso approach are

slightly more extreme than those resulting from the RE-SkN-Lasso-CV model. However, the difference is not visible.

The RE-SkN-Lasso-CV and RE-SkN-BLasso models are also compared with the traditional GARCH and EGARCH models that use the skewed Student-*t* distribution for the return distribution, which are called the GARCH-Sk and EGARCH-Sk models. The 1% ES forecasts for these models are presented in Figure 4.21 to facilitate visual analysis of the ES forecasts. The figure shows that the RE-SkN-Lasso-CV and RE-SkN-BLasso models rapidly recover to follow the tail of the return series, whereas the GARCH-Sk and EGARCH-Sk models adjust slowly due to high persistence. The GARCH-Sk and EGARCH-Sk models also produce less extreme tail risks than the RE-SkN-Lasso models.

The most extreme tail risk forecasts in both figures are produced around the period of the 2015 stock market crisis. On 25 August 2015, one day after the August 24 flash card when the S&P 500 return fell by 4.02%, the 1 % ES forecast produced by the RE-SkN-BLasso model is -9.0078%, the RE-SkN-Lasso-CV model -8.8520%, RE-SkN-MCMC model -9.0650%, RE-SkN-ML model -9.2992%, GARCH-SkN model -4.5161%, and EGARCH-SkN model -5.3160%. The RE-SkN-ML model produces the most extreme 1% ES forecast, followed by the RE-SkN-BLasso model, then RE-SkN-MCMC model.

FIGURE 4.20: 1% Expected shortfall forecasts for S&P500: RE-SkN type models

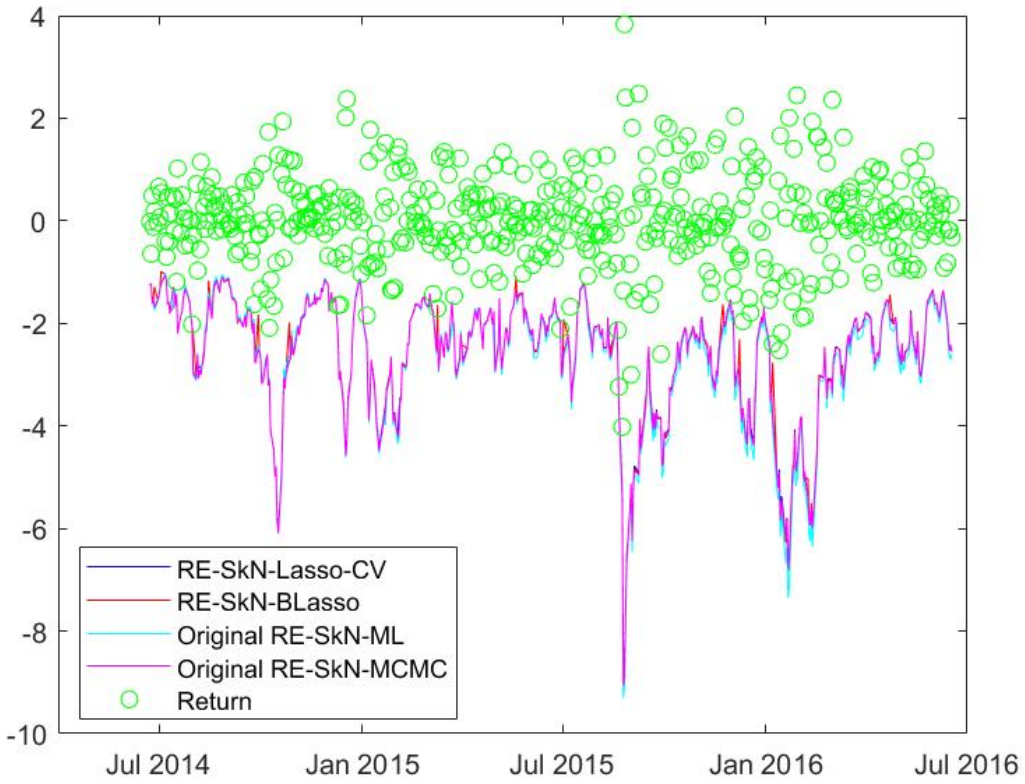
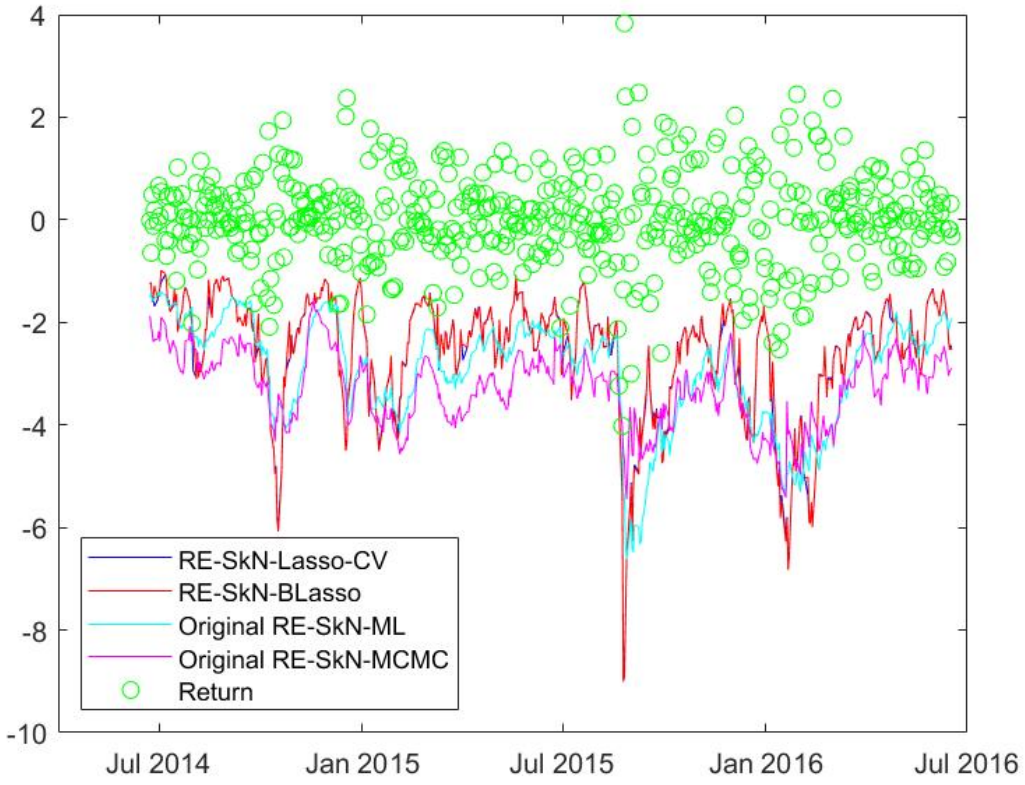


FIGURE 4.21: 1% Expected shortfall forecasts for S&P500: RE-SkN-Lasso, GARCH-SkN, and EGARCH-SkN models



The result of three types of VaR backtests (UC, CC, and DQ tests) conducted at the 5% significance level is presented in Table 4.11. All models, except the GARCH-SkN model, pass all the VaR backtest tests at the 2.5% forecast level. At the 1% forecast level, only the RE-SkN-Lasso CV, RE-SkN-MCMC, and RE-SkN-BLasso models pass the three VaR backtest tests for all market indices.

TABLE 4.11: VaR Backtest: UC, CC, and DQ tests

Model	**UC**	**CC**	**DQ4**
$\alpha = 2.5\%$			
GARCH-Sk	4	5	5
EGARCH-Sk	0	0	0
RE-SkN-ML	0	0	0
RE-SkN-Lasso-CV	0	0	0
RE-SkN-MCMC	0	0	0
RE-SkN-Blasso	0	0	0
$\alpha = 1\%$			
GARCH-Sk	1	3	3
EGARCH-Sk	1	3	3
RE-SkN-ML	1	0	0
RE-SkN-Lasso-CV	0	0	0
RE-SkN-MCMC	0	0	0
RE-SkN-Blasso	0	0	0

To further investigate the difference in the prediction accuracy of the RE-SkN-Lasso-CV and the RE-SkN-BLasso models, the quantile losses and the AL log scores based on the 500 one-step ahead forecasts of the tail risks of the seven market indices are calculated and compared. Smaller values indicate superior accuracy. Tables 4.12 and 4.13 show the quantile loss function values and the AL log score values of various models in the seven market indices, respectively. For individual stock index, bold indicates the most favored model, yieding the smallest quantile loss function values. Meanwhile, bold italic indicates the second most favored model.

TABLE 4.12: Quantile loss function values

Model	**ASX 200**	**DAX**	**FTSE 100**	**Hang Seng**	**NASDAQ**	**SMI**	**S&P 500**
$\alpha = 2.5\%$							
GARCH-Sk	30.9300	46.5099	36.2997	45.5327	35.3512	42.7327	30.3542
EGARCH-Sk	31.5470	46.2680	37.0056	42.1774	35.2145	41.7843	30.7087
RE-SkN-ML	30.4986	45.0899	33.9853	40.7662	33.6400	39.2348	27.0946
RE-SkN-Lasso-CV	***30.2873***	44.3423	33.7808	40.2724	***32.4917***	39.1057	***26.9753***
RE-SkN-MCMC	30.3533	***44.3174***	***33.7680***	***40.2771***	32.5896	***39.0585***	27.0908
RE-SkN-Blasso	**30.2275**	**44.2733**	**33.7494**	**40.2610**	**32.4828**	**38.9262**	**26.8262**
$\alpha = 1\%$							
GARCH-Sk	14.3674	21.7946	17.3091	22.9700	16.1093	23.3753	14.7492
EGARCH-Sk	14.1508	21.5017	17.6354	20.2403	16.1387	22.5368	15.0860
RE-SkN-ML	14.3342	21.4328	15.9113	19.0011	15.6540	21.3844	12.0711
RE-SkN-Lasso-CV	***13.8910***	19.3425	15.7629	**18.3414**	15.1741	***21.2936***	11.9226
RE-SkN-MCMC	13.9315	***19.3409***	***15.6983***	18.3632	**15.1330**	21.3229	***11.9171***
RE-SkN-Blasso	**13.8660**	**19.2923**	**15.6904**	***18.3454***	***15.1334***	**21.2534**	**11.8927**

Note: For individual stock index, bold indicates the 1st ranked model, bold italics indicates the 2nd ranked model.

TABLE 4.13: AL log score values

Model	**ASX 200**	**DAX**	**FTSE 100**	**Hang Seng**	**NASDAQ**	**SMI**	**S&P 500**
$\alpha = 2.5\%$							
GARCH-Sk	1.8980	2.3656	2.1134	2.3678	2.0370	2.1956	1.8920
EGARCH-Sk	1.9225	2.3358	2.1220	2.2076	2.0403	2.1896	1.9144
RE-SkN-ML	1.8611	2.2808	1.9980	2.1704	2.0308	2.1431	1.7279
RE-SkN-Lasso-CV	***1.8545***	2.2781	1.9920	***2.1468***	***1.9662***	***2.1103***	***1.7212***
RE-SkN-MCMC	1.8569	***2.2768***	***1.9907***	2.1480	1.9680	2.1162	1.7285
RE-SkN-Blasso	**1.8525**	**2.2748**	**1.9902**	**2.1465**	**1.9628**	**2.1033**	**1.7124**
$\alpha = 1\%$							
GARCH-Sk	2.0258	2.4959	2.2546	2.5685	2.1471	2.4717	2.0725
EGARCH-Sk	2.0113	2.4629	2.2656	2.3482	2.1638	2.4295	2.0985
RE-SkN-ML	2.0102	2.4357	2.1437	2.3137	2.1842	2.4596	1.8143
RE-SkN-Lasso-CV	***1.9702***	***2.3453***	2.1391	***2.2717***	2.1164	***2.4067***	1.8021
RE-SkN-MCMC	1.9719	2.3444	***2.1334***	2.2762	***2.1065***	2.4193	***1.8005***
RE-SkN-Blasso	**1.9672**	**2.3411**	**2.1324**	**2.2725**	**2.1033**	**2.4017**	**1.7977**

Note: For individual stock index, bold indicates the 1st ranked model, bold italics indicates the 2nd ranked model.

It is evident that the RE-SkN-BLasso model has the best tail risk prediction accuracy compared to other competing models, followed by the RE-SkN-MCMC model, then the RE-SkN-Lasso-CV model. The difference in the quantile loss function values and the AL log score values is quite marginal between those three models.

At the 2.5% forecast level, the RE-SkN-BLasso model produces the lowest quantile loss value than all other models for all market indices. The RE-SkN-MCMC model is the second-most favored model for four market indices, whereas the RE-SkN-Lasso-CV model is the second-most-favored model for three market indices. At the 1% forecast level, the RE-SkN-BLasso model generates the lowest quantile loss values for five market indices, while the RE-SkN-Lasso-CV and RE-SkN-MCMC models are the most favored model for only one market index, respectively. The RE-SkN-MCMC model is the second-best favored model for three market indices, while the RE-SkN-Lasso-CV model is the second-best favored model for two market indices.

With regard to the joint VaR and ES loss function, the RE-SkN-BLasso model has the best prediction accuracy, followed by the RE-SkN-Lasso-CV model. Therefore the Lasso approach can help improve the tail risk forecasting performance of the RE-SkN model. To be more specific, the RE-SkN-BLasso model generates the smallest AL log score values compared to other competing models for all market indices at both the 2.5% and 1% forecast levels. At the 2.5% forecast level, the RE-SkN-Lasso-CV model produces the second lowest AL log score values for five market indices, whereas the RE-SkN-MCMC model produces the second lowest AL log score values only for two market indices. At the 1% forecast level, the RE-SkN-Lasso-CV model generates the second lowest AL log score values for four market

indices, while the RE-SkN-MCMC model produces the second lowest AL log score values only for three market indices.

TABLE 4.14: 90% and 75% Model confidence set with R and SQ methods

Model	alpha=2.5%				alpha=1%			
	90% MCS		75% MCS		90% MCS		75% MCS	
	R Method	SQ Method	R Method	SQ Method	R Method	SQ Method	R Method	SQ Method
GARCH-Sk	6	6	6	4	6	6	4	5
EGARCH-Sk	6	6	5	4	6	6	5	6
RE-SkN-ML	7	7	7	6	7	7	4	6
RE-SkN-Lasso-CV	7	7	7	6	7	7	6	7
RE-SkN-MCMC	7	7	7	6	7	7	6	7
RE-SkN-Blasso	7	7	7	7	7	7	7	7

Note: For each MCS method, a green highlight indicates favored models for all market indices.

The AL log score function for the 2.5% and 1% forecast levels are incorporated as the loss function in the calculation of model confidence set (MCS) (Hansen et al., 2011). Tests are carried out using both R and SQ methods at 90% and 75% confidence levels. The results of the tests are presented in Table 4.14. The models that consistently enter the MCS for both MCS methods and across market indices are highlighted in green.

The RE-Skn-BLasso are consistently included in the set of superior models. The use of the standardized skewed Student t distribution evidently improves GARCH and EGARCH models. These two models are rarely excluded in the MCS. The SQ method shows the difference in performance of the models in Table 4.14, especially for the test at 75% confidence level. As expected, RE-SkN models have a superior performance compared to the GARCH-Sk and EGARCH-Sk models. The superiority of the RE-Skn-BLasso model is especially highlighted by much higher MCS p-values of the model at both confidence levels using both R and SQ methods compared to other competing models. For details regarding the MCS p-values for each model in Table 4.14, please see Appendix B.2.

4.8 Chapter Summary

This chapter develops a methodology for selecting realized volatility measures in the realized EGARCH model by using the Lasso approach. A frequentist approach is developed by adopting a CV method. A Bayesian Lasso framework is also proposed. This study concludes that the proposed Lasso approaches could help in selecting informative realized volatility measures in the realized EGARCH model, thus resulting in a sparse and parsimonious model. However, the tuning parameter values chosen by those two methods are quite different based on both the simulated data and the empirical data. The Bayesian Lasso approach is consistent in producing small values of tuning parameters across market indices.

The application of both proposed methods to historical data shows that the RRSS is the most informative volatility measures compared to the RK, RVSS, and Range. In several market indices, the methods suggest that the RK and range variables can be eliminated from the

realized EGARCH model. While the VaR backtests are passed by both methods, the quantile loss function and the AL log score values confirm that the Bayesian Lasso approach has a better forecasting performance than the CV approach.

The grid search to select the tuning parameter value under the CV approach takes a substantial amount of computation time, since the proposed set of tuning parameter values covers a very wide range of values. This high computation burden could be reduced by focusing on a narrower range of values, as found in both the simulation and the empirical studies.

In this study's approach, each γ coefficient is equally penalized in the l_1 penalty. Further research avenues could explore the adaptive Lasso approach of Zou (2006) that assigns different weight to different γ coefficients as follows:

$$\mathcal{L}(\theta) + \lambda_l \sum_{j=1}^{p} \hat{w}_j |\gamma_j|, \tag{4.22}$$

where w is a known weights vector.

Zou (2006) shows that data-dependent weights that are cleverly chosen will result in a weighted Lasso that has oracle properties. The adaptive Lasso does not suffer from the multiple local optima issues and the global minimizer can be solved efficiently since it is a convex optimization problem.

For those working in Bayesian modeling, further research could also explore the Bayesian adaptive Lasso (BaLasso) method (Leng et al., 2014). This method assigns different penalty parameters for the regression coefficients, allowing different λ_j, one for each coefficient. Another version of the BaLasso method is proposed by Mallick and Yi (2014) who, instead of using the scale mixture of normal, used a different Gibbs sampler by utilizing the scale mixture of uniform (SMU) representation of the Laplace density.

Although this study suggests that the regularization method can help in finding useful realized measures to use in tail risk forecasting using the realized EGARCH model, there are only four types of realized volatility measures included in the empirical study and they are all robust measures, especially RRSS and RVSS. Future works may consider including and testing a higher number of realized volatility measures and including less robust measures such as simple realized variance, bi-power variation (Barndorff-Nielsen & Shephard, 2004), and median realized variance (Andersen et al., 2012).

Chapter 5

Bayesian Realized EGARCH Model with the Two-sided Weibull Distribution

5.1 Introduction

There has been a growing interest in the use of the Weibull distribution (Weibull, 1951) in finance. The Weibull distribution is a special case of an extreme value distribution and of the generalized gamma distribution and has been extensively used as a model of time-to-failure (Chen et al., 2008). It is even a super-exponential distribution (Rolski et al., 2009, pp. 23–78). The Weibull distribution was first introduced in the article by Waloddi Weibull (1939a), "A Statistical Theory of the Strength of Material". The moments of the Weibull distribution, as well as graphical and tabular aids for parameter estimation, can be found in 'The phenomenon of rupture in solids' (Weibull, 1939b). The Weibull distribution is a more flexible extension of the Laplace distribution and is also the limit distribution of a geometric summation scheme (Mittnik & Rachev, 1989). In modern statistics, the Weibull distribution is one of the most popular models, along with normal, exponential, χ^2, t, and F distributions.

There are several interesting properties of the Weibull distribution discussed by Mittnik and Rachev (1989) as follows.

- The Weibull distribution has exponential tails in addition to fatter tails, therefore guaranteeing the existence of all moments.
- The estimation of the parameters of the Weibull distribution is straightforward via maximum likelihood methods or regression-type estimation, and the existence of closed-form expression makes easier the construction of estimated Weibull distribution functions.
- The shape parameter of the Weibull distribution is invariant with respect to the sampling interval when fitting daily and monthly data over the same sampling period.

- The simulation result shows that the estimated shape parameter α of the stable Paretian distribution fitted to the data drawn from the Weibull distribution increases when the length of the sampling interval increases.
- In modeling asset return, the Weibull distribution has a *stabilizing tendency* when $\alpha \geq 1$ and a *destabilizing tendency* when $\alpha \leq 1$.

Due to its versatility, the Weibull distribution has been widely applied not only in the fields of material science and engineering but also in finance and climatology. In the study of asset return, Mittnik and Rachev (1989) found it to fit substantially better than any other model in their study, including the stable Paretian distribution, to the unconditional return distribution for the S&P500 index when fitted separately to positive and negative returns. It was employed as an error distribution in autoregressive conditional duration (ACD) models (Engle & Russell, 1998) and in a range-based threshold heteroskedastic model (Chen et al., 2008). Furthermore, Gebizlioglu et al. (2011) considered the Weibull distribution and its quantiles in estimating VaR.

Sornette et al. (2000) further developed a symmetric, two-sided 'modified' Weibull distribution, which was then used as an unconditional distribution of asset return by Malevergne and Sornette (2004). Chen and Gerlach (2013) proposed a more flexible asymmetric two-sided Weibull distribution as a conditional return distribution. Since a conditional error in a GARCH-type model should have a mean 0 and variance 1, Chen and Gerlach (2013) developed the standardized two-sided Weibull distribution (STW). They considered four popular volatility models (GARCH, GJR-GARCH, T-GARCH, and ST-GARCH) and found that the STW distribution performs at least as well as Gaussian, Student-t, skewed t, and asymmetric Laplace distribution for VaR forecasting, but performs most favorably for ES forecasting prior, during, and after crisis.

Several studies have adopted the STW distribution for tail risk forecasting. Silahli et al. (2019) applied the STW distribution to investigate the volatility dynamics of four major coins in the cryptocurrency market. In their study, they used a combination of historical simulation and GARCH model with innovations following the STW distribution. Wang et al. (2019) employed the STW distribution in the realized GARCH (Hansen et al., 2012) for tail risk forecasting purposes. They found that the realized GARCH models incorporating the STW distribution, especially those employing the sub-sampled realized variance and sub-sampled realized range, are more favorable than a range of well-known parametric GARCH and the original realized GARCH models.

n Chapter 3, the realized EGARCH model (Hansen & Huang, 2016) is extended by incorporating a standardized skewed Student-t distribution (Hansen, 1994), that is, the RE-SkN model. Meanwhile, the use of STW distribution in the realized EGARCH model is very limited or even not available up to the writing of this study. Therefore, this chapter proposes another variant of the realized EGARCH model by adopting the STW distribution (Chen & Gerlach, 2013) for the return error equation, hereafter referred as the RE-STW model. The RE-STW is used to forecast the VaR and ES of seven market indices, including during the

global financial crisis 2008. The forecast performance of the RE-STW model is then compared with that of the RE-SkN models and several other competing models.

This chapter is structured as follows. Section 5.2 provides an overview of the standardized two-sided Weibull distribution, and Section 5.3 reviews the VaR and ES for the standardized two-sided Weibull distribution. The proposed RE-STW model and the Bayesian framework for estimating the RE-STW models are discussed in Section 5.4. The simulation and empirical studies are discussed in Sections 5.5 and 5.6. Section 5.7 concludes this chapter.

5.2 A Standardized Two-sided Weibull Distribution

Wang et al. (2019) provides a general definition of a two-sided Weibull (TW) distribution, $Y \sim TW(\lambda_1, k_1, \lambda_2, k_2)$ if:

$$\begin{aligned} -Y &\sim \text{Weibull}(\lambda_1, k_1); Y < 0 \\ Y &\sim \text{Weibull}(\lambda_2, k_2); Y \geq 0, \end{aligned}$$

where $k_1, k2$ are the shape parameters and λ_1, λ_2 are the scale parameters.

An STW distribution is defined via $X = Y/\sqrt{\text{Var}(Y)}$, where $Y \sim TW(\lambda_1, k_1, \lambda_2, k_2)$. The probability density function (PDF) for an STW random variable is:

$$f(x|\lambda_1, k_1, k_2) = \begin{cases} b_p \left(\frac{-b_p x}{\lambda_1}\right)^{k_1 - 1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]; & x < 0, \\ b_p \left(\frac{b_p x}{\lambda_2}\right)^{k_2 - 1} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]; & x \geq 0. \end{cases} \tag{5.1}$$

The shape parameters satisfy $k_1, k_2 > 0$, the scale parameters satisfy $\lambda_1, \lambda_2 > 0$, and the constant $b_p^2 = \text{Var}(Y)$ is given by

$$b_p^2 = \frac{\lambda_1^3}{k_1}\Gamma\left(1 + \frac{2}{k_1}\right) + \frac{\lambda_2^3}{k_2}\Gamma\left(1 + \frac{2}{k_2}\right) - \left[-\frac{\lambda_1^2}{k_1}\Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{k_2}\Gamma\left(1 + \frac{1}{k_2}\right)\right]^2, \tag{5.2}$$

where $\Gamma(\bullet)$ denotes the gamma function.

To ensure that the pdf integrates to 1, then:

$$\frac{\lambda_1}{k_1} + \frac{\lambda_2}{k_2} = 1. \tag{5.3}$$

For parsimony and simplification, Chen and Gerlach (2013) set $k_1 = k_2$. Following (Wang et al., 2019), this constraint is relaxed in the proposed model in this chapter, thus $k_1 \neq k_2$ is considered. By this, we have an STW(λ_1, k_1, k_2) and need only to estimate three parameters, since $\lambda_2 = k_2(1 - \lambda_1/k_1)$. Since $Pr(X < 0) = \lambda_1/k_1$, then $0 < \lambda_1 \leq k_1$ and $\lambda_2 < k_2$ should hold. Positive (or right) skewness occurs if $\frac{\lambda_1}{k_1} < 0.5$, while when $\frac{\lambda_1}{k_1} > 0.5$, the density is negatively (or left) skewed. The symmetry is achieved when $\frac{\lambda_1}{k_1} = 0.5$.

The cumulative distribution function (CDF) of the $STW(\lambda_1, k_1, k_2)$ is

$$F(x|\lambda_1, k_1, k_2) = \begin{cases} \frac{\lambda_1}{k_1} \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]; & x < 0, \\ 1 - \frac{\lambda_2}{k_2} \exp\left[-\left(\frac{b_p x}{\lambda_2}\right)^{k_2}\right]; & x \geq 0. \end{cases} \tag{5.4}$$

The mean of an STW is given by

$$\mu_X = \frac{-\lambda_1^2}{b_p k_1}\Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2}\Gamma\left(1 + \frac{1}{k_2}\right). \tag{5.5}$$

The distribution of $Z = X - \mu_X$ is a shifted STW(λ_1, k_1, k_2) distribution with mean 0 and variance 1.

5.3 VaR and ES for the Standardized Two-sided Weibull Distribution

The general definition of VaR and ES has been discussed in Chapter 2 Section 2.2. In this chapter, the loss distribution F_L follows a STW distribution. The VaR is simply the quantile or inverse CDF given by

$$F^{-1}(\alpha|\lambda_1, k_1, k_2) = \begin{cases} -\frac{\lambda_1}{b_p}\left[-\ln\left(\frac{k_1}{\lambda_1}\alpha\right)\right]^{1/k_1}; & 0 \leq \alpha < \frac{\lambda_1}{k_1}, \\ \frac{\lambda_2}{b_p}\left[-\ln\left(\frac{k_2}{\lambda_2}(1-\alpha)\right)\right]^{1/k_2}; & \frac{\lambda_1}{k_1} \leq \alpha < 1, \end{cases} \tag{5.6}$$

Since returns only exhibit a mild skewness in practice, the estimated values for λ_1/k_1 are closer to 0.5 than α. In risk management, we focus only on the extreme tails of the return distribution, therefore the only relevant case is $\alpha < \lambda_1/k_1$ in Equation 5.6. Thus, the ES is

$$\begin{aligned} \text{ES}_\alpha &= \frac{-\lambda_1^2}{\alpha b_p k_1}\int_{(-b_p \text{VaR}_\alpha/\lambda_1)^{k_1}}^{\infty} \left[\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right]^{1/k_1+1-1} \\ &\quad \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}\right] \mathrm{d}\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \\ &= \frac{-\lambda_1^2}{\alpha b_p k_1}\Gamma\left(1 + \frac{1}{k_1}, \left(\frac{-b_p \text{VaR}_\alpha}{\lambda_1}\right)^{k_1}\right); 0 \leq \alpha < \frac{\lambda_1}{k_1}, \end{aligned} \tag{5.7}$$

where $\Gamma(s, x) = \int_x^\infty t^{s-1}e^{-t}\mathrm{d}t$ is the upper incomplete gamma function. See Chen and Gerlach (2013) for derivation details.

5.4 Bayesian Estimation of the Proposed Realized EGARCH-STW Model

5.4.1 Likelihood

The log-likelihood function for the proposed RE-STW model when $r_t < 0$ and $J = \sum_{t=1}^{n} I(r_t < 0)$ is:

$$l(r,x;\theta) = \underbrace{J\log(b_p) + \sum_{t=1}^{J}(k_1-1)\log\left(\frac{-b_p r_t}{\lambda_1}\right) - \sum_{t=1}^{J}\left(\frac{-b_p r_t}{\lambda_1}\right)^{k_1} - \frac{1}{2}\sum_{t=1}^{J}\log h_t}_{l(r;\theta)} \tag{5.8}$$
$$\underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[K\log(2\pi) + \log(|\Sigma|) + u_t'\Sigma^{-1}u_t\right]}_{l(x|r;\theta)},$$

and the log-likelihood function when $r_t \geq 0$ and $M = \sum_{t=1}^{n} I(r_t > 0)$ is

$$l(r,x;\theta) = \underbrace{M\log(b_p) + \sum_{t=1}^{M}(k_2-1)\log\left(\frac{b_p r_t}{\lambda_2}\right) - \sum_{t=1}^{M}\left(\frac{b_p r_t}{\lambda_2}\right)^{k_2} - \frac{1}{2}\sum_{t=1}^{M}\log h_t}_{l(r;\theta)} \tag{5.9}$$
$$\underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[K\log(2\pi) + \log(|\Sigma|) + u_t'\Sigma^{-1}u_t\right]}_{l(x|r;\theta)}.$$

Here, $r_t = Y/b_p - \mu_{r_t}$ is standardized and shifted with mean 0, where $b_p^2 = \text{VaR}(Y)$ is defined in Equation 5.2.

5.4.2 Prior specification

A combination of mostly uninformative and Jeffreys-type priors is chosen over the possible region for the parameters.

$$\pi(\theta) \propto \frac{1}{\sigma^2_{u,k=1}}\frac{1}{\sigma^2_{u,k=2}}..\frac{1}{\sigma^2_{u,k=K}}I(A) \tag{5.10}$$

where A is the region described by the parameter restriction in Table 5.1 and equation 5.11. The indicator function $I(A) = 1$ over region A and zero otherwise.

TABLE 5.1: Parameter Restriction

Parameter	Bounds
$\mu, \omega, \tau_1, \tau_2, \xi, \delta_1, \delta_2$	$\pm\infty$
σ_u^2	> 0
$\beta - \gamma^{'}\varphi$	< 1
λ_1, λ_2	> 0
k_1, k_2	> 0

Under the constraint for the STW parameters in Section 5.2, λ_1 and k_1 have a flat prior over their permissible region:

$$\pi(\lambda_1) \propto I(0 < \lambda_1 < k_1).$$

The joint posterior distribution is obtained by multiplying the likelihood in Equation 5.8 or Equation 5.9 and the joint prior in Equation 5.10.

5.4.3 Adaptive MCMC method

The likelihood in Equation 5.8 or Equation 5.9 involves 14 unknown parameters for the RE-STW model employing only one realized measure. With an additional realized volatility measure in the measurement equations, another six parameters of the RE-STW model need to be estimated. Since the performance and finite sample properties of Maximum likelihood (ML) estimates of these likelihoods are not yet studied, a powerful numerical and computational algorithms in a Bayesian framework is considered as a competing estimator for these models.

The adaptive MCMC method proposed in Section 3.5.3 is employed to numerically perform an inference on the model parameters, by drawing samples from the joint posterior distribution of the model parameters. Under the adaptive MCMC approach, the Robust Adaptive Metropolis (RAM) algorithm of Vihola (2012) is employed in the burn-in period and the standard RWM algorithm is employed in the post-burn-in period. The RAM algorithm is run for 20,000 iterations for each model parameter, the last 10,000 of which are used to calculate the sample mean (X_n) and the covariance matrix (Σ) for each parameter as the proposal distribution in the post burn-in period. The RWM algorithm is then run for 10,000 iterations. Each iterate for each parameter is used to form MCMC iterates for VaR and ES forecasts at the 2.5% and 1% level, as recommended by the Basel Committee (2012, 2016, 2019). The final VaR and ES forecasts are calculated as the posterior mean of the MCMC iterates for those VaR and ES forecasts. The one-step ahead forecast is produced by estimating the models with a fixed size in-sample data using the adaptive MCMC method in combination with a rolling window approach.

The comparison of two settings of parameter blocks are again conducted to choose an optimal setting. The first setting is to put all parameters as one block. The second setting is adapted from Wang et al. (2019) and adjusted to the proposed RE-STW models by considering the possible correlation among the parameters in the posterior (or likelihood). The

second setting consists of 4 blocks: $\theta_1 = (\mu)$, $\theta_2 = (\omega, \beta, \tau', \gamma', \varphi_k)$, $\theta_3 = (\xi_k, \delta_k', \Sigma)$, and $\theta_4 = (\lambda_1, k_1, k_2)$. The target mean acceptance rate follows the recommendation of Roberts et al. (1997), that is, $\alpha^* = 0.234$ (if $d > 4$), or 0.35 (if $2 \leq d \leq 4$), or 0.44 (if $d = 1$).

The convergence of the adaptive MCMC method is assessed via the Gelman-Rubin diagnostic, the effective sample size test (Gelman et al., 2014) and the autocorrelation time (Chib & Jeliazkov, 2001), as discussed in Section 1.6.

5.5 Simulation Study

The performance of the MCMC estimators of the proposed RE-STW models in comparison with the Maximum Likelihood (ML) estimators is assessed via a simulation study. The MCMC estimators are obtained using the adaptive MCMC method, while the ML estimators are obtained through the direct optimization approach by minimizing the negative of the log-likelihood function in Equation 5.8 or Equation 5.9. This is conducted by using both the standard 'fmincon' constrained optimization function and the 'MultiStart' function in Matlab. Under the 'MultiStart' function setting, 50 different starting points were chosen to generate 50 local solutions, among which the global optimum was selected.

5.5.1 Simulation set up and model

A total of 1000 replicated datasets are simulated from the proposed RE-STW model in Equation 5.11 using one realized measure, with a sample size of 2000.

$$\begin{aligned} r_t &= -0.007 + \sqrt{h_t}\epsilon_t, \\ \log h_t &= 0.008 + 0.92 \log h_{t-1} - 0.04\epsilon_{t-1} + 0.02(\epsilon_{t-1}^2 - 1) + 0.25u_{t-1}, \\ \log x_t &= -0.28 + 0.99 \log h_t - 0.07\epsilon_t + 0.08(\epsilon_t^2 - 1) + u_t,\ u_t = \sigma_u z_t, \end{aligned} \tag{5.11}$$

where $\epsilon_t \sim STW(0.69, 1.16, 1.1)$, $z_t \sim N(0,1)$, $\sigma_u^2 = 0.25$, and $h_0 = 0.0025$. During estimation, all initial parameter values were arbitrarily set equal to 0.5.

The estimation of the simuated dataset is run on the University of Sydney's Ártemis High-Performance Computing (HPC). The simulation job is divided into 40 chunks to reduce processing time, each of which estimates the RE-STW model and forecasts the tail risk for time $t+1$ for 25 datasets. All 40 processing jobs are run simultaneously in Artemis, each of which takes approximately 8 hours 50 minutes to complete.

5.5.2 MCMC convergence and efficiency

Using $m = 10$ independent MCMC chains, the convergence of the proposed adaptive MCMC method under two different parameter block settings as explained in Section 5.4.3 is assessed. The summary of the the Gelman-Rubin diagnostic $\hat{R}$, the number of effective sample size n_{eff} (Gelman et al., 2014), and the autocorrelation time $\hat{t}$ (Chib & Jeliazkov, 2001)

for the one parameter-block and four parameter-blocks settings are presented in Table 5.2. Favorable results are indicated by the values printed in bold.

The convergence and efficiency of the four-block setting are found to be more satisfactory than those of the one-block setting. The $\hat{R}$ for each parameter is very close to 1 and all $\hat{R}$ are < 1.1, the threshold recommended by Gelman et al. (2014). The $\hat{R}$ in the four-block settings is closer to 1 for all parameters, suggesting an improved convergence property under this block setting of parameters. The n_{eff} in the four-block setting is 2.4 times (for μ parameter) to 6.7 times (for the λ_1 parameter) higher than the n_{eff} in the one-block setting. As for the correlation time, the parameters under the 4 blocks setting have a lower autocorrelation time than those under the 1 block setting, for example, the ratio ranges from 0.146 times (for λ_1 parameter) to 0.401 times (for μ parameter).

TABLE 5.2: Summary statistics of the Gelman-Rubin diagnostic, effective sample size and autocorrelation time of the RE-STW model estimated using Bayesian method for simulated dataset

Parameter	**1 block**			**4 blocks**		
	$\hat{R}$	n_{eff}	$\hat{t}$	$\hat{R}$	n_{eff}	$\hat{t}$
μ	1.0319	649.8	68.7	**1.0036**	**1579.4**	**27.6**
ω	1.0157	588.1	83.6	**1.0053**	**1489.1**	**32.0**
β	1.0137	879.4	53.8	**1.0010**	**2702.2**	**17.5**
γ	1.0206	515.1	94.4	**1.0063**	**1624.7**	**29.1**
τ_1	1.0166	828.0	56.7	**1.0024**	**2478.2**	**18.0**
τ_2	1.0123	699.7	62.9	**1.0049**	**2102.2**	**22.3**
ξ	1.0302	375.9	126.5	**1.0084**	**1061.7**	**45.7**
φ	1.0206	443.1	108.1	**1.0073**	**1430.0**	**33.5**
δ_1	1.0089	686.7	70.2	**1.0026**	**2373.9**	**16.3**
δ_2	1.0233	744.4	57.9	**1.0042**	**2184.1**	**21.9**
σ_u^2	1.0107	795.3	59.4	**1.0019**	**4233.7**	**10.9**
λ_1	1.0212	450.5	105.4	**1.0016**	**3023.5**	**15.4**
k_1	1.0157	521.4	93.5	**1.0021**	**3202.4**	**14.3**
k_2	1.0177	413.8	89.5	**1.0035**	**2327.1**	**15.6**

The trace plots of the MCMC chains using the RAM algorithm during the burn-in period (Figure 5.1) and the RWM algorithm during the post-burn-in period (Figure 5.2) also show a very rapid convergence and the effective tuning process of the covariance matrix. With an initial value of 0.5, some selected parameters, that is $\mu, \omega, \beta, \gamma$, and λ_1, during the burn-in period quickly converge to their stationary values. The acceptance rates of the 4 blocks are 0.4456, 0.1838, 0.3426, and 0.3477, which are very close to the target acceptance rates discussed in Section 5.4.3.

FIGURE 5.1: RAM iterations with simulated data from the RE-STW model

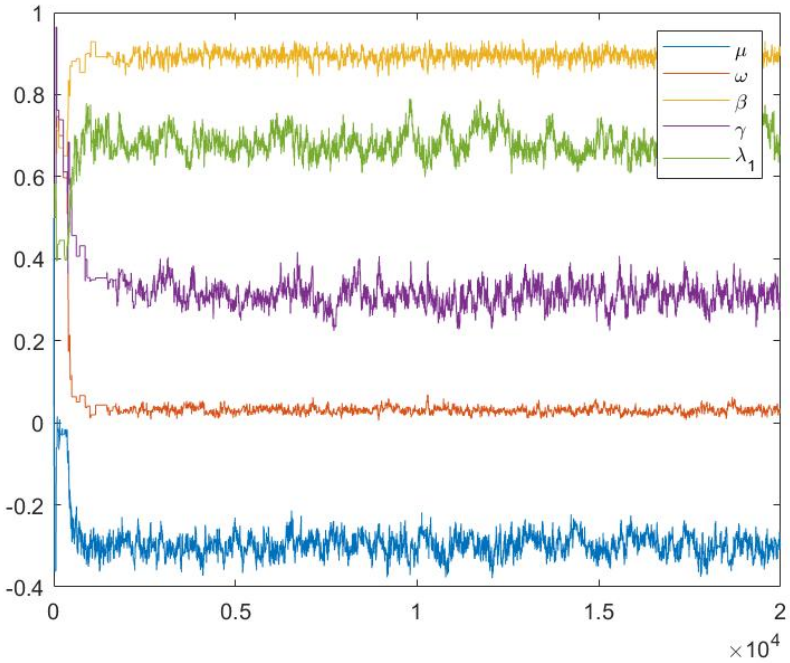
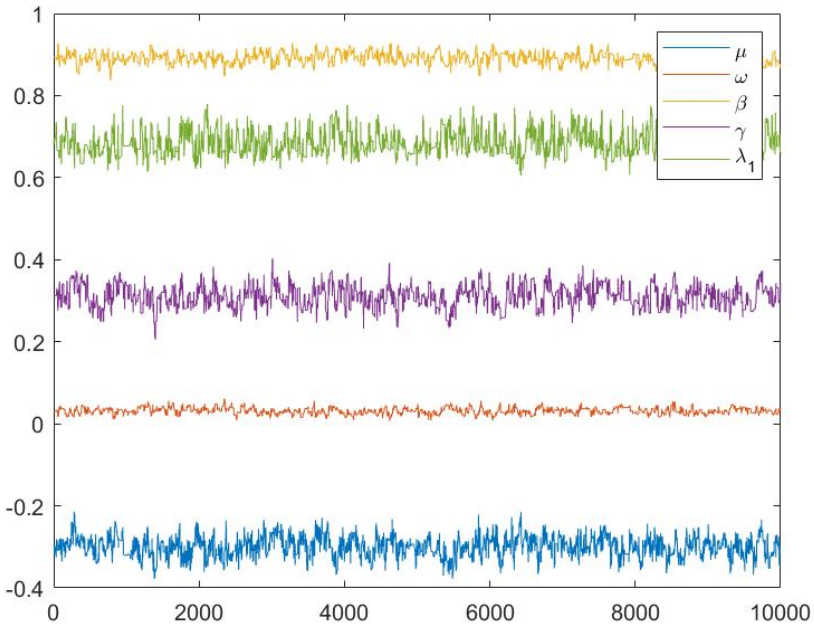


FIGURE 5.2: RWM iterations with simulated data from the RE-STW model



5.5.3 Simulation results

The results of the estimation of the proposed RE-STW model based on the two ML methods and the adaptive MCMC method using the four-parameter block setting are summarized in Table 5.3. The parameters closed to the true values are printed in bold. True one-step-ahead

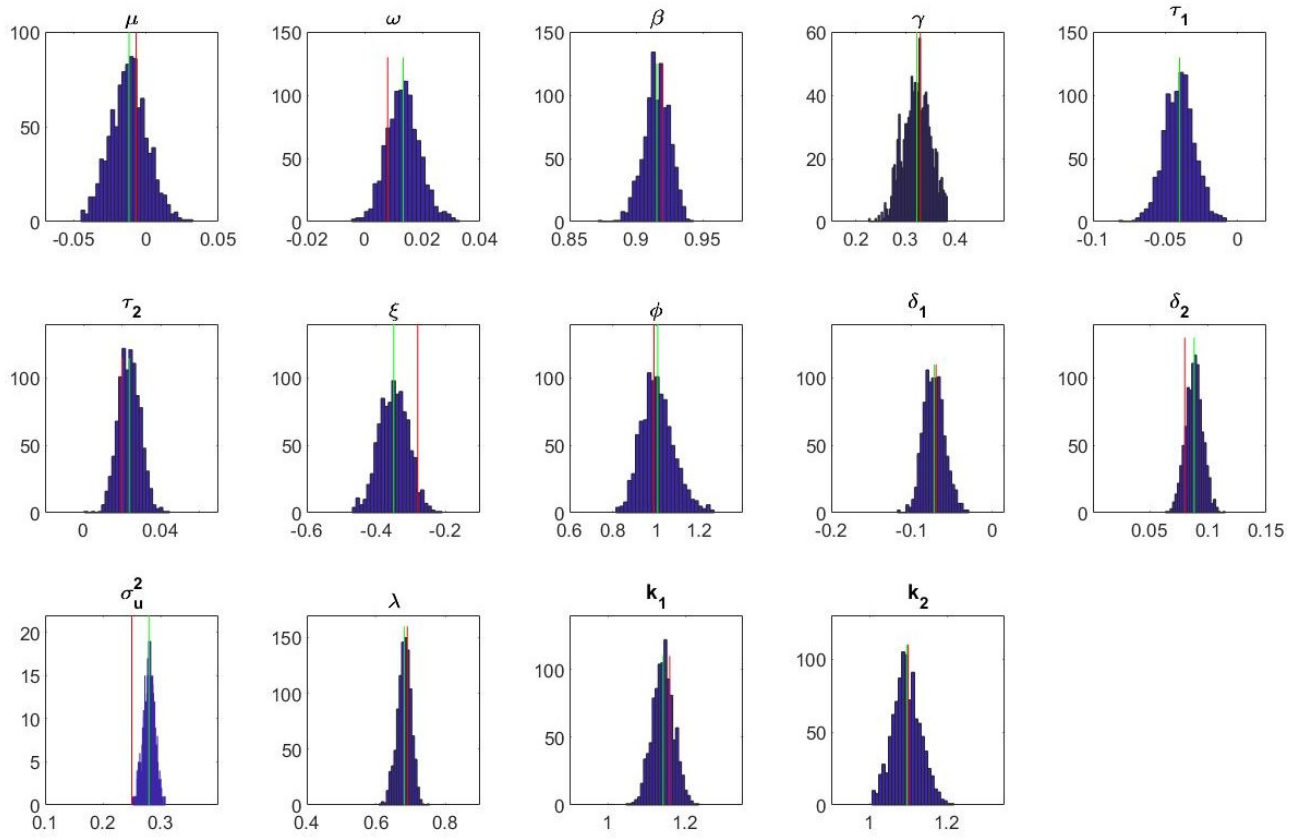
VaR and ES forecasts are calculated, and their average over the 1000 datasets is given in the 'True' column. The results show that the MCMC estimators are more precise (lowest RMSE) and less unbiased than the estimators of both ML methods. The only exception is the δ_1 parameter, which has the smallest bias under ML with MultiStart setting.

TABLE 5.3: Summary statistics of the RE-STW model estimators based on Maximum Likelihood and the MCMC methods

n=200	True	ML		Multi-ML		MCMC	
Parameter		Mean	RMSE	Mean	RMSE	Mean	RMSE
μ	-0.007	-0.0146	0.0831	0.0194	0.1027	**-0.0119**	**0.0137**
ω	0.008	0.0430	0.1273	0.0703	0.1141	**0.0134**	**0.0079**
β	0.92	0.8861	0.9032	0.8547	0.1663	**0.9154**	**0.0107**
γ	0.33	0.3473	0.3632	0.3989	0.1494	**0.3228**	**0.0288**
τ_1	-0.04	-0.0424	0.0541	-0.0461	0.0635	**-0.0401**	**0.0105**
τ_2	0.02	0.0283	0.0475	0.0368	0.0530	**0.0237**	**0.0068**
ξ	-0.28	-0.4256	0.5718	-0.5114	0.2789	**-0.3494**	**0.0812**
φ	0.99	1.0170	1.3468	0.9176	0.2339	**1.0070**	**0.0777**
δ_1	-0.07	-0.0769	0.0790	**-0.0716**	0.0655	-0.0724	**0.0134**
δ_2	0.08	0.0951	0.1051	0.1098	0.0486	**0.0880**	**0.0109**
σ_u^2	0.25	0.2936	0.3144	0.3644	0.2074	**0.2796**	**0.0309**
λ_1	0.69	0.6413	0.6632	0.8738	0.3341	**0.6810**	**0.0219**
k_1	1.16	1.1100	1.1219	1.2149	0.2015	**1.1430**	**0.0327**
k_2	1.1	1.1167	1.1308	0.9877	0.2312	**1.0962**	**0.0364**
2.5% VaR	-2.3847	-2.5579	0.4221	-2.8449	0.6336	**-2.4657**	**0.1383**
1% VaR	-3.0426	-3.2853	0.5947	-3.6538	0.8425	**-3.1559**	**0.1883**
2.5% ES	-3.0913	-3.3448	0.6141	-3.7141	0.8602	**-3.2083**	**0.1936**
1% ES	-3.7293	-4.0598	0.7964	-4.4992	1.0740	**-3.8799**	**0.2457**

The histogram of the posterior mean of each parameter of the proposed RE-SkN model over the 1000 datasets (Figure 5.3) shows an accurate and approximately unbiased estimation for all parameters. While the estimated σ_u^2 based on the MCMC method is slightly higher than the true value, it is still closer to the true value than those of both ML methods.

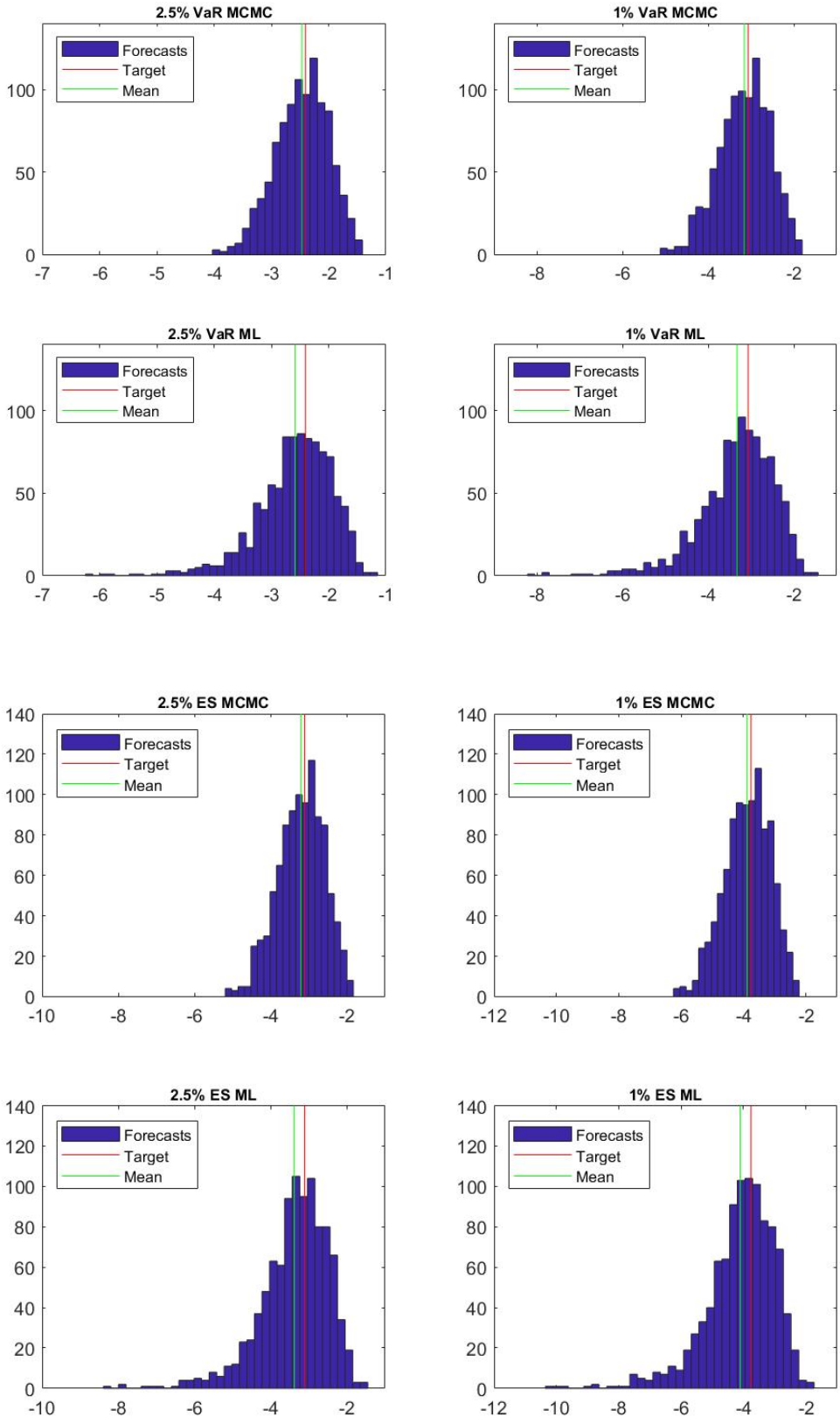
FIGURE 5.3: Histogram of 1000 posterior means for the RE-STW Model



Note: Red line indicates the true parameter and the green line indicates the mean estimate.

The performance of the MCMC and ML estimators in terms of unbiasedness and precision of the risk forecasts is also illustrated by using the distribution of 1000 forecasts of 2.5% and 1% VaR and ES in Figure 5.4. Although both the MCMC and ML methods generate VaR and ES forecasts relatively close to the true value of VaR and ES (represented by the red horizontal lines), it can be seen that the ML method produces more outlying forecasts than the MCMC method. The mean values of both VaR and ES forecasts generated using the MCMC method are also very close to their respective true values, compared to those of the ML method. This could be due to the convergence issue in the ML method, which results in a higher RMSE of the risk forecasts. This suggests that the MCMC method is more preferable compared to the ML method, as it produces more accurate risk forecasts.

FIGURE 5.4: Histogram of 1000 forecasts of 2.5% and 1% VaR and ES: MCMC and ML estimation



5.6 Empirical Study

The estimation of historical data in this chapter comprises the estimation of the proposed RE-STW model and the competing model proposed in Chapter 3, that is, the RE-SkN model. Both models are estimated using the adaptive MCMC approach proposed in Section 5.4 for the RE-STW model and in Section 3.5 for the RE-SkN model using 4 blocks of parameters. This chapter also includes some other competing models, such as the GARCH, EGARCH, GJR-GARCH, and GARCH-EVT models, all with Student-*t* distribution. The GARCH-EVT-t model combines the application of the peaks over threshold of the extreme value theory (EVT) method to the standardized residuals of the GARCH model with Student-*t* distribution (McNeil & Frey, 2000). The threshold is set for 10% unconditional quantile for the lower tail (Taylor, 2008). A filtered GARCH approach (GARCH-HS-t) is also estimated according to Wang et al. (2019). The last model used to forecast VaR and ES using the historical data is the symmetric absolute value conditional autoregressive expectiles (CARE-SAV) model of Taylor (2008).

The GARCH, EGARCH, and GJR-GARCH models are estimated using the Econometrics toolbox in Matlab, while the GARCH-EVT-t, CARE-SAV, RE-SkN, and RE-STW models are estimated using the Matlab code developed for this thesis. The parameters of the models are used to generate m one-step-ahead VaR and ES forecasts for the forecast period specified for each market index in Table 5.4.

Several tables are presented in the following sections. They contain the abbreviations of various model names. In each model name, RE refers to the realized EGARCH model, RV refers to the subsumpled RV, RR refers to the subsampled RR, and RK refers to the Realized Kernel. The last term of the model name refers to the type of error distributions: N for the Normal distribution, t for the standardized Student-t distribution, Sk for the standardized Skewed Student-t distribution, and STW for the Standardized Two-sided Weibull distribution.

5.6.1 Data

The data used in this chapter is the same historical dataset used in Chapter 3, sourced from the dataset used in Wang et al. (2019) and the Oxford-Man Institute's Realized Library, Library Version: 0.3. The daily return data for seven market indices are collected; these are the ASX 200 (Australia), DAX (Germany), FTSE 100 (UK), Hang Seng (Hong Kong), NASDAQ (US), SMI (Swiss), and S&P500 (US). This chapter employs subsampled RV (RVSS) and subsampled RR (RRSS) the most robust realized measures of volatility, and also includes the Realized Kernel (RK) for comparison purposes.

The forecast period in this chapter is different from that in Chapter 3. The tail risk forecasts in Chapter 3 cover the last 1000 of the sample for each market index. In this chapter, the tail risk forecasts for each index start from 1 January 2008, to take into account the effect of the global financial crisis (GFC) on each market. The descriptive statistics of the return series for each market index for both the estimation and the forecast periods are presented in Table 5.4. Most market indices have a positive average return in the estimation period,

except the NASDAQ. During the forecast period, three of the seven market indices generate a negative average return. The returns are negatively skewed, both in the estimation and forecast periods, except for the Hang Seng in the estimation period and the NASDAQ in the forecast period. The Hang Seng, in particular, is the most volatile among the market indices during the period of this study. Its maximum return over the study period is both the highest and the lowest. The ASX 200 is the most stable market index during the study period. The test for normality (Jarque & Bera, 1987) gives JB statistics higher than the 5% critical JB test value, that is, 5.99. This suggests that all return series are not normally distributed.

TABLE 5.4: Descriptive statistics for each return series

Index	Period	T	Total	Mean	Stdv	Skewness	Kurtosis	Min	Max	JB
ASX 200	Estimate (n)	1857	3966	0.033	0.755	-0.427	6.179	-4.806	4.510	838.6
	Forecast (m)	2109		-0.010	1.191	-0.347	6.567	-7.713	5.628	1160.4
DAX	Estimate (n)	1936	4066	0.009	1.544	-0.035	5.866	-6.500	7.553	662.9
	Forecast (m)	2130		0.008	1.517	-0.053	7.201	-7.434	10.685	1567.3
FTSE 100	Estimate (n)	1943	4081	0.000	1.127	-0.253	6.352	-5.885	5.903	934.2
	Forecast (m)	2138		-0.001	1.310	-0.105	10.200	-9.266	9.384	4604.8
Hang Seng	Estimate (n)	1879	3937	0.024	1.320	-0.306	6.545	-9.285	5.759	1006.1
	Forecast (m)	2058		-0.016	1.701	0.034	12.155	-13.582	13.407	7147.7
NASDAQ	Estimate (n)	1890	4001	-0.024	1.774	0.251	7.760	-10.168	13.255	1804.3
	Forecast (m)	2111		0.028	1.468	-0.537	12.219	-12.452	11.159	7577.5
SMI	Estimate (n)	649	2730	0.031	0.923	-0.515	6.254	-5.407	4.451	314.9
	Forecast (m)	2081		0.000	1.255	-0.549	14.835	-12.735	10.788	12250.0
S&P 500	Estimate (n)	1905	4018	0.002	1.110	-0.085	6.309	-7.257	5.573	871.2
	Forecast (m)	2113		0.016	1.385	-0.421	13.256	-10.003	10.957	9323.7

Note: $T = n + m$ is the number of observation, JB is the Jarque-Bera statistic to test the null hypothesis of normality.

5.6.2 Estimation of the RE-STW and RE-SkN models

The estimation of the RE-STW and RE-SkN model to forecast VaR and ES is performed using the University of Sydney's Artemis High-Performance Computing (HPC). With a single and combined use of three realized measures, there are seven types of RE-SkN models and seven types of RE-STW models to estimate for each market index. For both models and seven market indices, there is a total of 98 models to estimate to produce the VaR and ES forecasts. For producing tail risk forecasts, fixed size in-sample data is combined with a rolling window approach to produce m one-step ahead forecasts. The forecasting load for each model and each market index is broken down into 22 separate chunks and runs on multiple computers in the Artemis HPC cluster. Each chunk takes approximately 5 to 7 hours to complete.

5.6.3 Model parameter estimates

This Section discusses the posterior mean of the parameters for all forecasting steps of the proposed RE-STW models. The average of posterior means for each parameter of the RE-STW models for all forecasting steps for each parameter over the seven market indices is presented in Table 5.5. The order of the realized volatility measures in the measurement equations is the RVSS, the RRSS, and the RK. It is interesting to see the coefficients of γ, which indicates how strong the signal about future volatility is contained in the corresponding realized volatility measure. On average, the RVSS and RRSS could give a stronger signal than tge RK, indicated by the higher γ_1 and γ_2 coefficients on average, with the RVSS being the highest on average over the forecast period. The RVSS also produces the lowest measurement errors $\sigma^2_{u,1}$, on average, compared to the RVSS and the RK.

It is valuable to more closely examine the coefficient of γ_3, which represents the signal contained by the RK. While it is relatively lower than the coefficients of γ_1 and γ_2, it is still relatively high for some market indices and not approaching zero, except for the case of the FTSE 100. In the RE-SkN models using the Lasso approach in Chapter 4, the coefficients of RK shrink to nearly 0 in three of the seven market indices. It would be interesting to see the use of the Cross Validation and the BLasso method proposed in Chapter 4 to the RE-STW.

The estimates of other parameters are quite consistent with those of the RE-SkN model in Chapter 3. The high value of β coefficient suggests a volatility persistence, although it is less than 1. The leverage functions have the expected signs, τ_1 and δ_1 are negative, while τ_2 and δ_2 are positive, except for $\delta_{2,3}$ for the ASX 200. The estimates of ϕ are close to 1 and the estimates of ξ are all negative, suggesting that the realized volatility measures are proportional to the conditional variance with negative corrections given by the ξ, which relatively varies across types of realized volatility measures and market indices. The average posterior means of each parameter for individual market index can be seen in Appendix C.

TABLE 5.5: The average of posterior means of RE-STW model parameters for all forecasting steps for each market index

Parameter	ASX 200	DAX	FTSE 100	Hang Seng	NASDAQ	SMI	SP500
μ	0.0004	0.0004	0.0000	0.0002	0.0003	-0.0002	0.0001
ω	-0.197	-0.205	-0.245	-0.192	-0.331	-0.465	-0.323
β	0.979	0.977	0.973	0.977	0.963	0.949	0.965
γ_1	0.115	0.174	0.214	0.173	0.240	0.229	0.260
γ_2	0.123	0.097	0.056	0.104	0.135	0.134	0.091
γ_3	0.055	0.049	0.007	0.042	0.072	0.020	0.029
τ_1	-0.110	-0.137	-0.136	-0.050	-0.162	-0.128	-0.189
τ_2	0.028	0.045	0.049	0.060	0.046	0.042	0.043
ξ_1	-1.234	-0.421	-0.287	-1.469	-1.340	-0.464	-0.845
ξ_2	-1.018	-0.461	-0.425	-1.419	-1.247	-0.967	-0.766
ξ_3	-0.448	-0.311	-0.157	-0.581	-0.819	-0.751	-0.620
φ_1	0.971	1.017	1.038	0.959	0.950	1.011	0.988
φ_2	0.981	1.016	1.019	0.973	0.964	0.931	1.000
φ_3	0.985	1.032	1.026	1.063	0.992	0.926	1.000
$\delta_{1,1}$	-0.098	-0.163	-0.133	-0.107	-0.156	-0.145	-0.143
$\delta_{1,2}$	-0.130	-0.147	-0.104	-0.100	-0.147	-0.134	-0.104
$\delta_{1,3}$	-0.176	-0.127	-0.080	-0.102	-0.134	-0.139	-0.049
$\delta_{2,1}$	0.107	0.084	0.095	0.066	0.051	0.054	0.088
$\delta_{2,2}$	0.044	0.103	0.137	0.083	0.070	0.054	0.140
$\delta_{2,3}$	-0.008	0.135	0.200	0.122	0.099	0.063	0.211
$\sigma^2_{u,1}$	0.340	0.196	0.172	0.228	0.211	0.184	0.221
$\sigma^2_{u,2}$	0.423	0.234	0.212	0.319	0.262	0.381	0.290
$\sigma^2_{u,3}$	0.689	0.306	0.292	0.485	0.350	0.682	0.405
λ_1	0.771	0.688	0.732	0.603	0.687	0.712	0.671
k_1	1.375	1.161	1.314	1.321	1.341	1.217	1.480
k_2	1.526	1.261	1.346	1.201	1.360	1.323	1.283

5.6.4 VaR and ES backtests

The tail risk forecast performance of the RE-STW model is evaluated based on several tests discussed in Section 2.3 and compared with those of the RE-SkN model and other competing models. The tail risks are considered for the quantile level $\alpha = 2.5\%$ and $\alpha = 1\%$, following the recommendation of the Basel Committee (2012; 2016; 2019).

Figure 5.5 shows the 1% VaR forecasts for S&P500 based on the RE-STW and RE-SkN models, both employing RVSS and RRSS in the measurement equations, as well as the 1% VaR forecasts of several competing models. The RE-STW model generates the most extreme 1% VaR forecast during the 2008 global crisis compared to other models, followed by the RE-SkN model. The 1% VaR forecast generated by the RE-STW model on 13 October 2008 is -20.28% and -17.11% on 21 November 2008. Meanwhile, the RE-SkN model produces the

most extreme 1% VaR forecast on 27 October 2008 (-14.49%) and on 21 November 2008 (-14.79%). Other competing models produce less extreme 1% VaR forecasts around October to November 2008. This suggests that the RE-STW and RE-SkN models could provide useful information to financial institutions about capital reserves to cover extreme losses such as the global financial crisis.

FIGURE 5.5: The 1% VaR forecasts for S&P500 with RE-STW and RE-SkN models employing RVSS and RRSS, and some competing models

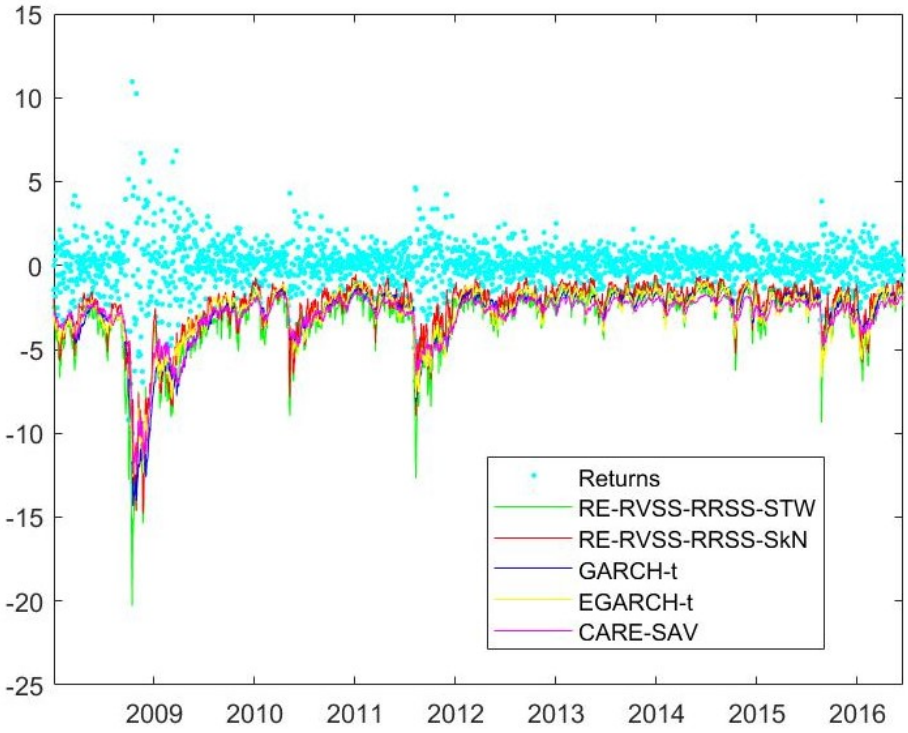


Table 5.6 shows the VaR violation rates (VRates) for each model for all market indices for the 2.5% and 1% forecast levels. In general, both the RE-STW and RE-SkN type models generate more optimal VaR forecast series than other models. At the 2.5% forecast level, six of the RE-STW type models produce a VRate that is closest to the nominal VRate 2.5% for four market indices, while there are only three RE-SkN models that produce a VRate closest to the nominal VRate 2.5%. However, at the 1% forecast level, the RE-SkN model is more conservative than the RE-STW. For all market indices, various RE-SkN models generate the VRate very close to 1%.

TABLE 5.6: VaR Forecasting violation rate

Model	ASX 200	DAX	FTSE 100	HK	NASDAQ	SMI	S&P500	Average
α = 2.5%								
GARCH-t	4.03	3.94	4.16	3.30	3.98	3.27	4.07	3.82
EGARCH-t	3.79	3.99	3.79	3.06	3.69	3.94	3.83	3.73
GJR-t	3.98	3.99	3.79	3.16	3.88	3.27	3.98	3.72
GARCH-HS-t	2.70	3.43	**2.53**	2.92	3.84	2.79	3.12	3.05
GARCH-EVT-t	2.32	0.05	0.09	0.10	0.05	0.53	0.57	0.53
CARE-SAV	**2.56**	2.77	2.85	2.92	2.98	3.08	2.79	2.85
RE-RV-SkN	2.37	2.86	2.62	3.40	2.98	6.92	3.50	3.52
RE-RR-SkN	2.18	2.72	2.25	**2.48**	**2.61**	2.84	2.93	**2.57**
RE-RK-SkN	**2.47**	8.17	2.90	3.21	2.89	**2.74**	3.12	3.64
RE-RV-RR-SkN	2.28	2.72	2.67	2.77	3.03	2.98	3.08	2.79
RE-RV-RK-SkN	2.37	2.72	**2.57**	3.26	5.16	2.79	3.31	3.17
RE-RR-RK-SkN	2.18	4.79	2.34	2.87	2.70	2.93	3.03	2.98
RE-RV-RR-RK-SkN	2.42	2.72	**2.53**	2.77	2.89	2.79	3.27	2.77
RE-RV-STW	1.66	2.25	2.15	2.33	**2.51**	2.07	2.70	2.24
RE-RR-STW	1.61	2.07	1.92	2.72	2.13	**2.31**	2.37	2.16
RE-RK-STW	2.75	2.21	2.15	2.14	2.70	2.21	2.65	**2.40**
RE-RV-RR-STW	1.47	**2.30**	1.87	**2.58**	2.23	2.11	**2.60**	2.17
RE-RV-RK-STW	2.42	**2.35**	2.10	2.19	**2.51**	2.11	2.79	2.35
RE-RR-RK-STW	2.13	2.11	1.92	2.62	2.27	2.16	**2.46**	2.24
RE-RV-RR-RK-STW	1.52	**2.35**	1.82	2.38	2.27	1.92	2.70	2.14
α = 1%								
GARCH-t	1.71	1.41	1.73	1.65	1.85	1.54	1.51	1.63
EGARCH-t	1.42	1.41	1.78	1.31	1.61	1.78	1.56	1.55
GJR-t	1.47	1.41	1.78	1.26	1.56	1.59	1.47	1.51
GARCH-HS-t	0.81	1.36	1.36	1.26	2.46	1.54	1.56	1.48
GARCH-EVT-t	2.23	0.05	0.09	0.10	0.05	0.53	0.52	0.51
CARE-SAV	0.90	1.41	1.26	1.12	1.56	1.30	1.47	1.29
RE-RV-SkN	1.14	1.08	**1.03**	**1.02**	1.28	3.17	1.33	1.43
RE-RR-SkN	0.85	0.89	0.80	0.73	**0.90**	1.06	**0.99**	0.89
RE-RK-SkN	**1.00**	4.13	**0.94**	**1.02**	1.18	**1.01**	1.14	1.49
RE-RV-RR-SkN	**1.04**	**0.94**	**0.94**	1.12	1.14	**0.96**	1.14	**1.04**
RE-RV-RK-SkN	**1.04**	**1.03**	**1.03**	1.17	2.61	**1.01**	1.14	1.29
RE-RR-RK-SkN	0.85	2.44	**0.94**	**0.92**	**0.99**	1.25	**0.90**	1.19
RE-RV-RR-RK-SkN	**1.04**	**0.94**	1.08	1.17	1.23	1.11	1.14	**1.10**
RE-RV-STW	0.38	0.42	0.56	0.78	0.71	0.62	0.66	0.59
RE-RR-STW	0.38	0.47	0.56	0.73	0.66	0.62	0.57	0.57
RE-RK-STW	1.28	0.61	0.61	0.83	0.76	0.77	0.76	0.80
RE-RV-RR-STW	0.33	0.47	0.56	0.73	0.66	0.43	0.57	0.54
RE-RV-RK-STW	0.85	0.42	0.56	0.83	0.71	0.77	0.76	0.70
RE-RR-RK-STW	0.57	0.47	0.51	0.73	0.66	0.72	0.57	0.60
RE-RV-RR-RK-STW	0.38	0.47	0.47	0.53	0.66	0.48	0.71	0.53

Note: For individual models, a green highlight indicates the favored model and bold indicates the 2nd ranked model.

With regard to the type of realized volatility measures, no distinctive pattern can be observed as to which combination of realized volatility measures gives the most optimal VRates. However, the RE-RK-SkN and RE-RV-RK-SkN models are most favorable for three

market indices at the 1% forecast level, followed by the RE-RV-SkN. At the 2.5% forecast level, the RE-RV-RK-STW is the most favorable, although only for two market indices.

Three types of VaR backtests are conducted based on the Unconditional Coverage (UC) test (Kupiec, 1995) discussed in Section 2.3.1, the Conditional Coverage (CC) test discussed in Section 2.3.2, and the Dynamic Quantile (DQ) test (Dumitrescu et al., 2012; Engle & Manganelli, 2004) using four lags discussed in Section 2.3.3 The result of those VaR backtests conducted at the 5% signifcance level is presented in Table 5.7.

In general the proposed RE-STW and RE-SkN type models have a lower number of rejections, across market indices as highlighted in green, than the GARCH-type models and CARE-SAV model. As expected, the standard GARCH-family models have relatively high rejection rates. At the 2.5% forecast level, both the RE-STW and the RE-SkN models have relatively low rejection rates, with the RE-STW type models performing slightly better than the RE-SkN model. At the 1% forecast level, the RE-SkN models have lower rejections in the UC and CC tests but higher rejections in the DQ test.

TABLE 5.7: VaR backtest: UC, CC, and DQ tests

Model	$\alpha = 2.5\%$			$\alpha = 1\%$		
	UC	**CC**	**DQ4**	**UC**	**CC**	**DQ4**
GARCH-t	7	5	5	6	6	6
EGARCH-t	6	6	6	4	4	6
GJR-t	6	6	3	5	3	5
GARCH-HS-t	2	2	4	3	3	3
GARCH-EVT-t	6	6	6	7	6	6
CARE-SAV	**1**	2	7	4	5	5
RE-RV-SkN	3	3	3	**1**	**1**	3
RE-RR-SkN	**0**	**0**	**0**	**0**	**0**	**1**
RE-RK-SkN	2	**1**	4	**1**	**1**	2
RE-RV-RR-SkN	**0**	**1**	3	**0**	**0**	2
RE-RV-RK-SkN	3	2	2	**1**	**1**	2
RE-RR-RK-SkN	**1**	**1**	4	**1**	**1**	2
RE-RV-RR-RK-SkN	**1**	**1**	3	**0**	**0**	**1**
RE-RV-STW	**1**	**1**	**1**	3	2	**0**
RE-RR-STW	**1**	**1**	**1**	4	2	**0**
RE-RK-STW	**0**	**1**	**1**	**1**	**0**	2
RE-RV-RR-STW	**1**	**1**	**1**	5	3	**0**
RE-RV-RK-STW	**0**	**0**	2	2	**1**	**1**
RE-RR-RK-STW	**0**	**0**	2	4	2	**0**
RE-RV-RR-RK-STW	2	2	**1**	5	4	**1**

Note: A green highlight indicates the favored model, bold indicates the 2nd ranked model

The ES backtests based on the multi-quantile regression (ES-MQR) approach of Couperier

and Leymarie (2019) in Equation 2.30 are performed by using the number of quantile $p = 6$ at the coverage level $\tau = 2.5\%$ across models and all market indices. The hypothesis testing of four ES-MQR test is conducted at the 5% significance level based on bootstrap critical values. The test consists of joint backtests that sum the intercept and slope coefficients (H_{0,J_1}) and the intercept backtests ($H_{0,I}$) and individual intercept backtest ($H_{0,I}$), and slope backtest ($H_{0,S}$).

The results of the ES-MQR test are presented in Table 5.8. Consistent with the ES-MQR test result in Chapter 3 Section 3.7.4, the individual ES-MQR tests are less sensitive, giving less clear distinction in the ES prediction accuracy accross models. However, it can observed that the RE-STW and RE-SkN models tend to have low rejections in all four ES backtests. The most favored models with only one rejection in all four types of backtests across market indices include four types of RE-SkN models and only one type of RE-STW model. Nevertheless, two types of RE-STW models are the second best models, producing ES forecasts with only five total ES-MQR test rejections. Therefore, it can be concluded that the RE-SkN type models have a slightly better performance in forecasting the ES.

TABLE 5.8: ES Backtests based on the Multi-Quantile Regression (ES-MQR) method

Model	**Hypothesis**				**Total**
	J_1	J_2	I	S	**Rejection**
GARCH-t	**1**	4	**1**	**1**	7
EGARCH-t	2	3	2	2	9
GJR-t	2	**2**	2	**1**	7
GARCH-HS-t	6	4	6	4	20
GARCH-EVT-t	**1**	4	**1**	2	8
CARE-SAV	**1**	5	**1**	3	10
RE-RV-SkN	2	**1**	2	**1**	6
RE-RR-SkN	2	**2**	**1**	3	8
RE-RK-SkN	**1**	**1**	**1**	**1**	**4**
RE-RV-RR-SkN	**1**	**1**	**1**	**1**	**4**
RE-RV-RK-SkN	**1**	**1**	**1**	**1**	**4**
RE-RR-RK-SkN	2	2	2	2	8
RE-RV-RR-RK-SkN	**1**	**1**	**1**	**1**	**4**
RE-RV-STW	**1**	2	**1**	2	6
RE-RR-STW	**1**	**2**	**1**	**1**	5
RE-RK-STW	**1**	**1**	**1**	**1**	**4**
RE-RV-RR-STW	**1**	2	**1**	**1**	5
RE-RV-RK-STW	2	2	2	2	8
RE-RR-RK-STW	2	3	2	3	10
RE-RV-RR-RK-STW	2	3	2	3	10

Note: A green highlight indicates the favored model

5.6.5 Quantile loss function values

The quantile loss function values at the 2.5% and 1% forecast levels based on Section 2.3.6 Equation 2.37 are presented in Table 5.9. Smaller values indicate superior accuracy. The RE-STW and RE-SkN models in general rank better than other competing models, generating smaller quantile loss function values.

TABLE 5.9: Quantile loss function values

Model	ASX 200	DAX	FTSE 100	Hang Seng	NASDAQ	SMI	S&P500	Mean Loss	Mean Rank
$\alpha = 2.5\%$									
GARCH-t	152.3	201.1	171.3	209.4	194.9	166.5	174.0	181.4	17.00
EGARCH-t	146.9	194.2	165.6	199.2	191.2	162.3	173.5	176.1	14.00
GJR-t	148.6	195.6	167.1	200.5	189.8	161.0	168.3	175.8	13.14
GARCH-HS-t	148.9	198.2	169.0	208.4	194.0	164.8	170.7	179.1	15.71
GARCH-EVT-t	149.0	420.6	239.4	302.7	354.7	197.1	199.7	266.2	19.00
CARE-SAV	184.5	225.5	187.2	240.5	228.2	219.5	221.2	215.2	19.00
RE-RV-SkN	141.3	189.1	**158.9**	192.5	185.6	204.1	168.5	177.1	8.43
RE-RR-SkN	141.0	189.0	159.7	197.7	182.7	783.2	161.3	259.2	8.43
RE-RK-SkN	145.1	260.3	160.5	192.9	182.9	167.0	169.6	182.6	12.43
RE-RV-RR-SkN	141.9	**188.9**	159.8	**192.2**	184.4	160.3	166.2	170.5	**5.86**
RE-RV-RK-SkN	**140.0**	**188.7**	**159.4**	194.3	225.5	**159.1**	164.5	175.9	**5.86**
RE-RR-RK-SkN	140.5	235.9	160.6	196.1	182.6	160.6	161.3	176.8	8.71
RE-RV-RR-RK-SkN	**139.9**	189.0	160.2	193.6	184.2	**160.3**	164.9	170.3	6.00
RE-RV-STW	142.3	190.9	159.8	**191.6**	179.3	161.1	159.6	**169.2**	**5.14**
RE-RR-STW	143.5	192.1	160.8	200.5	180.6	161.2	159.7	171.2	10.00
RE-RK-STW	157.4	190.7	160.1	195.5	**178.2**	171.8	161.4	173.6	9.43
RE-RV-RR-STW	144.4	191.3	160.4	197.4	180.1	162.4	159.4	170.8	8.71
RE-RV-RK-STW	144.8	190.9	160.0	192.9	**178.9**	161.4	**159.3**	**169.7**	**5.86**
RE-RR-RK-STW	143.5	192.1	161.0	198.0	179.6	161.1	159.5	170.7	9.00
RE-RV-RR-RK-STW	142.7	191.2	161.0	196.3	179.5	162.5	**159.4**	170.4	8.29
$\alpha = 1\%$									
GARCH-t	69.3	93.4	81.5	98.6	92.6	84.6	81.9	86.0	16.43
EGARCH-t	67.1	92.2	77.1	90.4	92.7	81.4	80.4	83.0	11.86
GJR-t	67.6	93.8	78.0	92.3	90.2	81.8	77.7	83.0	13.86
GARCH-HS-t	69.0	93.0	79.9	96.9	96.5	84.1	81.5	85.9	15.00
GARCH-EVT-t	72.9	169.5	96.7	121.7	143.6	85.0	84.2	110.5	19.14
CARE-SAV	84.5	112.7	89.1	100.8	107.6	104.7	99.3	99.8	18.86
RE-RV-SkN	64.2	89.1	**74.6**	**86.9**	89.1	101.2	76.2	83.0	7.71
RE-RR-SkN	64.0	89.3	75.6	88.9	86.5	391.5	**73.7**	124.2	6.86
RE-RK-SkN	67.7	129.7	75.1	89.1	87.0	82.1	78.3	87.0	10.86
RE-RV-RR-SkN	64.6	**89.0**	75.2	**87.7**	88.5	77.6	75.1	**79.7**	6.14
RE-RV-RK-SkN	63.8	**88.8**	**74.9**	89.6	114.3	77.2	74.1	83.2	5.86
RE-RR-RK-SkN	**63.7**	116.7	75.6	89.6	86.7	**77.0**	74.0	83.3	6.57
RE-RV-RR-RK-SkN	**63.6**	89.3	75.5	88.8	88.2	78.0	**73.9**	**79.6**	**5.14**
RE-RV-STW	67.2	93.1	75.1	88.4	85.8	77.6	74.8	80.3	**5.71**
RE-RR-STW	68.2	93.9	76.2	91.4	87.2	**76.9**	75.6	81.3	10.86
RE-RK-STW	76.8	93.2	76.1	91.9	**84.3**	84.3	75.9	83.2	12.00
RE-RV-RR-STW	69.1	93.4	75.9	89.5	87.1	78.0	75.0	81.1	10.14
RE-RV-RK-STW	67.0	92.9	75.3	90.2	**85.5**	78.0	74.5	80.5	6.43
RE-RR-RK-STW	67.2	94.1	76.8	90.6	87.0	77.3	75.0	81.1	10.14
RE-RV-RR-RK-STW	68.9	93.2	76.3	90.5	87.0	78.6	74.9	81.3	10.43

Note: For individual models, a green highlight indicates the favored model, bold indicates the 2nd ranked mode

At 2.5% forecast level, there are four different RE-SkN type models that have the lowest quantile loss function value for one market index, respectively. Meanwhile, there are only three RE-STW type models that have the lowest quantile loss function values. However, the RE-RV-STW is most favorable among the RE-STW and RE-SkN models, generating the lowest quantile loss function values for two market indices and also the lowest average quantile loss value.

Overall, the RE-SkN type models rank better than the RE-STW type models at 1% forecast level. In particular, the RE-RV-SkN and RE-RV-RR-RK-SkN models generates the lowest quantile loss function value for two market indices, respectively. Additionally, the RE-RV-RR-RK-SkN model ranks best in terms of the lowest average quantile loss value.

5.6.6 Joint loss function values

In addition to the quantile loss function, the joint evaluation of VaR and ES forecasts across models are conducted using the asymmetric Laplace (AL) log score in Section 2.3.7 Equation 2.40. Smaller values indicate better prediction accuracy.

Both the RE-STW and RE-SkN models tend to have favorable forecast accuracy. There are three types of RE-STW models and four RE-SkN type models with the lowest AL log score value for different market indices. This is the case for both the 2.5% and 1% forecast levels. Of all types of RE-STW and RE-SkN models, only the RE-SkN model using all three realized volatility measures produces the smallest AL log score values for two market indices at the 2.5% forecast level. Other models rank best only in one market index. At the 1% forecast level, no model has the lowest AL log score value more than once.

Overall, the RE-SkN type models rank slightly better than the RE-STW type models for both the 2.5% and 1% forecast levels. Nevertheless, one particular type of RE-STW model, that is the RE-RV-STW model, has a favorable performance, on average. This finding is consistent with the evaluation of the quantile loss function values, where the RE-RV-STW also has the best forecast accuracy at the 2.5% forecast level. In particular, the the RE-RV-STW has the lowest average AL log score value and best average AL log score rank across market indices. The model also has the best average AL log score rank at the 1% forecast level, although its average AL log score value is not the smallest.

TABLE 5.10: Joint evaluation for VaR and ES forecasts based on AL log score

Model	ASX 200	DAX	FTSE 100	Hang Seng	NASDAQ	SMI	S&P500	Mean Loss	Mean Rank
$\alpha = 2.5\%$									
GARCH-t	2.033	2.331	2.127	2.342	2.266	2.128	2.137	2.195	17.57
EGARCH-t	1.988	2.298	2.100	2.287	2.241	2.122	2.123	2.166	14.29
GJR-t	2.001	2.307	2.097	2.293	2.224	2.102	2.093	2.160	14.14
GARCH-HS-t	1.998	2.307	2.094	2.329	2.259	2.106	2.102	2.171	15.43
GARCH-EVT-t	2.008	3.031	2.455	2.733	2.870	2.297	2.269	2.523	19.29
CARE-SAV	2.054	2.316	2.105	2.360	2.252	2.192	2.164	2.206	17.86
RE-RV-SkN	1.938	2.271	**2.030**	2.285	2.163	2.616	2.027	2.190	10.14
RE-RR-SkN	1.936	2.265	2.035	2.273	**2.145**	20.644	2.004	4.758	7.86
RE-RK-SkN	1.959	2.982	2.048	**2.256**	2.152	2.087	2.049	2.219	11.43
RE-RV-RR-SkN	1.941	2.255	2.036	2.312	2.157	2.068	2.013	2.112	8.71
RE-RV-RK-SkN	**1.930**	2.251	2.035	2.311	2.503	**2.058**	2.028	2.159	8.29
RE-RR-RK-SkN	1.932	2.641	2.045	2.271	2.146	2.069	2.003	2.158	7.86
RE-RV-RR-RK-SkN	**1.928**	2.252	2.043	2.328	2.154	**2.055**	2.015	2.111	8.29
RE-RV-STW	1.947	**2.250**	**2.031**	**2.258**	2.146	2.071	1.996	**2.100**	**3.71**
RE-RR-STW	1.956	2.253	2.037	2.295	2.148	2.078	**1.993**	2.109	7.86
RE-RK-STW	2.051	2.253	2.041	2.276	**2.142**	2.160	2.015	2.134	9.86
RE-RV-RR-STW	1.965	**2.251**	2.038	2.284	2.148	2.084	1.994	2.109	7.29
RE-RV-RK-STW	1.960	2.251	2.032	2.266	2.146	2.078	**1.993**	**2.104**	**5.29**
RE-RR-RK-STW	1.955	2.254	2.040	2.288	2.146	2.077	1.994	2.108	7.57
RE-RV-RR-RK-STW	1.951	2.251	2.041	2.285	2.147	2.084	1.994	2.108	7.29
$\alpha = 1\%$									
GARCH-t	2.140	2.481	2.279	2.500	2.403	2.341	2.271	2.345	16.43
EGARCH-t	2.127	2.476	2.268	2.423	2.407	2.337	2.275	2.330	14.43
GJR-t	2.113	2.493	2.244	2.434	2.356	2.333	2.209	2.312	13.86
GARCH-HS-t	2.131	2.470	2.250	2.475	2.460	2.330	2.259	2.339	15.14
GARCH-EVT-t	2.188	3.024	2.449	2.722	2.863	2.346	2.288	2.554	19.00
CARE-SAV	2.185	2.492	2.276	2.473	2.441	2.461	2.320	2.379	17.29
RE-RV-SkN	2.057	2.448	2.184	2.446	2.321	2.782	2.156	2.342	10.86
RE-RR-SkN	**2.052**	2.438	2.196	**2.394**	**2.289**	21.343	2.136	4.978	7.43
RE-RK-SkN	2.105	3.097	2.208	**2.392**	2.303	2.283	2.189	2.368	10.71
RE-RV-RR-SkN	2.059	2.421	2.193	2.516	2.313	2.244	2.137	2.269	8.43
RE-RV-RK-SkN	2.055	**2.413**	2.189	2.496	2.662	**2.234**	2.157	2.315	8.29
RE-RR-RK-SkN	**2.050**	2.803	2.198	2.406	2.295	2.238	2.136	2.304	6.86
RE-RV-RR-RK-SkN	2.054	**2.418**	2.203	2.547	2.310	**2.226**	2.134	**2.270**	7.71
RE-RV-STW	2.098	2.428	**2.176**	2.398	2.294	2.248	**2.128**	2.253	**4.00**
RE-RR-STW	2.113	2.433	2.194	2.431	2.305	2.245	2.136	2.265	8.71
RE-RK-STW	2.245	2.431	2.200	2.423	**2.285**	2.386	2.160	2.304	11.00
RE-RV-RR-STW	2.129	2.428	2.192	2.417	2.306	2.263	2.130	**2.266**	8.14
RE-RV-RK-STW	2.105	2.426	**2.180**	2.410	2.294	2.262	**2.126**	2.258	**5.00**
RE-RR-RK-STW	2.098	2.434	2.202	2.426	2.305	2.252	2.135	2.265	8.57
RE-RV-RR-RK-STW	2.114	2.425	2.198	2.425	2.305	2.262	2.130	2.265	8.14

Note: For individual models, a green highlight indicates the favored model and bold indicates the 2nd ranked model.

5.6.7 Model confidence set

The joint loss function values based on the AL log score are further evaluated by using the model confidence set (MCS) (Hansen et al., 2011), which determines a set of statistically superior models. The MCS procedure involves a sequential testing procedure to eliminate

the worst-performing model at each step. This test is continued until the hypothesis of Equal Predictive Ability (EPA) is no longer rejected for all remaining models in a set of superior models.

TABLE 5.11: 75% and 90% model confidence set with R and SQ methods

Model	$\alpha = 2.5\%$				$\alpha = 1\%$			
	90% MCS		**75% MCS**		**90% MCS**		**75% MCS**	
	R	**SQ**	**R**	**SQ**	**R**	**SQ**	**R**	**SQ**
GARCH-t	1	1	1	1	3	5	2	2
EGARCH-t	3	5	2	1	4	5	4	3
GJR-t	2	4	1	1	5	6	4	3
GARCH-HS-t	2	2	1	1	3	4	3	2
GARCH-EVT-t	0	0	0	0	0	1	0	0
CARE-SAV	1	0	0	0	2	2	1	0
RE-RV-SkN	5	6	5	5	6	6	5	6
RE-RR-SkN	6	6	6	6	6	6	6	6
RE-RK-SkN	6	6	6	6	6	6	6	6
RE-RV-RR-SkN	6	**7**	6	6	**7**	**7**	6	**7**
RE-RV-RK-SkN	5	6	5	5	5	6	5	5
RE-RR-RK-SkN	5	6	5	5	6	6	5	6
RE-RV-RR-RK-SkN	6	**7**	6	6	**7**	**7**	6	**7**
RE-RV-STW	6	**7**	6	**7**	**7**	**7**	**7**	**7**
RE-RR-STW	4	**7**	4	6	5	6	5	5
RE-RK-STW	3	6	3	3	5	6	4	4
RE-RV-RR-STW	4	**7**	4	5	5	7	5	4
RE-RV-RK-STW	5	**7**	4	6	**7**	**7**	**7**	6
RE-RR-RK-STW	5	**7**	5	6	6	6	6	5
RE-RV-RR-RK-STW	5	**7**	5	6	5	6	5	5

Note: For individual models, a green highlight indicates the favored model.

The MCS tests are performed using both R and SQ methods at 90% and 75% confidence levels. The results of the MCS tests are presented in Table 5.11. The models that consistently enter the MCS for both MCS methods and across market indices are highligted in green. Overall, the proposed RE-STW type models are consistently included in the superior set of models for most market indices. In particular, all RE-STW models for all market indices are included in the set of superior models according to the 90% confidence level of the MCS test conducted on the AL log score at the 2.5% forecast level using the SQ method. The only exception is the RE-RK-STW model, which enters the superior set of models for six of seven market indices for this tail risk forecast level using the SQ method.

There are two best performing-models that belong to the RE-STW type models, that is the

RE-RV-STW model, followed by the RE-RV-RK-STW model. The RE-RV-STW model is especially excluded from the superior set of models for only two cases across all market indices, MCS test confidence levels, and MCS test methods. A majority of the RE-SkN type models also enter the set of superior models. The best two are the RE-SkN model that jointly employs the RVSS and the RRSS in the measurement equations and the RE-SkN model using all three realized volatility measures in the measurement equations. However, their forecast accuracy based on the MCS procedure is slightly lower than that of the RE-STW models, being excluded from the superior set of models in five cases. As for other competing models, they are excluded from the set of superior models in most cases.

5.7 Chapter Summary

Another extension of the realized EGARCH model is proposed in this chapter by allowing the return error distribution to follow a standardized two-sided Weibull distribution, called the RE-STW model. A Bayesian approach is proposed to estimate the RE-STW. The results of the simulation study suggest that the proposed Bayesian method produces less biased and more precise parameter estimates than does the Maximum Likelihood estimation. The adaptive MCMC procedure is found to produce MCMC iterates with an excellent convergence and chain mixing property.

An empirical study using seven market indices is performed by employing the subsampled realized variance (RVSS), subsampled realized range (RRSS), and realized kernel (RK), either individually or jointly. The prediction accuracy of the RE-STW model in the forecasting of VaR and ES at the 2.5% and 1% quantiles is compared to that of several competing models, especially the proposed realized EGARCH model using the standardized Skewed Student-t distribution for the return error (RE-SkN model) proposed in Chapter 3.

Overall, the RE-STW and RE-SkN models have better tail risk prediction accuracy than other models across market indices. The forecast performance of the RE-STW model is relatively the same as that of the RE-SkN models, satisfying most of the VaR backtests and the ES backtests. However, SkN-type models have a slightly better tail risk forecast accuracy than the RE-STW in forecasting the ES. The SkN-type models also produce relatively lower quantile loss function values and joint loss function values. The type of RE-SkN models that consistently have low rejections in VaR and ES backtests and produces low quantile loss function and joint loss function values in most market indices is the RE-SkN model that employs all three realized measures.

Nevertheless, one type of RE-STW model has quite consistent favorable forecasting performance, that is, the RE-STW model that only uses the RVSS. This is particularly the case when evaluating the VaR and ES joint loss function values using the AL log score. Contrary to the results of the VaR and ES backtests, the RE-STW type models are more favorable than the RE-SkN type models based on the Model Confidence Set procedure, being included in the set of superior models in most cases. However, the difference is only a small margin.

It could then be argued that using either a standardized two-sided Weibull distribution or a standardized skewed Student-t distribution could equally improve the tail forecast of financial asset return using the realized EGARCH model and other relevant parametric volatility models, for example, in the realized Threshold GARCH model (Chen & Watanabe, 2019a). It would be interesting to see the performance of the volatility models using these two distributions to forecast the tail risk during the recent COVID-19 pandemic. In addition, the use of the two-sided Weibull distribution in forecasting asset volatility in other highly volatile markets, such as derivative, commodity, and foreign exchange markets, is also highly recommended.

The realized EGARCH model can be extended by using a Gumbel distribution, which is also called the extreme value distribution. The Gumbel distribution is an asymmetric distribution, where the tail skews to the direction of the extreme values. It assigns a greater likelihood to more extreme events. Therefore, the use of this distribution may be beneficial, especially for study periods covering extreme events such as the Global Financial Crisis of 2008 and the recent COVID-19 pandemic. This distribution is particularly attractive because the location parameters (μ) and the scale parameter (σ) are easy to interpret.

Chapter 6

Conclusion and Future Research

6.1 Conclusion

Financial regulators and institutions have to navigate a rapidly changing market environment. Academics and practitioners are continually developing financial volatility models and methods for forecasting risks as part of daily risk management and to satisfy the requirements of the Basel Capital Accord. In the past two decades, financial crises have taken place with more frequency and on a global scale. Extreme left-tail events, which are difficult to predict, can have a disastrous impact on portfolio returns. Therefore, accurate tail risk forecasting is critical for prudent risk management in order to monitor risk exposure and prepare efficient capital allocation, which can prevent devastating losses while maximizing return on investment. This thesis proposes Bayesian parametric approaches to forecast financial tail risks, that is, the Value at Risk (VaR) and the Expected Shortfall (ES), by employing multiple high-frequency realized volatility measures in the realized EGARCH model. An adaptive Markov Chain Monte Carlo (MCMC) is proposed to address issues in parameter uncertainty and is proven to have an advantage over a frequentist maximum likelihood (ML) approach in simulation studies. A Least Absolute Shrinkage and Selection Operator (Lasso) approach is also adapted to facilitate the selection of realized volatility measures. Various tests are conducted to compare the performance of the proposed models with various well-known volatility models.

Chapter 1 provides a brief review of the types of volatility models, an introduction to the Realized EGARCH model, which employs multiple realized measures, and a short overview of the advantages of using Bayesian estimation for GARCH-type models. This chapter also discusses various adaptive MCMC algorithms, as well as MCMC convergence diagnostics.

Chapter 2 presents a literature review on the distribution of financial asset returns and provides an introduction to VaR and ES, as well as their formal definitions. This chapter also provides a review of the literature on various tests to evaluate tail risk prediction accuracy. These tests are used throughout this thesis to measure the relative forecast performance of various models proposed in comparison with other models.

Chapter 3 proposes an extension of the Realized EGARCH model by allowing the return equation errors to follow a standardized Student-t distribution and a standardized skewed

Student-*t* distribution. Three types of robust realized measures of volatility used in this thesis are briefly discussed, that is, subsampled realized variance (RVSS), subsampled realized range (RRSS), and realized kernel (RK). An adaptive MCMC approach is developed and used to estimate the proposed models. The superiority of the adaptive MCMC approach over the ML approach, in terms of unbiasedness and efficiency, is demonstrated in a simulation study. The proposed models are used in an empirical study using seven market indices. The RVSS, RRSS, and RK are used, either individually or jointly, in the measurement equations of the proposed models. The results of the empirical study show that the proposed extension of the realized EGARCH model employing the standardized Student-*t* distribution or the standardized skewed Student-*t* distribution have a more favorable tail risk forecasting performance compared to standard GARCH-type models, CARE models, realized GARCH models employing the Student-*t* distribution for the observation equation errors, and the original realized EGARCH model employing Gaussian distribution estimated using the proposed adaptive MCMC approach. In regard to the choice of realized measures, the RRSS (either employed individually or jointly with RVSS and/or RK) is found to be crucial in improving the accuracy of the tail risk forecasts.

Chapter 4 builds upon the proposed model in Chapter 3, that is realized EGARCH model incorporating a standardized skewed Student-*t* distribution, and adapts a Lasso approach to perform the selection of the realized volatility measures in the model. The proposed Lasso approach considers both the frequentist cross Validation (CV) and Bayesian Lasso (BLasso) techniques in choosing the tuning parameter. This study finds that the proposed Lasso approach on the realized EGARCH model could help in selecting informative realized volatility measures. In addition to RVSS, RRSS, and RK, the empirical study in this chapter also includes another realized measure, that is the Range variable, which is assumed to have a lower future volatility forecast performance than the RVSS and RRSS. The BLasso approach is found to be consistent in producing smaller values of tuning parameters than the CV approach, both in simulation and empirical studies across market indices. Nevertheless, the results of both confirm similar conclusions that the RK and Range variables are less important than RRSS and RVSS. The empirical study also suggests that the BLasso approach has a superior performance than the CV approach, producing lower quantile loss function and joint loss function values.

Chapter 5 further proposes another extreme-value distribution called a standardized two-sided Weibull distribution for modeling the return equation errors in the realized EGARCH model. The adaptive MCMC is also developed to estimate the proposed model and is shown to have less biased and more precise parameter estimates than the Maximum Likelihood estimation in a simulation study. An empirical study using seven market indices is performed by employing the RVSS, RRSS, and RK, either individually or jointly. The proposed model is used to forecast tail risks, including during the Global Financial Crisis of 2008. The forecast performance of the proposed model is compared with that of the extension of the realized EGARCH model proposed in Chapter 3, several GARCH-type models, and the CARE model. The results of the empirical study suggest that both types of distributions, the

standardized two-sided Weibull distribution and the standardized skewed Student-*t* distribution, could equally improve tail risk prediction accuracy. The realized EGARCH model employing these two distributions to model the return distribution have a superior forecast accuracy compared to other competing models.

Overall, the use of the standardized two-sided Weibull distribution and the standardized skewed Student-*t* distribution to model financial asset return volatility is highly encouraged. These distributions can capture various specific elements of highly volatile financial time series data, such as extreme volatility, volatility clustering, skewness, and fat tails. Using multiple robust realized volatility measures is also recommended, especially the RVSS and RRSS. The Bayesian method has continued to perform very well, either applied to simulated data or historical data, and therefore is highly recommended.

6.2 Possible Future Research

This research offers several potential future research avenues that could extend this study. The realized EGARCH model could be extended by considering two other extreme and asymmetric distributions, that is, the asymmetric Laplace distribution and the Gumbel distribution. The use of these distributions has the potential to improve tail forecast accuracy, especially when the study period covers extreme events. Further, while the standardized skewed Student-*t* has gained more popularity in modeling financial asset return, the use of the standardized two-sided Weibull distribution is limited. Future research may use the standardized two-sided Weibull distribution to forecast volatility in derivative, commodity, and foreign exchange markets.

Another interesting future research avenue is to develop some semi-parametric models, including for Realized EGARCH model, using the proposed Bayesian method used in this thesis. Semi-parametric models offer an a flexible and accurate representation of data and are useful for capturing non-linear relationships and addressing the limitations of fully parametric models that may not adequately capture complex patterns in tail behavior.

Semi-parametric models combine parametric and non-parametric approaches to provide a flexible and accurate representation of data, particularly in extreme or tail regions. These models are useful for capturing non-linear relationships and addressing the limitations of fully parametric models that may not adequately capture complex patterns in tail behavior.

The empirical study of the Lasso approach to select realized volatility measures in Chapter 4 only explores a small, yet robust, subset of available realized measures. Future research could consider including and testing a larger subset of realized volatility measures and including less robust measures such as simple realized variance, bi-power variation, and median realized variance. Further, an adaptive Lasso approach, which assigns different parameter weights is a very promising and rapidly growing research field on regularization that can be applied in the selection of realized measures. Finally, it would also be interesting

to assess the performance of the models and methods proposed in this thesis to forecast tail risk during the recent COVID-19 pandemic.

Appendix A

Chapter 3 Appendix

A.1 MCMC Burn In and Post Burn In Plots

FIGURE A.1: MCMC Burn In and Post Burn In Plots for μ

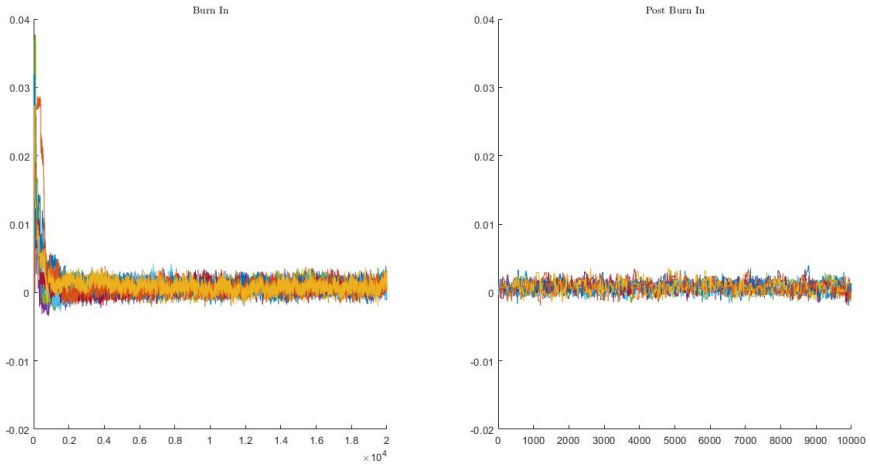


FIGURE A.2: MCMC Burn In and Post Burn In Plots for ω

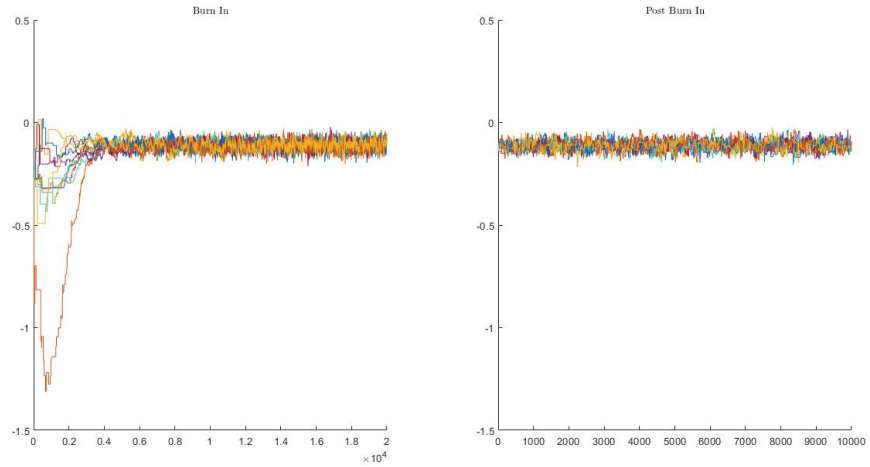


FIGURE A.3: MCMC Burn In and Post Burn In Plots for β

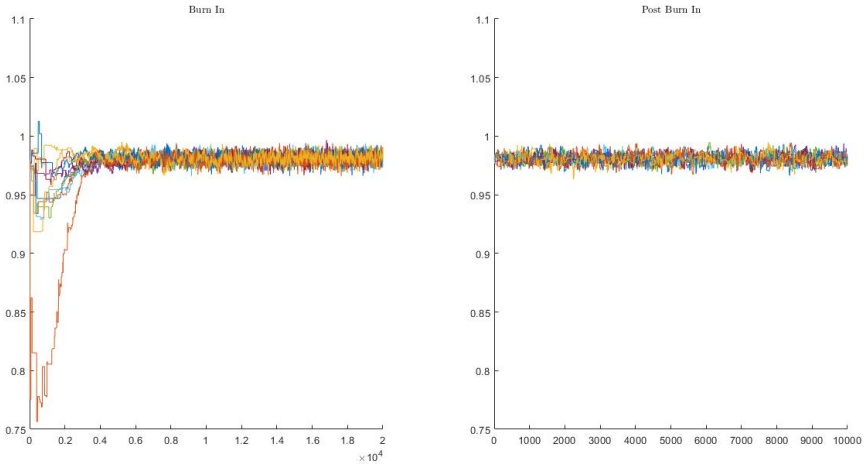


FIGURE A.4: MCMC Burn In and Post Burn In Plots for γ

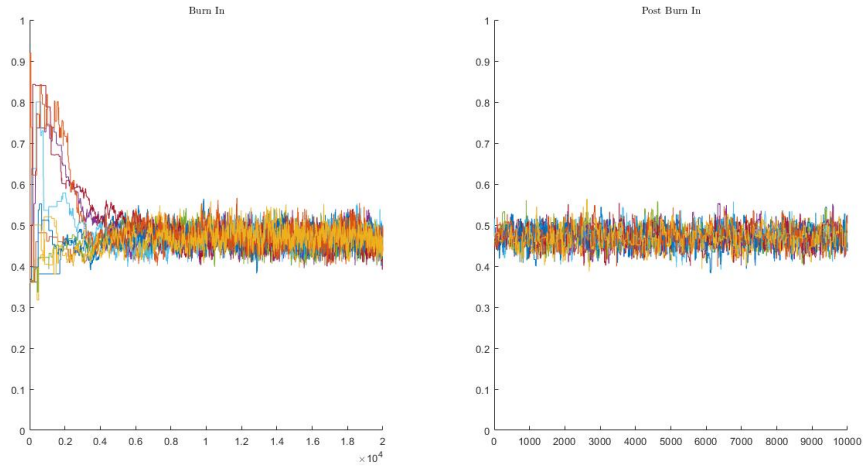


FIGURE A.5: MCMC Burn In and Post Burn In Plots for τ_1

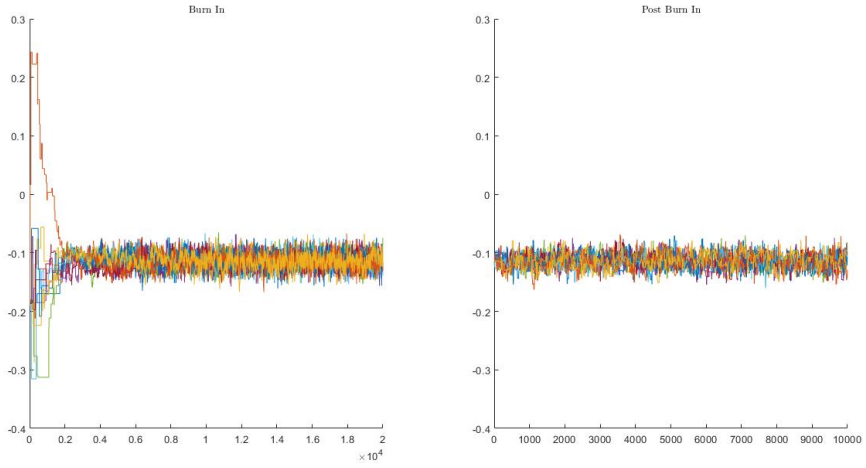


FIGURE A.6: MCMC Burn In and Post Burn In Plots for τ_2

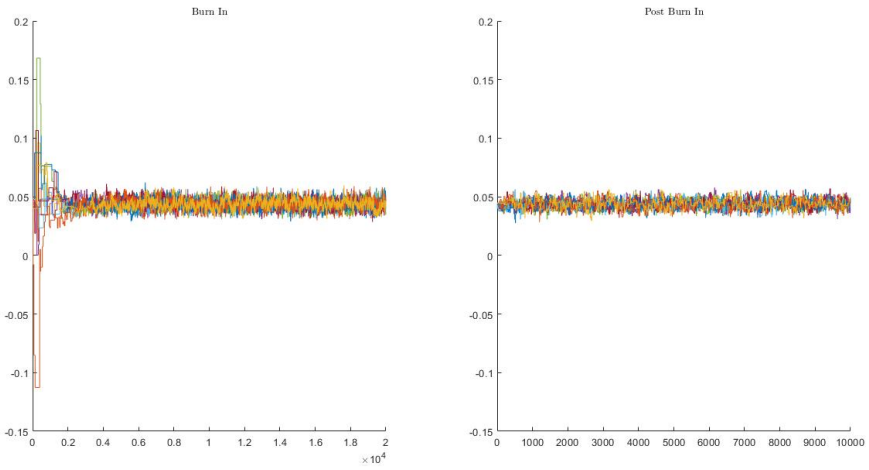


FIGURE A.7: MCMC Burn In and Post Burn In Plots for ξ

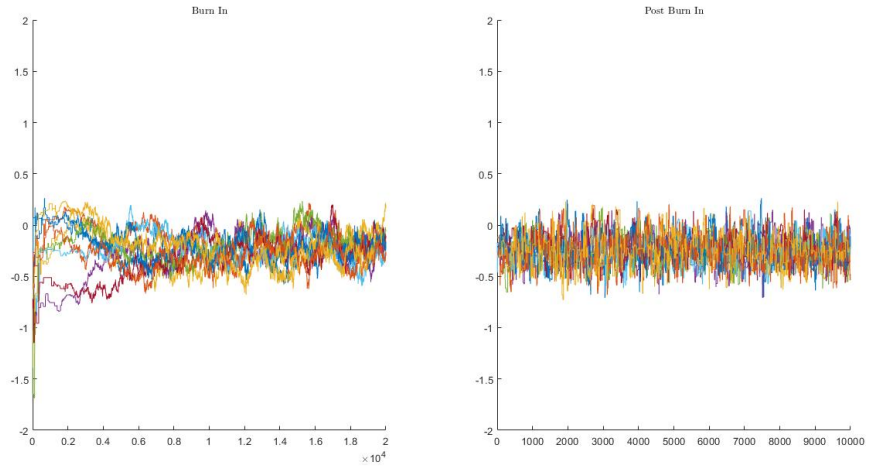


FIGURE A.8: MCMC Burn In and Post Burn In Plots for φ

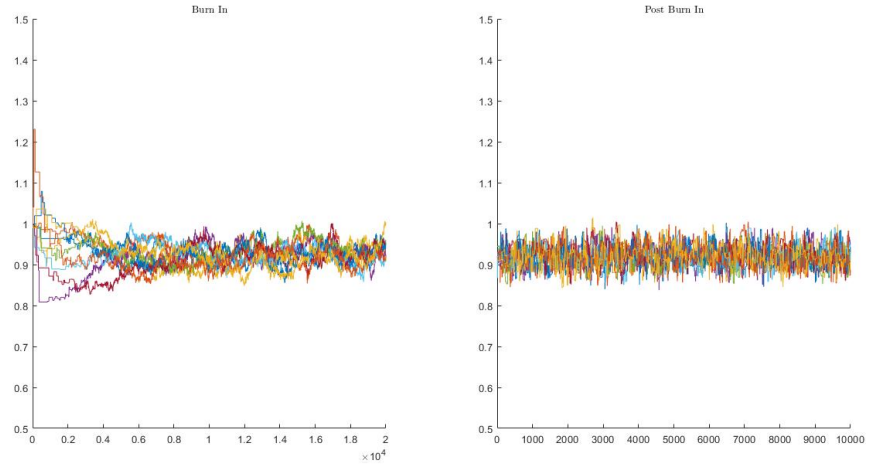


FIGURE A.9: MCMC Burn In and Post Burn In Plots for δ_1

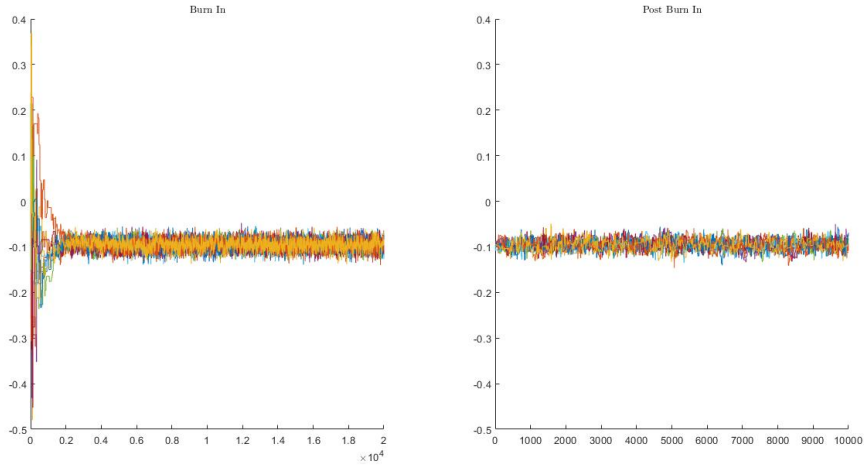


FIGURE A.10: MCMC Burn In and Post Burn In Plots for δ_2

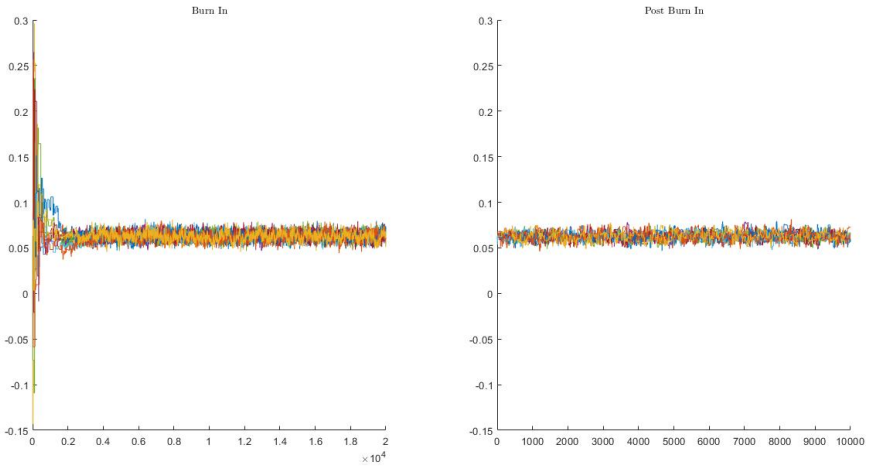


FIGURE A.11: MCMC Burn In and Post Burn In Plots for σ_u^2

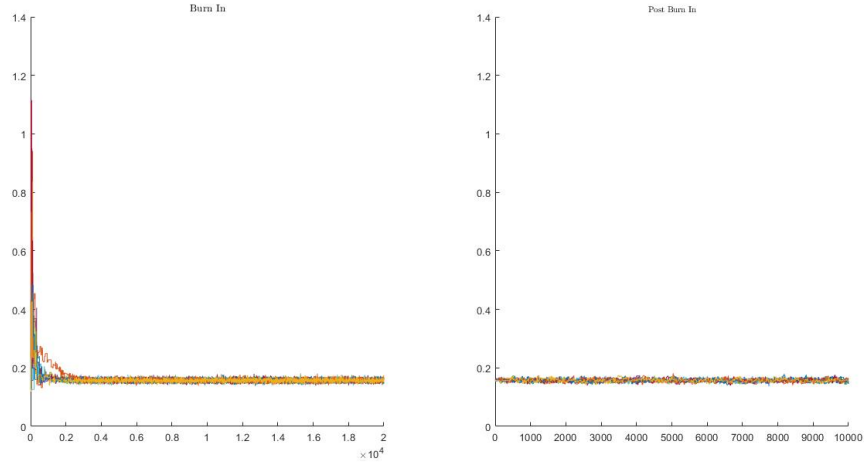


FIGURE A.12: MCMC Burn In and Post Burn In Plots for ν

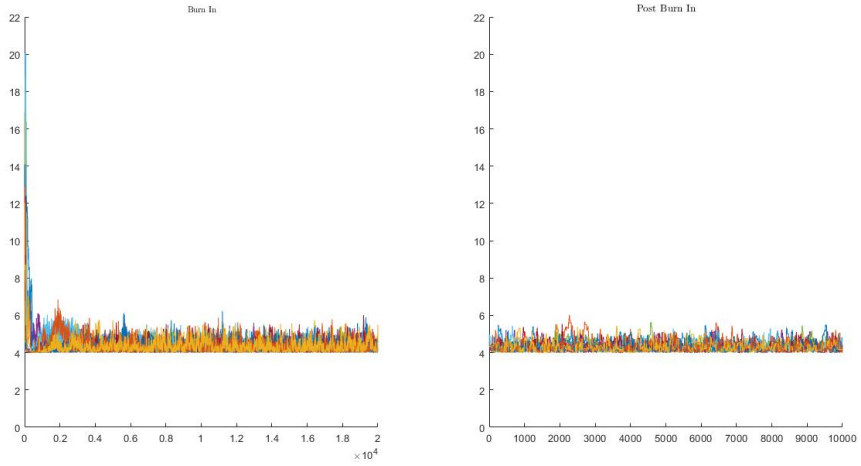
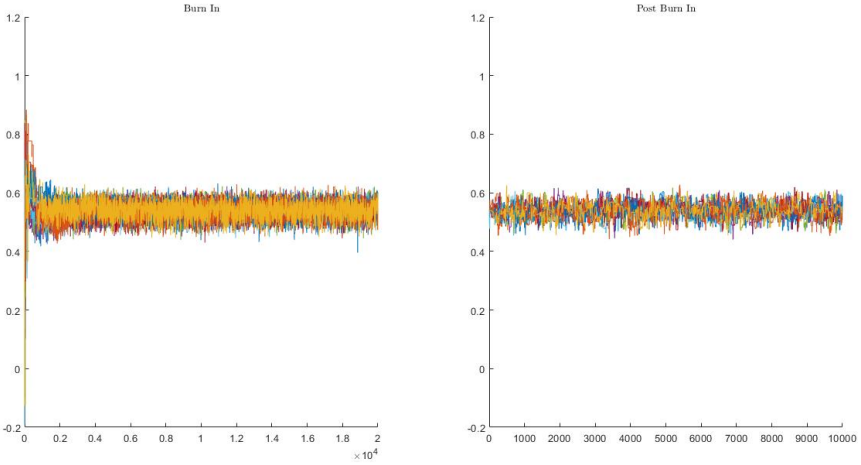


FIGURE A.13: MCMC Burn In and Post Burn In Plots for λ



A.2 Posterior Mean of Realized EGARCH Models

TABLE A.1: Estimated parameter mean for all forecasting steps for each parameter: ASX 200-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.000	-0.182	0.981	0.152			-0.109	0.036	-1.568			0.925		
RE-RV-SkN	0.000	-0.180	0.981	0.148			-0.108	0.035	-1.552			0.927		
RE-RV-tN	0.000	-0.170	0.982	0.150			-0.107	0.036	-1.611			0.921		
RE-RR-NN	0.000	-0.172	0.982		0.174		-0.104	0.032		-2.215			0.907	
RE-RR-SkN	0.000	-0.173	0.982		0.169		-0.104	0.032		-2.190			0.910	
RE-RR-tN	0.000	-0.162	0.983		0.172		-0.103	0.031		-2.270			0.901	
RE-RK-NN	0.000	-0.230	0.976			0.161	-0.090	0.018			1.812			1.221
RE-RK-SkN	0.000	-0.217	0.977			0.156	-0.088	0.016			1.904			1.232
RE-RK-tN	0.000	-0.210	0.978			0.159	-0.087	0.016			1.888			1.231
RE-RV-RR-NN	0.000	-0.200	0.979	-0.099	0.261		-0.101	0.038	-1.074	-1.885		0.978	0.942	
RE-RV-RR-SkN	0.000	-0.182	0.981	-0.100	0.263		-0.101	0.038	-1.275	-2.012		0.957	0.929	
RE-RV-RR-tN	0.000	-0.177	0.981	-0.099	0.263		-0.100	0.038	-1.262	-1.999		0.958	0.930	
RE-RV-RK-NN	0.000	-0.229	0.976	0.135		0.100	-0.110	0.023	-1.414		-0.029	0.943		1.016
RE-RV-RK-SkN	0.000	-0.224	0.977	0.132		0.098	-0.109	0.022	-1.429		-0.067	0.941		1.012
RE-RV-RK-tN	0.000	-0.217	0.978	0.133		0.100	-0.108	0.022	-1.421		-0.063	0.943		1.013
RE-RR-RK-NN	0.000	-0.210	0.978		0.157	0.080	-0.105	0.022		-1.846	-0.097		0.946	1.008
RE-RR-RK-SkN	0.000	-0.206	0.978		0.153	0.079	-0.105	0.022		-1.889	-0.143		0.942	1.003
RE-RR-RK-tN	0.000	-0.203	0.979		0.158	0.081	-0.104	0.022		-1.952	-0.202		0.936	0.997
RE-RV-RR-RK-NN	0.000	-0.207	0.978	-0.112	0.265	0.039	-0.098	0.033	-0.634	-1.380	-0.319	1.024	0.995	0.984
RE-RV-RR-RK-SkN	0.000	-0.206	0.978	-0.105	0.256	0.038	-0.099	0.033	-0.633	-1.372	-0.332	1.024	0.996	0.983
RE-RV-RR-RK-tN	0.000	-0.194	0.980	-0.107	0.257	0.038	-0.097	0.032	-0.632	-1.352	-0.261	1.024	0.998	0.990
Average	**0.000**	**-0.198**	**0.979**	**0.019**	**0.212**	**0.094**	**-0.102**	**0.028**	**-1.209**	**-1.864**	**0.341**	**0.964**	**0.944**	**1.057**

TABLE A.2: Estimated parameter mean for all forecasting steps for each parameter: ASX 200-Part 2

Model	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}	δ_{23}	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RV-NN	-0.090			0.122			0.326					
RE-RV-SkN	-0.092			0.121			0.326				71.505	-0.166
RE-RV-tN	-0.088			0.121			0.326			29.293		
RE-RR-NN		-0.085			0.104			0.246				
RE-RR-SkN		-0.088			0.105			0.247			68.856	-0.163
RE-RR-tN		-0.082			0.104			0.247		28.327		
RE-RK-NN			-0.152			0.002			0.483			
RE-RK-SkN			-0.151			0.004			0.482		14.145	-0.114
RE-RK-tN			-0.152			0.003			0.482	11.058		
RE-RV-RR-NN	-0.092	-0.089		0.121	0.104		0.323	0.244				
RE-RV-RR-SkN	-0.094	-0.088		0.122	0.105		0.323	0.245			63.246	-0.163
RE-RV-RR-tN	-0.089	-0.084		0.122	0.105		0.323	0.245		29.692		
RE-RV-RK-NN	-0.080		-0.164	0.136		-0.031	0.385		0.618			
RE-RV-RK-SkN	-0.079		-0.163	0.135		-0.030	0.384		0.618		51.428	-0.147
RE-RV-RK-tN	-0.076		-0.164	0.136		-0.030	0.385		0.616	27.158		
RE-RR-RK-NN		-0.077	-0.164		0.115	-0.028		0.284	0.660			
RE-RR-RK-SkN		-0.077	-0.163		0.114	-0.028		0.283	0.663		57.027	-0.153
RE-RR-RK-tN		-0.075	-0.165		0.114	-0.029		0.284	0.660	26.984		
RE-RV-RR-RK-NN	-0.087	-0.082	-0.165	0.122	0.106	-0.011	0.334	0.254	0.761			
RE-RV-RR-RK-SkN	-0.090	-0.084	-0.164	0.123	0.107	-0.012	0.333	0.254	0.763		73.325	-0.159
RE-RV-RR-RK-tN	-0.084	-0.080	-0.165	0.123	0.106	-0.012	0.333	0.254	0.761	31.838		
Average	**-0.087**	**-0.083**	**-0.161**	**0.125**	**0.107**	**-0.017**	**0.342**	**0.257**	**0.631**	**26.336**	**57.076**	**-0.152**

TABLE A.3: Estimated parameter mean for all forecasting steps for each parameter: DAX-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.000	-0.210	0.976	0.222			-0.128	0.038	-0.031			1.051		
RE-RV-SkN	0.000	-0.203	0.977	0.224			-0.129	0.038	-0.104			1.042		
RE-RV-TN	0.000	-0.193	0.978	0.225			-0.127	0.038	-0.114			1.042		
RE-RR-NN	0.000	-0.215	0.976		0.241		-0.125	0.039	-0.063			1.056		
RE-RR-SkN	0.000	-0.210	0.976		0.245		-0.126	0.040	-0.156			1.046		
RE-RR-TN	0.000	-0.205	0.977		0.246		-0.124	0.040	-0.156			1.046		
RE-RK-NN	0.000	-0.177	0.980			0.139	-0.126	0.033			-0.024			1.058
RE-RK-SkN	0.000	-0.167	0.981			0.140	-0.127	0.033			-0.098			1.049
RE-RK-TN	0.000	-0.158	0.983			0.140	-0.125	0.033			-0.095			1.050
RE-RV-RR-NN	0.000	-0.219	0.975	0.107	0.128		-0.123	0.042	0.021	-0.061		1.057	1.056	
RE-RV-RR-SkN	0.000	-0.209	0.976	0.111	0.127		-0.123	0.042	-0.078	-0.159		1.045	1.045	
RE-RV-RR-TN	0.000	-0.204	0.977	0.110	0.129		-0.122	0.042	-0.057	-0.143		1.049	1.047	
RE-RV-RK-NN	0.000	-0.196	0.978	0.161		0.042	-0.125	0.038	0.004	-0.003		1.055	1.060	
RE-RV-RK-SkN	0.000	-0.190	0.979	0.163		0.042	-0.126	0.038	-0.103	-0.048		1.042	1.054	
RE-RV-RK-TN	0.001	-0.182	0.980	0.163		0.042	-0.124	0.038	-0.087	-0.047		1.045	1.055	
RE-RR-RK-NN	0.000	-0.200	0.977		0.171	0.048	-0.124	0.039	-0.033	-0.002		1.059	1.060	
RE-RR-RK-SkN	0.000	-0.195	0.978		0.174	0.048	-0.125	0.039	-0.208	-0.084		1.039	1.050	
RE-RR-RK-TN	0.001	-0.188	0.979		0.171	0.049	-0.123	0.039	-0.160	-0.041		1.045	1.055	
RE-RV-RR-RK-NN	0.000	-0.203	0.977	0.083	0.104	0.039	-0.123	0.041	-0.075	-0.170	-0.043	1.046	1.044	1.055
RE-RV-RR-RK-SkN	0.000	-0.200	0.978	0.082	0.103	0.039	-0.123	0.041	-0.090	-0.191	-0.030	1.044	1.041	1.056
RE-RV-RR-RK-TN	0.001	-0.193	0.979	0.081	0.104	0.040	-0.121	0.041	-0.084	-0.159	-0.027	1.045	1.045	1.057
Average	**0.000**	**-0.192**	**0.978**	**0.118**	**0.135**	**0.067**	**-0.124**	**0.039**	**-0.079**	**-0.092**	**-0.053**	**1.048**	**1.051**	**1.054**

TABLE A.4: Estimated parameter mean for all forecasting steps for each parameter: DAX-Part 2

Model	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}	δ_{23}	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RK-NN	-0.165			0.074			0.187					
RE-RK-SkN	-0.165			0.074			0.187				17.261	-0.138
RE-RK-tN	-0.161			0.074			0.187			15.069		
RE-RR-NN	-0.160			0.071			0.163					
RE-RR-SkN	-0.160			0.071			0.163				19.352	-0.135
RE-RR-tN	-0.156			0.071			0.163			16.811		
RE-RV-NN	-0.129			0.127			0.315					
RE-RV-SkN	-0.128			0.127			0.315				17.027	-0.137
RE-RV-tN	-0.123			0.126			0.315			14.211		
RE-RR-RK-NN	-0.164	-0.159		0.077	0.070		0.190	0.166				
RE-RR-RK-SkN	-0.163	-0.159		0.076	0.070		0.190	0.166			18.606	-0.135
RE-RR-RK-tN	-0.160	-0.155		0.077	0.070		0.190	0.165		15.698		
RE-RV-RK-NN	-0.164	-0.127		0.074	0.128		0.187	0.314				
RE-RV-RK-SkN	-0.164	-0.124		0.073	0.127		0.188	0.314			17.947	-0.138
RE-RV-RK-tN	-0.160	-0.119		0.074	0.127		0.187	0.314		14.648		
RE-RV-RR-NN	-0.159	-0.126		0.070	0.133		0.164	0.317				
RE-RV-RR-SkN	-0.159	-0.126		0.070	0.133		0.164	0.318			19.050	-0.133
RE-RV-RR-tN	-0.155	-0.119		0.069	0.133		0.164	0.317		15.841		
RE-RV-RR-RK-NN	-0.162	-0.158	-0.123	0.075	0.070	0.131	0.189	0.166	0.316			
RE-RV-RR-RK-SkN	-0.162	-0.157	-0.123	0.076	0.069	0.130	0.189	0.166	0.315		17.994	-0.133
RE-RV-RR-RK-tN	-0.159	-0.154	-0.118	0.075	0.070	0.131	0.189	0.166	0.315	15.719		
Average	**-0.154**	**-0.140**	**-0.121**	**0.084**	**0.100**	**0.131**	**0.209**	**0.241**	**0.315**	**15.224**	**18.125**	**-0.135**

TABLE A.5: Estimated parameter mean for all forecasting steps for each parameter:FTSE 100-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.000	-0.218	0.977	0.237			-0.121	0.042	-0.111			1.049		
RE-RV-SkN	0.000	-0.216	0.977	0.238			-0.121	0.042	-0.131			1.047		
RE-RV-tN	0.000	-0.214	0.977	0.237			-0.119	0.042	-0.122			1.047		
RE-RR-NN	0.000	-0.235	0.975		0.275		-0.123	0.043		-0.583			1.017	
RE-RR-SkN	0.000	-0.233	0.975		0.276		-0.124	0.043		-0.632			1.012	
RE-RR-tN	0.000	-0.235	0.975		0.274		-0.122	0.043		-0.556			1.020	
RE-RK-NN	0.000	-0.205	0.978			0.140	-0.130	0.043			-0.051			1.029
RE-RK-SkN	0.000	-0.203	0.978			0.139	-0.131	0.043			-0.039			1.031
RE-RK-tN	0.000	-0.202	0.978			0.140	-0.130	0.043			-0.063			1.028
RE-RV-RR-NN	0.000	-0.242	0.974	0.133	0.122		-0.117	0.043	0.035	-0.394		1.065	1.038	
RE-RV-RR-SkN	0.000	-0.241	0.974	0.132	0.123		-0.118	0.044	0.009	-0.462		1.063	1.031	
RE-RV-RR-tN	0.000	-0.238	0.974	0.131	0.125		-0.118	0.044	0.009	-0.472		1.063	1.030	
RE-RV-RK-NN	0.000	-0.220	0.976	0.210		0.010	-0.128	0.041	-0.071		0.012	1.053		1.035
RE-RV-RK-SkN	0.000	-0.224	0.976	0.210		0.010	-0.128	0.041	-0.104		0.007	1.050		1.035
RE-RV-RK-tN	0.000	-0.222	0.976	0.211		0.009	-0.128	0.041	-0.103		-0.001	1.049		1.034
RE-RR-RK-NN	0.000	-0.238	0.974		0.250	0.000	-0.129	0.043		-0.536	0.056		1.022	1.040
RE-RR-RK-SkN	0.000	-0.239	0.974		0.252	-0.001	-0.130	0.043		-0.623	0.009		1.013	1.036
RE-RR-RK-tN	0.000	-0.238	0.974		0.251	0.001	-0.130	0.043		-0.599	0.015		1.016	1.037
RE-RV-RR-RK-NN	0.000	-0.238	0.974	0.130	0.120	-0.006	-0.124	0.043	-0.072	-0.489	0.068	1.053	1.027	1.042
RE-RV-RR-RK-SkN	0.000	-0.236	0.974	0.128	0.122	-0.006	-0.125	0.043	-0.073	-0.522	0.054	1.054	1.024	1.041
RE-RV-RR-RK-tN	0.000	-0.239	0.974	0.129	0.121	-0.006	-0.124	0.043	-0.063	-0.492	0.070	1.055	1.028	1.043
Average	**0.000**	**-0.227**	**0.975**	**0.177**	**0.193**	**0.036**	**-0.125**	**0.043**	**-0.066**	**-0.530**	**0.011**	**1.054**	**1.023**	**1.036**

TABLE A.6: Estimated parameter mean for all forecasting steps for each parameter: FTSE 100-Part 2

Model	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}	δ_{23}	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RV-NN	-0.141			0.073			0.157					
RE-RV-SkN	-0.141			0.073			0.157				30.018	-0.152
RE-RV-tN	-0.140			0.073			0.156			22.160		
RE-RR-NN		-0.122			0.072			0.130				
RE-RR-SkN		-0.123			0.072			0.130			26.894	-0.152
RE-RR-tN		-0.122			0.072			0.130		20.857		
RE-RK-NN			-0.073			0.185			0.295			
RE-RK-SkN			-0.074			0.185			0.295		23.444	-0.149
RE-RK-tN			-0.072			0.186			0.295	18.934		
RE-RV-RR-NN	-0.141	-0.123		0.074	0.072		0.158	0.131				
RE-RV-RR-SkN	-0.141	-0.123		0.074	0.072		0.158	0.131			29.153	-0.151
RE-RV-RR-tN	-0.141	-0.124		0.074	0.072		0.158	0.131		21.086		
RE-RV-RK-NN	-0.141		-0.074	0.073		0.185	0.157		0.290			
RE-RV-RK-SkN	-0.141		-0.073	0.073		0.186	0.158		0.289		29.083	-0.154
RE-RV-RK-tN	-0.140		-0.073	0.072		0.184	0.157		0.290	22.440		
RE-RR-RK-NN		-0.123	-0.073		0.071	0.186		0.130	0.290			
RE-RR-RK-SkN		-0.123	-0.072		0.072	0.188		0.131	0.290		28.487	-0.154
RE-RR-RK-tN		-0.123	-0.072		0.072	0.188		0.131	0.290	20.640		
RE-RV-RR-RK-NN	-0.141	-0.122	-0.072	0.073	0.071	0.186	0.159	0.131	0.290			
RE-RV-RR-RK-SkN	-0.142	-0.123	-0.072	0.074	0.072	0.188	0.158	0.131	0.290		31.207	-0.153
RE-RV-RR-RK-tN	-0.141	-0.123	-0.072	0.074	0.072	0.188	0.158	0.131	0.291	21.476		
Average	**-0.141**	**-0.123**	**-0.073**	**0.073**	**0.072**	**0.186**	**0.158**	**0.131**	**0.291**	**21.085**	**28.327**	**-0.152**

TABLE A.7: Estimated parameter mean for all forecasting steps for each parameter: Hang Seng-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.000	-0.191	0.978	0.230			-0.050	0.048	-1.322			0.958		
RE-RV-SkN	0.000	-0.174	0.980	0.244			-0.051	0.050	-1.603			0.927		
RE-RV-tN	0.000	-0.173	0.980	0.245			-0.050	0.050	-1.634			0.923		
RE-RR-NN	0.000	-0.191	0.978		0.276		-0.051	0.039		-1.799			0.928	
RE-RR-SkN	0.000	-0.182	0.979		0.288		-0.051	0.040		-2.017			0.905	
RE-RR-tN	0.000	-0.177	0.980		0.287		-0.049	0.040		-2.011			0.905	
RE-RK-NN	0.000	-0.132	0.985			0.111	-0.050	0.049			-1.406			0.956
RE-RK-SkN	0.000	-0.119	0.986			0.119	-0.049	0.050			-1.640			0.930
RE-RK-tN	0.000	-0.115	0.987			0.119	-0.048	0.050			-1.643			0.929
RE-RV-RR-NN	0.000	-0.239	0.973	0.054	0.188		-0.044	0.047	-0.445	-1.324		1.062	0.981	
RE-RV-RR-SkN	0.000	-0.227	0.974	0.054	0.196		-0.045	0.048	-0.674	-1.536		1.037	0.958	
RE-RV-RR-tN	0.000	-0.227	0.974	0.053	0.197		-0.043	0.048	-0.634	-1.479		1.041	0.964	
RE-RV-RK-NN	0.000	-0.203	0.977	0.153		0.042	-0.046	0.052	-1.078		-0.790	0.985		1.027
RE-RV-RK-SkN	0.000	-0.184	0.979	0.161		0.044	-0.047	0.054	-1.438		-1.137	0.946		0.989
RE-RV-RK-tN	0.000	-0.180	0.979	0.163		0.044	-0.045	0.054	-1.451		-1.145	0.944		0.987
RE-RR-RK-NN	0.000	-0.217	0.975		0.197	0.028	-0.047	0.046		-1.461	-0.676		0.965	1.043
RE-RR-RK-SkN	0.000	-0.205	0.976		0.207	0.028	-0.047	0.047		-1.733	-0.942		0.935	1.015
RE-RR-RK-tN	0.000	-0.205	0.976		0.204	0.028	-0.046	0.047		-1.630	-0.836		0.947	1.026
RE-RV-RR-RK-NN	0.000	-0.243	0.972	0.039	0.148	0.030	-0.042	0.048	-0.323	-1.218	-0.315	1.075	0.992	1.084
RE-RV-RR-RK-SkN	0.000	-0.245	0.972	0.037	0.150	0.031	-0.042	0.049	-0.323	-1.222	-0.317	1.076	0.993	1.085
RE-RV-RR-RK-tN	0.000	-0.242	0.972	0.035	0.151	0.032	-0.041	0.049	-0.349	-1.214	-0.310	1.073	0.994	1.086
Average	**0.000**	**-0.194**	**0.978**	**0.122**	**0.208**	**0.055**	**-0.047**	**0.048**	**-0.939**	**-1.554**	**-0.930**	**1.004**	**0.956**	**1.013**

TABLE A.8: Estimated parameter mean for all forecasting steps for each parameter: Hang Seng-Part 2

Model	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}	δ_{23}	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RV-NN	-0.105			0.055			0.199					
RE-RV-SkN	-0.106			0.056			0.199				9.052	-0.059
RE-RV-tN	-0.105			0.056			0.200			9.035		
RE-RR-NN		-0.090			0.044			0.144				
RE-RR-SkN		-0.091			0.045			0.145			7.798	-0.055
RE-RR-tN		-0.089			0.045			0.144		7.687		
RE-RK-NN			-0.100			0.105			0.448			
RE-RK-SkN			-0.100			0.106			0.448		9.692	-0.061
RE-RK-tN			-0.098			0.105			0.449	9.558		
RE-RV-RR-NN	-0.107	-0.091		0.063	0.041		0.207	0.151				
RE-RV-RR-SkN	-0.106	-0.092		0.064	0.042		0.208	0.150			8.164	-0.063
RE-RV-RR-tN	-0.105	-0.091		0.064	0.041		0.208	0.150		8.190		
RE-RV-RK-NN	-0.106		-0.100	0.054		0.107	0.202		0.458			
RE-RV-RK-SkN	-0.107		-0.101	0.055		0.108	0.202		0.459		9.343	-0.062
RE-RV-RK-tN	-0.105		-0.098	0.055		0.108	0.202		0.459	9.261		
RE-RR-RK-NN		-0.091	-0.096		0.040	0.116		0.151	0.504			
RE-RR-RK-SkN		-0.092	-0.099		0.041	0.119		0.151	0.505		8.425	-0.062
RE-RR-RK-tN		-0.090	-0.095		0.041	0.118		0.151	0.504	8.215		
RE-RV-RR-RK-NN	-0.107	-0.091	-0.099	0.059	0.039	0.113	0.201	0.159	0.489			
RE-RV-RR-RK-SkN	-0.107	-0.092	-0.097	0.060	0.039	0.115	0.201	0.159	0.490		8.569	-0.064
RE-RV-RR-RK-tN	-0.106	-0.091	-0.098	0.061	0.039	0.115	0.202	0.159	0.488	8.554		
Average	**-0.106**	**-0.091**	**-0.098**	**0.058**	**0.041**	**0.111**	**0.203**	**0.151**	**0.475**	**8.643**	**8.720**	**-0.061**

TABLE A.9: Estimated parameter mean for all forecasting steps for each parameter: NASDAQ-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.000	-0.344	0.961	0.326			-0.151	0.043	-1.183			0.957		
RE-RV-SkN	0.000	-0.345	0.961	0.341			-0.157	0.044	-1.404			0.932		
RE-RV-tN	0.000	-0.334	0.963	0.337			-0.153	0.044	-1.305			0.943		
RE-RR-NN	0.000	-0.365	0.959		0.354		-0.150	0.044		-1.337			0.971	
RE-RR-SkN	0.000	-0.376	0.958		0.373		-0.157	0.045		-1.549			0.948	
RE-RR-tN	0.000	-0.361	0.960		0.368		-0.152	0.045		-1.454			0.958	
RE-RK-NN	0.000	-0.235	0.974			0.181	-0.145	0.030			-1.335			0.929
RE-RK-SkN	0.000	-0.231	0.974			0.185	-0.150	0.031			-1.518			0.908
RE-RK-tN	0.001	-0.222	0.976			0.187	-0.147	0.032			-1.478			0.913
RE-RV-RR-NN	0.000	-0.393	0.956	0.174	0.158		-0.142	0.046	-0.618	-0.888		1.020	1.021	
RE-RV-RR-SkN	0.000	-0.390	0.956	0.183	0.171		-0.150	0.049	-1.039	-1.281		0.973	0.977	
RE-RV-RR-tN	0.000	-0.375	0.958	0.181	0.168		-0.145	0.048	-0.953	-1.209		0.983	0.985	
RE-RV-RK-NN	0.000	-0.332	0.963	0.233		0.047	-0.142	0.041	-0.695		-0.647	1.010		1.006
RE-RV-RK-SkN	0.000	-0.323	0.964	0.259		0.050	-0.155	0.043	-1.313		-1.230	0.942		0.941
RE-RV-RK-tN	0.001	-0.314	0.966	0.255		0.050	-0.151	0.043	-1.241		-1.166	0.950		0.949
RE-RR-RK-NN	0.000	-0.347	0.961		0.250	0.066	-0.149	0.043		-1.234	-0.909		0.982	0.977
RE-RR-RK-SkN	0.000	-0.344	0.962		0.268	0.068	-0.157	0.045		-1.554	-1.221		0.946	0.943
RE-RR-RK-tN	0.001	-0.335	0.963		0.262	0.068	-0.152	0.045		-1.408	-1.103		0.963	0.956
RE-RV-RR-RK-NN	0.000	-0.376	0.958	0.119	0.130	0.055	-0.139	0.045	-0.513	-0.776	-0.406	1.031	1.033	1.034
RE-RV-RR-RK-SkN	0.000	-0.386	0.957	0.119	0.135	0.055	-0.141	0.045	-0.541	-0.774	-0.389	1.027	1.032	1.035
RE-RV-RR-RK-tN	0.001	-0.374	0.959	0.116	0.137	0.055	-0.138	0.046	-0.532	-0.774	-0.389	1.030	1.034	1.036
Average	**0.000**	**-0.338**	**0.962**	**0.220**	**0.231**	**0.089**	**-0.149**	**0.043**	**-0.945**	**-1.186**	**-0.983**	**0.983**	**0.987**	**0.969**

TABLE A.10: Estimated parameter mean for all forecasting steps for each parameter: NASDAQ-Part 2

Model	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}	δ_{23}	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RV-NN	-0.159			0.043			0.194					
RE-RV-SkN	-0.159			0.043			0.193				13.605	-0.154
RE-RV-tN	-0.157			0.043			0.193			12.572		
RE-RR-NN		-0.156			0.040			0.175				
RE-RR-SkN		-0.157			0.041			0.175			13.026	-0.158
RE-RR-tN		-0.154			0.040			0.175		12.308		
RE-RK-NN			-0.139			0.097			0.354			
RE-RK-SkN			-0.139			0.097			0.354		15.626	-0.155
RE-RK-tN			-0.134			0.097			0.354	13.722		
RE-RV-RR-NN	-0.157	-0.156		0.042	0.039		0.195	0.176				
RE-RV-RR-SkN	-0.158	-0.156		0.044	0.040		0.195	0.176			13.690	-0.151
RE-RV-RR-tN	-0.155	-0.154		0.043	0.040		0.194	0.176		12.718		
RE-RV-RK-NN	-0.157		-0.136	0.040		0.095	0.193		0.349			
RE-RV-RK-SkN	-0.157		-0.137	0.042		0.097	0.193		0.349		14.289	-0.152
RE-RV-RK-tN	-0.156		-0.132	0.042		0.097	0.193		0.348	12.778		
RE-RR-RK-NN		-0.156	-0.134		0.038	0.099		0.175	0.352			
RE-RR-RK-SkN		-0.155	-0.136		0.040	0.100		0.175	0.352		13.696	-0.152
RE-RR-RK-tN		-0.154	-0.130		0.039	0.101		0.175	0.351	12.988		
RE-RV-RR-RK-NN	-0.156	-0.155	-0.134	0.042	0.038	0.097	0.193	0.176	0.350			
RE-RV-RR-RK-SkN	-0.156	-0.155	-0.134	0.041	0.037	0.096	0.194	0.176	0.351		14.499	-0.145
RE-RV-RR-RK-tN	-0.155	-0.154	-0.132	0.042	0.037	0.097	0.194	0.176	0.351	13.028		
Average	**-0.157**	**-0.155**	**-0.135**	**0.042**	**0.039**	**0.098**	**0.194**	**0.176**	**0.351**	**12.874**	**14.062**	**-0.152**

TABLE A.11: Estimated parameter mean for all forecasting steps for each parameter: SMI-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.000	-0.288	0.969	0.290			-0.119	0.031	-0.016			1.061		
RE-RV-SkN	0.000	-0.277	0.970	0.285			-0.119	0.031	-0.039			1.058		
RE-RV-tN	0.000	-0.270	0.971	0.285			-0.118	0.031	-0.047			1.057		
RE-RR-NN	0.000	-0.292	0.968		0.323		-0.105	0.030		-0.067			1.048	
RE-RR-SkN	0.000	-0.284	0.969		0.321		-0.104	0.030		-0.089			1.046	
RE-RR-tN	0.000	-0.280	0.970		0.323		-0.104	0.030		-0.106			1.044	
RE-RK-NN	0.000	-0.403	0.956			0.159	-0.129	0.029			0.234			1.030
RE-RK-SkN	0.000	-0.409	0.955			0.168	-0.132	0.026			0.245			1.035
RE-RK-tN	0.000	-0.403	0.956			0.169	-0.132	0.027			0.215			1.031
RE-RV-RR-NN	0.000	-0.255	0.972	0.042	0.281		-0.109	0.031	0.132	-0.209		1.078	1.034	
RE-RV-RR-SkN	0.000	-0.267	0.971	0.042	0.282		-0.110	0.032	0.076	-0.404		1.073	1.013	
RE-RV-RR-tN	0.000	-0.259	0.972	0.040	0.283		-0.109	0.032	0.092	-0.380		1.075	1.016	
RE-RV-RK-NN	0.000	-0.320	0.965	0.268		0.025	-0.122	0.031	0.115		-1.877	1.075		0.788
RE-RV-RK-SkN	0.000	-0.316	0.966	0.266		0.025	-0.122	0.031	0.050		-2.064	1.068		0.767
RE-RV-RK-tN	0.000	-0.309	0.967	0.263		0.025	-0.121	0.031	0.099		-1.952	1.073		0.779
RE-RR-RK-NN	0.000	-0.328	0.964		0.301	0.022	-0.107	0.030		0.099	-1.369		1.067	0.843
RE-RR-RK-SkN	0.000	-0.322	0.965		0.299	0.021	-0.106	0.030		0.089	-1.307		1.066	0.850
RE-RR-RK-tN	0.000	-0.318	0.966		0.299	0.022	-0.106	0.030		0.025	-1.349		1.059	0.845
RE-RV-RR-RK-NN	0.000	-0.283	0.969	0.031	0.280	0.012	-0.110	0.031	0.210	-0.100	-0.520	1.088	1.046	0.935
RE-RV-RR-RK-SkN	0.000	-0.286	0.969	0.029	0.279	0.013	-0.110	0.032	0.243	-0.151	-0.736	1.092	1.042	0.913
RE-RV-RR-RK-tN	0.000	-0.275	0.970	0.026	0.282	0.013	-0.110	0.032	0.187	-0.177	-0.664	1.087	1.039	0.921
Average	**0.000**	**-0.307**	**0.967**	**0.156**	**0.296**	**0.056**	**-0.114**	**0.030**	**0.092**	**-0.123**	**-0.929**	**1.074**	**1.043**	**0.895**

TABLE A.12: Estimated parameter mean for all forecasting steps for each parameter: SMI-Part 2

Model	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}	δ_{23}	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RV-NN	-0.147			0.049			0.133					
RE-RV-SkN	-0.148			0.049			0.134				10.233	-0.145
RE-RV-tN	-0.147			0.049			0.134			9.744		
RE-RR-NN		-0.122			0.043			0.095				
RE-RR-SkN		-0.123			0.043			0.095			9.979	-0.140
RE-RR-tN		-0.122			0.043			0.095		9.473		
RE-RK-NN			-0.126			0.036	0.575					
RE-RK-SkN			-0.128			0.039	0.571				5.786	-0.123
RE-RK-tN			-0.127			0.038	0.571			5.724		
RE-RV-RR-NN	-0.146	-0.124		0.051	0.043		0.134	0.095				
RE-RV-RR-SkN	-0.147	-0.123		0.052	0.044		0.133	0.096			9.201	-0.143
RE-RV-RR-tN	-0.146	-0.123		0.051	0.044		0.133	0.095		9.045		
RE-RV-RK-NN	-0.147		-0.117	0.051		0.045	0.137		0.806			
RE-RV-RK-SkN	-0.147		-0.117	0.051		0.044	0.137		0.802		10.009	-0.148
RE-RV-RK-tN	-0.147		-0.120	0.050		0.044	0.137		0.800	9.788		
RE-RR-RK-NN		-0.123	-0.116		0.044	0.044		0.97	0.807			
RE-RR-RK-SkN		-0.123	-0.114		0.044	0.045		0.097	0.805		9.757	-0.141
RE-RR-RK-tN		-0.121	-0.116		0.044	0.046		0.097	0.809	9.326		
RE-RV-RR-RK-NN	-0.147	-0.124	-0.112	0.052	0.044	0.048	0.135	0.096	0.844			
RE-RV-RR-RK-SkN	-0.147	-0.124	-0.114	0.052	0.045	0.048	0.134	0.096	0.841		9.391	-0.144
RE-RV-RR-RK-tN	-0.148	-0.124	-0.115	0.053	0.045	0.049	0.135	0.096	0.843	8.783		
Average	**-0.147**	**-0.123**	**-0.118**	**0.051**	**0.044**	**0.044**	**0.222**	**0.096**	**0.817**	**8.840**	**9.194**	**-0.141**

TABLE A.13: Estimated parameter mean for all forecasting steps for each parameter: S&P 500-Part 1

Model	μ	ω	β	γ_1	γ_2	γ_3	τ_1	τ_2	ξ_1	ξ_2	ξ_3	φ_1	φ_2	φ_3
RE-RV-NN	0.0001	-0.322	0.965	0.319			-0.179	0.037	-0.934			0.967		
RE-RV-SkN	0.0000	-0.324	0.965	0.336			-0.186	0.039	-1.165			0.943		
RE-RV-tN	0.0003	-0.305	0.968	0.335			-0.183	0.038	-1.158			0.944		
RE-RR-NN	0.0000	-0.326	0.965		0.344		-0.177	0.039		-1.066			0.977	
RE-RR-SkN	0.0000	-0.327	0.965		0.357		-0.183	0.041		-1.227			0.961	
RE-RR-tN	0.0003	-0.313	0.967		0.356		-0.179	0.040		-1.210			0.963	
RE-RK-NN	0.0001	-0.264	0.972			0.141	-0.174	0.031			-0.538			1.003
RE-RK-SkN	0.0001	-0.272	0.971			0.144	-0.181	0.032			-0.694			0.987
RE-RK-tN	0.0003	-0.254	0.973			0.144	-0.178	0.032			-0.662			0.990
RE-RV-RR-NN	0.0001	-0.355	0.962	0.185	0.125		-0.162	0.042	-0.239	-0.354		1.041	1.054	
RE-RV-RR-SkN	0.0001	-0.348	0.963	0.198	0.133		-0.172	0.044	-0.707	-0.817		0.991	1.004	
RE-RV-RR-tN	0.0003	-0.334	0.965	0.199	0.128		-0.168	0.043	-0.587	-0.702		1.005	1.018	
RE-RV-RK-NN	0.0002	-0.330	0.965	0.256		0.011	-0.173	0.033	-0.220		-0.150	1.043		1.043
RE-RV-RK-SkN	0.0001	-0.339	0.964	0.271		0.013	-0.182	0.034	-0.539		-0.422	1.009		1.015
RE-RV-RK-tN	0.0004	-0.318	0.967	0.266		0.013	-0.177	0.034	-0.425		-0.337	1.022		1.024
RE-RR-RK-NN	0.0001	-0.339	0.964		0.282	0.011	-0.174	0.034		-0.408	-0.205		1.048	1.038
RE-RR-RK-SkN	0.0001	-0.347	0.963		0.295	0.012	-0.182	0.036		-0.700	-0.457		1.016	1.011
RE-RR-RK-tN	0.0003	-0.333	0.965		0.291	0.013	-0.178	0.036		-0.592	-0.360		1.028	1.021
RE-RV-RR-RK-NN	0.0002	-0.349	0.963	0.158	0.120	0.015	-0.167	0.039	-0.244	-0.360	-0.117	1.041	1.053	1.047
RE-RV-RR-RK-SkN	0.0002	-0.353	0.962	0.163	0.118	0.015	-0.169	0.039	-0.275	-0.407	-0.162	1.037	1.048	1.042
RE-RV-RR-RK-tN	0.0004	-0.337	0.965	0.161	0.117	0.016	-0.165	0.040	-0.246	-0.368	-0.138	1.042	1.053	1.045
Average	**0.0002**	**-0.323**	**0.966**	**0.237**	**0.222**	**0.046**	**-0.176**	**0.037**	**-0.562**	**-0.684**	**-0.353**	**1.007**	**1.019**	**1.022**

TABLE A.14: Estimated parameter mean for all forecasting steps for each parameter: S&P 500-Part 2

Model	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}	δ_{23}	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	$\sigma^2_{u_3}$	ν	ν_{skw}	λ
RE-RV-NN	-0.159			0.064			0.194					
RE-RV-SkN	-0.160			0.065			0.194				10.266	-0.167
RE-RV-tN	-0.156			0.064			0.194			9.594		
RE-RR-NN		-0.153			0.065			0.175				
RE-RR-SkN		-0.154			0.066			0.175			10.688	-0.165
RE-RR-tN		-0.150			0.066			0.175		10.187		
RE-RK-NN			-0.053			0.205			0.413			
RE-RK-SkN			-0.056			0.204			0.413		10.599	-0.166
RE-RK-tN			-0.042			0.205			0.413	9.750		
RE-RV-RR-NN	-0.158	-0.151		0.063	0.066		0.194	0.176				
RE-RV-RR-SkN	-0.159	-0.153		0.064	0.067		0.195	0.176			10.568	-0.162
RE-RV-RR-tN	-0.155	-0.147		0.064	0.067		0.195	0.176		9.766		
RE-RV-RK-NN	-0.158		-0.050	0.062		0.202	0.194		0.402			
RE-RV-RK-SkN	-0.159		-0.053	0.063		0.202	0.194		0.402		10.797	-0.161
RE-RV-RK-tN	-0.156		-0.041	0.063		0.203	0.194		0.401	10.093		
RE-RR-RK-NN		-0.152	-0.054		0.064	0.204		0.176	0.402			
RE-RR-RK-SkN		-0.153	-0.056		0.065	0.203		0.176	0.402		11.078	-0.160
RE-RR-RK-tN		-0.150	-0.043		0.065	0.204		0.176	0.401	10.370		
RE-RV-RR-RK-NN	-0.157	-0.150	-0.050	0.063	0.066	0.203	0.195	0.176	0.403			
RE-RV-RR-RK-SkN	-0.158	-0.151	-0.051	0.063	0.065	0.203	0.195	0.176	0.403		11.047	-0.157
RE-RV-RR-RK-tN	-0.154	-0.147	-0.038	0.063	0.066	0.204	0.195	0.175	0.403	9.762		
Average	**-0.157**	**-0.151**	**-0.049**	**0.064**	**0.066**	**0.203**	**0.194**	**0.176**	**0.405**	**9.932**	**10.720**	**-0.162**

A.3 Model Confidence Set Results

TABLE A.15: 90% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-t	0.096	0.001	0.159	0.248	0.120	0.513	0.017	4
EGARCH-t	0.096	0.001	0.427	0.248	0.382	0.513	0.280	5
GJR-t	0.096	0.001	0.313	0.368	0.323	0.513	0.094	4
GARCH-HS	0.076	0.001	0.257	0.248	0.382	0.836	0.032	4
GARCH-t-EVT	0.002	0.001	0.000	0.248	0.071	0.836	0.000	2
CARE-SAV	0.002	0.988	0.427	0.041	0.289	0.252	0.000	4
RG-RV-tN	0.096	0.031	0.201	0.809	0.146	0.189	0.482	5
RG-RR-tN	0.096	0.031	0.313	0.979	0.623	0.252	0.482	5
RG-RK-tN	0.096	0.112	0.235	0.809	0.213	0.513	0.187	6
RE-RV-NN	0.096	0.160	0.313	0.809	0.323	0.836	0.689	6
RE-RV-SkN	0.809	0.947	0.634	0.961	0.698	0.836	0.959	7
RE-RV-tN	0.596	0.940	0.634	0.809	0.009	0.252	0.689	6
RE-RR-NN	0.596	0.302	0.427	1.000	0.757	0.513	0.689	7
RE-RR-SkN	0.096	0.947	0.853	0.893	0.907	0.890	0.689	6
RE-RR-tN	0.096	0.940	0.634	0.248	0.907	0.890	0.482	6
RE-RK-NN	0.096	0.160	0.257	0.809	0.623	0.782	0.482	6
RE-RK-SkN	0.596	0.947	0.634	0.961	0.757	1.000	0.689	7
RE-RK-tN	0.809	0.302	0.634	0.809	0.880	0.890	0.689	7
RE-RV-RR-NN	0.096	0.160	0.634	0.893	0.880	0.836	0.482	6
RE-RV-RR-SkN	0.096	1.000	0.634	0.979	0.757	0.836	0.689	6
RE-RV-RR-tN	0.809	0.581	1.000	0.248	0.907	0.836	1.000	7
RE-RV-RK-NN	0.096	0.192	0.313	0.809	0.623	0.836	0.482	6
RE-RV-RK-SkN	0.809	0.988	0.709	0.893	0.757	0.932	0.959	7
RE-RV-RK-tN	0.809	0.581	0.634	0.307	0.757	0.932	0.959	7
RE-RR-RK-NN	0.096	0.237	0.444	0.893	0.757	0.513	0.482	6
RE-RR-RK-SkN	1.000	0.302	0.634	0.809	0.880	0.890	0.959	7
RE-RR-RK-tN	0.809	0.302	0.634	0.307	0.884	0.890	0.959	7
RE-RV-RR-RK-NN	0.096	0.160	0.427	0.961	0.757	0.782	0.482	6
RE-RV-RR-RK-SkN	0.809	0.581	0.634	0.893	1.000	0.932	0.959	7
RE-RV-RR-RK-tN	0.596	0.648	0.634	0.368	0.907	0.890	0.862	7

TABLE A.16: 90% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-t	0.065	0.000	0.032	0.312	0.026	0.256	0.000	2
EGARCH-t	0.516	0.000	0.182	0.452	0.114	0.277	0.000	5
GJR-t	0.149	0.000	0.089	0.743	0.082	0.321	0.000	3
GARCH-HS	0.245	0.000	0.078	0.366	0.130	0.632	0.000	4
GARCH-t-EVT	0.000	0.000	0.003	0.241	0.019	0.753	0.000	2
CARE-SAV	0.356	0.985	0.141	0.166	0.072	0.000	0.000	4
RG-RV-tN	0.684	0.000	0.047	0.847	0.035	0.000	0.000	2
RG-RR-tN	0.907	0.000	0.093	0.979	0.272	0.200	0.000	4
RG-RK-tN	0.858	0.000	0.061	0.851	0.046	0.298	0.000	3
RE-RV-NN	0.626	0.000	0.102	0.899	0.097	0.632	0.000	4
RE-RV-SkN	0.767	0.983	0.415	0.978	0.325	0.632	0.953	7
RE-RV-tN	0.745	0.983	0.382	0.875	0.014	0.224	0.000	5
RE-RR-NN	0.853	0.000	0.196	1.000	0.396	0.325	0.773	6
RE-RR-SkN	0.945	0.983	0.853	0.965	0.916	0.756	0.869	7
RE-RR-tN	0.947	0.983	0.423	0.458	0.916	0.756	0.000	6
RE-RK-NN	0.786	0.000	0.082	0.904	0.180	0.508	0.000	4
RE-RK-SkN	0.947	0.983	0.382	0.978	0.429	1.000	0.761	7
RE-RK-tN	0.907	0.000	0.362	0.896	0.773	0.788	0.000	5
RE-RV-RR-NN	0.858	0.000	0.342	0.961	0.822	0.566	0.000	5
RE-RV-RR-SkN	1.000	1.000	0.437	0.979	0.617	0.753	0.000	6
RE-RV-RR-tN	0.940	0.000	1.000	0.523	0.916	0.753	1.000	6
RE-RV-RK-NN	0.808	0.000	0.105	0.914	0.214	0.651	0.000	5
RE-RV-RK-SkN	0.947	0.985	0.702	0.965	0.555	0.928	0.936	7
RE-RV-RK-tN	0.907	0.658	0.382	0.598	0.509	0.928	0.942	7
RE-RR-RK-NN	0.936	0.000	0.252	0.960	0.456	0.376	0.000	5
RE-RR-RK-SkN	0.864	0.000	0.437	0.921	0.898	0.756	0.953	6
RE-RR-RK-tN	0.936	0.000	0.437	0.660	0.900	0.756	0.953	6
RE-RV-RR-RK-NN	0.947	0.000	0.154	0.978	0.353	0.458	0.000	5
RE-RV-RR-RK-SkN	0.947	0.592	0.523	0.954	1.000	0.928	0.953	7
RE-RV-RR-RK-tN	0.947	0.827	0.415	0.745	0.916	0.756	0.914	7

TABLE A.17: 75% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-t	0.020	0.001	0.156	0.225	0.132	0.559	0.019	1
EGARCH-t	0.612	0.001	0.420	0.225	0.355	0.559	0.256	5
GJR-t	0.080	0.001	0.271	0.334	0.313	0.559	0.095	4
GARCH-HS	0.081	0.001	0.271	0.225	0.355	0.842	0.034	3
GARCH-t-EVT	0.000	0.001	0.000	0.225	0.066	0.842	0.000	1
CARE-SAV	0.139	0.990	0.420	0.043	0.259	0.259	0.001	4
RG-RV-tN	0.766	0.030	0.195	0.813	0.181	0.192	0.492	3
RG-RR-tN	0.966	0.030	0.299	0.974	0.624	0.259	0.492	6
RG-RK-tN	0.922	0.109	0.232	0.813	0.234	0.559	0.174	3
RE-RV-NN	0.766	0.168	0.299	0.813	0.354	0.842	0.681	6
RE-RV-SkN	0.799	0.950	0.641	0.966	0.710	0.842	0.956	7
RE-RV-tN	0.766	0.935	0.641	0.813	0.008	0.259	0.681	6
RE-RR-NN	0.922	0.299	0.420	1.000	0.732	0.559	0.681	7
RE-RR-SkN	0.966	0.950	0.849	0.899	0.907	0.900	0.681	7
RE-RR-tN	0.981	0.935	0.641	0.225	0.907	0.900	0.492	6
RE-RK-NN	0.910	0.168	0.271	0.813	0.624	0.799	0.492	6
RE-RK-SkN	0.981	0.950	0.641	0.966	0.732	1.000	0.681	7
RE-RK-tN	0.966	0.299	0.641	0.813	0.886	0.900	0.681	7
RE-RV-RR-NN	0.922	0.168	0.641	0.899	0.886	0.842	0.492	6
RE-RV-RR-SkN	1.000	1.000	0.641	0.974	0.732	0.842	0.681	7
RE-RV-RR-tN	0.966	0.589	1.000	0.225	0.907	0.842	1.000	6
RE-RV-RK-NN	0.910	0.199	0.299	0.813	0.624	0.842	0.492	6
RE-RV-RK-SkN	0.966	0.990	0.704	0.899	0.827	0.930	0.956	7
RE-RV-RK-tN	0.966	0.589	0.641	0.275	0.732	0.930	0.956	7
RE-RR-RK-NN	0.966	0.259	0.451	0.899	0.732	0.559	0.492	7
RE-RR-RK-SkN	0.922	0.299	0.641	0.813	0.886	0.900	0.956	7
RE-RR-RK-tN	0.966	0.299	0.641	0.275	0.890	0.900	0.956	7
RE-RV-RR-RK-NN	0.966	0.168	0.420	0.966	0.732	0.799	0.492	6
RE-RV-RR-RK-SkN	0.981	0.589	0.641	0.899	1.000	0.930	0.956	7
RE-RV-RR-RK-tN	0.981	0.634	0.641	0.334	0.907	0.900	0.866	7

TABLE A.18: 75% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-t	0.074	0.000	0.025	0.323	0.039	0.283	0.000	2
EGARCH-t	0.534	0.000	0.172	0.460	0.128	0.301	0.000	3
GJR-t	0.150	0.000	0.078	0.762	0.097	0.341	0.000	2
GARCH-HS	0.245	0.000	0.072	0.377	0.153	0.645	0.000	2
GARCH-t-EVT	0.000	0.000	0.003	0.237	0.026	0.764	0.000	1
CARE-SAV	0.369	0.989	0.130	0.161	0.082	0.000	0.000	2
RG-RV-tN	0.700	0.000	0.044	0.853	0.048	0.000	0.000	2
RG-RR-tN	0.914	0.000	0.089	0.980	0.285	0.218	0.000	3
RG-RK-tN	0.874	0.000	0.057	0.862	0.059	0.318	0.000	3
RE-RV-NN	0.642	0.000	0.098	0.909	0.110	0.645	0.000	3
RE-RV-SkN	0.777	0.986	0.396	0.980	0.349	0.645	0.947	7
RE-RV-tN	0.756	0.986	0.375	0.882	0.016	0.244	0.000	4
RE-RR-NN	0.872	0.000	0.187	1.000	0.414	0.348	0.775	5
RE-RR-SkN	0.944	0.986	0.849	0.969	0.925	0.770	0.864	7
RE-RR-tN	0.946	0.986	0.403	0.463	0.925	0.770	0.000	6
RE-RK-NN	0.803	0.000	0.080	0.913	0.205	0.529	0.000	3
RE-RK-SkN	0.946	0.986	0.375	0.980	0.439	1.000	0.755	7
RE-RK-tN	0.914	0.000	0.355	0.903	0.768	0.779	0.000	5
RE-RV-RR-NN	0.874	0.000	0.333	0.969	0.826	0.576	0.000	5
RE-RV-RR-SkN	1.000	1.000	0.419	0.980	0.559	0.764	0.000	6
RE-RV-RR-tN	0.940	0.000	1.000	0.530	0.925	0.764	1.000	6
RE-RV-RK-NN	0.823	0.000	0.100	0.919	0.238	0.665	0.000	3
RE-RV-RK-SkN	0.946	0.989	0.692	0.969	0.658	0.927	0.932	7
RE-RV-RK-tN	0.914	0.679	0.375	0.607	0.510	0.927	0.932	7
RE-RR-RK-NN	0.934	0.000	0.237	0.969	0.463	0.398	0.000	4
RE-RR-RK-SkN	0.880	0.000	0.419	0.928	0.899	0.770	0.947	6
RE-RR-RK-tN	0.934	0.000	0.419	0.671	0.904	0.770	0.947	6
RE-RV-RR-RK-NN	0.946	0.000	0.143	0.980	0.372	0.479	0.000	4
RE-RV-RR-RK-SkN	0.946	0.601	0.523	0.957	1.000	0.927	0.947	7
RE-RV-RR-RK-tN	0.946	0.830	0.396	0.762	0.925	0.770	0.912	7

TABLE A.19: 90% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 1\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-t	0.151	0.440	0.402	0.135	0.025	0.369	0.092	5
EGARCH-t	0.694	0.299	0.451	0.135	0.561	0.369	0.092	6
GJR-t	0.502	0.294	0.625	0.534	0.561	0.369	0.092	6
GARCH-HS	0.502	0.931	0.439	0.135	0.728	0.369	0.079	6
GARCH-t-EVT	0.000	0.183	0.046	1.000	0.025	1.000	0.001	3
CARE-SAV	0.694	0.941	0.451	0.070	0.025	0.369	0.002	4
RG-RV-tN	0.883	0.240	0.402	0.534	0.025	0.369	0.092	5
RG-RR-tN	1.000	0.440	0.625	0.981	0.516	0.369	0.092	6
RG-RK-tN	1.000	0.203	0.402	0.135	0.025	0.369	0.092	5
RE-RV-NN	0.694	0.195	0.451	0.135	0.025	0.369	0.092	5
RE-RV-SkN	0.959	0.931	0.625	0.534	0.561	0.369	0.801	7
RE-RV-tN	0.922	0.931	0.625	0.897	0.561	0.369	0.605	7
RE-RR-NN	0.959	0.299	0.451	0.534	0.516	0.369	0.605	7
RE-RR-SkN	1.000	1.000	0.650	0.534	0.561	0.940	0.092	6
RE-RR-tN	1.000	0.941	0.625	0.135	0.599	0.940	0.092	6
RE-RK-NN	0.959	0.195	0.451	0.135	0.025	0.369	0.092	5
RE-RK-SkN	1.000	0.941	0.625	0.981	0.561	0.940	0.605	7
RE-RK-tN	1.000	0.812	0.625	0.534	0.561	0.620	0.801	7
RE-RV-RR-NN	0.883	0.208	0.625	0.506	0.561	0.369	0.092	6
RE-RV-RR-SkN	1.000	0.941	0.625	0.981	0.561	0.774	0.092	6
RE-RV-RR-tN	0.922	0.931	1.000	0.506	0.599	0.940	0.801	7
RE-RV-RK-NN	0.694	0.183	0.451	0.135	0.025	0.369	0.092	5
RE-RV-RK-SkN	0.989	0.941	0.625	0.981	0.561	0.940	0.801	7
RE-RV-RK-tN	0.956	0.931	0.625	0.534	0.561	0.774	0.801	7
RE-RR-RK-NN	1.000	0.240	0.625	0.506	0.516	0.369	0.092	6
RE-RR-RK-SkN	0.959	0.931	0.625	0.534	0.599	0.774	1.000	7
RE-RR-RK-tN	1.000	0.440	0.625	0.220	0.599	0.940	0.801	7
RE-RV-RR-RK-NN	1.000	0.183	0.451	0.506	0.516	0.369	0.092	6
RE-RV-RR-RK-SkN	0.998	0.941	0.625	0.897	1.000	0.940	0.801	7
RE-RV-RR-RK-tN	1.000	0.931	0.625	0.534	0.599	0.369	0.605	7

TABLE A.20: 90% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 1\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-t	0.323	0.000	0.150	0.494	0.176	0.571	0.000	5
EGARCH-t	0.638	0.000	0.284	0.579	0.548	0.183	0.000	5
GJR-t	0.456	0.000	0.431	0.873	0.460	0.439	0.000	5
GARCH-HS	0.512	0.000	0.194	0.558	0.748	0.439	0.000	5
GARCH-t-EVT	0.043	0.000	0.071	1.000	0.056	1.000	0.000	2
CARE-SAV	0.782	0.860	0.254	0.383	0.208	0.229	0.000	6
RG-RV-tN	0.882	0.000	0.167	0.877	0.042	0.144	0.000	4
RG-RR-tN	0.997	0.000	0.504	0.977	0.404	0.131	0.000	5
RG-RK-tN	0.997	0.000	0.177	0.579	0.070	0.149	0.000	4
RE-RV-NN	0.584	0.000	0.209	0.579	0.083	0.199	0.000	4
RE-RV-SkN	0.967	0.000	0.568	0.901	0.548	0.571	0.821	6
RE-RV-tN	0.912	0.000	0.547	0.966	0.636	0.153	0.000	5
RE-RR-NN	0.986	0.000	0.301	0.912	0.290	0.137	0.806	6
RE-RR-SkN	0.997	1.000	0.650	0.925	0.677	0.827	0.000	6
RE-RR-tN	0.997	0.926	0.579	0.581	0.748	0.827	0.000	6
RE-RK-NN	0.979	0.000	0.215	0.558	0.118	0.133	0.000	5
RE-RK-SkN	0.997	0.860	0.568	0.971	0.466	0.827	0.000	6
RE-RK-tN	0.995	0.000	0.568	0.925	0.548	0.756	0.821	6
RE-RV-RR-NN	0.847	0.000	0.464	0.769	0.439	0.345	0.000	5
RE-RV-RR-SkN	0.997	0.926	0.568	0.977	0.591	0.827	0.000	6
RE-RV-RR-tN	0.923	0.000	1.000	0.794	0.748	0.827	0.821	6
RE-RV-RK-NN	0.711	0.000	0.234	0.579	0.140	0.271	0.000	5
RE-RV-RK-SkN	0.992	0.860	0.596	0.977	0.502	0.827	0.821	7
RE-RV-RK-tN	0.941	0.695	0.554	0.894	0.527	0.827	0.821	7
RE-RR-RK-NN	0.997	0.000	0.338	0.755	0.371	0.160	0.000	5
RE-RR-RK-SkN	0.988	0.786	0.579	0.917	0.748	0.827	1.000	7
RE-RR-RK-tN	0.997	0.000	0.579	0.660	0.703	0.827	0.821	6
RE-RV-RR-RK-NN	0.996	0.000	0.264	0.785	0.260	0.230	0.000	5
RE-RV-RR-RK-SkN	0.995	0.853	0.579	0.966	1.000	0.827	0.821	7
RE-RV-RR-RK-tN	0.997	0.738	0.568	0.916	0.748	0.571	0.818	7

TABLE A.21: 75% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 1\%$

GARCH-t	0.154	0.473	0.418	0.139	0.024	0.398	0.096	3
EGARCH-t	0.686	0.326	0.449	0.139	0.555	0.398	0.096	5
GJR-t	0.470	0.326	0.646	0.513	0.555	0.398	0.096	6
GARCH-HS	0.470	0.918	0.449	0.139	0.728	0.398	0.076	5
GARCH-t-EVT	0.000	0.207	0.045	1.000	0.024	1.000	0.002	2
CARE-SAV	0.875	0.945	0.418	0.513	0.024	0.398	0.096	5
RG-RV-tN	0.686	0.271	0.449	0.070	0.024	0.398	0.002	4
RG-RR-tN	1.000	0.473	0.646	0.986	0.527	0.398	0.096	6
RG-RK-tN	1.000	0.215	0.418	0.139	0.024	0.398	0.096	3
RE-RV-NN	0.686	0.212	0.449	0.139	0.024	0.398	0.096	3
RE-RV-SkN	0.965	0.918	0.646	0.513	0.555	0.398	0.809	7
RE-RV-tN	0.928	0.807	0.646	0.905	0.555	0.398	0.596	7
RE-RR-NN	0.965	0.326	0.449	0.513	0.527	0.398	0.596	7
RE-RR-SkN	1.000	1.000	0.646	0.513	0.555	0.943	0.096	6
RE-RR-tN	1.000	0.945	0.646	0.139	0.588	0.943	0.096	5
RE-RK-NN	0.965	0.215	0.449	0.139	0.024	0.398	0.096	3
RE-RK-SkN	1.000	0.945	0.646	0.986	0.555	0.943	0.596	7
RE-RK-tN	1.000	0.807	0.646	0.513	0.555	0.645	0.809	7
RE-RV-RR-NN	0.875	0.224	0.646	0.513	0.555	0.398	0.096	5
RE-RV-RR-SkN	1.000	0.945	0.646	0.986	0.555	0.757	0.096	6
RE-RV-RR-tN	0.928	0.918	1.000	0.513	0.588	0.943	0.809	7
RE-RV-RK-NN	0.686	0.207	0.449	0.139	0.024	0.398	0.096	3
RE-RV-RK-SkN	0.989	0.945	0.646	0.986	0.555	0.943	0.809	7
RE-RV-RK-tN	0.962	0.918	0.646	0.513	0.555	0.757	0.809	7
RE-RR-RK-NN	1.000	0.252	0.646	0.513	0.527	0.398	0.096	6
RE-RR-RK-SkN	0.965	0.918	0.646	0.513	0.588	0.757	1.000	7
RE-RR-RK-tN	1.000	0.473	0.646	0.235	0.588	0.943	0.809	6
RE-RV-RR-RK-NN	1.000	0.207	0.449	0.513	0.527	0.398	0.096	5
RE-RV-RR-RK-SkN	0.998	0.945	0.646	0.905	1.000	0.943	0.809	7
RE-RV-RR-RK-tN	1.000	0.918	0.646	0.513	0.588	0.398	0.596	7
Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**

TABLE A.22: 75% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 1\%$

GARCH-t	0.327	0.000	0.151	0.488	0.175	0.584	0.000	3
EGARCH-t	0.660	0.000	0.283	0.574	0.557	0.192	0.000	4
GJR-t	0.472	0.000	0.431	0.872	0.463	0.446	0.000	5
GARCH-HS	0.533	0.000	0.194	0.555	0.729	0.446	0.000	4
GARCH-t-EVT	0.039	0.000	0.070	1.000	0.056	1.000	0.000	2
CARE-SAV	0.798	0.866	0.246	0.375	0.206	0.230	0.000	3
RG-RV-tN	0.893	0.000	0.165	0.876	0.072	0.143	0.000	2
RG-RR-tN	0.997	0.000	0.471	0.980	0.413	0.131	0.000	4
RG-RK-tN	0.997	0.000	0.175	0.574	0.086	0.140	0.000	2
RE-RV-NN	0.603	0.000	0.209	0.574	0.098	0.231	0.000	2
RE-RV-SkN	0.969	0.000	0.573	0.900	0.557	0.584	0.823	6
RE-RV-tN	0.920	0.000	0.556	0.960	0.631	0.160	0.000	4
RE-RR-NN	0.986	0.000	0.303	0.910	0.299	0.149	0.805	5
RE-RR-SkN	0.997	1.000	0.634	0.922	0.661	0.836	0.000	6
RE-RR-tN	0.997	0.924	0.574	0.577	0.729	0.836	0.000	6
RE-RK-NN	0.980	0.000	0.214	0.555	0.124	0.135	0.000	2
RE-RK-SkN	0.997	0.866	0.573	0.975	0.475	0.836	0.000	6
RE-RK-tN	0.995	0.000	0.573	0.922	0.557	0.764	0.823	6
RE-RV-RR-NN	0.856	0.000	0.503	0.765	0.443	0.355	0.000	5
RE-RV-RR-SkN	0.997	0.924	0.573	0.980	0.588	0.836	0.000	6
RE-RV-RR-tN	0.931	0.000	1.000	0.784	0.729	0.836	0.823	6
RE-RV-RK-NN	0.731	0.000	0.233	0.574	0.146	0.285	0.000	3
RE-RV-RK-SkN	0.992	0.866	0.600	0.980	0.505	0.836	0.823	7
RE-RV-RK-tN	0.947	0.700	0.559	0.893	0.533	0.836	0.823	7
RE-RR-RK-NN	0.997	0.000	0.344	0.744	0.386	0.173	0.000	4
RE-RR-RK-SkN	0.987	0.790	0.574	0.914	0.729	0.836	1.000	7
RE-RR-RK-tN	0.997	0.000	0.574	0.655	0.693	0.836	0.823	6
RE-RV-RR-RK-NN	0.997	0.000	0.264	0.759	0.268	0.208	0.000	4
RE-RV-RR-RK-SkN	0.994	0.845	0.574	0.963	1.000	0.836	0.823	7
RE-RV-RR-RK-tN	0.997	0.743	0.573	0.914	0.729	0.584	0.822	7
Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**

Appendix B

Chapter 4 Appendix

B.1 Distribution of the posterior mean of γ_1 and γ_4 based on historical data over the forecast period

FIGURE B.1: Distribution of the posterior mean of γ_1 and γ_4 over the forecast period: ASX 200

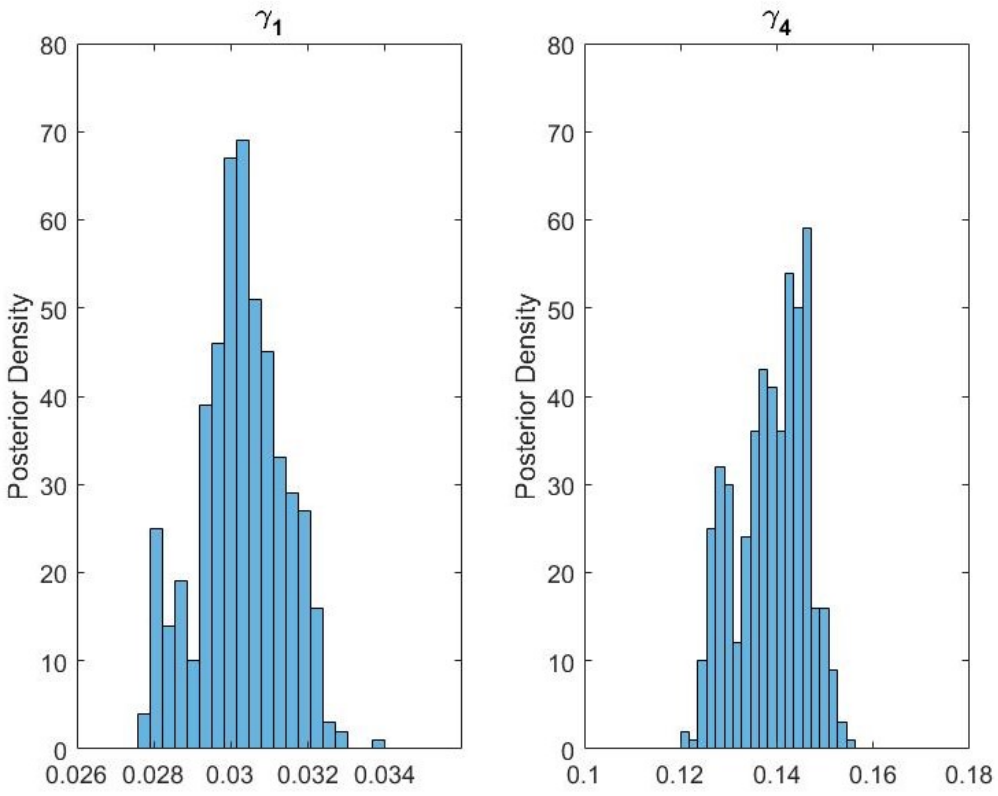


FIGURE B.2: Distribution of the posterior mean of γ_1 and γ_4 over the forecast period: DAX

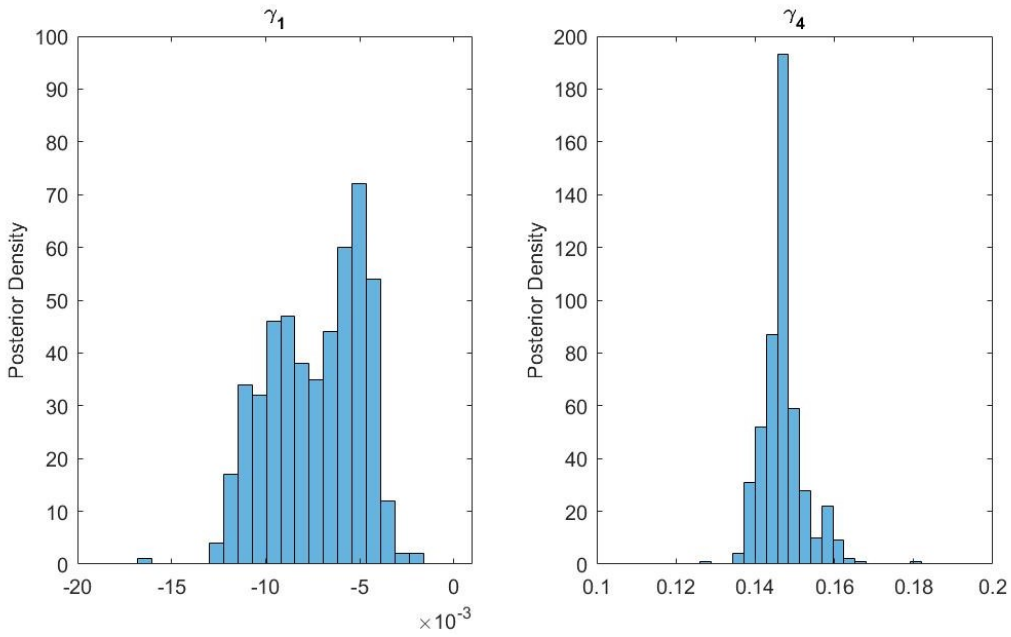


FIGURE B.3: Distribution of γ_1 and γ_4 over the forecast period: FTSE 100

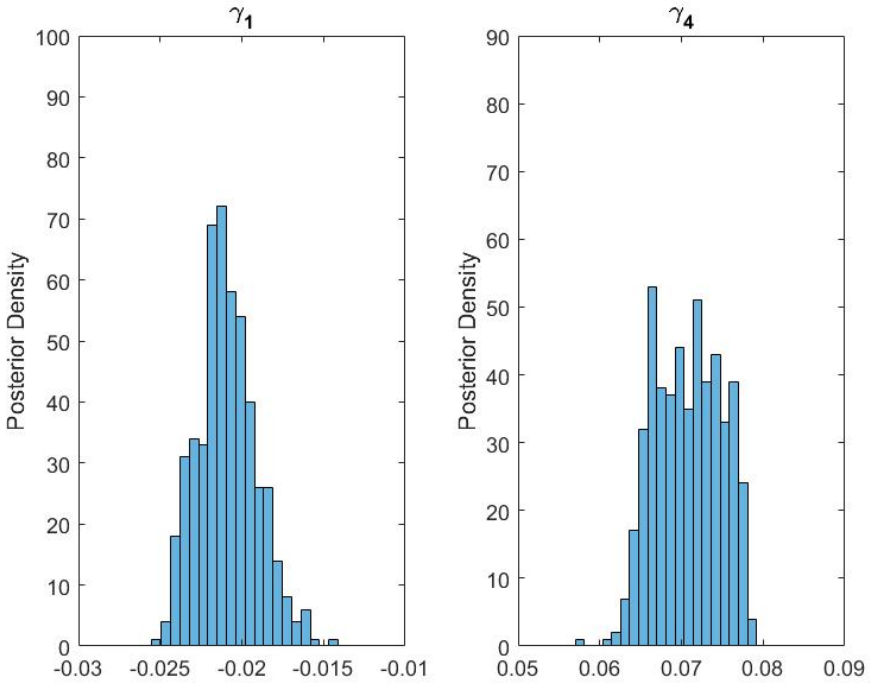


FIGURE B.4: Distribution of the posterior mean of γ_1 and γ_4 over the forecast period: Hang Seng

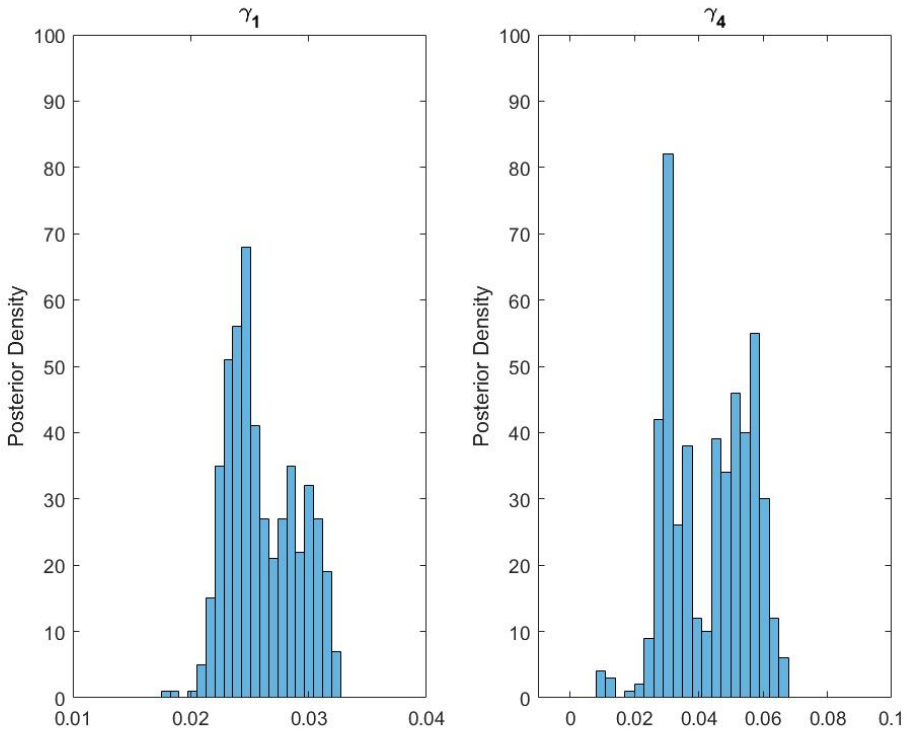


FIGURE B.5: Distribution of the posterior mean of γ_1 and γ_4 over the forecast period: Nasdaq

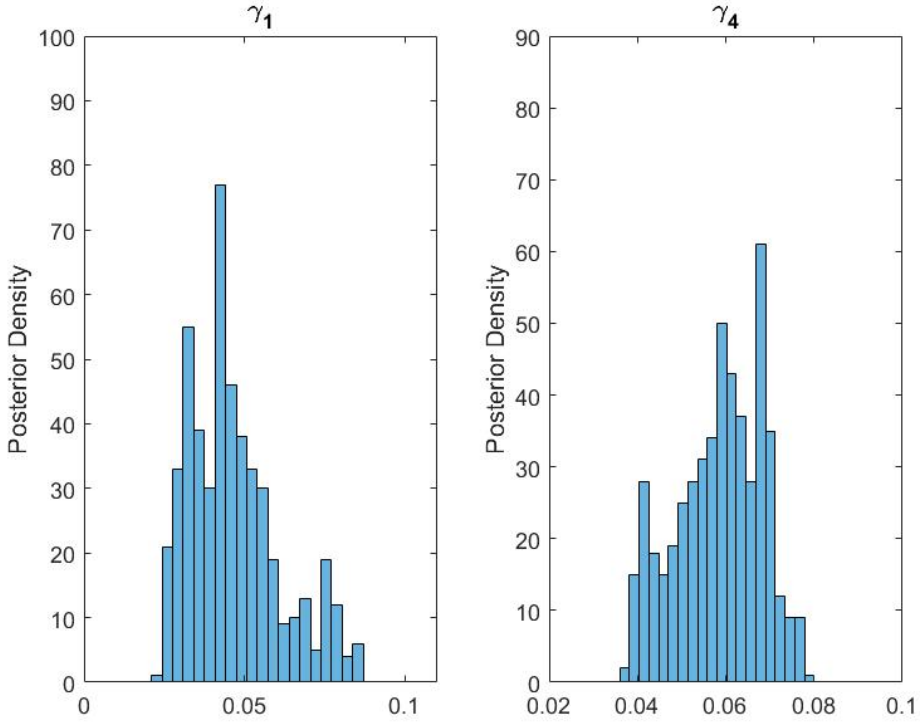


FIGURE B.6: Distribution of the posterior mean of γ_1 and γ_4 over the forecast period: SMI

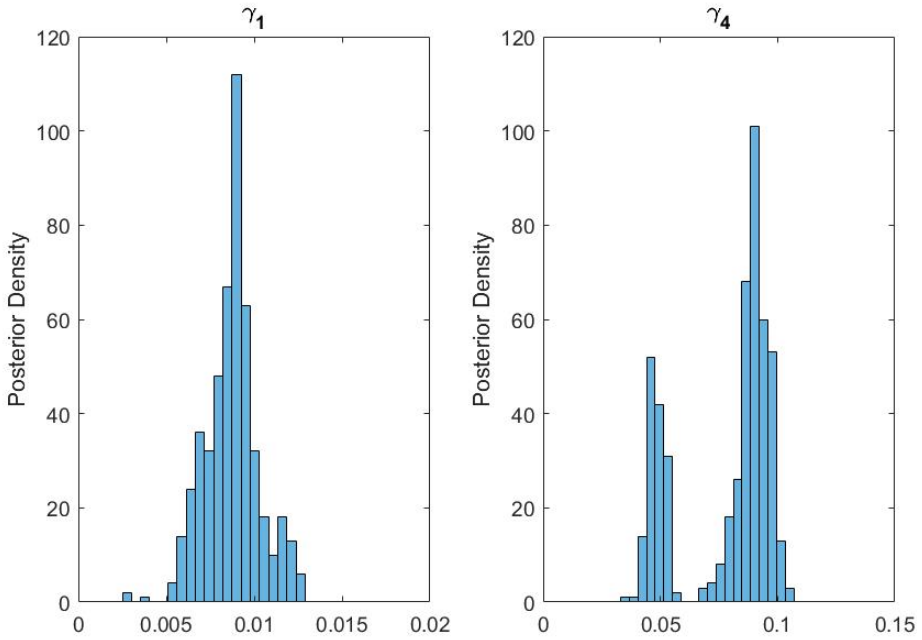
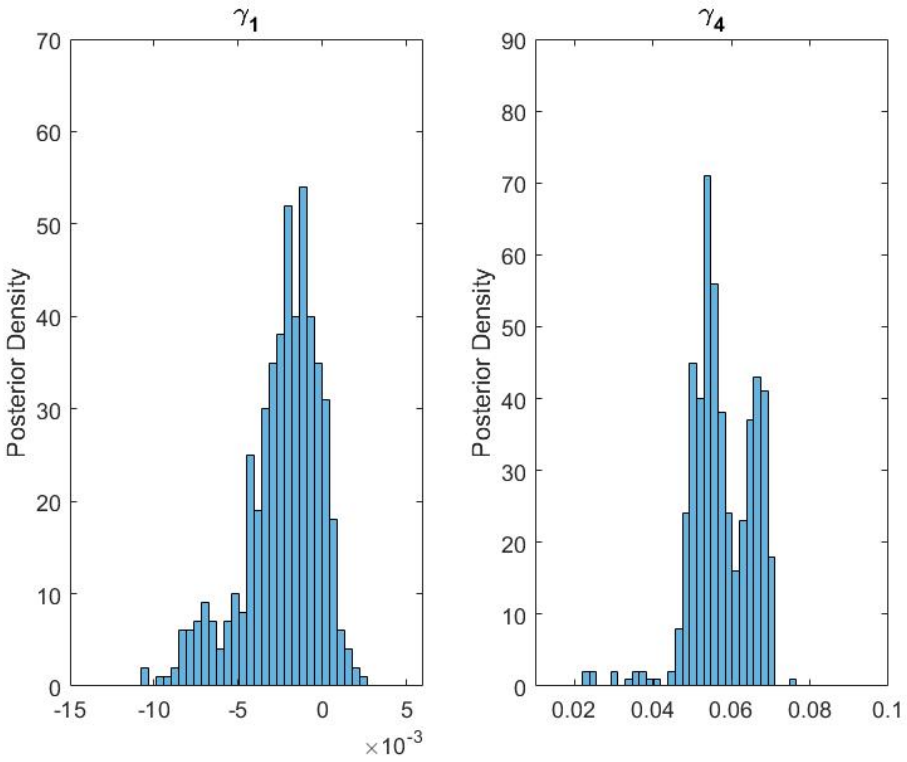


FIGURE B.7: Distribution of the posterior mean of γ_1 and γ_4 over the forecast period: S&P 500



B.2 Model Confidence Set Results

TABLE B.1: 90% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.413	0.337	0.676	0.686	0.472	0.328	0.021	6
EGARCH-Sk	0.203	0.337	0.417	0.420	0.454	0.328	0.022	6
RE-SkN-ML	0.413	0.337	0.676	0.686	0.454	0.328	0.276	7
RE-SkN-Lasso-CV	0.842	0.337	0.751	0.747	0.472	0.328	0.276	7
RE-SkN-MCMC	0.413	0.337	0.846	0.848	0.472	0.328	0.276	7
RE-SkN-Blasso	1.000	1.000	1.000	1.000	1.000	1.000	1.000	7

TABLE B.2: 90% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.428	0.306	0.696	0.511	0.455	0.335	0.028	6
EGARCH-Sk	0.206	0.306	0.444	0.634	0.438	0.335	0.029	5
RE-SkN-ML	0.428	0.306	0.696	0.634	0.438	0.335	0.294	7
RE-SkN-Lasso-CV	0.850	0.306	0.752	0.806	0.455	0.335	0.294	7
RE-SkN-MCMC	0.428	0.306	0.855	0.806	0.455	0.335	0.294	7
RE-SkN-Blasso	1.000	1.000	1.000	1.000	1.000	1.000	1.000	7

TABLE B.3: 75% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.236	0.200	0.374	0.439	0.772	0.721	0.003	6
EGARCH-Sk	0.598	0.200	0.374	0.439	0.772	0.757	0.020	6
RE-SkN-ML	0.236	0.200	0.374	0.439	0.695	0.757	0.179	7
RE-SkN-Lasso-CV	0.778	0.200	0.374	1.000	0.772	0.757	0.706	7
RE-SkN-MCMC	0.598	0.200	0.663	0.439	0.772	0.757	0.706	7
RE-SkN-Blasso	1.000	1.000	1.000	0.721	1.000	1.000	1.000	7

TABLE B.4: 75% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 2.5\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.436	0.405	0.415	0.249	0.480	0.238	0.006	4
EGARCH-Sk	0.202	0.442	0.243	0.473	0.355	0.320	0.015	4
RE-SkN-ML	0.560	0.442	0.541	0.473	0.355	0.320	0.225	6
RE-SkN-Lasso-CV	0.850	0.442	0.749	0.786	0.480	0.320	0.225	6
RE-SkN-MCMC	0.560	0.442	0.855	0.768	0.480	0.320	0.225	6
RE-SkN-Blasso	1.000	1.000	1.000	1.000	1.000	1.000	1.000	7

TABLE B.5: 90% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 1\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.236	0.200	0.374	0.439	0.772	0.721	0.003	6
EGARCH-Sk	0.598	0.200	0.374	0.439	0.772	0.757	0.020	6
RE-SkN-ML	0.236	0.200	0.374	0.439	0.695	0.757	0.179	7
RE-SkN-Lasso-CV	0.778	0.200	0.374	1.000	0.772	0.757	0.706	7
RE-SkN-MCMC	0.598	0.200	0.663	0.439	0.772	0.757	0.706	7
RE-SkN-Blasso	1.000	1.000	1.000	0.721	1.000	1.000	1.000	7

TABLE B.6: 90% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 1\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.343	0.279	0.362	0.242	0.803	0.688	0.002	6
EGARCH-Sk	0.477	0.279	0.345	0.347	0.803	0.768	0.011	6
RE-SkN-ML	0.343	0.279	0.362	0.347	0.686	0.696	0.158	7
RE-SkN-Lasso-CV	0.778	0.279	0.362	1.000	0.803	0.768	0.707	7
RE-SkN-MCMC	0.672	0.279	0.663	0.371	0.803	0.768	0.707	7
RE-SkN-Blasso	1.000	1.000	1.000	0.721	1.000	1.000	1.000	7

TABLE B.7: 75% Model Confidence Set of ALS with R method across the markets and assets, $\alpha = 1\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.236	0.206	0.373	0.408	0.786	0.735	0.003	4
EGARCH-Sk	0.588	0.206	0.373	0.408	0.786	0.761	0.018	5
RE-SkN-ML	0.236	0.206	0.373	0.408	0.703	0.761	0.172	4
RE-SkN-Lasso-CV	0.767	0.206	0.373	1.000	0.786	0.761	0.710	6
RE-SkN-MCMC	0.588	0.206	0.679	0.408	0.786	0.761	0.710	6
RE-SkN-Blasso	1.000	1.000	1.000	0.719	1.000	1.000	1.000	7
Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**

TABLE B.8: 75% Model Confidence Set of ALS with SQ method across the markets and assets, $\alpha = 1\%$

Model	**ASX 200**	**DAX**	**FTSE 100**	**HK**	**NASDAQ**	**SMI**	**S&P 500**	**Included in MCS**
GARCH-Sk	0.334	0.279	0.362	0.223	0.811	0.688	0.002	5
EGARCH-Sk	0.473	0.279	0.350	0.337	0.811	0.762	0.010	6
RE-SkN-ML	0.334	0.279	0.362	0.337	0.690	0.695	0.152	6
RE-SkN-Lasso-CV	0.767	0.279	0.362	1.000	0.811	0.762	0.708	7
RE-SkN-MCMC	0.661	0.279	0.679	0.371	0.811	0.762	0.708	7
RE-SkN-Blasso	1.000	1.000	1.000	0.719	1.000	1.000	1.000	7

Appendix C

Chapter 5 Appendix

C.1 Posterior Mean of the RE-STW Models

TABLE C.1: Estimated parameter mean of all RE-STW models for all forecasting steps for each parameter for ASX 200

Parameter	RE-STW models employing:							Average
	RV	RR	RK	RV, RR	RV, RK	RR, RK	RV, RR, RK	
μ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ω	-0.183	-0.177	-0.225	-0.176	-0.220	-0.209	-0.192	-0.197
β	0.980	0.981	0.976	0.981	0.977	0.978	0.980	0.979
γ_1	0.155	0.175	0.175	0.009	0.130	0.153	0.009	0.115
γ_2				0.149	0.111	0.096	0.138	0.123
γ_3							0.055	0.055
τ_1	-0.115	-0.111	-0.101	-0.105	-0.118	-0.114	-0.108	-0.110
τ_2	0.036	0.032	0.018	0.037	0.021	0.020	0.029	0.028
ξ_1	-1.460	-2.132	1.092	-1.160	-1.641	-2.212	-1.122	-1.234
ξ_2				-1.972	-0.076	-0.139	-1.884	-1.018
ξ_3							-0.448	-0.448
φ_1	0.940	0.920	1.156	0.974	0.922	0.911	0.976	0.971
φ_2				0.938	1.025	1.017	0.946	0.981
φ_3							0.985	0.985
$\delta_{1,1}$	-0.092	-0.087	-0.161	-0.094	-0.081	-0.078	-0.089	-0.098
$\delta_{1,2}$				-0.088	-0.174	-0.175	-0.084	-0.130
$\delta_{1,3}$							-0.176	-0.176
$\delta_{2,1}$	0.127	0.110	0.007	0.129	0.135	0.115	0.126	0.107
$\delta_{2,2}$				0.112	-0.022	-0.021	0.109	0.044
$\delta_{2,3}$							-0.008	-0.008
σ^2_{u1}	0.330	0.252	0.486	0.326	0.373	0.280	0.335	0.340
σ^2_{u2}				0.252	0.579	0.604	0.258	0.423
σ^2_{u3}							0.689	0.689
λ_1	0.781	0.774	0.701	0.799	0.773	0.784	0.787	0.771
k_1	1.276	1.271	1.197	1.698	1.603	1.293	1.284	1.375
k_2	1.366	1.364	1.288	1.795	1.759	1.543	1.568	1.526

TABLE C.2: Estimated parameter mean of all RE-STW models for all forecasting steps for each parameter for DAX

Parameter	RE-STW models							Average
	RV	**RR**	**RK**	**RV, RR**	**RV, RK**	**RR, RK**	**RV, RR, RK**	
μ	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
ω	-0.215	-0.219	-0.184	-0.213	-0.200	-0.199	-0.205	-0.205
β	0.976	0.975	0.979	0.976	0.977	0.977	0.977	0.977
γ_1	0.245	0.267	0.159	0.111	0.173	0.196	0.069	0.174
γ_2				0.153	0.055	0.053	0.128	0.097
γ_3							0.049	0.049
τ_1	-0.140	-0.135	-0.137	-0.135	-0.139	-0.139	-0.134	-0.137
τ_2	0.044	0.045	0.038	0.049	0.044	0.046	0.048	0.045
ξ_1	-0.389	-0.315	-0.220	-0.456	-0.523	-0.614	-0.428	-0.421
ξ_2				-0.519	-0.390	-0.440	-0.496	-0.461
ξ_3							-0.311	-0.311
φ_1	1.018	1.033	1.041	1.011	1.003	0.999	1.014	1.017
φ_2				1.011	1.022	1.017	1.013	1.016
φ_3							1.032	1.032
$\delta_{1,1}$	-0.170	-0.165	-0.132	-0.170	-0.170	-0.165	-0.168	-0.163
$\delta_{1,2}$				-0.165	-0.129	-0.129	-0.164	-0.147
$\delta_{1,3}$							-0.127	-0.127
$\delta_{2,1}$	0.077	0.074	0.129	0.080	0.076	0.073	0.079	0.084
$\delta_{2,2}$				0.073	0.131	0.137	0.073	0.103
$\delta_{2,3}$							0.135	0.135
σ^2_{u1}	0.186	0.161	0.308	0.187	0.186	0.161	0.186	0.196
σ^2_{u2}				0.162	0.306	0.308	0.162	0.234
σ^2_{u3}							0.306	0.306
λ_1	0.692	0.688	0.684	0.693	0.690	0.686	0.687	0.688
k_1	1.162	1.160	1.159	1.164	1.161	1.160	1.161	1.161
k_2	1.244	1.244	1.236	1.379	1.244	1.239	1.243	1.261

TABLE C.3: Estimated parameter mean of all RE-STW models for all forecasting steps for each parameter for FTSE 100

Parameter	RE-STW models							Average
	RV	**RR**	**RK**	**RV, RR**	**RV, RK**	**RR, RK**	**RV, RR, RK**	
μ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ω	-0.227	-0.257	-0.229	-0.249	-0.239	-0.258	-0.254	-0.245
β	0.975	0.972	0.975	0.972	0.974	0.971	0.972	0.973
γ_1	0.251	0.282	0.155	0.168	0.226	0.259	0.157	0.214
γ_2				0.103	0.014	0.009	0.098	0.056
γ_3							0.007	0.007
τ_1	-0.128	-0.130	-0.143	-0.131	-0.140	-0.141	-0.138	-0.136
τ_2	0.046	0.048	0.049	0.051	0.048	0.050	0.051	0.049
ξ_1	-0.116	-0.433	-0.097	-0.180	-0.234	-0.687	-0.260	-0.287
ξ_2				-0.706	-0.150	-0.095	-0.747	-0.425
ξ_3							-0.157	-0.157
φ_1	1.054	1.040	1.030	1.049	1.042	1.013	1.041	1.038
φ_2				1.012	1.025	1.031	1.008	1.019
φ_3							1.026	1.026
$\delta_{1,1}$	-0.146	-0.131	-0.075	-0.150	-0.148	-0.131	-0.151	-0.133
$\delta_{1,2}$				-0.132	-0.075	-0.076	-0.133	-0.104
$\delta_{1,3}$							-0.080	-0.080
$\delta_{2,1}$	0.078	0.078	0.195	0.080	0.078	0.077	0.080	0.095
$\delta_{2,2}$				0.078	0.196	0.197	0.079	0.137
$\delta_{2,3}$							0.200	0.200
σ^2_{u1}	0.159	0.133	0.299	0.160	0.160	0.133	0.160	0.172
σ^2_{u2}				0.133	0.292	0.292	0.133	0.212
σ^2_{u3}							0.292	0.292
λ_1	0.731	0.727	0.718	0.766	0.727	0.728	0.728	0.732
k_1	1.232	1.819	1.219	1.255	1.227	1.224	1.223	1.314
k_2	1.310	1.305	1.259	1.692	1.260	1.324	1.268	1.346

TABLE C.4: Estimated parameter mean of all RE-STW models for all forecasting steps for each parameter for Hang Seng

Parameter	RE-STW models							Average
	RV	**RR**	**RK**	**RV, RR**	**RV, RK**	**RR, RK**	**RV, RR, RK**	
μ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ω	-0.178	-0.194	-0.120	-0.226	-0.179	-0.206	-0.244	-0.192
β	0.979	0.977	0.986	0.973	0.979	0.976	0.971	0.977
γ_1	0.263	0.304	0.134	0.082	0.169	0.206	0.052	0.173
γ_2				0.187	0.053	0.041	0.136	0.104
γ_3							0.042	0.042
τ_1	-0.053	-0.055	-0.051	-0.049	-0.049	-0.051	-0.046	-0.050
τ_2	0.062	0.052	0.062	0.060	0.066	0.059	0.061	0.060
ξ_1	-1.600	-2.129	-1.605	-0.919	-1.497	-1.967	-0.568	-1.469
ξ_2				-1.892	-1.225	-0.999	-1.561	-1.419
ξ_3							-0.581	-0.581
φ_1	0.935	0.901	0.939	1.017	0.947	0.919	1.058	0.959
φ_2				0.928	0.985	1.014	0.966	0.973
φ_3							1.063	1.063
$\delta_{1,1}$	-0.112	-0.097	-0.101	-0.113	-0.113	-0.098	-0.114	-0.107
$\delta_{1,2}$				-0.099	-0.101	-0.100	-0.100	-0.100
$\delta_{1,3}$							-0.102	-0.102
$\delta_{2,1}$	0.062	0.049	0.114	0.068	0.061	0.047	0.065	0.066
$\delta_{2,2}$				0.047	0.117	0.125	0.045	0.083
$\delta_{2,3}$							0.122	0.122
σ^2_{u1}	0.208	0.152	0.449	0.214	0.209	0.157	0.209	0.228
σ^2_{u2}				0.157	0.458	0.498	0.164	0.319
σ^2_{u3}							0.485	0.485
λ_1	0.616	0.593	0.602	0.595	0.610	0.604	0.600	0.603
k_1	2.167	1.438	1.131	1.111	1.131	1.159	1.110	1.321
k_2	1.218	1.182	1.174	1.158	1.212	1.298	1.162	1.201

TABLE C.5: Estimated parameter mean of all RE-STW models for all forecasting steps for each parameter for NASDAQ

Parameter	RE-STW models							Average
	RV	**RR**	**RK**	**RV, RR**	**RV, RK**	**RR, RK**	**RV, RR, RK**	
μ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ω	-0.340	-0.361	-0.235	-0.371	-0.316	-0.333	-0.359	-0.331
β	0.962	0.959	0.974	0.958	0.964	0.962	0.960	0.963
γ_1	0.351	0.381	0.203	0.152	0.249	0.265	0.081	0.240
γ_2				0.214	0.071	0.080	0.175	0.135
γ_3							0.072	0.072
τ_1	-0.165	-0.166	-0.161	-0.159	-0.164	-0.166	-0.154	-0.162
τ_2	0.047	0.047	0.035	0.051	0.047	0.047	0.049	0.046
ξ_1	-1.379	-1.478	-1.576	-1.149	-1.392	-1.528	-0.879	-1.340
ξ_2				-1.322	-1.320	-1.285	-1.060	-1.247
ξ_3							-0.819	-0.819
φ_1	0.938	0.960	0.905	0.965	0.936	0.953	0.994	0.950
φ_2				0.977	0.935	0.939	1.005	0.964
φ_3							0.992	0.992
$\delta_{1,1}$	-0.159	-0.160	-0.138	-0.159	-0.159	-0.158	-0.158	-0.156
$\delta_{1,2}$				-0.158	-0.135	-0.135	-0.158	-0.147
$\delta_{1,3}$							-0.134	-0.134
$\delta_{2,1}$	0.044	0.042	0.098	0.045	0.044	0.040	0.043	0.051
$\delta_{2,2}$				0.041	0.100	0.101	0.039	0.070
$\delta_{2,3}$							0.099	0.099
σ^2_{u1}	0.194	0.176	0.355	0.193	0.193	0.175	0.192	0.211
σ^2_{u2}				0.175	0.349	0.350	0.175	0.262
σ^2_{u3}							0.350	0.350
λ_1	0.683	0.692	0.690	0.689	0.684	0.697	0.676	0.687
k_1	1.183	2.178	1.183	1.188	1.184	1.290	1.181	1.341
k_2	1.192	1.311	1.208	1.610	1.261	1.745	1.192	1.360

TABLE C.6: Estimated parameter mean of all RE-STW models for all forecasting steps for each parameter for SMI

Parameter	RE-STW models							Average
	RV	**RR**	**RK**	**RV, RR**	**RV, RK**	**RR, RK**	**RV, RR, RK**	
μ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ω	-0.416	-0.433	-0.626	-0.418	-0.459	-0.468	-0.433	-0.465
β	0.955	0.953	0.932	0.954	0.950	0.949	0.952	0.949
γ_1	0.313	0.353	0.186	0.095	0.268	0.308	0.081	0.229
γ_2				0.241	0.035	0.031	0.229	0.134
γ_3							0.020	0.020
τ_1	-0.131	-0.118	-0.148	-0.122	-0.133	-0.120	-0.124	-0.128
τ_2	0.042	0.042	0.037	0.044	0.042	0.042	0.044	0.042
ξ_1	-0.527	-0.784	-0.404	-0.278	-0.352	-0.595	-0.307	-0.464
ξ_2				-0.877	-1.249	-0.968	-0.774	-0.967
ξ_3							-0.751	-0.751
φ_1	1.013	0.979	0.965	1.044	1.033	1.000	1.042	1.011
φ_2				0.971	0.868	0.900	0.984	0.931
φ_3							0.926	0.926
$\delta_{1,1}$	-0.156	-0.128	-0.130	-0.159	-0.156	-0.129	-0.160	-0.145
$\delta_{1,2}$				-0.132	-0.136	-0.135	-0.133	-0.134
$\delta_{1,3}$							-0.139	-0.139
$\delta_{2,1}$	0.054	0.048	0.054	0.057	0.056	0.049	0.058	0.054
$\delta_{2,2}$				0.049	0.058	0.060	0.050	0.054
$\delta_{2,3}$							0.063	0.063
σ^2_{u1}	0.133	0.094	0.564	0.132	0.135	0.096	0.133	0.184
σ^2_{u2}				0.094	0.669	0.666	0.095	0.381
σ^2_{u3}							0.682	0.682
λ_1	0.725	0.717	0.672	0.716	0.725	0.717	0.712	0.712
k_1	1.333	1.247	1.156	1.180	1.224	1.205	1.173	1.217
k_2	1.260	1.236	1.180	1.261	1.627	1.465	1.235	1.323

TABLE C.7: Estimated parameter mean of all RE-STW models for all forecasting steps for each parameter for S&P500

Parameter	RE-STW models							Average
	RV	**RR**	**RK**	**RV, RR**	**RV, RK**	**RR, RK**	**RV, RR, RK**	
μ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ω	-0.321	-0.326	-0.269	-0.335	-0.332	-0.338	-0.339	-0.323
β	0.965	0.964	0.971	0.963	0.964	0.963	0.963	0.965
γ_1	0.354	0.376	0.148	0.195	0.287	0.310	0.153	0.260
γ_2				0.165	0.025	0.022	0.150	0.091
γ_3							0.029	0.029
τ_1	-0.191	-0.191	-0.182	-0.185	-0.193	-0.194	-0.185	-0.189
τ_2	0.043	0.045	0.037	0.048	0.039	0.042	0.045	0.043
ξ_1	-1.066	-1.148	-0.501	-0.930	-0.737	-0.827	-0.703	-0.845
ξ_2				-1.000	-0.636	-0.649	-0.778	-0.766
ξ_3							-0.620	-0.620
φ_1	0.957	0.974	1.012	0.973	0.993	1.009	0.998	0.988
φ_2				0.990	0.997	0.997	1.014	1.000
φ_3							1.000	1.000
$\delta_{1,1}$	-0.159	-0.157	-0.049	-0.159	-0.159	-0.157	-0.159	-0.143
$\delta_{1,2}$				-0.155	-0.051	-0.055	-0.155	-0.104
$\delta_{1,3}$							-0.049	-0.049
$\delta_{2,1}$	0.067	0.069	0.209	0.068	0.067	0.069	0.068	0.088
$\delta_{2,2}$				0.069	0.208	0.214	0.069	0.140
$\delta_{2,3}$							0.211	0.211
σ^2_{u1}	0.194	0.176	0.418	0.194	0.194	0.176	0.194	0.221
σ^2_{u2}				0.176	0.405	0.405	0.175	0.290
σ^2_{u3}							0.405	0.405
λ_1	0.666	0.680	0.653	0.674	0.665	0.696	0.663	0.671
k_1	1.155	2.174	1.136	1.149	1.264	2.337	1.145	1.480
k_2	1.160	1.250	1.158	1.509	1.175	1.574	1.151	1.283

Bibliography

Aït-sahali, Y., & Brandt, M. W. (2001). Variable selection for portfolio choice. *The Journal of Finance*, *56*(4), 1297–1351.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle,[w:] proceedings of the 2nd international symposium on information, bn petrow, f. *Czaki, Akademiai Kiado, Budapest*.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Akgiray, V., & Booth, G. G. (1988). The siable-law model of stock returns. *Journal of Business & Economic Statistics*, *6*(1), 51–57.

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*(3), 1047–1091.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2000). Great realizations. *Risk*, 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*(453), 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.

Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, *169*(1), 75–93.

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

Ardia, D., Baştürk, N., Hoogerheide, L., & van Dijk, H. K. (2012). A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Computational Statistics & Data Analysis*, *56*(11), 3398–3414.

Ardia, D., & Hoogerheide, L. F. (2010). Efficient Bayesian estimation and combination of GARCH-type models. In K. Bocker (Ed.), *Rethinking risk measurement and reporting: Examples and applications from finance*. RiskBooks.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, *4*, 40–79.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1997). Thinking coherently. risk, 10. *November, 68*, *71*.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Ausín, M. C., & Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, *51*(5), 2636–2652.

Bachelier, L. (1900). Theory of speculation. *Dimson, E. and M. Mussavian (1998), A brief history of market efficiency, European Financial Management*, *4*(1), 91–193.

Bae, K., & Mallick, B. K. (2004). Gene selection using a two-level hierarchical bayesian model. *Bioinformatics*, *20*(18), 3423–3430.

Bandi, F. M., & Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, *75*(2), 339–369.

Barndorff–Nielsen, O. E., Hansen, P. R., Lunde, A., & 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.

Barndorff–Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, *12*(3), C1–C32.

Barndorff–Nielsen, O. E., & 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.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, *2*(1), 1–37.

Basel Committee on Banking Supervision. (2012). Consultative document: Fundamental review of the trading book.

Basel Committee on Banking Supervision. (2016). Minimum capital requirements for market risk. *Consultative Document*.

Basel Committee on Banking Supervision. (2019). *Minimum capital requirements for market risk*. Bank for International Settlements. https://www.bis.org/bcbs/publ/d457.pdf

Bauwens, L., Laurent, S., & Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, *21*(1), 79–109.

Bauwens, L., & Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. *The Econometrics Journal*, *1*(1), 23–46.

Bayer, S., & Dimitriadis, T. (2018). Regression based expected shortfall backtesting. *arXiv preprint arXiv:1801.04112*.

Berger, J. O. (2013). *Statistical decision theory and bayesian analysis*. Springer Science & Business Media.

Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, *57*(12), 2213–2227.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 542–547.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. handbook of econometrics. *Vol. IV, North Holland*.

Brada, J., Ernst, H., & Tassel, J. V. (1966). Letter to the editor—the distribution of stock price differences: Gaussian after all? *Operations Research*, *14*(2), 334–340.

Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer Science & Business Media.

Casella, G. (2001). Empirical bayes gibbs sampling. *Biostatistics*, *2*(4), 485–500.

Cawley, G. C., & Talbot, N. L. (2006). Gene selection in cancer classification using sparse logistic regression with bayesian regularization. *Bioinformatics*, *22*(19), 2348–2355.

Chatterjee, S. (2014). A new perspective on least squares under convex constraint. *The Annals of Statistics*, *42*(6), 2340–2381.

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

Chen, C. W. S., Gerlach, R., & Lin, E. M. H. (2014). Bayesian estimation of smoothly mixing time-varying parameter GARCH models. *Computational Statistics & Data Analysis*, *76*, 194–209.

Chen, C. W. S., Gerlach, R., & So, M. K. P. (2006). Comparison of nonnested asymmetric heteroskedastic models. *Computational Statistics & Data Analysis*, *51*(4), 2164–2178.

Chen, C. W. S., Hsu, H.-Y., & Watanabe, T. (2023). Tail risk forecasting of realized volatility caviar models. *Finance Research Letters*, *51*, 103326.

Chen, C. W. S., Lin, E. M. H., & Huang, T. F. J. (2022). Bayesian quantile forecasting via the realized hysteretic garch model. *Journal of Forecasting*, *41*(7), 1317–1337.

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

Chen, C. W. S., & Watanabe, T. (2019a). Bayesian modeling and forecasting of value-at-risk via threshold realized volatility. *Applied Stochastic Models in Business and Industry*, *35*(3), 747–765.

Chen, C. W. S., & Watanabe, T. (2019b). Bayesian modeling and forecasting of value-at-risk via threshold realized volatility. *Applied Stochastic Models in Business and Industry*, *35*(3), 747–765.

Chen, C. W. S., Watanabe, T., & Lin, E. M. H. (2021). Bayesian estimation of realized garch-type models with application to financial tail risk management. *Econometrics and Statistics*.

Chen, Q., Gerlach, R., & Lu, Z. (2012). Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics & Data Analysis*, *56*(11), 3498–3516.

Chen, Q., & Gerlach, R. H. (2013). The two-sided Weibull distribution and forecasting financial tail risk. *International Journal of Forecasting*, *29*(4), 527–540.

Chetverikov, D., Liao, Z., & Chernozhukov, V. (2021). On cross-validated lasso in high dimensions. *The Annals of Statistics*, *49*(3), 1300–1317.

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*, *96*(453), 270–281.

Christensen, K., & Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, *141*(2), 323–349.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 841–862.

Contino, C., & Gerlach, R. H. (2017). Bayesian tail-risk forecasting using realized GARCH. *Applied Stochastic Models in Business and Industry*. https://doi.org/10.1002/asmb.2237

Couperier, O., & Leymarie, J. (2019). Backtesting expected shortfall via multi-quantile regression. (halshs-01909375).

Cowles, M. K. (2013). *Applied Bayesian statistics: With R and OpenBUGS examples* (Vol. 98). Springer Science & Business Media.

Dimitriadis, T., & Bayer, S. (2017). A joint quantile and expected shortfall regression framework. *arXiv preprint arXiv:1704.02213*.

Dumitrescu, E.-I., Hurlin, C., & Pham, V. (2012). Backtesting value-at-risk: From dynamic quantile to dynamic binary tests. *Finance*, *33*(1), 79–112.

DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, *1*(5), 948–957.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407–499.

Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, *3*(2), 30–41. https://doi.org/10.1080/10920277.1999.10595797

Emmer, S., Kratz, M., & Tasche, D. (2013). What is the best risk measure in practice? a comparison of standard measures. *IDEAS Working Paper Series from RePEc*.

Enders, W. (2014). *Applied econometric time series*. John Wiley & Sons.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

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

Engle, R. F., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, *131*(1), 3–27.

Engle, R. F., & Manganelli, S. (2001). Value at risk models in finance. *IDEAS Working Paper Series from RePEc*.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381. https://doi.org/10.1198/073500104000000370

Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 1127–1162.

Fama, E. F. (1963). Mandelbrot and the stable Paretian hypothesis. *The Journal of Business*, *36*(4), 420–429.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1), 34–105.

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

Fang, Y. (1996). *Volatility modeling and estimation of high-frequency data with Gaussian noise* (Thesis).

Fernández, C., & Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, *93*(441), 359–371.

Figueiredo, M. A. (2003). Adaptive sparseness for supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, *25*(9), 1150–1159.

Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, *44*(4), 1680–1707.

Friedl, H., Stampfer, E., El-Shaarawi, A. H., & Piegorsch, W. W. (2002). Cross-validation. *Encyclopedia of environmetrics*, *1*, 452–460.

Gaglianone, W. P., Lima, L. R., Linton, O., & Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, *29*(1), 150–160.

Gebizlioglu, O. L., Şenoğlu, B., & Kantar, Y. M. (2011). Comparison of certain value-at-risk estimation methods for the two-parameter Weibull loss distribution. *Journal of Computational and Applied Mathematics*, *235*(11), 3304–3314.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, *70*(350), 320–328.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd). CRC Press.

Gelman, A., Roberts, G. O., & Gilks, W. R. (1996). Efficient Metropolis jumping rules. *Bayesian Statistics*, *5*(599-608), 42.

Gerlach, R., Walpole, D., & Wang, C. (2017). Semi-parametric Bayesian tail risk forecasting incorporating realized measures of volatility. *Quantitative Finance*, *17*(2), 199–215.

Gerlach, R., & Wang, C. (2016). Forecasting risk via realized GARCH, incorporating the realized range. *Quantitative Finance*, *16*(4), 501–511.

Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics*, *23*(4), 416–431.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1996). *Markov chain Monte Carlo in practice*. Chapman; Hall/CRC.

Gilks, W. R., Roberts, G. O., & Sahu, S. K. (1998). Adaptive markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association*, *93*(443), 1045–1054.

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

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

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Haario, H., Saksman, E., & Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, *14*(3), 375–396.

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, *7*(2), 223–242.

Hans, C. (2009). Bayesian lasso regression. *Biometrika*, *96*(4), 835–845.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 705–730.

Hansen, P. R., & Huang, Z. (2016). Exponential GARCH modeling with realized measures of volatility. *Journal of Business & Economic Statistics*, *34*(2), 269–287. https://doi.org/10.1080/07350015.2015.1038543

Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, *24*(2), 127–161.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The modal confidence set. *Econometrica*, *79*(2), 453–497. https://doi.org/10.3982/ECTA5771

Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906. https://doi.org/10.1002/jae.1234

Harvey, C. R., & Siddique, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 465–487.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity. *Monographs on statistics and applied probability*, *143*, 143.

Hinkley, D. V., & Revankar, N. S. (1977). Estimation of the Pareto law from underreported data: A further analysis. *Journal of Econometrics*, *5*(1), 1–11.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hsu, D.-A., Miller, R. B., & Wichern, D. W. (1974). On the stable Paretian behavior of stock-market prices. *Journal of the American Statistical Association*, *69*(345), 108–113.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163–172.

Jin, X., & Maheu, J. M. (2012). Modeling realized covariances and returns. *Journal of Financial Econometrics*, *11*(2), 335–369.

Jorion, P. (1996). Risk2: Measuring the risk in value at risk. *Financial analysts journal*, *52*(6), 47–56.

Jorion, P. (2001). *Value at risk : The new benchmark for managing financial risk* (2nd. ed.) [Includes bibliographical references (p. 521-530) and index.]. McGraw-Hill.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*(2), 78–84.

Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, *5*(2), 369–411.

Lau, A. H.-L., Lau, H.-S., & Wingender, J. R. (1990). The distribution of stock returns: New evidence against the stable model. *Journal of Business & Economic Statistics*, *8*(2), 217–223.

Laurent, A. (1959). Letter to the editor-comments on "brownian motion in the stock market". *Operations Research*, *7*(6), 806–807.

Lee, S.-W., & Hansen, B. E. (1994). Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, *10*(1), 29–52.

Lei, J. (2020). Cross-validation with confidence. *Journal of the American Statistical Association*, *115*(532), 1978–1997.

Leng, C., Tran, M.-N., & Nott, D. (2014). Bayesian adaptive lasso. *Annals of the Institute of Statistical Mathematics*, *66*(2), 221–244.

Li, Q., & Lin, N. (2010). The bayesian elastic net. *Bayesian analysis*, *5*(1), 151–170.

Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates. *Economic Review*, *2*, 3–17.

Madan, D. B., & Seneta, E. (1990). The variance gamma (vg) model for share market returns. *Journal of Business*, 511–524.

Malevergne, Y., & Sornette, D. (2004). Value-at-risk-efficient portfolios for a class of super- and sub-exponentially decaying assets return distributions. *Quantitative Finance*, *4*(1), 17–36.

Mallick, H., & Yi, N. (2014). A new Bayesian lasso. *Statistics and its interface*, *7*(4), 571–582.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, *40*(4), 393–413.

Markowitz, H. (1952). Portfolio selection (vol. 7 (1)).

Martens, M., & van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, *138*(1), 181–207.

McNeil, A. J., & 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.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 141–183.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of Chemical Physics*, *21*(6), 1087–1092.

Mittnik, S., & Rachev, S. T. (1989). Stable distributions for asset returns. *Applied Mathematics Letters*, *2*(3), 301–304.

Mittnik, S., & Rachev, S. T. (1993). Modeling asset returns with alternative stable distributions. *Econometric Reviews*, *12*(3), 261–330.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Ntzoufras, I. (2011). *Bayesian modeling using winbugs* (Vol. 698). John Wiley & Sons.

Osborne, M. F. (1959). Brownian motion in the stock market. *Operations Research*, *7*(2), 145–173.

Osborne, M. R., Presnell, B., & Turlach, B. A. (2000). A new approach to variable selection in least squares problems. *IMA journal of numerical analysis*, *20*(3), 389–403.

Paap, R. (2002). What are the advantages of mcmc based inference in latent variable models? *Statistica Neerlandica*, *56*(1), 2–22.

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 61–65.

Praetz, P. D. (1972). The distribution of share price changes. *Journal of Business*, 49–55.

Robert, C. B., & Nicholas, J. G. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *The Journal of Business*, *47*(2), 244–280. https://doi.org/10.1086/295634

Robert, C. P. (2007). *The bayesian choice: From decision-theoretic foundations to computational implementation* (Vol. 2). Springer.

Roberts, G. O., Gelman, A., & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, *7*(1), 110–120. https://doi.org/10.1214/aoap/1034625254

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, *18*(2), 349–367.

Rolski, T., Schmidli, H., Schmidt, V., & Teugels, J. L. (2009). *Stochastic processes for insurance and finance* (Vol. 505). John Wiley & Sons.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, *88*(422), 486–494.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Shevade, S. K., & Keerthi, S. S. (2003). A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, *19*(17), 2246–2253.

Silahli, B., Dingec, K. D., Cifter, A., & Aydin, N. (2019). Portfolio value-at-risk with two-sided Weibull distribution: Evidence from cryptocurrency markets. *Finance Research Letters*, 101425.

Silvapulle, M. J., & Sen, P. K. (2011). *Constrained statistical inference: Order, inequality, and shape constraints* (Vol. 912). John Wiley & Sons.

Sornette, D., Simonetti, P., & Andersen, J. V. (2000). q-field theory for portfolio optimization:"fat tails" and nonlinear correlations. *Physics Reports*, *335*(2), 19–92.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, *36*(2), 111–133.

Takahashi, M., Omori, Y., & Watanabe, T. (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis*, *53*(6), 2404–2426.

Takahashi, M., Watanabe, T., & Omori, Y. (2016). Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution. *International Journal of Forecasting*, *32*(2), 437–457.

Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, *6*(2), 231–252. https://doi.org/10.1093/jjfinc/nbn001

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.

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

Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543). John Wiley & Sons.

Van de Schoot, R., & Miocević, M. (2020). *Small sample size solutions: A guide for applied researchers and practitioners*. Taylor & Francis.

Van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology*, *89*, 31–50.

Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, *22*(5), 997–1008.

Virbickaite, A., Ausin, M. C., & Galeano, P. (2015). Bayesian inference methods for univariate and multivariate GARCH models: A survey. *Journal of Economic Surveys*, *29*(1), 76–96. https://doi.org/10.1111/joes.12046

Wang, C., Chen, Q., & Gerlach, R. (2019). Bayesian realized-GARCH models for financial tail risk forecasting incorporating the two-sided Weibull distribution. *Quantitative Finance*, *19*(6), 1017–1042.

Wang, H., Li, G., & Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the lad-lasso. *Journal of Business & Economic Statistics*, *25*(3), 347–355.

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

Weibull, W. (1939a). The phenomenon of rupture in solids. *Swedish Institute of Engineering Research (Ingenioersvetenskaps Akad. Handl.)*, *153*.

Weibull, W. (1939b). A statistical theory of the strength of materials. *Swedish Institute of Engineering Research (Ingenioersvetenskaps Akad. Handl.)*, *151*.

Weibull, W. (1951). A statistical distribution function of wide applicability, asme. *Journal of Applied Mechanics*.

Wu, X., Xia, M., & Zhang, H. (2019). Forecasting var using realized EGARCH model with skewness and kurtosis. *Finance Research Letters*.

Wu, X., & Xie, H. (2021). A realized egarch-midas model with higher moments. *Finance Research Letters*, *38*, 101392.

Yamai, Y., & Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, *29*(4), 997–1015.

Yuan, M., & Lin, Y. (2005). Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, *100*(472), 1215–1225.

Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, *100*(472), 1394–1411.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429.