Analysis of filtering and smoothing algorithms for Lévy-driven stochastic volatility models

Drew D. Creal

Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Abstract

Filtering and smoothing algorithms that estimate the integrated variance in Lévy-driven stochastic volatility models are analyzed. Particle filters are algorithms designed for nonlinear, non-Gaussian models while the Kalman filter remains the best linear predictor if the model is linear but non-Gaussian. Monte Carlo experiments are performed to compare these algorithms across different specifications of the model including different marginal distributions and degrees of persistance for the instantaneous variance. The use of realized variance as an observed variable in the state space model is also evaluated. Finally, the particle filter's ability to identify the timing and size of jumps is assessed relative to popular nonparametric estimators.

Key words: Lévy process, stochastic volatility, particle filter, Kalman filter

1. Introduction

This paper contributes to the literature on parametric modeling of volatility by comparing various filtering and smoothing algorithms for the non-Gaussian Ornstein-Uhlenbeck (OU) stochastic volatility models introduced by Barndorff-Nielsen and Shephard (2001a) and extended in Barndorff-Nielsen and Shephard (2003). The instantaneous variance in these models is driven by a non-Gaussian Lévy process, which requires approximate recursions for filtering and smoothing. I examine how well the particle filter compares to the Kalman filter at estimating the integrated variance under a number of different specifications of the model. I also consider estimation of jumps by the particle filter and smoother. The goal of the article is to build on some of the comments in the original

Email address: dcreal@feweb.vu.nl (Drew D. Creal).

[1] Tel: +31 20 598 60 10; Fax: +31 20 598 60 20

paper. The filtering algorithms considered here are applicable regardless of the method used to estimate the model's parameters. Parameters for the BNS-SV model have been estimated using quasi-maximum likelihood in Barndorff-Nielsen and Shephard (2002) and Barndorff-Nielsen et al. (2004); the smooth particle filter of Pitt (2001); Markov chain Monte Carlo (MCMC) in Roberts et al. (2004), Gander and Stephens (2005), and Griffin and Steel (2006); minimum distance estimation in Todorov (2006); and sequential Monte Carlo samplers in Del Moral et al. (2006). MCMC can also be used to compute smoothed estimates, but that is beyond the scope of this article.

When there are no jumps, both filtering algorithms estimate increments of the integrated variance almost exactly, even at moderate intra-daily frequencies. When returns are sampled at low frequencies, the Kalman filter's performance is more robust than the particle filter. These differences can be large in practice depending on the marginal distribution of the instantaneous variance. If returns are available at higher frequencies, the particle filter offers a moderate improvement over the Kalman filter. The size of the improvement also depends on the marginal distribution of the instantaneous variance. The particle filter struggles when the variance of the stationary distribution is larger. Small misspecifications of the marginal distribution can potentially lead to poorer estimates from the particle filter. These differences can be large enough to make the Kalman filter and smoother competitive at estimating the integrated variance at any frequency with at which returns are sampled.

Finally, I also compare the particle filter and smoother's ability to estimate jumps versus the jump statistics developed in Barndorff-Nielsen and Shephard (2004). Results indicate that the particle filter identifies the jump times and sizes well.

2. Lévy-driven stochastic volatility models

2.1. *Price processes with stochastic volatility and jumps*

I assume that the log-price process $y^*(t)$ takes the form

$$\mathrm{d}y^*(t) = \left(\mu + \beta\sigma^2(t)\right)\mathrm{d}t + \sigma(t)\,\mathrm{d}B(t) + \rho\mathrm{d}\bar{z}(\lambda t), \tag{1}$$

where μ is the drift and β is the risk premium. The instantaneous variance $\sigma^2(t)$ is assumed to be independent of the standard Brownian motion $B(t)$. Following Barndorff-Nielsen and Shephard (2001a), the instantaneous variance is specified as a non-Gaussian OU process

$$\mathrm{d}\sigma^2(t) = -\lambda\sigma^2(t)\,\mathrm{d}t + \mathrm{d}z(\lambda t), \tag{2}$$

where $z(t)$ is a Lévy process called the background driving Lévy process. The process $z(t)$ is also known as a subordinator. The timing of the subordinator ($\mathrm{d}z(\lambda t)$ versus $\mathrm{d}z(t)$) is intentionally designed to separate the marginal distribution of the instantaneous variance from its autocovariance function. Lévy processes are a general group of continuous time stochastic processes that have independent and stationary increments, see e.g. Cont and Tankov (2004). The compensated version of the subordinator, $\bar{z}(t) = z(t) - E\{z(t)\}$, in (1) accounts for the leverage effect with ρ determining the direction of the jumps in the price.

Regardless of the model for the instantaneous variance, the integrated variance is defined as

$$\sigma^{2*}(t) = \int_0^t \sigma^2(u)\,\mathrm{d}u.$$

Defining $\hbar$ to be the length of time between two periods of interest (e.g. 1 day), increments of the integrated variance

$$\sigma_n^2 = \sigma^{2*}(\hbar n) - \sigma^{2*}(\hbar(n-1)),$$

are known as the actual variance and they are a critical ingredient in option pricing and risk management. For this model, the actual variance has a simple structure

$$\sigma_n^2 = \lambda^{-1}\left\{z(\lambda\hbar n) - \sigma^2(\hbar n) - z(\lambda\hbar(n-1)) + \sigma^2(\hbar(n-1))\right\}. \tag{3}$$

It can be calculated by simulating the continuous time process

$$\begin{bmatrix} \sigma^2(\hbar n) \\ z(\lambda\hbar n) \end{bmatrix} = \begin{bmatrix} \exp(-\lambda\hbar)\,\sigma^2(\hbar(n-1)) \\ z(\lambda\hbar(n-1)) \end{bmatrix} + \eta_n, \tag{4}$$

where

$$\eta_n = \begin{bmatrix} \exp(-\lambda\hbar)\int_0^{\hbar}\exp(\lambda s)\,\mathrm{d}z(\lambda s) \\ \int_0^{\hbar}\mathrm{d}z(\lambda s) \end{bmatrix}. \tag{5}$$

There are numerous methods for simulating Lévy processes, see e.g. Rosínski (2001b) and Cont and Tankov (2004). In this paper, I use the series representation from Barndorff-Nielsen and Shephard (2001a) to simulate $\sigma^2(t)$ when it has gamma marginal. For the instantaneous variance with tempered stable marginal, I use the series representation developed by Rosínski (2001a) and described in Barndorff-Nielsen and Shephard (2001b).

Aggregate returns over the nth day are defined as

$$y_n = y^*(\hbar n) - y^*(\hbar(n-1)).$$

Conditional on the actual variance and jumps, returns are normally distributed

$$p\left(y_n|\sigma_n^2, z_n\right) = N\left(y_n|\mu\hbar + \beta\sigma_n^2 + \rho z_n\right), \tag{6}$$

with $z_n = \int_{\hbar(n-1)}^{\hbar n}\mathrm{d}z(\lambda t) - \hbar\lambda E\{z(1)\}$. The OU model in (2) unfortunately has limited flexibility because its autocovariance function $r(s) = \exp(-\lambda s)$ decays at a constant rate. Barndorff-Nielsen and Shephard (2001a) suggested using superpositions of independent Lévy processes. Superpositions amount to combining independent processes $\left\{\sigma_i^2(t)\right\}_{i=1}^m$ and calculating the conditional variance as a weighted sum $\sigma^2(t) = \sum_{i=1}^m w^i\sigma_i^2(t)$ where

the weights sum to one. The autocovariance function is now more flexible, see Barndorff-Nielsen and Shephard (2001a) and Griffin and Steel (2006) for more details on superpositions.

In Section 4, I evaluate how well filtering and smoothing algorithms estimate the actual variance and jumps for this model under a number of different specifications. To better evaluate the performance of the algorithms, it is helpful to compare them to several popular nonparametric methods known as realized variance and realized bipower variation. Barndorff-Nielsen and Shephard (2007) provide a detailed survey of this literature. Briefly, the realized variance process is defined as

$$\{y_M^*\}_n^{[2]} = \sum_{j=1}^{M} y_{j,n}^2, \tag{7}$$

where $y_{j,n}$ is the jth intra-$\hbar$ return on the nth day. This makes for a total of M equally-spaced returns in each day. Under the assumptions on the price process given above, realized variance converges to the quadratic variation process. Increments of the quadratic variation for a process with continuous sample paths (i.e. $\rho = 0$ in (1)) are equal to increments of the integrated variance.

If there are jumps in the price process $y^*(t)$, the overall quadratic variation will be the sum of the quadratic variation from the continuous and discontinuous components. Realized variance remains a consistent estimator of quadratic variation in models with jumps, but it is no longer a consistent estimator of the integrated variance. Realized variance cannot separately identify the quadratic variation coming from the discontinuous and continuous components. Barndorff-Nielsen and Shephard (2004) introduced the concept of realized bipower variation, which they proved converges to the integrated variance in the presence of finite activity jumps. The realized bipower variation process of order $[r, s]$ is defined as

$$\{y_M^*\}_n^{[r,s]} = \left\{ \left(\frac{\hbar}{M} \right)^{1-(r+s)/2} \right\} \sum_{j=1}^{M-1} |y_{j,n}|^r \, |y_{j+1,n}|^s \quad r, s \geq 0.$$

The difference between realized variance and realized bipower variation provides an estimate of the discontinuous component of quadratic variation. Barndorff-Nielsen and Shephard (2004) suggested using

$$\max \left\{ \{y_M\}_n^{[2]} - \mu_1 \{y_M\}_n^{[1,1]}, 0 \right\}, \quad \mu_r = 2^{r/2} \frac{\Gamma\left(\frac{1}{2}(r+1)\right)}{\Gamma\left(\frac{1}{2}\right)}, \tag{8}$$

which I will compare with estimates from the particle filter and smoother.

2.2. *Marginal distributions for the instantaneous variance*

In this paper, I consider different marginals for $\sigma^2(t)$ in order to compare the performance of both filters across distributions. The particle filter uses knowledge of the entire marginal distribution whereas the Kalman filter does not. The models I consider have gamma and tempered stable marginal distributions. In this paper, the gamma distribution $Ga(\nu, \alpha)$ has mean $\xi = \nu/\alpha$ and variance $\omega^2 = \nu/\alpha^2$. The tempered stable

distribution $TS(\kappa, \delta, \gamma)$ is a special case of the modified stable distribution, and is discussed in detail in Barndorff-Nielsen and Shephard (2001b). An important special case of the tempered stable is the inverse Gaussian distribution $IG(\delta, \gamma)$ when $\kappa = 0.5$. Nicolato and Venardos (2003) develop option pricing formulas for the BNS-SV model and demonstrate how to price options when $\sigma^2(t)$ has an IG or gamma marginal.

3. Filtering and smoothing

3.1. *State space forms*

The SV model given by (1) and (2) may be placed in state space form and the actual variance estimated by filtering and smoothing algorithms. General nonlinear, non-Gaussian state space models do not admit closed-form solutions for exact filtering. The particle and Kalman filters both provide approximations. When applied directly to returns, the state space form for the particle filter has measurement density given by (6) and the transition density for the models with jumps is $p\left(\sigma_n^2, z(\lambda\hbar n) | \sigma_{n-1}^2, z(\lambda\hbar(n-1))\right)$, where the jump size ρz_n is a function of $z(\lambda\hbar n)$ and $z(\lambda\hbar(n-1))$. This density is generally unknown for this class of models but σ_n^2 and z_n can still be simulated recursively using (4) and (5) and then by applying (3). To run the Kalman filter on returns, the model can be placed in linear state space form as described in Barndorff-Nielsen and Shephard (2001a).

If a researcher is only interested in estimating volatility at a frequency $\hbar$, an alternative strategy proposed by Barndorff-Nielsen and Shephard (2002) is to decompose realized variance from (7) into the actual variance σ_n^2 and a measurement error

$$\{y_M\}_n^{[2]} = \sigma_n^2 + u_n. \tag{9}$$

The presence of u_n is intended to correct for the fact that realized variance has yet to converge. The distribution of u_n will depend upon the assumptions of the price process. Equation (9) is the measurement equation for an alternative state space representation. Barndorff-Nielsen and Shephard (2002) derived a finite sample approximation to the distribution of u_n as

$$var\left(\sqrt{M}u_n\right) \to 2\hbar^2\left(\omega^2 + \xi^2\right), \tag{10}$$

where $\xi = E\{\sigma^2(t)\}$ and $\omega^2 = V\{\sigma^2(t)\}$ are the mean and variance of the stationary distribution. This approximation is valid when the mean process in (1) is zero and is only mildly affected by the addition of a drift term. Using the realized variance approximation is advantageous because it reduces the number of observations with which one needs to work. The errors u_n are assumed to be normally distributed in the measurement equation of the state space representation for the particle filter, i.e. $p\left(\{y_M\}_n^{[2]} | \sigma_n^2\right) \approx N\left(\sigma_n^2, 2\hbar^2 M^{-1}\left(\omega^2 + \xi^2\right)\right)$. Barndorff-Nielsen and Shephard (2002) and Barndorff-Nielsen et al. (2004) describe how to place the realized variance approximation in linear state space form in order to use the Kalman filter. In Section 4, I consider both state space representations.

3.2. *Particle filtering and smoothing*

Particle filters and smoothers are simulation methods for approximating the filtering and smoothing distributions $\sigma_n^2, z_n \mid \mathcal{F}_n$ and $\sigma_n^2, z_n \mid \mathcal{F}_T$. The particle filters applied in this paper are standard and a more detailed coverage of them can be found in the books by Doucet et al. (2001), Ristic et al. (2004), and Cappé et al. (2005). Cappé et al. (2007) is a recent survey of the field.

Rather than representing $\sigma_n^2, z_n \mid \mathcal{F}_n$ analytically, a particle filter represents a continuous density using a collection of N particles. The first component of each particle is a value representing a point on the support of $\sigma_n^2, z_n \mid \mathcal{F}_n$ and the second component of say the ith particle is an importance weight $W_n^{(i)}$ denoting the probability mass at that point. The collection of particles and their weights therefore approximate the distribution by scaling a Dirac mass at each particle's location by that particle's normalized importance weight. As the number of particles increases and $N \to \infty$, the particle filter's approximation should become more accurate. The particle filter estimates the actual variance and jumps at each iteration using the importance weights as in standard importance sampling.

The particle filter computes the importance weights recursively in time. A standard recursion for the particle filter applied to returns is as follows

Step 1: At $n = 0$, initialize N particles $\left\{\sigma^{2(i)}(\lambda\hbar)\right\}_{i=1}^{N}$ from the stationary distribution of $\sigma^2(t)$ and set $W_0^{(i)} = \frac{1}{N}$.

Step 2: For $i = 1, .., N$, draw $\left\{\sigma_n^{2(i)}, z_n^{(i)}\right\}_{i=1}^{N}$ from an importance density

$$q\left(\sigma_n^{2(i)}, z_n^{(i)} \mid y_n, \sigma_{n-1}^2, \sigma^{2(i)}(\lambda\hbar(n-1))\right).$$

Step 3: For $i = 1, .., N$, compute importance weights

$$W_n^{(i)} \propto W_{n-1}^{(i)} \frac{p\left(y_n \mid \sigma_n^{2(i)}, z_n^{(i)}\right) p\left(\sigma_n^{2(i)}, z^{(i)}(\lambda\hbar n) \mid \sigma_{n-1}^{2(i)}, z^{(i)}(\lambda\hbar(n-1))\right)}{q\left(\sigma_n^{2(i)}, z_n^{(i)} \mid y_n, \sigma_{n-1}^2, \sigma^{2(i)}(\lambda\hbar(n-1))\right)}.$$

Step 4: Normalize the importance weights and compute estimates of σ_n^2 and ρz_n.

Step 5: Resample the particles $\left\{\sigma^{2(i)}(\lambda\hbar n)\right\}_{i=1}^{N}$ using the normalized importance weights. Set the weights to $W_n^{(i)} = \frac{1}{N}$ and return to *Step 2*.

In most applications of the BNS-SV model, the transition density is unknown. This will either force the user to implement a particle filter that uses the transition density as the importance density $q(\cdot)$ in order to compute the importance weights. Otherwise, the user must reparameterize the model. Reparameterizing the model is usually more computationally intensive and its feasibility will depend upon the marginal distribution of $\sigma^2(t)$. Roberts et al. (2004) and Griffin and Steel (2006) use reparameterizations in their work developing MCMC algorithms for the model. As explained in Barndorff-Nielsen and Shephard (2001a), draws from the transition density are available from the recursions in (4) and (5). I use this approach below.

In preliminary work, I compared several standard particle filtering algorithms on the BNS-SV model with different marginals for $\sigma^2(t)$. These included the bootstrap filter,

SISR filter, and regularized particle filter. Results indicated that the regularized particle filter was slightly more efficient in terms of mean square error. This particle filter adds another step to the above recursion. After resampling the particles in *Step 5*, the particles are moved according to a normal kernel with an 'optimal' bandwidth, see e.g. Musso et al. (2001) or Ristic et al. (2004) for details. In this application, use of the transition density as the importance density means that particles are only moved upward via a jump or are moved downward at the same rate as the decay parameter λ. Adding the regularization step allows particles to be moved downward beyond the parameter λ. This is relevant when a resampling step causes the majority of the cloud of particles to be replicated above the 'true' actual variance.

Kitagawa (1996) introduced the first particle smoothing algorithm and his method is reviewed in Kitagawa and Sato (2001). This algorithm does not require evaluating the transition density, whereas other particle smoothing algorithms typically do. The algorithm works by resampling both the particles of the current generation and past generations at each iteration of the algorithm for some fixed number of lags. In practice, this smoother requires truncating the lag length, resulting in a fixed-lag particle smoother.

4. Simulation results

This section contains Monte Carlo experiments to compare the performance of the particle filter with the Kalman filter (see, e.g., Durbin and Koopman (2001)) under a number of different scenarios. For simplicity, I assume that the parameters of the models are known. Throughout this section, I define $\hbar = 1$ as one day and use M=1, 12, 72, and 288 which correspond to daily, two-hour, half-hour, and 5-minute returns, respectively. All of the particle filters use the systematic resampling algorithm of Carpenter et al. (1999). The lag length on the fixed-lag particle smoothers was set to $L = 20$.

4.1. *Assessing the realized variance approximation*

This paper mainly compares the particle and Kalman filters using the realized variance approximation due to the considerable computational savings it provides. Prior to this analysis, I investigate the accuracy of the Kalman filter applied to the realized variance approximation relative to when it is applied to high-frequency returns. This can be done by summing the smoothed estimates of actual variance over the day after the Kalman filter and smoother are run on high-frequency returns.

For simplicity, I consider the model without jumps (i.e. $\rho = 0$). I simulated 500 realizations of 500 days from three models with different marginal distributions. Table 1 contains the mean square error (MSE) from the Kalman smoother using both state space representations. The first four columns of Table 1 are computed from a process with gamma marginal whereas columns 5-6 are estimates from a process with IG marginal. Comparing columns 1 and 2 of Table 1 demonstrates that the realized variance approximation only introduces a small error in the estimates. Estimates from the Kalman smoother applied to high-frequency returns (then summed over the day) and estimates from the realized variance approximation are shown in Figure 1 (a) and (b), respectively. These graphs are examples of the process in columns 1-2 of Table 1. The graphs show that the two estimates are close to being indistinguishable.

Table 1
Mean square error from the Kalman smoother applied to both returns and the realized variance approximation when $\sigma^2(t) \sim Ga(\nu, \alpha)$ (1st-4th columns) and $\sigma^2(t) \sim IG(\delta, \gamma)$ (5th-6th columns).†

	$\xi = 0.5$		$\xi = 0.5$		$\xi = 0.5$	
	$\omega^2 = 0.0625$		$\omega^2 = 0.4$		$\omega^2 = 0.15$	
	(1)	(2)	(3)	(4)	(5)	(6)
M	KS-RV	KS-HF	KS-RV	KS-HF	KS-RV	KS-HF
	$\lambda = 0.01$		$\lambda = 0.01$		$\lambda = 0.10$	
1	0.0129	0.0125	0.0486	0.0475	0.0660	0.0652
12	0.00379	0.00374	0.0153	0.0142	0.0184	0.0176
72	0.00151	0.00146	0.00510	0.00489	0.00600	0.00553
288	0.00066	0.00063	0.00212	0.00198	0.00208	0.00199

† Estimators: Kalman smoothing on the realized variance approximation (KS-RV). Kalman smoothing directly on high-frequency returns (KS-HF). $\mu = 0.05$, $\beta = 0$, $\rho = 0$.

By comparing columns 1-2 of Table 1 with columns 3-4, we can see that as the variance of the marginal distribution increases the MSE increases. Notice that this happens regardless of whether one uses the realized variance approximation or high-frequency returns. Columns 5-6 of Table 1 contain estimates when $\sigma^2(t) \sim IG$ and $\lambda = 0.1$, where again the realized variance approximation only introduces a small error despite changing the marginal. The main conclusions from this exercise are that the realized variance approximation works well for the Kalman smoother, it is relatively stable across marginals, but it increases for processes with larger variance.

Whereas the Kalman smoother's estimates from both state space representations are close, differences between the two particle smoothers' estimates are more noticeable. Figure 1 (c) and (d) provides some visual evidence. These estimates are computed using $N = 5,000$ particles. Although the Monte Carlo experiment above did not include the particle filter, informal comparisons like this indicate that using the realized variance approximation actually provides better estimates from the particle filter. Estimates from the particle filter on both state space representations get closer to one another as M increases. Meanwhile, the realized variance estimates in Figure 1 (e) are still considerably far from converging to the actual variance (notice the different scale on the y-axis).

Figure 1 (f) plots the effective sample size (ESS) over time. The ESS, see e.g. Cappé et al. (2005), is an estimator commonly used in the particle filtering literature to measure the number of particles that represent the current filtering distribution. The picture demonstrates how the number of particles degenerates when the instantaneous variance jumps. During periods with large jumps, the distribution is only estimated by a handful of particles. This occurs for two reasons. The importance density does not include the current observation when it moves the particles. Secondly, jumps correspond to small tail events and the importance density typically only simulates a few particles of that size. Below, I assess the effects of the particle size on the estimates.

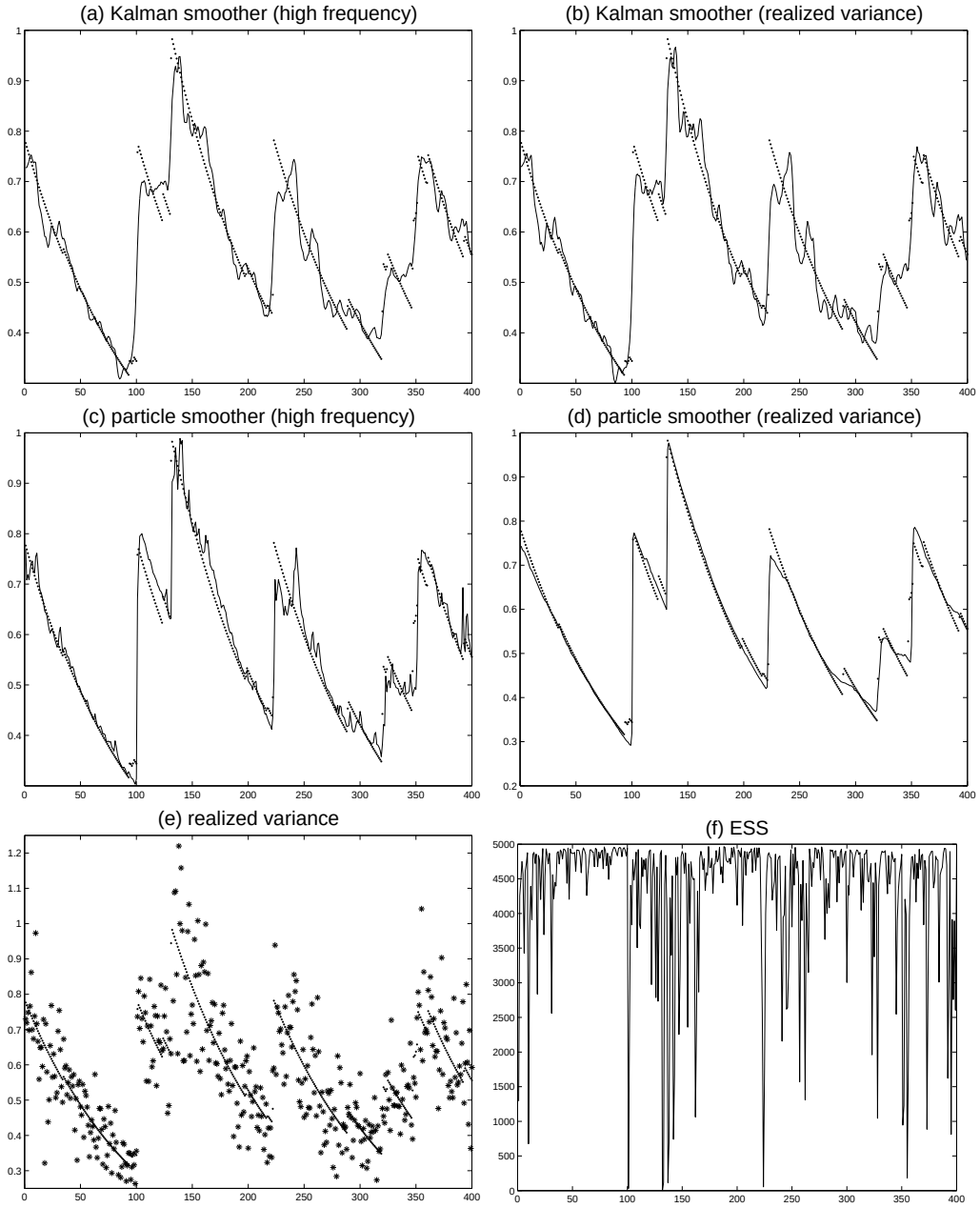

Fig. 1. Estimates of the actual variance over 400 days (a) Kalman smoother on high-frequency returns (b) Kalman smoother on the realized variance approximation (c) particle smoother on high-frequency returns (d) particle smoother on the realized variance approximation (e) realized variance (f) effective sample size. $\xi = 0.5$, $\omega^2 = 0.0625$, $M = 72$, $\mu = 0.05$, $\beta = 0$, $\rho = 0$.

Table 2
Mean square error of the estimators of actual variance when $\sigma^2(t) \sim Ga(\nu, \alpha)$.†

	$\xi = 0.5$, $\omega^2 = 0.0625$					$\xi = 0.5$, $\omega^2 = 0.25$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
M	RV	KP	KS	PP	PS	RV	KP	KS	PP	PS
	$\lambda = 0.01$									
1	0.621	0.0224	0.0129	0.0283	0.0159	1.031	0.0645	0.0351	0.241	0.156
12	0.0518	0.00791	0.00379	0.00993	0.00381	0.0839	0.0215	0.00951	0.0648	0.0416
72	0.00887	0.00379	0.00151	0.00364	0.00095	0.0158	0.0114	0.00390	0.00968	0.00259
288	0.00216	0.00209	0.00066	0.00156	0.00025	0.0032	0.00598	0.00147	0.00588	0.00132
	$\lambda = 0.10$									
1	0.619	0.0460	0.0351	0.0516	0.0391	0.987	0.153	0.0960	0.235	0.145
12	0.0517	0.0235	0.0111	0.0260	0.0114	0.0828	0.0696	0.0257	0.0880	0.0330
72	0.00882	0.0135	0.00390	0.0128	0.00313	0.0140	0.0421	0.00825	0.0422	0.00707
288	0.00217	0.00968	0.00145	0.00911	0.00112	0.00351	0.0340	0.00285	0.0338	0.00249

† Estimators: Realized variance (RV), Kalman prediction (KP), Kalman smoothing (KS), particle prediction (PP), particle smoothing (PS). The estimators are computed for two different marginals with mean ξ and variance ω^2. $\mu = 0.05$, $\beta = 0$, $\rho = 0$.

4.2. *Comparison using realized variance*

Using the realized variance approximation, this section compares the two filtering algorithms under several specifications of the model to understand if and when differences are significant. I only report the one-step ahead forecasts and smoothed estimates from the algorithms as these are of primary interest to economists. Table 2 presents the MSE computed from the Kalman and particle filtering algorithms on a model with gamma marginal for two different values of the variance ω^2 and two different values of the persistence λ. The table also includes the MSE from realized variance in columns 1 and 6. The simulation exercise used 250 realizations of 500 days and the particle filter was run with $N = 5,000$ particles. We can see that the Kalman filter performs considerably better than the particle filter for small values of M across each of the four processes. Differences between the two algorithms' estimates gets larger as ω^2 increases. This can be seen by comparing columns 1-5 with 6-10 of Table 2 for a fixed value of λ. Both algorithms outperform realized variance until roughly $M = 288$. Of course, this neglects the fact that both algorithms are being run on the true model. In practice, we would expect the MSE from the model-based estimators to be closer to realized variance due to the effects of parameter estimation and model misspecification.

The impact of the number of particles on the estimates from the particle filter can also be assessed. Table 3 provides estimates when the particle filter is run on 3 of the 4 series that were used to construct Table 2 only with $N = 1,000$ and $10,000$ particles. This allows for the comparison of three different particle sizes. Adding more particles at low frequencies makes the particle filter perform slightly worse whereas at higher frequencies increasing the number of particles improves the estimates. The difference between the estimates can unfortunately be significant relative to the differences between the Kalman

Table 3
Mean square error of the estimators of actual variance for two additional particle sizes when $\sigma^2(t) \sim Ga(\nu, \alpha)$.†

	$\xi = 0.5,\ \lambda = 0.01$				$\xi = 0.5,\ \lambda = 0.01$				$\xi = 0.5,\ \lambda = 0.1$			
	$\omega^2 = 0.0625$				$\omega^2 = 0.25$				$\omega^2 = 0.25$			
N	1,000		10,000		1,000		10,000		1,000		10,000	
M	PP	PS	PP	PS	PP	PS	PP	PS	PP	PS	PP	PS
1	0.0273	0.0132	0.0312	0.0173	0.211	0.144	0.273	0.162	0.221	0.133	0.247	0.147
12	0.00982	0.00360	0.0112	0.00407	0.0635	0.0404	0.0659	0.0439	0.0897	0.0367	0.0911	0.0361
72	0.00420	0.00152	0.00366	0.00096	0.0113	0.00691	0.00912	0.00213	0.0433	0.00825	0.0398	0.00686
288	0.00165	0.00044	0.00152	0.00022	0.00602	0.00147	0.00560	0.00108	0.0347	0.00391	0.0325	0.00201

† See Table 2 for a description of the column headings. $\mu = 0.05$, $\beta = 0$, $\rho = 0$.

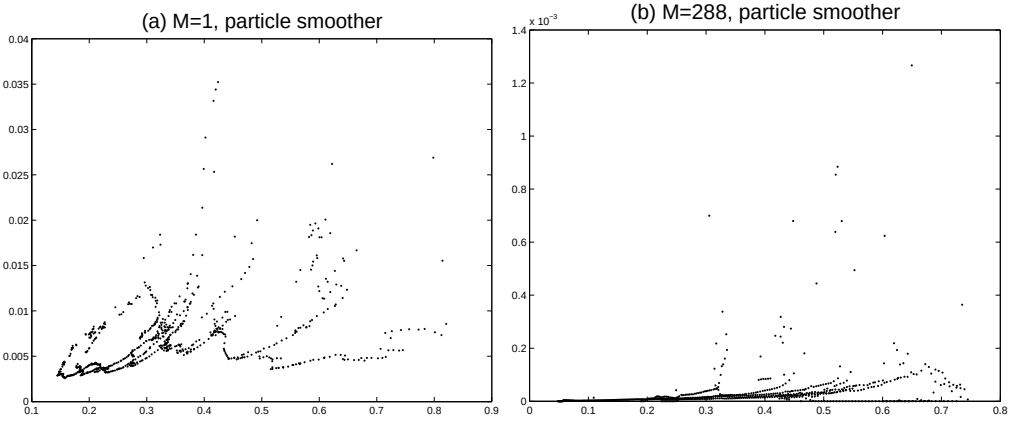

Fig. 2. Graph of $V\left(\sigma_n^2 \mid \mathcal{F}_{n+20}\right)$ versus $E\left(\sigma_n^2 \mid \mathcal{F}_{n+20}\right)$ for the fixed-lag particle smoother over 750 days (a) M=1 (b) M=288.

filter and particle filter. For the remainder of this paper, I run all the particle filtering algorithms with $5,000$ particles.

In the experiments above and in those that follow, the relative quality of the particle filter's estimates depends on the frequency at which returns are observed. Barndorff-Nielsen and Shephard (2001a) noticed that the particle filter's estimates of $E\left(\sigma_n^2 \mid \mathcal{F}_n\right)$ tend to be positively correlated with $V\left(\sigma_n^2 \mid \mathcal{F}_n\right)$. Figure 2 (a) depicts this correlation pattern for the smoothed estimates when the realized variance approximation is used and $M = 1$. As noted by Barndorff-Nielsen and Shephard (2001a), this correlation pattern does not exist for the Kalman filter's estimates of the model. Although this type of behavior is pronounced when the particle filter is applied to the BNS-SV model, this relationship also occurs when it is used to estimate the conditional volatility of returns in a standard discrete time log-normal SV model. In a SV model, the ratio of measurement noise to transition noise is a function of the level of the state variable. The ratio is larger at higher levels of volatility and it is of course time-varying. Conversely, when an SV model is approximated in a linear state space form for use with the Kalman filter, this ratio is constant through time. Importantly, this behavior becomes less pronounced as the

Table 4
Mean square error of the estimators of actual variance when $\sigma^2(t) \sim IG(\delta, \gamma)$.†

	$\xi = 0.5$, $\omega^2 = 0.0625$					$\xi = 0.5$, $\omega^2 = 0.3$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
M	RV	KP	KS	PP	PS	RV	KP	KS	PP	PS
	$\lambda = 0.01$									
1	0.625	0.0244	0.0149	0.0370	0.0207	1.036	0.0872	0.0480	0.661	0.586
12	0.0520	0.00772	0.00369	0.0115	0.00485	0.0901	0.0248	0.0113	0.141	0.117
72	0.00922	0.00414	0.00164	0.00320	0.00097	0.0177	0.0140	0.0045	0.0102	0.00235
288	0.00218	0.00237	0.00071	0.00181	0.00031	0.00408	0.00846	0.00183	0.00692	0.00129
	$\lambda = 0.10$									
1	0.634	0.0479	0.0357	0.0583	0.0415	1.098	0.179	0.114	0.383	0.243
12	0.0519	0.0235	0.0111	0.0269	0.0113	0.0940	0.0834	0.0315	0.112	0.0461
72	0.00911	0.0138	0.00400	0.0132	0.00297	0.0154	0.0493	0.0094	0.0438	0.00491
288	0.00218	0.00968	0.00147	0.00876	0.00099	0.00392	0.0395	0.00326	0.0377	0.00226

† See Table 2 for a description of the column headings. $\mu = 0.05$, $\beta = 0$, $\rho = 0$.

frequency of returns increases, see Figure 2 (b) where the same realization is estimated when $M = 288$.

Changing marginal distributions also has an impact on the results. Table 4 contains estimates from a similar experiment as in Table 2 only when $\sigma^2(t) \sim IG$. The inverse Gaussian distribution has a different shape than the gamma distribution. For the same mean and variance, the inverse Gaussian distribution has a longer right-hand tail. Adding more probability mass in the tail means that when the instantaneous variance jumps the size of the jumps can be larger. Both the Kalman filter and particle filter perform slightly worse for a given value of M under this marginal. For example, compare the processes with $\xi = 0.5$ and $\omega^2 = 0.0625$ for each value of λ in Tables 2 and 4. At smaller values of M, the magnitude of the differences between the Kalman filter and particle filter are also much larger for the IG marginal. Notice from columns 7-10 of Table 4 that the particle filter performs considerably worse than the Kalman filter when $\omega^2 = 0.3$ and M is small.

Next, I investigate if the above observations continue to hold when the instantaneous variance is composed of a superposition of two OU processes. As mentioned in Section 2, superpositions of OU processes are intended to allow more flexibility in the autocorrelation function of the instantaneous variance. The dynamics of the individual processes $\sigma_i^2(t)$ for $i = 1, 2$ are chosen to be a mixture of a slowly decaying component $\lambda_1 = 0.01$ and a component that decays faster with $\lambda_2 = 0.5$. The second process has a larger weight with $w_2 = 0.8$, which means the overall instantaneous variance has less persistence than any of the processes considered in Tables 2-4. Both of the individual processes have gamma marginal with the same mean and variance. Table 5 contains the MSE calculated from this experiment. With less persistence, the MSE estimates in Table 5 are larger for both algorithms at all values of M and at each level of ω^2. For the superposition model, the differences between these model-based estimators and realized variance continues to be significant at small values of M. As the persistence of the instantaneous variance increases, neither algorithm provides a considerable advantage over realized variance at

Table 5
Mean square error of the estimators of actual variance when $\sigma^2(t)$ is a superposition of two OU processes with $Ga(\nu,\alpha)$ marginal.†

	$\xi=0.5$, $\omega^2=0.0625$					$\xi=0.5$, $\omega^2=0.25$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
M	RV	KP	KS	PP	PS	RV	KP	KS	PP	PS
1	0.619	0.0547	0.0453	0.0499	0.0435	1.020	0.180	0.130	0.191	0.164
12	0.0511	0.0354	0.0176	0.0343	0.0181	0.0830	0.119	0.043	0.121	0.0496
72	0.00884	0.0256	0.00596	0.0250	0.00588	0.0145	0.0938	0.0113	0.0962	0.0131
288	0.00220	0.0219	0.00197	0.0211	0.00196	0.00331	0.0814	0.00314	0.0816	0.0038

† See Table 2 for a description of the column headings. The superposition combines two processes $\sigma_i^2(t)$ where $\lambda_1=0.01$ with $w_1=0.2$ and $\lambda_2=0.5$ with $w_2=0.8$. $\mu=0.05$, $\beta=0$, $\rho=0$.

Table 6
Mean square error of the estimators of actual variance under misspecification of the marginal distribution.†

	$\xi=0.5$		$\xi=0.5$		$\xi=0.5$		$\xi=0.5$	
	$\omega^2=0.0625$		$\omega^2=0.3$		$\omega^2=0.0625$		$\omega^2=0.25$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
M	PP	PS	PP	PS	PP	PS	PP	PS
	$\lambda=0.10$		$\lambda=0.10$		$\lambda=0.10$		$\lambda=0.10$	
1	0.0518	0.0392	0.298	0.185	0.0637	0.0472	0.426	0.296
12	0.0250	0.0108	0.112	0.0436	0.0286	0.0130	0.114	0.0466
72	0.0131	0.00318	0.0507	0.0103	0.0130	0.0034	0.0451	0.00555
288	0.00965	0.00141	0.0463	0.0074	0.00913	0.00115	0.0382	0.00259

† Columns 1-4 were simulated with $IG(\delta,\gamma)$ as the true distribution. Columns 5-8 were simulated from the $Ga(\nu,\alpha)$ as the true distribution. See Table 2 for a description of the remaining column headings. $\mu=0.05$, $\beta=0$, $\rho=0$.

$M=288$, 5-minute returns.

The particle filter outperforms the Kalman filter at higher frequencies when the marginal distribution is correctly specified. However, the marginal distribution of the instantaneous variance is unknown in practice, see Barndorff-Nielsen and Shephard (2003) for a discussion on relevant distributions and Gander and Stephens (2005) for related work. Models for the instantaneous variance are typically chosen for computational convenience; the gamma distribution being the leading example. It is interesting to see how the performance of the algorithms changes when the marginal distribution is misspecified. I have simulated 250 realizations of 500 days from models with both gamma and IG marginals and then estimated them using the opposite marginal (with the same mean and variance). The marginals are consequently slightly misspecified for the particle filter. The Kalman filter's estimates only depend on ξ,ω^2,μ, and λ. I use the same values for ξ,ω^2,μ, and λ as in Tables 2 and 4 which means that the Kalman filter's estimates (and the realized variance estimates) from each marginal do not change. In other words, the Kalman filter's estimates of data simulated from the gamma marginal (IG marginal) are the same as in Table 2 (Table 4).

Columns 1-4 of Table 6 contain the estimates for two processes with different variances when the 'true' model has an IG marginal distribution. Comparing the estimates in columns 1-4 of Table 6 with the particle filter's estimates in columns 4-5 and 9-10 of Table 4, the particle filter performs slightly better at small values of M when the process has been misspecified and estimated as a gamma marginal rather than when it was estimated with the correct marginal. Whereas at higher frequencies, the particle filter performs slightly worse for both levels of the variance. The Kalman filter's estimates of the same models (in columns 2-3 and 7-8 of Table 4) are now better than or roughly equivalent to the particle filter's estimates.

When the 'true' marginal of $\sigma^2(t)$ is a gamma distribution and has been misspecified as an IG marginal, the particle filter tends to perform worse at all frequencies M. The differences are considerable at low frequencies; compare the Kalman filter's estimates in columns 2-3 and 7-8 of Table 2 with the particle filter's estimates in columns 5-8 of Table 6. The main conclusion drawn from this exercise is that the relative performance of the two model-based estimators has gotten closer. Relatively small deviations from the true marginal can be important in practice. Consequently, the Kalman filter is competitive with the particle filter at all frequencies. These conclusions assume that one can estimate the mean, variance, and persistence of the instantaneous variance reasonably well. Results in Barndorff-Nielsen and Shephard (2002) indicate that as M gets larger, estimates of these parameters (using quasi-maximum likelihood via the Kalman filter) are typically good.

4.3. *Estimation of jumps by particle filter and smoother*

Estimation of jumps and the separation of quadratic variation into continuous and discontinuous components is a critical area of research in financial econometrics. A particle filter can separate the two sources of volatility. Griffin and Steel (2006) estimate the BNS-SV model with jumps by particle filter on daily data but do not report estimates of the jumps. Related work includes Raggi and Bordignon (2006) and Johannes et al. (2006) who apply particle filters to jump-diffusions on daily data while Bos (2006) estimates jump-diffusions on intra-daily data via MCMC. Here, I compare how well the particle filter estimates the discontinuous component of quadratic variation relative to the nonparametric measure in (8). Daily estimates of the discontinuous component of quadratic variation from the particle smoother are obtained by running the algorithm on high-frequency returns and summing the squared estimates of the jumps over each day.

I simulated 50 days worth of returns at a 5-minute interval from the BNS-SV model with gamma marginal and $\lambda = 0.3$, $\mu = 0.05$, and $\rho = -0.5$. This means that all jumps in the price are negative. More flexible models for jumps can be built using superpositions of OU processes with a different ρ for each subordinator, see Barndorff-Nielsen and Shephard (2001a) and also Todorov and Tauchen (2006) for another alternative. The negative correlation between jumps in the price and the instantaneous variance is typical of the leverage effect. The value of $E\{z(1)\}$ needed to compute the jumps for the gamma process is $E\{z(1)\} = \lambda\hbar\nu E(v)/\alpha$ with $v \sim$ exponential(1). Figure 3 (a) and (b) show that smoothed estimates of quadratic variation are much more well defined than the nonparametric estimator at $M = 12$. When $M = 288$, the particle smoother's estimates of quadratic variation in (c) are quite good and potentially much clearer than

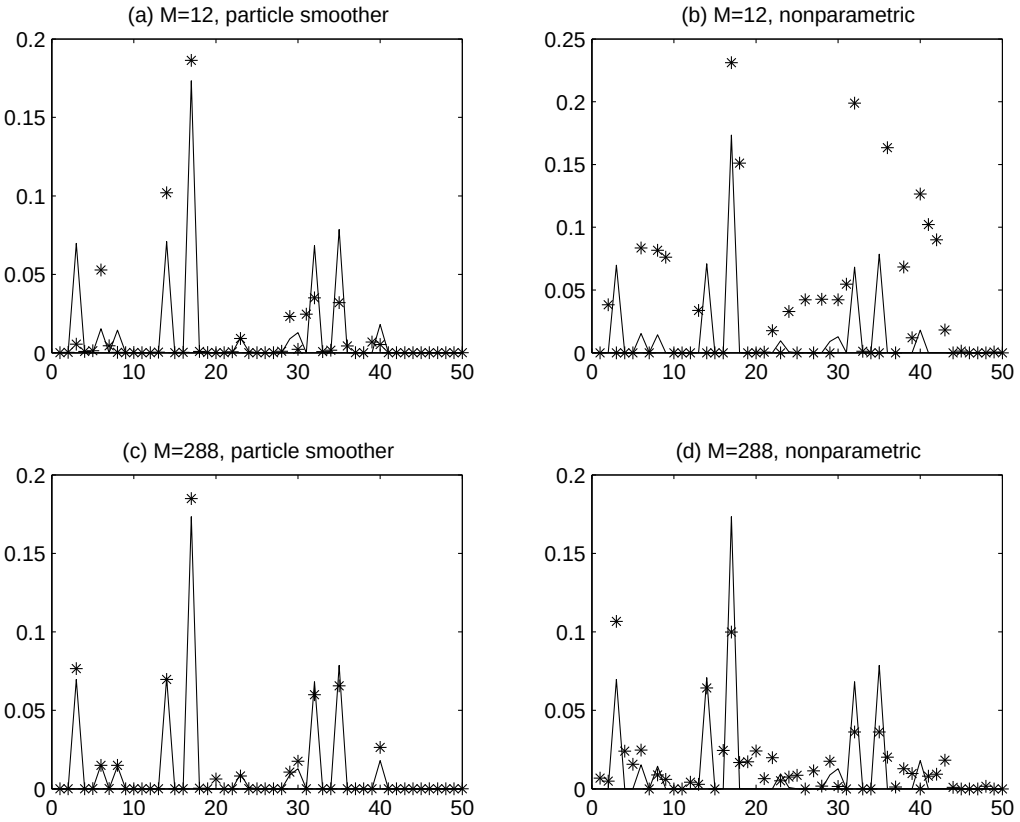

Fig. 3. Estimates of the discontinuous component of quadratic variation for different values of M. The * are the estimates and the continuous line is the true value. (a) and (c) particle smoother. (b) and (d) nonparametric estimator. $\xi = 0.5$, $\omega^2 = 0.25$, $\lambda = 0.3$, $\mu = 0.05$, $\beta = 0$, and $\rho = -0.50$.

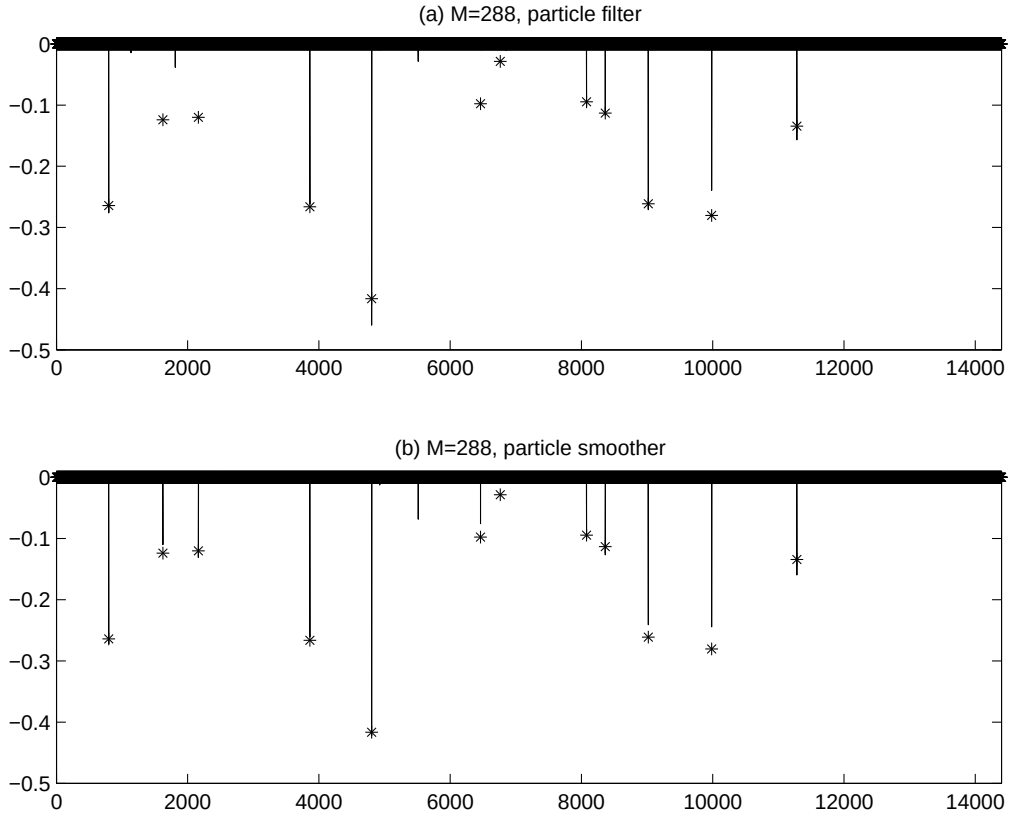

Fig. 4. Estimates of jumps from (a) the particle filter and (b) the particle smoother. $M = 288$ corresponding to 5-minute returns over 50 days. The * are the true values of the jumps. $\xi = 0.5$, $\omega^2 = 0.25$, $\lambda = 0.3$, $\mu = 0.05$, $\beta = 0$, and $\rho = -0.50$.

the nonparametric measure's estimates in (d). Of course, the particle smoother has an advantage over the nonparametric estimator as it is run on the true model with the true parameter values.

Jump statistics like (8) are typically applied on a daily basis, whereas the particle filter can also estimate jumps intra-daily. Figure 4 compares the particle filter and smoothers' estimates of the jumps rather than the quadratic variation directly on the high-frequency returns with $M = 288$. The filtered or 'real-time' estimates of the jumps in (a) are quite good with the smoothed estimates in (b) picking up some of the smaller jumps.

5. Conclusion

This paper evaluated the performance of filtering and smoothing algorithms for Lévy-driven SV models. The performance of the particle filter versus the Kalman filter depends upon the marginal distribution of $\sigma^2(t)$ but more importantly on the number of intra-daily returns that are available. The Kalman filter's estimates appear to be more robust, especially at lower frequencies. At higher frequencies, the differences between the two estimators is small with the particle filter performing better. If the marginal distribution

of the model is slightly misspecified, the Kalman filter will be competitive with the particle filter if not potentially outperforming it at all frequencies. One advantage of a particle filter that I did not emphasize above is its ability to estimate the likelihood of the model and provide model diagnostics; see Andrieu et al. (2004).

Another particle filtering and smoothing algorithm a researcher might consider is a sequential Monte Carlo sampler from Del Moral et al. (2006); albeit, this will likely increase the computational time as it requires reparameterizing the model as discussed above. Particle filtering algorithms are also relevant for estimating the time-change in models of time-deformed Lévy processes as in Carr et al. (2003) and Carr and Wu (2004) as well as models built from CARMA processes as in Brockwell (2001) and Tsai and Chan (2005). Work on estimating the unobserved time-change using the Kalman filter has been completed by Barndorff-Nielsen and Shephard (2006).

References

Andrieu, C., Doucet, A., Singh, S. S., Tadić, V., 2004. Particle methods for change detection, system identification, and control. Proceedings of the IEEE 92 (3), 423–438.

Barndorff-Nielsen, O. E., Nielsen, B., Shephard, N., Ysusi, C., 2004. Measuring and forecasting financial variability using realised variance. A. C. Harvey, S. J. Koopman, and N. Shephard, Eds. State Space and Unobserved Components Models: Theory and Applications, Cambridge, UK: Cambridge University Press.

Barndorff-Nielsen, O. E., Shephard, N., 2001a. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). Journal of the Royal Statistical Society, Series B 63 (2), 167–241.

Barndorff-Nielsen, O. E., Shephard, N., 2001b. Normal modified stable processes. Theory of Probability and Mathematical Statistics 65, 1–19.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64 (2), 253–280.

Barndorff-Nielsen, O. E., Shephard, N., 2003. Integrated OU processes and non-Gaussian OU-based stochastic volatility models. Scandinavian Journal of Statistics 30, 277–295.

Barndorff-Nielsen, O. E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2 (1), 1–37.

Barndorff-Nielsen, O. E., Shephard, N., 2006. Impacts of jumps on returns and realised variances: econometric analysis of time-deformed Lévy processes. Journal of Econometrics 131, 217–252.

Barndorff-Nielsen, O. E., Shephard, N., 2007. Variation, jumps, market frictions and high frequency data in financial econometrics. R. Blundell, P. Torsten, and W. Newey, Eds. Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Econometric Society Monographs, Cambridge University Press, Cambridge, UK.

Bos, C. S., 2006. Model-based estimation of high frequency jump diffusions with microstructure noise and stochastic volatility, mimeo, Department of Econometrics, Vrije Universiteit Amsterdam.

Brockwell, P. J., 2001. Lévy-driven CARMA processes. Annals of the Institute of Statistical Mathematics 53 (1), 113–124.

Cappé, O., Godsill, S., Moulines, E., 2007. An overview of existing methods and recent advances in sequential Monte Carlo. IEEE Proceedings in Signal Processing, Forthcoming.

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. Springer, New York, NY.

Carpenter, J., Clifford, P., Fearnhead, P., 1999. An improved particle filter for non-linear problems. IEE Proceedings. Part F: Radar and Sonar Navigation 146 (1), 2–7.

Carr, P., Geman, H., Madan, D. B., Yor, M., 2003. Stochastic volatility for Lévy processes. Mathematical Finance 13, 345–382.

Carr, P., Wu, L., 2004. Time-changed Lévy processes and option pricing. Journal of Financial Economics 71, 113–141.

Cont, R., Tankov, P., 2004. Financial Modeling with Jump Processes. Chapman & Hall, London, UK.

Del Moral, P., Doucet, A., Jasra, A., 2006. Sequential Monte Carlo for Bayesian computation. Bayesian Statistics 8, Oxford University Press, Oxford, UK.

Doucet, A., de Freitas, N., Gordon, N. J., 2001. Sequential Monte Carlo Methods in Practice. Springer, New York, NY.

Durbin, J., Koopman, S. J., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford, UK.

Gander, M., Stephens, D. A., 2005. Stochastic volatility modeling with general marginal distributions: inference, prediction, and model selection for option pricing, mimeo, Department of Mathematics, Imperial College London.

Griffin, J., Steel, M., 2006. Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. Journal of Econometrics 134, 605–644.

Johannes, M., Polson, N., Stroud, J., 2006. Optimal filtering of jump-diffusions: extracting latent states from asset prices, mimeo, Graduate School of Business, Columbia University.

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics 5, 1–25.

Kitagawa, G., Sato, S., 2001. Monte Carlo smoothing and self-organizing state-space model. A. Doucet, N. de Freitas, and N. Gordon, Eds. Sequential Monte Carlo Methods in Practice, Springer, New York, NY.

Musso, C., Oudjane, N., Le Gland, F., 2001. Improving regularised particle filters. A. Doucet, N. de Freitas, and N. Gordon, Eds. Sequential Monte Carlo Methods in Practice, Springer, New York, NY.

Nicolato, E., Venardos, E., 2003. Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. Mathematical Finance 13, 445–466.

Pitt, M. K., 2001. Comment on 'Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics'. Journal of the Royal Statistical Society, Series B 63 (2), 228–229.

Raggi, D., Bordignon, S., 2006. Comparing stochastic volatility models through Monte Carlo simulations. Computational Statistics and Data Analysis 50, 1678–1699.

Ristic, B., Arulampalam, S., Gordon, N. J., 2004. Beyond the Kalman filter: particle filters for tracking applications. Artech House Press, Boston, MA.

Roberts, G. O., Papaspiliopoulos, O., Dellaportas, P., 2004. Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility models. Journal of the Royal Statistical Society, Series B 66 (2), 369–393.

Rosínski, J., 2001a. Comment on 'Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics'. Journal of the Royal Statistical Society, Series B 63 (2), 230–231.

Rosínski, J., 2001b. Series representations of Lévy processes from the perspective of point processes. O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick, Eds. Lévy Processes - Theory and Applications, Birkhauser, Boston, MA.

Todorov, V., 2006. Econometric analysis of jump-driven stochastic volatility models, mimeo, Department of Economics, Duke University.

Todorov, V., Tauchen, G., 2006. Simulation methods for Lévy-driven CARMA stochastic volatility models. Journal of Business and Economic Statistics 24 (4), 450–469.

Tsai, H., Chan, K. S., 2005. A note on non-negative continuous time processes. Journal of the Royal Statistical Society, Series B 67 (4), 589–597.