THE UNIVERSITY OF
SYDNEY

Doctoral Thesis

Bayesian parametric and semi-parametric financial tail-risk forecasting incorporating range and realized measures

Author:
Chao Wang

Supervisor:
Prof. Richard Gerlach

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

Discipline of Business Analytics
Business School
The University of Sydney

April 2016

Statement of Originality

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

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

Chao Wang

Abstract

Now we are in a world saturated with data and information, and numerous quantitative methods for financial risk management are proposed and used by many financial research institutions and organizations within recent years. Quantitative financial risk measurement is now a fundamental tool for investment decisions, capital allocation and external regulation. The Global Financial Crisis (GFC) has once again emphasized the importance of accurate risk measurement and prediction for financial organizations, which require accurate volatility estimation and forecasting.

The intra-day range has been frequently used in the literature and proven its superiority compared to return in volatility estimation and forecasting. Furthermore, high frequency econometrics has been gaining more popularity in the last decade and has developed into a major area in econometrics, driven by the increasing availability of high frequency data and algorithm-based high frequency trading in seconds or even milliseconds. The data recorded on a high frequency level contain much more information than the conventional daily financial data, and thus the volatility measures calculated based on high frequency data are much more efficient than the daily return and range.

In this thesis, we aim to develop a series of volatility and tail risk models employing intra-day and high frequency volatility measures. Firstly, the Realized GARCH framework is extended to incorporate the realized range, as potentially more efficient series of information than realized variance. Furthermore, we propose an innovative sub-sampled realized range and also adopt an existing scaling scheme, in order to deal with the micro-structure noise of the high frequency volatility measures. In addition, a Bayesian estimator is developed for the Realized GARCH type models, and presents favourable results compared to the frequentist estimator. Through empirical studies on various market indices that consider predictive likelihoods as well as 1% VaR and ES forecasting, results clearly indicate that the realized range and sub-sampled realized range in a Realized GARCH framework, with Student-t errors, lead to more accurate volatility and predictive density forecasts.

Further, a new framework called Realized Conditional Autoregressive Expectile (Realized CARE) is proposed, through incorporating a measurement equation into

the conventional CARE model, in a manner analogous to the Realized GARCH model. The intra-day range and realized measures (e.g. realized variance and realized range, etc.) are employed as the dependent variable in the measurement equation. The measurement equation here models the contemporaneous dependence between the realized measure and the latent conditional expectile. In addition, a targeted search based on a quadratic approximation is proposed, which improves the computational speed of estimation of the expectile level parameter. Bayesian adaptive Markov Chain Monte Carlo methods and likelihood-based frequentist methods are proposed for estimation, whilst their properties are compared via a simulation study. Furthermore, the methods of sub-sampling and scaling are applied to the realized variance and realized range, to help deal with the inherent micro-structure noise of the realized volatility measures. In a real forecasting study applied to 6 market indices and 3 individual assets, compared to the original CARE, the parametric GARCH and Realized GARCH models, one-day-ahead Value-at-Risk and Expected Shortfall forecasting results favor the proposed Realized CARE model, especially the Realized CARE model incorporating the realized range and the sub-sampled realized range.

Finally, we propose a new intra-day volatility estimator named signed range, which incorporates open, high and low prices for its calculation. A high frequency simulation study is conducted to analyze the relationship between signed range volatility and return volatility. An adaptive MCMC is developed for the parameter estimation and is compared with the maximum likelihood approach through simulation study. Then we propose the symmetric and asymmetric Conditional Autoregressive Signed Range (CARSR) type models, and the proposed models demonstrate their superiority compared to GARCH and Conditional Autoregressive Range (CARR) models in the 1% VaR and ES forecasting study.

Acknowledgements

On this occasion of finishing my PhD thesis, first of all I would like to sincerely thank my supervisor Professor Richard Gerlach. He introduced me to the field of Bayesian econometrics and gave me countless help and suggestions to both my research and life in Australia during the last 4 years. I really appreciate it. His profound knowledge, rigorous research style and unselfish attitude give me deep impression and have a profound impact on my future life.

I would like to thank Dr. Boris Choy, who gave me the precious opportunity to work with him in QBUS5001 as a tutor. He provided me with numerous instructions during the last 6 semesters, and it is my great pleasure and honor to work with him. His professional attitude towards teaching and students set up a role model to my future tutoring and teaching work.

I would like to thank all the staff in the discipline of Business Analytics. They are all really kind to me and gave me many suggestions and numerous help during my PhD study.

I would like to thank my dear wife Rongrong Hou. She married to me without any hesitation when I was in the hardest time of my life. Over the years, she took care of my daily life and helped me go through all the hard time. I am greatly indebted to her. All I can do is trying my best in the future life to give her a better life.

I would like to thank my dear father-in-law and mother-in-law. They raise such an excellent and understanding girl and gave me full support and understanding during the last few years.

I would like to thank my dear parents. They gave me life, and they provide me with sufficient material and spiritual support so that I can study in such a nice university and country. Their countless calls and numerous care really support me to get through some confusion and disappointment time during my study in Australia. I wish them happiness and health forever, and I will try my best to be a good son.

Chao Wang

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Financial risk measurement and forecasting

In recent decades, quantitative financial risk measurement has become a fundamental tool for investment decisions, capital allocation and external regulation. Steinherr (1998) proposed that risk management is one of most important innovations of the 20th century. From the perspective of financial risk, risk refers to an action that will adversely affect a company's capability of achieving its goals and implementing its business strategies, and risk can be quantified or expressed as the probability of loss or much less than expected returns (Dorfman, 2007). In addition, quantitative financial risk measurement has become a fundamental tool for investment decisions, capital allocation and external regulation, etc (Engle and Manganelli, 2004). Furthermore, the Global Financial Crisis (GFC) in 2008 has once again emphasized the importance of accurate risk measurement and prediction for financial organizations. After the famous 'Black Monday' in 1987, on which a major stock market crash happened, the G-10 group agreed to set up and form the original Basel Capital Accord in 1988, in order to better control financial risk of financial institutions and protect them from unexpected financial losses (Chen and Gerlach, 2013). However, the financial crisis continued to occur in the 1990s, such as Orange Country lost 1.6 billion dollars in 1994, the Barings Bank's 1.4 billion money loss in 1995, etc (Chen *et al.*, 2011). All these financial crises and losses forced the market regulators and financial market risk assessment institutions to establish a new benchmark for financial risk measurement. Then

Value-at-Risk was introduced and established in 1993 by JP Morgan to describe and measure daily financial risk.

1.2 Value-at-Risk and Expected Shortfall

The 2008 GFC challenged market participators' risk management abilities and brought into question of financial risk management methods and practice. More and more worldwide financial institutions and corporations now employ Value-at-Risk (VaR) to assist their decisions on capital allocation and risk management. The G30 Group published a report named "Derivative Products Practices and Rules" in 1993 and proposed VaR as a measurement tool to evaluate the market risk. J.P. Morgan introduced and presented VaR in the RiskMetrics model at 1993, as a part of the "Weatherstone 4:15pm" daily risk assessment report (Jorion, 1996). Value-at-Risk is a quantitative tool to measure and control financial risk. It represents the market risk as one number and has become a standard measurement for capital allocation and risk management of financial institutions. There are three basic approaches to compute VaR: non-parametric; semi-parametric and parametric, although there are numerous variations within each approach.

As a commonly used financial risk measurement, Value-at-Risk (VaR) summarizes risk through a single number. VaR refers to: under the normal fluctuation conditions in the market, the maximum possible loss of a financial asset or portfolio at probability level α. More precisely, VaR means the maximum possible loss of a financial asset or portfolio value within a specific period of time in the future, under a certain probability (confidence level) (Jorion, 1996). From perspective of statistics, VaR is just a number, shows the value under risk facing a "normal" market volatility state. It illustrates the amount of the maximum expected loss (which can be an absolute value, may also be a relative value), within a given confidence level and a certain holding period. For example, for a combination of securities held by an investment company within the next 24 hours, under the confidence level of 95%, in the case of normal fluctuations in the securities markets, the VaR value is 1 million dollars. It means the probability of the company portfolio's maximum loss being more than 1 million dollars is 5% in one day (24 hours), due to changes in the market price, which demonstrates this situation is likely to emerge once on average every 20 trading days. In another way, we have 95% certainty to say that the loss of this investment company within the next

trading day is less than 1 million dollars. 5% reflect the risk aversion level of the managers of financial assets, which can be decided according to different investors' degree of risk appetite and affordability.

However, VaR has been criticised, because VaR only considers the distribution quantile and ignores the extreme loss beyond the VaR level. Expected Shortfall (ES), which was proposed by Artzner *et al.* (1997, 1999), gives the expected loss conditional on returns exceeding a VaR threshold. There, by definition, ES does give the expected loss (magnitude) conditional on exceeding a VaR threshold and is coherent and has been used widely for tail risk measurement. Both VaR and ES are recommended tail risk measures in the Basel III Capital Accord, thus they are both included in the thesis as the tail risk measures.

1.3 Parametric and non-parametric volatility modelling

Volatility is an important characteristic of financial time series (Engle, 1982), thus a key aspect of parametric VaR or ES measurement is return volatility estimation. Statistically speaking, volatility equals to the square root of conditional variance for the given past:

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

Where F_{t-1} stands for the available information before time t. For the financial time series, there is sufficient empirical evidence to show that the financial returns are fat-tailed and negatively skewed, and has conditional heteroscedastic property. Thus the traditional mean and variance time series model cannot describe and capture daily assets returns properly. Models of conditional heteroskedasticity for time series have become an increasingly hot research field in today's financial risk management field. Engle (1982) proposed the most popular and well known volatility models: autoregressive conditionally heteroscedastic (ARCH) model, and Bollerslev (1986) presented the generalized ARCH (GARCH) model. The GARCH (1,1) is specified as Model (1.1). These models consider heteroscedastic property of the financial time series in the model forecasting update equation. Till recently, ARCH and GARCH models are recognized as one of the most important class for

financial time series analysis as it has the ability to capture the commonly observed change in variance of the observed stock index or exchange rate over time.

$$\begin{aligned} r_t &= \mu + \sigma_t \varepsilon_t = \mu + a_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{1.1}$$

where r_t stands for the return for day t, $\varepsilon_t \overset{\text{i.i.d.}}{\sim} D_1(0,1)$. α_1 is the news parameter and β_1 is the decay parameter. The parameters in this model should satisfy $\alpha_0 > 0$, $\alpha_1 > 0$ and $\beta_1 > 0$ to guarantee that $\sigma_t^2 > 0$. $\alpha_1 + \beta_1 < 1$ is required to ensure that σ_t is stationary.

Ghysels *et al.* (1996) pointed out some important features of volatility which play a crucial role in predicting model construction and selection, such as heavy tails, asymmetric effect, volatility clustering. Besides, numerical works show that the shape of daily return distribution is fat-tailed and skewed. Thus the Gaussian distribution cannot describe the financial time series conditional return distribution properly. In order to capture the daily return distribution's fat-tail and skewness property, different GARCH models are applied with diverse error distributions to overcome these problems, such as Student-t, Two-sided Weibull (Chen and Gerlach, 2013), Asymmetric Laplace (Chen *et al.*, 2011). The Exponential GARCH model (Nelson, 1991) and GJR-GARCH model (Glosten *et al.*, 1993) were proposed in order to capture the well-known leverage effect. Furthermore, substantial variants of GARCH volatility models are proposed, such as Integrated GARCH (IGARCH, Engle and Bollerslev (1986)), Threshold GARCH (TGARCH, Zakoian (1994)) and Double Threshold GARCH (DTGARCH, Li and Li (1996)).

However, the performance of parametric GARCH type models heavily depend on the choice of error distribution. A semi-parametric model named Conditional Autoregressive Expectile (CARE) was proposed by Taylor (2008). The expectile can be estimated with Asymmetric Least Square (ALS), then it is transformed into ES through a connection between expectile and ES (Newey and Powell, 1987). Gerlach, Chen and Lin (2012) developed non-linear family of the CARE model and the Bayesian estimation framework. Further, Gerlach and Chen (2016) extended

the CARE type models with employing daily high-low range as an input.

1.4 Intra-day and high frequency volatility measures

The high-low range has been known for a long time and recently was frequently used as a latent volatility estimator. Based on the assumption that return follows a Brownian motion with zero drift during the day, Parkinson (1980) derived the relationship between range and return volatility estimator and proved that range is a much more efficient and less noisy volatility estimator than return. Garman and Klass (1980) proposed a volatility estimator that incorporated high, low, open and close prices, which is even more efficient than range. In addition, an estimator which allows for arbitrary drift was devised by Rogers and Satchell (1991), and Yang and Zhang (2000) derived another drift-independent estimator.

Through a proper dynamic structure of the conditional mean of range, Chou (2005) proposed the conditional autoregressive range (CARR) model, which successfully demonstrated range's superiority compared with return in empirical volatility forecasting. Brandt and Jones (2006) formulated the range-based EGARCH model which is analogous to the EGARCH (Nelson, 1991) models, while used the square root of high-low range to replace the absolute return. The advantages of range encourage me to test the range related models and also add range into the GARCH type model to improve the volatility estimation performance. In addition, Chen *et al.* (2008) proposed the volatility forecasting using range-based threshold conditional autoregressive model (TARR), which allows us to capture size and sign asymmetry through a nonlinear specification.

Since nowadays the high frequency and ultra high frequency tick by tick data are available in a number of databases, a voluminous literature has discussed about various realized volatility measures. In particular, realized variance (RV) is rapidly gaining popularity for estimating daily volatility, and it is basically the sum of squared returns over non-overlapping intervals within a sampling period (Andersen *et al.*, 2003, and Barndorff-Nielsen and Shephard 2002). Furthermore, considering the superiority of range with respect to return, a volatility estimator named realized range (RR) was proposed by Martens and van Dijk (2007) and Christensen and Podolskij (2007). Empirical analysis demonstrated the potential and superiority of the realized range. The relative efficiency between realized range

and realized variance are identical to the result of range square and return square, which means RR is 5 times more efficient than RV theoretically.

1.5 Bayesian inference

Maximum likelihood (ML) is quite commonly used for the parameter estimation of the GARCH type model, and is included in most statistical software packages. Nevertheless, parameter estimation with ML has certain drawbacks. Specifically, in ML estimation of GARCH type models, constrained estimation is normally required to ensure the models' stationarity (e.g. $\alpha_1 + \beta_1 < 1$ for Model (1.1)), and this can cause problems in the parameters estimation and the standard error calculation (see e.g. Silvapulle and Sen, 2004). This problem can be even severe when the model becomes more complex, e.g. the Realized GARCH framework (Hansen *et al.*, 2011). Therefore, the Bayesian methodology is employed for parameter estimation in this thesis work.

Generally, Bayesian decision theory is a fundamental method in statistical model of decision-making, and its the basic idea and procedure is:
a. Calculate the conditional probability density expression and select a prior probability;
b. Use the Bayesian formula to convert the conditional and prior probability into posterior probability;
c. Make classification decisions according to the result of the posterior probability.

Basically, the Bayesian model is a probability model that consists of a likelihood function and a prior distribution. Firstly, denoting samples of size t as $\mathbf{y} = (y_1, y_2, \ldots, y_t)$, and m parameters need to be estimated as $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$, the Bayes' rule is presented in Equation (1.2). $P(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution. $P(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function, and $P(\boldsymbol{\theta})$ is the prior distribution. In Bayesian inference, we start with the prior distribution for parameter $\boldsymbol{\theta}$ that will be estimated. Before seeing the data, the prior distribution demonstrates our degree of belief about unobservable parameter $\boldsymbol{\theta}$. Then our degree of belief of parameter $\boldsymbol{\theta}$ can be updated through Bayesian calculation (posterior probability) after seeing the data.

$$P(\boldsymbol{\theta}|\mathbf{y}) = \frac{P(\boldsymbol{\theta})P(\mathbf{y}|\boldsymbol{\theta})}{P(\mathbf{y})} \propto P(\boldsymbol{\theta})P(\mathbf{y}|\boldsymbol{\theta}) \tag{1.2}$$

Inference and estimation of Bayesian approach may require advanced Bayesian computation methods. Markov Chain Monte Carlo (MCMC) scheme is a Bayesian inference method that is well established and widely used. MCMC firstly constructs a Markov Chain that contains a sequence of unobservable parameters $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \boldsymbol{\theta}_n$. In MCMC scheme, the next state $\boldsymbol{\theta}_{n+1}$ is sampled from one step ahead conditional probability $P(\boldsymbol{\theta}_{n+1}|\boldsymbol{\theta}_n)$, which forms a Markov Chain. After running MCMC sampling scheme for required iterations, the estimated parameter $\boldsymbol{\theta}$ can be approximated by the posterior mean (Gilks, 2005).

1.5.1 The Metropolis algorithm

Introduced by Metropolis *et al.* (1953), the Metropolis algorithm works quite efficiently, while it requires a symmetric proposal distribution, meaning $g(\theta_a|\theta_b) = g(\theta_b|\theta_a)$. The process of Metropolis algorithm is:
a. Choose $\boldsymbol{\theta}_1$ as the starting value of the algorithm, which may be a choice based on previous experience or just a random guess, or a random draw from a particular distribution),
b. then for $i = 1, 2, \ldots, N$, draw a candidate sample $\boldsymbol{\theta}^*$ (or called proposed sample) from a symmetric proposal distribution $g(\boldsymbol{\theta})$,
c. Calculate the acceptance probability as:

$$A^i = min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{i-1}|\mathbf{y})}\right\} \tag{1.3}$$

d. Draw a threshold value B^i from a uniform distribution $[0, 1]$, and compare A^i and B^i. If $A^i > B^i$, accept proposed sample and set $\boldsymbol{\theta}^i = \boldsymbol{\theta}^*$; otherwise, discard the proposed sample and set $\boldsymbol{\theta}^i = \boldsymbol{\theta}^{i-1}$
e. The N proposed and accepted samples $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^N$ will converge to the target distribution $p(\boldsymbol{\theta}|\mathbf{y})$, and their average can be then used as the parameter estimates.

1.5.2 The Metropolis Hastings algorithm

The basic Metropolis algorithm requires the proposal distribution to be symmetric, while Hastings (1970) proposed an improvement on this technique and shows that

proposal distribution g does not need to be symmetric. Therefore, the Metropolis Hastings (MH) algorithm is generalised version of the Metropolis algorithm and is used more frequently in reality since the less restrictions on the proposal distribution.

The implementation process of the MH algorithm is similar to that of the Metropolis algorithm, but the acceptance probability is slightly modified. Specifically, step c. in the Metropolis algorithm is modified for the Metropolis Hastings algorithm as following:

$$A^i = min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})g(\boldsymbol{\theta})^{i-1}}{p(\boldsymbol{\theta}^{i-1}|\mathbf{y})g(\boldsymbol{\theta}^*)}\right\} \tag{1.4}$$

As can be seen, the proposed distribution density value with proposal samples and accepted samples are included in the acceptance probability calculation. In addition, the acceptance rate, which is ratio between the number of accepted proposed sample and the total number of proposed samples, is an important metric to evaluate the quality of the Metropolis or MH algorithms, because it measures the suitability of the proposal distribution. Specifically, a MH algorithm with a acceptance close to 1 can either indicate the proposal distribution is perfectly close to the target distribution or indicate the variance of the proposal distribution is too low, and cannot capture the tails of the target distribution $p(\boldsymbol{\theta}|\mathbf{y})$. On the other hand, if the acceptance rate is close to 0, meaning most of the proposed samples are discarded, the proposed distribution might be inappropriate or its variance might be too large.

1.5.3 Adaptive MCMC algorithm

In this thesis work, we employ an adaptive MCMC method, adapted from that in Contino and Gerlach (2014), extended from work originally in Chen and So (2006). This algorithm is a two step process. In step 1, also called the burn-in period, a Metropolis algorithm employing a Gaussian proposal distribution, with a random walk (RM) mean vector, is utilized for parameters $\boldsymbol{\theta}$ to be estimated. In each iteration of RW Metropolis, the candidate sample is generated from the random walk kernel, $\boldsymbol{\theta}^* = \boldsymbol{\theta} + \epsilon$, where ϵ follows a proposal distribution with zero mean and changing variance/covariance. The variance/covariance is subsequently tuned, aiming towards a target acceptance rate of 23.4% (Roberts, Gelman and

Gilks, 1997). As discussed in Section 1.5.1, step c. decides whether the proposal samples are accepted. Through the RWM process, we will get the adjusted variance/covariance and mean of the burnin samples, as initial values of next step of the algorithm.

In step 2 (the MCMC sampling period), a mixture of three Gaussians proposal distribution is employed in an "independent" Metropolis-Hastings (IMH) algorithm. Proposal samples are drawn from a proposal distribution, and the sample is accepted or rejected with the probability in Equation (1.4). Afterwards we use the sample mean as the estimation value of the parameter. After running the loop for certain number of iterations, we will get the estimation results of each parameter. The adaptive MCMC algorithm specifications will be discussed in each chapter separately.

1.5.4 MCMC convergence and efficiency testing

In this thesis, we employ the Gelman-Rubin diagnostic (Gelman *et al.*, 2014) to diagnose the convergence of the adapted MCMC method. Further, an effective sample size testing is incorporated to evaluate the efficiency of the MCMC (Gelman *et al.*, 2014). Firstly, The Gelman-Rubin statistics is calculated as below (for each parameter):

a. Run m MCMC chains, each with length N. For example, for each of the 8 parameters in the Realized-GARCH-Gaussian-Gaussian frame, we run the adaptive MCMC for $m = 5$ times (each run with random starting values) with burn-in RWM iterations $15,000$ and IMH iterations $n_1 = 5,000$.

b. Then we only use the IMH iterations $n_1 = 5,000$ in each chain for Gelman-Rubin statistics calculation, since the MCMC posterior mean and RMSE is calculated based on the IMH results.

c. Supposing $\psi_{i,j}$ is the i-th sample in j-th chain ($i = 1, \ldots, n_1$; $j = 1, \ldots, m$), we can calculate the between-chain (B) and within-chain (W) variance as below (similar to a classical ANOVA):

$$B = \frac{n_1}{m-1}\sum_{j=1}^{m}(\overline{\psi}_{.j} - \overline{\psi}_{..})^2, \tag{1.5}$$

where $\overline{\psi}_{.j} = \frac{1}{n_1}\sum_{i=1}^{n_1}\psi_{i,j}$ and $\overline{\psi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\overline{\psi}_{.j}$. Then

$$W = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{n_1-1}\sum_{i=1}^{n_1}(\psi_{i,j} - \overline{\psi}_{.j})^2\right]. \tag{1.6}$$

d. Compute the estimated marginal posterior variance $\hat{Var}(\theta)$ for each parameter as a weighted sum of B and W: $\frac{n_1-1}{n_1}W + \frac{1}{n_1}B$.

e. The Gelman-Rubin statistics (potential scale reduction factor) can be calculated as:

$$\hat{R} = \sqrt{\frac{\hat{Var}(\theta)}{W}} \tag{1.7}$$

The potential scale reduction factor needs to be calculated separately for each parameter. Intuitively speaking, the convergence of MCMC chain is good when the chains have "forgotten" their starting values, and the output iterations from different chains are indistinguishable. Statistically when interpreting the Gelman-Rubin diagnostic results, Values of $\hat{R}$ close 1 suggest convergence, while high value of $\hat{R}$ (probbaly greater than 1.1 or 1.2), then we should run the MCMC chains longer to improve convergence to the stationary distribution.

After the diagnostic of the convergence of the MCMC iterations, we can compute an approximate "effective number of independent simulation draws". In order to calculate the effective sample size, we need an estimate of the sum of the correlations ρ (refer to Gelman *et al.*, 2014, page 286-287 for details), which is computed based on between and within chains information. Firstly, we calculate the variogram V_t at each lag t:

$$W = \frac{1}{m(n_1-t)}\sum_{j=1}^{m}\sum_{i=t+1}^{n}(\psi_{i,j} - \psi_{i-t,j})^2 \tag{1.8}$$

Then the correlation estimate is computed as:

$$\hat{\rho}_t = 1 - \frac{V_t}{2\hat{Var}(\theta)}, \tag{1.9}$$

here $\hat{Var}(\theta)$ is the estimated marginal posterior variance in Gelman-Rubin diagnostic. However, since at large values of lag t the sample correlation is quite noisy, we cannot sum all $\hat{\rho}_t$ values to calculate effective sample size. Instead, Gelman *et al.* (2014) computed a partial sum, starting from lag 0 and continuing until the sum of autocorrelation estimates for two continuous correlation estimates is negative. Then the effective sample size is:

$$\hat{n}_{eff} = \frac{mn_1}{1 + 2\sum_{t=1}^{T} \hat{\rho}_t} \tag{1.10}$$

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

1.6 Tail risk forecast calculation and assessment

Both Value-at-Risk (VaR) and Expected Shortfall (ES) are recommended tail risk measures in the Basel III Capital Accord. ES is defined as the expected value of an r.v. Y, conditional on Y being more extreme than its α-level quantile: i.e. $\text{ES}_\alpha = E(Y|Y < Q_\alpha)$, where Q_α is the quantile of Y. Value-at-Risk is here defined as the α-level quantile of Y: Q_α. Here we consider only $\alpha < 0.5$ and thus restrict this work to left-tail or negative risk on long positions, as is standard in the literature.

We employ Gaussian and Student-t distributions in the proposed models for the volatility estimation in this thesis work, and the VaR and ES calculations are calculated differently based on different distributions. For a Gaussian distribution, the VaR forecasts at give confidence level $\alpha = 1\%$ can be calculated as:

$$\text{VaR}_{\alpha,t} = \sigma_t \Phi^{-1}(\alpha),\ t = 1, 2, \ldots m, \tag{1.11}$$

where Φ^{-1} is the inverse of standardized Gaussian distribution's CDF, σ_t is conditional volatility estimated and m stands for the forecasting step. The VaR with standardized t distribution is calculated as below:

$$\text{VaR}_{\alpha,t} = \sigma_t t_\nu^{-1}(\alpha)\sqrt{\frac{\nu-2}{\nu}},\ t = 1, 2, \ldots m, \tag{1.12}$$

where t^{-1} is the inverse of Student-t's CDF with the ν degree of freedom estimated.

The calculation process of ES with Gaussian and t distributions are:

$$\mathrm{ES}_{\alpha,t} = -\sigma_t \frac{\phi(\Phi^{-1}(\alpha))}{\alpha},\ t = 1, 2, \ldots m, \tag{1.13}$$

where $\phi()$ is the pdf of standard Gaussian distribution.

$$\mathrm{ES}_{\alpha,t} = -\sigma_t \left(\frac{g_\nu(t_\nu^{-1}(\alpha))}{\alpha} \right) \left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu - 1} \right) \sqrt{\frac{\nu - 2}{\nu}},\ t = 1, 2, \ldots m, \tag{1.14}$$

where g_ν is the Student-t pdf (McNeil *et al.*, 2005).

While various common tests can be applied to directly assess VaR quantile forecasts: e.g. the unconditional coverage (UC) and conditional coverage (CC) tests of Kupiec (1995) and Christoffersen(1998) respectively, as well as the dynamic quantile (DQ) test of Engle and Manganelli (2004) and the VQR test of Gaglianone *et al.* (2011), proper or optimal assessment of a set of ES forecasts is still an issue under investigation. The most common method applied to assess ES forecasts is based on the fact that it is a conditional expectation beyond a VaR quantile; an aspect which can be tested directly or indirectly. The direct test examines the residuals, observations minus forecast ES level, for data that are violations, i.e. more extreme than the corresponding VaR predictions, and tests whether these residuals have mean 0. Since the ES predictions are usually not independent over time, the residuals are often scaled by predicted volatility, e.g. see McNeil *et al.* (2000), or by the predicted VaR levels, as in Taylor (2008).

Following Kerkhof and Melenberg (2004), Chen, Gerlach and Lu (2012) illustrated how to treat ES forecasts as quantile forecasts in parametric models, where the quantile level that ES falls at can be deduced exactly. Gerlach and Chen (2016) further presented that across a range of non-Gaussian distributions, when applied to real daily financial return data, the quantile where the 1% ES is estimated to fall was $\approx 0.35\%$. Nominal levels for ES with various distributions are presented in Table 1.1. Their approaches are followed to assess and test ES forecasts here, treating them as quantile forecasts at appropriate quantile levels, as discussed in Gerlach and Chen (2016), and applying the UC, CC, DQ and VQR tests.

1.6.1 UC and CC tests

The testing process of unconditional coverage (UC) and conditional coverage (CC) tests of Kupiec (1995) and Christoffersen(1998) are presented now. The unconditional coverage test is a likelihood ratio test. Supposing in the $\alpha = 1\%$ VaR

TABLE 1.1: Nominal levels for ES for the Gaussian, Student-t, AL and TW distributions

		δ_α		
α	$N(0,1)$	$t^*(10)$	$t^*(6)$	$t^*(4)$
0.01	0.0038	0.0036	0.0034	0.0032
$ES_{0.01}$	-2.665	-3.008	-3.293	-3.692
	$Sk-t^*(6)$	$Sk-t^*(4)$	AL	TW
0.01	0.0034	0.0032	0.0037	0.0037
$ES_{0.01}$	-3.365	-3.857	-3.544	-3.433

Note: AL is the Asymmetric Laplace distribution specified in Chen, Gerlach, and Lu (2012). TW is the two-sided Weibull distribution of Malevergne and Sornette (2004); δ_α is independent of the single parameter of the AL; the TW δ_α is based on the data employed in Gerlach and Chen (2016).

forecasting empirical study with T steps, we have the number of violations as m_1, and the number of non-violations as m_0. Therefore, the violation rate π under such condition will be $\frac{m_1}{m_1+m_0}$. The UC log likelihood ratio test statistics under the null (of a correct unconditional coverage level) and the alternative hypothesis are:

$$\ell_{UC,H_0} = m_0 * log(1-\alpha) + m_1 log(\alpha)$$
$$\ell_{UC,H_1} = m_0 * log(1-\pi) + m_1 log(\pi)$$

The UC log likelihood ratio test statistics is asymptotically χ_1^2 and is calculated as below:

$$\ell_{UC,H_0} = 2(\ell_{UC,H_1} - \ell_{UC,H_0}) \overset{n\to\infty}{\to} \chi_1^2$$

In order to derive the log likelihood ratio of conditional coverage, a test for independence comparing two different models for the probability of a violation at time t needs to be established. This independence test introduces two Bernoulli process, supposing one is independent with parameter p, and one is a two state Markov dependent process with parameter δ_{ij}. The log likelihood for the independent model is:

$$\ell_{ind,H_0} = log(p^{m_1}(1-p)^{T-m_1})$$

With the two-state process dependent model, the likelihood of a violation given whether or not a violation occurred in the last period is calculated. Denoting $m_{ij} = \sum_{t=1}^{T}(Hit_{t-1} = i)(Hit_t = j)$, $\hat{\delta}_{ij} = \frac{m_{ij}}{m_{i0}+m_{i1}}$, and $\hat{p} = \frac{m_{01}+m_{11}}{T}$. Then the log likelihood of the dependent model is:

$$\ell_{ind,H_1} = log[\delta_{01}^{m_{01}}(1-\delta_{01}^{m_{00}})\delta_{11}^{m_{11}}(1-\delta_{11}^{m_{10}})]$$

Now the log likelihood ratio under the null of independence of the violation process can be calculated as:

$$LR_{ind} = 2(\ell_{ind,H_1} - \ell_{ind,H_0}) \overset{n\to\infty}{\to} \chi_1^2$$

Finally, the conditional coverage test statistics (the joint test of coverage and independence) is given by:

$$LR_{CC} = 2(\ell_{ind,H_1} - \ell_{UC,H_0}) \overset{n\to\infty}{\to} \chi_2^2$$

1.6.2 DQ test

The dynamic quantile (DQ) test was proposed by Engle and Manganelli (2004). Both the in-sample and out-of-sample tests are proposed in their paper, while we only focus on the out-of-sample approach. To begin with, still supposing VaR forecasts are generated at given confidence level α, variable H_t with zero expectation is constructed as below:

$$H_t = I(r_t < VaR_t) - \alpha,$$

where $I(r_t < VaR_t)$ indicates that it equals to 1 when $r_t < VaR_t$.

Then an artificial regression is built:

$$H_t = \delta_0 + \delta_1 H_{t-1} + \ldots + \delta_k H_{t-k} + \delta_{k+1} VaR_t + \mu_t$$

$$H_t = \mu_t + X\delta = \begin{cases} -\alpha & , \quad P = 1-\alpha \\ 1-\alpha & , \quad P = \alpha \end{cases}$$

The reason of building this artificial regression is H_t must be uncorrelated with constant μ_t, its lagged values (k is the number of lags) and the VaR forecasts in

order to make H_t have the correct unconditional and conditional coverage, and be uncorrelated. Then we can have the OLS estimator $\hat{\delta}_{OLS}$ of this regression as:

$$\hat{\delta}_{OLS} = (X^{'}X)^{-1}X'H \overset{n\to\infty}{\to} N(0, \alpha(1-\alpha)(X^{'}X)^{-1})$$

Finally, with $H_0 : \delta = 0$, the DQ test statistics is given by:

$$DQ = \frac{\hat{\delta}'_{OLS}X^{'}X\hat{\delta}_{OLS}}{\alpha(1-\alpha)} \overset{n\to\infty}{\to} \chi^2_{k+2}$$

The authors recommend the use of 1 and 4 lags for the DQ test, while Chen, Gerlach and Lu (2012) tested and found that little sensitivity of the DQ test statistics to the number of lags selected.

1.6.3 VQR test

Based on the definition of VaR, we can treat it as the α quantile of the returns r_t, implied by $P(r_t < \text{VaR}_t|F_{t-1}) = \alpha$.

Proposed by Gaglianone *et al.* (2011), VQR tests the following regression model employing α conditional quantile as dependent variable and VaR as independent variable:

$$Q_{r_t}(\alpha|F_{t-1}) = \beta_0 + \beta_1 VaR_t$$

Then the following null hypothesis $\beta_0 = 0$ and $\beta_1 = 1$ is established to test the how the VaR fit the data, or the H_0 can be represented as: $\theta = 0$, where $\theta = [\beta_0 \ \beta_1]'$. The null hypothesis should be interpreted as a Mincer and Zarnowitz (1969) type-regression framework.

Let $\hat{\theta}$ be the quantile regression estimator of θ, Gaglione *et al.* (2011) derives the asymptotic distribution of θ as:

$$\sqrt{T}(\hat{\theta} - \theta) \sim N(0, \alpha(1-\alpha)H_\alpha^{-1}JH_\alpha^{-1})$$

where J and H are defined as:

$$J = p \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t x_t^{'}, \; x_t = [1 \; \text{VaR}_t]$$

$$H_\alpha = p \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t x_t^{'} [f_t(Q_{r_t}(\alpha|F_{t-1}))]$$

here the term $f_t(Q_{r_t}(\alpha|F_{t-1}))$ stands for the the conditional density of r_t at the quantile level α.

Now the test statistics of VQR testing is defined as:

$$VQR = T\big[\hat{\theta}^{'}\big(\alpha(1-\alpha)H_\alpha^{-1}JH_\alpha^{-1}\big)^{-1}\hat{\theta}\big] \overset{n\to\infty}{\to} \chi_2^2$$

1.7 Structure of the thesis

Chapter 1 briefly discusses the financial risk management, volatility modelling, Bayesian inference and tail risk forecast calculation and assessment.

Chapter 2 incorporates the realized range volatility estimator into the Realized GARCH framework. To help deal with the inherent micro-structure noise of the realized volatility measures, an existing scaling procedures is employed to account for the impact of micro-structure noise on realized range and realized variance, and the methods of sub-sampling are proposed to be applied on the realized range. A Bayesian adaptive Markov Chain Monte Carlo method is developed and employed for estimation and forecasting, and demonstrates its superiority in simulation study. The proposed realized range GARCH is studied with tail risk forecasting experiment across different indices.

A new tail risk forecasting framework named Realized Conditional Auotregressive Expectile (Realized-CARE) is proposed in Chapter 3, through incorporating a measurement equation into the conventional CARE model, in a manner analogous to the Realized GARCH model. In addition, a targeted search based on a quadratic approximation is proposed that improves the computational speed of estimation of the expectile level parameter. Bayesian adaptive Markov Chain Monte Carlo methods and likelihood-based frequentist methods are proposed for estimation, whilst their properties are compared via a simulation study. In a real forecasting

study applied to 6 market indices and 3 individual assets, the performance of the Realized-CARE model incorporating various realized measures are studied and compared with the original CARE, the parametric GARCH and Realized GARCH models.

In Chapter 4, we propose a new volatility estimator named signed range. Through incorporating open, high and low prices, the proposed signed range possesses the characteristics of both return and high-low range. Further, the relationship between signed range volatility and return volatility are studied through high frequency simulation study. The symmetric and asymmetric Conditional Autoregressive Signed Range (CARSR) type models are proposed and tested.

Chapter 5 concludes the thesis and discusses about the future work.

1.8 Chapter Summary

This chapter is a review of financial risk management and volatility modelling in the literature. Both parametric and non-parametric volatility models are discussed, and the intra-day and high frequency volatility measured are reviewed. Regarding the parameters estimation of volatility models, maximum likelihood and Bayesian approaches are considered and discussed.

Chapter 2

Bayesian tail-risk forecasting with Realized GARCH employing the realized range and scaled & sub-sampled realized measures

This chapter is an extended version of paper "Gerlach and Wang, 2016: forecasting risk via realized GARCH, incorporating the realized range", Quantitative Finance, 16(4), 501-511.

2.1 Introduction

The 2008 Global Financial Crisis (GFC) challenged market participators' risk management abilities and brought the concern on the efficiency of financial risk management methods and practice. It is also acknowledged that accurate volatility estimation and forecasting can significantly facilitate the risk management practice. Modern market risk measurement and management incorporates tail risk measures: e.g. Value-at-Risk (VaR), pioneered by JP Morgan in 1993, and conditional VaR, or expected shortfall (ES), proposed by Artzner *et al.* (1997, 1999). VaR is the maximum loss expected on an investment, over a given time period at a specific quantile level and is an important regulatory tool, recommended by the Basel Committee on Banking Supervision in Basel II, to control the risk of financial institutions, by helping to set minimum capital requirements to protect against large unexpected losses. VaR has been criticised, as it does not measure the magnitude of the loss for violating returns and Artzner *et al.* (1999) found

that VaR is not a 'coherent' measure: i.e. it is not sub-additive; VaR can (sometimes) lead to portfolio concentration rather than diversification. ES does give the expected loss (magnitude) conditional on exceeding a VaR threshold and is coherent and has been incorporated into the Basel III Capital Accord. Thus, both tail risk measures, VaR and ES, are considered here.

Among the volatility estimation models, the Autoregressive Conditional Heteroskedasticity model (ARCH) and Generalized ARCH (GARCH) gained high popularity in the recent years, proposed by Engle (1982) and Bollerslev (1986) respectively. Numerous GARCH type models had been developed during the past decades. Especially, GJR-GARCH (Glosten *et al.* 1993) and EGARCH (Nelson, 1991) were introduced to capture and describe the well known leverage effect (Black, 1976).

The standard GARCH type models only employ daily returns for the daily volatility estimation, and in recent years, various volatility estimators were proposed and applied to improve the volatility estimation, such as range and realized measures. High-low range has been proven to be a much more efficient and less noisy volatility estimator compared to return, see Parkinson (1980), Garman and Klass (1980), and Alizadeh, Brandt, and Diebold (2002). The Conditional Autoregressive Range model (Chou 2005) and Range-based EGARCH model (Brandt and Jones, 2006) demonstrated the superiority of range.

Furthermore, since nowadays the high frequency and ultra high frequency tick by tick data is available in a number of databases, a voluminous literature has discussed various realized volatility measures, including realized variance (RV) and realized kernel (RK), etc, see Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), Andersen *et.al* (2003) and Barndorff-Nielsen *et.al* (2004), (2008). These realized measures consider the tick by tick intra-day process, thus they are more informative than the daily return and provide a more accurate volatility estimation. Hansen *et.al* (2011) introduced a new framework named Realized GARCH (Re-GARCH), which adds a measurement equation compared to the conventional GARCH. An important feature of the Realized GARCH is the measurement equation which captures the joint dependence between volatility and realized measures. The Realized GARCH is also closely related to the multiplicative error model (Engle and Gallo, 2006) and the HEAVY model (Shephard and Sheppard, 2010). Hansen and Huang (2016) extended Realized GARCH into the Realized EGARCH framework, and allowed the Realized EGARCH model to have more than one measurement equation. Takahashi, Omori and Watanabe (2009)

proposed the realized stochastic volatility model and Takahashi, Watanabe and Omori (2016) extended it with generalized hyperbolic distribution. Different RV measures with various intra-day frequency, RK and their combinations were evaluated through both in-sample and out-of-sample testing. Based on their results, the Realized GARCH and Realized EGARCH dominate the conventional GARCH and EGARCH models. Realized EGARCH with 2 measurement equations (use daily range square and RK respectively) gave the best out-of-sample result, and the results generated from RV and RK are quite similar. Watanabe(2012) showed that RK does not present improved results compared with RV in the Realized GARCH framework, which demonstrates that Realized GARCH has the potential of correcting the RV bias caused by microstructure noise.

In addition, Martens and van Dijk (2007) and Christensen and Podolskij (2007) extended the daily high-low range with the high frequency intra-day data and proposed the realized range (RR), which replaces every square return terms of realized variance with scaled squared range (scaling factor 4log(2) presented by Parkinson in 1980). Through both simulation and experiments with real-world data sets, they proved that realized range has much lower mean squared error than realized variance. Martens and van Dijk (2007) also presented and discussed a few different bias-correction approaches for realized range. However, although RR was already shown to be an efficient return variance estimator, the literature has much less work on RR than RV.

In this chapter, we propose a new framework named Realized Range GARCH (RR-GARCH), which incorporates realized range into the Realized GARCH by Hansen *et.al* (2011), where both Gaussian and Student-t errors are considered for the observation equation. The bias-correction process by Martens and van Dijk (2007) is employed to get the scaled RV and scaled RR, and they are also included in the Realized GARCH model (form Scaled-RV-GARCH and Scaled-RR-GARCH respectively). Also, we develop a sub-sampled RR estimator, which is inspired by the sub-sampled-RV (Zhang, Mykland and Aït-Sahalia, 2005), and they are both incorporated in the Realized GARCH framework (form sub-sampled-RV-GARCH and sub-sampled-RR-GARCH respectively). Finally, the daily range square Realized GARCH (Ra-Realized-GARCH) is also proposed and tested in this work. Further, the MCMC estimation methods in Contino and Gerlach (2014) are extended to estimate these models. The sampling properties of this estimator

are compared to that for the usual maximum likelihood (ML) estimator, via a simulation study, highlighting favourable performance for the MCMC estimator.

To assess the benefits of the proposed extended class of Realized GARCH models, accuracy in terms of predictive likelihood and tail risk forecasting will be assessed, and compared across a range of competing models, for five international market index return series. The tail risk forecast combination methods of Chang *et al.* (2011) and McAleer *et al.* (2013) are also incorporated into this study.

This chapter is structured as follows: Section 2.2 briefly reviews modern volatility estimators, including Ra, RV and RR. The specifications for RR-GARCH and other Re-GARCH type models are briefly presented in Section 2.3. Section 2.4 discusses parameter estimation via Bayesian MCMC. Section 2.5 presents some simulation results comparing Bayesian and maximum likelihood estimation for Realized GARCH models. Section 2.6 describes the data and presents the results of the empirical study. Section 2.7 concludes, and discusses possible future work.

2.2 Realized measures

This section gives a brief introduction to various volatility estimators included in the models employed in this chapter. First, for day t denote the intra-day high, low and closing prices as H_t, L_t and C_t. The daily log return is then:

$$r_t = \log(C_t) - \log(C_{t-1}) \tag{2.1}$$

Assuming the mean return is zero, as standard, a constant daily return variance can be estimated by:

$$V = \frac{1}{n}\sum_{t=1}^{n} r_t^2 \tag{2.2}$$

Based on the distribution of range derived by Feller(1951), Parkinson (1980) proposed the high-low intra-day range (squared), with scaling factor 4log(2) as an approximately unbiased variance estimator:

$$Ra_t^2 = \frac{(\log H_t - \log L_t)^2}{4\log 2} \tag{2.3}$$

Through theoretical derivation and a simulation study, Parkinson showed that this is a more efficient estimator than the traditional squared return. Garman-Klass

(1980), Rogers and Satchell (1991) and Yang and Zhang (2000) derived other range based estimators; a full study and comparison on the properties of different volatility estimators is presented in Molnár (2012).

Gerlach and Chen (2015) incorporated overnight price movements into this measure, defining range plus overnight as:

$$RaO_t = (\log(\max(C_{t-1}, H_t)) - \log(\min(C_{t-1}, L_t))) \times 100 \ , \tag{2.4}$$

where C_{t-1} is the closing price on day $t-1$.

Extending into the high frequency intra-day framework, each day t can be divided into N equally sized intervals of length $\triangle$, each intra-day time subscripted as $i = 0, 1, 2, ..., N$. The log closing price at the *i-th* interval of day t is denoted $P_{t-1+i\triangle}$. Then, the high and low prices during this time interval are $H_{t,i} = \sup_{(i-1)\triangle<j<i\triangle} P_{t-1+j}$ and $L_{t,i} = \inf_{(i-1)\triangle<j<i\triangle} P_{t-1+j}$ respectively. Realized variance (RV) has proven an efficient volatility estimator and gained popularity in recent years. RV is simply the sum of the N intra-day squared returns, at frequency $\triangle$, for day t, i.e.:

$$RV_t^{\triangle} = \sum_{i=1}^{N} [log(P_{t-1+i\triangle}) - log(P_{t-1+(i-1)\triangle})]^2 \tag{2.5}$$

Proposed by Barndorff-Nielsen and Shephard (2002), the realized kernel is a more robust volatility estimator compared to realized variance, especially when the returns are contaminated with micro-structure noise.

The Realized Range (RR), proposed by Martens and van Dijk (2007) and Christensen and Podolskij (2007), has the following specification, which simply replaces the intra-day squared returns with intra-day squared ranges, and scales:

$$RR_t^{\triangle} = \frac{\sum_{i=1}^{N} (\log H_{t,i} - \log L_{t,i})^2}{4 \log 2} \tag{2.6}$$

Theoretically, the RR may contain more information about volatility, in the same way as the intra-day range contains more information than squared returns: it uses all the price movements in a time period to form the high and low price, not just the price at each end of each time period. Results in Martens and van Dijk (2007) lend support to this hypothesis. Only when $N \to \infty$, the scaling factor $4 \log 2$ makes the RR as an unbiased volatility estimator.

Of course, both RV and RR have been criticized as being subject to micro-structure noise bias and inefficiency, more so than daily returns or daily ranges. This issue has been studied extensively, see Rogers and Satchell (1991), Barndorff-Nielsen *et al.* (2004) and Christensen and Podolskij (2007) for discussion. In response, Martens and van Dijk (2007) presented a scaling process, as in Equations (2.7) and (2.8).

$$ScRV_t^{\triangle} = \frac{\sum_{l=1}^{q} RV_{t-l}}{\sum_{l=1}^{q} RV_{t-l}^{\triangle}} RV_t^{\triangle}, \tag{2.7}$$

$$ScRR_t^{\triangle} = \frac{\sum_{l=1}^{q} RR_{t-l}}{\sum_{l=1}^{q} RR_{t-l}^{\triangle}} RR_t^{\triangle}, \tag{2.8}$$

where RV_{t-1} and RR_{t-1} represent the daily return square and range square at day $t-1$. This scaling process is motivated by the fact that the daily return and range are less affected by micro-structure noise and thus can be used to help reduce bias. Recently jumps in returns have also attracted attention in the analysis of high-frequency data (Andersen, Bollerslev and Diebold, 2007), but are not tackled in this thesis.

Further, Zhang, Mykland and Aït-Sahalia (2005) proposed a sub-sampled process to further smooth out micro-structure noise. For day t, N equally sized samples are grouped into M non-overlapping subsets $X^{(m)}$ with size $N/M = n_k$, which means:

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

Then sub-sampling will be implemented on the subsets $X^{(i)}$ with n_k interval:

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

Representing the log closing price at the *i-th* interval of day t as $C_{t,i} = P_{t-1+i\triangle}$, the RV with the subsets X^i is:

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

We have the T/M RV with T/N sub-sampling for day t as (supposing there are T minutes per trading day):

$$SSRV^{\triangle}_{t,T/M,T/N} = \frac{\sum_{i=0}^{n_k-1} RV_{t,i}}{n_k}, \tag{2.9}$$

Then, denoting the high and low prices during the interval $i+n_k(m-1)$ and $i+n_km$ as $H_{t,i} = \sup_{(i+n_k(m-1))\triangle<j<(i+n_km)\triangle} P_{t-1+j}$ and $L_{t,i} = \inf_{(i+n_k(m-1))\triangle<j<(i+n_km)\triangle} P_{t-1+j}$ respectively, we propose the T/M RR with T/N sub-sampling as:

$$RR_{t,i} = \sum_{m=1}^{M}(H_{t,i} - L_{t,i})^2; \text{ where } i = 0, 1, 2..., n_k - 1. \tag{2.10}$$

$$SSRR^{\triangle}_{t,T/M,T/N} = \frac{\sum_{i=0}^{n_k-1} RR_{t,i}}{4\log 2 n_k}, \tag{2.11}$$

For example, the 5 mins RV and RR with 1 min sub-sampling for any day can be calculated, respectively, as below :

$$\begin{aligned}
RV_{5,1,0} &= (\log C_{t5} - \log C_{t0})^2 + (\log C_{t10} - \log C_{t5})^2 + ... \\
RV_{5,1,1} &= (\log C_{t6} - \log C_{t1})^2 + (\log C_{t11} - \log C_{t6})^2 + ... \\
&\vdots \\
RV_{5,1,4} &= (\log C_{t9} - \log C_{t4})^2 + (\log C_{t14} - \log C_{t9})^2 + ... \\
SSRV^{\triangle}_{5,1} &= \frac{\sum_{i=0}^{4} RV_{5,1,i}}{5}
\end{aligned}$$

$$\begin{aligned}
RR_{5,1,0} &= (\log H_{t0<t<t5} - \log L_{t0<t<t5})^2 + (\log H_{t5<t<t10} - \log L_{t5<t<t10})^2 + ... \\
RR_{5,1,1} &= (\log H_{t1<t<t6} - \log L_{t1<t<t6})^2 + (\log H_{t6<t<t11} - \log L_{t6<t<t11})^2 + ... \\
&\vdots \\
RR_{5,1,4} &= (\log H_{t4<t<t9} - \log L_{t4<t<t9})^2 + (\log H_{t9<t<t14} - \log L_{t9<t<t14})^2 + ... \\
SSRR^{\triangle}_{5,1} &= \frac{\sum_{i=0}^{4} RR_{5,1,i}}{4\log(2)5}
\end{aligned}$$

Only intra-day returns on the 5 minute frequency, additionally with 1 minute sub-sampling when employed, are considered in this thesis work.

2.3 Realized Range GARCH

This section reviews the literature on Realized GARCH and proposes the Realized Range GARCH model.

2.3.1 Model description

The Realized GARCH model of Hansen *et al.* (2011) can be written as:

$$\begin{aligned} r_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \beta\sigma_{t-1}^2 + \gamma x_{t-1} \ , \\ x_t &= \xi + \varphi\sigma_t^2 + \tau_1 z_t + \tau_2(z_t^2 - 1) + \sigma_\varepsilon \varepsilon_t \ , \end{aligned} \tag{2.12}$$

where $r_t = [\log(C_t) - \log(C_{t-1})] \times 100$ is the percentage log-return for day t, $z_t \overset{\text{i.i.d.}}{\sim} D_1(0,1)$ and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} D_2(0,1)$ and x_t is a realized measure, e.g. RV; $D_1(0,1), D_2(0,1)$ indicate distributions that have mean 0 and variance 1. The three equations in Model (2.12) are, in order: the *return equation*, the *volatility equation* and the *measurement equation*, respectively. The measurement equation is a second observation equation that captures the contemporaneous dependence between latent volatility and the realized measure. The term $\tau_1 z_t + \tau_2(z_t^2 - 1)$ is used to capture a leverage-type effect.

Hansen *et al.* (2011) utilized the RV (among others) as the realized measure (i.e. x_t) in Model (2.12); and chose Gaussian errors, i.e. $D_1(0,1) = D_2(0,1) \equiv N(0,1)$. Watanabe (2012) allowed $D_1(0,1)$ to be a standardized Student-t and skew Student-t distribution of Fernández and Steel (1998); Contino and Gerlach (2014) allowed it to be the skewed-t of Hansen (1994) and also allowed $D_2(0,1)$ to be a standardized Student-t.

The following RG specifications are proposed in this thesis:

Realized Range GARCH (RR-RG): $x_t = RR_t^{\triangle}$
Range-squared Realized GARCH model (Ra-RG): $x_t = Ra_t^2$
Range Overnight-squared Realized GARCH model (RaO-RG): $x_t = RaO_t^2$
Scaled Realized Variance GARCH (ScRV-RG): $x_t = ScRV_t^{\triangle}$
Scaled Realized Range GARCH (ScRR-RG): $x_t = ScRR_t^{\triangle}$

Sub-sampled Realized Variance GARCH (SSRV-RG): $x_t = SSRV_t^{\triangle}$
Sub-sampled Realized Range GARCH (SSRV-RG): $x_t = SSRR_t^{\triangle}$

As discussed in Section 2.2, the scaling factor $4\log 2$ makes the RR as an unbiased volatility estimator, only when $N \to \infty$. Although in the empirical study we employ the 5 minute intra-day frequency, meaning N is finite and Equation (2.6) is biased to calculate realized range, RR in the Re-GARCH or Re-CARE (proposed in Chapter 3) models is not required to be unbiased, because the coefficient of the RR in the model can adjust such bias. Therefore, this is another advantage of using the Realized GARCH model with the realized-range based volatility as a realized measure.

2.3.2 Stationarity and positivity

Stationarity is an important issue in time series modelling in general. In this context it is important to understand the conditions or parameter restrictions required so that the long-run unconditional variance exists and is positive, as well as sufficient conditions ensuring each σ_t^2 is also positive.

Substituting the measurement equation into the volatility equation in 2.12 leads to:

$$\sigma_t^2 = (\omega + \gamma\xi) + (\beta + \gamma\varphi)\sigma_{t-1}^2 + a_t, \tag{2.13}$$

where $a_t = \gamma[\tau_1 z_{t-1} + \tau_2(z_{t-1}^2 - 1) + \varepsilon_{t-1}]$, so that $\mathrm{E}(a_t) = 0$. Taking expectations of both sides of (2.13), the long-run variance is $(\omega + \gamma\xi)/[1 - (\beta + \gamma\varphi)]$. To ensure this is finite and positive, the required conditions for the general Realized GARCH model are:

$$\begin{aligned} &\omega + \gamma\xi > 0, \\ &0 < \beta + \gamma\varphi < 1 \end{aligned} \tag{2.14}$$

Further, to ensure positivity of each σ_t^2, it is sufficient that ω, β, γ are all positive. This set of conditions are subsequently enforced during estimation of all Realized GARCH models in this chapter.

2.4 Bayesian and likelihood estimation

The model specification for the general Realized GARCH is in (2.12).

2.4.1 Likelihood

Following Hansen *et al.* (2011), where $D_1 = D_2 \equiv N(0,1)$, the log-likelihood function for model (2.12) is:

$$\ell(r,x;\theta) = \underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[log(2\pi) + log(\sigma_t^2) + r_t^2/\sigma_t^2\right]}_{\ell(r;\theta)} \underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[log(2\pi) + log(\sigma_\varepsilon^2) + \varepsilon_t^2/\sigma_\varepsilon^2\right]}_{\ell(x|r;\theta)} \tag{2.15}$$

where $\varepsilon_t = x_t - \xi - \varphi\sigma_t^2 - \tau_1 z_t - \tau_2(z_t^2 - 1)$; the parameter vector to be estimated is $\theta = (\omega, \beta, \gamma, \xi, \varphi, \tau_1, \tau_2, \sigma_\varepsilon)'$, under the constraints in (2.14) and positivity on (ω, β, γ). Hansen *et al.* (2011) derived the 1st and 2nd derivative of this log-likelihood function, allowing calculation of asymptotic standard errors of estimation, via a Hessian matrix. Subsequently, this model is denoted RG-GG (Realized GARCH with Gaussian-Gaussian errors).

Under the choice $D_1 \sim t^*(0,1,\nu)$; $D_2 \equiv N(0,1)$, as in Watanabe (2012) and Contino and Gerlach (2014), the log-likelihood function for model (2.12) is now:

$$\begin{aligned}\ell(r,x;\theta) \;=\; & \underbrace{-\sum_{t=1}^{n}\left[A(\nu) + 0.5\log(\sigma_t^2) + \frac{\nu+1}{2}(1 + \frac{r_t^2}{\sigma_t^2(\nu-2)})\right]}_{\ell(r;\theta)} \\ & \underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[log(2\pi) + log(\sigma_\varepsilon^2) + \varepsilon_t^2/\sigma_\varepsilon^2\right]}_{\ell(x|r;\theta)}\end{aligned} \tag{2.16}$$

where $\varepsilon_t = x_t - \xi - \varphi\sigma_t^2 - \tau_1 z_t - \tau_2(z_t^2 - 1)$ and $t^*(0,1) \equiv t(0,1,\nu) \times \sqrt{\frac{\nu-2}{\nu}}$, which is a Student-t distribution with ν degrees of freedom, scaled to have variance 1; and $A(\nu) = \log(\Gamma\left(\frac{\nu+1}{2}\right)) - \log(\Gamma\left(\frac{\nu}{2}\right)) + \log(\pi(\nu-2))$. The parameter vector to be estimated is now $\boldsymbol{\theta} = (\omega, \beta, \gamma, \nu, \xi, \varphi, \tau_1, \tau_2, \sigma_\varepsilon)'$, under the constraints in (2.14) and positivity on (ω, β, γ); further we restrict $\nu > 4$ to ensure the first four moments of the error distribution are finite. Subsequently, this model is denoted RG-tG.

2.4.2 Bayesian estimation

The likelihoods in (2.15) and (2.16) involve 8 and 9 unknown parameters respectively; most of which are part of equations involving latent, unobserved variables. The performance and finite sample properties of ML estimates of these likelihoods are not yet well studied. As such, we also consider powerful numerical and computational algorithms in a Bayesian framework, under weak or uninformative priors, as a competing estimator for these models.

2.4.2.1 Priors

The prior is chosen to be close to uninformative over the possible stationarity and positivity region for the model parameters $\boldsymbol{\theta}$, with two exceptions. We add a Jeffreys prior for the scale parameter σ_ε in the measurement equation, and also a Jeffreys-type prior for the intercept parameter ξ in this equation, i.e.:

$$\pi(\boldsymbol{\theta}) \propto I(A)\frac{1}{\sigma_\varepsilon}\frac{1}{\xi},$$

for the RG-GG model, and

$$\pi(\boldsymbol{\theta}) \propto I(A_2)\frac{1}{\sigma_\varepsilon}\frac{1}{\xi}\frac{1}{\nu^2},$$

for the RG-tG model. This is a mostly flat prior on the parameters in $\boldsymbol{\theta}$, restricted by the indicator function being non-zero only over the region A (or A_2), where A is the region defined by (2.14) plus positivity for ω, β, γ and A_2 is A intersected with $\nu > 4$. For the degrees of freedom parameter ν, the prior is equivalent to a uniform prior on $\nu^{-1} \sim \text{Unif}(0, 0.25)$, as used by Chen, Gerlach and So (2006), among others.

2.4.2.2 Adaptive MCMC

An adaptive MCMC method, adapted from that in Contino and Gerlach (2014), is employed, extended from work originally in Chen and So (2006). For the burn-in period, a Metropolis algorithm employing a Gaussian proposal distribution, with a random walk mean vector, is utilised for each block of parameters. The var-cov matrix of each block is initially set to $\frac{2.38}{\sqrt{(d_i)}} I_{d_i}$, where d_i is the dimension of the block (i) of parameters being generated, and I_{d_i} is the identity matrix of

dimension d_i. This covariance matrix is subsequently tuned, aiming towards a target acceptance rate of 23.4% (if $d_i > 1$, or 40% if $d_i = 1$), as standard, via the algorithm of Roberts, Gelman and Gilks (1997).

During the MCMC sampling period, a mixture of three Gaussian proposal distribution is employed in an "independent" Metropolis-Hastings algorithm. The mean vector for each block is the sample mean of the last 50% of the burn-in iterates for that block; i.e. it is the same for each of the three mixture elements. The proposal var-cov matrix in each element is $C_i\Sigma$, where $C_1 = 1; C_2 = 10; C_3 = 100$ and Σ is the sample covariance matrix of the last 50% of the burn-in iterates for that block.

As an example, for the RG-GG model, two blocks were employed: $\boldsymbol{\theta_1} = (\omega, \beta, \gamma, \varphi)'$ and $\boldsymbol{\theta_2} = (\xi, \tau_1, \tau_2, \sigma)'$ via motivations that parameters within the same equation are likely to be more correlated in the posterior (likelihood) than those in separate equations and the sampling scheme will mix faster when highly correlated parameters are generated together, with the exception that the stationarity condition may cause correlation between iterates of β, γ, φ, thus they are kept together. For the RG-tG model a third block containing only ν^{-1} was added, with $\boldsymbol{\theta_1}, \boldsymbol{\theta_2}$ remaining unchanged.

2.5 Simulation study

A simulation study is now presented to illustrate the comparative performance of the MCMC and ML estimators, in terms of parameter estimation, quantile and expected shortfall forecasting, accuracy. The aim is to illustrate the bias and precision properties for these two methods, highlighting the comparative performance of the MCMC estimator. The results presented focus on the RG-GG and RG-tG model specifications.

Samples of size $n = 1500$ and $n = 3000$ are simulated from two specific models, specified as:

$$
\begin{aligned}
\textbf{Model 1} \quad & r_t = \sigma_t z_t, \;\; z_t \sim N(0,1) \\
& \sigma_t^2 = 0.02 + 0.75\sigma_{t-1}^2 + 0.25x_{t-1} \;, \\
& x_t = 0.1 + 0.95\sigma_t^2 + 0.1z_t - 0.1(z_t^2 - 1) + \varepsilon_t \\
& \varepsilon_t \sim N(0, 0.5^2) \\
\textbf{Model 2} \quad & r_t = \sigma_t z_t, \;\; z_t \sim t_8^*(0,1) \\
& \sigma_t^2 = 0.01 + 0.7\sigma_{t-1}^2 + 0.29x_{t-1} \;, \\
& x_t = 0.01 + 0.99\sigma_t^2 + 0.25z_t - 0.25(z_t^2 - 1) + \varepsilon_t \\
& \varepsilon_t \sim N(0, 2^2)
\end{aligned}
$$

In each model r_t is analogous to a daily log-return and x_t is analogous to the daily realized measure. The persistence level $(\beta + \gamma\varphi)$ is deliberately chosen very close to 1 in each case; with true values chosen close to those estimated from real data. Here, t^* represents the Student-t distribution, standardised to have variance 1. For each model the forecast α-level quantile is then $q_\alpha(r_{t+1}|\boldsymbol{\theta}) = \sigma_{t+1}\Phi^{-1}(\alpha)$ (Model 1), where Φ^{-1} is the inverse standard Gaussian cdf, and $q_\alpha(r_{t+1}|\boldsymbol{\theta}) = \sigma_{t+1}t_\nu^{-1}(\alpha)$ (Model 2), where t_ν^{-1} is the inverse standardised Student-t cdf. Following Basel II and Basel III risk management guidelines, quantile levels of $\alpha = 0.01, 0.05$ are considered.

A total of 5000 replicated datasets are simulated from model 1 and from model 2, for each sample size $n = 1500, 3000$. The RG-GG model is fit to each dataset from Model 1, once using the MCMC method and once using the ML estimator, the latter employing the 'fmincon' constrained optimisation routine in Matlab software. The MCMC sampler is run for $N = 20000$, with a burn-in of $M = 15000$, iterations; in each case all iterations after burn-in are used to calculate the posterior mean estimates. For both estimation methods, all initial parameter values were arbitrarily set equal to 0.25. MCMC convergence was checked extensively by running the sampler from different starting points and visually observing convergence to the same posterior well inside the burn-in period, for multiple simulated (and real) datasets from each model; such convergence almost always occurs within one thousand iterations.

Estimation results are summarised in Tables 2.1 and 2.2. Boxes indicate the optimal measure comparing MCMC and ML for both bias (Mean) and precision

(RMSE). For $n = 1500$, the results are fairly mixed across the methods. Both methods generate close to unbiased and quite reasonably precise parameter estimates and quantile forecasts. The bias results slightly favour the ML method, with 6 out of 8 parameter estimates and both quantile forecasts averaging closer to their true value; whilst the precision is slightly lower for the MCMC method in 6 out of 8 parameter estimates, but slightly higher for the quantile forecasts.

For $n = 3000$, the results are more in favour of the MCMC method overall. Again both methods generate close to unbiased and quite reasonably precise parameter estimates and quantile forecasts. The bias results are mixed, with 4 out of 8 parameter estimates favouring each method, though both quantile forecasts favour the ML; whilst the precision is slightly lower for the MCMC method for 8 parameters and also for the quantile forecasts.

The typical increase in precision in the MCMC estimator is small in most cases, but is notably larger for the parameters ω, ξ, σ. The latter two of these, which also have smaller bias than the MLE, have Jeffreys-type priors, that shrink estimates towards 0; clearly these priors have had an effect in this case at both sample sizes. The increased RMSE for the ML estimator of ω is partly due to a few datasets inducing large MLEs for that parameter, whilst the MCMC estimator was not at all large in those cases, and further that often the MLE was very, very close to the boundary at $\omega = 0$ (i.e. $> 20\%$ of the MLEs were < 0.000001) whilst the corresponding MCMC estimates were never similarly close to 0; this clearly reduces the bias for the MLE in this case as well.

Estimation results for Model 2 are summarised in Table 2.2. For $n = 1500$, the results are mostly in favour of the MCMC method. Both methods generate close to unbiased and quite reasonably precise parameter estimates and quantile forecasts, except the ML method for ν. This is because about 0.5% of MLEs for ν were above 30, and some of those were in the tens or hundreds of thousands, leading to large Mean and RMSE results. We would like to point out that the parameter ν of the Student-t distribution is not identified when it is large. For example, the likelihood cannot distinguish between $\nu = 100000$ and $\nu = 100001$, which practically makes the Student-t distribution the same as Gaussian distribution. As discussed in Geweke (1993), the posterior with flat prior is non-integrable and improper in this case. Therefore, through incorporating the $1/\nu^2$ prior in the MCMC algorithm, the highest ν estimate was 75 with the same simulated data sets. Therefore, here it again demonstrates the superiority of MCMC. The bias results are evenly spread

TABLE 2.1: Summary statistics for the two estimators of the RG-GG model, data simulated from Model 1.

$n = 1500$		MCMC		ML	
Parameter	True	Mean	RMSE	Mean	RMSE
ω	0.02	0.0299	0.0215	0.0216	0.0325
β	0.75	0.7420	0.0206	0.7471	0.0232
γ	0.25	0.2577	0.0233	0.2528	0.0237
ξ	0.10	0.1266	0.0664	0.1359	0.1206
φ	0.95	0.9367	0.0459	0.9406	0.0512
τ_1	0.10	0.1003	0.0132	0.1000	0.0131
τ_2	-0.10	-0.1008	0.0100	-0.1003	0.0098
σ_ε	0.50	0.5002	0.0092	0.4991	0.0093
1% VaR	-4.386	-4.3987	0.0904	-4.3864	0.0888
5% VaR	-3.101	-3.1101	0.0639	-3.1014	0.0628
$n = 3000$	True	Mean	RMSE	Mean	RMSE
ω	0.02	0.0244	0.0173	0.0212	0.0343
β	0.75	0.7460	0.0145	0.7479	0.0305
γ	0.25	0.2540	0.0162	0.2516	0.0285
ξ	0.10	0.1136	0.0517	0.1175	0.1136
φ	0.95	0.9428	0.0325	0.9467	0.0462
τ_1	0.10	0.1000	0.0092	0.0998	0.0092
τ_2	-0.10	-0.1000	0.0070	-0.0997	0.0070
σ_ε	0.50	0.5001	0.0064	0.4998	0.0095
1% VaR	-4.382	-4.3873	0.0612	-4.3796	0.0682
5% VaR	-3.098	-3.1020	0.0432	-3.0966	0.0482

between methods, though the ML quantile and ES forecasts average closer to their true value; whilst the precision is slightly lower for the MCMC method in almost all cases.

For $n = 3000$, the results are almost all in favour of the MCMC method. Again both methods generate close to unbiased and quite reasonably precise parameter estimates and quantile forecasts, except for the MLEs for ν. The bias and precision results almost all favour the MCMC estimator.

The increase in precision in the MCMC estimator is small in most cases, but larger for the parameters ν, ξ. Both of these have shrinkage priors; clearly these have had a positive effect in this case at both sample sizes. Similar increases in precision for Bayesian estimates over frequentist optimisation were found in Gerlach and Chen

TABLE 2.2: Summary statistics for the two estimators of the RG-tG model, data simulated from Model 2.

$n = 1500$		MCMC		ML	
Parameter	True	Mean	RMSE	Mean	RMSE
ω	0.01	0.0877	0.0856	0.0686	0.0983
β	0.70	0.6952	0.0227	0.7008	0.0236
γ	0.29	0.2815	0.0268	0.2767	0.0321
ξ	0.01	0.3649	0.3840	0.3914	0.4616
φ	0.99	0.9451	0.0762	0.9480	0.0918
τ_1	0.25	0.2019	0.0682	0.2003	0.0713
τ_2	-0.25	-0.1898	0.0676	-0.1875	0.0694
σ_ε	2.00	1.8749	0.1323	1.8707	0.1364
ν	8.00	8.4472	2.6691	1815.0	30559.0
1% VaR	-5.362	-5.4108	0.2036	-5.3629	0.2012
5% VaR	-3.442	-3.4522	0.0915	-3.4397	0.0949
1% ES	-6.625	-6.7557	0.3804	-6.6655	0.5154
5% ES	-4.659	-4.6925	0.1629	-4.6591	0.3357
$n = 3000$	True	Mean	RMSE	Mean	RMSE
ω	0.01	0.0783	0.0764	0.0671	0.0790
β	0.70	0.7008	0.0151	0.7035	0.0156
γ	0.29	0.2749	0.0232	0.2729	0.0265
ξ	0.01	0.3361	0.3518	0.3570	0.4104
φ	0.99	0.9599	0.0566	0.9598	0.0710
τ_1	0.25	0.2014	0.0595	0.2006	0.0608
τ_2	-0.25	-0.1872	0.0666	-0.1861	0.0677
σ_ε	2.00	1.8760	0.1275	1.8741	0.1294
ν	8.00	8.2107	8.2938	1057.8	18671.7
1% VaR	-5.371	-5.3814	0.1494	-5.3578	0.1553
5% VaR	-3.448	-3.4471	0.0737	-3.4411	0.0783
1% ES	-6.636	-6.6897	0.2574	-6.6439	0.2591
5% ES	-4.666	-4.6694	0.1221	-4.6527	0.1751

(2014) and Gerlach, Chen and Chan (2011) for different classes of financial time series models.

As discussed in Section 1.5.4, in order to evaluate the convergence and efficiency of the employed MCMC method, we employ the Gelman-Rubin diagnostic (Equation (1.7)) and an effective sample size test (Equation (1.10)) (Gelman *et al.*, 2014). For each parameter in the Realized-GARCH, we run the adaptive MCMC for $m = 5$ times (each run with random starting values and different simulated data sets) with burn-in RWM iterations $15,000$ and IMH iterations $n_1 = 5,000$, then

only the IMH iterations $n_1 = 5,000$ in each chain are used for Gelman-Rubin statistics and effective sample size calculation. For example, Figure 2.1 visualizes the 5 MCMC chains with 5 random starting points for φ parameter of the RG-GG model with simulated data generated from Model 1, and only the IMH iterations $n_1 = 5,000$ (last 5000 iterations in the figure) are used for the convergence and efficiency tests. Table 2.3 summarizes the Gelman-Rubin statistics and effective sample for 8 parameters of the RG-GG model with simulated data of sample size $n = 1500$ and $n = 3000$ respectively. As can be seen, the $\hat{R}$ is very close to 1 (e.g. < 1.1) for each parameter (except for ξ with $n = 1500$ has $\hat{R}$ slightly larger than 1.1) in the RG-GG model, meaning excellent convergency testing results for each parameter. Through closer check of the between- and within-chain variances, we observer very small within-chain variances for each parameter, which leads to a close to 1 $\hat{R}$ and suggests good convergence property. In addition, Gelman *et al.* (2014) suggested, as a default rule, running the simulation until total $\hat{n}_{eff}$ is $5m$. From Table 2.3, we can clearly see that the total effective sample sizes for all parameters are larger than $5m = 25$ (the average $\hat{n}_{eff}$ for each chain is presented as well), which proves the MCMC algorithm is efficient. Comparing the $\hat{R}$ and $\hat{n}_{eff}$ results with $n = 1500$ and $n = 3000$, generally the $\hat{R}$ and $\hat{n}_{eff}$ results of $n = 3000$ are better compared with that of $n = 1500$, which is consistent with the results in Table 2.1 (bigger in-sample size leads to better estimation accuracy, convergence and efficiency). Similar observations are found in the Gelman-Rubin diagnostic and effective sample size test in the following chapters. Finally, Gelman *et al.* (2014) also suggested that each chain can be split into 2 parts with $n_1/2$ length, so that $\hat{R}$ can assess stationary as well as mixing. We implemented this test as well and still observe close to 1 $\hat{R}$ and good $\hat{n}_{eff}$ results, while the results are not shown here.

Now we run the Gelman-Rubin diagnostic and effective sample size test with the simulated data generated from Model 2. As presented in Table 2.4, all 9 parameters in RG-tG framework still have close to 1 $\hat{R}$ (except for ξ with $n = 1500$ has $\hat{R}$ slightly larger than 1.1) and satisfiable $\hat{n}_{eff}$. Therefore, based on the MCMC results from Tables 2.1, 2.2, 2.3 and 2.4, we can confirm that the adapted MCMC algorithm for the Realized-GARCH framework has better bias (Mean) and precision (RMSE) results compared with ML, and the MCMC chains are proved to have good convergence and efficiency performance given the employed steps of

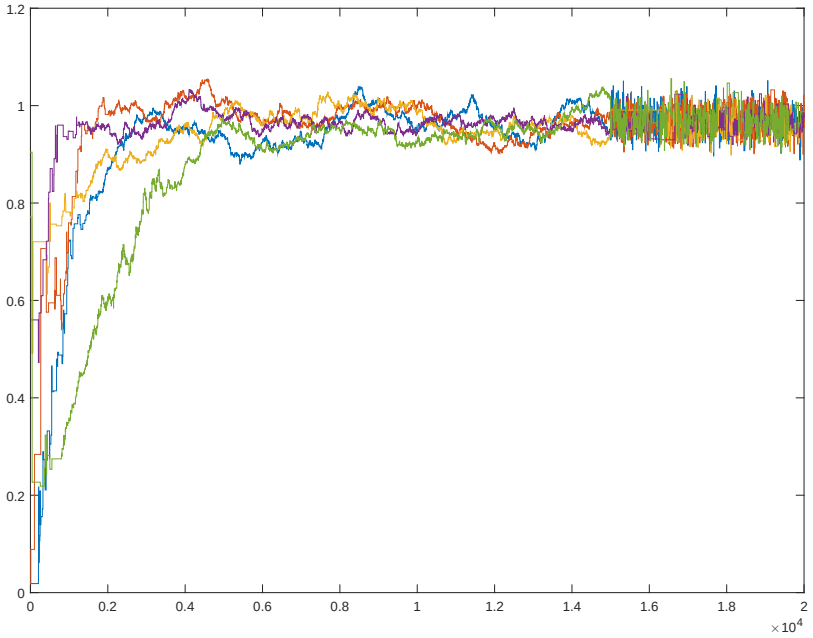

FIGURE 2.1: 5 MCMC chains with the simulated date set ($n = 3000$) and random starting points for φ of the RG-GG model.

TABLE 2.3: Summary statistics of the Gelman-Rubin diagnostic and effective sample size with RG-GG and simulated data set (simulated from Model 1).

$n = 1500$			
Parameter	$\hat{R}$	Total $\hat{n}_{eff}$	Average $\hat{n}_{eff}$
ω	1.07168	387	77
β	1.01295	750	150
γ	1.02517	672	134
ξ	1.12133	83	17
φ	1.03229	351	70
τ_1	1.00286	1316	263
τ_2	1.00299	1035	207
σ_ε	1.00045	1,500	300
$n = 3000$			
ω	1.05747	351	70
β	1.00674	891	178
γ	1.00475	824	165
ξ	1.07498	107	21
φ	1.00581	702	140
τ_1	1.00060	2118	424
τ_2	1.00120	1412	282
σ_ε	1.00076	1986	397

TABLE 2.4: Summary statistics of the Gelman-Rubin diagnostic and effective sample size with RG-tG and simulated data set (simulated from Model 2.

$n = 1500$			
Parameter	$\hat{R}$	Total $\hat{n}_{eff}$	Average $\hat{n}_{eff}$
ω	1.06029	295	59
β	1.00028	1170	234
γ	1.02908	731	146
ξ	1.11055	104	21
φ	1.07682	367	73
τ_1	1.00090	1820	364
τ_2	1.00201	1295	259
σ_ε	1.00146	2009	402
ν	1.00013	4242	848
$n = 3000$			
ω	1.02884	353	71
β	1.00267	1152	230
γ	1.00897	783	157
ξ	1.03266	132	26
φ	1.02217	530	106
τ_1	1.00054	1929	386
τ_2	1.00082	1596	319
σ_ε	1.00173	2395	479
ν	1.00197	2578	516

iterations.

2.6 Data and empirical study

2.6.1 Data description and cleaning

Five daily international stock market indices are analyzed: the S&P 500 (US); NASDAQ (US); Hang Seng(Hong Kong); FTSE 100 (UK); DAX (Germany). Daily closing price index data from Jan 3, 2000 to Sep 18, 2014 are obtained from Thomson-Reuters Tick history, along with 1 minute and 5 minute open, close, high and low prices for each day. The daily percentage log return series were generated as $y_t = (\ln(C_t) - \ln(C_{t-1})) \times 100$, where C_t is the closing price index or closing exchange rate on day t. Realized measures are obtained using the formulas in Section 2.2. Three months' data are used for the scaling process ($q = 66$ in Equations (2.7) and (2.8)), so the the final start date of the data used is April 6, 2000. Market-specific non-trading days were removed from each series.

The full data period is divided into an estimation sample: Apr 6, 2000 to Dec 31, 2007, of $\approx n = 1900$ days; and a forecast sample: approximately $m = 1660$ trading days from Jan 1, 2008 to Sep 18, 2014. The latter period includes most, if not all, of the effects of the global financial crisis (GFC) on each market. Small differences in forecast sample sizes and end-dates occurred across markets, due to market-specific non-trading days. The exact in-sample sizes n and forecast sample sizes m are given in Table 2.8. All series display the standard properties of daily asset returns: positive excess kurtosis, persistent heteroskedasticity and mostly mild, negative skewness. Figures 2.2, 2.3 and 2.4 visualize the S&P 500 absolute return versus the RV & RR, scaled RV & scaled RR and sub-sampled RV & sub-sampled RR respectively, for exposition.

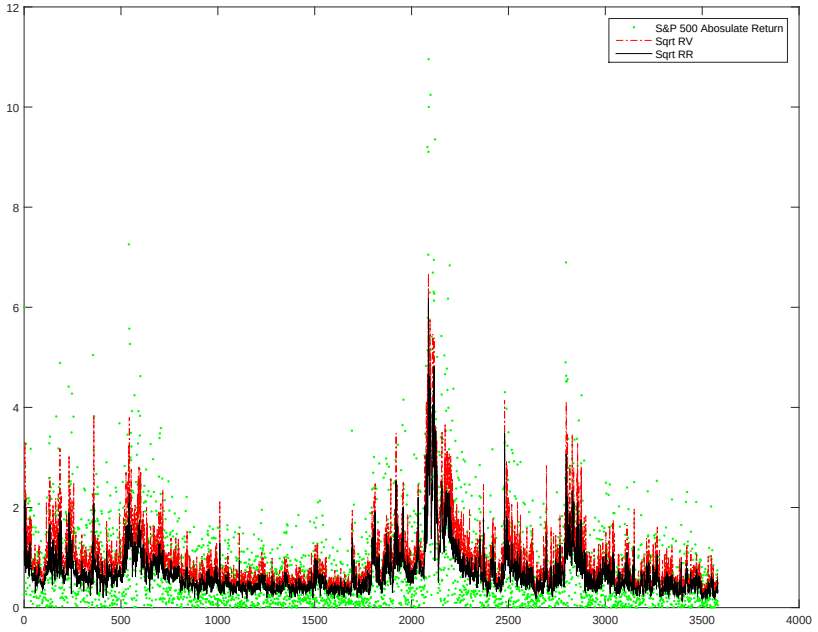

FIGURE 2.2: S&P 500 absolute return, square root of RV and square root of RR.

2.6.2 In-sample parameter estimation results

This section presents the in-sample results to see estimated parameter of various proposed models and how the adaptive MCMC works in the empirical study. As presented in Section 2.6.1, the in-sample and out-of-sample sizes for different markets are around 1900 and 1660 respectively. Now we focus on the estimation results with first S&P 500 in-sample data set: observation 1 to 1960. To begin

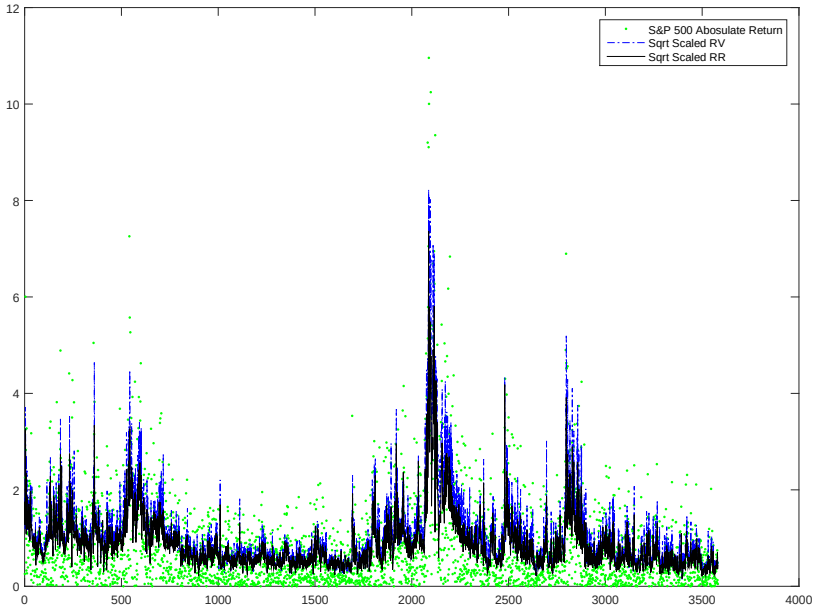

FIGURE 2.3: S&P 500 absolute return, square root of scaled RV and square root of scaled RR.

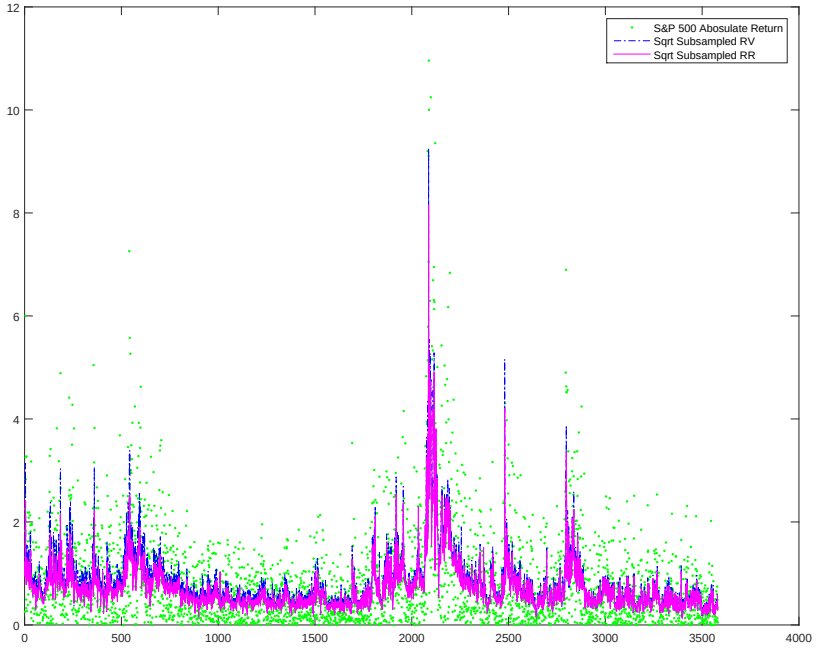

FIGURE 2.4: S&P 500 absolute return, square root of sub-sampled RV and square root of sub-sampled RR.

with, the RR-RG-GG and RR-RG-tG RWM iterates for each block with the 1st S&P 500 in-sample data set are plot in Figures 2.5 and 2.7 respectively. As can be seen, in the 15000 RWM iterations, the parameters started to converge after 1000 iterations, with the acceptance rates that are very close to the target 23.4%. Then all IMH 5000 iteration values for the parameters θ, as in Figure 2.6 and 2.8, are plugged into the predictive density formula. The adaptive MCMC iterates plots are very similar for the other proposed models.

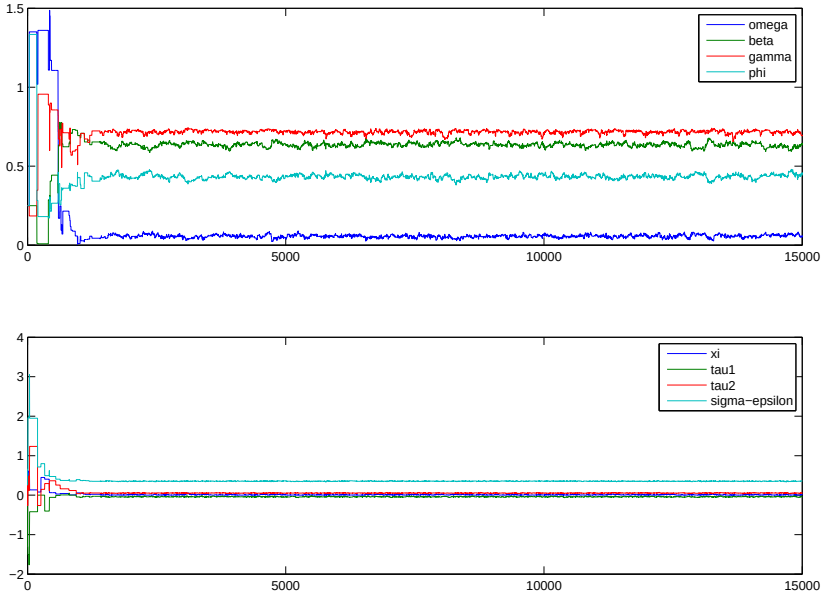

FIGURE 2.5: Plots of 15000 RWM iterations of 2 blocks with RG-RR-GG and S&P500. Acceptance rates for block 1 and 2: 21.0%, 20.9%.

Further, with the 1st S&P 500 in-sample data set and MCMC. The estimated parameters of 16 different Realized GARCH type models are presented in Table 2.5. No matter employing the Gaussian or t distributions for the volatility equation of Re-GARCH, we can clearly see the much smaller estimated σ_ε with RR-RG compared to RV-RG. This results is consistent with the findings in Martins and van Dijk (2007), Christensen and Podolskij (2007): RR has much lower mean squared error than RV, which might provide RR with higher accuracy and efficiency in volatility estimation and forecasting. Through looking at the σ_ε values estimated from RG employing scaled and sub-sampled realized measures, the SSRV-RG provides clearly smaller σ_ε compared to RV-RG, and SSRR-RG have similar σ_ε estimated as RR-RG. However, we see increased σ_ε through incorporating ScRV or ScRR.

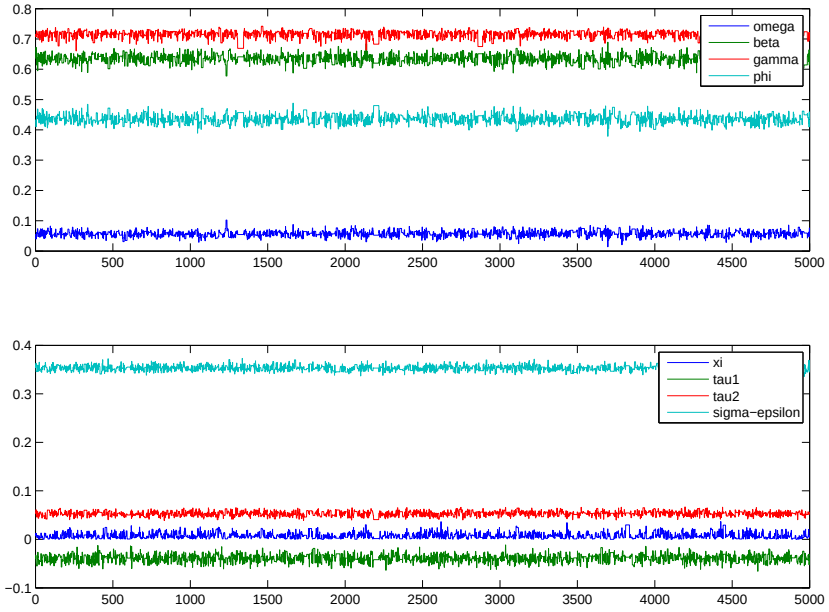

FIGURE 2.6: Plots of 5000 IMH iterations of 2 blocks with RG-RR-GG and S&P500. Acceptance rates for block 1 and 2: 38.0%, 42.0%.

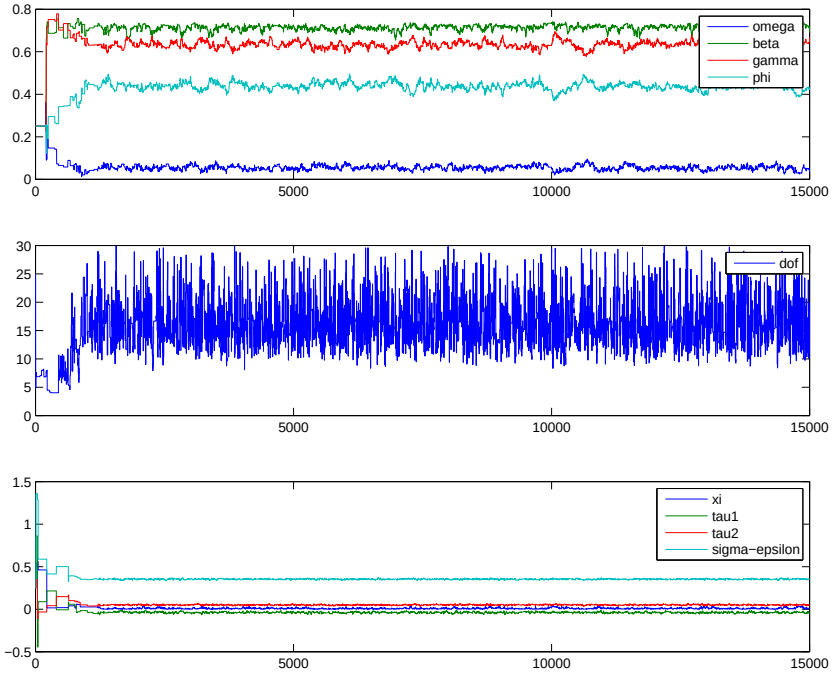

FIGURE 2.7: Plots of 15000 RWM iterations of 3 blocks with RG-RR-tG and S&P500. Acceptance rates for block 1, 2 and 3: 21.3%, 20.4%, 21.0%.

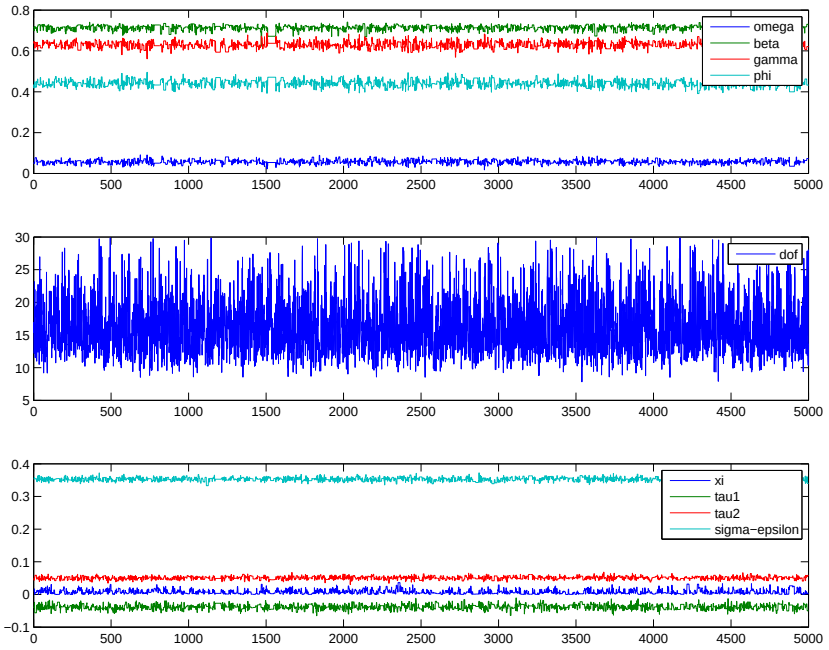

FIGURE 2.8: Plots of 15000 IMH iterations of 3 blocks with RG-RR-tG and S&P500. Acceptance rates for block 1, 2 and 3: 31.8%, 80.6%, 34.7%.

TABLE 2.5: In-sample estimated parameters for 16 RG type models with S&P 500.

Models	ω	β	γ	ξ	φ	τ_1	τ_2	σ_ε	ν
Ra-RG-GG	0.1125	0.8222	0.1233	0.0070	0.8242	0.0228	0.5765	0.8196	
RaO-RG-GG	0.6548	0.4855	0.0006	0.6092	0.2761	-0.0217	0.5712	0.7630	
RV-RG-GG	0.0665	0.6430	0.3455	0.0218	0.8665	-0.0513	0.1696	0.9040	
RR-RG-GG	0.0579	0.5332	1.0267	0.0204	0.3923	-0.0405	0.0512	0.3514	
ScRV-RG-GG	0.0746	0.6326	0.2918	0.0214	1.0570	-0.0578	0.1900	1.1721	
ScRR-RG-GG	0.0686	0.5498	0.5129	0.0183	0.7595	-0.0768	0.0908	0.6816	
SSRV-RG-GG	0.0901	0.4584	0.6922	0.0348	0.6504	-0.0960	0.0882	0.6522	
SSRR-RG-GG	0.0903	0.3956	1.1276	0.0227	0.4520	-0.0660	0.0504	0.4011	
Ra-RG-tG	0.1501	0.8459	0.1193	0.0092	0.7219	0.0301	0.7700	0.7938	4.0859
RaO-RG-tG	0.0142	0.9870	0.0072	0.0130	0.7289	0.0061	0.7962	0.7582	4.0747
RV-RG-tG	0.0673	0.6398	0.3490	0.0243	0.8639	-0.0487	0.1702	0.9027	16.0763
RR-RG-tG	0.0603	0.5214	1.0549	0.0193	0.3911	-0.0401	0.0513	0.3517	15.2417
ScRV-RG-tG	0.0741	0.6300	0.2961	0.0209	1.0506	-0.0574	0.1895	1.1730	14.6134
ScRR-RG-tG	0.0667	0.5421	0.5293	0.0221	0.7498	-0.0771	0.0900	0.6822	14.9609
SSRV-RG-tG	0.0897	0.4510	0.7067	0.0328	0.6478	-0.0979	0.0862	0.6539	14.5389
SSRR-RG-tG	0.0848	0.3812	1.1738	0.0283	0.4449	-0.0660	0.0486	0.4016	13.8718

Finally, in order to study the convergence and efficiency performance of the employed MCMC algorithm with the real world data set, we also perform the Gelman-Rubin diagnostic and effective sample size test with the 1st S&P 500 in-sample data set. Similarly, 5 random starting points ($m = 5$ and $n_1 = 5000$) are used for the RR-RG-GG and RR-RG-tG models respectively. As can be seen in Tables 2.6 and 2.7, both the $\hat{R}$ and $\hat{n}_{eff}$ tests produce values that support good convergency and efficiency results for both the RR-RG-GG and RR-RG-tG frameworks.

TABLE 2.6: Summary statistics of the Gelman-Rubin diagnostic and effective sample size with RR-RG-GG and S&P 500.

Parameter	$\hat{R}$	Total $\hat{n}_{eff}$	Average $\hat{n}_{eff}$
ω	1.00044	1838	368
β	1.00271	1624	325
γ	1.00071	2563	513
ξ	1.00225	867	173
φ	1.00565	1179	236
τ_1	1.00003	4084	817
τ_2	1.00167	2933	587
σ_ε	1.00021	3981	796

TABLE 2.7: Summary statistics of the Gelman-Rubin diagnostic and effective sample size with RR-RG-tG and S&P 500.

Parameter	$\hat{R}$	Total $\hat{n}_{eff}$	Average $\hat{n}_{eff}$
ω	1.00523	1318	264
β	1.00158	1475	295
γ	1.00127	2400	480
ξ	1.00051	12899	2580
φ	1.00291	634	127
τ_1	1.00319	807	161
τ_2	1.00077	3405	681
σ_ε	1.00251	2304	461
ν	1.00037	3619	724

2.6.3 Out-of-sample forecasting: predictive log-likelihood

Approximately 1600-1700 one-step-ahead volatility, VaR and ES forecasts are generated separately for the RG-GG and RG-tG specifications using Ra square, RaO square, RV, RR, ScRV, ScRR, SSRV and SSRR as the measurement equation input. The period from April 6, 2000 to Dec 31, 2007 is used as the initial learning period to generate the first day's forecasts, being for Jan 3, 2008. This is approximately 1900-2000 days in each market, with small differences due to trading day and holiday variations. This estimation period window is then moved ahead by one day to estimate each model and generate the next day's set of forecasts, this process continuing until forecasts are generated for each day in the forecast sample period Jan 3, 2008 - Sep 18, 2014 for each model.

To assess and compare volatility forecasting accuracy between models we consider the predictive likelihood, as in Hansen *et al.* (2011). Based on the sample period

data, $r_1, \ldots, r_n$, the ML method estimates the model parameters $\hat{\theta}$, which are plugged in to form forecasts of $\hat{\sigma}^2_{n+1}$. Such forecasts are plugged into the one-step-ahead return density function, usually logged, to form the one-step-ahead predictive density estimate. For example, the 1st day predictive log-likelihood for the RG-GG model is given by:

$$\ell_{n+1} = -\frac{1}{2}\left[log(2\pi) + log(\hat{\sigma}^2_{n+1}) + r^2_{n+1}/\hat{\sigma}^2_{n+1}\right]$$

These log-density estimates are calculated for each day in the forecast period, and subsequently summed to estimate the log-predictive likelihood for each model.

Alternatively, under the MCMC approach, each IMH iterate of values for the parameters θ are plugged into the one day predictive density formula, as above, giving an MCMC iterate of this quantity. These density iterates are subsequently averaged over the MCMC sampling period to estimate each day's predictive log-density. The single day predictive density estimates are then summed over all the days in the forecast period to give an MCMC estimate of the log-predictive likelihood for each model.

Both approaches give predictive likelihoods that are equal to at least one decimal place and give qualitatively the same order ranking of models. As such, only the MCMC predictive likelihood estimates are reported here. Table 2.8 reports these estimates across the parametric models considered here: being the RG-GG and RG-tG models, each using Ra square and RaO square, RV, RR, ScRV, ScRR, SSRV and SSRR as the input measurement, as well as standard GARCH-Gaussian (G-G) and GARCH-t (G-t) models.

Firstly, we can see that Realized GARCH employing the realized range always has improved predictive log-likelihood compared to the Realized GARCH with RV, no matter using Gaussian or Student-t errors for the volatility equation. The in-sample results might partially explain such improvement. In addition, we can see the scaling and sub-sampling process contribute to the further improved predictive log-likelihood, especially the proposed sub-sampled RR. In three out of five markets the RG model with Student-t errors that uses the SSRR as measurement input, is clearly favoured among the 18 models presented.

In four markets the RG that employs RaO as an input and Gaussian errors is the least favoured model. In four out of five markets, the RV-RG-GG model beats the GARCH-G. Clearly, by the measure of predictive likelihood, the RR and scaled

and sub-sampled realized measures are more informative in all markets, compared to the squared daily returns, Ra and RaO and RV. The RaO is the least favoured measure.

The results here clearly indicate the superior predictive power of the RR-RG-tG, SSRV-RG-tG and SSRR-RG-tG models over standard GARCH models and RG models that employ RV, Ra or RaO as the realized measures. This suggests that employing RR leads to significant gains in information and predictability of both volatility and the predictive return distribution, over RV, Ra and RaO, and that the RR-RG-tG, SSRV-RG-tG and SSRR-RG-tG models should be considered for financial applications that require volatility or distributional forecasts, e.g. option pricing and tail-risk forecasting (as illustrated in the next section).

Table 2.8: Log-predictive likelihoods; Jan 2008 - Sep 2014.

Model	S&P500	NASDAQ	Hang Seng	FTSE	DAX
G-G	-2,390.2	-2,694.9	-2,846.6	-2,533.3	-2,851.0
G-t	-2,357.1	-2,667.6	-2,832.2	-2,515.7	-2,822.2
Ra-RG-GG	-2,440.0	-2,649.1	-2,855.1	**-2,624.7**	-2,853.2
RaO-RG-GG	**-2,461.4**	**-2,741.6**	**-3,278.1**	-2,624.5	**-2,875.2**
RV-RG-GG	-2,343.0	-2,656.5	-2,928.4	-2,496.7	-2,803.3
RR-RG-GG	-2,311.2	-2,616.6	-2,893.0	-2,489.4	-2,783.6
ScRV-RG-GG	-2,345.1	-2,659.7	-2,837.5	-2,500.5	-2,813.4
ScRR-RG-GG	-2,322.3	-2,618.8	-2,822.9	-2,488.9	-2,788.6
SSRV-RG-GG	-2,308.8	-2,613.4	-2,830.6	-2,479.3	-2,785.2
SSRR-RG-GG	-2,299.0	-2,615.9	-2,868.5	-2,477.4	-2,778.5
Ra-RG-tG	-2,464.2	-2,638.8	-2,863.5	-2,645.2	-2,842.1
RaO-RG-tG	-2,459.6	-2,735.9	-3,066.3	-2,646.1	-2,872.5
RV-RG-tG	-2,329.9	-2,647.0	-2,896.1	-2,490.0	-2,790.4
RR-RG-tG	-2,299.5	-2,608.5	-2,877.8	-2,482.4	-2,775.2
ScRV-RG-tG	-2,327.7	-2,646.6	-2,829.0	-2,492.5	-2,797.7
ScRR-RG-tG	-2,307.4	-2,609.9	-2,816.0	-2,481.1	-2,778.9
SSRV-RG-tG	-2,294.3	-2,605.1	-2,823.2	-2,479.7	-2,776.5
SSRR-RG-tG	-2,287.2	-2,607.1	-2,862.4	-2,472.2	-2,770.9
m	1621	1672	1631	1697	1691
n	1960	1892	1890	1944	1936

Note: A box indicates the favored model in each market, based on minimum predictive log-likelihood, whilst bold indicates the least favoured model.

2.6.4 Out-of-sample forecasting: tail risk

The Basel II and III Capital Accords favour VaR and ES as tail risk measures for financial institutions to employ in market risk management. Thus, it is highly important for institutions to have access to highly accurate VaR and ES forecast models, allowing accurate capital allocation, both to avoid default and over-allocation of funds.

The same estimation sample period, forecast sample period and fixed, moving window approach in the last section are employed in this section that focuses on VaR and ES forecasting at 1% risk levels for five daily financial indices. Popular non-parametric methods for forecasting VaR and ES, including Historical Simulation (HS), using the last 100 (HS100) and the last 250 (HS250) days of returns, are added to the competing models.

2.6.4.1 Value-at-Risk

Table 2.9 presents the estimation period sample size for each forecast n, and the forecast sample size m, in each market; also presented are the numbers of returns in the forecast period that are more extreme than the forecasted VaR (called VaR violations) at the 1% quantile for each model in each market, and the mean & median VaR violations of the 5 markets for each model. These numbers are expected to be $0.01m$: boxes indicate the model that has a violations closest to that baed on the mean & median; bold indicates the model with VRate furthest away from expected. Results for the MCMC estimated RG models are shown, whilst ML methods were used for the standard GARCH models.

Although we can see that, based on mean and median VaR violations for 5 markets, the Ra-RG-tG and RaO-RG-tG are favoured and they are the only two models generated the conservative VaR forecasting results, e.g. violations are less than $0.01m$, some of their VaR forecasts are too conservative, especially the RaO-RG-tG model. This can be explained by the small value of degree of freedom generated by Ra-RG-tG and RaO-RG-tG, refer to the in-sample estimation results Table 2.5. Also, the Ra-RG-tG and RaO-RG-tG are rejected in almost every markets through back testing, which will be explained later on.

Besides the Ra-RG-tG and RaO-RG-tG models, clearly, the 2 models with VaR violations typically closest to the expected $0.01m$ across the five markets are the

TABLE 2.9: Counts of 1% VaR violations during the forecast period in each market.

Model	S&P500	NASDAQ	Hang Seng	FTSE	DAX	Mean	Median
G-G	41	42	33	33	33	**36.4**	33.0
G-t	27	32	26	26	24	27.0	26.0
HS100	23	27	27	28	30	27.0	27.0
HS250	24	26	25	24	21	24.0	24.0
Ra-RG-GG	38	36	32	24	30	32.0	32.0
RaO-RG-GG	40	24	40	24	24	30.4	24.0
RV-RG-GG	37	34	49	24	33	35.4	**34.0**
RR-RG-GG	34	29	45	20	29	31.4	29.0
ScRV-RG-GG	38	36	26	28	32	32.0	32.0
ScRR-RG-GG	38	33	28	27	31	31.4	31.0
SSRV-RG-GG	41	36	30	30	31	33.6	31.0
SSRR-RG-GG	39	30	40	22	26	31.4	30.0
Ra-RG-tG	17	26	21	12	18	18.8	18.0
RaO-RG-tG	21	4	32	12	6	15.0	12.0
RV-RG-tG	26	25	34	19	25	25.8	25.0
RR-RG-tG	27	22	29	14	23	23.0	23.0
ScRV-RG-tG	31	23	20	25	26	25.0	25.0
ScRR-RG-tG	30	27	23	21	25	25.2	25.0
SSRV-RG-tG	35	26	22	25	27	27.0	26.0
SSRR-RG-tG	25	23	27	15	21	22.2	23.0
Mean	26	21	21	16	22	21.2	21.0
Median	31	27	25	21	26	26.0	26.0
Min	2	2	6	3	2	3.0	2.0
Max	72	63	73	56	58	64.4	63.0
m	1621	1672	1631	1697	1691	1662.4	1672.0
n	1960	1892	1890	1944	1936	1924.4	1936.0

Note: Boxes indicate the model with mean or median VaR violations closest to its nominal level only for individual models, whilst bold indicates the least favoured individual models.

RR-RG-tG and SSRR-RG-tG. All models, except Ra-RG-tG and RaO-RG-tG models, have higher than expected average VRates across the markets, which may not be too surprising given that the GFC is at the start of the forecast sample; this issue is examined further later.

Chang *et al.* (2011) and McAleer *et al.* (2013) proposed employing forecast combinations of the VaR series from different models, potentially as a robust to the GFC combined VaR forecast. We incorporated this approach since our forecasting period includes the GFC. Specifically, four combinations of 1% VaR forecasts from individual models are considered: the mean, median, minimum and maximum, of each of the VaR forecasts from the 20 models in Table 2.9. The VaR forecasts are negative in this chapter, so "Min" is the most extreme of the 11 forecasts (i.e. furthest from 0) and "Max" is the least extreme. The violations for "Mean" and "Median", "Min" and "Max" series are also presented in Table 2.9. As expected, the "Min" approach is too conservative in each series, while the "Max" series produces anti-conservative VaR forecasts that produce far too many violations. The "Mean" of the 20 models produced a series that generated closest to the nominal violation rate on average overall; which was closely matched by the the RR-RG-tG and SSRR-RG-tG models for 1% VaR forecasting.

Having a VRate close to 1% on average is not sufficient to guarantee an accurate forecast model. Several tests exist in the literature to statistically assess forecast accuracy and independence of violations, a requirement of a proper risk model. These include the unconditional coverage (UC) Kupiec (1995), conditional coverage (CC) of Christoffersen(1998), dynamic quantile (DQ) of Engle and Manganelli (2004) and VaR quantile regression (VQR) test of Gaglianone *et al.* (2011). The UC tests the hypothesis that the true VRate is $\alpha(= 1\%)$; the CC and DQ are joint tests of that plus the independence of the violations over time; whilst the VQR conducts a Mincer-Zarnawicz quantile regression of forecasted quantiles on the forecast returns, whose parameters are jointly tested to be intercept zero and slope one, respectively, as would indicate an accurate quantile forecasting model. See the referenced chapter for more details.

Table 2.10 counts the number of markets in which each 1% VaR forecast model is rejected, for each test, all conducted at a 5% significance level. Clearly, for 1% VaR forecasting from 2008-2014, the RR-RG-tG and SSRR-RG-tG models have forecast the most accurately and can be least rejected overall. In addition, the

1% VaR series calculated by the mean of 20 VaR forecasts is rejected in only 2 markets and is thus ranked as the best model by this criterion.

TABLE 2.10: Counts of rejections for each test and 1% VaR model during the forecast period over the five markets, $\alpha = 0.05$.

$\alpha = 0.01$	UC	CC	DQ4	VQR	Total
G-G	5	5	5	5	**5**
G-t	4	2	5	3	**5**
HS100	4	4	5	1	**5**
HS250	2	3	5	1	**5**
Ra-RG-GG	4	4	5	4	**5**
RaO-RG-GG	2	2	3	5	**5**
RV-RG-GG	4	4	4	4	4
RR-RG-GG	4	4	3	3	4
ScRV-RG-GG	5	4	5	4	**5**
ScRR-RG-GG	5	4	5	4	**5**
SSRV-RG-GG	5	5	5	4	**5**
SSRR-RG-GG	4	3	3	3	4
Ra-RG-tG	1	0	5	1	**5**
RaO-RG-tG	3	3	3	4	**5**
RV-RG-tG	2	1	1	2	3
RR-RG-tG	2	2	1	2	3
ScRV-RG-tG	2	1	3	1	4
ScRR-RG-tG	2	2	2	0	3
SSRV-RG-tG	3	2	2	3	4
SSRR-RG-tG	2	1	1	3	3
Mean	1	0	1	1	2
Median	4	2	3	3	4
Min	5	5	3	5	5
Max	5	5	5	5	5

Note: Boxes indicate the model with lowest number of total rejections only for individual models, whilst bold indicates the individual models with highest number of total rejections.

Overall, RR-RG-tG and SSRR-RG-tG models are the best performing models at 1% VaR forecasting. In terms of employing forecast combination to produce a VaR series that is robust to the GFC, the "Mean" series showed the most potential, having the closest overall violation rate to nominal and ranking first in the diagnostic testing. However, the "Median", "Min" and "Max" series were not competitive with many of the individual models, on these criteria.

2.6.4.2 Expected Shortfall

The same set of models are employed to generate 1-step-ahead forecasts of 1% ES during the forecast sample in each market. Chen, Gerlach and Lu (2012) discuss how to treat ES forecasts as quantile forecasts in parametric models, where the quantile level that ES falls at can be deduced exactly. Gerlach and Chen (2015) illustrate that across a range of non-Gaussian distributions, when applied to real daily financial return data, the quantile level that the 1% ES was estimated to fall was $\approx 0.36\%$. Their approaches are followed to assess and test ES forecasts, by treating them as quantile forecasts and employing the UC, CC, DQ and VQR tests. As such, the expected number of violations from ES models are expected to be $= 0.0038m$ (exact for models with Gaussian errors); $\approx 0.0036m$ (for non-parametric models) and estimated by the quantile level implied by the degrees of freedom estimates for models with Student-t errors (also $\approx 0.0036m$ for the data considered here), refer to Table 1.1 for details. Thus, based on the actual sizes of m in Table 2.9, all models have an expected or target ES violation number of between 6 and 6.5 in each market.

Figures 2.10 and 2.9 show the forecast sample returns from the S&P 500 and some associated forecasted ES series. The models shown are the G-G, G-t, HS250,RR-RG-tG and RR-RG-tG (estimated by MCMC). The violation numbers from these four models are, respectively, 23, 11, 14, 6 and 5; the expected is $5.8(\approx 6)$. Despite the large differences in number of violations, the ES forecasts from the RR-RG-tG model, which has the lowest number of violations, are, visually from Figures 2.10 and 2.9 , often less extreme than those from the other three models shown. In fact the ES forecasts from the RR-RG-tG model are less extreme than the G-G on 26.3% of the forecast sample days, less extreme than the G-t model on 63.8% of days and on 64.0% of the days less extreme than the RV-RG-tG model, which can be clearly demonstrated through Figure 2.10. Further, at times where there is a persistence of extreme returns (e.g. the GFC), close inspection of Figure 2.9 reveals that the RG-RR-tG model's ES forecasts "recover" the fastest, in terms of being marginally the fastest to produce forecasts that again follow the tail of the data; GARCH models are well-known to over-react to extreme events and to be subsequently very slow to recover, due to their oft-estimated very, very high persistence. As an example, from August, 2008-January, 2009 (114 observations), the most volatile of the GFC period, the RG-RR-tG model's forecasts are less extreme than the G-G model's on 36.0% of the forecast sample days; including

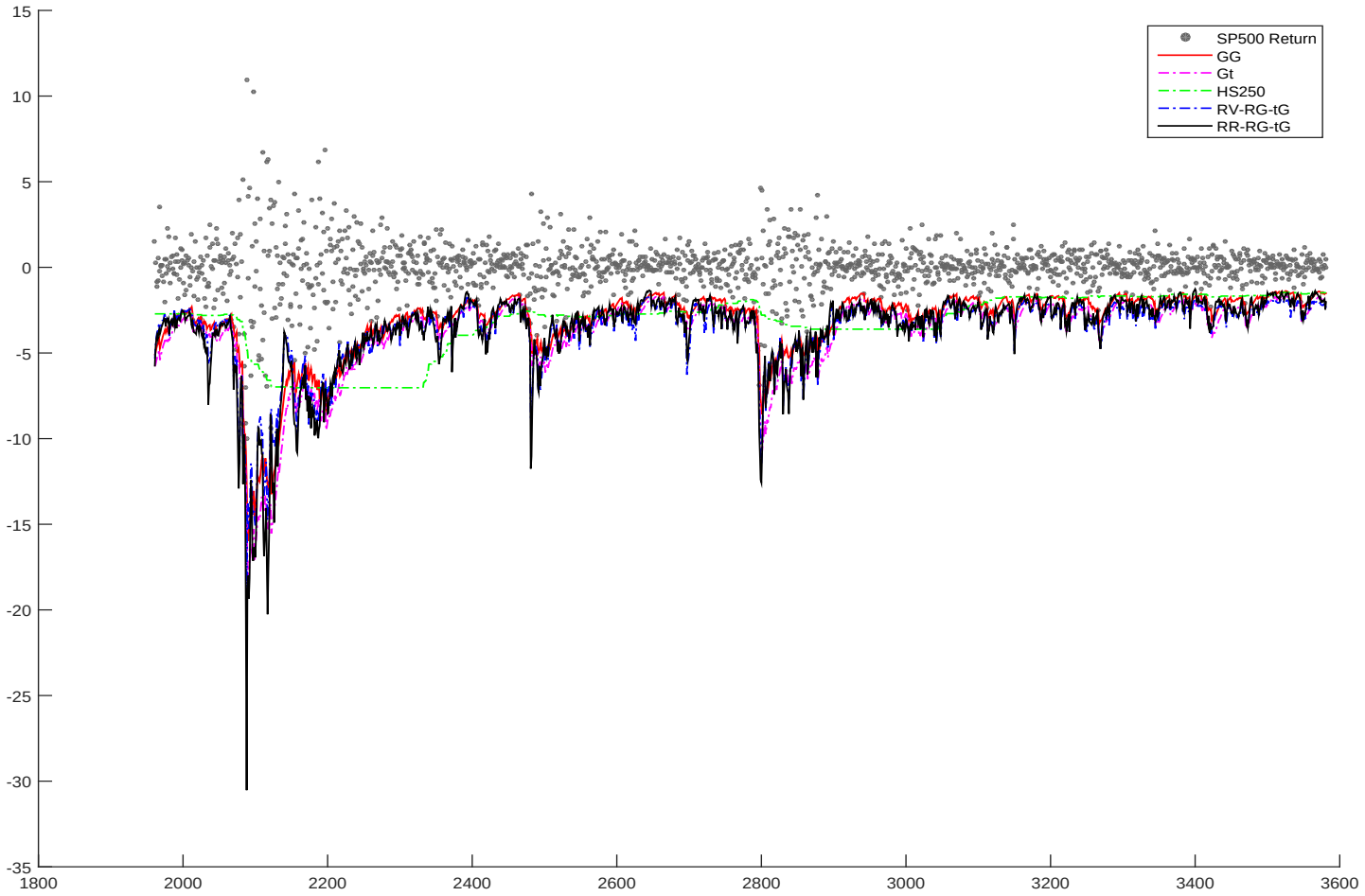

FIGURE 2.9: S&P 500 1% ES forecasts plot with GG, Gt, HS250, RR-RG-tG and RR-RG-tG.

every day in the period from 16/12/08 - 13/01/09; this percentage is 39.5% from 27/04/10-13/10/10 and 47.1% in the period 16/08/11 - 10/01/12. These are three persistent high volatility periods in the S&P500 market during the forecast sample period.

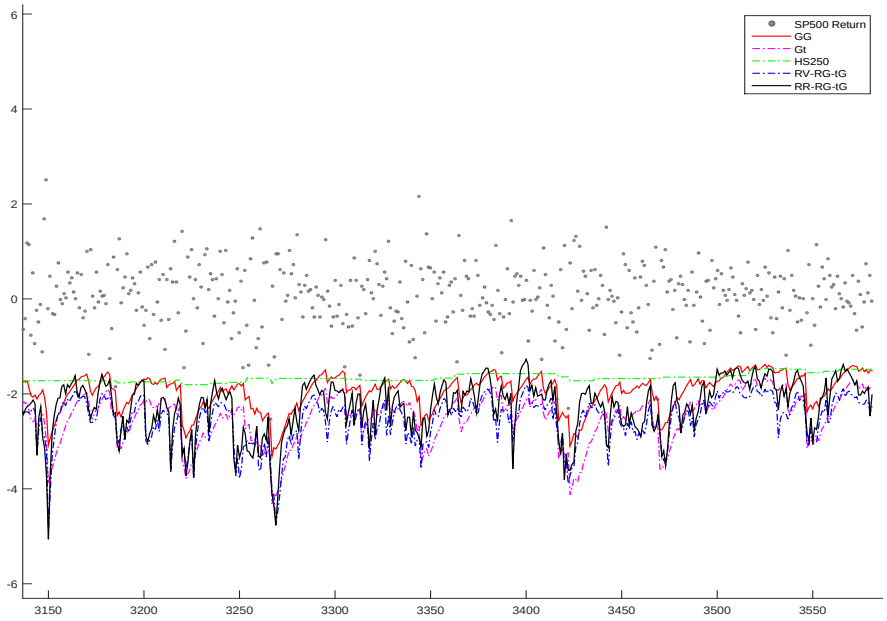

FIGURE 2.10: Zoomed in S&P 500 1% ES forecasts plot with GG, Gt, HS250, RR-RG-tG and RR-RG-tG.

To summarise, Figures 2.10 and 2.9 highlight the extra efficiency that can be gained by employing an RG model, specifically one that employs RR as an input. The efficiency here can be deduced in that this model can produce ES forecasts that have far fewer violations but are simultaneously less extreme than those of the traditional GARCH model. Since the capital set aside by financial institutions, to cover extreme losses, should be directly proportional to the ES forecast, the RG-RR-tG model is saving the company money, by giving more accurate and often less extreme ES forecasts, compared to GARCH models. More evidence for this statement, and how it applies to other markets considered, is now presented.

Table 2.12 presents the numbers of returns in the forecast period that are more extreme than the forecasted ES (called ES violations) at the 1% quantile for each model in each market; called ES violations. The mean and median of ES violations of 5 markets for each model are calculated as well. Similar to Table 2.9, boxes indicate the model in each market that has an ES violations closest to that desired;

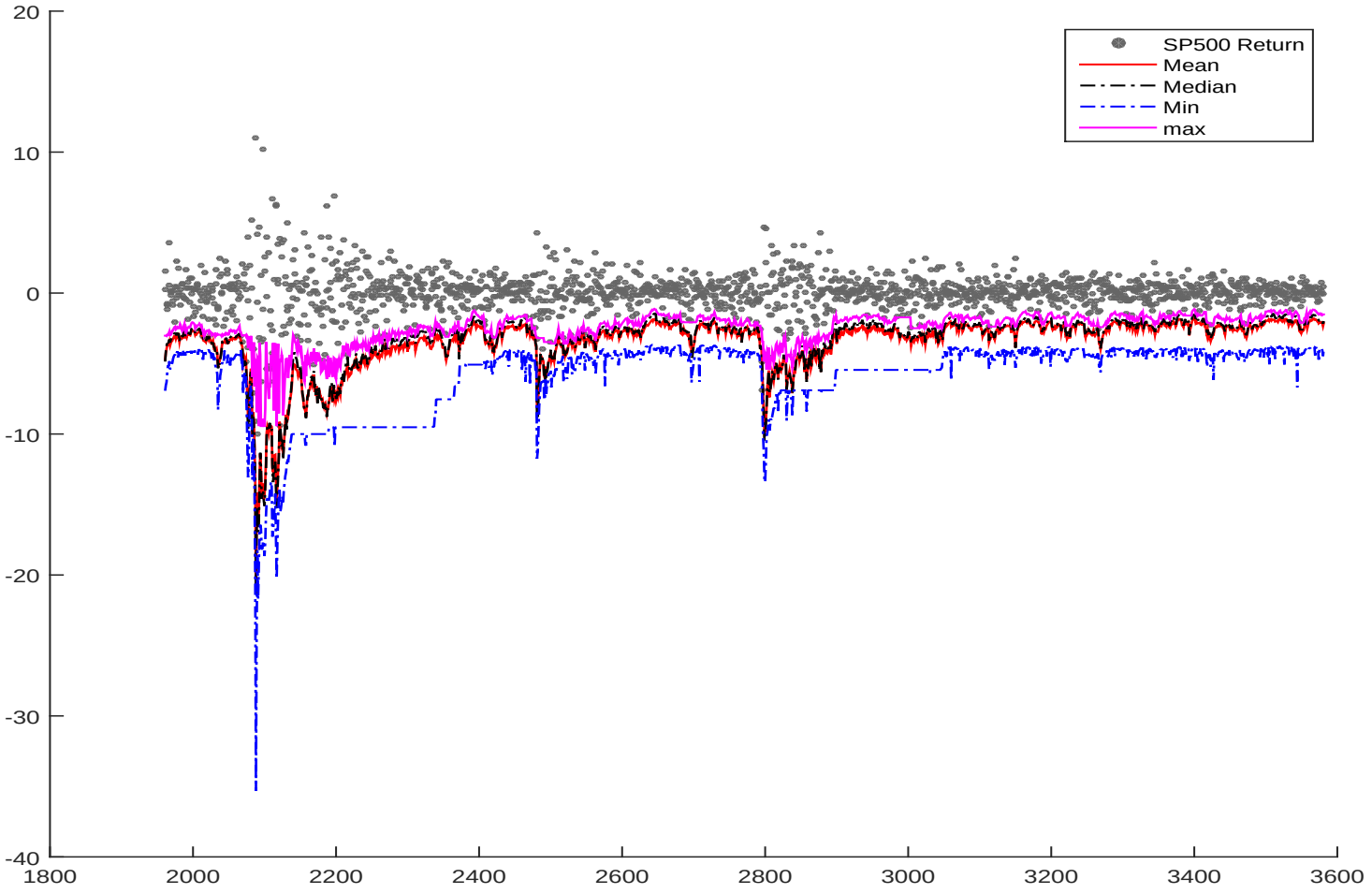

FIGURE 2.11: S&P 500 1% ES forecasts with "Mean", "Median", "Min" and "Max".

bold indicates the model with ES violations furthest from that expected. All RG type models are estimated with MCMC.

Clearly, the best model with ESRates typically closest to that expected for the 1% ES across the five markets is the RR-RG-tG, closely followed by ScRR-RG-tG, SSRV-RG-tG and SSRR-RG-tG. All models have higher than expected average ESRates across the markets, which is not surprising given that the GFC is at the start of the forecast sample; this issue is examined further later.

The mean, median, minimum and maximum of the 20 models' 1% ES forecasts are again calculated and visualised in Figure 2.11 and their ES violations are also shown in Table 2.12 as well. The "Mean" approach is again optimal among the four combination methods, but with average violations of 6.8, which is actually the closest to nominal ES violations and slightly better than the RR-RG-tG.

Having an average ESRate close to that expected is not sufficient to guarantee an accurate forecast model. Following Chen *et al.* (2012) and Gerlach and Chen (2014) the UC, CC, DQ and VQR quantile accuracy tests are applied to the ES violations from each model, using that model's nominal (or an estimate of) 1% ES quantile level. The quantile level corresponding to the median for the estimated ν during the forecast sample is used for models with Student-t errors (the actual estimated range across the t-distributed error models in all markets is $(0.0033, 0.00375)$); 0.0036 is used for non-parametric models, 0.0038 for Gaussian error models.

Table 2.11 counts the number of markets in which each model is rejected, for each test, all conducted at a 5% significance level. For 1% ES forecasting from 2008-2014, the ScRR-RG-tG and SSRV-RG-tG model has forecast got 0 rejection in five markets overall. The next best is the SSRR-RG-tG, rejected in one out of the five markets, and the RR-RG-tG got rejected twice. Of the forecast combination series, the "Mean" is only rejected once.

Figure 2.12 plots the averages of the 1% ES forecast residuals, standardised by the 1% VaR forecasts, for each of the five markets, plus the average of these averages, for each individual forecast model/method. In this figure, RR-RG-GG is represented as RR, and RR-RG-tG is represented as RRt. An accurate 1% ES forecast model should produce standardised residuals that average approximately 0. Table 2.11 illustrates that a bootstrap test on whether these averages differ from 0 is not very powerful, compared to the UC, CC and DQ tests. Agreeing with those results,

TABLE 2.11: Counts of 1% ES model rejections for each test and model during the forecast period over the five markets, $\alpha = 0.05$.

$\alpha = 0.01$	UC	CC	DQ4	VQR	Bootstrap	Total
G-G	5	5	5	2	4	**5**
G-t	1	1	4	0	0	4
HS100	5	5	5	2	5	**5**
HS250	4	4	5	1	2	**5**
Ra-RG-GG	5	5	5	1	3	**5**
RaO-RG-GG	4	4	5	2	3	**5**
RV-RG-GG	5	4	4	1	1	**5**
RR-RG-GG	4	2	2	1	1	4
ScRV-RG-GG	5	5	4	1	3	**5**
ScRR-RG-GG	5	5	4	2	2	**5**
SSRV-RG-GG	4	4	3	1	1	4
SSRR-RG-GG	3	2	2	1	1	3
Ra-RG-tG	0	0	1	0	1	2
RaO-RG-tG	2	1	2	1	1	4
RV-RG-tG	1	1	1	1	1	2
RR-RG-tG	0	0	1	1	1	2
ScRV-RG-tG	0	0	1	0	1	2
ScRR-RG-tG	0	0	0	0	0	0
SSRV-RG-tG	0	0	0	0	0	0
SSRR-RG-tG	0	0	1	0	1	1
Mean	0	0	1	1	0	1
Median	0	0	1	2	0	2
Min	3	3	0	5	1	5
Max	5	5	5	3	5	5

Note: Boxes indicate the model with lowest number of total rejections only for individual models, whilst bold indicates the individual models with highest number of total rejections.

TABLE 2.12: Counts of 1% ES violations during the forecast period in each market.

Model	S&P500	NASDAQ	Hang Seng	FTSE	DAX	Mean	Median
G-G	23	23	18	17	17	19.6	18.0
G-t	11	8	8	13	9	9.8	9.0
HS100	19	21	22	22	21	**21.0**	**21.0**
HS250	14	16	15	15	10	14.0	15.0
Ra-RG-GG	20	16	16	15	18	17.0	16.0
RaO-RG-GG	25	10	28	15	16	18.8	16.0
RV-RG-GG	21	18	32	13	16	20.0	18.0
RR-RG-GG	15	13	23	8	13	14.4	13.0
ScRV-RG-GG	19	14	17	16	18	16.8	17.0
ScRR-RG-GG	21	14	17	15	15	16.4	15.0
SSRV-RG-GG	20	15	18	12	15	16.0	15.0
SSRR-RG-GG	16	13	21	9	12	14.2	13.0
Ra-RG-tG	5	4	9	2	3	4.6	4.0
RaO-RG-tG	6	1	12	2	2	4.6	2.0
RV-RG-tG	6	9	16	7	7	9.0	7.0
RR-RG-tG	5	5	10	7	8	7.0	7.0
ScRV-RG-tG	6	9	10	9	8	8.4	9.0
ScRR-RG-tG	8	7	6	7	9	7.4	7.0
SSRV-RG-tG	7	8	7	7	7	7.2	7.0
SSRR-RG-tG	7	5	9	7	8	7.2	7.0
Mean	7	4	11	7	5	6.8	7.0
Median	8	7	11	8	9	8.6	8.0
Min	1	1	0	2	2	1.2	1.0
Max	51	42	51	33	40	43.4	42.0

Note: Boxes indicate the model with mean or median ES violations closest to its nominal level only for individual models, whilst bold indicates the least favoured individual models.

it is clear that the G-t, RV-RG-tG and RR-RG-tG and RG-tG employing scaled and sub-sampled realized measures are the most accurate; whilst G-G, HS100, HS250, RG-GG incorporating various realized measures clearly, consistently and significantly under-estimate the 1% ES levels, causing negative average residuals to result in all five series.

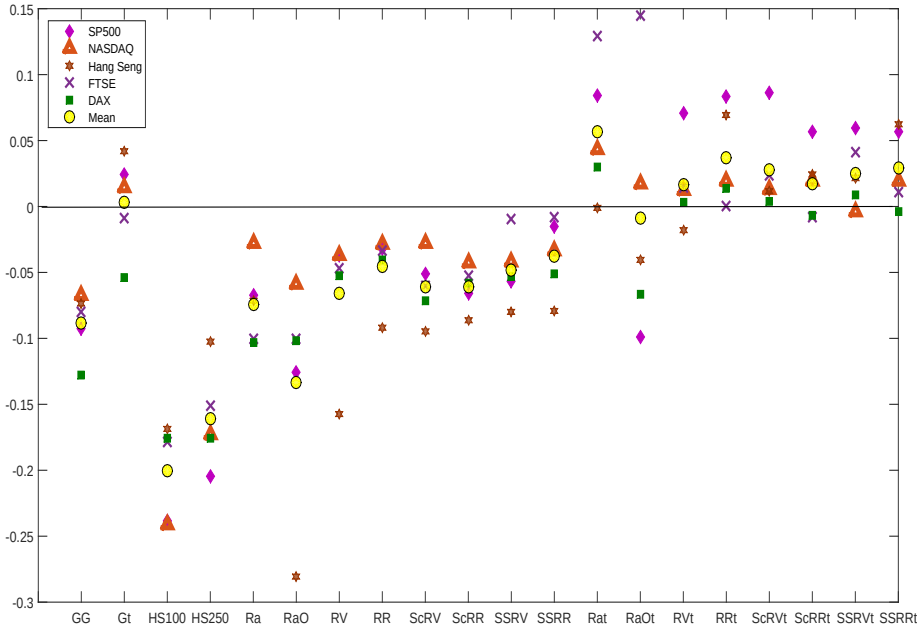

FIGURE 2.12: Residuals for 1% ES forecasts, standardised by 1% VaR forecasts, averaged. For each model the five averages are shown, one for each data series, as well as the average of these averages. A reference line is drawn at 0.

In summary, the RR-RG-tG model, estimated by MCMC, is the most accurate at forecasting 1% ES for the forecast period from Jan, 2008 to Sep, 2014 across five market index return series. It consistently displays ES violation rates closest to the nominal rate predicted by the estimated Student-t error distribution, and has competitive performance with standard diagnostic tests of quantile forecast accuracy. The next best models are ScRR-RG-tG, SSRV-RG-tG, SSRR-RG-tG, and they have really close tail risk forecasting performance. The series produced by the mean of the 20 forecast models also presents accurate forecasting results. Clearly, in the context of RG models, the use of RR led to greater efficiency in ES forecasting; as it did also for predictive density forecasting. Finally, the RR-RG-tG model was also highly competitive in 1% VaR forecasting, marginally outperformed only by the "Mean" series and SSRR-RG-tG.

In terms of employing forecast combination to produce an ES series that is robust to the GFC, the "Mean" series showed the most potential, having the closest overall ES violation rate to nominal. However, the "Median", "Min" and "Max" series were not competitive with many of the individual models, on these criteria.

2.7 Chapter summary

In this chapter, the realized range, observed at a 5 minute frequency, was proposed as an alternative realized measure for use in the Realized GARCH modelling framework. This choice led to significant improvements in the out-of-sample predictive likelihood and the forecasting of tail risk measures VaR and ES, compared to RG models employing realized volatility or intra-day range, and traditional GARCH models, as well as forecast combinations of these models; when combined with Student-t errors in the observation equation. In addition, we employ an existing scaling process and propose a sub-sampling process for RR to consider the micro-structure noise, then the scaled and sub-sampled realized measures are employed and tested in the Realized GARCH framework and demonstrate accurate tail forecasting results, especially with SSRR. The Realized GARCH model with RR and SSRR employing Student-t error should be considered for financial applications requiring volatility or tail risk forecasting, and should allow financial institutions to more accurately allocate capital under the Basel Capital Accord to protect their investments from extreme market movements. This work could be extended by alternative frequencies of observation for the realized measures and considering jumps in returns. In addition, we would apply and test different skew-t distributions (Aas and Haff, 2006) in the Realized-GARCH framework. Further, we could extend the adapted MCMC algorithm to employ more flexible and fat-tail mixture distribution as the MCMC proposal distribution, such as Adaptive Mixture of Student-t (AdMit) distribution (Hoogerheide, Kaashoek and van Dijk HK, 2007), to further improve the MCMC performance.

3

Chapter 3

Bayesian semi-parametric Realized CARE models for tail-risk forecasting incorporating range and realized measures

3.1 Introduction

In recent decades, quantitative financial risk measurement has provided a fundamental toolkit for investment decisions, capital allocation and external regulation. Value-at-Risk (VaR) and Expected Shortfall (ES) are tail risk measures that are employed, as part of this toolkit, to help measure and control financial risk. VaR represents the market risk as one number, a quantile of the risk distribution, and has become a standard measurement for capital allocation and risk management, since it was proposed in 1993. However, VaR has been criticised because it cannot measure the expected loss for violating returns and is not mathematically coherent, in that it can favour non-diversification. ES, proposed by Artzner *et al.* (1997, 1999), gives the expected loss, conditional on returns exceeding a VaR threshold, and is a coherent measure, thus in recent years it has become more widely employed for tail risk measurement.

Volatility estimation can play a key role in calculating accurate VaR or ES forecasts. Since the introduction of the Auto-Regressive Conditionally Heteroskedastic (ARCH) model of Engle (1982) and the generalised (G)ARCH of Bollerslev

(1986), both employing squared returns as model input, many different volatility estimators and volatility models have been developed. However, Parkinson (1980) and Garman and Klass (1980) considered the daily high-low range as a more efficient volatility estimator compared to the daily return. The availability of high frequency intra-day data has generated several more popular and efficient realized measures of volatility, including realized variance (RV): Andersen and Bollerslev (1998), Andersen *et al.* (2003); and realized range (RR): Martens and van Dijk (2007), Christensen and Podolskij (2007). In order to further deal with the well-known, inherent micro-structure noise accompanying high frequency volatility measures, Zhang, Mykland and Aït-Sahalia (2005) and Martens and van Dijk (2007) designed the sub-sampling and scaling processes, respectively, aiming to provide smoother and more efficient realized measures.

Hansen *et al.* (2011) extended the GARCH model framework by proposing the Realized GARCH (Re-GARCH), adding a measurement equation that contemporaneously links unobserved volatility with a realized measure. Gerlach and Wang (2016) extended the Re-GARCH model through employing RR as the realized measure (called RR-GARCH) and illustrated that the proposed RR-GARCH framework can generate more accurate and efficient volatility, VaR and ES forecasts compared to traditional GARCH and Re-GARCH models. However, the tail-risk forecast performance of these parametric volatility models heavily depends on the choice of error distribution. A semi-parametric model that directly estimates quantiles and expectiles, and implicitly ES, called the Conditional Autoregressive Expectile (CARE) model is proposed by Taylor (2008). The relevant expectile can be estimated with Asymmetric Least Square (ALS), which is transformed to be an estimate of ES through a connection discovered by Newey and Powell (1987). Gerlach, Chen and Lin (2012) developed the non-linear family of CARE models and an associated Bayesian estimation framework. Further, Gerlach and Chen (2016) extended CARE type models through employing daily high-low range as input.

In this chapter, a Realized Conditional Autoregressive Expectile (Re-CARE) framework is proposed, which is roughly analogous to the Re-GARCH framework. The Re-CARE includes the CARE model but adds a measurement equation that links the latent conditional expectile with the realized measure. The work in Gerlach and Chen (2016) allows a likelihood formulation for CARE models, giving an MLE that is equivalent to the ALS estimator. A standard parametric assumption on the

errors of the Re-CARE measurement equation allows this formulation to be extended, permitting an Re-CARE likelihood to be developed and an ML estimator to be explored. Further, an adaptive Bayesian MCMC algorithm is developed. To evaluate the performance of the proposed Re-CARE models, employing the range and various realized measures as inputs, the accuracy of VaR and ES forecasts will be assessed and compared with competitors such as the CARE, GARCH and Re-GARCH models.

The CARE model includes a nuisance parameter, currently not estimable by standard methods, for which Taylor (2008) employed a grid search estimator. A quadratic approximation method, followed by a refined grid search, is proposed as an alternative, substantially reducing the computing time in estimating this parameter, whilst maintaining an equivalent level of accuracy.

This chapter is organized as follows: Section 3.2 reviews some realized measures. Expectile and their connection with existing CARE type models, as well as a review of Re-GARCH type models comprises Section 3.3. Section 3.4 proposes the Realized CARE type models. The associated likelihood and the adaptive Bayesian MCMC algorithm for parameter estimation are presented in Section 3.5. The simulation and empirical studies are discussed in Section 3.6 and Section 3.7 respectively. Section 3.8 concludes the chapter and discusses future work.

3.2 Realized measures

The motivation and calculation details of different volatility estimators have been presented and discussed in Section 2.2, thus we only briefly review them in this section.

The most commonly used daily log return r_t is calculated as Equation (2.1). The high-low (squared) range (Equation (2.3)), proposed by Parkinson (1980), proved to be a much more efficient volatility estimator than r_t^2, based on the range distribution theory (see e.g. Feller, 1951). 4log(2) scales Ra to be an approximately unbiased volatility estimator. Several other range-based estimators, e.g. Garman and Klass (1980); Rogers and Satchell (1991); Yang and Zhang (2000) were subsequently proposed; see Molnár (2012) for a full review regarding their properties.

The range allowing for overnight price jumps (Equation (2.4)) is proposed in Gerlach and Chen (2016), where again the associated volatility estimator squares Rao_t, then divides by 4log(2).

If each day t is divided into N equally sized intervals of length Δ, several high frequency volatility measures can be calculated. Then RV is proposed by Andersen and Bollerslev (1998) as presented in Equation (2.5). Martens and van Dijk (2007) and Christensen and Podolskij (2007) developed the Realized Range, which sums the squared intra-period ranges (Equation(2.6)).

Through theoretical derivation and simulation, Martijns and van Dijk (2007) showed that RR is a competitive, and sometimes more efficient, volatility estimator than RV under some micro-structure conditions and levels. Gerlach and Wang (2016) confirm that RR can provide extra efficiency in empirical tail risk forecasting, when employed as the measurement equation variable in an Re-GARCH model. To further reduce the effect of microstructure noise, Martens and van Dijk (2007) presented a scaling process, as in Equations (2.7) and (2.8). This scaling process is inspired by the fact that the daily squared return and range are each less affected by micro-structure noise than their high frequency counterparts, thus can be used to scale and smooth RV and RR, creating less micro-structure sensitive measures.

Further, a sub-sampled process is proposed by Zhang, Mykland and Aït-Sahalia (2005), also to deal with micro-structure effects. The sub-sampled RV and sub-sampled RR are calculated with Equations (2.9) and (2.11) respectively.

3.3 Expectile and CARE type models

3.3.1 Expectile

The τ level expectile μ_τ, defined by Aigner, Amemiya and Poirier (1976), can be estimated through minimising the following expectation:

$$\mathrm{E}(|\tau - I(Y < \mu_\tau)|(Y - \mu_\tau)^2)$$

where Y is a continuous r.v., $\tau \in [0, 1]$, $I(Y < \mu_\tau)$ equals 1 when $Y < \mu_\tau$ and 0 otherwise. If $Y = y_1, y_2, ...y_n$, the following asymmetric sum of squares equation

is employed for μ_τ estimation in Taylor (2008):

$$\sum_{t=1}^{n}(|\tau - I(y_t < \mu_\tau)|(y_t - \mu_\tau)^2) \ , \tag{3.1}$$

minimising this equation results in the Asymmetric Leaste Squares (ALS) estimator. No distributional assumption is required to estimate μ_τ here.

Figure 3.1 plots the α level quantiles and unconditional τ level expectiles versus α and τ , respectively. As can be seen, when the unconditional α quantile and τ expectile are identical, the value of τ is more extreme than the value of α. Further, we would like to emphasize that α, τ and the estimated μ_τ all have one-to-one relationship as presented in Figure 3.1 and Equation (3.1).

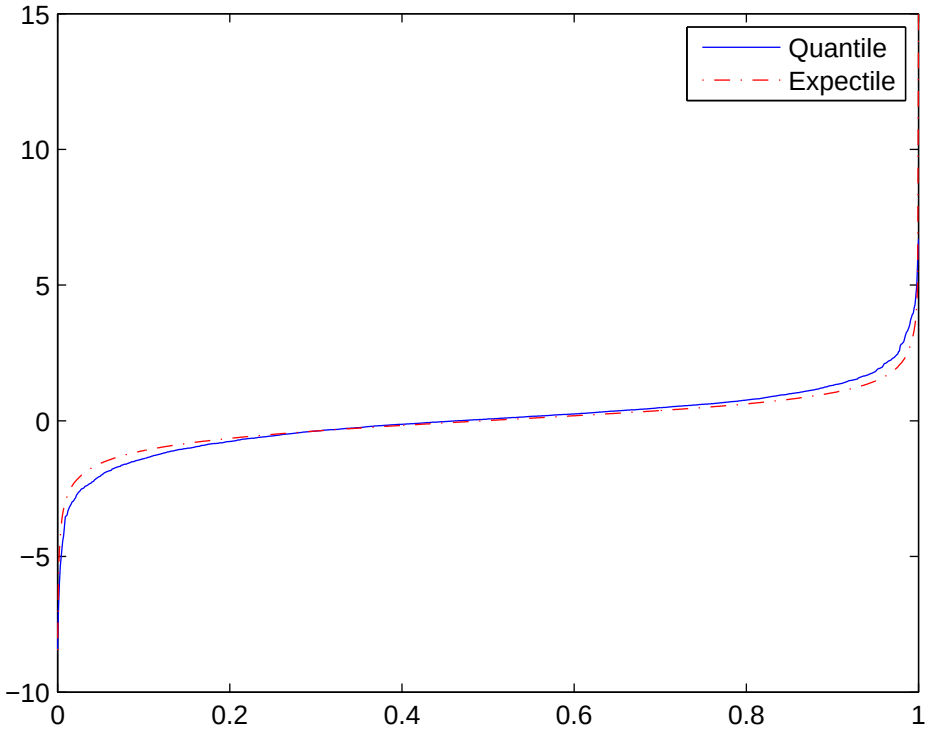

FIGURE 3.1: α level quantile and τ level expectile versus corresponding α and τ.

As discussed in Section 1.6, ES is defined as $\text{ES}_\alpha = E(Y|Y < Q_\alpha)$, which stands for the expected value of Y conditional on the set of Y that is more extreme than Q_α. Newey and Powell (1987) found a relationship between the expectile and ES:

If $\mathrm{E}(Y) = 0$, Taylor (2008) showed this relationship can be formulated as:

$$\mathrm{ES}_\alpha = (1 + \frac{\tau}{(1-2\tau)\alpha_\tau})\mu_\tau \ , \tag{3.2}$$

where $\mu_\tau = Q_\alpha$. Thus, μ_τ can be used to estimate the α level quantile Q_α, and scaled to estimate the associated ES.

3.3.2 CARE type models and Realized GARCH

Taylor (2008) proposed the CARE type models that have the similar form as the CAViaR type models (Engle and Manganelli 2004), i.e. symmetric absolute value (SAV), asymmetric (AS) and indirect GARCH (IG). The CARE type models were extended into fully nonlinear family in Gerlach, Chen and Lin (2012). Here we only present the CARE-SAV model:

CARE-SAV:

$$\mu_t = \beta_1 + \beta_2 \mu_{t-1} + \beta_3 |r_t|$$

where r_t is the day t return, and μ_t is the τ level expectile for day t, while τ is removed from the notation for the reason of brevity. Further, Gerlach and Chen (2016) employed the Range in the CARE framework, and the Ra-CARE type models demonstrated superiority compared to the CARE using return in the tail risk forecasting. The Range-CARE-SAV is specified as:

Range-CARE-SAV

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t \\ \mu_t &= \beta_1 + \beta_2 \mu_{t-1} + \beta_3 Ra_{t-1} \\ \varepsilon_t &\sim AG(\tau, 0, \sigma) \end{aligned} \tag{3.3}$$

where AG is the Asymmetric Gaussian distribution. Both the CARE-SAV and Range-CARE-SAV can be estimated by ALS, or by maximum likelihood (ML) assuming the AG error distribution: these estimators are mathematically equivalent. Thus, the AG is only employed so as to construct a likelihood function that subsequently allows a Bayesian estimator as in Gerlach, Chen, and Lin (2012) and Gerlach and Chen (2016).

These models can all produce one-step-ahead forecasts of μ_t (expectiles), which can be directly used as the VaR estimates. Then, Equation (3.2) can be employed to produce forecasts of ES simultaneously.

An innovative Realized GARCH framework was developed in Hansen *et al.* (2011). Comparing to the conventional GARCH model, Re-GARCH employed a measurement equation which captures the contemporaneous connection between unobserved volatility and the realized variance. The superiority of Re-GARCH compared to GARCH has been demonstrated by several authors, including Hansen *et al.* (2011) and Watanabe (2012).

Re-GARCH

$$\begin{aligned} r_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \beta\sigma_{t-1}^2 + \gamma x_{t-1}, \\ x_t &= \xi + \varphi\sigma_t^2 + \tau_1 z_t + \tau_2(z_t^2 - 1) + \sigma_\varepsilon \varepsilon_t \end{aligned} \tag{3.4}$$

where the 3rd equation is the measurement equation. Here $z_t \overset{\text{i.i.d.}}{\sim} D_1(0,1)$ and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} D_2(0,1)$; Hansen *et al.* (2011) suggested $x_t = RV_t$ and $D_1(0,1) = D_2(0,1) \equiv N(0,1)$. Watanabe (2012), Contino and Gerlach (2014) further extended the model through incorporating the Student-t or skewed-t (Hansen, 1994) for either or both D_1, D_2. Gerlach and Wang (2016) also allowed $x_t = RR_t, Ra_t^2$.

3.4 Model proposed

Inspired by the CARE type models and the Re-GARCH framework, the Realized-CARE-SAV is proposed, as follows:

Realized-CARE-SAV (Re-CARE-SAV)

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t \\ \mu_t &= \beta_1 + \beta_2\mu_{t-1} + \beta_3 x_{t-1} \\ x_t &= \xi + \phi|\mu_t| + u_t \end{aligned} \tag{3.5}$$

where $r_t = [\log(C_t) - \log(C_{t-1})] \times 100$ is the percentage log-return for day t, $\varepsilon_t \overset{\text{i.i.d.}}{\sim} AG(\tau, 0, \sigma), u_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2)$. The three equations in the Realized CARE are named as: the *return equation*, the *CARE equation* and the *measurement equation*, respectively. The measurement equation here captures the contemporaneous dependence between the expectile μ_t and realized measure x_t, analogous to capturing that between unobserved volatility and the realized measure in the Re-GARCH framework.

Through choosing x_t as Ra_t, $\sqrt{RV_t}$ and $\sqrt{RR_t}$ respectively, we propose the Realized-CARE-SAV-Range (Re-CARE-SAV-Ra), Realized-CARE-SAV-Realized Variance (Re-CARE-SAV-RV) and Realized-CARE-SAV-Realized Range (Re-CARE-SAV-RR) models.

The Re-CARE framework can be easily extended into the asymmetric, indirect GARCH and other nonlinear CARE versions (Engle and Manganelli, 2004; Taylor, 2008; Gerlach, Chen and Lin, 2012; Gerlach and Chen, 2016), though we focus solely on the Re-CARE-SAV in this chapter. For example, the Re-CARE with indirect GARCH specification can be written:

Realized-CARE-IG (Re-CARE-IG)

$$
\begin{aligned}
& r_t = \mu_t + \varepsilon_t \\
& \mu_t = -\sqrt{\beta_1 + \beta_2 \mu_{t-1}^2 + \beta_3 x_{t-1}^2} \\
& x_t^2 = \xi + \phi \mu_t^2 + u_t \\
& \varepsilon_t \overset{\text{i.i.d.}}{\sim} AG(\tau, 0, \sigma), u_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2) \\
& x_t = Ra_t, \sqrt{RV_t}, \sqrt{RR_t}, \\
& \beta 1 > 0, \beta 2 > 0, \beta 3 > 0
\end{aligned}
$$

In order to guarantee that the μ_t does not diverge, it is logical that a necessary condition for Re-CARE-SAV type models is $\beta_2 + \beta_3 \phi < 1$. This can be derived through substituting the measurement equation into the CARE equation in Model (3.5). The CARE equation in Re-CARE framework can produce one-step-ahead expectile forecasts (VaR), which can be mapped to ES forecasts directly through employing Equation (3.2).

In this chapter, the Realized GARCH (Hansen *et.al*, 2011) is also adapted by setting the volatility equation as an absolute value GARCH specification (Taylor (1986); Schwert (1989)), as follows:

Realized-GARCH-Abs (Re-GARCH-Abs)

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t^* \\
\sigma_t &= \beta_1^* + \beta_2^* \sigma_{t-1} + \beta_3^* x_{t-1} \\
x_t &= \xi^* + \phi^* \sigma_t + u_t^* \\
\varepsilon_t^* &\overset{\text{i.i.d.}}{\sim} D_1(0,1), u_t^* \overset{\text{i.i.d.}}{\sim} D_2(0, \sigma_{u^*}^2) \\
x_t &= R_t, \sqrt{RV_t}, \sqrt{RR_t},
\end{aligned}
$$

This model allows us to simulate from the Re-CARE-SAV model for the purpose of comparing the likelihood and Bayesian estimators for that model.

3.5 Likelihood and Bayesian estimation

3.5.1 CARE likelihood function with AG

With $\mathbf{r} = (r_1, r_2, ..., r_n)'$, the ALS as specified in Equation (3.6) is employed by Taylor (2008) to estimate μ_τ, after the expectile level τ is estimated through a grid search: τ is chosen to make the violation rate (VRate) of μ_τ closest to the quantile level α.

$$\sum_{t=1}^{n} (|\tau - I(r_t < \mu_\tau)|(r_t - \mu_\tau)^2) \tag{3.6}$$

Gerlach, Chen, and Lin (2012), Gerlach and Chen (2016) developed an asymmetric Gaussian (AG) distribution and included it as the error distribution in an observation equation for a CARE model, i.e. $\varepsilon_t \sim AG(\tau, 0, \sigma)$ in (3.3). This makes the construction of a likelihood function feasible. The scale factor σ is a nuisance parameter and can be integrated out, with a Jeffreys prior. Gerlach, Chen, and Lin (2012) showed that maximizing the resulting integrated likelihood function produces identical estimation results to the ALS approach. However, the likelihood formulation allows access to powerful computational Bayesian approaches, such as adaptive MCMC algorithms, for estimation.

The CARE likelihood in this setting is:

$$L(\mathbf{r};\theta) = \Big(\sum_{t=1}^{n} |\tau - I(r_t < \mu_t(\beta)|(r_t - \mu_t(\beta))^2\Big)^{-n/2} \tag{3.7}$$

3.5.2 Realized CARE log-likelihood

Because the Re-CARE framework has a measurement equation, with $u_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2)$, the full log-likelihood function for Re-CARE (as in Model (3.5)) is the sum of the log-likelihood $\ell(\mathbf{r};\theta)$ for the CARE equation and the log-likelihood $\ell(\mathbf{x}|\mathbf{r};\theta)$ from the measurement equation. In Re-GARCH framework, the measurement equation variable contributes to volatility estimation, thus the GARCH equation in-sample and predictive log-likelihood values are improved compared to the traditional GARCH. Thus, we expect the measurement equation in the Re-CARE to also facilitate an improved estimate τ and of μ_τ.

$$\begin{aligned}
&\ell(\mathbf{r},\mathbf{x};\theta) = \ell(\mathbf{r};\theta) + \ell(\mathbf{x}|\mathbf{r};\theta) \\
&= \underbrace{(-n/2)log\Big(\sum_{t=1}^{n} |\tau - I(r_t < \mu_t(\beta)|(r_t - \mu_t(\beta))^2\Big)}_{\ell(\mathbf{r};\theta)} \\
&\underbrace{-\frac{1}{2}\sum_{t=1}^{n}\big(log(2\pi) + log(\sigma_u^2) + u_t^2/\sigma_u^2\big)}_{\ell(\mathbf{x}|\mathbf{r};\theta)}
\end{aligned}$$

where $\mathbf{u}$ is the measurement equation residual series, e.g. in the Re-CARE-SAV $u_t = x_t - \xi - \phi|\mu_t|$, $t = 1, \ldots, n$. Further, as discussed in Section 3.3, α, τ and the estimated μ_τ all have one-to-one relationship. Therefore, incorporating one grid searched τ (corresponding to α) and estimated μ_τ into the CARE or Re-CARE likelihood will generate unique solution to the maximum likelihood estimation and produce a unique corresponding set of estimated parameters.

For the Re-GARCH model framework, Hansen *et.al* (2011) studied the asymptotic properties of the quasi-maximum likelihood estimator, conjecturing a central limit theorem. Yao and Tong (1996) considered the asymptotics of ALS estimation for expectile regression and showed consistency of the estimator. Results from both

these papers allow us to conjecture the consistency and asymptotic normality of the ML estimator obtained by numerically maximising the log-likelihood function above. We leave the proofs for future work.

3.5.3 Bayesian Estimation

Given a likelihood function, Bayesian algorithms can be employed to estimate the parameters of an Re-CARE model. A two-step adaptive Bayesian MCMC method, extended from that in Contino and Gerlach (2014) is employed. First, the parameters are dived into two blocks: $\boldsymbol{\theta_1} = (\beta_1, \beta_2, \beta_3, \phi)'$ and $\boldsymbol{\theta_2} = (\xi, \sigma)'$, where groupings are chosen to maximise within group correlation of MCMC iterates; e.g. here the stationarity constraint $\beta_2 + \beta_3\phi < 1$ induces some correlation among these parameters, whilst in GARCH models the equivalent of β_1, β_2 are known to be highly correlated.

Priors are chosen to be uninformative over the possible stationarity (and positivity, where relevant) regions, e.g. $\pi(\boldsymbol{\theta}) \propto I(A)$, which is a flat prior for $\boldsymbol{\theta}$ over the region A.

An adaptive MCMC algorithm, adapted from that in Contino and Gerlach (2014) and based on Chen and So (2006), employs a random walk Metropolis (RW-M) for the burn-in period, and independent kernel Metropolis-Hastings (IK-MH) algorithm (Metropolis et al., 1953; Hastings, 1970) for the sampling period. The burn-in period uses a Gaussian proposal distribution for the random walk process of each parameter group. The covariance matrix of the proposal distribution in each block is tuned towards a target accept ratio of 23.4% (Roberts, Gelman and Gilks, 1997). Then the IK-MH sampling period incorporates a mixture of three Gaussian proposal distributions. The sample mean of last 10% of the burn-in period samples are used as the proposal mean vector, while the sample variance-covariance matrix Σ is employed so that the three Gaussian proposal var-cov matrices are: Σ, 10Σ, 100Σ respectively, where Σ is calculated as the covariance of the last 10% of the burn-in period samples, for each block.

3.5.4 Quadratic fitting for the expectile level search

As discussed, the estimation of CARE type models relies on a full grid search of the optimal expectile level τ, e.g employing M equally spaced trial values of τ on $[0, \alpha]$. Our investigations discovered that the relationship between τ and the corresponding violation rate of μ_τ, i.e. $\hat{\alpha} = \text{VRate}$ is close to monotonic and linear, while the relationship between τ and $|\text{VRate} - \alpha|$ is close to a V-shape; see top and bottom plots of Figure 3.2. To assist in finding a smaller and more refined region than $[0, \alpha]$ on which to do a grid-search, a quadratic is fit to a small number of points in this V-shaped curve and the area close to the estimated minimum value of the quadratic is employed as the refined search area. Thus, the following two-step quadratic fitting approach is proposed to accelerate the grid search process.

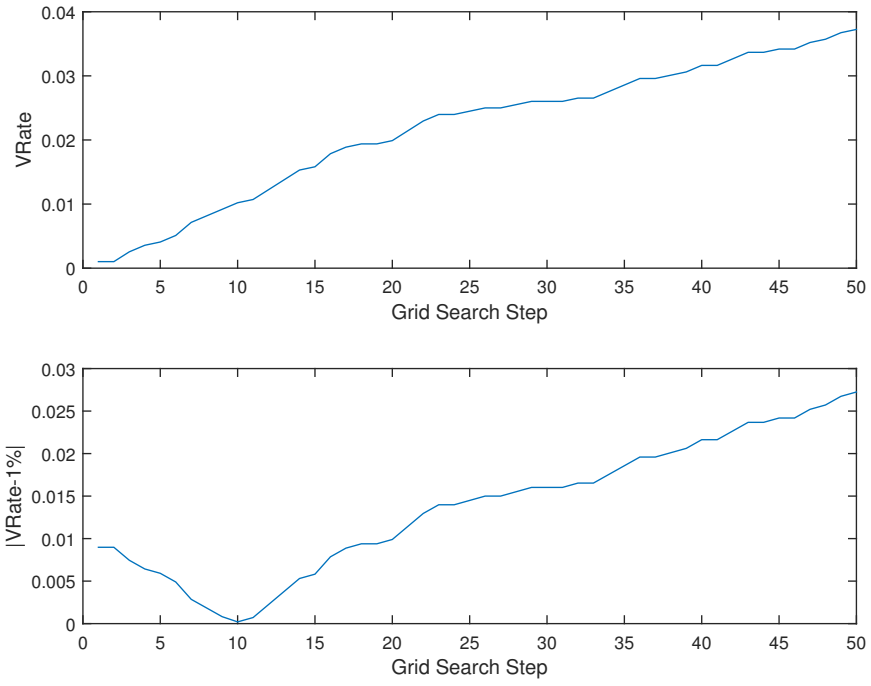

FIGURE 3.2: Expectile grid search VRate plot.

Step 1: Choose 4 equally spaced values for τ, generated in $[0.0001, \alpha/1.5]$, e.g. 0.0001, 0.0023, 0.0045, 0.0067, when $\alpha = 0.01$. This region is used because the empirical study shows $\hat{\tau}$ is always inside it. Employing $|\text{VRate} - \alpha|$ as the objective function, a quadratic function is fit to the 4 calculated points, as in Figure 3.3. Then the minimum, stationary point c is calculated, e.g. say $c = 0.0025$.

Step 2:

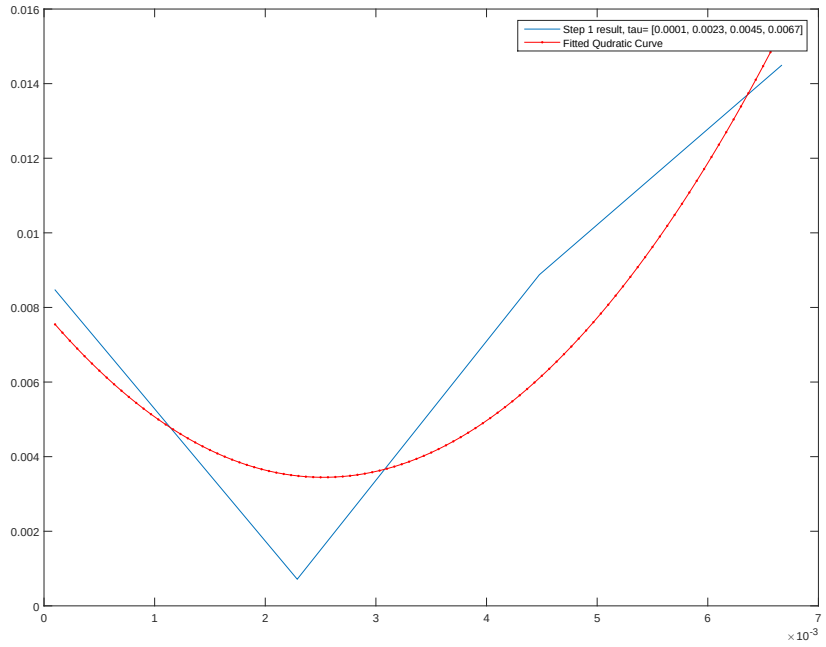

FIGURE 3.3: Expectile quadratic grid search step one results.

A focused, refined grid search around the minimum c is then conducted. Using a grid search step size of 0.0002, 8 equally spaced points on either side of c are used (as in Table 3.1 below) and then the final $\hat{\tau}$ (say 0.0021) is selected as that value minimising $|\text{VRate} - \alpha|$.

TABLE 3.1: Step 2 τ trial values of the quadratic fitting of τ grid search.

0.0009	0.0011	0.0013	0.0015	0.0017	0.0019	0.0021	0.0023	0.0025
0.0027	0.0029	0.0031	0.0033	0.0035	0.0037	0.0039	0.0041	

This approach makes savings on the original grid search time of approximately 50% to 60%, while producing estimates $\hat{\tau}$ that have the same properties as those from the full grid search, and in fact usually producing the exact same estimated value; confirmed in both our simulation and empirical studies, and highlighted in Table 3.2.

In order to study the validity of the proposed quadratic target grid search, it was examined together with the full grid search method for 5000 simulated datasets of sample size 1500. The simulated data are generated from Model (3.8) in the subsequent Section 3.6 in which we also explain how to derive the 1% true value of expectile level as $\tau = 0.001452$ for that model. The targeted grid search and

full grid search methods are employed on each dataset, generating 5000 estimated expectile levels in each case. The full grid search incorporates 50 equally distributed steps between 0.0001 and $\alpha = 1\%$, with grid search step size 0.0002. The experiments are implemented on a standard PC with Inter(R) Core(TM) i5-3470 @ 3.20 GHz CPU and 8 GB memory. For the targeted and full grid search, their mean, median and standard deviation (Std) of computation time t on 5000 replicated data sets, and mean & RMSE of the 5000 estimated τ values are calculated and presented in Table 3.2. Clearly the targeted search generated estimates with almost exactly the same properties as the full grid search, but in significantly less time, e.g. saving around 62.5% computation time on average.

TABLE 3.2: Quadratic target grid search and full grid search comparison with 5000 simulated datasets (time in seconds).

	Target Grid Search	Full Grid Search
t Mean	30.18	80.65
t Median	30.50	79.65
t Std	5.21	8.23
τ Est	0.001303	0.001304
τ RMSE	0.000398	0.000397
τ True	0.001452	0.001452

3.6 Simulation study

A simulation study is conducted to compare the properties and performance of the MCMC and ML estimation approaches for the Re-CARE model, with respect to parameter estimation and one-step-ahead VaR and ES forecasting accuracy. Both the mean and Root Mean Square Error (RMSE) values are calculated for the MCMC and ML estimates to illustrate their respective bias and precision. $N = 5000$ simulated datasets were generated from the Re-GARCH-Abs model, specified as Model (3.8). The Re-CARE-SAV model was fit to each data set, once using ML and once using MCMC. Two sample sizes for the simulated data sets are employed: n=1500 and n=3000 respectively.

Data replications are simulated from:

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t^* \\
\sigma_t &= 0.02 + 0.75\sigma_{t-1} + 0.25x_{t-1} \\
x_t &= 0.1 + 0.9\sigma_t + u_t \\
\varepsilon_t^* &\overset{\text{i.i.d.}}{\sim} N(0,1), u_t \overset{\text{i.i.d.}}{\sim} N(0, 0.3^2)
\end{aligned} \tag{3.8}
$$

In order to calculate the corresponding Re-CARE-SAV true values, a parameter mapping between from the Re-GARCH-Abs to the Re-CARE-SAV is required. With $\text{VaR}_t = \mu_t = \sigma_t \Phi^{-1}(\alpha)$, then $\sigma_t = \frac{\mu_t}{\Phi^{-1}(\alpha)} = \frac{\text{VaR}_t}{\Phi^{-1}(\alpha)}$, where $\Phi^{-1}(\alpha)$ is the standard Normal inverse cdf at α quantile level. Substituting back into the GARCH and measurement equations of Model (3.8), the corresponding Re-CARE-SAV specification can be written:

$$
\begin{aligned}
\mu_t &= 0.02\Phi^{-1}(\alpha) + 0.75\mu_{t-1} + 0.25\Phi^{-1}(\alpha)x_{t-1} \\
x_t &= 0.1 - \frac{0.9}{\Phi^{-1}(\alpha)}|\mu_t| + u_t
\end{aligned} \tag{3.9}
$$

allowing true parameter values to be read off.

In each model the true one-step-ahead α level VaR forecast is then $\text{VaR}_{n+1} = \sigma_{n+1}\Phi^{-1}(\alpha)$, and the true one-step-ahead α level ES forecast is $\text{ES}_{n+1} = \sigma_{n+1}\Phi^{-1}(\delta_\alpha)$, where δ_α is the quantile level that ES occurs at for the standard normal distribution (Gerlach and Chen, 2016). Following Basel II and Basel III risk management guidelines, the 1% quantile level is employed (corresponding $\delta_\alpha = 0.38\%$ with the standard normal distribution), then the true value of VaR_{n+1} and ES_{n+1} can be calculated for each dataset; the averages of these, over the 5000 datasets, are given in the "True" column of Table 3.3. Through the one-to-one relationship between VaR and ES (Equation (3.2)), the true value of τ is 0.001452 for this model. In addition, the quadratic fitting targeted grid search of τ is incorporated in the MCMC process, while there is no target search for τ before the ML estimation, to testify the accuracy of target search. In addition, we would like to point out that the expectile level τ would be time varying with the real data sets. However, τ is

assumed to be fixed in our study. An assumption that is also appeared in many parametric time series models with fixed parameters.

The Re-CARE-SAV model is fit to the 5000 datasets generated, once using the MCMC method and once using the ML estimator (the 'fmincon' constrained optimisation routine in Matlab software is employed). The MCMC sampler has 10000 iterations for each data set, with a burn-in of 5000 iterations. All iterations in the independent MH sampling period are used to calculate the posterior estimates.

Estimation results are summarised in Tables 3.3. Boxes indicate the preferred measure comparing MCMC and ML for both bias (Mean) and precision (RMSE). Regarding the simulation results with $n = 1500$, both methods generate close to unbiased and quite reasonably precise parameter estimates and VaR and ES forecasts. The bias results slightly favour the ML method, for 4 out of 7 parameter estimates and the ES forecasts. However, the precision clearly favors the MCMC method in 7 out of 9 parameter estimates and both VaR and ES forecasts. Extending the sample size to 3000, first it can be seen that both MCMC and ML show improved bias and precision in estimation. Second, the results are even more in favour of the MCMC method compared to $n = 1500$. Again both methods generate close to unbiased and quite reasonably precise parameter estimates and tail risk forecasts. The bias results favor the MCMC approach in 6 out of 9 parameter estimates and VaR& ES forecasts, whilst the precision is clearly lower for the MCMC method for 7 parameters and both tail risk forecasts. Finally, the estimation results for τ highlight the validity of the quadratic fitting targeted search approach. Note that the estimation of τ is neither MCMC nor ML, but the targeted procedure was only used for the MCMC results. Lastly, in order to study the convergence and efficiency performance of the employed MCMC algorithm for the simulated (and real word) data set, in this chapter we also implemented the Gelman-Rubin diagnostic and effective sample size test as discussed in Sections 1.5.4 and 2.5, with the data set simulated from Re-CARE-SAV model (also 1st in-sample S&P 500 data set). We still observe quite good convergence and efficiency results, which is not surprising since the Re-CARE-SAV framework is analogous

TABLE 3.3: Summary statistics for the two estimators of the Re-CARE-SAV model, with data simulated from model (3.8).

$n = 1500$		MCMC-Target Search		ML	
Parameter	True	Mean	RMSE	Mean	RMSE
$\beta 1$	-0.0465	**-0.0544**	**0.1923**	-0.0709	0.3258
$\beta 2$	0.7500	0.7335	0.0417	**0.7378**	**0.0393**
$\beta 3$	-0.5816	-0.6238	**0.1485**	**-0.6045**	0.2048
ξ	0.1000	0.0839	**0.2830**	**0.1018**	0.7508
φ	0.3869	**0.3879**	**0.0723**	0.3852	0.1734
σ_u	0.3000	0.03010	0.0057	**0.3005**	**0.0056**
τ	0.001452	**0.001304**	**0.0004**	0.001303	**0.0004**
1% VaR_{n+1}	-4.1872	**-4.2392**	**0.2920**	-4.2416	0.3241
1% ES_{n+1}	-4.7970	-4.7911	**0.3241**	**-4.7935**	0.3608
$n = 3000$	True	Mean	RMSE	Mean	RMSE
$\beta 1$	-0.0465	**-0.0499**	**0.1287**	-0.0577	0.2059
$\beta 2$	0.7500	**0.7422**	0.0272	0.7411	**0.0257**
$\beta 3$	-0.5816	-0.6014	**0.0976**	**-0.5925**	0.1328
ξ	0.1000	**0.0919**	**0.1947**	0.0880	0.4542
φ	0.3869	**0.3876**	**0.0503**	0.3891	0.1066
σ_u	0.3000	0.3005	0.0040	**0.3002**	**0.0039**
τ	0.001452	**0.001378**	**0.0003**	**0.001378**	**0.0003**
1% VaR_{n+1}	-4.1784	-4.1970	**0.1965**	**-4.1969**	0.2103
1% ES_{n+1}	-4.7869	**-4.7759**	**0.2255**	**-4.7759**	0.2411

Note:A box indicates the favored estimators, based on mean and RMSE.

to Re-GARCH, thus the results are not shown here.

3.7 Data and empirical study

3.7.1 Data description

Daily and high frequency data, observed at 1-minute and 5-minute frequency, including open, high, low and closing prices, are downloaded from Thomson Reuters Tick History. Data are collected for 6 market indices: S&P500, NASDAQ (both US), Hang Seng (Hong Kong), FTSE 100 (UK), DAX (Germany) and SMI (Swiss), with time range Jan 2000 to Sep 2014; as well as for 3 individual assets: IBM, GE (both US) and BHP (AU), with time range Jan 2000 to Dec 2014. The starting data collection time for BHP is July 2001 since it had a 2 : 1 Stock Split in June 2001 and the starting time for GE is May 2000, for a similar reason.

The data are used to calculate the daily return, daily range and daily range considering overnight price jump. Further, the 5-minute data are employed to calculate the daily RV and RR measures, while both 5 and 1-minute data are employed to produce daily scaled and sub-sampled versions of these measures, as in Section 3.2; $q = 66$ is employed for the scaling process, i.e. around 3 months. Thus, the final starting time is 3 months from the starting time of data collection.

The full data period is divided into an estimation sample: April 6, 2000 to December 31, 2007, of $\approx n = 1900$ days (the starting dates for IBM, GE and BHP are different); and a forecast sample: approximately $m = 1660$ trading days from January 1, 2008 to Sept 18, 2014. The forecasting period includes most, if not all, of the effects of the global financial crisis (GFC) on each market.There are small differences in forecast sample sizes and end-dates occurred across markets, due to market-specific non-trading days. Table 3.5 presents the exact in-sample sizes n and forecast sample sizes m.

3.7.2 In-sample parameters estimation results

Further, with the 1st S&P 500 in-sample data set (first 1960 observations) and MCMC, the estimated parameters of 8 different Re-CARE-SAV type models are presented in Table 3.4.

Similar to the observation in Table 2.5, we can clearly see the much smaller estimated σ_u with Re-CARE-SAV-RR compared to Re-CARE-SAV-RV. As discussed in Martins and van Dijk (2007), Christensen and Podolskij (2007), RR has much lower mean squared error than RV which might provide RR with higher accuracy and efficiency in volatility estimation and forecasting. Through looking at the σ_u values estimated from Re-CARE-SAV employing scaled and sub-sampled realized measures, the Re-CARE-SAV-SSRV provides smaller σ_u compared to Re-CARE-SAV-RV, and Re-CARE-SAV-SSRR has similar σ_u as Re-CARE-SAV-RR. However, we see slightly increased σ_u through incorporating scaled RV or RR.

3.7.3 Tail risk forecasting

Both daily Value-at-Risk (VaR) and Expected Shortfall (ES) are estimated for the 6 indices and the 3 asset series, as recommended in Basel II and III Capital

TABLE 3.4: In-sample parameters estimation for 8 models with S&P 500.

Models	β_1	β_2	β_3	ξ	ϕ	σ_u
RC-Ra	-0.0855	0.7508	-0.4296	-0.0299	0.5176	0.6025
RC-RaO	-0.0707	0.7598	-0.4152	-0.0109	0.5192	0.6140
RC-RV	-0.0564	0.6774	-0.8516	0.0072	0.3496	0.2912
RC-RR	-0.0015	0.6483	-1.4320	0.0408	0.2293	0.1720
RC-ScRV	-0.0743	0.6837	-0.7437	-0.0158	0.3922	0.3258
RC-ScRR	-0.0931	0.6507	-0.9177	-0.0427	0.3573	0.2364
RC-SSRV	-0.0793	0.6298	-1.0767	-0.0100	0.3190	0.2308
RC-SSRR	-0.0294	0.6098	-1.4335	0.0275	0.2536	0.1798

Note: RC represents the Re-CARE-SAV type models.

Accord. As discussed in Section 3.3, in the CARE setting the VaR, which is the α level quantile, can be estimated by the corresponding τ level expectile. Then through employing the one-to-one relationship between expectile and ES (as in Equation (3.2)), ES can subsequently be calculated.

A rolling window with fixed size in-sample data is employed for estimation to produce each one-step-ahead forecast, the in-sample size n is given in Table 3.5 for each series. In order to see the performance during the GFC period, the initial date of the forecast sample is chosen as the beginning of 2008. On average, 1684 VaR and ES forecasts are generated with the proposed Re-CARE type models (estimated with MCMC) with different input measures of volatility: including range, range considering overnight jump, RV & RR, scaled RV & RR and sub-sampled RV & RR. The conventional GARCH with Student-t distribution, CARE-SAV and Re-GARCH with Gaussian or Student-t as the error distributions for its volatility equation (estimated with ML), are also included, for the purpose of comparison. The actual forecast sample sizes m are given in Table 3.5. Figure 3.4 visualises $m = 1621$ estimated τ values for each forecasting step with S&P 500, for exposition.

The VaR violation rate (VRate) and ES violation rate (ESRate) are employed to evaluate the VaR and ES forecasting accuracy. VRate and ESRate are simply the proportion of returns that exceed the forecasted VaR or ES level in the forecasting period (Equations (3.10) and (3.11)). Models with VRate closest to nominal quantile level $\alpha = 1\%$ are preferred.

In addition, Gerlach and Chen (2016) presented the quantile levels where the 1% ES is estimated to fall at over many different distributional choices. For the

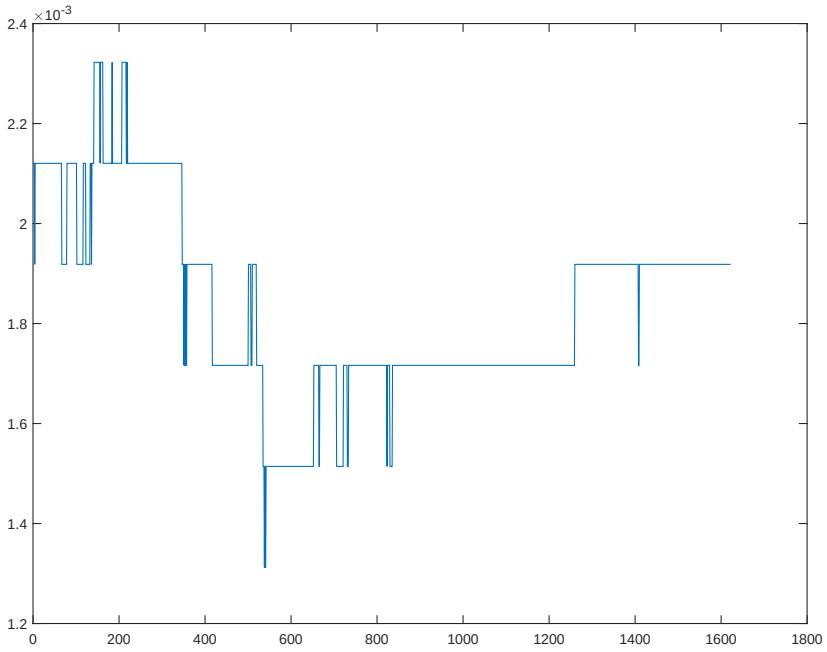

FIGURE 3.4: 1621 estimated expectile levels with S&P 500.

Gaussian this is 0.38%. For other distributions estimated in GARCH models for dailr return data, the 1% ES quantile is estimated in 0.35% to 0.37%, for a range of non-Gaussian distributions. As such, following Gerlach and Chen (2016), 0.36% is chosen as the approximate nominal expected ESRate for the ES forecasting study for CARE-type models.

$$\text{VRate} = \frac{1}{m}\sum_{t=n+1}^{n+m} I(r_t < \text{VaR}_t), \tag{3.10}$$

$$\text{ESRate} = \frac{1}{m}\sum_{t=n+1}^{n+m} I(r_t < \text{ES}_t), \tag{3.11}$$

where n is the in-sample size and m is the forecasting sample size.

However, having a VRate or ESRate close to the expected level is necessary but not sufficient to guarantee an accurate forecasting model. Thus several standard quantile accuracy and independence tests are employed: e.g. the unconditional coverage (UC) and conditional coverage (CC) tests of Kupiec (1995) and Christoffersen (1998) respectively, as well as the dynamic quantile (DQ) test of Engle and

Manganelli (2004) and the VQR test of Gaglianone *et al.* (2011). With the approach of Gerlach and Chen (2016), the derived expected ES level can be used so as to treat ES forecasts as quantile forecasts at appropriate quantile levels and these same tests can be applied; in addition the standard bootstrap t-test that the ES residuals for VaR violations have mean 0 is applied for each model's ES forecast series.

3.7.3.1 Value-at-Risk

Table 3.5 presents the VRate at the 1% quantile for each model for 9 market or assets (also mean and median of the 9 VRates for each model). The estimation period sample size for each forecast is denoted as n, and the forecast sample size is represented with m, in each market. Box indicates the model in each market that has a violation rate (VRate) closest to 1%, while bold indicates the model with VRate furthest away from expected. G-t, CARE-SAVE, Re-GARCH-GG with RV and Re-GARCH-tG with RV are estimated with ML, and the Realized CARE type models are estimated with MCMC incorporating the quadratic fitting target search. The VRate in Table 3.8 is plot in Figiure 3.5 as well, with a reference line drawn at 1%.

Chang *et al.* (2011) and McAleer et al. (2013) proposed using forecast combinations of the VaR series from different models, potentially as a robust combined VaR forecast to the GFC . This approach is also employed in our empirical study since our forecasting period includes the GFC. Specifically, the mean, median, minimum and maximum of each of the VaR forecasts from the 12 models in Table 3.5 are considered. We consider the lower tail VaR forecasts in this chapter, so "Min" is the most extreme of the 12 forecasts (i.e. furthest from 0) and "Max" is the least extreme. The violation rate for "Mean", "Median", "Min" and "Max", series are also presented in Table 3.5.

Clearly, Re-CARE-SAV employing sub-sampled RR has VRate that is the closest to the 1% quantile level based on the mean of VRates on 9 time series studied, and Re-CARE-RR has the closest to the expected VRate with the median. In addition, we can apparently observe the generally improved performance of Re-CARE compared to GARCH, Re-GARCH or CARE-SAV, while Re-GARCH-GG had the VRate that is furthest from that expected, which is not surprising since it is the only parametric model employing the Gaussian error. Further, regarding

the combination approach, the "Min" approach is too conservative in each series, while the "Max" series produce anti-conservative VaR forecasts that generate far too many violations. The "Mean" and "Median" of the 12 models produced series that generated competitive violation rates.

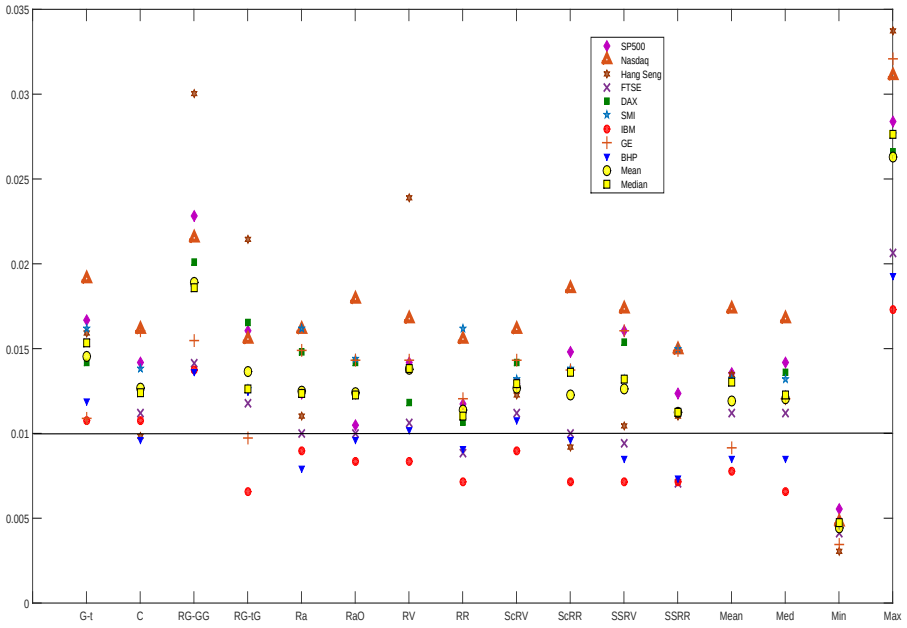

FIGURE 3.5: 1% VaR Forecasting VRates.

Several tests are employed to statistically assess the forecast accuracy and independence of violations from each VaR model. Table 3.6 shows the number of return series (out of 9) in which each 1% VaR forecast model is rejected for each test, conducted at a 5% significance level. The Re-CARE type models are generally less likely to be rejected by these various back tests compared to other models, while the Re-CARE with RaO achieved the least number of rejections, following by the Re-CARE-RR, Re-CARE-ScRV, Re-CARE-ScRR and "Mean" combination approach (rejected 3 times in total respectively). The "Min" and "Max" combinations are rejected in all 9 series, and G-t and Re-GARCH-GG models are rejected 8 times respectively.

3.7.3.2 Expected Shortfall

One-step-ahead daily ES forecasts are generated for the same 12 models and 9 series during the forecast sample periods. Regarding the expected level of ESRate

TABLE 3.5: 1% VaR Forecasting VRate with different models on 6 indices and 3 assets.

Model	S&P 500	NASDAQ	HK	FTSE	DAX	SMI	IBM	GE	BHP	Mean	Median
G-t	1.67%	1.91%	1.59%	1.53%	1.42%	1.62%	1.07%	1.09%	1.19%	1.454%	1.532%
CARE	1.42%	1.61%	0.98%	1.12%	1.24%	1.38%	1.07%	1.60%	0.97%	1.267%	1.242%
RG-RV-GG	2.28%	2.15%	3.00%	1.41%	2.01%	1.86%	1.37%	1.55%	1.36%	**1.888%**	**1.861%**
RG-RV-tG	1.60%	1.56%	2.15%	1.18%	1.66%	1.26%	0.66%	0.97%	1.25%	1.366%	1.261%
RC-Ra	1.23%	1.61%	1.10%	1.00%	1.48%	1.62%	0.90%	1.49%	0.80%	1.248%	1.234%
RC-RaO	1.05%	1.79%	1.23%	1.00%	1.42%	1.44%	0.84%	1.43%	0.97%	1.241%	1.226%
RC-RV	1.42%	1.67%	2.39%	1.06%	1.18%	1.38%	0.84%	1.43%	1.02%	1.377%	1.381%
RC-RR	1.17%	1.56%	1.10%	0.88%	1.06%	1.62%	0.72%	1.20%	0.91%	1.136%	1.104%
RC-ScRV	1.30%	1.61%	1.23%	1.12%	1.42%	1.32%	0.90%	1.43%	1.08%	1.268%	1.295%
RC-ScRR	1.48%	1.85%	0.92%	1.00%	1.36%	1.38%	0.72%	1.37%	0.97%	1.228%	1.360%
RC-SSV	1.60%	1.73%	1.04%	0.94%	1.54%	1.32%	0.72%	1.60%	0.85%	1.260%	1.321%
RC-SSRR	1.23%	1.50%	1.10%	0.71%	1.12%	1.50%	0.72%	1.49%	0.74%	1.123%	1.124%
Mean	1.36%	1.73%	1.35%	1.12%	1.30%	1.32%	0.78%	0.92%	0.85%	1.192%	1.301%
Median	1.42%	1.67%	1.23%	1.12%	1.36%	1.32%	0.66%	1.20%	0.85%	1.204%	1.226%
Min	0.56%	0.48%	0.31%	0.41%	0.47%	0.48%	0.48%	0.34%	0.45%	0.442%	0.473%
Max	2.84%	3.11%	3.37%	2.06%	2.66%	2.76%	1.73%	3.21%	1.93%	2.631%	2.761%
m	1621	1672	1631	1697	1691	1666	1675	1746	1760	1684.33	1675
n	1960	1892	1890	1944	1936	1930	1916	1839	1569	1875.11	1916

Note:A box indicates the favored individual model based on mean or median VRate, whilst bold indicates the least favoured model. m is the out-of-sample size, and n is in-sample size. SAV stands for the CARE-SAV model, RG stands for the Realized GARCH type models, and RC represents the Re-CARE-SAV type models.

TABLE 3.6: Counts of 1% VaR rejections with UC, CC, DQ and VQR tests for different models on 6 indices and 3 assets.

Model	UC	CC	DQ4	VQR	Total
G-t	5	3	7	3	**8**
CARE	2	1	6	0	6
RG-RV-GG	6	5	5	6	**8**
RG-RV-tG	4	2	2	4	5
RC-Ra	2	2	4	1	4
RC-RaO	1	1	2	1	2
RC-RV	2	2	3	2	5
RC-RR	2	1	3	1	3
RC-ScRV	1	2	2	0	3
RC-ScRR	1	2	3	0	3
RC-SSRV	4	2	4	1	6
RC-SSRR	0	1	3	1	4
Mean	1	2	3	2	3
Median	1	2	2	3	4
Min	9	4	2	5	9
Max	9	9	9	8	9

Note:A box indicates the individual model with least number of rejections, whilst bold indicates that with the highest number os rejections. All tests are conducted at 5% significance level.

for different models and distributions, Chen, Gerlach and Lu (2012) discussed how to treat ES forecasts from parametric models as quantile forecasts, where the quantile level that ES falls at can be deduced exactly (e.g. 0.38% for a Gaussian). Gerlach and Chen (2016) illustrate that the quantile level that the 1% ES is estimated to fall at is between 0.35% and 0.37%. Specifically, they present the expected violation rate of ES as exactly 0.38% for models with Gaussian errors, estimated by the quantile level dependent on the degrees of freedom estimated for models with Student-t errors ($\approx$ 0.36% for the time series considered here), and $\approx$ 0.36% for non-parametric models. With this approach, we can then treat ES forecasts as quantile forecasts and employ the UC, CC, DQ and VQR tests with corresponding ES nominal level to test the ES forecasting accuracy and independence of violations.

Table 3.8 presents the ESRate in the forecast period at the 1% quantile for each model in 9 time series. Similar to the VaR study, box indicates the model in each market that has an ES violation rate closest to that desired, and bold indicates the model with ESRate furthest from the corresponding nominal level. Figure 3.6 visualises the ESRate in Table 3.8, with a reference line drawn at 0.36%.

Clearly, the Re-CARE-RR and Re-CARE-SSRV generate ES forecasts with ESRates closest to that expected for the 1% ES across the 6 markets and 3 assets. Their overall ESRate is also marginally lower than the 0.36% nominal level, i.e. the ES forecasts from these two models are marginally conservative, which could be a plus from a risk management point of view. Further, the Re-CARE type models have clearly better performance at 1% ES forecasting than the GARCH, Re-GARCH and CARE-SAV models. The mean, median, min and max of the 12 models' 1% ES forecasts are again calculated and their ES violations are also shown in Table 3.8. The "Mean" and "Median" approaches are again optimal among the four combination methods, and competitive overall, though their ESRates are still somewhat above nominal overall and thus anti-conservative.

Furthermore, Figure 3.7 demonstrates the extra efficiency that can be gained by employing the Re-CARE framework with RR. Specifically, the ESRate of the Gt, CARE-SAV and Re-CARE-RR are 0.68%, 0.37% and 0.31% respectively for this series. These violation rates mean Gt generated anti-conservative ES forecasts that produced too many violations, while CARE-SAV is more conservative and close to nominal level ESRate and the Re-CARE-RR is conservative here. Through close inspection of Figure 3.7, e.g. the close to end of the forecasting period, CARE-SAV has an obviously lower (i.e. more extreme in the negative direction) level of ES forecasts than Gt, in order to be conservative, but this also means the capital set aside by financial institutions to cover extreme losses, based on such ES forecasts, is more with the CARE-SAV than with Gt, as expected. However, we can clearly observe the Re-CARE-RR produces ES forecasts that are less extreme than both the CARE-SAV and Gt models here, meaning that lower amounts of capital are needed to protect against market risk, while simultaneously producing fewer ES violations. This suggests a higher level of efficiency of the Re-CARE-RR model, in that this model can produce ES forecasts that have far fewer and close to expected violations, but are simultaneously less extreme than those of the traditional GARCH and CARE-SAV model. Since the capital set aside by financial institutions should be directly proportional to the ES forecast, the Re-CARE-RR model is saving the company money, by giving more accurate and often less extreme ES forecasts, and this extra efficiency is also often observed for Re-CARE type models in the other markets/assets.

Further, at times of GFC when there is a persistence of extreme returns, close inspection of Figure 3.7 reveals that the Re-CARE-RR ES forecasts "recover" the

fastest among the 3 models presented, in terms of being marginally the fastest to produce forecasts that again follow the tail of the data. GARCH models tend to over-react to extreme events and to be subsequently very slow to recover, due to their oft-estimated very high level of persistence.

Back testing is conducted on all the ES forecasts and results are shown in Table 3.7. The UC, CC, DQ and VQR quantile accuracy tests are applied to the ES violations from each model, using that model's nominal 1% ES quantile level. In addition, the averages of the 1% ES forecast residuals, standardised by the 1% VaR forecasts are also calculated. Given an accurate 1% ES forecast model should produce standardised residuals that average approximately 0, a bootstrap test on whether these averages differ from 0 is also performed and presented in Table 3.7. As can be seen, models with least number of rejections are Re-CARE-RaO and Re-CARE-RV, only rejected 2 out 9 time series. They are followed by Re-GARCH-tG, Re-CARE-Ra, Re-CARE-ScRV, Re-CARE-ScRR, Re-CARE-SubRV, "Mean", "Median" and "Min" approaches (rejected 3 times in total respectively).

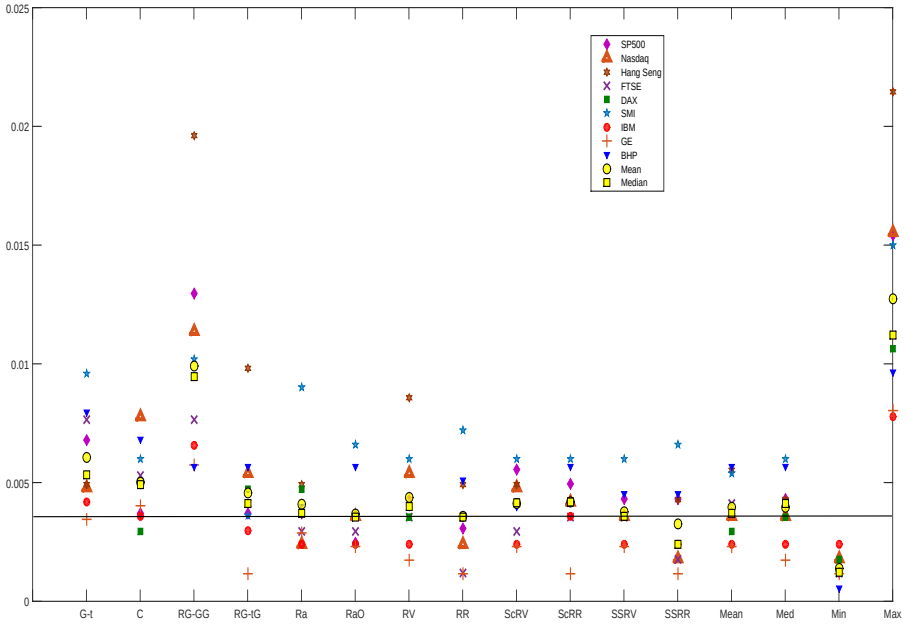

FIGURE 3.6: 1% ES forecasting ESRates.

TABLE 3.7: Counts of 1% ES rejections with UC, CC, DQ, VQR and bootstrap tests for different models on 6 indices and 3 assets.

Model	UC	CC	DQ4	VQR	Bootstrap	Total
GARCH-t	3	3	6	3	2	7
CARE	2	1	4	1	0	5
RG-RV-GG	6	6	6	2	2	**8**
RG-RV-tG	2	1	2	1	0	3
RC-Ra	1	1	3	0	0	3
RC-RaO	0	0	2	0	0	2
RC-RV	1	1	2	1	1	2
RC-RR	2	0	2	1	1	4
RC-ScRV	0	0	3	0	0	3
RC-ScRR	1	0	2	0	0	3
RC-SSRV	0	0	2	0	1	3
RC-SSRR	1	0	2	0	1	4
Mean	0	0	3	0	0	3
Median	0	0	3	0	0	3
Min	2	1	0	1	2	3
Max	9	9	8	6	3	9

Note:A box indicates the individual model with least number of rejections, whilst bold indicates that with the highest number os rejections. All tests are conducted at 5% significance level.

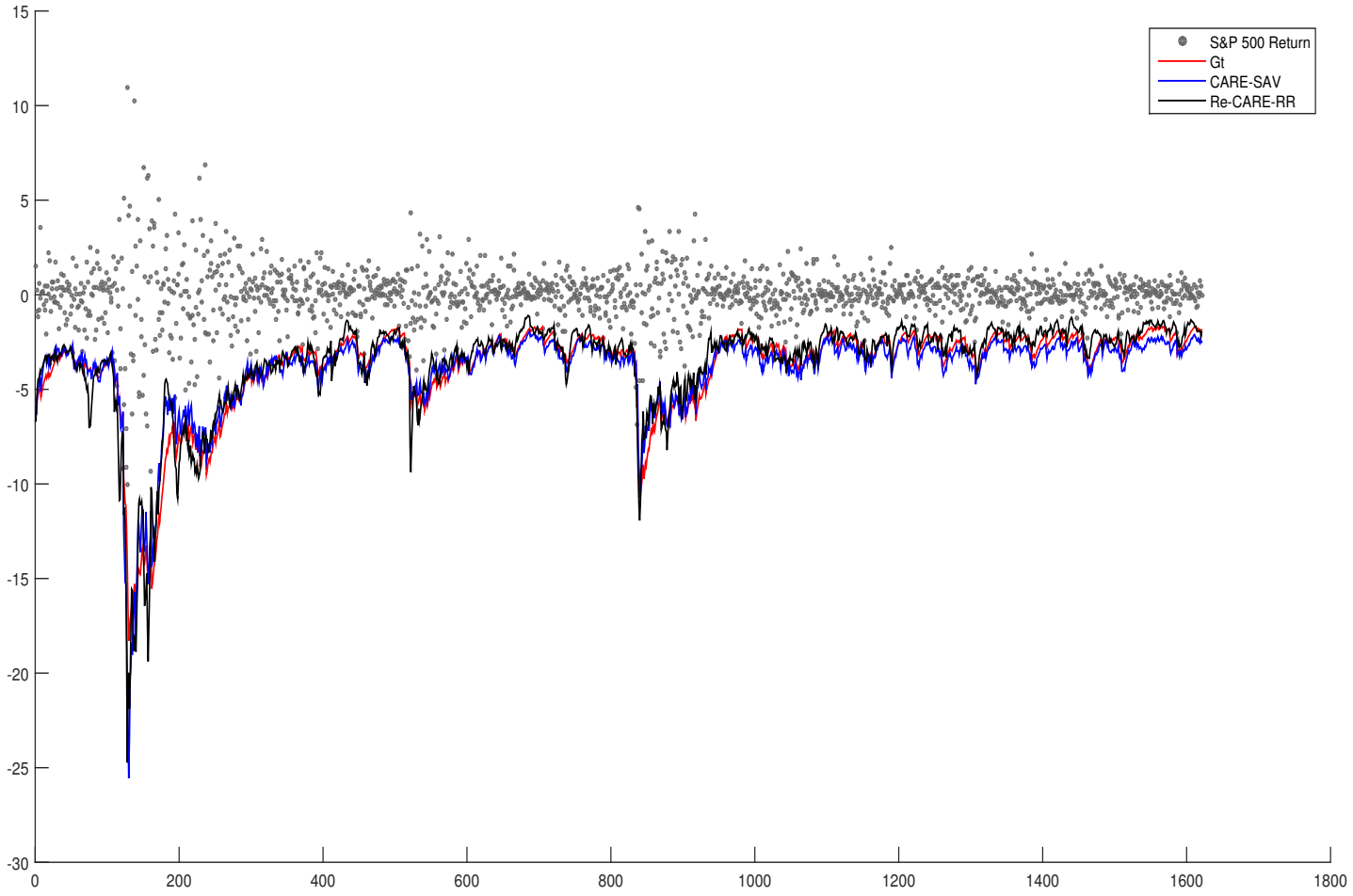

FIGURE 3.7: S&P 500 1% ES Forecasts with Gt, CARE-SAV and Re-CARE-RR.

TABLE 3.8: 1% ES Forecasting ESRates with different models on 6 indices and 3 assets.

Model	S&P 500	NASDAQ	HK	FTSE	DAX	SMI	IBM	GE	BHP	Mean	Median
G-t	0.68%	0.48%	0.49%	0.77%	0.53%	0.96%	0.42%	0.34%	0.80%	0.608%	0.532%
CARE	0.37%	0.78%	0.49%	0.53%	0.30%	0.60%	0.36%	0.40%	0.68%	0.501%	0.490%
RG-RV-GG	1.30%	1.14%	1.96%	0.77%	0.95%	1.02%	0.66%	0.57%	0.57%	**0.993%**	**0.946%**
RG-RV-tG	0.37%	0.54%	0.98%	0.41%	0.47%	0.36%	0.30%	0.11%	0.57%	0.457%	0.412%
RC-Ra	0.37%	0.24%	0.49%	0.29%	0.47%	0.90%	0.24%	0.29%	0.40%	0.410%	0.370%
RC-RaO	0.25%	0.36%	0.37%	0.29%	0.35%	0.66%	0.24%	0.23%	0.57%	0.369%	0.355%
RC-RV	0.43%	0.54%	0.86%	0.35%	0.35%	0.60%	0.24%	0.17%	0.40%	0.438%	0.398%
RC-RR	0.31%	0.24%	0.49%	0.12%	0.35%	0.72%	0.36%	0.11%	0.51%	0.357%	0.355%
RC-ScRV	0.56%	0.48%	0.49%	0.29%	0.41%	0.60%	0.24%	0.23%	0.40%	0.411%	0.414%
RC-ScRR	0.49%	0.42%	0.43%	0.35%	0.41%	0.60%	0.36%	0.11%	0.57%	0.416%	0.419%
RC-SSRV	0.43%	0.36%	0.37%	0.35%	0.35%	0.60%	0.24%	0.23%	0.45%	0.376%	0.359%
RC-SSRR	0.43%	0.18%	0.43%	0.18%	0.24%	0.66%	0.24%	0.11%	0.45%	0.324%	0.239%
Mean	0.37%	0.36%	0.55%	0.41%	0.30%	0.54%	0.24%	0.23%	0.57%	0.396%	0.370%
Median	0.43%	0.36%	0.43%	0.41%	0.35%	0.60%	0.24%	0.17%	0.57%	0.396%	0.412%
Min	0.12%	0.18%	0.12%	0.12%	0.18%	0.12%	0.24%	0.11%	0.06%	0.139%	0.123%
max	1.54%	1.56%	2.15%	1.12%	1.06%	1.50%	0.78%	0.80%	0.97%	1.275%	1.120%

Note:A box indicates the favored individual model based on mean or median ESRate, whilst bold indicates the least favoured model.

3.7.3.3 Loss function

Quantiles are elicitable, in the sense defined by Gneiting and Ranjan (2012), since the standard quantile loss function is strictly consistent, i.e. the expected loss is a minimum at the true quantile series. Thus, the most accurate quantile forecasting model should minimise the quantile loss function, given as:

$$\sum_{t=n+1}^{n+m} (\alpha - I(y_t < q_t))(y_t - q_t) \ , \tag{3.12}$$

where $q_{n+1}, \ldots, q_{n+m}$ is a series of quantile forecasts at level α for the observations $y_{n+1}, \ldots, y_{n+m}$. Each model in this study produced both a series of quantile forecasts and ES forecasts. Table 3.9 gives the calculated quantile loss function values for each return series and each model's VaR forecasts. Table 3.10 gives the loss function values for each model's ES forecasts, using the appropriate quantile level, e.g. $\alpha = 0.36\%$. Both tables are in order from highest to lowest in terms of average loss across the nine return series.

From Table 3.9, Re-CARE type models' 1% VaR forecast series have the lowest loss in 8 of the 9 series, excepting IBM where the "Median" combined series had the lowest loss. On average and by median, over the 9 return series, the Re-CARE model employing sub-sampled RR had the lowest loss function values; marginally ahead of the "Mean" and "Median" forecast combination series. Table 3.5 shows that this is also the most accurate model in terms of average VRate. Table 3.10 shows that an Re-CARE model's 1% ES forecast series again has the lowest loss in 8 of the 9 series, excepting BHP where the CARE model's ES forecast series had the lowest loss. On average the lowest loss was for the "Median" combined series, marginally ahead of the Re-CARE model employing sub-sampled RR, which had the lowest median loss across the 9 series. In both tables, Re-CARE models clearly had consistently lower loss than all other models considered and were at least highly competitive with the forecast combined series "Mean" and "Median".

3.8 Chapter summary

In this chapter, the Realized CARE, a new model framework to estimate and forecast financial tail risk, is proposed. Through incorporating intra-day and high frequency volatility measures, e.g. Ra, RaO, RV, RR, Scaled RV, Scaled RR, Sub-sampled RV and Sub-sampled RR, significant improvements in the out-of-sample

TABLE 3.9: Quantile loss function values for the VaR forecast series at $\alpha = 1\%$ over different models on the 6 indices and 3 assets.

Model	S&P 500	NASDAQ	HK	FTSE	DAX	SMI	IBM	GE	BHP	Mean	Median
Max	72.66	85.53	103.35	67.50	80.19	71.09	83.04	127.35	113.92	**89.40**	**83.04**
CARE	70.19	81.55	75.13	65.85	75.55	69.39	82.89	127.94	100.44	83.21	75.55
GARCH-t	67.88	78.05	79.09	64.73	75.90	67.51	82.74	114.38	100.66	81.22	78.05
RG-RV-GG	65.89	73.56	101.21	61.69	74.63	62.42	78.42	102.37	102.60	80.31	74.63
Min	62.83	71.22	77.03	62.09	75.14	64.95	83.16	109.38	101.14	78.55	75.14
RG-RV-tG	63.40	71.67	92.56	61.39	73.19	60.93	78.63	99.90	103.10	78.31	73.19
RC-Ra	60.82	71.25	75.26	60.55	75.85	67.77	78.47	106.14	102.49	77.62	75.26
RC-RV	60.55	73.11	86.34	61.20	73.41	61.87	77.30	105.54	94.72	77.12	73.41
RC-ScRV	61.44	73.39	76.79	61.75	75.41	63.67	78.85	107.26	95.30	77.10	75.41
RC-RaO	60.94	72.73	72.59	60.44	72.95	64.54	77.66	104.41	101.33	76.40	72.73
RC-ScRR	61.66	72.61	74.44	61.20	73.24	62.54	77.39	99.55	102.81	76.16	73.24
RC-RR	59.46	69.77	79.06	61.17	71.78	61.36	78.81	98.52	103.88	75.98	71.78
RC-SSRV	59.95	71.66	72.53	59.47	73.25	60.92	78.30	102.95	103.46	75.83	72.53
Median	60.96	71.56	74.78	60.62	72.55	61.84	77.14	100.71	98.36	75.39	72.55
Mean	61.07	71.33	75.86	60.26	72.04	62.14	77.29	100.45	97.92	75.37	72.04
RC-SSRR	59.72	69.66	75.94	59.79	71.29	61.11	78.05	101.65	100.89	75.35	71.29

Note:A box indicates the favored model based on mean or median minimum loss, whilst bold indicates the least favoured model.

TABLE 3.10: Quantile loss function values for the ES forecasts, at $\alpha = 0.36\%$ over different models on the 6 indices and 3 assets.

Model	S&P 500	NASDAQ	HK	FTSE	DAX	SMI	IBM	GE	BHP	Mean	Median
Max	30.48	38.69	54.47	29.54	33.40	32.58	41.13	53.00	55.17	**40.94**	**38.69**
RG-RV-GG	28.90	32.65	55.02	27.68	33.51	29.28	41.39	47.22	49.57	38.36	33.51
CARE	30.75	39.19	30.24	28.34	33.37	28.84	39.20	58.91	39.04	36.43	33.37
GARCH-t	29.29	34.75	32.97	27.14	34.16	29.64	40.37	54.22	40.62	35.91	34.16
Min	27.56	32.71	32.29	25.25	33.63	26.78	40.89	52.75	41.28	34.79	32.71
RG-RV-tG	25.26	29.93	40.03	25.98	31.10	25.54	39.31	46.37	48.40	34.66	31.10
RC-ScRV	25.36	32.78	31.24	24.98	32.80	27.45	37.93	50.38	40.45	33.71	32.78
RC-RV	24.53	32.19	34.98	24.95	31.80	26.53	38.01	47.84	40.53	33.48	32.19
RC-Ra	25.14	31.28	31.45	24.38	31.57	29.20	38.59	45.78	42.25	33.29	31.45
RC-RR	24.31	30.98	32.93	24.56	30.52	24.62	39.11	45.04	43.80	32.87	30.98
RC-ScRR	24.80	32.05	31.07	24.60	31.08	24.99	38.36	45.30	43.55	32.87	31.08
RC-SSRV	24.79	30.72	29.54	23.87	31.48	24.01	38.62	47.48	44.59	32.79	30.72
RC-RaO	25.07	31.86	28.86	24.39	31.32	26.82	38.41	45.65	40.07	32.49	31.32
Mean	24.85	31.33	30.42	24.55	31.08	25.01	38.28	46.50	39.85	32.43	31.08
RC-SSRR	24.49	30.70	30.97	24.19	30.33	23.86	38.44	45.14	43.23	32.37	30.70
Median	24.71	31.29	30.39	24.39	31.11	25.19	38.35	45.70	39.99	32.35	31.11

Note:A box indicates the favored model based on mean or median minimum loss, whilst bold indicates the least favoured model.

forecasting of tail risk measures is observed, compared to Re-GARCH models employing realized volatility, and traditional GARCH and CARE models, as well as forecast combinations of these models. Specifically, Re-CARE models with RR and Sub-sampled RR generate the most accurate VaR forecasts, while Re-CARE models employing RR, SubRV, RaO are the most accurate for ES forecasting in the empirical study of nine financial return series. Regarding back testing, the Re-CARE type models are also less likely to be rejected than their counterparts: Re-CARE with RaO is rejected the least for VaR forecasting, and Re-CARE models with RaO and RV are rejected the least for ES forecasting. With respect to the quantile loss function, Re-CARE type models' VaR and ES forecasts consistently have lower loss than all other models considered and were at least highly competitive with the forecast combined series "Mean" and "Median". In addition to being more accurate, the Re-CARE models generated less extreme tail risk forecasts, regularly allowing smaller amounts of capital allocation without leading to unexpected violations. The Re-CARE type models with RaO, RR and SSRR should be considered for financial applications when forecasting tail risk, and should allow financial institutions to more accurately allocate capital under the Basel Capital Accord, to protect their investments from extreme market movements. This work could be extended by developing asymmetric and non-linear Re-CARE specifications, and using alternative frequencies of observation for the realized measures.

Chapter 4

Signed range: a new volatility estimator and its application on tail-risk forecasting

4.1 Introduction

As discussed in Section 1.2, since the introduction of Value-at-Risk (VaR) and Expected Shortfall (ES), more and more worldwide financial institutions and corporations started to employ VaR and ES to assist their decision making on capital allocation and risk management, while accurate volatility estimation plays a crucial role in the parametric VaR and ES calculation. Among the volatility estimation models, the Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized ARCH (GARCH) models gained high popularity in the recent years, proposed by Engle (1982) and Bollerslev (1986) respectively. Numerous GARCH types models had been developed during the past decades. Especially, EGARCH (Nelson, 1991) and GJR-GARCH (Glosten *et al.*, 1993) were introduced to capture and describe the well known leverage effect of volatility (Black, 1976). In addition, since the well-known fact that the conditional return distribution is heavy-tailed, voluminous literature considers the error distribution in GARCH type models in order to improve the volatility estimation, see Bollerslev (1987), Chen *et al.* (2012), Chen and Gerlach (2013), Gerlach *et al.* (2013).

The standard GARCH type models only employ daily log returns for the volatility estimation. In recent years, various volatility estimators were proposed and applied

to improve the volatility estimation, such as range and realized measures. High-low range has been proven to be a much more efficient and less noisy volatility estimator compared to return, see Parkinson (1980), Garman and Klass (1980), Andersen and Bollerslev (1998), and Alizadeh, Brandt, and Diebold (2002). The Conditional Autoregressive Range model (CARR) was introduced to describe the evolution of conditional range (Chou, 2005). The CARR(1,1) is specified as Model (4.1). However, since range only has positive values, it can not directly consider the leverage effect compared to return. Different asymmetric range type models are proposed in recent years, such as range-based EGARCH model (REGARCH, Brandt and Jones (2006)), range-based threshold conditional autoregressive model (TARR, Chen *et al.* (2008)), asymmetric smooth transition dynamic range model (Lin *et al.*, 2012). See Chou *et al.* (2010) for a review of range type models and their applications.

$$\begin{aligned} Ra_t &= \lambda_{Ra,t}\varepsilon_t \\ \lambda_{Ra,t} &= \alpha_0 + \alpha_1 Ra_{t-1} + \beta_1 \lambda_{Ra,t-1}, \end{aligned} \tag{4.1}$$

where $\lambda_{Ra,t}$ is the conditional mean of the range. ε_t is assumed to be distributed with a density function with a unit mean, e.g. Exponential, Weibull distributions.

As discussed, range demonstrates its advantages compared with return, since it includes the intra-day price moving process instead of only the closing price. However, the high-low range only has positive value so it can not directly reflect the leverage effect of volatility, while return can consider this aspect (Nelson, 1991; Glosten *et al.*, 1993). In addition, the intra day range only calculates the difference between the high and low prices of the day, while it can not distinguish the "good" days and "bad" days, e.g. day (a) and day (b) in Figure 4.1. Therefore, we proposed a new volatility estimator, named Singed Range (SR), that can consider more intra day information than both return and range. The proposed SR has the ability of considering both the intra-day price process and the leverage effect, and that is why we anticipate the signed range will lead to improved volatility estimation results compared with return and range.

The chapter is structured as follows: Section 4.2 defines the calculation rule of

signed range, and also briefly introduces the return and range. Through high frequency simulation, Section 4.3 presents the relationship between signed range and return volatility, and also discusses the impact of micro-structure noise on various volatility estimator. The symmetric and asymmetric Conditional Autoregressive Signed Range (CARSR) models are proposed in Section 4.4. Section 4.5 discusses the Quasi Maximum Log-Likelihood approach for the parameter estimation, followed by the Adaptive MCMC discussion in Section 4.6. The simulation study is conducted in Section 4.7, which compares the ML and MCMC estimators. Section 4.8 describes the data sets used in the empirical study, and presents the 1% VaR and ES forecasting results. Section 4.9 concludes the chapter and discusses about the future work.

4.2 Signed range

In this section, we propose the calculation rule of signed range. First, as discussed in Section 2.2, for day t, the most commonly used daily log return is calculated in Equation (2.1). Then assuming the mean of return is zero, the variance of return is presented in Equation (2.2).

Based on the distribution of high-low range derived by Feller(1951), Parkinson (1980) proposed range with scaling factor $\frac{1}{\sqrt{4\log(2)}}$ as an unbiased estimator for the return volatility, refer to Equation (2.3) for the calculation details.

We propose a new volatility estimator named signed-range as:

$$sr_t = (\log H_t - \log O_t) - (\log O_t - \log L_t).$$

Now we discuss in detail about the motivation of devising signed range. Range demonstrates its advantages compared with return as it considers the intra-day process instead of only closing price. However, range only has positive value so it can not directly reflect the leverage effect of volatility. Figure 4.1 presents the open, high, low and close prices for two days with completely different characteristics. As is shown Figure 4.1, starting from the opening price, panel (a) has more "decrease" $(\log O_t - \log L_t)$ than "increase" $(\log H_t - \log O_t)$ during the day, and panel (b) displays an opposite scenario. But for each panel, the range provides us with almost same values while the signed range has totally opposite values. Based on

the return values and from a stock market point of view, day (b) is definitely a better day than day (a) which will cause higher volatility. However, the range cannot distinguish the difference between day (a) and day (b). Compared with return and range, the signed range has the ability of considering both the intra-day price process and the leverage effect, that is why we anticipate the signed range will lead to significantly improved volatility estimation results compared with return and range.[1]

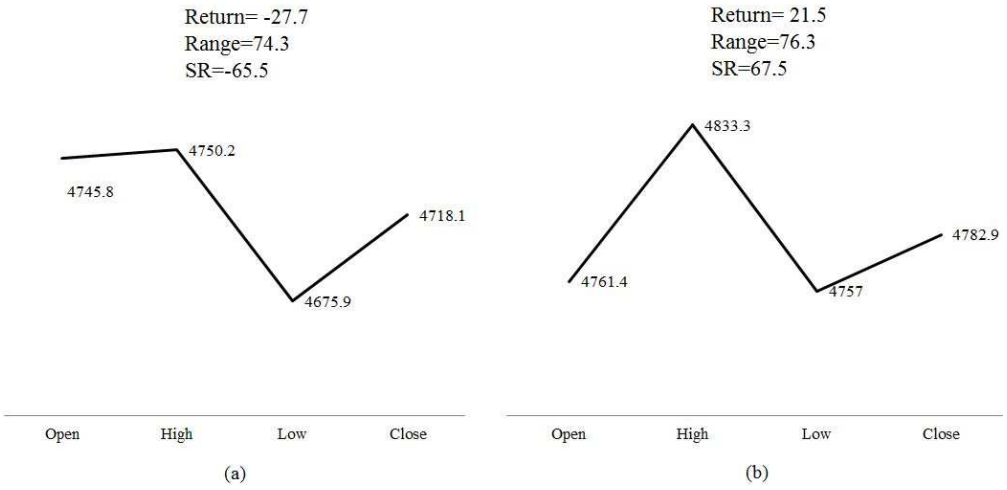

FIGURE 4.1: Motivation of proposing signed range.

4.3 Signed range and return volatility relationship

Parkinson (1980) derived the high-low range and return volatility ($\sigma_{r,t}$) relationship, which is $Ra_t^2 \approx 4\log 2\sigma_{r,t}^2$, based on the property of Brownian motion and distribution of range (Feller, 1951). Similarly, as we are interested in forecasting return volatility using the signed range, the relationship between signed range volatility and return volatility is needed to be quantified. Afterwards, we can construct signed range type models to estimate the signed range volatility, and transform it into return volatility with this relationship. Here we have $\mathrm{E}(r_t^2) = \sigma_{r,t}^2$, $\mathrm{E}(sr_t^2) = \sigma_{sr,t}^2$ and $\mathrm{E}(R_t^2) = \sigma_{R,t}^2 \approx 4\log 2\sigma_{r,t}^2$.

The relationship between signed range and return volatility is analyzed through high frequency intra-day simulation, following the Brownian simulation approach

[1] As Figure 4.1 is used for the demonstration purpose, we do not use the log prices when calculating return, range and signed range, and return is calculated as the close to open return.

of Alizadeh, Brandt, and Diebold (2002), which employs a driftless Brownian motion process. The intra-day prices can be simulated by the Gaussian random walk:

$$s_{t,i} = s_{t,i-1} + z_{t,i},\ t = 1, 2, \ldots T;\ i = 1, 2, \ldots N$$

where $z_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_z^2)$. For each day, different intra-day frequencies were implemented in our simulation study, including 1-second ($N = 21,600$ observations per day), 5-second, 30-second, 1-minute, 1.25-minute and 5-minute, given 360 minutes per trading day. We set $\sigma_z = \sqrt{1/N}$ so that the daily true volatility (TV) equals to 1. This one day process is replicated for $T = 10,000$ times in order to calculate the relationship between true volatility and expected values of return, range and signed range respectively. Then we can derive the ratio between return volatility and signed range, and also testify the return volatility and range relationship ($\sigma_{r,t}^2 = \frac{1}{4\log 2}\mathrm{E}(R_t^2) = 0.3607\mathrm{E}(R_t^2)$). Figure 4.2 describes one simulated intra-day price movement process (21,600 observations) with Brownian motion, and 100 such simulated series are presented in Figure 4.3.

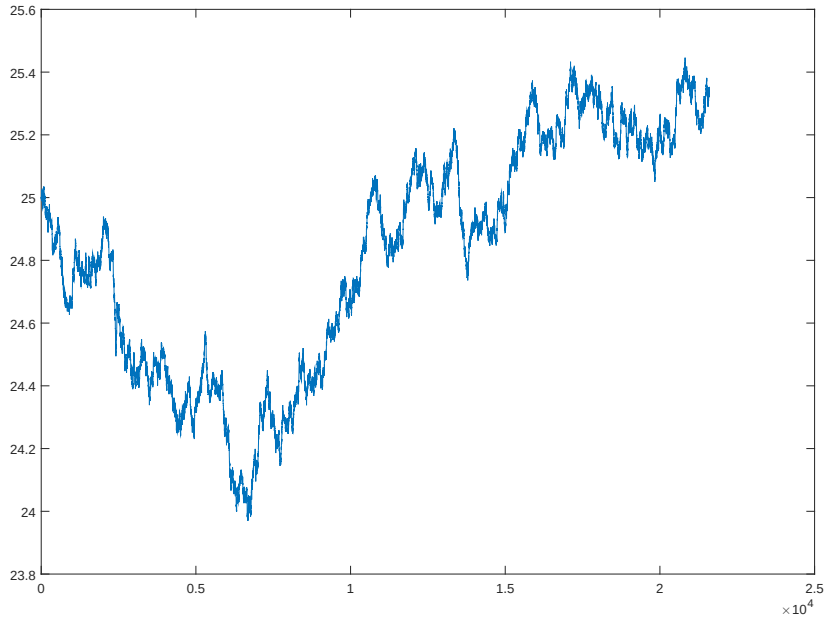

FIGURE 4.2: One simulated intra-day prices movement with Brownian motion. Without bid-ask price.

Without any micro-structure noise, the simulation results between the true volatility (TV) and daily return, range and signed range are summarized in Table 4.1. $\mathrm{E}(r_t^2)$, $\mathrm{E}(R_t^2)$ and $\mathrm{E}(sr_t^2)$ stand for the expected values of return square, range square and signed range square for the T replications. Through taking the average

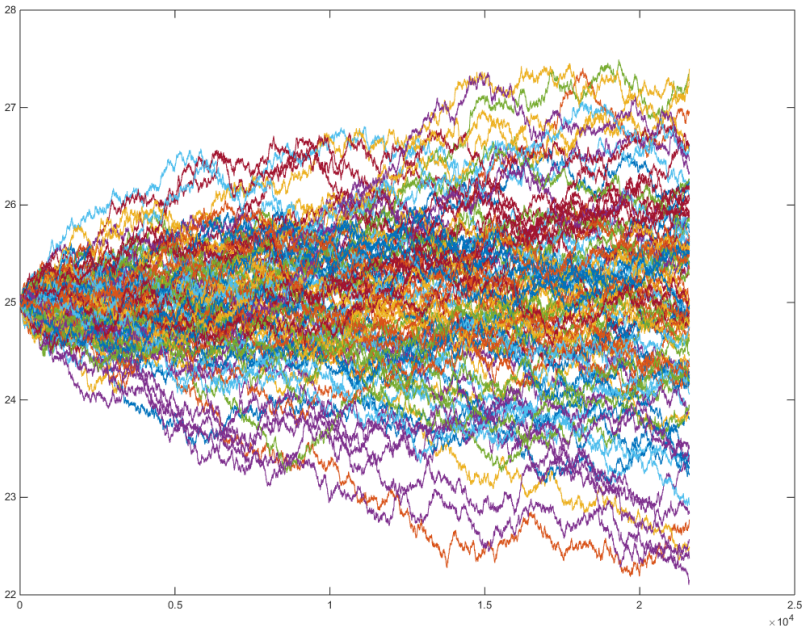

FIGURE 4.3: 100 simulated intra-day prices movement with Brownian motion. Without bid-ask price.

value of different frequencies, we can get that $\mathrm{E}(r_t^2)/\mathrm{E}(sr_t^2) = \sigma_{r,t}^2/\sigma_{sr,t}^2 \approx 0.811$. As can be seen in Table 4.1, $TV^2/\mathrm{E}(r_t^2)$ and $TV^2/\mathrm{E}(sr_t^2)$ stay stably around 1 and 0.81 respectively with various trading frequencies. However, $TV^2/\mathrm{E}(R_t^2)$ can be seriously affected by the intra-day trading frequency. This ratio is close the theoretical value (0.3607) at 1-second frequency, but increases to 0.4256 at 5-minute frequency. Thus return and signed range demonstrate their advantages compared with range from this perspective.

Furthermore, the MSE and MAE (Equations (4.2) and (4.3) between various measured variances (MV, which equals to r_t^2, $0.3607R_t^2$ and $0.811sr_t^2$ respectively) and true volatility square ($TV^2 = 1$) are presented in Table 4.2. The MSE results from return and range is consistent with the theoretical value, which shows that the variance of squared return is five times larger than that of range square. The MSE value from signed range is smaller than that from return, but is around 4 times larger than that from range. Therefore, range demonstrates its superiority in this experiment and has higher efficiency compared with return or proposed signed range. Employing range as an exogenous variable in volatility modelling could potentially improve the volatility estimation accuracy, e.g. CARE-SAV-Range (Gerlach and Chen, 2016). However, if range were directly employed in the

volatility modelling, such as the CARR model, the scaling factor issue as presented in Table 4.1 could affect the estimation accuracy, which will be discussed in the empirical study section.

Based on the high frequency simulations, signed range displays more stable expected values compared with range, and it has less MSE and MAE compared with return. In addition, as we discussed in Section 4.2, signed range has the ability to consider both the leverage effect and intra-day process. Thus all these simulation results and hypothesis demonstrate the potential of signed range.

$$MSE = T^{-1} \sum_{t=1}^{T} (MV_t - TV_t^2)^2 \tag{4.2}$$

$$MAE = T^{-1} \sum_{t=1}^{T} |MV_t - TV_t^2|. \tag{4.3}$$

TABLE 4.1: The ratios between true variance and expected values of return square, range square and signed range square, without bid-ask price.

Trading Frequency	$TV^2/\mathrm{E}(r_t^2)$	$TV^2/\mathrm{E}(R_t^2)$	$TV^2/\mathrm{E}(sr_t^2)$
1-second	0.9976	0.3598	0.8029
5-second	1.0001	0.3676	0.8122
30-second	0.9948	0.3788	0.8217
1-minute	0.9926	0.3858	0.8147
1.25-minute	1.0111	0.3918	0.8152
5-minute	1.0237	0.4265	0.8153

Furthermore, in order to study the impact of micro-structure noise on the high frequency simulation results of various volatility estimates, the bid-ask price is added into the previous simulations. In the experiments, we select the ticksize as \$0.01, and implement the bid price as $B_t = s_t - ticksize$, and ask price as $A_t = s_t + ticksize$. The observed price is $S_t^{obs} = A_t q_t + B_t(1 - q_t)$, where $q_t =$ Bernoulli(0.5) .

Figure 4.4 visualizes one simulated intra-day prices movement process with Brownian motion, including the bid-ask price. We present the difference between one simulated intra-day prices movement process with and without bid-ask price in Figure 4.5 (only first 1000 observations out of 21,600), through which we can

TABLE 4.2: MSE and MAE between various MVs and TV with simulated data sets, without bid-ask price.

MSE			
Trading Frequency	$MV = r^2$	$MV = 0.3607R^2$	$MV = 0.811sr^2$
1-second	1.9905	0.4105	1.6340
5-second	2.0422	0.4151	1.6155
30-second	2.0597	0.3929	1.5570
1-minute	1.9104	0.3746	1.5206
1.25-minute	1.9247	0.3842	1.5815
5-minute	1.9572	0.3765	1.5320
MAE			
Trading Frequency	$MV = r^2$	$MV = 0.3607R^2$	$MV = 0.811sr^2$
1-second	0.9663	0.4680	0.9032
5-second	0.9800	0.4704	0.9011
30-second	0.9709	0.4684	0.8896
1-minute	0.9668	0.4684	0.8876
1.25-minute	0.9530	0.4698	0.8954
5-minute	0.9513	0.4819	0.8774

clearly see the observed price fluctuates randomly between the bid and the ask prices.

The results with simulated intra-day prices including bid-ask jump are shown in Table 4.3 and 4.4. Overall, we get quite similar observations between the simulations with and without bid-ask price. Finally, we also implement these simulations for realized variance and realized range, and the results are consistent with that of Martens and van Dijk (2007) and are not presented in this thesis.

TABLE 4.3: The ratios between true volatility and expected values of return, range and signed range, with bid-ask price.

Trading Frequency	$TV^2/\mathrm{E}(r_t^2)$	$TV^2/\mathrm{E}(R_t^2)$	$TV^2/\mathrm{E}(sr_t^2)$
1-second	1.0105	0.3571	0.8130
5-second	0.9917	0.3621	0.7993
30-second	0.9835	0.3751	0.8011
1-minute	1.0015	0.3875	0.8124
1.25-minute	1.0060	0.3899	0.8301
5-minute	1.0269	0.4230	0.8214

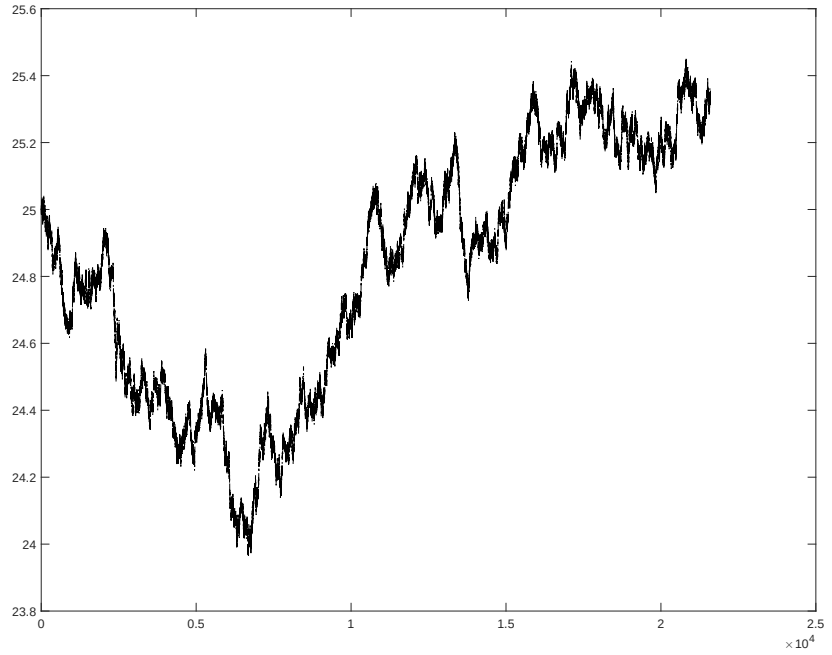

FIGURE 4.4: One simulated intra-day prices movement process with Brownian motion. With bid-ask price.

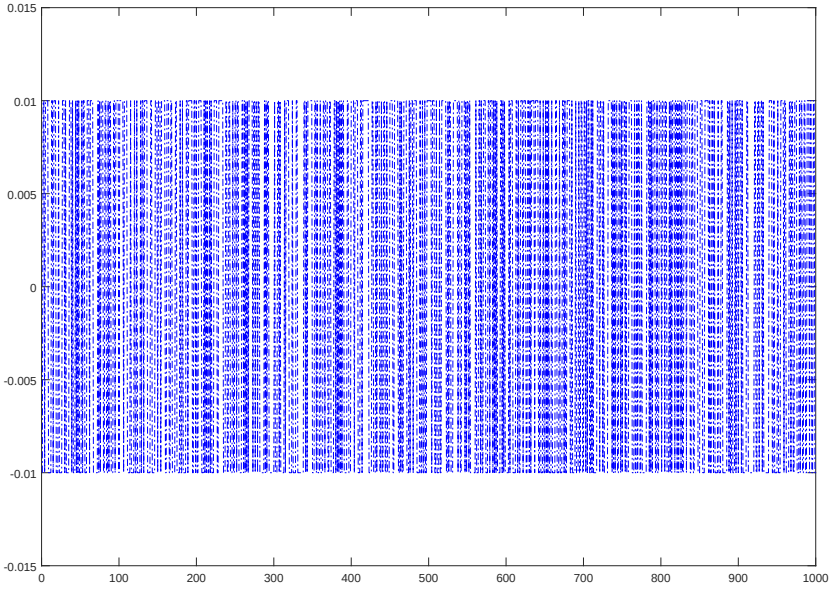

FIGURE 4.5: Difference between one simulated intra-day prices movement process with and without bid-ask price. First 1000 observations.

TABLE 4.4: MSE and MAE between various MVs and TV with simulated data sets, with bid-ask price.

MSE			
Trading Frequency	$MV = r^2$	$MV = 0.3607R^2$	$MV = 0.811sr^2$
1-second	1.9196	0.4047	1.5703
5-second	2.0426	0.4044	1.6371
30-second	1.9886	0.3874	1.5733
1-minute	1.9778	0.3850	1.5662
1.25-minute	1.9646	0.3820	1.5348
5-minute	1.9317	0.3711	1.5362
MAE			
Trading Frequency	$MV = r^2$	$MV = 0.3607R^2$	$MV = 0.811sr^2$
1-second	0.9616	0.4667	0.8975
5-second	0.9772	0.4646	0.9098
30-second	0.9682	0.4647	0.8909
1-minute	0.9686	0.4675	0.8923
1.25-minute	0.9660	0.4672	0.8891
5-minute	0.9559	0.4798	0.8842

4.4 Model proposed

Firstly, since signed range can be used as a return volatility estimator, we completed autocorrelation and partial autocorrelation studies for signed range square and return square, to justify whether we could propose a signed range autoregressive framework that is analogous to GARCH. As can be seen in Figures 4.6 and 4.7, S&P 500 signed range square and return display similar results regarding the autocorrelation and partial autocorrelation. Therefore, we propose an innovative Conditional Autoregressive Signed Range (1,1) (CARSR) model (analogous to GARCH) with the specification as (4.4):

$$\begin{aligned} sr_t &= \mu + \sigma_{sr,t}\varepsilon_t = \mu + a_t, \\ \sigma^2_{sr,t} &= \alpha_0 + \alpha_1 a^2_{t-1} + \beta_1 \sigma^2_{sr,t-1}, \end{aligned} \tag{4.4}$$

where $sr_t = [(\log H_t - \log O_t) - (\log O_t - \log L_t)] \times 100$ is percentage signed range for day t, $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1)$ or $\varepsilon_t \overset{\text{i.i.d.}}{\sim}$ Student-t. Basically, the model follows the structure of GARCH(1,1) and is able to estimate the conditional signed range volatility through this dynamic structure. The parameters are constrained by:

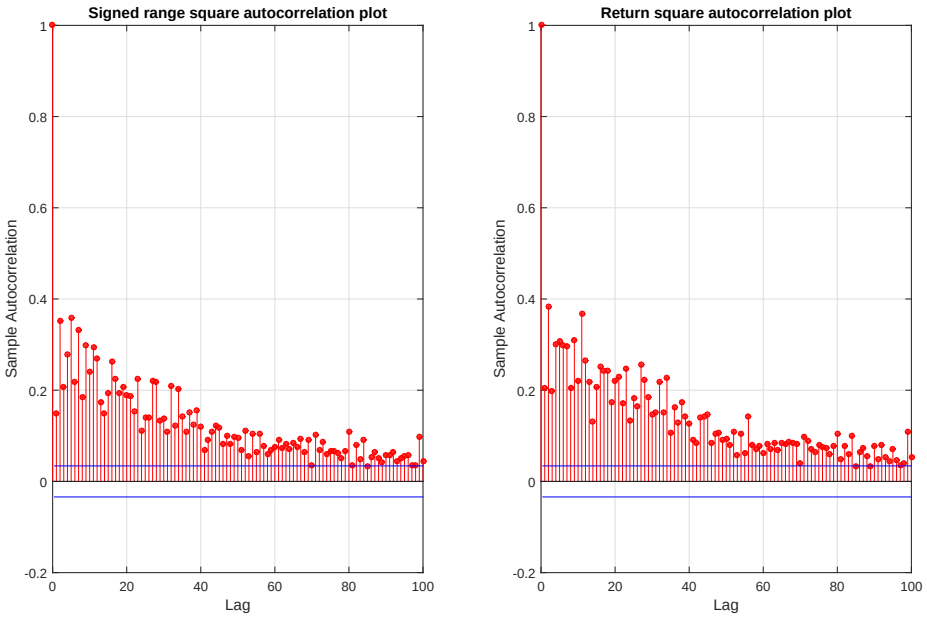

FIGURE 4.6: Autocorrelation plots of signed range square and return square.

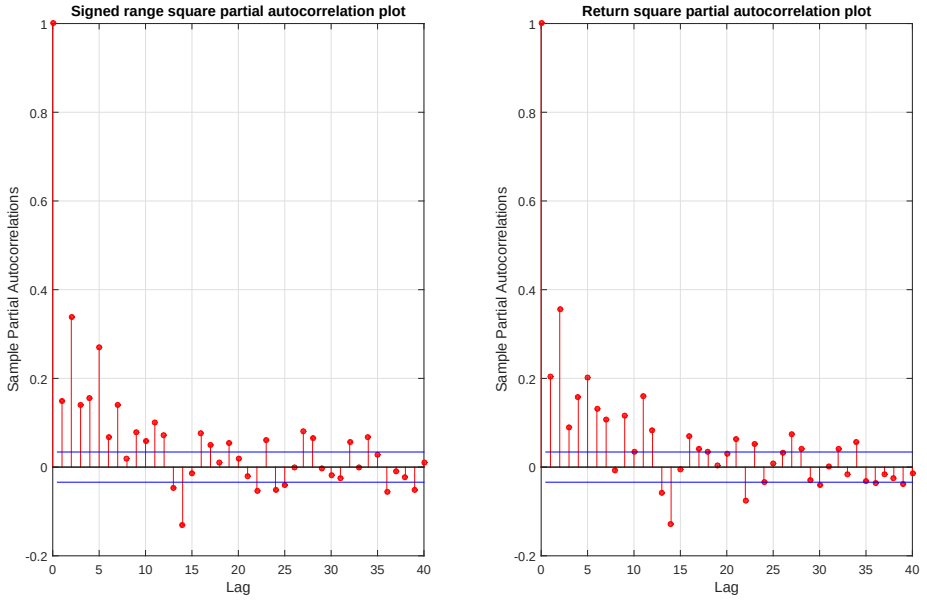

FIGURE 4.7: Partial autocorrelation plots of signed range square and return square.

$\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$, in order to ensure the positivity of $\sigma_{sr,t}$ equation and satisfy the stationary condition. Further, through combining the least squares (LS) estimating function and the least absolute deviation (LAD) estimating function, Ghahramani and Thavaneswaran (2009) developed a process to identify the error distribution of the GARCH model. In the future, we could develop an similar identification procedure of the error distribution for CARSR type models, instead of directly assuming $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1)$ or $\varepsilon_t \overset{\text{i.i.d.}}{\sim}$ Student-t.

We can easily extend the model into the CARSR (p,q) structure. In addition, as a key feature of signed range is its ability to describe the leverage effect, we can have the following ECARSR(1,1) model:

$$\begin{aligned} sr_t &= \mu + \sigma_{sr,t}\varepsilon_t = \mu + a_t, \\ \log\sigma^2_{sr,t} &= \omega + \alpha[|\varepsilon_{t-1}| - \mathrm{E}(|\varepsilon_{t-1}|)] + \gamma\varepsilon_{t-1} + \beta\log\sigma^2_{sr,t-1}. \end{aligned} \tag{4.5}$$

In addition, the following Conditional Autoregressive Signed Range Square (CARSR2) model is proposed and estimated in this chapter, in order to compare its performance with CARR (Model (4.1), Chou 2005).

$$\begin{aligned} sr^2_t &= \lambda_{sr2,t}\varepsilon_t \\ \lambda_{sr2,t} &= \alpha_0 + \alpha_1 sr^2_{t-1} + \beta_1\lambda_{sr2,t-1}, \end{aligned} \tag{4.6}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim}$ Exponential(1). The parameter constraints of CARSR2 are the same as CARSR. Thus we can estimate the conditional mean of signed range square through CARSR2, and have $E(sr^2_t) = \lambda_{sr2,t} = \sigma^2_{sr,t}$. The conditional mean of range is estimated with CARR, and $E(R_t) = \lambda_{R,t} = \sigma_{R,t}$.

4.5 Parameters estimation with quasi-maximum log-likelihood

The log-likelihood function of CARSR with Normal distribution is shown in Equation (4.7). Its specification is identical to the log-likelihood of GARCH with Gaussian distribution.

$$\ell(sr;\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left[log(2\pi) + log(\sigma^2_{sr,t}) + (sr_t - \mu)^2/\sigma^2_{sr,t}\right]. \tag{4.7}$$

Engle and Russell (1998) proved that quasi-maximum likelihood estimate (QMLE) method can give a consistent estimation of the parameters for Autoregressive Conditional Duration (ACD) model with unit mean Exponential distribution. Since CARR and CARSR2 models have analogous specification compared to ACD, Chou (2005) pointed out that QLME can generate a consistent estimation of parameters in CARR (CARSR2 as well), and the log-likelihood of CARR with Exponential distribution is identical to the log-likelihood of GARCH with Gaussian distribution, but without the conditional mean. Log-likelihood of CARSR2 is presented in Equation (4.8). Thus we can estimated CARR or CARSR2 simply with a GARCH specification without the conditional mean term in the mean equation, with signed range or square root of range as input respectively. Furthermore, CARR and CARSR2 possess all the asymptotic properties of GARCH.

$$\ell(sr^2;\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left[log(\lambda_{sr2,t}) + sr_t^2/\lambda_{sr2,t}\right]. \tag{4.8}$$

4.6 Bayesian estimation

After the construction of the log-likelihood function, now we are developing the Bayesian algorithm to estimate the parameters of GARCH/CARSR type models. As discussed in Section 1.5.3, a two-step adaptive Bayesian method that is adapted from Chen and So (2006) is employed. To begin with, all 4 parameters are estimated simultaneously in one group: $\boldsymbol{\theta} = (\mu, \alpha_0, \alpha_1, \beta_1)'$. Further, uninformative priors are chosen over the possible stationarity and positivity region, with one exception. A Jeffreys-type prior is used for α_0, i.e.:

$$\pi(\boldsymbol{\theta}) \propto I(A)\frac{1}{\alpha_0},$$

where region A is defined by $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$.

The adaptive MCMC algorithm employs random walk Metropolis (RW-M) for burn-in period, and independent kernel Metropolis-Hastings (IK-MH) algorithm (Metropolis *et al.*, 1953; Hastings, 1970) for the sampling period. During the burn-in period, a Gaussian proposal distribution is incorporated for the mean vector random walk process. The covariance matrix of the proposal distribution for parameter block $\boldsymbol{\theta}$ is tuned towards a target accept ratio of 23.4% (Roberts, Gelman and Gilks, 1997). Then the IK-MH sampling period is commenced after

the burn-in process, incorporating a mixture of three Gaussian proposal distributions. The average of last 10% of the burn-in period samples is calculated as the input of the IMH period, and the varaice-covariance matrices of three Gaussian proposal distribution are Σ, 10Σ, 100Σ respectively, where Σ is calculated as the covariance of the last 10% of the burn-in period samples for $\boldsymbol{\theta}$. All iterations in the independent MH are used to calculate the posterior mean estimate.

4.7 Simulation Study

In order to the test the performance of the proposed MCMC estimator, a simulation study is conducted. Firstly, $N = 5000$ simulated datasets with sample size $n = 3000$ were generated from Model (4.9). The Bayesian and ML estimation approaches are then employed to estimate the parameters with the 5000 simulated datasets. Therefore, 5000 sets of estimated parameters are respectively collected for MCMC and ML, and both the mean and Root Mean Square Error (RMSE) values are calculated. The Matlab *garch* and related functions in the econometrics toolbox are employed as the ML estimator. For MCMC, the starting values of RWM is randomly chosen as $(0.5, 0.5, 0.5, 0.5)$, and the number of iterations for the RWM and IMH are both set as 10000.

$$\begin{aligned} sr_t &= 0.05 + \sigma_{sr,t}\varepsilon_t = 0.05 + a_t, \\ \sigma^2_{sr,t} &= 0.01 + 0.04a^2_{t-1} + 0.94\sigma^2_{sr,t-1}, \end{aligned} \tag{4.9}$$

Figure 4.8 and 4.9 visualize the 10000 RWM and IMH iterates respectively, for 1 simulated dataset. As can be seen from Figure 4.8, starting from $(0.5, 0.5, 0.5, 0.5)$, $\mu, \alpha 0, \alpha 1, \beta_1$ converge to values that are close to their true values after about 2000 iterations, and the accept rate for this RWM step is 19.07%, which is quite close to the target accept rate of 23.4% and proves the validity of the covariance matrix tuning process. Then with the mean vector and mixture of three Gaussian proposal distributions calculated with the RWM estimates, Figure 4.9 presents 10000 IMH iterations (acceptance rate 46.09%).

Estimation results are summarised in Tables 4.5, boxes indicate the preferred measure comparing MCMC and ML for both bias (Mean) and precision (root

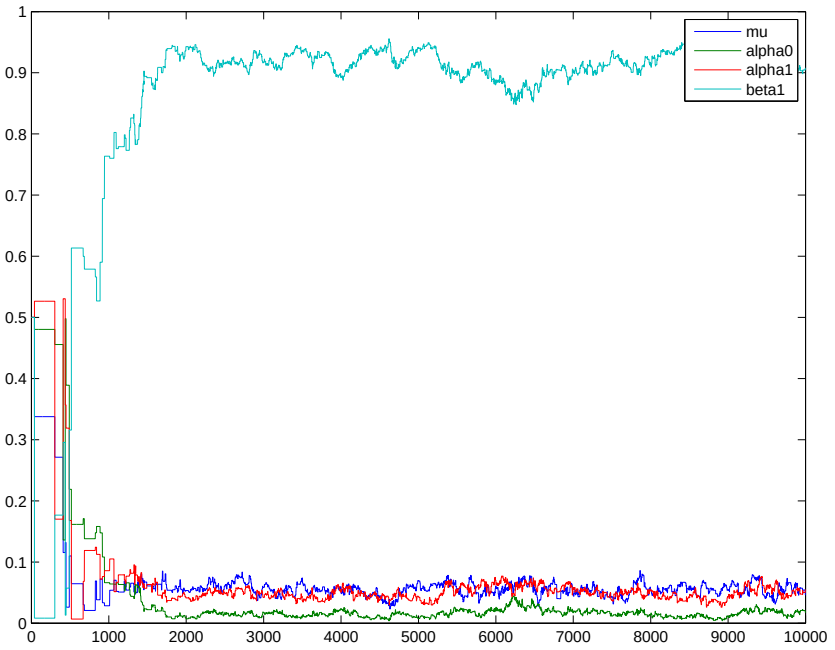

FIGURE 4.8: RWM iterations plot with the simulated data from CARSR-Gaussian. Acceptance rate: 19.07 %.

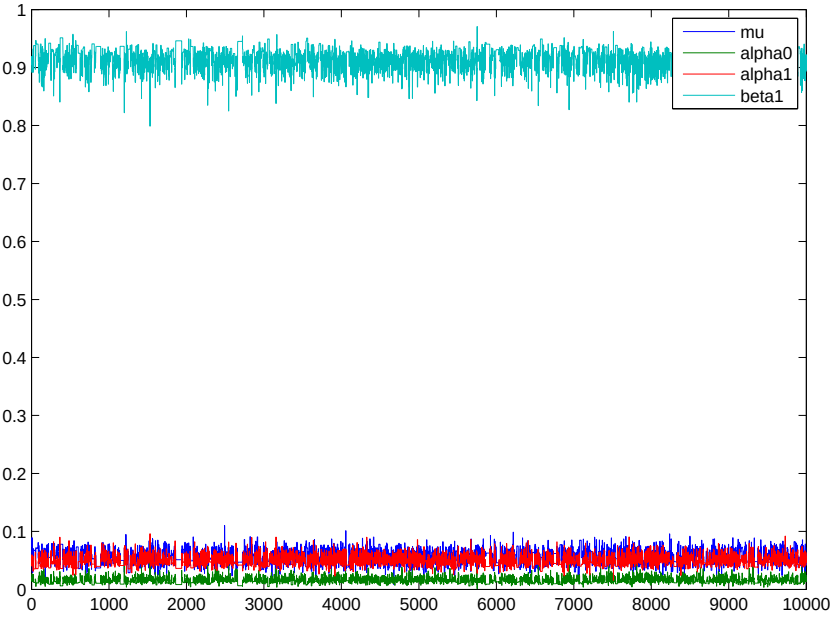

FIGURE 4.9: IMH iterations plot with the simulated data from CARSR-Gaussian. Acceptance rate: 46.09%.

mean square error, RMSE). Overall, both methods generate close to unbiased and quite reasonably precise parameter estimates. Although the bias results slightly favour the ML method for 3 out of 4 parameter estimates, the MCMC approach is favoured by the precision results in 3 out of 4 parameter estimates.

TABLE 4.5: Summary statistics for the two estimators of the CARSR type model with Gaussian Error, with data simulated from Model (4.9).

$n = 3000$		MCMC		ML	
Parameter	True	Mean	RMSE	Mean	RMSE
μ	0.05	0.0489	0.0126	0.0488	0.0127
$\alpha 0$	0.01	0.0136	0.0076	0.0119	0.0087
$\alpha 1$	0.04	0.0439	0.0092	0.0405	0.0083
$\beta 1$	0.94	0.9291	0.0214	0.9358	0.0219

Note:A box indicates the favored estimators, based on mean and RMSE.

As we discussed in Sections 1.5.4 and 2.5, Gelman-Rubin diagnostic and effective sample size test are employed here as well to evaluate the convergence and efficiency performance of the adapted MCMC method on the CARSR frame work. 5 IMH chains each with size $n_1 = 10000$ with simulated data set are used to calculated the $\hat{R}$ and $\hat{n}_{eff}$, which are shown in Table 4.6. Apparently, both the $\hat{R}$ and $\hat{n}_{eff}$ results are quite satisfiable, meaning good convergence and efficiency results.

TABLE 4.6: Summary statistics of the Gelman-Rubin diagnostic and effective sample size with CARSR type model with Gaussian Error and simulated data set.

$n = 1500$			
Parameter	$\hat{R}$	Total $\hat{n}_{eff}$	Average $\hat{n}_{eff}$
$n = 3000$			
μ	1.00268	4577	915
$\alpha 0$	1.00070	3240	648
$\alpha 1$	1.00047	4292	858
$\beta 1$	1.00039	3188	638

4.8 Data and empirical study

4.8.1 Data description

Daily open, high, low and closing prices are downloaded from Oxford-Man Institute of Quantitative Finance Realized Library (http://realized.oxford-man.ox.ac.uk/), from Jan 2000 to Nov 2013. Indices from various countries are considered in our experiments, including S&P500, AORD, FTSE, Hang Seng and DAX. The summary statistics for return, range and signed range on these 5 indices is presented in Table 4.7. Among all 3 volatility estimator, signed range has the smallest kurtosis, but is still heavy-tailed. In the future work, it will be interesting to derive the distribution of signed range and signed range square. Figure 4.10 visualizes the S&P500 return, range and signed range.

TABLE 4.7: Summary statistics of return, range and signed range from Jan 2000 to Nov 2013 for 5 data sets.

Data set	Estimator	Obs.	Mean	Min	Max	Std	Skewness	Kurtosis
SP500	Return	3456	0.0062	-9.3511	10.2202	1.2651	-0.1467	10.0200
	Range	3456	1.4711	0.2474	10.9041	1.0769	3.0601	18.6004
	Signed Range	3456	-0.0583	-9.5522	9.8199	1.3837	-0.3926	8.8353
AORD	Return	3475	-0.0007	-6.4383	3.8912	0.8163	-0.5426	7.5828
	Range	3475	0.9609	0.1396	7.3281	0.6416	2.6217	14.6854
	Signed Range	3475	-0.0027	-7.3281	5.6035	0.9942	-0.4312	6.4520
FTSE100	Return	3476	-0.0386	-5.7603	7.0441	1.0017	-0.1396	6.9341
	Range	3476	1.3492	0.1859	9.1958	0.9215	2.3270	11.8651
	Signed Range	3476	-0.0730	-7.6855	6.1263	1.1641	-0.3110	6.4433
Hang Seng	Return	3147	-0.0403	-11.6162	12.1553	1.0550	0.0957	16.1510
	Range	3147	1.3809	0.2851	17.6474	0.8969	4.5629	53.9500
	Signed Range	3147	-0.1318	-11.8468	11.9702	1.1671	-0.1362	12.0645
DAX	Return	3510	-0.0373	-9.4122	9.9934	1.3833	-0.0792	7.7542
	Range	3510	1.8363	0.2171	11.9869	1.2450	2.2276	10.8840
	Signed Range	3510	-0.0987	-8.3142	10.6839	1.5681	-0.0923	6.5962

In addition, with the 1st S&P 500 in-sample data set and MCMC, we also implement the Gelman-Rubin diagnostic and effective sample size test, and the results are shown in Table 4.8, similar as the results in the previous chapters, $\hat{R}$ is quite close to 1, and the average $\hat{n}_{eff}$ values are around 1500 for $\alpha 0$, $\alpha 1$ and $\beta 1$ which confirm the high level of efficiency of the MCMC algorithm on CARSR estimation.

TABLE 4.8: Summary statistics of the Gelman-Rubin diagnostic and effective sample size with CARSR type model with Gaussian Error and S&P 500.

$n = 1500$			
Parameter	$\hat{R}$	Total $\hat{n}_{eff}$	Average $\hat{n}_{eff}$
$n = 3000$			
μ	1.00107	3749	750
$\alpha 0$	1.00020	7031	1406
$\alpha 1$	1.00105	7474	1495
$\beta 1$	1.00064	7105	1421

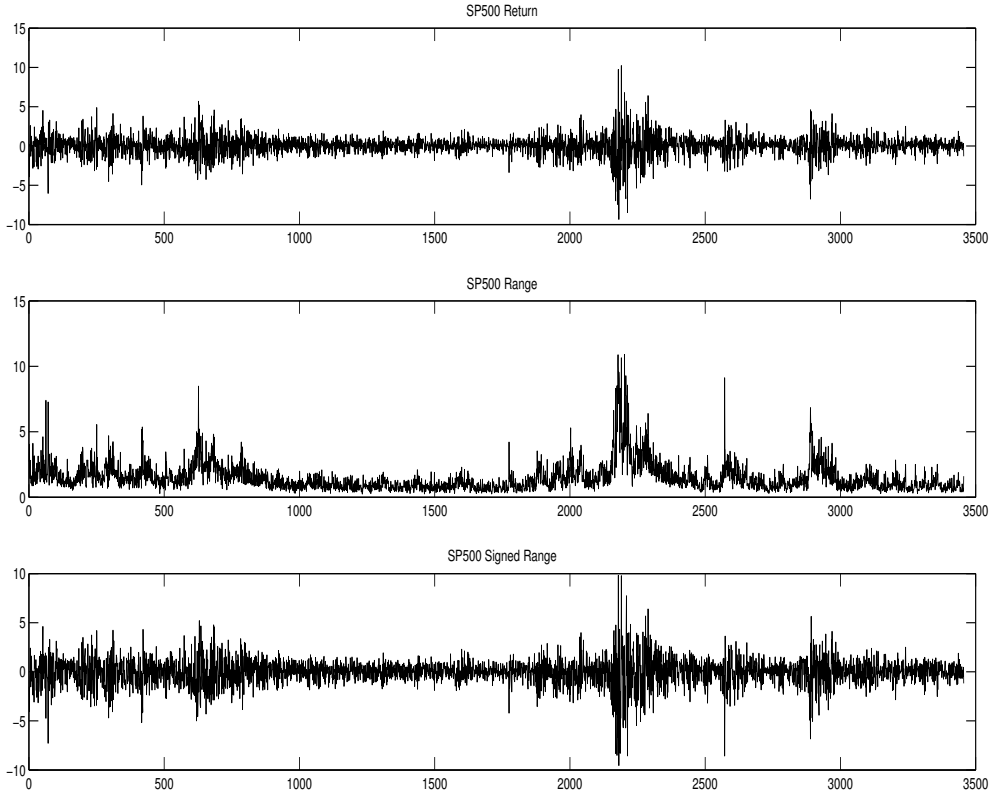

FIGURE 4.10: SP500 return, range and signed range.

4.8.2 VaR and ES forecasting with symmetric type models

The $\alpha = 1\%$ VaR and ES forecasting results with symmetric type models (GARCH, CARSR, CARR and CARSR2) are presented in this section. In order to include the 2008 global financial crisis period in the forecasting experiments, $m = 1500$ one-step-ahead volatility or conditional mean forecasts are calculated based on the 4 different models on 5 data sets, with fixed-size rolling sample n. Thus we have

$n = N - m$, given the size data set is N (number of obs. in Table 4.7). With CARSR model, we can estimate the signed range volatility $\sigma_{sr,t}$, then transform it into an estimator of return volatility with $\hat{\sigma}_{r,t} = \sqrt{0.811}\sigma_{sr,t}$. As to the CARR and CARSR2 models, we can transform the conditional mean of range and signed range square into return volatility estimates, with $\hat{\sigma}_{r,t} = 0.6006\sigma_{R,t} = 0.6006\lambda_{R,t}$ and $\hat{\sigma}_{r,t} = \sqrt{0.811}\sigma_{sr,t} = \sqrt{0.811\lambda_{sr2,t}}$. These volatility or conditional mean forecasts are employed to calculate VaR and ES.

To evaluate the performance of VaR and ES forecasts through different models, we employ the VaR violation rate (VRate) and ES violation rate (ESRate), which are presented in Equations (3.10) and (3.11).

As we discussed in Section 4.5, four models (GARCH-Gaussian, CARSR-Gaussian, CARR-Exponential and CARSR2-Exponential) are on equal comparison ground, and they are employed in the 1% VaR and ES forecasting study, in order to see how the signed range type models work in the empirical study. Table 4.9 presents the VRate, for each model (estimated with ML and MCMC respectively) in each market. The expected VRate should be 1%: boxes represent the model in each market that has a VRate closest to that; bold indicates the model with VRate furthest away from expected.

To begin with, all the VRate values in Table 4.9 is clearly higher than 1%, because of the error distribution selection. Nevertheless, the VRate results already demonstrate the superiority of CARSR-G and CARSR2-Exp compared to GARCH-G and CARR-Exp, while results generated from ML and MCMC are quite close to each other, which is expected based on the simulation study. The mean VRate favours the CARSR2 model, and CARSR and CARSR2 both produced 1.67 % VRate considering the median of violations of 5 VaR forecasting series.

Furthermore, the ES tail risk forecasting results are shown in Table 4.10. As presented before, Gerlach and Chen (2016) illustrated that the quantile level that the 1% ES was estimated to fall was between 0.34% and 0.38%, depending on the error distribution selection. Still, we observe all the ESRate values are clearly higher than the expected violation rate. However, CARSR and CARSR2 generates the ES forecasts that have the ESRate closest to the nominal level among the 4 models, based on mean or median of violations of 5 ES forecasting series.

Lastly, 1500 ES forecasts by GARCH-G and CARSR-G for SP500 are plotted in Figure 4.12, which shows that ES forecasts from CARSR-G recovers faster than

that of GARCH-G during the 2008 GFC time period, meaning the extra efficiency can be gained with SR compared to return. In addition, Figure 4.12 presents the 1500 one-step-ahead ES forecasts by CARR-Exp and CARSR2-Exp with S&P500. It is shown that CARR estimates too low risk level, and leads to high VRate/α ratio. This can be related to the simulation results that we observed in Section 4.3. The expected value of range is affected by the trading frequency. At 5-min frequency, we have $\sigma^2_{r,t} = 0.4265\sigma^2_{R,t}$. Therefore, the theoretical coefficient $1/4\log 2 = 0.3607$ might be too small in real-world data sets, thus provides much reduced risk level. However, the expected value of signed range is quite robust to the different trading frequencies, and it also considers the intra-day process compared to return. Therefore, CARSR type models provide the most satisfiable VaR and ES forecasting results, compared to that of GARCH or CARR type models.

TABLE 4.9: VRate for 5 data sets with 4 symmetric models estimated with ML and MCMC.

Data sets	S&P 500	AORD	DAX	FTSE	Hang Seng	Mean	Median
GARCH-G-ML	2.67%	2.13%	2.07%	2.20%	1.47%	2.11%	2.13%
GARCH-G-MCMC	2.67%	2.13%	2.07%	2.27%	1.53%	2.13%	2.13%
CARSR-G-ML	2.67%	1.33%	1.67%	1.73%	1.47%	1.77%	1.67%
CARSR-G-MCMC	2.67%	1.40%	1.67%	1.87%	1.53%	1.83%	1.67%
CARR-Exp-ML	4.27%	4.20%	2.87%	2.73%	2.93%	3.40%	2.93%
CARR-Exp-MCMC	4.20%	4.40%	3.20%	3.00%	3.27%	**3.61%**	**3.27%**
CARSR2-Exp-ML	2.40%	1.13%	1.67%	1.87%	1.53%	1.72%	1.67%
CARSR2-Exp-MCMC	2.33%	1.20%	1.67%	1.87%	1.53%	1.72%	1.67%

Note:A box indicates the favored model based on mean or median VRate, whilst bold indicates the least favoured model.

TABLE 4.10: ESRate for 5 data sets with 4 symmetric models employing ML and MCMC.

Data sets	S&P 500	AORD	DAX	FTSE	Hang Seng	Mean	Median
GARCH-G-ML	1.27%	1.20%	1.20%	1.27%	0.93%	1.17%	1.20%
GARCH-G-MCMC	1.27%	1.13%	1.13%	1.33%	1.00%	1.17%	1.13%
CARSR-G-ML	1.33%	0.67%	1.07%	0.93%	0.87%	0.97%	0.93%
CARSR-G-MCMC	1.33%	0.67%	1.13%	1.00%	1.00%	1.03%	1.00%
CARR-Exp-ML	3.00%	2.53%	1.33%	1.47%	1.87%	**2.04%**	1.87%
CARR-Exp-MCMC	3.13%	2.40%	1.47%	1.33%	1.93%	2.05%	**1.93%**
CARSR2-Exp-ML	1.33%	0.60%	1.07%	1.00%	1.00%	1.00%	1.00%
CARSR2-Exp-MCMC	1.27%	0.60%	1.07%	1.00%	1.00%	0.99%	1.00%

Note:A box indicates the favored model based on mean or median ESRate, whilst bold indicates the least favoured model.

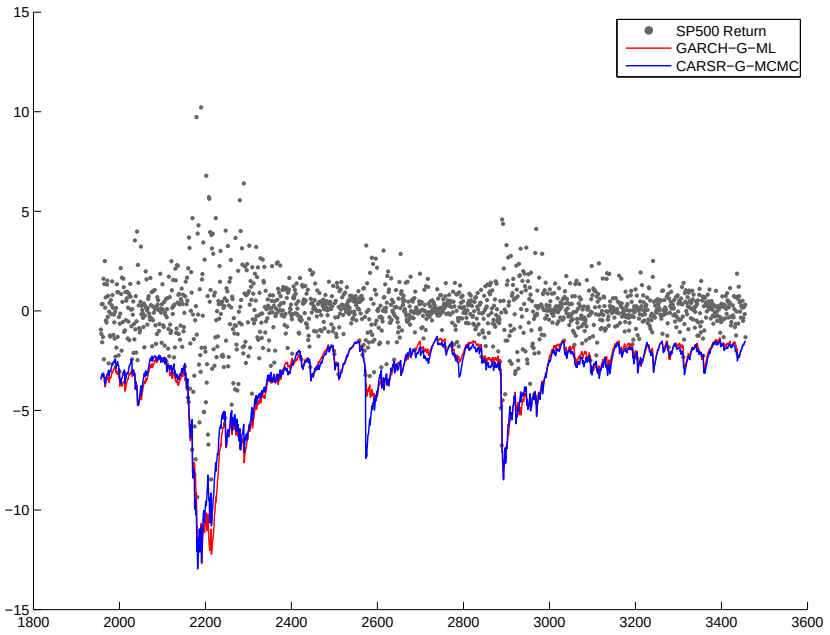

FIGURE 4.11: 1500 S&P500 ES forecasts with GARCH-G-ML and CARSR-Exp-MCMC.

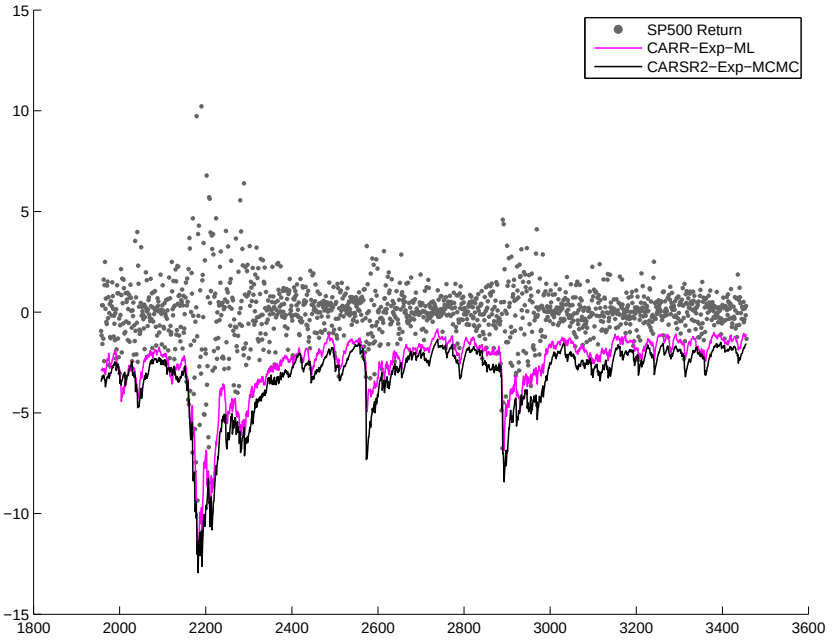

FIGURE 4.12: 1500 SP500 ES forecasts with CARR-Exp-ML and CARSR2-Exp-MCMC.

4.8.3 VaR and ES forecasting with asymmetric type models

As is presented in Section 4.8.1, the shape of signed range is also heavy-tailed. Thus the exponential CARSR with Student-t distribution is implemented to capture the leverage effect and the fat-tail, anticipating much improved volatility and VaR & ES forecasting results compared to the symmetric type models with Gaussian distribution. Using Student-t distribution, VaR and ES can be calculated as Equations (1.12) and (1.14) respectively.

Employing EGARCH-t and ECARSR-t, the VRate are presented in Table 4.11, for 1500 one-step-ahead VaR forecasts. However, we have not implemented the asymmetric CARR (REGARCH) and CARSR2 type models yet, e.g. REGARCH (Brandt and Jones, 2006); TARR (Chen *et al.*, 2008), which should be incorporated in the future work. Firstly, comparing Table 4.9 and Table 4.11, much improved VRate ratios are observed with the asymmetric type models with Student-t distribution, and proof our assumptions at the beginning of this section. In addition, based on Table 4.11, VaR forecasting results from 4 data sets, including mean and median VRate, favor the ECARSR-t.

The ESRate mean and median results actually favor the EGARCH-t model. However, through close inspection of ES forecasts with different data sets, we observe that the ECARSR-t model can over estimate ES level in empirical study, thus this highlights the potential extra efficiency that can be gained by employing an ECARSR-t model. For example, although EGARCH-t and ECARSR-t present quite close ESRate ratio for S&P500 in Table 4.12, Figure 4.13 demonstrates that EGARCH-t has the tendency to over estimate the risk level, especially during the high volatility time period, e.g. times of GFC when there is a persistence of extreme returns.

TABLE 4.11: VRate for 5 data sets with 2 asymmetric models estimated with ML.

Data sets	S&P 500	AORD	DAX	FTSE	Hang Seng	Mean	Median
EGARCH-t-ML	2.27%	1.53%	1.67%	1.80%	1.13%	1.68%	1.67%
ECARSR-t-ML	1.93%	1.00%	1.80%	1.27%	1.13%	1.43%	1.27%

TABLE 4.12: ESRate for 5 data sets with 2 asymmetric models estimated with ML.

Data sets	S&P 500	AORD	DAX	FTSE	Hang Seng	Mean	Median
EGARCH-t-ML	0.87%	0.60%	0.53%	0.73%	0.47%	0.64%	0.60%
ECARSR-t-ML	0.80%	0.67%	0.87%	0.73%	0.47%	0.71%	0.73%

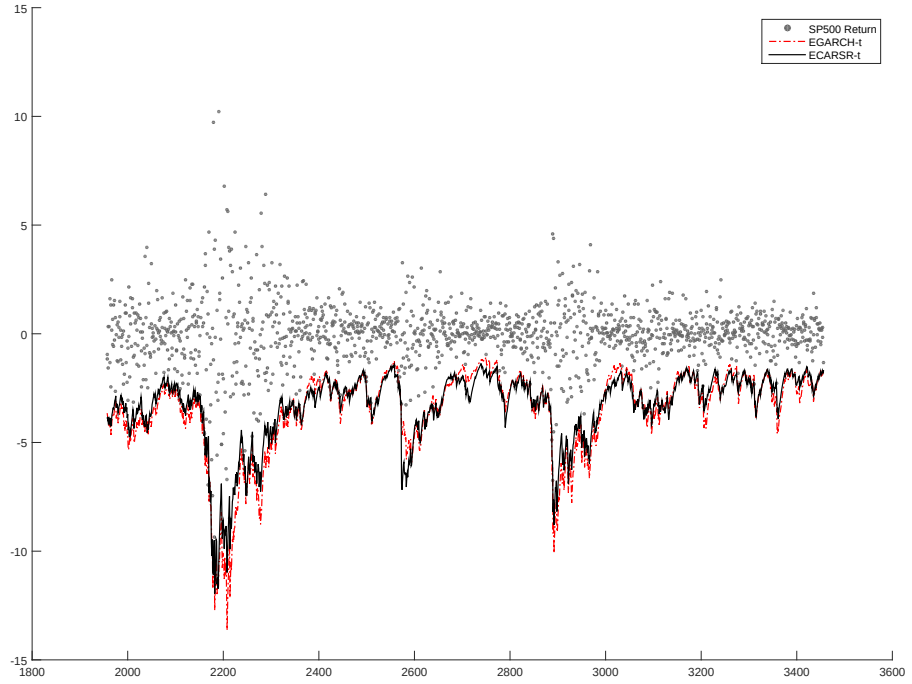

FIGURE 4.13: SP500 ES forecasts with EGARCH-t and ECARSR-t.

4.9 Chapter summary

A new volatility estimator named signed range is proposed in this chapter, which combines the advantages of both return and range. High frequency intra-day price simulations are implemented with or without micro-structure noise, in order quantify the relationship between signed range volatility estimator and return volatility. The simulation results prove that the expected value of signed range is quite robust to the high frequency trading and micro-structure noise, while range is not. In addition, the simulations also demonstrate signed range is more efficient than return. Symmetric and asymmetric signed range type models are proposed and tested with Normal and Student-t distribution. Experimental results

demonstrate the superiority of signed range type models when forecasting the VaR and ES.

The signed range work can be extended in a number of ways. Firstly, we get signed range and return volatility relationship through high frequency intra-day simulation. It will be interesting to have further study on the distribution of signed range and derive this relationship theoretically.

Further, the CARR & CARSR2 model employing Weibull distribution and asymmetric Range type model should be incorporated in the empirical study. Also, we could develop an identification procedure of the error distribution for CARSR type models.

Finally, since nowadays the high frequency tick-by-tick data are widely available, the signed range can be extended into the high frequency framework so that we can have the realized signed range. The realized signed range can be employed in the Realized GARCH or Realized CARE frame work as discussed in Chapter 2 and 3.

Chapter 5

Conclusion & Future Works

5.1 Conclusion

This thesis proposes a series of parametric and semi-parametric dynamic tail risk models for financial tail-risk forecasting, incorporating intra-day and high frequency volatility measures. An adaptive MCMC process is employed for parameter estimation and demonstrates its superiority compared to the frequentist approach for the realized GARCH and proposed realized CARE framework. All the proposed models are tested with VaR and ES forecasting and are compared with a range of famous volatility models.

Chapter 1 discusses the financial tail risk management and measurement, volatility modelling, various volatility measures, parameter estimation techniques, and emphasises the importance of accurate volatility forecasting for various market activities of financial organizations.

Chapter 2 extends the Realized GARCH framework to incorporate the realized range, and the intra-day range, as potentially more efficient series of information than realized variance or daily returns, for the purpose of volatility and tail risk forecasting in a financial time series. Furthermore, we propose an innovative sub-sampled realized range and also adopt an existing scaling scheme, in order to deal with the micro-structure noise of the high frequency volatility measures. A Bayesian adaptive Markov Chain Monte Carlo method is developed and employed for estimation and forecasting. Compared to a range of well known parametric GARCH and Realized GARCH models, predictive log-likelihood results across five market index return series clearly favor the realized GARCH models incorporating

the realized range and sub-sampled realized range, over a six year period that includes the global financial crisis. Further, these same models, when combined with Student-t errors, also compare favourably for tail risk forecasting, both during and after the crisis period.

In Chapter 3, a new framework named Realized Conditional Autoregressive Expectile (Realized CARE) is proposed, through incorporating a measurement equation into the conventional CARE model, in a framework analogous to Realized GARCH. The range and realized measures (realized variance and realized range) are employed as the dependent variables of the measurement equation, since they were proved to be more efficient than return for volatility estimation. The dependence between range & realized measures and expectile can be modelled with this measurement equation. The grid search accuracy of the expectile level will be potentially improved with introducing this measurement equation. In addition, through employing a quadratic fitting targeted search, the speed of grid search is significantly improved. Bayesian adaptive Markov Chain Monte Carlo is used for estimation, and demonstrates its superiority compared to maximum likelihood in a simulation study. Compared to the CARE, the parametric GARCH and the Realized GARCH models, Value-at-Risk and Expected Shortfall forecasting results of 6 indices and 3 assets series favor the proposed Realized CARE model, especially the Realized CARE model with realized Rrange and sub-sampled realized range.

In Chapter 4, we propose a new intra-day volatility estimator named signed range. Through incorporating open, high and low prices, the proposed signed range possesses the characteristics of both return and high-low range. Hence the key features of signed range are its ability to consider both the leverage effect and intra-day process. We implement the high frequency intra-day simulation with and without bid-ask price respectively to simulate signed range, return, and range, in order to quantify the relationship between signed range and return volatility and analyze how the micro-structure noise can affect the efficiency of these 3 volatility estimators. Then the symmetric and asymmetric Conditional Autoregressive Signed Range (CARSR) type models are proposed. An adaptive MCMC is developed for the parameter estimation and is compared with the frequentist through simulation study. Finally, based on 5 data sets across various markets, the out-of-sample VaR and ES forecasting results demonstrate the superiority of CARSR type models compared to Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Conditional Autoregressive Range (CARR) models.

5.2 Future Works

The projects in this thesis work can be extended in the following suggested directions:

a. The asymmetric and non-linear Realized CARE type should be implemented and tested in the future research, to see their performance compared to the Re-CARE-SAV studied in this thesis;

b. The realized CARE framework can be extended to a multivariate specification or regime switching specification. Once a multivariate framework is established, it would be interesting to investigate how to develop an efficient expectile level targeted grid search algorithm on a 2 or more dimensions space;

c. It would be interesting to conduct study on the distribution of signed range, so that it can be better utilized in the volatility estimation and forecasting;

d. More error distributions and model specifications should be considered for the future works of Chapter 4. For example, the Weibull distribution can be employed for the CARR and CARSR2 models, which could potentially improve their VaR and ES forecasting performance. In addition, the asymmetric Range type models, e.g. REGARCH and TARR, should be incorporated in the empirical study;

e. The realized signed range should be tested and compared with realized variance and realized range, and it can be employed in the Realized GARCH or Realized CARE frameworks to forecast tail risk in the future work.

References

Aas, K. and Haff, I. H. (2006). The Generalized Hyperbolic Skew Student's t-Distribution. *Journal of Financial Econometrics*, 4(2), 275–309.

Aigner, D.J. ,Amemiya, T., and Poirier, D. J. (1976). On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function. *International Economic Review*, 377-396.

Andersen, T. G. and 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. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579-625.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, 89(4), 701-720.

Alizadeh, S., Brandt, M. W. and Diebold, F. X. (2002). Range based estimation of stochastic volatility models. *Journal of Finance*, 57(3), 1047-1091.

Anderson, H. M., Nam, K. and Vahid, F. (1999). Asymmetric nonlinear smooth transition GARCH models. In *Nonlinear Time Series Analysis of Economic and Financial Data*, 191-207. Springer US.

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1997). Thinking coherently. *Risk*, 10, 68-71.

Artzener, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-228.

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2004). Regular and modified kernel-based estimators of integrated variance: The case with independent noise. CAF, Centre for Analytical Finance.

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

Black, F. (1976). Studies in stock price volatility changes. In *American Statistical Association Proceedings of the Business and Economic Statistics Section.* 177–181.

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.

Brandt, M. W. and Jones, C. S. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business and Economic Statistics*, 24(4), 470-486.

Chang, C. L., Jiménez-Martín, J. Á., McAleer, M., and Pérez-Amaral, T. (2011). Risk management of risk under the Basel Accord: Forecasting value-at-risk of VIX futures. *Managerial Finance*, 37(11), 1088-1106.

Chen, C.W.S., Gerlach, R. and Lin, E.M.H. (2008). Volatility Forecast Using Threshold Heteroskedastic Models of the Intra-day Range. *Computational Statistics and Data Analysis, on Statistics and Computational Methods in Finance*, 52(6), 2990-3010.

Chen, Q. and Gerlach, R. H. (2013). The two-sided Weibull distribution and forecasting financial tail risk. *International Journal of Forecasting*, 29(4), 527-540.

Chen, Q, Gerlach, R. and Lu, Z. (2012). Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics and Data Analysis*, 56(11), 3498-3516.

Chen, C. W. S. and So, M. K. P. (2006). On a Threshold Heteroscedastic Model. *International Journal of Forecasting*, 22(1), 73-89.

Chen, C.W.S., Gerlach, R. and So, M.K.P. (2008). Comparison of nonnested asymmetric heteroskedastic models. *Computational Statistics and Data Analysis*, 51(4), 2164-2178.

Christensen, K. and Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141(2), 323-349.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39, 841-862.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking*, 561-582.

Chou, R. Y., Chou, H. and Liu, N. (2010). Range volatility models and their applications in finance. In Handbook of Quantitative Finance and Risk Management (pp. 1273-1281). Springer US.

Contino, C. and Gerlach, R. (2014). Bayesian Tail Risk Forecasting using Realised GARCH. The University of Sydney Business School working paper BAWP-2014-05 http://ses.library.usyd.edu.au/handle/2123/12060.

Dorfman, M. S. (2007). Introduction to Risk Management and Insurance (9 ed.). Englewood Cliffs, N.J: Prentice Hall.

Engle, R. F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations. *Econometrica*, 987-1007.

Engle, R.F. and Bollerslev, T. (1986). Modeling the Persistence of Conditional Variances. *Econometric Reviews*, 5(1), 1-50.

Engle R. F. and Gallo G. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1), 3-27.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics*, 22, 367-381.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127-1162.

Feller, W. (1951). The Asymptotic Distribution of the Range of Sums of Random Variables. *Annals of Mathematical Statistics*, 22, 427-32.

Fernández, C. and Steel, M. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93(441), 359–371.

Gaglianone, W. P., Lima, L. R., Linton, O. and Smith, D. R. (2011), Evaluating Value-at-Risk models via quantile regression. *Journal of Business and Economic Statistics*, 29, 150-160.

Garman, M. B. and Klass, M. J. (1980). On the Estimation of Security Price Volatilities from historical data. *The Journal of Business*, 67-78.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2014). Bayesian data analysis (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

Gerlach, R and Chen, C.W.S. (2016). Bayesian Expected Shortfall Forecasting Incorporating the Intraday Range, *Journal of Financial Econometrics*, 14(1), 128-58.

Gerlach, R, Chen, C. W. S. and Chan, N. (2011). Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets. *Journal of Business and Economic Statistics*, 29(4), 481-492.

Gerlach, R, Chen, C. W. S. and Lin, L. (2012). Bayesian Semi-parametric Expected Shortfall Forecasting in Financial Markets. The University of Sydney working paper, available at http://ses.library.usyd.edu.au/handle/2123/8169.

Gerlach, R. and Chen, Q. (2010), Forecasting risk measures via smooth transition threshold GARCH models and the two-sided Weibull distribution. 4th International Conference on Computational and Financial Econometrics CFE 2010, London, United Kingdom, 12th December 2010.

Gerlach, R., Lu, Z. and Huang, H. (2013). Exponentially Smoothing the Skewed Laplace Distribution for Value-at-Risk Forecasting. *Journal of Forecasting*, 32(6), 534-550.

Gerlach, R, and Wang, C. (2016). Forecasting risk via realized GARCH, incorporating the realized range. *Quantitative Finance*, 16(4), 501-511.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, 8(S1), S19-S40.

Ghahramani, M. and Thavaneswaran, A. (2009). Combining estimating functions for volatility. *Journal of Statistical Planning and Inference*, 139(4), 1449-1461.

Gilks, W. R. (2005). Markov chain monte carlo. John Wiley & Sons, Ltd.

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

Gneiting, T. and Ranjan, R. (2012). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35, 705-730.

Hansen, P. R., Huang, Z. and Shek, H. H. (2011). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6), 877-906.

Hansen, P. R. and Huang, Z. (2016). Exponential GARCH Modeling with Realized Measures of Volatility. *Journal of Business and Economic Statistics*. forthcoming.

Hastings, W. K. (1970). Monte-Carlo Sampling Methods Using Markov Chains And Their Applications. *Biometrika*, 57(1), 97-109.

Hoogerheide L. F., Kaashoek J. F. and van Dijk H. K. (2007). On the Shape of Posterior Densities and Credible Sets in Instrumental Variable Regression Models with Reduced Rank: An Application of Flexible Sampling Methods using Neural Networks. *Journal of Econometrics*, 139(1), 154-180.

Jorion, P. (1996), Risk: measuring the risk in Value at Risk. *Financial Analysis Journal*, 52(6), 47-56.

Jorion, P. (1999). Risk management lessons from long-term capital management. Available at SSRN 169449.

Kerkhof, J. and Melenberg, B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking & Finance*, 28(8), 1845-1865.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, 3(2), 73-84.

Li, C.W. and Li, W.K. (1996). On a Double Threshold Autoregressive Heterskedastic Time Series Model. *Journal of Applied Econometrics*, 11(3), 253-274.

Lin, E. M., Chen, C. W. and Gerlach, R. (2012). Forecasting volatility with asymmetric smooth transition dynamic range models. *International Journal of Forecasting*, 28(2), 384-399.

Malevergne, Y. and Sornette, D. (2004). VaR-efficient portfolios for a class of super and sub-exponentially decaying assets return distributions. *Quantitative Finance*, 4(1), 17-36.

Martens, M. and van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, 138(1), 181-207.

McAleer, M., Jiménez-Martín, J. Á. and Pérez-Amaral, T. (2013), GFC-robust risk management strategies under the Basel Accord, *International Review of Economics and Finance*, 27 , pp. 97-111.

McNeil, A., Frey, R., and Embrechts, P. (2005). Quantitative risk management. Princeton University Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The journal of chemical physics*, 21(6), 1087-1092.

Mincer, J. A., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance (pp. 3-46). NBER.

Molnár, P. (2011). High-low range in GARCH models of stock return volatility. *EFMA Annual Meetings, Barcelona.*

Molnár, P. (2012). Properties of range-based volatility estimators. *International Review of Financial Analysis*, 23, 20-29.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59, 347-370.

Newey, W. K. and Powell, J.L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica*, 55, 819-847.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53(1), 61.

Rogers, L. C. G. and Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *Annals of Applied Probability*, 1, 504-512.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability*, 7(1), 110-120.

Schwert, G.W. (1989). Why Does Stock Market Volatility Change Over Time? *The journal of Finance*, 44(5), 1115-1153.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high frequency based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2), 197-231.

Silvapulle, M.J. and Sen, P.K. (2004). Constrained Statistical Inference. New York, USA: John Wiley & Sons.

Steinherr, A. (1998). Derivatives The Wild Beast of Finance. Wiley, New York.

Takahashi, M., Omori, Y. and 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. and 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. (2008). Estimating Value at Risk and Expected Shortfall Using Expectiles. *Journal of Financial Econometrics*, 6(2), 231-252.

Taylor, S.J. (1986). Modeling Financial Time Series. Chichester, UK: John Wiley and Sons.

Tsionas, E. G. (2003). Bayesian quantile inference. *Journal of Statistical Computation and Simulation*, 73(9), 659-674.

Watanabe, T. (2012). Quantile Forecasts of Financial Returns Using Realized GARCH Models. *Japanese Economic Review*, 63(1), 68-80.

Yang, D. and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business*, 73(3), 477-492.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics*, 6(2-3), 273-292.

Yu, K. and Moyeed, RA. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, 54(4), 437-447.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5), 931-955.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales. *Journal of the American Statistical Association*, 100(472), 1394-1411.