CIRJE-F-949

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

December 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 by employing a wider class distribution, the generalized hyperbolic skew Student's t-distribution, for financial returns. The extension makes it possible 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 the US index, Dow Jones Industrial Average, show that the extended model fits the data better and improves the volatility and quantile 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.

*We would like to thank seminar participants at Hitotsubashi University and Osaka University, Esther Ruiz (editor), the associate editor, and the anonymous referee for their helpful comments and suggestions. Financial support from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government through Grant-in-Aid for Scientific Research (No. 21243018; 22243021; 26245028), the Global COE program "Research Unit for Statistical and Empirical Analysis in Social Science" at Hitotsubashi University and the Joint Usage and Research Center, Institute of Economic Research, Hitotsubashi University (IERPK1206) is gratefully acknowledged. All remaining errors are solely our responsibility.

†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.[1] Another well-known family is the stochastic volatility (SV) models, developed by Taylor (1986). Shephard (1996) provides a comprehensive explanation of the SV models. The unconditional distribution of financial returns are known to be leptokurtic. This leptokurtosis can fully or partly be captured by time-varying volatility, but the distribution conditional on volatility may still be leptokurtic and also be skewed. Aas and Haff (2006) introduce the general distribution class, called generalized hyperbolic distribution, which can take a flexible form to fit the return characteristics.

On the other hand, 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.[2]

The RV, 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 9:30 a.m.

[1] See, for example, Andersen, Bollerslev, Christoffersen, and Diebold (2013).

[2] More detail properties of the RV can be found in Andersen, Bollerslev, and Diebold (2010) and references therein.

to 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,[3] which also causes a bias in the estimated RV. Hansen and Lunde (2006) and Ubukata and Oya (2008) study the effects of the market microstructure noise on the RV estimator. In general, the MMN effect becomes larger at the higher sampling frequency while the information loss becomes larger at the lower frequency.

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, a contemporaneous modeling of financial returns and the RV estimators, which can adjust the bias within the models, have been proposed. For example, Takahashi, Omori, and Watanabe (2009) propose to model daily returns and the RV estimator simultaneously under the framework of the SV model. Additionally, 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 SV model extended with the RV estimator as the realized SV (RSV) model. On the other hand, Hansen, Huang, and Shek (2011) propose to extend GARCH models incorporating them with the RV, which is called the realized GARCH model. The contemporaneous models can adjust the bias in the RV estimator 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 in the RV estimator caused by the MMN. The RSV model,

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

however, has not been fully applied to quantile forecasts.[4]

In this paper, we extend the RSV model of Takahashi, Omori, and Watanabe (2009) with 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 in 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 the US index, Dow Jones Industrial Average (DJIA). The estimation results show that the RSV model with the GH skew Student's t-distribution fits the data well in the sense that the model gives the higher marginal likelihood than the RSV models with normal and Student's t-distributions. Moreover, the prediction results show that the extended model improves both volatility and quantile forecasts.

The paper is organized as follows. In Section 2, we present the basic and extended RSV models with a brief description of the SV model and RV estimators. Then, we explain the estimation and prediction scheme to estimate the parameters, volatility and quantile forecasts jointly via the MCMC technique in Section 3. Further, we introduce several methods to evaluate the volatility and quantile forecasts in Section 4. We present the empirical results using the DJIA data in Section 5. Finally, we conclude the paper in Section 6.

2 Realized Stochastic Volatility Model

In this section, we describe the RSV model, which incorporates the asymmetric SV model with the RV estimator. In Section 2.1, we introduce the SV model and then briefly describe the RV estimator in Section 2.2. We introduce the basic RSV model proposed by Takahashi, Omori, and Watanabe (2009) and present its extension in Section 2.3.

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

2.1 Stochastic Volatility Model

The asymmetric SV model is written as

$$r_t = \exp(h_t/2)\epsilon_t, \quad t = 1, \ldots, n, \tag{1}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, \ldots, n-1, \tag{2}$$

where r_t is an asset return and h_t is an unobserved log-volatility. It is common to assume that $|\phi| < 1$ for a stationarity of the log-volatility process. For the moment, we assume the normality for the return and volatility innovations as follows,

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix}. \tag{3}$$

The parameter ρ in (3) represents the correlation between ϵ_t and η_t, which 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 a well known phenomenon in stock markets and referred to as a volatility asymmetry.[5] Additionally, we assume the following initial conditions,

$$h_0 = \mu, \quad \eta_0 \sim N\left(0, \frac{\sigma_\eta^2}{1-\phi^2}\right). \tag{4}$$

2.2 Realized Volatility

We first consider a simple continuous time process,

$$dp(s) = \sigma(s)dw(s), \tag{5}$$

where $p(s)$ denotes the log price of a financial asset at time s, and $\sigma^2(s)$ is the instantaneous or spot volatility, which is assumed to be stochastically independent of the Wiener process $w(s)$. Then, the true volatility for a day t is defined as

$$\sigma_t^2 = \int_t^{t+1} \sigma^2(s)ds, \tag{6}$$

which is called an integrated volatility.

Andersen and Bollerslev (1998) propose a model-free estimator of the true volatility σ_t^2, which is called a RV estimator. Suppose that we have m intraday returns during the day t,

[5] See, for example, Black (1976) and Christie (1982).

$\{r_{t,i}\}_{i=1}^{m}$, then a simple RV estimator is defined as a sum of squared returns,

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2, \tag{7}$$

which converges to the true volatility σ_t^2 as $m \to \infty$. That is, RV_t is a consistent estimator of σ_t^2 and thus may provide a precise estimate of the true volatility when there are sufficient number of intraday returns.

There are, however, some problems in computing the RV estimator using the high frequency data. First, the high frequency asset price contains the market microstructure noise (MMN) such as a bid-ask bounce and non-synchronous trading.[6] With the presence of the MMN, the RV estimator is biased and is not a consistent estimator of the true volatility. Hansen and Lunde (2006) and Ubukata and Oya (2008) study the MMN effects on the RV estimator. In general, the MMN effect becomes larger at the higher sampling frequency while the information loss becomes larger at the lower frequency.

There are several methods available for mitigating the MMN effects on the RV estimators. For example, Aït-Sahalia, Mykland, and Zhang (2005) and Bandi and Russell (2006, 2008) derive an optimal sampling frequency to balance the trade off between the MMN effect and the information loss. Additionally, Zhang, Mykland, and Aït-Sahalia (2005) propose a two (multi) scale estimator, which combines two (multiple) RV estimators calculated from returns with different sampling frequencies. Moreover, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) propose a realized kernel (RK) estimator,

$$RK_t = \sum_{q=-Q}^{Q} k\left(\frac{q}{Q+1}\right)\gamma_q, \quad \gamma_q = \sum_{i=|q|+1}^{m} r_{t,i} r_{t,i-|q|}, \tag{8}$$

where $k(\cdot) \in [0,1]$ is a non-stochastic weight function. As for the choice of $k(\cdot)$, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) suggests the Parzen kernel given by

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \le x \le 1/2 \\ 2(1-x)^3 & 1/2 \le x \le 1 \\ 0 & x > 1, \end{cases} \tag{9}$$

which satisfies the smoothness conditions, $k'(0) = k'(1) = 0$, and is guaranteed to produce a non-negative estimate.

The second problem in computing the RV estimator is the presence of the non-trading hours. For example, New York Stock Exchange is open only for six and a half hours from

[6] See, for example, Campbell, Lo, and MacKinlay (1997) for details.

9:30 a.m. to 4 p.m. (in Eastern Time). If we calculate the RV estimator using the intraday returns only in the market open period, it may underestimate the true volatility σ_t^2. To avoid this underestimation, Hansen and Lunde (2005) propose to scale the RV calculated from returns for the market open period as

$$RV_t^{scale} = cRV_t, \quad c = \frac{\sum_{t=1}^n (r_t - \bar{r})^2}{\sum_{t=1}^n RV_t}, \tag{10}$$

where r_t is the daily return and $\bar{r} = \sum_{t=1}^n r_t/n$. This ensures that the mean of the scaled RV (RV_t^{scale}) is equal to the variance of daily returns.[7]

2.3 Realized Stochastic Volatility Model

Takahashi, Omori, and Watanabe (2009) propose modeling daily returns and the RV estimator simultaneously as follows,

$$r_t = \exp(h_t/2)\epsilon_t, \quad t = 1, \ldots, n, \tag{11}$$

$$x_t = \xi + h_t + u_t, \quad t = 1, \ldots, n, \tag{12}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, \ldots, n-1, \tag{13}$$

where x_t is a logarithm of the RV estimator. The parameter ξ in (12) is designed to correct 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. We assume that the disturbance u_t and other disturbances (ϵ_t, η_t) are not correlated, that is,

$$\begin{bmatrix} \epsilon_t \\ u_t \\ \eta_t \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{bmatrix}. \tag{14}$$

Dobrev and Szerszen (2010) and Koopman and Scharth (2013) also propose the joint modeling of daily returns and the realized volatility based on the SV model. Following Koopman and Scharth (2013), we refer the model consisting of (11)-(14) as a realized stochastic volatility (RSV) model.

We extend the RSV model in (11)-(14) with more general distribution for daily returns. Following Nakajima and Omori (2012), we employ the general hyperbolic (GH) skew Stu-

[7] One may consider including returns for the non-trading hours (overnight interval) but this can make the RV estimator less precise since such returns contain much discretization noise.

dent's t-distribution for the return distribution.[8] Specifically, the return equation (11) is extended as follows,

$$r_t = \frac{\beta(z_t - \mu_z) + \sqrt{z_t}\epsilon_t}{\sqrt{\beta^2\sigma_z^2 + \mu_z}} \exp(h_t/2), \quad t = 1, \ldots, n, \tag{15}$$

where

$$z_t \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \quad \mu_z = E[z_t] = \frac{\nu}{\nu - 2}, \quad \sigma_z^2 = Var[z_t] = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}, \tag{16}$$

and $IG(\cdot, \cdot)$ denotes the inverse gamma distribution. We assume that $\nu > 4$ 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 remains $\exp(h_t)$. This specification includes the Student's t-distribution as a special case when $\beta = 0$ as well as the normal distribution when $\beta = 0$ and $\nu \to \infty$ (that is, $z_t = 1$ for all t). 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.[9]

3 Estimation and Prediction Scheme

In this section, we describe the estimation and prediction scheme for the RSVskt model. In Section 3.1, we present a Bayesian estimation procedure via Markov chain Monte Carlo method. Then, we explain how to obtain the volatility and quantile forecasts within the Bayesian estimation procedure in Section 3.2.

[8]The GH skew Student's t-distribution is a subclass of the GH distribution. 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.

[9]We can extend the RV equation (12) as follows,

$$x_t = \xi + \psi h_t + u_t, \quad t = 1, \ldots, n.$$

Hansen, Huang, and Shek (2011) first consider this type of specification in their realized GARCH framework which is the joint modeling of daily returns and the RV estimator based on the GARCH type models. We estimate the RSV models with this specification but it turns out that this extension does not improve the volatility forecasts nor quantile forecasts. Therefore, we focus on the RSV models with $\psi = 1$ in this paper.

3.1 Bayesian Estimation Procedure

The RSVskt model is written as

$$r_t = \frac{\beta(z_t - \mu_z) + \sqrt{z_t}\epsilon_t}{\sqrt{\beta^2\sigma_z^2 + \mu_z}} \exp(h_t/2), \quad t = 1, \ldots, n, \tag{17}$$

$$x_t = \xi + h_t + u_t, \quad t = 1, \ldots, n, \tag{18}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, \ldots, n-1, \tag{19}$$

where

$$z_t \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \quad \mu_z = E[z_t] = \frac{\nu}{\nu - 2}, \quad \sigma_z^2 = Var[z_t] = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}, \tag{20}$$

and

$$\begin{bmatrix} \epsilon_t \\ u_t \\ \eta_t \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{bmatrix}. \tag{21}$$

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 $\theta = (\phi, \sigma_\eta, \rho, \mu, \beta, \nu, \xi, \sigma_u)$, $y = \{r_t, x_t\}_{t=1}^n$, $h = \{h_t\}_{t=1}^n$, 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 $\phi|\sigma_\eta, \rho, \mu, \beta, \nu, \xi, \sigma_u, h, z, y$.
2. Generate $(\sigma_\eta, \rho)|\phi, \mu, \beta, \nu, \xi, \sigma_u, h, z, y$.
3. Generate $\mu|\phi, \sigma_\eta, \rho, \beta, \nu, \xi, \sigma_u, h, z, y$.
4. Generate $\beta|\phi, \sigma_\eta, \rho, \mu, \nu, \xi, \sigma_u, h, z, y$.
5. Generate $\nu|\phi, \sigma_\eta, \rho, \mu, \beta, \xi, \sigma_u, h, z, y$.
6. Generate $\xi|\phi, \sigma_\eta, \rho, \mu, \beta, \nu, \sigma_u, h, z, y$.
7. Generate $\sigma_u|\phi, \sigma_\eta, \rho, \mu, \beta, \nu, \xi, h, z, y$.
8. Generate $z|\theta, h, y$.
9. Generate $h|\theta, z, y$.

10. 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 8. We can also easily modify the sampling scheme by Takahashi, Omori, and Watanabe (2009) for steps 6, 7, and 9. The detail procedures are given in Appendix A.

3.2 Volatility and Quantile Forecasts

To obtain the one-day-ahead log-volatility and daily return, we implement the following sampling scheme for each sample of (θ, h, z) generated from the MCMC algorithm described above.

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

$$\mu_{n+1} = \mu + \phi(h_n - \mu) + \rho\sigma_\eta \frac{\sqrt{\beta^2\sigma_z^2 + \mu_z} r_n - \beta(z_n - \mu_z)\exp(h_n/2)}{\sqrt{z_n}\exp(h_n/2)}, \tag{22}$$

$$\sigma_{n+1}^2 = (1-\rho^2)\sigma_\eta^2. \tag{23}$$

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

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

$$\hat{\mu}_{n+1} = \frac{\beta(z_{n+1} - \mu_z)\exp(h_{n+1}/2)}{\sqrt{\beta^2\sigma_z^2 + \mu_z}}, \tag{24}$$

$$\hat{\sigma}_{n+1}^2 = \frac{z_{n+1}\exp(h_{n+1})}{\beta^2\sigma_z^2 + \mu_z}. \tag{25}$$

The quantile forecasts, VaR and ES, can easily be computed from the predictive distribution of financial returns obtained above. Let $\mathrm{VaR}_t(\alpha)$ 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

$$\Pr[r_t < \mathrm{VaR}_t(\alpha)|\mathcal{I}_{t-1}] = \alpha, \tag{26}$$

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 the 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, $\mathrm{ES}_t(\alpha)$, satisfies

$$\mathrm{ES}_t(\alpha) = \mathrm{E}[r_t | r_t < \mathrm{VaR}_t(\alpha), I_{t-1}]. \tag{27}$$

Let n and T be the number of samples for the estimation and prediction, respectively. Then, the one-day-ahead forecasts of the VaR $(\mathrm{VaR}_{n+1}(\alpha), \ldots, \mathrm{VaR}_{n+T}(\alpha))$ and the ES $(\mathrm{ES}_{n+1}(\alpha), \ldots, \mathrm{ES}_{n+T}(\alpha))$ 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 $(y_i, \ldots, y_{n+i-1})$.
3. Compute $\mathrm{VaR}_{n+i}(\alpha)$ as the α-percentile of the MCMC sample of r_{n+i}.
4. Compute $\mathrm{ES}_{n+i}(\alpha)$ as a sample average of r_{n+i} conditional on $r_{n+i} < \mathrm{VaR}_{n+i}(\alpha)$.
5. Set $i = i + 1$ and return to 1 while $i < T$.

4 Evaluation of Volatility and Quantile Forecasts

In this section, we describe how to evaluate the predictive ability of the RSV models with different specifications. Since there is no single measure which ranks the models thoroughly, we compare the model performance from the various perspectives. In Section 4.1, we introduce two loss functions for the volatility forecasts. In Section 4.2, we describe various backtesting methods for the VaR forecasts. In Section 4.3, we present a backtesting measure of the expected shortfall forecasts.

4.1 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 by[10]

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

[10] Both loss functions are normalized to be the robust and homogeneous loss functions proposed by Patton (2011). For instance, $QLIKE$ is normalized to yield a distance of zero when $\hat{\sigma}_t^2 = h_t$.

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 the volatility, that is, $\mathrm{E}_{t-1}[\hat{\sigma}_t^2] = \sigma_t^2$.

4.2 Backtesting Value-at-Risk

To describe the backtesting methods for the VaR forecasts formally, let T be the number of VaR forecasts and T_1 be the number of times when the VaR is violated, that is, $r_t < \mathrm{VaR}_t(\alpha)$. 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,

$$L(p) = p^{T_1}(1-p)^{T-T_1}. \tag{29}$$

The LR statistic of the unconditional coverage test is then

$$LR_{uc} = 2\{\ln L(\hat{\pi}_1) - \ln L(\alpha)\}, \tag{30}$$

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

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}, \tag{31}$$

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

$$L(\pi_{01}, \pi_{11}) = (1-\pi_{01})^{T_0-T_{01}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_1-T_{11}}\pi_{11}^{T_{11}}, \tag{32}$$

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

$$LR_{ind} = 2\{\ln L(\hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(\hat{\pi}_1)\}, \tag{33}$$

which is again asymptotically distributed as a $\chi^2(1)$ under the null hypothesis.[11]

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

$$LR_{cc} = LR_{uc} + LR_{ind} = 2\{\ln L(\hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(\alpha)\}, \tag{34}$$

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

$$D_i = t_i - t_{i-1}, \tag{35}$$

where t_i denotes the day of the 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

$$f_{\exp}(D; \alpha) = \alpha \exp(-\alpha D). \tag{36}$$

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

$$f_W(D; a, b) = a^b b D^{b-1} \exp\{-(aD)^b\}, \tag{37}$$

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.

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

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

$$\psi_i = \mathrm{E}[D_i] = c + dD_{i-1}, \tag{38}$$

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

$$f_{EACD}(D_i|\psi_i) = \frac{1}{\psi_i}\exp\left(-\frac{D_i}{\psi_i}\right). \tag{39}$$

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

$$\begin{aligned}\ln L(D;\Theta) = {} & C_1 \ln S(D_1) + (1 - C_1)\ln f(D_1) + \sum_{i=2}^{N(T)-1} \ln f(D_i) \\ & + C_{N(T)} \ln S(D_{N(T)}) + (1 - C_{N(T)}) \ln f(D_{N(T)}),\end{aligned} \tag{40}$$

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).

4.3 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

$$V(\alpha) = \frac{|V_1(\alpha)| + |V_2(\alpha)|}{2}, \tag{41}$$

where

$$V_1(\alpha) = \frac{1}{T_1} \sum_{t \in \kappa(\alpha)} \delta_t(\alpha), \quad V_2(\alpha) = \frac{1}{T_2} \sum_{t \in \tau(\alpha)} \delta_t(\alpha), \tag{42}$$

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 $V(\alpha)$.

5 Empirical Studies

We apply the RSV model to daily (close-to-close) returns and RKs of DJIA index obtained from Oxford-Man Institute. The sample contains 2884 trading days from January 4, 2000 through July 29, 2011. Figure 1 shows the time-series plot of the daily returns and logarithms of the RKs.

Table 1 shows the descriptive statistics of the daily returns (r) and logarithms of RKs ($\ln RK$). The mean of r is not statistically significant from zero and its Ljung-Box (LB) statistic does not reject the null hypothesis of no autocorrelation up to 10 lags, which allows us to estimate the RSV models using the daily returns without adjustment of mean and autocorrelation. The kurtosis of r shows that its distribution is leptokurtic as commonly observed in the financial returns and the Jacque-Bera (JB) statistic rejects its normality whereas its skewness is not statistically significant from zero. In the RSVskt model, the leptokurtosis of r_t may be explained by stochastic volatility but the distribution of $\beta(z_t -$

$\mu_z) + \sqrt{z_t}\epsilon_t$ may also be leptokurtic. Additionally, the LB statistic of $\ln RK$ rejects the null of no autocorrelation, which is consistent with the high persistence of volatility known as the volatility clustering. The skewness of $\ln RK$ is significantly positive and its kurtosis shows that the distribution of $\ln RK$ is leptokurtic. Consequently, the JB statistic rejects the normality of $\ln RK$. This contradicts the normality assumption for u_t and η_t in (21) but we stick to the normality assumption in this paper and leave alternative specifications for future research.

In the following sections, we present the estimation and prediction results. In Section 5.1, we show the estimation results of the RSV models using all samples and compare the models by the marginal likelihood. In Section 5.2, we show the results of the volatility and quantile forecasts obtained by the rolling estimation.

5.1 Estimation Results

We estimate the RSV models with the priors for the parameters as follows,

$$\mu \sim N(0, 10), \quad \beta \sim N(0, 1), \quad \nu \sim Gamma(5, 0.5)I(\nu > 4), \tag{43}$$

$$\xi \sim N(0, 1), \quad \sigma_u^{-2} \sim Gamma(2.5, 0.1), \tag{44}$$

$$\frac{\phi + 1}{2} \sim Beta(20, 1.5), \quad \sigma_\eta^{-2} \sim Gamma(2.5, 0.025), \quad \frac{\rho + 1}{2} \sim Beta(1, 2). \tag{45}$$

Table 2 summarizes the MCMC estimation results of the RSV models with normal, Student's t, and skew t distributions obtained by 20,000 samples recorded after discarding 5,000 samples from MCMC iterations.[12] CD is the p-value of the convergence diagnostic test by Geweke (1992). All values indicate that the convergence of the posterior samples is not rejected at 5% level. The inefficiency factor measures how well the MCMC chain mixes.[13] Its values show that the chain is reasonably efficient and the 20,000 posterior samples are large enough to give a statistical inference.

The parameters in the latent volatility equation (19) 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

[12] All calculations in this paper are done by using Ox of Doornik (2009).

[13] The inefficiency factor is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$, where ρ_s is the sample autocorrelation at lag s. It is the ratio of the numerical variance of the posterior sample mean to the variance of the posterior sample mean from uncorrelated draws. The inverse of the inefficiency factor is also known as relative numerical efficiency (See, for example, Chib (2001)). When the inefficiency factor is equal to x, we need to draw MCMC samples x times as many as uncorrelated samples to obtain the same accuracy.

negative and the 95% credible interval does not contain zero for all models, which confirms the volatility asymmetry. The posterior mean of μ is similar among models.

The posterior mean of β for the RSVskt model is negative and the 95% credible interval does not contain zero, which seems to contradict the insignificant skewness of the daily returns shown in Table 1. We attribute this seemingly contradictory result to the difference between unconditional and conditional distribution of the daily returns. In Figure 1, large negative returns followed by large positive returns are observed in some periods such as the Lehman crisis in 2008, which results in the (almost) symmetric unconditional distribution of the returns, that is, the insignificant skewness. On the other hand, the initial negative return (r_t) may not be fully explained by the stochastic volatility component (h_t) and the remaining part may be explained by the return shock component ($\beta(z_t - \mu_z) + \sqrt{z}_t\epsilon_t$). The negative return increases the subsequent volatilities (h_{t+1}, $h_{t+2}, \ldots$), which may explain the most part of the subsequent large positive returns. As a result, the large positive returns are mostly explained by the stochastic volatility component whereas the large negative returns are largely explained by the return shock component. Consequently, the conditional distribution of returns becomes negatively skewed, which is captured by the negative value of β.

The posterior mean of ν is around 23 for the RSVt and RSVskt models, which implies that the fat tail is mostly explained by the stochastic volatility component. The large value of ν is not consistent with the previous studies. For example, Nakajima and Omori (2012) estimate the SV model with the GH skew Student's t-distribution and report that the posterior mean of ν is around 13 for the S&P500 returns from January 1970 to December 2003. We attribute such a difference to the persistence of the return shock in the data. The data in Nakajima and Omori (2012) contains a quite large but temporal shock, Black Monday shock in 1987, whereas our dataset contains the Lehman crisis, which persists for relatively longer period as shown in Figure 1. As a result, the temporal shock in the former data is explained by the small value of ν while the persistent shock in the latter is explained by the stochastic volatility component.

The parameters in the RV equation (18) are qualitatively similar among models. 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.

For model comparisons, we compute the marginal likelihoods of the RSV models.[14]

[14]See Appendix B for a brief description of the procedure to calculate the marginal likelihood.

Table 3 shows the marginal likelihood estimates. The RSVskt model provides the highest marginal likelihood whereas the RSVt model does not improve the model fit compared to the RSVn model. This is consistent with the negative estimate of β for the RSVskt model and larger values of ν for the RSVt and RSVskt models. Overall, the results show that the RSVskt model improves the model fit and the importance of the negative skewness in the return distribution.

5.2 Prediction Results

We implement the rolling estimation using 1,989 samples each time in the following way. First, we estimate the RSV models using 1,989 samples from January 4, 2000 to December 31, 2007 and compute one-day-ahead (January 2, 2008) forecasts of volatility, VaR, and ES from 15,000 posterior samples recorded after discarding 5,000 burn-in samples. Second, we implement the estimation using 1,989 samples from January 5, 2000 to January 2, 2008 and compute the one-day-ahead (January 3, 2008) forecasts. We continue the rolling estimation until we obtain the one-day-ahead forecasts on July 29, 2011 with the 1,989 estimation samples from August 20, 2003 to July 28, 2011. Finally, we obtain 895 prediction samples from January 2, 2008 to July 29, 2011.[15]

Table 4 shows the MSE and QLIKE of the volatility forecasts with the RK as a proxy of the latent volatility. Following Hansen and Lunde (2005), we adjust the effect of the non-trading hours on the RK as in (10). The volatility forecasts and the adjusted RKs are shown in Figure 2. The RSVskt model performs better than the RSVn and RSVt models, which implies that the extended model improves the volatility forecasts.

Tables 5 and 6 show the empirical failure rates (both conditional and unconditional) and the finite sample p-values of the Kupiec (1995)'s likelihood ratio test, Markov, Weibull, and EACD tests for the VaR forecasts, respectively.[16] The empirical failure rates are higher than the null probabilities in most cases due to the VaR violations in a volatile period from 2008 through 2009 as depicted in Figure 3. The failure rates of the RSVskt model are slightly closer to the null probabilities than those of the RSVn and RSVt models. Consequently, the

[15]From the second estimation, we use the posterior means obtained from the previous period as the initial values and generate 15,000 posterior samples after discarding 1,500 burn-in samples.

[16]Note that Kupiec (1995)'s LR test implicitly assumes the independent VaR violations whereas Christoffersen (1998)'s Markov test considers the very restrictive first order Markov process as the alternative. These assumptions are not usually satisfied in practice and the corresponding tests are not suited for the model comparison. We show the test results only for reference here. We thank the editor, Esther Ruiz, for pointing out this concern.

RSVskt model shows marginally better p-vlaues for most cases in Table 6. This result shows that the the extension of the return distribution marginally improves the VaR forecasts.

Table 7 shows the backtesting measures of the ES forecasts proposed by Embrechts, Kaufmann, and Patie (2005). The RSVskt model shows the best performance, followed by the RSVt model, for all null probabilities $\alpha \in \{0.5\%, 1\%, 5\%, 10\%\}$. This indicates the importance of the fat tail and skewness in the return distribution. That is, the extended model also improves the ES forecasts.

6 Conclusion

The RSV model of Takahashi, Omori, and Watanabe (2009), which incorporates the asymmetric SV model with the RV estimator, is extended with the GH skew Student's t-distribution for financial returns. The extension makes it possible 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 the DJIA index. The estimation results show that the extended model improves the model fit evaluated by the marginal likelihood. Moreover, the prediction results show that the extended model improves both volatility and quantile forecasts.

The RSV model can be extended further in several directions. Recently, Trojan (2013) proposes a regime switching RSVskt model and confirms several regimes in the S&P 500 index data. We can also consider different types of skew Student's t-distribution such as Fernández and Steel (1998) and Azzalini and Capitanio (2003). Additionally, extending the univariate RSV model to the multivariate model enables the portfolio risk management and optimal portfolio selection. Moreover, modeling multiple RV estimators 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 estimators, and show that the model with multiple RV estimators dominates the one with a single RV estimator. We leave these extensions for future research.

Appendices

A MCMC Sampling Procedure

Consider the RSVskt model in (17)-(21). Let $\theta = (\phi, \sigma_\eta, \rho, \mu, \beta, \nu, \xi, \sigma_u)$, $y = \{r_t, x_t\}_{t=1}^n$, $h = \{h_t\}_{t=1}^n$, $z = \{z_t\}_{t=1}^n$, 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

$$
\begin{aligned}
f(\Theta|y) &\propto f(r|\Theta) \times f(x|\theta, h) \times f(h|z,\theta) \times f(z|\theta) \times f(\theta) && (46)\\
&= \prod_{t=1}^{n-1} f(r_t|\theta, h_t, h_{t+1}, z_t) f(r_n|\theta, h_n) \times \prod_{t=1}^{n} f(x_t|\theta, h_t) && (47)\\
&\quad \times f(h_1|\theta) \prod_{t=1}^{n-1} f(h_{t+1}|\theta, h_t) \times \prod_{t=1}^{n} f(z_t|\theta) \times f(\theta) && (48)\\
&= \prod_{t=1}^{n-1} f(h_{t+1}|\theta, h_t, z_t, r_t) \times f(h_1|\theta) \times \prod_{t=1}^{n} f(r_t|\theta, h_t, z_t) \times \prod_{t=1}^{n} f(x_t|\theta, h_t) && (49)\\
&\quad \times \prod_{t=1}^{n} f(z_t|\theta) \times f(\theta) && (50)\\
&\propto (1-\rho^2)^{-(n-1)/2} \sigma_\eta^{-(n-1)} \prod_{t=1}^{n-1} \exp\left\{ -\frac{(\bar{h}_{t+1} - \phi \bar{h}_t - \bar{r}_t)^2}{2(1-\rho^2)\sigma_\eta^2} \right\} && (51)\\
&\quad \times (1-\phi^2)^{1/2} \sigma_\eta^{-1} \exp\left\{ -\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma_\eta^2} \right\} && (52)\\
&\quad \times (\beta^2\sigma_z^2 + \mu_z)^{n/2} \prod_{t=1}^{n} z_t^{-1/2} \exp\left(-\frac{h_t}{2} - \frac{\tilde{r}_t^2}{2} \right) && (53)\\
&\quad \times \sigma_u^{-n} \prod_{t=1}^{n} \exp\left\{ -\frac{(x_t - \xi - h_t)^2}{2\sigma_u^2} \right\} && (54)\\
&\quad \times \left(\frac{\nu}{2}\right)^{n\nu/2} \Gamma\left(\frac{\nu}{2}\right)^{-n} \prod_{t=1}^{n} z_t^{-\nu/2+1} \exp\left(-\frac{\nu}{2z_t} \right) \times f(\theta), && (55)
\end{aligned}
$$

where

$$
\tilde{r}_t = \frac{\sqrt{\beta^2\sigma_z^2 + \mu_z}\, r_t \exp(-h_t/2) - \beta \bar{z}_t}{\sqrt{z_t}}, \quad \bar{h}_t = h_t - \mu, \quad \bar{r}_t = \rho\sigma_\eta \tilde{r}_t, \quad \bar{z}_t = z_t - \mu_z \qquad (56)
$$

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 $\theta_3 = (\xi, \sigma_u)$. We describe how to sample θ_1, θ_2, θ_3, z, and h in the following subsections.

A.1 Generation of θ_1

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

$$f(\theta_1|\theta_2, h, z, y) \propto (1-\rho^2)^{-(n-1)/2}\sigma_\eta^{-(n-1)} \prod_{t=1}^{n-1} \exp\left\{-\frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \bar{r}_t)^2}{2(1-\rho^2)\sigma_\eta^2}\right\} \tag{57}$$

$$\times (1-\phi^2)^{1/2}\sigma_\eta^{-1} \exp\left\{-\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma_\eta^2}\right\} \times f(\theta_1), \tag{58}$$

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 (56).

A.2 Generation of θ_2

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

$$f(\theta_2|\theta_1, h, z, y) \propto \prod_{t=1}^{n-1} \exp\left\{-\frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \bar{r}_t)^2}{2(1-\rho^2)\sigma_\eta^2}\right\} \times (\beta^2\sigma_z^2 + \mu_z)^{n/2} \prod_{t=1}^{n} \exp\left(-\frac{\tilde{r}_t^2}{2}\right) \tag{59}$$

$$\times \left(\frac{\nu}{2}\right)^{n\nu/2} \Gamma\left(\frac{\nu}{2}\right)^{-n} \prod_{t+1}^{n} z_t^{-\nu/2+1} \exp\left(-\frac{\nu}{2z_t}\right) \times f(\theta_2). \tag{60}$$

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

$$f(\theta_3|\theta_1, \theta_2, h, z, y) \propto \sigma_u^{-n} \prod_{t=1}^{n} \exp\left\{-\frac{(x_t - \xi - h_t)^2}{2\sigma_u^2}\right\} \times f(\theta_3). \tag{61}$$

Let the prior distributions of parameters in θ_3 be

$$\xi \sim N(m_\xi, s_\xi^2), \quad \sigma_u^{-2} \sim Gamma(n_u, S_u). \tag{62}$$

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

$$\xi|\sigma_u, h, y \sim N(\tilde{m}_\xi, \tilde{s}_\xi^2), \tag{63}$$

$$\sigma_u^2|\xi, h, y \sim Gamma(\tilde{n}_u, \tilde{S}_u), \tag{64}$$

where

$$\tilde{m}_\xi = \frac{s_\xi^2 \sum_{t=1}^n (x_t - h_t) + \sigma_u^2 m_\xi}{n s_\xi^2 + \sigma_u^2}, \quad \tilde{s}_\xi^2 = \frac{\sigma_u^2 s_\xi^2}{n s_\xi^2 + \sigma_u^2}, \tag{65}$$

$$\tilde{n}_u = \frac{n}{2} + n_u, \quad \tilde{S}_u = \frac{1}{2}\sum_{t=1}^n (x_t - \xi - h_t)^2 + S_u. \tag{66}$$

A.4 Generation of z

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

$$f(z_t|\theta_1, \theta_2, \theta_3, h, y) \propto g(z_t) \times z_t^{-\frac{\nu+1}{2}+1} \exp\left(-\frac{\nu}{2z_t}\right), \tag{67}$$

where

$$g(z_t) = \exp\left\{-\frac{\tilde{r}_t^2}{2} - \frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \bar{r}_t)^2}{2(1-\rho^2)\sigma_\eta^2} I(t<n)\right\}, \tag{68}$$

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 RSVskt model in (17)-(19) as

$$r_t = \{\beta(z_t - \mu_z) + \sqrt{z_t}\epsilon_t\}\exp(\alpha_t/2)\gamma, \quad t = 1, \ldots, n \tag{69}$$

$$x_t = c + \alpha_t + u_t, \quad t = 1, \ldots, n \tag{70}$$

$$\alpha_{t+1} = \phi\alpha_t + \eta_t, \quad t = 0, \ldots, n-1 \tag{71}$$

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

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,

$$f(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) \propto \prod_{t=s}^{s+m} f(y_t|\alpha_t, \alpha_{t+1}) \prod_{t=s}^{s+m-1} f(\eta_t), \tag{72}$$

for $s+m<n$, and

$$f(\eta_s,\dots,\eta_{s+m-1}|\alpha_s,y_s,\dots,y_{s+m}) \propto \prod_{t=s}^{s+m-1} f(y_t|\alpha_t,\alpha_{t+1}) f(y_n|\alpha_n) \prod_{t=s}^{s+m-1} f(\eta_t), \tag{73}$$

for $s+m=n$. The logarithm of $f(y_t|\alpha_t,\alpha_{t+1})$ or $f(y_n|\alpha_n)$ in (72) and (73) (excluding constant term) is given by

$$l_t = -\frac{\alpha_t}{2} - \frac{(r_t-\mu_t)^2}{2\sigma_t^2} - \frac{(x_t-c-\alpha_t)^2}{2\sigma_u^2}, \tag{74}$$

where

$$\mu_t = \begin{cases} \{\beta\bar{z}_t + \sqrt{z_t}\rho\sigma_\eta^{-1}(\alpha_{t+1}-\phi\alpha_t)\}\exp(\alpha_t/2)\gamma, & t<n, \\ \beta\bar{z}_n\exp(\alpha_n/2)\gamma, & t=n, \end{cases} \tag{75}$$

and

$$\sigma_t^2 = \begin{cases} (1-\rho^2)z_t\exp(\alpha_t)\gamma^2, & t<n, \\ z_n\exp(\alpha_n)\gamma^2, & t=n. \end{cases} \tag{76}$$

Then the logarithm of the conditional density in (72) and (73) is given by (excluding a constant term)

$$\sum_{t=s}^{s+m-1} \log f(\eta_t) + L, \tag{77}$$

where

$$L = \begin{cases} \sum_{t=s}^{s+m} l_t - \dfrac{(\alpha_{s+m+1}-\phi\alpha_{s+m})^2}{2\sigma_\eta^2}, & s+m<n, \\ \sum_{t=s}^{s+m} l_t, & s+m=n. \end{cases} \tag{78}$$

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

$$\alpha = (\alpha_{s+1},\dots,\alpha_{s+m})', \tag{79}$$

$$d = (d_{s+1},\dots,d_{s+m})', \quad d_t = \frac{\partial L}{\partial \alpha_t}, \quad t=s+1,\dots,s+m, \tag{80}$$

$$Q = -E\left[\frac{\partial^2 L}{\partial\alpha\partial\alpha'}\right] = \begin{bmatrix} A_{s+1} & B_{s+2} & 0 & \cdots & 0 \\ B_{s+2} & A_{s+2} & B_{s+3} & \cdots & 0 \\ 0 & B_{s+3} & A_{s+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B_{s+m} \\ 0 & \cdots & 0 & B_{s+m} & A_{s+m} \end{bmatrix}, \tag{81}$$

$$A_t = -E\left[\frac{\partial^2 L}{\partial\alpha_t^2}\right], \quad t=s+1,\dots,s+m, \tag{82}$$

$$B_t = -E\left[\frac{\partial^2 L}{\partial\alpha_t\partial\alpha_{t-1}}\right], \quad t=s+2,\dots,s+m, \quad B_{s+1}=0. \tag{83}$$

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

$$d_t = -\frac{1}{2} + \frac{(r_t - \mu_t)^2}{2\sigma_t^2} + \frac{r_t - \mu_t}{\sigma_t^2}\frac{\partial \mu_t}{\partial \alpha_t} + \frac{r_{t-1} - \mu_{t-1}}{\sigma_{t-1}^2}\frac{\partial \mu_{t-1}}{\partial \alpha_t} + \frac{(x_t - c - \alpha_t)}{\sigma_u^2} + \kappa(\alpha_t), \quad (84)$$

where

$$\frac{\partial \mu_t}{\partial \alpha_t} = \begin{cases} \left\{ \frac{\beta \bar{z}_t}{2} + \sqrt{z_t}\rho\sigma_\eta^{-1}\left(-\phi + \frac{\alpha_{t+1} - \phi\alpha_t}{2}\right)\right\} \exp(\alpha_t/2)\gamma, & t < n, \\ \frac{\beta \bar{z}_n \exp(\alpha_n/2)\gamma}{2}, & t = n, \end{cases} \quad (85)$$

$$\frac{\partial \mu_{t-1}}{\partial \alpha_t} = \begin{cases} 0, & t = 1, \\ \sqrt{z_{t-1}}\rho\sigma_\eta^{-1}\exp(\alpha_{t-1}/2)\gamma, & t = 2, \ldots, T. \end{cases} \quad (86)$$

$$\kappa(\alpha_t) = \begin{cases} \frac{\phi(\alpha_{t+1} - \phi\alpha_t)}{\sigma_\eta^2}, & t = s + m < n, \\ 0, & \text{otherwise.} \end{cases} \quad (87)$$

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,

$$A_t = \frac{1}{2} + \sigma_t^{-2}\left(\frac{\partial \mu_t}{\partial \alpha_t}\right)^2 + \sigma_{t-1}^{-2}\left(\frac{\partial \mu_{t-1}}{\partial \alpha_t}\right)^2 + \frac{1}{\sigma_u^2} + \kappa'(\alpha_t), \quad (88)$$

$$B_t = \sigma_{t-1}^{-2}\frac{\partial \mu_{t-1}}{\partial \alpha_{t-1}}\frac{\partial \mu_{t-1}}{\partial \alpha_t}, \quad (89)$$

where

$$\kappa'(\alpha_t) = \begin{cases} \frac{\phi^2}{\sigma_\eta^2}, & t = s + m < n, \\ 0, & \text{otherwise.} \end{cases} \quad (90)$$

Applying the second order Taylor expansion to the conditional density (72) 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),

$$\log f(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) \quad (91)$$

$$\approx \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t^2 + \hat{L} + \left.\frac{\partial L}{\partial \eta}\right|_{\eta=\hat{\eta}}(\eta - \hat{\eta}) + \frac{1}{2}(\eta - \hat{\eta})' \left. E\left[\frac{\partial^2 L}{\partial \eta \partial \eta'}\right]\right|_{\eta=\hat{\eta}}(\eta - \hat{\eta}) \quad (92)$$

$$= \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t^2 + \hat{L} + \hat{d}'(\alpha - \hat{\alpha}) - \frac{1}{2}(\alpha - \hat{\alpha})'\hat{Q}(\alpha - \hat{\alpha}) \quad (93)$$

$$= \text{const} + \log f^*(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}), \quad (94)$$

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 (73).

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,

$$D_t = \hat{A}_t - D_{t-1}^{-1}\hat{B}_t^2, \quad D_{s+1} = \hat{A}_{s+1}, \tag{95}$$

$$K_t = \sqrt{D_t}, \tag{96}$$

$$J_t = \hat{B}_t K_{t-1}^{-1}, \quad J_{s+1} = 0, \quad J_{s+m+1} = 0, \tag{97}$$

$$b_t = \hat{d}_t - J_t K_{t-1}^{-1} b_{t-1}, \quad b_{s+1} = \hat{d}_{s+1}. \tag{98}$$

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

$$\hat{\gamma}_t = \hat{\alpha}_t + K_t^{-1} J_{t+1} \hat{\alpha}_{t+1}, \quad t = s+1, \ldots, s+m. \tag{99}$$

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

$$\hat{y}_t = Z_t \alpha_t + G_t \zeta_t, \quad t = s+1, \ldots, s+m, \tag{100}$$

$$\alpha_{t+1} = \phi \alpha_t + H_t \zeta_t, \quad t = s, s+1, \ldots, s+m, \quad \zeta_t \sim \mathrm{N}(0, I), \tag{101}$$

where

$$Z_t = 1 + K_t^{-1} J_{t+1} \phi, \quad G_t = K_t^{-1}(1,\, J_{t+1}\sigma_\eta), \quad H_t = (0,\, \sigma_\eta). \tag{102}$$

We can sample $(\eta_s, \ldots, \eta_{s+m-1})$ from the full posterior distribution in (72) and (73) by applying the simulation smoother[17] 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 details of the ARMH algorithm.

B Marginal Likelihood

The marginal likelihood $m(y)$ is defined as the integral of the likelihood with respect to the prior density of the parameter,

$$m(y) = \int_\Theta f(y|\Theta) f(\Theta) d\Theta = \frac{f(y|\Theta) f(\Theta)}{f(\Theta|y)}, \tag{103}$$

[17] See, for example, de Jong and Shephard (1995) and Durbin and Koopman (2002).

where Θ is a parameter, $f(y|\Theta)$ is a likelihood, $f(\Theta)$ is a prior probability density, and $f(\Theta|y)$ is a posterior probability density.[18] Following Chib (1995), we estimate the logarithm of the marginal likelihood as

$$\log m(y) = \log f(y|\Theta) + \log f(\Theta) - \log f(\Theta|y). \tag{104}$$

Although the equality holds for any values of Θ, we use the posterior mean of Θ to obtain a stable estimate of $m(y)$.

Given the posterior sample of Θ, the prior density $f(\Theta)$ is easily calculated. However, the likelihood and posterior components must be evaluated by simulation. The likelihood is estimated using the auxiliary particle filter with 10,000 particles, which provides an unbiassed estimator at a particular ordinate Θ for $f(y|\Theta)$.[19] 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 50,000 reduced MCMC samples.

References

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

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005): "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, 18(2), 351–416.

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):

[18]The last equality is obtained by the Bayes' rule,

$$f(\Theta|y) = \frac{f(y|\Theta)f(\Theta)}{\int_\Theta f(y|\Theta)f(\Theta)d\Theta}.$$

[19]See, for example, Pitt and Shephard (1999) and Omori, Chib, Shephard, and Nakajima (2007) for the details.

"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.

Azzalini, A., and A. Capitanio (2003): "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society B*, 65(2), 367–389.

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.

——— (2009): "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12(3), C1–C32.

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

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

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *The Eonometrics of Financial Markets*. Princeton University Press, Princeton.

Chib, S. (1995): "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90(432), 1313–1321.

——— (2001): "Markov chain Monte Carlo methods: Computation and inference," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, pp. 3569–3649. Elsevier.

Chib, S., and E. Greenberg (1995): "Understanding the Metropolis-Hastings algorithm," *American Statistician*, 49(4), 327–335.

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

Christie, A. A. (1982): "The stochastic behavior of common stock variances: Value, leverage, and interest rate effects," *Journal of Financial Economics*, 10(4), 407–432.

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

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

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," FRB Working Paper 2010-45.

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.

Fernández, C., and M. F. J. Steel (1998): "On Bayesian modeling of fat tails and skewness," *Journal of the American Statistical Association*, 93(441), 359–371.

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 realized 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*, 27(6), 877–906.

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.

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.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007): "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140(2), 425–449.

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 dissertation, 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): "Markov chains for exploring posterior distributions," *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

Table 1: Descriptive statistics of the daily return (r) and logarithm of the realized kernel ($\ln RK$) for DJIA from January 4, 2000 to July 29, 2011 (2884 samples). The data is obtained from Oxford-Man Institute. Standard errors of skewness and kurtosis are 0.0456 and 0.0911, respectively. JB is the p-value of the Jaque-Bera statistic to test the null hypothesis of normality. 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.

Variable	Mean	SE	SD	Skew	Kurt	Min	Max	JB	LB
r	0.002	0.024	1.264	0.008	10.414	-8.615	10.532	0.00	0.14
$\ln RK$	-0.361	0.018	0.982	0.640	3.860	-2.958	4.514	0.00	0.00

Table 2: MCMC estimation results of RSV models with normal, student's t, and skewed t distributions obtained by 20,000 samples recorded after discarding 5,000 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 by Chib (2001). Priors are set as $\mu \sim N(0, 10)$, $(\phi + 1)/2 \sim Beta(20, 1.5)$, $\sigma_\eta^{-2} \sim Gamma(2.5, 0.025)$, $(\rho + 1)/2 \sim Beta(1, 2)$, $\beta \sim N(0, 1)$, $\nu \sim Gamma(5, 0.5)I(\nu > 4)$, $\xi \sim N(0, 1)$, $\sigma_u^{-2} \sim Gamma(2.5, 0.1)$.

Model		Mean	SD	95%L	Median	95%U	CD	Inef.
RSVn	ϕ	0.9650	0.0044	0.9560	0.9652	0.9736	0.480	4.84
	σ_η	0.2186	0.0079	0.2036	0.2183	0.2345	0.331	16.50
	ρ	-0.4825	0.0300	-0.5401	-0.4832	-0.4218	0.731	10.45
	μ	-0.0764	0.1100	-0.2908	-0.0778	0.1441	0.826	4.46
	ξ	-0.2032	0.0327	-0.2688	-0.2028	-0.1404	0.489	46.68
	σ_u	0.3964	0.0077	0.3817	0.3962	0.4120	0.812	5.20
RSVt	ϕ	0.9655	0.0045	0.9565	0.9656	0.9740	0.725	3.43
	σ_η	0.2165	0.0082	0.2010	0.2163	0.2330	0.242	15.42
	ρ	-0.5010	0.0314	-0.5617	-0.5015	-0.4376	0.761	12.19
	μ	-0.0579	0.1087	-0.2668	-0.0589	0.1563	0.348	6.27
	ν	22.5909	4.5608	14.7718	22.2752	32.1559	0.705	173.51
	ξ	-0.2139	0.0329	-0.2798	-0.2140	-0.1500	0.066	48.66
	σ_u	0.3980	0.0076	0.3832	0.3979	0.4132	0.950	5.57
RSVskt	ϕ	0.9660	0.0043	0.9573	0.9660	0.9742	0.145	6.62
	σ_η	0.2181	0.0078	0.2031	0.2179	0.2339	0.370	25.92
	ρ	-0.5248	0.0306	-0.5831	-0.5253	-0.4637	0.057	20.33
	μ	-0.0668	0.1080	-0.2761	-0.0684	0.1474	0.480	3.32
	β	-0.6292	0.2101	-1.0870	-0.6147	-0.2691	0.429	87.91
	ν	23.0995	4.2464	15.8961	22.7920	32.4364	0.142	221.19
	ξ	-0.2160	0.0346	-0.2847	-0.2155	-0.1501	0.826	42.33
	σ_u	0.3980	0.0075	0.3838	0.3979	0.4131	0.057	6.63

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 10,000 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) with 50,000 reduced MCMC samples. The numbers in the parentheses show the standard errors.

Model	Likelihood		Prior	Posterior		Marginal	
RSVn	-5911.83	(0.69)	-1.18	20.30	(0.01)	-5933.30	(0.69)
RSVt	-5914.83	(0.64)	-6.69	15.99	(0.05)	-5937.50	(0.64)
RSVskt	-5906.16	(0.82)	-8.10	15.25	(0.05)	-5929.51	(0.82)

Table 4: Mean squared error (MSE) and quasi-likelihood (QLIKE) of volatility forecasts from January 2, 2008 to July 29, 2011 (895 samples). Realized kernel with the adjustment of Hansen and Lunde (2005) are used as a proxy of latent volatility.

Model	MSE	QLIKE
RSVn	9.0459	0.2121
RSVt	8.9489	0.2103
RSVskt	8.9106	0.2084

Table 5: Empirical failure rates for the VaR forecasts from January 2, 2008 to July 29, 2011 (895 samples). 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	0.0156	0.0159	0.0000	0.0268	0.0276	0.0000
RSVt	0.0089	0.0090	0.0000	0.0223	0.0229	0.0000
RSVskt	0.0067	0.0067	0.0000	0.0190	0.0194	0.0000
	5%			10%		
RSVn	0.0771	0.0775	0.0580	0.1162	0.1226	0.0577
RSVt	0.0782	0.0788	0.0571	0.1207	0.1271	0.0648
RSVskt	0.0737	0.0736	0.0606	0.1196	0.1269	0.0561

Table 6: Finite sample p-values of the Kupiec (1995)'s likelihood ratio test, Markov, Weibull, and EACD tests for the VaR forecasts from January 2, 2008 to July 29, 2011 (895 samples). We compute the finite sample p-values based on the Monte Carlo testing technique of Dufour (2006).

Model	Kupiec	Markov	Weibull	EACD	Kupiec	Markov	Weibull	EACD
	0.5%				1%			
RSVn	0.0112	0.0121	0.2794	0.6712	0.0003	0.0006	0.4407	0.9830
RSVt	0.1252	0.1462	0.8325	0.6735	0.0017	0.0047	0.1563	0.9109
RSVskt	0.6072	0.6163	0.4623	0.9237	0.0151	0.0298	0.1922	0.5754
	5%				10%			
RSVn	0.0011	0.0003	0.0644	0.0658	0.1059	0.0076	0.1425	0.0542
RSVt	0.0003	0.0003	0.0752	0.0838	0.0497	0.0043	0.1544	0.0239
RSVskt	0.0029	0.0012	0.0567	0.1014	0.0545	0.0022	0.1256	0.0289

Table 7: Backtesting measure of Embrechts, Kaufmann, and Patie (2005) for the expected shortfall forecasts from January 2, 2008 to July 29, 2011 (895 samples).

Model	0.5%	1%	5%	10%
RSVn	0.2688	0.2621	0.2721	0.2492
RSVt	0.1130	0.1743	0.2010	0.1948
RSVskt	0.0326	0.1437	0.1712	0.1454

Figures

Figure 1: Time series plot of close-to-close returns (top) and logarithms of realized kernels (bottom) for DJIA from January 4, 2000 though July 29, 2011 (2884 samples).

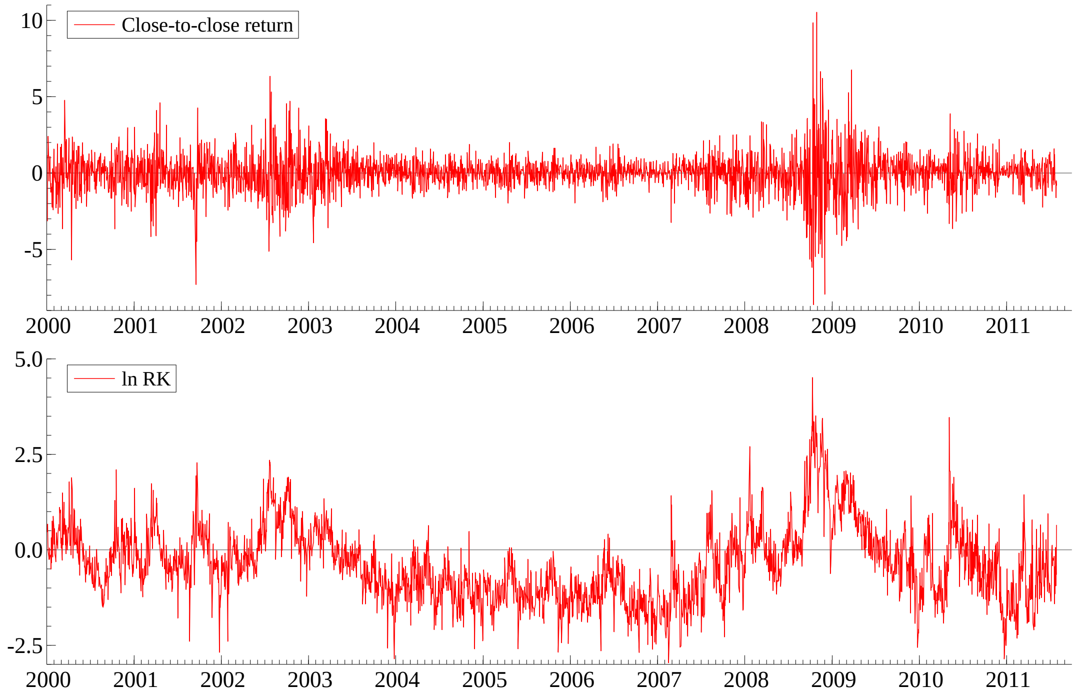
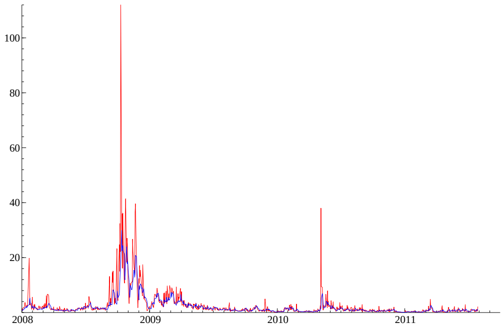
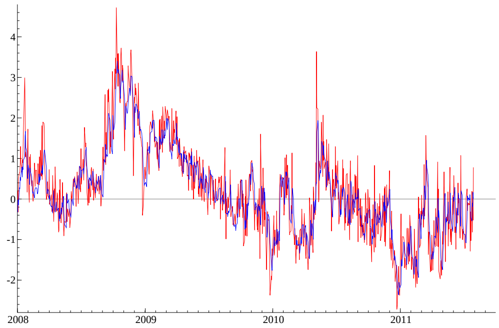


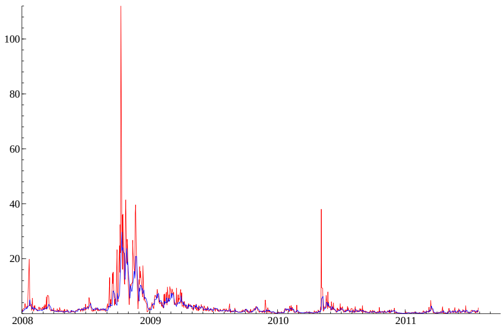
Figure 2: Volatility forecasts (blue line) and realized kernel with the adjustment of Hansen and Lunde (2005) (red line) from January 2, 2008 to July 29, 2011 (895 samples).



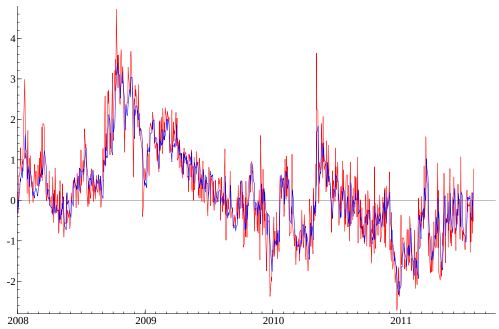
RSVn (level)



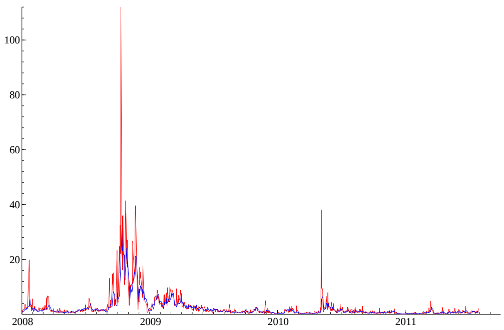
RSVn (log)



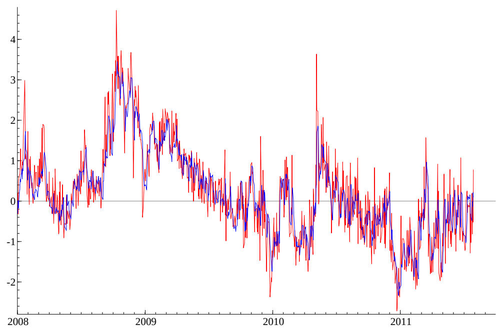
RSVt (level)



RSVt (log)

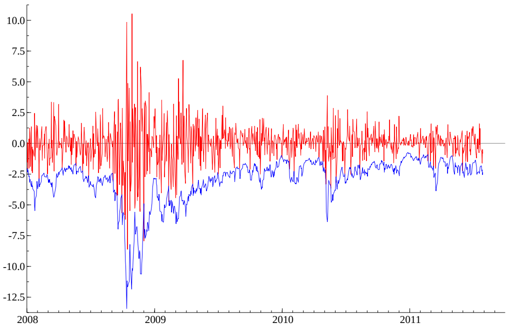


RSVskt (level)

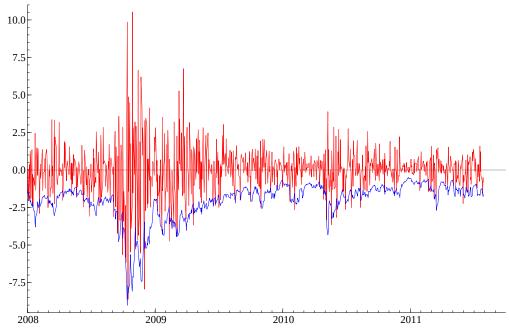


RSVskt (log)

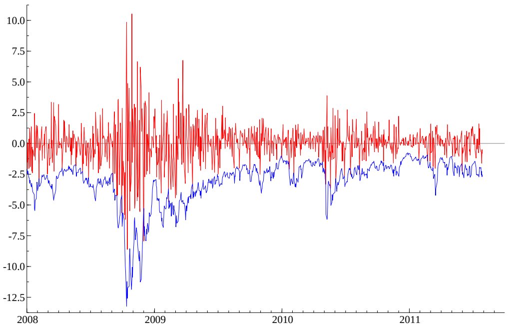
Figure 3: VaR forecasts (blue line) and daily returns (red line) from January 2, 2008 to July 29, 2011 (895 samples).



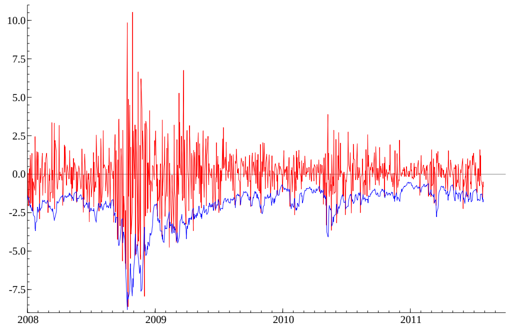
RSVn (1%)



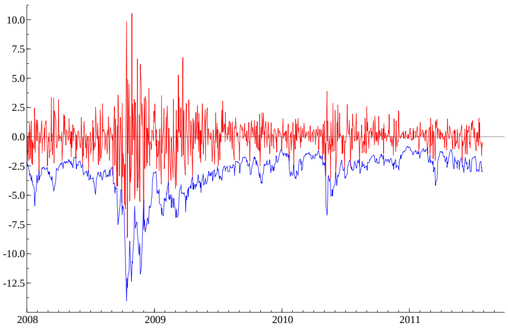
RSVn (5%)



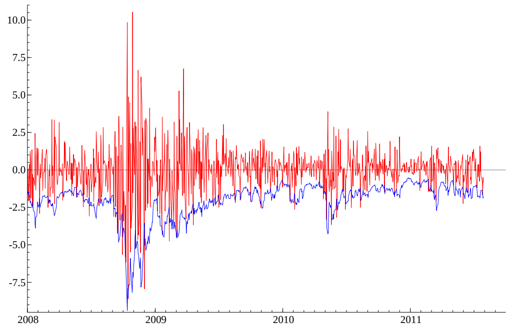
RSVt (1%)



RSVt (5%)



RSVskt (1%)



RSVskt (5%)