Received: 18 September 2015 | Revised: 14 March 2017 | Accepted: 29 April 2017
DOI: 10.1002/asmb.2258

DISCUSSION PAPER

WILEY

Sequential Bayesian learning for stochastic volatility with variance-gamma jumps in returns

Samir P. Warty[1] | Hedibert F. Lopes[2] | Nicholas G. Polson[3]

[1]Analysis Group, Inc., 180 N. Stetson Ave. Chicago, IL, 60601, USA

[2]Insper - Institute of Education and Research, São Paulo, Brazil

[3]The University of Chicago Booth School of Business, Chicago, IL, USA

Correspondence
Hedibert Lopes, Insper, Rua Quatá, 300. Vila Olímpia, São Paulo/SP. Zip 04546-042, Brazil.
Email: HedibertFL@insper.edu.br

In this work, we investigate sequential Bayesian estimation for inference of stochastic volatility with variance-gamma (SVVG) jumps in returns. We develop an estimation algorithm that combines the sequential learning auxiliary particle filter with the particle learning filter. Simulation evidence and empirical estimation results indicate that this approach is able to filter latent variances, identify latent jumps in returns, and provide sequential learning about the static parameters of SVVG. We demonstrate comparative performance of the sequential algorithm and off-line Markov Chain Monte Carlo in synthetic and real data applications.

KEYWORDS
auxiliary particle filtering, Bayesian learning, sequential Monte Carlo, stochastic volatility, variance gamma

1 | INTRODUCTION

Stochastic volatility (SV) models with jumps have become standard tool in asset pricing finance literature.[1] Lévy-driven SV models provide a highly adaptable modeling framework that can handle many asset pricing problems regularly encountered. Models with Lévy jumps offer great flexibility for modeling returns. The stochastic volatility variance-gamma (SVVG) model is particularly useful with a wide range of finance applications. The goal of our paper is to show sequential Bayesian methods[2] can be applied to this class of models. Specifically, a Lévy-driven jump diffusion with a leverage effect for the price process is used where the jump component of the return process obtains as a Brownian motion with drift subordinated to a gamma subordinator. This is the so-called variance gamma (VG) process of Madan and Seneta.[3] The latent SV is a square-root diffusion that can be correlated with diffusion in the price process. A direct implication and benefit of SVVG's modeling structure is that the leverage effect can be modeled explicitly.

SVVG's combination of flexibility and parsimony makes it an attractive modeling choice for researchers and practitioners alike. The VG process can have infinite jump arrival rate within finite time intervals. The infinite-activity jump process has been proposed as a model for markets with high liquidity and high activity. Carr et al[4] later extended the VG process by adding a parameter to permit finite or infinite activity and finite or infinite variation.

Unlike Brownian motion with infinite activity, the sum of absolute jumps within a finite time interval is bounded for the standard VG process.[3] Infinite activity with finite variation gives SVVG an advantage over the finite-activity compound Poisson process of competing affine jump diffusion (AJD) models. Prior studies have shown that traditional AJD models cannot adequately account for infinite-activity jumps in index returns, even at daily sampling frequencies. In particular, Li et al[6] find that infinite activity jumps in returns are better able to model high kurtosis as is often observed in asset return data at daily and higher sampling frequencies.

Our work focuses attention on developing algorithms for SVVG estimation using sequential importance sampling/resampling scheme of Carvalho et al.[2] Despite the modeling conveniences it offers, SVVG has not received great attention in the statistical finance literature because of computational challenges in estimation, filtering, smoothing, and prediction. Much of the literature on the VG process tends to focus on modeling returns with a VG distribution directly rather than embedding VG via jumps in an SV model.

Likelihood-based inference for the VG model is given in Carr et al[4] and Seneta.[5] However, intractability of the likelihood complicates frequentist analysis of SVVG. The gamma-subordinated Gaussian jump process produces a nonlinearity in the SVVG likelihood, which cannot be marginalized analytically. Furthermore, analysis of the marginal SVVG jump distribution is complicated and requires computationally costly evaluations of special functions.

Li et al[6] develop Markov chain Monte Carlo (MCMC) methods for Bayesian inference of SVVG. They show that diffusion, latent SV, and infinite-activity jumps in returns can be identified jointly using discretely observed asset prices. In particular, enough information exists in daily index returns to identify parameters and latent states of SVVG. They also show that Lévy jump models can outperform traditional AJD models for daily returns. They conclude that infinitely many small jumps are too large to be modeled by Brownian motion and too small for the compound Poisson process.

Sequential simulation–based estimation has the potential to offer computational efficiency gains over bulk estimation techniques such as MCMC. In addition, sequential, or "online," techniques yield tangible benefits over off-line methods in high-frequency and low-latency settings where the arrival rate of new information requires very rapid updating of posteriors to perform inference. We show that estimation via sequential Bayesian learning is feasible and produces adequate estimates for SVVG, see also Jasra et al[7] who propose a sequential Monte Carlo (SMC) estimation algorithm for SVVG based on the SMC scheme of Del Moral et al.[8]

The rest of the paper is outlined as follows. Section 2 provides technical discussion of the SVVG modeling framework. Section 3 describes previous approaches to posterior inference of SVVG via MCMC. Section 4 develops our sequential learning algorithm. Section 5 presents simulation results on the performance of our algorithm and shows its comparison to traditional MCMC methods. Section 6 illustrates our methodology in a real-data setting on SP500 returns. Section 7 discusses limitations of our filter and related work. Details of prior distributions are provided in a technical appendix.

2 | METHODOLOGY

2.1 | Model specification

The SVVG model consists of a jump diffusion price process with gamma-subordinated Brownian motion jumps, a correlated Cox-Ingersoll-Ross (CIR) variance process[9] and leverage in the price process. To make the model specification more precise, consider the following description of the model's dynamics. Let Y_t be the natural logarithm of a stock's price, J_t is the jump component of returns, and $p_t = Y_t - J_t$. The sample paths of the SVVG model satisfy the system of stochastic differential equations:

$$dY_t = \mu dt + \sqrt{v_t} dW_1(t) + dJ_t \tag{1}$$

$$dv_t = \kappa(\theta - v_t)dt + \rho\sigma_v\sqrt{v_t}dW_1(t) + \sigma_v\sqrt{(1-\rho^2)v_t}dW_2(t) \tag{2}$$

$$dJ_t = \gamma dg(t; 1, \nu) + \sigma_J dW_3(g(t; 1, \nu)) \tag{3}$$

where $W_1(t)$, $W_2(t)$, and $W_3(t)$ are independent Wiener processes evaluated at time t; μ is the drift in log price or expected return; v_t is the instantaneous variance of returns; θ is the stationary mean of the variance process; κ is the speed of mean-reversion of the variance process; σ_v is the volatility of volatility, or σ_v^2 is the conditional variance of the variance process; ρ measures instantaneous correlation of returns and variance, ie, leverage; γ is the conditional drift in jumps; σ_J is the volatility of jumps conditional on the random time change; and $g(t; 1, \nu)$ is a gamma process time change with unit mean rate and variance rate ν.

Bayesian inference can be performed by using an Euler approximation of the continuous-time model with daily sampling frequency $\Delta = 1$. Discretization is a modeling convenience for SVVG that makes the model more tractable with standard sampling techniques. The continuous-time transition density of the latent variance $p(v_{t+\tau}|v_t, \Theta)$ is known to be noncentral gamma.[10] The conditional transition density of the observation equation $p(Y_{t+\tau}|Y_t, J_{t+\tau}, \dots, J_t, v_{t+\tau}, \dots, v_t, \Theta)$ is Gaussian only if the entire paths of the latent variance process v_t and latent jump process J_t are known from t to $t + \tau$. Analytical marginalization of the jump process presents an intractable problem. As an alternative to numerical or Monte Carlo integration, discretization provides a reasonable approximation to continuous-time SVVG and is commonly used in the literature. Discretization bias, see Eraker et al,[1] is typically small with daily data.

The discretized SVVG model is described by the state-space model:

$$r_t = \mu\Delta + \sqrt{v_{t-1}\Delta}\varepsilon_{1,t} + J_t \tag{4}$$

$$\begin{aligned} v_t = v_{t-1} + \kappa(\theta - v_{t-1})\Delta + \rho\sigma_v\sqrt{v_{t-1}\Delta}\varepsilon_{1,t} \\ + \sigma_v\sqrt{(1-\rho^2)v_{t-1}\Delta}\varepsilon_{2,t} \end{aligned} \tag{5}$$

$$J_t = \gamma G_t + \sigma_J\sqrt{G_t}\varepsilon_{3,t} \tag{6}$$

$$G_t \sim \mathcal{G}\left(G_t; \frac{\Delta}{\nu}, \nu\right) \tag{7}$$

where r_t is the log return $(Y_t - Y_{t-1})$; $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{3,t}$ are independent standard Gaussian random variables; J_t is the jump in returns; v_{t-1} is the return variance; G_t is the gamma-distributed time change with mean Δ and variance $\Delta\nu$; and the model's static parameters $\Theta = \{\mu, \kappa, \theta, \sigma_v, \rho, \gamma, \sigma_J, \nu\}$ have the same interpretations as in the continuous-time model. In the discrete time specification, the observation equation is given by Equation 4 and the latent states' evolutions are described by Equations 5, 6, and 7.

Stochastic volatility models can account for excess kurtosis in returns through autocorrelation in the volatility series and feedback effects between the return and volatility processes. The jump process J_t provides the SVVG model an additional mechanism whereby excess kurtosis and skewness in the return distribution can be modeled. The marginal moments of the the jump process J_t are functions of γ, σ_J^2, and ν. In particular, the sign of γ determines the sign of the skewness of J_t, and ν indicates percentage excess kurtosis relative to Gaussian kurtosis[11] when $\gamma = 0$.

2.2 | Prior specification

Completion of the Bayesian model structure requires specification of prior distributions for the static parameter vector Θ and the initial latent variance state v_0. The prior dependence structure is as follows:

$$\pi(\Theta, v_0) = \pi(\mu)\ \pi(\kappa)\ \pi(\theta)\ \pi(\sigma_v, \rho)\ \pi(\gamma)\ \pi(\sigma_J)\ \pi(\nu)\ \pi(v_0).$$

The Bayesian model considered here follows closely the general specification proposed by Li et al.[6] All prior distributions are proper and conjugate where conjugacy exists. This simplifies posterior inference somewhat by providing a more intuitive analytical framework. Conjugacy also allows for the use of efficient direct sampling of full conditional posterior distributions. Prior hyperparameters are chosen to be rather uninformative and, where applicable, consistent with unconditional moments of historical data. For most parameters of SVVG, the likelihood overwhelms the contribution from the priors rather quickly for the sample sizes used in the simulation and empirical studies considered here. In limited testing on synthetic data, these prior choices often provide good results for fitting SVVG to individual asset returns.

The initial variance state v_0 is assumed to be gamma on the positive real line. Prior beliefs of the static parameters are modeled as follows. Expected drift in log returns μ is Gaussian, with mean value equal to historical mean return of the asset being analyzed. Mean reversion of the variance process is governed by κ (speed of mean reversion in variance) and θ (long-run variance). Truncated Gaussian priors on both parameters ensure the nonnegativity constraint of the CIR variance process holds. The prior belief for κ favors slow mean reversion in the variance process (ie, small values of κ). The prior mean for θ is chosen to be consistent with historical unconditional return variance of the asset being analyzed. Priors for static parameters of the jump and time-change processes are informed by previous SVVG calibration studies where available.

The volatility of volatility parameter σ_v and leverage parameter ρ are modeled jointly in the prior structure. Under reparameterization, the estimation of ρ and σ_v is framed as estimation of the slope and error variance of regression of the volatility shock ε_2 on the return shock ε_1. This regression naturally elicits the common conditional Gaussian-inverse gamma prior on the transformed parameter space. Operationally, this modeling choice induces a diffuse prior on the leverage parameter ρ. The prior choice follows the treatment of the correlated errors SV model of Jacquier et al.[12]

An alternate Bayesian analysis of SVVG by Jasra et al [7] takes a different route to prior elicitation for parameters governing the variance process. The prior dependence structure keeps ρ independent of the other parameters a priori. Rather, an improper joint prior links the 3 variance parameters κ, θ, and σ_v. Jasra et al[7] impose CIR regularity conditions on these parameters to satisfy the nonnegativity constraint for the variance process. In the sequential Bayesian algorithm developed here, this restriction is imposed through truncated Gaussian priors for volatility mean-reversion parameters κ and θ and the joint prior structure for

σ_v and ρ. As discussed above, Jasra et al[7] also do not pursue estimation of the Gamma time-change variance rate parameter ν, instead fixing its value in their simulation and empirical studies.

Appendix A.1 provides further technical detail regarding the specific prior structure used in the analysis.

3 | POSTERIOR INFERENCE VIA MCMC

The Bayesian analysis of Li et al[6] uses Gibbs sampling for all static parameters and latent states and incorporates adaptive rejection Metropolis sampling (ARMS) steps to sample difficult-to-sample (and potentially non–log-concave) posteriors. This includes ARMS steps for the gamma process variance rate ν and sequences of latent variances v^T and time-changes G^T.

The gamma process variance rate ν requires special attention in developing the estimation routine. The posterior of ν is not log-concave. This phenomenon arises directly from the nonlinear fashion in which ν enters the likelihood of G_t at Equation 7 through both the shape and rate/scale parameters of the gamma distribution. This non–log-concavity complicates the Gibbs sampling step for ν in the MCMC algorithm of Li et al.[6] To overcome this issue, the algorithm uses ARMS steps to sample posteriors that cannot be sampled directly.[13]

ARMS is an extension of adaptive rejection sampling (ARS).[14] At each sampling step, ARS constructs an envelope of the target log density, which is used as the proposal density in a rejection sampling scheme. The envelope is progressively updated whenever a point is rejected. When the target log density is not concave, ARMS extends ARS by including Metropolis steps for accepted proposal points. The procedure is exact for sampling from the target density regardless of the degree of convexity, although it can have a high-computational cost when the target log density is complicated.

4 | SEQUENTIAL BAYESIAN LEARNING OF SVVG

We estimate the parameters of the SVVG model with an SMC sampling scheme based on Carvalho et al.[2] An overview of the algorithm's implementation follows.

Let $x_t = (v_t, v_{t-1}, J_t, G_t)$ be the vector of latent states entering the augmented likelihood at time t. This algorithm approximates the filtering distribution of states and parameters $p(x_t, \Theta|r^t)$ sequentially using the discrete representation $\hat{p}(x_t, \Theta|r^t) = \sum_{i=1}^{M} \delta_{(x_t,\Theta)^{(i)}}(x_t, \Theta)$, where $\{(x_t, \Theta)^{(i)}\}_{i=1}^{M}$ is the set of particles in the discrete approximation at time t and $\delta_{(x_t,\Theta)^{(i)}}(\cdot)$ is the Dirac delta function located at $(x_t, \Theta)^{(i)}$. The particle distribution is updated sequentially as new information arrives. The algorithm presented here uses a resample-propagate-type approach similar to the auxiliary particle filter developed by Pitt and Shephard.[15]

For the state-space model described at Equations 4 to 7, the joint posterior for the static parameters conditional on the full state space $p(\Theta|x^t, r^t)$ can be written as $p(\Theta|z_t)$, where z_t is a vector of parameter conditional sufficient statistics computed deterministically according to an updating rule $z_t = \mathcal{Z}(z_{t-1}, x_t, r_t)$. Conditioning on conditional sufficient statistics achieves variance reduction in resampling weights.

The algorithm can be summarized as follows:

1. **Blind propagation of J_t.** Draw candidate jump state particles $\{\tilde{J}_t^{(i)}\}_{i=1}^{M}$ with proposal density $p\left(\tilde{J}_t|\left(\gamma, \sigma_J^2, \nu\right)^{(i)}\right)$.
2. **Resample particles.** Draw $\{(J_t, x_{t-1}, \Theta, z_{t-1})^{(i)}\}_{i=1}^{M}$ from $\{(\tilde{J}_t, x_{t-1}, \Theta, z_{t-1})^{(i)}\}_{i=1}^{M}$ with weights $w_t^{(i)} \propto p\left(r_t|(v_{t-1}, J_t, \mu)^{(i)}\right)$.
3. **Propagate v_t and G_t.** Draw $G_t^{(i)}$ from density $p(G_t^{(i)}|(J_t, \Theta)^{(i)})$. Draw $v_t^{(i)}$ from density $p\left(v_t^{(i)}|r_t, (v_{t-1}, J_t, \Theta)^{(i)}\right)$.
4. **Update** z_t. Compute $z_t^{(i)} = \mathcal{Z}\left(z_{t-1}^{(i)}, x_t^{(i)}, r_t\right)$.
5. **Sample** Θ. Draw $\Theta^{(i)} \sim p\left(\Theta^{(i)}|z_t^{(i)}\right)$.

Details of the algorithm, including sampling densities, posterior densities, and conditional sufficient statistics, are given in Appendices A.2 and A.3.

The filter presented here is not fully adapted because of difficulties presented by the form of the likelihood. Work is ongoing to address this deficiency. For SVVG, the likelihood based only on observable r_t is not available in closed form. There is no analytical closed form when integrating both the contemporaneous latent jump J_t and time-change G_t from the observation equation at Equation 4. Further, numerical and Monte Carlo integration methods are not practical in this application. The fully adapted filter that resamples according to weights $w_t^* \propto p(r_t|(x_{t-1}, z_{t-1}, \Theta)^{(i)})$ does not appear to be readily available. The filter is developed as a hybrid of the extended bootstrap filter of Storvik [16] and the sequential learning filter of Carvalho et al. [2] This filter has produced good results in synthetic and real-data applications.

Storvik's filter addresses the partial adaption problem by generating proposal draws for the problem latent states to evaluate the likelihood on an augmented space of observables. The current set of particles is resampled according to weights proportional to the conditional likelihood scaled by a factor to adjust for proposal draws that do not come from the "correct" posterior distribution. This adjustment takes the form of the the ratio of the true predictive density to the propagation proposal density:

$$\begin{aligned} w_t^{(i)} &\propto p\left(r_t|\tilde{J}_t^{(i)}, v_{t-1}^{(i)}, \Theta^{(i)}\right) \frac{p\left(\tilde{J}_t^{(i)}|J_{t-1}^{(i)}, v_{t-1}^{(i)}, \Theta^{(i)}\right)}{q\left(\tilde{J}_t^{(i)}|J_{t-1}^{(i)}, v_{t-1}^{(i)}, \Theta^{(i)}, r^t\right)} \\ &= p\left(r_t|\tilde{J}_t^{(i)}, v_{t-1}^{(i)}, \Theta^{(i)}\right) \frac{p\left(\tilde{J}_t^{(i)}|\Theta^{(i)}\right)}{q\left(\tilde{J}_t^{(i)}|r_t, \Theta^{(i)}\right)} \end{aligned}$$

where the simplification obtains because J_t has no serial dependence.

Sampling directly from the posterior $p(\tilde{J}_t^{(i)}|r_t, \Theta^{(i)})$ as a proposal density is not trivial. Instead, the blind propagation proposal $q(\tilde{J}_t^{(i)}|\Theta^{(i)})$, ie, the marginal of J_t conditional on the static parameter set, given above is used. The proposal and evolution densities are identical, which yields resampling weights proportional to the conditional likelihood of $(r_t|J_t)$. The algorithm using blind propagation gives good performance.

The sequential estimation algorithm approaches sampling of static parameters similarly to Li et al.[6] Static parameters for which conditional conjugacy exists are sampled efficiently from their full-conditional distribution in Gibbs sampler steps. It does not appear that there are any efficient, specialized algorithms for sampling from the full-conditional posterior distribution of the Gamma time-change variance rate parameter ν. Following Li et al,[6] we retain an ARMS step to sample ν.

ARMS allows for sampling from the exact target density even in the absence of log-concavity. However, ARMS also requires initial construction of an envelope and multiple evaluations of the log-density, which can be very expensive computationally for complicated log-density functions. The computational toll is large and especially so for sequential algorithms due to the multiplicity of evaluations during each update step. As such, sampling ν represents somewhat of a computational bottleneck in sequential estimation of SVVG. We have considered alternate methods, including discrete sampling of ν and alternate prior distributions. However, alternate approaches to estimation of ν do not prove to be better than ARMS. These are discussed in greater detail below with the estimation results.

We propagate the latent variance v_t and latent time-change G_t from their respective conditional posterior distributions. The latent volatility is conditionally Gaussian. The time-change is distributed generalized inverse Gaussian (GIG). Li et al[6] sample G_t with an ARMS step in their MCMC algorithm. However, the overhead of this sampling choice is reduced by implementing efficient sampling of the GIG density using the algorithm of Leydold and Hörmann,[17] which updates a commonly used algorithm proposed by Dagpunar.[18] Further detail is given at A.2.11.

The sequential estimation algorithm outputs a sequence of particle distributions that constitute discrete approximations to filtered posteriors of the latent states and static parameters given by $p(x_t, \Theta|r^t)$. For inference of the latent states and static parameters, we rely on summaries of the sequential particle distributions that are by-products of the sampling. Time series of point estimates for static parameters and latent states obtain from cross-sectional means of the sequential particle distributions. We measure posterior uncertainty by cross-sectional standard deviations and extreme quantiles of the particle sampling distributions.

Availability of the sequence of particle distributions also allows the user to approximate integrals of interest numerically. In particular, one can obtain approximations to the sequential marginal likelihoods, which can be used to compute sequential Bayes factors for model comparison.[19] The marginal predictive density $p(r_{t+1}|r^t)$ can be approximated using the algorithm's output at time t:

$$\hat{p}(r_{t+1}|r^t) = \frac{1}{M}\sum_{i=1}^{M} p\left(r_{t+1}|(x_t, \Theta)^{(i)}\right)$$

where $\{(x_t, \Theta)^{(i)}\}_{i=1}^{M}$ is the filtered particle distribution at time t.

An alternate filter specification also considered consisted of blind propagation of G_t and resampling with weights proportional to $p(r_t|G_t, \Theta)$. However, this alternate filter performed very poorly compared to the approach described above in detail. It is unclear why this is the case, although it appears to be an issue of differing conditioning information sets. Blind propagation of G_t relies on filtered information on ν in the initial propagation step. Blind propagation of J_t uses information encoded in the particle distributions of γ and σ_J^2 in addition to ν. The 3-tuple $\left(\gamma, \sigma_J^2, \nu\right)$ fully specifies the first 4 marginal moments of the jump process,[11] so posterior samples of these parameters should carry more information about the posterior distribution of interest than using only samples of ν. It is possible that the filter used here may not perform as well as the alternate filter when a considerably smaller number of particles is used and the posterior samples of $\left(\gamma, \sigma_J^2, \nu\right)$ do not adequately explore the parameter space.

The aforementioned assumption of Jasra et al[7] regarding the gamma process variance rate ν implies G_t is distributed exponential with scale parameter Δ. This assumption does not appear to be validated by estimation results for daily data, although the authors estimate their model on hourly returns. This assumption has implications for the marginal moments of the jump process. When the magnitude of γ is small, the marginal distribution of J_t would be near mesokurtic. This would understate the jump process's variance. This could potentially impair the model's ability to adequately identify large jump events, which tend to happen at frequencies suggested by a leptokurtic distribution. The sequential algorithm presented here models ν as an unknown static parameter and allows the user to learn about it sequentially. In particular, the claim that $\nu = \Delta$ can be tested in a sequential manner using the algorithm developed here.

5 | SIMULATION STUDY

To illustrate the performance of the sequential filter for SVVG, we apply the methodology to synthetic data simulated from the discrete-time SVVG model at Equations 4 to 7. The simulated data consist of approximately 20 years of data ($T = 5000$). Figure 1 demonstrates one realization from the model. As a comparison, we also report the performance of the MCMC algorithm of Li et al.[6] All estimation routines were written in C. Sequential particle filtering used $M = 10\,000$ particles; tables and figures report summaries over 10 particle learning (PL) runs. MCMC retained 20 000 sampler iterations after burn-in of 30 000 iterations.

Figure 2 shows the time series of true instantaneous variances v_t, sequential filtered estimates $E[v_t|r^t]$, and smoothed estimates $E[v_t|r^T]$ from MCMC. Sequential filtering of v_t appears to perform well and compares favorably with the performance of MCMC. However, there are clear problem spots where the filtered variances fail to track the latent states closely. In particular, sequential estimation suffers from a deficiency common to many filtering algorithms. Sequential filtering tends to underestimate latent variance when the true latent state attains very large values for short periods. During short episodes of high-latent variance, the mass of the posterior distribution shifts substantially. It is promising, although, that the particle filter can move its particles

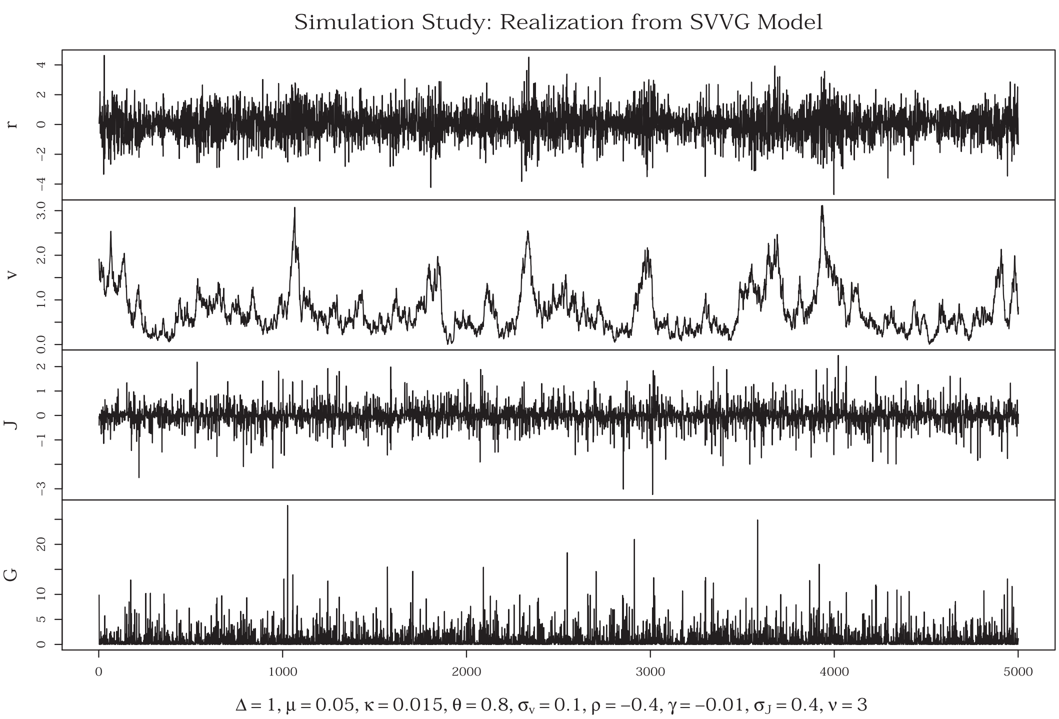

FIGURE 1 Simulated realization from the stochastic volatility with variance-gamma (SVVG) model ($T = 5000$)

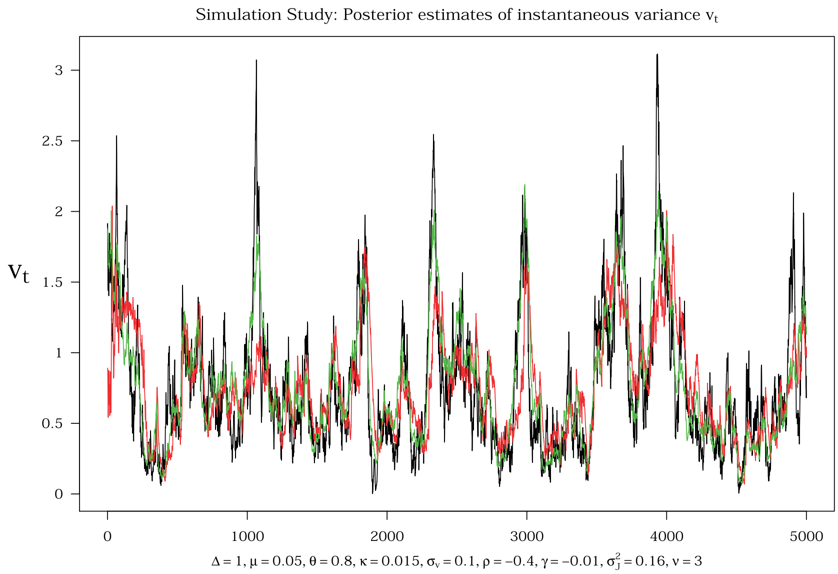

FIGURE 2 Posterior estimates of instantaneous variance v_t for simulated data. True series shown in black, sequential posterior mean estimates $E[v_t|r^t]$ in red, and Markov chain Monte Carlo posterior mean estimates $E[v_t|r^T]$ in green [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Posterior inference of stochastic volatility with variance gamma for 20 years of simulated daily returns ($T = 5000$)

		Sequential			MCMC		
	Truth	**Mean**	**SD**	**RMSE**	**Mean**	**SD**	**RMSE**
μ	0.050	0.0577	0.0099	0.0166	0.0598	0.0168	0.0195
θ	0.800	0.7129	0.0327	0.0930	0.7985	0.0919	0.0919
κ	0.015	0.0055	0.0023	0.0096	0.0125	0.0029	0.0039
σ_v	0.100	0.1539	0.0022	0.0540	0.0828	0.0089	0.0194
ρ	−0.400	−0.3749	0.0183	0.0310	−0.3504	0.0725	0.0878
γ	−0.010	0.0137	0.0034	0.0239	−0.0156	0.0149	0.0159
σ_J^2	0.160	0.2020	0.0059	0.0425	0.1442	0.0255	0.0300
ν	3.000	3.1813	0.0505	0.1882	3.1186	0.0866	0.1468

Abbreviation: MCMC, Markov chain Monte Carlo; RMSE, root mean squared error.

sufficiently fast across the support of the distribution to avoid subsequent overestimation of latent variance once high variance periods subside.

Table 1 and Figure 3 demonstrate each algorithm's performance for the model's static parameters. These results indicate that the sequential algorithm is able to recover many of the static parameters and is able to do so quickly. In most cases, the particle approximations are quite good after only 500 returns have been processed. Sequential estimation performance in learning diffusion parameters is generally better than in learning jump parameters. Notable is the performance of the sequential filter for γ and κ, both of which are learned very poorly. Further, the very small posterior standard deviations and narrow credible bands for certain parameters suggest particle degeneracy might be a problem.

The particle distributions obtained here generally exhibit low contemporaneous correlations between parameters after roughly 500 to 1000 returns have been processed. The exception is κ and θ, which routinely have correlations in the range $[-0.8, -0.5]$. Correlation in the particle distributions could have a deleterious effect on some estimation runs if the initial particle distribution is chosen poorly. The results presented here attempt to mitigate this effect by running estimation in parallel with multiple chains each starting from individually sampled initial particle distributions. Results are then aggregated across sampling chains. This approach also minimizes certain computational issues associated with the numerical precision of weight and sampling calculations in the presence of a very large number of particles with large variance in particle weights.

Similar results obtain across alternate realizations from the SVVG model and when using different sets of initial particles. Implementing a more optimal propagation rule in the first step of the algorithm might mitigate this poor performance. Another

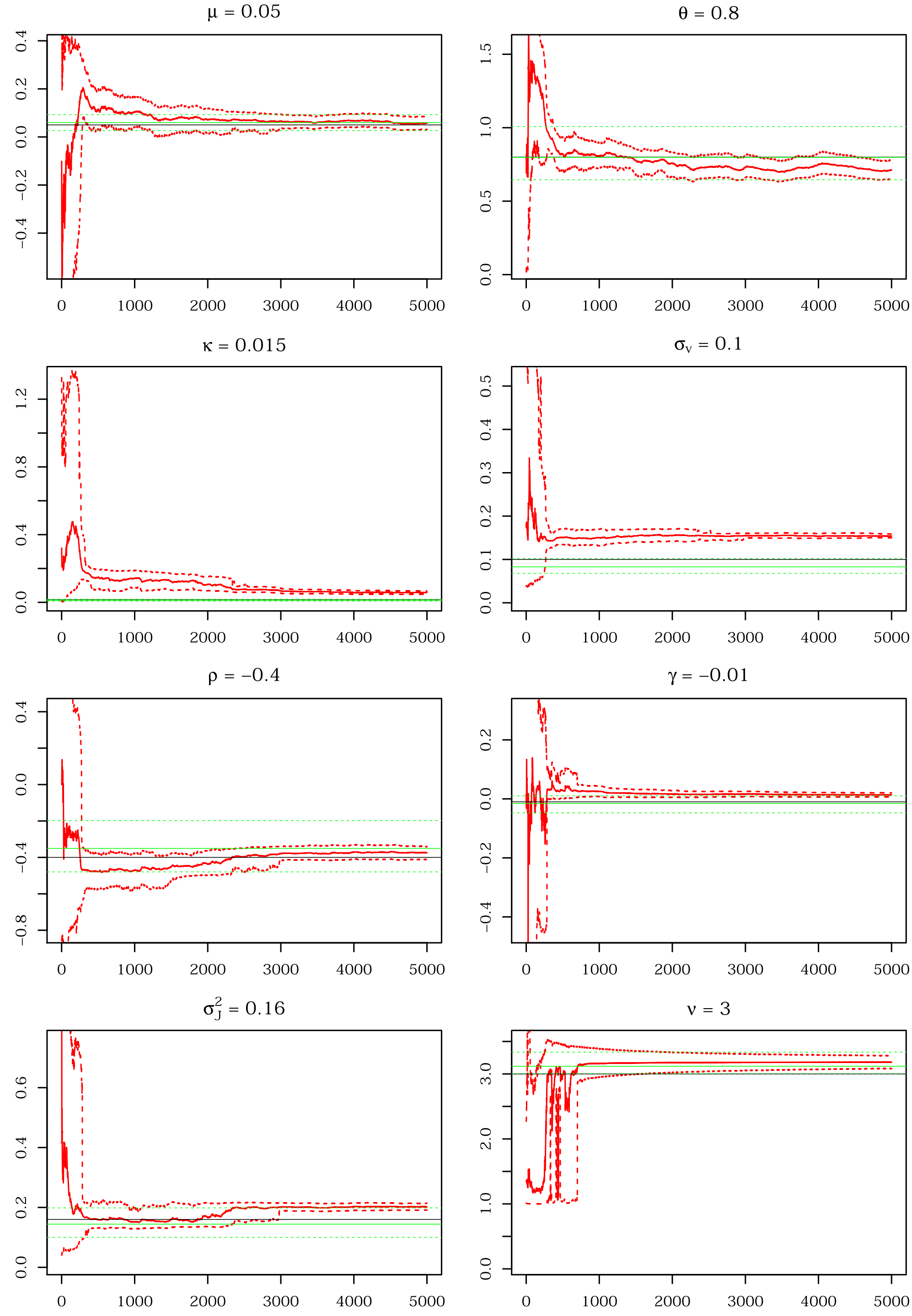

FIGURE 3 Sequential learning of static parameters for simulated data. True series shown in black, sequential posterior mean estimates $[E\Theta|r^t]$ in red, and Markov chain Monte Carlo posterior mean estimates $[E\Theta|r^T]$ in green. Dashed lines show posterior 95% credible intervals [Colour figure can be viewed at wileyonlinelibrary.com]

alternative would be to include a smoothing step to backward sample v_{t-1} at time t rather than simply resampling old particles. This latter strategy would nullify benefits of an online estimation strategy. However, this may be necessary to deal with degeneracy of the particle distribution when the sample size T grows very large, and the number of particles is limited by practical concerns such as computation time.

The performance of the sequential algorithm in parameter learning can be attributed largely to its adequate performance in filtering variance v_t and poor performance filtering jump J_t and time-change G_t for this set of static parameters. Running the estimation on realizations from alternate configurations of static parameters suggests that this sequential estimation algorithm struggles to identify the jump process when jump drift γ is near 0. We observe similarly poor performance when the gamma time-change variance rate ν is small. These conditions result in reduced skewness and reduced excess kurtosis in the marginal distribution of J_t. When this is the case, sample realizations of the J_t process appear very noisy with few large jump events. In turn, this provides little information with which to identify the jump process against random noise in the return process.

Sequential estimation's performance improves greatly as the marginal distribution of J_t becomes more skewed and heavier tailed, which are features that correspond to larger values of $|\gamma|$ and ν.

6 | REVISITING LI ET AL

In this section, we examine the empirical performance of the sequential Bayesian algorithm. In particular, we benchmark sequential estimation against the MCMC approach proposed by Li et al.[6] The results show that the sequential filter performs well when applied to actual data, although there appear to be some deficiencies that require additional investigation.

The data comprise 21 years of log returns for the S&P 500 Index from January 12, 1980, to December 29, 2000 ($T = 5307$). The sample period is the same as that of Li et al.[6] Table 2 reports summary statistics for the data sample. Figure 4 provides a time series plot of the log return series. Salient features of the data include what appear to be a few very large jumps in October 1987 (Black Monday), January 1989 and 1997 (Asian financial crisis), August 1998 (Russian default), and April 2000 (tech bubble crash), and periods of apparent volatility clustering in the early 1980s and late 1990s. The measure of the sequential algorithm's performance will depend on its ability to identify such stylized facts of the data. Our experience in estimation SV in periods of crisis[20] makes us believe that our model ought to perform very well in the European/Greek crisis. We feel however that this is an interesting topic for further exploration.

Figures 5 and 6 visualize sequential and MCMC posterior estimates of instantaneous variances and large jumps, respectively. As noted previously, we can expect the sequential variance estimates to be noisier than those produced by the MCMC smoother. The sequential estimates track the smoother estimates well. In particular, sequential learning does identify periods of high volatility. Further, sequential learning and MCMC are both able to identify the times of many of the same large jumps. However, the sequential estimates of jump sizes tend to be smaller and large positive jumps occur with greater frequency than in the MCMC estimates. This may be attributable to the 2 algorithms' disparate performance in estimating the model's static parameters.

Table 3 shows summaries of estimation of Θ. Relative to MCMC, sequential learning struggles with the mean-reversion speed of volatility κ and parameters associated with the jump process. The magnitudes of the posterior standard errors for sequential

TABLE 2 Summary statistics for 21-year sample of log returns of S&P 500 Index for January 2, 1980, to December 29, 2000

Sample Size	Mean	Variance	Skewness	Kurtosis	Minimum	Maximum
5307	0.0476	1.0437	−2.3578	55.5870	−22.8997	8.7089

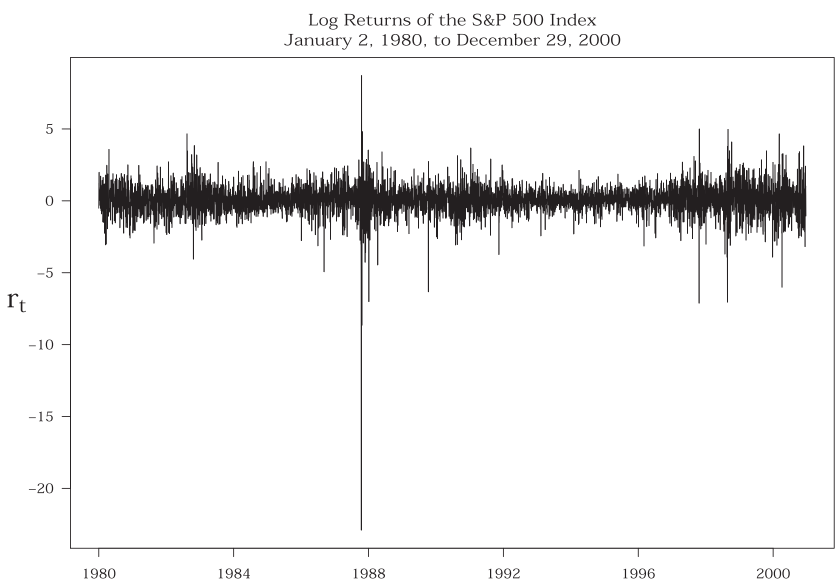

FIGURE 4 Log returns of the S&P 500 Index for January 2, 1980, to December 29, 2000 ($T = 5307$)

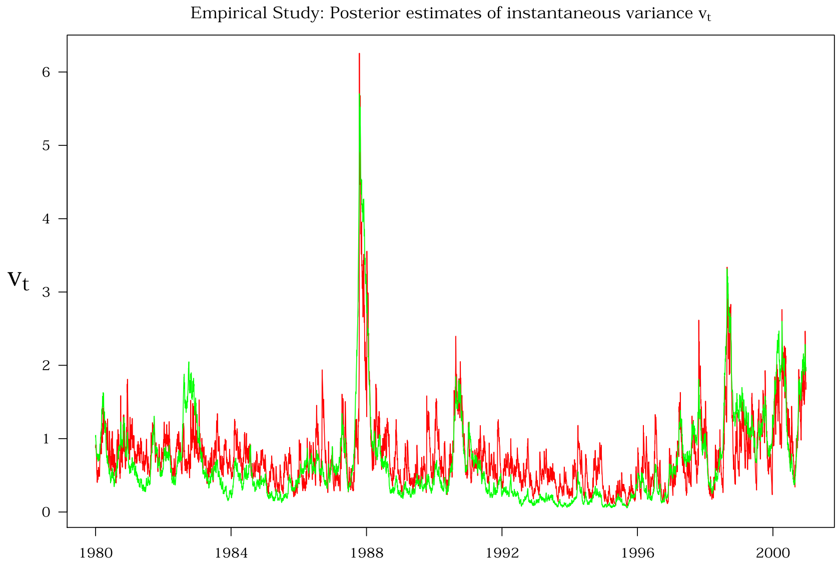

FIGURE 5 Posterior estimates of instantaneous variance v_t for S&P 500 Index data. Sequential posterior mean estimates $E[v_t|r^t]$ in red and Markov chain Monte Carlo posterior mean estimates $E[v_t|r^T]$ in green [Colour figure can be viewed at wileyonlinelibrary.com]

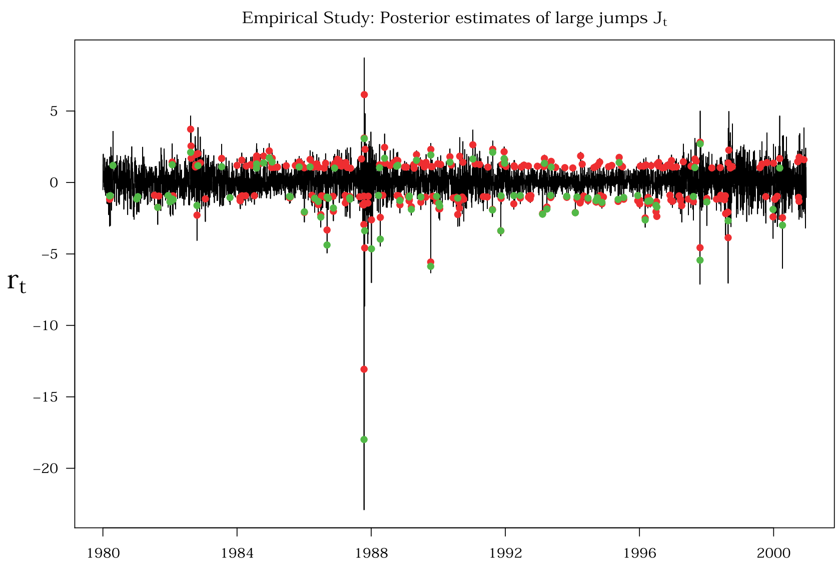

FIGURE 6 Posterior estimates of large jumps J_t for S&P 500 Index data. Log return series shown in black, sequential posterior mean estimates $E[v_t|r^t]$ in red, and Markov chain Monte Carlo posterior mean estimates $E[v_t|r^T]$ in green [Colour figure can be viewed at wileyonlinelibrary.com]

learning again suggest particle degeneracy might be an issue. Despite these issues, filtering of variances v_t and jumps J_t does not appear to suffer greatly. As documented in Figures 5 and 6, the sequential filter recovers similar latent variances as MCMC and manages to identify many of the same large jumps in returns.

It appears that identification of γ in both algorithms greatly influences the ability to identify jumps of either sign. The data may not bear enough information to deal with extreme jumps. Both algorithms have difficulty identifying large jumps with opposite sign as γ. Such jumps could conceivably be thought of as outliers as they occur far in the "short tail" of the marginal distribution of J_t, ie, in the opposite direction of the skewness. In this implementation of MCMC as well as in Li et al,[6] γ is estimated to be negative for the S&P 500 data, and negative jumps are identified with much greater frequency and intensity than positive jumps. The most notable instance is the apparent large positive jump on October 21, 1987, following the large decline on October 19, 1987. This phenomenon would also explain why the sequential approach may underestimate the magnitude of large negative jumps. The sequential estimates of average jump size γ are positive.

The time-change variance rate parameter ν represents a significant challenge to sequential estimation and inference of SVVG. As in the simulation study, sequential estimation of ν is very poor for this empirical example. The particle distribution

TABLE 3 Posterior inference of stochastic volatility with variance gamma for 21 years of daily S&P 500 Index returns ($T = 5,307$)

	Sequential		MCMC	
Parameter	**Mean**	**SD**	**Mean**	**SD**
μ	0.0722	0.0009	0.0768	0.0245
θ	0.8173	0.0643	0.8244	0.1720
κ	0.0225	0.0018	0.0096	0.0025
σ_v	0.1362	0.0010	0.1060	0.0085
ρ	−0.5455	0.0033	−0.5737	0.0697
γ	0.0519	0.0113	−0.0441	0.0261
σ_J^2	0.4210	0.0088	0.2442	0.0263
ν	1.6817	0.0274	5.7455	0.0968

Abbreviation: MCMC, Markov chain Monte Carlo.

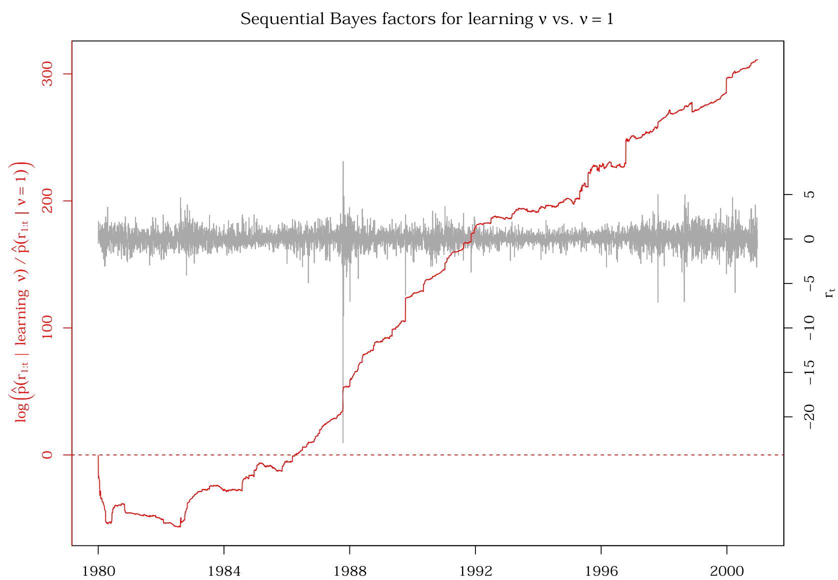

FIGURE 7 Sequential Bayes factors for learning of ν versus $\nu = 1$ for S&P 500 Index data (red line and left vertical axis). Observed return series is superimposed (grey line and right vertical axis) [Colour figure can be viewed at wileyonlinelibrary.com]

degenerates to a near point mass quickly. Poor estimation of this parameter is closely linked to difficulties in filtering the time-change G_t, which may stem from the absence of any serial dependence in the time-change process.

Another driver of the poor filtering performance for J_t and G_t is the blind propagation rule used in the first step of the algorithm. Blind propagation in this case introduces orthogonal noise into the particle approximation and has been shown to lead to degeneracy of the particle distribution. We are currently working to determine whether a fully adapted filter is available or if a more optimal proposal density would improve the activity of the filter.

Despite poor estimation performance of the time change, the sequential algorithm performs reasonably well filtering the latent variances and large jumps and learning the other static parameters. Estimation results with such robustness to poor identification of the time-change parameters and latent state could suggest model misspecification for these data. That is, treating the time-change variance rate ν as an unknown parameter may introduce an unnecessary parameter that is poorly estimated. The sequential algorithm's output offers a direct method of comparison to the case of Jasra et al[7] where ν is fixed and known. This comparison may be assessed with a sequential Bayes factor:

$$B_t = \frac{p(r_{1:t}|\text{ learning } \nu)}{p(r_{1:t}|\text{ fixed } \nu = 1)}$$

One would expect that evidence for $\nu \neq$)1 would increase when observing returns consistent with return distribution with high excess kurtosis indicated by large values of ν. Within the framework of SVVG, such returns may be attributable to very large jumps as well as high volatility in jumps. Results are summarized by Figure 7. In spite of poor identification, the estimation output suggests that evidence for $\nu \neq 1$ grows exponentially over time. As expected, evidence against $\nu = 1$ increases greatly when large magnitude returns are observed, which coincides with identification of large jumps in Figure 6. These results agree with MCMC inference for SVVG for these data here and in Li et al,[6] suggesting that poor estimation of ν is indicative of the sequential algorithm's performance rather than model misspecification.

7 | DISCUSSION

A more sophisticated observation model for SVVG may be introduced by generalizing the mean function from the constant conditional mean model specified at Equation 4. We may consider the following general specification:

$$r_t = f(\cdot)\Delta + \sqrt{v_{t-1}\Delta}\varepsilon_{1,t} + J_t, \tag{8}$$

where $f(\cdot)$ represents the generalized conditional mean return function. The choice of $f(\cdot)$ provides modeling flexibility by allowing the practitioner to incorporate other sources of information to solve the SVVG filtering and estimation problem. Perhaps the simplest such mean model extension is a "regression-in-mean" specification in which $f(\cdot)$ consists of a linear regression on an exogenous regressor x:

$$f(x) = \mu' + \beta_x x,$$

where μ' is, in general, different from μ considered earlier.

Well-chosen explanatory variables should improve modeling of the cross section of returns, thereby increasing the filtering algorithm's efficiency in learning the latent states and static parameters from the sequence of observed returns. This may mitigate model identification difficulties that arise from using information derived solely from observed returns of the asset under study. For example, observable measures of market volatility could potentially help to differentiate true jump events from large returns associated with periods of high volatility.

The use of the latent volatility v_{t-1} as a predictive regressor presents a particularly compelling regression-in-mean specification for SVVG:

$$f(v_{t-1}) = \mu' + \beta_v v_{t-1},$$

where β_v represents the equity risk premium associated with v_{t-1}. Such a specification would allow the algorithm to learn sequentially about the potentially changing market price of risk. Details of posterior inference for this SVVG extension are included at Appendix A.4.

Alternate explanatory variables may also prove to be good candidates for risk premium estimation in the regression-in-mean framework. Candidates for inclusion in the model include VIX index levels, option-based implied volatilities, or realized volatility measures for the asset under study. This latter model extension should be tractable within the sequential estimation setup but requires further evaluation of the likelihood and Bayesian model. This approach could also be extended further to the inclusion of exogenous regressors in the model of SV at Equation 5.

The regression-in-mean specification can also be used to mimic GARCH-in-mean modeling by including the contemporaneous latent volatility v_t in the observation equation (Equation 8) at time t. However, this could potentially introduce practical difficulty in model identification. The SVVG model so far considered attempts to extract all information about the latent states and static parameters from the single observable return series r^T. This can pose a difficult state filtering problem for SVVG. The posterior for G_t depends only on the parameter vector Θ and contemporaneous information. As such, the filter has very little information to identify G_t simultaneously with the other latent states and static parameters. Finally, our SMC scheme could be extended to conditionally Gaussian errors, including generalized volatility models with nonparametric innovation as in Jensen and Maheu.[21]

REFERENCES

1. Eraker B, Johannes M, Polson N. The impact of jumps in volatility and returns. *The J Finance*. 2003;58(3):1269-1300.
2. Carvalho C, Johannes M, Lopes H, Polson N. Particle learning and smoothing. *Stat Sci*. 2010;25(1):86-106.
3. Madan DB, Seneta E. The variance gamma (V.G.) model for share market returns. *J Bus*. 1990;63(4):511-524.

4. Carr P, Geman H, Madan DB, Yor M. The fine structure of asset returns: an empirical investigation. *J Bus*. 2002;75(2):305-332.
5. Seneta E. Fitting the variance-gamma model to financial data. *J Appl Probab*. 2004;41:177-187.
6. Li H, Wells MT, Yu CL. A Bayesian analysis of return dynamics with Lévy jumps. *The Rev Financial Stud*. 2008;21(5):2345-2378.
7. Jasra A, Stephens DA, Doucet A, Tsagaris T. Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand J Stat*. 2011;38(1):1-22.
8. Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. *J Royal Stat Soc Series B (Stat Method)*. 2006;68(3):411-436.
9. Cox JC, Ingersoll JE Jr., Ross SA. A theory of the term structure of interest rates. *Econometrica*. 1985;53(2):385-407.
10. Gourieroux C, Jasiak J. Autoregressive gamma processes. *J Forecasting*. 2006;25(2):129-152.
11. Madan DB, Carr P, Chang EC. The variance gamma process and option pricing. *Eur Finance Rev*. 1998;2:79-105.
12. Jacquier E, Polson NG, Rossi P. Stochastic volatility: univariate and multivariate extensions. *Rodney L. White Center for Financial Research Working Papers 19-95*, Wharton School Rodney L. White Center for Financial Research, 1995.
13. Gilks WR, Best NG, Tan KKC. Adaptive rejection Metropolis sampling within Gibbs sampling. *Appl Stat*. 1995;44(4):455-472.
14. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *J Royal Stat Soc Series C (Appl Stat)*. 1992;41(2):337-348.
15. Pitt M, Shephard N. Filtering via simulation: auxiliary particle filters. *J Am Stat Assoc*. 1999;94(446):590-599.
16. Storvik G. Particle filters for state space models with the presence of static parameters. *IEEE Trans Signal Proc*. 2002;50:281-289.
17. Leydold J, Hörmann W. Generating generalized inverse Gaussian random variates by fast inversion. *Comput Stat Data Anal*. 2011;55(1):213-217.
18. Dagpunar JS. An easily implemented generalised inverse Gaussian generator. *Commun Stat - Simul Comput*. 1989;18(2):703-710.
19. West M. Bayesian model monitoring. *J Royal Stat Soc (Series B)*. 1986;48(1):70-78.
20. Lopes HF, Polson NG. 2010. Extracting SP500 and NASDAQ volatility: the credit crisis of 2007-2008. In: *The Oxford Handbook of Applied Bayesian Analysis*; Oxford University Press, Oxford, England 319-342.
21. Jensen M, Maheu J. Bayesian semiparametric stochastic volatility modeling. *J Econometrics*. 2010;157:306-316.
22. Geweke J. 1991. Efficient simulation from the multivariate Normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*; Interface Foundation of North America, Fairfax Station, VA, USA 571-578.
23. Geweke J, Zhou G. Measuring the price of the arbitrage pricing theory. *The Rev Financial Stud*. 1996;9(2):557-587.
24. Hörmann W, Leydold J. Generating generalized inverse Gaussian random variates. *Stat Comput*. 2013;24(4):1-11.
25. Borzadaran GRM, Borzadaran HAM. Log-concavity property for some well-known distributions. *Surv Math Appl*. 2011;6:203-219.

How to cite this article: Warty S, Lopes H, Polson N. Sequential bayesian learning for stochastic volatility with variance-gamma jumps in returns. *Appl Stochastic Models Bus Ind*. 2018;34:460–479. https://doi.org/10.1002/asmb.2258

TECHNICAL APPENDIX ??

A.1 Prior specification

The parametrizations of standard statistical distributions used here are as follows. The Gaussian distribution $\mathcal{N}\left(x;\mu,\sigma^2\right)$ has mean μ and variance σ^2. The Gamma distribution $\mathcal{G}(x;\alpha,\beta)$ has mean $\alpha\beta$ and variance $\alpha\beta^2$. The inverse gamma distribution $\mathcal{IG}(x;\alpha,\beta)$ has mean $\frac{1}{(\alpha-1)\beta}$ for $\alpha > 1$ and variance $\frac{1}{(\alpha-1)^2(\alpha-2)\beta^2}$ for $\alpha > 2$.

The prior structure used in this study is given by the following set of distributions:

$$\begin{aligned}
\pi(\mu) &\sim \mathcal{N}(\mu;0,1)\\
\pi(\kappa) &\sim \mathcal{N}(\kappa;0,0.25)\ \mathbf{1}_{(0,\infty)}(\kappa),\\
\pi(\theta) &\sim \mathcal{N}(\theta;0,1)\ \mathbf{1}_{(0,\infty)}(\theta),\\
\pi(\gamma) &\sim \mathcal{N}(\gamma;0,1),\\
\pi(\sigma_v,\rho) = \pi(w_v)\pi(\phi_v|w_v) &\sim \mathcal{IG}(w_v;2,200)\ \mathcal{N}(\phi_v;0,w_v/2),\\
\pi(\sigma_J^2) &\sim \mathcal{IG}(\sigma_J^2;2.5,5),\\
\pi(\nu) &\sim \mathcal{IG}(\nu;10,0.1),\\
\pi(v_0) &\propto \mathbf{1}_{(0,\infty)}(v_0),
\end{aligned}$$

where $(\phi_v,w_v) = g(\rho,\sigma_v) = \left(\sigma_v\rho,\sigma_v^2(1-\rho^2)\right)$ and $\mathbf{1}_A(x)$ is the indicator function equal to 1 when $x\in A$ and 0 otherwise.

The joint prior and reparametrization of (σ_v, ρ) follows the treatment of the correlated errors stochastic volatility model of Jacquier et al.[12] The inverse mapping is given by $(\rho, \sigma_v) = g^{-1}(\phi_v, w_v) = \left(\frac{\phi_v}{\sqrt{w_v+\phi_v^2}}, w_v + \phi_v^2\right)$.

A.2 Posterior distributions

Let x_t refer to observation of x at time t and x^t the set of all observations of x from time 1 up to and including time t. At time $T > 0$, the posterior distribution for the static parameter vector Θ and the initial latent volatility state v_0 conditional on the full observed and latent information set $\mathcal{F}_T = (r^T, v^T, J^T, G^T)$ is

$$\begin{aligned}
p(\Theta, v_0|\mathcal{F}_T) &\propto p(\Theta, v_0, r^T, v^T, J^T, G^T)\\
&= p(r^T, v^T|J^T, \Theta, v_0)p(J^T|\Theta, G^T)\\
&\quad \times p(G^T|\Theta)\pi(v_0)\pi(\Theta)\\
&\propto \exp\Big(-T\log(\sigma_v\Delta(1-\rho^2))\\
&\quad - \sum_{t=1}^{T}(A_t/2 + \log v_{t-1})\Big)\\
&\quad \times \exp\Big(-T\log\sigma_J - \frac{1}{2}\\
&\quad \times \sum_{t=1}^{T}\left(\log G_t + \frac{J_t^2}{\sigma_J^2 G_t} - \frac{2\gamma J_t}{\sigma_J^2} + \frac{\gamma^2 G_t}{\sigma_J^2}\right)\Big)\\
&\quad \times \exp\Big(-T\log\left(\nu^{\frac{\Delta}{\nu}}\Gamma\left(\frac{\Delta}{\nu}\right)\right)\\
&\quad \times \sum_{t=1}^{T}\left(\left(\frac{\Delta}{\nu} - 1\right)\log G_t - \frac{G_t}{\nu}\right)\Big)\\
&\quad \times \pi(\mu)\ \pi(\kappa)\ \pi(\theta)\ \pi(\sigma_v, \rho)\ \pi(\gamma)\ \pi(\sigma_J)\ \pi(\nu),
\end{aligned}$$

where

$$\begin{aligned}
A_t &= \left(1-\rho^2\right)^{-1}\left(\left(\varepsilon_t^r\right)^2 - 2\rho\varepsilon_t^r\varepsilon_t^v + \left(\varepsilon_t^v\right)^2\right),\\
\varepsilon_t^r &= \frac{r_t - \mu\Delta - J_t}{\sqrt{v_{t-1}\Delta}},\\
\varepsilon_t^v &= \frac{v_t - v_{t-1} - \kappa(\theta - v_{t-1})\Delta}{\sigma_v\sqrt{v_{t-1}\Delta}}.
\end{aligned}$$

A.2.1 Conditional posterior for μ

The posterior of μ at time T is Gaussian:

$$\begin{aligned}
p(\mu|\theta, \sigma_v, \rho, \mathcal{F}_T) &\sim \mathcal{N}\left(\mu; \frac{\mathcal{S}_\mu}{\mathcal{W}_\mu}, \frac{1}{\mathcal{W}_\mu}\right),\\
\mathcal{W}_\mu &= \frac{\Delta}{1-\rho^2}\sum_{t=1}^{T}\frac{1}{v_{t-1}} + \frac{1}{M^2},\\
\mathcal{S}_\mu &= \frac{1}{1-\rho^2}\sum_{t=1}^{T}\frac{1}{v_{t-1}}\left(C_t - \rho\frac{D_t}{\sigma_v}\right) + \frac{m}{M^2},\\
C_t &= r_t - J_t,\\
D_t &= v_t - v_{t-1} - \kappa(\theta - v_{t-1})\Delta,
\end{aligned}$$

where m and M are the hyperparameters for the prior of μ.

A.2.2 Conditional posterior for θ

The posterior of θ at time T is truncated Gaussian on the nonnegative real line:

$$p(\theta|\mu,\kappa,\sigma_v,\rho,\mathcal{F}_T) \sim \mathcal{N}\left(\theta; \frac{\mathcal{S}_\theta}{\mathcal{W}_\theta}, \frac{1}{\mathcal{W}_\theta}\right) \mathbf{1}_{(0,\infty)}(\theta),$$

$$\mathcal{W}_\theta = \frac{\kappa^2\Delta}{\sigma_v^2(1-\rho^2)} \sum_{t=1}^{T} \frac{1}{v_{t-1}} + \frac{1}{M^2},$$

$$\mathcal{S}_\theta = \frac{\kappa}{\sigma_v(1-\rho^2)} \sum_{t=1}^{T} \frac{1}{v_{t-1}} \times \left(\frac{D_t}{\sigma_v} - \rho C_t\right) + \frac{m}{M^2},$$

$$C_t = r_t - \mu\Delta - J_t,$$

$$D_t = v_t - v_{t-1} + \kappa v_{t-1}\Delta,$$

where m and M are the hyperparameters for the prior of θ. Truncated Gaussian sampling of θ and κ uses the normal-exponential rejection sampling algorithm of Geweke [22] and Geweke and Zhou.[23] This algorithm is more efficient than simple rejection sampling with a Gaussian proposal. It also avoids potential arithmetic underflow problems of random uniform quantile inversion methods when the mean of the left-truncated distribution is far to the left of the truncation point.

A.2.3 Conditional posterior for κ

The posterior of κ at time T is

$$p(\kappa|\mu,\theta,\sigma_v,\rho,\mathcal{F}_T) \sim \mathcal{N}\left(\frac{\mathcal{S}_\kappa}{\mathcal{W}_\kappa}, \frac{1}{\mathcal{W}_\kappa}\right) \mathbf{1}_{(0,\infty)}(\kappa),$$

$$\mathcal{W}_\kappa = \frac{\Delta}{\sigma_v^2(1-\rho^2)} \sum_{t=1}^{T} \frac{1}{v_{t-1}}(\theta - v_{t-1})^2 + \frac{1}{M^2},$$

$$\mathcal{S}_\kappa = \frac{1}{\sigma_v(1-\rho^2)} \sum_{t=1}^{T} \frac{1}{v_{t-1}}(\theta - v_{t-1}) \times \left(\frac{D_t}{\sigma_v} - \rho C_t\right) + \frac{m}{M^2},$$

$$C_t = r_t - \mu\Delta - J_t,$$

$$D_t = v_t - v_{t-1},$$

where m and M are the hyperparameters for the prior of κ.

A.2.4 Conditional posterior for σ_v and ρ

After reparametrizing to $(\phi_v, w_v) = (\sigma_v\rho, \sigma_v^2(1-\rho^2))$, the posterior of (ϕ_v, w_v) at time T is

$$
\begin{aligned}
p(\phi_v|w_v,\mu,\theta,\kappa,\mathcal{F}_T) &\sim \mathcal{N}\left(\frac{\mathcal{S}_v}{\mathcal{W}_v},\frac{w_v}{\mathcal{W}_v}\right),\\
p(w_v|\mu,\theta,\kappa,\mathcal{F}_T) &\sim \mathcal{IG}\left(\frac{T}{2}+m,\left(\frac{1}{2}\sum_{t=1}^{T}D_t^2\ +\frac{1}{M}-\frac{\mathcal{S}_v^2}{2\mathcal{W}_v}\right)^{-1}\right),\\
\mathcal{W}_v &= 2+\sum_{t=1}^{T}C_t^2,\\
\mathcal{S}_v &= \sum_{t=1}^{T}C_tD_t,\\
C_t &= \frac{1}{\sqrt{v_{t-1}\Delta}}(r_t-\mu\Delta-J_t),\\
D_t &= \frac{1}{\sqrt{v_{t-1}\Delta}}(v_t-v_{t-1}-\kappa(\theta-v_{t-1})\Delta),
\end{aligned}
$$

where m and M are the hyperparameters for the prior of w_v.

A.2.5 Conditional posterior for γ

The posterior of γ at time T is

$$
\begin{aligned}
p(\gamma|\sigma_J,\mathcal{F}_T) &\sim \mathcal{N}\left(\frac{\mathcal{S}_\gamma}{\mathcal{W}_\gamma},\frac{1}{\mathcal{W}_\gamma}\right),\\
\mathcal{W}_\gamma &= \frac{1}{\sigma_J^2}\sum_{t=1}^{T}G_t+\frac{1}{M^2},\\
\mathcal{S}_\gamma &= \frac{1}{\sigma_J^2}\sum_{t=1}^{T}J_t+\frac{m}{M^2},
\end{aligned}
$$

where m and M are the hyperparameters for the prior of γ.

A.2.6 Conditional posterior for σ_J

The posterior of σ_J^2 at time T is

$$
p\left(\sigma_J^2|\gamma,\mathcal{F}_T\right) \sim \mathcal{IG}\left(\frac{T}{2}+m,\left(\frac{1}{2}\sum_{t=1}^{T}\frac{(J_t-\gamma G_t)^2}{G_t}+\frac{1}{M}\right)^{-1}\right),
$$

where m and M are the hyperparameters for the prior of σ_J.

A.2.7 Conditional posterior for ν

The posterior of ν at time T is

$$
\begin{aligned}
p(\nu|\cdot) \propto &\left(\frac{1}{\nu^{\frac{\Delta}{\nu}}\Gamma\left(\frac{\Delta}{\nu}\right)}\right)^T\left(\prod_{t=1}^{T}G_t\right)^{\frac{\Delta}{\nu}}\\
&\times\exp\left(-\frac{1}{\nu}\left(\sum_{t=1}^{T}G_t+\frac{1}{M}\right)\right)\left(\frac{1}{\nu}\right)^{m+1},
\end{aligned}
$$

where m and M are the hyperparameters for the prior of ν.

Sampling of this posterior is complicated by non–log-concavity. In the Gibbs sampling step of the sequential algorithm, this parameter is sampled via an adaptive rejection Metropolis sampling (ARMS) step with convexity correction parameter set to 2.0. We also investigated sampling v on a discrete grid with weights proportional to the conditional posterior. However, this discrete sampling scheme is very inefficient and requires a large number of evaluations of the log density function. Further, the ARMS approach has yielded far superior results in the applications considered to date.

A.2.8 Conditional posterior for v_0

The posterior for v_0 at time $T = 1$ is

$$\begin{aligned}
p(v_0|\cdot) &\propto \exp\left(-D_1/2 - \log v_0\right), \\
D_1 &= \left(1-\rho^2\right)^{-1}\left((\varepsilon_1^r)^2 - 2\rho\varepsilon_1^r\varepsilon_1^v + (\varepsilon_1^v)^2\right), \\
\varepsilon_1^r &= \frac{r_1 - \mu\Delta - J_1}{\sqrt{v_0\Delta}}, \\
\varepsilon_1^v &= \frac{v_1 - v_0 - \kappa(\theta - v_0)\Delta}{\sigma_v\sqrt{v_0\Delta}}.
\end{aligned}$$

A.2.9 Conditional posterior for v_T

For $T > 0$, the posterior for v_T at time T is

$$\begin{aligned}
p(v_T|\mu,\theta,\kappa,\sigma_v,r_T,J_T,v_{T-1}) &\sim \mathcal{N}(v_T;\mathcal{S}_v,\mathcal{W}_v)\mathbf{1}_{(0,\infty)}(v_t), \\
\mathcal{S}_v &= v_{T-1} + \kappa(\theta - v_{T-1})\Delta \\
&\quad + \rho\sigma_v(r_T - \mu\Delta - J_T), \\
\mathcal{W}_v &= (1-\rho^2)\sigma_v^2 v_{T-1}\Delta.
\end{aligned}$$

The prior structures for κ, θ, and σ_v^2 help to ensure the nonnegativity and stationarity of the variance process. However, when the ex-jump excess return $(r_T - \mu\Delta - J_T)$ is very large, there is a greater probability that the sampler generates negative values for v_t(ρ tends to be negative empirically). To guard against this case, propagation of v_t is truncated Gaussian.

A.2.10 Conditional posterior for J_T

For $T > 0$, the posterior of J_T at time T is

$$\begin{aligned}
p(J_T|\mu,\theta,\kappa,\sigma_v,\rho,\sigma_J,r_T,v_T,v_{T-1},G_T) &\sim \mathcal{N}\left(\frac{\mathcal{S}_J}{\mathcal{W}_J},\frac{1}{\mathcal{W}_J}\right), \\
\mathcal{W}_J &= \frac{1}{v_{T-1}\Delta(1-\rho^2)} + \frac{1}{\sigma_J^2 G_T}, \\
\mathcal{S}_J &= \frac{1}{v_{T-1}\Delta(1-\rho^2)}\left(C_T - \frac{\rho D_T}{\sigma_v}\right) + \frac{\gamma}{\sigma_J^2}, \\
C_T &= r_T - \mu\Delta, \\
D_T &= v_T - v_{T-1} - \kappa(\theta - v_{T-1})\Delta.
\end{aligned}$$

At present, this conditional distribution is not sampled as new particles for J_t are obtained via blind propagation from the marginal $p(J_t|\Theta)$.

A.2.11 Conditional posterior for G_T

For $T > 0$, the posterior of G_T at time T is generalized inverse gaussian (GIG):

$$p(G_T|\nu,\gamma,\sigma_J,J_T) \propto \mathcal{GIG}\Bigg(G_T;\chi=\frac{J_T^2}{\sigma_J^2},\psi=\frac{\gamma^2}{\sigma_J^2} + \frac{2}{\nu},\lambda=\frac{\Delta}{\nu}-0.5\Bigg).$$

Draws of G_t can be obtained, for example, via ARMS or GIG sampling algorithms such as those developed by Dagpunar[18] or Leydold and Hörmann[17] and Hörmann and Leydold.[24] Both methods produce similar results. However, ARMS can have a considerably higher computation cost than GIG samplers when the region from which proposals are drawn is large. In this application, the GIG samplers tend to run 30% faster than ARMS. The efficient algorithm of Leydold and Hörmann[17] and Hörmann and Leydold[24] is used here. The authors document deterioration in the performance of Dagpunar's method when $\lambda < 1$ and $\sqrt{\chi\psi}$ are close to 0, which is often the case in this application. Further, the GIG distribution is potentially log-concave over certain regions of the parameter space,[25] which would increase the overhead of ARMS because the Metropolis step would be unnecessary.

A.3 Sequential learning algorithm

A.3.1 Propagate latent state J_t

As the stochastic volatility with variance-gamma (SVVG) likelihood is not available in closed form, the sequential algorithm presented here relies on proposal draws of J_t to evaluate the conditional likelihood $p(r_t|J_t,\Theta)$ in the resampling step. The set of proposal draws $\left\{\tilde{J}_t^{(i)}\right\}_{i=1}^{M}$ is obtained from the marginal distribution of J_t. The propagation proposal density is given by Madan et al[11, theorem 1]:

$$\begin{aligned} p\left(\tilde{J}_t|(\gamma,\sigma_J^2,\nu)^{(i)}\right) &= \int_0^{\infty} p\left(\tilde{J}_t|G_t,\gamma^{(i)},\left(\sigma_J^{(i)}\right)^2\right)p(G_t|\nu^{(i)})dG_t \\ &\propto \exp\left(\gamma\sigma_J^{-2}\tilde{J}_t\right)\left(\tilde{J}_t^2\right)^{\frac{\Delta}{2\nu}-\frac{1}{4}} \\ &\quad\times K_{\frac{\Delta}{\nu}-\frac{1}{2}}\left(\sigma_J^{-2}\sqrt{\tilde{J}_t^2\left(\frac{2\sigma_J^2}{\nu}+\gamma^2\right)}\right), \end{aligned}$$

where $p(G_t|\nu)$ is the gamma distribution density function $\mathcal{G}(G_t;\Delta/\nu,\nu)$ and $K_\alpha(\cdot)$ is the modified Bessel function of the second kind with order α.

The marginal density of J_t is easily sampled in a 2-step procedure using the conditional Gaussian-gamma property of the subordinated Brownian motion given at Equations 6 and 7. When evaluating the likelihood, the intermediate gamma draws can be discarded because r_t is conditionally independent of G_t when J_t is known.

A.3.2 Resample particles after observing r_t

Propagation of J_t is blind in that it does not condition on the current observation r_t. In this implementation, the proposal density and marginal density of J_t are identical. Therefore, resampling weights equal to the conditional likelihood $p(r_t|J_t,\Theta)$ require no adjustment. The set of particles $\left\{(J_t,x_{t-1},\Theta,z_{t-1})^{(i)}\right\}_{i=1}^{M}$ is reweighted to reflect the likelihood of parameters given $\left(r_t,J_t^{(i)}\right)$. Resampling weights are given by

$$\begin{aligned} w_t^{(i)} &= p\left(r_t|(v_{t-1},J_t,\mu)^{(i)}\right) \\ &\propto \exp\left(-\frac{\left(r_t-\mu^{(i)}\Delta-J_t^{(i)}\right)^2}{2v_{t-1}^{(i)}\Delta}\right). \end{aligned}$$

This method requires no evaluations of the marginal density of J_t given at A.3.1, although this may not be the case when using an alternative proposal density.

A.3.3 Propagate latent states v_t and G_t

The latent states v_t and G_t are propagated with draws from their respective conditional posterior distributions:

$$p\left(v_t^{(i)}|r_t,(v_{t-1},J_t,\Theta)^{(i)}\right)\sim\mathcal{N}\left(v_t^{(i)};\mathcal{S}_v^{(i)},\mathcal{W}_v^{(i)}\right)\mathbf{1}_{(0,\infty)}(v_t),$$
$$\mathcal{S}_v^{(i)}=v_{t-1}^{(i)}+\kappa^{(i)}\left(\theta^{(i)}-v_{t-1}^{(i)}\right)\Delta$$
$$+\rho^{(i)}\sigma_v^{(i)}\left(r_t-\mu^{(i)}\Delta-J_t^{(i)}\right),$$
$$\mathcal{W}_v^{(i)}=\left(1-\left(\rho^{(i)}\right)^2\right)\left(\sigma_v^2\right)^{(i)}v_{t-1}^{(i)}\Delta,$$

$$p\left(G_t^{(i)}|\left(J_t,\gamma,\sigma_J^2,\nu\right)^{(i)}\right)\sim\mathcal{GIG}\left(G_t^{(i)};\left(\frac{J_t^{(i)}}{\sigma_J^{(i)}}\right)^2,\left(\frac{\gamma^{(i)}}{\sigma_J^{(i)}}\right)^2\right.$$
$$\left.+\frac{2}{\nu^{(i)}},\frac{\Delta}{\nu^{(i)}}-0.5\right).$$

Details of obtaining draws of G_t are given at A.2.11.

A.3.4 Update conditional sufficient statistics vector z_t

The form of the posterior suggests 14 parameter- and state-conditional sufficient statistics. For each particle, the vector z_t is updated as follows:

$$z_t^{(i)}=\mathcal{Z}(z_{t-1}^{(i)},x_t^{(i)},r_t)=z_{t-1}^{(i)}$$
$$+\left(1,r_t,v_{t-1},v_t,\frac{1}{v_{t-1}},\frac{v_t}{v_{t-1}},\frac{v_t^2}{v_{t-1}},J_t,G_t,\log G_t,\right.$$
$$\left.\frac{r_t-J_t}{v_{t-1}},\frac{(r_t-J_t)v_t}{v_{t-1}},\frac{(r_t-J_t)^2}{v_{t-1}},\frac{J_t^2}{G_t}\right)^{(i)}.$$

A.3.5 Sample static parameter vector Θ

Sampling of static parameters comprises Gibbs sampling iterating over the elements of Θ. Each element of the static parameter vector Θ is drawn in turn from its respective conditional posterior given $z_t^{(i)}$. Full-conditional posterior distributions are given in Appendix A.2.

A.4 SVVG with volatility in mean

Here, we consider the extended SVVG model with conditional mean equation as specified at Equation 8 and volatility-in-mean conditional mean function given by

$$f(v_{t-1})=\mu'+\beta_v v_{t-1}.$$

Posterior distributions for μ' and β_v are provided below. We assume Gaussian priors for μ' and β_v to obtain posterior conjugacy. Inference for the latent states and the remaining static parameters follows the methodology presented in Appendix A.2.

A.4.1 Conditional posterior for μ' for volatility in mean

The posterior of μ' at time T is Gaussian:

$$
\begin{aligned}
p(\mu' | \beta_v, \theta, \sigma_v, \rho, \mathcal{F}_T) \sim \mathcal{N}\left(\mu'; \frac{\mathcal{S}_{\mu'}}{\mathcal{W}_{\mu'}}, \frac{1}{\mathcal{W}_{\mu'}}\right), \\
\mathcal{W}_{\mu'} = \frac{\Delta}{1-\rho^2} \sum_{t=1}^{T} \frac{1}{v_{t-1}} + \frac{1}{M^2}, \\
\mathcal{S}_{\mu'} = \frac{1}{1-\rho^2} \sum_{t=1}^{T} \frac{1}{v_{t-1}} \left(C_t - \rho \frac{D_t}{\sigma_v}\right) + \frac{m}{M^2}, \\
C_t = r_t - J_t - \beta_v v_{t-1} \Delta, \\
D_t = v_t - v_{t-1} - \kappa(\theta - v_{t-1})\Delta,
\end{aligned}
$$

where m and M are the hyperparameters for the prior of μ'.

A.4.2 Conditional posterior for β_v for volatility in mean

The posterior of β_v at time T is Gaussian:

$$
\begin{aligned}
p(\beta_v | \mu, \theta, \sigma_v, \rho, \mathcal{F}_T) \sim \mathcal{N}\left(\beta_v; \frac{\mathcal{S}_{\beta_v}}{\mathcal{W}_{\beta_v}}, \frac{1}{\mathcal{W}_{\beta_v}}\right), \\
\mathcal{W}_{\beta_v} = \frac{\Delta}{1-\rho^2} \sum_{t=1}^{T} v_{t-1} + \frac{1}{M^2}, \\
\mathcal{S}_{\beta_v} = \frac{1}{1-\rho^2} \sum_{t=1}^{T} \left(C_t - \rho \frac{D_t}{\sigma_v}\right) + \frac{m}{M^2}, \\
C_t = r_t - J_t - \mu' \Delta \\
D_t = v_t - v_{t-1} - \kappa(\theta - v_{t-1})\Delta,
\end{aligned}
$$

where m and M are the hyperparameters for the prior of β_v.