CrossMark

VIX Forecast Under Different Volatility Specifications

Ying Wang[1] · **Hoi Ying Wong**[1]

Abstract Stochastic volatility (SV) models are theoretically more attractive than the GARCH type of models as it allows additional randomness. The classical SV models deduce a continuous probability distribution for volatility so that it does not admit a computable likelihood function. The estimation requires the use of Bayesian approach. A recent approach considers discrete stochastic autoregressive volatility models for a bounded and tractable likelihood function. Hence, a maximum likelihood estimation can be achieved. This paper proposes a general approach to link SV models under the physical probability measure, both continuous and discrete types, to their processes under a martingale measure. Doing so enables us to deduce the close-form expression for the VIX forecast for the both SV models and GARCH type models. We then carry out an empirical study to compare the performances of the continuous and discrete SV models using GARCH models as benchmark models.

Keywords Stochastic volatility · GARCH · DSARV · Continuous SV · VIX · Forecast

1 Introduction

There has been a growing interest in VIX computed from market prices for S&P 500 options, as a volatility derivative, since the Chicago Board Options Exchange (CBOE) introduced trading in futures and options on S&P 500 implied volatility

✉ Hoi Ying Wong
hywong@sta.cuhk.edu.hk

Ying Wang
rachelwang415@gmail.com

[1] Department of Statistics, The Chinese University of Hong Kong, Hong Kong, Hong Kong

🐾 Springer

index (VIX) in 2004 and 2006, repectively, allowing volatility to be treated as an asset class. The VIX-related literature includes Carr and Wu (2006), Dupire (2006), Psychoyios and Skiadopoulos (2006), Dotsis et al. (2007), Becker et al. (2009), and Lin and Chang (2010). The growing interest in forecasting on the VIX index brings the necessity of choosing a better model for volatility. Kanniainen et al. (2014) valuate S&P 500 options using three popular GARCH models with VIX data. They find a joint maximum likelihood estimation (MLE) with returns and VIX improves option pricing performance, in contrast to traditional returns-based MLE. They also find that non-affine models clearly outperform affine models, consistently with the existing research.

GARCH models of Bollerslev (1986) are gaining popularity in practice probably because the likelihood function can be expressed in close form, making the maximum likelihood estimation (MLE) of the model parameters possible. Therefore, GARCH models is a natural candidate to predict VIX as VIX is essentially related to the options' implied volatility. However, Nelson (1991) finds that a random oscillatory behavior of the conditional variance process is missing in the GARCH type models. Besides, these types of models can only generate a volatility skew in some option markets whereas a U-shaped volatility smile is observed in practice.

Heston (1993) extends the SV model to have a leverage effect by allowing a non-zero correlation between the asset return and its volatility. Hull and White (1987) adds an additional stochastic process in the volatility of GARCH-type models which is unobserved, introducing the stochastic volatility (SV) model. Heston (1993) extends the SV model by allowing a leverage effect, a non-zero correlation, between the asset return and its volatility.

Unfortunately, the estimation of the SV models is a highly challenging task as the models do not have the likelihood function in close form. Therefore, the Bayesian framework becomes a useful alternative. The implementation of Bayesian methods usually requires the construction of a Markov Chain Monte Carlo (MCMC) simulation. Jacquier et al. (1994) analyze SV models with a leverage effect by adopting the Gibbs sampling scheme. Shephard and Pitt (1997) employ the Metropolis–Hastings scheme for the same problem. Meyer and Yu (2000), Yu (2005) also make some generalizations. Wang et al. (2016) propose an SV model with scale mixture of normal type of error distributions, satisfying that the historical model can be calibrated using Bayesian inference from historical data and that it can easily be transformed into a risk neutral model useful to estimate option prices. The implementation of Bayesian approach on continuous SV models could be time consuming and is criticized for the use of a prior information. Yet there are frequentist approaches of estimating the continuous SV in mean (SVM) models based on Monte Carlo simulation methods like Koopman and Hol-Uspensky (2002).

Recent advances attempt to discretize the continuous volatility to a finite set of volatility states. Cordis and Kirby (2014) develop a class of discrete stochastic autoregressive volatility (DSARV) models using Markov chain methods, allowing a low-dimensional state space for the volatility, which greatly improves the computational tractability. They can be easily estimated through a recursively computed likelihood function shown by Hamilton (1989, 1990) at first, and can accommodate features such as volatility asymmetry and time-varying volatility persistence. The

DSARV models share some structural similarities with both the Markov-switching multi-fractal (MSM) models of Calvet and Fisher (2004) and the component-driven regime-switching (CDRS) models of Fleming and Kirby (2013) but dispute the tight parameterization and let the volatility follows a low-order autoregressive process, making it feasible to be developed into various multivariate versions. In the paper, several variations within the class of DSARV models are compared by their performances of forecasting volatility. However, they solely consider the models under physical measure.

In this paper, we examine continuous SV model, DSARV model, and the GARCH type models in forecasting the VIX index. We propose a transformation to bridge between the physical and risk-neutral probability measures so that the three types of models can predict the VIX as the risk-neutral expectation on the future volatility. We examine the models' performances through out-of-sample tests. By "out-of-sample" here, we refer to the tex sample which is out of the estimation sample, instead of which from the time series perspective. Our goal is to determine which model explains the volatility better. Given a time series of stock prices sampled from time points $t_1, t_2, \ldots, t_n$, we predict the VIX at time t_n and compare it with the market observed VIX at t_n. This is important for derivatives pricing purpose. For example, the European call option at time t_n needs the volatility at t_n as an input. When the underlying asset does not have a liquid option market for calibration purpose, the volatility at t_n predicted by the time series of the underlying asset prices becomes important. We use the liquid index option market which has observed VIX data to test for volatility prediction of stochastic model based on historical index values. Unlike Cordis and Kirby (2014) who use VIX data as input to forecast volatility, we do not use the VIX data as our sample input but, rather, output the VIX to compare with the observed VIX.

Therefore, we derive the risk-neutral dynamic of the DSARV model, so that the model can be applied to option pricing and VIX forecast directly. We also deduce the close-form expression for the VIX with the continuous SV model. Our empirical study suggest that the continuous SV model has the best performance among the three models although MCMC method has to be used in the estimation procedure. The DSARV model gives a rather unsatisfactory result, even compared to GARCH model. Although the implementation of the MLE method avoid subjectivity and may reduce computational time for the DSARV model, the computation of VIX is rather tedious.

This paper is organized as followed. Sect. 2 presents the the GARCH-type model, the continuous SV model, and the generalized DSARV model that we are to compare, under the physical probability measure. Then we give the risk-neutral equivalence of the DSARV model, in order to conduct volatility forecast under the risk-neutral measure. The GARCH and continuous SV model in the risk-neutral measure are also shown. In Sect. 3, the estimation methods of the models are discussed. The methods of VIX prediction with the three models are given. An empirical study based on the historical S&P 500 data is done in Sect. 4 to show the performances of the three models on predicting the VIX index. Finally concluding remarks are made in Sect. 5.

2 Models

We consider three volatility specifications in this paper: the DSARV models, the classical continuous SV model with leverage, and the non-affine GARCH(1, 1) models. The NGARCH-specifications proposed by Engle and Ng (1993) is adopted in our GARCH benchmarking model. In later part of this section, we deduce the DSARV model under the risk-neutral measure, which is new to the literature, so that the volatility forecast can be done straightforwardly.

2.1 NGARCH Model

The NGARCH model is specified to be:

$$R_{t+1} = \log\left(\frac{S_{t+1}}{S_t}\right) = r + \lambda\sigma_{t+1} - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}z_{t+1}, \tag{1}$$

$$\sigma_{t+1}^2 = \beta_0 + \sigma_t^2\left[\beta_1 + \beta_2\left(z_t - \beta_3\right)^2\right], \tag{2}$$

where $\beta_0 > 0$, $\beta_1, \beta_2 \geq 0$ since the conditional variance has to be positive. z_t is the *iid* standard normal random variable. λ stands for the risk-premium. According to Duan (1995), the dynamics of the return and volatility under the risk-neutral measure can be expressed as followed:

$$R_{t+1} = \log\left(\frac{S_{t+1}}{S_t}\right) = r - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\tilde{z}_{t+1}, \tag{3}$$

$$\sigma_{t+1}^2 = \beta_0 + \sigma_t^2\left[\beta_1 + \beta_2\left(\tilde{z}_t - \tilde{\beta}_3\right)^2\right], \tag{4}$$

where $\tilde{\beta}_3 \equiv \beta_3 + \lambda$. $\tilde{z}_t$ is the corresponding normal process under measure $\mathbb{Q}$. The shift cauzed by the change of measure is absorbed by the parameter β_3, which therefore becomes $\tilde{\beta}_3$.

2.2 Continuous SV Model

For the continuous SV model, the return and log-volatility take the following dynamics:

$$R_{t+1} = r + \lambda e^{h_{t+1}/2} - \frac{1}{2}e^{h_{t+1}} + e^{h_{t+1}/2}z_{t+1}, \tag{5}$$

$$h_{t+1} = \mu + \phi\left(h_t - \mu\right) + \tau\eta_t, \tag{6}$$

$$\eta_t = \rho\varepsilon_t + \sqrt{1 - \rho^2}z_t, \tag{7}$$

where h_t is the log-volatility at time t, τ is the standard variance of h_t. We assume the persistence in the volatility, i.e., $|\phi| < 1$, so that h_t is stationary. To reflect leverage

effect, volatility and return are correlated through (7). z_t and η_t follow standard normal distributions.

The risk-neutral dynamics become

$$R_{t+1} = r - \frac{1}{2}e^{h_{t+1}} + e^{h_{t+1}/2}\tilde{z}_{t+1}, \tag{8}$$

$$h_{t+1} = \tilde{\mu} + \phi\left(h_t - \mu\right) + \tau\tilde{\eta}_t, \tag{9}$$

where $\tilde{z}_{t+1}$ and $\tilde{\eta}_t$ are standard normal process under the $\mathbb{Q}$ measure. $\tilde{\mu}$ is the corresponding μ under $\mathbb{Q}$,

$$\tilde{\mu} = \mu - \rho\lambda\tau, \tag{10}$$

absorbing the shift from the physical measure by the error term of volatility through Eq. (7). Therefore, the volatility is a continuous random variable for any future time points.

2.3 Discrete Stochastic Autoregressive Volatility (DSARV) Model

Cordis and Kirby (2014) develop discrete stochastic autoregressive volatility (DSARV) models as a discrete version of SV models. They consider several variations of the original DSARV model in their paper. We take the generalized first-order DSARV model, denoted as DSARV(1, N), for the volatility process for analysis, where N stands for N volatility states in total, the length of the vector $\boldsymbol{\sigma}$. In our paper, the DSARV(1, N) model for the volatility takes the following form:

$$R_{t+1} = r + \lambda v_{t+1} - \frac{1}{2}v_{t+1}^2 + v_{t+1}z_{t+1}, \tag{11}$$

$$v_{t+1} = \boldsymbol{\sigma}'\mathbf{x}_{t+1}, \boldsymbol{\sigma} = (\sigma_1, \sigma_2, ... \sigma_N), \tag{12}$$

$$\mathbf{x}_{t+1} = \mathbf{P}_t'\mathbf{x}_t + \mathbf{e}_{t+1}, \tag{13}$$

$$\mathbf{P}_t = \phi\mathbf{I}_N + (1 - \phi)\mathbf{1}_N\boldsymbol{\pi}_t'. \tag{14}$$

In the model setting, volatility dynamics are described in Eq. (12) by a first-order Markov chain instead of a continuous diffusion process setting. $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)'$ is an $N \times 1$ vector that specifies the volatility mass points. $\mathbf{x}_t$ is an $N \times 1$ vector that represents the states of the N-state Markov chain at time t, whose jth element equals 1 if the process is in state $j \in 1, 2, \ldots, N$ at time t and 0 otherwise (Hamilton 1994). Equation (13) describes the state transitions specifically. $\mathbf{e}_{t+1}$ is a vector martingale difference sequence, where the expectation of $\mathbf{e}_{t+1}$ given state transitions up to time t equals 0. The $N \times N$ time-changing transition matrix $\mathbf{P}_t$ stands for the transition probabilities of the Markov chain $\{\mathbf{x}_t\}$, whose typical element is $p_{kjt} = Pr(v_{t+1} = \sigma_j|v_t = \sigma_k)$, and is modeled by Eq. (20). $\mathbf{I}_N$ denotes $N \times N$ identity matrix. $\mathbf{1}_N$ denotes an $N \times 1$ vector of all 1. z_{t+1} in the return process also follows a standard normal distribution.

To capture the feature of asymmetric volatility by making the volatility correlated with the returns and to reduce heavy parameterization, Cordis and Kirby (2014) allow

the transition matrix $\mathbf{P}_t$ to be changing with time and the past return series. Of the many ways of specifying the elements within $\mathbf{P}_t$ to make the volatility process correlated with the returns, we follow one possible specification proposed by Cordis and Kirby (2014) by modeling $\boldsymbol{\pi}_t$ in (20) as

$$\pi_{jt} = \frac{(N-1)!}{(j-1)!(N-j)!} w_t^{j-1}(1-w_t)^{N-j}, \quad j = 1, 2, \ldots, N, \qquad (15)$$

$$w_t = \frac{\exp(\eta + \psi R_t)}{1 + \exp(\eta + \psi R_t)}, \qquad (16)$$

where parameter w_t is time-varying, and a function of the past returns. $0 < w_t < 1$. The direction and strength of the volatility asymmetry effect is modeled by the sign and the magnitude of ψ.

Under the risk-neutral measure, the model becomes

$$\tilde{R}_{t+1} = r - \frac{1}{2}v_{t+1}^2 + v_{t+1}\tilde{z}_{t+1}, \qquad (17)$$

$$v_{t+1} = \boldsymbol{\sigma}'\mathbf{x}_{t+1}, \qquad (18)$$

where $\tilde{z}_{t+1}$ is the standard normal process under probability measure $\mathbb{Q}$. $\tilde{R}_{t+1}$ is the corresponding risk-neutral return process that to be generated. Therefore the transition matrix is also changing to $\tilde{\mathbf{P}}_t$ as in the following process

$$\mathbf{x}_{t+1} = \tilde{\mathbf{P}}_t'\mathbf{x}_t + \mathbf{e}_{t+1}. \qquad (19)$$

Further, $\tilde{\boldsymbol{\pi}}_t$ and $\tilde{w}_t$ are the risk-neutral equivalence of $\boldsymbol{\pi}_t$ and w_t in (15) and (16):

$$\tilde{\mathbf{P}}_t = \phi\mathbf{I}_N + (1-\phi)\mathbf{1}_N\tilde{\boldsymbol{\pi}}_t', \qquad (20)$$

$$\tilde{\pi}_{jt} = \frac{(N-1)!}{(j-1)!(N-j)!}\tilde{w}_t^{j-1}(1-\tilde{w}_t)^{N-j}, \quad j = 1, 2, \ldots, N, \qquad (21)$$

$$\tilde{w}_t = \frac{\exp(\eta + \psi\tilde{R}_t)}{1 + \exp(\eta + \psi\tilde{R}_t)}. \qquad (22)$$

Here for the DSARV model, the change in the volatility through the measure transformation is not linear as for the continuous SV model. The shift from the transformation procedure is absorbed by (22), and have effect on the risk-neutral volatility through a series of equations by (21), (20), and (19).

To further reduce the heavy parametrization of the DSARV(1, N) model, we adopt the extended possibility of 'log-linear' DSARV specification for $\boldsymbol{\sigma}$ by Cordis and Kirby (2014):

$$\log\sigma_j = \delta + \gamma j, \quad j = 1, 2, \ldots, N, \qquad (23)$$

where $\gamma > 0$ and no restriction is given to δ. In this specification, log-volatility are distributed evenly along a line.

3 Model Estimation and VIX Prediction

3.1 Parameter Estimation of NGARCH and Continuous SV Model

For the NGARCH models, parameters can be estimated through the maximum likelihood estimation.

One tricky point of the continuous SV models is the impossibility of writing the likelihood function of the model in close form. One has to use Bayesian approach via MCMC sampling method to do the estimation in this study. Under the Bayesian framework, the sampling from posterior distribution of parameters needs us to specify the prior distributions for the parameters concerned in this model.

We adopt the following prior distributions:

$$\mu \sim N(a_\mu, b_\mu)$$
$$\lambda \sim N(a_\lambda, b_\lambda)$$
$$\tau^2 \sim IG(a_\tau, b_\tau)$$
$$\phi^* \sim Be(a_\phi, b_\phi)$$
$$\rho \sim U(a_\rho, b_\rho)$$

where $\phi^* = \frac{\phi+1}{2}$, $Be(a, b)$ is the beta distribution with density

$$f(x|a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1},$$

$B(\cdot, \cdot)$ is the beta function, and $IG(a, b)$ is the inverse gamma distribution with density

$$f(x|a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-\frac{b}{x}}.$$

Specifically, a vague normal prior distribution is assigned to μ, a uniform prior is assigned to ρ, a non-informative inverse gamma prior distribution is assigned to τ^2, a beta prior distribution is assigned to ϕ^* and a normal prior distribution is assigned to λ.

3.2 Parameter Estimation of the DSARV(1, N) Models

Cordis and Kirby (2014) gives in their paper the log likelihood function of the DSARV(1, N) model as

$$L(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \mathbf{1}'_N (\mathbf{x}_{t|t-1} \odot \boldsymbol{\eta}_t), \qquad (24)$$

where $\odot$ denotes element-by-element multiplication, and $\boldsymbol{\eta}_t = (\eta_{1t}, ..., \eta_{Nt})'$ with $\eta_{jt} = f(R_t | v_t = \sigma_j, \mathcal{F}_t; \boldsymbol{\theta})$. Here the density is Gaussian as the error distribution

of return process is normal. $\boldsymbol{\theta}$ denotes all the unknown parameters within the model, including parameters in the transition matrix and volatility states. In this likelihood function, $\mathbf{x}_{t+1|t}$ denote the expectation of the $N \times N$ state vector $\mathbf{x}_{t+1}$ given information up to time t, i.e., $\mathbf{x}_{t+1|t} = \mathrm{E}(\mathbf{x}_{t+1}|\mathcal{F}_t)$. T denote the number of observations within a dataset. $\mathbf{x}_{t+1|t}$ is a recursive algorithm given by Hamilton (1989, 1990)

$$\mathbf{x}_{t+1|t} = \tilde{\mathbf{P}}'_t \left(\frac{\mathbf{x}_{t|t-1} \odot \boldsymbol{\eta}_t}{\mathbf{1}'_N (\mathbf{x}_{t|t-1} \odot \boldsymbol{\eta}_t)} \right). \tag{25}$$

Thus MLE can be applied to the log likelihood function in (24) to get the result of the estimation.

3.3 VIX Derivation

This paper mainly intends to compare the volatility forecast with the NGARCH, continuous SV model, DSARV$(1, N)$ model, by the help of the VIX index. Considering the fact that there is relatively not much literature on directly predicting the VIX, and that we need to figure out a way to use the three models to conveniently make VIX forecast, we take Kanniainen et al. (2014)'s expression for the VIX as the risk-neutral expectation of integrated variance within a month:

$$\frac{1}{\tau} \left(\frac{\mathrm{VIX}_t}{100} \right)^2 \cong \frac{1}{T} \tilde{\mathrm{E}}_t \sum_{j=1}^{T} \sigma^2_{t+j}, \tag{26}$$

in discrete time, where $\tilde{\mathrm{E}}(\cdot)$ is an expectation under the risk-neutral measure, and σ is the volatility. We follow Hao and Zhang (2013) and take the annualizing parameter τ as 252, and $T = 30$.

3.3.1 VIX for NGARCH

Kanniainen et al. (2014) shows for NGARCH models, the VIX$_t$ can be computed by iteratively computing the volatility as

$$\frac{1}{\tau} \left(\frac{\mathrm{VIX}^G_t}{100} \right)^2 = \frac{1}{T} \sum_{j=1}^{T} \left(h_{t+1} \tilde{\Psi}^{j-1} + \tilde{h} \left(1 - \tilde{\Psi}^{j-1} \right) \right) \tag{27}$$

$$= h_{t+1} \frac{1 - \tilde{\Psi}^T}{(1 - \tilde{\Psi})T} + \tilde{h} \left(1 - \frac{1 - \tilde{\Psi}^T}{(1 - \tilde{\Psi})T} \right), \tag{28}$$

with

$$\tilde{h} = \frac{\beta_0}{1 - \tilde{\Psi}},$$

where Ψ and $\tilde{\Psi}$ denote the volatility persistence of the corresponding model under the physical and risk-neutral measures, respectively. VIX_t^G denotes the predicted VIX series from the NGARCH models. Under this model, $\Psi = \beta_1 + \beta_2(1 + \beta_3^2)$, and $\tilde{\Psi} = \beta_1 + \beta_2(1 + \tilde{\beta}_3^2)$.

3.3.2 VIX for Continuous SV Models

We derive the close-form expression for the VIX with the continuous type of SV models by applying VIX basic formula (26) to the model in Eqs. (5), (6), and (7) directly. Then we get

$$\frac{1}{\tau}\left(\frac{\text{VIX}_t}{100}\right)^2 \cong \frac{1}{T}\tilde{E}_t\sum_{j=1}^{T}e^{h_{t+j}}. \tag{29}$$

From Eq. (10), if we denote the linear shift from measure $\mathbb{P}$ to measure $\mathbb{Q}$ as Δ, i.e., $\tilde{\mu} - \mu = \rho\lambda\tau\lambda\Delta$, the conditional n step prediction under the risk-neutral measure can be expressed as

$$\tilde{E}_t e^{h_{t+j}} = e^{\mu(1-\phi^j) + \phi^j h_t - \frac{\Delta(1-\phi^j)}{\phi(1-\phi)}} \cdot \tilde{E}_t e^{\tau(\phi^{j-1}\tilde{\eta}_{t+1} + \phi^{j-2}\tilde{\eta}_{t+2} + \ldots + \tilde{\eta}_{t+j})} \tag{30}$$

$$= e^{\mu(1-\phi^j) + \phi^j h_t - \frac{\phi\Delta(1-\phi^j)}{(1-\phi)}} \cdot e^{\frac{1}{2}\tau^2\frac{1-\phi^{2j}}{1-\phi^2}} \tag{31}$$

with $\tilde{\eta}_{t+j}$ being the random process driving the volatility under the risk-neutral measure at time $t + j$. Therefore taking (31) back into (29), the VIX_t for the continuous SV models can be written as

$$\text{VIX}_t^C = 100 \cdot \sqrt{\frac{\tau}{T}\sum_{j=1}^{T}e^{\mu(1-\phi^j) + \phi^j h_t - \frac{\phi\Delta(1-\phi^j)}{(1-\phi)}} \cdot e^{\frac{1}{2}\tau^2\frac{1-\phi^{2j}}{1-\phi^2}}}. \tag{32}$$

Here VIX_t^C denotes the VIX index predicted from the continuous SV model.

We can see from (32) that for continuous SV model, although estimation could be troublesome with MCMC sampling under the Baysian framework, the forecast on the VIX can be written fully in close form. Thus the prediction on the VIX is direct and convenient.

3.3.3 VIX for DSARV Models

For the DSARV$(1, N)$ models, unfortunately we do not have a close form of calculating the VIX index. So we simply follow the definition of the VIX in Eq. (26), simulate v_{t+j} according to Eq. (18) given information up to time t, and then compute $\tilde{E}_t v_{t+j}^2$ from

$$\text{VIX}_t^D = 100 \cdot \sqrt{\frac{\tau}{T} \sum_{j=1}^{T} \tilde{\mathbb{E}}_t v_{t+j}^2}, \tag{33}$$

where VIX_t^D denotes the predicted VIX index from the DSARV(1, N).

3.4 AR(1) Model for VIX

When we actually conduct the VIX forecast, we need to first estimate the parameters from each model under measure $\mathbb{P}$. After estimation, we transform each model is into the measure $\mathbb{Q}$ and get the corresponding VIX^G, VIX^C, and VIX^D from (28), (32), and (33). To compare the performances of different models, there are many statistical approaches. One way is to form a model with VIX from the model and market VIX.

Kanniainen et al. (2014) take the AR(1) specification for the VIX index to describe autoregressive disturbances to associate with the VIX market price. They denote the bias as

$$u_t = \text{VIX}_t^{Mkt} - \text{VIX}_t^{Mdl}, \tag{34}$$

where VIX^{Mkt} represents the market VIX index, VIX^{Mdl} denotes the VIX that we get from the above three models, respectively. Further the bias takes an AR(1) form:

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{35}$$

where $\varepsilon_t \sim NID(0, \sigma^2)$. We follow Beach and MacKinnon (1979) in applying MLE with autoregressive disturbances with the VIX, where u_t is granted a normal distribution with a zero mean and a contemporaneous variance Σ. Hence we have the log likelihood for the parameters ρ and Σ in the AR(1) model in Eqs. (34) and (35) as

$$\begin{aligned}
\log L^{VIX}(\text{VIX}; \boldsymbol{\theta}, \rho) = &-\frac{n}{2} \left((\log(2\pi) + \log(\Sigma(1 - \rho^2))) \right) \\
&+ \frac{1}{2} \left(\log(\Sigma(1 - \rho^2)) - \log(\Sigma) \right) \\
&- \frac{1}{2\Sigma} \left(u_1^2 + \sum_{t=2}^{n} \frac{(u_t - \rho u_{t-1})^2}{1 - \rho^2} \right).
\end{aligned}$$

where $\boldsymbol{\theta}$ is the parameter vector that contains all the parameters that have already been estimated in the estimation procedure.

Using this AR(1) model for the difference between the market observed VIX index and the model predicted VIX, we empirical compare the prediction errors for the three models. Our goal is to test if the DSARV model implemented with MLE really outperforms its continuous counterpart.

4 Empirical Studies

We collect historical data of S&P 500 on a daily basis, estimate the models with the stock data using methods mentioned in Sect. 3, and calculate the predicted VIX index using the corresponding representation of each model. The empirical performance is based on the comparison between the model-predicted VIX and the market-observed VIX, during which the testing sample is the VIX data from the market, out of our estimation sample.

4.1 The Empirical Framework

Specifically, for the discrete type of SV models, we examine the estimation behavior of DSARV models with different volatility state. Cordis and Kirby (2014) find the estimation for the first order DSARV models achieves the best result at $N = 8$, for there is not a decrease in the DIC as obvious as the DSARV models with smaller N. In our study, we focus on DSARV$(1, 4)$, DSARV$(1, 6)$, DSARV$(1, 8)$, and DSARV$(1, 10)$. Maximum likelihood estimation (MLE) is carried out directly for the aforementioned models with different numbers of volatility mass points.

The MLE of the parameters is also computed directly for NGARCH models. We implement the estimation of the continuous SV model using the WinBUGS package in R. The WinBUGS software mainly implement the Gibbs sampler. A single Markov chain is run for 15000 iterations. To ensure better convergence, the initial 5000 of all the iterations are discarded as burn-in period. Prior distributions are chosen according to Sect. 3.1. After estimation, we start simulating the VIX from the starting date where we take the S&P 500 data. We generate a series of VIX index for each model under the risk-neutral measure according to Sect. 3.2, and calculate predictive bias by dividing the market VIX by the model VIX. The VIX data simulated from the three models are further used to compute the autoregressive coefficient and variance of the VIX model itself, for the reference of comparison.

The data points we take are the S&P 500 dailies from January 2009 to December 2014. VIX index of the same period are collected for out-of-sample comparison. Table 1 and Fig. 1 present the summary statistics and the plot for the source data of market VIX, respectively. We use the Libor rates in US dollar as the constant interest rate for both the estimation, while a zero risk-free interest rate in the predicting. The choice of risk-free interest rate do not affect the result of comparison. During the estimation of DSARV model, we treat v_1, volatility at state 1, as a parameter to be estimated through the MLE. Table 2 gives the result of estimation through MLE for the DSARV models with different number of volatility mass points. BIC is calculated for each model for reference.
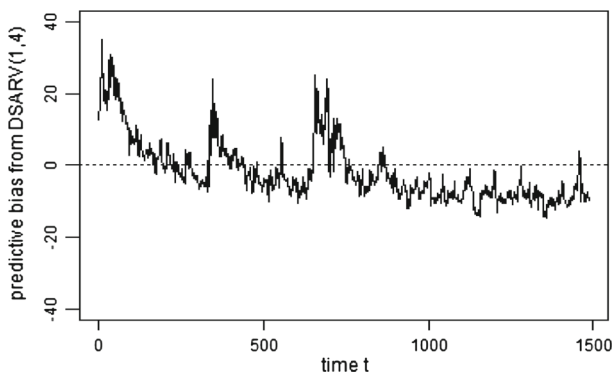
Table 1 Summary statistics of market VIX

| Min	25% quantile	Median	75% quantile	Max
10.32	14.74	18.02	24.32	56.65
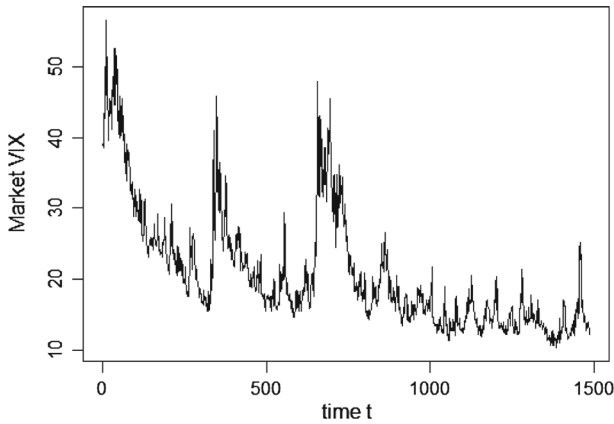
Fig. 1 Plot of market VIX

Table 2 Estimation result of DSARV(1, N) model for VIX

No. of volatility states	δ	γ	ψ	η	ϕ	β
$N = 4$	−0.6897	0.556	−0.334	−1.468	0.795	0.0514
$N = 6$	−0.3480	0.2787	−0.3314	−1.3843	0.67780	0.0662
$N = 8$	−0.4187	0.2138	−0.5105	−1.2201	0.8972	0.1616
$N = 10$	−0.2728	0.1214	−0.2868	−1.2052	0.9063	0.0389

Fig. 2 Plot of predictive bias from DSARV(1, 4) model

For the continuous SV model, the Bayesian approach treats all the volatilities as latent variables, making them all parameters to be estimated, altogether as augmented parameter space. Predictive bias for DSARV(1, 4), DSARV(1, 6), DSARV(1, 8), and DSARV(1, 10) are calculated and plotted in Figs. 2, 3, 4, and 5. Figures 6 and 7 are the predictive bias for the continuous SV model and NGARCH model. Summary statistics of the predictive bias for each model are also given in Tables 3 and 4. In order
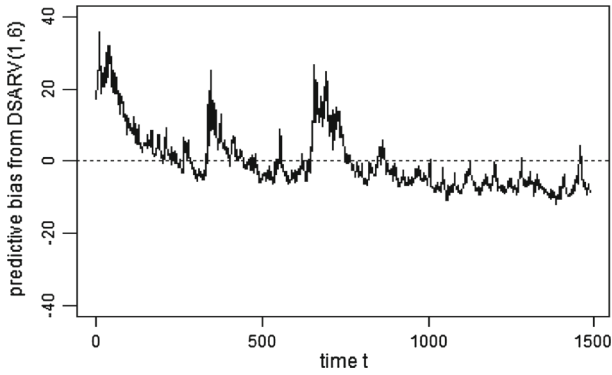
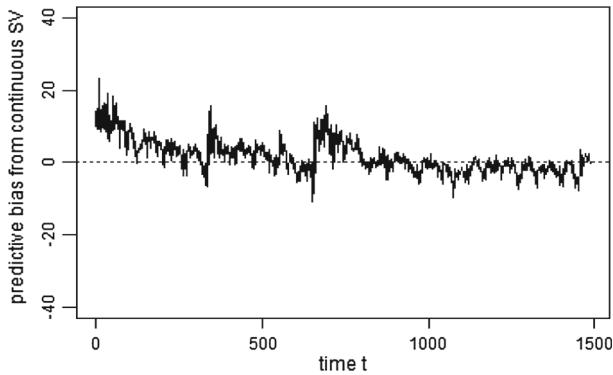**Fig. 3** Plot of predictive bias from DSARV(1, 6) model
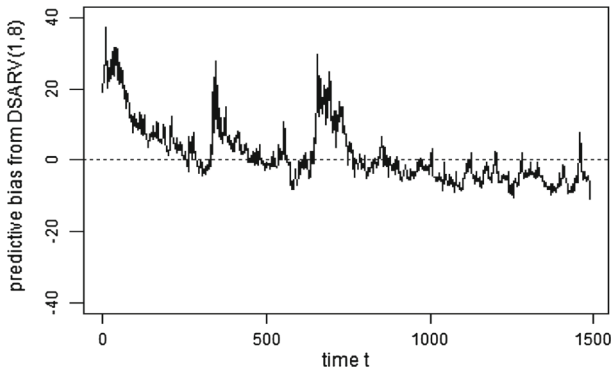
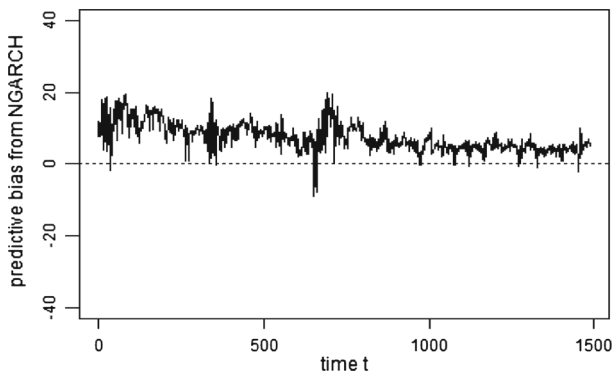**Fig. 4** Plot of predictive bias from DSARV(1, 8) model
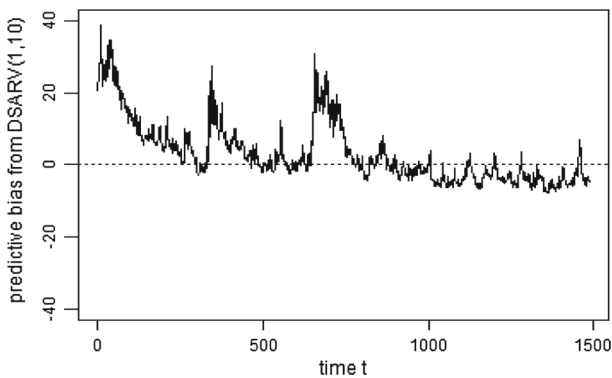
Fig. 5 Plot of predictive bias from DSARV(1, 10) model

to visualize the bias and better describe it, boxplots of the bias from DSARV(1, 4), DSARV(1, 6), DSARV(1, 8), DSARV(1, 10), continuous SV model, and NGARCH are presented in Figs. 8, 9, 10 and 11.

Fig. 6 Plot of predictive bias from continuous SV model

Fig. 7 Plot of predictive bias from NGARCH model

**Table 3** Summary statistics of predictive bias

| Model	Min	25% quantile	Median	75% quantile	Max
NGARCH	−8.95	4.79	6.67	9.84	19.76
Continuous SV	−10.80	−1.48	0.63	4.15	23.37
DSARV(1, 4)	−14.70	−7.85	−4.43	1.99	34.87
DSARV(1, 6)	−11.87	−6.12	−2.94	3.38	35.90
DSARV(1, 8)	−10.89	−4.46	−1.04	5.06	37.33
DSARV(1, 10)	−7.99	−3.16	0.34	6.56	38.81

4.2 Results

We can see that through the comparison of the plot of predictive bias, the bias of NGARCH and continuous SV models are obviously closer to the horizontal dotted line in the graph as a reference of ideal case of zero bias, of which the bias of continuous SV is relatively more around the zero line than the NGARCH model, which tends to

Table 4 Summary statistics of predictive bias

| Model	Mean	Absolute mean	Variance	Percentage mean
NGARCH	7.44	7.50	14.53	36.3
Continuous SV	1.63	3.52	20.41	2.91
DSARV(1, 4)	−1.92	7.08	71.38	−23.56
DSARV(1, 6)	−0.16	6.47	70.66	−14.16
DSARV(1, 8)	1.49	6.06	69.64	−4.97
DSARV(1, 10)	2.99	6.15	72.09	2.85
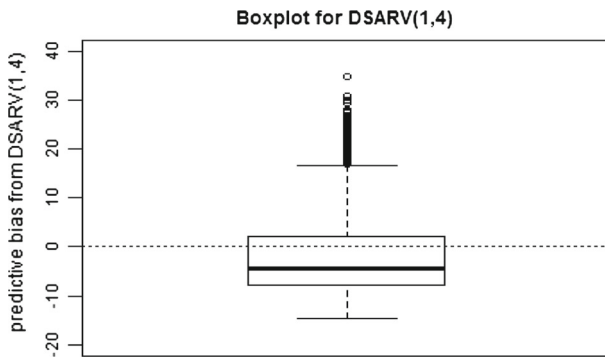
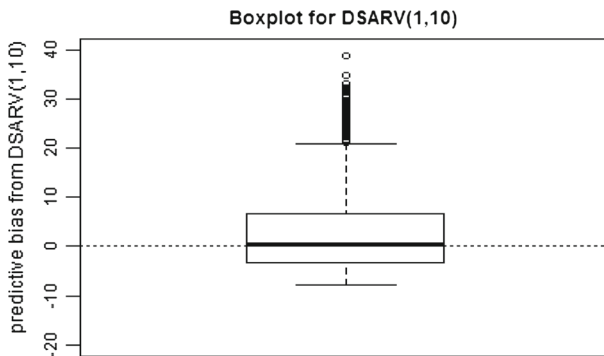**Fig. 8** Boxplot of predictive bias from DSARV(1, 4) model
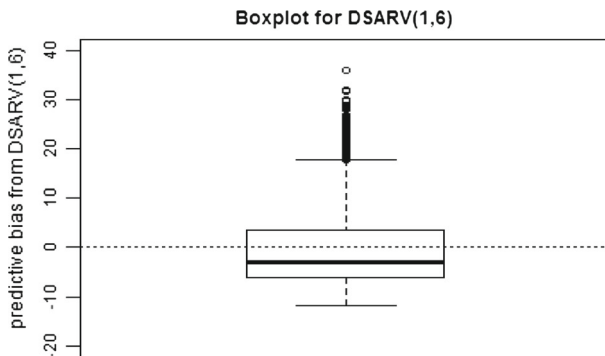
Fig. 9 Boxplot of predictive bias from DSARV(1, 6) model

slightly under-predict the VIX. While the plots of the bias from the DSARV models show a few big jumps from time to time, at which points the models do not make a fairly good forecast. In contrast, the forecast of continuous SV and NGARCH models are more stable.

Table 3 shows from a digital facet that the median of predictive bias from continuous SV and DSARV models with higher N are closest to zero while the NGARCH model suffer from a bigger bias median, which might come from the under-prediction. Similarly for the mean of the bias, NGARCH model has a larger bias than the other
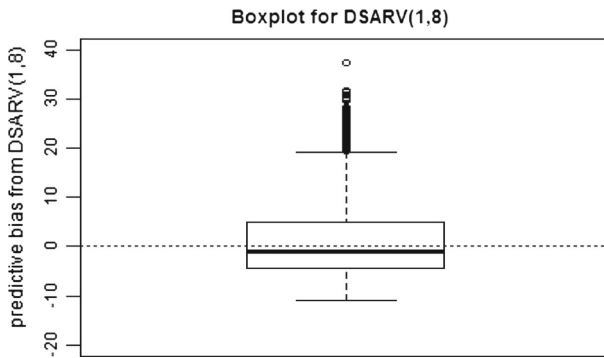
Fig. 10 Boxplot of predictive bias from DSARV(1, 8) model

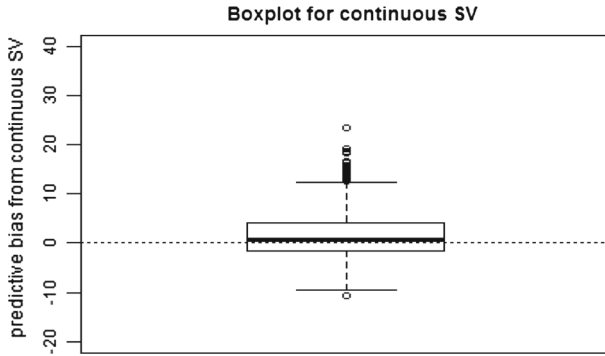**Fig. 11** Boxplot of predictive bias from DSARV(1, 10) model

models in the sense of magnitude. But for the absolute mean, which is calculated by taking the mean of the absolute value of the bias, the continuous type of SV model seems to the outperform the other models notably. The same case is for the mean of percentage bias, which is calculated by taking the mean of the quotient of the bias divided by the market VIX. From the perspective of variance, from Table 4, we see the variances are relatively smaller for NGARCH and continuous SV models, while the DSARV models produce rather large variance, which can be explained by the simulation procedure of VIX as there is no close-form formula for the VIX forecast for the DSARV models. The above results are in favor of the continuous SV model, in making the best forecast among the three models, though the variance of the bias may not the smallest.

We further take the VIX predicted from the three models into an AR(1) model in Sect. 3.4. The result is presented in Table 5, showing that the next forecast conditional on the last one from continuous SV is less volatile compared to the other two. Also from the boxplots, it is clearly seen that the VIX forecast from continuous SV model greatly outperforms the other two type of models in the sense of a closer to zero median for bias and a relatively smaller interquartile range. Although there is an outlier from the continuous SV model, the predictions from NGARCH and DSARV are either over-predict or under-predict the VIX quite much (Figs. 8, 9, 10, 11, 12, 13).
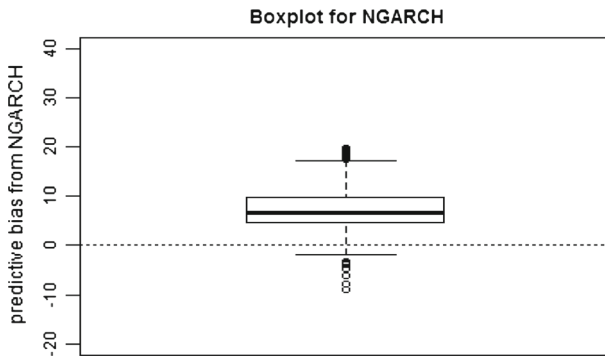
Table 5 Parameter estimation of AR(1) model for VIX

Model	NGARCH	CSV	DSARV
Conditional var	70.14	23.43	241.84

Fig. 12 Boxplot of predictive bias from continuous SV model

Fig. 13 Boxplot of predictive bias from NGARCH model

5 Conclusion

We generalize the DSARV$(1, N)$ model by proposing a risk-neutral transformation for VIX forecast. We further deduce the close-form formulas for the VIX under the continuous SV model with normal innovations. Using these analytical results, we empirically examine the VIX forecasting ability of NGARCH, the continuous SV model and the DSARV model with different number of volatility states. The empirical study based on the S&P 500 data shows that the continuous SV model performs the best among the three models and the little subjectivity in the prior for parameter estimation does not materially affects its performance. Discrete SV is attractive for its convenient estimation with MLE. However, the DSARV$(1, N)$ model gives a rather unsatisfactory prediction in the VIX.

References

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

Becker, R., Clements, A. E., & McClelland, A. (2009). The jump component of S&P 500 volatility and the VIX index. *Journal of Banking and Finance*, *33*, 1033–1038.

Bollerslev, T., Gibson, M., & Zhou, H. (2011). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, *160*, 102–118.

Beach, C. M., & MacKinnon, J. G. (1979). Maximum likelihood estimation of singular eqtation systems with autoregressive disturbances. *International Economic Review*, *20*, 459–464.

Calvet, L., & Fisher, A. (2004). How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, *2*, 49–83.

Cordis, A. S., & Kirby, C. (2014). Discrete stochastic autoregressive volatility. *Journal of Banking and Finance*, *43*, 160–178.

Carr, P., & Wu, L. (2006). A tale of two indices. *Journal of Derivatives*, *13*, 13–29.

Duan, J. C. (1995). The GARCH option pricing model. *Mathematical Finance*, *5*, 13–32.

Dupire, B. (2006). Model free results on volatility derivatives. Bloomberg.

Dotsis, G., Psychoyios, D., & Skiadopoulos, G. S. (2007). An empirical comparison of continuous-time models of implied volatility indices. *Journal of Banking and Finance*, *31*, 3584–3603.

Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, *68*, 1343–1376.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, *48*, 1749–1778.

Fleming, J., & Kirby, C. (2013). Component-driven regime-switching volatility. *Journal of Financial Econometrics*, *11*, 263–301.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*, 357–384.

Hamilton, J. (1990). Analysis of time series subject to change in regime. *Journal of Econometrics*, *45*, 39–70.

Hamilton, J. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hao, J., & Zhang, J. E. (2013). GARCH option pricing models, the CBOE VIX, and variance risk premium. *Journal of Financial Econometrics*, *11*, 556–580.

Heston, S. L. (1993). A closed-form solutions for options with stochastic volatility. *Review of Financial Studies*, *6*, 327–343.

Hull, J. C., & White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, *42*, 281–300.

Jacquier, E., Polson, N., & Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, *12*, 371–417. (with discussion).

Kanniainen, J., Lin, B., & Yang, H. (2014). Estimating and using GARCH models with VIX data for option valuation. *Journal of Banking and Finance*, *43*, 200–211.

Koopman, S. J., & Hol-Uspensky, E. (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of Applied Econometrics*, *17*, 667–689.

Lin, Y. N., & Chang, C. H. (2010). Consistent modeling of S&P500 and vix derivatives. *Journal of Economic Dynamics and Control*, *34*, 2302–2319.

Meyer, R., & Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal Royal Economic Society*, *3*, 198–215.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*, 347–370.

Psychoyios, D., & Skiadopoulos, G. (2006). Volatility options: hedging effectiveness, pricing, and model error. *Journal of Futures Markets*, *26*, 1–31.

Shephard, N., & Pitt, M. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika, 84*, 653–667.

Wang, Y., Choy, B., & Wong, H. Y. (2016). Bayesian option pricing framework with stochastic volatility. *Risks*, *4*(4), 51. doi:10.3390/risks4040051.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, *127*, 165–178.