

# An Empirical Study of Quantile Forecasts for the S&P500 Stock Index

Chao LU

## Abstract

This paper estimates the one-day-ahead Value-at-Risk (VaR) and the one-day-ahead expected shortfall (ES) of daily returns for the Standard&Poor's 500 (S&P500) stock index. The predictive ability of the VaR forecasts and the ES forecasts are evaluated via the associate backtesting methods. Return volatility defined as the standard deviation or variance of daily return, which is essential to the calculation of VaR and ES, is estimated by applying the observations to the stochastic volatility (SV) model and the realized stochastic volatility (RSV) model.

Although many empirical studies show that with the introduction of high frequency intraday returns data, the RSV model performs better over the basic SV model, the difference in time spans should also be taken into consideration. To investigate the sensitivity to the choice of the time span of observations, this paper estimates the volatilities of a relatively long time period from the years 2000 to 2016, along with two short periods before and after the financial crisis of the year 2008 separately. Empirical evidence demonstrates that volatility forecasts generated from the RSV model outperform those from the SV model for the short time periods of the years 2000-2007 and 2009-2016, while the conclusion for the long time period of years 2000-2016 is the opposite. Therefore, with respect to the S&P500 stock index, the RSV model fits the data of shorter time spans better, while the SV is more capable for the data of longer time span.

## 1. Introduction

The risk measure of VaR is typically used as quantile forecast of financial returns in financial risk management. To obtain the VaR forecast, it is also required to predict the unobservable variable of volatility. In practice, parametric approaches such as autoregressive conditional heteroskedasticity (ARCH) models and SV models, and nonparametric estimators such as realized volatility (RV) and realized kernel (RK) are commonly applied to estimate the volatility.

The parametric approaches aim to establish certain models to describe the dynamics of the latent volatility. The ARCH-type models, including the underlying ARCH model proposed by Engle (1982), the GARCH (generalized ARCH) model proposed by Bollerslev (1986) and a variety of other extension models, consider volatility at date  $t$  to be a deterministic function of the known variable at date  $t-1$ . This assumption, however, is not in keeping with the situations of real markets. Alternatively, this paper applies the returns data to the SV model developed by Taylor (1986) where volatility is taken as an unknown parameter. It is notable that only the financial daily returns data are used either in an ARCH model or in a SV model, and thus some intraday information may be lost for such models.

On the other hand, nonparametric methods acquire the estimate of latent volatility based on the high frequency intraday returns data, among which RV proposed by Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2001) is widely used. This realized measure, however, usually turns out to be an inconsistent estimator of the latent volatility due to the influence of non-trading hours and market microstructure noise (MMN) that appear in real markets.

Takahashi, Omori, and Watanabe (2009) propose to model daily returns and RVs simultaneously to a RSV framework to account for the information loss of daily returns and the bias of RV caused by non-trading hours and MMN. This paper also estimates this RSV model to explore the model fitting.

The VaR can be computed easily given the volatility forecast and the returns distribution. However, VaR only focuses on the quantile and thus it is insensitive to the shape of the tail of the returns distribution. To cope with this problem, ES is also estimated to capture the further information of tail distribution. These VaR and ES forecasts are evaluated with their corresponding backtesting procedures, in addition to computing several loss functions on predictive performance of the volatility forecasts.

This paper is organized as follows. Section 2 briefly introduces the SV and RSV models for volatility forecasts and the risk measures for quantile forecasts. Section 3 illustrates the estimation approaches and the evaluation methods for the one-day-ahead VaR and ES forecasts. Empirical results are analyzed in Section 4. Conclusions are given in section 5.

## 2. Models and Quantile Forecasts

Return volatility of date  $t$  is defined as the integral of the instantaneous volatility  $\sigma^2(s)$  over the interval of  $(t, t+1)$ :

$$IV_t = \int_t^{t+1} \sigma^2(s) ds. \quad (1)$$

The SV model specified as follows is established as a parametric method to estimate latent variable in equation (1):

$$R_t = \exp(h_t / 2) \varepsilon_t, \quad (2)$$

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

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix}.$$

where  $R_t$  denotes the daily return and  $h_t$  denotes the logarithm of latent volatility. Suppose that  $(h_t - \mu)$  follows a stationary AR(1) process with  $|\phi| < 1$  and the initial value  $h_1$  follows a Gaussian distribution with an unconditional mean

and variance of  $h_t$ , that is,  $h_t \sim N(\mu, \sigma_\eta^2 / (1 - \phi^2))$ . This paper also assumes that  $\varepsilon_t$  and  $\eta_t$  follow a bivariate normal distribution.  $\rho$ , which represents the correlation coefficient of  $R_t$  and  $h_{t+1}$ , can explain the asymmetric phenomenon and is usually evaluated to be negative mainly due to the leverage effect and the volatility feedback effect.

On the other hand, RV defined as the sum of squared intraday returns over day  $t$ , is also widely used as a nonparametric estimator of latent volatility.

$$RV_t = \sum_{i=0}^{n-1} r_{t,i}^2, \quad (4)$$

where  $\{r_{t,i}\}_{i=0}^{n-1}$  represents the intraday returns data of date  $t$ . As stated in Section 1, this nonparametric estimator is usually inconsistent with latent volatility in real markets where non-trading hours and MMN exist.

Takahashi, Omori and Watanabe (2009) develop the following RSV model to incorporate the SV model with the RV:

$$R_t = \exp(h_t / 2) \varepsilon_t, \quad (5)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad (6)$$

$$x_t = \xi + h_t + u_t, \quad (7)$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \\ u_t \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho\sigma_\eta & 0 \\ \rho\sigma_\eta & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}.$$

where  $x_t$  denotes the logarithm of realized volatility of date  $t$ . Equations (5) and (6) are the basic SV model. The sign of  $\xi$  in equation (7) shows the dominant effect of MMN and non-trading hours. A positive estimate of  $\xi$  indicates an upward bias due to MMN, while a negative  $\xi$  suggests a downward bias due to non-trading hours.

Let  $\boldsymbol{\theta}$  denote the parameter vector  $(\mu, \phi, \rho, \sigma_\eta^2, \xi, \sigma_u^2)'$  and  $\mathbf{h}$  denote the vol-

atility vector  $(h_1, h_2, \dots, h_T)'$ . Given the observations  $\mathbf{R} = (R_1, R_2, \dots, R_T)'$  and  $\mathbf{X} = (x_1, x_2, \dots, x_T)'$ , where  $T$  is the sample size of the observations, samples of  $\boldsymbol{\theta}$  and  $\mathbf{h}$  can be drawn recursively from the joint conditional posterior density via MCMC-based Bayesian method.

Based on the estimation results, the one-day-ahead VaR forecast of date  $t+1$  given the information up to date  $t$  ( $\mathbf{I}_t$ ) with probability  $\alpha$  is defined as

$$Pr(R_{t+1} < VaR_{t+1}(\alpha) \mid \mathbf{I}_t) = \alpha. \quad (8)$$

In spite of its extensive utilization in the area of financial risk management, VaR is a yet imperfect measure for the reason that it simply discusses the quantile of the returns distribution and overlooks the information of the tail distribution. This problem is settled by developing another measure referred to as ES to account for the returns distribution that is beyond the quantile. The one-day-ahead ES forecast of  $R_{t+1}$  with probability  $\alpha$  is given by

$$ES_{t+1}(\alpha) = E \left[ R_{t+1} \mid R_{t+1} < VaR_{t+1}(\alpha), \mathbf{I}_t \right]. \quad (9)$$

### 3. Estimation and Evaluation

This section interprets the computational algorithms and evaluation methods for one-day-ahead VaR and ES forecasts. Additionally, two loss functions are also introduced to evaluate the predictive ability of the volatilities to make comparisons between the two SV models.

#### 3.1 Estimation of the VaR and ES forecasts

By applying the previous  $T_0$  observations to the RSV model recursively, the one-day-ahead VaR forecasts  $\{VaR_{T_0+i}\}_{i=1}^{T_f}$  and one-day-ahead ES forecasts  $\{ES_{T_0+i}\}_{i=1}^{T_f}$  can be calculated according to the following algorithm, where  $T_f = T - T_0$  represents the number of VaR or ES forecasts.

Step 1. Set  $i = 1$ .

Step 2. Generate  $\boldsymbol{\theta}^i = (\mu^i, \phi^i, \rho^i, \sigma_\eta^{i2}, \xi^i, \sigma_u^{i2})'$  and  $\mathbf{h}^i = (h_1^i, h_2^i, \dots, h_{T_0}^i)'$  from the

joint conditional posterior density  $\pi(\boldsymbol{\theta}^i, \mathbf{h}^i | \mathbf{R}, \mathbf{X})$ .

Step 3. Generate the sample of logarithm of one-day-ahead volatility from the posterior predictive distribution:

$$h_{T_0+1}^i | \cdot \sim N(\mu_{T_0+1}^i, \sigma_{T_0+1}^{i^2}),$$

where the conditional mean and variance are given by

$$\mu_{T_0+1}^i = \mu^i + \phi^i (h_{T_0}^i - \mu^i) + \rho^i \sigma_\eta^i R_t \exp(-h_{T_0}^i / 2), \quad \sigma_{T_0+1}^{i^2} = \sigma_\eta^{i^2} (1 - \rho^{i^2}).$$

Step 4. Compute  $V\alpha R_{T_0+i}(\alpha) = \exp(h_{T_0+1}^i / 2) z_i(\alpha)$ , where  $z_i(\alpha)$  is the  $\alpha$ -quantile of returns distribution.

Step 5. Randomly generate  $M$  samples of  $z_i$  from the returns distribution, where  $M$  is set to be a sufficiently large enough number, and compute the corresponding  $\{R_{T_0+i}^m\}_{m=1}^M$  using  $R_{T_0+i} = \exp(h_{T_0+1}^i / 2) z_i$ .

Step 6. Calculate  $ES_{T_0+i}(\alpha) = E[R_{T_0+i} | R_{T_0+i}^m < V\alpha R_{T_0+i}(\alpha), \mathbf{I}_{T_0+i-1}]$ , which is the average of the violated  $R_{T_0+i}$ .

Step 7. Set  $i = i + 1$ , repeat from Step 2 if  $i < T_f$  and end if  $i = T_f$ .

The Bayesian based MCMC method is used to generate the RSV model in Step 2 on account that the nonlinear specified measurement equation (5) will make it difficult to carry out the maximum likelihood estimation (MLE). The details of generation algorithm for RSV model are described in appendix A. This algorithm is also available for SV model simply by substituting the parameter vector  $(\mu, \phi, \rho, \sigma_\eta^2, \xi, \sigma_u^2)'$  to  $(\mu, \phi, \rho, \sigma_\eta^2)'$ .

### 3.2 Evaluation of the VaR and ES forecasts

This subsection explains the principles of several backtesting methods used to investigate the robustness of the predicted VaR and ES estimates.

Suppose that the number of the violated VaR forecasts among  $\{VaR_{T_0+i}(\alpha)\}_{i=1}^{T_f}$  is  $T_v$ . Then the empirical failure rate (EFR) can be written as  $EFR = T_v / T_f$ . This paper tests the null hypothesis of  $EFR = \alpha$  by calculating the following likelihood ratio (LR) statistic (Kupiec (1995)):

$$LR = 2 \left\{ \log \left[ EFR^{T_v} (1 - EFR)^{T_f - T_v} \right] - \log \left[ \alpha^{T_v} (1 - \alpha)^{T_f - T_v} \right] \right\}. \quad (10)$$

This statistic is asymptotically distributed as a  $\chi^2(1)$  under the condition that the null hypothesis is true.

The measure  $D(\alpha)$  proposed by Embrechts, Kaufmann, and Patie (2005) is computed to backtest the ES forecasts. Specifically, define  $\delta_t(\alpha) = R_t - ES_t(\alpha)$  and denote  $q(\alpha)$  as the empirical  $\alpha$ -quantile of  $\delta_t(\alpha)$ . Let  $\kappa_1(\alpha)$  and  $\kappa_2(\alpha)$  be the sets of time points for which  $\delta_t(\alpha) < 0$  occurs and  $\delta_t(\alpha) < q(\alpha)$  occurs, respectively. Suppose that the respective number of  $\kappa_1(\alpha)$  and  $\kappa_2(\alpha)$  are counted to be  $T_1$  and  $T_2$ , then the measure is written as

$$D(\alpha) = \frac{1}{2} \left( |D_1(\alpha)| + |D_2(\alpha)| \right), \quad (11)$$

where

$$D_1(\alpha) = \frac{1}{T_1} \sum_{t \in \kappa_1(\alpha)} \delta_t(\alpha), \quad D_2(\alpha) = \frac{1}{T_2} \sum_{t \in \kappa_2(\alpha)} \delta_t(\alpha). \quad (12)$$

This measure considers to average the absolute values of the standard backtesting measure  $D_1(\alpha)$  and the penalty term  $D_2(\alpha)$ . A lower value of  $D(\alpha)$  indicates better ES forecasts.

Moreover, two loss functions, which are the mean squared error (MSE) and the quasi-likelihood (QLIKE), defined as equations (13) and (14), respectively, are calculated to measure the precision of the volatility forecasts, of which values can provide further information about the predictive performance.

$$MSE = \frac{1}{T_f} \sum_{t=T_0}^{T-1} \left( \hat{\sigma}_{t+1|t}^2 - \sigma_{t+1}^2 \right)^2, \quad (13)$$

$$QLIKE = \frac{1}{T_f} \sum_{t=T_0}^{T-1} \left( \log \hat{\sigma}_{t+1|t}^2 + \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1|t}^2} \right)^2, \quad (14)$$

where  $\hat{\sigma}_{t+1|t}^2$  and  $\sigma_{t+1}^2$  denote one-day-ahead volatility forecast and volatility proxy of date  $t+1$ . As stated in Patton (2011), these two loss functions can lead to  $E[\hat{\sigma}_{t+1|t}^2 | \mathbf{I}_t] = \sigma_{t+1}^2$ , which is a necessary condition for a loss function to be robust to noise in the volatility proxy.

This paper uses the scaled realized kernel (SRK) proposed by Hansen and Lunde (2005) as the proxy of latent volatility. The SRK is specified as the

product of the RK and the ratio of the variance of daily return to the mean of RKs. This specification makes it possible to adjust the underestimation caused by non-trading hours.

$$SRK_t = cRK_t, \quad c = \frac{\sum_{t=1}^T (R_t - \bar{R})^2}{\sum_{t=1}^T RK_t}. \quad (15)$$

MMN, which is the primary reason that causes significant increase of bias in RV when the time interval is close to zero, is accounted by the estimator of RK proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008).

$$RK_t = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \gamma_h. \quad (16)$$

where  $H$  is the bandwidth of autocovariance lag,  $\gamma_h$  represents the autocovariance at lag  $h$  with high frequency data  $\{r_j\}$ :

$$\gamma_h = \sum_{j=|h|+1}^n r_j r_{j-|h|}, \quad (17)$$

and weight  $k(\cdot)$  represents the Parzen kernel function defined as:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 \leq x \leq 1/2 \\ 2(1-x)^3, & 1/2 \leq x \leq 1. \\ 0, & x > 1 \end{cases} \quad (18)$$

## 4. Empirical Results

In this section, daily returns and 5-minute RVs of S&P500, provided by the Oxford-Man Institute's Realized Library, are applied for empirical estimation. The sample period is from January 4, 2000 to December 30, 2016, and the sample size is 4247.

In order to investigate the model fitting with respect to different time spans, this paper also estimates the volatilities of the years 2000-2007 and 2009-2016, respectively. The year 2008 is excluded to avoid the contaminated time period of financial crisis. The computational results were generated by using the MATLAB\_R2015b software.



### 4.1 Descriptive Statistics

Table 1 shows the descriptive statistics of daily returns (in percentage) and the logarithm of RVs for the respective time spans.  $LB_p(m)$  represents the  $p$ -value of the Ljung-Box (LB) statistics, whose null hypothesis is no autocorrelation up to  $m$  lags. Following Diebold (1988), the  $LB(m)$  statistics of returns and logRVs modified for heteroscedasticity are described as

$$LB_{return}(m) = T(T + 2) \sum_{k=1}^m \left( \frac{\hat{\sigma}_R^4}{\hat{\sigma}_R^4 + \gamma_{R^2}(k)} \right) \frac{\hat{\rho}_R^2(k)}{T - k}, \tag{19}$$

$$LB_{RV}(m) = T(T + 2) \sum_{k=1}^m \left( \frac{\hat{\sigma}_{RV}^4}{\hat{\sigma}_{RV}^4 + \gamma_{RV^2}(k)} \right) \frac{\hat{\rho}_{RV}^2(k)}{T - k}, \tag{20}$$

where  $\hat{\sigma}_R^4$  and  $\hat{\sigma}_{RV}^4$  are the squared sample variance of daily returns and logRVs,  $\gamma_{R^2}(k)$  and  $\gamma_{RV^2}(k)$  are the  $k$ -th order autocovariance of sequence  $\{(R_t - \bar{R})^2\}_{t=1}^T$  and  $\{(\log RV_t - \overline{\log RV})^2\}_{t=1}^T$ ,  $\hat{\rho}_R^2(k)$  and  $\hat{\rho}_{RV}^2(k)$  are the squares of  $k$ -th order autocovariance of returns and logRVs. The statistic is asymptotic to a  $\chi^2(m)$  distribution on the condition that the null hypothesis is true.

Table 1. Descriptive Statistics

	Mean	Stdev	Skew	Kurt	Min	Max	$LB_p(10)$
<i>2000-2016</i>							
<i>return</i>	0.0090	1.1874	-0.1673	10.6668	-9.3511	10.2202	0.10
<i>logRV</i>	-0.5161	1.0835	0.4414	3.3805	-4.1221	4.3500	0.00
<i>2000-2007</i>							
<i>return</i>	-0.0030	1.1000	-0.0027	5.7102	-6.0443	5.6786	0.63
<i>logRV</i>	-0.4696	0.9383	0.2860	2.8427	-2.8711	2.6959	0.00
<i>2009-2016</i>							
<i>return</i>	0.0404	1.0255	-0.2831	7.7355	-6.7627	6.3971	0.23
<i>logRV</i>	-0.7209	1.0878	0.3752	3.0414	-4.1221	3.6196	0.00

NOTE: The sample periods are 2000-2016, 2000-2007 and 2009-2016 where the respective sample sizes are 4247, 1987 and 2014.  $LB_p(10)$  shows the  $p$  value of the Ljung-Box statistic up to 10 lags for the returns and realized measures where the heteroskedasticity is corrected following Diebold (1988).

The means of daily returns are not statistically significant from zero, and the  $LB(10)$  statistics do not reject the null hypothesis of no autocorrelation at the 10% significance level. As a result, there is no need to adjust the means and autocorrelations of the daily returns. In contrast, the null hypothesis that no autocorrelation exists at the 1% significance level are rejected for the cases of logRV, which indicate the high serial correlation of logarithm of volatilities.

## 4.2 Estimation Results

Tables 2-3 describe the posterior means, the standard deviations of the posterior means, the 95% Bayesian credible intervals, the  $p$  values of the convergence diagnostic ( $CD_p$ ) statistics, and the inefficiency factors (IF) of the SV and RSV models for the three investigated time periods, respectively. The results are based on  $N=5000$  samples generated from the respective posterior distributions of the parameters after discarding 5000 samples of the very beginning as the burn-in period.

Table 2. Estimation results of SV model

	<i>Mean</i>	<i>Stdev</i>	<i>95% interval</i>	$CD_p$	<i>IF</i>
<i>2000-2016</i>					
$\mu$	0.4812	0.0243	[0.4338, 0.5285]	0.79	1.30
$\phi$	0.9853	0.0004	[0.9845, 0.9859]	0.49	1.12
$\rho$	-0.5588	0.0113	[-0.5806, -0.5366]	0.35	1.46
$\sigma_\eta^2$	0.0562	0.0014	[0.0534, 0.0590]	0.44	4.32
<i>2000-2007</i>					
$\mu$	0.5632	0.0711	[0.4058, 0.6850]	0.71	2.29
$\phi$	0.9883	0.0017	[0.9839, 0.9905]	0.31	1.64
$\rho$	-0.5243	0.0280	[-0.5774, -0.4671]	0.48	10.54
$\sigma_\eta^2$	0.0347	0.0013	[0.0323, 0.0373]	0.15	1.49
<i>2009-2016</i>					
$\mu$	0.6323	0.0315	[0.5713, 0.6953]	0.34	1.12
$\phi$	0.9883	0.0002	[0.9879, 0.9886]	0.82	11.84
$\rho$	-0.6165	0.0161	[-0.6469, -0.5846]	0.87	9.56
$\sigma_\eta^2$	0.0479	0.0019	[0.0442, 0.0518]	0.91	5.97

The CD statistic specified by Geweke (1992) is computed to test whether the samples converge to an invariant distribution after the burn-in period. This statistic converges in distribution to the standard normal when the samples are stationary. The empirical results of the  $p$  values of the CD statistic demonstrate that during the sample period, the null hypothesis that the samples of the posterior distribution is converged after burn-in period is not rejected at the 10% significance level for all parameters.

The IF, which is defined as  $1+2\sum_{k=1}^{\infty}\rho(k)$ , quantifies the relative efficiency loss of the samples.  $\rho(k)$  is the sample autocorrelation at lag  $k$  corresponds to the ratio of the numerical variance of posterior sample mean to the variance

**Table 3.** Estimation results of RSV model

	<i>Mean</i>	<i>Stdev</i>	<i>95% interval</i>	<i>CD<sub>p</sub></i>	<i>IF</i>
<i>2000-2016</i>					
$\mu$	0.5039	0.0321	[0.4396, 0.5664]	0.99	2.34
$\phi$	0.9840	0.0009	[0.9821, 0.9855]	0.70	7.29
$\rho$	-0.5352	0.0173	[-0.5684, -0.4995]	0.63	7.13
$\xi$	-0.2906	0.0270	[-0.3423, -0.2374]	0.72	1.05
$\sigma_{\eta}^2$	0.0515	0.0025	[0.0472, 0.0570]	0.62	9.57
$\sigma_u^2$	0.4699	0.0194	[0.4312, 0.5074]	0.77	1.85
<i>2000-2007</i>					
$\mu$	0.4053	0.0318	[0.3440, 0.4691]	0.86	1.18
$\phi$	0.9897	0.0001	[0.9895, 0.9899]	0.59	1.34
$\rho$	-0.5742	0.0151	[-0.6040, -0.5448]	0.66	1.19
$\xi$	-0.3724	0.0335	[-0.4390, -0.3077]	0.42	1.13
$\sigma_{\eta}^2$	0.0279	0.0009	[0.0259, 0.0294]	0.84	1.29
$\sigma_u^2$	0.6031	0.0193	[0.5664, 0.6420]	0.78	1.00
<i>2009-2016</i>					
$\mu$	0.6656	0.0454	[0.5770, 0.7539]	0.28	4.07
$\phi$	0.9863	0.0006	[0.9851, 0.9873]	0.61	7.37
$\rho$	-0.6328	0.0209	[-0.6720, -0.5911]	0.78	9.76
$\xi$	-0.2720	0.0371	[-0.3463, -0.2000]	0.85	1.80
$\sigma_{\eta}^2$	0.0438	0.0018	[0.0405, 0.0475]	0.34	2.26
$\sigma_u^2$	0.4961	0.0235	[0.4501, 0.5438]	0.42	2.57

of the sample mean based on independent draws (Chib (2001)). For instance, the value of IF equals to  $M$  means that  $M$  times of the samples should be drawn in order to keep the equivalent precision to the independent samples. The low IFs (less than 30) of Tables 2-3 suggest that the generation method applied in this paper is quite efficient.

The posterior means of the persistence parameter  $\phi$  are close to 1 for all the cases, which indicates the existence of high persistence of latent volatility. The negative correlations between the return at date  $t$  and the volatility at date  $t+1$  are explained by the negative results of  $\rho$ . This common asymmetric phenomenon is relatively stronger in 2009-2016 than in 2000-2007.

The posterior means of  $\sigma_{\eta}^2$  are larger for the SV model than for the RSV model, which suggests that the RSV model can estimate the latent volatility with higher precision. In addition, the negative intervals of the bias-correction parameter  $\xi$  show underestimations of the 5-minute RV estimator major caused by MMN, and the relatively large  $\sigma_{\epsilon}^2$  indicate high levels of noises in RVs.

Comparisons of the logarithm of latent volatilities generated from the SV model (h-SV) and the RSV model (h-RSV) with the logarithm of RVs (logRV) are depicted in Figures 1-3, where full lines denote the logarithm of gener-

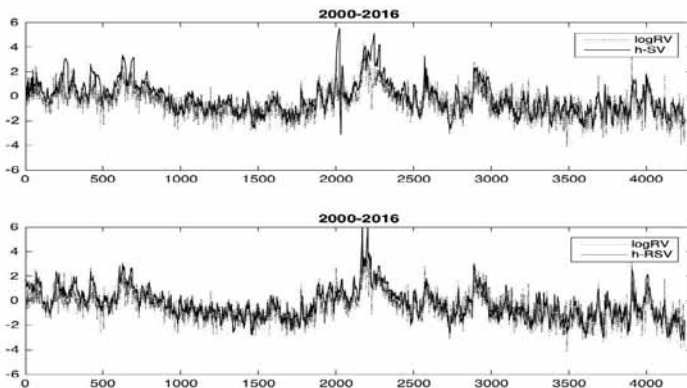


Figure 1. Realized volatilities and volatilities of 2000-2016

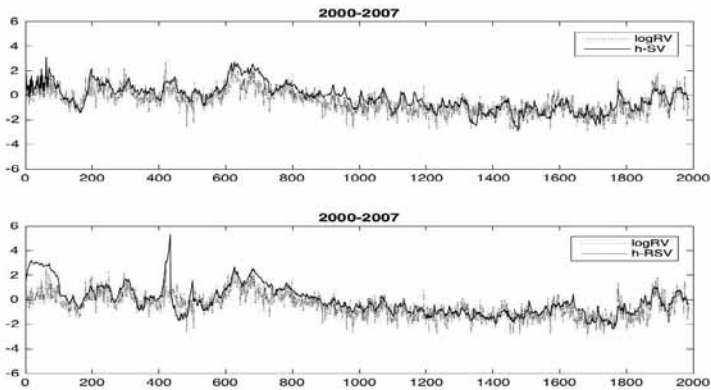


Figure 2. Realized volatilities and volatilities of 2000-2007

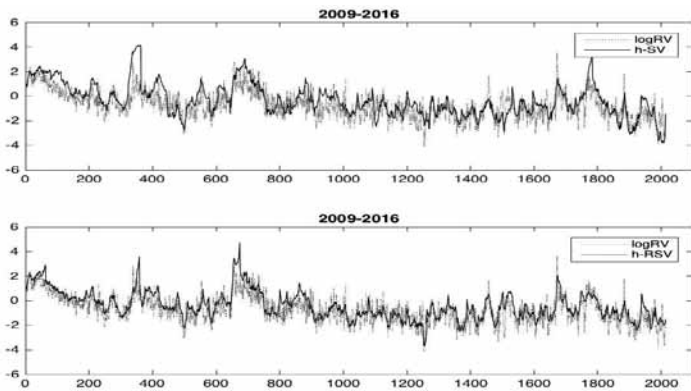


Figure 3. Realized volatilities and volatilities of 2009-2016

ated volatilities and dotted lines denote the logarithm of RVs.

By applying the previous 3000 observations of 2000-2016 and the previous 1000 observations of 2000-2007 and 2009-2016 to the SV and RSV models recursively, this paper generated one-day-ahead volatilities from the posterior predictive distributions with the respective forecast numbers 1247, 987 and 1014. Then, one-day-ahead VaR forecasts and one-day-ahead ES forecasts are estimated based on these volatility estimates.

Table 4 shows the results of EFR,  $p$ -values of the Kupiec LR test ( $LR_p$ ) for the VaR forecasts of all the investigated time periods at  $\alpha = 1\%$ ,  $5\%$ ,  $10\%$ . For  $\alpha = 10\%$  at the  $10\%$  significance level, the estimates of EFR are not significantly different from the corresponding probability  $\alpha$ , which imply that the models are robust to predict the  $10\%$  VaR forecasts. In the RSV model, the null hypothesis that  $EFR = \alpha$  is rejected for  $\alpha = 1\%$  and  $\alpha = 5\%$  even at the  $10\%$  significance level during the long time period of years 2000-2016, but not rejected at the  $5\%$  significance level in the cases of 2000-2007 and 2009-2016. In contrast, although the  $LR_p$  for  $\alpha = 5\%$  of the years 2000-2016 suggests a superior performance of the SV model, the RSV model outperforms the SV model in most other cases of  $\alpha = 1\%$  and  $\alpha = 5\%$  at the  $5\%$  significance level. It is also noticeable that the values of EFR are more inclined to rise for 2000-2016 than for 2000-2007 and 2009-2016 due to the shaky financial markets of the year 2008 as shown in Figure 1.

The results of evaluation criteria  $D(\alpha)$  are summarized in Table 5. Generally, the lower  $D(\alpha)$  indicate superior predictive ability of the RSV model for 2000-2007 and 2009-2016. On the other hand, the SV model is more suitable when the time span is expanded to 2000-2016.

Table 6 computes the loss functions MSE and QLIKE for the models. The

Table 4. Results for VaR

$\alpha$	SV			RSV		
	1%	5%	10%	1%	5%	10%
<i>2000-2016</i>						
<i>EFR</i>	0.03	0.06	0.10	0.04	0.07	0.11
<i>LR</i>	0.00	0.14	0.84	0.00	0.00	0.13
<i>2000-2007</i>						
<i>EFR</i>	0.02	0.06	0.10	0.02	0.06	0.11
<i>LR</i>	0.02	0.13	0.74	0.07	0.13	0.24
<i>2009-2016</i>						
<i>EFR</i>	0.02	0.07	0.11	0.02	0.06	0.10
<i>LR</i>	0.00	0.00	0.24	0.06	0.52	0.71

**Table 5.** Results for ES

$\alpha$	SV			RSV		
	1%	5%	10%	1%	5%	10%
<i>2000-2016</i>	0.19	0.25	0.23	0.29	0.30	0.26
<i>2000-2007</i>	0.26	0.19	0.17	0.25	0.14	0.14
<i>2009-2016</i>	0.31	0.26	0.20	0.31	0.21	0.18

**Table 6.** MSE and QLIKE

	<i>2000-2016</i>		<i>2000-2007</i>		<i>2009-2016</i>	
	SV	RSV	SV	RSV	SV	RSV
<i>MSE</i>	0.56	0.62	0.22	0.14	4.20	0.91
<i>QLIKE</i>	0.04	0.05	0.20	0.12	0.09	0.05

SRK (Hansen and Lunde (2005)) described as equation (15) is used as the proxy of latent volatility. The results of Table 6, which are in agreement with those of Table 5, state that the RSV model is more capable than the basic SV model for 2000-2007 and 2009-2016, but does not improve in predictive ability for 2000-2016.

## 5. Conclusion

This paper estimates the one-day-ahead VaR and ES forecasts for the S&P500 stock index, where latent volatilities are estimated by using the SV and RSV models. Observations of different time spans are chosen to investigate the model fitting.

Empirical studies show that the RSV model holds the advantage of describing the biases of RVs due to non-trading hours and MMN, whereas this model does not always improve the predictive performance with respect to different time spans. Consequently, on the subject of the investigated stock index, RSV model fits data of short time spans better, while SV is more appropriate for the long time span.

## Appendix A. Generation Algorithm for RSV Model

Substitute  $\alpha_t \equiv h_t - \mu$ ,  $\sigma_e \equiv \exp(\mu/2)$ ,  $c \equiv \xi + \mu$  into (5) – (7), such the RSV model could be equivalently expressed as

$$R_t = \exp(\alpha_t / 2) e_t, \quad (\text{A1})$$

$$\alpha_{t+1} = \phi \alpha_t + \eta_t, \quad (\text{A2})$$

$$x_t = c + \alpha_t + u_t, \quad (\text{A3})$$

$$\begin{pmatrix} e_t \\ \eta_t \\ u_t \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Sigma}), \mathbf{\Sigma} = \begin{pmatrix} \sigma_e^2 & \rho \sigma_e \sigma_\eta & 0 \\ \rho \sigma_e \sigma_\eta & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}.$$

In the light of the work discussed in Omori and Watanabe (2008), this specification will facilitate the generation steps through generating the latent  $\alpha_t$  instead of  $h_t$  on the condition that  $\rho \neq 0$ .

Provided that the prior distributions of parameters are the following:

$$\begin{aligned} \mu &\sim N(\mu_0, \sigma_0^2), \quad (\phi + 1) / 2 \sim \text{Beta}(a_{\phi_0}, b_{\phi_0}), \\ (\rho + 1) / 2 &\sim \text{Beta}(a_{\rho_0}, b_{\rho_0}), \quad \sigma_\eta^2 \sim \text{IG}(\alpha_{\eta_0} / 2, \beta_{\eta_0} / 2), \\ \xi &\sim N(\mu_{\xi_0}, \sigma_{\xi_0}^2), \quad \sigma_u^2 \sim \text{IG}(\alpha_{u_0} / 2, \beta_{u_0} / 2). \end{aligned}$$

$(\phi + 1) / 2$  follows a beta distribution which ensures the stationarity assumption that  $|\phi| < 1$ .  $\text{IG}(\alpha / 2, \beta / 2)$  denotes an inversed gamma distribution with shape parameter  $\alpha / 2$  and scale parameter  $\beta / 2$ .

Given the prior distributions, samples of the parameters and latent volatilities can be drawn recursively from the respective posterior distributions via MCMC-based Bayesian method as follows.



## A.1 Generation of $\theta$

### (1) Generation of $\phi$

The conditional posterior density of  $\phi$  is

$$\begin{aligned} & \pi(\phi | \mu, \rho, \sigma_\eta^2, \alpha, \mathbf{R}) \\ & \propto \pi(\phi) \sqrt{1 - \phi^2} \exp \left\{ - \frac{(1 - \phi^2) \alpha_1^2}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{T-1} [\alpha_{t+1} - \phi \alpha_t - \rho \sigma_\eta \sigma_e^{-1} \exp(-\alpha_t / 2) R_t]^2}{2\sigma_\eta^2 (1 - \rho^2)} \right\} \\ & \propto (1 + \phi)^{a_0 - 1} (1 - \phi)^{b_0 - 1} \sqrt{1 - \phi^2} \exp \left\{ - \frac{(\phi - \mu_\phi)^2}{2\sigma_\phi^2} \right\}, \end{aligned} \quad (A4)$$

where

$$\mu_\phi = \frac{\sum_{t=1}^{T-1} [\alpha_{t+1} - \rho \sigma_\eta \sigma_e^{-1} \exp(-\alpha_t / 2) R_t] \alpha_t}{\rho^2 \alpha_1^2 + \sum_{t=2}^{T-1} \alpha_t^2}, \quad (A5)$$

$$\sigma_\phi^2 = \frac{\sigma_\eta^2 (1 - \rho^2)}{\rho^2 \alpha_1^2 + \sum_{t=2}^{T-1} \alpha_t^2}. \quad (A6)$$

The Metropolis-Hastings (MH) algorithm is applied since the samples of  $\phi$  are not easy to obtain from this conditional posterior density of (A4). Specifically, the candidate  $\phi^*$  is generated from  $TN_{(-1,1)}(\mu_\phi, \sigma_\phi^2)$  in consideration of the restriction that  $|\phi| < 1$ , where  $TN_{(a,b)}(\mu, \sigma^2)$  represents a truncated normal distribution within the interval  $(a, b)$ . Given the current value  $\phi$  and accept  $\phi^*$  with probability

$$\min \left\{ \frac{(1 + \phi^*)^{a_0 - 1} (1 - \phi^*)^{b_0 - 1} \sqrt{1 - \phi^{*2}}}{(1 + \phi)^{a_0 - 1} (1 - \phi)^{b_0 - 1} \sqrt{1 - \phi^2}}, 1 \right\}. \quad (A7)$$

### (2) Generation of $\Gamma$

Let  $\Gamma$  represents the  $2 \times 2$  variance-covariance matrix of  $e_t$  and  $\eta_t$ ,

$$\Gamma = \begin{pmatrix} \sigma_e^2 & \rho \sigma_e \sigma_\eta \\ \rho \sigma_e \sigma_\eta & \sigma_\eta^2 \end{pmatrix}, \quad (A8)$$

The logarithm of conditional posterior density about  $\mathbf{\Gamma}$  is

$$\begin{aligned} \log \pi(\mathbf{\Gamma} | \phi, \mathbf{a}, \mathbf{R}) = \text{const} - \log \sigma_\eta - \frac{\alpha_1^2(1-\phi^2)}{2\sigma_\eta^2} - \log \sigma_e - \frac{R_T^2}{2\sigma_e^2 \exp(\alpha_T)} \\ - \frac{\nu_1 + 3}{2} \log |\mathbf{\Gamma}| - \frac{1}{2} \text{tr}(\mathbf{\Gamma}^{-1} \mathbf{\Gamma}_1^{-1}), \end{aligned} \quad (\text{A9})$$

where

$$\nu_1 = \nu_0 + T - 1, \quad (\text{A10})$$

$$\mathbf{\Gamma}_1^{-1} = \mathbf{\Gamma}_0^{-1} + \sum_{t=1}^{T-1} \tau_t \tau_t', \quad (\text{A11})$$

$$\tau_t = \begin{pmatrix} R_t \exp(-\alpha_t / 2) \\ \alpha_{t+1} - \phi \alpha_t \end{pmatrix}. \quad (\text{A12})$$

Generate the candidate matrix  $\mathbf{\Gamma}^{*-1} \sim W(\mathbf{\Gamma}_1, \nu_1)$ , where  $W(S, \nu)$  denotes a Wishart distribution with parameters  $(S, \nu)$ , and accept it with probability

$$\min \left\{ \frac{\sigma_\eta \sigma_e \exp \left[ -\frac{(1-\phi^2)\alpha_1^2}{2\sigma_\eta^2} - \frac{R_T^2}{2\sigma_e^2 \exp(\alpha_T)} \right]}{\sigma_\eta^* \sigma_e^* \exp \left[ -\frac{(1-\phi^2)\alpha_1^2}{2\sigma_\eta^{*2}} - \frac{R_T^2}{2\sigma_e^{*2} \exp(\alpha_T)} \right]}, 1 \right\}. \quad (\text{A13})$$

If  $\mathbf{\Gamma}^{*-1}$  is accepted, new draws of  $\sigma_e^{*2}, \rho^*, \sigma_\eta^{*2}$  are obtained from  $\mathbf{\Gamma}^*$  and therefore  $\mu^* = \log \sigma_e^{*2}$ ; else, samples will stay as the previous  $\sigma_e^2, \rho, \sigma_\eta^2, \mu$ .

### (3) Generation of $\xi$

The conditional posterior distribution of  $\xi$  is Gaussian with mean  $\mu_{\xi_1}$  and variance  $\sigma_{\xi_1}^2$  where

$$\mu_{\xi_1} = \frac{\sigma_{\xi_0}^2 \sum_{t=1}^T (x_t - h_t) + \sigma_u^2 \mu_{\xi_0}}{T \sigma_{\xi_0}^2 + \sigma_u^2}, \quad (\text{A14})$$

$$\sigma_{\xi_1}^2 = \frac{\sigma_{\xi_0}^2 \sigma_u^2}{T \sigma_{\xi_0}^2 + \sigma_u^2}. \quad (\text{A15})$$

Generate  $\xi \sim N(\mu_{\xi_1}, \sigma_{\xi_1}^2)$ .

#### (4) Generation of $\sigma_u^2$

It is also straightforward to generate  $\sigma_u^2$  from the conditional posterior distribution  $IG(\alpha_{u_1}/2, \beta_{u_1}/2)$  where

$$\alpha_{u_1} = \alpha_{u_0} + T, \quad (\text{A16})$$

$$\beta_{u_1} = \beta_{u_0} + \sum_{t=1}^T (x_t - \xi - h_t)^2. \quad (\text{A17})$$

### A.2 Generation of $h$

Since the existence of high autocorrelation in the latent variables  $\alpha$ , using single-move sampler, where only one  $\alpha_t$  is generated at one time from full conditional distribution given all the other parameters, may bring about inefficient MCMC samples, this paper applies block sampler developed by Omori and Watanabe (2008) to improve the generation efficiency.

Equation (A18) shows the operation of how to divide  $(\alpha_1, \alpha_2, \dots, \alpha_T)$  into  $K+1$  blocks by selecting  $K$  knots randomly in accordance with Shephard and Pitt (1997):

$$k_i = \text{int} \left\{ T \times (i + U_i) / (K + 2) \right\}, \quad i = 1, 2, \dots, K, \quad (\text{A18})$$

where  $\text{int}(x)$  rounds  $x$  to the nearest integer number and  $U_i$  is a uniformly distributed random number in the interval  $(0, 1)$ .

Suppose that there are  $m$  elements  $\alpha^{(i)} = (\alpha_{s+1}, \alpha_{s+2}, \dots, \alpha_{s+m})'$  included in the  $i$ -th block. Given  $\alpha_s$  and  $\alpha_{s+m+1}$ , generating the normalized  $\eta^{(i)} = (\eta_s, \eta_{s+1}, \dots, \eta_{s+m+1})'$ , which are easier to work with, is equivalent to generate the dependent  $(\alpha_{s+1}, \alpha_{s+2}, \dots, \alpha_{s+m})$  due to equation (A2).

The logarithm of joint conditional posterior density of  $\eta^{(i)}$  (ignore the constant term) is

$$\log f \left( \eta^{(i)} \mid \alpha_s, \alpha_{s+m+1}, R_s, \dots, R_{s+m}, x_s, \dots, x_{s+m}, \theta \right) = -\frac{1}{2\sigma_\eta^2} \sum_{t=s}^{s+m-1} \eta_t^2 + L, \quad (\text{A19})$$

where

$$L = \begin{cases} \sum_{t=s+1}^{s+m} l_t - \frac{1}{2\sigma_\eta^2} (\alpha_{s+m+1} - \phi\alpha_{s+m})^2, & s+m < T, \\ \sum_{t=s+1}^{s+m} l_t, & s+m = T, \end{cases} \quad (\text{A20})$$

$$l_t \equiv \log f(R_t, X_t | \mathbf{a}, \boldsymbol{\theta}) = -\frac{\alpha_t}{2} - \frac{(R_t - \mu_t)^2}{2\sigma_t^2} - \frac{(x_t - c - \alpha_t)^2}{2\sigma_u^2}, \quad (\text{A21})$$

$$\mu_t = \begin{cases} \rho\sigma_e\sigma_\eta^{-1}(\alpha_{t+1} - \phi\alpha_t)\exp(\alpha_t/2), & t < T, \\ 0, & t = T, \end{cases} \quad (\text{A22})$$

$$\sigma_t^2 = \begin{cases} (1 - \rho^2)\sigma_e^2\exp(\alpha_t), & t < T, \\ \sigma_e^2\exp(\alpha_T), & t = T. \end{cases} \quad (\text{A23})$$

Consider that  $L$  is a continuously differentiable function of  $\alpha_t$ , a  $(m \times 1)$  vector  $\mathbf{d} = (d_{s+1}, d_{s+2}, \dots, d_{s+m})'$  is the first derivatives of  $L$  and a positive definite  $(m \times m)$  matrix  $Q$  is  $-1$  times the matrix of second derivatives of  $L$  with respect to  $\alpha_t$ . Let  $\hat{L}$ ,  $\hat{\mathbf{d}}$  and  $\hat{Q}$  represent  $L$ ,  $\mathbf{d}$  and  $Q$  evaluated at  $\boldsymbol{\alpha}^{(i)} = \hat{\boldsymbol{\alpha}}^{(i)}$ , respectively. Approximate  $\log f(\boldsymbol{\eta}^{(i)} | \cdot)$  with a second-order Taylor expansion around the mode  $\hat{\boldsymbol{\eta}}^{(i)}$  (or, equivalently, around  $\hat{\boldsymbol{\alpha}}^{(i)}$ ):

$$\begin{aligned} & \log f\left(\boldsymbol{\eta}^{(i)} \mid \alpha_s, \alpha_{s+m+1}, R_s, \dots, R_{s+m}, x_s, \dots, x_{s+m}, \boldsymbol{\theta}\right) \\ & \cong -\frac{1}{2\sigma_\eta^2} \sum_{t=s}^{s+m-1} \eta_t^2 + \hat{L} + \frac{\partial L}{\partial \boldsymbol{\eta}^{(i)'}} \Bigg|_{\boldsymbol{\eta}^{(i)} = \hat{\boldsymbol{\eta}}^{(i)}} (\boldsymbol{\eta}^{(i)} - \hat{\boldsymbol{\eta}}^{(i)}) \\ & \quad + \frac{1}{2} (\boldsymbol{\eta}^{(i)} - \hat{\boldsymbol{\eta}}^{(i)})' E \left[ \frac{\partial^2 L}{\partial \boldsymbol{\eta}^{(i)} \partial \boldsymbol{\eta}^{(i)'}} \right] \Bigg|_{\boldsymbol{\eta}^{(i)} = \hat{\boldsymbol{\eta}}^{(i)}} (\boldsymbol{\eta}^{(i)} - \hat{\boldsymbol{\eta}}^{(i)}) \\ & = -\frac{1}{2\sigma_\eta^2} \sum_{t=s}^{s+m-1} \eta_t^2 + \hat{L} + \hat{\mathbf{d}}'(\boldsymbol{\alpha}^{(i)} - \hat{\boldsymbol{\alpha}}^{(i)}) - \frac{1}{2}(\boldsymbol{\alpha}^{(i)} - \hat{\boldsymbol{\alpha}}^{(i)})' \hat{Q}(\boldsymbol{\alpha}^{(i)} - \hat{\boldsymbol{\alpha}}^{(i)}) \\ & \equiv \log g(\boldsymbol{\eta}^{(i)} \mid \alpha_s, \alpha_{s+m+1}, R_s, \dots, R_{s+m}, x_s, \dots, x_{s+m}, \boldsymbol{\theta}), \end{aligned} \quad (\text{A24})$$

where  $g(\boldsymbol{\eta}^{(i)} | \cdot)$  is the proposal density function and

$$d_t = \frac{\partial L}{\partial \alpha_t} = -\frac{1}{2} + \frac{(R_t - \mu_t)^2}{2\sigma_t^2} + \frac{R_t - \mu_t}{\sigma_t^2} \frac{\partial \mu_t}{\partial \alpha_t} + \frac{R_{t-1} - \mu_{t-1}}{\sigma_{t-1}^2} \frac{\partial \mu_{t-1}}{\partial \alpha_t} + \frac{x_t - c - \alpha_t}{\sigma_u^2} + \kappa(\alpha_t), \quad (\text{A25})$$

$$Q = \begin{pmatrix} A_{s+1} & B_{s+2} & 0 & \cdots & 0 \\ B_{s+2} & A_{s+2} & B_{s+3} & \cdots & 0 \\ 0 & B_{s+3} & A_{s+3} & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B_{s+m} \\ 0 & \cdots & 0 & B_{s+m} & A_{s+m} \end{pmatrix}, \quad (\text{A26})$$

$$\frac{\partial \mu_t}{\partial \alpha_t} = \begin{cases} \frac{\rho \sigma_e}{\sigma_\eta} \left( -\phi + \frac{\alpha_{t+1} - \phi \alpha_t}{2} \right) \exp\left(\frac{\alpha_t}{2}\right), & t < T, \\ 0, & t = T, \end{cases} \quad (\text{A27})$$

$$\frac{\partial \mu_{t-1}}{\partial \alpha_t} = \begin{cases} 0, & t = 1, \\ \frac{\rho \sigma_e}{\sigma_\eta} \exp\left(\frac{\alpha_{t-1}}{2}\right), & t > 1, \end{cases} \quad (\text{A28})$$

$$\kappa(\alpha_t) = \begin{cases} \phi(\alpha_{t+1} - \phi \alpha_t) / \sigma_\eta^2, & t = s + m < T, \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A29})$$

$$A_t = -E \left[ \frac{\partial^2 L}{\partial \alpha_t^2} \right] = \frac{1}{2} + \frac{1}{\sigma_t^2} \left( \frac{\partial \mu_t}{\partial \alpha_t} \right)^2 + \frac{1}{\sigma_{t-1}^2} \left( \frac{\partial \mu_{t-1}}{\partial \alpha_t} \right)^2 + \frac{1}{\sigma_u^2} + \kappa'(\alpha_t), \quad (\text{A30})$$

$$B_t = -E \left[ \frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}} \right] = \frac{1}{\sigma_{t-1}^2} \frac{\partial \mu_{t-1}}{\partial \alpha_{t-1}} \frac{\partial \mu_{t-1}}{\partial \alpha_t}, \quad (\text{A31})$$

$$\kappa'(\alpha_t) = \begin{cases} \frac{\phi^2}{\sigma_\eta^2}, & t = s + m < T, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A32})$$

The expectations are taken with respect to  $R_t$ 's conditional on  $\alpha_t$ 's. Hence

the logarithm of proposal density function  $\log g(\boldsymbol{\eta}^{(i)}|\cdot)$  can be written as

$$\begin{aligned} \log g\left(\boldsymbol{\eta}^{(i)} \mid \alpha_s, \alpha_{s+m+1}, R_s, \dots, R_{s+m}, x_s, \dots, x_{s+m}, \boldsymbol{\theta}\right) \\ = -\frac{1}{2\sigma_\eta^2} \sum_{t=s}^{s+m-1} \eta_t^2 + \hat{L} + \sum_{t=s+1}^{s+m} (\alpha_t - \hat{\alpha}_t) \hat{d}_t \\ - \frac{1}{2} \left\{ \sum_{t=s+1}^{s+m} (\alpha_t - \hat{\alpha}_t)^2 \hat{A}_t + 2 \sum_{t=s+2}^{s+m} (\alpha_t - \hat{\alpha}_t)(\alpha_{t-1} - \hat{\alpha}_{t-1}) \hat{B}_t \right\}, \end{aligned} \quad (\text{A33})$$

where  $\hat{A}_t$  and  $\hat{B}_t$  are values that  $A_t$  and  $B_t$  evaluated at  $\alpha_t = \hat{\alpha}_t$ .

### (I) Generation of $\hat{\boldsymbol{\eta}}^{(i)}$

Omori and Watanabe (2008) proposed to apply the Kalman filter (Anderson and Moore (1979), de Jong (1991)) and disturbance smoother (Koopman (1993)) to the linear Gaussian state space model to find the posterior modes of  $\hat{\boldsymbol{\eta}}^{(i)}$  and  $\hat{\boldsymbol{\alpha}}^{(i)}$ . This segment explicates the computation procedures by iterating the following algorithm several times until  $\hat{\boldsymbol{\eta}}^{(i)}$  converge to the posterior modes.

For  $t = s+1, \dots, s+m$ :

Step 1. Initialize  $\hat{\alpha}_t$ .

Step 2. Compute  $\hat{d}_t, \hat{A}_t, \hat{B}_t$  at  $\alpha_t = \hat{\alpha}_t$ .

Step 3. Calculate the following variables:

$$E_t = \begin{cases} A_t, & t = s+1, \\ A_t - E_{t-1}^{-1} B_t^2, & t = s+2, \dots, s+m, \end{cases}$$

$$F_t = \sqrt{E_t},$$

$$M_t = \begin{cases} 0, & t = s+1, s+m+1, \\ \hat{B}_t F_{t-1}^{-1}, & t = s+2, \dots, s+m, \end{cases}$$

$$b_t = \begin{cases} \hat{d}_t, & t = s+1, \\ \hat{d}_t^{(j)} - M_t F_{t-1}^{-1} b_{t-1}, & t = s+2, \dots, s+m, \end{cases}$$

$$\hat{\gamma}_t = \hat{\alpha}_t + F_t^{-1} M_{t+1} \hat{\alpha}_{t+1},$$

$$\hat{y}_t = \hat{\gamma}_t + E_t^{-1}b_t,$$

Step 4. The linear Gaussian state space representation is given by

$$\hat{y}_t = Z_t\alpha_t + G_t\mathbf{v}_t, \quad (\text{A34})$$

$$\alpha_{t+1} = \phi\alpha_t + H_t\mathbf{v}_t, \quad (\text{A35})$$

$$\mathbf{v}_t \sim i.i.d.N(\mathbf{0}, \mathbf{I}),$$

where  $\mathbf{I}$  is a  $2 \times 2$  identity matrix and

$$Z_t = 1 + F_t^{-1}M_{t+1}\phi, \quad G_t = [F_t^{-1}, F_t^{-1}M_{t+1}\sigma_\eta], \quad H_t = [0, \sigma_\eta]$$

Step 4-a. Apply the Kalman filter to (A34) and (A35), and generate the series  $\{a_t\}_{t=s+1}^{s+m}$  and  $\{P_t\}_{t=s+1}^{s+m}$  recursively:

$$a_{t+1} = \begin{cases} 0, & t = s, \\ \phi a_t + K_t \zeta_t, & t = s+1, \dots, s+m-1, \end{cases} \quad (\text{A36})$$

$$P_{t+1} = \begin{cases} \sigma_\eta^2, & t = s, \\ \phi P_t L_t + H_t J_t', & t = s+1, \dots, s+m-1, \end{cases} \quad (\text{A37})$$

where

$$\zeta_t = \hat{y}_t - Z_t\alpha_t, \quad D_t = Z_t^2 P_t + G_t G_t', \quad K_t = (\phi P_t Z_t + H_t G_t') D_t^{-1},$$

$$L_t = \phi - K_t Z_t, \quad J_t = H_t - K_t G_t.$$

Step 4-b. Compute the sequences  $\{r_t\}_{t=s}^{s+m-1}$  and  $\{U_t\}_{t=s}^{s+m-1}$  in succession by backward starting based on the disturbance smoother method:

$$r_{t-1} = \begin{cases} 0, & t = s+m, \\ Z_t D_t^{-1} \zeta_t + L_t r_t, & t = s+m-1, \dots, s, \end{cases} \quad (\text{A38})$$

$$U_{t-1} = \begin{cases} 0, & t = s+m, \\ Z_t^2 D_t^{-1} + L_t^2 U_t, & t = s+m-1, \dots, s. \end{cases} \quad (\text{A39})$$

Then

$$\mathbf{v}_t \mid \boldsymbol{\theta}, \mathbf{R}, \mathbf{X} \sim N\left(E\left[\mathbf{v}_t \mid \boldsymbol{\theta}, \mathbf{R}, \mathbf{X}\right], \text{Var}\left[\mathbf{v}_t \mid \boldsymbol{\theta}, \mathbf{R}, \mathbf{X}\right]\right),$$

where

$$E\left[\mathbf{v}_t \mid \boldsymbol{\theta}, \mathbf{R}, \mathbf{X}\right] = \mathbf{G}'_t D_t^{-1} \boldsymbol{\zeta}_t + \mathbf{J}'_t \boldsymbol{\gamma}_t, \quad (\text{A40})$$

$$\text{Var}\left[\mathbf{v}_t \mid \boldsymbol{\theta}, \mathbf{R}, \mathbf{X}\right] = \mathbf{I} - \mathbf{G}'_t \mathbf{G}_t D_t^{-1} - \mathbf{J}'_t \mathbf{J}_t U_t, \quad (\text{A41})$$

Therefore, the state variables  $(\hat{\boldsymbol{\alpha}}_{s+1}, \hat{\boldsymbol{\alpha}}_{s+2}, \dots, \hat{\boldsymbol{\alpha}}_{s+m})'$  given the initial  $\hat{\boldsymbol{\alpha}}_s$ :

$$\hat{\boldsymbol{\alpha}}_{t+1} = \phi \hat{\boldsymbol{\alpha}}_t + \mathbf{H}_t (\mathbf{G}'_t D_t^{-1} \boldsymbol{\zeta}_t + \mathbf{J}'_t \boldsymbol{\gamma}_t). \quad (\text{A42})$$

Step 5. Go to Step 2.

## (2) Generation of $\boldsymbol{\eta}^{(i)}$

Reserve  $\{\mathbf{G}_t, \mathbf{J}_t, D_t, Z_t, L_t, \zeta_t, \boldsymbol{\gamma}_t\}_{t=s}^{s+m-1}$  and generate  $\boldsymbol{\eta}^{(i)}$  from the joint posterior distribution of (A24) with the following generation algorithm:

Step 1. Set  $r_{s+m}^* = 0, U_{s+m}^* = 0$ . Calculate the sequences  $\{C_t, \boldsymbol{\kappa}_t, V_t, r_t^*, U_t^*, \boldsymbol{\eta}_t^*\}_{t=s}^{s+m-1}$  by backward starting with  $t = s+m$ :

$$C_t = \mathbf{H}_t (\mathbf{I} - \mathbf{G}'_t \mathbf{G}_t D_t^{-1} - \mathbf{J}'_t \mathbf{J}_t U_t^*) \mathbf{H}'_t, \quad (\text{A43})$$

$$\boldsymbol{\kappa}_t \sim N(0, C_t), \quad (\text{A44})$$

$$V_t = \mathbf{H}_t (\mathbf{G}'_t D_t^{-1} Z_t + \mathbf{J}'_t U_t^* L_t), \quad (\text{A45})$$

$$r_{t-1}^* = Z_t D_t^{-1} \boldsymbol{\zeta}_t + L_t \boldsymbol{\gamma}_t - V_t C_t^{-1} \boldsymbol{\kappa}_t, \quad (\text{A46})$$

$$U_{t-1}^* = Z_t^2 D_t^{-1} + L_t^2 U_t^* + V_t^2 C_t^{-1}, \quad (\text{A47})$$

$$\boldsymbol{\eta}_t^* = \mathbf{H}_t (\mathbf{G}'_t D_t^{-1} \boldsymbol{\zeta}_t + \mathbf{J}'_t \boldsymbol{\gamma}_t^*) + \boldsymbol{\kappa}_t. \quad (\text{A48})$$

Step 2. Let  $\boldsymbol{\eta}^{*(i)} = (\boldsymbol{\eta}_s^*, \boldsymbol{\eta}_{s+1}^*, \dots, \boldsymbol{\eta}_{s+m-1}^*)'$  be the samples of proposal density function  $g(\boldsymbol{\eta}^{(i)} | \cdot)$  and accept it with probability



$$\min \left\{ \frac{f(\boldsymbol{\eta}^{*(i)} | \cdot)g(\boldsymbol{\eta}^{(i)} | \cdot)}{f(\boldsymbol{\eta}^{(i)} | \cdot)g(\boldsymbol{\eta}^{*(i)} | \cdot)}, 1 \right\}. \quad (\text{A49})$$

## References

- Andersen, T. G. and Bollerslev, T. (1998). "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts", *International Economic Review*, 39, 885-905.
- Barndorff-Nielsen, O. E. Hansen, P. R., Lunde, A. and Shephard, N. (2008). "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise", *Econometrica*, 76, 1482-1536.
- Barndorff-Nielsen, O. E. and Shephard, N. (2001). "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society*, B 63, 167-241.
- Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.
- Chib, S. (2001). "Markov chain monte carlo methods: computation and inference", *Handbook of Econometrics*, Volume 5, edited by J. J. Heckman and E. Leaper, Amsterdam: North-Holland, Amsterdam, 3569-3649.
- Diebold, F. X. (1988). *Empirical Modeling of Exchange Rate Dynamics*, Springer-Verlag, New York.
- Embrechts, P., Kaufmann, R. and Patie, P. (2005). "Strategic long-term financial risks: single risk factors", *Computational Optimization and Applications*, 32, 61-90.
- Engle, R. F. (1982). "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, 50, 987-1007.
- Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion)", *Bayesian Statistics*, Volume 4, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford University Press, Oxford.
- Hansen, P. R. and Lunde, A. (2005). "A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?" *Journal of Applied Econometrics*, 20, 873-889.
- Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics and econometrics of securities trading*. Oxford: Oxford University Press.
- Heber, G., Lunde, N., Shephard, N. and Shepard, K. (2009). "Oxford-Man Institute's Realized Library, Version 0.1", Oxford-Man Institute, University of Oxford.
- Koopman, S. J. (1993). "Disturbance smoother for state space models", *Biometrika*, 80, 117-126.
- Kupiec, P. (1995). "Techniques for verifying the accuracy of risk measurement models", *Journal of Derivatives*, 3, 73-84.
- O'Hara, M. (1995). *Market Microstructure Theory*. Blackwell.
- Omori, Y. and Watanabe, T. (2008). "Block sampler and posterior mode estimation for asymmetric stochastic volatility models", *Computational Statistics & Data Analysis*, 52, 2892-2910.
- Patton, A. J. (2011). "Volatility forecast comparison using imperfect volatility proxies", *Journal of Econometrics*, 160, 246-256.
- Takahashi, M., Omori, Y. and Watanabe, T. (2009). "Estimating stochastic volatility models using daily returns and realized volatility simultaneously", *Computational Statistics & Data Analysis*,

53, 2404-2426.

Takahashi, M., Watanabe, T. and Omori, Y. (2016). "Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution", *International Journal of Forecasting*, 32, 437-457.

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

Watanabe, T. (2012). "Quantile forecasts of financial returns using realized GARCH models", *Japanese Economic Review*, 63, 68-80.