


Article

# Generalized Method of Moments Estimation of Realized Stochastic Volatility Model

Luwen Zhang <sup>†</sup> and Li Wang <sup>\*,†</sup> 

School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau 999078, China; zlw15133532117@163.com

\* Correspondence: liwang-fi@must.edu.mo; Tel.: +853-8897-3008

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** The purpose of this paper is to study the generalized method of moments (GMM) estimation procedures of the realized stochastic volatility model; we give the moment conditions for this model and then obtain the estimation of parameters. Then, we apply these moment conditions to the realized stochastic volatility model to improve the volatility prediction effect. This paper selects the Shanghai Composite Index (SSE) as the original data of model research and completes the volatility prediction under a realized stochastic volatility model. Markov chain Monte Carlo (MCMC) estimation and quasi-maximum likelihood (QML) estimation are applied to the parameter estimation of the realized stochastic volatility model to compare with the GMM method. And the volatility prediction accuracy of these three different methods is compared. The results of empirical research show that the effect of model prediction using the parameters obtained by the GMM method is close to that of the MCMC method, and the effect is obviously better than that of the quasi-maximum likelihood estimation method.

**Keywords:** realized stochastic volatility model; generalized method of moments (GMM); Markov chain Monte Carlo (MCMC); quasi-maximum likelihood (QML); high-frequency data



**Citation:** Zhang, Luwen, and Li Wang. 2023. Generalized Method of Moments Estimation of Realized Stochastic Volatility Model. *Journal of Risk and Financial Management* 16: 377. <https://doi.org/10.3390/jrfm16080377>

Academic Editors: Liqun Wang and Klaus Pötzelberger

Received: 28 June 2023

Revised: 10 August 2023

Accepted: 13 August 2023

Published: 16 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Stock markets are not only one of the most important economic and financial markets in each country today, but their immaturity and institutional weaknesses can lead to serious divergences between their development and macroeconomic developments, mainly referring to the high volatility of stock prices, which makes asset pricing and effective portfolios subject to a lot of uncertainty. Therefore, the study of stock market volatility and more accurate estimation and prediction of stock market fluctuations play an important role and significance in reducing stock market risks, maintaining the safe and stable development of the stock market and ensuring the healthy and stable operation of the macro economy; refer to Brooks and Persaud (2003); Giot and Laurent (2004). With the rapid advancement of computer technology, accessing high-frequency financial data has become easier. Using high-frequency data, we can estimate realized volatility; refer to Andersen et al. (2003); Barndorff-Nielsen and Shephard (2003a, 2003b); Jacod et al. (2009), etc. By incorporating high-frequency financial data, it provides a more accurate measure of market volatility compared to traditional methods.

The original stochastic volatility (SV) model was proposed by Taylor (1986) and others. Taylor (1986) proposed a discrete-time SV model, White (1984) proposed a continuous-time SV model, and Harvey and Shephard (1996) discussed an asymmetric SV model with leverage effects between the return process and the stochastic volatility process in the SV model using the quasi-maximum likelihood estimation method. Han et al. (2016) described an asymmetric stochastic volatility model using Gaussian regression with parameter estimation using the sequential Monte Carlo method.

The inclusion of unobservable potential random variables in the SV model makes the implementation of parameter estimation methods for the SV model very complicated. Commonly used parameter estimation methods include the generalized method of moments (Taylor (1986); Andersen and Sørensen (1996)), quasi-maximum likelihood method (Harvey et al. (1994); Harvey and Shephard (1996)), Markov chain Monte Carlo method (Kim et al. (1998); Yu (2005); Takahashi et al. (2021)), Bayesian method (Jacquier et al. (2002, 2004); Liu (2021); Bormetti et al. (2020)), and efficient method of moments (Andersen et al. (1999); Bansal et al. (1995); Gallant and Tauchen (1996)).

Financial return volatility is defined as the standard deviation of returns and plays a central role in modern finance. Realized volatility is the sum of the squares of intra-day returns over an interval and is used by modern financial economists and econometricians as a measure of true volatility. Andersen and Bollerslev (1997) thought the realized volatility proposed would provide a stable estimate of the potential volatility under the assumption of an ideal market. However, in real markets, measuring daily realized volatility based on high-frequency return data raises problems related to the presence of microstructure noise of the trading market. There are many noise-robust approaches of realized volatility (see Zhang et al. (2005); Barndorff-Nielsen et al. (2008); Xiu (2010); Jacod et al. (2009) and references therein). We apply the pre-averaging method (Jacod et al. (2009)) to estimate the realized volatility using high-frequency data.

Realized volatility reveals some important information of volatility; combining realized volatility into a traditional volatility model can improve the forecasting effect. Hansen et al. (2012, 2014) incorporated realized volatility with a generalized autoregressive conditional heteroscedasticity model. Takahashi et al. (2009) explores a stochastic volatility model with realized volatility, selecting the new sampling method and using the Markov chain Monte Carlo method for parameter estimation. Chaussé and Xu (2018) used four generalized methods of moments for the generalized asymmetric stochastic volatility with a realized volatility model (GASV-RV) and concluded that the efficiency of the GMM was improved by automatic moment selection through the principal component GMM and regularized GMM procedures.

This paper uses realized volatility constructed from high-frequency data and adds it to the stochastic volatility model to improve the prediction of volatility; the new model is called the realized stochastic volatility model. We employ the GMM method to estimate the parameters of the realized stochastic volatility model. The paper presents the theoretical moment conditions of the realized stochastic volatility model; the research contribution is providing the moment conditions for realized volatility. Furthermore, we explore the accuracy of GMM by comparing it with other two methods, MCMC and QML, which are utilized for parameter estimation in the realized stochastic volatility model.

We introduce the realized stochastic volatility model in Section 2. Estimation of realized volatility is given in Section 3. Three parameter estimation methods are introduced in Section 4. Section 5 provides an empirical illustration and demonstrates the effectiveness of three different parameter estimation methods. Section 6 contains the conclusion.

## 2. Realized Stochastic Volatility Model

Considering the realized volatility measure in the traditional SV model, the realized stochastic volatility (RSV) model is constructed by Takahashi et al. (2009). Compared with the traditional SV model, the RSV model contains more intra-day information, which is helpful to improve the prediction performance of the model inside and outside the sample. The specific RSV model is expressed as follows:

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim N(0,1), \quad (1)$$

$$z_t = \zeta + h_t + u_t, \quad u_t \sim N(0, \sigma_u^2), \quad (2)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \eta_t \sim N(0, \sigma_\eta^2), \tag{3}$$

$$h_1 = \mu + \epsilon_0, \epsilon_0 \sim N(0, \frac{\sigma_\eta^2}{1 - \phi^2}). \tag{4}$$

In the yield equation  $y_t$ , the volatility  $\sigma_t = \exp(h_t/2)$  plays the role of a constant scale factor, and  $h_t$  is the unobserved potential volatility. To ensure the strict stationarity and iterative nature of the stochastic process, the persistence parameter  $|\phi| < 1$  is assumed in the logarithmic volatility equation  $h_t$  and set  $h_1$ .  $\epsilon_t$  and  $\eta_t$  are random error terms. Theoretically, when the error term  $\eta_t$  obeys the standard normal distribution,  $h_t$  is a stationary process of AR(1), following the normal distribution with the mean value of  $\mu$  and the variance of  $\sigma_\eta^2 / (1 - \phi^2)$ . The RSV model is composed by adding a metric Equation (2) to the rate of return equation and the state equation of the SV model. Where  $z_t$  is the realized volatility at time  $t$ , the pre-averaging method can be chosen for the estimation of realized volatility,  $\sigma_u^2$  is the variance of the new interest  $u_t$ , the smaller the  $\sigma_u^2$ , the better the fit of the model,  $\xi$  is the bias correction term of the realized volatility measure. We use the pre-averaging method to estimate the realized volatility, because this method can handle the microstructure noise problem when using high-frequency data.

### 3. Realized Volatility

The term volatility comes from mathematical statistics, and it is an indicator used to measure the level of price volatility and reflects the extent to which prices deviate from their average value. Realized volatility is an estimation of integrated volatility. When using high-frequency data, the traditional realized volatility estimator will be dominated by noise and will not have convergence to the integrated volatility. In this work, the pre-averaging method is used to calculate the realized volatility, which is proposed by [Jacod et al. \(2009\)](#). This method can reduce the effect of microstructure noise, and the estimator is a consistent estimator for the integrated volatility. Precisely, the latent price is  $X_i^n = X_{i\Delta_n}$ ,  $\Delta_n = 1/n$ , the noise price is  $\epsilon_i^n = \epsilon_{i\Delta_n}$ , and the observed contaminated data are represented by  $Z_i^n = Z_{i\Delta_n}$ ,

$$Z_i^n = X_i^n + \epsilon_i^n. \tag{5}$$

We choose a sequence  $k_n$  and a number  $\theta$  that satisfies  $k_n \sqrt{\Delta_n} = \theta + o(\Delta_n^{1/4})$ .  $g(x)$  be a function defined in  $[0, 1]$  which satisfies  $g$  is continuous, piecewise  $C^1$  with a piecewise Lipschitz derivative  $g'$ . Denote  $g_j^n := g(\frac{j}{k_n})$  for  $j = 0, 1, \dots, k_n$ , then we define the pre-averaged increments:  $\bar{Z}_i^n = \sum_{j=1}^{k_n-1} g_j^n \Delta_{i+j}^n Z$ ,  $\Delta_i^n Z = Z_i^n - Z_{i-1}^n$ ,  $i = 0, 1, \dots, n - k_n + 1$ .

The pre-averaged estimator is

$$\hat{C}_t^n = \frac{\sqrt{\Delta_n}}{\theta \psi_2} \sum_{i=0}^{[t/\Delta_n] - k_n + 1} (\bar{Z}_i^n)^2 - \frac{\psi_1 \Delta_n}{2\theta^2 \psi_2} \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n Z)^2. \tag{6}$$

In this case, we choose

$$g(x) = x \wedge (1 - x), \psi_1 = 1, \psi_2 = \frac{1}{12}, \theta = \frac{1}{3}. \tag{7}$$

When using Equation (2), we employ the value of  $\hat{C}_t^n$  to replace  $z_t$ , which represents the realized volatility.

#### 4. Parameter Estimation Methods

##### 4.1. GMM Method Based on RSV Model

The GMM method was first proposed by Hansen (1982). It is a generic method for estimation parameters in semiparametric models. The method requires a certain number of moment conditions that are specified for the model. In this work, we refer to the method used in Jacquier et al. (2002) to construct moment conditions for the rate of return and have the following theorem.

**Theorem 1.** Given the RSV model given in Equations (1)–(3), for  $0 \leq j \leq 10$ , the first four order moments and the cross-moment expressions for  $y_t$  and  $y_{t+j}$  are

$$E|y_t| = \sqrt{2/\pi}E(h_t), \tag{8}$$

$$E(y_t^2) = E(h_t^2), \tag{9}$$

$$E|y_t^3| = 2\sqrt{2/\pi}E(h_t^3), \tag{10}$$

$$E(y_t^4) = 3E(h_t^4), \tag{11}$$

$$E|y_t y_{t+j}| = (2/\pi)E(h_t h_{t+j}), j = 1, 2, \dots, 10, \tag{12}$$

$$E(y_t^2 y_{t+j}^2) = E(h_t^2 h_{t+j}^2), j = 1, 2, \dots, 10, \tag{13}$$

where,  $E(h_t^r) = \exp(\frac{r\mu}{2} + \frac{r^2\sigma^2}{8})$ ,  $E(h_t^r h_{t+j}^s) = E(h_t^r)E(h_t^s)\exp(\frac{rs\phi^j\sigma^2}{4})$ ,  $\sigma^2 = \frac{\sigma_\eta}{1-\phi^2}$ .

Referring to the proof of the moment condition in Chaussé and Xu (2018), this paper gives the moment condition that the RSV model has the realized volatility term, which is proved as follows.

**Proposition 1.** Given the RSV model specified in Equations (1)–(3), the first two order moments and the cross-moment expressions for  $z_t$  and  $z_{t+j}$  ( $0 \leq j \leq 10$ ) are

$$E(z_t) = \zeta + \frac{\mu}{1-\phi}, \tag{14}$$

$$E(z_t^2) = \zeta^2 + \frac{\mu^2}{(1-\phi)^2} + \frac{\sigma_\eta^2}{1-\phi^2} + \sigma_u^2 + 2\frac{\zeta\mu}{1-\phi}, \tag{15}$$

$$E(z_t z_{t+j}) = \zeta^2 + 2\frac{\zeta\mu}{1-\phi} + \frac{\mu^2}{1-\phi} \sum_{i=1}^j \phi^{i-1} + \phi^j (\frac{\sigma_\eta^2}{1-\phi^2} + \frac{\mu^2}{(1-\phi)^2}), \tag{16}$$

$j = 1, 2, \dots, 10.$

**Proof of Proposition 1.** Given  $z_t$  and  $h_t$  specified in (2),

$$E(z_t) = \zeta + E(h_t) = \zeta + \frac{\mu}{1-\phi}, \tag{17}$$

$$\begin{aligned} E(z_t^2) &= \zeta^2 + E(h_t^2) + E(u_t^2) + 2\frac{\zeta\mu}{1-\phi} + 2E(h_t u_t) \\ &= \zeta^2 + \frac{\mu^2}{(1-\phi)^2} + \frac{\sigma_\eta^2}{1-\phi^2} + \sigma_u^2 + 2\frac{\zeta\mu}{1-\phi}, \end{aligned} \tag{18}$$

$$\begin{aligned}
 E(z_t z_{t+j}) &= \zeta^2 + 2 \frac{\xi \mu}{1 - \phi} + E[h_t (\mu \sum_{i=1}^j \phi^{i-1} + \phi^j h_t + \sum_{i=1}^j \phi^{j-i} \eta_{t+i-1})] \\
 &= \zeta^2 + 2 \frac{\xi \mu}{1 - \phi} + \frac{\mu^2}{1 - \phi} \sum_{i=1}^j \phi^{i-1} + \phi^j (\frac{\sigma_\eta^2}{1 - \phi^2} + \frac{\mu^2}{(1 - \phi)^2}), \\
 &j = 1, 2, \dots, 10.
 \end{aligned}
 \tag{19}$$

□

Let  $\psi_t$  be a  $q \times 1$  vector with typical element  $(y_t^n y_{t+j}^m, z_t^n z_{t+j}^m)$  for some  $m, j$ , and  $n \in (0, 1, 2, 3, \dots)$ , and let  $\psi(\theta_0) = E(\psi_t(\theta_0))$  be the theoretical moments of the RSV model. Let  $g_t(\theta) = [\psi_t - \psi(\theta)]$ ; then, the GMM estimator  $\hat{\theta}$  of the true vector of coefficients  $\theta_0$  is based on the following moment conditions:

$$E[g_t(\theta_0)] = 0, \tag{20}$$

and is the solution to:

$$\operatorname{argmin}_{\theta \in \Theta} \bar{g}(\theta)' \hat{\Omega}^{-1} \bar{g}(\theta). \tag{21}$$

where  $\Theta$  is the admissible parameter space implied by the model,  $\bar{g}(\theta) = [\sum_{t=1}^T \frac{\psi_t}{T} - \psi(\theta)]$  and  $\hat{\Omega}$  is a consistent estimate of the auto-correlation matrix of  $\sqrt{n} \bar{g}(\theta_0)$ .

Therefore, the estimator defined by Equation (21) is a one-step GMM with the estimate of the auto-correlation consistent (HAC) matrix given by:

$$\hat{\Omega} = \sum_{i=-T+1}^{T-1} \omega_h(i) \hat{\Gamma}_i, \tag{22}$$

where  $\omega_h(i)$  is a kernel, and  $h$  is the bandwidth, which can be chosen using the procedures proposed by Newey and West (1986) and Andrews (1991),

$$\hat{\Gamma}_i = \frac{1}{T} \sum_t (\psi_t - \bar{\psi})(\psi_{t+i} - \bar{\psi})'. \tag{23}$$

In order to improve the properties of the two-step GMM, Hansen (1982) suggested two other methods. The first one is the iterative version of the two-step GMM and can be computed as follows:

1. Compute  $\theta^{(0)} = \operatorname{argmin}_\theta \bar{g}(\theta)' \bar{g}(\theta)$ ;
2. Compute the HAC matrix  $\hat{\Omega}(\theta^{(0)})$ ;
3. Compute the  $\theta^{(1)} = \operatorname{argmin}_\theta \bar{g}(\theta)' [\hat{\Omega}(\theta^{(0)})]^{(-1)} \bar{g}(\theta)$ ;
4. If  $\|\theta^{(0)} - \theta^{(1)}\| < tol$  stops, else  $\theta^{(0)} = \theta^{(1)}$  and go to step 2;
5. Define the two-step GMM estimator  $\hat{\theta}$  as  $\theta^{(1)}$ ;

where  $tol$  can be set as small as we want to increase the precision.

#### 4.2. MCMC Method Based on RSV Model

In this paper, the MCMC method is used to estimate the parameters of the RSV model as a comparison with the GMM method. In the estimation, the prior distribution of the parameters is estimated and the conditional distribution of the combined sample information is given first, and then, the posterior distribution of the parameters to be estimated is calculated and the parameters of the models can be estimated for specific problems using the WinBUGS 1.4.3 software package.

Consider the RSV model described in (1)–(4). When given  $h = (h_1, \dots, h_T)$ , referring to Takahashi et al. (2009), we can compute the conditional likelihood of the RSV model as:

$$f(y_1, z_1, \dots, y_T, z_T | \bar{\theta}, h) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi} \exp(h_t/2)} \exp\left\{-\frac{y_t^2}{2\exp(h_t)}\right\} \times \frac{1}{\sqrt{2\pi} \sigma_u} \exp\left\{-\frac{(z_t - \xi - h_t)^2}{2\sigma_u^2}\right\}, \tag{24}$$

where  $\bar{\theta} = (\xi, \sigma_u^2, \mu, \phi, \sigma_\eta^2)$  denotes the parameters. Therefore, we use a Bayesian approach to estimate the posterior distribution of the parameters of the RSV model, considering  $h$  as an additional latent variable. In this setup, the most important thing is how to sample  $h$  efficiently. Therefore, we first describe the sampling algorithm for  $h$ .

When selecting the parameter’s prior distributions, we refer to the setting in Yu (2005), and we set priors as:  $\xi \sim N(0, 27)$ ,  $\sigma_u^2 \sim IG(2.5, 0.027)$ ,  $\mu \sim N(0, 25)$ ,  $\frac{1+\phi}{2} \sim Beta(20, 1.5)$ ,  $\sigma_\eta^2 \sim IG(2.5, 0.025)$ . Then, denoting  $Y = (y_1, \dots, y_T)$  and  $Z = (z_1, \dots, z_T)$ , the posterior density for  $\bar{\theta} = (\xi, \sigma_u^2, \mu, \phi, \sigma_\eta^2)$  and  $h$  becomes

$$f(\bar{\theta}, h | Y, Z) \propto \exp\left[-\frac{1}{2} \sum_{t=1}^T y_t^2 \exp(-h_t)\right] (\sigma_u^2)^{-T/2} \exp\left\{-\frac{1}{\sigma_u^2} \sum_{t=1}^T (z_t - \xi - h_t)^2\right\} \times \sqrt{1 - \phi^2} (\sigma_\eta^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma_\eta^2} (1 - \phi^2) (h_1 - \mu)^2\right. \\ \left. - \frac{1}{2\sigma_\eta^2} \sum_{t=1}^{T-1} (h_{t+1} - (1 - \phi)\mu - \phi h_t)^2\right\} \times \exp\left\{-\frac{(\xi)^2}{54}\right\} (\sigma_u^2)^{-3.5} \exp\left(-\frac{0.027}{\sigma_u^2}\right) \exp\left\{-\frac{(\mu)^2}{50}\right\} \\ \times \left(\frac{1 + \phi}{2}\right)^{19} \left(\frac{1 - \phi}{2}\right)^{0.5} (\sigma_\eta^2)^{-3.5} \exp\left(-\frac{0.025}{\sigma_\eta^2}\right). \tag{25}$$

To implement the Markov chain Monte Carlo simulation, we sample from the posterior distribution as follows:

- Simulate  $h$  from  $f(h | \mu, \phi, \sigma_\eta^2, Z, Y)$ .
- Simulate  $\xi$  from  $f(\xi | h, \sigma_u^2, Z)$ .
- Simulate  $\sigma_u^2$  from  $f(\sigma_u^2 | h, \xi, Z)$ .
- Simulate  $\mu$  from  $f(\mu | h, \phi, \sigma_\eta^2)$ .
- Simulate  $\sigma_\eta^2$  from  $f(\sigma_\eta^2 | h, \mu, \phi)$ .
- Simulate  $\phi$  from  $f(\phi | h, \mu, \sigma_\eta^2)$ .

### 4.3. QML Method Based on RSV Model

In this work, QML estimation is also performed for the RSV model. Due to the nonlinear relationship between the daily returns and the log of latent volatility in the Equations (1)–(3), we cannot compute the likelihood of these models by the Kalman filter. But given the parameter vector of the RSV model is  $\bar{\theta} = (\mu, \phi, \sigma_\eta^2, \xi, \sigma_u^2)$ , the log latent volatility is  $h = (h_1, \dots, h_T)$ , and by referencing Takahashi et al. (2009), we can compute the conditional likelihood of the RSV model as:

$$f(y_1, z_1, \dots, y_T, z_T | \bar{\theta}, h) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi} \exp(h_t/2)} \exp\left\{-\frac{y_t^2}{2\exp(h_t)}\right\} \times \frac{1}{\sqrt{2\pi} \sigma_u} \exp\left\{-\frac{(z_t - \xi - h_t)^2}{2\sigma_u^2}\right\}. \tag{26}$$

Then, the RSV model log-likelihood function can be written as:

$$\log \hat{L}(\bar{\theta}) = \log f(y_1, z_1, \dots, y_T, z_T | \bar{\theta}, h) = \sum_{t=1}^T \log f(y_t, z_t | \bar{\theta}, h). \tag{27}$$

The log-likelihood estimation obtained from the above equation is a continuous function of the RSV model parameter  $\bar{\theta}$ . Then, the parameter  $\bar{\theta}$  of this model can be estimated by virtue of the classical proposed maximum likelihood estimation method obtained as follows.

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmax}} \log \hat{L}(\bar{\theta}), \tag{28}$$

where  $\Theta$  is the admissible parameter space implied by the model.

### 5. Empirical Research

In this part, an empirical study will be conducted using the data of the Shanghai Stock Exchange (SSE) Composite Index from 4 January 2005 to 15 December 2022. The GMM method is used for estimating the parameters in the RSV model. The QML method and MCMC methods are also used in the RSV model for a comparative study.

#### 5.1. Loss Functions

To measure different methods' estimation and prediction performance, loss functions, also known as objective functions, are needed for measuring the errors between the actual volatility and predicted volatility. For regression data, the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) are often used. The mean square error refers to the expected value of the square of the difference between the estimated value and the true value. The root mean square error is the arithmetic square root of the mean square error, which can directly observe the direct difference between the predicted value and the real value. The mean absolute error can better reflect the actual error between the predicted value and the actual value. The mean absolute percentage error is a measure of relative error, which uses the absolute value to avoid the positive error and negative error canceling each other. The relative error can be used to compare the prediction accuracy of various time-series models. The loss functions mentioned above are defined as follows:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (\hat{h}_t - RV_t)^2, \tag{29}$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{h}_t - RV_t)^2}, \tag{30}$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |\hat{h}_t - RV_t|, \tag{31}$$

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{h}_t - RV_t}{RV_t} \right|, \tag{32}$$

where  $\hat{h}_t$  is the predicted volatility at time  $t$ ,  $T$  is the count number of the model forecast, and  $RV_t$  is the realized volatility estimated by the pre-averaged estimator.

#### 5.2. Data Selection and Processing

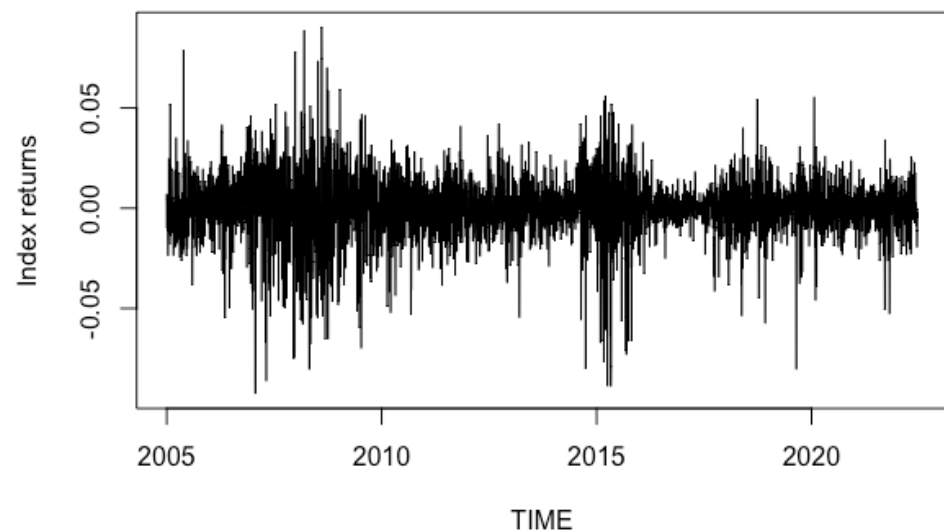
We used high-frequency data for the SSE Composite Index for the period from 4 January 2005 to 15 December 2022. The sample length is 4363, where the first 3963 trading days of data are selected for in-sample fitting and the last 400 trading days of data are selected for out-of-sample prediction. The frequency of our observed stock data is every



five minutes. For a normal trading day, there are 48 observations. The data used in the empirical analysis are sourced from the Oxford-Man Institute of Quantitative Finance Realized Library and the Wind database. Prior to conducting the empirical analysis, certain processing steps are required for the return variable,  $r_t$ .

1. The logarithm of the stock index closing price data for each stock market trading day is  $\log(p_t)$ ,  $t = 1, \dots, T$ , forming a logarithmic price series  $(\log(p_1), \dots, \log(p_T))$ ;
2. The logarithmic price series are differenced to obtain the return  $r_t = \log(p_t) - \log(p_{t-1})$ ,  $t = 1, \dots, T$  for the  $t$ -th trading day and then constitute the return sequence  $(r_1, r_2, \dots, r_T)$ .

Figure 1 below is the index returns of SSE. We can see the irregular and aggregation of the SSE stock index return volatility. In the three phases 2007–2009, 2015–2016 and 2018–2020, the SSE composite index return volatility is large, and extreme values are more prominent. As we know, there are relatively large stock price fluctuations during these three periods since the financial crisis and economic market downturn.



**Figure 1.** Index returns.

### 5.3. Model Parameter Estimation

We use the daily return series of the SSE Composite Index to represent  $y_t$  in Equation (1). In addition, we use the five-minute high-frequency return series of the SSE Composite Index to estimate the pre-averaged realized volatility in (6), and we use it as  $z_t$  in Equation (2). The GMM method and QML method are used to estimate the parameters of the RSV model by R 4.1.3 language software. The MCMC method is used to estimate the parameters of the RSV model using WinBUGS software. WinBUGS' basic principle is to sample from the complete conditional probability distribution through Gibbs sampling and the Metropolis algorithm, so as to generate a Markov chain, and finally estimate the model parameters through iteration. The obtained parameter estimation results are shown in Table 1. The advantage of introducing Gibbs sampling and MCMC is self-evident: that is, to avoid calculating a complete joint posterior probability publication with high-dimensional integral form and instead calculate the univariate conditional probability distribution of each estimated parameter.

Observing the persistence parameter  $\phi$ , the parameter  $\phi$ 's value of the RSV model of the SSE index is close to 1, indicating that the estimation results show that the time series of the SSE index has high persistent volatility characteristics. Next, observing the bias correction term  $\zeta$ , the parameter  $\zeta$  of the RSV model is positive, indicating that the effect of market microstructure noise still persists.



**Table 1.** Parameter estimation results for RSV model using different methods.

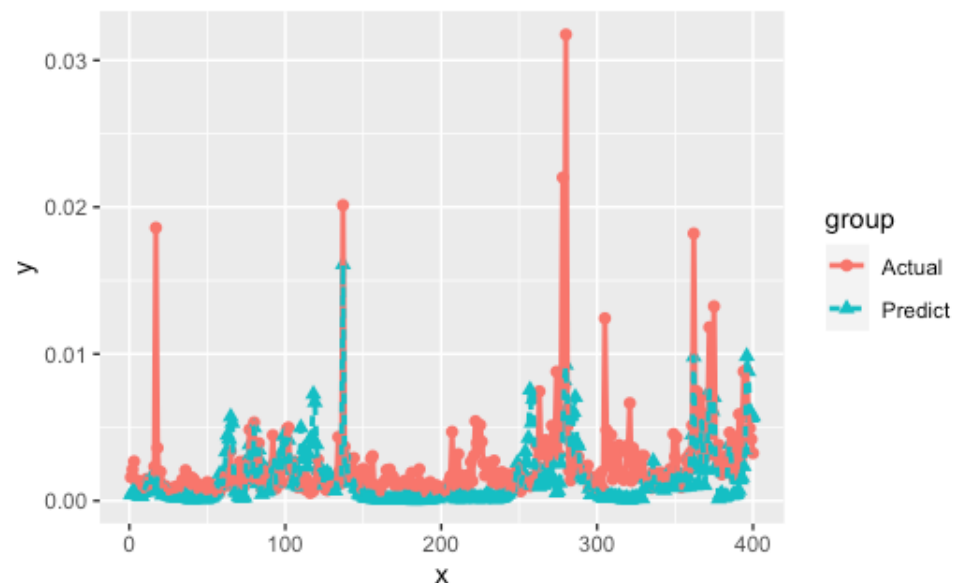
Method	$\mu$	$\phi$	$\sigma_\eta$	$\xi$	$\sigma_u$
RSV-GMM	−0.2630743 (0.14369)	0.9585373 (0.02479)	0.1408710 (0.08535)	1.1870245 (0.09746)	0.9432676 (0.00936)
RSV-MCMC	−0.1254 (0.15841)	0.9473 (0.02923)	0.2261 (0.05545)	1.135 (0.11521)	0.1211 (0.00724)
RSV-QML	−0.22454762 (0.149782)	0.9776021 (0.02637)	0.1461197 (0.01725)	0.7612748 (0.09839)	0.9253961 (0.00893)

Note: The number in parenthesis is the standard error.

The results of the parameters of the GMM method do not differ much from those of the QML method. The  $\phi$  values are still close to 1 and the persistence of volatility is still high.

From Figures 2–4, it is evident that the GMM method exhibits a notable ability to identify significant changes in volatility, particularly when volatility levels are high. The GMM method outperforms the MCMC, QML method in accurately predicting large volatility. The MCMC method performs well in forecasting, as it closely aligns the predicted volatility with the actual volatility. The predictive performance of the QML method in volatility estimation is satisfactory, yet it is not on par with the superior performance demonstrated by the GMM and MCMC approaches. The four loss functions are used to test the accuracy of the forecasting results.

The efficiency of three parameter estimation methods was investigated, and the results presented in Table 2 demonstrate that, under the RSV model, using parameters obtained from the MCMC method yields the most effective predictions of volatility, followed by the GMM method, while the QML method performs relatively weaker. When the RSV model is used for volatility prediction, the error of predicting volatility using the parameters estimated by the GMM method is almost the same as that predicted by the MCMC method. It is worth mentioning that the MCMC method requires more computation time compared to the GMM method, yet the predictive performance remains comparable. This finding substantiates the effectiveness and utility of the GMM method of RSV model proposed in this study.



**Figure 2.** Volatility prediction obtained from GMM method applied to RSV model.

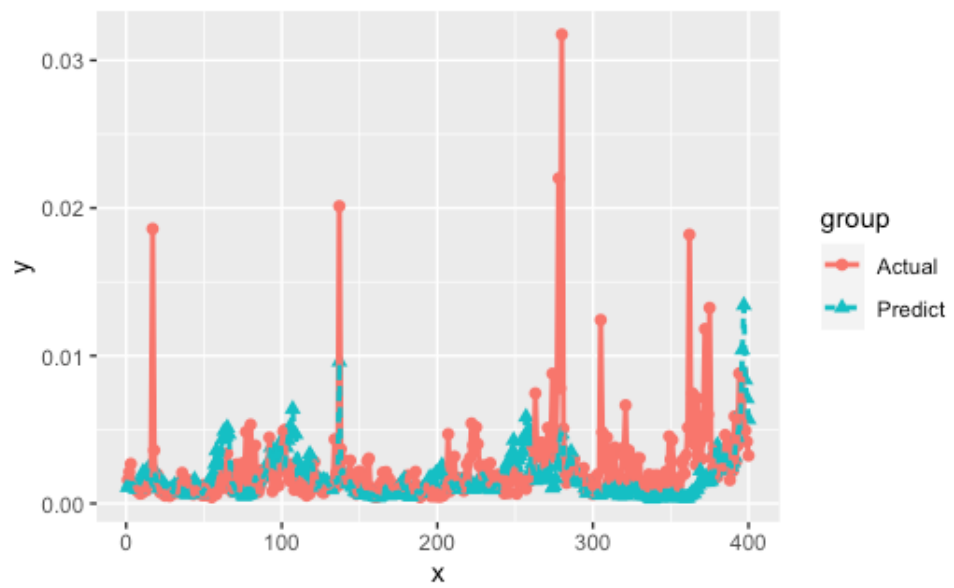


Figure 3. Volatility prediction obtained from MCMC method applied to RSV model.

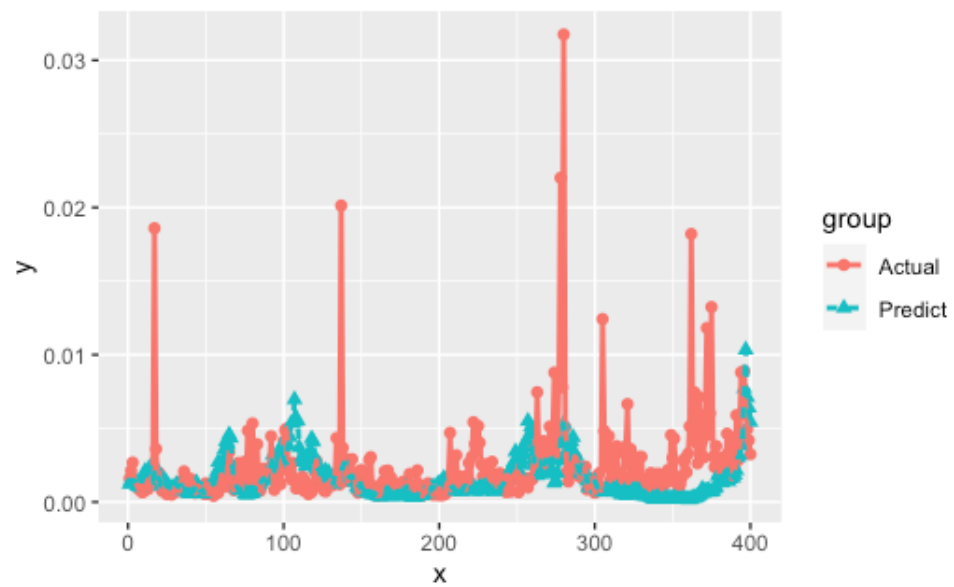


Figure 4. Volatility prediction obtained from QML method applied to RSV model.

Table 2. Prediction errors of volatility forecasting using different methods for RSV model parameter estimation.

Model	MSE	RMSE	MAE	MAPE
RSV-GMM	0.02319271	0.1522915	0.04333079	0.0610987
RSV-MCMC	0.02319251	0.1522909	0.04308118	0.0610579
RSV-QML	0.03319375	0.1821915	0.05093191	0.0637402

### 6. Conclusions

With the development of science and technology, people’s research in the field of stochastic volatility-type models parameter estimation is becoming more and more in-depth, and new parameter estimation methods are bound to appear. In this paper, GMM, MCMC and QML methods are used for realized stochastic volatility model parameter estimation, and we use these parameters and the realized stochastic volatility model to

predict volatility. Empirical data are analyzed in this paper. We use the five-minute high-frequency return series of the SSE Composite Index, and we apply the pre-averaging method to estimate the realized volatility. The prediction results illustrate that the GMM method is very effective and the calculation speed is faster, while the MCMC method is also effective, and the QML method is less accurate.

Although the realized volatility is introduced on the basis of the random volatility model, this paper still assumes that the disturbance term obeys normal distribution. According to the research in recent years, it is shown that the model disturbance term obeys the generalized hyperbolic distribution, which may improve the prediction effect of the model. For the improved model, we can consider using the efficient generalized method of moments to estimate the unknown parameters.

**Author Contributions:** All authors contributed equally to this article. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research of L. Wang was funded by MUST Faculty Research Grants (FRG), grant number FRG-21-003-FI.

**Data Availability Statement:** Data are obtained from the Oxford-Man Institute of Quantitative Finance Realized Library and the Wind database.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Andersen, Torben G., and Bent E. Sørensen. 1996. GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business & Economic Statistics* 14: 328–52.
- Andersen, Torben G., and Tim Bollerslev. 1997. Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts. *Social Science Electronic Publishing* 4: 885–905.
- Andersen, Torben G., Hyung Jin Chung, and Bent E. Sørensen. 1999. Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics* 91: 61–87. [\[CrossRef\]](#)
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. [\[CrossRef\]](#)
- Andrews, Donald W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society* 59: 817–58. [\[CrossRef\]](#)
- Bansal, Ravi, Gallant A. Ronald, Hussey Robert, and Tauchen George. 1995. Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics* 66: 251–87. [\[CrossRef\]](#)
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2003a. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64: 253–80. [\[CrossRef\]](#)
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2003b. Realized power variation and stochastic volatility models. *Bernoulli* 9: 243–65. [\[CrossRef\]](#)
- 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: 1481–536. [\[CrossRef\]](#)
- Bormetti, Giacomo, Roberto Casarin, Fulvio Corsi, and Giulia Livieri. 2020. A stochastic volatility model with realized measures for option pricing. *Journal of Business & Economic Statistics* 38: 856–71.
- Brooks, Chris, and Gita Persaud. 2003. Volatility forecasting for risk management. *Journal of Forecasting* 22: 1–22. [\[CrossRef\]](#)
- Chaussé, Pierre, and Dinghai Xu. 2018. GMM estimation of a realized stochastic volatility model: A Monte Carlo study. *Econometric Reviews* 37: 719–43. [\[CrossRef\]](#)
- Gallant, A. Ronald, and George Tauchen. 1996. Which moments to match? *Econometric Theory* 12: 657–81. [\[CrossRef\]](#)
- Giot, Pierre, and Sébastien Laurent. 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11: 379–98. [\[CrossRef\]](#)
- Han, Jianan, Xiao-Ping Zhang, and Fang Wang. 2016. Gaussian process regression stochastic volatility model for financial time series. *IEEE Journal of Selected Topics in Signal Processing* 10: 1015–28. [\[CrossRef\]](#)
- Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society* 50: 1029–54. [\[CrossRef\]](#)
- Hansen, Peter Reinhard, Asger Lunde, and Valeri Voev. 2014. Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics* 29: 774–99. [\[CrossRef\]](#)
- Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek. 2012. Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27: 877–906. [\[CrossRef\]](#)

- Harvey, Andrew C., and Neil Shephard. 1996. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics* 14: 429–34.
- Harvey, Andrew, Esther Ruiz, and Neil Shephard. 1994. Multivariate stochastic variance models. *The Review of Economic Studies* 61: 247–64. [[CrossRef](#)]
- Jacod, Jean, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter. 2009. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications* 119: 2249–76. [[CrossRef](#)]
- Jacquier, Eric, Nicholas G. Polson, and Peter Rossi. 2002. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* 20: 69–87.
- Jacquier, Eric, Polson Nicholas G., and Peter Rossi. 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122: 185–212. [[CrossRef](#)]
- Kim, Sangjoon, Neil Shephard, and Siddhartha Chib. 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies* 65: 361–93. [[CrossRef](#)]
- Liu, Jia. 2021. A Bayesian semiparametric realized stochastic volatility model. *Journal of Risk and Financial Management* 14: 617. [[CrossRef](#)]
- Newey, Whitney K., and Kenneth D. West. 1986. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelationconsistent Covariance Matrix. *The Econometric Society* 55: 703–8. [[CrossRef](#)]
- Takahashi, Makoto, Toshiaki Watanabe, and Yasuhiro Omori. 2021. Forecasting daily volatility of stock price index using daily returns and realized volatility. *Econometrics and Statistics*. [[CrossRef](#)]
- 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.
- Taylor, Stephen J. 1986. *Modelling Financial Time Series*. New York: John Wiley.
- White, Halbert. 1984. *Asymptotic Theory for Econometricians*. New York: Academic Press.
- Xiu, Dacheng. 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159: 235–50. [[CrossRef](#)]
- Yu, Jun. 2005. On leverage in a stochastic volatility model. *Journal of Econometrics* 127: 165–78. [[CrossRef](#)]
- Zhang, Lan, Per A. Mykland, and Yacine Aït-Sahalia. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100: 1394–411. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.