Methods for Estimating Discrete-Time Stochastic Volatility Models

XIAOBIN LIU *Zhejiang University*

This chapter reviews alternative methods proposed in the literature for estimating discrete-time stochastic volatility models and illustrates the details of their application. The methods reviewed are classified as either frequentist or Bayesian. The methods in the frequentist class include generalized method of moments, quasi-maximum likelihood, empirical characteristic function, efficient method of moments, simulated maximum likelihood based on Laplace-based importance sampler. The Bayesian methods include single-move Markov chain Monte Carlo, multi-move Markov chain Monte Carlo, and sequential Monte Carlo.

9.1 Introduction

For an asset, when only returns are available, measuring the volatility of the asset at the same frequency (for example, daily) is challenging. While it is possible to use the square of the return to measure the volatility, the squared return is known to be a very noisy estimator of the volatility. To better estimate the volatility, one approach is to use returns at a higher frequency to construct the so-called realized volatility. This approach has been very popular in the last twenty years. The literature was excellently reviewed in Andersen and Benzoni (2010) and a good textbook review of the approach is given in Aït-Sahalia and Jacod (2012).

However, information about the returns of assets is not always available at a higher frequency than the intended frequency for estimating the volatility. In this case, can the volatility be better estimated than by the use of the squared return? The answer is yes. There are two popular approaches in the literature. Both are motivated by the fact that the volatilities of financial assets tend to cluster over time.

The first approach uses a class of models called generalized autoregressive conditional heteroskedasticity (GARCH). The other adopts models from the stochastic volatility (SV) class. The GARCH-type models, initially proposed by Bollerslev (1986), are extensions of the autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982). The properties of the GARCH model have been well studied. In addition to the basic model, many specifications have been developed to accommodate important features, such as leverage effects and fat-tails in return distributions. In the GARCH-type models, the current volatility is a deterministic function of past returns and past volatilities. In other words, in the appropriately defined probability space, the volatility at period t is measurable given the information set at period t. This assumption makes the estimation of the model easier as the likelihood function of the model can be decomposed as the product of a sequence of univariate conditional densities, thus facilitating the maximum likelihood estimation.

Unlike GARCH-type models, the SV models assume volatilities evolve as a random process and hence, are latent. The first SV model allowing for volatility clustering was proposed by Taylor (1986), who suggested modeling the log-volatilities as a first-order autoregressive process. The properties of the SV model have been investigated thoroughly ((Ghysels et al., 1996)). The advantages of the SV model relative to GARCH-type models are now well understood (See, for example, (Broto and Ruiz, 2004) and (Kim et al., 1998)). For example, compared with GARCH-type models, the SV model fits the asset return series better. Moreover, it has a stronger connection to the continuous-time SV models widely used for pricing options. Consequently, many methods have been developed to estimate the SV models. We classify them into two categories: frequentist and Bayesian.

In the frequentist domain, all methods rely on carefully specifying a criterion function and then maximizing or minimizing that criterion function. The frequentist methods reviewed include generalized method of moments (GMM), quasi-maximum likelihood (QML), empirical characteristic function (ECF), efficient method of moments (EMM), and simulated maximum likelihood (SML). GMM, QML, ECF, and EMM avoid calculating the full likelihood function, whereas SML relies on the likelihood information by employing a simulation-based method to approximate the value of likelihood given the parameter values. The asymptotic behaviors of those estimators depend on the properties of the targeted model and the criterion function. They are usually asymptotically normally distributed. Unfortunately, since volatilities of most financial assets are highly persistent, asymptotic normality may not work well in finite samples ((Wang and Yu, 2022)), especially for the persistency parameter. Therefore, the asymptotic theory-based statistical inference may be misleading.

Unlike the asymptotic theory-based inference used by frequentist methods, with a Bayesian approach, finite-sample inference is conducted via the posterior analysis. There are several Bayesian methods available in the literature for estimating SV models. We discuss two classes of Bayesian methods based on whether the data augmentation technique is used. Data augmentation treats

the latent variables (or the log-volatilities) as parameters to estimate. In this case, the complete data likelihood function for both the parameters and latent variables has a closed-form expression. Therefore, we can use the classic Markov chain Monte Carlo method (MCMC) to sample from the posterior distribution (see, e.g., (Jacquier et al., 2002) (Shephard and Pitt, 1997), (Kim et al., 1998), (Chib et al., 2002), (Liesenfeld and Richard, 2003) and (Omori et al., 2007)). Specifically, we review two MCMC algorithms. The fist is the single-move MCMC, which updates the latent volatility step by step. The other is the multi-move MCMC, which updates the volatility in a block and is more efficient than the single-move MCMC. A smoothed estimate of volatility is obtained as a by-product in both algorithms.

Without data augmentation, we can use a Bayesian parameter learning algorithm that consists of the so-called sequential Monte Carlo (SMC) method and a marginal resample-move step. The SMC method is used to filter the hidden state (log-volatility). We can use it to approximate the likelihood given the values of parameters by marginalizing the hidden states. The resample-move step keeps the parameter from being static all the time and hence avoids particle depletion ((Fulop and Li, 2013)). The algorithm is sequential compared to the particle MCMC (PMCMC, see (Andrieu et al., 2010)) and it is more suitable for financial time-series applications since the belief is updated recursively as new data become available. The filtered estimate of volatility is obtained as a by-product in the algorithm.

However, Bayesian methods require sampling from the posterior. Consequently, the computation is very intensive, especially for latent variable models. In addition, the model must be fully specified since the likelihood function is required.

The rest of this chapter proceeds as follows. Section 9.2 introduces the SV model of Taylor (1986) to illustrate alternative estimation methods. This section also discusses some basic properties of the model. In Section 9.3, the methods in the frequentist domain are reviewed and the motivations and details of each method are presented. Section 9.4 provides a review of the methods in the Bayesian domain and explains the motivations and limitations of each method. A conclusion is provided in Section 9.5.

9.2 The Basic Model

Let y_t be the observed return at period t. In the literature, the log-normal SV model proposed by Taylor (1986) is defined as

$$y_t = \varepsilon_t \exp\left(\frac{1}{2}h_t\right), \tag{9.2.1}$$

$$h_t = \mu + \phi\left(h_{t-1} - \mu\right) + \sigma_\eta \eta_t, \tag{9.2.2}$$

where $t = 1, \ldots, T$, h_t is the log-volatility and $|\phi| < 1$, $\eta_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, η_t and ε_t are independent, $h_1 \sim \mathcal{N}\left(0, \frac{\sigma_\eta^2}{1-\phi^2}\right)$. The parameters of interest are $\boldsymbol{\theta} = (\mu, \phi, \sigma_\eta)' \in \boldsymbol{\Theta} \subset \mathbb{R}^3$, where $\boldsymbol{\Theta}$ is a compact parameter space. For the simplicity of notation, we use the subscript $1:t$ to represent the data ranging from period 1 to period t, that is, $y_{1:t} = (y_1, y_2, \ldots, y_t)'$. Denote the information set up to period t as $\mathcal{I}_t$, the model defined in Eq. (9.2.1) - (9.2.2) has the following properties.

Theorem 9.2.1 *The conditional variance and kurtosis of y_{t+1} are*

$$Var\left(y_{t+1}|\mathcal{I}_t\right) = \exp\left(\mu + \phi\left(h_t - \mu\right) + \frac{1}{2}\sigma_\eta^2\right). \tag{9.2.3}$$

$$Kurt\left(y_t\right) = 3\exp\left(\frac{\sigma_\eta^2}{1-\phi}\right) \tag{9.2.4}$$

Theorem 9.2.2 *The moments of y_t and $|y_t|$ satisfy*

$$\mathbb{E}\left(|y_t|^m\right) = 2^{\frac{m}{2}} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \exp\left[\frac{m\mu}{2\left(1-\phi\right)} + \frac{m^2\sigma_\eta^2}{8\left(1-\phi^2\right)}\right], \textit{for } m > -1,$$

$$\mathbb{E}\left(y_t^m\right) = \begin{cases} 0 & m \textit{ is odd}, \\ (m-1)!! \exp\left(\frac{m}{2}\mu + \frac{m^2\sigma_\eta^2}{8(1-\phi^2)}\right) & m \textit{ is even}, \end{cases}$$

where $n!!$ denotes the double factorial.

Corollary 9.2.1 *The model given by Eq. (9.2.1) - (9.2.2) has the following properties,*

$$\mathbb{E}\left(|y_t y_{t-k}|\right) = \frac{2}{\pi} \exp\left[\mu + \frac{\left(1+\phi^k\right)\sigma_\eta}{4\left(1-\phi^2\right)}\right],$$

$$\mathbb{E}\left(y_t^2 y_{t-k}^2\right) = \exp\left[2\mu + \frac{\left(1+\phi^k\right)\sigma_\eta}{1-\phi^2}\right].$$

Theorem 9.2.1 shows that the simple SV model can capture the fat tail and volatility clustering properties that are common in return time series. Theorem 9.2.2 and Corollary 9.2.1 imply that the moments of powers of both y_t and $|y_t|$ are finite when some restrictions of the powers m are satisfied, which is helpful for developing various tools to estimate the parameters. However, the simple structure raises a challenge to use the maximum likelihood estimation directly. Specifically, the complete-data likelihood of the model including the unobservable log-volatility $h_{1:T}$ is

$$\begin{aligned} p\left(y_{1:T}, h_{1:T}|\boldsymbol{\theta}\right) = & \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}e^{\frac{h_t}{2}}} \exp\left[-\frac{y_t^2}{2e^{h_t}}\right] \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left[-\frac{\left(h_t - \mu - \phi\left(h_{t-1}-\mu\right)\right)^2}{2\sigma_\eta^2}\right] \\ & \times \sqrt{\frac{1-\phi^2}{2\pi\sigma_\eta^2}} \exp\left[-\frac{\left(1-\phi^2\right)\left(h_1-\mu\right)^2}{2\sigma_\eta^2}\right], \end{aligned} \tag{9.2.5}$$

which implies the likelihood of the observed data and the associated maximum likelihood estimator (MLE) is

$$L\left(\boldsymbol{\theta}\right) = \int p\left(y_{1:T}, h_{1:T}|\boldsymbol{\theta}\right) dh_{1:T} = \int p\left(y_{1:t}|h_{1:T}, \boldsymbol{\theta}\right) p\left(h_{1:T}|\boldsymbol{\theta}\right) dh_{1:T}, \tag{9.2.6}$$

$$\widehat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log L\left(\boldsymbol{\theta}\right). \tag{9.2.7}$$

where $\log L\left(\boldsymbol{\theta}\right)$ is not analytically available since the integration is very complicated and suffers from the curse of dimensionality. The larger the sample size is, the greater the computational challenge faced by MLE. Therefore, one cannot directly compute MLE.

Lastly, the model is closely related to the well-known state-space model since it can be rewritten as

$$\begin{aligned} \log\left(y_t^2\right) &= h_t + \log\left(\varepsilon_t^2\right), \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t. \end{aligned} \tag{9.2.8}$$

where $\log\left(\varepsilon_t^2\right)$ is the logarithm of a chi-squared random variable with 1 degree of freedom. Therefore, the model belongs to the non-normal state-space models. One can refer to Durbin (2000) for more discussion of the state-space model. Even though we cannot apply the usual Kalman filter and smooth technique, the model expression (9.2.8) inspires some innovative alternative estimation approaches. Those approaches for the log-normal SV model shed new light on the estimation of other types of SV models.

9.3 Methods in the Frequentist Domain

As mentioned previously, the frequentist estimators are all based on the minimizing or maximizing a carefully designed criterion function. We define the estimator as the one minimizing a distance measure closely related to the criterion function,

$$\widehat{\boldsymbol{\theta}}_{freq} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Dist\left(\boldsymbol{\theta}\right), \tag{9.3.1}$$

where $Dist\left(\boldsymbol{\theta}\right)$ is the distance measure defined in different estimation approaches. The asymptotic behavior of the estimators depends on this distance measure and the properties of the model, and they are usually asymptotically normally distributed.

9.3.1 Generalized Method of Moments

Hansen and Singleton (1982) firstly formalized the generalized method of moments (GMM). GMM has become one of the most widely used econometric methods in economics and finance. Unlike MLE, which requires full knowledge of the distribution of the data, GMM only uses a set of moment conditions. Classic GMM estimation proceeds relying on the convergence of the sample moments to their population counterparts. Following Hansen and Singleton (1982) and Andersen and Sørensen (1996), by denoting the true value of parameters as $\boldsymbol{\theta}_0$, we define the GMM estimator as

$$\widehat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Dist^{GMM}\left(\boldsymbol{\theta}\right) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[\widehat{M}\left(\boldsymbol{\theta}\right) - A\left(\boldsymbol{\theta}\right)\right]' W^{-1} \left[\widehat{M}\left(\boldsymbol{\theta}\right) - A\left(\boldsymbol{\theta}\right)\right], \tag{9.3.2}$$

where the function in the right hand side of the equation is the criterion function, $\widehat{M}\left(\boldsymbol{\theta}\right) = \left(\widehat{M}_1\left(\boldsymbol{\theta}\right), \ldots, \widehat{M}_q\left(\boldsymbol{\theta}\right)\right)'$, $\widehat{M}_i\left(\boldsymbol{\theta}\right) = \sum_{t=j+1}^{T} \frac{\widehat{m}_{it}(\boldsymbol{\theta})}{T-j}$ for $j = 1, \ldots, q$, $\widehat{m}_{it}\left(\boldsymbol{\theta}\right)$ is the sample realization of the moments at time t, q is the number of selected moments, which is usually larger than the number of parameters, $A\left(\boldsymbol{\theta}\right)$ are the analytical moments corresponding to the sample counterparts, $\widehat{M}\left(\boldsymbol{\theta}\right)$, W is the weighting matrix. The optimal choice of W minimizing the asymptotic covariance matrix of $\boldsymbol{\theta}$ is given by

$$W = \lim_{T\to\infty} E\left[\sum_{t,\tau=1}^{T} \frac{\left(m_t\left(\boldsymbol{\theta}_0\right) - A\left(\boldsymbol{\theta}_0\right)\right)\left(m_\tau\left(\boldsymbol{\theta}_0\right) - A\left(\boldsymbol{\theta}_0\right)\right)'}{T}\right],$$

where $m_t\left(\boldsymbol{\theta}_0\right) = \left(m_{1t}\left(\boldsymbol{\theta}_0\right), \ldots, m_{qt}\left(\boldsymbol{\theta}_0\right)\right)'$. When a model satisfies some regular conditions, the associated GMM estimator $\widehat{\boldsymbol{\theta}}_{GMM}$ is consistent and asymptotically normal.

For the SV model, the implementation of GMM is straightforward as the moments have closed-form expressions. Furthermore, GMM is advantageous compared with MLE since the computation of the latter is very burdensome ((Melino and Turnbull, 1990)). However, there are many challenges to GMM. One issue is that the finite sample and large sample properties of GMM depend on the selected moments but in general we do not know which or how many moments should be used.

Jacquier et al. (2002) and Andersen and Sørensen (1996) investigated the performance of GMM for the log-normal SV model defined in Eq. (9.2.1) and (9.2.2). The basic set of moments they used was inspired by Melino and Turnbull (1990), that is,

$$\left\{\left\{\mathbb{E}\left(|y_t|^m\right)\right\}_{m=1}^{4}, \left\{\mathbb{E}\left(|y_t y_{t-k}|\right)\right\}_{k=1}^{10}, \left\{\mathbb{E}\left(y_t^2 y_{t-k}^2\right)\right\}_{k=1}^{10}\right\}, \tag{9.3.3}$$

among which the moments of the marginal distribution serve to identify μ, and the autocovariance of the squares and absolute values help to identify ϕ. Thanks to Theorem 9.2.2 and Corollary

9.2.1, these moments have analytical expressions. The weighting matrix recommended is the kernel estimator of the spectral density matrix,

$$\widehat{W} = \sum_{j=-T+1}^{T+1} \mathcal{K}(j)\,\widehat{\Gamma}(j), \widehat{\Gamma}(j) = \frac{1}{T}\sum_{t=j+1}^{T}\left[m_t\left(\widehat{\boldsymbol{\theta}}\right) - A\left(\widehat{\boldsymbol{\theta}}\right)\right]\left[m_{t-j}\left(\widehat{\boldsymbol{\theta}}\right) - A\left(\widehat{\boldsymbol{\theta}}\right)\right]', \tag{9.3.4}$$

where $\mathcal{K}(j)$ is a weighted kernel function.

Jacquier et al. (2002) used the full set of (9.3.3) and found that the GMM estimators of the parameters were substantially biased. Andersen and Sørensen (1996) investigated the finite sample performance of GMM extensively. They reported several interesting findings. First, the selection of moments was limited due to the requirement for analytical expression. Second, they postulated that the number of moments used should depend on the samples size. When the sample size is small, an excessive number of moments may introduce more bias and a larger root mean square error (RMSE) since the highly correlated moments can result in a poor estimate for the weighting matrix in the criterion function. Third, when the sample is large, a fairly large number of lags (L_T) in the kernel function should be used. Fourth, the paper recommends using lower-order moments of y_t because the higher-order moments can cause erratic finite sample behavior. Based on the work of Andersen and Sørensen (1996), one may use the Bartlett kernel function with fixed bandwidth $L_T = 1.2T^{\frac{1}{3}}$,

$$\mathcal{K}(j) = \begin{cases} 1 - \frac{j}{L_T}, & j \leq L_T, \\ 0, & j > L_T, \end{cases}$$

and define another two sets of moments for estimation as below,

$$\mathcal{M}9 = \left\{\left\{\mathbb{E}\left(|y_t|^m\right)\right\}_{m=1}^{4}, \left\{\mathbb{E}\left(|y_t y_{t-k}|\right)\right\}_{k=1,3,5}, \left\{\mathbb{E}\left(y_t^2 y_{t-k}^2\right)\right\}_{k=2,4}\right\},$$

$$\mathcal{M}14 = \left\{\left\{\mathbb{E}\left(|y_t|^m\right)\right\}_{m=1}^{4}, \left\{\mathbb{E}\left(|y_t y_{t-k}|\right)\right\}_{k=2,4,6,8,10}, \left\{\mathbb{E}\left(y_t^2 y_{t-k}^2\right)\right\}_{k=1,3,5,7,9}\right\}.$$

We can summarize the estimation procedure in the Algorithm 1.

9.3.2 Quasi-maximum Likelihood

The quasi-maximum likelihood (QML) is another popular estimation approach for SV models. The advantage of the QML approach is its speed and adaptability to various situations. Specifically, we use an approximated model instead of the true one to obtain the MLE,

$$\widehat{\boldsymbol{\theta}}_{QML} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Dist^{QML}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left[-\log L^{quasi}(\boldsymbol{\theta})\right],$$

where $\log L^{quasi}(\boldsymbol{\theta})$ is the likelihood function of the approximated model. This is why we call it quasi-likelihood and the estimator is denoted as $\widehat{\boldsymbol{\theta}}_{QML}$. Although the QML method is inefficient, it leads to consistent estimators and is easy to implement numerically since the quasi-likelihood function is easy to evaluate.

In this section we discuss how to use QML to estimate the log-normal SV model. According to the properties of the log-normal SV model discussed in Section 9.2, the likelihood is intractable and hence the MLE is difficult to obtain. Thanks to the close relation to the linear state-space model as shown in Section 9.2, Ruiz (1994) proposed to estimate the SV model using QML based on the Kalman filter. To see how it works, following the derivation in Section 9.2, the model (9.2.1) - (9.2.2) can be rewritten as

$$\log\left(y_t^2\right) = h_t + \log\left(\varepsilon_t^2\right),$$

$$h_t = \mu + \phi\left(h_{t-1} - \mu\right) + \sigma_\eta \eta_t,$$

Algorithm 1 GMM estimation

1: Initialize $\boldsymbol{\theta}^{(0)}$.
2: Select the moments to use. When $T \geq 1000$, adopt the moments $\mathcal{M} = \mathcal{M}14$. Otherwise adopt $\mathcal{M} = \mathcal{M}9$.
3: Compute

$$\boldsymbol{\theta}^{(1)} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[\widehat{M}(\boldsymbol{\theta}) - A(\boldsymbol{\theta})\right]' \left[\widehat{M}(\boldsymbol{\theta}) - A(\boldsymbol{\theta})\right].$$

4: Define $L_T = 1.2T^{\frac{1}{3}}$ and compute

$$\mathcal{K}(j) = \begin{cases} 1 - \frac{j}{L_T}, & j \leq L_T, \\ 0, & j > L_T. \end{cases}$$

5: Compute the weighting matrix

$$\widehat{W} = \sum_{j=-T+1}^{T+1} \mathcal{K}(j)\,\widehat{\Gamma}(j), \widehat{\Gamma}(j) = \frac{1}{T}\sum_{t=j+1}^{T} \left[m_t\left(\boldsymbol{\theta}^{(1)}\right) - A\left(\boldsymbol{\theta}^{(1)}\right)\right]\left[m_{t-j}\left(\boldsymbol{\theta}^{(1)}\right) - A\left(\boldsymbol{\theta}^{(1)}\right)\right]',$$

where the moments used are from $\mathcal{M}$.
6: Compute

$$\widehat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[\widehat{M}(\boldsymbol{\theta}) - A(\boldsymbol{\theta})\right]' \widehat{W}^{-1} \left[\widehat{M}(\boldsymbol{\theta}) - A(\boldsymbol{\theta})\right],$$

where the moments used are from $\mathcal{M}$.

where $\log\left(\varepsilon_t^2\right)$ is the logarithm of the chi-squared random variable with 1 degrees of freedom.

If $\log\left(\varepsilon_t^2\right)$ is indeed normally distributed, the model above is the state-space model widely used in economics and finance. We can obtain the MLE of state-space models with Gaussian error terms via the Kalman filter. However, we know that $\log\left(\varepsilon_t^2\right)$ does not follow a Gaussian white noise process, therefore, the Kalman filter can only approximate the outcome of the true nonlinear optimal filter. Specifically, in the QML method, we use a normal distribution to approximate $\log\left(\varepsilon_t^2\right)$. However, as pointed out by Harvey et al. (1994) and Ruiz (1994), the performance of QML depends on the magnitude of σ_η. If σ_η is large, the $\log(h_t)$ will dominate the error term $\log\left(\varepsilon_t^2\right)$ in the measurement equation. As a result, a normal approximation may be adequate and the Kalman filter may be close to the optimal filtering. However, for small values of σ_η, the normal approximation may break down and the QML method may perform poorly compared to the full-likelihood-based approach. Jacquier et al. (2002) also showed that the QML estimator has similar performance relative to the GMM estimator only for high values of σ_η. Following the steps in Yu (2002), we use a normal distribution with mean -1.27 and variance $\pi^2/2$ to approximate $\log\left(\varepsilon_t^2\right)$ and rewrite the model as

$$\begin{aligned} y_t &= A + B\tilde{h}_t + \tilde{\varepsilon}_t, \tilde{\varepsilon}_t \sim \mathcal{N}(0, R), \\ \tilde{h}_{t+1} &= \phi\tilde{h}_t + v_{t+1}, v_{t+1} \sim \mathcal{N}(0, Q), \end{aligned}$$

where $A = -1.27 + \mu$, $B = 1$, $R = \pi^2/2$, $\tilde{h}_t = h_t - \mu$, $Q = \sigma_\eta^2$. We describe the algorithm employing the Kalman filter to evaluate the quasi-likelihood of the log-normal SV model in Algorithm 2.

9.3.3 Empirical Characteristic Function

There is a one-to-one correspondence between the cumulative distribution function (CDF) and the characteristic function (CF). As a result, the empirical CF (ECF) should incorporate the same information as the empirical cumulative distribution (ECDF). Therefore, we can use the ECF

Algorithm 2 Computation of $\log L^{quasi}(\boldsymbol{\theta})$

1: Initialization: $\widehat{\tilde{h}}_{1|0} = 0$, $\tilde{\Sigma}_{1|0} = \sigma_\eta^2/\left(1-\phi^2\right)$.
2: Sequential updating:

$$\widehat{\tilde{h}}_{t|t} = \widehat{\tilde{h}}_{t|t-1} + \tilde{\Sigma}_{t|t-1}\frac{y_t - A - \widehat{\tilde{h}}_{t|t-1}}{\tilde{\Sigma}_{t|t-1}+R},$$

$$\tilde{\Sigma}_{t|t} = \tilde{\Sigma}_{t|t-1} - \tilde{\Sigma}_{t|t-1}\left(\tilde{\Sigma}_{t|t-1}+R\right)^{-1}\tilde{\Sigma}_{t|t-1}.$$

3: In-sample sequential predition

$$\widehat{\tilde{h}}_{t+1|t} = \phi\widehat{\tilde{h}}_{t|t-1} + \phi\frac{y_t - A - \widehat{\tilde{h}}_{t|t-1}}{1+R\tilde{\Sigma}_{t|t-1}^{-1}},$$

$$\tilde{\Sigma}_{t+1|t} = \phi^2\tilde{\Sigma}_{t|t} + \sigma_\eta^2.$$

4: Compute the quasi-likelihood

$$\log L^{quasi}(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\log\left(\tilde{\Sigma}_{t|t-1}+R\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{\left(y_t - A - \widehat{\tilde{h}}_{t|t-1}\right)^2}{\tilde{\Sigma}_{t|t-1}+R}.$$

to estimate parameters for a model. Essentially, the method approximates the likelihood by the product of joint densities of overlapping blocks with a fixed size. We briefly outline the ECF method below for general models. For more details, one can refer to Knight and Yu (2002) and Knight et al. (2002).

Given a random variable X and its realization x, we assume the cumulative distribution function (CDF) as $F(x|\boldsymbol{\theta})$, which depends on a parameter $\boldsymbol{\theta}$. The CF is defined as

$$C(m,\boldsymbol{\theta}) = \mathbb{E}\left(e^{imx}\right) = \int e^{imx}dF(x|\boldsymbol{\theta}), \tag{9.3.5}$$

and the ECF is

$$C_T(m) = \frac{1}{T}\sum_{t=1}^{T}e^{imx_t} = \int e^{imx}dF_T(x), \tag{9.3.6}$$

where $F_T(x)$ is the ECDF and m is the transform variable. Apparently, Eq. (9.3.5) only involves the parameter $\boldsymbol{\theta}$ and Eq. (9.3.6) contains only data. Hence we can estimate $\boldsymbol{\theta}$ by minimizing the distance between the CF and ECF, i.e., given a grid of transform variable $m_1,\ldots,m_q$ and the weight function $w(\cdot)$, the ECF estimator minimizes

$$\sum_{j=1}^{q}\left|C_T(m_j) - C(m_j,\boldsymbol{\theta})\right|^2 w(m_j),$$

or we can minimize the distance given the continuous transformation m,

$$\int\left|C_T(m) - C(m,\boldsymbol{\theta})\right|^2 w(m)\,dm.$$

The marginal ECDF cannot capture the structure of dependence in time-series data. Knight and Yu (2002) and Knight et al. (2002) recommended to use the joint ECF involving moving blocks of data. Given the observations $\{x_1,\ldots,x_T\}$, the overlapping blocks are

$$\boldsymbol{r}_j = (x_j,\ldots,x_{j+p}) \quad (j=1,\ldots,T-p),$$

which imply that each block has p periods overlapping with its adjacent blocks. The CF of each block is basically a joint CF and is defined as

$$C(\boldsymbol{m}, \boldsymbol{\theta}) = \mathbb{E}\left(e^{i\boldsymbol{m}'\boldsymbol{r}_j}\right), \tag{9.3.7}$$

where the transformation variable $\boldsymbol{m} = (m_1, \ldots, m_{p+1})'$. Hence, the joint ECF is

$$C_n(\boldsymbol{m}) = \frac{1}{n}\sum_{j=1}^{n} e^{i\boldsymbol{m}'\boldsymbol{r}_j}, \text{ where } n = T - p. \tag{9.3.8}$$

It is obvious that $\boldsymbol{r}_j$ is dependent since it is a sequence of overlapping moving blocks. Feuerverger (1990) proved that under standard regularity conditions, $C_n(\boldsymbol{m})$ converges almost surely to $C(\boldsymbol{m}, \boldsymbol{\theta})$ for any $\boldsymbol{m}$. Matching the joint CF in Eq. (9.3.7) with the ECF in Eq. (9.3.8) over a grid of discrete points is an alternative method for MLE ((Knight and Satchell, 1997)). Knight et al. (2002) defined the estimator for time-series data as

$$\widehat{\boldsymbol{\theta}}_{ECF} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Dist^{ECF}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \int \left|C(\boldsymbol{m}, \boldsymbol{\theta}) - C_n(\boldsymbol{m})\right|^2 w(\boldsymbol{m})\, d\boldsymbol{m}, \tag{9.3.9}$$

or equivalently, the parameter that solving

$$\int \left[C(\boldsymbol{m}, \widehat{\boldsymbol{\theta}}_{ECF}) - C_n(\boldsymbol{m})\right] g\left(\boldsymbol{m}, \widehat{\boldsymbol{\theta}}_{ECF}\right) d\boldsymbol{m} = 0,$$

where $C(\boldsymbol{m}, \boldsymbol{\theta})$ and $C_n(\boldsymbol{m})$ are from Eq. (9.3.7) and (9.3.8), respectively. $g(\boldsymbol{m}, \boldsymbol{\theta})$ and $w(\boldsymbol{m})$ are weighted functions. In fact, if $\boldsymbol{m}$ is a discrete grid of points (discrete ECF, DECF), the estimation is equivalent to matching a finite number of moments like GMM. (Knight and Yu, 2002) and Knight et al. (2002) proposed minimizing the integral in Eq. (9.3.9), i.e., matching all the moments continuously. To do this, the weight function should be continuous and hence we do not need to choose the transform variable, $\boldsymbol{m}$, which will be integrated out in Eq. (9.3.9). One appropriate proposal for the weight function is the exponential function, $\exp(-\boldsymbol{m}'\boldsymbol{m})$. The exponential function can put more weight on the interval around the origin and also has the advantage of convenient computation of the integration. Specifically, we can use the Hermitian quadrature or Monte Carlo integration to calculate the integral in Eq. (9.3.9). Theoretically, an optimal choice for the weight function $\omega(\boldsymbol{m}, \boldsymbol{\theta})$ is

$$\omega^*(\boldsymbol{m}, \boldsymbol{\theta}) = \int \exp\left(-i\boldsymbol{m}'\boldsymbol{r}_j\right) \frac{\partial \log p\left(x_{j+p}|x_{j:j+p-1}, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} dx_{j:j+p},$$

which is not available for some time-series models where the conditional score function does not have an analytical expression ((Knight and Yu, 2002)).

The estimator in Eq. (9.3.9) is consistent and asymptotically normal for strictly stationary processes under some mild assumptions ((Knight and Yu, 2002)). For the log-normal SV model, we firstly transform the return data following the logarithm operation in Eq. (9.2.8),

$$x_t = \log(y_t) = h_t + \log\left(\varepsilon_t^2\right), \tag{9.3.10}$$

and then plug the transformed data into the general estimation procedure. Specifically, the joint CF for $x_{t:t+p-1}$ is

$$\begin{aligned} C(m_1, \ldots, m_{p+1}, \boldsymbol{\theta}) = \exp&\left[i\mu\sum_{j=1}^{p+1} m_j - \frac{\sigma_\eta^2}{2(1-\phi^2)}\left(\sum_{j=1}^{p+1} m_j^2 + 2\phi\sum_{l=1}^{p+1}\sum_{j=l+1}^{p+1}\phi^{j-l-1} m_l m_j\right)\right] \\ &\times \frac{\prod_{j=1}^{p+1}\Gamma\left(\frac{1}{2} + i m_j\right)}{\Gamma\left(\frac{1}{2}\right)^k} 2^{i\sum_{j=1}^{p+1} m_j}. \end{aligned} \tag{9.3.11}$$

As pointed out by Knight et al. (2002), the autocorrelation for kth lag for $\{x_t\}_{t=1}^T$ is

$$\varrho_k = \phi^k \frac{\sigma_\eta^2/\left(1-\phi^2\right)}{\sigma_\eta^2/\left(1-\phi^2\right)+c}, k = 1, 2, \ldots, \text{ and } c = \frac{\Gamma''\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} - \left[\frac{\Gamma'\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\right]^2 \approx 0.4948.$$

Therefore, we only need $k = 1$ to identify the parameter ϕ. We briefly summarize the evaluation of criterion function in Algorithm 3.

Algorithm 3 Computation of $Dist^{ECF}(\boldsymbol{\theta})$

1: Compute

$$C(m_1, m_2, \boldsymbol{\theta}) = \exp\left[i\mu(m_1+m_2) - \frac{\sigma_\eta^2}{2\left(1-\phi^2\right)}\left(m_1^2 + 2\phi m_1 m_2 + m_2^2\right)\right] \times \frac{\Gamma\left(\frac{1}{2}+im_1\right)\Gamma\left(\frac{1}{2}+im_2\right)}{\Gamma\left(\frac{1}{2}\right)^2} 2^{im_1+im_2}$$

by extracting the real and imaginary parts $Re\left(C\left(m_1, m_2|\boldsymbol{\theta}\right)\right)$ and $Im\left(C\left(m_1, m_2|\boldsymbol{\theta}\right)\right)$.

2: Compute

$$C_n(m_1, m_2) = \frac{1}{n}\sum_{j=1}^{n} \exp\left(im_1 x_j + im_2 x_{j+1}\right),$$

by extracting the real and imaginary parts $Re\left(C_n\left(m_1, m_2\right)\right)$ and $Im\left(C_n\left(m_1, m_2\right)\right)$, where $n = T-1$. Specifically

$$Re\left(C_n\left(m_1, m_2\right)\right) = \frac{1}{n}\sum_{j=1}^{n} \cos\left(m_1 x_j + m_2 x_{j+1}\right),$$

$$Im\left(C_n\left(m_1, m_2\right)\right) = \frac{1}{n}\sum_{j=1}^{n} \sin\left(m_1 x_j + m_2 x_{j+1}\right).$$

3: Compute the criterion function

$$Dist^{ECF}(\boldsymbol{\theta}) = \int\int \Big\{\left[Re\left(C\left(m_1, m_2|\boldsymbol{\theta}\right)\right) - Re\left(C_n\left(m_1, m_2\right)\right)\right]^2 + \left[Im\left(C\left(m_1, m_2|\boldsymbol{\theta}\right)\right) - Im\left(C_n\left(m_1, m_2\right)\right)\right]^2\Big\} \exp\left(-m_1^2 - m_2^2\right) dm_1 dm_2.$$

9.3.4 Efficient method of moments

Unless the moment conditions correspond to the scores, GMM leads to, in general, an inefficient estimator. The efficient method of moments (EMM) provides more efficient estimators than GMM. Specifically, EMM relies on an auxiliary quasi-likelihood to generate a set of moments. Like GMM, EMM accommodates a variety of models, and its asymptotic property has been studied in Gallant and Tauchen (1996)

Following Andersen et al. (1999), the estimation procedure includes two major steps. In the first step, we choose an auxiliary model and estimate it by QML. Denote the conditional density of the auxiliary model as $f\left(y_t \mid \mathcal{I}_{t-1}, \boldsymbol{\vartheta}\right)$, where $\mathcal{I}_{t-1}$ is the information set up to time $t-1$ and $\boldsymbol{\vartheta}$ is the parameter. EMM firstly obtains the QML estimator for $\boldsymbol{\vartheta}$, $\widehat{\boldsymbol{\vartheta}}$, which satisfies the first-order

condition based on the quasi-score function $s\left(y_t, \boldsymbol{\vartheta}\right) = \frac{\partial \log f\left(y_t|\mathcal{I}_{t-1}, \boldsymbol{\vartheta}\right)}{\partial \boldsymbol{\vartheta}'}$, that is,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\vartheta}}\log f\left(y_t \mid \mathcal{I}_{t-1}, \widehat{\boldsymbol{\vartheta}}\right) = \frac{1}{T}\sum_{t=1}^{T} s\left(y_t, \widehat{\boldsymbol{\vartheta}}\right) = 0. \tag{9.3.12}$$

After that, we use GMM to estimate the original parameter $\boldsymbol{\theta}$. The moment conditions used to identify $\boldsymbol{\theta}$ are denoted as

$$\begin{aligned} m\left(\boldsymbol{\theta}, \boldsymbol{\vartheta}_0\right) &= \mathbb{E}_{\boldsymbol{\theta}}\left[s\left(y_t, \boldsymbol{\vartheta}_0\right)\right] = \int s\left(y_t, \boldsymbol{\vartheta}_0\right) dP\left(y_t, \boldsymbol{\theta}\right) & (9.3.13)\\ &= \int \frac{\partial \log f\left(y_t \mid \mathcal{I}_{t-1}, \boldsymbol{\vartheta}_0\right)}{\partial \boldsymbol{\vartheta}} dP\left(y_t, \boldsymbol{\theta}\right), & (9.3.14)\end{aligned}$$

where the number of moment conditions should be larger than $\dim\left(\boldsymbol{\theta}\right)$ for identification and $P\left(y_{1:t}, \boldsymbol{\theta}\right)$ is the probability measure induced by the targeted model.

However, the presence of the latent log-volatility makes the moment condition analytically intractable. Following Duffie and Singleton (1993), the corresponding sample moments for Eq. (9.3.14) at the fixed QML estimate $\widehat{\boldsymbol{\vartheta}}$ are

$$\widehat{m}_S\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right) = \frac{1}{S}\sum_{s=1}^{S}\frac{\partial}{\partial \boldsymbol{\vartheta}}\log f\left(\widehat{y}_s\left(\boldsymbol{\theta}\right) \mid \mathcal{I}_{s-1}\left(\boldsymbol{\theta}\right), \widehat{\boldsymbol{\vartheta}}\right), \tag{9.3.15}$$

where $\widehat{y}_s\left(\boldsymbol{\theta}\right)$ is the simulated data from the targeted model given the value of $\boldsymbol{\theta}$, S is the sample size. As $S \to \infty$, $m_S\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right) \to m\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right)$ almost surely. Conditional on Eq. (9.3.15),

$$\widehat{\boldsymbol{\theta}}_{EMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Dist^{EMM}\left(\boldsymbol{\theta}\right) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left[\widehat{m}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right)' \widehat{W}^{-1}\widehat{m}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right)\right], \tag{9.3.16}$$

where the weighting matrix $\widehat{W}$ is one of the remaining problems. Andersen et al. (1999) proposed to use the quasi-information matrix

$$\widehat{W} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\vartheta}}\log f\left(y_t \mid \mathcal{I}_{t-1}, \widehat{\boldsymbol{\vartheta}}\right)\frac{\partial}{\partial \boldsymbol{\vartheta}'}\log f\left(y_t \mid \mathcal{I}_{t-1}, \widehat{\boldsymbol{\vartheta}}\right), \tag{9.3.17}$$

which can be obtained directly from the QML step. Clearly, the second step is simulation-based.

EMM uses the scores as moment conditions. However, the scores do not come from the original model since those are analytically intractable. Instead, the moment conditions are based on the quasi-score of the auxiliary model. Clearly, the choice of the auxiliary model determines the efficiency of the EMM estimator and hence, is critical. If the quasi-score of the auxiliary model asymptotically spans the true score vector, the estimation achieves the full asymptotic efficiency. In other words, as the score generator gets close to the true one, the covariance matrix estimated converges that of the MLE.

Gallant et al. (1997) proposed using EMM to estimate stochastic volatility models. Andersen et al. (1999) performed an extensive Monte Carlo study of the performance of EMM to estimate the log-normal SV model. Specifically, they examined the sensitivity to different choices of the auxiliary model such as ARCH, GARCH, and EGARCH models for the score and found that the inference is sensitive to the choice of the auxiliary model when the sample size is small but robust in large samples. They further showed that the efficiency of EMM approaches that of MLE as the sample size increased. Following Andersen et al. (1999), one may choose GARCH(1,1) as the auxiliary model given by

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega + \alpha\sigma_{t-1}^2 + \beta\varepsilon_{t-1}^2. \end{aligned}$$

where we denote the parameters of the GARCH(1,1) as $\boldsymbol{\vartheta} = (\omega, \alpha, \beta)'$. We describe the details of evaluating the criterion function of EMM to estimate the log-normal SV model in Algorithm 4.

Algorithm 4 Computation of $Dist^{EMM}(\boldsymbol{\theta})$

1: Estimate the parameters of GARCH(1,1) model given the data $y_{1:T}$

$$\widehat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}} \log \tilde{L}(\boldsymbol{\vartheta}),$$

where

$$\log \tilde{L}(\boldsymbol{\vartheta}) = -\frac{T-1}{2}\log 2\pi - \frac{1}{2}\sum_{t=2}^{T}\log\sigma_t^2 - \sum_{t=2}^{T}\frac{y_t^2}{2\sigma_t^2}$$

2: Compute

$$\widehat{W} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\vartheta}}\log f\left(y_t \mid \mathcal{I}_{t-1}, \widehat{\boldsymbol{\vartheta}}\right)\frac{\partial}{\partial\boldsymbol{\vartheta}'}\log f\left(y_t \mid \mathcal{I}_{t-1}, \widehat{\boldsymbol{\vartheta}}\right),$$

where

$$\frac{\partial \log f(y_t \mid \mathcal{I}_{t-1}, \boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} = \begin{pmatrix} \frac{1}{\sigma_t^2} + \frac{y_t^2}{2\sigma_t^4} \\ \frac{\sigma_{t-1}^2}{\sigma_t^2} + \frac{y_t^2\sigma_{t-1}^2}{2\sigma_t^4} \\ \frac{\varepsilon_{t-1}^2}{\sigma_t^2} + \frac{y_t^2\varepsilon_{t-1}^2}{2\sigma_t^4} \end{pmatrix}$$

3: Simulate data from the log-normal SV model given $\boldsymbol{\theta}$ with $S = 20000$. The first one uses the simulated $\{\widehat{\varepsilon}_t, \widehat{\eta}_t\}$ to construct the $\left\{\widehat{y}_t^1\right\}_{t=1}^S$, while the latter uses the simulated $\{-\widehat{\varepsilon}_t, -\widehat{\eta}_t\}$ to generate the $\left\{\widehat{y}_t^2\right\}_{t=1}^S$.

4: Compute

$$\widehat{m}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right) = \frac{1}{2}\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial}{\partial\boldsymbol{\vartheta}}\log f\left(\widehat{y}_s^1(\boldsymbol{\theta}) \mid \mathcal{I}_{s-1}(\boldsymbol{\theta}), \widehat{\boldsymbol{\vartheta}}\right) + \frac{1}{S}\sum_{s=1}^{S}\frac{\partial}{\partial\boldsymbol{\vartheta}}\log f\left(\widehat{y}_s^2(\boldsymbol{\theta}) \mid \mathcal{I}_{s-1}(\boldsymbol{\theta}), \widehat{\boldsymbol{\vartheta}}\right)\right]$$

5: Compute

$$Dist^{EMM}(\boldsymbol{\theta}) = \widehat{m}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right)' \widehat{W}^{-1}\widehat{m}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\vartheta}}\right).$$

9.3.5 Simulated Maximum Likelihood

The GMM, QML, ECF, and EMM methods reviewed above are alternative approaches to the classic MLE when the likelihood of the model does not have an analytical expression. These methods lead to estimators that are consistent but less efficient compared to the MLE. With careful design, ECF and EMM can approach the efficiency of the MLE for only a limited set of models. Under mild regularity conditions, the MLE is consistent, asymptotically normal, and efficient. Further, due to the invariance principle, any function of the MLE is still an MLE, inheriting all its asymptotic properties. Therefore, it is worthwhile to obtain the MLE. One popular approach to resolve the intractable likelihood function is to use simulation-based methods. Shephard and Pitt (1997) and Durbin and Koopman (1997) introduced SML methods for nonlinear non-Gaussian state space models. Durham (2006), Durham (2007), Koopman et al. (2009), Skaug and Yu (2009) and Yu (2012a) used the SML method to estimate SV models and showed its reliable performance in different cases. There are other variations, like the Monte Carlo likelihood (MCL) method proposed by Durbin and Koopman (2000) and the efficient importance sampler (EIS) of Richard and Zhang (2007). For more details, one can refer to Yu (2012b). In the section, following Yu (2012b), we focus on how to use the simulated maximum likelihood (SML) to estimate the log-normal SV model. The methodology can also accommodate other complex models.

Recall the discussion in Section 9.2, the likelihood function for the log-normal SV model does not have a closed-form expression. By treating the log-volatility $h_{1:T}$ as missing data, the so-called "complete-data likelihood " is

$$p\left(y_{1:T},h_{1:T}|\boldsymbol{\theta}\right)=\prod_{t=1}^{T}\frac{1}{2\pi e^{\frac{h_t}{2}}}\exp\left[-\frac{y_t^2}{2e^{h_t}}\right]\prod_{t=2}^{T}\frac{1}{2\pi\sigma_\eta}\exp\left[-\frac{\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2}{2\sigma_\eta^2}\right]$$
$$\times\sqrt{\frac{1-\phi^2}{2\pi\sigma_\eta^2}}\exp\left[-\frac{\left(1-\phi^2\right)\left(h_1-\mu\right)^2}{2\sigma_\eta^2}\right],$$

which implies that the likelihood of the observed data is

$$L\left(\boldsymbol{\theta}\right)=\int p\left(y_{1:T},h_{1:T}|\boldsymbol{\theta}\right)dh_{1:T}=\int p\left(y_{1:T}|h_{1:T},\boldsymbol{\theta}\right)p\left(h_{1:T}|\boldsymbol{\theta}\right)dh_{1:T}.$$

Although the integration cannot be solved analytically, the integral can be approximated by the Monte Carlo integration, that is,

$$L^{SML}\left(\boldsymbol{\theta}\right)=\frac{1}{S}\sum_{s=1}^{S}p\left(y_{1:T}|h_{1:T}^{(s)},\boldsymbol{\theta}\right),\tag{9.3.18}$$

where $h_{1:T}^{(s)}$ is the s-th sample generated from the targeted model given $\boldsymbol{\theta}$. The associated SML estimator is

$$\widehat{\boldsymbol{\theta}}_{SML}=\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}Dist^{SML}\left(\boldsymbol{\theta}\right)=\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left[-L^{SML}\left(\boldsymbol{\theta}\right)\right].\tag{9.3.19}$$

In practice, the direct simple average in Eq. (9.3.18) is not efficient and very hard for convergence. Therefore, we need to find a proposal distribution $q\left(h_{1:T}|\boldsymbol{\theta}\right)$ to replace the original one and then based on importance sampling, the integral can be approximated by

$$L^{SML}\left(y_{1:T}|\boldsymbol{\theta}\right)=\frac{1}{S}\sum_{s=1}^{S}p\left(y_{1:T}|h_{1:T}^{(s)},\boldsymbol{\theta}\right)\frac{p\left(h_{1:T}^{(s)}|\boldsymbol{\theta}\right)}{q\left(h_{1:T}^{(s)}|\boldsymbol{\theta}\right)}=\frac{1}{S}\sum_{s=1}^{S}p\left(y_{1:T}|h_{1:T}^{(s)},\boldsymbol{\theta}\right)\omega^{(s)},\tag{9.3.20}$$

where $h_{1:T}^{(s)}$ is now the sth sample generated from proposal distribution $q\left(h_{1:T}|\boldsymbol{\theta}\right)$ and the weights $\omega^{(s)}$ compensate for discrepancy between $p\left(h_{1:T}^{(s)}|\boldsymbol{\theta}\right)$ and $q\left(h_{1:T}^{(s)}|\boldsymbol{\theta}\right)$. Apparently, as $S\to\infty$,

$$\frac{1}{S}\sum_{s=1}^{S}p\left(y_{1:T}|h_{1:T}^{(s)},\boldsymbol{\theta}\right)\omega^{(s)}\to\int p\left(y_{1:T}|h_{1:T},\boldsymbol{\theta}\right)\frac{p\left(h_{1:T}|\boldsymbol{\theta}\right)}{q\left(h_{1:T}|\boldsymbol{\theta}\right)}q\left(h_{1:T}|\boldsymbol{\theta}\right)dh_{1:T}.$$

In the following, we discuss one of the popular approaches to specify a "good" proposal distribution, $q\left(h_{1:T}|\boldsymbol{\theta}\right)$. Here we introduced a multivariate normal distribution based on the Laplace approximation. We then draw a sequence of independent variables from the multivariate normal distribution and approximate the integral by the sample mean of a function of the independent draws. Here the carefully selected multivariate normal distribution is served as the proposal distribution in the importance sampler. This method is called the Laplace-based importance sampler.

The Laplace approximation is to choose a multivariate normal distribution as the proposal distribution given the values of parameter $\boldsymbol{\theta}$, namely $\mathcal{MN}\left(\boldsymbol{h}_{\boldsymbol{\theta}}^{*}, -\Omega_{\boldsymbol{\theta}}^{*-1}\right)$. To select the desired multivariate normal density, we need to pin down the mean and the covariance matrix, which can be achieved by

$$\boldsymbol{h}_{\boldsymbol{\theta}}^{*} = \arg\max_{\boldsymbol{h}\in\mathbb{R}^{T}} \log p\left(y_{1:T}|\boldsymbol{h},\boldsymbol{\theta}\right), \tag{9.3.21}$$

where $p\left(y_{1:T},\boldsymbol{h}|\boldsymbol{\theta}\right)$ is defined in Eq. (9.2.5) and $\boldsymbol{h} = h_{1:T}$. Besides, the covariance matrix can be obtained by

$$\Omega_{\boldsymbol{\theta}}^{*} = \left.\frac{\partial^{2}\log p\left(y_{1:T}|\boldsymbol{h},\boldsymbol{\theta}\right)}{\partial\boldsymbol{h}\partial\boldsymbol{h}'}\right|_{\boldsymbol{h}=\boldsymbol{h}_{\boldsymbol{\theta}}^{*}}. \tag{9.3.22}$$

The SV model in Eq. (9.2.1) - (9.2.2) does not have analytical expressions for $\boldsymbol{h}_{\boldsymbol{\theta}}^{*}$ and hence numerical methods are needed. Following Shephard and Pitt (1997), Durham (2006), Hans and Yu (2014), we can use Newton's method for calculation, which is very straightforward to use in widely available programming languages such as MATLAB, Python, and R. In particular, the evaluation of $L^{SML}\left(y_{1:T}|\boldsymbol{\theta}\right)$ can be summarized in Algorithm 5.

Algorithm 5 SML estimation based on Laplace approximation

1: Initialize $\boldsymbol{h}^{0}$.

2: Calculate the $\boldsymbol{h}_{\boldsymbol{\theta}}^{*}$ according to the first-order condition

$$\frac{\partial}{\partial\boldsymbol{h}}\log p\left(y_{1:T}|\boldsymbol{h},\boldsymbol{\theta}\right) = \mathbf{0},$$

which can be achieved by Newton method in a recursive algorithm. For $j = 1$ to J

a: $\boldsymbol{h}^{j} = \boldsymbol{h}^{j-1} - \left[\Omega(\boldsymbol{h}^{j-1})\right]^{-1}\boldsymbol{h}^{j-1}$, where $\Omega(\cdot)$ is defined below,

$$\Omega\left(\boldsymbol{h}\right) = \begin{pmatrix} -\frac{1}{\sigma_{\eta}^{2}} - \frac{1}{2}\varepsilon_{1}^{2} & \frac{\phi}{\sigma_{\eta}^{2}} & \cdots & 0 & 0 \\ \frac{\phi}{\sigma_{\eta}^{2}} & -\frac{1+\phi^{2}}{\sigma_{\eta}^{2}} - \frac{1}{2}\varepsilon_{2}^{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1+\phi^{2}}{\sigma_{\eta}^{2}} - \frac{1}{2}\varepsilon_{T-1}^{2} & \frac{\phi}{\sigma_{\eta}^{2}} \\ 0 & 0 & \cdots & \frac{\phi}{\sigma_{\eta}^{2}} & -\frac{1}{\sigma_{\eta}^{2}} - \frac{1}{2}\varepsilon_{T}^{2} \end{pmatrix},$$

and $\varepsilon_{t} = y_{t}\exp\left(-\frac{h_{t}}{2}\right)$, for $t = 1,\ldots,T$.

b: Stop when $\left|\boldsymbol{h}^{j} - \boldsymbol{h}^{j-1}\right| < \delta$, and $\boldsymbol{h}_{\boldsymbol{\theta}}^{*} = \boldsymbol{h}^{j}$, where δ is some stop criterion otherwise $\boldsymbol{h}_{\boldsymbol{\theta}}^{*} = \boldsymbol{h}^{J}$.

3: Calculate $\Omega_{\boldsymbol{\theta}}^{*} = \Omega\left(\boldsymbol{h}_{\boldsymbol{\theta}}^{*}\right)$ and obtain $q\left(h_{1:T}|\boldsymbol{\theta}\right)$, which is $\mathcal{MN}\left(\boldsymbol{h}_{\boldsymbol{\theta}}^{*}, -\Omega_{\boldsymbol{\theta}}^{*-1}\right)$.

4: Simulate $h_{1:T}^{(s)} = \boldsymbol{h}^{(s)} \sim \mathcal{MN}\left(\boldsymbol{h}_{\boldsymbol{\theta}}^{*}, -\Omega_{\boldsymbol{\theta}}^{-1}\right)$ for $s = 1,\ldots,S$, and compute

$$L^{SML}\left(\boldsymbol{\theta}\right) = \frac{1}{S}\sum_{s=1}^{S} p\left(y_{1:T}|\boldsymbol{h}^{(s)},\boldsymbol{\theta}\right)\omega^{(s)}, \omega^{(s)} = \frac{p\left(\boldsymbol{h}^{(s)}|\boldsymbol{\theta}\right)}{q\left(\boldsymbol{h}^{(s)}|\boldsymbol{\theta}\right)}. \tag{9.3.23}$$

Under the assumption of finite variance, Kolmogorov's strong law of large numbers ensures the convergence of Eq. (9.3.23) to the true likelihood as $S \to \infty$. By the central limit theorem, the square root rate of convergence achieves if

$$Var\left[\frac{p\left(y_{1:T}|\boldsymbol{h}^{(s)},\boldsymbol{\theta}\right)}{q\left(\boldsymbol{h}^{(s)}|\boldsymbol{\theta}\right)}\right] < \infty, \forall s.$$

Hence, we can apply Eq. (9.3.23) to compute the likelihood value given parameter value $\boldsymbol{\theta}$. The Laplace approximation is straightforward to implement, but it will result in approximation errors. The magnitude of errors is determined not only by the distance between the true distribution and the proposal but also the value of S. The approximation error decreases as S increases. To ensure that the likelihood surface is smooth, when simulating the random samples over the parameter space, the same set of random seeds should be used.

For the estimators discussed above, the inference relies on the asymptotic normality properties of the estimators. The asymptotic behavior of the estimators may not work well in finite samples, especially when the time series is highly persistent. Therefore, inference based on asymptotic theory may lead to a misleading conclusion.

9.4 Methods in the Bayesian Domain

In the Bayesian paradigm, when we combine the likelihood function $p\left(\boldsymbol{y}|\boldsymbol{\theta}\right)$ with prior information, which is expressed as a probability distribution with density $\pi\left(\boldsymbol{\theta}\right)$, we obtain the posterior, which is another probability distribution, $\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$. The posterior is obtained by the Bayes formula,

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p\left(\boldsymbol{y}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)}{p\left(\boldsymbol{y}\right)},$$

where $p\left(\boldsymbol{y}\right)$ is the marginal density of $\boldsymbol{y}$ (see, e.g., (Robert, 2007)).

In the Bayesian domain, the estimation of the parameter $\boldsymbol{\theta}$ is based on the decision theory framework. To achieve this we firstly need to specify a loss function $\rho\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)$, which represents the loss caused by choosing the estimates of $\boldsymbol{\theta}$ as $\boldsymbol{\delta}$. The Bayesian estimator is the one minimizing posterior risk defined below,

$$\widehat{\boldsymbol{\theta}}^{Bayes} = \arg\min_{\boldsymbol{\delta}\in\boldsymbol{\Theta}} \int\int \rho\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)d\boldsymbol{\theta}d\boldsymbol{y}. \tag{9.4.1}$$

Specifically, we usually choose the quadratic loss function

$$\rho(\boldsymbol{\delta},\boldsymbol{\theta}) = \|\boldsymbol{\theta}-\boldsymbol{\delta}\|^2.$$

Then the Bayes estimator is the posterior mean $\widehat{\boldsymbol{\theta}}^{Bayes} = \bar{\boldsymbol{\theta}} = \mathbb{E}\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$.

Unlike frequentist methods, Bayesian approaches estimate the parameter by performing posterior analysis. Given the likelihood function $p\left(\boldsymbol{y}|\boldsymbol{\theta}\right)$ and the prior $\pi\left(\boldsymbol{\theta}\right)$ for parameters, the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is also available. If we can obtain random samples from the posterior distribution, we can use sample moments to estimate parameters and make statistical inference about parameters. For example, we can use the sample mean (the posterior mean) and the sample quantiles (the posterior quantiles) to calculate the point estimate and credible intervals, respectively.

While the inference of frequentist methods often relies on asymptotic theory that may or may not provide accurate approximations to the finite sample distribution, Bayesian inference is based on the posterior distribution. The exact finite sample inference provided by $\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$ can be performed in the Bayesian domain via posterior analysis. However, the Bayesian approach always

requires numerical integration or sampling techniques. Consequently, the computational cost is generally high. Besides, the model must be fully specified since the likelihood function is required.

The log-normal SV model belongs to the class of latent variable models. For latent variable models, one way to implement Bayesian estimation is to treat the unobserved variables $h_{1:T}$ as additional parameters to estimate. This technique is known as data augmentation. Another way is to directly calculate the likelihood function without data augmentation, which can be achieved by applying the sequential Monte Carlo method (SMC). We discuss how to estimate the parameters in the log-normal SV model with and without data augmentation. The procedures can accommodate other types of latent variable models. Specifically, when using data augmentation, we can estimate the model either through the single-move Markov chain Monte Carlo method (single-move MCMC, see, (Jacquier et al., 2002)) or multi-move MCMC method (multi-move MCMC). In the former, only one variable at one time is updated and in the latter, the latent vector $\boldsymbol{h}$ is sampled in a single block, which is more efficient (see, e.g., (Shephard and Pitt, 1997), (Kim et al., 1998), (Chib et al., 2002), (Liesenfeld and Richard, 2003) and (Omori et al., 2007)).

9.4.1 Single-move MCMC

When applying data augmentation for the log-normal SV model, we treat the unobserved log-volatility $\boldsymbol{h} = h_{1:T}$ as a parameter vector. By doing this, we avoid the problem of intractable likelihood function. Instead, according to the discussion in Section 9.2, the analytical expression for $(\boldsymbol{h}, \boldsymbol{\theta})$, the so-called complete-data likelihood, is given by

$$p\left(y_{1:T}|\boldsymbol{h},\boldsymbol{\theta}\right) = \prod_{t=1}^{T}\frac{1}{\sqrt{2\pi}e^{\frac{h_t}{2}}}\exp\left[-\frac{y_t^2}{2e^{h_t}}\right]\prod_{t=2}^{T}\frac{1}{\sqrt{2\pi}\sigma_\eta}\exp\left[-\frac{\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2}{2\sigma_\eta^2}\right] \times\sqrt{\frac{1-\phi^2}{2\pi\sigma_\eta^2}}\exp\left[-\frac{\left(1-\phi^2\right)\left(h_1-\mu\right)^2}{2\sigma_\eta^2}\right].$$

To perform Bayesian estimation, the prior for $\boldsymbol{\theta}$ is specified as

$$\pi\left(\boldsymbol{\theta}\right) = \pi\left(\mu\right)\pi\left(\phi\right)\pi\left(\sigma_\eta^2\right).$$

Therefore, the joint posterior density function is

$$\pi\left(\boldsymbol{\theta},\boldsymbol{h}|y_{1:T}\right) \propto \pi\left(\mu\right)\pi\left(\phi\right)\pi\left(\sigma_\eta^2\right)p\left(y_{1:T}|\boldsymbol{h},\boldsymbol{\theta}\right). \tag{9.4.2}$$

which implies that we only need to focus on the posterior of $(\boldsymbol{\theta}, \boldsymbol{h})$ and the next step is to draw samples of $(\boldsymbol{\theta}, \boldsymbol{h})$ from the posterior distribution in Eq. (9.4.2).

How to draw samples from the targeted posterior distribution is the key problem in Bayesian estimation. MCMC is one of the popular approaches in practice. Essentially, MCMC is a category of algorithms which generate a correlated sample from a Markov chain with stationary distribution being the same as the posterior distribution. The single-move MCMC was first proposed by Jacquier et al. (2002) and Kim et al. (1998) summarized the algorithm clearly. Here we use the priors in Kim et al. (1998) to illustrate how the single-move MCMC works. In the following, s indicates the s-th MCMC draws and we define the notation $h_{-t}^{(s-1)} = \left(h_1^{(s)},\ldots,h_{t-1}^{(s)},h_{t+1}^{(s-1)},h_{t+2}^{(s-1)},\ldots,h_T^{(s-1)}\right)'$. The Gibbs sampler with Metropolis-Hasting step is illustrated in Algorithm 6.

We discuss the posterior distributions used in Algorithm 6 in detail.

The Posterior for h_t, $\pi\left(h_t|\cdot\right)$

The posterior density function has the following property,

$$\pi\left(h_t|h_{-t}^{(s-1)},y_{1:T},\boldsymbol{\theta}^{(s-1)}\right) \propto f_{\mathcal{N}}\left(h_t|h_t^{*(s)},\upsilon^{2(s)}\right)p\left(y_t|h_t,\boldsymbol{\theta}^{(s-1)}\right),$$

Algorithm 6 The single-move MCMC algorithm

1: Initialize $\boldsymbol{\theta}^{(1)} = \left(\mu^{(1)}, \phi^{(1)}, \sigma_\eta^{2(1)}\right)$ and $h_t = 0$ for $t = 1, \ldots, T$.
2: Pre-specify the number of effective samples S and the burn-in samples S_{burnin}.
3: **for** $s = 2$ to $S + S_{burnin}$ **do**
4: Draw $h_t^{(s)} \sim \pi\left(h_t | h_{-t}^{(s-1)}, y_{1:T}, \boldsymbol{\theta}^{(s-1)}\right)$, for $t = 1, \ldots, T$.
5: Draw $\sigma_\eta^{2(s)} \sim \pi\left(\sigma_\eta^2 | h_{1:T}^{(s)}, \phi^{(s-1)}, \mu^{(s-1)}\right)$.
6: Draw $\phi^{(s)} \sim \pi\left(\phi | h_{1:T}^{(s)}, \mu^{(s-1)}, \sigma_\eta^{2(s)}\right)$.
7: Draw $\mu^{(s)} \sim \pi\left(\mu | h_{1:T}^{(s)}, \phi^{(s)}, \sigma_\eta^{2(s)}\right)$.
8: Collect $\left\{h_{1:T}^{(s)}, \phi^{(s)}, \mu^{(s)}, \sigma_\eta^{2(s)}\right\}_{s=S_{burnin}+1}^{S+S_{burnin}}$ as the effective samples.

where $f_{\mathcal{N}}$ is the normal density function with mean and variance as

$$h_t^{(s)*} = \mu^{(s-1)} + \frac{\phi^{(s-1)}\left[\left(h_{t-1}^{(s)} - \mu^{(s-1)}\right) + \left(h_{t+1}^{(s-1)} - \mu^{(s-1)}\right)\right]}{\left(1 + \phi^{(s-1)2}\right)},$$

$$\upsilon^{2(s)} = \frac{\sigma_\eta^{2(s-1)}}{1 + \phi^{(s-1)2}}.$$

Besides, since

$$\begin{aligned}\log p\left(y_t | h_t, \boldsymbol{\theta}\right) &= -\log 2\pi - \frac{h_t}{2} - \frac{y_t}{2e^{h_t}} \\ &\leq -\log 2\pi - \frac{1}{2}h_t - \frac{y_t^2}{2}\left[\exp\left(-h_t^*\right)\left(1 + h_t^*\right) - h_t \exp\left(-h_t^*\right)\right],\end{aligned}$$

we can find

$$\pi\left(h_t | h_{-t}^{(s-1)}, y_{1:T}, \boldsymbol{\theta}^{(s-1)}\right) \leq f_{\mathcal{N}}\left(h_t | \mu_t^{*(s)}, \upsilon^{2(s)}\right),$$

where

$$\mu_t^{*(s)} = h_t^{*(s)} + \frac{\upsilon^{2(s)}}{2}\left[y_t^2 \exp\left(-h_t^{*(s)}\right) - 1\right].$$

One may say that it is possible to directly sample the latent variable h_t from the normal proposal distribution and implement the Reject/Accept algorithm to obtain the samples. However, in practice, we usually use the Metropolis-Hasting step to ensure the effectiveness of samples. The sampling of $h_t^{(s)}$ is performed as follows.

1. Draw $\tilde{h} \sim \mathcal{N}\left(\mu_t^{*(s)}, \upsilon^{2(s)}\right)$.

2. Compute

$$\alpha = \min\left\{1, \frac{f_{\mathcal{N}}\left(h_t^{(s-1)} | \mu_t^{*(s)}, \upsilon^{2(s)}\right)}{f_{\mathcal{N}}\left(\tilde{h} | \mu_t^{*(s)}, \upsilon^{2(s)}\right)} \times \frac{f_{\mathcal{N}}\left(\tilde{h} | h_t^{*(s)}, \upsilon^{2(s)}\right) p\left(y_t | \tilde{h}, \boldsymbol{\theta}^{(s-1)}\right)}{f_{\mathcal{N}}\left(h_t^{(s-1)} | h_t^{*(s)}, \upsilon^{2(s)}\right) p\left(y_t | h_t^{(s-1)}, \boldsymbol{\theta}^{(s-1)}\right)}\right\}.$$

3. Generate $u \sim U(0, 1)$,

$$h_t^{(s)} = \begin{cases} \tilde{h}, & \text{if } u \leq \alpha, \\ h_t^{(s-1)}, & \text{if } u > \alpha. \end{cases}$$

The Posterior for σ_η, $\pi(\sigma_\eta|\cdot)$

Following Kim et al. (1998), we impose a conjugate prior

$$\sigma_\eta^2 \mid \phi, \mu \sim \mathcal{IG}(s_r/2, S_\sigma/2),$$

then σ_η^2 is sampled from an inverse-Gamma distribution,

$$\sigma_\eta^2 \mid y_{1:T}, h_{1:T}^{(s)}, \phi^{(s-1)}, \mu^{(s-1)} \sim \mathcal{IG}\left\{ \frac{T+s_r}{2}, \frac{S_\sigma + \left(h_1^{(s)} - \mu^{(s-1)}\right)^2 \left(1-\phi^{(s-1)2}\right)}{2} + \frac{\sum_{t=1}^{T-1} \left(\left(h_{t+1}^{(s)} - \mu^{(s-1)}\right) - \phi^{(s-1)}\left(h_t^{(s)} - \mu^{(s-1)}\right)\right)^2}{2} \right\}$$

where $\mathcal{IG}$ denotes the inverse-gamma distribution. Here we set $\sigma_r = 5$ and $S_\sigma = 0.01 \times \sigma_r$.

The Posterior for ϕ, $\pi(\phi|\cdot)$

For ϕ, denote $\phi = 2\phi^* - 1$, where ϕ^* follows a Beta distribution with parameters (γ_1, γ_2). Hence, the prior used for ϕ is

$$\pi(\phi) \propto \left[\frac{(1+\phi)}{2}\right]^{\gamma_1 - 1} \left[\frac{(1-\phi)}{2}\right]^{\gamma_2 - 1}, \quad \gamma_1, \gamma_2 > \frac{1}{2}$$

and has support on the interval $(-1, 1)$. Here we set $\gamma_1 = 20$ and $\gamma_2 = 1.5$. Then the posterior of ϕ is

$$\pi\left(\phi|y_{1:T}, h_{1:T}^{(s)}, \mu^{(s-1)}, \sigma_\eta^{2(s)}\right) \propto \pi(\phi)\, p\left(h_{1:T}^{(s)}|\mu^{(s-1)}, \phi, \sigma_\eta^{2(s)}\right).$$

The Metropolis-Hasting sampling procedure is illustrated as below,

1. Draw $\tilde{\phi} \sim \mathcal{N}\left(\phi^{*(s)}, S_\phi^{(s)}\right)$, where

$$\phi^{*(s)} = \frac{\sum_{t=1}^{T-1}\left(h_{t+1}^{(s)} - \mu^{(s-1)}\right)\left(h_t^{(s)} - \mu^{(s-1)}\right)}{\sum_{t=1}^{T-1}\left(h_t^{(s)} - \mu^{(s-1)}\right)^2},$$

$$S_\phi^{(s)} = \sigma_\eta^{2(s)} \left[\sum_{t=1}^{T-1}\left(h_t^{(s)} - \mu^{(s-1)}\right)^2\right]^{-1}.$$

2. Compute

$$\alpha_\phi = \min\left\{1, \frac{f_\mathcal{N}\left(\phi^{(s-1)}|\phi^{*(s)}, S_\phi^{(s)}\right)}{f_\mathcal{N}\left(\tilde{\phi}|\phi^{*(s)}, S_\phi^{(s)}\right)} \times \frac{\pi\left(\tilde{\phi}\right) p\left(h_{1:T}^{(s)}|\mu^{(s-1)}, \tilde{\phi}, \sigma_\eta^{2(s)}\right)}{\pi\left(\phi^{(s-1)}\right) p\left(h_{1:T}^{(s)}|\mu^{(s-1)}, \phi^{(s-1)}, \sigma_\eta^{2(s)}\right)}\right\}.$$

3. Generate $u \sim U(0,1)$,

$$\phi^{(s)} = \begin{cases} \tilde{\phi}, & \text{if } u \le \alpha_\phi, \\ \phi^{(s-1)}, & \text{if } u > \alpha_\phi. \end{cases}$$

The Posterior for μ, $\pi(\mu|\cdot)$

We can impose a normal prior for μ ,

$$\pi\left(\mu|h_{1:T}^{(s)},\phi^{(s)},\sigma_{\eta}^{2(s)}\right)\propto\pi\left(h_{1:T}^{(s)}|\mu,\phi^{(s)},\sigma_{\eta}^{2(s)}\right)\pi\left(\mu\right),$$

where $\pi(\mu)$ is the prior $\mathcal{N}\left(0,\upsilon_{\mu}^{2}\right)$. It can be derived that

$$\mu|h_{1:T}^{(s)},\phi^{(s)},\sigma_{\eta}^{2(s)}\sim\mathcal{N}\left(\mu^{*(s)},\sigma_{\mu}^{2(s)}\right),$$

where

$$\mu^{*(s)}=\sigma_{\mu}^{2(s)}\left\{\frac{\left(1-\phi^{(s)2}\right)}{\sigma_{\eta}^{2(s)}}h_{1}^{(s)}+\frac{\left(1-\phi^{(s)}\right)}{\sigma_{\eta}^{2(s)}}\sum_{t=1}^{T-1}\left(h_{t+1}^{(s)}-\phi^{(s)}h_{t}^{(s)}\right)\right\},$$

$$\sigma_{\mu}^{2(s)}=\frac{\upsilon_{\mu}^{2}\sigma_{\eta}^{2(s)}}{\sigma_{\eta}^{2(s)}+\upsilon_{\mu}^{2}\left[(T-1)(1-\phi^{(s)})^{2}+\left(1-\phi^{(s)2}\right)\right]}.$$

There we can directly sample μ from the normal distribution.

After iterating until $S+S_{burnin}$ we can obtain a series of correlated samples from the posterior. To remove the impact of initial values, we must perform the burn-in step, that is, remove the first S_{burnin} samples. The final remaining samples are treated as the effective samples from the posterior distribution and can be used for inference. In particular, we can directly compute the estimator according to Eq. (9.4.1). If we choose the quadratic loss function, the estimator is the posterior mean, that is,

$$\bar{\boldsymbol{\theta}}=\frac{\sum_{s=1}^{S}\boldsymbol{\theta}^{(s)}}{S},\tag{9.4.3}$$

and we can use the posterior samples to construct posterior intervals for parameters.

9.4.2 Multi-move MCMC

As pointed out by Kim et al. (1998), the samples generated by single-move MCMC have high correlation and hence, are not efficient. Multi-move algorithms are more efficient. Shephard and Pitt (1997), Kim et al. (1998), Chib et al. (2002), Liesenfeld and Richard (2003) and Omori et al. (2007) have developed various multi-move algorithms to estimate univariate and multivariate SV models. Essentially, the multi-move algorithms sample the latent volatility $h_{1:T}$ in a single block. Here we use the algorithm in Kim et al. (1998) for illustration.

As discussed in Section 9.2, the observation equation can be rewritten as

$$\log\left(y_{t}^{2}\right)=h_{t}+\log\left(\varepsilon_{t}^{2}\right),\tag{9.4.4}$$

where $\log\left(\varepsilon_{t}^{2}\right)$ is the logarithm of the chi-squared random variable with 1 degree of freedom and hence, the Kalman filter cannot be applied. In Section 9.3.2, we use a normal distribution with mean -1.27 and variance $\pi^{2}/2$ to approximate $\log\left(\varepsilon_{t}^{2}\right)$ and the Kalman filter to conduct the QML estimation. However, a better approximation of the logarithm of the chi-squared distribution is a mixture of normal distributions given by

$$\sum_{i=1}^{K}w_{i}f_{\mathcal{N}}\left(\varepsilon_{t}\mid m_{i},v_{i}^{2}\right),$$

which is a mixture of K normal densities $f_{\mathcal{N}}$ with component weights w_i. The constants $\left\{w_{i},m_{i},v_{i}^{2}\right\}$ are pre-determined (see, e.g., (Kim et al., 1998)) and summarized in Table 9.1.

Table 9.1: Selection of $\left\{w_i, m_i, v_i^2\right\}$

i	w_i	m_i	v_i^2
1	0.00730	−10.12999	5.79596
2	0.10556	−3.97281	2.61369
3	0.00002	−8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	−1.08819	1.26261

We then rewrite the density of the mixture of normal distributions in terms of a component indicator variable q_t at each time t, such that Eq. (9.4.4) is now,

$$\begin{aligned} \log\left(y_t^2\right) &= h_t + \tilde{\varepsilon}_t, && (9.4.5)\\ \tilde{\varepsilon}_t | q_t = i &\sim \mathcal{N}\left(m_i - 1.2704, v_i^2\right), i = 1, \ldots, K, && (9.4.6)\\ P\left(q_t = i\right) &= w_i. && (9.4.7) \end{aligned}$$

Therefore, by introducing the indicator vector $\boldsymbol{q} = (q_1, \ldots, q_T)'$ and the mixture normal distribution, we can directly use the Kalman filter and smoother since we can use a normal random variable to replace the $\log \varepsilon_t^2$ at every time t given the values of q_t. The log-volatility $h_{1:T}$ can now be treated as state variables in a Gaussian state-space model and updated in a block. In other words, we only need to add the updating of indicator vector $\boldsymbol{q}$ and use the Kalman filter and smoother to update $h_{1:T}$ in the single-move MCMC. By doing this, the samples obtained are much more efficient relative to those obtained as in Section 9.4.1. The algorithm is summarized in Algorithm 7.

Algorithm 7 The multi-move MCMC algorithm

1: Initialize $\boldsymbol{\theta}^{(1)} = \left(\mu^{(1)}, \phi^{(1)}, \sigma_\eta^{2(1)}\right)$ and $q_t^{(1)} = 1$ for $t = 1, \ldots, T$.
2: Pre-specify the number of effective samples S and the burn-in samples S_{burnin}.
3: **for** $s = 2$ to $S + S_{burnin}$ **do**
4: Draw $\boldsymbol{h}^{(s)} \sim \pi\left(\boldsymbol{h}|\tilde{\boldsymbol{x}}, \boldsymbol{q}^{(s-1)}, \boldsymbol{\theta}^{(s-1)}\right)$.
5: Draw $\boldsymbol{q}^{(s)} \sim \pi\left(\boldsymbol{q}|\tilde{\boldsymbol{x}}, \boldsymbol{h}^{(s)}\right)$.
6: Draw $\sigma_\eta^{2(s)} \sim \pi\left(\sigma_\eta^2|\boldsymbol{h}^{(s)}, \phi^{(s-1)}, \mu^{(s-1)}\right)$.
7: Draw $\phi^{(s)} \sim \pi\left(\phi|\boldsymbol{h}^{(s)}, \mu^{(s-1)}, \sigma_\eta^{2(s)}\right)$.
8: Draw $\mu^{(s)} \sim \pi\left(\mu|\boldsymbol{h}^{(s)}, \phi^{(s)}, \sigma_\eta^{2(s)}\right)$.
9: Collect $\left\{\boldsymbol{h}^{(s)}, \phi^{(s)}, \mu^{(s)}, \sigma_\eta^{2(s)}\right\}_{s=S_{burnin}+1}^{S+S_{burnin}}$ as the effective samples.

In the following, we discuss how to sample the $\boldsymbol{h}$ and $\boldsymbol{q}$.

The Posterior for $\boldsymbol{h}$, $\pi\left(\boldsymbol{h}|\cdot\right)$

The posterior for $\boldsymbol{h}$ conditional on $\boldsymbol{q}$ can be sampled by the Kalman filter. Specifically, given $\boldsymbol{q}$, the state-space model is

$$\begin{aligned} \tilde{x}_t - \mu^{(s)} &= z_t + \tilde{\varepsilon}_t,\\ z_t &= \phi^{(s)} z_{t-1} + \sigma_\eta^{(s)} \eta_t. \end{aligned}$$

Therefore, conditional on $\boldsymbol{q}$, we can sample $\boldsymbol{h}^{(s)}$ at the s-th sampling as below.

1. At the s-th sampling, initialize $\boldsymbol{h} = \mu^{(s)}$, i.e., $z_t = 0$, for $t = 1, \ldots T$.
2. Kalman filter: let $z_t|y_{1:t-1} \sim \mathcal{N}(c_{t-1}, C_{t-1})$, conditional on $q_t = i$, $\tilde{\varepsilon}_t \sim \mathcal{N}\left(m_i, v_i^2\right)$,
$$a_t = \phi^{(s)} m_{t-1}, R_t = \phi^{(s)2} C_{t-1} + \sigma_\eta^{2(s)},$$
$$f_t = a_t + m_i - 1.2704, Q_t = R_t + v_i^2,$$
$$c_t = a_t + R_t Q_t^{-1}\left(\tilde{x}_t - \mu^{(s)} - f_t\right), C_t = R_t - R_t^2/Q_t.$$
3. Smoother: let $z_t|y_{1:T} \sim \mathcal{N}(d_t, D_t)$, conditional on $q_t = i$, $\tilde{\varepsilon}_t \sim \mathcal{N}\left(m_i, v_i^2\right)$,
$$d_t = c_t + \phi^{(s)} C_t \left(c_{t+1} - a_{t+1}\right)/R_{t+1}, D_t = C_t - \phi^{(s)2} C_t^2 \left(R_{t+1} - D_{t+1}\right)/R_{t+1}^2.$$
4. Draw $z_t \sim \mathcal{N}(d_t, D_t)$ for $t = 1, \ldots, T$.
5. Recover $h_t^{(s)} = z_t + \mu^{(s)}$.

The Posterior for $\boldsymbol{q}$, $\pi(\boldsymbol{q}|\cdot)$

The posterior for q_t has the property,

$$Pr\left(q_t^{(s)} = i \middle| \tilde{x}_t, h_t^{(s)}\right) \propto w_i f_{\mathcal{N}}\left(\tilde{x}_t | h_t^{(s)} + m_i, v_i^2\right), i = 1, \ldots, K.$$

Therefore, it can be sampled by using the probability mass function.

In summary, the above-reviewed MCMC methods perform well in estimating the SV model. Moreover, the posterior mean of $\boldsymbol{h}$ is easy to obtain, which serves as the smooth estimate of log-volatility as it is conditional on $\boldsymbol{y}$. However, these MCMC methods require full knowledge of the likelihood function and are time-consuming. Besides, we must usually design the Gibbs MCMC estimation procedure case-by-case for a specific data generating process.

9.4.3 Parameter Learning via Sequential Monte Carlo Methods

The single-move MCMC and multi-move MCMC algorithms are all based on data augmentation, which makes the complete-data likelihood function analytically tractable. A class of alternative Bayesian approaches uses the particle filter, or the sequential Monte Carlo (SMC) method. SMC is mainly used to analyze state-space models, in which there is a hidden state of interest. Inference regarding the states depends on the known parameter $\boldsymbol{\theta}$ and the noisy observations $\boldsymbol{y}$. Different from the classic Kalman filter, which is optimally designed for linear and Gaussian state-space models, the particle filter also works well for nonlinear and non-Gaussian state-space models. Essentially, the particle filter uses a set of random samples ("particles") with associated weights to represent the density in the prediction and updating steps.

However, the target here is to estimate the parameter $\boldsymbol{\theta}$ instead of $\boldsymbol{h}$. To estimate parameters, the particle MCMC (PMCMC) proposed by Andrieu et al. (2010) combines the MCMC method and SMC. They provided a generic solution to estimate parameters for a variety of models. While it is flexible, Fulop and Li (2013) and Fulop et al. (2015) emphasized that sequential learning (i.e., filtering as opposed to smoothing) is preferred in financial and economic applications, in which the states and parameters are updated as new observations arrive. Therefore, we explain how to use the sequential learning algorithm combined with the SMC method. For convenience, we divide the estimation procedure into two parts. The first is the bootstrap filter (see, e.g., (Gordon, 1993), (Kitagawa, 1996)) for the inference of hidden state $\boldsymbol{h}$, the other is the parameter learning as in Fulop and Li (2013). Estimates are computed based on the particles and weights. As the number of samples generated becomes large, this simulation-based empirical distribution is equivalent to the true distribution.

Sequential Monte Carlo Method

In this section, we show how to use the sequential Monte Carlo method, or particle filter to filter the log-volatility in the log-normal SV model. This filter necessitates simulation from the transition density $p\left(h_{t+1}|h_t, y_t\right)$ and computation of the measurement density $p\left(y_t|h_t\right)$. Suppose we have a set of "particles", $h_t^{(1)}, \ldots, h_t^{(M)}$ with associated discrete probability masses $w_t^{(1)}, \ldots, w_t^{(M)}$. Based on the Bayes rule, SMC algorithms propagate and update particles to yield a sample from $p\left(h_t|y_{1:t}, \boldsymbol{\theta}\right)$ according to the following filter density,

$$p\left(h_t|y_{1:t}, \boldsymbol{\theta}\right) \propto p\left(y_t|h_t, \boldsymbol{\theta}\right) \int p\left(h_t|h_{t-1}, y_{t-1}, \boldsymbol{\theta}\right) p\left(h_{t-1}|y_{1:t-1}, \boldsymbol{\theta}\right) dh_{t-1}.$$

Following Chopin and Papaspiliopoulos (2020), the bootstrap particle filter algorithm is outlined below in Algorithm 8. For the log-normal SV model Eq. (9.2.1) - (9.2.2), we can define

$$p(h_1|\boldsymbol{\theta}) \equiv \mathcal{N}\left(0, \frac{1}{1-\phi^2}\sigma_\eta^2\right),$$

$$p\left(h_t|h_{t-1}, \boldsymbol{\theta}\right) \equiv \mathcal{N}\left(\mu + \phi\left(h_{t-1} - \mu\right), \sigma_\eta^2\right),$$

$$p\left(y_t|h_t, \boldsymbol{\theta}\right) \equiv \frac{1}{\sqrt{2\pi}e^{\frac{h_t}{2}}} \exp\left[-\frac{y_t^2}{2e^{h_t}}\right].$$

The estimator for the log-likelihood is

$$\log \widehat{L}_{SMC}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \widehat{p}\left(y_t|y_{1:t-1}, \boldsymbol{\theta}\right) = \sum_{t=1}^{T} \log\left(\sum_{i=1}^{M} \frac{\tilde{\omega}_t^{(i)}}{\sum_{i=1}^{M} \tilde{\omega}_{t-1}^{(i)}}\right),$$

where $\tilde{\omega}_t^{(i)}$ are the weights for the particles at time t.

The particle filter is flexible for a variety of models. However, the accuracy depends on the number of particles (M). The larger M is, the better the approximation. Given M, if the sample size is large, the computation will be intensive. There are many other variations of the particle filter, which are similar to the bootstrap particle filter, like sampling importance resampling filter of Gordon et al. (1993), auxiliary sampling importance resampling filter of Pitt and Shephard (1999), and regularized particle filter of Smith (2013). They can improve certain aspects of the performance, but face some trade-offs between the computational burden and restricted model forms.

Parameter Learning via Marginal Resample-move

SMC for state filtering is relatively straightforward, but the goal of drawing $\boldsymbol{\theta}$ from $\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$ sequentially is still difficult. If we simply apply a particle filter over $\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$ while keeping $\boldsymbol{\theta}$ invariant, the algorithm will quickly fail and result in particle depletion. Following Fulop and Li (2013), by approximately integrating out the states using a SMC as in Section 9.4.3, we can break up the interdependence of the log-volatility $\boldsymbol{h}$ and the static parameters $\boldsymbol{\theta}$. In what follows, we apply a Bayesian resample-move algorithm according to Gilks and Berzuini (2001) and Chopin (2002) after the marginalization. For more details, one can also refer to Chopin et al. (2013).

Conditional on $\boldsymbol{\theta}$, we denote the particles for the log-volatility h_t at time t as $\mathcal{S}_t = \left\{h_t^{(i)}, \omega_t^{(i)}, \tilde{\omega}_t^{(i)}, i = 1, \ldots, M\right.$
We denote $\mathcal{S}_t$ associated with the n-th particles $\boldsymbol{\theta}^{(n)}$ as $\mathcal{S}_t^{(n)}$ and outline the procedure in Algorithm 9. In the algorithm, the proposal distribution $q_t\left(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}^{(n)}\right)$ is used to keep the particles $\boldsymbol{\theta}^{(n)}$ away from being static all the time. In practice, following Fulop and Li (2013), we can define $q_t\left(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}^{(n)}\right) \equiv \mathcal{MN}\left(\boldsymbol{\mu}, \Sigma\right)$ where $\boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N} \tilde{\boldsymbol{\theta}}^{(n)}$, $\Sigma = \frac{1}{N}\sum_{n=1}^{N}\left(\tilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\mu}\right)\left(\tilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\mu}\right)'$.

Algorithm 8 The Bootstrap Particle Filter

Draw $h_1^{(i)} \sim p(h_1|\boldsymbol{\theta})$ for $i = 1, \ldots, M$ and initialize

$$\omega_1^{(i)} = \frac{\tilde{\omega}_1^{(i)}}{\sum_{i=1}^{M} \tilde{\omega}_1^{(i)}}, \quad \tilde{\omega}_1^{(i)} = p\left(y_1|\widetilde{h}_1^{(i)}, \boldsymbol{\theta}\right)$$

for $t = 2$ to T **do**

  Calculate the effective sample size (ESS),

$$ESS_h = \frac{1}{\sum_{i=1}^{M} \omega_{t-1}^{(i)2}},$$

  if $ESS_h < c_h$ **then**

    Draw $h_{t-1}^{(i)}$ by resampling according to $\omega_{t-1} = \left(\omega_{t-1}^{(1)}, \ldots, \omega_{t-1}^{(M)}\right)$.

    Change

$$\tilde{\omega}_{t-1}^{(i)} = 1, \omega_{t-1}^{(i)} = \frac{1}{M},$$

  else

    $h_{t-1}^{(i)}, \omega_{t-1}^{(i)}$ and $\tilde{\omega}_{t-1}^{(i)}$ are unchanged.

  Draw $\widetilde{h}_t^{(i)} \sim p\left(h_t|h_{t-1}^{(i)}, \boldsymbol{\theta}\right)$, for $i = 1, \ldots, M$.

  Calculate normalized weights,

$$\omega_t^{(i)} = \frac{\tilde{\omega}_t^{(i)}}{\sum_{i=1}^{M} \tilde{\omega}_t^{(i)}}, \quad \tilde{\omega}_t^{(i)} = p\left(y_t|\widetilde{h}_t^{(i)}, \boldsymbol{\theta}\right) \tilde{\omega}_{t-1}^{(i)}$$

  Calculate

$$\widehat{p}\left(y_t|y_{1:t-1}, \boldsymbol{\theta}\right) = \sum_{i=1}^{M} f\left(y_t|\widetilde{h}_t^{(i)}, , \boldsymbol{\theta}\right) \frac{\tilde{\omega}_{t-1}^{(i)}}{\sum_{i=1}^{M} \tilde{\omega}_{t-1}^{(i)}} = \sum_{i=1}^{M} \frac{\tilde{\omega}_t^{(i)}}{\sum_{i=1}^{M} \tilde{\omega}_{t-1}^{(i)}}.$$

Finally we can collect the joint particles from $p\left(\boldsymbol{\theta}, \boldsymbol{h}|y_{1:T}\right)$, namely, $\left\{\boldsymbol{h}^{(n,i)}, \boldsymbol{\theta}^{(n)}\right\}_{n=1,\ldots,N, i=1,\ldots,M}$. For any function $g\left(\boldsymbol{\theta}, \boldsymbol{h}\right)$, its posterior expectation can be approximated by

$$\mathbb{E}\left(g\left(\boldsymbol{\theta}, \boldsymbol{h}\right)|y_{1:t}\right) \approx \sum_{n=1}^{N}\sum_{i=1}^{M} W_t^{(n)} \omega_t^{(n,i)} g\left(\boldsymbol{\theta}^{(n)}, \boldsymbol{h}^{(n,i)}\right). \tag{9.4.8}$$

Therefore, based on Eq. (9.4.8), given the quadratic loss function, the Bayesian estimator of parameters is

$$\bar{\boldsymbol{\theta}} = \sum_{n=1}^{N} W_T^{(n)} \boldsymbol{\theta}^{(n)}.$$

Algorithm 9 Bayesian Parameter Learning

Draw $\boldsymbol{\theta}^{(n)} \sim \pi\left(\boldsymbol{\theta}\right)$, for $n = 1, \ldots, N$ and initialize the particle weights $\tilde{W}_1^{(n)} = 1$.
Conditional on y_1, run Algorithm 8 to obtain $\mathcal{S}_1^{(n)}$, $n = 1, \ldots, N$.
for $t = 2$ to T **do**
    At the beginning, $\left\{\boldsymbol{\theta}^{(n)}, \mathcal{S}_{t-1}^{(n)}, \tilde{W}_{t-1}^{(n)}, W_{t-1}^{(n)}\right\}$ are known for $n = 1, \ldots N$.
    Run Algorithm 8 to obtain $\mathcal{S}_t^{(n)}$.
    Calculate

$$\tilde{W}_t^{(n)} = \tilde{W}_{t-1}^{(n)} \widehat{p}\left(y_t|y_{1:t-1}, \boldsymbol{\theta}^{(n)}\right), W_t^{(n)} = \frac{\tilde{W}_t^{(n)}}{\sum_{n=1}^{N} \tilde{W}_t^{(n)}},$$

$$\widehat{p}\left(y_{1:t}|\boldsymbol{\theta}^{(n)}\right) = \widehat{p}\left(y_{1:t-1}|\boldsymbol{\theta}^{(n)}\right) \widehat{p}\left(y_t|y_{1:t-1}, \boldsymbol{\theta}^{(n)}\right).$$

    Calculate the effective sample size

$$ESS_{\theta,t} = \frac{1}{\sum_{n=1}^{N} W_t^{(n)2}}.$$

    if $ESS_{\theta,t} < B_1$ **then**
        Resample the particles proportional to $W_t^{(n)}$, denoted as $\tilde{\boldsymbol{\theta}}^{(n)}$.
        Update the weights $\tilde{w}_t^{(n)} = 1$, for $n = 1, \ldots, N$.
        Move the particles via the Metropolis-Hasting step:
        for $n = 1$ to N **do**
            Draw $\boldsymbol{\theta}^{(n)*} \sim q_t\left(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}^{(n)}\right)$.
            Calculate

$$\alpha = \min\left\{1, \frac{p\left(\boldsymbol{\theta}^{(n)*}\right) \hat{p}\left(y_{1:t} \mid \boldsymbol{\theta}^{(n)*}\right)}{p\left(\tilde{\boldsymbol{\theta}}^{(n)}\right) \hat{p}\left(y_{1:t} \mid \tilde{\boldsymbol{\theta}}^{(n)}\right)} \frac{q_t\left(\tilde{\boldsymbol{\theta}}^{(n)} \mid \boldsymbol{\theta}^{(n)*}\right)}{q_t\left(\boldsymbol{\theta}^{(n)*} \mid \tilde{\boldsymbol{\theta}}^{(n)}\right)}\right\}.$$

            Generate $u \sim U\left(0, 1\right)$,

$$\boldsymbol{\theta}^{(n)} = \begin{cases} \boldsymbol{\theta}^{(n)*}, & \text{if } u \leq \alpha, \\ \tilde{\boldsymbol{\theta}}^{(n)}, & \text{if } u > \alpha. \end{cases}$$

9.5 Conclusion

In this chapter, we reviewed and illustrated several methods for estimating the discrete-time SV model. The methods reviewed were either in the frequentist domain or in the Bayesian domain.

Due to the intractable likelihood function, classic MLE estimation is difficult, and many alternatives have been proposed in the literature. The frequentist methods include GMM, QML, ECF, EMM, and SML. The frequentist estimators are extreme estimates obtained by minimizing a distance measure function. Additionally, inference regarding frequentist estimators relies on the asymptotic normality. Among these methods, GMM, QML, ECF, and EMM avoid calculating the likelihood. They generally suffer from a loss of efficiency. The SML method instead approximates the likelihood via simulation. The asymptotic normality of the frequentist methods may not hold in finite samples, especially when the volatility process is highly persistent.

In contrast to frequentist methods, Bayesian inference is conducted based on the posterior distribution. We reviewed three Bayesian methods for estimating the log-normal SV model. The first two methods use the data augmentation technique and the third one applies SMC. Data augmentation treats the log-volatility as parameters and hence, the likelihood is analytically tractable, facilitating posterior sampling via MCMC. Specifically, we reviewed two types of data-augmentation-based MCMC methods. One updates the log-volatility step-by-step and is referred to as single-move MCMC. The other draws the log-volatility in a block and is known as multi-move MCMC. Without data augmentation, we introduced the Bayesian parameter learning algorithm combining the SMC for the hidden states (log-volatility) and a marginal resample-move step for the parameters. Bayesian parameter learning is recursive and generic and hence is suitable for financial time series. However, compared with frequentist methods, the computation of Bayesian estimation is always very intensive.

Although we reviewed methods that estimate volatility using return observed at low frequency, many of the reviewed methods are applicable when we use realized volatility to estimate integrated volatility or spot volatility. This is because log realized volatility is an estimate of log integrated volatility and hence, a nonlinear state-space model is needed.

Bibliography

Aït-Sahalia, Y. and J. Jacod (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature 50*(4), 1007–50.

Andersen, T. G. and L. Benzoni (2010). Stochastic volatility. *CREATES Research Paper* (2010-10).

Andersen, T. G., H.-J. Chung, and B. E. Sørensen (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of econometrics 91*(1), 61–87.

Andersen, T. G. and B. E. Sørensen (1996). Gmm estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business & Economic Statistics 14*(3), 328–352.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 269–342.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Broto, C. and E. Ruiz (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys 18*(5), 613–649.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*(2), 281–316.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika 89*(3), 539–552.

Chopin, N., P. E. Jacob, and O. Papaspiliopoulos (2013). SMC^2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society Series B 75*(3), 397–426.

Chopin, N. and O. Papaspiliopoulos (2020). *An Introduction to Sequantial Monte Carlo*. Springer Series in Statistics.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*(4), 929–952.

Durbin, J. (2000). The foreman lecture: The state space approach to time series analysis and its potential for official statistics (with discussion). *Australian & New Zealand Journal of Statistics 42*(1), 1–23.

Durbin, J. and S. J. Koopman (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika 84*(3), 669–684.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(1), 3–56.

Durham, G. B. (2006). Monte Carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models. *Journal of Econometrics 133*(1), 273–305.

Durham, G. B. (2007). SV mixture models with application to S&P 500 index returns. *Journal of Financial Economics 85*(3), 822–856.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society 50*(4), 987–1007.

Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique 18*(2), 155–161.

Fulop, A. and J. Li (2013). Efficient learning via simulation: A marginalized resample-move approach. *Journal of Econometrics 176*(2), 146–161.

Fulop, A., J. Li, and J. Yu (2015). Self-exciting jumps, learning, and asset pricing implications. *The Review of Financial Studies 28*(3), 876–912.

Gallant, A. R., D. Hsieh, and G. Tauchen (1997). Estimation of stochastic volatility models with diagnostics. *Journal of econometrics 81*(1), 159–192.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric theory 12*, 657–681.

Ghysels, E., A. C. Harvey, and E. Renault (1996). 5 stochastic volatility. *Handbook of statistics 14*, 119–191.

Gilks, W. R. and C. Berzuini (2001). Following a moving target—Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(1), 127–146.

Gordon, N. (1993). *Bayesian methods for tracking*. Doctoral Dissertation, Imperial College London (University of London).

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F-radar and signal processing 140*(2), 107–113.

Hans, S. and J. Yu (2014). A flexible and automatic likelihood based framework for inference in stochastic volatility models. *Computational Statistics and Data Analysis 76*, 642–654.

Hansen, L. P. and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society 50*(5), 1269–1286.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *The Review of Economic Studies 61*(2), 247–264.

Jacquier, E., N. G. Polson, and P. E. Rossi (2002). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 20*(1), 69–87.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies 65*(3), 361–393.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of computational and graphical statistics 5*(1), 1–25.

Knight, J. L. and S. E. Satchell (1997). The cumulant generating function estimation method: implementation and asymptotic efficiency. *Econometric Theory 13*, 170–184.

Knight, J. L., S. E. Satchell, and J. Yu (2002). Estimation of the stochastic volatility model by the empirical characteristic function method. *Australian & New Zealand Journal of Statistics 44*(3), 319–335.

Knight, J. L. and J. Yu (2002). Empirical characteristic function in time series estimation. *Econometric Theory 18*, 691–721.

Koopman, S. J., N. Shephard, and D. Creal (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics 149*(1), 2–11.

Liesenfeld, R. and J.-F. Richard (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance 10*(4), 505–531.

Melino, A. and S. M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of econometrics 45*(1-2), 239–265.

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

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association 94*(446), 590–599.

Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*(2), 1385–1411.

Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Volume 2. Springer Science & Business Media.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics 63*(1), 289–306.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika 84*(3), 653–667.

Skaug, H. and J. Yu (2009). Automated likelihood based inference for stochastic volatility models. *Working Paper*.

Smith, A. (2013). *Sequential Monte Carlo methods in practice*. Springer Science & Business Media.

Taylor, S. J. (1986). *Modelling financial time series*. New York: John Wiley & Sons.

Wang, X. and J. Yu (2022). Latent local-to-unity models. *Working Paper, Singapore Management University*.

Yu, J. (2002). Forecasting volatility in the new zealand stock market. *Applied Financial Economics 12*(3), 193–202.

Yu, J. (2012a). A semiparametric stochastic volatility model. *Journal of Econometrics 167*(2), 473–482.

Yu, J. (2012b). Simulation-based estimation methods for financial time series models. In *Handbook of Computational Finance*, pp. 401–435. Research Collection School Of Economics.