

CIRJE-F-880

**Realized Stochastic Volatility with Leverage
and Long Memory**

Shinichiro Shirota
Graduate school of Economics, University of Tokyo

Takayuki Hizu
Mitsubishi UFJ Trust and Banking

Yasuhiro Omori
University of Tokyo

March 2013

CIRJE Discussion Papers can be downloaded without charge from:

<http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html>

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

Realized stochastic volatility with leverage and long memory

Shinichiro Shirota

Graduate school of Economics, The University of Tokyo, Japan.

Takayuki Hizu

Mitsubishi UFJ Trust and Banking, Japan.

Yasuhiro Omori*

Faculty of Economics, The University of Tokyo, Japan

Abstract

The daily return and the realized volatility are simultaneously modeled in the stochastic volatility model with leverage and long memory. The dependent variable in the stochastic volatility model is the logarithm of the squared return, and its error distribution is approximated by a mixture of normals. In addition, we incorporate the logarithm of the realized volatility into the measurement equation, assuming that the latent log volatility follows an Autoregressive Fractionally Integrated Moving Average (ARFIMA) process to describe its long memory property. Using a state space representation, we propose an efficient Bayesian estimation method implemented using Markov chain Monte Carlo method (MCMC). Model comparisons are performed based on the marginal likelihood, and the volatility forecasting performances are investigated using S&P500 stock index returns.

Keywords: ARFIMA, Leverage effect, Long memory, Markov Chain Monte Carlo, Mixture sampler, Realized volatility, Realized stochastic volatility model, State space model

1. Introduction

The realized volatility is defined as the sum of the squared intraday returns over a specified time interval such as a day (e.g., Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2001)). This measure would provide a consistent estimator of the latent volatility under the ideal market assumption. The theory of the realized volatility is discussed in Barndorff-Nielsen and Shephard (2002) and Meddahi (2002), and there have been extensive studies on its time series structure and performance in volatility prediction (e.g., Andersen et al. (2003), Andersen et al. (2007), Andersen et al. (2004), Koopman et al. (2005) and Maheu and McCurdy (2007)).

In the real market, however, two major problems arise in measuring the daily realized volatility using high frequency return data: (1) the presence of non-trading hours and (2) market microstructure noise in transaction prices. The first problem arises because the stock market is usually open for only part of the day. For example, the Tokyo Stock Exchange (TSE) is open for 4.5 hours a day and there is a lunch break. If we calculate the realized volatility as the sum of the squared intraday returns when the market is open, we may underestimate the latent one-day volatility. To avoid this underestimation, Hansen and Lunde (2005)

*Corresponding author. Tel.: +81 3 5841 5516; fax: +81 3 5841 5521.

Email addresses: 6614272319@mail.ecc.u-tokyo.ac.jp (Shinichiro Shirota), omori@e.u-tokyo.ac.jp (Yasuhiro Omori)

proposed a scale realized volatility that adjusts the realized volatility by the ratio of the variance of the daily return to the mean of the realized volatility.

Market microstructure noise has various causes, including bid-ask spread and variation in trade sizes (see O'Hara (1995) and Hasbrouck (2007) for details) and can cause the realized volatility to be a biased estimator of the latent volatility. As the sample time interval approaches zero, the bias owing to microstructure noise is expected to increase significantly. At-Sahalia et al. (2005) and Bandi and Russell (2008) propose a procedure to determine the optimal sampling interval, and Zhang et al. (2005) propose a bias adjusting method by assigning different weights to the realized volatilities calculated using different time intervals. In addition, Barndorff-Nielsen et al. (2008) derive the Realized Kernel (RK) as a consistent estimator of the latent volatility using high frequency data with noise.

Whereas, the intraday returns are heavily contaminated by microstructure noise, the daily returns are less subject to the noise. The daily returns could, therefore, provide additional information to eliminate the bias owing to microstructure noise and non-trading hours simultaneously. Takahashi et al. (2009) propose an extension of the stochastic volatility (SV) model to include such simultaneous modeling of the daily returns and realized volatility known as the Realized Stochastic Volatility (RSV) model. Hansen et al. (2012) implement a similar simultaneous modeling approach within the GARCH framework, called the Realized GARCH model, and demonstrate the superior performance of the proposed model compared to GARCH (using daily returns only). Maheu and McCurdy (2011) consider the simultaneous modeling of S&P500 and IBM data and show that this approach outperforms the conventional EGARCH model.

Two important properties of the stochastic volatility and realized volatility have been discussed in previous empirical studies: (i) the leverage effect and (ii) long memory. The leverage effect refers to the correlation between the return at time t and the logarithm of the volatility at time $t + 1$ and has been well established in empirical studies of stock returns (see, e.g., the survey by Shephard (2005)). To account for leverage effects, Melino and Turnbull (1990), for example, use the GMM (generalised methods of moments), and Harvey and Shephard (1996) use the QML (quasi-maximum likelihood method) with the Kalman filter for their estimation. Bayesian estimations have been described in various studies (e.g., Jacquier et al. (2004), Omori et al. (2007), Omori and Watanabe (2008)). Takahashi et al. (2009) further propose a Bayesian estimation method for the RSV model with leverage where they use a single realized measure, while multiple realized measures are used in Venter and de Jongh (2013) and Koopman and Scharth (2013). Superposition model, in which the logarithm of the volatility is a sum of latent factor processes, is proposed to describe the long-range dependence of the volatility in Dobrev and Szerszen (2010) with jumps in latent processes, and in Koopman and Scharth (2013) with a correlation between returns and measurement errors.

The long memory property of the realized volatility has also been investigated in many empirical studies using the high frequency data (e.g., Andersen, Bollerslev, Diebold, and Ebens (2001)) and Raggi and Bordignon (2012) modeled the realized volatility with long memory and Markov switching dynamics using a Bayesian estimation method for the state space model. The SV model with long memory is discussed in Breidt et al. (1998) using the frequency domain approach (spectral likelihood estimator) and in So (2002) using a Bayesian approach with the state space model (So (1999)). Ruiz and Veiga (2008) investigate the statistical property of the stochastic volatility model with leverage and long memory (but without using the realized volatility), and compare with those of FIEGARCH models. Further, the autocorrelation function of powered absolute returns and their cross-correlations with original returns are derived in Pérez et al. (2009).

This paper extends the RSV model by incorporating both the leverage effects in the SV model and the long memory property of the realized volatility, and proposes a highly efficient Bayesian estimation method with a Markov Chain Monte Carlo (MCMC) implementation. Instead of the block sampler used in Takahashi

et al. (2009), we employ the mixture sampler, a highly efficient Bayesian estimation method proposed by Kim et al. (1998) and Omori et al. (2007). In these methods, we take the logarithm of the squared asset return as a dependent variable to obtain linear measurement equations and approximate the error distribution by a mixture of normal distributions. In addition to the transformed stochastic volatility model, we assume an Autoregressive Fractionally Integrated Moving Average (ARFIMA) process for the logarithm of the log volatility to describe the long memory property of the realized volatility.

The paper is organized as follows. In Section 2, we introduce our model and its motivation. Section 3 describes the Bayesian estimation procedure based on the state space representation and Markov chain Monte Carlo methods. We illustrate our proposed method through numerical examples using simulated data in Section 4. In Section 5, we present our empirical studies using S&P500 realized volatility and realized kernels, perform model comparisons based on the marginal likelihood, and investigate the volatility forecast performances. We conclude in Section 6.

2. Realized stochastic volatility with leverage and long memory

2.1. Realized stochastic volatility with leverage

The simple stochastic volatility model with leverage is given by

$$y_{1t} = \exp(h_t/2)\epsilon_t, \quad t = 1, 2, \dots, n, \quad (1)$$

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

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \text{i.i.d. } N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix}\right), \quad |\phi| < 1, \quad (3)$$

$$h_1 \sim (\mu, \sigma^2/(1 - \phi^2)), \quad (4)$$

where y_{1t} is a stock return at time t . The parameter ρ measures the correlation between ϵ_t and η_t and, when negative, captures the increase in volatility following a drop in equity returns (e.g., Black (1976), Nelson (1991), Yu (2005)). The volatility clustering is described by the first order autoregressive process (8) with mean μ for the h_{t+1} (the log volatility at time $t + 1$). Because it is difficult to evaluate the likelihood function using the high dimensional numerical integration, Bayesian efficient estimation methods have been proposed in previous studies (e.g., Omori et al. (2007), Omori and Watanabe (2008)).

Furthermore, to incorporate the information contained in the realized volatility, Takahashi et al. (2009) propose simultaneous modeling of the daily returns and realized volatility by introducing an additional measurement equation

$$y_{2t} = \xi + h_t + u_t, \quad t = 1, 2, \dots, n, \quad u_t \sim \text{i.i.d. } N(0, \sigma_u^2), \quad (5)$$

where y_{2t} is the logarithm of the realized volatility at time t , and u_t is assumed to be independent of ϵ_t and η_t . This model makes it possible to use the realized volatility calculated from all available returns without any additional adjustment such as selecting the optimal sampling frequency to compute the realized volatility. The bias adjustment term, ξ , accounts for the effects of the market microstructure noise and non-trading hours simultaneously. When it is negative (positive), the realized volatility is considered to underestimate (overestimate) the latent volatility. We refer to this model as the Realized Stochastic Volatility (RSV) model (e.g., Koopman and Scharth (2013), Dobrev and Szerszen (2010), Venter and de Jongh (2013)). Although we could extend the model by replacing h_t with ψh_t in (5), where ψ is another adjustment coefficient, this extension does not necessarily improve the model fit in the empirical studies in Section 5.3. We therefore adopt the measurement equation (5) for the logarithm of the realized volatility by setting $\psi = 1$.

Takahashi et al. (2009) compared the simultaneous models using naive and scaled realized volatilities based on the marginal likelihood, and demonstrated that the effect of non-trading hours is more important than that of microstructure noise.

2.2. Realized stochastic volatility with superposition

Koopman and Scharth (2013) consider superpositions of independent ARMA processes as the volatility process in RSV framework.

$$y_{1t} = \exp\left(\sum_{j=1}^K h_{j,t}/2\right)\epsilon_t, \quad t = 1, 2, \dots, n, \quad (6)$$

$$y_{2t} = \xi + \sum_{j=1}^K h_{j,t} + u_t, \quad t = 1, 2, \dots, n, \quad (7)$$

$$h_{1,t+1} = \mu + \phi_1(h_{1,t} - \mu) + \eta_{1,t}, \quad t = 1, \dots, n, \quad (8)$$

$$h_{j,t+1} = \phi_j h_{j,t} + \eta_{j,t}, \quad j = 2, \dots, K, \quad t = 1, \dots, n, \quad (9)$$

$$h_1 \sim (\mu, \sigma_{\eta_1}^2/(1 - \phi_1^2)), \quad h_j \sim (0, \sigma_{\eta_j}^2/(1 - \phi_j^2)), \quad j = 2, \dots, K \quad (10)$$

This model can describe the long memory property of the log volatility process. We will explain about this property in the next subsection. We also include the leverage effect to the each independent volatility processes. In this literature, we call this model as realized stochastic model with superposition (RSV-SP).

Koopman and Scharth (2013) argue that the dependence between ϵ_t and u_t may not be negligible due to the discretization effects and jumps when the realized measure is based on a finite sample. We could consider, for example, a linear dependence by incorporating an additional correlation between these errors, but it is still not clear how to model the nonlinear dependence structure. Detail investigation of such a dependence problem is left for our future work.

2.3. The long memory property of the realized volatility

In empirical studies, the realized volatilities often display long memory properties, and the ARFIMA process is frequently used to express this characteristic (Andersen et al. (2003), Giot and Laurent (2004), Koopman et al. (2005), Raggi and Bordignon (2012)). The ARFIMA(p,d,q) process is defined by

$$(1 - L)^d \Phi(L)(h_{t+1} - \mu) = \Theta(L)\eta_t, \quad t = 0, 1, \dots, n, \quad (11)$$

where η_t denotes white noise, L is the lag operator such that $L^i h_t = h_{t-i}$, $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$, and $\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$. When $d = 1$, it is referred to as the ARIMA(p,1,q) process and is nonstationary. We observe that in general

$$(1 - L)^d = 1 + \sum_{j=1}^{\infty} \frac{d(d-1)\cdots(d-j+1)}{j!} (-L)^j = 1 + \sum_{j=1}^{\infty} \gamma_j L^j, \quad (12)$$

$$\gamma_0 = 1, \quad \gamma_{j+1} = \frac{j-d}{j+1} \gamma_j, \quad j \geq 0. \quad (13)$$

We assume that $0 < d < 1$ because the estimates of the memory parameter, d , are typically found to fall between 0 and 1 in empirical studies of the realized volatilities. The process is stationary if $d < 0.5$ and nonstationary if $d \geq 0.5$.

Long memory stochastic volatility models where the latent log volatility is assumed to follow an ARFIMA process have also been discussed in several studies without explicitly utilizing the information contained in

the realized volatilities (e.g., Breidt et al. (1998), So (2002), Ruiz and Veiga (2008)). Koopman and Scharth (2013) implement a superposition model in the RSV framework, describing the long range behavior of the log volatility process. However, in this paper, we consider a straightforward description of the long range behavior by adapting the ARFIMA process directly in the RSV framework. To incorporate the long memory property of the realized volatility into the stochastic volatility model, we consider the following state space model, referred to as the Realized Stochastic Volatility with Long Memory model (RSV-LM(p,d,q)):

$$y_{1t} = \exp(h_t/2)\epsilon_t, \quad t = 1, 2, \dots, n, \quad (14)$$

$$y_{2t} = \xi + h_t + u_t, \quad t = 1, 2, \dots, n, \quad (15)$$

$$(1-L)^d \Phi(L)(h_{t+1} - \mu) = \Theta(L)\eta_t, \quad t = 0, 1, \dots, n, \quad (16)$$

$$\begin{pmatrix} \epsilon_t \\ u_t \\ \eta_t \end{pmatrix} \sim \text{i.i.d. } N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{pmatrix} \right). \quad (17)$$

In this paper, we focus on three specific cases of appearing frequently in empirical studies: RSV-LM(0,d,0), RSV-LM(0,d,1) and RSV-LM(1,d,0). We therefore assume that $\Phi(L) = 1 - \phi L$ and $\Theta(L) = 1 - \theta L$, where $|\phi| < 1$ and $|\theta| < 1$. We assume that u_t is independent of (ϵ_t, η_t) because the measurement error is dominated by the computing the realized volatilities. However, it is straightforward to extend our model to incorporate a correlation between u_t and the other error terms.

3. Highly efficient Bayesian estimation

This section describes the highly efficient Bayesian estimation of the parameter using the MCMC method. Following the mixture sampler approach (e.g., Kim et al. (1998), Omori et al. (2007)), we first represent the RSV-LM model in linear Gaussian state space form to utilize efficient estimation procedures such as filtering, smoothing and prediction.

3.1. The efficient auxiliary mixture sampler

The mixture sampler proposed by Kim et al. (1998) and Omori et al. (2007) is well-known as an efficient MCMC sampling method for SV models. Its basic idea is to transform the nonlinear measurement equation into a linear equation and to approximate the distribution of the non-normal disturbances by a mixture of normals. We first transform y_{1t} to (y_{1t}^*, δ_t) in Equation (14) as follows.

$$y_{1t}^* = \log y_{1t}^2 = h_t + \epsilon_t^*, \quad \epsilon_t^* = \log \epsilon_t^2, \quad (18)$$

$$\delta_t = I(y_{1t} \geq 0) - I(y_{1t} < 0), \quad (19)$$

for $t = 1, 2, \dots, n$. Because ϵ_t^* is the logarithm of the chi-square random variable with one degree of freedom, its probability density is given by

$$f(\epsilon_t^*) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\epsilon_t^* - \exp(\epsilon_t^*)}{2}\right). \quad (20)$$

We approximate this density by mixtures of normal densities as follows:

$$g(\epsilon_t^*) = \sum_{j=1}^K p_j f_N(\epsilon_t^* | m_j, v_j^2), \quad (21)$$

where $f_N(\epsilon_t^*|m, v^2)$ denotes the probability density of a normal distribution with mean m and variance v^2 , $N(m, v^2)$. As the conditional distribution of η_t given ϵ_t^* and δ_t is

$$\eta_t|\rho, \sigma_\eta, \delta_t, \epsilon_t^* \sim N(\delta_t\rho\sigma_\eta\exp(\epsilon_t^*/2), \sigma_\eta^2(1 - \rho^2)), \quad (22)$$

where the mean is a nonlinear function of ϵ_t^* , we furthermore employ a linear approximation of $\exp(\epsilon_t^*/2)$. We therefore approximate the distribution of (ϵ_t^*, η_t) by a bivariate mixture normal distribution given δ_t ,

$$g(\epsilon_t^*, \eta_t|\rho, \sigma_\eta, \delta_t) = \sum_{j=1}^K p_j f_N(\epsilon_t^*|m_j, v_j^2) \times f_N[\eta_t|\delta_t\rho\sigma_\eta\exp(m_j/2)\{a_j + b_j(\epsilon_t^* - m_j)\}, \sigma_\eta^2(1 - \rho^2)]. \quad (23)$$

This approximation is generally quite accurate and there is little difference between the true and approximated probability densities. Omori et al. (2007) propose the approximation with $K = 10$ and provide their selected $p_j \equiv Pr(s_t = j)$ and mixture component parameters (m_j, v_j, a_j, b_j) for $j = 1, \dots, 10$; we reproduced this parameter list in Table 1.

Table 1: The values of $p_j, m_j, v_j^2, a_j, b_j$

j	p_j	m_j	v_j^2	a_j	b_j
1	0.00609	1.92677	0.11265	1.01418	0.50710
2	0.04775	1.34744	0.17788	1.02248	0.51124
3	0.13057	0.73504	0.26768	1.03403	0.51701
4	0.20674	0.02266	0.40611	1.52070	0.52604
5	0.22715	-0.85173	0.62699	1.08153	0.54076
6	0.18842	-1.97278	0.98583	1.13114	0.56557
7	0.12047	-3.46788	1.57469	1.21754	0.60877
8	0.05591	-5.55246	2.54498	1.37454	0.68728
9	0.01575	-8.68384	4.16491	1.68327	0.84163
10	0.00115	-14.65000	7.33342	2.50097	1.25049

3.2. Linear Gaussian state space representation

Given $s = \{s_1, \dots, s_n\}$, we obtain a linear Gaussian measurement equation. Noting that $h_t = E(h_t|h_{t-1}, h_{t-2}, \dots) + \eta_{t-1}$, we define a dependent vector and a state vector,

$$y_t^* = \begin{pmatrix} y_{1t}^* \\ y_{2t}^* \end{pmatrix}, \quad \alpha_t = \begin{pmatrix} \eta_{t-1} \\ E(h_t|h_{t-1}, h_{t-2}, \dots) \\ E(h_{t+1}|h_{t-1}, h_{t-2}, \dots) \\ \vdots \\ E(h_{t+M-1}|h_{t-1}, h_{t-2}, \dots) \end{pmatrix}, \quad (24)$$

so that h_t is a sum of the first two elements of α_t . Let 0_k , 1_l and I_m denote a $k \times 1$ zero vector, a $l \times 1$ vector with all elements equal to one, and an $m \times m$ identity matrix. Then, using the MA model of an ARFIMA

process for the state equation based on the finite truncation ¹,

$$\begin{aligned}
h_{t+1} &= \mu + \Phi(L)^{-1}(1-L)^{-d}\Theta(L)\eta_t \\
&= \mu + \sum_{j=0}^{\infty} \psi_j \eta_{t-j} \\
&\approx \mu + \sum_{j=0}^M \psi_j \eta_{t-j}, \quad \psi_0 = 1, \quad \psi \equiv (\psi_1, \dots, \psi_M)',
\end{aligned} \tag{25}$$

we can represent the RSV-LM model as the following linear Gaussian state space model:

$$y_t^* = X_t \beta + Z_t \alpha_t + G_t u_t^*, \quad t = 1, 2, \dots, n, \tag{26}$$

$$\alpha_{t+1} = W_t \beta + T_t \alpha_t + H_t u_t^*, \quad t = 0, 1, \dots, n-1, \tag{27}$$

$$\alpha_0 \equiv 0, \quad u_t^* \sim \text{i.i.d. } N(0_3, I_3), \tag{28}$$

where

$$\begin{aligned}
X_t &= \begin{pmatrix} m_{s_t} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 \\ \xi \\ \mu \end{pmatrix}, \quad Z_t = \begin{pmatrix} 1 & 1 & 0'_{M-1} \\ 1 & 1 & 0'_{M-1} \end{pmatrix}, \\
W_t &= \left(\begin{array}{c|c|c} \delta_t \rho \sigma_\eta a_{s_t} \exp(m_{s_t}/2) & & 0_M \\ & 0_{M+1} & 1 \end{array} \right), \quad T_t = \left(\begin{array}{c|c|c} 0 & & 0'_{M-1} \\ \psi & 0_{M+1} & I_{M-1} \\ & & 0'_{M-1} \end{array} \right), \\
G_t &= \begin{pmatrix} v_{s_t} & 0 & 0 \\ 0 & \sigma_u & 0 \end{pmatrix}, \quad H_t = \begin{pmatrix} \delta_t \rho \sigma_\eta b_{s_t} v_{s_t} \exp(m_{s_t}/2) & 0 & \sigma_\eta \sqrt{1-\rho^2} \\ & 0_M & 0_M \end{pmatrix}.
\end{aligned}$$

For the initial latent log volatility, h_1 , we assume that

$$W_0 = \left(\begin{array}{c|c|c} & & 0 \\ 0_{M+1} & 0_{M+1} & 1_M \end{array} \right), \quad H_0 = (0_{M+1} \quad 0_{M+1} \quad \sigma_\eta 1_{M+1}).$$

for simplicity. If h_t follows an ARFIMA(1,d,0) process, then the coefficient ψ_j is given by

$$\psi_i = \phi^i + \tilde{\gamma}_1 \phi^{i-1} + \dots + \tilde{\gamma}_{i-1} \phi + \gamma_i, \tag{29}$$

where

$$\tilde{\gamma}_0 = 1, \quad \tilde{\gamma}_{j+1} = \frac{j+d}{j+1} \tilde{\gamma}_j, \quad j \geq 0.$$

Conditional on s , we obtain the linear Gaussian state space representation, and can therefore generate samples from the conditional posterior distributions using the simulation smoother and augmented Kalman filter (de Jong (1991)). As we shall see in the next subsection, by integrating out the latent state variables and mean parameter μ using the augmented Kalman filter, we generate posterior samples in a highly efficient way.

¹The state equation expressed by the ARFIMA process contains infinite past disturbance terms. Even if we truncate it using M ($\geq \sqrt{n}$) past disturbances to implement the maximum likelihood estimation, the consistency and asymptotic normality of the maximum likelihood estimator are established for a stationary ARFIMA process (Chan and Palma (1998)).

3.3. MCMC implementation

Let $s = \{s_t\}_{t=1}^n$, $y^* = \{y_t^*\}_{t=1}^n$, $h = \{h_t\}_{t=0}^n$, $\delta = \{\delta_t\}_{t=1}^n$ and set the prior probability densities $\pi(\zeta)$ and $\pi(\varphi)$ for $\zeta = (\phi, \theta, \sigma_\eta^2, d, \rho, \sigma_u^2)'$ and $\varphi = (\xi, \mu)'$. We draw samples from the posterior distribution with probability density $\pi(\zeta, \varphi, h, s|y)$ using the MCMC technique. We summarize the sampling steps as follows.

1. Set the initial value of ζ, φ, s .
2. Generate $s|\zeta, \varphi, h, y^*, \delta$.
3. Generate $(\zeta, \varphi, h)|s, y^*, \delta$.
 - (a) Generate $\zeta|s, y^*, \delta$.
 - (b) Generate $\varphi|\zeta, s, y^*, \delta$.
 - (c) Generate $h|\zeta, \varphi, s, y^*, \delta$.
4. Return to step 2.

We will describe each sampling step in detail below.

Generation of s . The posterior probability mass function of s_t given $\zeta, \varphi, h, y^*, \delta$ is given by

$$\pi(s_t = j|\varphi, \zeta, h, y^*, \delta) \propto \Pr(s_t = j)v_j^{-1} \exp\left\{-\frac{(\epsilon_t^* - m_j)^2}{2v_j^2} - \frac{[\eta_t - \delta_t \rho \sigma_\eta \exp(m_j/2)\{a_j + b_j(\epsilon_t^* - m_j)\}]^2}{2\sigma_\eta^2(1 - \rho^2)}\right\}$$

where

$$\epsilon_t^* = y_{1t}^* - h_t, \quad \eta_t = h_{t+1} - \mu - \phi(h_t - \mu).$$

We can generate a sample from this discrete distribution using the inverse distribution method.

Generation of (ζ, φ, h) . The conditional posterior probability density function of (ζ, φ, h) is

$$\pi(\zeta, \varphi, h|s, y^*, \delta) \propto \pi(\zeta|s, y^*, \delta)\pi(\varphi|\zeta, s, y^*, \delta)\pi(h|\varphi, \zeta, s, y^*, \delta), \quad (30)$$

where

$$\pi(\zeta|s, y^*, \delta) \propto f(y^*|\zeta, s, \delta)\pi(\zeta), \quad (31)$$

and f is the conditional likelihood of the approximated model. We note that the conditional posterior probability density $\pi(\zeta|s, y^*, \delta)$ is marginalized over both h and φ . By implementing the augmented Kalman Filter (de Jong (1991)), we can integrate out (h, φ) to obtain the conditional likelihood $f(y^*|\zeta, s, \delta)$ (see Appendix Appendix B for the details). Using this likelihood and the prior probability density, we use the Metropolis-Hastings algorithm (MH, e.g., Chib and Greenberg (1995)) to generate posterior samples of parameters as follows.

- (a) Generate $\zeta \sim \pi(\zeta|s, y^*, \delta)$.

To generate ζ in the region $R = \{\gamma : |\phi| < 1, |\theta| < 1, \sigma_\eta^2 > 0, -1/2 < d < 1, |\rho| < 1, \sigma_u^2 > 0\}$, we consider the transformation $\zeta^\dagger = (\log(1 + \phi) - \log(1 - \phi), \log(1 + \theta) - \log(1 - \theta), \log(\sigma_\eta^2), \log(1 + 2d) - \log(1 -$

$d), \log(1 + \rho) - \log(1 - \rho), \log(\sigma_u^2))'$. First we compute the mode, $\hat{\zeta}^\dagger$, of the conditional posterior density of ζ^\dagger , $\pi(\zeta^\dagger|s, y^*, \delta)$, numerically, and then construct the proposal density based on the Taylor expansion around the mode:

$$\begin{aligned} \log \pi(\zeta^\dagger|s, y^*, \delta) &\approx \log \pi(\hat{\zeta}^\dagger|s, y^*, \delta) + g'_{\hat{\zeta}^\dagger}(\zeta^\dagger - \hat{\zeta}^\dagger) + \frac{1}{2}(\zeta^\dagger - \hat{\zeta}^\dagger)' g''_{\hat{\zeta}^\dagger}(\zeta^\dagger - \hat{\zeta}^\dagger) \\ &= \text{const.} - \frac{1}{2}(\zeta^\dagger - \mu_\zeta)' \Sigma_\zeta^{-1} (\zeta^\dagger - \mu_\zeta), \\ \Sigma_\zeta^{-1} &= -g_{\zeta^\dagger \zeta^\dagger}, \quad \mu_\zeta = \hat{\zeta}^\dagger + \Sigma_\zeta g_{\zeta^\dagger}, \end{aligned}$$

where

$$g_{\zeta^\dagger} \equiv \left. \frac{d \log \pi(\zeta^\dagger|s, y^*, \delta)}{d\zeta^\dagger} \right|_{\zeta^\dagger = \hat{\zeta}^\dagger}, \quad g_{\zeta^\dagger \zeta^\dagger} \equiv \left. \frac{d^2 \log \pi(\zeta^\dagger|s, y^*, \delta)}{d\zeta^\dagger d\zeta^\dagger} \right|_{\zeta^\dagger = \hat{\zeta}^\dagger}.$$

We generate a candidate $\zeta^\dagger \sim N(\mu_\zeta, \Sigma_\zeta)$ and conduct MH algorithm.

- (b) Given ζ , generate $\varphi = (\xi, \mu)' \sim N(\varphi_1, C_1)$ where φ_1 and C_1 are defined in Appendix Appendix B.
- (c) Given ζ and φ , generate h simultaneously using a simulation smoother by Durbin and Koopman (2002)), which is known to be stable when the dimension of the state vector is high. We generate state disturbances, $\{u_t^*\}$, from the posterior distribution and substitute them into the state equation to obtain h recursively.

4. Illustrative example using simulated data

To illustrate our proposed estimation method, we consider the RSV-LM (1,d,0) model with the true parameters are set equal to

$$\begin{aligned} \phi &= 0.2, \quad \sigma_\eta^2 = \sigma_u^2 = 0.16, \quad d = 0.6, \\ \rho &= -0.4 \quad \mu = 1, \quad \xi = 0, \end{aligned}$$

and generate 2,000 observations. As suggested by Kim et al. (1998) and Omori et al. (2007), we take $y_{1t}^* = \log(y_{1t}^2 + c)$ where $c = 10^{-4}$ is the offset used to handle very small values of y_{1t}^2 . Let Beta(a,b) denote the beta distribution with parameters (a,b) and mean $a/(a+b)$, and let IG(a,b) denote the inverse Gamma distribution with parameters (a,b) and mean $b/(a-1)$. The prior densities are assumed to be as follows:

$$\begin{aligned} \mu &\sim N(0, 1), \quad \frac{1+\phi}{2} \sim \text{Beta}(1, 1), \quad \sigma_\eta^2 \sim \text{IG}(5/2, 0.05/2) \\ \xi &\sim N(0, 1), \quad \frac{1+\rho}{2} \sim \text{Beta}(1, 1), \quad \sigma_u^2 \sim \text{IG}(5/2, 0.05/2), \\ &\quad \frac{1+2d}{3} \sim \text{Beta}(1, 1), \end{aligned} \tag{32}$$

based on previous empirical studies, (e.g., Takahashi et al. (2009)) where we also account for the nonstationary case by setting $-1/2 < d < 1$ as, d is sometimes observed to vary between 0.5 and 0.6 in empirical studies.

The number of iterations in our MCMC implementation is set to 1,500 after 500 samples are discarded as the burn-in period. The number of truncation lags, M , is set to 50.

Table 2: RSV-LM (1,d,0) model for the simulated data

Param.	True	Mean	Stdev	95% interval	IF
ϕ	0.2	0.268	0.083	[0.112, 0.438]	4.9
σ_η^2	0.16	0.146	0.022	[0.104, 0.193]	6.6
d	0.6	0.592	0.029	[0.532, 0.647]	1.7
ρ	-0.4	-0.350	0.033	[-0.414, -0.284]	3.2
σ_u^2	0.16	0.170	0.014	[0.139, 0.198]	6.0
μ	1	0.949	0.121	[0.722, 1.194]	1.6
ξ	0	0.005	0.033	[-0.060, 0.070]	5.9

Table 2 shows the true values, posterior means, posterior standard deviations, posterior 95% credible intervals and inefficiency factors (IF)². The estimation result shows that our estimates are close to the true values of the parameters and that all of the 95% credible intervals include these true values. We note that the inefficiency factors are extremely low (1 ~ 7) in comparison to the values using other approaches (see, e.g., Omori and Watanabe (2008), where the inefficiency factors are 68 ~ 433 for the multi-move sampler and 103 ~ 3507 for the single move sampler). This result demonstrates that our proposed estimation method is highly efficient and that we have successfully extended the work of Omori et al. (2007) to the RSV-LM model without loss of sampling efficiency.

5. Application to S&P500 returns data

5.1. Data

We apply our proposed model to the daily returns and realized volatility (or the realized kernel) of the S&P500 stock index. The sample period is from January 3, 1996 to February 27, 2009, and the number of observations is $n = 3,263$.³ Figure 1 shows a time series plot of the $\log RV$, and the mean level appears to be slowly changing. The mean level began to increase at the lower level and remained near zero for the first half of the sample period. It later began decreasing again but then increased sharply toward the end of second half. Figure 2 shows a autocorrelation function of $\log RV$. The autocorrelation of $\log RV$ maintain high level even in long lags. These behaviors suggests that the logarithm of the realized volatility has the long memory property.

The summary statistics are also shown in Table 3. The distributions of the $\log RV$ and $\log RK$ are much closer to a normal distribution than those of the RV and RK . This indicates that our normality assumption for the error term in (5) is plausible for our empirical analysis, and we defer the extension to a non-normal error distribution to future work.

²The inefficiency factor (IF, see, e.g., Chib (2001)) is defined by $1 + 2 \sum_{s=1}^{\infty} \rho_s$, where ρ_s is the sample autocorrelation at lag s . It measures how well the MCMC chain mixes. It is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. When the inefficiency factor is equal to N , we must draw MCMC samples N times to generate one uncorrelated sample.

³These are obtained from the Oxford-Man Institute's Realized Library (Heber et al. (2009)). The realized kernel is calculated using the method of Barndorff-Nielsen et al. (2008) to account for microstructure noise.

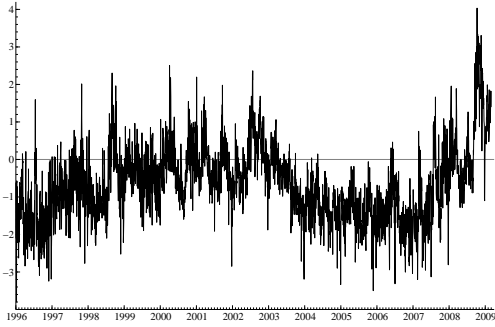


Figure 1: Time series plot of log RV.

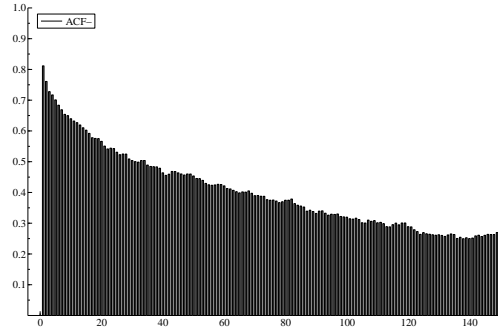


Figure 2: Sample autocorrelation function of log RV.

Table 3: Summary statistics. S&P500 returns.

Variable	Mean	Stdev	Skew	Kurt	Max	Min
R_t	0.005	1.312	-0.258	8.038	10.956	-9.469
RV_t	0.964	2.088	10.981	199.500	56.482	0.022
RK_t	1.002	2.141	10.410	178.260	55.961	0.022
$\log RV_t$	-0.656	0.992	0.556	3.830	4.033	-3.800
$\log RK_t$	-0.621	1.000	0.530	3.777	4.024	-3.791

5.2. Estimation results

We estimate the following eight models:

- RSV model. The prior distributions are the same as in in Section 4 except that we assume $(1 + \phi)/2 \sim \text{Beta}(20,1.5)$ as in the previous literature.
- RSV-SP model. As in Section 2.2, we consider the RSV model with superposition where the log volatility is assumed to be the sum of two independent stationary AR(1) processes with leverage effects (see Omori et al. (2007)):

$$\begin{aligned}
 h_t &= h_{1t} + h_{2t}, \\
 h_{1t} &= \mu + \phi_1(h_{1t} - \mu) + \eta_{1t}, \quad |\phi_1| < 1, \\
 h_{2t} &= \phi_2 h_{2t} + \eta_{2t}, \quad |\phi_2| < 1, \quad \phi_1 > \phi_2, \\
 \begin{pmatrix} \epsilon_t \\ u_t \\ \eta_{1t} \\ \eta_{2t} \end{pmatrix} &\sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho_1 \sigma_{\eta_1} & \rho_2 \sigma_{\eta_2} \\ 0 & \sigma_u^2 & 0 & 0 \\ \rho_1 \sigma_{\eta_1} & 0 & \sigma_{\eta_1}^2 & 0 \\ \rho_2 \sigma_{\eta_2} & 0 & 0 & \sigma_{\eta_2}^2 \end{pmatrix} \right), \quad |\rho_1| < 1, \quad |\rho_2| < 1, \quad \rho_1^2 + \rho_2^2 < 1,
 \end{aligned}$$

The prior distributions for $(\mu, \phi_1, \sigma_{\eta_1}^2, \xi, \rho_1, \sigma_u^2)$ are the same as in the RSV model. For the parameters of the second AR(1) process, we assume $(1 + \phi_2)/2 \sim \text{Beta}(10,10)$, $(1 + \rho_2)/2 \sim \text{Beta}(10,10)$, $\sigma_{\eta_2}^2 \sim \text{IG}(5/2, 0.05/2)$ with side constraints $\phi_2 < \phi_1$ and $\rho_1^2 + \rho_2^2 < 1$ following Omori et al. (2007).

- RSV-LM models: RSV-LM-MA (0,d,0), (1,d,0), (0,d,1) and RSV-LM-AR (0,d,0), (1,d,0), (0,d,1) models. The prior distributions of the parameters are assumed to be the same as in Section 4. For the prior distribution of θ , we assume $(1 + \theta)/2 \sim \text{Beta}(1,1)$.

The number of iterations is set to 1,500 and the initial 500 samples are discarded as the burn-in period, taking account of the inefficiency factors. As in the previous section, these factors are extremely small ($1 \sim 6$),

and our MCMC estimation method is highly efficient. This is because we use the mixture sampler, which integrates out all of the latent volatility variables to compute the conditional likelihood and, furthermore, uses the additional information based on the logarithm of the realized volatilities. We note that these inefficiency factors are even smaller overall than those obtained using the mixture sampler approach for the stochastic volatility model with leverage, without using the realized volatilities (Omori et al. (2007)). We can therefore perform the statistical inference even with a small number of the iterations and the short burn-in period compared to previous studies.

The estimation results for the RSV model are shown in Tables 4 using $\log RV_t$ and $\log RK_t$ for y_{2t} . The estimation results of RSV model are quite similar using $\log RV_t$ and $\log RK_t$. The persistence parameter in the volatility is found to be high ($\phi = 0.965$ (0.965)), and the negative ρ indicates the existence of leverage effects ($\rho = -0.534$ (-0.539)). We note that the posterior probability of a negative bias, ($\xi = -0.625$ (-0.588)), in the realized volatility (or the realized kernel) is greater than 0.975. This implies that the realized volatilities underestimate the integrated volatilities because they do not account for the presence of non-trading hours and microstructure noise. By introducing the term ξ , the bias can be estimated and eliminated.

Hansen and Lunde (2005), for example, propose to correct the bias by computing the scaled realized volatility (SRV). This ensures that the mean of the scaled realized volatilities (or the scaled realized kernels) is equal to the sample variance of the daily returns. This may therefore be used to substantially reduce the bias owing to overnight price changes. However, this is not necessary in our model because the bias itself is estimated and eliminated. This result is consistent with the preceding empirical study by Takahashi et al. (2009).

Table 4: Estimation results of RSV model

log RM	log RV				log RK			
	Mean	Stdev	95%	IF	Mean	Stdev	95%	IF
Parameter								
ϕ	0.965	0.004	[0.956, 0.973]	1.6	0.965	0.004	[0.957, 0.973]	1.6
σ_η^2	0.043	0.003	[0.036, 0.051]	1.7	0.043	0.003	[0.037, 0.051]	1.6
ρ	-0.534	0.036	[-0.601, -0.461]	2.4	-0.539	0.035	[-0.607, -0.466]	2.4
σ_u^2	0.183	0.006	[0.171, 0.196]	1.3	0.189	0.006	[0.176, 0.202]	1.4
μ	0.105	0.096	[-0.079, 0.303]	1.6	0.105	0.097	[-0.081, 0.305]	1.6
ξ	-0.625	0.027	[-0.679, -0.573]	6.0	-0.588	0.027	[-0.643, -0.538]	5.7

Table 5: Estimation results of RSV-SP model

log RM	log RV				log RK			
	Mean	Stdev	95%	IF	Mean	Stdev	95%	IF
Parameter								
ϕ_1	0.983	0.003	[0.977, 0.988]	1.9	0.983	0.003	[0.977, 0.989]	2.3
ϕ_2	0.434	0.062	[0.314, 0.553]	3.2	0.436	0.062	[0.317, 0.553]	3.8
$\sigma_{\eta_1}^2$	0.018	0.002	[0.013, 0.024]	1.4	0.018	0.002	[0.013, 0.024]	1.6
$\sigma_{\eta_2}^2$	0.079	0.010	[0.060, 0.101]	5.8	0.081	0.010	[0.060, 0.103]	5.4
ρ_1	-0.547	0.056	[-0.651, -0.431]	0.9	-0.551	0.056	[-0.655, -0.432]	1.5
ρ_2	-0.220	0.044	[-0.308, -0.141]	3.1	-0.223	0.044	[-0.312, -0.145]	2.9
σ_u^2	0.131	0.010	[0.110, 0.151]	5.4	0.135	0.010	[0.113, 0.156]	4.8
μ	0.219	0.127	[-0.025, 0.483]	2.0	0.219	0.129	[-0.031, 0.485]	2.0
ξ	-0.613	0.025	[-0.664, -0.565]	6.2	-0.576	0.025	[-0.627, -0.528]	5.7

The estimation results for the RSV-SP model are shown in Tables 5. For the RSV-SP model, the estimation results are similar to those of RSV. The posterior mean of the persistence parameter for the second AR(1) process, ϕ_2 , is smaller than that of first one. This indicate the second process, h_{2t} , corresponds to the transitory component, while the first process, h_{1t} , corresponds to the persistent component. The degree of the leverage effect in second AR(1) process is moderate but still seems to exist ($\rho_2 = -0.220$ (-0.223)).

Table 6: Estimation results of RSV-LM-MA and RSV-LM-AR models

log RM	log RV				log RK			
	Mean	Stdev	95%interval	IF	Mean	Stdev	95%interval	IF
RSV-LM-MA								
(0,d,0)								
σ_η^2	0.127	0.010	[0.108, 0.148]	3.4	0.127	0.010	[0.110, 0.150]	3.5
d	0.610	0.016	[0.577, 0.640]	2.0	0.610	0.016	[0.576, 0.640]	2.2
ρ	-0.385	0.030	[-0.443, -0.327]	2.3	-0.389	0.030	[-0.446, -0.332]	2.3
σ_u^2	0.127	0.008	[0.109, 0.143]	3.3	0.132	0.008	[0.114, 0.149]	3.3
μ	0.060	0.082	[-0.095, 0.222]	1.4	0.058	0.082	[-0.097, 0.224]	1.5
ξ	-0.614	0.026	[-0.663, -0.561]	7.6	-0.578	0.026	[-0.628, -0.525]	8.2
(1,d,0)								
ϕ	0.103	0.062	[-0.012, 0.225]	1.6	0.100	0.062	[-0.017, 0.225]	1.7
σ_η^2	0.111	0.012	[0.088, 0.138]	3.0	0.112	0.012	[0.089, 0.139]	3.3
d	0.594	0.020	[0.556, 0.635]	1.4	0.594	0.020	[0.555, 0.636]	1.5
ρ	-0.402	0.031	[-0.468, -0.343]	1.7	-0.405	0.031	[-0.470, -0.345]	2.1
σ_u^2	0.137	0.009	[0.116, 0.154]	2.5	0.141	0.009	[0.120, 0.158]	3.2
μ	0.056	0.080	[-0.100, 0.216]	1.5	0.055	0.079	[-0.098, 0.217]	1.4
ξ	-0.614	0.026	[-0.666, -0.565]	4.9	-0.578	0.026	[-0.630, -0.529]	4.9
(0,d,1)								
θ	-0.095	0.063	[-0.225, 0.019]	2.2	-0.093	0.063	[-0.224, 0.022]	2.6
σ_η^2	0.112	0.012	[0.090, 0.139]	3.0	0.114	0.012	[0.090, 0.141]	4.2
d	0.599	0.018	[0.564, 0.635]	1.0	0.599	0.018	[0.563, 0.638]	1.2
ρ	-0.398	0.029	[-0.460, -0.340]	2.0	-0.402	0.030	[-0.464, -0.344]	2.2
σ_u^2	0.136	0.009	[0.117, 0.153]	3.0	0.140	0.009	[0.121, 0.157]	3.4
μ	0.057	0.079	[-0.100, 0.214]	1.5	0.055	0.080	[-0.100, 0.214]	1.5
ξ	-0.613	0.025	[-0.664, -0.565]	4.6	-0.577	0.026	[-0.628, -0.528]	4.6
RSV-LM-AR								
(0,d,0)								
σ_η^2	0.107	0.009	[0.090, 0.127]	3.1	0.109	0.009	[0.090, 0.129]	3.8
d	0.639	0.021	[0.599, 0.679]	2.9	0.638	0.020	[0.599, 0.679]	3.0
ρ	-0.433	0.033	[-0.499, -0.372]	2.4	-0.437	0.033	[-0.503, -0.376]	2.5
σ_u^2	0.141	0.008	[0.124, 0.158]	2.9	0.146	0.008	[0.128, 0.163]	3.3
μ	0.066	0.161	[-0.249, 0.385]	0.9	0.064	0.160	[-0.253, 0.381]	1.0
ξ	-0.615	0.026	[-0.666, -0.562]	8.8	-0.579	0.026	[-0.629, -0.525]	8.6
(1,d,0)								
ϕ	0.053	0.063	[-0.066, 0.178]	1.5	0.056	0.064	[-0.063, 0.178]	1.7
σ_η^2	0.100	0.011	[0.078, 0.123]	3.3	0.101	0.011	[0.079, 0.125]	3.6
d	0.629	0.023	[0.583, 0.679]	1.8	0.628	0.023	[0.582, 0.677]	1.7
ρ	-0.440	0.033	[-0.505, -0.373]	2.4	-0.445	0.033	[-0.509, -0.377]	2.7
σ_u^2	0.145	0.009	[0.126, 0.163]	3.2	0.150	0.009	[0.130, 0.168]	3.3
μ	0.071	0.155	[-0.242, 0.375]	1.2	0.068	0.155	[-0.245, 0.371]	1.1
ξ	-0.614	0.025	[-0.665, -0.566]	4.9	-0.578	0.025	[-0.631, -0.530]	5.4
(0,d,1)								
θ	-0.053	0.069	[-0.193, 0.068]	2.5	-0.057	0.068	[-0.193, 0.067]	2.9
σ_η^2	0.101	0.012	[0.077, 0.126]	5.3	0.101	0.012	[0.078, 0.127]	6.1
d	0.630	0.023	[0.586, 0.678]	2.7	0.629	0.023	[0.584, 0.678]	2.9
ρ	-0.438	0.033	[-0.503, -0.373]	4.5	-0.443	0.033	[-0.508, -0.375]	5.1
σ_u^2	0.144	0.009	[0.126, 0.162]	4.9	0.150	0.009	[0.129, 0.168]	5.3
μ	0.070	0.155	[-0.237, 0.377]	1.0	0.068	0.155	[-0.245, 0.371]	1.1
ξ	-0.614	0.026	[-0.665, -0.566]	5.0	-0.578	0.026	[-0.630, -0.530]	5.8

For RSV-LM models, we use two different types of models such as AR and MA models. The estimation results for the RSV-LM models are shown in Table 6 (the MA model (26)-(28) and the AR model (A.2)-(A.4)).⁴ The results using the logarithm of RV_t and RK_t are very similar, as in the RSV models. The AR and MA models also give similar results, although σ_η^2 and ρ are slightly smaller and d and σ_u^2 are a slightly larger in the AR models.

The bias adjustment terms ξ are estimated to be negative, and the posterior probability of negative bias is greater than 0.975 as in the RSV models. Furthermore, although the absolute values of the posterior means of the correlation parameter, ρ , are slightly smaller than those of the RSV models, they are nevertheless found to be negative in all models, indicating the existence of leverage effects in the long memory process.

The posterior means of the fractional parameter d are greater than 0.5, which suggests that log volatility process may have the long memory and nonstationary properties. On the other hand, the RSV-LM models are found to have lower values of the persistence parameter compared to the RSV models. For example, the posterior means of ϕ are 0.05 (AR models) \sim 0.10 (MA models) in RSV-LM(1,d,0) models and those of θ in RSV-LM(0,d,1) models are -0.10 (MA models) \sim -0.05 (AR models). The high autoregressive impact (expressed by ϕ) in the log volatility process in the RSV models therefore appears to be replaced by a dependence on the long past disturbance terms (expressed by d) in the RSV-LM models.

5.3. Model comparison

In this section, we perform a model comparison of the RSV, RSV-SP and RSV-LM models based on the logarithm of the marginal likelihood. The marginal likelihood is defined as the integral of the likelihood with respect to the prior density of the parameter(s). Following Chib (1995), we estimate the logarithm of the marginal likelihood, $\log m(y)$, as

$$\log m(y) = \log f(y|\vartheta) + \log \pi(\vartheta) - \log \pi(\vartheta|y), \quad (33)$$

where $\vartheta = (\zeta, \varphi)$ and $\log f(y|\vartheta)$, $\log \pi(\vartheta)$ and $\log \pi(\vartheta|y)$ denote the likelihood, prior density and posterior density. The prior density can be computed in a straightforward manner, but we must evaluate the likelihood and posterior density using a Monte Carlo method for our models. To compute the likelihood, we use the auxiliary particle filter of Pitt and Shephard (1999) with 8,000 particles. We repeat the computation of the particle filter 10 times to calculate the numerical standard errors of the estimated likelihood. The posterior density at ϑ is evaluated using the MCMC method as in Chib (1995) and Chib and Jeliazkov (2001), with the number of the reduced run is set to 1,000. The estimation results are shown in Tables 7 and 8 for the log RV_t and log RK_t .

Similar results are obtained in both cases, using the log RV_t and log RK_t . The RSV-LM models and the RSV-SP model outperform the RSV model with respect to the marginal likelihood. This suggests that the logarithm of the latent volatility process may have long memory properties. Among the RSV-LM and RSV-SP models, the AR model of RSV-LM (RSV-LM-AR) outperforms other models and the RSV-SP model outperforms the MA model of RSV-LM (RSV-LM-MA) overall. The RSV-LM-AR models may be able to capture the long range dependence better than other models through their lagged log volatilities. On the other hand, the marginal likelihoods for the three RSV-LM models (i.e., (0,d,0), (1,d,0), (0,d,1) models) are quite similar taking account of the standard errors, for RSV-LM models using log RV_t and RSV-LM-AR models using log RK_t . The RSV-LM-AR models appear to fit the data during this period equally well, and we could not identify any clear difference among these models. We therefore further compare the above models using another criterion based on the volatility forecasting performance in the next subsection.

Remark 1. We also estimated the RSV model with a multiplicative coefficient, λ , in front of the log volatility h_t (hence, $y_{2t} = \xi + \lambda h_t + u_t$) using the logarithm of RV_t . We found that the posterior mean of λ ($= 0.9516$) is close to one and that the logarithm of its marginal likelihood ($= -7131.6$) is not improved. The data, therefore, do not motivate an extension of the RSV model to include such a multiplicative parameter.

⁴For the truncation lag, $M = 60$ is used in the state space representations for (26)-(28) and (A.2)-(A.4).

Remark 2. When the number of observations is small, the RSV model may outperform the RSV-LM models. To outperform the RSV models with respect to marginal likelihood, we would need to take account of sufficiently large lags in the RSV-LM models.

Table 7: Logarithm of marginal likelihood (using log RV).
Standard errors are in parentheses.

Model	Likelihood	Prior	Posterior	Marg. Lik.	Ranking
RSV	-7107.2 (0.4)	-12.08	10.90 (0.1)	-7130.3 (0.5)	8
RSV-SP	-7061.9 (0.6)	-17.06	13.67 (0.1)	-7092.6 (0.7)	4
RSV-LM-MA					
(0,d,0)	-7088.5 (0.6)	-15.08	12.58 (0.1)	-7116.1 (0.7)	7
(1,d,0)	-7087.7 (0.7)	-15.28	12.26 (0.1)	-7115.2 (0.8)	6
(0,d,1)	-7086.7 (0.8)	-15.26	11.89 (0.1)	-7113.9 (0.9)	5
RSV-LM-AR					
(0,d,0)	-7052.3 (0.6)	-15.02	11.41 (0.1)	-7078.7 (0.7)	1
(1,d,0)	-7052.6 (0.7)	-14.97	12.71 (0.1)	-7080.2 (0.8)	3
(0,d,1)	-7051.7 (0.7)	-14.97	12.59 (0.1)	-7079.2 (0.8)	2

Table 8: Logarithm of marginal likelihood (using log RK).

Model	Likelihood	Prior	Posterior	Marg. Lik.	Ranking
RSV	-7143.0 (0.4)	-12.13	10.81 (0.1)	-7165.9 (0.5)	8
RSV-SP	-7093.3 (0.6)	-17.10	13.70 (0.1)	-7124.1 (0.7)	4
RSV-LM-MA					
(0,d,0)	-7124.6 (0.5)	-15.16	12.56 (0.1)	-7152.3 (0.6)	7
(1,d,0)	-7120.1 (0.5)	-15.36	12.48 (0.1)	-7147.9 (0.6)	6
(0,d,1)	-7118.3 (0.7)	-15.35	11.90 (0.1)	-7145.6 (0.8)	5
RSV-LM-AR					
(0,d,0)	-7088.2 (0.4)	-15.07	11.15 (0.1)	-7114.4 (0.5)	3
(1,d,0)	-7085.8 (0.5)	-15.02	12.71 (0.1)	-7113.5 (0.6)	2
(0,d,1)	-7085.7 (0.5)	-15.02	12.58 (0.1)	-7113.3 (0.6)	1

Model comparison result. S&P500 returns.

Table 9: Bayes factors. The figure are log (base 10) of Bayes factors for the row model against the column model.

Model	MA			SP	AR		
	(0,d,0)	(1,d,0)	(0,d,1)		(0,d,0)	(1,d,0)	(0,d,1)
log RV							
RSV	-6.16	-6.55	-7.12	-16.37	-22.40	-21.75	-22.19
MA(0,d,0)		-0.39	-0.95	-10.20	-16.24	-15.59	-16.02
MA(1,d,0)			-0.56	-9.81	-15.85	-15.20	-15.63
MA(0,d,1)				-9.25	-15.28	-14.63	-15.07
RSV-SP					-6.03	-5.38	-5.81
AR(0,d,0)						0.65	0.21
AR(1,d,0)							-0.43
log RK							
RSV	-5.90	-7.81	-8.81	-18.15	-22.36	-22.75	-22.84
MA(0,d,0)		-1.91	-2.90	-12.24	-16.45	-16.85	-16.93
MA(1,d,0)			-0.99	-10.33	-14.54	-14.93	-15.02
MA(0,d,1)				-9.33	-13.55	-13.94	-14.02
RSV-SP					-4.21	-4.60	-4.69
AR(0,d,0)						-0.39	-0.47
AR(1,d,0)							-0.08

5.4. Volatility forecast performance

In addition to the above model comparison using marginal likelihoods which measures the goodness of fit for the in-sample period, we investigate the predictive performance of our models based on the volatility forecasting for the out-of-sample period. By generating the volatilities from their posterior predictive distribution, we compare the performance of the models with respect to several loss functions as described in Patton (2011) (see, e.g., Poon and Granger (2003) and Andersen et al. (2006) for recent comprehensive reviews of the volatility forecast).

Generation of volatilities from the posterior predictive distribution. Let N denote the number of MCMC iterations used in the parameter estimation, and let $(\theta^{(i)}, \{h_t^{(i)}\}_{t=1}^n)$ denote the posterior sample of $(\theta, \{h_t\}_{t=1}^n)$ at the i -th iteration ($i = 1, \dots, N$). Then, the K -step-ahead volatility forecast is obtained by adding several steps to each MCMC iteration:

1. For each MCMC iteration $i = 1, \dots, N$:
 - (a) Initialise $k = 1$.
 - (b) Generate $h_{n+k}^{(i)} | \{y_t\}_{t=1}^n, \{h_t^{(i)}\}_{t=1}^{n+k-1}, \{y_{n+s}^{(i)}\}_{s=1}^{k-1}, \theta^{(i)}$.
 - (c) Generate $y_{n+k}^{(i)} | \{y_t\}_{t=1}^n, \{h_t^{(i)}\}_{t=1}^{n+k}, \{y_{n+s}^{(i)}\}_{s=1}^{k-1}, \theta^{(i)}$.
 - (d) If $k < K$, let $k + 1 \rightarrow k$ and go to Step (b). Otherwise, save $g_{n+K}^{(i)} \equiv \exp(h_{n+k}^{(i)})$ as a random sample of the K -step-ahead variance from its posterior predictive distribution.
2. The estimate of the conditional variance σ_{n+K}^2 is obtained by

$$\bar{g}_{n+K} = \frac{1}{N} \sum_{i=1}^N g_{n+K}^{(i)}.$$

Robust loss functions based on volatility proxies. Patton (2011) derived the functional form of the loss function for comparing volatility forecasts using imperfect volatility proxies, such that the forecasts are robust to the presence of noise in the proxies. A loss function, L , is called “robust” if the ranking of any two (possibly imperfect) volatility forecasts, g_{1t} , and g_{2t} , by expected loss is the same whether the ranking is performed

using the true conditional variance, σ_t^2 , or some conditionally unbiased volatility proxy, $\hat{\sigma}_t^2$. Patton (2011) showed that RMSE⁵ and QLIKE type loss functions,

$$\text{RMSE} : L(\hat{\sigma}_t^2, g_t) = \sqrt{(\hat{\sigma}_t^2 - g_t)^2}, \quad (34)$$

$$\text{QLIKE} : L(\hat{\sigma}_t^2, g_t) = \frac{\hat{\sigma}_t^2}{g_t} - \log \frac{\hat{\sigma}_t^2}{g_t} - 1, \quad (35)$$

are robust with respect to the forecast error, $\hat{\sigma}_t^2 - g_t$, and standardized forecast error, $\hat{\sigma}_t^2/g_t$ respectively, where g_t denotes a volatility forecast of the conditional variance σ_t^2 . We note that these loss functions are also invariant with respect to a rescaling of the data. Therefore, even if we use a conditionally unbiased volatility proxy (e.g., the realized variances computed using 5 minute returns for liquid stocks and 30 minute returns for less liquid stocks), these loss function can produce a correct ranking of the volatility forecasts.

We use the sum of the realized volatility (or the realized kernel), computed using 5 minute returns and squared overnight returns, as a conditionally unbiased volatility proxy. Further sensitivity analysis using other types of realized measures may be interest because microstructure noise could decrease the predictive accuracy (see, e.g., Andersen et al. (2011), Ghysels and Sinko (2011), Asai et al. (2012), Koopman and Scharth (2013)), but we defer this analysis to future work.

Volatility forecast comparisons. We consider eight models: RSV, RSV-SP, RSV-LM-MA-(1,d,0), RSV-LM-MA-(0,d,1), RSV-LM-MA-(0,d,0), RSV-LM-AR-(1,d,0), RSV-LM-AR-(0,d,1) and RSV-LM-AR-(0,d,0) (using the truncation lag $M = 40$). Two prediction periods are considered: (I) April 30, 2007 – July 31, 2008 (the number of predictions, I , is 318) and (II) April 30, 2007 – February 27, 2009 ($I = 462$), where prediction period II includes the financial crisis of September 2008 with high volatilities. Using a rolling window of $n = 1,500$ (i.e., using the 1,500 most recent observations for each estimation and forecast), we estimate the model parameters and compute the K -step-ahead volatility forecasts ($K = 1, 5, 10$), with the number of MCMC iterations set to $N = 300$.

⁵in Patton (2011), he showed MSE loss function $L = \frac{1}{2}(\hat{\sigma}_t^2 - g_t)^2$ as a robust loss function

Table 10: RMSE of volatility forecasts (30/Apr/2007 – 31/Jul/2008): The highest performance for each period and proxy is shown in bold font.

Volatility proxy Forecast horizon	RV			RK		
	1	5	10	1	5	10
Model						
RSV	0.963	1.231	1.584	1.000	1.268	1.612
RSV-SP	0.881	1.074	1.151	0.922	1.119	1.197
RSV-LM-MA						
(0,d,0)	0.860	1.023	1.109	0.905	1.074	1.160
(1,d,0)	0.879	0.977	1.032	0.925	1.031	1.085
(0,d,1)	0.844	1.009	1.089	0.890	1.061	1.140
RSV-LM-AR						
(0,d,0)	0.905	1.092	1.183	0.946	1.137	1.228
(1,d,0)	0.897	1.073	1.162	0.938	1.120	1.208
(0,d,1)	0.892	1.082	1.170	0.934	1.127	1.215

Table 11: QLIKE of volatility forecasts (30/Apr/2007 – 31/Jul/2008): The highest performance for each period and proxy is shown in bold font.

Volatility proxy Forecast horizon	RV			RK		
	1	5	10	1	5	10
Model						
RSV	0.231	0.333	0.441	0.237	0.342	0.449
RSV-SP	0.205	0.295	0.346	0.213	0.304	0.356
RSV-LM-MA						
(0,d,0)	0.202	0.294	0.354	0.209	0.305	0.367
(1,d,0)	0.204	0.284	0.329	0.211	0.295	0.340
(0,d,1)	0.200	0.291	0.348	0.207	0.301	0.359
RSV-LM-AR						
(0,d,0)	0.211	0.302	0.357	0.217	0.311	0.367
(1,d,0)	0.209	0.300	0.351	0.216	0.309	0.362
(0,d,1)	0.211	0.303	0.355	0.217	0.312	0.365

For prediction period I, Tables 10 and 11 show the Root Mean Square Error (RMSE)⁶ and the values of the QLIKE type loss function. For both the RMSE and the QLIKE function, the RSV-LM models outperform the RSV model, suggesting that modelling the long memory property may improve the accuracy of the volatility forecast. The RSV-SP model, which is also expected to describe the long memory property, outperforms the RSV model for 1, 5, 10-step ahead forecasts, but the RSV-LM-MA models still outperform the RSV-SP model. Among the RSV-LM models, the RSV-LM-MA models outperform the RSV-LM-AR models in contrast to the model comparison results in Section 5.3. The performances of the RSV-LM-MA (0,d,1), (1,d,0) and (0,d,0) models are quite similar. For one-step-ahead volatility forecasts, the RSV-LM-MA (0,d,1) model appears to perform slightly better than other RSV-LM-MA models, whereas, for the longer forecast horizons ($K = 5, 10$), the RSV-LM-MA (1,d,0) model outperforms the other RSV-LM-MA models.

⁶The RMSE is computed as $\sqrt{1/I \sum_{j=1}^I (\hat{\sigma}_{n+K,j}^2 - \bar{g}_{n+K,j})^2}$ where $\hat{\sigma}_{n+K,j}^2$ and $\bar{g}_{n+K,j}$ are the K -step ahead volatility proxy and the estimate of the conditional variance for the j -th prediction ($j = 1, \dots, I$).

Table 12: RMSE of volatility forecasts (30/Apr/2007 – 27/Feb/2009): The highest performance for each period and proxy is shown in bold font.

Volatility proxy Forecast horizon	RV			RK		
	1	5	10	1	5	10
Model						
RSV	2.973	4.190	7.578	2.991	4.197	7.526
RSV-SP	2.998	3.279	3.464	3.004	3.321	3.518
RSV-LM-MA						
(0,d,0)	2.695	2.898	3.373	2.735	3.007	3.491
(1,d,0)	2.686	2.903	3.408	2.730	3.018	3.527
(0,d,1)	2.640	2.867	3.390	2.686	2.979	3.510
RSV-LM-AR						
(0,d,0)	3.189	3.472	3.598	3.185	3.350	3.491
(1,d,0)	3.142	3.368	3.539	3.142	3.398	3.564
(0,d,1)	3.176	3.438	3.647	3.177	3.464	3.672

Table 13: QLIKE of volatility forecasts (30/Apr/2007 – 27/Feb/2009): The highest performance for each period and proxy is shown in bold font.

Volatility proxy Forecast horizon	RV			RK		
	1	5	10	1	5	10
Model						
RSV	0.217	0.341	0.500	0.222	0.347	0.504
RSV-SP	0.199	0.309	0.407	0.204	0.315	0.416
RSV-LM-MA						
(0,d,0)	0.188	0.313	0.474	0.194	0.324	0.488
(1,d,0)	0.189	0.309	0.469	0.195	0.320	0.483
(0,d,1)	0.186	0.312	0.471	0.192	0.323	0.485
RSV-LM-AR						
(0,d,0)	0.207	0.306	0.399	0.211	0.313	0.407
(1,d,0)	0.204	0.304	0.394	0.209	0.311	0.402
(0,d,1)	0.207	0.306	0.402	0.212	0.312	0.409

For prediction period II, Tables 12 and 13 show the values for the RMSE and QLIKE functions. The values for the RMSE are higher than those in period I because they are sensitive to the volatile values of RV_t ⁷. In the case of the RMSE loss function, the RSV-LM-MA (0,d,1) model performs better than other RSV-LM models for one-step-ahead forecasts, but the RSV and RSV-SP models outperform the RSV-LM-AR models. This result appears to be inconsistent with the result below for the QLIKE loss function. However, we must recall that a few outliers can severely deteriorate the values for the RMSE loss function. The result for the QLIKE loss function may be more reliable because the QLIKE loss function is less sensitive to these volatile values. For the longer forecast horizons, the values of the AR and MA models become more similar compared to those for the one-step-ahead forecast, and the RSV-SP model performs as well as the RSV-LM models for the 10-step ahead forecast. In the case of the QLIKE loss function, the RSV-LM models outperform the RSV model, and, among the RSV-LM models, the RSV-LM-MA (0,d,1) model performs better than other RSV-LM models for one-step-ahead volatility forecasts. However, for the longer forecast horizons ($K = 5, 10$), the RSV-LM-AR (1,d,0) model outperform the other models. The RSV-SP model performs as well as RSV-LM-AR models.

Overall, the RSV-LM-MA models show high predictive performance in both periods. In high volatility

⁷We omitted the value of RV_t on October 10, 2008 (56.482, see Figure 1) because it is too large to evaluate the loss functions.

period (period II), RSV-SP and RSV-LM-AR also perform well for the longer forecast horizons. The RSV-LM models can describe the long memory property well via relatively small number of parameters while maintaining the short step ahead predictive performance.

6. Conclusion

In this paper, we propose simultaneous modeling of the daily returns and realized volatility (or realized kernel) including leverage and long memory. The state space representation of the new model is described, and a highly efficient sampling algorithm is proposed to implement the MCMC estimation. We have shown that the biases in the realized variances owing to both non-trading hours and market microstructure noise can be estimated within our modeling framework. In empirical studies, the posterior distribution of the leverage parameter, ρ , supports the presence of the leverage effect for the RSV and RSV-LM models. The estimated fractional parameter, d , of the ARFIMA process suggests that the volatilities have long memory and nonstationary properties. It corresponds to the long range persistence of the realized volatilities and realized kernels.

The RSV-LM models and RSV (RSV and RSV-SP) models are compared based on the marginal likelihood and their volatility forecasts. In the marginal likelihood comparison, if we adopt sufficiently large lags and a high enough number of observations, the RSV-LM-AR models outperform the RSV models. Among the RSV-LM models, the RSV-LM-AR models provide a superior fit to the data compared to the RSV-LM-MA models. In the volatility forecast comparison based on RMSE and QLIKE type loss functions, the RSV-LM models again outperform the RSV models. For the one-step-ahead forecast, the RSV-LM-MA (0,d,1) model outperforms other models, but for the longer forecast horizons, the performance appears to depend on the prediction period.

Acknowledgements

The authors would like to thank Erricos Kontogiorghes, the associate editor and anonymous referees for their helpful comments and suggestions. Furthermore, we appreciate the remarks by Torben Andersen, Tim Bollerslev, and Peter Reinhard Hansen at the 2011 Second International Conference "High-Frequency Data Analysis in Financial Markets" in Osaka, Japan. We are also grateful to Cathy Chen, Mei-Hui Guo, Mike So, Toshiaki Watanabe and Philip Yu for their comments at the 7th IASC-ARS Conference and its satellite meeting in Taiwan and to Haruhisa Nishino for useful discussion. The computational results were generated using the Ox 6.20 software (see Doornik (2007)).

This work is supported by the Research Fellowship (DC1) from the Japan Society for the Promotion of Science and the Grants-in-Aid for Scientific Research (A) 21243018 from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

Appendix

Appendix A. RSV-LM-AR model

Since

$$\Theta(L)^{-1}\Phi(L)(1-L)^d(h_{t+1}-\mu) = \eta_t,$$

we let

$$\begin{aligned} h_{t+1} &= \mu + \sum_{j=1}^{\infty} \psi_j(h_{t-j+1} - \mu) + \eta_t \\ &\approx \left(1 - \sum_{j=1}^M \psi_j\right)\mu + \sum_{j=1}^M \psi_j h_{t-j+1} + \eta_t, \quad \psi \equiv (\psi_1, \dots, \psi_M)'. \end{aligned} \quad (\text{A.1})$$

Then the RSV-LM-AR model is given by

$$y_t^* = X_t\beta + Z_t\alpha_t + G_t u_t^*, \quad t = 1, 2, \dots, n, \quad (\text{A.2})$$

$$\alpha_{t+1} = W_t\beta + T_t\alpha_t + H_t u_t^*, \quad t = 0, 1, \dots, n-1, \quad (\text{A.3})$$

$$\alpha_0 \equiv 0, \quad u_t^* \sim \text{i.i.d. } N(0_3, I_3), \quad (\text{A.4})$$

where $\alpha_t = (h_t, h_{t-1}, \dots, h_{t-M+1})'$ and

$$\begin{aligned} X_t &= \begin{pmatrix} m_{s_t} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 \\ \xi \\ \mu \end{pmatrix}, \quad Z_t = \begin{pmatrix} 1 & 0'_{M-1} \\ 1 & 0'_{M-1} \end{pmatrix}, \\ W_t &= \left(\begin{array}{c|c} \delta_t \rho \sigma_\eta a_{s_t} \exp(m_{s_t}/2) & \\ \hline 0_{M-1} & \end{array} \middle| \begin{array}{c} 0_M \\ \hline 1 - \sum_{j=1}^M \psi_j \\ \hline 0_{M-1} \end{array} \right), \quad T_t = \begin{pmatrix} \psi' \\ I_{M-1} & 0_{M-1} \end{pmatrix}, \\ G_t &= \begin{pmatrix} v_t & 0 & 0 \\ 0 & \sigma_u & 0 \end{pmatrix}, \quad H_t = \begin{pmatrix} \delta_t \rho \sigma_\eta b_{s_t} v_{s_t} \exp(m_{s_t}/2) & 0 & \sigma_\eta \sqrt{1-\rho^2} \\ \hline 0_{M-1} & 0_{M-1} & 0_{M-1} \end{pmatrix}. \end{aligned}$$

For the initial latent log volatility, h_1 , we assume

$$W_0 = (0_M \mid 0_M \mid 1_M), \quad H_0 = \begin{pmatrix} 0 & 0 & \sigma_\eta \\ 0_{M-1} & 0_{M-1} & 1_{M-1} \end{pmatrix}.$$

for simplicity. If h_t follows ARFIMA(1,d,0) process, the coefficient ψ_j is given by

$$\psi_1 = \phi + d, \quad \psi_{j+1} = \tilde{\gamma}_{j+1} - \phi \tilde{\gamma}_j, \quad j \geq 0, \quad (\text{A.5})$$

where

$$\tilde{\gamma}_1 = d, \quad \tilde{\gamma}_{j+1} = \frac{j-d}{j+1} \tilde{\gamma}_j, \quad j \geq 0,$$

or, using $(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-L)^k$,

$$\psi_j = (-1)^{j+1} \left\{ \frac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)} + \phi \frac{\Gamma(d+1)}{\Gamma(j)\Gamma(d-j+2)} \right\}, \quad j = 1, 2, \dots, M.$$

Appendix B. Augmented Kalman filter

We first implement Kalman filter to compute the logarithm of the likelihood given φ, ζ, y^*, s and δ .

1. Set $a_1 = W_0\beta$, $P_1 = H_0H_0'$.
2. For $t = 1, 2, \dots, n$, compute

$$a_{t+1} = W_t\beta + T_t a_t + K_t e_t, \quad P_{t+1} = T_t P_t L_t' + H_t J_t' \quad (\text{B.1})$$

and

$$\begin{aligned} e_t &= y_t^* - X_t\beta - Z_t a_t, \quad D_t = Z_t P_t Z_t' + G_t G_t', \\ K_t &= (T_t P_t Z_t' + H_t G_t') D_t^{-1}, \quad L_t = T_t - K_t Z_t, \quad J_t = H_t - K_t G_t, \end{aligned}$$

where e_t 's are independent predictive errors and $e_t \sim N(0, D_t)$. Then the log likelihood is given by

$$\log f(y^*|\varphi, \zeta, s, \delta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n \log |D_t| - \frac{1}{2} \sum_{t=1}^n e_t' D_t^{-1} e_t. \quad (\text{B.2})$$

Next we apply the augmented Kalman Filter to compute the log likelihood marginalized over φ (see de Jong (1991), Nakajima and Omori (2009)). Noting that $\beta = b + B\varphi$ where $\varphi = (\xi, \mu)'$ and

$$b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},$$

we implement the augmented Kalman Filter as follows.

1. Set $a_1^* = W_0 b$, $A_1^* = -W_0 B$.
2. For $t = 1, 2, \dots, n$, compute

$$\begin{aligned} f_t &= y_t^* - X_t b - Z_t a_t^*, \quad a_{t+1}^* = W_t b + T_t a_t^* + K_t f_t, \\ F_t &= X_t B - Z_t A_t^*, \quad A_{t+1}^* = -W_t B + T_t A_t^* + K_t F_t. \end{aligned} \quad (\text{B.3})$$

Since the log likelihood given (φ, ζ) is given by

$$\begin{aligned} \log f(y^*|\varphi, \zeta, s, \delta) &= -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n \log |D_t| \\ &\quad - \frac{1}{2} \left\{ \sum_{t=1}^n f_t' D_t^{-1} f_t - 2 \left(\sum_{t=1}^n F_t' D_t^{-1} f_t \right)' \varphi + \varphi' \left(\sum_{t=1}^n F_t' D_t^{-1} F_t \right) \varphi \right\}, \end{aligned}$$

the conditional posterior distribution of φ given ζ, s, δ, y^* is $\varphi|\zeta, s, \delta, y^* \sim N(\varphi_1, C_1)$, where

$$\begin{aligned} C_1 &= \left(C_0^{-1} + \sum_{t=1}^n F_t' D_t^{-1} F_t \right)^{-1} \\ \varphi_1 &= C_1 \left(C_0^{-1} \varphi_0 + \sum_{t=1}^n F_t' D_t^{-1} f_t \right). \end{aligned}$$

where φ_0, C_0 are the mean and variance of the normal prior for φ . We set each of them as follows.

$$\varphi_0 = (0, 0)', \quad C_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Thus the log likelihood marginalized over φ is computed as

$$\begin{aligned}\log f(y^*|\zeta, s, \delta) &\equiv \log f(y^*|\varphi, \zeta, s, \delta) + \log \pi(\varphi|\zeta, s, \delta) - \log \pi(\varphi|\zeta, s, \delta, y^*) \\ &= -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n \log |D_t| - \frac{1}{2} \log |C_0| + \frac{1}{2} \log |C_1| \\ &\quad - \frac{1}{2} \left\{ \sum_{t=1}^n f_t' D_t^{-1} f_t + \varphi_0' C_0^{-1} \varphi_0 - \varphi_1' C_1^{-1} \varphi_1 \right\}.\end{aligned}\tag{B.4}$$

Appendix C. Auxiliary particle filter

Appendix C.1. RSV model

We describe the auxiliary particle filter (e.g. Pitt and Shephard (1999)) to calculate the log likelihood for the RSV model:

$$y_{1,t} = \exp(h_t/2)\epsilon_t, \quad t = 1, 2, \dots, n \tag{C.1}$$

$$y_{2,t} = \xi + h_t + u_t, \quad t = 1, 2, \dots, n \tag{C.2}$$

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

$$\begin{pmatrix} \epsilon_t \\ u_t \\ \eta_t \end{pmatrix} \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{\Omega}), \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{\Omega} = \begin{pmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{pmatrix} \tag{C.4}$$

Let $\vartheta = (\phi, \sigma_\eta^2, \rho, \xi, \sigma_u^2)'$. Then the conditional probability density functions of the observation equation and the state equation are

$$f(y_t|h_t, \vartheta) = \frac{1}{2\pi\sigma_u} \exp\left\{-\frac{1}{2}h_t - \frac{1}{2}y_{1,t}^2 \exp(-h_t) - \frac{1}{2\sigma_u^2}(y_{2,t} - \xi - h_t)^2\right\}, \tag{C.5}$$

$$f(h_{t+1}|y_t, h_t, \mu, \vartheta) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_\eta} \exp\left(-\frac{(h_{t+1} - \mu_{t+1})^2}{2(1-\rho^2)\sigma_\eta^2}\right), \tag{C.6}$$

$$\text{where } \mu_{t+1} = \mu + \phi(h_t - \mu) + \rho\sigma_\eta \exp(-h_t/2)y_t.$$

We generate i -th particle ($i = 1, \dots, I$) using an importance function given by

$$\begin{aligned}g(h_{t+1}, h_t^i|y_{t+1}, \mu, \vartheta) &\propto f(y_{t+1}|\mu_{t+1}^i, \vartheta) f(h_{t+1}|y_t, h_t^i, \mu, \vartheta) \hat{f}(h_t^i|y_t, \mu, \vartheta) \\ &\propto f(h_{t+1}|y_t, h_t^i, \mu, \vartheta) g(h_t^i|y_{t+1}, \mu, \vartheta),\end{aligned}$$

where

$$\begin{aligned}g(h_t^i|y_{t+1}, \mu, \vartheta) &= \frac{f(y_{t+1}|\mu_{t+1}^i, \vartheta) \hat{f}(h_t^i|y_t, \mu, \vartheta)}{\sum_{j=1}^I f(y_{t+1}|\mu_{t+1}^j, \vartheta) \hat{f}(h_t^j|y_t, \mu, \vartheta)}, \\ f(y_{t+1}|\mu_{t+1}^i, \vartheta) &= \frac{1}{2\pi\sigma_u} \exp\left\{-\frac{1}{2}\mu_{t+1}^i - \frac{1}{2}y_{1,t+1}^2 \exp(-\mu_{t+1}^i) - \frac{1}{2\sigma_u^2}(y_{2,t+1} - \xi - \mu_{t+1}^i)^2\right\}, \\ \mu_{t+1}^i &= \mu + \phi(h_t^i - \mu) + \rho\sigma_\eta \exp(-h_t^i/2)y_t.\end{aligned}$$

To compute the likelihood via the auxiliary particle filter,

1. For $t = 1$, generate $h_1^i \sim N(\mu, \sigma^2/(1-\phi^2))$, $i = 1, 2, \dots, I$.

- (a) Calculate $w_i = f(y_1|h_1^i, \vartheta)$ and $W_i = F(y_1|h_1^i, \vartheta)$, (where F is the distribution function of h_t given y_t), and record

$$\bar{w}_1 = \frac{1}{I} \sum_{i=1}^I w_i, \quad \bar{W}_1 = \frac{1}{I} \sum_{i=1}^I W_i. \tag{C.7}$$

(b) Let $\hat{f}(h_t^i|y_1, \mu, \vartheta) = \pi_1^i = w_i / \sum_{j=1}^I w_j$, $i = 1, 2, \dots, I$.

2. Generate (h_t^i, h_{t+1}^i) ($i = 1, \dots, I$) using $g(h_{t+1}, h_t|y_{t+1}, \mu, \vartheta)$.

(a) Calculate

$$w_i = \frac{f(y_{t+1}|h_{t+1}^i, \vartheta)f(h_{t+1}^i|y_t, h_t^i, \mu, \vartheta)\hat{f}(h_t^i|y_t, \mu, \vartheta)}{g(h_{t+1}^i, h_t^i|y_{t+1}, \mu, \vartheta)} = \frac{f(y_{t+1}|h_{t+1}^i, \vartheta)\hat{f}(h_t^i|y_t, \mu, \vartheta)}{g(h_t^i|y_{t+1}, \mu, \vartheta)}, \quad (\text{C.8})$$

$$W_i = \frac{F(y_{t+1}|h_{t+1}^i, \vartheta)\hat{f}(h_t^i|y_t, \mu, \vartheta)}{g(h_t^i|y_{t+1}, \mu, \vartheta)}, \quad i = 1, \dots, I, \quad (\text{C.9})$$

and record

$$\bar{w}_t = \frac{1}{I} \sum_{i=1}^I w_i, \quad \bar{W}_t = \frac{1}{I} \sum_{i=1}^I W_i. \quad (\text{C.10})$$

(b) Let $\hat{f}(h_{t+1}^i|y_{t+1}, \mu, \vartheta) = \pi_{t+1}^i = w_i / \sum_{j=1}^I w_j$, $i = 1, 2, \dots, I$.

3. If $t < n$, let $t \leftarrow t + 1$ and return to Step 2.

As $I \rightarrow \infty$, $\bar{w}_{t+1} \xrightarrow{p} f(y_{t+1}|y_t, \mu, \vartheta)$ and $\bar{W}_{t+1} \xrightarrow{p} F(y_{t+1}|y_t, \mu, \vartheta)$. Thus $\sum_{t=1}^n \log \bar{w}_t$ is the consistent estimator of $\sum_{t=1}^n \log f(y_t|y_{t-1}, \mu, \vartheta)$. The sequence of \bar{W}_t and its reflected version $2|\bar{W}_t - 1/2|$ can be used to check for model fit as these are approximately i.i.d. standard uniform if the model is correctly specified.

Appendix C.2. RSV-LM model

RSV-LM model given by

$$y_{1,t} = \exp(h_t/2)\epsilon_t, \quad t = 1, 2, \dots, n \quad (\text{C.11})$$

$$y_{2,t} = \xi + h_t + u_t, \quad t = 1, 2, \dots, n \quad (\text{C.12})$$

$$\Theta(L)^{-1}\Phi(L)(1-L)^d(h_{t+1} - \mu) = \eta_t, \quad t = 0, 1, \dots, n \quad (\text{C.13})$$

$$\begin{pmatrix} \epsilon_t \\ u_t \\ \eta_t \end{pmatrix} \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{\Omega}), \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{\Omega} = \begin{pmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{pmatrix}. \quad (\text{C.14})$$

For RSV-LM-MA model, we use an MA representation of the state equation:

$$h_{t+1} = \mu + \sum_{j=0}^M \psi_j \eta_{t-j} = \mu + \psi' \nu_t + \eta_t, \quad \nu_t = (\eta_{t-1}, \dots, \eta_{t-M})'. \quad (\text{C.15})$$

To implement the auxiliary particle filter, we generate i -th particle given $\nu_t^i = (\eta_{t-1}^i, \dots, \eta_{t-M}^i)'$ using the importance function as in the previous subsection where

$$\mu_{t+1}^i = \mu + \rho\sigma \exp(-h_t^i/2)y_t + \psi' \nu_t^i. \quad (\text{C.16})$$

We can calculate the weight (C.8) by using the conditional probability density functions of the observation and the state equation (C.5), (C.6).

For RSV-LM-AR models, we use a AR representation of the state equation:

$$h_{t+1} = \left(1 - \sum_{j=1}^M \psi_j\right)\mu + \sum_{j=1}^M \psi_j h_{t-j+1} + \eta_t = \left(1 - \sum_{j=1}^M \psi_j\right)\mu + \psi' \nu_t + \eta_t, \\ \nu_t = (h_t, \dots, h_{t-M+1})'. \quad (\text{C.17})$$

We generate i -th particle given $\nu_t^i = (h_t^i, \dots, h_{t-M+1}^i)'$ using the importance function in the previous subsection where

$$\mu_{t+1}^i = \left(1 - \sum_{j=1}^M \psi_j\right)\mu + \rho\sigma \exp(-h_t^i/2)y_t + \psi' \nu_t^i. \quad (\text{C.18})$$

References

- Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006). Volatility and correlation forecasting. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*. Amsterdam: North Holland Press.
- Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics* 89, 701–720.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Andersen, T. G., T. Bollerslev, and N. Meddahi (2004). Analytic evaluation of volatility forecasts. *International Economic Review* 45, 1079–1110.
- Andersen, T. G., T. Bollerslev, and N. Meddahi (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* 160, 220–234.
- Asai, M., M. McAleer, and M. C. Medeiros (2012). Modelling and forecasting noisy realized volatility. *Computational Statistics and Data Analysis* 56, 217–230.
- At-Sahalia, Y., P. 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, 351–416.
- Bandi, F. M. and J. R. Russell (2008). Microstructure noise, realized volatility, and optimal sampling. *Review of Economic Studies* 75, 339–364.
- 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, 1481–1536.
- Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B* 63, 167–241.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253–280.
- Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 177–181.
- Breidt, F. J., N. Crato, and P. D. Lima (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 325–348.
- Chan, N. H. and W. Palma (1998). State space modeling of long-memory processes. *Annals of Statistics* 26, 719–740.
- Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association* 90, 1313–1321.
- Chib, S. (2001). Markov chain monte carlo methods: computation and inference. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North Holland Press.
- Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327–335.
- Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* 96, 270–281.
- de Jong, P. (1991). The diffuse kalman filter. *Annals of Statistics* 19, 1073–1083.

- Dobrev, D. P. and P. J. Szerszen (2010). The information content of high-frequency data for estimating equity return models and forecasting risk. Discussion Paper.
- Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox, 5th ed.* London: Timberlake Consultants Press and Oxford. www.doornik.com.
- Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89, 603–616.
- Ghysels, E. and A. Sinko (2011). Volatility forecasting and microstructure noise. *Journal of Econometrics* 160, 257–271.
- Giot, P. and S. Laurent (2004). Modeling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance* 11, 379–398.
- Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27, 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, 873–889.
- Harvey, A. C. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, 429–434.
- Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics and econometrics of securities trading.* Oxford: Oxford University Press.
- Heber, G., A. Lunde, N. Shephard, and K. Sheppard (2009). *Oxford-Man Institute's realized library, version 0.1.* Oxford-Man Institute, University of Oxford.
- Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility with fat-tails and correlated errors. *Journal of Econometrics* 122, 185–212.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with arch models. *Review of Economic Studies* 65, 361–393.
- Koopman, S. J., B. Jungbacker, and E. Hol (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445–475.
- Koopman, S. J. and M. Scharth (2013). The analysis of stochastic volatility in the presence of daily realized measures. *Journal of Financial Econometrics*, 76–115.
- Maheu, J. M. and T. H. McCurdy (2007). Components of market risk and return. *Journal of Financial Econometrics* 5, 560–590.
- Maheu, J. M. and T. H. McCurdy (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics* 160, 69–76.
- Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics* 17, 479–508.
- Melino, A. and S. Turnbull (1990). The pricing of foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–265.
- Nakajima, J. and Y. Omori (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics and Data Analysis* 53, 2335–2353.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica* 59, 347–370.
- O'Hara, M. (1995). *Market Microstructure Theory.* Blackwell.
- Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics* 140-2, 425–449.
- Omori, Y. and T. Watanabe (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis* 52-6, 2892–2910.

- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.
- Pérez, A., E. Ruiz, and H. Veiga (2009). A note on the properties of power-transformed returns in long-memory stochastic volatility models with leverage effect. *Computational Statistics and Data Analysis* 53, 3593–3600.
- Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association* 94, 590–599.
- Poon, S. H. and C. W. J. Granger (2003). Forecasting volatility in financial markets: a review. *Journal of Economic Literature* 41, 478–539.
- Raggi, D. and S. Bordignon (2012). Long memory and nonlinearities in realized volatility: A markov switching approach. *Computational Statistics and Data Analysis* 56, 3730–3742.
- Ruiz, E. and H. Veiga (2008). Modeling long-memory volatilities with leverage effect: A-LMSV versus FIEGARCH. *Computational Statistics and Data Analysis* 52, 2846–2862.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- So, M. K. P. (1999). Time series with additive noise. *Biometrika* 86, 474–482.
- So, M. K. P. (2002). Bayesian analysis of long memory stochastic volatility models. *Computational Statistics and Data Analysis* 64, 1–10.
- Takahashi, M., Y. Omori, and T. Watanabe (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics and Data Analysis* 53-6, 2404–2426.
- Venter, J. H. and P. J. de Jongh (2013). Extended stochastic volatility models incorporating realized measures. *Computational Statistics and Data Analysis* forthcoming.
- Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165–178.
- Zhang, L., P. A. Mykland, and T. At-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association* 100, 1394–1411.