Article

Estimation of Realized Asymmetric Stochastic Volatility Models Using Kalman Filter

Manabu Asai

Faculty of Economics, Soka University, Tokyo 192-8577, Japan; m-asai@soka.ac.jp

Abstract: Despite the growing interest in realized stochastic volatility models, their estimation techniques, such as simulated maximum likelihood (SML), are computationally intensive. Based on the realized volatility equation, this study demonstrates that, in a finite sample, the quasi-maximum likelihood estimator based on the Kalman filter is competitive with the two-step SML estimator, which is less efficient than the SML estimator. Regarding empirical results for the S&P 500 index, the quasi-likelihood ratio tests favored the two-factor realized asymmetric stochastic volatility model with the standardized t distribution among alternative specifications, and an analysis on out-of-sample forecasts prefers the realized stochastic volatility models, rejecting the model without the realized volatility measure. Furthermore, the forecasts of alternative RSV models are statistically equivalent for the data covering the global financial crisis.

Keywords: realized volatility; stochastic volatility; asymmetry; heavy-tailed distribution; quasi-maximum likelihood estimation

JEL Classification: C2; C22

1. Introduction

Over the last two decades, research on realized volatility has received significant attention in modeling and forecasting the volatility of financial returns. For the generalized autoregressive conditional heteroskedasticity (GARCH) class models, Engle and Gallo (2006) and Shephard and Sheppard (2010) incorporated realized volatility for modeling and forecasting volatility. Using the information of return and realized volatility measure simultaneously, Hansen et al. (2012) and Hansen and Huang (2016) developed the "realized GARCH" and "realized exponential GARCH" models, respectively.

The literature on stochastic volatility models considers a realized volatility measure to be an estimate of latent volatility. As highlighted by Barndorff-Nielsen and Shephard (2002), there is a gap between true volatility and its consistent estimate, referred to as the "realized volatility error". Since the aforementioned error is nonnegligible, Barndorff-Nielsen and Shephard (2002); Bollerslev and Zhou (2006); Takahashi et al. (2009), and Asai et al. (2012a, 2012b) accommodate a homoscedastic disturbance as an ad hoc approach. Analogous to the realized GARCH model, Takahashi et al. (2009) suggested the realized stochastic volatility (RSV) model, which is based on the information of return and realized volatility measure.

As in the GARCH model, it is useful to accommodate asymmetric effects and heavy-tailed conditional distributions in stochastic volatility models. For the former, a typical approach is to assume a negative correlation between the return and the disturbance for the one-step-ahead log-volatility (see Harvey and Shephard 1996; Yu 2005), known as the leverage effect. Instead of the standard normal distribution for the conditional distributions for return, we may consider the standardized t distribution and the generalized error distribution. Harvey et al. (1994), Sandmann and Koopman (1998), Liesenfeld and Jung (2000), and Asai (2008, 2009), among others, assumed the (standardized) t distribution. Furthermore, the empirical results in Liesenfeld and Jung (2000) and Asai (2009) indicate

that the standardized t distribution yields better fits than the generalized error distribution. By the statistical property of the stochastic volatility models, we can analytically obtain the fourth moment of the return series.

For estimating various RSV models, Koopman and Scharth (2013) and Shirota et al. (2014) used the simulated maximum likelihood (SML) estimation and the Bayesian Markov chain Monte Carlo (MCMC) technique, respectively. Both approaches are computationally demanding. This study reconsiders the quasi-maximum likelihood (QML) method of Harvey et al. (1994) as it is straightforward to include the realized volatility equation, and it is expected to contribute toward improving the efficiency of the QML estimator.

Interest in modeling volatility using the information on return and realized volatility measure is simultaneously growing. Extending the class of the GARCH, Hansen et al. (2012) and Hansen and Huang (2016) developed the "realized GARCH" and the "realized exponential GARCH" models, respectively. Using the stochastic volatility (SV) models (e.g., see Chib et al. 2009), Takahashi et al. (2009) and Koopman and Scharth (2013) considered the "realized SV" (RSV) and the "realized asymmetric SV" (RSV-A) models, respectively.

While the realized GARCH class models can be estimated by the maximum likelihood estimation (MLE) technique, RSV class models require computationally demanding techniques. While Takahashi et al. (2009) and Shirota et al. (2014) suggested the Bayesian MCMC method, Koopman and Scharth (2013) developed the SML technique. Note that Koopman and Scharth (2013) suggested a two-step SML (2SML) estimator that is less efficient but less computationally intensive than the SML estimator. The current study applies the QML estimation Harvey et al. (1994) to the RSV models to show the practical usefulness.

The remainder of this paper is organized as follows. The RSV-A model with the standardized t distribution (RSVt-A) is outlined in Section 2. The asymptotic property of the QML estimator using the Kalman filter is discussed in Section 3, and its finite sample properties are examined and compared with that of the 2SML method of Koopman and Scharth (2013). The empirical results for the Standard and Poor's (S&P) 500 index using the return and realized volatility measure are reported in Section 4. Finally, concluding remarks are presented in Section 5.

2. Realized Stochastic Volatility Models

2.1. Model

Let y_t and x_t denote open-to-close return of a financial asset and the log of a realized measure of its volatility on day t, respectively. The pair of close-to-close return and the realized volatility measure accommodating overnight volatility can be used as in Koopman and Scharth (2013).

Consider the realized asymmetric SV model with the standardized t distribution as follows:

$$y_t = z_t \exp\left(\frac{1}{2}h_t\right), \quad z_t = \frac{\varepsilon_t}{\sqrt{w_t/(\nu-2)}}, \quad w_t \sim \chi^2(\nu), \tag{1}$$

$$h_t = c + \alpha_t, \quad \alpha_{t+1} = \phi\alpha_t + \eta_t, \tag{2}$$

$$x_t = \xi + h_t + u_t, \tag{3}$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \\ u_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_\eta & 0 \\ \rho\sigma_\eta & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix}\right), \tag{4}$$

where c, ϕ, σ_η, ρ, ν, ξ, and σ_u are parameters. According to the structure, z_t follows a standardized t distribution with the degree-of-freedom parameter ν. Note that $E(z_t) = 0$, $E(z_t^2) = 1$, $E(z_t^3) = 0$, and $E(z_t^4) = 3(\nu-2)/(\nu-4) > 3$. To guarantee the stationarity of the process and the existence of the fourth moment, we assume $|\phi| < 1$ and $\nu > 4$, respectively. As ρ is the correlation coefficient between ε_t and η_t, it satisfies $|\rho| < 1$. If the realized volatility measure, x_t, is a consistent estimate of h_t, ξ is expected to be zero. By

specification, non-zero ξ implies the (finite sample) bias in x_t. We denote the model (1)–(4) as the "RSVt-A". The model reduces to the asymmetric SV model with the standardized t distribution when we omit Equation (3) (see Harvey and Shephard 1996; Asai 2008). Without the leverage effect, that is, $\rho = 0$, we obtain the RSVt model. By setting $\nu \to \infty$, the model reduces to the RSV-A model. As in Asai (2008) and Koopman and Scharth (2013), we can consider multi-factor models by allowing multiple factors in log-volatility as $h_t = c + \sum_{i=1}^{m} \alpha_{it}$.

Following Harvey et al. (1994) and Harvey and Shephard (1996), we obtain the state space for the RSVt-A model. The logarithmic transformation of the squared y_t yields

$$
\begin{pmatrix} \log y_t^2 \\ x_t \end{pmatrix} = \begin{pmatrix} c + \mu_{\log z^2} \\ c + \xi \end{pmatrix} + \iota \alpha_t + \begin{pmatrix} \zeta_t \\ u_t \end{pmatrix}, \tag{5}
$$

where ι is the 2×1 vector of ones, $\mu_{\log z^2} = E(\log z_t^2)$, and $\zeta_t = \log z_t^2 - \mu_{\log z^2}$. By Equation (26.3.46) in Abramovits and Stegun (1970), we obtain $\mu_{\log z^2} = \psi(1/2) - \psi(\nu/2) + \log(\nu - 2)$ and $V(\log z_t^2) = V(\zeta_t) = \psi'(1/2) + \psi'(\nu/2)$, where $\psi(x)$ is the digamma function defined by $\psi(x) = \frac{d \log \Gamma(x)}{dx}$.

By the transformation $\log y_t^2$, we lose the information of the sign of y_t. To recover such information, we define the sign of y_t as $s_t = I(y_t > 0) - I(y_t \le 0)$, where $I(A)$ is the indicator function, which takes one if the condition A holds, and zero otherwise. As in Harvey and Shephard (1996), we can modify Equation (2) as follows:

$$
\alpha_{t+1} = \phi \alpha_t + \eta_t^*, \tag{6}
$$

with

$$
E\left[\begin{pmatrix} \zeta_t \\ \eta_t^* \\ u_t \end{pmatrix} \middle| s_t \right] = \begin{pmatrix} 0 \\ a s_t \\ 0 \end{pmatrix}, \quad V\left[\begin{pmatrix} \zeta_t \\ \eta_t^* \\ u_t \end{pmatrix} \middle| s_t \right] = \begin{pmatrix} \sigma_\zeta^2 & b s_t & 0 \\ b s_t & \sigma_\eta^2 - a^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}, \tag{7}
$$

where $\sigma_\zeta^2 = \psi'(1/2) + \psi'(\nu/2)$, $a = E(\eta_t | s_t = 1) = \rho \sigma_\eta \sqrt{2/\pi} = 0.7979 \rho \sigma_\eta$, and $b = \rho \sigma_\eta E(|\varepsilon_t| \log \varepsilon_t^2) - \rho \sigma_\eta \sqrt{2/\pi} E(\log \varepsilon_t^2) = 1.1061 \rho \sigma_\eta$. Hence, the measurement in Equation (5) and the transition in Equation (6) form the state-space model. By applying the Kalman filter to the state-space form in (5) and (6), it is straightforward to construct the quasi-log-likelihood function (see Appendix A).

As in Koopman and Scharth (2013), the assumption on the covariance matrix in Equation (4) can be relaxed by considering dependence between conditional return and measurement noise. This modification requires a change in Equation (3) using s_t to construct the state-space model, as in Equations (6) and (7).

2.2. Realized Kernel Estimator

As a realized volatility measure, we adopt the realized kernel (RK) estimator developed by Barndorff-Nielsen et al. (2008), since it is a consistent estimator of the quadratic variation and is robust to microstructure noise and jumps. This subsection explains the RK estimator concisely.

Consider that the latent log-price p^* follows a Brownian semimartingale plus jump process given by

$$
p_t^* = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + J_t,
$$

where μ_t is a predictable locally bounded drift; σ_t is a càdlàg process of volatility; W_t is a process of Brownian motion; and $J_t = \sum_{i=1}^{N_t} C_i$ is a finite activity jump process, which has a finite number of jumps in any bounded interval of time. More precisely, N_t counts the number of jumps that have occurred in the interval $[0, t]$ and $N_t < \infty$ for any t.

The quadratic variation of p^* is given by

$$[p^*] = \int_0^\tau \sigma_s^2 ds + \sum_{i=1}^{N_t} C_i^2,$$

where $\int_0^\tau \sigma_s^2 ds$ is the integrated variance. The estimator of Barndorff-Nielsen et al. (2008) for the quadratic variation is based on the noisy observation of log-price, $p_t = p_t^* + v_t$ ($t = \tau_0, \tau_1, \ldots, \tau_n$) with $\tau_0 = 0$ and $\tau_n = \tau$, where $E(v_t) = 0$ and $\text{Var}(v_t) = \omega^2$.

Barndorff-Nielsen et al. (2008) suggested a non-negative estimator that takes the following form:

$$K(p) = \sum_{-G}^{G} k\left(\frac{g}{G+1}\right)\gamma_g, \quad \gamma_g = \sum_{j=|g|+1}^{n} \hat{y}_j \hat{y}_{j-|g|},$$

where $\hat{y}_j$ is the jth high-frequency return calculated over the interval $[\tau_{j-1}, \tau_j]$, and $k(x)$ is a kernel weight function. For practical purposes, Barndorff-Nielsen et al. (2009) focused on the Parzen kernel function, defined by

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

Barndorff-Nielsen et al. (2008, 2009) provided an estimator for the bandwidth of G. Under mild regularity conditions, Barndorff-Nielsen et al. (2008) demonstrated that $K(p)$ converges to $[p^*]$ in probability, as $n \to \infty$.

For practical purposes, it is convenient to work with the log of the RK estimator for its stability. Even though the RK estimator is consistent, it has finite sample bias and noise, and Equation (3) accommodates the constant term and the disturbance.

3. QML Estimation via Kalman Filter

3.1. QML Estimation

Define $\boldsymbol{\theta} = (c, \phi, \sigma_\eta^2, \rho, \nu, \xi, \sigma_u^2)'$. For the state-space form (5) and (6), applying the Kalman filtering algorithm produces the quasi-log-likelihood function:

$$L(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \quad l_t(\boldsymbol{\theta}) = -\ln(2\pi) - \frac{1}{2}\ln|F_t| - \frac{1}{2}v_t' F_t^{-1} v_t.$$

where v_t (2×1) and F_t (2×2) can be obtained as described in Appendix A. Maximizing the quasi-log-likelihood derives the QML estimator, $\hat{\boldsymbol{\theta}}$. Although the vector of errors has a non-Gaussian distribution, the state-space form has the martingale property and at least the finite fourth moment. Based on the results in Dunsmuir (1979), it is straightforward to demonstrate the consistency and the asymptotic normality of the QML estimator, as follows:

$$\text{plim } \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0, \quad \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, C(\boldsymbol{\theta}_0)),$$

where $\boldsymbol{\theta}_0$ is the vector of true parameters. The covariance matrix $C(\boldsymbol{\theta}_0)$ is equivalent to that of the MLE based on the Whittle likelihood. Using the equivalence, we can demonstrate that the quasi-likelihood ratio (QLR) statistic has an asymptotic χ^2 distribution under the null hypothesis (e.g., see the proof of Theorem 3.1.3 in Taniguchi and Kakizawa 2000).

Koopman and Scharth (2013) developed the SML approach and its two-step version, while Durbin and Koopman (2001) suggested a general approach for obtaining the simulated likelihood function for non-Gaussian state-space models and Koopman and Scharth (2013) applied it to the RSVA-*t* model. For the SML method, the likelihood function can be arbitrarily approximated precisely by decomposing it into a Gaussian part, constructed with a Kalman filter, and a remainder function, for which the expectation is evaluated through simulation. The decomposition reduces the computational time, but the SML is still

computer-intensive since it requires a *T*-dimensional Monte Carlo integration for the latent variables whenever the log-likelihood function is evaluated. To avoid this problem, Koopman and Scharth (2013) developed the 2SML estimator. The first-step estimation provides estimates via the Kalman filter for the state-space model consisting of Equations (2) and (3), while the second maximizes the remaining part of the likelihood function to obtain the remaining parameter and to correct the bias caused by neglecting (1) in the first step (see Appendix B for details).

Compared with the SML estimator, the 2SML and QML estimators are less efficient owing to the fluctuations caused by the asymptotic variance of the first-step estimator and the non-Gaussianity in ζ_t in Equation (5), respectively. Hence, the inefficiency of these two estimators derives from different reasons, and the finite sample properties of the 2SML and QML estimators are worth examining. Note that the QML estimation is faster than the 2SML method, since the former has no additional step using the simulated quantity. Recently, Asai et al. (2017) developed a two-step estimation method based on the Whittle likelihood for the RSV-A model. While their approach requires two steps, the QML estimation using the Kalman filter needs no additional step, implying that the two-step Whittle likelihood estimator is less efficient than the QML estimator.

3.2. Finite Sample Property of QML Estimator

In this subsection, we conducted a Monte Carlo experiment to investigate the finite sample properties of the QML estimator, with a comparison with the 2SML estimator. Although the QML and 2SML estimator are asymptotically less efficient than the SML estimator, these estimation methods are computationally faster than the SML approach. Hence, it is worth comparing the QML and 2SML estimators. The experiments follow the framework in Koopman and Scharth (2013) with the two data-generating processes (DGP) based on Equations (1)–(4). The first DGP was the RSV-A model with the true parameters reported in Table 1, while the second was the RSV*t* model based on the true parameters in Table 2. The sample size was $T = 2500$. While Koopman and Scharth (2013) set the number of replications as 250 for the computationally intensive SML method, we considered 2000 replications for fast estimation procedures, that is, the QML and 2SML methods. For details regarding the 2SML estimation, see Appendix B. All the experiments were run on MATLAB R2022b, using the interior-point algorithm and starting from the true parameter values.

Table 1. Monte Carlo results for QML and 2SML estimators for RSV-A.

Parameter	True	QML			2SML		
		Mean	Std. Dev.	RMSE/$|\theta_i|$	Mean	Std. Dev.	RMSE/$|\theta_i|$
ϕ	0.98	0.9786	(0.0042)	[0.0045]	0.9784	(0.0045)	[0.0048]
σ_η^2	0.05	0.0501	(0.0034)	[0.0675]	0.0501	(0.0035)	[0.0703]
ξ	0.10	0.1002	(0.0444)	[0.4442]	0.0928	(0.0290)	[0.2988]
σ_u^2	0.05	0.0500	(0.0027)	[0.0545]	0.0500	(0.0029)	[0.0572]
c	0.40	0.3998	(0.2021)	[0.5055]	0.4092	(0.2216)	[0.5545]
ρ	-0.30	-0.3020	(0.0298)	[0.0994]	-0.2999	(0.0280)	[0.0932]

Table 2. Monte Carlo results for QML and 2SML estimators for RSV*t*.

Parameter	True	QML			2SML		
		Mean	Std. Dev.	RMSE/$|\theta_i|$	Mean	Std. Dev.	RMSE/$|\theta_i|$
ϕ	0.98	0.9786	(0.0044)	[0.0048]	0.9787	(0.0045)	[0.0047]
σ_η^2	0.05	0.0500	(0.0033)	[0.0653]	0.0500	(0.0033)	[0.0658]
ξ	0.10	0.0899	(0.0645)	[0.6523]	0.0584	(0.4538)	[4.5541]
σ_u^2	0.05	0.0500	(0.0028)	[0.0564]	0.0500	(0.0028)	[0.0569]
c	0.40	0.4022	(0.2268)	[0.5671]	0.4289	(0.5024)	[1.2579]
ν	10.00	10.365	(4.0983)	[0.4114]	10.547	(0.8452)	[0.0998]

Regarding the QML and 2SML estimates for the RSV-A model, Table 1 reports the sample means, standard deviations, and root mean squared errors (RMSEs) divided by the absolute value of the corresponding true parameters. As presented in Table 1, the sample mean of the QML estimates were close to the true value, implying that the finite sample biases are negligible. The RMSE/$|\theta_i|$ ($i = 1, \ldots, 6$) takes a values less than 0.51. The finite sample biases for the 2SML estimator are negligible, except for c. The bias is caused by the fluctuations in the first-step estimator, and it will disappear as sample size increases. Compared with the result of the 2SML estimator, the values of the QML estimator are close to those corresponding to the former except for ξ. The QML estimator for ξ has a higher value of RMSE/$|\theta_i|$, implying an inefficiency when estimating the constant term in the volatility equation caused by including the $\log y_t^2$ equation. Regarding c and ρ, the QML estimator has smaller standard deviations and RMSEs, implying an inefficiency caused by the first-step estimate on 2SML.

The simulation results for the RSVt model are reported in Table 2, implying that the finite sample biases are negligible. The results of RMSE/$|\theta_i|$ for the QML estimator indicate that the QML and 2SML estimators are competitive. Owing to the inefficiency caused by the two-step estimation and bias correction, the 2SML estimator has greater values for RMSE/$|\theta_i|$ for (c, ξ). On the other hand, the 2SML estimator has a smaller value of RMSE/$|\theta_i|$ for ν, indicating the inefficiency caused by approximating the log of the squared t variable as a normal distribution on the QML estimation.

The 2SML method estimates ($\phi, \sigma_\eta^2, \xi^*, \sigma_u^2$) using the information of x_t in the first step, while the QML estimation uses that of ($\log(y_t^2), x_t, s_t$) to obtain the estimates of all parameters. These estimations are based on the Kalman filtering algorithm and common parameters are ($\phi, \sigma_\eta^2, \sigma_u^2$). Hence, it is possible to examine the contributions of $\log(y_t^2)$ for estimating ($\phi, \sigma_\eta^2, \sigma_u^2$) via the Kalman filter. According to Tables 1 and 2, the differences are negligible for ($\phi, \sigma_\eta^2, \sigma_u^2$), indicating that the contribution of $\log(y_t^2)$ in the state-space form is negligible for estimating these parameters. The difference in the variance between ζ_t and u_t supports the results, 4.93 and 0.05, respectively.

The Monte Carlo results imply that the QML estimator based on the Kalman filter is competitive with the 2SML estimator in Koopman and Scharth (2013). Owing to computational simplicity, the QML estimator is a fast and useful alternative to the 2SML estimator.

4. Empirical Analysis

4.1. Estimation Results

To estimate alternative RSV models using the QML method based on the Kalman filter, we use a daily return and realized volatility measure for the S&P 500 index. For the realized volatility measure, we selected the RK estimator in Barndorff-Nielsen et al. (2008) as it is robust to microstructure noise and jumps, as explained above. As the realized volatility is calculated using intraday data, the open-to-close return is used for y_t, as in Hansen et al. (2012). The data are obtained from the Oxford Man Institute of Quantitative Finance, and the sample period is from 22 December 2005 and to 4 December 2017, giving 3000 observations. The first $T = 2500$ observations are used for estimating parameters, and the remaining $F = 500$ are reserved for forecasting. The descriptive statistics for the whole sample are presented in Table 3. The standardized variable was calculated using $y_t \exp(-0.5x_t)$, as x_t is the log of RK. The return and RK have heavy tails, whereas the kurtoses of x_t and the standardized variable are close to three. Compared with the return, the standardized variable is close to the Gaussian distribution.

Table 3. Descriptive statistics for S&P 500.

Data	Mean	Std. Dev.	Skewness	Kurtosis
Return	0.0222	1.3138	−0.2989	14.419
RK	1.0535	8.5195	14.260	359.30
log(RK)	−0.8240	1.3799	0.5697	3.5821
Std. Var.	0.1316	1.1649	−0.0053	2.6944

Note: The standardized variable is calculated by dividing the return by the square root of the RK.

This section compares six models: SV, RSV, RSV-A, RSVt, RSVt-A, and two-factor RSVt-A (2fRSVt-A). Among these, the last model is defined by Equations (1) and (3), with

$$h_t = c + \alpha_{1t} + \alpha_{2t}, \quad \alpha_{1,t+1} = \phi\alpha_{1t} + \eta_{1t}, \quad \alpha_{2,t+1} = \phi_2\alpha_{2t} + \eta_{2t}, \tag{8}$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_{1t} \\ \eta_{2t} \\ u_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_\eta & \rho_2\sigma_{\eta,2} & 0 \\ \rho\sigma_\eta & \sigma_\eta^2 & 0 & 0 \\ \rho_2\sigma_{\eta,2} & 0 & \sigma_{\eta,2}^2 & 0 \\ 0 & 0 & 0 & \sigma_u^2 \end{bmatrix} \right). \tag{9}$$

As discussed in the previous section, the model comparison is based on the QLR test.

The QML estimates for the six models are reported in Table 4, indicating that all parameters are significant at the five percent level. The estimates of (c, ϕ, σ_η^2) for the SV model are typical values in empirical analysis. For the RSV model, the estimates of (c, ϕ, σ_η^2) are similar to those of the SV model. As discussed in Section 3.2, the contribution of the $\log y_t^2$ equation is negligible for estimating (ϕ, σ_η^2) in the RSV model. In other words, the finite sample bias of the QML estimator for (ϕ, σ_η^2) in the SV model is corrected in the RSV model owing to the contribution of the realized volatility equation to construct the RSV model. As the estimate of ϕ decreases, the value of σ_η^2 increases, keeping the variance of α_t, $\sigma_\eta^2/(1 - \phi^2)$ at a similar level. The estimate of ξ is negative and significant, which may be caused by finite sample bias in the RK estimates. Note that it is inappropriate to compare the quasi-log-likelihood of the SV and RSV models, since the former excludes the information of x_t. The estimates for the RSV-A model are close to those of the RSV model. The estimate of ρ is negative and significant, implying the existence of the leverage effect. The QLR test rejects the null hypothesis $\rho = 0$. For the RSVt model, the estimate of ν is 11.2. In contrast, the QLR test failed to reject the null hypothesis of the Gaussian distribution, $\nu = \infty$. As implied by the Monte Carlo results in Table 2, there is an inefficiency on estimating ν, which may derive an ambiguous result for the inference on ν. Note that the descriptive statistics for the standardized variable in Table 3 support the results of the QLR tests. The QLR tests in Table 4 indicate that the RSVt-A is preferred to the RSV and RSVt model. For the 2fRSVt-A model, the estimates of ϕ and $|\rho|$ in the first factor are larger than the corresponding values in the second factor. The QLR test rejects the null hypothesis of the one factor model; the tests selected the 2fRSVt-A model among the six models.

Table 4. QML estimates via the Kalman filter for S&P 500.

Parameter	SV	RSV	RSV-A	RSVt	RSVt-A	2fRSVt-A
c	−0.4605	−0.4588	−0.3243	−0.3843	−0.2946	−0.2113
	(0.0045)	(0.0029)	(0.0023)	(0.0033)	(0.0028)	(0.0029)
ϕ	0.9820	0.9539	0.9583	0.9542	0.9583	0.9714
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
σ_η^2	0.0411	0.0989	0.0761	0.0982	0.0760	0.0482
	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
ρ			−0.6034		−0.6048	−0.5737
			(0.0007)		(0.0007)	(0.0009)
ϕ_2						0.2188
						(0.0015)

Table 4. *Cont.*

Parameter	SV	RSV	RSV-A	RSV*t*	RSV*t*-A	2fRSV*t*-A
$\sigma^2_{\eta,2}$						0.2128
						(0.0005)
ρ_2						−0.1216
						(0.0006)
ξ		−0.1807	−0.1927	−0.2553	−0.2207	−0.1950
		(0.0009)	(0.0009)	(0.0018)	(0.0017)	(0.0011)
σ^2_u		0.1567	0.1839	0.1572	0.1840	0.0026
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0005)
ν				15.0751	37.8286	102.1949
				(0.2884)	(1.9073)	(6.6828)
QLogLike	−5734.4	−7756.8	−7641.5	−7756.3	−7641.4	−7590.8
H_0			$\rho = 0$	$\nu = \infty$	$\rho = 0$	1 factor
QLR test			230.61	0.9698	229.80	101.27
			[0.0000]	[0.3247]	[0.0000]	[0.0000]

Note: Standard errors are in parentheses. *p*-values are in brackets.

4.2. Forecasting Performance

We compare out-of-sample forecasts of the SV and the five RSV models. For these six models, the Kalman filter prediction for the state-space form in (5) and (6) provides the one-step-ahead forecast, $\hat{x}_{T+1}$, and $\hat{\sigma}^2_{T+1} = \exp(\hat{x}_{T+1})$ is a forecast of the quadratic variation for day $T + 1$. By updating the parameter estimates, we calculate the forecasts $\hat{\sigma}^2_{T+j}$ ($j = 1, \ldots, F$) using the rolling window of the model for recent $T = 2500$ observations. An alternative forecast can be considered as follows. Under the Gaussianity of ζ_t, the distribution of α_{t+1} conditional on the past observations is $N(a_{t+1}, P_{t+1})$, where a_t and P_t are defined in Appendix A. Then, the conditional distribution of $\sigma^2_{t+1} = \exp(\xi + c + \alpha_{t+1})$ follows the log-normal distribution, which gives the conditional mean $\exp(\xi + c + a_{t+1} + 0.5P_{t+1})$. Define $\hat{\sigma}^{2*}_{T+1} = \exp(\hat{x}_{T+1} + 0.5P_{T+1})$ where $\hat{x}_{t+1} = \hat{\xi} + \hat{c} + a_{t+1}$. Then, $\hat{\sigma}^{2*}_{T+1}$ can be a forecast as an alternative to $\hat{\sigma}^2_{T+1}$.

For comparison, we obtain forecasts based on the RSV, RSV-A, and RSV*t* models using the 2SML method, as explained in Koopman and Scharth (2013) For the RSV-A model, we first compute the Kalman filter prediction of x_{T+1} and subsequently add the leverage effect $E(\eta_t|\varepsilon_t) = \rho\sigma_\eta\varepsilon_t$ by calculating $\hat{\varepsilon}_t = y_t \times E(\exp(-0.5h_t)|x_1, \ldots, x_t)$. Since the RSV and RSV*t* have no leverage effect, they are free from the latter part.

For comparing the out-of-sample forecasts, Patton (2011) suggested an approach using imperfect volatility proxies. Patton (2011) examined the functional form of the loss function for comparing volatility forecasts, such that the forecasts are robust to the presence of noise in the proxies. According to the definition in Patton (2011), a loss function is "robust" if the ranking of any two forecasts of the co-volatility matrix, $\hat{\sigma}^{2(1)}_{T+j}$ and $\hat{\sigma}^{2(2)}_{T+j}$, by expected loss is the same whether the ranking is performed using the true covariance matrix or an unbiased volatility proxy, $\check{\sigma}^2_t$. Patton (2011) demonstrated that squared forecast error and quasi-likelihood type loss functions, defined by

$$\text{MSFE} : LF^m_{i,T+j} = LF^m(\check{\sigma}^2_{T+j}, \hat{\sigma}^{2(i)}_{T+j}) = \left(\check{\sigma}^2_{T+j} - \hat{\sigma}^{2(i)}_{T+j}\right)^2, \quad (10)$$

$$\text{QLIKE} : LF^q_{i,T+j} = LF^q(\check{\sigma}^2_{T+j}, \hat{\sigma}^{2(i)}_{T+j}) = \frac{\check{\sigma}^2_{T+j}}{\hat{\sigma}^{2(i)}_{T+j}} + \log\hat{\sigma}^{2(i)}_{T+j}, \quad (11)$$

are robust to the forecast error $\check{\sigma}^2_{T+j} - \hat{\sigma}^{2(i)}_{T+j}$ and the standardized forecast error $\check{\sigma}^2_{T+j}/\hat{\sigma}^{2(i)}_{T+j}$, respectively.

For these loss functions in (10) and (11), the MCS procedure in Hansen et al. (2011) enables us to determine the set of models, $\mathcal{M}^*$, that consist of the best model(s) from a collection of models, $\mathcal{M}_0$. Define $d_{ik,t} = LF_{i,t} - LF_{k,t}$ and $\bar{d}_{ik} = F^{-1}\sum_{j=1}^{F} d_{ik,T+j}$ as the

difference in the loss functions of two competitive models and its sample mean, respectively. Under the null hypothesis $H_0 : E(d_{ik,t}) = 0$ $(i > k, \forall i, k \in \mathcal{M}_k)$, Hansen et al. (2011) considered two kinds of test statistics:

$$t_R = \max_{i,k \in \mathcal{M}_k} \left| \frac{\bar{d}_{ik}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d}_{ik})}} \right|, \qquad t_{SQ} = \sum_{i,k \in \mathcal{M}_k, i>k} \left(\frac{\bar{d}_{ik}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d}_{ik})}} \right)^2, \tag{12}$$

where $\widehat{\mathrm{Var}}(\bar{d}_{ik})$ is a bootstrap estimate of the variance of $\bar{d}_{ik}$, and the *p*-values of the test statistics are determined using a bootstrap approach. If the null hypothesis is rejected at a given confidence level, the worst performing model is excluded (rejection is determined on the basis of bootstrap *p*-values under the null hypothesis). Such a model is identified as follows:

$$i = \arg \max_{i \in \mathcal{M}_k} \left(\sum_{k \in \mathcal{M}_k} \bar{d}_{ik} \right) \left\{ \widehat{\mathrm{Var}} \left(\sum_{k \in \mathcal{M}_k} \bar{d}_{ik} \right) \right\}^{-1/2}, \tag{13}$$

where the variance is computed using a bootstrap method.

The means of the MSFE and QLIKE, defined by (10) and (11), for the six models for the two volatility forecasts, that is, $\hat{\sigma}^2_{T+j}$ and $\hat{\sigma}^{2*}_{T+j}$ $(j = 1, \ldots, F)$, based on the QML estimation are presented in Table 5, which reports three models for the 2SML method. Generally, the Kalman filtering prediction, $\hat{\sigma}^2_{T+j}$, has a smaller MSFE, while adjusted forecast, $\hat{\sigma}^{2*}_{T+j}$, has the smaller QLIKE. MSFE selects the 2fRSV-*t* (QML) based on the Kalman filtering prediction, while QLIKE chooses the simple RSV (QML) using the adjusted value. The *p*-values of the MCS for the best model based on the t_R statistic are presented in brackets in Table 5. We omitted the results for t_{SQ} as these are similar. The differences in model performance are explained using the *p*-values. The forecast made by the SV model is significantly different from those of alternative RSV models for both loss functions. Among the RSV models, the differences may be negligible for the datasets. In general, the QML method produces better forecasts than the 2SML estimations, but the differences are statistically insignificant.

Table 5. Out-of-Sample Forecast Evaluation.

	MSFE		**QLIKE**	
Model	$\hat{\sigma}^2_{T+1}$	$\hat{\sigma}^{2*}_{T+1}$	$\hat{\sigma}^2_{T+1}$	$\hat{\sigma}^{2*}_{T+1}$
SV (QML)	5.7010 [0.000]	9.0803 [0.000]	1.0754 [0.000]	1.2688 [0.000]
RSV (QML)	0.0816 [0.750]	0.0817 [0.750]	−0.5990 [0.337]	**−0.6043** [1.000]
RSV-A (QML)	0.0800 [0.967]	0.0813 [0.750]	−0.5937 [0.337]	−0.5978 [0.337]
RSV*t* (QML)	0.0817 [0.750]	0.0817 [0.750]	−0.5989 [0.337]	−0.6042 [0.426]
RSV*t*-A (QML)	0.0800 [0.967]	0.0812 [0.750]	−0.5938 [0.337]	−0.5979 [0.426]
2fRSV*t*-A (QML)	**0.0799** [1.000]	0.0841 [0.750]	−0.5905 [0.337]	−0.5956 [0.337]
RSV (2SML)	0.0829 [0.750]	0.0824 [0.750]	−0.5964 [0.337]	−0.6017 [0.426]
RSV-A (2SML)	0.0815 [0.750]	0.0821 [0.750]	−0.5904 [0.242]	−0.5972 [0.337]
RSV*t* (2SML)	0.0829 [0.750]	0.0824 [0.750]	−0.5964 [0.337]	−0.6017 [0.426]

Note: The table reports the means of MSFE and QLIKE, defined by (10) and (11), respectively. $\hat{\sigma}^2_{T+1}$ and $\hat{\sigma}^{2*}_{T+1}$ are the one-step-ahead volatility forecast and its adjusted value based on the log-normal assumption. The minimum values for MSFE and QLIKE are highlighted in bold. The values in brackets are the *p*-value obtained via the model confidence set (MCS) procedure (Equations (12) and (13)) on the 18 sets of forecasts for the two loss functions.

The out-of-sample forecast performance indicates that the data prefer the RSV models to the SV model. For the datasets, there are no statistical differences among the two forecasts of the six RSV models. The data contain the period of the global financial crisis caused by the collapse of the Lehman Brothers, starting from 15 September 2008. The statistical indifference among the RSV models may be caused by the effects of turbulence in the data.

5. Conclusions

This study examined the QML method using the Kalman filter for RSV*t*-A models. The Monte Carlo experiments reveal that the finite sample property of the QML estimator is competitive with the 2SML estimator in Koopman and Scharth (2013). The QML estimation is useful for its computational speed and simplicity. The empirical results for the S&P 500 index indicate that the 2fRSV*t*-A model is preferred over alternative RSV models, while the analysis of the out-of-sample forecasts favors the RSV models rejecting the simple SV model. Furthermore, the forecasts of alternative RSV models are statistically equivalent to the simple RSV for the data covering the global financial crisis.

Compared with the SML, 2SML, and Bayesian MCMC methods, the computational load of the QML estimation is negligible and useful for practical purposes. There are several directions for extending the current research. First, we can consider estimating a multivariate model with dynamic correlations, extending the work of Asai and McAleer (2009). Second, we can develop the QML technique for the long-memory volatility model, as in Shirota et al. (2014). Third, it is straightforward to include multiple components for the volatility equation, as in Engle and Gallo (2006). We leave such tasks for future research.

Data Availability Statement: The data were obtained from the Oxford Man Institute of Quantitative Finance on 29 June 2022.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

2fRSV*t*-A	Two-factor realized asymmetric stochastic volatility with standardized *t* distribution
2SML	Two-step simulated maximum likelihood
GARCH	Generalized autoregressive conditional heteroskedasticity
MCMC	Markov chain Monte Carlo
MSFE	Mean squared forecast error
QLIKE	Quasi-likelihood
QLR	Quasi-likelihood ratio
QML	Quasi-maximum likelihood
RK	Realized kernel
RSV	Realized stochastic volatility
RSV-A	Realized asymmetric stochastic volatility
RSV*t*	Realized stochastic volatility with standardized *t* distribution
RSV*t*-A	Realized asymmetric stochastic volatility with standardized *t* distribution
S&P	Standard and Poor's
SML	Simulated maximum likelihood
SV	Stochastic volatility

Appendix A. Kalman Filtering and Smoothing

Consider a linear state-space model for $m \times 1$ vector y_t:

$$y_t = c + Z\alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t),$$
$$\alpha_{t+1} = d_t + T\alpha_t + R\eta_t, \quad \eta_t \sim N(0, Q_t), \quad t = 1, \dots, T,$$

where $\alpha_1 \sim N(a_1, P_1)$. The Kalman filter recursion is given by

$$v_t = y_t - c - Z\alpha_t, \qquad\qquad F_t = ZP_tZ' + H_t,$$
$$K_t = TP_tZ'F_t^{-1}, \qquad\qquad L_t = T - K_tZ,$$
$$a_{t+1} = d_t + Ta_t + K_tv_t, \qquad\qquad P_{t+1} = TP_tL_t' + RQ_tR'.$$

Note that $a_{t+1} = E(\alpha_{t+1}|y_1, \ldots, y_t; \boldsymbol{\theta})$ and $P_{t+1} = \text{Var}(\alpha_{t+1}|y_1, \ldots, y_t; \boldsymbol{\theta})$. Since $v_t \sim N(0, F_t)$, we obtain the log-likelihood function as follows:

$$L(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \quad l_t(\boldsymbol{\theta}) = -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln|F_t| - \frac{1}{2}v_t'F_t^{-1}v_t.$$

If the distribution of ε_t is non-Gaussian, $L(\boldsymbol{\theta})$ becomes the quasi-log-likelihood function.

We can compute the smoothed estimate, $\hat{\alpha}_t = E(\alpha_t|y_1, \ldots, y_T; \boldsymbol{\theta})$, and the corresponding covariance matrix, $V_t = \text{Var}(\alpha_t|y_1, \ldots, y_t; \boldsymbol{\theta})$, using the backward state-smoothing equations:

$$r_{t-1} = Z'F_t^{-1}v_t + L_t'r_t, \qquad\qquad N_{t-1} = Z'F_{t-1}Z + L_t'N_tL_t,$$
$$\hat{\alpha}_t = a_t + P_tr_{t-1}, \qquad\qquad V_t = P_t - P_tN_{t-1}P_t,$$

with the starting values $r_T = 0$ and $N_T = O$.

de Jong (1989) developed the deletion-smoothing algorithm to obtain the estimate $\grave{\alpha}_t = E(\alpha_t|\boldsymbol{y}_{\setminus t}; \boldsymbol{\theta})$ and the associated covariance matrix $\check{V}_t = \text{Var}(\alpha_t|\boldsymbol{y}_{\setminus t}; \boldsymbol{\theta})$, where $\boldsymbol{y}_{\setminus t}$ is the interpolation set $\{y_1, \ldots, y_{t-1}, y_{t+1}, \ldots, y_T\}$. Define forward equations as follows:

$$w_t = F_t^{-1}v_t - K_t'r_t, \quad W_t = F_t^{-1} + K_t'N_tK_t, \quad M_t = L_tN_tK_t - Z'F_t^{-1}. \tag{A1}$$

As demonstrated by Theorem 5 in de Jong (1989), we obtain the following:

$$\grave{\alpha}_t = \hat{\alpha}_t + P_tM_tW_t^{-1}w_t, \quad \check{V}_t = \hat{V}_t + P_tM_tW_t^{-1}M_t'P_t. \tag{A2}$$

We can obtain the deletion-smoothing estimate of $\alpha_t^* = (\alpha_t, \eta_t)'$ by redefining the state space model,

$$\begin{aligned} y_t &= c + Z^*\alpha_t^* + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t), \\ \alpha_{t+1}^* &= d_t^* + T^*\alpha_t^* + R^*\eta_t^*, \quad \eta_t^* \sim N(0, Q_t), \end{aligned} \tag{A3}$$

where

$$d_t^* = \begin{bmatrix} d_t \\ O_{q\times 1} \end{bmatrix}, \quad Z^* = \begin{bmatrix} Z & O_{m\times q} \end{bmatrix}, \quad T^* = \begin{bmatrix} T & R \\ O_{p\times r} & O_{q\times q} \end{bmatrix}, \quad R^* = \begin{bmatrix} O_{r\times q} \\ I_q \end{bmatrix},$$

to apply the deletion-soothing algorithms in (A1) and (A2) to the model in (A3).

Appendix B. Two-Step SML (2SML) Estimation

Appendix B.1. Framework

Let $\boldsymbol{y} = (y_1, \ldots, y_T)$, $\boldsymbol{x} = (x_1, \ldots, x_T)$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$. For the vector of unknown parameters, define $\boldsymbol{\psi} = (\boldsymbol{\psi}_y', \boldsymbol{\psi}_x^{*\prime}, \boldsymbol{\psi}_\alpha')'$, where $\boldsymbol{\psi}_y = (c, \rho, \nu)'$, $\boldsymbol{\psi}_x^* = (\xi^*, \sigma_u^2)'$, $\boldsymbol{\psi}_\alpha = (\phi, \sigma_\eta^2)$ with $\xi^* = c + \xi$. The density of $(\boldsymbol{y}, \boldsymbol{x})$ is expressed as follows:

$$p(\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\psi}) = \int p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha}; \boldsymbol{\psi}) d\boldsymbol{\alpha}$$

$$= \int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\alpha}; \boldsymbol{\psi}_y) p(\boldsymbol{x}|\boldsymbol{\alpha}; \boldsymbol{\psi}_x^*) p(\boldsymbol{\alpha}; \boldsymbol{\psi}_\alpha) d\boldsymbol{\alpha} \tag{A4}$$

$$= \int \prod_{t=1}^T p(y_t|x_t, \alpha_t, \eta_t; \boldsymbol{\psi}_y) p(x_t|\alpha_t; \boldsymbol{\psi}_x^*) p(\alpha_t|\alpha_{t-1}; \boldsymbol{\psi}_\alpha) d\boldsymbol{\alpha}.$$

Koopman and Scharth (2013) documented that the likelihood function based on the density (A4) can be approximated via

$$L(\boldsymbol{\psi}; \boldsymbol{y}, \boldsymbol{x}) = p(\boldsymbol{x}; \boldsymbol{\psi}_x^*, \boldsymbol{\psi}_\alpha) \times p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\psi}_y, \boldsymbol{\psi}_\alpha), \tag{A5}$$

where

$$p(\boldsymbol{x}; \boldsymbol{\psi}_x^*, \boldsymbol{\psi}_\alpha) = \int \prod_{t=1}^T p(x_t|\alpha_t; \boldsymbol{\psi}_x^*) p(\alpha_t|\alpha_{t-1}; \boldsymbol{\psi}_\alpha) d\boldsymbol{\alpha},$$

$$p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\psi}) = \prod_{t=1}^T p(y_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi}), \quad p(y_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi}) = \int p(y_t|\alpha_t, \eta_t; \boldsymbol{\psi}) p(\alpha_t, \eta_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi}) d\alpha_t d\eta_t, \tag{A6}$$

with the interpolation set $\boldsymbol{x}_{\backslash t} = \{x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_T\}$. Intuitively, the idea of Koopman and Scharth (2013) is to evaluate the marginal density $p(\boldsymbol{x}; \boldsymbol{\psi}_x^*, \boldsymbol{\psi}_\alpha)$ via the Kalman filter and to estimate the remaining part $p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\psi})$ via numerical integration or quasi-Monte Carlo integration. Maximizing $\log p(\boldsymbol{x}; \boldsymbol{\psi}_x^*, \boldsymbol{\psi}_\alpha)$ gives the first-step estimator for $\boldsymbol{\psi}_x$ and $\boldsymbol{\psi}_\alpha$. Conditional on the estimate, maximizing the remaining part with respect to $\boldsymbol{\psi}_y$ gives the second-step estimate. The details of the second-step estimation for the models with and without the asymmetric effect are explained in the rest of the appendix.

Appendix B.2. Estimation for Model without Asymmetric Effect

For the model without asymmetric effects, the conditional density for y_t in (A6) reduces to

$$p(y_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi}) = \int p(y_t|\alpha_t; \boldsymbol{\psi}_y) p(\alpha_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi}) d\alpha_t.$$

As discussed in Koopman and Scharth (2013), $p(\alpha_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi})$ is the normal density function with mean $E(\alpha_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi})$ and variance $\text{Var}(\alpha_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi})$, which can be obtained via the deletion smoothing algorithm (see Appendix A). Koopman and Scharth (2013) recommended using the Gaussian quadrature for approximating $p(y_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi}_y, \hat{\boldsymbol{\psi}}_x, \hat{\boldsymbol{\psi}}_\alpha)$, where $\hat{\boldsymbol{\psi}}_x$ and $\hat{\boldsymbol{\psi}}_\alpha$ are estimates obtained via the first-step. This current study adopts the Gaussian–Legendre quadrature based on six points since there is no major improvement from increasing the number of points after six.

Using the approximated density

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\psi}_y, \hat{\boldsymbol{\psi}}_x^*, \hat{\boldsymbol{\psi}}_\alpha) = \prod_{t=1}^T \hat{p}(y_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi}_y, \hat{\boldsymbol{\psi}}_x^*, \hat{\boldsymbol{\psi}}_\alpha),$$

it is straightforward to obtain the second-step estimator by maximizing the log of the simulated likelihood function for the 2SML estimation. Note that the constant term in (3) needs correcting by $\hat{\zeta} = \hat{\zeta}^* - \hat{c}$ using $\hat{c}$ in the second step.

Appendix B.3. Estimation for Model with Asymmetric Effect

For the model without asymmetric effects, Koopman and Scharth (2013) suggested approximating $p(y_t|\boldsymbol{x}_{\backslash t}; \boldsymbol{\psi})$ via quasi-Monte Carlo integration using the Halton sequence (see Train 2003 for instance):

$$\hat{p}(y_t|\boldsymbol{x}_{\setminus t}; \boldsymbol{\psi}) = \frac{1}{S} \sum_{s=1}^{S} p(y_t|\alpha_t^{(s)}, \eta_t^{(s)}; \boldsymbol{\psi}_y, \hat{\boldsymbol{\psi}}_\alpha),$$

where $(\alpha_t^{(s)}, \eta_t^{(s)})$ is obtained via the two-dimensional Halton sequence, which is controlled by the mean and the covariance matrix of $p(\alpha_t, \eta_t|\boldsymbol{x}_{\setminus t}; \hat{\boldsymbol{\psi}}_x^*, \hat{\boldsymbol{\psi}}_\alpha)$. The deletion-smoothing algorithm for the redefined state (α_t, η_t) is explained in Appendix A. For the case of the normal distribution for ε_t, $p(y_t|\alpha_t, \eta_t; \boldsymbol{\psi}_y, \hat{\boldsymbol{\psi}}_\alpha)$ is the normal distribution with mean $(\rho\eta_t/\hat{\sigma}_\eta)\exp(0.5h_t)$ and variance $(1-\rho^2)\exp(0.5h_t)$, with $h_t = c + \alpha_t$, respectively. For the case of non-normal distribution, Koopman and Scharth (2013) used a copula function for the dependence. Maximizing the log of the simulated likelihood gives the second-step SML estimator, as above.

The approximation error of the quasi-Monte Carlo integration is of the order $O(S^{-1}(\log S)^2)$. As discussed in Asmussen and Glynn (2007), the convergence rate of the quasi-Monte Carlo method in practice is usually much faster than its theoretical upper bound. The Monte Carlo experiments in Koopman and Scharth (2013) set $S = 100$, and the current study follows this approach.

References

Abramowitz Milton and Irene A. Stegun 1970. *Handbook of Mathematical Functions*. Mineola : Dover Publications.

Asai, Manabu. 2008. Autoregressive stochastic volatility models with heavy-tailed distributions: A comparison with multifactor volatility models. *Journal of Empirical Finance* 15: 332–41. [CrossRef]

Asai, Manabu. 2009. Bayesian analysis of stochastic volatility models with mixture-of-normal distributions. *Mathematics and Computers in Simulation* 79: 2579–96. [CrossRef]

Asai, Manabu, Chia-Lin Chang, and Michael McAleer. 2017. Realized stochastic volatility with general asymmetry and long memory. *Journal of Econometrics* 199: 202–12. [CrossRef]

Asai, Manabu, and Michael McAleer. 2009. The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics* 150: 182–92. [CrossRef]

Asai, Manabu, Michael McAleer, and Marcelo C. Medeiros. 2012a. Asymmetry and long memory in volatility modeling. *Journal of Financial Econometrics* 10: 495–512. [CrossRef]

Asai, Manabu, Michael McAleer, and Marcelo C. Medeiros. 2012b. Estimation and forecasting with noisy realized volatility. *Computational Statistics & Data Analysis* 56: 217–30.

Asmussen, Søren, and Peter W. Glynn. 2007 *Stochastic Simulation: Algorithms and Analysis*. New York: Springer.

Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76: 1481–36. [CrossRef]

Barndorff-Nielsen, Ole E., P. Reinhard Hansen, Asger Lunde, and Neil Shephard. 2009. Realised kernels in practice: Trades and quotes. *Econometrics Journal* 12: C1–C32. [CrossRef]

Barndorff-Nielsen, Ole E., and Neil Shephard. 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64: 253–80. [CrossRef]

Bollerslev, Tim, and Hao Zhou. 2006. Volatility puzzles: A simple framework for gauging return-volatility regressions. *Journal of Econometrics* 131: 123–50. [CrossRef]

Chib, S., Y. Omori, and M. Asai. 2009. Multivariate stochastic volatility. In *Handbook of Financial Time Series*. Edited by T. G. Andersen, R. A. Davis, J. P. Kreiss and T. Mikosch. New York: Springer, pp. 365–400.

De Jong, Piet. 1989. Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association* 84: 1085–88. [CrossRef]

Dunsmuir, W. 1979. A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed with noise. *Annals of Statistics* 7: 490–506. [CrossRef]

Durbin, James, and Siem Jan Koopman. 2001. *Time Series Analysis by State-Space Methods*. Oxford: Oxford University Press.

Engle, Robert F., and Giampiero M. Gallo. 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131: 3–27. [CrossRef]

Hansen, Peter Reinhard, and Zhuo Huang. 2016. Exponential GARCH modeling with realized measures of volatility. *Journal of Business & Economic Statistics* 34: 269–87.

Hansen, P. R., Z. Huang, and H. H. Shek. 2012. Realized GARCH: A complete model of returns and realized measures of volatility. *Journal of Applied Econometrics* 27: 877–906. [CrossRef]

Hansen, Peter R., Asger Lunde, and James M. Nason. 2011. The model confidence set. *Econometrica* 79: 453–97. [CrossRef]

Harvey, Andrew, Esther Ruiz, and Neil Shephard. 1994. Multivariate stochastic variance models. *Review of Economic Studies* 61: 247–64. [CrossRef]

Harvey, Andrew C., and Neil Shephard. 1996. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14: 429–34.

Koopman, Siem Jan, and Marcel Scharth. 2013. The analysis of stochastic volatility in the presence of daily realized measures. *Journal of Financial Econometrics* 11: 76–115. [CrossRef]

Liesenfeld, Roman, and Robert C. Jung. 2000. Stochastic volatility models: Conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* 15: 137–60. [CrossRef]

Patton, Andrew J. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160: 246–56. [CrossRef]

Sandmann, Gleb, and Siem Jan Koopman. 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* 87: 271–301. [CrossRef]

Shephard, Neil, and Kevin Sheppard. 2010. Realising the future: Forecasting with high frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 25: 197–231. [CrossRef]

Shirota, Shinichiro, Takayuki Hizu, and Yasuhiro Omori. 2014. Realized stochastic volatility with leverage and long memory. *Computational Statistics & Data Analysis* 76: 618–41.

Takahashi, Makoto, Yasuhiro Omori, and Toshiaki Watanabe. 2009. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis* 53: 2404–26.

Taniguchi, Masanobu, and Yoshihide Kakizawa. 2000. *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer.

Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Yu, Jun. 2005. On leverage in a stochastic volatility model. *Journal of Econometrics* 127: 165–78. [CrossRef]