Nonlinear Filtering of Stochastic Differential Equations with Jumps

Michael Johannes Nicholas Polson Jonathan Stroud*

October 8, 2002

Abstract

In this paper, we develop an approach for filtering state variables in the setting of continuous-time jump-diffusion models. Our method computes the filtering distribution of latent state variables conditional only on discretely observed observations in a manner consistent with the underlying continuous-time process. The algorithm is a combination of particle filtering methods and the "filling-in-the-missing-data" estimators which have recently become popular. We provide simulation evidence to verify that our method provides accurate inference. As an application, we apply the methodolgy to the multivariate jump models in Duffie, Pan and Singleton (2000) using daily S&P 500 returns from 1980-2000 and we investigate option pricing implications.

*Johannes is at the Graduate School of Business, Columbia University, 3022 Broadway, NY, NY, 10027, mj335@columbia.edu. Polson is at the Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago IL 60637, ngp@gsbngp.uchicago.edu. Stroud is at The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6302, stroud@wharton.upenn.edu. We thank Mark Broadie, Mike Chernov and Mike Pitt who suggested the application of Storvik's (2002) parameter learning algorithm. Special thanks go to Neil Shephard for his extensive comments.

1 Introduction

Since the work of Merton (1969, 1971) and Black and Scholes (1973), researchers in economics and finance commonly specify their models in continuous-time and assume that state variables solve stochastic differential equations (SDEs). The first generation of these models were diffusions, but research is increasingly focussed on more general jump-diffusion models (see, e.g., Duffie, Pan and Singleton (2000)). These models often lead to closed form solutions or easy to solve differential equations for objects of interest such as prices or Bellman equations. Moreover, through judicious choice of the characteristics (the drift, diffusion, jump intensity and jump distribution) of the SDEs, the models provide enough flexibility to accommodate a wide range of dynamics for the state variables.

Due to their popularity, a number of methods have been developed for estimating characteristics of diffusion[1] and jump-diffusion models.[2] In this paper, we focus on a related problem which has received less attention: latent state variable filtering in continuous-time jump-diffusion models. The motivation for this problem comes from practical finance applications such as option pricing and portfolio allocation where popular models incorporate both diffusive and jump components.[3] In these contexts, parameter estimation is only the first step and in the next stage reliable estimates of unobserved state variables such as volatility are required to price options or to compute portfolio allocations.

In certain special cases, the filtering problem can be solved. For example, in linear and Gaussian models, we can solve the SDE's and convert the system into a discrete-time Gaussian model and apply the Kalman filter. In stochastic volatility models, it is possible to

[1] See, for example, the review article by Ait-Sahalia, Hansen and Scheinkman (2002) and also Gouriéroux, Monfort and Renault (1993), Pedersen (1995), Ait-Sahalia (1996a, 1996b), Gallant and Long (1997), Conley, Hansen, Luttmer and Scheinkman (1998), Brandt and Santa-Clara (2001), Elerian, Shephard and Chib (2001) and Eraker (2001).

[2] See, Piazzesi (2001), Johannes (2001), Andersen, Benzoni and Lund (2002), Chib, Nardari and Shephard (2002), Eraker, Johannes and Polson (2002) and Pan (2002).

[3] See, for example, Duffie, Pan and Singleton (2000) or Liu, Longstaff and Pan (2002).

estimate volatility if there are no jumps in the observed state variable (Foster and Nelson (1996), and Andersen, Bollerslev and Diebold (2002)). Del Moral , Jacod and Protter (2001) provide a Monte Carlo solution for estimating states in the setting of more general diffusion models. However, none of these methods have been applied in models with jumps and there are no existing methods available for separating out jump and stochastic volatility components based on contemporaneously available data.[4] The difficulty in separating out these components can be seen from the following example. Suppose there is a large, negative equity return: how do we determine if it is a jump in returns, a shock due to high volatility or a combination of the two?

This paper provides an efficient simulation based solution to the filtering problem in the general setting where the state variables solve parameterized stochastic differential equations with jumps. Our approach consists of two steps. First, we augment the state space by high frequency augmentation or "filling-in-missing data" by simulating M additional data points between the discrete observation times. This insures that the conditional densities we use are consistent with the continuous-time specification and is a common method used for parameter estimation.[5] Second, we use the particle filter and its extension with auxiliary variables on the augmented system to estimate the filtering densities.

Our approach offers a number of practical advantages. First, the class of models in which our approach applies is quite general. We only require that the model can be discretized and simulated using standard techniques such as the Euler scheme. Thus our method applies in models with nonlinear characteristics, multivariate diffusion models with Poisson driven jumps, models with embedded continuous-time, discrete state Markov chains. Second, unlike estimators that rely on increasingly fine observations, our filters do not require high frequency data, although, of course, our methodology does not preclude it. Third,

[4] It is possible to estimate the *smoothed* distribution of volatility and jumps: see Eraker, Johannes and Polson (2002).

[5] See, e.g, Pedersen (1995), Elerian, Shephard and Chib (2001), Eraker (2001), Brandt and Santa-Clara (2002) and Durham and Gallant (2002).

our method uses the structure imposed by the model under consideration, in contrast to recent nonparametric methods. Since most practical applications use specific models, our approach is efficient in the sense that we use the information contained in the model. Fourth, our approach delivers the entire filtering distribution. Fifth, the particle filter adapts in a straightforward manner to address the issue of sequential parameter learning. Finally, the algorithm is computationally very fast and easy to implement when compared to alternative methods such as MCMC or naive Monte Carlo.

To demonstrate the algorithm, we consider three special cases of the double-jump model of Duffie, Pan and Singleton (2000): the square-root stochastic volatility (SV, Heston (1993)) model, the square-root stochastic volatility model with jumps in returns (SVJ, Bates (1996)) and a square-root stochastic volatility model with correlated jumps in returns and volatility (SVCJ, Duffie, Pan and Singleton (2000)). This model is an excellent example because these factors are important features of the data[6] and the model contains all of the cases of practical interest: a time-varying state variable (volatility), jumps in the observed state variable and jumps in the unobserved state.

We provide simulation results to document the performance of the filtering algorithm as a function of the number of additional augmented data points. In most cases, filling a small number of data points results in a very accurate filter. Using daily S&P 500 index data from 1980-2000, we apply the filtering methodology to the SV, SVJ and SVCJ models to obtain filtered estimates of jump times, jump sizes and volatility. We also extend our methodology to cover the case of sequential parameter learning. Building on the algorithm of Storvik (2002), we provide sequential parameter estimates consistent with the continuous-time model for the SV model. While the algorithm performs well on simulations, it does not perform well on S&P 500 data as the algorithm degenerates after the Crash of 1987. The poor performance is likely due to model misspecification.

Using the filtered estimates, we address an important issue related to Merton (1976b).

[6]See, e.g., Bakshi, Cao and Chen (1997), Pan (2002), Chernov, et al. (2002) and Eraker, Johannes and Polson (2002).

Merton (1976b) quantifies the error that an investor incurs if they incorrectly apply the Black-Scholes model when the true model is Merton's (1976a) mixed jump-diffusion. In the setting of models with time-varying volatility and jumps a similar issue occurs and was mentioned above. How does one know when a jump occurs? Our filter provides a statistical mechanism for identifying jump and volatility components. We show how the different models generate drastically different implied volatility curves in periods of market stress. Moreover, the volatility estimates remain different for long periods of time. This quantifies a new dimension along which jumps have an important economic impact on option prices.

There are related papers using particle filters which focus on parameter estimation or specification testing. Durham and Gallant (2002) and Pitt (2002) use particle filtering and data augmentation to construct likelihood functions for maximum likelihood parameter estimation in diffusion based stochastic volatility models. Pitt and Shephard (1999), Durham (2002) and Chib, Nardari and Shephard (2002) use particle filtering in discrete time stochastic volatility models to construct likelihood functions for estimation or testing. Chib, Nardari and Shephard (2002) also consider a discrete-time model with t-distributed errors and jumps in returns. Barndorff-Nielson and Shephard (2002) use particles to filter volatility in a stochastic volatility model driven by Lévy processes. Our approach extends these papers into continuous-time and/or jump-diffusion models.

The rest of the paper is outlined as follows. The next section introduces the models under consideration, reviews general filtering theory and introduces our algorithms. Section 3 provides empirical results. Section 4 concludes.

2 General Filtering Theory

This section discusses the class of models that we consider and reviews standard discrete and continuous-time approaches for filtering.

2.1 Model Specification

We assume that the vector of observed state variables, S_t, and the vector of unobserved state variables, X_t, jointly solve

$$S_t = S_0 + \int_0^t \mu^s\left(\Theta, S_s, X_s\right) ds + \int_0^t \sigma^s\left(\Theta, S_{s-}, X_{s-}\right) dW_s^s + \sum_{j=1}^{N_t^s} f^s\left(\xi_j^s, S_{\tau_j^s-}, X_{\tau_j^s-}\right) \quad (1)$$

$$X_t = X_0 + \int_0^t \mu^x\left(\Theta, S_s, X_s\right) ds + \int_0^t \sigma^x\left(\Theta, S_{s-}, X_{s-}\right) dW_s^x + \sum_{j=1}^{N_t^x} f^x\left(\xi_j^x, S_{\tau_j^x-}, X_{\tau_j^x-}\right) \quad (2)$$

where $s-$ denotes the limit taken from the left, W_t^s and W_t^s are vectors of potentially correlated Brownian motions; $\{N_t^s, N_t^x\}_{t>0}$ are a pair of point processes with stochastic intensities $\{\lambda^j\left(\Theta, S_t, X_t\right)\}_{j=s,x}$ and associated jump times $\left\{\tau_j^s, \tau_j^x\right\}_{j=1}^{\infty}$; and $\left\{\xi_j^s, \xi_j^x\right\}_{j=1}^{\infty}$ are the marks of the point process with an $\mathcal{F}_{\tau_j-}$ conditional distribution $\Pi\left(\Theta\right)$. We refer to (1) as the observation equation and (2) as the state evolution. The jump impact function f translates the marks of the point process $\left\{\xi_j^s, \xi_j^x\right\}$ into jumps in the state variables. In most cases the jump impact function is either independent of S_t and X_t or is a simple function of one or the other. The interpretation of the process is as follows. At a jump time τ_j^s or τ_j^x, there is a discontinuity in the sample path,

$$S_{\tau_j^s} - S_{\tau_j^s-} = f^s\left(\xi_j^s, S_{\tau_j^s-}, X_{\tau_j^s-}\right) \text{ and } X_{\tau_j^x} - X_{\tau_j^x-} = f^x\left(\xi_j^x, S_{\tau_j^x-}, X_{\tau_j^x-}\right),$$

and between jump times, $\tau_j \leq t < \tau_{j+1}$, the state variables diffuse:

$$dS_t = \mu^s\left(\Theta, S_t, X_t\right) dt + \sigma^s\left(\Theta, S_t, X_t\right) dW_t^s$$
$$dX_t = \mu^x\left(\Theta, S_t, X_t\right) dt + \sigma^x\left(\Theta, S_t, X_t\right) dW_t^x.$$

Throughout, we assume that the characteristics in (1) and (2) have sufficient regularity for a well defined strong solution. Sufficient conditions for a strong solution include linear growth and Lipschitz on the drift, diffusion and jump impact function and integrability conditions on the initial values. If we define $y = (s, x)$ and $z = (s', x')$, these conditions state that for $j = s, v$

$$Growth : \left\| \mu^j \left(\Theta, y \right) \right\| + \left\| \sigma^j \left(\Theta, y \right) \right\| + \int \left\| f^j \left(\xi^j, y \right) \right\| \Pi^j \left(\Theta, d\xi^j \right) \leq k^j \left(1 + \left\| y \right\| \right)$$

$$\begin{aligned} Lipschitz : & \left\| \mu^j \left(\Theta, s, x \right) - \mu^j \left(\Theta, s', x' \right) \right\| + \left\| \sigma^j \left(\Theta, s, x \right) - \sigma^j \left(\Theta, s', x' \right) \right\| + \\ & \int \left\| f^j \left(\xi^j, s, x \right) - f^j \left(\xi^j, s', x' \right) \right\| \Pi^j \left(\Theta, d\xi^j \right) \leq k^j \left(1 + \left\| s - s' \right\| + \left\| x - x' \right\| \right). \end{aligned}$$

For integrability, we assume that the initial values, S_0 and X_0 are square-integrable. Jacod and Shiryaev (1987) and Gihkman and Skorohod (1972) provide details on the regularity conditions. In some cases, such as models with square-root volatility, the growth condition is violated at the boundary of the state space. In this case, there are often alternative conditions to guarantee existence and uniqueness (see, e.g., Duffie, Filipovic and Schachermayer (2002)).

Our specification is general enough to cover nearly all of the special cases of interest in economics and finance: multivariate diffusion models, multivariate jump-diffusion models (Duffie, Pan and Singleton (2000)) and regime-switching models (see Dai and Singleton (2002) and Landen (2000)). The only class of models that falls outside this class are those driven by a Lévy processes. While these models are not typically used in finance applications, these models are promising for modeling stochastic volatility. The methods developed below extend straightforwardly to this case.

2.2 Classical Filtering Methods

We now turn to the issue of filtering the unobserved states, jump times and jump sizes from observations on S_t. We assume that S_t is observed at discretely spaced intervals that, without any loss of generality, are equally spaces apart. The vector of observed data is $S_{0:t} = [S_1, ..., S_t]$. We first consider the filtering problem with known parameters and later relax this assumption.

The filtering distribution is the conditional distribution of the latent state variables given contemporaneous information on the observed state: if X_t is the latent state, the filtering density is $p(X_t|S_{0:t})$ and our goal is to compute this density sequentially for large t. We make note of the difference between the filtering density, $p(X_t|S_{0:t})$, and the smoothing density, $p(X_t|S_{0:T})$ where $S_{0:T}$ is the entire sample. The smoothing density also provides state estimates, but it uses future data to estimate the states. For practical problems such as option pricing or portfolio allocation, this is not relevant. We will first discuss existing approaches to filtering with discrete and continuous observations.

2.2.1 Discrete time filtering

With discrete observations, the filtering problem can be solved analytically in the class of linear, Gaussian state space models (the Kalman filter) and in certain other cases in which the state variables have a discrete distribution. For the models we consider, the discretely sampled distributions are generally neither linear nor Gaussian. If the conditional distribution of the observed state given the unobserved state and the state evolution equation were known in closed form, generalizations of the Kalman filter such as the extended Kalman filter could apply. Unfortunately, the conditionals are not known, but even if they were known analytically, this approach is not likely to work well as the transitions can be extremely non-normal.

Since the problem cannot be solved analytically, there are a number of related numerical approaches which rely on approximating the filtering density with a discrete distribution. This approach, initially developed by Gordon, Salmond and Smith (1993) has wide applicability to many nonlinear and non-normal discrete time models. In general, the method can be directly applied if one can sample from the state evolution and evaluate the conditional distribution of the observable given the latent states. In the case of continuous-time jump-diffusions, this is not analytically possible, but below, we adapt the method to handle this case.

2.2.2 Continuous time filtering

The classical approach to filtering in continuous-time models is to assume that S_t is observed continuously in time, the continuous-record case (see Pugachev (1987) and Liptser and Shiryaev (2001)). In this case, the sample path is observed at every time point and this information is summarized by the filtration $\mathcal{F}^s_{0,t} = \sigma(S_u, 0 \leq u \leq t)$. As before, the solution to the filtering problem is the posterior distribution of the unobserved variables $\{X_t\}_{t\geq 0}$ and jump times and sizes,

$$\left(\left\{\xi^s_j, \tau^s_j\right\}_{j=1}^{N^s_t}, \left\{\xi^x_j, \tau^x_j\right\}_{j=1}^{N^x_t}\right),$$

given $\mathcal{F}^s_{0,t}$.

The assumption of continuous-record observations implies that much of the filtering problem disappears. To see this, note first that with continuous-record observations, both S_t and S_{t-} are known implying that the jump times and sizes, $\left\{\tau^s_j, f^s\left(\xi^s_{\tau^s_j}, S_{\tau^s_j-}, X_{\tau^s_j-}\right)\right\}_{j=1}^{N^s_t}$, are observed as discontinuities in the sample path. Moreover, in most cases of interest, the jump impact function is independent of X_t and f^s can be inverted to solve for the mark, $\xi^s_{\tau_j}$.

Given that jump times and sizes in S_t are known, we now consider the information contained in continuous component of the observed state variables, S^c_t, which is given by

$$S^c_t = S_t - \sum_{j=1}^{N^s_t} f^s\left(\xi^s_j, X_{\tau^s_j-}, S_{\tau^s_j-}\right) = S_{t-1} + \int_{t-1}^t \mu\left(\Theta, S_s, X_s\right) ds + \int_{t-1}^t \sigma\left(\Theta, S_{s-}, X_{s-}\right) dW^s_s.$$

With observations on S^c_t, consider a partition of the interval $[0,t]$ into n intervals, $0 = t_1 < ... < t_n = t$, and set $\Delta_n = \sup_{1\leq j\leq n} |t_j - t_{j-1}|$. The quadratic variation of the continuous component is defined as

$$QV(t) = \int_0^t \sigma^2\left(\Theta, S_s, X_s\right) ds = \lim_{\Delta_n \to 0} \sum_{j=1}^n \left|S^c_{t_j} - S^c_{t_{j-1}}\right|^2$$

where the limit is taken in probability. The derivative of this expression with respect to t and taken from the left is

$$\frac{dQV(t)}{dt} = \frac{d}{dt}\int_0^t \sigma^2\left(\Theta, S_s, X_s\right) ds = \sigma\left(\Theta, S_t, X_t\right) \ a.s. \tag{3}$$

and implies that the diffusion coefficient, $\sigma(\Theta, S_t, X_t)$, is also observed. In most cases of practical interest, this implies that X_t is also observed. For example, this occurs if σ is an invertible function of X_t.

For a concrete example, consider the SVCJ model of Duffie, Pan and Singleton (2000):

$$S_{t+1} = S_t + \int_t^{t+1} S_s (r_s + \eta_v V_s)\, ds + \int_t^{t+1} S_{s-} \sqrt{V_{s-}} dW_t^s + \sum_{j=N_t+1}^{N_{t+1}} S_{\tau_{j-}} \left(e^{\xi_j^s} - 1\right)$$

$$V_{t+1} = V_t + \int_t^{t+1} \kappa_v (\theta_v - V_s)\, ds + \int_t^{t+1} \sigma_v \sqrt{V_{s-}} dW_t^v + \sum_{j=N_t+1}^{N_{t+1}} \xi_j^v$$

where $\xi^s | \xi^v \sim N(\mu_s + \rho_J \xi^v, \sigma_s^2)$, $\xi^v \sim \exp(\mu_v)$ and $N_t \sim Poi(\lambda t)$. Continuous-record observations on the equity price imply that N_t^s and $\xi_{\tau_j}^s$ are known. Additionally, the quadratic variation theorem implies that V_t is observed. This in turn implies that N_t^v and $\xi_{\tau_j}^v$ are also observed. Thus, there is nothing to filter.

In continuous-time models, the only state variables that are not observed are those that appear in the drift of the observation equation. In this case, the problem can be solved analytically or numerically in only the simplest possible cases: models with constant diffusion functions $\sigma^s(S_t, X_t) = \sigma_s$ and linear drifts, $\mu^s = \alpha^s + \alpha_s^s S_t + \alpha_x^s X_t$ and $\mu^x = \beta + \beta_s X_t + \beta_x X$. In this case, the continuous-time Kalman filter applies. In other cases, such as those with state dependent diffusion coefficients or nonlinearities, the filtering problem is solved by set of partial differential equations. These multi-variate partial differential equations are intractable to solve in practice. [7]

2.2.3 Filtering with increasingly fine observations

Beginning with Merton (1980) and continuing with a number of recent papers (see, e.g., Foster and Nelson (1995). Barndorff-Nielson and Shephard (2001) and Andersen, et al.

[7] For example, Pugachev (1987) argues that "there exists no method for the exact solution of the non-linear filtering problem," and moreover that "the numerical solution of the (partial differential) equations is also impossible in practice."

(2002)), a class of filters have been constructed using high frequency data to estimate either the quadratic variation (integrated volatility) or the derivative of the quadratic volatility (spot volatility). These estimators have a number of practical and theoretical advantages. First, if high frequency data is available, the estimators are easy to compute as they just rely on squaring and adding increments. Second, the methods apply equally well in univariate and multivariate models. Third, the methods are nonparametric in the sense that the estimators are independent of the specification for the drift and diffusion. Last, the estimators have attractive asymptotic properties (consistency and asymptotic normality).

These estimators also have some disadvantages from our perspective. First and foremost, if the observed state contains jumps, these methods cannot separate volatility and jump components. The reason is that these methods approximate the quadratic variation which, in the presence of jumps, is a combination of volatility and jumps. By definition, with increasingly fine observations on S_t we have that the quadratic variation estimates

$$\lim_{\Delta_n \to 0} \sum_{j=1}^{n} \left| S_{t_j} - S_{t_{j-1}} \right|^2 = \int_0^t \sigma\left(\Theta, S_s, X_s\right) ds + \sum_{j=1}^{N_t^s} f^s \left(\xi_j^s, S_{\tau_j^s -}, X_{\tau_j^s -} \right)^2 .$$

While computing this is straightforward, there is no way to separate out the diffusive and jump contributions to the quadratic variation. Since there is strong evidence that many time series (exchange rates, equity returns and interest rates) contain jump components, this severely limits the applicability of these estimators. Second, for financial applications, researchers typically work with parametric models and these estimators typically do not use the information contained in the parametric model (due to their nonparametric nature). Third, it is difficult if not impossible to obtain high frequency data for many financial time series. At one extreme, for example, are LIBOR rates which are recorded only at 11:00 a.m. London time. For these reasons, we must develop alternative estimators when the state variables solve jump-diffusions.

2.2.4 Time-Discretization and MCMC

One promising alternative to the previously mentioned methods is Markov Chain Monte Carlo which has been used extensively to estimate latent states in time-discretizations (see, e.g., Eraker, Johannes and Polson (2002)) or latent diffusive variables in a manner consistent with the continuous-time model (Elerian , Shephard and Chib (2001) and Eraker (2001)). Elerian, Shephard and Chib (2001) study the case where there are no latent state variables. In the context of single-factor diffusions, they use standard MCMC techniques and augment the state space by simulating additional data points between observation times. Eraker's (2001) MCMC methodology also augments the state space, but he considers more general models with latent variables such as stochastic volatility.

The key to this approach, which builds on Pedersen (1995), is a high-frequency augmentation of the state space. Our discussion generalizes Elerian, Shephard and Chib (2001) and Eraker (2001) to the case of jump-diffusions. Given the model in (1) $-$ (2) and with observations at points t and $t+\Delta$, the method simulates additional missing data points via an Euler discretization at length $1/M$:

$$S_{t_{j+1}} = S_{t_j} + \mu^s\left(\Theta, S_{t_j}, X_{t_j}\right)/M + \sigma^s\left(\Theta, S_{t_j}, X_{t_j}\right)\varepsilon^s_{t_{j+1}} + f\left(\xi^s_{t_{j+1}}, S_{t_j}, X_{t_j}\right)J^s_{t_{j+1}} \quad (4)$$

$$X_{t_{j+1}} = X_{t_j} + \mu^x\left(\Theta, S_{t_j}, X_{t_j}\right)/M + \sigma^x\left(\Theta, S_{t_j}, X_{t_j}\right)\varepsilon^x_{t_{j+1}} + f\left(\xi^x_{t_{j+1}}, S_{t_j}, X_{t_j}\right)J^x_{t_{j+1}} \quad (5)$$

where $t_j = t+\frac{j}{M}$, $\varepsilon^x_{t_{j+1}} = W^x_{t+\frac{j}{M}} - W^x_{t+\frac{j-1}{M}}$ and $\varepsilon^s_{t_{j+1}} = W^s_{t+\frac{j}{M}} - W^s_{t+\frac{j-1}{M}}$. The jump sizes retain their distributional structure and the jumps times are distributed Bernoulli with probability $\frac{\lambda^j(S_t,X_t)}{M}$. Mikulevicius and Platen (1988) and Liu and Li (2000) prove convergence of this discretization scheme for calculating expectations of sufficiently smooth functions of the state. An alternative scheme that requires simulation of the jump times via an exact solution of the doubly stochastic point process also converges, see Platen and Rebollado (1985). In our setting, the jump probabilities are small and these two schemes are very similar. For future reference, define

$$S^M_{t+1} = \left[S_{t+\frac{\Delta}{M}}, ..., S_{t+\frac{j}{M}\Delta}, ..., S_{t+\frac{M-1}{M}\Delta}\right]$$

as the vector containing the augmented values of S_t,

$$X_{t+1}^{M}=\left[X_t,X_{t+\frac{\Delta}{M}},...,X_{t+\frac{j}{M}\Delta},...,X_{t+\frac{M-1}{M}\Delta}\right]$$

as the vector containing the augmented values of X_t, and define J_t^M and ξ_t^M similarly. We let a capital letter without subscript denote the vector of stacked latent variables (e.g., $X=\left[X_0^M,...,X_{t-1}^M\right]$).

Given the discretization, it is straightforward build an MCMC algorithm which generates samples from the joint posterior distribution of the latent states and parameters, $p\left(\Theta,X,J,\xi,S|S_{0:t}\right)$. This algorithm provides inference consistent with the continuous-time model as M gets large. Using samples from the joint posterior, it is straightforward to form smoothed estimates of the latent states, for example, $p\left[X_t|S_{0:T}\right]$ for $t\leq T$. Moreover, this algorithm, when repeatedly applied, provides filtered estimates of the states and parameters, $E\left[X_t|S_{0:t}\right]$ and $E\left[\Theta|S_{0:t}\right]$. However, sequential implementation of this procedure is computationally intractable. For example, it would not be uncommon for a single MCMC algorithm to take a couple of hours to run.[8] Repeatedly performing this estimation for realistic data sets that are used in practical problems is typically not computationally feasible as the smoothing algorithm must be re-estimated thousands of times. Due to this, we turn to an alternative particle filtering approach.

3 Particle Filtering States and Parameters

In this section, we adapt the particle filtering methodology of Gordon, Salmond and Smith (1993) and its extensions in Carpenter, Clifford, and Fearnhead (1999) and Pitt and Shephard (1999) to state filtering and parameter learning in continuous-time jump-diffusion models. We first describe particle filtering in discrete time, develop extensions that apply

[8] Efficiently programmed in C, the algorithm in Eraker, Johannes and Polson (2002) to estimate jump times, jump sizes, volatility and parameters in the double-jump model of Duffie, et al. (2000) takes over an hour to perform 100,000 iterations.

with the augmented state space and then describe sequential parameter estimation.

3.1 Particle Filtering States and Parameters

We first focus on the problem of state filtering, conditional on a fixed set of parameters Θ. For our application, we use the sampling/importance resampling (SIR) version of the auxiliary particle filter. Our brief discussion of the particle and auxiliary particle filter follows Pitt and Shephard (2001).

To understand the mechanics of the particle filter, consider a discrete time setting where we refer to X_t as the latent variables, S_t the observed data, and $S_{0:t}$ as the vector of observed states up to time t. There are a number of densities associated with the filtering problem which we now define:

$$\begin{aligned} p\left(X_t|S_{0:t}\right) &: \text{ filtering density} \\ p\left(X_{t+1}|S_{0:t}\right) &: \text{ predictive density} \\ p\left(S_t|X_t\right) &: \text{ likelihood} \\ p\left(X_{t+1}|X_t\right) &: \text{ state transition} \end{aligned}$$

Bayes rule links the predictive and filtering densities through the identity

$$p\left(X_{t+1}|S_{0:t+1}\right) = \frac{p\left(S_{t+1}|X_{t+1}\right) p\left(X_{t+1}|S_{0:t}\right)}{p\left(S_{t+1}|S_{0:t}\right)}$$

where

$$p\left(X_{t+1}|S_{0:t}\right) = \int p\left(X_{t+1}|X_t\right) dP\left(X_t|S_{0:t}\right).$$

The key to particle filtering is an approximation of the (continuous) distribution of the random variable X_t conditional on $S_{0:t}$ by a discrete probability distribution, that is, the distribution $X_t|S_{0:t}$ is approximated by a set of particles, $\left\{X_t^{(i)}\right\}_{i=1}^{N}$ with probability $\pi_t^1, ..., \pi_t^N$. By assuming the distribution is approximated with particles, we can estimate the filtering and predictive densities via: ($\widehat{p}$ refers to an estimated density)

$$p^N\left(X_t|S_{0:t}\right)=\sum_{i=1}^{N}\delta_{X_t^{(i)}}\pi_t^i$$
$$p^N\left(X_{t+1}|S_{0:t}\right)=\sum_{i=1}^{N}p\left(X_{t+1}|X_t^{(i)}\right)\pi_t^i,$$

where δ is the Dirac function. As the number of particles increases, the accuracy of the discrete approximation to the continuous random variable improves. When combined with the conditional likelihood, the filtering density at time $t+1$ is defined via the recursion:

$$p^N\left(X_{t+1}|S_{0:t+1}\right)\propto p\left(S_{t+1}|X_{t+1}\right)\sum_{i=1}^{N}p\left(X_{t+1}|X_t^{(i)}\right)\pi_t^i.$$

The key to particle filtering is developing an efficient algorithm to propagated the particles forward from time t to time $t+1$.

There are a number of different methods to that can be used to update or propagate the particles. For example, importance sampling, rejection sampling and MCMC are all possible methods. In this paper, we use importance sampling (Smith and Gelfand (1992), Gordon, Salmond and Smith (1993)) to propagate the particles. To understand how the importance sampler works, we can view $p^N\left(X_{t+1}|S_{0:t}\right)$ as the prior and $p\left(S_{t+1}|X_{t+1}\right)$ as the likelihood:

$$p^N\left(X_{t+1}|S_{0:t+1}\right)\propto\underbrace{p\left(S_{t+1}|X_{t+1}\right)}_{Likelihood}\underbrace{\sum\nolimits_{i=1}^{N}p\left(X_{t+1}|X_t^{(i)}\right)\pi_t^i}_{Prior}.$$

The algorithm draws, $\left\{X_{t+1}^{(i)}\right\}_{i=1}^N$ from prior distribution and then re-samples $\left\{X_{t+1}^{(i)}\right\}_{i=1}^N$ with weights $\omega_i\propto p\left(S_{t+1}|X_{t+1}^{(i)}\right)$. Alternatively, we could use a pure rejection sampling algorithm which, if the acceptance probabilities are high, will be faster and provides a direct draw from the the target. The auxiliary particle filter, which we implement, is a straightforward extension of the particle filter and is described in Pitt and Shephard (1999). The auxiliary particle filter "peaks" forward via an initial resampling step and then propagates these higher likelihood samples forward with the particle filter.

As an example, consider a time-discretization ($M = 1$) of the SV model:

$$S_{t+1} = \log\left(\frac{P_{t+1}}{P_t}\right) = \mu + \sqrt{V_t}\left(W^s_{t+1} - W^s_t\right)$$

where

$$V_{t+1} = V_t + \kappa_v\left(\theta_v - V_t\right) + \sigma_v\sqrt{V_t}\left(W^v_{t+1} - W^v_t\right) \tag{6}$$

In this case, the state variable is $X_t = V_{t-1}$ (the timing convention will be clear below when we augment the state space). To begin, assume that we have a sample $\left\{(V_{t-1})^i\right\}_{i=1}^N$ from $p\left(V_{t-1}|S_{0:t}\right)$. We propagate these samples forward by drawing a Brownian increment in equation (6) which results in a sample $\left\{(V_t)^i\right\}_{i=1}^N$. Given a new observation, S_{t+1} we resample $\left\{(V_t)^i\right\}_{i=1}^N$ with weights proportional to $p\left(S_{t+1}|V_t^i\right) = \phi\left(\mu, (V_t)^i\right)$ where ϕ is the normal density. This provides the particles for the next iteration. Provided the number of particles N is large enough, this will deliver an accurate approximation to the filtering distribution.

Next, we show how to adapt the particle filter to handle jump components. Consider a time-discretization of the SVCJ model mentioned above at frequency $M = 1$. In this model, the likelihood and evolution are given by:

$$S_{t+1} = \log\left(\frac{P_{t+1}}{P_t}\right) = \mu + \sqrt{V_t}\left(W^s_{t+1} - W^s_t\right) + \xi^s_{t+1}J_{t+1}$$

$$V_{t+1} = V_t + \kappa_v\left(\theta_v - V_t\right) + \sigma_v\sqrt{V_t}\left(W^v_{t+1} - W^v_t\right) + \xi^v_{t+1}J_{t+1}$$

where ξ^s_t is normally distributed, ξ^v_t is exponentially distributed and J_t is Bernoulli with probability λ. In this model, since the arrivals and sizes are i.i.d., their transitions are state independent which makes simulating jump time and size particle easy. Moreover, the likelihood function is just a mixture of two normals depending on whether $J_{t+1} = 1$ or 0. The state vector at time t is $X_t = (V_{t-1}, \xi^s_t, \xi^v_t, J_t)$ and assume that we have a samples from $\left(V_{t-1}, \xi^s_t, \xi^v_t, J_t\right)|S_{0:t}$,

$$\left\{(V_{t-1})^i, (\xi^s_t)^i, (\xi^v_t)^i, (J_t)^i\right\}_{i=1}^N.$$

Again, to propagate the particles forward, we first draw Brownian increments, jump times and jump sizes. Since the jump times and sizes i.i.d., these draws are simple: $(J_{t+1})^i \sim$

$Ber(\lambda)$, $\left(\xi_{t+1}^s\right)^i \sim N\left(\mu_\xi, \sigma_\xi^2\right)$ and $\left(\xi_{t+1}^v\right)^i \sim \exp\left(\mu_v\right)$. Given the sample,

$$\left\{ (V_t)^i, (\xi_t^s)^i, (\xi_t^v)^i, (J_{t+1})^i \right\}_{i=1}^N,$$

again evaluate the density weights:

$$p\left(S_{t+1} | (V_t)^i, \left(\xi_{t+1}^v\right)^i, \left(\xi_{t+1}^v\right)^i, (J_{t+1})^i\right) = \phi\left(\mu + J_{t+1}\xi_{t+1}, (V_t)^i\right)$$

and resample as before. At this stage, it is also clear how to deal with state dependent jump intensities. Pan (2002) and Andersen, Benzoni and Lund (2002) consider models where the jump intensity depends on the V_t: $\lambda_t = \lambda_0 + \lambda_1 V_t$. In this case, the jump time update depends on the current $(V_{t-1})^i$. Kim and Shephard (1998) and Chib, Nardari and Shephard (2002) have used particle filtering to compute likelihood functions in discrete-time settings.

While we use importance sampling, it is always important to consider alternatives such as rejection sampling or MCMC. In general, no one of these three algorithms is generically preferred to the others, although in certain cases one may provide advantages in terms of computational time or efficiency. For example, if one can bound the likelihood as a function of X_t, rejection sampling can lead to substantial improvements in certain cases. However, this is not generically the case for the models that we consider. In the discretization of the SVJ model above, we have that

$$f\left(S_{t+1} | V_t, \xi_{t+1}, J_{t+1}\right) = \frac{1}{\sqrt{2\pi V_t}} \exp\left(-\frac{1}{2}\left(\frac{S_{t+1} - \mu - \xi_{t+1}J_{t+1}}{\sqrt{V_t}}\right)^2\right)$$

whose maximum does not exist (set $J_{t+1} = 1$, $\xi_{t+1} = S_{t+1} - \mu$ and drive V_t to zero). This is a general problem in many mixture models (see Kiefer (1978)). In order to check that our importance sampler works well, we have been careful to run extensive simulations.

3.2 Particle Filtering with missing data

This section extends particle filtering to the case $M > 1$, where additional observations are simulated at equally spaced intervals. For notational convenience, we focus on the case where the observable states do not appear on the right hand side (this will be the case with our equity price applications below). To develop intuition, we first consider the pure diffusion case,

$$dS_t = \mu^s(X_t)\,dt + \sigma^s(X_t)\,dW_t^s \tag{7}$$

$$dX_t = \mu^x(X_t)\,dt + \sigma^x(X_t)\,dW_t^x. \tag{8}$$

We have suppressed any parametric dependence in the drifts and diffusions.

A discretization with M points between observations implies that

$$S_{t+1} = S_t + M^{-1}\sum_{j=1}^{M}\mu^s\left(X_{t+\frac{j-1}{M}}\right) + \sum_{j=1}^{M}\sigma^s\left(X_{t+\frac{j-1}{M}}\right)\left(W^s_{t+\frac{j}{M}} - W^s_{t+\frac{j-1}{M}}\right)$$

where

$$X_{t+\frac{j}{M}} = X_{t+\frac{j-1}{M}} + \mu^x\left(X_{t+\frac{j-1}{M}}\right)M^{-1} + \sigma^x\left(X_{t+\frac{j-1}{M}}\right)\left(W^v_{t+\frac{j}{M}} - W^v_{t+\frac{j-1}{M}}\right),$$

$W^s_{t+\frac{j}{M}} - W^s_{t+\frac{j-1}{M}} \sim N(0, M^{-1})$ and $W^v_{t+\frac{j}{M}} - W^v_{t+\frac{j-1}{M}} \sim N(0, M^{-1})$. Again, recall the definition of $X^M_{t+1} = \left[X_t, X_{t+\frac{\Delta}{M}}, ..., X_{t+\frac{j}{M}\Delta}, ..., X_{t+\frac{M-1}{M}\Delta}\right]$. The distribution of the observable states conditional on X^M_{t+1} is

$$S_{t+1}|X^M_{t+1}, S_t \sim N\left(S_t + \frac{1}{M}\sum_{j=1}^{M}\mu^s\left(X_{t+\frac{j-1}{M}}\right), \frac{1}{M}\sum_{j=1}^{M}\sigma^s\left(X_{t+\frac{j-1}{M}}\right)^2\right). \tag{9}$$

To start the algorithm, suppose that we have samples $\left\{\left(X_t^M\right)^i\right\}_{i=1}^N$ from $p\left(X_t^M|S_{0:t}\right)$. Next, we take $\left\{\left(X_t^M\right)^i\right\}_{i=1}^N$ and simulate each of these particles forward using the stochastic differential equation discretized at frequency $\frac{1}{M}$ to obtain $\left\{\left(X_{t+1}^M\right)^i\right\}_{i=1}^N$. Given these samples, we generate the likelihood weights from (9) and reweight as before. This generates a sample, $\left\{\left(X_{t+1}^M\right)^i\right\}_{i=1}^N$ from $p\left(X_{t+1}^M|S_{0:t}\right)$. As was the case in the previous section, incorporating jumps in either X_t or S_t is straightforward if the jump size distribution and arrival intensities do not depend on S_t. If the jump size distribution or intensity depends on S_t we have to simulate values of S_t in between observations. This algorithm is closely related to the approach of Durham and Gallant (2001) and Pitt (2002) who use particle filters for maximum likelihood estimation in diffusion based models.

In our equity price applications, the observed state variable is the continuously compounded returns and in this case the price or return does not appear on the right hand side of the SDE. In other cases we may need to simulate data points for the observed variables between time t and $t+1$: $S_t^M = \left\{ S_{t+\frac{j}{M}} \right\}_{j=1}^{M-1}$. When this occurs, an Euler discretization at frequency M implies that

$$S_{t+1} = S_t + \sum_{j=1}^{M} \mu^s \left(S_{t+\frac{j-1}{M}}, X_{t+\frac{j-1}{M}} \right) M^{-1} + \sum_{j=1}^{M} \sigma^s \left(S_{t+\frac{j-1}{M}}, X_{t+\frac{j-1}{M}} \right) \left(W^s_{t+\frac{j}{M}} - W^s_{t+\frac{j-1}{M}} \right)$$

where

$$X_{t+\frac{j}{M}} = X_{t+\frac{j-1}{M}} + \mu^x \left(X_{t+\frac{j-1}{M}} \right) M^{-1} + \sigma^x \left(X_{t+\frac{j-1}{M}} \right) \left(W^v_{t+\frac{j}{M}} - W^v_{t+\frac{j-1}{M}} \right).$$

The particle filter still applies in this case by propagating the S_t^M and X_t^M forward. Note now, however, that the likelihood is more complicated since

$$\begin{aligned} & p\left(S_{t+1}|S_t, S_t^M, X_t^M\right) = \\ & p\left(S_{t+1}|X_{t+\frac{M-1}{M}}, S_{t+\frac{M-1}{M}}\right) \prod_{i=2}^{M} p\left(S_{t+\frac{i}{M}}|X_{t+\frac{i-1}{M}}, S_{t+\frac{i-1}{M}}\right) p\left(S_{t+\frac{1}{M}}|X_t, S_t\right). \end{aligned}$$

In this case, we have found that the gains to auxiliary particle filtering can be quite substantial.

3.3 Convergence of Particle Filters

The previous sections discussed particle filters for a given N and M. In this section we discuss convergence issues that arise when N and M are large.

First, assume that the $M = 1$ Euler discretization of the continuous-time model is the true model. In this case, we can appeal to a number of limiting results that are summarized in Crisan and Doucet (2002). If we denote the particle approximation to the filtering density for a fixed N as $p^N(X_t|S_{0:t})$, then we are interested in the behavior of this density for large N. Under sufficient regularity on the state transition and the likelihood, the particle filter is consistent in the sense that $\lim_{N\to\infty} p^N(X_t|S_{0:t}) = p(X_t|S_{0:t})$, where $p(X_t|S_{0:t})$ is the optimal filter. The regularity conditions require that the state variable transition density is Feller continuous and the likelihood is positive, bounded and continuous.

Second, consider the more general case with $M > 1$. If we let $p^{M,N}(X_t|S_{0:t})$ denote the particle filter approximation for a given M and N and $p^M(X_t|X_{t-1})$ the state transition for a given M, then we are interested in the convergence of the filter for large M and N. From the previous results, we know that for a fixed M, the particle filter is consistent and thus we need to focus on the convergence of the approximate transition density $p^M(X_t|X_{t-1})$ to the true transition density, $p(X_t|X_{t-1})$. While it is straightforward to provide convergence of Euler discretizations for computing expectations of smooth functions of the state, $E[f(X_T)|X_t = x]$ (Kloeden and Platen (2001)), pointwise convergence of the density function is a more difficult as it applies in the case where the function f is an indicator function (and is not smooth nor continuous).

In the case of diffusions, Bally and Talay (1996) prove pointwise convergence of $p^M(X_t|X_{t-1})$ using the Malliavin calculus. Brandt and Santa-Clara (2002) use this result to prove convergence of simulated maximum likelihood estimators (see also the appendix in Pedersen (1995)). Jacod and Del Moral (2001) and Jacod, Del Moral and Protter (2002) prove that the pointwise convergence of $p^M(X_t|X_{t-1})$ to $p(X_t|X_{t-1})$ implies that the particle filter is consistent provided that M increases slower than N (for example, $M = \sqrt{N}$). This result is similar to the limiting behavior of parametric and nonparametric estimators where the index that approximates the true function increases at a slower rate than new data arrives (see, e.g., Aït-Sahalia (2002), Gallant and Long (1995)).

In the context of jump-diffusions, Rebollado and Platen (1985), Mikulevicius and Platen (1988) and Liu and Li (2002) prove convergence of the Euler scheme for smooth functions, f. To our knowledge, we know of no analog to Bally and Talay (1996) for jump-diffusions or for Lévy processes. Although we conjecture that transitions generated by the Euler scheme in jump-diffusion models converge, it is an open question for future research.

3.4 Parameter Learning

The particle filter can also be adapted to perform on-line parameter estimation. There are two ways to do this, both of which can be used in conjunction with our state space expansion via high frequency augmentation. The first approach, discussed in Gordon, Salmond and Smith (1993), is a direct particle filtering approach. The advantage of this approach is that it applies in virtually all models, the disadvantage is that the algorithm often degenerates in that the particles can cluster. One way to avoid this issue is to use an extremely large number of particles. The second approach, which we discuss, was developed by Storvik (2002). While it applies only in certain cases, it provides a drastic improvement over naive particle filtering when it can be applied.

In the presence of parameter uncertainty, we need to update the parameter and state particles, $\left\{\Theta^i, \left(X_t^M\right)^i\right\}_{i=1}^N$, which are samples from $p\left(\Theta, X_t^M | S_{0:t}\right)$. If we there exists a sufficient statistic $T_t^M = T\left(S_{0:t}, X_t^M\right)$ for updating Θ, we can draw new particles by simulating forward $p\left(\Theta | T_t^M\right)$ and $p\left(X_t^M | \Theta, S_{0:t}\right)$. Given the new particles, we resample $\left\{\Theta^i, \left(X_{t+1}^M\right)^i\right\}_{i=1}^N$ with weights $p\left(S_{t+1} | \Theta, X_{t+1}^M\right)$. The key requirement to this algorithm is that we can represent the parameter posterior conditional on observations up to time t and any latent variables as a function of a low dimensional sufficient statistic. If this is not the case, much of the computational gain will be lost. We provide an example of the Storvik algorithm, with an augmented state space below.

3.5 Additional extensions and applications

There are additional applications and extensions of our methodology. Essentially anything the particle filter can do with discrete-time models, our state-augmented methodology can do for continuous-time models. Pitt and Shephard (1999) review a number of other extensions, all of which apply in our setting. We briefly review two of these applications.

First, we can use the particle filtering methodology to perform model specification. For example, in a discrete-time SV

$$S_{t+1} = \log\left(\frac{P_{t+1}}{P_t}\right) = \mu + \sqrt{V_t}\left(W^s_{t+1} - W^s_t\right)$$
$$V_{t+1} = V_t + \kappa_v\left(\theta_v - V_t\right) + \sigma_v\sqrt{V_t}\left(W^v_{t+1} - W^v_t\right)$$

the discrete-time particle filter provide a specification test via the empirical distributions of the assumed normal Brownian increments. In principle, a test at the $M = 1$ discretization interval could erroneously reject a model because the discrete-time model does not properly account for the fact that the true continuous-time model has fatter tails. Our high frequency augmentation in conjunction with particle filtering allows us to address this issue (see also Elerian, Shephard and Chib (2001)) as we have Monte Carlo samples of the errors.

Second, particle filtering can be used to compute the likelihood function. For example, in the SV model, the likelihood function for the continuous-time model is given by:

$$f\left(S_1, ..., S_T|\Theta\right) = \prod_{t=1}^{T}\int p\left(S_t|S_{t-1}, V_{t-1}, \Theta\right)p\left(V_{t-1}|S_{1:t-1}, \Theta\right)dV_{t-1}.$$

Direct maximum likelihood estimation is not feasible because neither $p\left(S_t|S_{t-1}, V_{t-1}, \Theta\right)$ nor $p\left(V_{t-1}|S_{1:t-1}, \Theta\right)$ are known in closed form. As noted by Durham and Gallant (2001), particle filtering provides a computationally tractable way of computing $p\left(V_{t-1}|S_{1:t-1}, \Theta\right)$ while $p\left(S_t|S_{t-1}, V_{t-1}, \Theta\right)$ can be computed using the simulation method of Pedersen (1995). If parameter estimation is the sole goal, the MCMC approach of Elerian, Shephard and Chib (2001) and Eraker (2001) may be more efficient (see, e.g., Eraker (2002b)). This extends the existing class of models to jump-diffusions on which simulated MLE can be efficiently performed (see, Pitt and Shephard (1999), Durham and Gallant (2002), Pitt (2002), Durham (2002) and Chib, Nardari and Shephard (2002)).

Third, it appears possible to extend many of these results to more general Lévy processes. Shephard and Barndorff-Nielson (2001) use a particle filter on integrated volatility in OU model driven by Lévy shocks and it appears that particle filtering will apply in models driven by different types of Lévy process. The case of general Lévy processes is more difficult to address because it is often impossible to directly simulated the increments of a general Lévy driving process. In this case, it is common to truncate the Lévy process by neglecting small jumps during the simulation. While this method is attractive from a simulation viewpoint (see Rubenthaler (2001)), it is less attractive from a modeling viewpoint. If the small jumps are truncated, the resulting process contains only "large" jumps, and thus the Lévy process looks more and more like a combination of Brownian motion and a compound Poisson process which is a special case of the models that we consider.

4 Models and Applications

As an example of our methodology, we consider special cases of the double jump model of Duffie, Pan and Singleton (2000):

$$d\log(S_t) = \mu dt + \sqrt{V_{t-}}dW_t^s + d\left(\sum_{j=1}^{N_t} \xi_{\tau_j}^s\right)$$

$$dV_t = \kappa_v(\theta_v - V_t)\,dt + \sigma_v\sqrt{V_{t-}}dW_t^v + d\left(\sum_{j=1}^{N_t} \xi_{\tau_j}^v\right)$$

where W_t^s and W_t^v are Brownian motions, $\xi^s = \mu_s + \rho_J \xi^v + \sigma_s \varepsilon$, $\varepsilon \sim N(0,1)$, $\xi^v \sim \exp(\mu_v)$ and $N_t \sim Poi(\lambda t)$. We let SV denote the special case with $\lambda = 0$ and SVJ the case with $\mu_v = 0$ and SVCJ denotes the general model. This model is an excellent example for two reasons. First, Eraker, Johannes and Polson (2002), Chernov, Gallant, Ghysels and Tauchen (2002) and Eraker (2002a) find support for this specification for modeling equity returns. This model is useful for practical applications because option prices and portfolio rule are known in closed form up to the numerical solution of an ordinary differential equation. Second, this model contains a very general factor structure. It includes an unobserved diffusive state (V_t), unobserved jumps in the observed state (ξ_t^s) and unobserved jumps in the unobserved state (ξ_t^v).

With observations at time t and $t+1$, the time-discretization of this model at frequency M^{-1} takes the form:

$$\log\left(\frac{S_{t+1}}{S_t}\right) = \mu + \sum_{j=1}^{M} \sqrt{V_{t+\frac{j-1}{M}}}\left(W^s_{t+\frac{j}{M}} - W^s_{t+\frac{j-1}{M}}\right) + \sum_{j=1}^{M} \xi^s_{t+\frac{j}{M}} J^s_{t+\frac{j}{M}}$$

where

$$V_{t+\frac{j}{M}} = V_{t+\frac{j-1}{M}} + \kappa_v\left(\theta_v - V_{t+\frac{j-1}{M}}\right)\frac{1}{M} + \sqrt{V_{t+\frac{j-1}{M}}}\left(W^v_{t+\frac{j}{M}} - W^v_{t+\frac{j-1}{M}}\right) + \xi^v_{t+\frac{j}{M}} J^s_{t+\frac{j}{M}}$$

where $P\left[J^s_{t+\frac{j}{M}} = 1\right] = \lambda M^{-1}$, $W^s_{t+\frac{j}{M}} - W^s_{t+\frac{j-1}{M}} \sim N\left(0, M^{-1}\right)$, $W^v_{t+\frac{j}{M}} - W^v_{t+\frac{j-1}{M}} \sim N\left(0, M^{-1}\right)$ and the jumps retain their distributional structure. As mentioned earlier, since we use log-returns, the equity price does not appear on the right hand side of the time-discretization which simplifies our algorithm as we do not need to augment the state space with additional S_t states.

4.1 Filtering Results

This section presents particle filtering results using both simulated and real data.

4.1.1 SV Model

For the SV model, we use the parameter estimates from Eraker, Johannes and Polson (2002): $\mu = 0.05$, $\theta_v = 0.9$, $\kappa_v = 0.02$ and $\sigma_v = 0.15$. To understand the nature of the discretization bias, we first perform a simulation study to determine the impact of M on the performance of the filter. To do this, we compare the performance of the filter for a given M to the true filter. Since it is not possible to compute the true filter analytically, we regard $M = 100$ as the true filter, in the sense that it is devoid of discretization error. We simulate 100 data sets on the daily, weekly and monthly basis by simulating 500 data points per day and then sampling the process daily, weekly or monthly. We then run the filter with $M = 1, 2, 5$,10, and 100 for $N = 100,000$.

Figures 1, 2 and 3 provide graphical summaries of the simulation. Figure 1 (daily), Figure 2 (weekly) and Figure 3 (monthly) show how sampling frequency affects the accuracy of the approximate filters. In each figure, the top panel plots absolute returns over the interval (to get a sense of observed variability); the middle panels plot the true simulated variance (dots) and the posterior filtered means for $M = 1, 2, 5$ and 10; the bottom plots the difference between the posterior mean for a given M and the true posterior mean. Table 1 summarizes the simulation via RMSE (square root of the mean-squared-error) and MAE (mean absolute error).

Table 1: Simulation results for the SV model.

	Monthly		Weekly		Daily	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
M=1	0.271	0.242	0.088	0.069	0.033	0.025
M=2	0.102	0.089	0.028	0.022	0.026	0.018
M=5	0.037	0.030	0.023	0.017	0.025	0.017
M=10	0.021	0.017	0.025	0.017	0.025	0.017
M=25	0.017	0.013	0.022	0.015	0.024	0.016

For daily data, it is clear that the $M = 1$ filter is not as accurate as the $M = 2, 5$ and 10 filters, but the differences are quite small. For the parameters given above, the daily variance is on average slightly less than 1 and the error in the filters is typically about 1 to 3 percent of the variance for $M = 1$ and less than 1 percent for $M = 2, 5$ and 10. This should not be a surprise as conditional equity returns and variance increments are approximately conditionally normal in the SV model over short time intervals (Das and Sundaram (1998)). On the other hand, at longer intervals such as weekly or monthly, the differences can be quite large for $M = 1$ but they decrease rapidly for $M = 2, 5$ and 10. Since conditional nonnormalities for typical parameters are maximized in the SV model horizons close to one month (Das and Sundaram (1998)), the greatest impact of filling in the missing data will occur at these frequencies. In general, $M = 1$ estimator underestimates volatility in periods of high volatility and overestimates volatility in periods of low volatility. This is due to the Gaussian approximation of the fat-tailed non-central χ^2 density.

As an example of the filters performance on real data, Figure 4 shows the $(10, 50, 90)$ percent quantiles of the filtering distribution for spot volatility using S&P 500 data from 1980-2000 for $M = 25$ and $N = 50,000$. The thick solid line is the filtered posterior median and the bands (thinner lines) are the 10 and 90 percent quantiles of the filtering distribution. The difference between the upper and lower quantiles of the distribution is typically about 5 percent. It is important to note that even with a large data set (over 5000 observations), the algorithm does not degenerate in the sense that the particles get stuck in a certain region of the state space.

4.1.2 SVJ Model

For the SVJ model, we use parameters estimated from Eraker, Johannes and Polson (2002): $\mu = 0.0496$, $\theta_v = 0.8152$, $\kappa_v = 0.0198$, $\sigma_v = 0.0954$, $\mu_y = -2.5862$, $\sigma_y = 4.0720$ and $\lambda_y = 0.0060$. As in the SV model, we performed detailed simulations on the performance of the algorithm for various time intervals. While the auxiliary particle filter performed well in identifying jumps at a daily frequency, it was not able to identify many of the jumps at the weekly or monthly frequency. This is not a surprise. Over longer time intervals, the variance contribution of the diffusion component increases while the jump component is constant. Effectively, the signal regarding the jumps is dwarfed by the noise from the diffusion coefficient. Since the simulation results with daily data were similar to the SV model, they did not warrant separate reporting.

To get a sense of the simulation results, Figure 5 displays a randomly selected simulation of 2000 days. The daily jump intensity of 0.006 implies that about 1.5 jumps arrive per

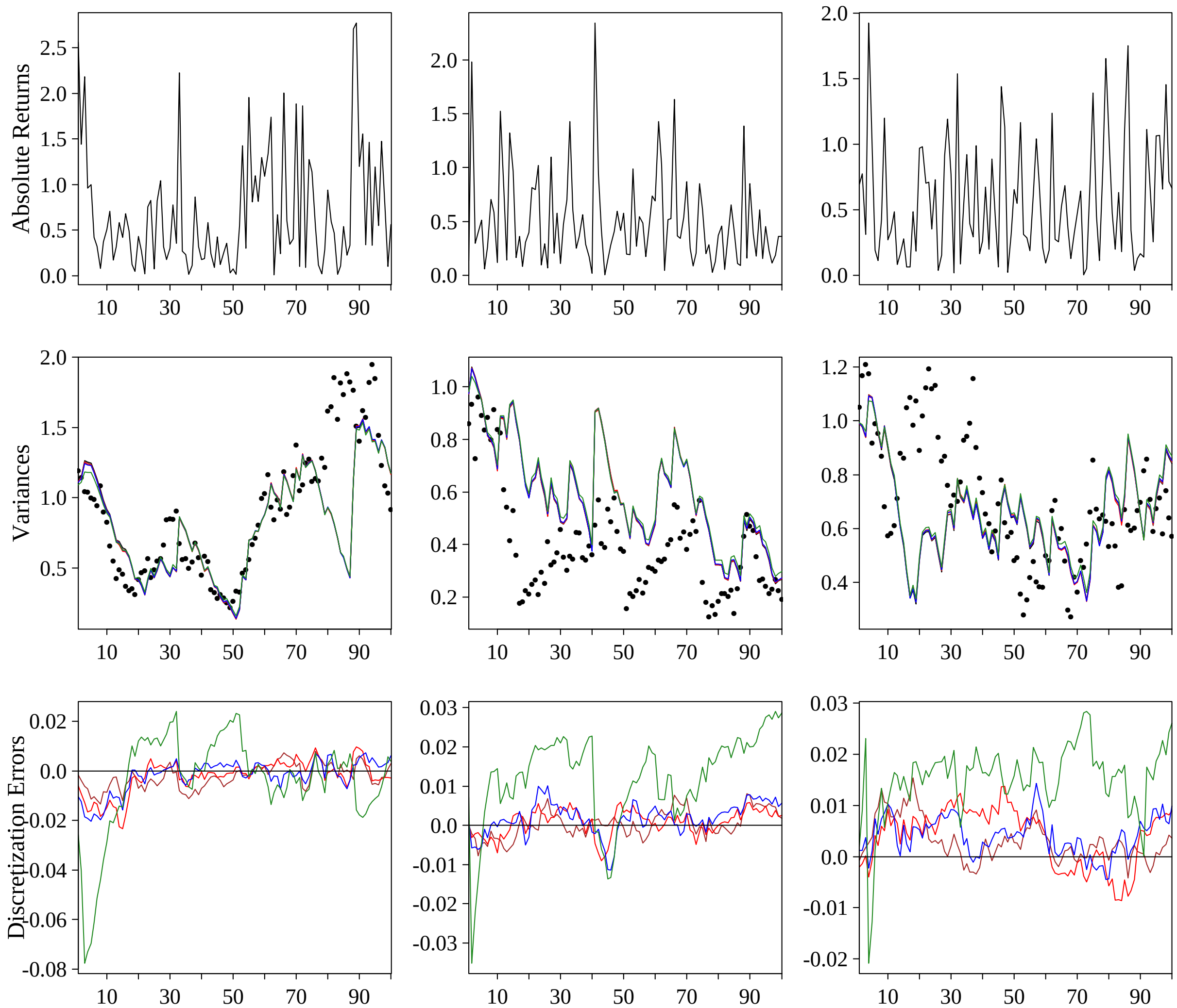

Figure 1: Simulation examples for the SV model with daily data. The top panel plots the absolute daily returns, the middle panel plots the true variances (dots) and filtered means for (M=1, 2, 5, 10) and the bottom panel plots the difference between the true filter (M=100) an the filtered means for M=1, 2, 5, 10.

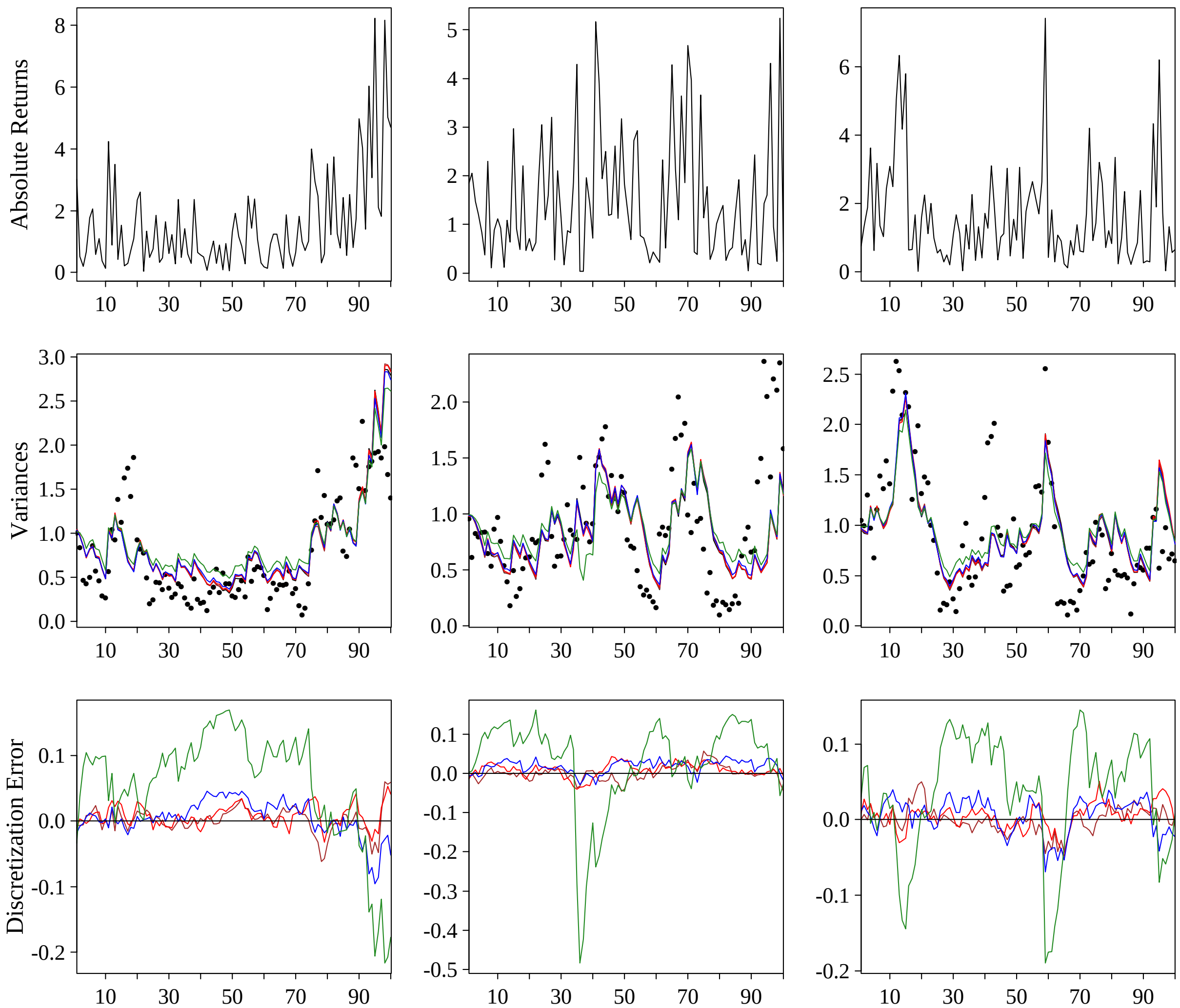

Figure 2: Simulation results for the SV model with weekly data. The top panel plots the absolute weekly returns, the middle panel plots the true variances (dots) and filtered means for (M=1, 2, 5, 10) and the bottom panel plots the difference between the true filter (M=100) an the filtered means for M=1, 2, 5, 10.

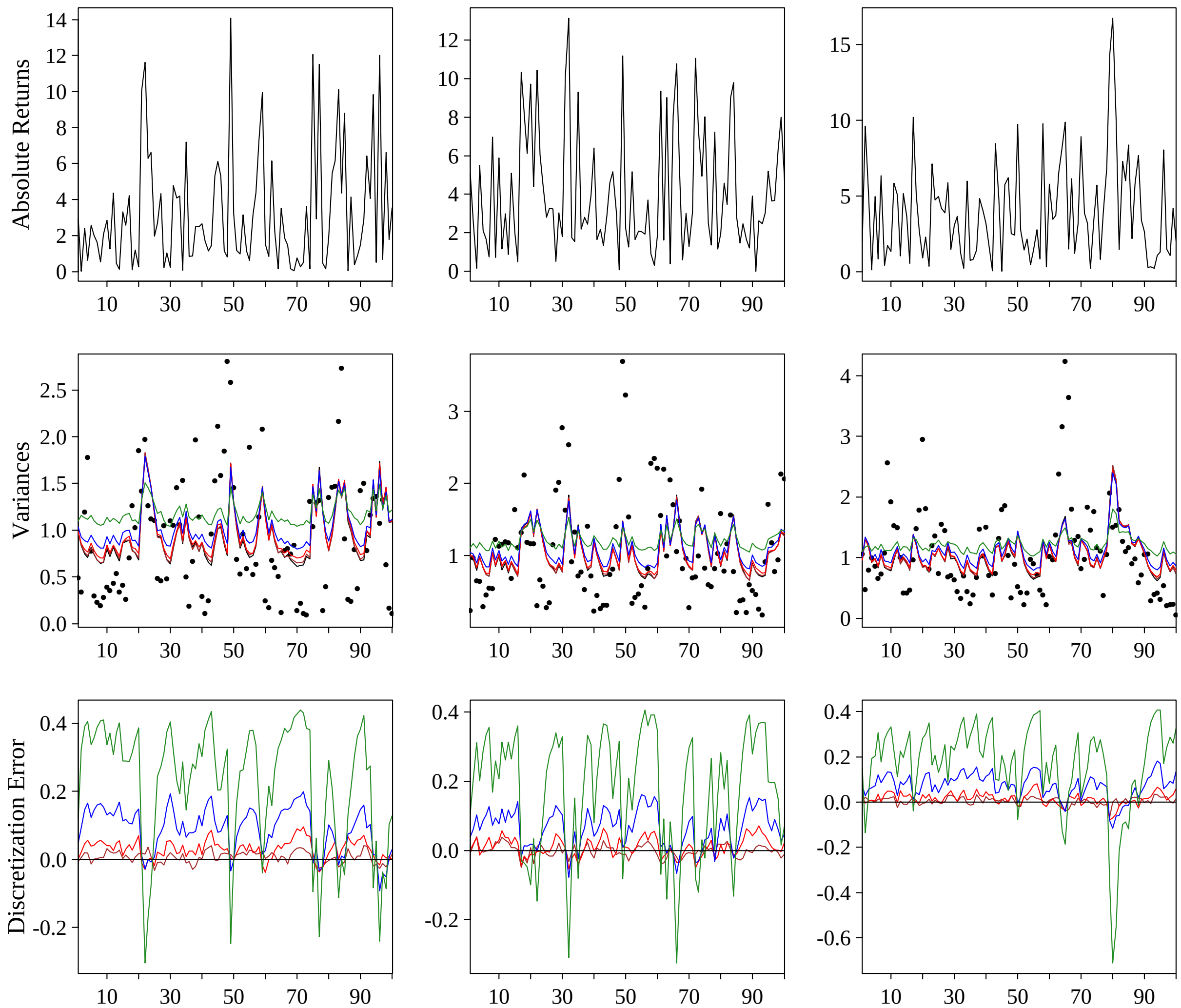

Figure 3: Simulation examples for the SV model with monthly data. The top panel plots the absolute monthly returns, the middle panel plots the true variances (dots) and filtered means for (M=1, 2, 5, 10) and the bottom panel plots the difference between the true filter (M=100) an the filtered means for M=1, 2, 5, 10.

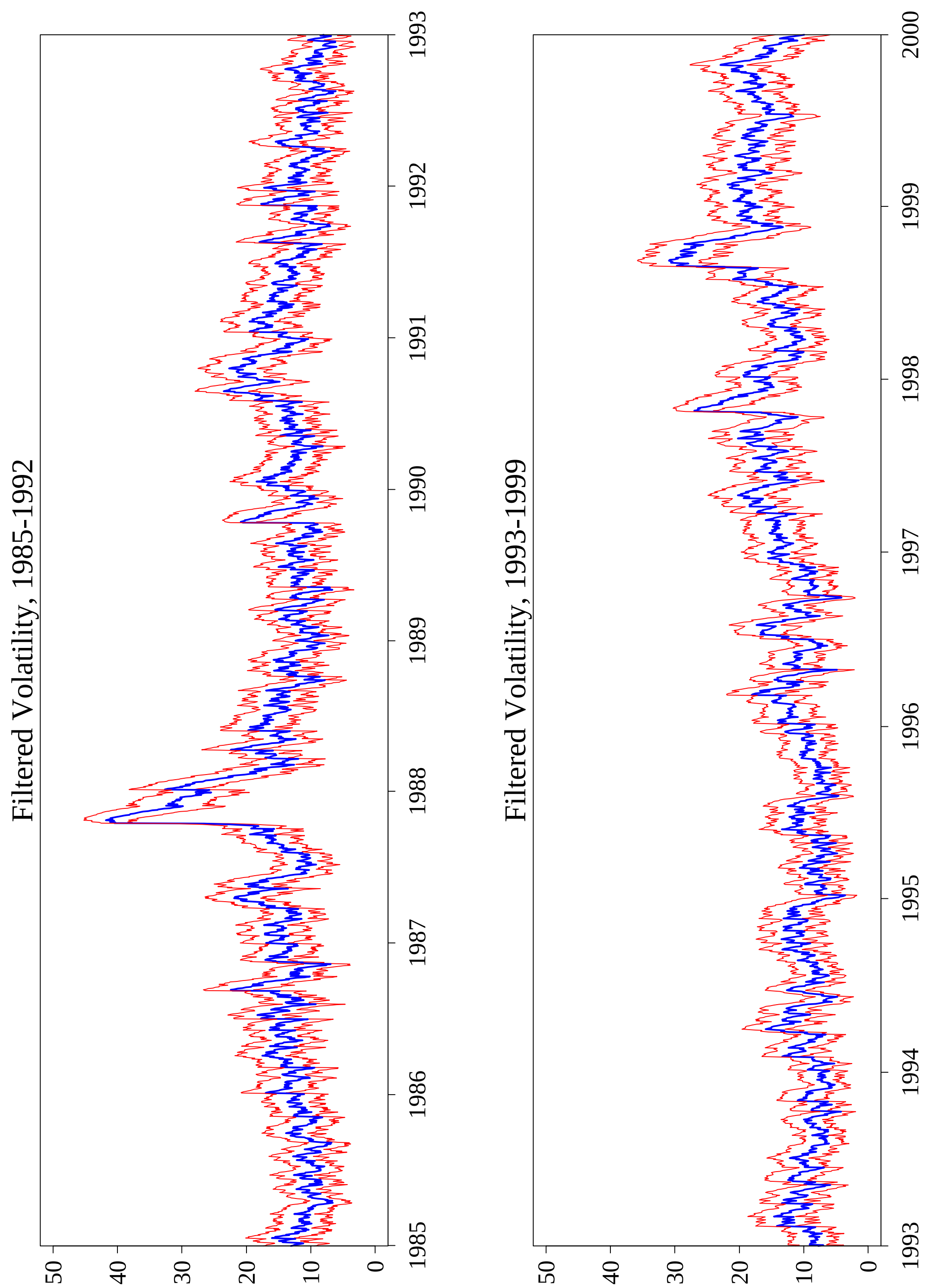

Figure 4: Filtered volatilities for the SV model, from 1985-2000 for M=10 and N=50,000.

year. The top panel displays the simulation returns; the second panel displays the true volatility path (dots), the $(10, 50, 90)$ percent quantiles; the third panel displays the true jump times (dots) and estimated jump probabilities; and the bottom panel displays the true jump sizes (dots) and the filtered jump size estimates. Note that the particle filter does an excellent job identifying the large jumps, but it does miss many of the smaller jumps. Again, this is expected. In the simulations, daily volatility is slightly less than 1 percent which implies that it is common to see diffusive moves of ± 3 percent. Since the distribution of the jump sizes is $N(-2.5, 4^2)$, more than half of the jumps will be be within the normal diffusive volatility range. Because of this, it is impossible for the filter to identify these small jumps. As a matter of modeling, this implies that truncated jump distributions such as a right truncated normal distribution might be more appropriate.

Figure 6 displays filtered volatility, jump times and sizes using daily S&P 500 returns from 1980-2000 for $M = 10$ and $N = 50,000$. For simplicity, we do not report quantiles of the filtering distribution. Note that, unlike the SV model where volatility increased from 20 to 40 percent during the crash, the SVJ model attributes the Crash to a jump and therefore volatility only modestly increases. There are a large number of jumps estimated after the Crash in 1987 which, given the i.i.d. arrival process indicates potential misspecification.

4.1.3 SVCJ Model

Figure 7 displays the filtered results for the SVCJ model for the S&P 500 for $M = 10$ and $N = 50,000$ using the parameter estimates from Eraker, Johannes and Polson (2002). Again, the algorithm is able to identify major movements as jumps. The addition of jumps in the variance removes much of the misspecification of the SVJ model as the jump times in periods of market stress (1987, 1997 and 1998) are not clustered. This result is similar to the smoothed estimates obtained by Eraker, Johannes and Polson (2002) which is not surprising given the parameters used. Also, in the SVCJ, model nearly all of the jumps in the return process are negative.

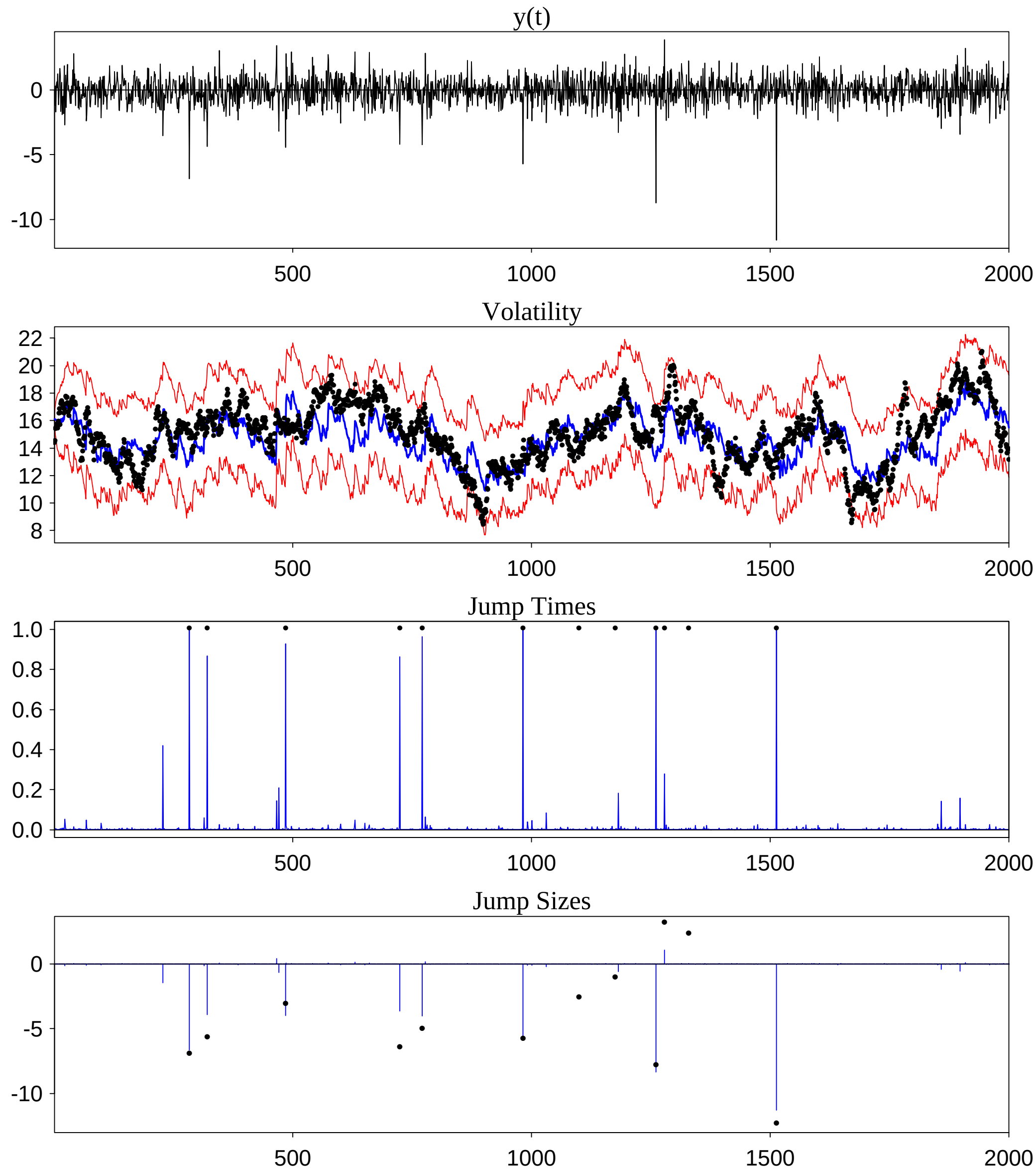

Figure 5: Particle filtering results for the SVJ model. The top panel displays 2000 daily returns simulating using the parameter estimates in Eraker, Johannes and Polson (2002); the second panel displays true volatility (dots) and quantiles (10,50,90) of the filtering distributions; the third panel displays the true jump times (dots) and the filtered jump times; and the bottom panel displays true jump sizes (dots) and filtered jump sizes.

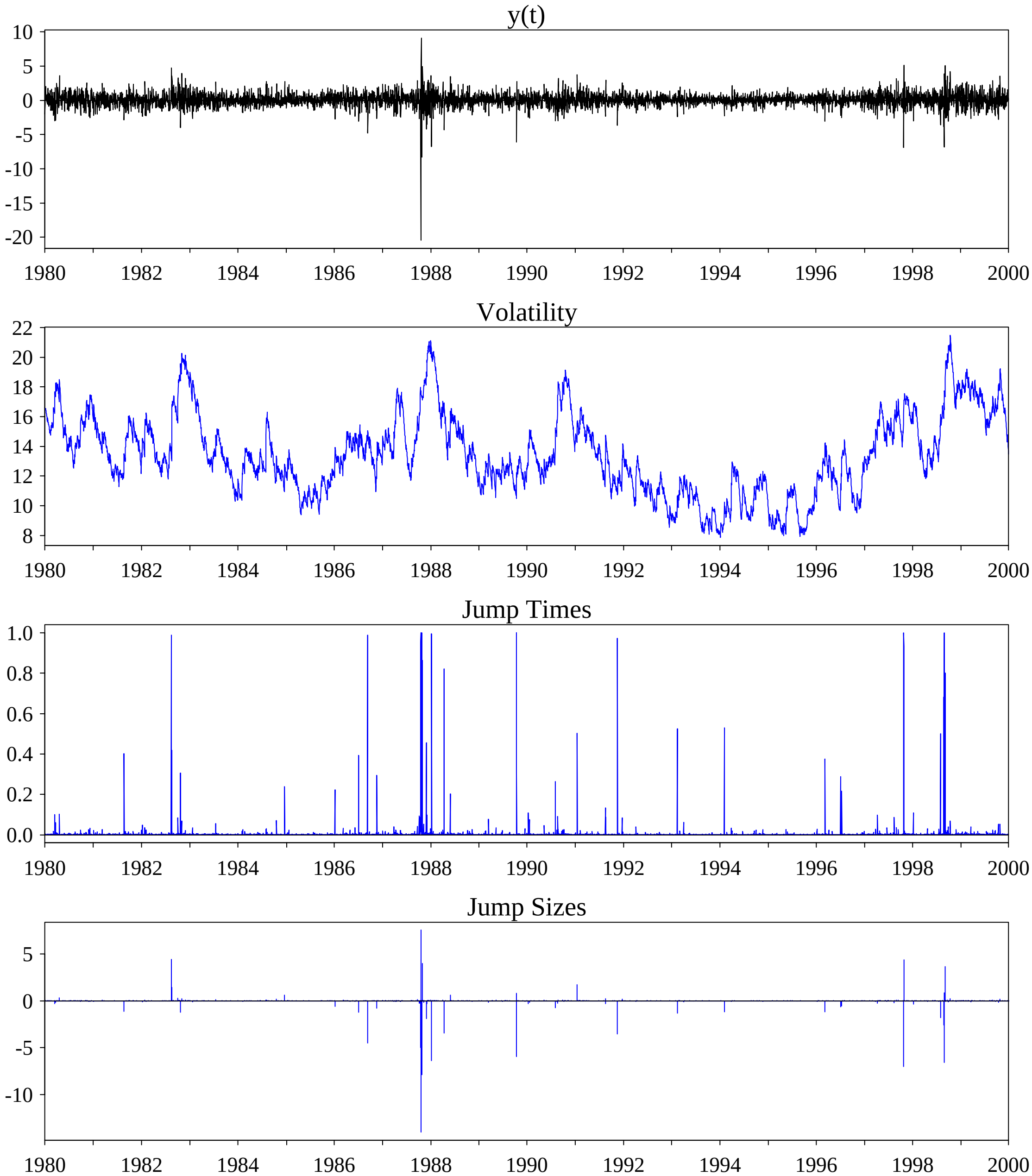

Figure 6: Particle filtering results for the SVJ model using S&P 500 returns from 1980-2000 with $M = 10$ and $N = 50,000$. The top panel displays daily returns; the second panel displays the median of the filtering distribution; the third panel displays the filtered jump times and the bottom panel displays filtered jump sizes.

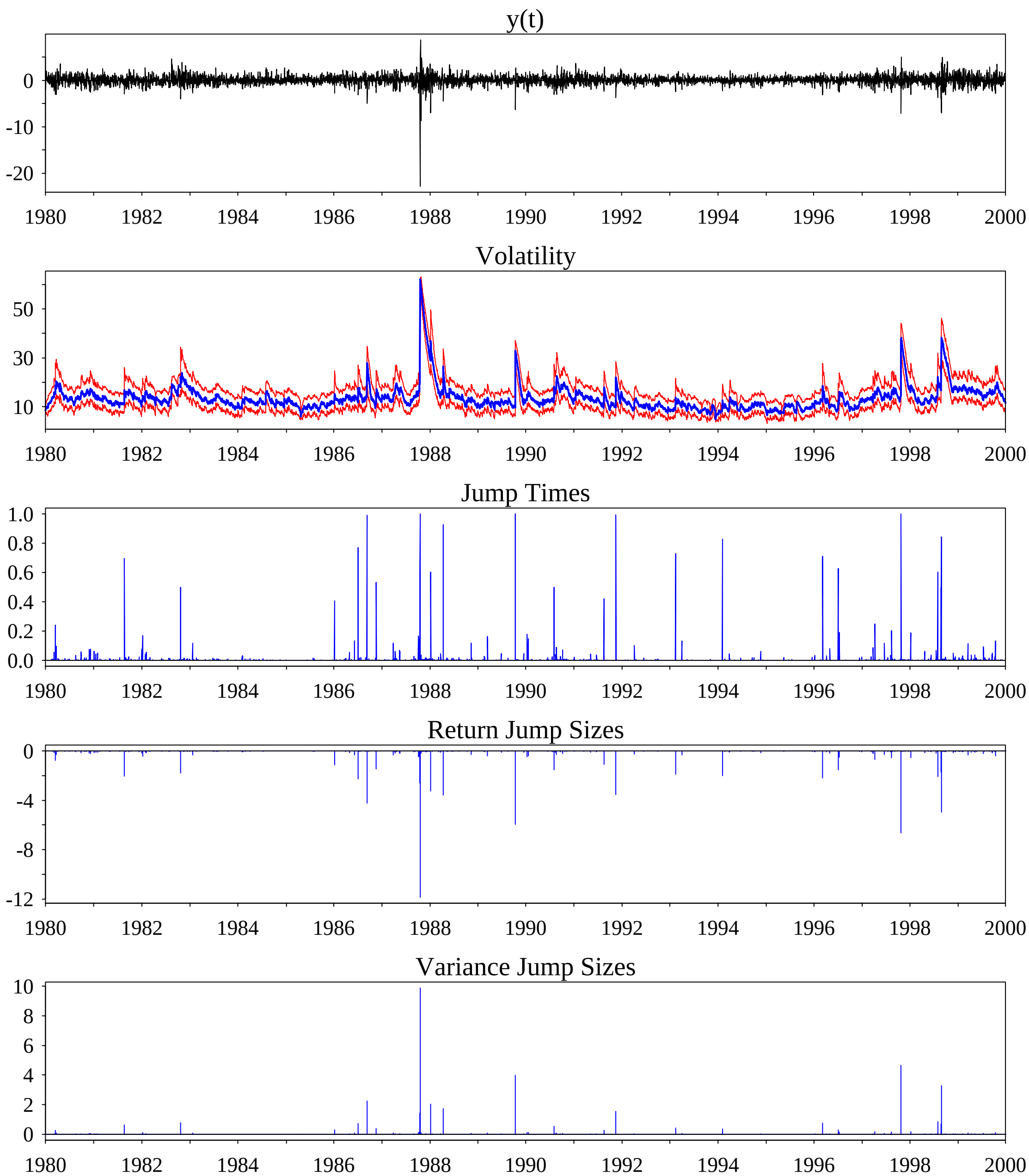

Figure 7: Particle filtering results for the SVCJ model using S&P 500 returns from 1980-2000 with $M = 10$ and $N = 50,000$. The top panel displays daily returns; the second panel displays the (10,50,90) percent quantiles of the filtering distribution; the third panel displays the filtered jump times and the bottom panel displays filtered jump sizes.

4.2 Parameter Learning

In the section, we apply the particle filtering to the joint problem of state and parameter estimation. We do this in the context of the SV model and use the algorithm of Storvik (2002) using simulated data. Figure 8 provide the estimates for the simulated data for $M = 10$ and $N = 100,000$. The upper left panel provides the simulated returns and the middle left panel provides quantiles of the filtered volatility distribution. The remaining panels provide the true parameters (solid) line and the $(10, 50, 90)$ percent quantiles of the sequential parameter distributions.

The first thing to note is that the in all cases, the sequential parameter posterior distributions contain the true parameter values and are gradually shrinking over time. Thus it appears as if the sequential parameter estimates are converging to the true values as expected. Also note that the speed at which the estimates converge to their true values varies: the stock return mean (μ) converges quickly while the volatility parameters (especially those in the drift) converge slower. This is not surprising as it is difficult to estimate the parameters of persistent processes such as volatility.

We also applied the Storvik's algorithm to the S&P 500 data set. Unfortunately, the algorithm degenerated in the sense that particles got clustered after the Crash of 1987 and never were able to uncluster. This is likely due to model misspecification as the SV model is incapable of handling events such as the -23% return that occurred during on October 19, 1987. We did not apply the parameter learning algorithm to the models with jumps. Sequential estimation in models with jumps is more difficult. Since jumps are rare, if there are no jumps in the first portion of the time series (as in our S&P 500 data set), the sequential parameters estimates for the jump parameters move toward one of two states: the first is with extremely high jump intensity and very small jumps and the second is with a jump intensity close to zero. One the particles are in these states, they do not appear to update well after jumps arrive (like the Crash of 1987). To remedy the problem, one needs to place an informative prior distribution on the jump parameters to avoid these

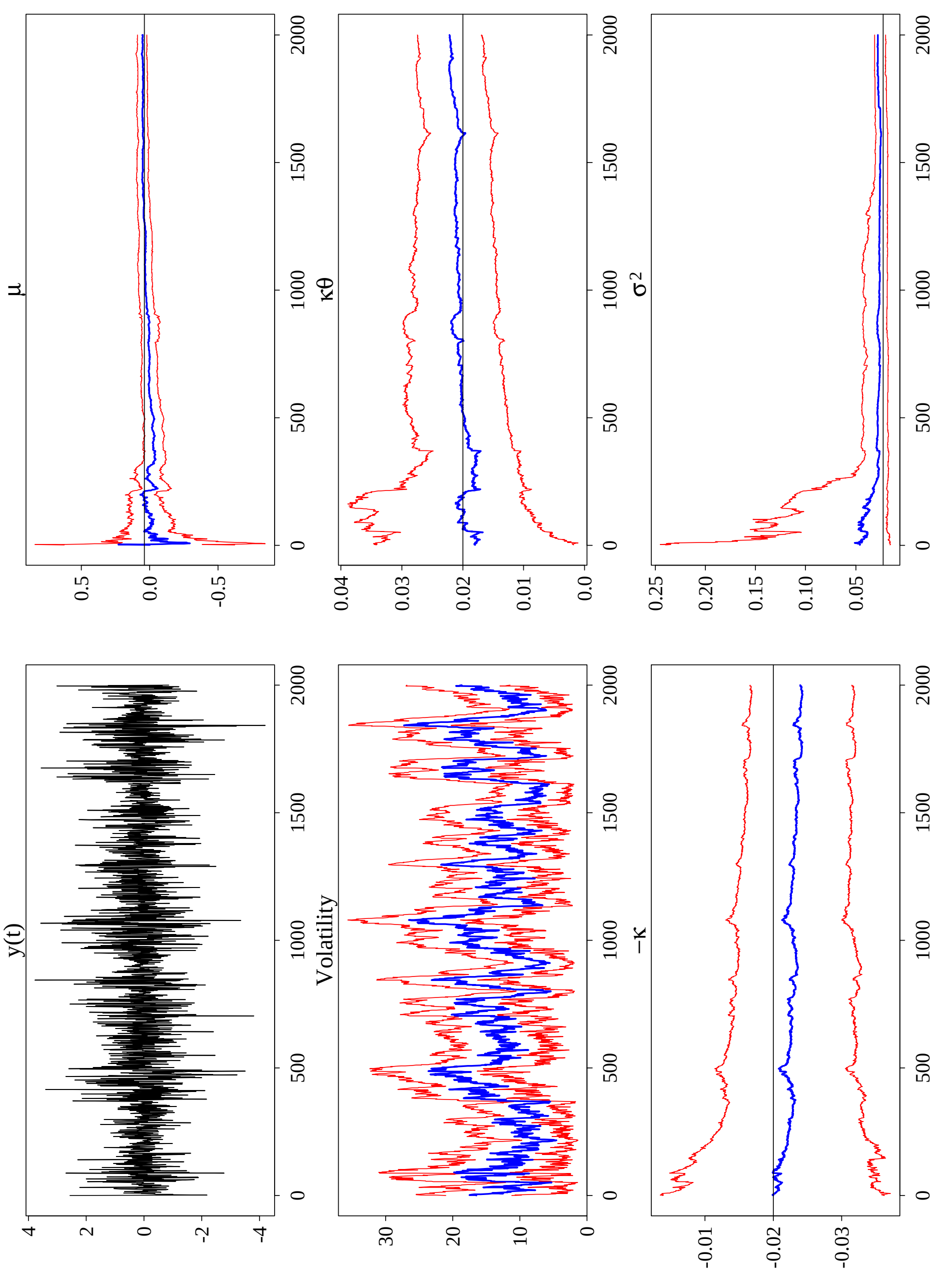

Figure 8: Sequential parameter and state variable estimation results for simulated data using the SV model.

degeneracies.

4.3 Option Pricing Implications

Last, we investigate an option pricing implication of our filtered volatility estimates. In models with jumps, a central issue is how to identify when a jump occurs and our algorithm provides a method to estimate the jump sizes. However, it is natural to ask what is the economic implication of the estimates: does it matter? In this section we show that it can have dramatic implications for option pricing. To demonstrate this, using the filtered volatilities we compute Black-Scholes implied volatility curves for the SV, SVJ and SVCJ models for two dates of market stress (October 20, 1987 and October 27, 1997), for an average volatility day and for a date, January 20, 1988, which is three months after the stock market crash.

The upper left panel indicates that model differences generate drastically different implied volatility curves. The SV model attributes the move to high volatility (see Figure 4), but since the increments of the model are normal, volatility can only increase gradually, and thus volatility is lower than the SVCJ model. The SVJ model attributes the Crash of 1987 to a jump in returns and thus spot and implied volatility curve remain low. Clearly, the difference between the SV and SVJ model is large. This issue has additional implications for individual equity options especially around dates such as earnings announcements, when a large move may be anticipated. If the move is in fact a jump, implied volatility should not change, although it typically does. The SVCJ model attributes the Crash to a jump in return with a very large contemporaneous jump in volatility. This results in a very high spot volatility. Interestingly, at this high level of volatility, the implied volatility curve is extremely flat. This occurs because the diffusive volatility dwarfs the contribution of jump components to volatility at these levels.

The implied volatility curves three months after the Crash of 1987 and show that the jump attribution issue persists long after the jumps arrive. This is due to the fact that

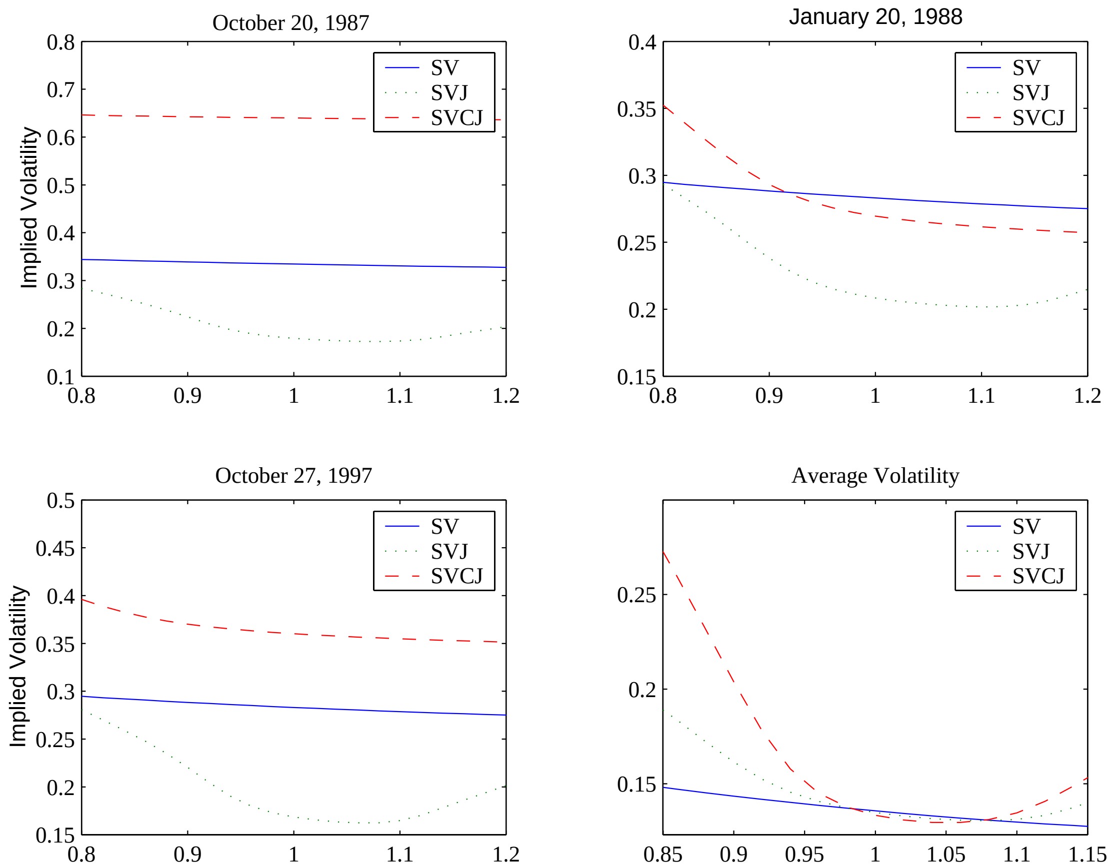

Figure 9: Black-Scholes implied volatility curves for various days using filtered volatility and the parameters from Eraker, Johannes and Polson (2002).

the models have very different speeds of mean reversion in volatility. Even though the SVCJ model has the highest mean reversion, volatility did not have enough time to revert down to the levels implied by the SVJ model. The lower left panel shows implied volatility curves on the day of the mini crash in 1997 and is similar to the results for 1987 which shows that it is not an anomaly, but is more indicative of filtered estimates during periods of market stress. For completeness, the lower right panel shows implied volatility on an average volatility day for the three models.

5 Conclusions

In this paper, we develop particle filtering algorithms for filtering and sequential parameter learning. The methods developed apply generically in multivariate jump-diffusion models. The algorithm performs well in simulations and we also apply the methodology to filter volatility, jumps in returns, and jumps in volatility from S&P 500 index returns.

In future work, we plan a number of methodological and empirical extensions. Empirically, we plan to further address the issue of jump attribution. If jumps in returns occur without corresponding jumps in volatility, we should not see an increase in implied volatility if the market is able to discern that the movement is a jump. Given the observation that implied volatility tends to increase, we can conclude either that the SVJ model is incorrect or that the market improperly estimates volatility in the presence of jumps. Second, we plan to augment the observed state vector with option prices to study the contemporaneous informational content of option prices regarding spot volatility.

Methodologically, we plan to investigate the advantages and disadvantages of particle filtering methods when compared to MCMC based fixed-lag filtering method (Johannes, Polson and Stroud (2002)). The motivation for this is the degeneracies that occur in the particle filtering algorithm in the presence of parameter learning. While MCMC based filtering is in general more computationally burdensome than particle filtering, there may be advantages in cases where the parameters are unknown due to its more efficient use of past information. In this regard, we also plan to implement and compare the methods to fixed-lag particle filtering.

It also appears possible to extend the methodology developed here to stochastic differential equations driven by Lévy processes. Barndorff-Nielson and Shephard (2001) use the particle filter to estimate integrated (as opposed to spot) volatility in a model where returns are driven by a Brownian motion and stochastic volatility is a Lévy Process.

References

Aït-Sahalia, Y. L. Hansen, J. Scheinkman (2002). Discretely-sampled diffusions. L.P. Hansen and Y. Ait-Sahalia (eds.), Handbook of Financial Econometrics . Amsterdam: North-Holland, forthcoming.

Aït-Sahalia, Y. (1996a). Nonparametric Pricing of Interest Rate Dependent Securities, *Econometrica*, 64, 527-560.

Aït-Sahalia, Y. (1996b). Testing Continuous Time Models of the Spot Interest Rate, *Review of Financial Studies*, 9, 385-426.

Aït-Sahalia, Y. (2002). Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, 70, 223-262.

Andersen, T. , L. Benzoni, and J. Lund. (2002). Towards an empirical foundation for continuous-time equity return models, *Journal of Finance* 57, 1239 - 1284.

Andersen, T., Bollerslev, T. and Diebold, F. (2002), Parametric and Nonparametric Volatility Measurement. L.P. Hansen and Y. Ait-Sahalia (eds.), Handbook of Financial Econometrics . Amsterdam: North-Holland, forthcoming.

Bakshi, G., C. Cao, and Z. Chen. (1997). Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003-2049.

Barndorff-Nielson, O. and N. Shephard (2001), Non-Gaussian OU based models and some of their uses in financial economics, *Journal of the Royal Statistical Society*, Series B, 63, 167–241.

Barndorff-Nielson, O. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models *Journal of the Royal Statistical Society*, Series B, volume 64, 2002, 253-280.

Bates, D. (1996). Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options, *Review of Financial Studies* 9, 69-107.

Bates, D., (2000). Post-'87 Crash fears in S&P 500 futures options, *Journal of Econometrics* 94, 181-238.

Black, F. and M. Scholes, (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81, no. 3, 637-654.

Brandt, M., and P. Santa-Clara. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *Journal of Financial Economics* 63, 161-210.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999). An Improved Particle Filter for Nonlinear Problems. *IEE Proceedings – Radar, Sonar and Navigation*, 1999, 146, 2–7.

Chernov, Mikhail, Eric Ghysels, A. Ronald Gallant, and George Tauchen, (2002), Alternative models for stock price dynamics. Forthcoming, *Journal of Econometrics*.

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

Conley, T., L. Hansen, E. Luttmer and J. Scheinkman, (1997), Short Term Interest Rates as Subordinated Diffusions, *Review of Financial Studies*, 10, 525-577.

Das, Sanjiv, and Randarajan Sundaram, 1999, Of smiles and smirks: a term structure perspective, *Journal of Financial and Quantitative Analysis* 34, 211-240.

Del Moral, P. and J. Jacod (2001). Interacting Particle Filtering with discrete observations. In, Sequential Monte Carlo Methods in Practice, pp. 43-77, Eds. A. Doucet, J. F. G. de Freitas, N. J. Gordon. Springer Verlag.

Del Moral, P. J. Jacod and P. Protter. (2001). The Monte Carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields,* 120, 346-368.

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10, 197–208.

Doucet, A., de Freitas, N. and Gordon, N. (2001). Sequential Monte Carlo methods in practice. Springer, New York.

Duffie, D., D. Filipovic and W. Schachermayer. (2002). Affine Processes and Applications in Finance. forthcoming, *Annals of Applied Probability.*

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

Durham, G. and R. Gallant. (2002). Numerical techniques for maximum likelihood Estimation of Continuous-time Diffusion Processes. Forthcoming, *Journal of Business and Economic Statistics.*

Elerian, O. S. Chib and N. Shephard. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, 69, 959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with applications to finance. *Journal of Business and Economic Statistics* 19-2, 177-191.

Eraker, B. (2002a) Do equity prices and volatility jump? Reconciling evidence from spot and option prices. Working paper, Duke University.

Eraker, B. (2002b) Comment on "Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes," Forthcoming, *Journal of Business and Economic Statistics* .

Eraker, B. , M. Johannes and N. Polson. (2002). The impact of jumps in volatility and returns." Forthcoming, *Journal of Finance.*

Foster, D. and D. Nelson, (1996). Continuous Record Asymptotics for Rolling Sample Variance Estimators. *Econometrica* , 64, 139 - 174.

Gallant, R., and J. Long (1997), Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared, *Biometrika* 84, 125-141.

Gihkman, I. and A.V. Skorohod, (1972). *Stochastic Differential Equations.* Springer-Verlag, New York.

Gourieroux, C, A. Monfort and E. Renault, (1993). Indirect Inference. *Journal of Applied Econometrics*, S85-118 Vol. 8 S85-118.

Gordon, N., Salmond, D. and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings*, F-140, 107–113.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327-343.

Kiefer, N., (1978). Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model, *Econometrica*, 46, 427-445.

Jacod, J. and A. Shiryaev, 1987, Limit Theorems for Stochastic Processes, Grundlehren der Mathematischne Wissenschaften, 288, Springer Verlag, New York.

Johannes, M. (2002) The Economic and Statistical Role of Jumps in Interest Rates. Working paper, Columbia University.

Johannes, M., N. Polson and J. Stroud. (2002) Sequential Optimal Portfolio Performance: Market and Volatility Timing. Working paper, Columbia University.

Landen, C. (2000). Bond pricing in a hidden Markov model of the short rate. *Finance and Stochastics*, 4, 371-389.

Liptser, R. and N. Shiriaev (2001). "Statistics of Random Processes." Springer-Verlag.

Liu, J., F. Longstaff, and J. Pan. (2002). Dynamic asset allocation with event risk. *Journal of Finance*, forthcoming.

Liu, X.Q. and C. Li. (2000). Weak Approximations and Extrapolation of Stochastic Differential Equations with Jumps. *Siam Journal of Numerical Analysis, 37, 1747-1767.*

Merton, R., (1969). Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case. *Review of Economics and Statistics, 51, 247-57.*

Merton, R., (1971). Optimum Consumption and Portfolio Rule in a Continuous-Time Model. *Journal of Economic Theory, 3, 373-413.*

Merton, R., (1976a). Option pricing when the underlying stock returns are discontinuous. *Journal of Financial Economics* 3, 1235-144.

Merton, R., (1976b). The Impact on Option Pricing of Specification Error in the Underlying Stock Price Returns. *Journal of Finance,* 31, 333-350.

Merton, R., (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics,* 8, 323-61.

Mikulevicius, R. , Platen, E. (1988). Time Discrete Taylor Approximations for Ito Processes with Jump Component, *Math. Nachr.*, 138 , 93-104.

Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study, *Journal of Financial Economics* 63, 3–50.

Pedersen, A. (1995). "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equationas Based on Discrete Observations." *Scandinavian Journal of Statistics 22*, 55-71.

Piazzesi, M. (2001). Macroeconomic jump effects and the yield curve. Working paper, UCLA.

Pitt, M. (2002). Smooth particle filters for likelihood evaluation and maximization. Working paper, University Warwick.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association*, 590–599.

Platen, E. and R. Rebolledo. (1985). Weak Convergence of semimartingales and discretization methods. *Stochastic Processes and Their Applications*, 20, 41-58.

Pugachev, V. (1987). "Stochastic differential systems: analysis and filtering." Wiley, New York.

Rubenthaler, S. (2001). Numerical Solution of the solution of a stochastic differential equation driven by a Lévy process. working paper, University of Paris.

Storvik, G. (2002). Particle filters in state space models with the presence of unknown static parameters. *IEEE Trans. on Signal Processing,* 50, 281–289.