

# Inferring Volatility Dynamics and Variance Risk Premia: An Efficient Bayesian Approach\*

Andras Fulop<sup>†</sup>

Junye Li<sup>‡</sup>

This Version: January 2015

## Abstract

We propose an efficient particle-based Bayesian method to investigate volatility dynamics and variance risk premia using the S&P 500 index returns and the term structure of variance swaps. This method can provide reliable and efficient inference over the model that takes into account co-jumps of prices and volatility, self-excitation, and non-affineness. We find that the non-affine models fit the data much better than affine ones, and self-excitation is an important feature in volatility dynamics. Our Bayesian method enables us to quantify uncertainty of variance risk premium estimates.

**Keywords:** Volatility Jumps, Non-affineness, Self-Exciting Jumps, Variance Risk Premia, Rao-Blackwellized Particle Filters, Density Tempering, Bayesian Methods

***JEL Classification:*** C11, C13, G12, G13

---

\*We would like thank Yacine Ait-Sahalia, Jun Yu, and seminar participants at Princeton-QUT-SMU Workshop in Financial Econometrics and European Seminar on Bayesian Econometrics (ESOB) for helpful comments.

<sup>†</sup>ESSEC Business School, Paris-Singapore. Email: fulop@essec.fr.

<sup>‡</sup>ESSEC Business School, Paris-Singapore. Email: li@essec.edu

# 1 Introduction

It is well known that volatility is time-varying, mean-reverting, and clustering. However, the 2008 global financial crisis and the recent European debt crisis in 2010-2012 reminds us of the importance to better understand the behavior of equity volatility during stressed times. Recent empirical studies find that a big jump in asset prices tends to be associated with an abrupt move in asset volatility, indicating co-jumps of prices and volatility (Eraker, Johannes, and Polson, 2003; Eraker, 2004; Jacod and Todorov, 2010; Todorov and Tauchen, 2011). It is also found that an extreme movement in markets tends to be followed by another extreme movement, resulting in self-exciting jump clustering (Carr and Wu, 2011; Ait-Sahalia, Cacho-Diaz, and Laeven, 2013; Fulop, Li, and Yu, 2014).<sup>1</sup> Furthermore, the literature documents that affine processes for volatility are misspecified (Jones, 2003; Ait-Sahalia and Kimmel, 2007; Christoffersen, Jacobs, and Mimouni, 2010).<sup>2</sup>

To accommodate these stylized facts, empirical researchers are resorting to ever richer frameworks for equity and volatility dynamics. At the same time, the development of derivatives markets provide us with rich information on volatility dynamics, which is very helpful to identify the parameters of such complex models. However, while rich model structure and introduction of derivatives data are useful, they also make model estimation a non-trivial undertaking, as one needs to simultaneously deal with multiple dynamic latent states following complex non-Gaussian processes and the fixed parameters

---

<sup>1</sup>Carr and Wu (2011) investigate a model in which diffusion volatility follows a square-root process and the jump intensity is self-exciting using both the underlying stock prices and options. They find strong evidence of the self-excitation. Ait-Sahalia, Cacho-Diaz, and Laeven (2013) propose a model that allows for the jump intensity following a Hawkes process while the diffusion volatility following a square-root process using the stock price data alone. They also find evidence of self-excitation. Fulop, Li, and Yu (2014) allow for jumps in diffusion volatility and self-excitation in the jump intensity. Using the stock price data alone, they find that the evidence of co-jumps of prices and volatility is quite robust. However, even though the data call for the self-exciting jump intensity, the self-exciting parameters are still hard to be pinned down.

<sup>2</sup>Jones (2003) investigates a constant elasticity of variance (CEV) model and find that the non-affine CEV model is favorable over the Heston stochastic volatility model. Ait-Sahalia and Kimmel (2007) find that the Heston model is misspecified, but the nature of the misspecification and the empirical findings are different from those found in Jones (2003). Christoffersen, Jacobs, and Mimouni (2010) provide a comprehensive investigation of alternatives to the Heston model, by comparing its empirical performance with that of five different but equally parsimonious stochastic volatility models. They find that the best volatility specification is one with linear rather than square root diffusion for variance.

driving the dynamics and embedded in the nonlinear pricing functions linking the latent states to the observed asset prices.

In this paper, we take a parametric Bayesian approach to the problem. Our purpose is two-fold. First, we propose a feasible way to estimate a rich dynamic asset pricing model, which allows for co-jumps of prices and volatility, self-exciting jump clustering, and non-affine volatility. Second, we investigate new features of volatility dynamics and variance risk premia. The specification is quite general and includes many models used in literature as special cases. Furthermore, it has closed-form conditional expectations of volatility components, making it convenient to use in volatility forecasting and pricing volatility derivatives. In a Bayesian context, Gibbs-type Markov chain Monte Carlo (MCMC) methods have been used to estimate jump-diffusion stochastic volatility models where the dynamic latent states and the fixed parameters are iteratively sampled assuming the other quantity is known (Eraker, Johannes, and Polson, 2003; Eraker, 2004; Johannes and Polson, 2009; Li, Wells, and Yu, 2008; Yu, Li, and Wells, 2011). However, such approaches can be hard to implement when the model structure becomes complex and derivative data are included in the dataset. First, the derivatives are very informative on the hidden states given the parameters and vice-versa, hence introducing strong correlation in the MCMC chain. This leads to a very slow convergence. Second, due to the complicated nonlinear derivative pricing functions, sufficient statistics for the conditional distributions of the parameters are hard to obtain and further the posteriors of the individual parameters can be highly correlated. Hence, it is hard to design efficient proposal distributions for the parameters. Particle Markov chain Monte Carlo (PMCMC) methods (Andrieu, Doucet, and Holenstein, 2010) could partially solve the problem of high correlation. However, they are hard to be parallelized and render very high computational cost, especially when derivatives data are introduced. Hence, these methods are not practically feasible for models in which we are interested in this paper.

In this paper, we propose a particle-based Bayesian method that aims to overcome the above-mentioned issues. First, we rely on pseudo-marginalization approach (Andrieu and Roberts, 2008; Andrieu, Doucet, and Holenstein, 2010) to break the correlation

between the the hidden states and the fixed parameters. We marginalize out the former using particle filters (PF) and then run a simulation routine targeting at the marginal distribution of the fixed parameters. In our case, the key to the successful application of such a method is the control of the estimation noise in the estimate of the likelihood of the data given the parameters. We find that with derivatives data a particle filter that uses the optimal proposal provides likelihood estimates that are too noisy for our purposes. Therefore, we instead propose the use of an approximate Rao-Blackwellized particle filter (RBPF) that aims to reduce the likelihood estimation noise. Second, to design efficient proposals targeting at the marginal distribution of the parameters, we employ the marginalized density-tempered sequential Monte Carlo sampler based on Del Moral, Doucet, and Jasra (2006) and Duan and Fulop (2014). This method represents the target with a simulated set of points and allows one to adapt the proposals to the simulation output in an iterative manner. More importantly, it allows a massively parallel implementation that helps to make this approach computationally feasible through the use of graphical processing units (GPUs).

We implement extensive Monte Carlo studies to check the accuracy, efficiency, and stability of the proposed approach. First, we show that our approximate RBPF performs quite similar to the exact particle filter in filtering the state variables. However, the noise of the likelihood estimate from the approximate RBPF is reduced by at least an order of magnitude with comparison to the exact particle filter. For example, with a small number of particles,  $M = 64$ , the standard deviation of the likelihood estimate in the approximate RBPF is only 0.29. However, even with a large number of particles,  $M = 2,048$ , the standard deviation of the likelihood estimate in the exact PF is as large as 2.9. Second, our particle-based Bayesian approach can accurately and efficiently identify most of the model parameters as their means are quite close to the true values and the root mean square errors (RMSEs) are small.

We estimate the model using the proposed particle-based Bayesian method on daily S&P 500 index returns and variance swap rates with fixed maturity 1-, 6-, and 12-month ranging from January 2, 2001 to July 15, 2013. First, an interesting question unex-

plored so far in the literature is whether non-affineness and self-excitation are substitutes to each other given that both are channels to allow for sudden bursts of volatility and extreme events. To investigate this issue, we compare the full model to nested specifications where we switch down non-affineness, self-excitation or both. Extending most of the extant literature that have used affine specifications (Carr and Wu, 2011; Fulop, Li, and Yu, 2014; Andersen, Fusari, and Todorov, 2014; Li and Zinna, 2014), we find support for self-excitation even when non-affineness is allowed for. Second, we find that non-affine specifications decisively beat affine ones. Furthermore, we find that the model with the CEV-type specification for diffusion volatility and jump intensity performs the best. These results resoundingly reinforce what have found by Jones (2003), Ait-Sahalia and Kimmel (2007), and Christoffersen, Jacobs, and Mimouni (2010) in our more flexible framework. Third, an important feature of the Bayesian approach is that we can easily quantify the statistical uncertainty about any quantity of interest, among else the dynamic variance risk premia (VRP). The variance risk premium provides an intuitive and straightforward measure of investors' risk aversion (Bekaert and Hoerova, 2014). Bollerslev, Tauchen, and Zhou (2008) show that the VRP is an important factor for short-term stock return predictability. Li and Zinna (2014) show that separate use of the VRP components further improves the short-term predictability. We find that investors' risk attitude is quite different between the diffusion variance risk and jump variance risk. The 90% credible interval is almost negative for the jump variance risk premia, no matter whether the short maturity (1-month) or the long maturity (1-year) is considered, whereas the upper 5% quantiles of the diffusion variance risk premia are always positive for all maturities considered. Furthermore, we find that despite our efficient likelihood-based inference approach, the variance swap data are not enough to reliably pin down these risk premia. Hence, either more informative priors or more derivatives data in the form of option panels are needed to precisely estimate the jump and diffusion risk premia.

The rest of the paper is organized as follows. Section 2 builds a parametric model that allows for co-jumps of prices and volatility, self-excitation, and non-affineness, and dis-

cusses pricing of variance swaps. Section 3 presents our particle-based Bayesian method and implement Monte Carlo studies. Section 4 implements model estimation using the real data on S&P 500 index returns and variance swap rates, and investigates the volatility dynamics and variance risk premia. Finally, Section 5 concludes the paper.

## 2 Model Setup

### 2.1 Stock Price and Volatility Dynamics

Under a given probability space  $(\Omega, \mathcal{F}, P)$  and the complete filtration  $\{\mathcal{F}_t\}_{t \geq 0}$ , the stock price,  $S_t$ , is modeled as follows

$$dS_t/S_t = \mu_t dt + \sqrt{V_t} dW_t + \int_R (e^x - 1) \tilde{\pi}(dx, dt), \quad (1)$$

where  $\mu_t$  is the instantaneous mean rate,  $W_t$  is a standard Brownian motion,  $V_t$  captures instantaneous diffusion variance, and the last term accounts for any price jumps with the return jump size,  $x$ , defined in the real line,  $R$ , through a compensated jump measure,  $\tilde{\pi}(dx, dt) = \pi(dx, dt) - \nu_t(dx)dt$ , in which  $\pi(dx, dt)$  is a random counting measure and  $\nu_t dx$  its compensator.

Equation (1) indicates that the stock price change consists of two orthogonal martingales: a purely continuous component and a purely discontinuous jump component. The jump component is important for generating the return non-normality and capturing short-maturity implied volatility smile/skew. We assume that the jump compensator takes the form of  $\nu_t(x) = \frac{\lambda_t}{\sqrt{2\pi}\sigma_J} \exp\{-\frac{(x-\mu_J)^2}{2\sigma_J^2}\}$ , indicating that the jump component follows a Compound Poisson process, where the number of jumps,  $N_t$ , arriving at any time interval,  $t$ , follows a Poisson process with the intensity  $\lambda_t$ , and the jump size,  $X_t$ , is identically and independently normally distributed with mean  $\mu_J$  and variance  $\sigma_J^2$ , *i.e.*,  $X_t \sim N(\mu_J, \sigma_J^2)$ .

Moreover, we propose to model the instantaneous diffusion variance,  $V_t$ , and the jump

intensity,  $\lambda_t$ , as follows,

$$dV_t = \kappa_v(\theta_v - V_t)dt + \sigma_v V_t^{\xi_1/2} dZ_{v,t} + dJ_{v,t}, \quad (2)$$

$$d\lambda_t = \kappa_\lambda(\theta_\lambda - \lambda_t)dt + \sigma_\lambda \lambda_t^{\xi_2/2} dZ_{\lambda,t} + \beta dN_t. \quad (3)$$

Equation (2) indicates that the diffusion variance follows a mean-reversion jump diffusion process, where  $Z_{v,t}$  is a standard Brownian motion and is allowed to be correlated with  $W_t$  in Equation (1),  $E[dW_t, dZ_t^v] = \rho dt$ , to accommodate the diffusion leverage effect, and  $J_{v,t}$  is a compound Poisson process whose Lévy density is given by  $\nu_{v,t}(x) = \lambda_t \frac{e^{-x/\mu_v}}{\mu_v}$ , indicating that the diffusion volatility jumps at the same time as stock returns with the same jump intensity,  $\lambda_t$ , and its jump size is independently exponentially distributed with mean  $\mu_v$ . The jump intensity,  $\lambda_t$ , follows a self-exciting process, as indicated in Equation (3), where  $Z_{\lambda,t}$  is an independent standard Brownian motion, and  $N_t$  is the same Poisson process as in the stock price jump and the diffusion volatility jump.

The above model (hereafter Model I) has some new features and can accommodate recently observed empirical facts of stock return volatility:

- *Co-jumps of stock prices and volatility:* Recent empirical studies find that a big jump, especially a big negative jump, in stock prices, tends to be associated with an abrupt move in variance (Eraker, Johanness, and Polson, 2003; Eraker, 2004; Jacod and Todorov, 2010; Todorov and Tauchen, 2011).
- *Self-exciting jump clustering:* A further intriguing empirical observation is that market turmoils seem to tell that an extreme movement in markets tends to be followed by another extreme movement (Carr and Wu, 2011; Ait-Sahalia, Cacho-Diaz, and Laeven, 2014; Fulop, Li, and Yu, 2014).
- *Non-affineness:* Jones (2003), Ait-Sahalia and Kimmel (2007), and Christoffersen, Jacobs, and Mimouni (2010) argue that the affine volatility models are misspecified.

The model has analytical conditional expectations of the variance components. The

conditional expectation of the jump intensity (3) can be found as follows

$$E[\lambda_t|\mathcal{F}_0] = e^{-(\kappa_\lambda - \beta)t}\lambda_0 + (1 - e^{-(\kappa_\lambda - \beta)t})\frac{\kappa_\lambda\theta_\lambda}{\kappa_\lambda - \beta}, \quad (4)$$

from which its long-run mean can be obtained by letting  $t \rightarrow +\infty$ ,

$$\bar{\lambda} = \frac{\kappa_\lambda\theta_\lambda}{\kappa_\lambda - \beta}. \quad (5)$$

Solutions (4) and (5) indicate that the conditional expectation of the jump intensity is a weighted average between the current intensity,  $\lambda_0$ , and its long-run mean,  $\bar{\lambda}$ , and the speed of mean-reversion is controlled by  $\kappa_\lambda - \beta$ . Using (4) and (5), the conditional expectation of diffusion variance (2) can also be found

$$E[V_t|\mathcal{F}_0] = e^{-\kappa_v t}V_0 + (1 - e^{-\kappa_v t})\theta_v + \mu_v \left[ \frac{1 - e^{-\kappa_v t}}{\kappa_v}\bar{\lambda} + \frac{e^{-(\kappa_\lambda - \beta)t} - e^{-\kappa_v t}}{\kappa_\lambda - \beta - \kappa_v}(\bar{\lambda} - \lambda_0) \right], \quad (6)$$

and its long-run mean is given by

$$\bar{V} = \theta_v + \frac{\mu_v}{\kappa_v}\bar{\lambda}. \quad (7)$$

The conditional expectation of the diffusion variance consists of two parts, one arising from the continuous part (the first two terms on the right-hand side in (6)) and the other from the jump component (the last term on the right-hand side in (6)).

The central questions we are concerned about in the present paper are the dynamic structures of jumps in stock returns and volatility. In order to explore these issues, we also investigate the following restricted models:

- Model II: both the diffusion variance and the jump intensity follow non-affine specifications, but there is no self-exciting effect, *i.e.*,  $\beta = 0$ .
- Model III: both the diffusion variance and the jump intensity follow affine specifications, and there is a self-exciting effect, *i.e.*,  $\xi_1 = \xi_2 = 1$ .
- Model IV: both the diffusion variance and the jump intensity follow affine speci-

cations, and there is no self-exciting effect, *i.e.*,  $\xi_1 = \xi_2 = 1$ , and  $\beta = 0$ .

## 2.2 Pricing Kernel and Variance Swap Valuation

A variance swap is a type of derivative that allows investors to trade variance as an asset class. At maturity, one leg of the swap pays an amount based upon realised variance (RV), and the other leg pays a fixed amount, which is called the variance swap rate, quoted at the inception. This contract has zero market value at entry and the payoff at maturity of the long position is equal to the difference between realized variance and the variance swap rate. Therefore, under the no-arbitrage condition, the variance swap rate should be the expected value of realized variance under the risk-neutral measure  $Q$ ,

$$VS_{t,T} = \frac{1}{T-t} E^Q[RV_{t,T} | \mathcal{F}_t]. \quad (8)$$

The no-arbitrage condition indicates that there exists at least one almost surely positive process,  $M_t$ , with  $M_0 = 1$ , such that the discounted gains process associated with any admissible trading strategy is a martingale (Harrison and Kreps, 1979).  $M_t$ , which is assumed to be a semimartingale, is the so-called stochastic discount factor, or the pricing kernel. If the market is complete,  $M_t$  is unique, otherwise there may exist many different pricing kernels. Following Pan (2002), Eraker (2004), and Broadie, Chernov, and Johannes (2007), we employ a class of models for the stochastic discount factor,  $M_t$ , such that the change-of-measure does not alter the model structure. Specifically,

$$M_t = \exp\left(-\int_0^t r_s ds\right) \mathcal{E}\left(-\int_0^t \gamma_W(s) dW_s\right) \mathcal{E}\left(-\int_0^t \gamma_v(s) dZ_{v,s}\right) \mathcal{E}\left(-\int_0^t \gamma_\lambda(s) dZ_{\lambda,s}\right) \prod_{j=1}^{N_t} \exp\left\{\frac{\mu_J^2 - (\mu_J^Q)^2}{2\sigma_J^2} + \frac{\mu_J^Q - \mu_J}{\sigma_J^2} J_{s,j} + \frac{\mu_v^Q - \mu_v}{\mu_v \mu_v^Q} J_{v,j}\right\}, \quad (9)$$

where  $r_t$  is the risk-free rate of interest, and  $\mathcal{E}(\cdot)$  denotes the stochastic (*Doleans-Dade*) exponential operator. The prices for diffusive risks,  $Z_{v,t}$  and  $Z_{\lambda,t}$ , are assumed to take the forms of  $\gamma_v(t) = \gamma_v V_t^{1-\xi_1/2}$  and  $\gamma_\lambda(t) = \gamma_\lambda \lambda_t^{1-\xi_2/2}$ , respectively, where  $\gamma_v$  and  $\gamma_\lambda$  are constants. For the jump components, the above change of measure indicates that

$\nu_t^Q(dx) = \frac{\lambda_t}{\sqrt{2\pi\sigma_J}} \exp\{-\frac{(x-\mu_J^Q)^2}{2\sigma_J^2}\}$  and  $\nu_{v,t}^Q(x) = \lambda_t \frac{e^{-x/\mu_v^Q}}{\mu_v^Q}$ . In contrast, we leave  $\gamma_W(t)$  unspecified as our main purpose in this paper is to investigate volatility dynamics and variance risk premium. We therefore have the following risk-neutral model,

$$dS_t/S_t = r_t dt + \sqrt{V_t} dW_t^Q + \int_R (e^x - 1) \tilde{\pi}^Q(dx, dt), \quad (10)$$

$$dV_t = \kappa_v^Q(\theta_v^Q - V_t) dt + \sigma_v V_t^{\xi_1/2} dZ_{v,t}^Q + dJ_{v,t}^Q, \quad (11)$$

$$d\lambda_t = \kappa_\lambda^Q(\theta_\lambda^Q - \lambda_t) dt + \sigma_\lambda V_t^{\xi_2/2} dZ_{\lambda,t}^Q + \beta dN_t, \quad (12)$$

where  $\kappa_v^Q = \kappa_v + \sigma_v \gamma_v$ ,  $\theta_v^Q = \kappa_v \theta_v / \kappa_v^Q$ ,  $\kappa_\lambda^Q = \kappa_\lambda + \sigma_\lambda \gamma_\lambda$ , and  $\theta_\lambda^Q = \kappa_\lambda \theta_\lambda / \kappa_\lambda^Q$ . The risk-neutral model has exactly the same structure as the objective one. We therefore can also obtain conditional expectations of  $V_t$  and  $\lambda_t$  in closed form similar to equations (6) and (4) under the risk-neutral measure. In particular, their risk-neutral long-run means become

$$\bar{V}^Q = \theta_v^Q + \frac{\mu_v^Q}{\kappa_v^Q} \bar{\lambda}^Q, \quad \bar{\lambda}^Q = \frac{\kappa_\lambda^Q \theta_\lambda^Q}{\kappa_\lambda^Q - \beta}. \quad (13)$$

The above risk-neutral model implies that realized variance, which can be approximated by quadratic variation, can be obtained as follows

$$\begin{aligned} RV_{t,T} &\doteq QV_{t,T} = \int_t^T V_s ds + \int_t^T \int_R x^2 \pi(ds, dx) \\ &= \int_t^T V_s ds + \int_t^T \int_R x^2 \nu^Q(dx) ds + \int_t^T \int_R x^2 \tilde{\pi}^Q(ds, dx). \end{aligned} \quad (14)$$

Taking expectation to equation (14) under the risk-neutral measure  $Q$  and plugging it into equation (8), we have the following variance swap pricing formula,

$$VS_{t,T} = \frac{1}{T-t} \int_t^T E^Q[V_s | \mathcal{F}_t] ds + \frac{1}{T-t} \left( (\mu_J^Q)^2 + \sigma_J^2 \right) \int_t^T E^Q[\lambda_s | \mathcal{F}_t] ds. \quad (15)$$

Similar to (6) and (4), the risk-neutral specifications of diffusion variance and the jump arrival rate result in the tractable conditional expectations of  $E^Q[V_s | \mathcal{F}_t]$  and  $E^Q[\lambda_s | \mathcal{F}_t]$ .

We therefore have the following variance swap pricing formula:<sup>3</sup>

$$VS_{t,T} = A(\tau) + B(\tau)V_t + C(\tau)\lambda_t, \quad (16)$$

where  $\tau = T - t$ ,  $A(\tau) = A_1(\tau) + A_2(\tau) + A_3(\tau)$ ,  $C(\tau) = C_1(\tau) + C_2(\tau)$ , and

$$\begin{aligned} A_1(\tau) &= (1 - B(\tau))\theta_v^Q, & A_2(\tau) &= ((\mu_J^Q)^2 + \sigma_J^2)(1 - H(\tau))\bar{\lambda}^Q, \\ A_3(\tau) &= \mu_v^Q \left[ \frac{1 - B(\tau)}{\kappa_v^Q} + \frac{H(\tau) - B(\tau)}{\kappa_\lambda^Q - \beta - \kappa_v^Q} \right] \bar{\lambda}^Q, \\ C_1(\tau) &= -\mu_v^Q \frac{H(\tau) - B(\tau)}{\kappa_\lambda^Q - \beta - \kappa_v^Q}, & C_2(\tau) &= ((\mu_J^Q)^2 + \sigma_J^2)H(\tau), \\ B(\tau) &= \frac{1 - e^{-\kappa_v^Q \tau}}{\kappa_v^Q \tau}, & H(\tau) &= \frac{1 - e^{-(\kappa_\lambda^Q - \beta)\tau}}{(\kappa_\lambda^Q - \beta)\tau}. \end{aligned}$$

### 3 A Particle-based Bayesian Method

Our model can be cast into a state-space model framework. After discretizing the return process for a small time interval  $\tau$  using the Euler method, we have the following observation equation for stock prices

$$\ln S_t = \ln S_{t-1} + \left( \mu - \frac{1}{2}V_{t-1} - k(1)\lambda_{t-1} \right) \tau + X_t \Delta N_t + \sqrt{\tau V_{t-1}} w_t \quad (17)$$

where  $k(1) = e^{\mu_J + \sigma_J^2/2} - 1$  is the convexity adjustment for the jump component,  $w_t$  is a standard normal noise, and  $X_t$  is the return jump size, which follows a normal distribution with mean  $\mu_J$  and variance  $\sigma_J^2$ . According to the existing literature, we approximate the increment in the jump counter by a Bernoulli variable, *i.e.*,  $\Delta N_t \equiv N_t - N_{t-1} \sim \text{Bernoulli}(\lambda_{t-1}\tau)$ .

The variance swap rates also enter into the observation equations. By assuming

---

<sup>3</sup>Following Britten-Jones and Neuberger (2000), Jiang and Tian (2005), and Carr and Wu (2009), the value of the variance swap can be synthesized from options as follows,  $\frac{2}{T-t} \sum_i e^{\tau(T-t)} \frac{\Delta K_i}{K_i^2} O(K_i, T-t)$ , where  $O(K_i, T-t)$  is an out-of-the-money put or call S&P 500 index option price with time to maturity  $T-t$  and strike  $K_i$ . As shown in Carr and Wu (2009), when the underlying price contains jumps, this replication does not hold exactly. However, based on their numerical experiments, they conclude that the error is negligible.

that the variance swap rates are collected with measurement errors, we have additional observation equations

$$\ln VS_{t,T}^O = \ln VS_{t,T}^M + \epsilon_t, \quad (18)$$

where  $VS_{t,T}^O$  is the market observed variance swap rate at time  $t$  with maturity  $T$ , and  $VS_{t,T}^M$  is the corresponding rate computed using the formula (16). The measurement errors,  $\epsilon_t$ , are assumed to be i.i.d normal with a mean vector of zero and a variance-covariance matrix of  $\sigma_\epsilon^2 I_n$ , where  $I_n$  is an  $n \times n$  identity matrix and  $n$  is the number of maturities.  $\sigma_\epsilon$  provides a measure of the degree of mispricing of variance swaps.

We have two state variables, which are the diffusion variance,  $V_t$ , and the jump intensity,  $\lambda_t$ . The discretized state equations are

$$V_t = V_{t-1} + \kappa_v(\theta_v - V_{t-1})\tau + \sigma_v V_{t-1}^{\xi_1/2} \sqrt{\tau} z_{v,t} + X_{v,t} \Delta N_t, \quad (19)$$

$$\lambda_t = \lambda_{t-1} + \kappa_\lambda(\theta_\lambda - \lambda_{t-1})\tau + \sigma_\lambda \lambda_{t-1}^{\xi_2/2} \sqrt{\tau} z_{\lambda,t} + \beta \Delta N_t, \quad (20)$$

where  $z_{v,t}$  is a standard normal noise, which is correlated to  $w_t$  in the return process (17) with a correlation parameter  $\rho$ ,  $X_{v,t}$  is the diffusion variance jump size, which is exponentially distributed with mean parameter  $\mu_v$ , and  $z_{\lambda,t}$  is an independent standard normal noise.

Denote the set of model parameters as  $\Theta$ , and all observations and the latent states up to time  $t$  as  $y_{1:t} = \{\ln S_s, \ln VS_s\}_{s=1}^t$ , and  $x_{1:t} = \{V_s, \lambda_s, \Delta N_t, J_{v,t}\}_{s=1}^t$ , respectively. Our aim is to find the joint posterior distribution of the parameters and the latent states,  $p(\Theta, x_{1:T} | y_{1:T})$ , which can be decomposed into

$$p(\Theta, x_{1:T} | y_{1:T}) = p(x_{1:T} | \Theta, y_{1:T}) p(\Theta | y_{1:T}), \quad (21)$$

where  $p(x_{1:T} | \Theta, y_{1:T})$  solves the state filtering issue, and  $p(\Theta | y_{1:T})$  addresses the parameter inference problem.

Bayesian Markov chain Monte Carlo (MCMC) methods could be used to estimate the above joint posterior distributions (Eraker, Johannes, and Polson, 2003; Eraker,

2004; Johannes and Polson, 2009; Li, Wells, and Yu, 2008; Yu, Li, and Wells, 2011). Typically, these algorithms use Gibbs sampling to move the latent states conditionally on the fixed parameters and vice-versa. While derivatives data are informative on the latent states, they pose problems to such MCMC methods. In particular, derivative observations introduce a tight link between the dynamic states and the fixed parameters, leading to high autocorrelation in the chain and very slow mixing. Furthermore, the variance swap rate pricing function is a complicated non-linear function of the parameters, leading to a loss of sufficient statistics and extra dependence between parameters. Hence, a generic good proposal over the parameters is hard to design.

In what follows, we come up with a particle-based Bayesian method, which aims to solve these issues. The key to successful application of this method is the control of the standard deviation of the likelihood estimate from a particle filter. We first propose an efficient Rao-Blackwellized particle filter for state filtering and likelihood estimation in Subsection 3.1. Then in Subsection 3.2, we tackle the issue of parameter estimation by relying on a sequential Monte Carlo sampler, which allows us to use the simulation output to adapt the proposal to our target in a principled way. Importantly, our method can be easily parallelized, making it computationally feasible and convenient to use in practice.

### 3.1 A Rao-Blackwellized Particle Filter

The above state-space model is clearly non-linear and non-Gaussian. State filtering can therefore be efficiently implemented using particle filters given the static parameters. For notational convenience, dependence on  $\Theta$  is suppressed in most of this subsection. Particle filters are a class of recursive algorithms that can be interpreted as simulation-based extensions of the Kalman Filter. The basic idea is to represent distributions of all random variables with a number of particles drawn directly from the state space and to approximate the posterior density  $p(x_t|y_{1:t})$  with the empirical point-mass estimate  $\hat{p}(x_t|y_{1:t})$

$$\hat{p}(x_t|y_{1:t}) = \sum_{i=1}^M \tilde{w}_t^{(i)} \delta(x_t - x_t^{(i)}) \quad (22)$$

where  $\tilde{w}_t^{(i)}$  is the normalized importance weight for each particle,  $x_t^{(i)}$  is the state particle, and  $\delta(\cdot)$  denotes the Dirac delta function.

Particle filters provide an estimate of the likelihood of the observations

$$\hat{p}(y_{1:t}|\Theta) = \prod_{l=2}^t \hat{p}(y_l|y_{1:l-1}, \Theta) \hat{p}(y_1|\Theta), \quad (23)$$

where

$$\hat{p}(y_l|y_{1:l-1}, \Theta) = \frac{1}{M} \sum_{i=1}^M w_l^{(i)}. \quad (24)$$

Importantly, the likelihood estimate (23) approximated by particle filters is unbiased,  $E[\hat{p}(y_{1:t}|\Theta)] = p(y_{1:t}|\Theta)$ , where the expectation is taken with respect to all the random quantities used in particle filters (Del Moral, 2004).

The most commonly used particle filter is the bootstrap filter of Gordon, Salmond, and Smith (1993), which simply takes the state transition law as the proposal density. However, the bootstrap filter is known to perform poorly when the observation is informative on the hidden states. Our model has this feature as derivatives contains rich information on volatility, and when we observe a large move in asset price, the jump can be largely pinned down by this observation. Alternatively, similar to Li (2011), we can design a more efficient filter that take the new observations into account in the proposal densities. However, in Monte Carlo studies below, we find that even such a particle filter leads to a large Monte Carlo noise in the likelihood estimates,  $\hat{p}(y_{1:T})$ , causing a breakdown in the parameter estimation algorithm proposed in Subsection 3.2. Hence, in order to decrease Monte Carlo noise in the likelihood estimates, we propose an approximate Rao-Blackwellized particle filter (RBPF) where the continuous states,  $z_t = \{V_t, \lambda_t\}$ , are dealt with analytically using the unscented Kalman filter (UKF) and the discrete particle filter is used over the jump variables,  $c_t = \{\Delta N_t, X_{v,t}\}$ .

The first fact to note is that conditional on the path of the jump variables,  $c_{1:t}$ , the states,  $z_t$ , follow multivariate constant elasticity of variance (CEV) processes. In such a system, the only feature that preempts the use of analytic filtering recursions is the nonlinearity in the transition and measurement equations. However, for similar settings,

it has been documented that in the presence of observations that are highly informative over the states, the unscented Kalman Filter provides a very good approximation to the true filtering densities (Li, 2013; Christoffersen et al., 2014). Given that derivatives are highly informative on the underlying hidden states, we are in exactly such a context and hence we use the UKF to provide us with an approximate analytic filter over  $z_t$ , conditionally on  $c_{1:t}$ . As a result, we can use particles over a lower dimensional system, solely tracking  $c_t$  while analytically marginalizing out  $z_t$ . Such a dimensional reduction approach can lead to a substantial decrease in the Monte Carlo noise (Chopin, 2004). Second, we approximate the volatility jumps,  $X_{v,t}$ , with a discrete random variable and apply the discrete particle filter (DPF) of Fearnhead (1998) to track the resulting discrete  $c_t$ . This approach avoids sampling from  $c_t$  by branching out all potential successor states and can outperform considerably alternative Rao-Blackwellized particle filters (Fearnhead and Clifford, 2003).<sup>4</sup>

The detailed algorithm of the proposed filter with  $M$  particles consists of the following steps. To simplify notation, when we write a superscript  $i$ , we always mean  $i = 1, 2, \dots, M$ .

**Initialization.** At time  $t = 0$ , set the initial discrete jump states to  $c_0^i$  and the mean and covariance matrix of  $(z_0 | c_0^i)$  to  $\mu_{z,0}^i$  and  $V_{z,0}^i$ , respectively, and give each set of particles a weight  $w_0^i = 1/M$ ;

**Time Recursion.** For  $t = 1, 2, \dots, T$ ,

- **Step 1:** we make the following two approximations to our system
  - **Approximation 1:** The jump counter depends on the filtered mean of the jump intensity,  $\hat{\lambda}_{t-1} = E(\lambda_{t-1} | c_{0:t-1}, y_{1:t-1})$ , *i.e.*, we have

$$\Delta N_t \sim \text{Bernoulli}(\hat{\lambda}_{t-1}\tau) \quad (25)$$

- **Approximation 2:** Approximate the continuous volatility jumps  $\Delta J_{v,t}$  with

---

<sup>4</sup>For previous use of the DPF in particle MCMC, see Whiteley et al. (2010, 2011)

a discrete random variable with equal weights on  $K$  support points,

$$X_{v,t} \in \left\{ \text{ExpCDF}_{\mu_v}^{-1} \left( \frac{j}{K+1} \right) \right\}, \quad j = 1, \dots, K, \quad (26)$$

where  $\text{ExpCDF}_{\mu_v}^{-1}$  denotes the inverse CDF of the exponential distribution with mean parameter  $\mu_v$ .

- **Step 2:** Now from each particle we branch out all  $K+1$  possible successor particles, leading to  $M \times (K+1)$  particles overall. First, all successor particle inherit the past discrete states  $c_{1:t-1}^{i,k} = c_{1:t-1}^i$  and hence the conditional moments of the continuous states:  $\mu_{z,t-1}^{i,k} = \mu_{z,t-1}^i$  and  $V_{z,t-1}^{i,k} = V_{z,t-1}^i$ . The weights and values of the new discrete states of the successor particles are

- For  $k = 1$ :  $\Delta N_t^{i,k} = 0$ ;  $X_{v,t}^k = 0$ ;  $w_{t-1}^{i,k} = \frac{1}{M}(1 - \hat{\lambda}_{t-1}^i \tau)$ ;
- For  $k = 2, \dots, K+1$ :  $\Delta N_t^{i,k} = 1$ ;  $X_{v,t}^k = \text{ExpCDF}_{\mu_v}^{-1} \left( \frac{k-1}{K+1} \right)$ ;  $w_{t-1}^{i,k} = \frac{1}{M \times K} \hat{\lambda}_{t-1}^i \tau$ .

- **Step 3:** Now we include the observations and update the continuous hidden states using the UKF:

- **Approximation 3:**  $(y_t, z_t \mid c_{1:t}, y_{1:t-1})$  is jointly normal and its moments are obtained by a UKF recursion starting from  $\mu_{z,t-1}$  and  $V_{z,t-1}$

This approximation allows us to update the weights to include the information in the new observation:

$$w_t^{i,k} = f(y_t \mid c_{1:t}^{i,k}, y_{1:t-1}) \times w_{t-1}^{i,k} \quad (27)$$

where  $f(y_t \mid c_{1:t}^{i,k}, y_{1:t-1})$  is the marginal normal density of the new observation implied by the UKF.

- **Step 4:** Approximate the likelihood of the new observation as follows

$$\hat{p}(y_t \mid y_{1:t-1}) = \sum_{i=1}^M \sum_{k=1}^{K+1} w_t^{i,k}. \quad (28)$$

Further, the UKF also provides the updated conditional moments of the continuous hidden states,  $\mu_{z,t}^{i,k}$  and  $V_{z,t}^{i,k}$ .

- **Step 5 (Stratified Resampling):** Resample  $M$  new particles out of the  $M \times (K + 1)$  proportional to normalized weights  $\pi^{i,k} \frac{w_t^{i,k}}{\hat{p}(y_t|y_{1:t-1})}$ . This produces  $M$  equally weighted particles,  $\{c_{1:t}^i, \mu_{z,t}^i, V_{z,t}^i\}$ . Notice that to iterate the algorithm forward, only  $\{\mu_{z,t}^i, V_{z,t}^i\}$  needs to be kept in the memory.

The importance of the first approximation is to make the jump probability measurable with respect to  $c_{0:t-1}$  such that  $z_t$  can be marginalized out. The second approximation allows us to work on the marginalized discrete state space over  $c_t$  and to use the discrete particle filter. Our justification for **Approximation 1** is that the observations are informative about  $\lambda_{t-1}$ , hence using the conditional expected value may not result in serious biases. In practice, the potential bias due to the discretization of the volatility jumps (**Approximation 2**) is easy to control as we simply increase the number of discretization points,  $K$ , in some preliminary runs until no significant change is observed in the results. In our applications, a small  $K$  suffices.<sup>5</sup> Last, the results in Li (2013) and Christoffersen et al. (2014) suggest that the UKF gives a very good analytic approximation to filtering CIR/square-root hidden states with informative observations, hence **Approximation 3** seems also quite reasonable.

## 3.2 Parameter Inference

We now move to the parameter inference issue. According to the Bayes rule, the posterior distribution of the model parameter is given by

$$p(\Theta|y_{1:T}) \propto p(y_{1:T}|\Theta)p(\Theta), \quad (29)$$

where the first term on the right-hand side is the likelihood and the second one is simply the prior. The decomposition suggests a hierarchical framework to the problem that

---

<sup>5</sup>A potential alternative to discretization would be to do stratified sampling over the continuous  $J_{v,t}$ . This would likely lead to similar results but would not introduce a bias to the algorithm.

targets at the posterior distribution of the fixed parameters—for a given set of model parameters proposed from some proposal, we can run a particle filter, which delivers the estimate of the likelihood,  $p(y_{1:T}|\Theta)$ , and empirical distribution of the hidden states,  $p(x_{1:T}|\Theta, y_{1:T})$ . This opens a way to use the pseudo-marginal approach of Andrieu and Roberts (2008) and Andrieu, Doucet, and Holenstein (2010).

The main idea of this approach is as follows. Define an auxiliary variable  $u_t$ , which include all the random variables produced by a particle filter at time  $t$  such as the state particles and resampling index, and denote the joint posterior distribution of auxiliary variables and the fixed parameters,  $\Theta$ , as  $\tilde{p}(\Theta, u_{1:T}|y_{1:T})$ . The unbiasedness property of the likelihood estimate from a particle filter indicates that this joint posterior distribution admits the target,  $p(\Theta|y_{1:T})$ , as a marginal

$$\int_{u_{1:T}} \tilde{p}(\Theta, u_{1:T}|y_{1:T}) du_{1:T} = p(\Theta|y_{1:T}). \quad (30)$$

Now instead of  $p(\Theta|y_{1:T})$ , we focus on the joint posterior distribution  $\tilde{p}(\Theta, u_{1:T}|y_{1:T})$ . In a recent paper, building on the tempered sequential Monte Carlo samplers of Del Moral, Doucet, and Jasra (2006), Duan and Fulop (2014) suggest a new sequential importance sampling algorithm targeting at  $\tilde{p}(\Theta, u_{1:T}|y_{1:T})$ . The key point is to begin with an easy-to-sample distribution and traverse through a sequence of densities to the ultimate target. For our case, we construct a sequence of  $I$  densities between the extended prior  $\pi_1(\Theta, u_{1:T})$  and the posterior  $\pi_I = \tilde{p}(\Theta, u_{1:T}|y_{1:T})$  using a tempering sequence  $\xi_i$ ,  $i = 1, 2, \dots, I$ , for  $\xi_1 = 0$ ,  $\xi_I = 1$ , and

$$\pi_i(\Theta, u_{1:T}) = \frac{\gamma_i(\Theta, u_{1:T})}{Z_i}, \quad (31)$$

$$\gamma_i(\Theta, u_{1:T}) = \hat{p}_i(y_{1:T}|\Theta)^{\xi_i} \tilde{p}_i(u_{1:T}|\Theta, y_{1:T}) p(\Theta), \quad (32)$$

where  $Z_i = \int \gamma_i(\Theta, u_{1:T}) d(\Theta, u_{1:T})$  is a normalized constant,  $\hat{p}_i(y_{1:T}|\Theta)$  is the estimated likelihood from the particle filter, and  $\tilde{p}_i(u_{1:T}|\Theta, y_{1:T})$  is the empirical distribution of the auxiliary variables.

Moving from  $\pi_i(\Theta, u_{1:T})$  to  $\pi_{i+1}(\Theta, u_{1:T})$  can be implemented by *reweighting* each

parameter particle by  $[\hat{p}_i(y_{1:T}|\Theta^{(n)})]^{\xi_{i+1}-\xi_i}$ , for  $n = 1, \dots, N$ . The tempering coefficients  $\xi_i$  can be chosen adaptively to ensure satisfactory particle diversity. Following Del Moral et al. (2012), we set the value of  $\xi_i$  to ensure that the effective sample size (ESS) stays close to some constant. This can be done by a simple grid search, where ESS is evaluated at the grid points of  $\xi_i$  and the one with the ESS closest to this constant is chosen.

With repeatedly reweighting and resampling, the support of the parameter particles would gradually deteriorate, leading to the well-known particle impoverishment problem. To solve this problem, periodically boosting the support becomes a must. Del Moral et al. (2006) suggest to first sample  $N$  points from  $\gamma_1(\Theta, u_{1:T})$ , which is the prior, and then recursively sample from  $\gamma_i(\Theta, u_{1:T})$  by *moving* these points using some MCMC kernels with the stationary distribution being  $\gamma_i(\Theta, u_{1:T})$ . The efficiency of the move step is measured by the acceptance rate, which captures how probable the proposed parameter set is accepted over the current set. Readers are referred to Del Moral et al. (2006) and Duan and Fulop (2014) for more detailed discussions.

Given that the likelihood function from the particle filter takes a complicated nonlinear function of the fixed parameters in this approach, conjugate priors are not available. Therefore, in this paper, we simply assume normal distributions for the priors, and simulation from the priors becomes straightforward. However, if a parameter under consideration has a finite support, we take a truncated normal as its prior. Furthermore, we take a normal mixture fitted on the particle population as the proposal and use the particle Marginal Metropolis-Hastings moves discussed in Andrieu, Doucet, and Holenstein (2010). Notice that fitting the proposal distribution using the information in the target is a key to the efficiency in our algorithm and is a key advantage of the SMC framework.

The marginal likelihood of the model can be computed as follows

$$\begin{aligned}
 p(y_{1:T}) &= \int p(y_{1:T}|\Theta)p(\Theta)d\Theta \\
 &= \int \hat{p}(y_{1:T})\tilde{p}(u_{1:T}|\Theta, y_{1:T})p(\Theta)d(u_{1:T}, \Theta) \\
 &= \prod_{i=2}^I [Z_i/Z_{i-1}], \tag{33}
 \end{aligned}$$

which can be used to construct the Bayes factor for model comparison. For any two models,  $\mathcal{M}_i$  and  $\mathcal{M}_j$ , the Bayes factor is given by the ratio of their marginal likelihoods, *i.e.*,

$$\mathcal{BF}_{i,j} = \frac{p(y_{1:T}|\mathcal{M}_i)}{p(y_{1:T}|\mathcal{M}_j)}, \quad (34)$$

which, different from the standard classical tests, does not rely on asymptotic distribution theory and provides an intuitive approach to evaluating the relative merits of competing models. It is important to note that the Bayes factor can be used to compare both nested and non-nested models, and furthermore it does not necessarily favor more complex model, as it contains a penalty for using more parameters due to its marginal nature.

This algorithm basically entails running  $N$  particle filters in parallel, each with  $M$  particles. It can be shown that the algorithm provides consistent inference to the extended target  $\tilde{p}(\Theta, u_{1:T}|y_{1:T})$  as  $N$  goes to infinity (Del Moral, Doucet, and Jasra, 2006). Given that the target is a marginal of the extended density, in the spirit of Andrieu and Roberts (2008), it also provides consistent inference to  $p(\Theta|y_{1:T})$  for any given number of state particles,  $M$ , as the number of parameter particles,  $N$ , goes to infinity. The algorithm can be easily parallelized in the parameter dimension. This is an important feature as it allows us to fully use the computational power of the modern graphical processing units (GPUs), resulting in a low computational cost.

### 3.3 Monte Carlo Studies

In this part, we implement three Monte Carlo studies. The first is to ascertain that the bias due to approximations in the proposed particle filter is indeed small for relevant parameter values. We compare the filtering performance of our approximate Rao-Blackwellized particle filter to an exact particle filter. This latter is similar in spirit to the approximate RBPF with two important differences: First, instead of using a discrete approximation to volatility jumps, it draws directly from the exponential density of the jumps,  $X_{v,t} \sim \text{Exp}(\mu_V)$ . Second, it keeps the state variables,  $\lambda_t$  and  $V_t$ , in the state space of the particle filter and uses the UKF to generate proposals over these variables. See Appendix for the detailed algorithm. The output of this PF converges to the optimal

filtering densities, and hence it can allow us to evaluate how far our approximate RBPF is from optimal filtering.

Table 1 reports results from a Monte Carlo exercise where we simulate 100 samples of length 4,000 from the model. We take the most general model, Model I, as our example. The true parameter values are chosen such that they are close to the empirical estimates in Section 4. As in the real data application, we use the underlying data and variance swaps with fixed maturities of 1-, 6-, and 12-month. For each simulated data sample, we run both the exact PF with the number of particles,  $M$ , equal to 512 and the approximate RBPF with  $M = 64$ . Then for each quantity of interest, we compute the root mean square errors (RMSEs) in the given sample between the true quantity and its filtered mean. The mean and interquartile range (IQR) of these RMSEs across the 100 samples are reported for both filters. The first two rows in the table show the results for the state variables,  $V_t$  and  $\lambda_t$ . One can see that while the performance of the two filters is virtually identical for filtering the jump intensity,  $\lambda_t$ , there seems to be a slight advantage from using the exact particle filter for filtering diffusion volatility,  $V_t$ . The main reason of this difference is that the diffusion volatility is much more variable and less persistent in the data generating process compared to the jump intensity. The larger local range of possible changes means that the UKF does a somewhat worse job in accounting for the underlying non-linearity for the diffusion compared to the jump intensity for the parameter configuration we use to generate the data. In addition to filtering the latent states, it is also of importance to see how well the different filters can identify variance swaps implied by these states. Hence row 3 to row 5 of Table 1 report the RMSE's over the three variance swap series. We can observe that the fitting performance of the two filters is almost identical across all maturities. Overall our results show that our approximate filter does a good job in approximating the optimal filtering recursion. This is in accord with the results in Christoffersen et al. (2014) who show that the UKF does a good job in approximating the full particle filter in affine term structure models at a much lower computational cost. Our approximate filtering procedure essentially extends those results to models also driven by discrete Poisson jumps.

— Table 1 around here —

As explained earlier, lots of recent estimation routines of the fixed parameters,  $\Theta$ , necessitate estimates of the likelihood of the data,  $p(y_{1:T}|\Theta)$ . Hence, in addition to filtering the dynamics states, it is also crucial to investigate the likelihood estimates,  $\hat{p}(y_{1:T}|\Theta)$ , implied by a given filtering routine. Table 2 sheds light on the performance of the full particle filter and the approximate Rao-Blackwellized particle filter from this angle. We simulate 20 data samples of length 4,000 from the model and run both filters on the each dataset at the fixed parameters we used to generate the data. The fixed parameters we use are identical to what we have used in the Monte Carlo exercise for filtering previously. For each data set, we run 256 independent filters at different number of particles. We use particle numbers  $M = 512, 1,024, 2,048$  in the full particle filter and  $M = 64, 128, 256$  particles in the approximate Rao-Blackwellized particle filter. These particle numbers are chosen to have roughly comparable computing times across the two filters. The first row reports the average estimate of the log likelihood of the data across the 20 data samples and 256 filtering runs per data samples. As expected, we can see that the exact particle filter using the true model delivers somewhat higher likelihood estimates compared to our approximate RBPF. While the difference in the loglikelihood values is visible, the almost identical filtering performance we saw in Table 1 gives rise to hope that it should not lead to large biases in the resulting parameter estimates.

— Table 2 around here —

The second row reports the averages across the 20 data samples of the standard deviations of the log likelihood estimates from both filters, where each time the standard deviation is computed as the sample standard deviation across the 256 filtering runs. One can directly see that the approximate filter has an order of magnitude smaller Monte Carlo noise at comparable computational cost. The recent literature on simulation-based inference routines that target at the fixed parameters,  $\Theta$ , and use an estimate of the likelihood,  $p(y_{1:T}|\Theta)$ , makes it clear that controlling the Monte Carlo noise of the likelihood estimates is key for the successful implementation of these routines. For example, for particle MCMC algorithms, Doucet et al. (2014) advise a standard deviation estimate of the

likelihood estimate of around 1.2-1.3 for general targets and proposal distributions in the MCMC. Looking at the second row of Table 2, we find that the approximate RBPF easily keeps the estimation noise below this value even for  $M = 64$ . In contrast, the standard deviation of the exact particle filter is well above this value, 2.9, even with  $M = 2,048$  particles. The usual asymptotic results posit that the variance of the loglikelihood estimate decreases linearly with the number of particles,  $M$  (Cerou et al., 2011). Using this asymptotic argument, the number of particles necessary to achieve a standard deviation of 1.3 in the exact particle filter would be approximately  $M = 2048 \times (\frac{2.9}{1.3})^2 \approx 10,000$ , an order of magnitude larger than the case for the approximate RBPF. Let us further note that as in this simulation we have run the particle filters at the parameters used to generate the data, it provides a best-case scenario for the filters. The estimation noise tends to increase both when the filters are run at parameter values not describing the data dynamics well, a must in the course of any parameter estimation routine, or when the data contains outliers which is a prevalent feature of real derivatives data. We have found that the approximate RBPF provides stable estimates of the loglikelihood, while the full particle filter can give rise to impractically large estimation noise. As a result, we can reliably use the former to do parameter estimation even with a small number of  $M$ , while with the latter the simulation routines over  $\Theta$  tend to get stuck even with thousands of particles.

We now move to the third Monte Carlo study, which aims to show that our proposed particle-based Bayesian method delivers reliable and efficient parameter estimates. This particle-based Bayesian method is initialized by the prior distributions or initial beliefs. As discussed above, in general, we assume (truncated) normal distributions for the priors. The choice of the hyper-parameters of the prior distributions is based on calibration using a training sample on S&P 500 index and VIX from January 1999 to December 2000. The training-sample approach is widely used to calibrate the objective prior distributions (O’Hagan, 1994). Notably, we find that most of parameters are not so sensitive to the selection of the priors. Therefore, we give quite flat priors to most of the model parameters. Table 3 presents the exact functional form, the support, and the hyper-

parameters for the prior distribution of each parameter. The same priors are used both in the simulation study below and in empirical investigations in the next section.

— Table 3 around here —

We implement a Monte Carlo simulation study to check accuracy, efficiency, and stability of the proposed econometric method. We take the most general model, Model I, as our example. The true values of model parameters are the same as those used in the above studies. We generate 30 sequences of daily observations on the underlying stock price and variance swap rates with maturity 1-, 6-, and 12-month, and then we run our particle-based Bayesian method for each simulated dataset. As before, the sample size is equal to 4,000.

In the implementation, we set the number of state particles,  $M$ , equal to 64 and the number of parameter particles,  $N$ , equal to 2,048. The number of discretization points for the variance jumps is chosen to be  $K = 5$ . Furthermore, we set the number of move step to 5. These tuning-parameters are chosen such that the acceptance rate is relatively high, the effective sample size fluctuates around  $N/2$ , and the computational cost is reasonable. To take advantage of the parallelisation property of our algorithm, we use graphical processor-based parallel architectures (GPUs) to speed up computations.<sup>6</sup> The priors for model parameters are the same as those in Table 3. Any values outside the support of a parameter are automatically discarded in simulations.

Table 4 presents the results of the simulation study. Means and RMSE's of the posterior means of the model parameters across 30 runs are reported. We find that the proposed particle-based Bayesian method can accurately and efficiently identify most of the model parameters, as for these parameters, their mean values are quite close to the true values, and their RMSE's are small. The only noticeable bias can be observed for the variance jump size parameters  $\mu_V$  and  $\mu_V^Q$ . One potential reason for this is that we have chosen  $K$  that is too small, hence leading to insufficient coverage of the right tail of the variance jump distribution.

---

<sup>6</sup>We program in MATLAB the main algorithm and offload the computational bottleneck of the algorithm, the particle filter, to the GPU, coded in CUDA. Relying on a Nvidia Titan Black GPU, our Bayesian algorithm is quite fast. Each run takes about 2/3 day.

## 4 Estimation and Empirical Results

In this section, we implement model estimation using the proposed particle-based Bayesian method on S&P 500 index returns and variance swap rates. The same  $M$ ,  $N$ , and the priors as in Subsection 3.3 are used. Furthermore, we set the number of move steps equal to 10 and  $K = 10$  in order to stabilize our estimates. Subsection 4.1 describes the data we use for model estimation; Subsection 4.2 implements statistical analysis and diagnosis; Subsection 4.3 discusses the volatility dynamics; and Subsection 4.4 investigates parameter uncertainty and variance risk premia.

### 4.1 The Data

The data used for estimation are S&P 500 index returns and variance swap rates with fixed maturity at 1-, 6-, and 12-month, ranging from January 2, 2001 to July 15, 2013 for a total of 3,148 business days. S&P 500 index values are obtained from *Datastream*, and the variance swap rates are provided by an investment bank. Notably, the data cover the market turmoils such as the 2002 dot-com bubble burst, the 2008 global financial crisis, and the recent European debt crisis of 2010-2012.

Table 5 reports the summary statistics of index returns and variance swap rates. We see that the annualized mean of index returns is about 6.9%, and the historical volatility is about 18.3%. The index returns are clearly left-skewed and leptokurtic as the skewness is negative (-0.24) and the kurtosis is by far larger than three (11.67). They are weakly autocorrelated with the first autocorrelation being about -0.09.

The mean values of variance swap rates increase with respect to maturity. In volatility measure, they are 21.2% at 1-month maturity, 22.5% at 6-month maturity, and 22.9% at 12-month maturity. In contrast, the standard deviations decrease with respect to maturity. Furthermore, the variance swap rates have positive skewness and excess kurtosis, both of which decrease with respect to maturity. The variance swap rates are highly

persistent. The first autocorrelation is 0.982 for 1-month maturity, 0.994 for 6-month maturity, and 0.996 for 12-month maturity.

— Table 5 around here —

Figure 1 presents the time series of index returns and the variance swap rates. The long-maturity variance swap rate is higher than the short-maturity rate when the market is tranquil. However, during the period of market turmoil, the term structure of variance swap rates reverts: the short-maturity variance swap rate move up quickly to even higher level than the long-maturity rate. This evidence is quite clear during the recent global financial crisis.

— Figure 1 around here —

## 4.2 Statistical Analysis and Diagnosis

Two main statistics in our particle-based Bayesian method are the effective sample size (ESS) and the acceptance rate, as discussed in Section 3. Figure 2 presents the acceptance rate (the left panel) and ESS (the right panel) for each model. There are a number of notable features.

First, for each of the four models, the acceptance rate remains high (above 30%) before the tempering coefficient,  $\xi_i$ , reaches the level of about 0.35. Then, it slowly goes down below 10%. This indicates that with respect to the tempering procedure, the information contained in the data is slowly absorbed. When the tempering goes to the end (*i.e.*,  $\xi_i$  is reaching the level of 1) and the full information in the dataset is included in the target density, the shape is becoming more complicated, making it harder to fit the proposal which in turn leads to lower acceptance probabilities.

— Figure 2 around here —

Second, when the tempering goes to the end, our estimates arrive at their posterior distributions. Figure 3 presents the tempering procedure for some selected parameters in Model I. The posterior means and (5, 95)% quantiles are plotted. We can see that

in the beginning when the tempering coefficient equal to 0, we only have the prior information, and the prior distributions have quite large dispersions. However, when the tempering procedure goes on, the information contained in the data is slowly reflected in the estimates. This can be see from the shrinkage of the (5, 95)% credible intervals. In the end when the tempering coefficient reaches one, the parameter estimates arrive at their posterior distributions. Basically, for all these selected parameters, they have quite narrow posterior (5, 95)% credible intervals.

— Figure 3 around here —

### 4.3 Volatility Dynamics

We now move to discuss volatility dynamics, which are quite important for risk management and derivatives pricing. Table 6 presents the log Bayes factors for model comparison, from which we have two important findings.

First, the non-affine models (Model I and Model II) perform much better than the affine models (Model III and Model IV). For example, the log Bayes factor between Model I and Model III (Model IV) is as large as 666.3 (759.1), and the log Bayes factor between Model II and Model III (Model IV) is about 593.0 (685.8). This result indicates that the usually used affine models are clearly misspecified, and non-affine specification needs to be used to capture volatility dynamics. With parsimoneous stochastic volatility specifications, Jones (2003), Ait-Sahalia and Kimmel (2007), and Christoffersen, Jacobs, and Mimouni (2010) also argue that affine models are misspecified.

— Table 6 around here —

Second, the jump intensity follows self-exciting dynamics. No matter whether non-affine specifications or affine specifications are concerned, the model with the self-exciting jump intensity always performs better than that excluding self-excitation. For example, for the non-affine models, the log Bayes factor between Model I and Model II is about 73.3, and for the affine models, the Bayes factor between Model III and Model IV is about 92.8, indicating that the self-excitation is an important feature in volatility dynamics. Under

affine specifications, Carr and Wu (2011), Ait-Sahalia, Cacho-Diaz, and Laeven (2014), and Fulop, Li, and Yu (2014) find evidence for the self-exciting jump intensity. Our investigation further indicates that even though both non-affineness and self-excitation are channels for sudden burst of volatility and extreme events, they are not substitutes to each other.

The above two findings can be also found from variance swap pricing errors, which are provided in the Internet Appendix. We find that in general the non-affine models result in smaller pricing errors than the affine models. However, the difference between Model I and Model II is not as big as that indicated by the Bayes factor.

The above model comparison indicates that we need to take into account both non-affineness and self-excitation in order to accurately capture volatility dynamics. Panel A of Table 7 presents the parameter estimates for Model I. There are several notable findings. First, the coefficients,  $\xi_1$  and  $\xi_2$ , which control the non-affineness in diffusion volatility and jump intensity, respectively, are well identified. Their posterior means are 1.97 and 2.19, respectively, and their posterior standard deviations are quite small, 0.03 and 0.10, respectively.

— Table 7 around here —

Second, the jump intensity is much more persistent than diffusion volatility because the estimated  $\kappa_\lambda$  is smaller than the estimated  $\kappa_v$  (1.16 and 4.76, respectively). The long-run mean of the jump intensity in Model I is given by formula (5), which is about 8.98. This value indicates that there are about 9 jumps, on average, in each year. The long-run mean of diffusion volatility is given by formula (7). According to our estimates, it is about 0.028.

Third, the parameter  $\beta$  controls the self-excitation of the jump intensity. Its posterior mean is about 1.09, and its posterior standard deviation is about 0.10. These values indicate that  $\beta$  is well identified and self-excitation is a key feature of the jump dynamics.

Fourth, the filtered diffusion volatility and jump intensity accounting for parameter uncertainty are plotted in Figure 4. The posterior means and (5, 95)% quantiles are presented. We can clearly see that diffusion volatility can be well identified, as the (5,

95)% credible intervals are quite narrow both in calm periods and in turmoil periods. However, the jump intensity has large credible intervals. This is particularly obvious during periods of market crash. The non-affineness and self-excitation mainly affect the estimates of the jump intensity as we see that all the four models result in very similar estimates of diffusion volatility, but the estimated jump intensities are different (see Internet Appendix).

— Figure 4 around here —

We have seen that the estimates of  $\xi_1$  and  $\xi_2$  are quite close to 2 in Model I. Hence, we further investigate a model with the restriction  $\xi_1 = \xi_2 = 2$  (hereafter Model CEV). Using the same estimation method, we find from Panel B of Table 7 that the parameter estimates are quite similar to those of Model I. Furthermore, from the last row of the table, we find that even though the pricing performance of Model I and Model CEV is quite similar, the log Bayes factor between Model I and Model CEV is about -3.5, indicating that the restricted model, Model CEV, is more capable of capturing the stock price and variance swap dynamics. Therefore, we conclude that the best model is the one with the self-exciting jumps and the CEV-type non-affine specification for diffusion volatility and jump intensity.

#### 4.4 Parameter Uncertainty and Variance Risk Premia

We implement the model estimation jointly using the underlying stock price data and the VIX data. Therefore, both the objective and the risk-neutral parameters can be obtained. The estimates of  $\gamma_v$ ,  $\gamma_\lambda$ ,  $\mu_J^Q$ , and  $\mu_v^Q$  in Table 7 indicate that there could be significant variance risk premia embedded in the data. The variance risk premium provides an intuitive and straightforward measure of investors' risk aversion (Bekaert and Hoerova, 2014). Recent empirical studies show that the variance risk premium is an important factor for short-term stock return predictability (Bollerslev, Tauchen, and Zhou, 2008; Drechsler and Yaron, 2011; Li and Zinna, 2014). Therefore, investigation of its properties and accurate measure are paramount for empirical study on return predictability.

Formally, the variance risk premium is defined as the difference between the expected objective and risk-neutral quadratic variations,

$$VRP(t, T) = \frac{1}{T-t} E_t [QV(t, T)] - \frac{1}{T-t} E_t^Q [QV(t, T)]. \quad (35)$$

The randomness of quadratic variation can either be induced by the randomness of the conditional future diffusion variance or by the jumps (with the random arrival rate) or by both.<sup>7</sup> We can thus decompose the variance risk premium into

$$VRP(t, T) = VRP_D(t, T) + VRP_J(t, T), \quad (36)$$

where the first term in equation (36) is due to variance risk from the diffusion,

$$VRP_D(t, T) = \frac{1}{T-t} E_t \left[ \int_t^T V_{1,s} ds \right] - \frac{1}{T-t} E_t^Q \left[ \int_t^T V_{1,s} ds \right], \quad (37)$$

and the second term in equation (36) reflects the compensation for variance risk due to the jump component,

$$VRP_J(t, T) = \frac{1}{T-t} E_t \left[ \int_t^T \int_{R^-} x^2 \pi(ds, dx) \right] - \frac{1}{T-t} E_t^Q \left[ \int_t^T \int_{R^-} x^2 \pi(ds, dx) \right]. \quad (38)$$

Both components in Equations (37) and (38) can be analytically computed under our model specifications in Section 2. By varying the maturity  $T$ , we can also obtain the term structures of variance risk premium and its components.

Here we focus on Model CEV. Table 8 presents summary statistics of the time series of the posterior means of variance risk premium and its components. The time-series average of the total variance risk premia is negative and it has a downward-sloping term structure. The negative sign of variance risk premia is in line with empirical findings in Bakshi and Kapadia (2003), Bollerslev, Gibson, and Zhou (2011), and Carr and Wu (2009). Ait-

---

<sup>7</sup>The jump component in our model is time-inhomogeneous, indicating that it contributes to the randomness of quadratic variance in two ways, one due to the jump size risk and the other due to risk related to the time-varying jump intensity. This is very different from the continuous diffusion, whose quadratic variation is random only because its variance is stochastic.

Sahalia, Karaman, and Mancini (2014) and Li and Zinna (2014) follow similar parametric approach and also find the downward-sloping term structure. The negative values of the total variance risk premia indicate that investors in the market dislike variance risk and they are willing to pay a premium to hedge against this risk. The longer the time horizon is, the larger such a premium is: investors' aggregate risk aversion is increasing with respect to the time horizon. The variance risk premium varies drastically. The variation of the total variance risk premium is increasing with respect to the maturity.

The total variance risk premium can be decomposed into diffusion variance risk premium and jump variance risk premium. Similar to the total variance risk premium, its components are also negative and has downward-sloping term structure. However, we can see that the jump variance risk premium has a much flatter term structure than the diffusion variance risk premium. We quantify the jump contribution to the total variance risk premium. It is found that for the one-month maturity, the jump contribution is as large as about 80%, and it is decreasing with respect to maturity, only about 34% at the one-year maturity. The jump contribution is quite stable, as its standard deviation is relatively small at each maturity.

— Table 8 around here —

Figure 5 plots the time series of the posterior means and (5, 95)% quantiles for the total variance risk premium and its components at maturity one-month and one-year accounting for parameter uncertainty. There are some interesting findings. First, no matter which maturity is concerned, the posterior means of the total variance risk premium and its components are always negative. This result is in contrast to the mode-free estimates, which could take positive values at times (Bekaert and Hoerova, 2014). Second, we find that during calm periods, the variance risk premium is quite small, whereas during the market crash, it drops to large values (in absolute). For example, after the Lehman Brothers' bankruptcy in September 2008, the variance risk premium (and its components) suddenly increases to very large values (in absolute).

Third, the 90% credible interval provides a measure of uncertainty of variance risk premium estimate. For the total variance risk premium, at one-month maturity, investors

seem to be quite certain on which sign it should take before the recent global financial crisis, as the 95% quantile is quite close to zeros and the 5% quantile is negative. However, after this crisis, even though investors' aversion becomes stronger, their uncertainty also becomes large, as now the 95% quantile can take positive values and the 5% quantile takes even small negative values. At one-year maturity, investors become even more uncertain, as the 90% credible interval become wider, and the 95% quantile always takes positive values. The longer the time horizon is, the more uncertain investors are.

— Figure 5 around here —

Fourth, when we compare the diffusion variance risk premium and the jump variance risk premium, we find that no matter which maturity is concerned, the 95% quantile of the diffusion variance risk premium always takes positive values, and this is particularly the case after the Lehman Brothers' bankruptcy. However, for the jump variance risk premium, its 95% quantile is quite close to zeros and its 5% quantile takes negative values for both one-month and one-year maturities. This indicates that investors are much more risk-averse of jump variance risk than diffusion variance risk. Furthermore, we find that the jump variance risk premium responds the market crash earlier, more quickly, and more dramatically than the diffusion variance risk premium. Putting all together, these results indicate that the jump variance risk premium measures investors fear of a market crash.

## 5 Concluding Remarks

In this paper, we propose a particle-based Bayesian method to estimate a rich asset pricing model that allows for co-jumps of prices and volatility, self-excitation, and non-affineness jointly using underlying and derivatives information. For efficient filtering, we use the UKF to analytically marginalize the Gaussian state variables, while using particle filtering to deal with the discrete jump variables. Crucially, the resulting approximate Rao-Blackwellized particle filter provides a low-variance estimate of the likelihood of the data given the fixed parameters. Then using this approximated likelihood, we run a

tempered sequential Monte Carlo routine targeting at posterior distribution of the fixed parameters. We provide extensive simulation evidence that our approach provides reliable and efficient inference over the model parameters.

We implement our methodology on daily S&P 500 return data and variance swap data at multiple maturities between 2001-2013. We have a few important substantive findings. First, we find there is evidence in the data for self-exciting jump behavior even when non-affineness is allowed for. Second, we reinforce the existing evidence that non-affine specifications dominate affine ones in a more flexible framework. Third, we find that despite using an efficient likelihood-based estimation method there is a large amount of uncertainty left about the variance risk premia.

There are interesting future research avenues that our results open up. First, we believe that the pseudo-marginalized Bayesian method based on the approximate Rao-Blackwellized particle filter is readily applicable to a wide range of dynamic asset pricing models with discrete jumps and informative derivatives data. Term structure models with jumps (Johannes, 2004; Piazzesi, 2005; Feldhutter et al, (2009); Jiang and Yan, 2009; Li and Song, 2013) or commodity and energy markets are interesting potential candidates. Second, an in-depth empirical study of diffusion and jump variance risk premia using option panels (Andersen, Fusari, and Todorov, 2014) and/or more informative economically motivated priors (Timmermann, Pettenuzzo, and Valkanov, 2014) would be insightful.

## REFERENCES

- Ait-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. (2014). “Modeling Financial Contagion Using Mutually Exciting Jump Processes.” *Journal of Financial Economics*, Forthcoming.
- Ait-Sahalia, Y., Karaman, M., and Mancini, L. (2014). “The Term Structure of Variance Swaps and Risk Premia.” Working Paper.
- Ait-Sahalia, Y. and Kimmel, R. (2007). “Maximum Likelihood Estimation of Stochastic Volatility Models.” *Journal of Financial Economics* 83, 413–452.
- Andersen, TG, Fusari, N., and Todorov, V. (2014). “Parametric Inference and Dynamic State Recovery from Option Panels.” *Econometrica*, forthcoming.

- Andrieu, C., Doucet, A., and Holenstein, R. (2010). “Particle Markov Chain Monte Carlo (with discussions).” *Journal of the Royal Statistical Society: Series B* 72, 269–342.
- Bakshi, G. and Kapadia, N. (2003). “Delt-Hedged Gains and the Negative Market Volatility Risk Premium.” *Review of Financial Studies* 16, 527–566.
- Bollerslev, T., Tauchen, G., and Zhou, H. (2009). “Expected Stock Returns and Variance Risk Premia.” *Review of Financial Studies* 22, 4463–4492.
- Bollerslev, T., Gibson, M., and 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.
- Bollerslev, T., and Todorov, V. (2011). “Tails, Fears and Risk Premia.” *Journal of Finance* 66, 2165–2211.
- Britten-Jones, M. and Neuberger, A. (2000). “Option Prices, Implied Price Processes, and Stochastic Volatility.” *Journal of Finance* 55, 839–866.
- Broadie, M., Chernov, M., and Johannes, M. (2007). “Model Specification and Risk Premia: Evidence from Futures Options.” *Journal of Finance* 62, 1453–1490.
- Carr, P., and Wu, L. (2009). “Variance Risk Premiums.” *Review of Financial Studies* 22, 1311–1341.
- Carr, P., and Wu, L. (2011). “Leverage effect, Volatility Feedback, and Self-Exciting Market Disruptions.” Working Paper.
- Cerou F., Del Moral, P., and Guyader, A. (2011). “A Nonasymptotic Theorem for Unnormalized Feynman-Kac Particle Models.” *Annales de l’Institut Henri Poincaré* 47, 629–649.
- Christoffersen, P., Dorion, C., Jacobs, K., and Karoui, L. (2014). “Nonlinear Kalman Filtering in Affine Term Structure Models.” *Management Science* 60, 2248–2268.
- Christoffersen, P., Jacobs, K., and Mimouni, K. (2010). “Volatility Dynamics for the S&P 500: Evidence from Realized Volatility, Daily Returns, and Option Prices.” *Review of Financial Studies* 23, 3141–3189.
- Chopin, N. (2004). “Central Limit Theorem for Sequential Monte Carlo Methods and its Applications to Bayesian Inference.” *Annals of Statistics* 32, 2385–2411.
- Del Moral, P., (2004). *Feynman-Kac Formulae Genealogical and Interacting Particle Systems with Applications*, Springer: New York.

- Del Moral, P., Doucet, A., and Jasra, A., (2006). “Sequential Monte Carlo Samplers.” *Journal of Royal Statistical Society B* 68, 411–436.
- Del Moral, P., Doucet, A., and Jasra, A., (2012). “An Adaptive Sequential Monte Carlo Method for Approximate Bayesian Computation.” *Statistics and Computing* 22, 1009–1020.
- Doucet, A., Pitt, M.K., Deligiannidis, G. and Kohn, R. (2014). “Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator.” *Biometrika*, forthcoming.
- Drechsler, I., and Yaron, A., (2011). “What’s Vol Got to Do with It.” *Review of Financial Studies* 24, 1–45.
- Duan, J., and Fulop, A. (2014). “Density-Tempered Marginalized Sequential Monte Carlo Samplers.” *Journal of Business and Economic Statistics*, Forthcoming.
- Eraker, B. (2004). “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices.” *Journal of Finance* 59, 1367–403.
- Eraker, B., Johannes, M., and Polson, N. (2003). “The Impact of Jumps in Equity Index Volatility and Returns.” *Journal of Finance* 58, 1269–1300.
- Fearnhead, P. and Clifford, P. (2003). “Online Inference for Well-Log Data.” *Journal of the Royal Statistical Society B* 65, 887–899.
- Feldhutter P, Schneider P and Trolle AB (2008) “Jumps in Interest Rates and pricing of Jump Risk: Evidence from teh Eurodollar Market” Working Paper
- Fulop, A., Li, J., and Yu, J. (2014). “Self-Exciting Jumps, Learning, and Asset Pricing Implications.” *Review of Financial Studies*, Forthcoming.
- Harrison, M. and Kreps, D., (1979). “Martingales and Arbitrage in Multiperiod Securities Markets.” *Journal of Economic Theory* 20, 381–408.
- Jiang, G., and Tian, Y. (2005). “The Model-Free Implied Volatility and Its Information Content.” *Review of Financial Studies* 18, 1305–1342.
- Jiang, G and Yan S (2009), “Linear-Quadratic Term Structure models: Toward the Understanding of Jumps in Interest Rates” *Journal of Banking and Finance*, 33, 473–485
- Johannes, M. (2004). “The Economic and Statistical Role of Jumps to Interest Rates.” *Journal of Finance* 59, 227–260.

- Johannes, M., and Polson, N. (2009). “MCMC Methods for Financial Econometrics.” In *Handbook of Financial Econometrics*, L. P. Hansen and Y. Ait-Sahalia, eds. Amsterdam: Elsevier.
- Jones, C. (2003). “The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets.” *Journal of Econometrics* 116, 181–224.
- Li, H. and Song Z. (2013). “Affine Jump Term Structure Models: Expectation Puzzles and Conditional Volatility.” Working Paper.
- Li, H., Wells, M.T., and Yu, C.L. (2008). “A Bayesian Analysis of Return Dynamics with Lévy Jumps.” *Review of Financial Studies* 21, 2345–2378.
- Li, J. (2011). “Sequential Bayesian Analysis of Time-Changed Infinite Activity Derivatives Pricing Models.” *Journal of Business & Economic Statistics* 29, 468–480.
- Li, J. (2013). “An Unscented Kalman Smoother for Volatility Extraction: Evidence from Stock Prices and Options.” *Computational Statistics & Data Analysis* 58, 15–26.
- Li, J. and Zinna, G. (2014). “The Variance Risk Premium: Components, Term Structures, and Stock Return Predictability.” Working Paper.
- Pan, J. (2002). “The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study.” *Journal of Financial Economics* 63, 3–50.
- Piazzesi, M. (2005). “Bond Yields and the Federal Reserve” *Journal of Political Economy* 113, 311–344
- Timmermann, A, Pettenuzzo, D., and Valkanov, R. (2014). “Forecasting Stock Returns under Economic Constraints.” *Journal of Financial Economics*, forthcoming.
- Todorov, V. (2010). “Variance Risk Premium Dynamics: The Role of Jumps.” *Review of Financial Studies* 23, 345–383.
- Whiteley, N., Doucet, A., and Andrieu, C. (2010). “Efficient Bayesian Inference for Switching State-Space Models Using Particle Markov Chain Monte Carlo Methods.” *Bristol Statistics Research Report*.
- Whiteley, N., Doucet, A., and Andrieu, C. (2011). “Bayesian Computational Methods for Inference in Multiple Change-points Models.” Working Paper.
- Yu, C.L., Li, H., and Wells, M.T. (2011) “MCMC Estimation of Levy Jumps Models Using Stock and Option Prices.” *Mathematical Finance* 21, 383–422.

Table 1: Monte Carlo on Filtering

	Full PF		Approximate RBPF	
	Mean of RMSEs	IQR of RMSEs	Mean of RMSEs	IQR of RMSEs
$V_t$	0.0008	0.0005	0.0010	0.0006
$\lambda_t$	0.1525	0.1027	0.1534	0.1037
$\ln VS1M$	0.0262	0.0009	0.0267	0.0012
$\ln VS6M$	0.0130	0.0012	0.0133	0.0012
$\ln VS1Y$	0.0121	0.0025	0.0122	0.0025

*Note:* This table compares the filtering performance of a full particle filter with the approximate Rao-Blackwellized particle filter proposed in this paper. We simulate 100 data samples of length 4,000 from the model and run both filters on the each dataset at the fixed parameters we used to generate the data. We use  $M = 512$  particles in the full particle filter and  $M = 64$  particles in the approximate Rao-Blackwellized particle filter. For each filter and each time point we compute the filtered mean of the state variables,  $V_t$  and  $\lambda_t$ , and also the filtered means of the log variance swap observations implied by the state variables. Then, for each dataset and filter, we compute the root mean square error (RMSE) of each filtered estimate versus the true counterpart. The first column report the mean RMSE for the full particle filter across the data sets and the second column the inter-quartile range (IQR) of the RMSE's across the 100 simulations. The third and fourth columns report the analogous quantities for the approximate Rao-Blackwellized particle filter.

Table 2: Monte Carlo on Likelihood Estimation

$M$	Full PF			Approximate RBPF		
	512	1,024	2,048	64	128	256
Mean of $\ln \hat{p}_i(y_{1:T} \Theta)$	27,339	27,343	27,345	27,218	27,218	27,218
Std of $\ln \hat{p}_i(y_{1:T} \Theta)$	5.11	3.67	2.90	0.29	0.20	0.16
Computational time	0.19	0.37	0.74	0.22	0.44	0.90

*Note:* This table compares the performance of a full particle filter with the approximate Rao-Blackwellized particle filter proposed in this paper to estimate the marginal log likelihood of the data. We simulate 20 data samples of length 4,000 from the model and run both filters on the each dataset at the fixed parameters we used to generate the data. The fixed parameters we use are identical we use for Table 1. For each data set we run 256 independent filters at different number of particles. We use particle numbers  $M = 512, 1,024, 2,048$  in the full particle filter and  $M = 64, 128, 256$  particles in the approximate Rao-Blackwellized particle filter. The first row reports the average estimate of the marginal log likelihood of the data across the 20 data samples and 256 filtering runs per data samples. The second row reports the average across the 20 data samples of the standard deviation estimate of the log likelihood estimate, where each time the standard deviation is computed as the sample standard deviation across the 256 filtering runs. The third row reports the computing time per filtering run. Here we run the 256 runs in parallel in one batch on an Nvidia Titan Black GPU, coded in CUDA. If the timing of such a run is  $T$ , we report the average across the 20 data samples of  $T/256$  to get a sense of the computational cost of evaluating the log likelihood for one parameter set and one run of the particle filter.

Table 3: **The Prior Distributions**

$\Theta$	F. Form	Support	$(\mu_0, \sigma_0)$	$\Theta$	F. Form	Support	$(\mu_0, \sigma_0)$
$\mu$	Normal	$(-\infty, \infty)$	(0.03, 0.10)	$\kappa_\lambda$	Tr. Normal	$(0, \infty)$	(1.50, 6.00)
$\mu_J$	Normal	$(-\infty, \infty)$	(-0.01, 0.05)	$\theta_\lambda$	Tr. Normal	$(0, \infty)$	(2.00, 6.00)
$\sigma_J$	Tr. Normal	$(0, \infty)$	(0.01, 0.05)	$\sigma_\lambda$	Tr. Normal	$(0, \infty)$	(0.50, 5.00)
$\kappa_v$	Tr. Normal	$(0, \infty)$	(8.00, 15.0)	$\xi_2$	Tr. Normal	$(0, \infty)$	(1.50, 3.00)
$\theta_v$	Tr. Normal	$(0, \infty)$	(0.01, 0.05)	$\beta$	Tr. Normal	$(0, \infty)$	(0.80, 5.00)
$\sigma_v$	Tr. Normal	$(0, \infty)$	(3.00, 6.00)	$\mu_J^Q$	Normal	$(-\infty, \infty)$	(-0.01, 0.05)
$\xi_1$	Tr. Normal	$(0, \infty)$	(1.50, 3.00)	$\mu_v^Q$	Tr. Normal	$(0, \infty)$	(0.02, 0.06)
$\rho$	Tr. Normal	$[-1, 1]$	(-0.50, 1.00)	$\gamma_v$	Normal	$(-\infty, \infty)$	(-8.00, 15.0)
$\mu_v$	Tr. Normal	$(0, \infty)$	(0.02, 0.05)	$\gamma_\lambda$	Normal	$(-\infty, \infty)$	(-0.50, 5.00)

*Note:* The table presents the functional form, its support, and the hyper-parameters of the prior distribution for each parameter. A normal distribution is usually assumed for the prior. However, if a parameter has a finite support, a truncated normal prior is attached to this parameter.

Table 4: Monte Carlo on Parameter Estimation

$\Theta$	True Value	Mean	RMSE	$\Theta$	True Value	Mean	RMSE
$\mu$	0.050	0.049	0.011	$\kappa_\lambda$	1.500	1.454	0.148
$\mu_J$	-0.015	-0.017	0.004	$\theta_\lambda$	0.500	0.491	0.102
$\sigma_J$	0.010	0.013	0.005	$\sigma_\lambda$	0.350	0.339	0.067
$\kappa_v$	5.000	4.825	0.691	$\xi_2$	2.000	2.152	0.421
$\theta_v$	0.015	0.017	0.003	$\beta$	1.000	0.823	0.195
$\sigma_v$	2.500	2.554	0.208	$\mu_J^Q$	-0.025	-0.024	0.004
$\xi_1$	2.000	2.004	0.038	$\mu_v^Q$	0.015	0.021	0.007
$\rho$	-0.700	-0.703	0.009	$\gamma_v$	-1.500	-1.190	0.390
$\mu_v$	0.015	0.020	0.017	$\gamma_\lambda$	-0.500	-0.605	0.218

*Note:* This table present the results of a small Monte Carlo study on parameter estimation using the approximate Rao-Blackwellized particle filter proposed in this paper. We simulate 30 dataset of length  $T = 4000$  using the parameters reported in the table. Then, for each dataset we run the proposed particle-based Bayesian method to estimate the fixed parameters. We use  $N = 2048$  fixed parameter particles,  $M = 64$  state particles, and after each resampling of the parameters, we use 5 move steps with an independent mixture of normal proposal. The parameter estimates are taken to be the posterior means. The table reports the mean parameter estimates across the simulation runs together with the root mean square errors of these estimates around the true values. A Nvidia Titan Black GPU is used.

Table 5: **Summary Statistics**

	Mean	Std.	Max.	Min.	Skew.	Kurt.	ACF
Returns	0.019	0.211	0.110	-0.095	-0.185	11.08	-0.092
VS1M	0.212	0.093	0.801	0.094	1.908	8.460	0.982
VS6M	0.225	0.073	0.614	0.123	1.345	5.735	0.994
VS1Y	0.229	0.067	0.536	0.130	1.013	4.453	0.996

*Note:* The table presents the summary statistics of the data used for model estimation. The data include S&P 500 index returns and variance swap rates with fixed maturity at 1-, 6-, and 12-month, ranging from January 2, 2001 to July 15, 2013, in total, 3,148 business days. Variance swap rates are reported in volatility measure.

Table 6: **The Log Bayes Factors**

	Model 1	Model 2	Model 3	Model 4
Model I	0.00	—	—	—
Model II	73.3	0.00	—	—
Model III	666.3	593.0	0.00	—
Model IV	759.1	685.8	92.8	0.00

*Note:* The table presents the log Bayes factor of the column model to the row model. For any two given models,  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , if the value of the log Bayes factor is between 0 and 1.1,  $\mathcal{M}_1$  is barely worth mentioning; if it is between 1.1 and 2.3,  $\mathcal{M}_1$  is substantially better than  $\mathcal{M}_2$ ; if it is between 2.3 and 3.4,  $\mathcal{M}_1$  is strongly better than  $\mathcal{M}_2$ ; if it is between 3.4 and 4.6,  $\mathcal{M}_1$  is very strongly better than  $\mathcal{M}_2$ ; and if it is larger than 4.6,  $\mathcal{M}_1$  is decisively better than  $\mathcal{M}_2$ .

Table 7: **Parameter Estimates**

$\Theta$	Mean	Std	90% C.I.	$\Theta$	Mean	Std	90% C.I.
<i>Panel A: Free <math>\xi_1</math> and <math>\xi_2</math></i>							
$\mu$	0.040	0.008	(0.026, 0.054)	$\kappa_\lambda$	1.156	0.107	(0.973, 1.336)
$\mu_J$	-0.015	0.001	(-0.017, -0.013)	$\theta_\lambda$	0.544	0.108	(0.358, 0.724)
$\sigma_J$	0.002	0.001	(0.000, 0.004)	$\sigma_\lambda$	0.351	0.065	(0.255, 0.470)
$\kappa_v$	4.761	0.146	(4.583, 5.041)	$\xi_2$	2.185	0.104	(2.008, 2.355)
$\theta_v$	0.009	0.001	(0.007, 0.011)	$\beta$	1.086	0.104	(0.912, 1.250)
$\sigma_v$	2.620	0.121	(2.433, 2.823)	$\mu_J^Q$	-0.016	0.001	(-0.018, -0.014)
$\xi_1$	1.965	0.027	(1.922, 2.011)	$\mu_v^Q$	0.010	0.001	(0.008, 0.012)
$\rho$	-0.827	0.007	(-0.838, -0.815)	$\gamma_v$	-0.069	0.053	(-0.171, -0.008)
$\mu_v$	0.010	0.001	(0.008, 0.013)	$\gamma_\lambda$	-0.156	0.116	(-0.395, -0.024)
<i>Panel B: <math>\xi_1 = \xi_2 = 2</math></i>							
$\mu$	0.039	0.008	(0.025, 0.052)	$\kappa_\lambda$	1.244	0.126	(1.037, 1.454)
$\mu_J$	-0.015	0.001	(-0.017, -0.013)	$\theta_\lambda$	0.522	0.103	(0.352, 0.693)
$\sigma_J$	0.002	0.001	(0.000, 0.003)	$\sigma_\lambda$	0.474	0.016	(0.447, 0.501)
$\kappa_v$	4.787	0.169	(4.585, 5.121)	$\xi_2$	2.000	—	—
$\theta_v$	0.011	0.001	(0.009, 0.012)	$\beta$	1.165	0.122	(0.973, 1.368)
$\sigma_v$	2.753	0.045	(2.680, 2.830)	$\mu_J^Q$	-0.016	0.001	(-0.018, -0.014)
$\xi_1$	2.000	—	—	$\mu_v^Q$	0.010	0.001	(0.008, 0.012)
$\rho$	-0.829	0.007	(-0.840, -0.818)	$\gamma_v$	-0.075	0.058	(-0.191, -0.009)
$\mu_v$	0.010	0.001	(0.008, 0.013)	$\gamma_\lambda$	-0.132	0.097	(-0.317, -0.021)
Model I vs. Model CEV						-3.53	

*Note:* The table presents the parameter estimates for Model I and Model CEV. For each parameter, the posterior mean, the posterior standard deviation, and the 90% credible interval (in brackets) are reported. The last row presents the log Bayes factor between Model I and Model CEV. The models are estimated using the particle-based Bayesian method discussed in Section 3. The data used for estimation are S&P 500 index returns and variance swap rates with fixed maturity 1-, 6-, and 12-month, ranging from January 2, 2001 to July 15, 2013, in total, 3,148 business days.

Table 8: **Variance Risk Premia**

	VRP	Mean	St. Dev.	Max	Min
Total VRP	VRPT1M	-0.139	0.102	-0.027	-0.803
	VRPT6M	-0.253	0.179	-0.060	-1.466
	VRPT1Y	-0.350	0.227	-0.088	-1.520
Diff. VRP	VRPD1M	-0.030	0.041	0.000	-0.468
	VRPD6M	-0.139	0.111	-0.035	-1.115
	VRPD1Y	-0.229	0.149	-0.064	-1.149
Jump VRP	VRPJ1M	-0.109	0.074	-0.021	-0.431
	VRPJ6M	-0.115	0.077	-0.022	-0.452
	VRPJ1Y	-0.122	0.082	-0.024	-0.477
Jump Ctr.	JCtr1M	0.800	0.114	1.000	0.220
	JCtr6M	0.454	0.072	0.591	0.124
	JCtr1Y	0.340	0.035	0.399	0.148

*Note:* The table reports the time-series properties of the total variance risk premia (Total VRP), the diffusion variance risk premia (Diff. VRP), and the jump variance risk premia (Jump VRP), as well as the jump contribution to the total variance risk premia (Jump Ctr.), which is computed as Jump VRP/Total VRP.

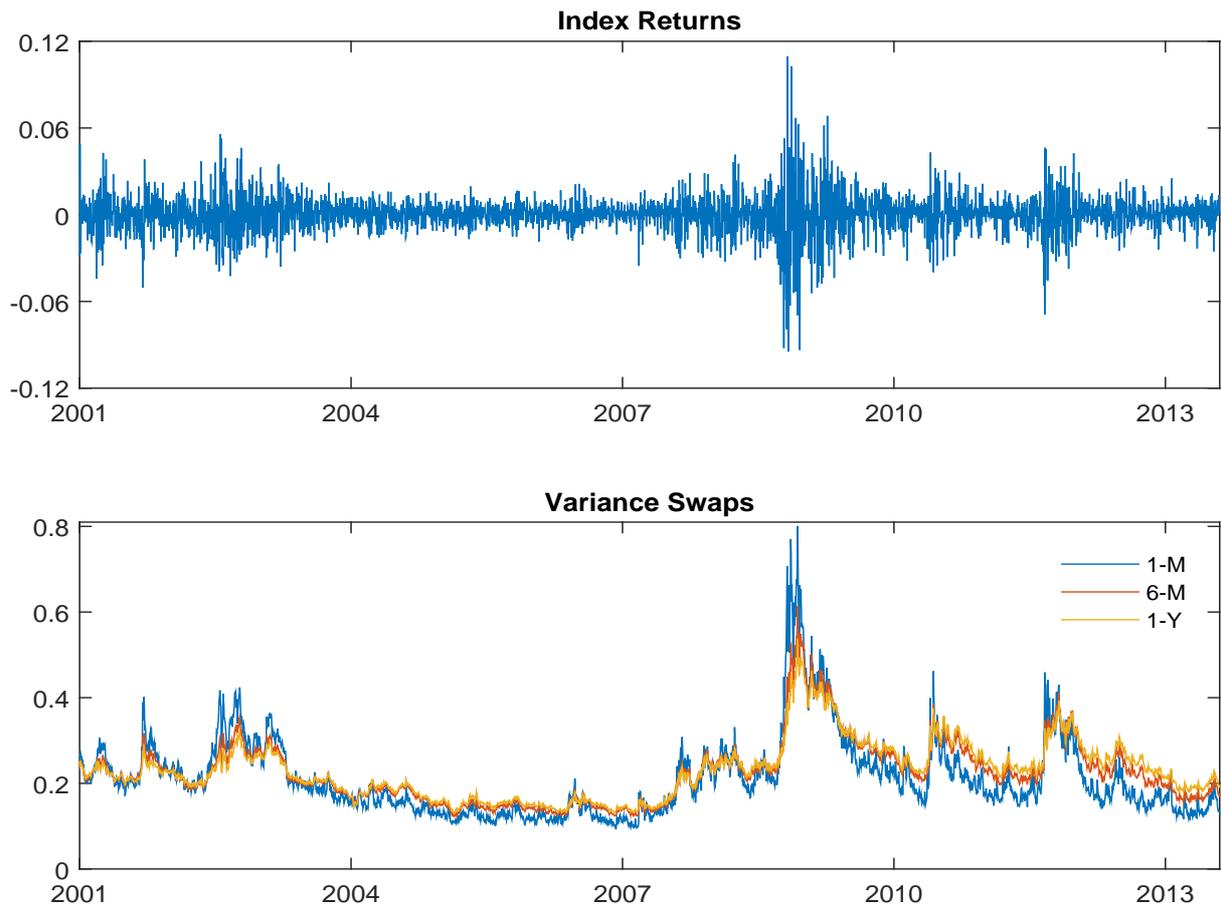


Figure 1: S&P500 Index Returns and Variance Swap Rates

*Note:* The figure plots the S&P500 index returns and variance swap rates with the fixed maturity 1-, 6-, and 12-month. The data range from January 2, 2001 to July 15, 2013, in total, 3,148 business days. Variance swap rates are presented in volatility measure.

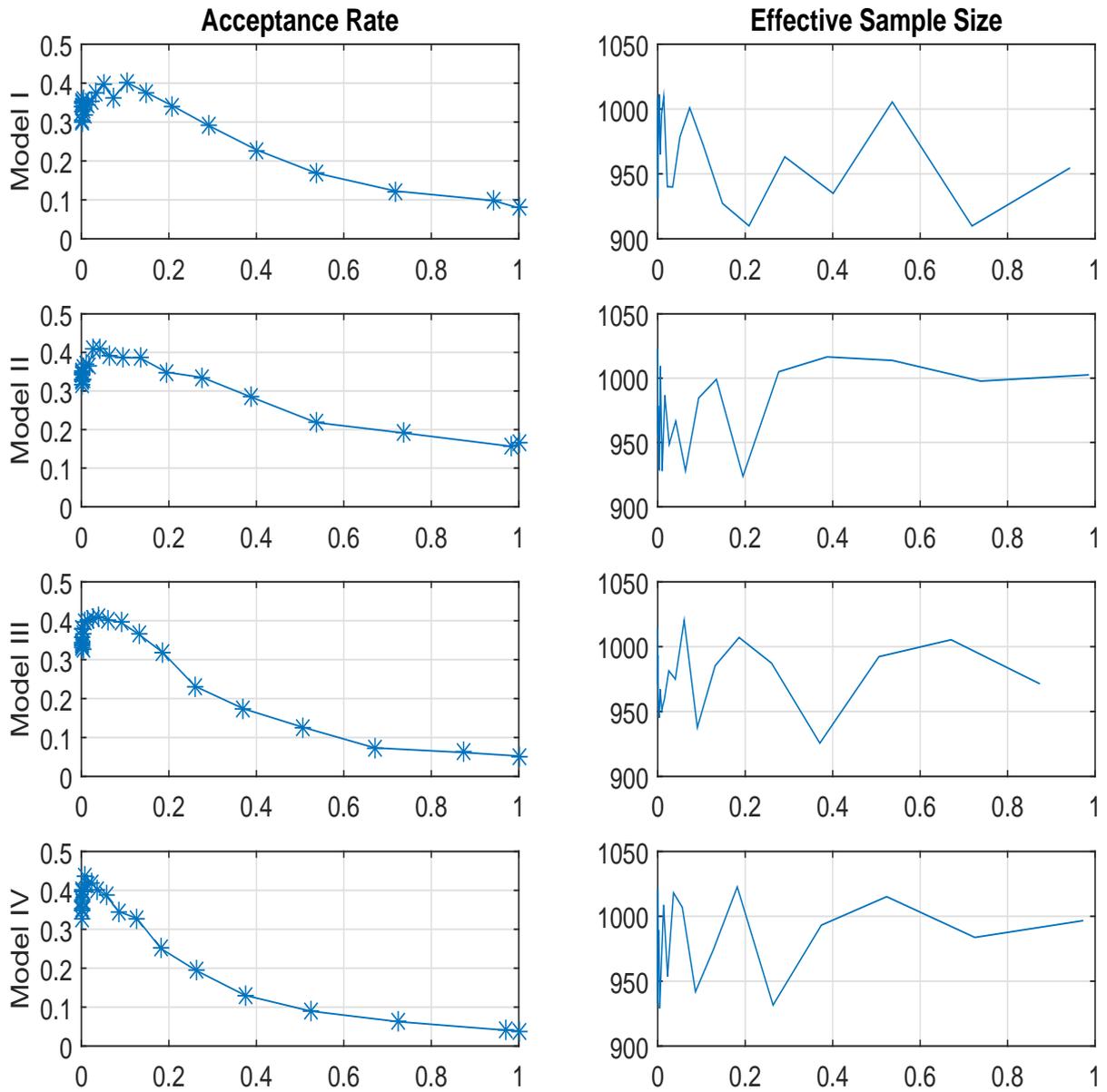


Figure 2: **Acceptance Rate and Effective Sample Size**

*Note:* The figure plots the acceptance rates and effective sample sizes (ESS) with respect to  $\xi_i$  for the four models considered. In our algorithm,  $\xi_i$  is automatically select. At each  $\xi_i$ , the algorithm computes the acceptance rate and effective sample size.

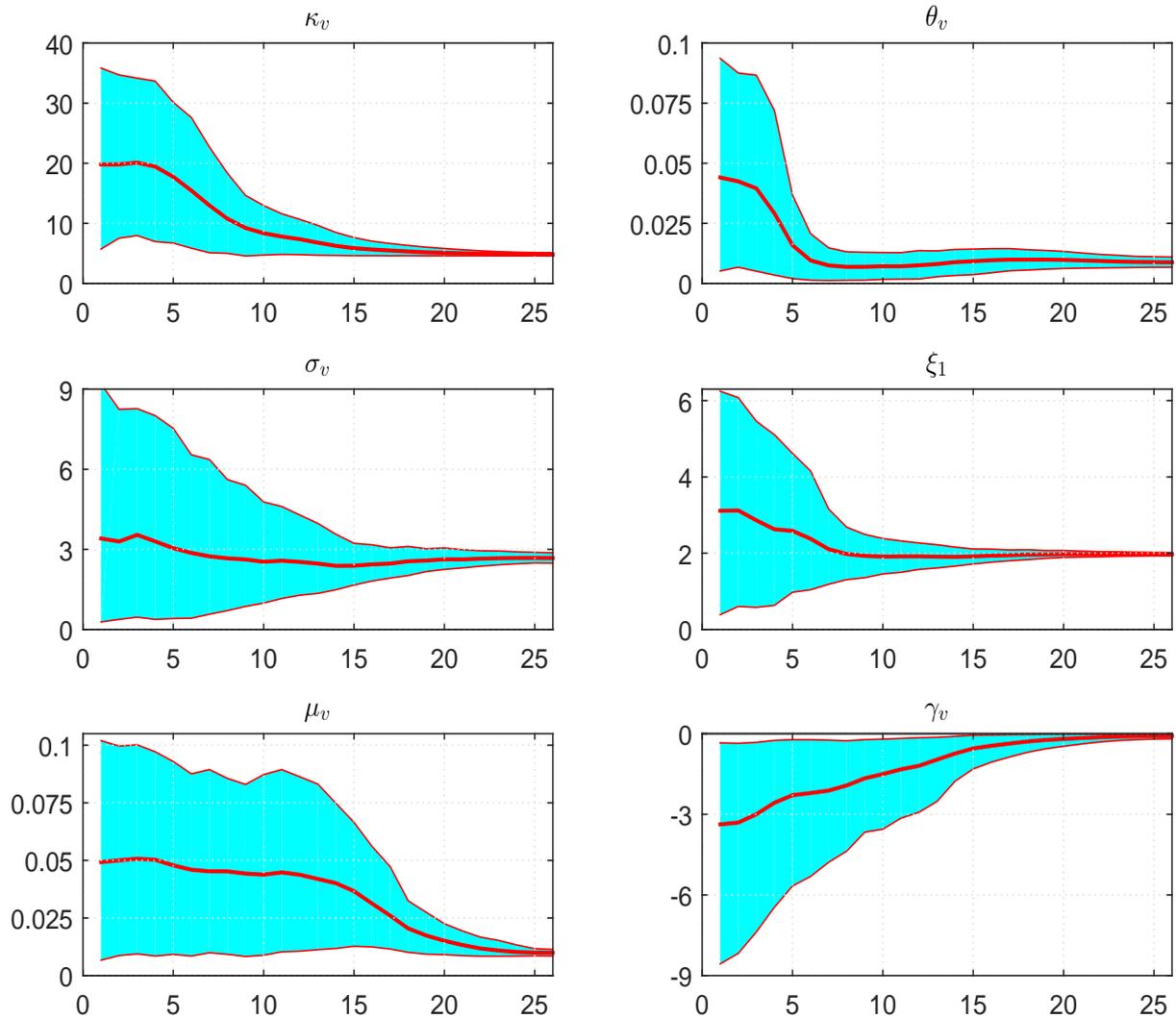


Figure 3: **Bridging the Priors and the Posteriors**

*Note:* The figure plots the tempering procedure for selected parameters in Model I. At each tempering stage, the mean and (5, 95)% quantiles of each of selected parameters are reported. The initial stage corresponds to the prior, and the last stage corresponds to the posterior.

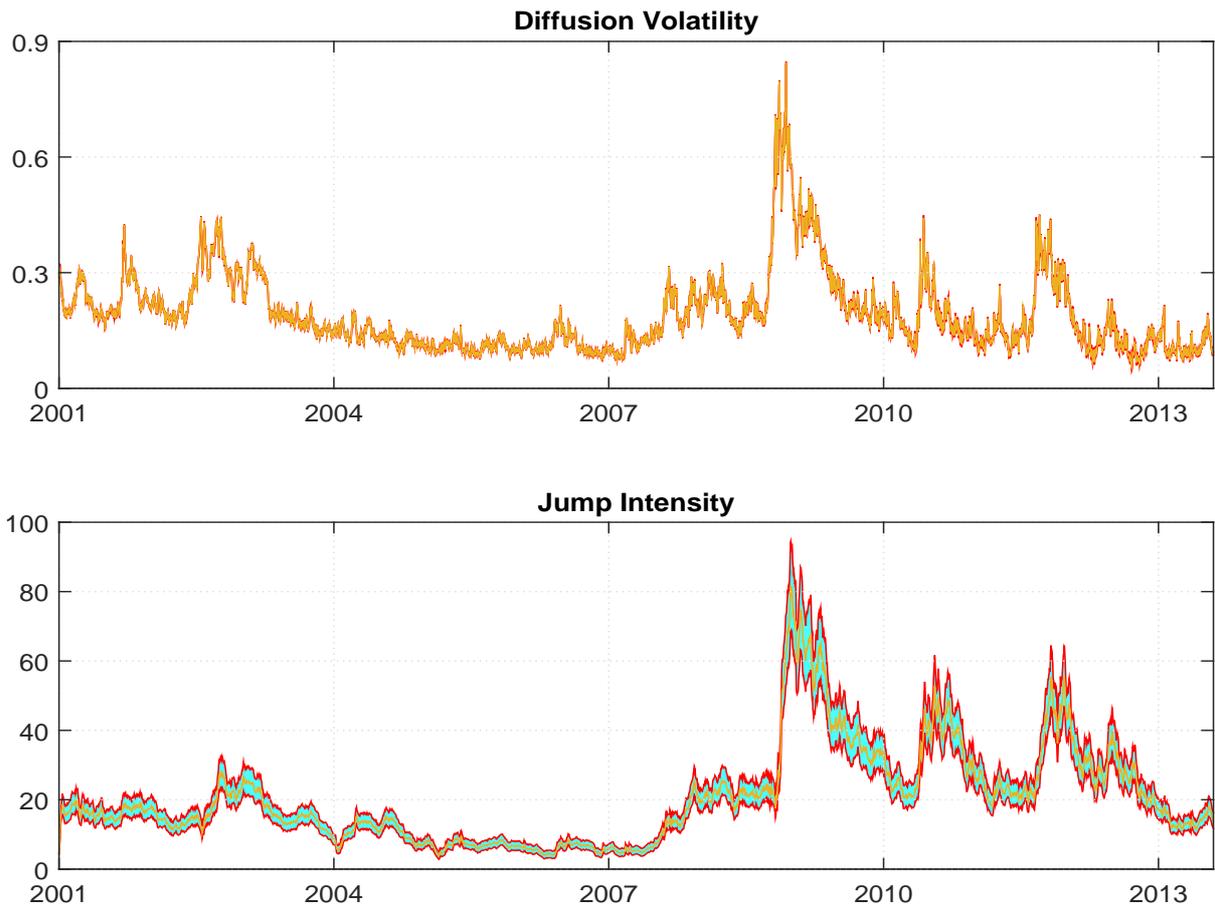


Figure 4: **Filtered Diffusion Volatility and Jump Intensity**

*Note:* The figure plots the filtered diffusion volatility and the filtered jump intensity in Model I using the real data on S&P 500 index returns and variance swap rates with the fixed maturity 1-, 6-, and 12-month, ranging from January 2, 2001 to July 15, 2013 for a total of 3.148 business days. At each time point, the mean and (5, 95)% quantiles are reported.

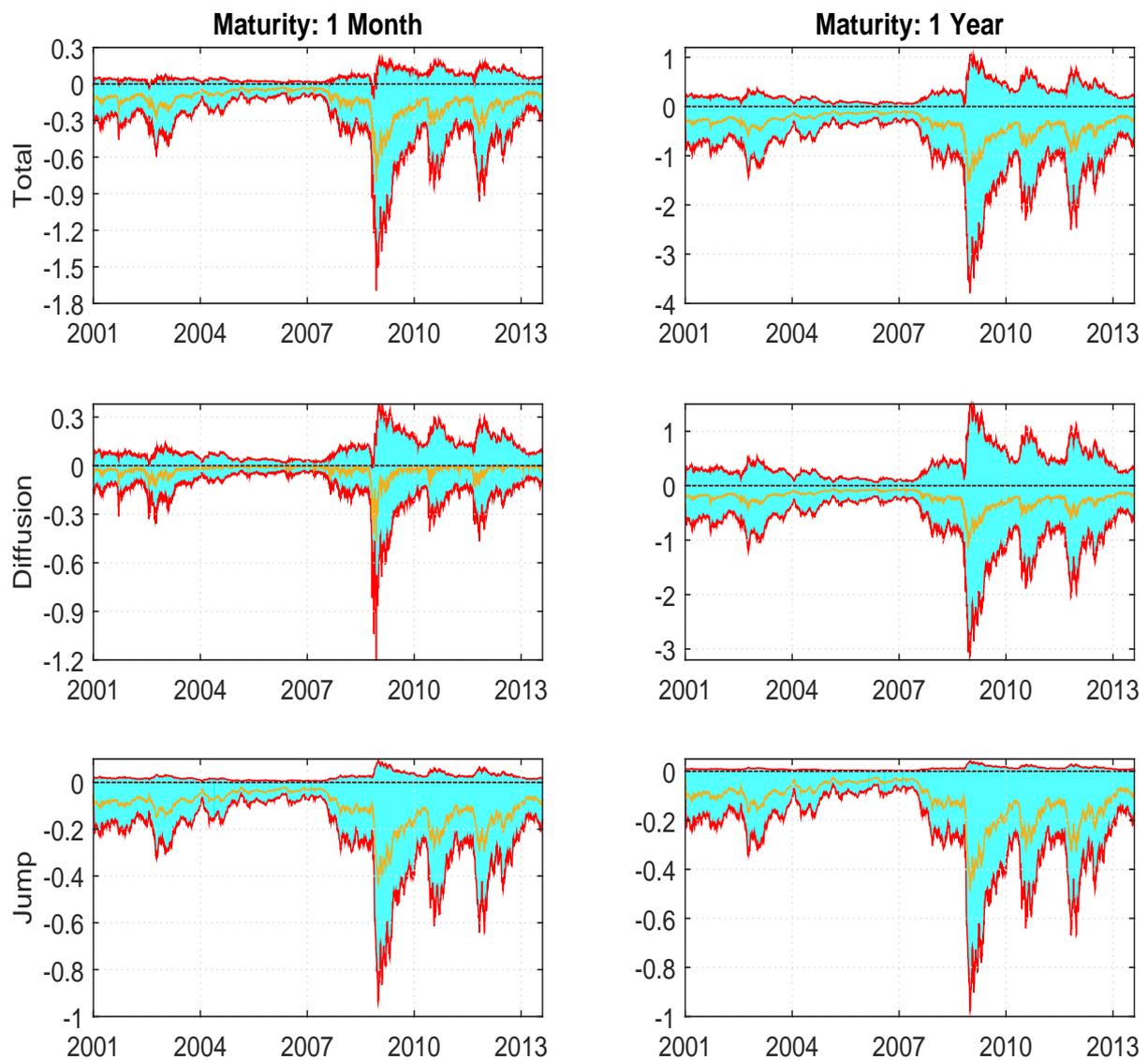


Figure 5: **Parameter Uncertainty and Variance Risk Premia**

*Note:* The figure plots the time series estimates of the total variance risk premia, the diffusion variance risk premia, and the jump variance risk premia for maturity 1 month and 1 year in Model CEV using the real data on S&P 500 index returns and variance swap rates with the fixed maturity 1-, 6-, and 12-month, ranging from January 2, 2001 to July 15, 2013 for a total of 3.148 business days. At each time point, the mean and (5, 95)% quantiles are reported.