The University of Sydney Business School
The University of Sydney

BUSINESS ANALYTICS WORKING PAPER SERIES

Forecasting risk via realized GARCH, incorporating the realized range

Richard Gerlach and Chao Wang
Discipline of Business Analytics, Business School,
The University of Sydney, Australia.

Abstract

The realized GARCH framework is extended 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. A Bayesian adaptive Markov chain Monte Carlo method is employed for estimation and forecasting. Compared to a range of well known parametric GARCH models, predictive log-likelihood results across six market index return series favor the realized GARCH models incorporating the realized range. Further, these same models also compare favourably for tail risk forecasting, both during and after the global financial crisis.

Keywords: Realized Range; Intra-day Range; Realized Variance; Realized GARCH; Predictive Likelihood; Tail Risk Forecasting.

November 2014

BA Working Paper No: BAWP-2014-06
http://sydney.edu.au/business/business_analytics/research/working_papers

Forecasting risk via realized GARCH, incorporating the realized range

Richard Gerlach and Chao Wang *
Discipline of Business Analytics, Business School,
The University of Sydney, Australia.

Abstract

The realized GARCH framework is extended 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. A Bayesian adaptive Markov chain Monte Carlo method is employed for estimation and forecasting. Compared to a range of well known parametric GARCH models, predictive log-likelihood results across six market index return series favor the realized GARCH models incorporating the realized range. Further, these same models also compare favourably for tail risk forecasting, both during and after the global financial crisis.

Keywords: Realized Range; Intra-day Range; Realized Variance; Realized GARCH; Predictive Likelihood; Tail Risk Forecasting.

*Corresponding author: Richard Gerlach. Fax: +61 2 93516409. Email: richard.gerlach@sydney.edu.au

1 INTRODUCTION

Volatility estimation for financial asset returns is highly important in many financial and investment applications. The auto-regressive conditionally heteroskedastic (ARCH) model of Engle (1982) and the generalised (G)ARCH of Bollerslev (1986) are early, seminal works in the time-varying volatility area. Here, volatility is driven by squared returns, and this work led to many variations of this model type: e.g. the EGARCH of Nelson (1991) and the GJR-ARCH of Glosten *et al.* (1993) capture volatility asymmetry as described in Black (1976); plus threshold nonlinear models such as the T-GARCH of Zakoian (1994) and smooth transition ST-GARCH of Anderson *et al.* (1999) which generalize that idea.

Squared daily returns are an inherently noisy volatility estimator, as illustrated by Andersen and Bollerslev (1998). Parkinson (1980) considered the daily intra-day range (Ra), finding this to be more efficient than squared returns. Chou (2005) and Chen, Gerlach and Lin (2008) developed conditionally autoregressive models for the range (CARR); whilst Brandt and Jones (2006) developed a range-EGARCH model. High frequency intra-day data has become now more and well used, with several more efficient realized measures of daily volatility proposed; including realized variance (RV), Andersen and Bollerslev (1998) and the realized kernel (RK), Barndorff-Nielsen and Shephard (2002) and Barndorff-Nielsen *et al.* (2008). Typically, these measures are either: i) employed in GARCH-X models, in the volatility equation only, as a replacement for, or in addition to, squared returns, e.g. see Molnár (2011); or (ii) the dynamics of the realized measure are modelled directly, but not necessarily linked to returns, e.g. see Andersen *et al.* (2003) for RV and Chou (2005) and Chou *et al.* (2010) for Ra. The latter models are employed for volatility prediction; the former can also be employed for tail risk estimation. A few papers directly modelled these measures and linked them to the conditional return distribution: e.g. the range-based stochastic volatility model of Alizadeh, Brandt, and Diebold (2002) and range-based EGARCH model of Brandt and Jones (2006).

Recently, Hansen *et al.* (2011) proposed an intuitive volatility model framework called realized GARCH. A measurement equation was added to GARCH-X model, directly linking each realized measure (X) as an observed variable dependent on contemporaneous

latent volatility. This measurement equation, when using RV or RK as X, was shown by Hansen *et al.* (2011) and Watanabe (2012) to lead to increases in the predictive likelihood, over standard GARCH and GARCH-X models; these models also allow a type of volatility asymmetry. The realized GARCH (RG) model is related to the multiplicative error model (Engle and Gallo, 2006) and the HEAVY model (Shephard and Sheppard, 2010). Watanabe (2012) extended the RG model by employing Student-t errors in the return equation, whilst Contino and Gerlach (2014) employed skewed-t errors in the observation, and Student-t errors in the measurement, equation, both further demonstrating that the RG framework allows more efficient volatility estimation and forecasts via employing RV and RK in the measurement equation.

Parkinson (1980) proposed the intra-day range as a measure of volatility and showed that this estimator was five times more efficient than squared returns for volatility estimation, in a certain setting. Martens and van Dijk (2007) and Christensen and Podolskij (2007) proposed a measure called the realized range (RR), which employs the sum of high frequency intra-period squared ranges. Their simulations showed that the realized range compared favourably to the RV in volatility estimation, across various micro-structure noise settings. Martens and van Dijk (2007) also discussed some bias-correction approaches for RR, thereby proposing the scaled RR, though we do not consider that in this paper. The RR seems to be barely considered in the literature since Martens and van Dijk (2007). We aim to fill that gap in this paper.

In this paper, the RR, and the Ra, will be incorporated into the Realized GARCH framework, creating the Realized Range GARCH (RR-RGARCH) and Range-GARCH (Ra-RGARCH) models; where both Gaussian and Student-t errors will be considered for the observation equation. Further, the MCMC estimation methods in Contino and Gerlach (2014) will be extended to cover these models and compared to the usual maximum likeihood (ML) estimator. Accuracy in terms of predictive likelihood and tail risk forecasting will be assessed across a range of competing models, for six international market index return series.

The paper is structured as follows. Section 2 briefly reviews modern volatility estimators, including Ra, RV, RK and RR. The specifications for the Ra-GARCH and

RR-GARCH are briefly presented in Section 3. Section 4 discusses parameter estimation via Bayesian MCMC. Section 5 reviewstail risk measures such as Value at Risk (VaR) and expected shortfall (ES). Section 6 presents some simulation results comparing Bayesian and maximum likelihood estimation for realized GARCH models. Section 7 describes the data and presents the results of the empirical study. Section 8 concludes, and discusses possible future work.

2 REALIZED MEASURES

This section gives a brief introduction to various volatility estimators included in the models employed in this paper. 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:

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

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

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

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:

$$R_t^2 = \frac{(\log H_t - \log L_t)^2}{4\log 2}$$

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).

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.:

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

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:

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

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 Martijns and van Dijk (2007) lend support to this hypothesis.

Of course, both RV and RR are subject to micro-structure noise bias and inefficiency, more so than daily returns or daily ranges. As such, realized measures have been criticized for this. This issue has been studied extensively, see Rogers and Satchell (1991), Barndorff-Nielsen *et al.* (2004) and Christensen and Podolskij (2007) for discussion.

3 REALIZED RANGE GARCH

This section reviews the literature on realized GARCH and proposes the realized range RG model.

3.1 Model Description

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

$$\begin{aligned} r_t &= \sqrt{h_t} z_t, \\ h_t &= \omega + \beta h_{t-1} + \gamma x_{t-1} \ , \\ x_t &= \xi + \varphi h_t + \tau_1 z_t + \tau_2 (z_t^2 - 1) + \sigma_\varepsilon \varepsilon_t \ , \end{aligned} \tag{3}$$

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 or RK; $D_1(0,1), D_2(0,1)$ indicate distributions that have mean 0 and variance 1. The three equations in 3 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 and RK as the realized measures (i.e. x_t) in model (3); 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 standardised Student-t; Contino and Gerlach (2014) allowed it to be the skewed-t of Hansen (1994) and also allowed $D_2(0,1)$ to be a standardised Student-t. We propose two RG specifications via the choice of: (i) $x_t = RR_t^{\triangle}$, called the realized range GARCH (RR-RG); and (ii) $x_t = Ra_t^2$, called the range-squared realized GARCH model (Ra-RG). The choice of RR as information to drive volatility is motivated by Martijns and van Dijk (2007).

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 h_t is also positive.

Substituting the measurement equation into the volatility equation in 3 leads to:

$$h_t = (\omega + \gamma\xi) + (\beta + \gamma\varphi) h_{t-1} + a_t, \tag{4}$$

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 (4), 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:

$$\begin{aligned} &\omega + \gamma\xi > 0, \\ &0 < \beta + \gamma\varphi < 1 \end{aligned} \tag{5}$$

Further, to ensure positivity of each h_t, 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 paper.

4 BAYESIAN and LIKELIHOOD ESTIMATION

The model specification for the general realized range GARCH (RR-RG) is in (3).

4.1 Likelihood

Following Hansen *et al.* (2011), where $D_1 = D_2 \equiv N(0,1)$, the log-likelihood function for model (3) is:

$$\ell(r, x; \theta) = \underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[log(2\pi) + log(h_t) + r_t^2/h_t\right]}_{\ell(r;\theta)} \underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[log(2\pi) + log(\sigma_\varepsilon^2) + \varepsilon_t^2/\sigma_\varepsilon^2\right]}_{\ell(r|x;\theta)} \tag{6}$$

where $\varepsilon_t = x_t - \xi - \varphi h_t - \tau_1 z_t - \tau_2(z_t^2 - 1)$; the parameter vector to be estimated is $\theta = (\omega, \beta, \gamma, \xi, \varphi, \tau_1, \tau_2, \sigma_\varepsilon)'$, under the constraints in (5) 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 (3) is now:

$$\ell(r,x;\theta) = \underbrace{-\sum_{t=1}^{n}\left[A(\nu)+\log(\pi(\nu-2))+0.5\log(h_t)+\frac{\nu+1}{2}(1+\frac{r_t^2}{h_t(\nu-2)})\right]}_{\ell(r;\theta)} \quad (7)$$

$$\underbrace{-\frac{1}{2}\sum_{t=1}^{n}\left[log(2\pi)+log(\sigma_\varepsilon^2)+\varepsilon_t^2/\sigma_\varepsilon^2\right]}_{\ell(r|x;\theta)} \quad (8)$$

where $\varepsilon_t = x_t - \xi - \varphi h_t - \tau_1 z_t - \tau_2(z_t^2-1)$ 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 $A(\nu) = \log(\Gamma\left(\frac{\nu+1}{2}\right)) - \log(\Gamma\left(\frac{\nu}{2}\right))$. 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 (5) 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.

4.2 Bayesian estimation

The likelihoods in (6) and (7) 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.

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 in the measurement equation, σ, and also a Jeffreys-type prior for the intercept parameter in this equation, ξ i.e.:

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

for the RV-GG model, and

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

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 (5) 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 $\nu^{-1} \sim \text{Unif}(0, 0.25)$, as used by Chen, Gerlach and So (2006), among others.

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 Gaussians 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 the motivation that parameters within the same equation are likely to be more correlated in the posterior (likelihood) than those in separate equations, 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.

5 TAIL RISK FORECAST 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 *alpha*-level quantile of Y, Q_α, itself. 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.

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 Gaglione *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 and Frey (2000), or by the predicted VaR levels, as in Taylor (2008).

Following Kerkhof and Melenberg (2004), Chen, Gerlach and Lu (2012) illustrate 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 (2014) further illustrate 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 $\approx 0.35\%$. 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 (2014), and applying the UC, CC, DQ and VQR tests.

6 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:

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

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 $q_\alpha(r_{t+1}|\boldsymbol{\theta}) = \sqrt{h_{t+1}}\Phi^{-1}(\alpha)$ (Model 1), where Φ^{-1} is the inverse standard Gaussian cdf, and $q_\alpha(r_{t+1}|\boldsymbol{\theta}) = \sqrt{h_{t+1}}T_\nu^{-1}(\alpha)$ (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 $N = 15000$, with a burn-in of $M = 10000$, iterations, which takes 1-2 seconds on a standard PC (depending on n); in each case all iterations after burn-in are used to calculate the posterior mean estimate. 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 1-2. 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

Table 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	0.25	0.2577	0.0233	0.2528	0.0237
γ	0.75	0.7420	0.0206	0.7471	0.0232
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	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	0.25	0.2540	0.0162	0.2516	0.0285
γ	0.75	0.7460	0.0145	0.7479	0.0305
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	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

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. 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; the highest MCMC estimate for ν was 75. The bias results are evenly spread 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. All 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 (2014) and Gerlach, Chen and Chan (2011) for different classes of financial time series models.

7 DATA and EMPIRICAL STUDY

7.1 Data Description and Cleaning

High frequency data at 5-minute intervals, as well as daily data, were downloaded from Thomson Reuters Tick History. This included the market indices for the S&P500, NASDAQ (both US), Hang Seng (Hong Kong), Nikkei 225 (Japan), DAX (Germany) and FTSE 100 (UK), collected from January 2000 to June 2014, within trading hours. Both the 5-minute and daily data consisted of open, high, low and closing prices, so that 5-minute and daily log-return and log-range can be calculated. For the realized measure calculations, the 1st and last 5-minute returns and ranges are removed for each day, as is common practice. Further, a small number of apparent extreme outlying 5-minute returns (e.g. representing price differences of more than 1 USD), likely data errors, were also deleted, as standard.

7.2 Out-of-sample Forecasting: Predictive Log-likelihood

Approximately 1500-1600 one-step-ahead volatility, VaR and ES forecasts are generated separately for the RG-GG and RG-tG specifications using 5 minute RV, 5 minute RR, and daily Ra as the measurement equation daily input. The period from Jan 3, 2000-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,

Table 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	0.29	0.2815	0.0268	0.2767	0.0321
γ	0.70	0.6952	0.0227	0.7008	0.0236
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	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	0.29	0.2749	0.0232	0.2729	0.0265
γ	0.70	0.7008	0.0151	0.7035	0.0156
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	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

this process continuing until forecasts are generated for each day in the forecast sample period Jan 03, 2008 - June 10, 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 $\hat{h}_{n+1}$. Such forecasts are plugged into the one-step-ahead density, 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:

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

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 MCMC 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 3 reports these estimates across the parametric models considered here: being the RG-GG and RG-tG models, each using RV, RR and Ra as the input measurement, as well as standard GARCH-Gaussian (G-G) and GARCH-t (G-t) models.

Table 3: Log-predictive likelihoods; Jan 2008 - June 2014

	S&P500	FTSE	HangSeng	Nikkei	DAX	Nasdaq
G-G	-2411.12	-2502.80	-2764.44	-2886.05	**-2774.86**	-2655.50
G-t	-2384.78	-2486.52	-2753.19	-2874.32	-2747.32	-2631.63
RG-RV-GG	-2379.26	-2460.25	-2840.59	-2900.48	-2724.25	-2618.62
RG-Ra-GG	**-2494.62**	**-2545.54**	**-2853.53**	**-2942.84**	-2773.16	**-2669.83**
RG-RR-GG	-2264.89	-2458.38	-2795.86	-2868.14	-2706.18	-2581.31
RG-RV-tG	-2367.22	-2453.47	-2812.65	-2886.89	-2713.03	-2611.74
RG-Ra-tG	-2440.86	-2510.19	-2811.11	-2873.96	-2749.39	-2640.53
RG-RR-tG	-2261.15	-2450.65	-2782.93	-2859.24	-2699.43	-2574.38

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

In five out of six markets the RG model with Student-t errors that uses the RR as measurement input, is clearly favoured. In each of these markets, the second favoured

model is the RG with Gaussian errors that uses RR as an input. The exception is the Hang Seng market, where the favoured model is the traditional GARCH-t, surprisingly followed by the traditional GARCH-Gaussian. In five markets the RG that employs Ra as an input and Gaussian errors is the least favoured model. In four out of six markets, the RG-RV-GG model beat the GARCH-G, and the RG-RV-Gt beat the GARCH-t in the same four markets. Clearly, by the measure of predictive likelihood, the RR is the most informative realized measure in all markets, compared to the RV and Ra, and in five out of six markets compared to squared daily returns. The Ra is the least favoured measure, whilst RV is favoured over daily returns in five markets.

7.3 Out-of-sample Forecasting: Tail risk

The same estimation sample period, forecast sample period and fixed, moving window approach in the last section are also employed in this section that focuses on VaR and ES forecasting at 1% risk levels. 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, as well as the conditional autoregressive expectile (CARE) indirect GARCH model of Taylor (2008), are added to the competing models.

7.3.1 Value at Risk

Table 4 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 at the 1% quantile for each model in each market; these numbers are expected to be $0.01m$: boxes indicate the model in each market that has a violation rate (VRate) closest to that; bold indicates the model with VRate furthest away from expected. Results for the MCMC estimated RG models are shown. The MCMC estimator of the CARE-IG model in Gerlach and Chen (2014) was also employed, whilst ML methods were used for the standard GARCH models.

Clearly, the three models with VRates typically closest to the expected 1% across the six markets are the CARE-IG, the RG with RR as input and Student-t errors and the HS250 method. All models have higher than expected average VRates across the markets,

Table 4: Counts of 1% VaR violations during the forecast period in each market, $\alpha = 0.01$

model	SP500	FTSE	HangSeng	Nasdaq	Nikkei	DAX	Mean
G-G	42	**33**	33	**33**	27	31	33.17
G-t	28	27	26	25	22	25	25.50
HS100	25	28	26	20	23	28	25.00
HS250	24	24	25	17	27	21	23.00
CARE-IG	23	15	17	24	24	20	20.50
RG-RV-GG	36	22	**49**	32	36	33	34.67
RG-Ra-GG	**49**	29	35	25	**40**	**34**	**35.33**
RG-RR-GG	34	19	45	24	27	27	29.33
RG-RV-tG	26	20	34	27	27	26	26.67
RG-Ra-tG	31	24	22	12	26	27	23.67
RG-RR-tG	27	15	30	19	24	19	22.33
m	1642	1626	1561	1601	1554	1619	1591.0
n	1944	2009	1931	1974	1951	2003	1968.7

Note: Boxes indicate the model with violations closest to its nominal violation rate, bold indicates the least favoured model, in each column.

which may not be too surprising given that the GFC is at the start of the forecast sample; this issue is examined further later.

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 test for forecast accuracy and also for 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 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 papers for more details.

Table 5 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 CARE-IG model has forecast the most accurately and can only be rejected in 1 market (Nikkei) overall. The next best is the RG-RR-tG which is rejected in three out of six markets.

Table 5: Counts of rejections for each test and 1% VaR model during the forecast period over the six markets, $\alpha = 0.01$

$\alpha = 0.01$	UC	CC	DQ	VQR	Total
G-G	**6**	**6**	**6**	5	**6**
G-t	5	3	**6**	4	**6**
HS100	4	4	5	0	5
HS250	2	4	**6**	0	**6**
CARE-IG	1	1	1	0	1
RG-RV-GG	4	4	4	4	4
RG-Ra-GG	5	5	**6**	**6**	**6**
RG-RR-GG	4	4	3	2	4
RG-RV-tG	5	3	2	3	5
RG-Ra-tG	3	4	**6**	5	**6**
RG-RR-tG	3	2	2	1	3

Note: Boxes indicate the model closest to its nominal violation rate, bold indicates the least favoured model, in each column.

7.3.2 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 (2014) 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 $\approx 0.35\%$. 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); $\approx 0.0035m$ (for non-parametric models) and estimated by the quantile level implied by the degrees of freedom

estimates for models with Student-t errors (also $\approx 0.0035m$ for the data considered here). Thus, based on the actual sizes of m in Table , all models have an expected or target ES violation number of between 5.5 and 6 in each market.

Figure 1 shows the forecast sample returns from the S&P 500 and some associated forecasted ES series. The models shown are the G-G, G-t, CARE-IG and RG-RR-tG (estimated by MCMC). The violation numbers from these four models are, respectively, 22, 10, 5 and 4; the expected is between 5 and 6. Despite the large differences in number of violations, the ES forecasts from the RG-RR-tG model, which have the lowest number of violations, are, visually from Figure 1, often less extreme than those from the other three models shown. In fact the ES forecasts from the RG-RR-tG model are less extreme than the G-G on 26% of the forecast sample days, 62% of days less extreme than the G-t model and on 70% of the days less extreme than the CARE-IG model. Further, at times where there is a persistence of extreme returns (e.g. the GFC), close inspection of Figure 1 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 high persistence. As an example, from August, 2008-January, 2009, the most volatile of the GFC period, the RG-RR-tG model's forecasts are less extreme than the G-G model's on 35% of the forecast sample days; including every day except two in the period from 13/12/08 - 13/01/09; this percentage is 39% from 27/04/10-13/10/10 and 44% 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.

To summarise, Figure 1 highlights 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.

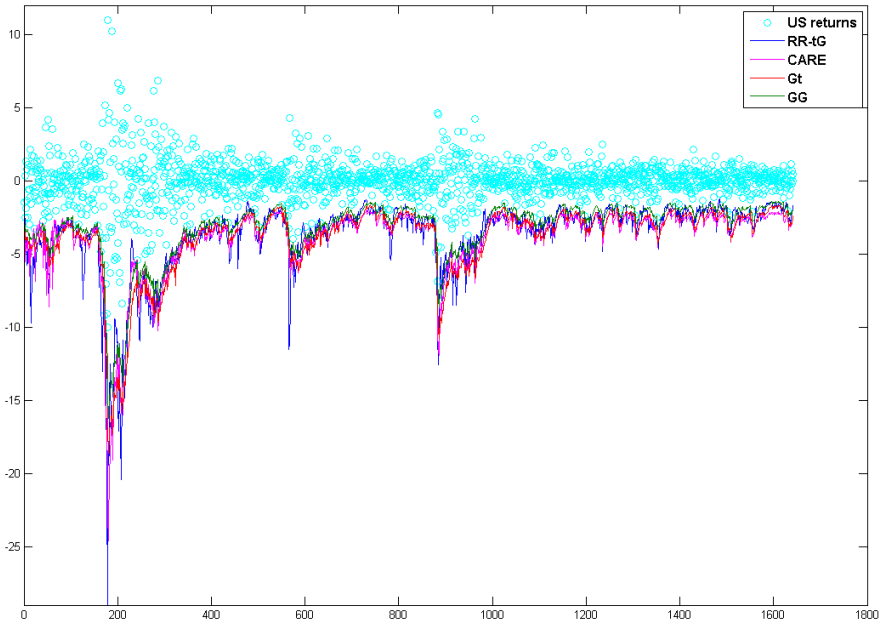

Figure 1: Some 1% ES forecasts and the US S&P500 forecast sample returns.

Table 6 presents the numbers of returns in the forecast period that are more extreme than the forecasted ES at the 1% quantile for each model in each market; called ES violations. Boxes indicate the model in each market that has an ES violation rate (ESRate) closest to that desired; bold indicates the model with ESRate furthest from that expected. Results for the MCMC estimated and ML estimated RG models are each shown separately (ML estimated RG models have -ML following their acronym; MCMC estimated ones have -B).

Clearly, the three models with ESRates typically closest to that expected for the 1% ES across the six markets are the RG-tG-RR-B (estimated by MCMC), RG-tG-RR-ML (estimated by ML) and CARE-IG. All models have higher than expected average ESRates 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. RG models estimated by MCMC all have average ES violations marginally closer to that expected, compared to the same

Table 6: Counts of 1% ES violations during the forecast period in each market, $\alpha = 0.01$

model	SP500	FTSE	HangSeng	Nasdaq	Nikkei	DAX	Mean
G-G	22	18	18	20	14	18	18.33
G-t	10	13	7	7	10	10	9.50
HS100	20	**22**	21	**21**	19	**19**	20.33
HS250	15	15	15	16	14	10	14.17
CARE-IG	5	9	7	11	7	7	7.67
RG-RV-GG-ML	22	12	**32**	16	15	14	18.50
RG-RV-GG-B	21	12	32	16	15	15	18.50
RG-Ra-GG-ML	**30**	20	21	11	**25**	19	**21.00**
RG-Ra-GG-B	**30**	20	20	11	24	19	20.67
RG-RR-GG-ML	15	8	26	12	15	12	14.67
RG-RR-GG-B	13	8	25	12	14	12	14.00
RG-RV-tG-ML	6	7	16	11	9	9	9.67
RG-RV-tG-B	5	7	16	10	9	8	9.17
RG-Ra-tG-ML	8	10	10	5	10	9	8.67
RG-Ra-tG-B	9	9	9	5	9	8	8.17
RG-RR-tG-ML	7	7	10	5	9	8	7.67
RG-RR-tG-B	4	7	9	4	8	8	6.67

Note: Boxes indicate the model closest to its nominal violation rate, bold indicates the least favoured model, in each column.

model estimated by ML.

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 $(0.0033, 0.00375)$); 0.0035 is used for non-parametric models, 0.0038 for Gaussian error models.

Table 7 counts the number of markets in which each model is rejected, for each test, all conducted at a 5% significance level. Only counts for RG models estimated by MCMC

are shown, since results did not qualitatively differ from the RG-ML models. Clearly, for 1% ES forecasting from 2008-2014, the RG-tG-RR model has forecast the most accurately and can only be rejected in three markets (S&P500, Hang Seng and Nikkei225) overall. The next best are the RG-tG-RV and the CARE-IG, both rejected in four out of the six markets.

Table 7: Counts of 1% ES model rejections for each test and model during the forecast period over the six markets, $\alpha = 0.01$

$\alpha = 0.01$	UC	CC	DQ	VQR	Bootstrap	Total
G-G	**6**	**6**	**6**	1	5	**6**
G-t	1	2	5	1	0	5
HS100	**6**	**6**	**6**	2	**6**	**6**
HS250	5	5	**6**	0	2	**6**
CARE-IG	0	0	4	0	0	4
RG-RV-GG	**6**	5	5	1	1	**6**
RG-Ra-GG	5	5	**6**	**5**	4	**6**
RG-RR-GG	5	2	2	1	2	5
RG-RV-tG	1	2	3	1	1	4
RG-Ra-tG	0	2	**6**	0	0	**6**
RG-RR-tG	0	0	2	1	1	3

Note: Boxes indicate the model with lowest number of rejections, bold indicates that with highest number of rejections, for each test.

Figure 2 plots the averages of the 1% ES forecast residuals, standardised by the 1% VaR forecasts, for each of the six markets, plus the average of these averages, for each forecast model/method. An accurate 1% ES forecast model should produce standardised residuals that average approximately 0. Table 7 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, it is clear that the G-t, CARE-IG, RG-RV-tG and RG-RR-tG (the latter two with both ML and MCMC estimation) are the most accurate; whilst G-G, HS100, HS250, RG-Ra-GG clearly, consistently and significantly under-estimated the 1% ES levels, causing negative average residuals to result in all six series. Whether an RG model is estimated by ML or MCMC does not affect the ES residual results, as is apparent in Figure 2.

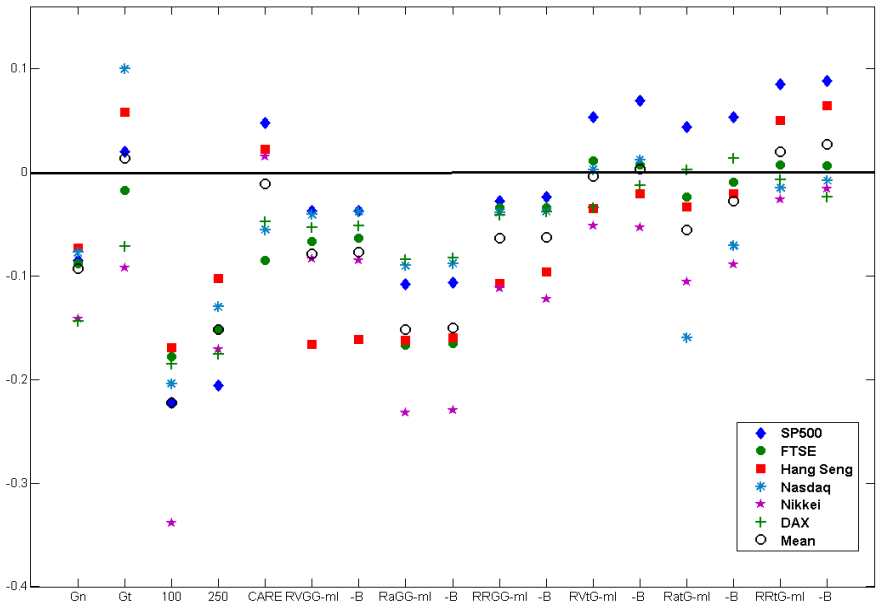

Figure 2: Residuals for 1% ES forecasts, standardised by 1% VaR forecasts, averaged. For each model the six averages are shown, one for each data series, as well as the average of these averages. A reference line is drawn at 0.

Figure 3 plots the ESRate ratio for the 1% ES forecasts, for each of the six markets, plus the average of these six ratios, for each forecast model/method. An accurate 1% ES forecast model should produce an ESRate ratio of 1. Table 7 illustrates that the UC test, of whether these ratios differ from 1, is quite powerful and distinguishes the models into two clear groups: those rejected in five or six markets and those rejected in only one or zero markets. Visually, however, it is clear that only the CARE-IG and RG-RR-tG models have ratios regularly, and on average, close to and surrounding 1, and as such are the most accurate; whilst the G-G, HS100, RG-RV-GG and RG-Ra-GG models clearly, consistently and significantly under-estimate the 1% ES levels, causing three or more times as many ES violations as expected. RG models estimated by MCMC in each case have an average ESRate ratio that is marginally closer to 1 than their corresponding RG model estimated via ML. In the case of the RG-RR-tG model, the distribution of ratios has clearly shifted down, towards 1, when using MCMC, compared to ML, estimation.

In summary, the RG-RR-tG model, estimated by MCMC, is the most accurate at

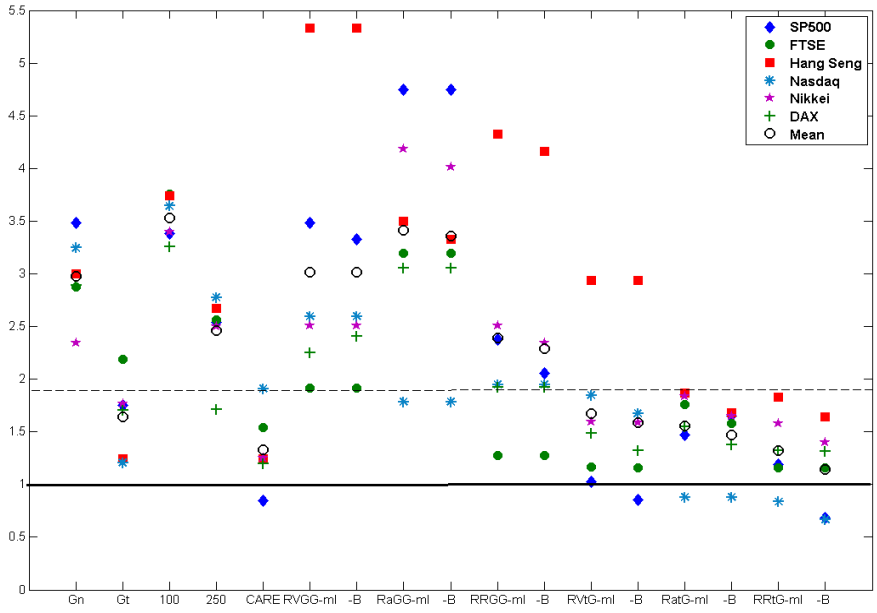

Figure 3: ESRate ratios for the 1% ES forecasts in each market for each model/method. For each model the six ratios are shown, one for each data series, as well as the average of these ratios. A solid reference line is drawn at the target ratio of 1; the dashed reference line represents the upper threshold for rejection by the UC test at the 5% significance level.

forecasting 1% ES for a long forecast period from Jan, 2008-June, 2014 across six market index return series. It consistently displayed rates of ES violation closest to the nominal rate predicted by the estimated Student-t error distribution, and was rejected the least by standard diagnostic tests of quantile forecast accuracy. The next best model was the RG-RR-tG model, estimated by ML, followed by the CARE-IG model. 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 RG-RR-tG model was also highly competitive in 1% VaR forecasting, marginally outperformed only by the CARE-IG model.

8 CONCLUSION

In this paper, 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 Value at Risk and expected shortfall, compared to RG models employing realized volatility or intra-day range, and traditional GARCH models. This work could be extended by considering scaled versions of the realized measures, as in Martens and van Dijk (2007) and also alternative frequencies of observation for the realized measures.

References

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.

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

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

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.* 177181.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

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.

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, 2990-3010.

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.

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, 3498-3516.

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

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

Engle, R. F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations. *Econometrica*, 50, 987-1007.

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.

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

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.

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 and Chen, C.W.S. (2014) Bayesian Expected Shortfall Forecasting Incorporating the Intraday Range, *Journal of Financial Econometrics*, to appear, doi: 10.1093/jjfinec/nbu022 .

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.

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.

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

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.

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

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

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.

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

Taylor, J. (2008). Estimating Value at Risk and Expected Shortfall Using Expectiles. *Journal of Financial Econometrics*, 6, 231-252.

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, 477-491.

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