Volatility and Quantile Forecasts by Realized Stochastic Volatility Models with Generalized Hyperbolic Distribution

Makoto Takahashi
Osaka University and Northwestern University

Toshiaki Watanabe
Hitotsubashi University

Yasuhiro Omori
The University of Tokyo

February 2014

Volatility and Quantile Forecasts
by Realized Stochastic Volatility Models
with Generalized Hyperbolic Distribution[*]

Makoto Takahashi[†]　　　Toshiaki Watanabe[‡]　　　Yasuhiro Omori[§]

Abstract

The realized stochastic volatility model of Takahashi, Omori, and Watanabe (2009), which incorporates the asymmetric stochastic volatility model with the realized volatility, is extended with more general form of bias correction in realized volatility and wider class distribution, the generalized hyperbolic skew Student's t-distribution, for financial returns. The extensions make it possible to adjust the bias due to the market microstructure noise and non-trading hours, which possibly depends on the level of the volatility, and to consider the heavy tail and skewness in financial returns. With the Bayesian estimation scheme via Markov chain Monte Carlo method, the model enables us to estimate the parameters in the return distribution and in the model jointly. It also makes it possible to forecast volatility and return quantiles by sampling from their posterior distributions jointly. The model is applied to quantile forecasts of financial returns such as value-at-risk and expected shortfall as well as volatility forecasts and those forecasts are evaluated by several backtesting procedures. Empirical results with SPDR, the S&P 500 exchange-traded fund, show that the heavy tail and skewness of daily returns are important for the model fit and the quantile forecasts but not for the volatility forecasts, and that the additional bias correction improves the quantile forecasts but does not substantially improve the model fit nor the volatility forecasts.

Key words: Backtesting; Bias correction; Expected shortfall; Generalized hyperbolic skew Student's t-distribution; Markov chain Monte Carlo; Realized volatility; Stochastic volatility; Value-at-risk.

[†]Center for the Study of Finance and Insurance, Osaka University and Department of Finance, Kellogg School of Management, Northwestern University. Email: m-takahashi@sigmath.es.osaka-u.ac.jp; m-takahashi@kellogg.northwestern.edu

[‡]Institute of Economic Research, Hitotsubashi University. Email: watanabe@ier.hit-u.ac.jp

[§]Faculty of Economics, University of Tokyo. Email: omori@e.u-tokyo.ac.jp

1 Introduction

Quantile forecasts of financial returns, as well as volatility forecasts, are important to assess the financial risk. For example, the value-at-risk (VaR) and expected shortfall (ES), computed from the quantile forecasts, have been widely known as measures of the financial tail risk. There are two important aspects for the quantile forecasts: the distribution of financial returns and the volatility. That is, it is essential for the quantile forecasts to model the distribution of financial return and the volatility jointly.

The volatility is unobservable and thus needed to be estimated from the available data. The volatility estimators are divided into two categories: model-based and model-free volatility estimators. The model-based estimator assumes a certain model or process on the dynamics of the latent volatility while the model-free estimator does not require any specific model on the volatility dynamics.

A number of volatility models have been studied for decades. One of the most widely used is the autoregressive conditional heteroskedasticity (ARCH) family including ARCH model of Engle (1982), generalized ARCH (GARCH) model by Bollerslev (1986), and their extensions (see, for example, Andersen, Bollerslev, Christoffersen, and Diebold (2013)). Another well-known family is the stochastic volatility (SV) models, developed by Taylor (1986). Shephard (1996) provides a comprehensive explanation of the SV models. Empirical studies with the volatility models have shown that the distribution of financial returns are well described by the fat tail and negative skewness. Aas and Haff (2006) introduce the general distribution class, called generalized hyperbolic distribution, which can take a flexible form to fit the return characteristics.

The model-free volatility estimator utilizes the recent availability of the high frequency data which contains the price and other asset characteristics sampled at a time horizon shorter than one day. Andersen and Bollerslev (1998) propose the so-called realized volatility (RV) as an accurate volatility measure computed from 5-minute returns. Under some assumptions, the RV is a consistent estimator of the true volatility. More detail properties of the RV can be found in Andersen, Bollerslev, and Diebold (2010) and references therein.

The realized volatility, however, has two practical problems in the real market, which results in a bias in the RV estimator. First, there are non-trading hours when we cannot obtain high frequency returns. For example, In New York Stock Exchange, assets are usually traded from 9 a.m. through 4 p.m. Therefore, computing the RV from only the available intraday returns results in the downward bias. Second, the RV is influenced by the

market microstructure noise (MMN) such as bid-ask bounce[1], which also causes a bias in the estimator. In general, the MMN effect becomes larger for the higher sampling frequency while the information loss becomes larger for the lower frequency. Hansen and Lunde (2006) and Ubukata and Oya (2008) study the effects of the market microstructure noise on the RV estimator.

There are various methods available to mitigate the bias in the RV estimator. For example, Hansen and Lunde (2005) suggest simple scaling methods to adjust the bias due to the non-trading hours. Additionally, Bandi and Russell (2006, 2008) propose an optimal sampling frequency which balance the trade off between the MMN effect and the information loss. Moreover, to alleviate the MMN effect, Zhang, Mykland, and Aït-Sahalia (2005) suggest a multi-scale approach which combines several RVs calculated from returns with different frequencies whereas Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) propose the kernel based estimator called the realized kernel (RK).

Instead of correcting the bias preemptively, contemporaneous modeling of financial returns and the RV measures, which can adjust the bias within the models, have been proposed. Takahashi, Omori, and Watanabe (2009) propose to model daily returns and the RV estimator simultaneously under the framework of the SV model. Dobrev and Szerszen (2010) and Koopman and Scharth (2013) propose the models in a similar manner. In line with Koopman and Scharth (2013), we call the contemporaneous model with the RV measure as the realized SV (RSV) model. Additionally, Hansen, Huang, and Shek (2011) propose to extend GARCH models incorporating them with the realized volatility, which is called the realized GARCH model. The contemporaneous models can adjust the RV bias within the models and make it possible to estimate parameters in return and volatility equations jointly.

The RV models have been applied to quantile forecasts. Giot and Laurent (2004) and Clements, Galvão, and Kim (2008) investigate the quantile forecast performance of GARCH models with the RV estimator although they are not fully contemporaneous models. Recently, Watanabe (2012) applies the realized GARCH model to quantile forecasts and show that the RV estimator improves the forecast performance and that the realized GARCH model can adjust the bias of the RV caused by the MMN. The RSV model, however, has not been fully applied to quantile forecasts[2].

---

[1]O'Hara (1995) and Hasbrouck (2007) provide a comprehensive description of the market microstructure theory and its applications

[2]Dobrev and Szerszen (2010) apply their model to the VaR forecasts but do not examine its performance

In this paper, we extend the RSV model of Takahashi, Omori, and Watanabe (2009) with a nonlinear bias correction in the RV measure and the generalized hyperbolic (GH) skew Student's t-distribution, which includes normal and Student's t-distributions as special cases, and apply the model to volatility and quantile forecasts. Bayesian estimation scheme via Markov chain Monte Carlo (MCMC) technique enables us to estimate the parameters in the distribution of financial returns and in the model jointly, which also makes it possible to adjust the bias of the RV estimator simultaneously. The MCMC technique samples the future volatility and return jointly from their posterior distributions. Using the samples of the future volatility and return, we can easily compute the volatility and quantile forecasts such as the VaR and ES.

We apply the model to daily returns and RKs of SPDR, the S&P 500 exchange-traded fund. The estimation results show that the RSV model with the GH skew Student's t-distribution fits the data well, provided that appropriate priors are used, but the additional bias correction does not improve the model fit. The prediction results show that the RSV models with the skew Student's t-distribution and the additional bias correction perform well for the ES forecasts but not for the volatility nor the VaR forecasts. Overall, the heavy tail and skewness of daily returns as well as the additional bias correction are important for the quantile forecasts, especially for the ES, but not for the volatility forecasts.

The paper is organized as follows. Section 2 describes the extended RSV model and explains the procedures to estimate the parameters, volatility and quantile forecasts jointly via the MCMC technique. Section 3 introduces the evaluation methods for the volatility and quantile forecasts followed by empirical studies in Section 4. Finally, Section 5 concludes the paper.

2　Realized Stochastic Volatility Model

We extend the RSV model proposed by Takahashi, Omori, and Watanabe (2009) with the GH skew Student's t-distribution as follows:

$$r_t = \frac{\beta(z_t - \mu_z) + \sqrt{z_t}\epsilon_t}{\sqrt{\beta^2\sigma_z^2 + \mu_z}} \exp(h_t/2), \quad t = 1, \ldots, n, \tag{1}$$

$$x_t = \xi + \psi h_t + u_t, \quad t = 1, \ldots, n, \tag{2}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, \ldots, n-1, \tag{3}$$

---

formally.

where r_t is an asset return, x_t is a logarithm of RV measure, h_t is an unobserved log-volatility, and

$$z_t \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \quad \mu_z = E[z_t] = \frac{\nu}{\nu - 2}, \quad \sigma_z^2 = Var[z_t] = \frac{2\nu^2}{(\nu - 2)^2(\nu - 4)}, \quad (4)$$

$$\begin{bmatrix} \epsilon_t \\ u_t \\ \eta_t \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{bmatrix}., \quad (5)$$

where $IG(\cdot, \cdot)$ denotes the inverse gamma distribution. We assume that $|\phi| < 1$ for a stationarity of the log-volatility process,

$$h_0 = \mu, \quad \eta_0 \sim N(0, \sigma_\eta^2/(1 - \phi^2)), \quad (6)$$

and $\nu > 4$ for the existence of the variance of z_t. The term $\sqrt{\beta^2 \sigma_z^2 + \mu_z}$ standardizes the return so that the variance of the return is $\exp(h_t)$.

The parameter ξ in (2) corrects the bias due to the MMN and non-trading hours. If ξ is positive, the RV estimator has an upward bias, which implies that the effect of the MMN dominates that of non-trading hours, and vice versa. Following Hansen, Huang, and Shek (2011), the parameter ψ, which is assumed to be one in Takahashi, Omori, and Watanabe (2009), is added to allow more general bias correction depending on the level of the volatility. With the bias correction parameters, the expectation of the RV measure conditional on h_t becomes

$$E[\exp(x_t)|h_t] = \exp\left(\xi + \psi h_t + \frac{1}{2}\sigma_u^2\right) = \exp\left(\xi + \frac{1}{2}\sigma_u^2\right)\sigma_t^{2\psi}, \quad (7)$$

where $\sigma_t^2 = \exp(h_t)$. That is, ξ and ψ correspond to the linear and non-linear bias correction terms, respectively.

The parameter ρ in (5) captures the correlation between r_t and h_{t+1}. A negative value of ρ implies a negative correlation between today's return and tomorrow's volatility, which is the well known phenomenon in stock markets and referred to as the volatility asymmetry.

The return distribution is important in its quantile forecasts. Following Nakajima and Omori (2012), we employ the subclass of the GH distribution, which includes the Student's t distribution as a special case of $\beta = 0$ as well as the normal distribution with $\beta = 0$ and $\nu \to \infty$ (i.e., $z_t = 1$ for all t). The GH distribution has a wider class of distribution but the parameters of the GH distribution are difficult to estimate as pointed out by Prause (1999) and Aas and Haff (2006). Nakajima and Omori (2012) also show that a wider class

of the GH distribution could lead to either the inefficient MCMC sampling or the over-parametrization. Thus, we focus on the GH skew Student's t-distribution throughout the paper. Following Nakajima and Omori (2012), we refer to the RSV model with the GH skew Student's t-distribution as the RSVskt model, hereafter. Similarly, the RSV models with the Student's t and normal distributions are referred to as the RSVt and RSVn models, respectively.

2.1 Estimation

To estimate the RSVskt model, we combine the MCMC algorithms for Bayesian estimation scheme of the SVskt model proposed by Nakajima and Omori (2012) and the RSV model by Takahashi, Omori, and Watanabe (2009). Let $\theta = (\phi, \sigma_\eta, \rho, \mu, \beta, \nu, \xi, \psi, \sigma_u)$, $y = \{r_t, x_t\}_{t=1}^n$, $h = \{h_t\}_{t=1}^n$, and $z = \{z_t\}_{t=1}^n$. Then, we draw random samples from the posterior distributions of (θ, h, z) given y for the RSVskt model using the MCMC method as follows:

   0. Initialize θ, h, and z.

   1. Generate $\phi | \sigma_\eta, \rho, \mu, \beta, \nu, \xi, \psi, \sigma_u, h, z, y$.

   2. Generate $(\sigma_\eta, \rho) | \phi, \mu, \beta, \nu, \xi, \psi, \sigma_u, h, z, y$.

   3. Generate $\mu | \phi, \sigma_\eta, \rho, \beta, \nu, \xi, \psi, \sigma_u, h, z, y$.

   4. Generate $\beta | \phi, \sigma_\eta, \rho, \mu, \nu, \xi, \psi, \sigma_u, h, z, y$.

   5. Generate $\nu | \phi, \sigma_\eta, \rho, \mu, \beta, \xi, \psi, \sigma_u, h, z, y$.

   6. Generate $\xi | \phi, \sigma_\eta, \rho, \mu, \beta, \nu, \psi, \sigma_u, h, z, y$.

   7. Generate $\psi | \phi, \sigma_\eta, \rho, \mu, \beta, \nu, \xi, \sigma_u, h, z, y$.

   8. Generate $\sigma_u | \phi, \sigma_\eta, \rho, \mu, \beta, \nu, \xi, \psi, h, z, y$.

   9. Generate $z | \theta, h, y$.

  10. Generate $h | \theta, z, y$.

  11. Go to 1.

Since u_t is independently and identically distributed, we can implement the same sampling scheme proposed by Nakajima and Omori (2012) for steps 1-5 and 9. We can also easily modify the sampling scheme by Takahashi, Omori, and Watanabe (2009) for steps 6-8 and 10. The detail procedures are given in Appendix.

2.2 Volatility and Quantile Forecasts

To obtain the one-day-ahead log-volatility, daily return, and log-RV measure, we add the following sampling scheme after the step 10 of the MCMC algorithm described above.

i. Generate $h_{n+1}|\theta, h, z, y \sim N(\mu_{n+1}, \sigma^2_{n+1})$, where

$$\mu_{n+1} = \mu + \phi(h_n - \mu) + \rho\sigma_\eta \frac{\sqrt{\beta^2\sigma_z^2 + \mu_z}r_n - \beta\bar{z}_n\exp(h_n/2)}{\sqrt{z}_n\exp(h_n/2)}, \tag{8}$$

$$\sigma^2_{n+1} = (1 - \rho^2)\sigma_\eta^2. \tag{9}$$

ii. Generate $z_{n+1} \sim IG(\nu/2, \nu/2)$.

iii. Generate $r_{n+1}|\theta, h_{n+1}, z_{n+1} \sim N(\hat{\mu}_{n+1}, \hat{\sigma}^2_{n+1})$, where

$$\hat{\mu}_{n+1} = \frac{\beta(z_{n+1} - \mu_z)\exp(h_{n+1}/2)}{\sqrt{\beta^2\sigma_z^2 + \mu_z}}, \tag{10}$$

$$\hat{\sigma}^2_{n+1} = \frac{z_{n+1}\exp(h_{n+1})}{\beta^2\sigma_z^2 + \mu_z}. \tag{11}$$

iv. Generate $x_{n+1}|\theta, h_{n+1} \sim N(\xi + \psi h_{n+1}, \sigma_\eta^2)$.

The quantile forecasts, VaR and ES, can easily be computed from the predictive distributions of financial returns obtained above. Let $\text{VaR}_t(\alpha)$ be the one-day-ahead forecast for the VaR of the daily return r_t with probability α. Then, assuming the long position, the VaR forecast satisfies

$$\Pr[r_t < \text{VaR}_t(\alpha)|\mathcal{I}_{t-1}] = \alpha, \tag{12}$$

where $\mathcal{I}_{t-1}$ is the available information up to $t-1$.

Although the VaR has been widely used to evaluate the quantile forecast of financial returns, it only measures a quantile of the distribution and ignores the important information of the tail beyond the quantile. To evaluate the quantile forecast with tail information, we compute the ES, which is defined as the conditional expectation of the return given that it violates the VaR. The one-day-ahead forecast of the ES with probability α, $\text{ES}_t(\alpha)$, satisfies

$$\text{ES}_t(\alpha) = \text{E}[r_t|r_t < \text{VaR}_t(\alpha), I_{t-1}]. \tag{13}$$

Let n and T be the number of samples for the estimation and prediction, respectively. Then, the one-day-ahead forecasts of the VaR $(\text{VaR}_{n+1}, \ldots, \text{VaR}_{n+T})$ and the ES $(\text{ES}_{n+1}(\alpha), \ldots, \text{ES}_{n+T}(\alpha))$ are computed repeatedly in the following way.

1. Set $i = 1$.

2. Generate the MCMC sample of the model parameters and one-day-ahead return r_{n+i} using the sample of $(y_1, \ldots, y_{n+i-1})$.

3. Compute $\text{VaR}_{n+i}(\alpha)$ as the α-percentile of the MCMC sample of r_{n+i}.

4. Compute $\text{ES}_{n+i}(\alpha)$ as a sample average of r_{n+i} conditional on $r_{n+i} < \text{VaR}_{n+i}(\alpha)$.

5. Set $i = i + 1$ and return to 1 while $i < T$.

3 Evaluation of Volatility and Quantile Forecasts

3.1 Backtesting Value-at-Risk

This section introduces several backtesting methods for the VaR forecasts obtained by the algorithm in Section 2.2. Let T be the number of VaR forecasts and T_1 be the number of times when the VaR is violated, i.e., $r_t < \text{VaR}_t(\alpha)$. Then the empirical failure rate is defined as $\hat{\pi}_1 = T_1/T$. Kupiec (1995) proposes the likelihood ratio (LR) test for the null hypothesis of $\pi_1 = \alpha$, where π_1 is the true failure rate. Since this is a test that on average the coverage is correct, Christoffersen (1998) refers to this as the correct unconditional coverage test. Let $L(p)$ be the likelihood function for an i.i.d. Bernoulli with probability p, that is,

$$L(p) = p^{T_1}(1-p)^{T-T_1}. \tag{14}$$

The LR statistic of the unconditional coverage test is then

$$LR_{uc} = 2\{\ln L(\hat{\pi}_1) - \ln L(\alpha)\}, \tag{15}$$

which is asymptotically distributed as a $\chi^2(1)$ under the null hypothesis of $\pi_1 = \alpha$. Note that this test implicitly assumes that the violations are independent.

To test the independence hypothesis explicitly, Christoffersen (1998) considers the alternative of the first-order Markov process with the switching probability matrix

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}, \tag{16}$$

where π_{ij} is the probability of an $i \in \{0, 1\}$ on day $t-1$ being followed by a $j \in \{0, 1\}$ on day t (1 represents a violation and 0 not). The likelihood under the alternative hypothesis is

$$L(\pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_0 - T_{01}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_1 - T_{11}} \pi_{11}^{T_{11}}, \tag{17}$$

where $T_0 = T - T_1$ and T_{ij} denotes the number of observations with a j following an i. The maximum likelihood estimates of π_{i1} are $\hat{f}_{i1} = T_{i1}/T_i$ for all i. The LR statistic for the null hypothesis of independence, $\pi_{01} = \pi_{11}$, is then

$$LR_{ind} = 2\{\ln L(\hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(\hat{\pi}_1)\}, \tag{18}$$

which is again asymptotically distributed as a $\chi^2(1)$ under the null hypothesis[3].

The two tests for the unconditional coverage and independence can be combined in one test with the null hypothesis of $\pi_{01} = \pi_{11} = \alpha$. Christoffersen (1998) refers to this test as the test of conditional coverage. The LR statistic of the conditional coverage is

$$LR_{cc} = LR_{uc} + LR_{ind} = 2\{\ln L(\hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(\alpha)\}, \tag{19}$$

which is asymptotically distributed as a $\chi^2(2)$ under the null hypothesis of $\pi_{01} = \pi_{11} = \alpha$. Although the above test considers the clustered violations, which is an important signal of risk model misspecification, the first-order Markov alternative represents a limited form of clustering.

Christoffersen and Pelletier (2004) propose more general tests for the clustering based on the duration of days between the violations of the VaR. Define the duration of time (the number of days) between two VaR violations as

$$D_i = t_i - t_{i-1}, \tag{20}$$

where t_i denotes the day of i-th violation. Under the null hypothesis of independent VaR violations, the duration has no memory and its mean of $1/\alpha$ days. The exponential distribution is the only continuous distribution with these properties. Under the null hypothesis, the likelihood of the durations is then

$$f_{\exp}(D; \alpha) = \alpha \exp(-\alpha D). \tag{21}$$

As a simple alternative of dependent durations, we consider the Weibull distribution which includes the null of exponential distribution as a special case. Under the Weibull alternative, the distribution of the duration is

$$f_W(D; a, b) = a^b b D^{b-1} \exp\{-(aD)^b\}, \tag{22}$$

---

[3]If the sample has $T_{11} = 0$, which may happen in small samples with small α, the likelihood is computed as $L((\pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_0 - T_{01}} \pi_{01}^{T_{01}}$.

which becomes the exponential one with probability parameter a when $b = 1$. The null hypothesis is then $b = 1$ in this case. This test can capture the higher-order dependence in the VaR violations by testing the unconditional distribution of the durations.

To test the conditional dependence of the VaR violations, we consider the exponential autoregressive conditional duration (EACD) framework of Engle and Russell (1998). The simple EACD(0,1) model characterizes the conditional expected duration, ψ_i, as

$$\psi_i = \mathrm{E}[D_i] = c + dD_{i-1}, \tag{23}$$

where $d \in [0, 1)$. Assuming the exponential distribution with mean one for the error term, $D_i - \psi_i$, the conditional distribution of the duration is

$$f_{EACD}(D_i|\psi_i) = \frac{1}{\psi_i} \exp\left(-\frac{D_i}{\psi_i}\right). \tag{24}$$

The null hypothesis of the independent durations is then $d = 0$ against the alternative of the conditional durations.

To implement the (un)conditional duration tests, we need to compute the likelihood of the durations with a different treatment for the first and last durations. Let C_i indicate if a duration is censored ($C_i = 1$) or not ($C_i = 0$). For the first observation, if the violation does not occur, then D_1 is the number of days until the first violation occurs and $C_1 = 1$ because the observed duration is left-censored. If instead the violation occurs at the first day, then D_1 is the number of days until the second violation and $C_1 = 0$. The similar procedure is applied to the last duration, $D_{N(T)}$. If the violation does not occur for the last observation, then $D_{N(T)}$ is the number of days after the last violation and $C_{N(T)} = 1$ because the observed duration is right-censored. If instead the violation occurs at the last day, then $D_{N(T)} = t_{N(T)} - t_{N(T)-1}$ and $C_{N(t)} = 0$. For the rest of observations, D_i is the number of days between each violation and $C_i = 0$. The log-likelihood under the distribution, f, is then

$$\ln L(D; \Theta) = C_1 \ln S(D_1) + (1 - C_1) \ln f(D_1) + \sum_{i=2}^{N(T)-1} \ln f(D_i)$$

$$+ C_{N(T)} \ln S(D_{N(T)}) + (1 - C_{N(T)}) \ln f(D_{N(T)}), \tag{25}$$

where we use the survival function $S(D_i) = 1 - F(D_i)$ for a censored observation since it is unknown whether the process lasts at least D_i days. The parameters of the likelihood under the alternative specifications (a and b of the Weibull distribution and c and d of the EACD(0,1) model) need to be estimated numerically since the maximum likelihood

estimates has no closed form solutions. Since the sample size is not large and EACD(0,1) model has a potential difficulty to obtain the asymptotic distribution, we take the Monte Carlo testing technique of Dufour (2006) and follow the specific testing procedure of the LR tests by Christoffersen and Pelletier (2004).

3.2 Backtesting Expected Shortfall

To backtest the ES forecasts with probability α, we use the measure proposed by Embrechts, Kaufmann, and Patie (2005). Define $\delta_t(\alpha) = r_t - ES_t(\alpha)$ and $\kappa(\alpha)$ as a set of time points for which a violation of the VaR occurs. Further, define $\tau(\alpha)$ as a set of time points for which $\delta_t(\alpha) < q(\alpha)$ occurs, where $q(\alpha)$ is the empirical α-quantile of $\delta_t(\alpha)$. The measure is then defined as

$$V(\alpha) = \frac{|V_1(\alpha)| + |V_2(\alpha)|}{2}, \tag{26}$$

where

$$V_1(\alpha) = \frac{1}{T_1} \sum_{t \in \kappa(\alpha)} \delta_t(\alpha), \quad V_2(\alpha) = \frac{1}{T_2} \sum_{t \in \tau(\alpha)} \delta_t(\alpha), \tag{27}$$

and T_1 and T_2 are the numbers of time points in $\kappa(\alpha)$ and $\tau(\alpha)$, respectively. $V_1(\alpha)$ evaluates excess of the VaR estimates and provides the standard backtesting measure of the ES estimates. Since only the values with the violations are considered, this measure strongly depends on the VaR estimates without adequately reflecting the correctness of these values. To correct this weakness, a penalty term $V_2(\alpha)$, which evaluates the values which should happen once every $1/\alpha$ days, is combined with $V_1(\alpha)$. Better ES estimates provide lower values of both $|V_1(\alpha)|$ and $|V_2(\alpha)|$ and so for $V(\alpha)$.

3.3 Loss Functions for Volatility Forecasts

To evaluate the volatility forecasts of different models, we use two loss functions, mean squared error (MSE) and quasi-likelihood (QLIKE) up to additive and multiplicative constants. Let $\hat{\sigma}_t^2$ and h_t be a volatility proxy and volatility forecast, respectively. Then, the two loss functions are given as

$$MSE = \frac{1}{T} \sum_{t=1}^{T} \frac{(\hat{\sigma}_t^2 - h_t)^2}{2}, \quad QLIKE = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\hat{\sigma}_t^2}{h_t} - \log \frac{\hat{\sigma}_t^2}{h_t} - 1 \right). \tag{28}$$

11

Since the true volatility is unobservable, the loss functions are computed using an imperfect volatility proxy, $\hat{\sigma}_t^2$. However, Patton (2011) shows that some class of loss functions including the above two provides a ranking consistent with the one using the true volatility as long as the volatility proxy is a conditionally unbiased estimator of volatility, i.e., $\mathrm{E}_{t-1}[\hat{\sigma}_t^2] = \sigma_t^2$.

4 Empirical Studies

4.1 Data

We use the daily return and RK of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) for SPDR, the S&P 500 ETF, obtained from Oxford-Man Institute[4]. The sample contains 1905 trading days from February 1, 2001 through August 29, 2008. We use the first 1121 samples to evaluate the in-sample performance of the extended RSV models and the remaining 784 samples for prediction. Table 1 shows the descriptive statistics of the daily return and RK for the two periods. The skewness of the daily return is significantly negative for the prediction period but not for the estimation period.

The variation of the data characteristics enables us to investigate the effectiveness and robustness of the extended RSV model. In Section 4.2, we show the estimation results of the RSV models. In Section 4.3, we show the results of the volatility and quantile forecasts for the prediction period. From the descriptive statistics, we expect that the GH skew Student's t distribution fits well for the prediction period and provides better forecasts than the normal distribution.

4.2 Estimation Results

Table 2 shows the MCMC estimation results of the RSV models with normal, Student's t, and skew t distributions obtained by 20000 samples recorded after discarding 5000 samples from MCMC iterations[5]. For ease of explanation, we label the RSV models with the restricted bias correction $\psi = 1$ as RSV models and the ones with the general bias correction $\psi \in \mathbb{R}$ as RSVC models. We estimate the RSV and RSVC models with the priors for the

---

[4]Noureldin, Shephard, and Sheppard (2012) provide the detailed explanation of the data.
[5]All calculations in this paper are done by using Ox of Doornik (2009).

parameters as follows:

$$\mu \sim N(0, 10), \quad \beta \sim N(0, 1), \quad \nu \sim Gamma(a, b) I(\nu > 4), \tag{29}$$

$$\xi \sim N(0, 1), \quad \psi \sim N(1, 1), \quad \sigma_u^{-2} \sim Gamma(2.5, 0.1), \tag{30}$$

$$\frac{\phi + 1}{2} \sim Beta(20, 1.5), \quad \sigma_\eta^{-2} \sim Gamma(2.5, 0.025), \quad \frac{\rho + 1}{2} \sim Beta(1, 2). \tag{31}$$

We check the sensitivity to the priors for ν by picking $(a, b) = (5, 0.5), (1, 0.1), (0.1, 0.01)$, which implies the mean of 10 for all cases and standard deviations of 4.47, 10, and 31.62, respectively.

In Table 2, the parameters specifying the stochastic volatility dynamics are consistent with the stylized features in the volatility literature. The posterior mean of ϕ is close to one for all models, which indicates the high persistence of volatility. Additionally, the posterior mean of ρ is negative and the 95% credible interval does not contain zero for all models, which confirms the volatility asymmetry. The posterior mean of μ are similar among models. These results hold irrespective of the prior for ν.

The parameters specifying the return distribution reflect the data characteristics in the estimation period in Table 1. The posterior mean of β is positive but the 95% credible interval contains zero for all models, which is consistent with that the skewness of daily returns is not significant for the estimation period. The posterior mean of ν increases from 25 to more than 100 as its prior becomes dispersed, which implies that the fat tail is not significant for the period.

The parameters specifying the realized volatility are qualitatively similar among models irrespective of the prior for ν. The posterior means of ξ are negative and the credible intervals do not contain zero for all models, showing the downward bias of the RK mainly due to the non-trading hours. Additionally, the posterior means of ψ are less than one and the credible intervals do not contain one, which indicates the statistically significant nonlinear bias of the RK.

For model comparisons, the marginal likelihood is computed by the method of Chib and Greenberg (1995). The likelihood is estimated using the auxiliary particle filter of Pitt and Shephard (1999) with 10000 particles. The likelihood estimate and its standard error are obtained as the sample mean and standard deviation of the likelihoods from 10 iterations. The posterior probability density and its numerical standard error are evaluated by the method of Chib and Greenberg (1995) and Chib and Jeliazkov (2001) with 50000 reduced MCMC samples.

Table 3 shows the marginal likelihood estimates. Irrespective of the prior for ν, the RSV models give higher marginal likelihoods than the RSVC models, which implies that the nonlinear bias correction does not improve the model fit. The RSVn model as well as the RSVt and RSVskt models with the most dispersed prior for ν provide the highest marginal likelihood whereas the RSVt and RSVskt models with less dispersed priors give significantly lower values. This result confirms that the fat tail and skewness does not improve the model fit for the estimation period.

The GH skew t-distribution, however, performs as good as the normal distribution and is expected to perform better than the normal when the returns are more volatile and skewed. Figure 1 shows the posterior mean and the 95% credible interval of the parameter β for the RSVskt and RSVCskt models obtained from the rolling estimation using 1121 samples each time. The first posterior mean and 95% interval are obtained from the MCMC estimation using 1121 samples from February 1, 2001 through July 20, 2005. The second ones correspond to the one using one day ahead 1121 samples from February 2, 2001 through July 21, 2005. We continue the rolling estimation until the 1121 samples from March 18, 2004 through August 28, 2008. As the rolling estimation proceeds, the posterior means and the 95% intervals become below zero. Thus, the RSVskt and RSVCskt models are expected to perform well in the prediction. We examine this conjecture in the following subsection.

4.3　Prediction Results

We implement the rolling estimation described above and obtain 784 samples of the one-day-ahead forecasts of the latent volatility and quantiles of daily returns from July 21, 2005 through August 29, 2008. We use the prior $\nu \sim Gamma(5, 0.5)I(\nu > 4)$[6] and compute the VaR and ES forecasts for the probabilities of $\alpha = 0.005, 0.01, 0.05$, and 0.1.

Table 4 shows the MSE and QLIKE of the volatility forecasts with the RK as a proxy of the latent volatility[7]. The RSV models outperform the RSVC models, which indicates that the additional bias correction with ψ is not effective for the volatility forecasts. Moreover, the RSVn model performs better than the RSVt and RSVskt models, which implies that the heavy-tail and the skewness do not improve the volatility forecasts.

Table 5 shows the backtesting measures of the ES forecasts proposed by Embrechts,

---

[6]We also implement the rolling estimation with the prior $\nu \sim Gamma(0.1, 0.01)I(\nu > 4)$ and obtain the similar results.

[7]Following Hansen and Lunde (2005), we adjust the effect of the non-trading hours by multiplying the RKs by the ratio of the sum of squared daily returns to the sum of RKs.

Kaufmann, and Patie (2005). The RSVCskt model shows the best performance for the cases of 0.5% and 1% whereas the RSVskt model does for 5$ and 10%. This indicates the effectiveness of the additional bias correction for the ES forecasts and the importance of the fat tail and skewness in the return distribution. That is, the general specifications of both the bias correction and return distribution are important for the ES forecasts.

Tables 6 and 7 show the empirical failure rates (both conditional and unconditional) and the finite sample p-values of the Markov, Weibull, and EACD tests for the VaR forecasts, respectively. The empirical failure rates are higher than the null probabilities in most cases due to the VaR violations in a volatile period from 2007 through 2008 as depicted in Figure 2. Moreover, the failure rates are similar in most cases and consequently we obtain the mixed results for the p-values.

5 Conclusion

The RSV model of Takahashi, Omori, and Watanabe (2009), which incorporates the asymmetric SV model with the RV estimator, is extended with the more general bias correction in the RV and the GH skew Student's t-distribution, for financial returns. The extensions make it possible to adjust the bias due to the MMN and non-trading hours, which possibly depends on the level of the volatility, and to consider the heavy tail and skewness in financial returns. With the Bayesian estimation scheme via Markov chain Monte Carlo method, the model enables us to estimate the parameters in the return distribution and in the model simultaneously. It also makes it possible to forecast the volatility and return quantiles by sampling from their posterior distributions jointly.

We apply the model to daily returns and RKs of SPDR, the S&P 500 exchange-traded fund. The estimation results show that the RSVskt model with the most dispersed prior for ν performs well whereas the additional bias correction does not improve the model fit. The prediction results show that the RSVtskt and RSVCskt models perform well for the ES forecasts but not for the volatility nor the VaR forecasts. Overall, the heavy tail and skewness of daily returns as well as the additional bias correction are important for the quantile forecasts, especially for the ES, but not for the volatility forecasts.

The RSV model has been extended in several directions. Recently, Trojan (2013) proposes a regime switching RSVCskt model and confirms several regimes in the S&P 500 index data. Additionally, extending the univariate RSV model to the multivariate model enables the portfolio risk management and optimal portfolio selection. Moreover, the RSV model

can be extended to utilize multiple realized measures. For example, modeling multiple realized measures with different frequencies and/or different computational methods may improve the volatility and quantile prediction as well as the model fit. In fact, Hansen and Huang (2012) introduce the realized exponential GARCH model which can utilize multiple RV measures and show that the model with multiple measures dominates the one with a single measure. We leave these extensions for future research.

Appendices

A MCMC Sampling Procedure

Let $\theta = (\phi, \sigma_\eta, \rho, \mu, \beta, \nu, \xi, \psi, \sigma_u)$, $y = \{r_t, x_t\}_{t=1}^n$, $h = \{h_t\}_{t=1}^n$, $z = \{z_t\}_{t=1}^n$, and $\Theta = (\theta, h, z)$. We denote a prior distribution of an arbitrary variable w as $f(w)$ and its (conditional) posterior as $f(w|\cdot)$. Given y, the full posterior density is

$$f(\Theta|y) \propto f(r|\Theta) \times f(x|\theta, h) \times f(h|z, \theta) \times f(z|\theta) \times f(\theta) \tag{32}$$

$$= \prod_{t=1}^{n-1} f(r_t|\theta, h_t, h_{t+1}, z_t) f(r_n|\theta, h_n) \times \prod_{t=1}^{n} f(x_t|\theta, h_t) \tag{33}$$

$$\times f(h_1|\theta) \prod_{t=1}^{n-1} f(h_{t+1}|\theta, h_t) \times \prod_{t=1}^{n} f(z_t|\theta) \times f(\theta) \tag{34}$$

$$= \prod_{t=1}^{n-1} f(h_{t+1}|\theta, h_t, z_t, r_t) \times f(h_1|\theta) \times \prod_{t=1}^{n} f(r_t|\theta, h_t, z_t) \times \prod_{t=1}^{n} f(x_t|\theta, h_t) \tag{35}$$

$$\times \prod_{t=1}^{n} f(z_t|\theta) \times f(\theta) \tag{36}$$

$$\propto (1-\rho^2)^{-(n-1)/2} \sigma_\eta^{-(n-1)} \prod_{t=1}^{n-1} \exp\left\{ -\frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \bar{r}_t)^2}{2(1-\rho^2)\sigma_\eta^2} \right\} \tag{37}$$

$$\times (1-\phi^2)^{1/2} \sigma_\eta^{-1} \exp\left\{ -\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma_\eta^2} \right\} \tag{38}$$

$$\times (\beta^2 \sigma_z^2 + \mu_z)^{n/2} \prod_{t=1}^{n} z_t^{-1/2} \exp\left(-\frac{h_t}{2} - \frac{\tilde{r}_t^2}{2} \right) \tag{39}$$

$$\times \sigma_u^{-n} \prod_{t=1}^{n} \exp\left\{ -\frac{(x_t - \xi - \psi h_t)^2}{2\sigma_u^2} \right\} \tag{40}$$

$$\times \left(\frac{\nu}{2} \right)^{n\nu/2} \Gamma\left(\frac{\nu}{2} \right)^{-n} \prod_{t+1}^{n} z_t^{-\nu/2+1} \exp\left(-\frac{\nu}{2z_t} \right) \times f(\theta), \tag{41}$$

16

where

$$\tilde{r}_t = \frac{\sqrt{\beta^2 \sigma_z^2 + \mu_z} r_t \exp(-h_t/2) - \beta \bar{z}_t}{\sqrt{z_t}}, \quad \bar{h}_t = h_t - \mu, \quad \bar{r}_t = \rho \sigma_\eta \tilde{r}_t, \quad \bar{z}_t = z_t - \mu_z \quad (42)$$

We can sample $w \in \Theta$ from the posterior density given other parameters and variables Θ_{-w}. Let $\theta_1 = (\phi, \sigma_\eta, \rho, \mu)$, $\theta_2 = (\beta, \nu)$, and $\theta_3 = (\xi, \psi, \sigma_u)$. The following sections describe how to sample θ_1, θ_2, θ_3, z, and h, respectively.

A.1 Generation of θ_1

Given θ_2, h, and z, the full conditional density of θ_1 is

$$f(\theta_1|\theta_2, h, y) \propto (1 - \rho^2)^{-(n-1)/2} \sigma_\eta^{-(n-1)} \prod_{t=1}^{n-1} \exp\left\{ -\frac{(\bar{h}_{t+1} - \phi \bar{h}_t - \bar{r}_t)^2}{2(1 - \rho^2)\sigma_\eta^2} \right\} \quad (43)$$

$$\times (1 - \phi^2)^{1/2} \sigma_\eta^{-1} \exp\left\{ -\frac{(1 - \phi^2)\bar{h}_1^2}{2\sigma_\eta^2} \right\} \times f(\theta_1), \quad (44)$$

which is similar to the one for the SVskt model of Nakajima and Omori (2012). Thus, we follow the same sampling procedure described in Nakajima and Omori (2012) with different specifications of $\bar{r}_t$ defined in equation (42).

A.2 Generation of θ_2

Given θ_1, h, and z, the full conditional density of θ_2 is

$$f(\theta_2|\theta_1, h, z, y) \propto \prod_{t=1}^{n-1} \exp\left\{ -\frac{(\bar{h}_{t+1} - \phi \bar{h}_t - \bar{r}_t)^2}{2(1 - \rho^2)\sigma_\eta^2} \right\} \times (\beta^2 \sigma_z^2 + \mu_z)^{n/2} \prod_{t=1}^{n} \exp\left(-\frac{\bar{r}_t^2}{2} \right) \quad (45)$$

$$\times \left(\frac{\nu}{2} \right)^{n\nu/2} \Gamma\left(\frac{\nu}{2} \right)^{-n} \prod_{t+1}^{n} z_t^{-\nu/2+1} \exp\left(-\frac{\nu}{2z_t} \right) \times f(\theta_2). \quad (46)$$

Since it is not easy to sample from this density, we apply the Metropolis-Hastings (MH) algorithm based on a normal approximation of the density around the mode. We implement the MH sampling for β and ν separately.

A.3 Generation of θ_3

Given θ_1, θ_2, h, and z, the full conditional density of θ_3 is

$$f(\theta_3|\theta_1, \theta_2, h, z, y) \propto \sigma_u^{-n} \prod_{t=1}^{n} \exp\left\{ -\frac{(x_t - \xi - \psi h_t)^2}{2\sigma_u^2} \right\} \times f(\theta_3). \quad (47)$$

Let the prior distributions of parameters in θ_3 be

$$\xi \sim N(m_\xi, s_\xi^2), \quad \psi \sim N(m_\psi, s_\psi^2), \quad \sigma_u^{-2} \sim Gamma(n_u, S_u). \tag{48}$$

Then, we can sample the parameters in θ_3 from the following posterior distributions:

$$\xi | \psi, \sigma_u, h, y \sim N(\tilde{m}_\xi, \tilde{s}_\xi^2), \tag{49}$$

$$\psi | \xi, \sigma_u, h, y \sim N(\tilde{m}_\psi, \tilde{s}_\psi^2), \tag{50}$$

$$\sigma_u^2 | \xi, \psi, h, y \sim Gamma(\tilde{n}_u, \tilde{S}_u), \tag{51}$$

where

$$\tilde{m}_\xi = \frac{s_\xi^2 \sum_{t=1}^n (x_t - \psi h_t) + \sigma_u^2 m_\xi}{n s_\xi^2 + \sigma_u^2}, \quad \tilde{s}_\xi^2 = \frac{\sigma_u^2 s_\xi^2}{n s_\xi^2 + \sigma_u^2}, \tag{52}$$

$$\tilde{m}_\psi = \frac{s_\psi^2 \sum_{t=1}^n (x_t - \xi) h_t + \sigma_u^2 m_\psi}{s_\psi^2 \sum_{t=1}^n h_t^2 + \sigma_u^2}, \quad \tilde{s}_\psi^2 = \frac{\sigma_u^2 s_\psi^2}{s_\psi^2 \sum_{t=1}^n h_t^2 + \sigma_u^2}, \tag{53}$$

$$\tilde{n}_u = \frac{n}{2} + n_u, \quad \tilde{S}_u = \frac{1}{2} \sum_{t=1}^n (x_t - \xi - \psi h_t)^2 + S_u. \tag{54}$$

A.4 Generation of z

Given θ_1, θ_2, θ_3, and h, the full conditional density of z_t is

$$f(z_t | \theta_1, \theta_2, \theta_3, h, y) \propto g(z_t) \times z_t^{-\frac{\nu+1}{2}+1} \exp\left(-\frac{\nu}{2z_t}\right), \tag{55}$$

where

$$g(z_t) = \exp\left\{ -\frac{\tilde{r}_t^2}{2} - \frac{(\bar{h}_{t+1} - \phi \bar{h}_t - \bar{r}_t)^2}{2(1-\rho^2)\sigma_\eta^2} I(t < n) \right\}, \tag{56}$$

and $I(\cdot)$ is an indicator function. Following Nakajima and Omori (2012), we use the MH algorithm. Specifically, we generate a candidate $z_t^* \sim IG((\nu+1)/2, \nu/2)$ and accept it with probability $\min\{g(z_t^*)/g(z_t), 1\}$.

A.5 Generation of h

We first rewrite the RSVCskt model in (1)-(3) as

$$r_t = \{\beta(z_t - \mu_z) + \sqrt{z_t}\epsilon_t\} \exp(\alpha_t/2)\gamma, \quad t = 1, \ldots, n \tag{57}$$

$$x_t = c + \psi \alpha_t + u_t, \quad t = 1, \ldots, n \tag{58}$$

$$\alpha_{t+1} = \phi \alpha_t + \eta_t, \quad t = 0, \ldots, n-1 \tag{59}$$

where $\alpha_t = h_t - \mu$, $\gamma = \exp(\mu/2)/\sqrt{\beta^2\sigma_z^2 + \mu_z}$, and $c = \xi + \psi\mu$.

To sample the latent variable $(\alpha_1, \ldots, \alpha_n)$ efficiently, we use the block sampler by Shephard and Pitt (1997) and Omori and Watanabe (2008). First, we divide $(\alpha_1, \ldots, \alpha_n)$ into $K + 1$ blocks $(\alpha_{k_{j-1}+1}, \ldots, \alpha_{k_j})$ for $j = 1, \ldots, K + 1$ with $k_0 = 0$ and $k_{K+1} = n$, where $k_j - k_{j-1} \geq 2$. We select K knots $(k_1, \ldots, k_K)$ randomly and sample the error term $(\eta_{k_{j-1}}, \ldots, \eta_{k_{j-1}})$, instead of $(\alpha_{k_{j-1}+1}, \ldots, \alpha_{k_j})$, simultaneously from their full conditional distribution.

Suppose that $k_{j-1} = s$ and $k_j = s + m$ for the jth block and let $y_t = (r_t, x_t)$. Then $(\eta_s, \ldots, \eta_{s+m-1})$ are sampled simultaneously from the following full conditional distribution:

$$f(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) \propto \prod_{t=s}^{s+m} f(y_t|\alpha_t, \alpha_{t+1}) \prod_{t=s}^{s+m-1} f(\eta_t), \qquad (60)$$

for $s + m < n$, and

$$f(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, y_s, \ldots, y_{s+m}) \propto \prod_{t=s}^{s+m-1} f(y_t|\alpha_t, \alpha_{t+1})f(y_n|\alpha_n) \prod_{t=s}^{s+m-1} f(\eta_t), \qquad (61)$$

for $s + m = n$. The logarithm of $f(y_t|\alpha_t, \alpha_{t+1})$ or $f(y_n|\alpha_n)$ in (60) and (61) (excluding constant term) is given by

$$l_t = -\frac{\alpha_t}{2} - \frac{(r_t - \mu_t)^2}{2\sigma_t^2} - \frac{(x_t - c - \alpha_t)^2}{2\sigma_u^2}, \qquad (62)$$

where

$$\mu_t = \begin{cases} \{\beta\bar{z}_t + \sqrt{z_t}\rho\sigma_\eta^{-1}(\alpha_{t+1} - \phi\alpha_t)\}\exp(\alpha_t/2)\gamma, & t < n, \\ \beta\bar{z}_n\exp(\alpha_n/2)\gamma, & t = n, \end{cases} \qquad (63)$$

and

$$\sigma_t^2 = \begin{cases} (1 - \rho^2)z_t\exp(\alpha_t)\gamma^2, & t < n, \\ z_n\exp(\alpha_n)\gamma^2, & t = n. \end{cases} \qquad (64)$$

Then the logarithm of the conditional density in (60) and (61) is given by (excluding a constant term)

$$\sum_{t=s}^{s+m-1} \log f(\eta_t) + L, \qquad (65)$$

where

$$L = \begin{cases} \sum_{t=s}^{s+m} l_t - \frac{(\alpha_{s+m+1} - \phi\alpha_{s+m})^2}{2\sigma_\eta^2}, & s + m < n, \\ \sum_{t=s}^{s+m} l_t, & s + m = n. \end{cases} \qquad (66)$$

19

Further, for $s \geq 0$, we define

$$\alpha = (\alpha_{s+1}, \ldots, \alpha_{s+m})', \tag{67}$$

$$d = (d_{s+1}, \ldots, d_{s+m})', \quad d_t = \frac{\partial L}{\partial \alpha_t}, \quad t = s+1, \ldots, s+m, \tag{68}$$

$$Q = -E\left[\frac{\partial^2 L}{\partial \alpha \partial \alpha'}\right] = \begin{bmatrix} A_{s+1} & B_{s+2} & 0 & \cdots & 0 \\ B_{s+2} & A_{s+2} & B_{s+3} & \cdots & 0 \\ 0 & B_{s+3} & A_{s+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B_{s+m} \\ 0 & \cdots & 0 & B_{s+m} & A_{s+m} \end{bmatrix}, \tag{69}$$

$$A_t = -E\left[\frac{\partial^2 L}{\partial \alpha_t^2}\right], \quad t = s+1, \ldots, s+m, \tag{70}$$

$$B_t = -E\left[\frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}}\right], \quad t = s+2, \ldots, s+m, \quad B_{s+1} = 0. \tag{71}$$

The first derivative of L with respect to α_t is given by

$$d_t = -\frac{1}{2} + \frac{(r_t - \mu_t)^2}{2\sigma_t^2} + \frac{r_t - \mu_t}{\sigma_t^2}\frac{\partial \mu_t}{\partial \alpha_t} + \frac{r_{t-1} - \mu_{t-1}}{\sigma_{t-1}^2}\frac{\partial \mu_{t-1}}{\partial \alpha_t} + \frac{\psi(x_t - c - \psi\alpha_t)}{\sigma_u^2} + \kappa(\alpha_t), \tag{72}$$

where

$$\frac{\partial \mu_t}{\partial \alpha_t} = \begin{cases} \left\{\frac{\beta \bar{z}_t}{2} + \sqrt{z_t}\rho\sigma_\eta^{-1}\left(-\phi + \frac{\alpha_{t+1} - \phi\alpha_t}{2}\right)\right\}\exp(\alpha_t/2)\gamma, & t < n, \\[2mm] \frac{\beta \bar{z}_n \exp(\alpha_n/2)\gamma}{2}, & t = n, \end{cases} \tag{73}$$

$$\frac{\partial \mu_{t-1}}{\partial \alpha_t} = \begin{cases} 0, & t = 1, \\[2mm] \sqrt{z_{t-1}}\rho\sigma_\eta^{-1}\exp(\alpha_{t-1}/2)\gamma, & t = 2, \ldots, T. \end{cases} \tag{74}$$

$$\kappa(\alpha_t) = \begin{cases} \frac{\phi(\alpha_{t+1} - \phi\alpha_t)}{\sigma_\eta^2}, & t = s+m < n \\[2mm] 0, & \text{otherwise.} \end{cases} \tag{75}$$

Taking expectations of second derivatives multiplied by -1 with respect to y_t's, we obtain the A_t's and B_t's as follows:

$$A_t = \frac{1}{2} + \sigma_t^{-2}\left(\frac{\partial \mu_t}{\partial \alpha_t}\right)^2 + \sigma_{t-1}^{-2}\left(\frac{\partial \mu_{t-1}}{\partial \alpha_t}\right)^2 + \frac{\psi^2}{\sigma_u^2} + \kappa'(\alpha_t), \tag{76}$$

$$B_t = \sigma_{t-1}^{-2}\frac{\partial \mu_{t-1}}{\partial \alpha_{t-1}}\frac{\partial \mu_{t-1}}{\partial \alpha_t}, \tag{77}$$

where

$$\kappa'(\alpha_t) = \begin{cases} \frac{\phi^2}{\sigma_\eta^2}, & t = s+m < n \\[2mm] 0, & \text{otherwise.} \end{cases} \tag{78}$$

20

Applying the second order Taylor expansion to the conditional density (60) will produce the approximating normal density $f^*(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m})$ as follows (see Omori and Watanabe (2008) for details):

$$\log f(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) \tag{79}$$

$$\approx \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t^2 + \hat{L} + \left.\frac{\partial L}{\partial \eta}\right|_{\eta=\hat{\eta}}(\eta - \hat{\eta}) + \frac{1}{2}(\eta - \hat{\eta})'\, E\left[\left.\frac{\partial^2 L}{\partial \eta \partial \eta'}\right]\right|_{\eta=\hat{\eta}}(\eta - \hat{\eta}) \tag{80}$$

$$= \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t^2 + \hat{L} + \hat{d}'(\alpha - \hat{\alpha}) - \frac{1}{2}(\alpha - \hat{\alpha})'\hat{Q}(\alpha - \hat{\alpha}) \tag{81}$$

$$= \text{const} + \log f^*(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}), \tag{82}$$

where $\eta = (\eta_s, \ldots, \eta_{s+m-1})'$, and $\hat{d}$, $\hat{L}$, and $\hat{Q}$ denote d, L, and Q evaluated at $\alpha = \hat{\alpha}$ (or, equivalently, at $\eta = \hat{\eta}$), respectively. The expectations are taken with respect to y_t's conditional on α_t's. Similarly, we can obtain the normal density which approximates the conditional density (61).

To make the linear Gaussian state-space model corresponding to the approximating density, we first compute the following D_t, K_t, J_t, and b_t for $t = s+2, \ldots, s+m$ recursively,

$$D_t = \hat{A}_t - D_{t-1}^{-1}\hat{B}_t^2, \quad D_{s+1} = \hat{A}_{s+1}, \tag{83}$$

$$K_t = \sqrt{D_t}, \tag{84}$$

$$J_t = \hat{B}_t K_{t-1}^{-1}, \quad J_{s+1} = 0, \quad J_{s+m+1} = 0, \tag{85}$$

$$b_t = \hat{d}_t - J_t K_{t-1}^{-1} b_{t-1}, \quad b_{s+1} = \hat{d}_{s+1}. \tag{86}$$

Second, we define auxiliary variables $\hat{y}_t = \hat{\gamma}_t + D_t^{-1}b_t$ where

$$\hat{\gamma}_t = \hat{\alpha}_t + K_t^{-1}J_{t+1}\hat{\alpha}_{t+1}, \quad t = s+1, \ldots, s+m. \tag{87}$$

Then the approximating density corresponds to the density of the linear Gaussian state-space model given by

$$\hat{y}_t = Z_t \alpha_t + G_t \zeta_t, \quad t = s+1, \ldots, s+m, \tag{88}$$

$$\alpha_{t+1} = \phi \alpha_t + H_t \zeta_t, \quad t = s, s+1, \ldots, s+m, \quad \zeta_t \sim \text{N}(0, I), \tag{89}$$

where

$$Z_t = 1 + K_t^{-1}J_{t+1}\phi, \quad G_t = K_t^{-1}(1,\, J_{t+1}\sigma_\eta), \quad H_t = (0,\, \sigma_\eta). \tag{90}$$

We can sample $(\eta_s, \ldots, \eta_{s+m-1})$ from the full posterior distribution in (60) and (61) by applying the simulation smoother (see, e.g., de Jong and Shephard (1995) and Durbin

and Koopman (2002)) to this state-space model and using Acceptance-Rejection (AR) MH algorithm proposed by Tierney (1994). See Omori and Watanabe (2008) and Takahashi, Omori, and Watanabe (2009) for the detail of the ARMH algorithm.

References

AAS, K., AND I. H. HAFF (2006): "The generalized hyperbolic skew Student's t-distribution," *Journal of Financial Econometrics*, 4(2), 275–309.

ANDERSEN, T. G., AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes , Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39(4), 885–905.

ANDERSEN, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2013): "Financial Risk Measurement for Financial Risk Management," *Handbook of the Economics of Finance*, 2(B), 1127–1220.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2010): "Parametric and nonparametric volatility measurement," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia, and L. P. Hansen, chap. 2, pp. 67–138. North Holland, Amsterdam.

BANDI, F. M., AND J. R. RUSSELL (2006): "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79(3), 655–692.

———— (2008): "Microstructure noise, realized variance, and optimal sampling," *Review of Economic Studies*, 75(2), 339–369.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

CHIB, S., AND E. GREENBERG (1995): "Understanding the Metropolis-Hastings algorithm," *American Statistician*, 49(4), 327.

CHIB, S., AND I. JELIAZKOV (2001): "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96(453), 270–281.

CHRISTOFFERSEN, P., AND D. PELLETIER (2004): "Backtesting value-at-risk: A duration-based approach," *Journal of Financial Econometrics*, 2(1), 84–108.

CHRISTOFFERSEN, P. F. (1998): "Evaluating interval forecasts," *International Economic Review*, 39(4), 841–862.

CLEMENTS, M. P., A. B. GALVÃO, AND J. H. KIM (2008): "Quantile forecasts of daily exchange rate returns from forecasts of realized volatility," *Journal of Empirical Finance*, 15(4), 729–750.

DE JONG, P., AND N. SHEPHARD (1995): "The simulation smoother for time series models," *Biometrika*, 82(2), 339–350.

DIEBOLD, F. X. (1988): *Empirical modeling of exchange rate dynamics*. Springer-Verlag, Berlin.

DOBREV, D. P., AND P. J. SZERSZEN (2010): "The Information Content of High-Frequency Data for Estimating Equity Return Models and Forecasting Risk," working paper.

DOORNIK, J. A. (2009): *An Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press and Oxford, London, 6 edn.

DUFOUR, J.-M. (2006): "Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics," *Journal of Econometrics*, 133(2), 443–477.

DURBIN, J., AND S. J. KOOPMAN (2002): "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89(3), 603–616.

EMBRECHTS, P., R. KAUFMANN, AND P. PATIE (2005): "Strategic long-term financial risks: Single risk factors," *Computational Optimization and Applications*, 32(1-2), 61–90.

ENGLE, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50(4), 987–1008.

ENGLE, R. F., AND J. R. RUSSELL (1998): "Autoregressive conditional duration : A new model for irregularly spaced transaction data," *Econometrica*, 66(5), 1127–1162.

GEWEKE, J. (1992): "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in *Bayesian Statistics 4*, ed. by J. M. Bernardo, J. O. Berger, A. Dawid, and A. Smith, pp. 169–193. Oxford.

GIOT, P., AND S. LAURENT (2004): "Modelling daily Value-at-Risk using realized volatility and ARCH type models," *Journal of Empirical Finance*, 11(3), 379–398.

HANSEN, P. R., AND Z. HUANG (2012): "Exponential GARCH modeling with realzied measures of volatility," working paper.

HANSEN, P. R., Z. HUANG, AND H. H. SHEK (2011): "Realized GARCH: A joint model for returns and realized measures of volatility," *Journal of Applied Econometrics*.

HANSEN, P. R., AND A. LUNDE (2005): "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?," *Journal of Applied Econometrics*, 20(7), 873–889.

HANSEN, P. R., AND A. LUNDE (2006): "Realized variance and market microstructure noise," *Journal of Business & Economic Statistics*, 24(2), 127–161.

HASBROUCK, J. (2007): *Empirical Market Microstructure: The Institutions, Economics and Econometrics of Securities Trading.* Oxford University Press, New York.

KOOPMAN, S. J., AND M. SCHARTH (2013): "The Analysis of Stochastic Volatility in the Presence of Daily Realized Measures," *Journal of Financial Econometrics*, 11(1), 76–115.

KUPIEC, P. (1995): "Techniques for verifying the accuracy of risk measurement models," *Journal of Derivatives*, 3(2), 73–84.

NAKAJIMA, J., AND Y. OMORI (2012): "Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution," *Computational Statistics & Data Analysis*, 56(11), 3690–3704.

NOURELDIN, D., N. SHEPHARD, AND K. SHEPPARD (2012): "Multivariate high-frequency-based volatility (HEAVY) models," *Journal of Applied Econometrics*, 27(6), 907–933.

O'HARA, M. (1995): *Market Microstructure Theory.* Blackwell, Oxford.

OMORI, Y., AND T. WATANABE (2008): "Block sampler and posterior mode estimation for asymmetric stochastic volatility models," *Computational Statistics & Data Analysis*, 52(6), 2892–2910.

PATTON, A. J. (2011): "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, 160(1), 246–256.

PITT, M. K., AND N. SHEPHARD (1999): "Filtering via simulation: Auxiliary particle filters," *Journal of the American Statistical Association*, 94(446), 590–599.

PRAUSE, K. (1999): "The generalized hyperbolic model: Estimation, financial derivatives, and risk measures," Phd dissertaion, University of Freiburg.

SHEPHARD, N. (1996): "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models in Econometrics, Fiannce and Other Fields*, ed. by D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, pp. 1–67. Chapman & Hall, New York.

SHEPHARD, N., AND M. K. PITT (1997): "Likelihood analysis of non-Gaussian measurement time series," *Biometrika*, 84(3), 653–667.

TAKAHASHI, M., Y. OMORI, AND T. WATANABE (2009): "Estimating stochastic volatility models using daily returns and realized volatility simultaneously," *Computational Statistics & Data Analysis*, 53(6), 2404–2426.

TAYLOR, S. J. (1986): *Modelling Financial Time Series*. Wiley, New York.

TROJAN, S. (2013): "Regime switching stochastic volatility with skew, fat tails and leverage using returns and realized volatility contemporaneously," Discussion Paper Series No. 2013-41, Department of Economics, School of Economics and Political Science, University of St. Gallen.

TIERNEY, L. (1994): "No Title," *Annals of Statistics*, 22(4), 1701–1762.

UBUKATA, M., AND K. OYA (2008): "Estimation and testing for dependence in market microstructure noise," *Journal of Financial Econometrics*, 7(2), 106–151.

WATANABE, T. (2012): "Quantile forecasts of financial returns using realized GARCH models," *Japanese Economic Review*, 63(1), 68–80.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, 100(472), 1394–1411.

Tables and Figures

Table 1: Descriptive statistics of the daily return (r), realized kernel (RK) for the estimation and prediction samples. The data is obtained from Oxford-Man Institute. JB is the p-value of the Jaque-Bera statistic to test the null hypothesis of normality. Standard errors of skewness and kurtosis are 0.0731 and 0.1460 for the estimation samples, respectively, whereas they are 0.0873 and 0.1744 for the prediction samples. LB is the p-value of the Ljung-Box statistic adjusted for heteroskedasticity following Diebold (1988) to test the null hypothesis of no autocorrelation up to 10 lags.

Estimation period: February 1, 2001 – July 20, 2005 (1121 samples)

Variable	Mean	SE	SD	Skew	Kurt	Min	Max	JB	LB
r	-0.000	0.035	1.185	0.132	5.106	-5.099	5.677	0.00	0.77
r^2	1.403	0.085	2.844	4.888	35.956	0.000	32.231	0.00	0.00
RK	0.991	0.036	1.208	4.731	37.018	0.049	13.145	0.00	0.00
$\ln r^2$	-1.288	0.073	2.433	-1.542	8.467	-17.941	3.473	0.00	0.00
$\ln RK$	-0.408	0.025	0.848	0.360	3.191	-3.023	2.576	0.00	0.00

Prediction period: July 21, 2005 – August 29, 2008 (784 samples)

Variable	Mean	SE	SD	Skew	Kurt	Min	Max	JB	LB
r	0.014	0.033	0.931	-0.182	4.979	-3.637	4.041	0.00	0.04
r^2	0.868	0.062	1.729	4.020	24.009	0.000	16.334	0.00	0.00
RK	0.629	0.027	0.756	3.903	26.045	0.049	7.396	0.00	0.00
$\ln r^2$	-1.927	0.091	2.547	-1.202	5.386	-13.649	2.793	0.00	0.00
$\ln RK$	-0.888	0.031	0.871	0.486	2.868	-3.011	2.001	0.00	0.00

Table 2: MCMC estimation results of RSV models with normal, student's t, and skewed t distributions obtained by 20000 samples recorded after discarding 5000 samples from MCMC iterations (all calculations in this paper are done by using Ox of Doornik (2009)). 95%L and 95%U are the lower and upper quantiles of 95% credible interval, respectively. The last two columns are the p-value of the convergence diagnostic test by Geweke (1992) and the inefficiency factor. Priors are set as $\mu \sim N(0, 10)$, $(\phi + 1)/2 \sim Beta(20, 1.5)$, $\sigma_\eta^{-2} \sim Gamma(2.5, 0.025)$, $(\rho + 1)/2 \sim Beta(1, 2)$, $\beta \sim N(0, 1)$, $\nu \sim Gamma(a, b)I(\nu > 4)$, $\xi \sim N(0, 1)$, $\psi \sim N(1, 1)$, $\sigma_u^{-2} \sim Gamma(2.5, 0.1)$.

Model		Mean	Stdev.	95%L	Median	95%U	CD	Inef.
RSVn	ϕ	0.9701	0.0068	0.9564	0.9702	0.9833	0.513	3.07
	σ_η	0.1789	0.0092	0.1616	0.1786	0.1976	0.013	14.92
	ρ	-0.3864	0.0488	-0.4794	-0.3871	-0.2883	0.008	15.85
	μ	-0.1937	0.1789	-0.5442	-0.1938	0.1664	0.280	2.27
	ξ	-0.2065	0.0414	-0.2891	-0.2060	-0.1281	0.889	21.55
	σ_u	0.3589	0.0100	0.3402	0.3586	0.3790	0.001	6.14
RSVCn	ϕ	0.9724	0.0066	0.9591	0.9725	0.9850	0.574	3.72
	σ_η	0.1990	0.0123	0.1763	0.1985	0.2238	0.197	34.56
	ρ	-0.4110	0.0501	-0.5043	-0.4126	-0.3077	0.367	18.66
	μ	-0.1690	0.2149	-0.5947	-0.1711	0.2674	0.375	1.45
	ξ	-0.2541	0.0427	-0.3390	-0.2539	-0.1723	0.625	26.74
	ψ	0.8485	0.0460	0.7629	0.8480	0.9420	0.240	32.77
	σ_u	0.3638	0.0102	0.3444	0.3636	0.3845	0.930	4.69

Table 2: MCMC estimation results of RSV models – Continued

Prior: $\nu \sim Gamma(5, 0.5)I(\nu > 4)$

Model		Mean	Stdev.	95%L	Median	95%U	CD	Inef.
RSVt	ϕ	0.9698	0.0069	0.9561	0.9699	0.9829	0.094	5.95
	σ_η	0.1791	0.0090	0.1624	0.1787	0.1978	0.474	22.08
	ρ	-0.3904	0.0496	-0.4832	-0.3913	-0.2905	0.357	14.45
	μ	-0.1867	0.1772	-0.5346	-0.1874	0.1674	0.109	2.17
	ν	24.7185	5.4374	15.9969	24.0579	37.4976	0.712	65.55
	ξ	-0.2135	0.0435	-0.3006	-0.2123	-0.1307	0.343	29.67
	σ_u	0.3590	0.0100	0.3403	0.3588	0.3792	0.865	6.96
RSVskt	ϕ	0.9697	0.0068	0.9561	0.9698	0.9828	0.238	4.60
	σ_η	0.1791	0.0090	0.1621	0.1789	0.1973	0.005	22.10
	ρ	-0.3910	0.0510	-0.4880	-0.3920	-0.2875	0.208	23.55
	μ	-0.1851	0.1786	-0.5344	-0.1872	0.1712	0.771	1.88
	β	0.1158	0.2687	-0.4163	0.1119	0.6571	0.082	25.72
	ν	25.3806	5.6189	15.9358	24.8142	38.1524	0.522	83.53
	ξ	-0.2176	0.0430	-0.3015	-0.2172	-0.1338	0.083	24.69
	σ_u	0.3590	0.0100	0.3401	0.3587	0.3793	0.003	9.07

Table 2: MCMC estimation results of RSV models – Continued

Prior: $\nu \sim Gamma(5, 0.5)I(\nu > 4)$

Model		Mean	Stdev.	95%L	Median	95%U	CD	Inef.
RSVCt	ϕ	0.9725	0.0066	0.9592	0.9727	0.9850	0.707	5.86
	σ_η	0.1987	0.0127	0.1750	0.1983	0.2250	0.608	48.13
	ρ	-0.4153	0.0504	-0.5113	-0.4164	-0.3145	0.860	14.99
	μ	-0.1642	0.2151	-0.5861	-0.1649	0.2592	0.749	3.26
	ν	23.6211	4.7493	15.7562	23.1892	34.1399	0.270	70.21
	ξ	-0.2583	0.0451	-0.3509	-0.2570	-0.1727	0.499	37.61
	ψ	0.8449	0.0475	0.7509	0.8444	0.9410	0.390	35.59
	σ_u	0.3641	0.0103	0.3443	0.3639	0.3852	0.125	5.90
RSVCskt	ϕ	0.9723	0.0065	0.9593	0.9724	0.9848	0.457	5.98
	σ_η	0.2012	0.0129	0.1776	0.2007	0.2287	0.591	43.09
	ρ	-0.4170	0.0513	-0.5125	-0.4183	-0.3145	0.454	26.18
	μ	-0.1592	0.2153	-0.5767	-0.1599	0.2758	0.001	3.13
	β	0.0523	0.2633	-0.4644	0.0506	0.5916	0.156	24.24
	ν	25.4040	5.5816	15.6377	24.8553	37.8619	0.811	107.91
	ξ	-0.2656	0.0448	-0.3565	-0.2645	-0.1804	0.155	29.57
	ψ	0.8376	0.0471	0.7410	0.8381	0.9267	0.473	47.12
	σ_u	0.3643	0.0104	0.3446	0.3640	0.3857	0.468	6.46

Table 2: MCMC estimation results of RSV models – Continued

Prior: $\nu \sim Gamma(1, 0.1)I(\nu > 4)$

Model		Mean	Stdev.	95%L	Median	95%U	CD	Inef.
RSVt	ϕ	0.9699	0.0068	0.9564	0.9700	0.9830	0.014	3.60
	σ_η	0.1789	0.0089	0.1618	0.1788	0.1965	0.001	18.73
	ρ	-0.3895	0.0486	-0.4832	-0.3904	-0.2912	0.005	15.18
	μ	-0.1915	0.1801	-0.5374	-0.1917	0.1624	0.118	2.22
	ν	45.3320	17.4236	20.6393	41.9753	87.8133	0.603	155.55
	ξ	-0.2119	0.0437	-0.2983	-0.2121	-0.1282	0.129	23.59
	σ_u	0.3591	0.0100	0.3402	0.3588	0.3799	0.004	7.21
RSVskt	ϕ	0.9697	0.0070	0.9556	0.9698	0.9831	0.905	6.59
	σ_η	0.1793	0.0091	0.1627	0.1789	0.1981	0.795	21.24
	ρ	-0.3896	0.0501	-0.4849	-0.3905	-0.2874	0.313	23.06
	μ	-0.1921	0.1794	-0.5474	-0.1912	0.1612	0.118	1.95
	β	0.2158	0.4620	-0.7191	0.2157	1.1484	0.538	24.45
	ν	49.3075	15.2241	25.5010	47.1330	84.3186	0.556	132.21
	ξ	-0.2122	0.0426	-0.2986	-0.2110	-0.1296	0.501	28.72
	σ_u	0.3586	0.0100	0.3400	0.3582	0.3790	0.916	4.09

Table 2: MCMC estimation results of RSV models – Continued

Prior: $\nu \sim Gamma(1, 0.1)I(\nu > 4)$

Model		Mean	Stdev.	95%L	Median	95%U	CD	Inef.
RSVCt	ϕ	0.9724	0.0066	0.9591	0.9726	0.9852	0.358	5.41
	σ_η	0.2011	0.0125	0.1780	0.2005	0.2272	0.863	47.16
	ρ	-0.4124	0.0505	-0.5080	-0.4131	-0.3119	0.187	27.24
	μ	-0.1645	0.2179	-0.5866	-0.1672	0.2743	0.271	1.79
	ν	43.5052	15.4304	21.3596	40.9325	81.3445	0.002	154.19
	ξ	-0.2581	0.0424	-0.3418	-0.2578	-0.1765	0.882	23.21
	ψ	0.8388	0.0460	0.7489	0.8381	0.9313	0.360	36.79
	σ_u	0.3638	0.0104	0.3444	0.3635	0.3853	0.815	6.77
RSVCskt	ϕ	0.9724	0.0067	0.9592	0.9726	0.9852	0.751	5.66
	σ_η	0.2003	0.0126	0.1770	0.1998	0.2263	0.848	44.86
	ρ	-0.4100	0.0512	-0.5077	-0.4109	-0.3080	0.037	16.41
	μ	-0.1665	0.2172	-0.5917	-0.1685	0.2641	0.236	2.21
	β	0.1167	0.4611	-0.8078	0.1091	1.0583	0.002	40.76
	ν	48.4294	16.5012	23.6108	45.5288	88.1198	0.598	122.70
	ξ	-0.2585	0.0428	-0.3465	-0.2575	-0.1760	0.874	39.76
	ψ	0.8412	0.0455	0.7519	0.8403	0.9322	0.821	42.17
	σ_u	0.3637	0.0103	0.3442	0.3634	0.3849	0.231	6.91

Table 2: MCMC estimation results of RSV models – Continued

Prior: $\nu \sim Gamma(0.1, 0.01)I(\nu > 4)$

Model		Mean	Stdev.	95%L	Median	95%U	CD	Inef.
RSVt	ϕ	0.9699	0.0069	0.9560	0.9701	0.9832	0.922	9.15
	σ_η	0.1793	0.0092	0.1622	0.1790	0.1982	0.949	36.72
	ρ	-0.3859	0.0491	-0.4805	-0.3869	-0.2863	0.249	22.22
	μ	-0.1893	0.1790	-0.5376	-0.1892	0.1643	0.449	1.19
	ν	167.7671	110.2217	39.6042	140.5874	461.6067	0.521	352.87
	ξ	-0.2110	0.0427	-0.2962	-0.2108	-0.1287	0.977	25.55
	σ_u	0.3588	0.0099	0.3401	0.3586	0.3785	0.948	9.39
RSVskt	ϕ	0.9697	0.0070	0.9559	0.9697	0.9832	0.454	7.37
	σ_η	0.1796	0.0092	0.1622	0.1793	0.1987	0.969	27.67
	ρ	-0.3870	0.0491	-0.4799	-0.3886	-0.2879	0.009	19.37
	μ	-0.1923	0.1782	-0.5459	-0.1932	0.1660	0.926	1.77
	β	0.2560	0.8128	-1.3915	0.2600	1.8558	0.497	39.57
	ν	170.8826	117.3749	42.0413	137.2567	479.2125	0.635	379.54
	ξ	-0.2103	0.0425	-0.2968	-0.2098	-0.1279	0.821	18.89
	σ_u	0.3585	0.0101	0.3394	0.3582	0.3792	0.034	7.12

Table 2: MCMC estimation results of RSV models – Continued

Prior: $\nu \sim Gamma(0.1, 0.01)I(\nu > 4)$

Model		Mean	Stdev.	95%L	Median	95%U	CD	Inef.
RSVCt	ϕ	0.9724	0.0065	0.9593	0.9725	0.9849	0.798	3.35
	σ_η	0.1990	0.0130	0.1751	0.1984	0.2265	0.699	35.18
	ρ	-0.4115	0.0496	-0.5057	-0.4128	-0.3102	0.359	8.99
	μ	-0.1689	0.2140	-0.5891	-0.1725	0.2679	0.220	3.10
	ν	127.2060	74.9582	35.4274	109.9902	310.5359	0.598	233.67
	ξ	-0.2537	0.0432	-0.3427	-0.2523	-0.1728	0.050	32.59
	ψ	0.8475	0.0480	0.7519	0.8476	0.9424	0.502	33.77
	σ_u	0.3640	0.0103	0.3449	0.3637	0.3850	0.944	5.76
RSVCskt	ϕ	0.9723	0.0067	0.9589	0.9724	0.9851	0.403	4.89
	σ_η	0.2010	0.0127	0.1775	0.2005	0.2273	0.330	59.16
	ρ	-0.4096	0.0522	-0.5082	-0.4109	-0.3038	0.044	29.17
	μ	-0.1687	0.2174	-0.5952	-0.1708	0.2697	0.685	3.38
	β	0.0940	0.8431	-1.5395	0.0854	1.8665	0.423	38.76
	ν	153.6488	92.9665	40.0863	132.3362	399.1619	0.171	310.08
	ξ	-0.2572	0.0442	-0.3458	-0.2561	-0.1732	0.678	35.12
	ψ	0.8416	0.0465	0.7472	0.8416	0.9316	0.759	33.33
	σ_u	0.3634	0.0103	0.3441	0.3631	0.3845	0.006	6.13

Table 3: Marginal likelihood and its components, likelihood, prior, and posterior (in logarithm). The likelihood is estimated using the auxiliary particle filter of Pitt and Shephard (1999) with 10000 particles. The likelihood estimate and its standard error are computed as the sample mean and standard deviation of the likelihoods from 10 iterations. The posterior probability density and its numerical standard error are evaluated by the method of Chib and Greenberg (1995) and Chib and Jeliazkov (2001). The numbers in the parentheses show the standard errors.

Model	Likelihood		Prior	Posterior		Marginal	
RSVn	-2179.16	(0.25)	0.29	18.20	(0.01)	-2197.07	(0.25)
RSVCn	-2176.05	(0.25)	-1.32	20.77	(0.02)	-2198.15	(0.25)

Prior: $\nu \sim Gamma(5, 0.5)I(\nu > 4)$							
RSVt	-2183.77	(0.25)	-5.83	13.94	(0.03)	-2203.54	(0.26)
RSVskt	-2183.29	(0.26)	-6.99	13.10	(0.04)	-2203.37	(0.26)
RSVCt	-2181.04	(0.23)	-7.07	16.57	(0.03)	-2204.68	(0.23)
RSVCskt	-2180.52	(0.31)	-8.67	15.56	(0.04)	-2204.75	(0.31)

Prior: $\nu \sim Gamma(1, 0.1)I(\nu > 4)$							
RSVt	-2181.44	(0.21)	-6.15	12.90	(0.05)	-2200.49	(0.21)
RSVskt	-2180.83	(0.23)	-7.49	11.59	(0.05)	-2199.91	(0.23)
RSVCt	-2179.11	(0.29)	-7.64	15.46	(0.05)	-2202.21	(0.29)
RSVCskt	-2178.53	(0.26)	-9.03	14.18	(0.05)	-2201.75	(0.26)

Prior: $\nu \sim Gamma(0.1, 0.01)I(\nu > 4)$							
RSVt	-2179.60	(0.28)	-7.30	10.66	(0.10)	-2197.56	(0.29)
RSVskt	-2179.50	(0.11)	-8.30	9.29	(0.09)	-2197.10	(0.14)
RSVCt	-2176.94	(0.16)	-8.25	13.79	(0.08)	-2198.98	(0.18)
RSVCskt	-2177.02	(0.19)	-9.66	11.90	(0.09)	-2198.58	(0.21)

Table 4: Mean squared error (MSE) and quasi-likelihood (QLIKE) of volatility forecasts from July 21, 2005 to August 29, 2008. Realized kernel with the adjustment of Hansen and Lunde (2005) are used as a proxy of latent volatility.

Model	MSE	QLIKE
RSVn	0.2150	0.1473
RSVt	0.2209	0.1515
RSVskt	0.2981	0.1672
RSVCn	0.3337	0.1594
RSVCt	0.3863	0.1746
RSVCskt	0.5316	0.1836

Table 5: Backtesting measure of Embrechts, Kaufmann, and Patie (2005) for the expected shortfall forecasts from July 21, 2005 to August 29, 2008. $|V_1|$ is an absolute value of the sample average of $\delta_t(\alpha) = r_t - ES_t(\alpha)$ given a violation of the VaR occurs, that is, $r_t < VaR(\alpha)$, whereas $|V_2|$ is an absolute value of the sample average of $\delta_t(\alpha)$ below its empirical α-quantile. V is an average of those two measures, $V = (|V_1| + |V_2|)/2$.

Model	$|V_1|$	$|V_2|$	V	$|V_1|$	$|V_2|$	V
		0.5%			1%	
RSVn	0.2366	0.8422	0.5394	0.2139	0.6614	0.4377
RSVt	0.1630	0.7106	0.4368	0.2223	0.5137	0.3680
RSVskt	0.1229	0.7561	0.4395	0.2163	0.5087	0.3625
RSVCn	0.1512	0.7881	0.4697	0.1435	0.5940	0.3687
RSVCt	0.0484	0.6520	0.3502	0.1221	0.4894	0.3058
RSVCskt	0.0062	0.6310	0.3186	0.1449	0.4572	0.3011
		5%			10%	
RSVn	0.1835	0.3878	0.2857	0.2177	0.2800	0.2489
RSVt	0.1192	0.3166	0.2179	0.1689	0.2466	0.2078
RSVskt	0.0523	0.2958	0.1741	0.0850	0.2262	0.1556
RSVCn	0.1617	0.3589	0.2603	0.1795	0.2611	0.2203
RSVCt	0.1044	0.3096	0.2070	0.1278	0.2417	0.1848
RSVCskt	0.0798	0.3198	0.1998	0.1434	0.2399	0.1917

Table 6: Empirical failure rates for the VaR forecasts from July 21, 2005 to August 29, 2008. π_1 is an empirical probability of VaR violations. π_{01} is the empirical probability of VaR violations conditional on no VaR violation on previous day while π_{11} is the one conditional on VaR violation on previous day.

Model	π_1	π_{01}	π_{11}	π_1	π_{01}	π_{11}
		0.5%			1%	
RSVn	0.0191	0.0195	0.0000	0.0268	0.0275	0.0000
RSVt	0.0153	0.0155	0.0000	0.0204	0.0208	0.0000
RSVskt	0.0153	0.0155	0.0000	0.0204	0.0208	0.0000
RSVCn	0.0204	0.0208	0.0000	0.0306	0.0316	0.0000
RSVCt	0.0191	0.0195	0.0000	0.0242	0.0248	0.0000
RSVCskt	0.0179	0.0182	0.0000	0.0230	0.0235	0.0000
		5%			10%	
RSVn	0.0765	0.0773	0.0667	0.1071	0.1100	0.0833
RSVt	0.0740	0.0744	0.0690	0.1097	0.1117	0.0814
RSVskt	0.0753	0.0759	0.0678	0.1135	0.1165	0.0787
RSVCn	0.0740	0.0744	0.0690	0.1122	0.1149	0.0795
RSVCt	0.0753	0.0759	0.0678	0.1186	0.1230	0.0753
RSVCskt	0.0753	0.0759	0.0678	0.1148	0.1182	0.0778

Table 7: Finite sample p-values of the Markov, Weibull, and EACD tests for the VaR forecasts from July 21, 2005 to August 29, 2008. We compute the finite sample p-values based on the Monte Carlo testing technique of Dufour (2006).

Model	Markov	Weibull	EACD	Markov	Weibull	EACD
		0.5%			1%	
RSVn	0.0192	0.2072	0.0001	0.0010	0.7013	0.0001
RSVt	0.0242	0.0255	0.0159	0.0241	0.0427	0.0039
RSVskt	0.0219	0.0250	0.0146	0.0240	0.2652	0.0043
RSVCn	0.0192	0.8418	0.0001	0.0005	0.5488	0.0012
RSVCt	0.0194	0.1449	0.0019	0.0024	0.9567	0.0075
RSVCskt	0.0205	0.9452	0.0021	0.0047	0.7245	0.0018
		5%			10%	
RSVn	0.0127	0.0355	0.0090	0.6398	0.2283	0.0048
RSVt	0.0289	0.0540	0.0106	0.1014	0.1098	0.0084
RSVskt	0.0236	0.0454	0.0066	0.0592	0.1006	0.0068
RSVCn	0.0309	0.0268	0.0093	0.0712	0.2014	0.0336
RSVCt	0.0215	0.0462	0.0051	0.0227	0.1230	0.0806
RSVCskt	0.0220	0.0180	0.0077	0.0502	0.1130	0.0239

Figure 1: The posterior mean and the 95% credible interval of the parameter β for the RSVskt and RSVCskt models obtained from the rolling estimation using 1121 samples each time. The first posterior mean and 95% interval are obtained from the MCMC estimation using 1121 samples from February 1, 2001 through July 20, 2005. The second ones correspond to the one using one day ahead 1121 samples from February 2, 2001 through July 21, 2005. We continue the rolling estimation until the 1121 samples from March 18, 2004 through August 28, 2008. The horizontal axis corresponds to the last day of samples for each MCMC estimation.

RSVskt

RSVCskt

39
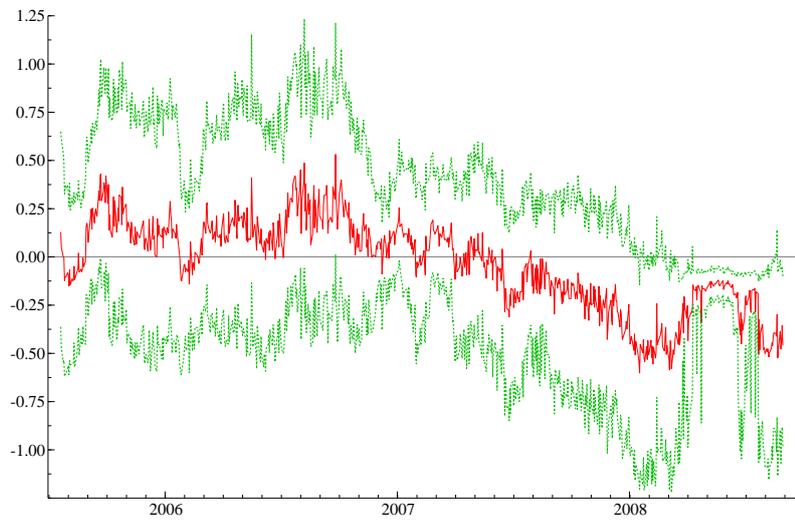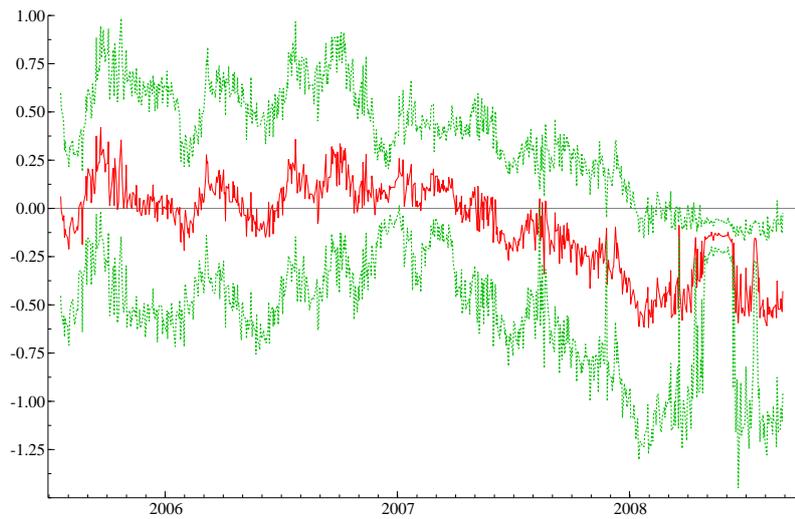
Figure 2: Daily returns and the VaR forecasts (5%) of the RSV models from July 21, 2005 to August 29, 2008.

RSVn

RSVCn

RSVt

RSVCt

RSVskt

RSVCskt