Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices

Michael Johannes Nicholas Polson Jonathan Stroud*

February 22, 2006

*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 for providing option pricing code and Mike Chernov, Neil Shephard and Mike Pitt and seminar participants at Cirano/University of Montreal and Columbia University for their comments.

Abstract

This paper provides a methodology for optimally filtering latent state variables in discretely observed jump-diffusion models. When prices are continuously observed, state variables such as volatility, jump times, and jump sizes are also observed, but with discrete observations these variables are unobserved. We combine discretization schemes with Monte Carlo methods to compute the optimal filtering distribution: the distribution of the latent states conditional on the observed prices. Our approach is very general, applying in models with nonlinear characteristics and even non-analytic observation equations. We use simulations to investigate how sampling frequency affects jump and volatility estimates and then extract information about volatility jointly from equity index options and index returns.

1 Introduction

Asset pricing models for term structure, option pricing, portfolio allocation and exchange rate applications are commonly specified in continuous-time. To apply these models, researchers require parameter estimates and estimates of any latent variables in order to compute the predictive distribution of prices. There are now a wide range of methods available for parameter estimation.[1] In this paper, we focus on the other estimation problem, estimating latent variables from observed prices sequentially.

Latent variable estimates are crucial for empirical applications. For example, in portfolio allocation or option pricing applications, volatility forecasts are required every time period to form portfolios or prices options. Similarly, a central problem in models with jumps is to separately identify or 'disentangle' the jumps from diffusive components and to identify the underlying economic events that generated jumps (see Aït-Sahalia (2004b)). These predictive and filtering problems are inherently *sequential*, as it must be repeated for every time period.[2] The sequential nature of the problem places severe constraints on filtering methods, as it must be implementable in real time as additional data arrives each time period.

This paper provides a general solution to the predictive and filtering problem in a wide

[1] For example, simulated method of moments (Duffie and Singleton, 1993, Gallant and Long 1997, and Gallant and Tauchen 2003), approximate analytical maximum likelihood (Aït-Sahalia, 2004a, Aït-Sahalia and Kimmel, 2005a,b), simulated maximum likelihood (Brandt and Santa-Clara, 2002, Piazzesi, 2005), implied state GMM (Pan, 2002), Markov Chain Monte Carlo methods (Jones, 1998, Eraker, 2001, Elerian, Shephard, and Chib, 2001, Chib, Pitt, and Shephard 2005), and nonparametric methods (Aït-Sahalia, 1996a,b, Conley, Hansen, Luttmer, and Scheinkman 1997).

[2] In principle, parameters could also be re-estimated every time period. However, since parameters are fixed and we typically have long time series datasets, parameters do not change substantially as additional data observations arrive. Johannes, Polson and Stroud (2005) consider the problem of sequential parameter estimation in discrete-time jump-diffusion models.

class of continuous-time models. The approach relies on an commonly used insight: to accurately approximate a conditional distribution arising from a continuous-time model, simulate the stochastic differential equations at very high frequencies. This reduces and can even eliminate the discretization bias and has been used extensively for parameter estimation. Given the high-frequency approximation, we apply particle filtering methods based on Gordon, Salmond and Smith (1993) and Pitt and Shephard (1999). By analogy, this approach is very similar to Jones (1998), Elerian, Shephard and Chib (2001) and Eraker (2001) who apply MCMC methods to a system augmented by high frequency simulation.

Our approach has a number of advantages. First, it applies in an extremely wide-range of settings, encompassing virtually all of the models and types of data encountered in practice. It handles diffusions, jump-diffusions, and continuous-time regime-switching models, models with nonlinear drift, diffusion, jump intensity or jump amplitude functions, and models in which the observed prices are non-analytical functions of the state variables, such as those arising when using option prices in stochastic volatility models.[3] These factors provide the flexibility to model predictable variation in conditional moments as well as importance economic events such as crashes or announcements.[4]

Second, our approach computes the *optimal* filtering distribution, that is, the posterior distribution of the latent state variables given the observed data. This distribution provides all of the sample information about the state variables, quantifying, for example, estimation risk. Third, our approach applies with multiple observed price series and allows the researcher to coherently combine different sources of information when computing the

[3]As noted in Barndorff-Nielson and Shephard (2004, 2006b), our approach also handles the important class of Levy-driven stochastic differential equations (see, e.g., Cont and Tankov, 2003).

[4]In portfolio settings, see Pan and Liu (2003), Liu, Longstaff, and Pan (2003), Das and Uppal (2004) or Aït-Sahalia, Chaco-Diaz, and Hurd (2006), in fixed income settings, see Andersen, Benzoni, and Lund (2002), Das (2002), Johannes (2004), Piazzesi (2005), Dai, Singleton, and Yang (2005), in option pricing settings see Bates (2000), Bakshi, Cao and Chen (1997), Pan (2002) and Duffie, Pan and Singleton (2000).

filtering distribution. For example, consider the case of filtering stochastic volatility from observations that include both stock and option prices. Our approach computes the filtering distribution by optimally combining the information based on returns ($\mathbb{P}-$measure) and option prices ($\mathbb{Q}-$measure).

Finally, our approach is extremely simple to implement and computationally fast. Unlike many filtering algorithms that rely on solving differential equation or developing deterministic approximations, our approach takes only a few lines of code and requires only simulation from well-known distributions such as the normal or multinomial distribution.[5] The algorithms are also extremely fast in terms of computational time. This is important for real-world examples where data arrives sequential and predictive distributions must be computed in real time. For these reasons, our approach provides a toolkit for implementing continuous-time models for finance applications.

We demonstrate our filtering approach with both simulation and real data using the double-jump model of Duffie, Pan and Singleton (2000). This model offers significant challenges as it has multiple latent factors (stochastic volatility and jumps in prices and/or volatility). Thus, price variation is driven by both factors and a central problem is separately estimating jump and volatility components[6]; Disentangling jumps and stochastic volatility, in particular, is important as prior research documents that jumps and diffusive components are compensated differently as the former earn significant risk premiums while the latter do not (see, e.g., Pan 2002 or Broadie, Chernov, and Johannes 2005). Moreover,

[5] For example, classical filtering approaches require computing the solutions to partial differential equations (see, e.g, Pugachev (1987)). In the Gaussian case, these PDE's reduce to ODE's that are solved to generate the Kalman filter. Bates (2005) use deterministic functional approximations to the characteristic function of affine and semi-affine processes. Chapter 8 of Anderson and Moore (2004) discusses other functional approximation methods for filtering.

[6] There is a growing literature analyzing nonparametric estimators of jumps and realized volatility using high-frequency data. For example, see, e.g., Barndorff-Nielson and Shephard (2004, 2006a,b), Huang and Tauchen (2005), and Andersen, Bollerslev, and Diebold (2005).

volatility forecasts depend crucially on the presence of jumps.

Based on simulations, we provide evidence to document the performance of the algorithm, the ability to estimate volatility and the ability to disentangle volatility from jumps at various observation frequencies. In a diffusive stochastic volatility model, we find that for realistic parameters, there is little discretization bias for high frequencies such as daily, but more for longer frequencies. For these lower frequencies, simulating a small number of data points between observations largely eliminates the discretization bias. In models with jumps, however, we find that it is possible to identify jumps at a daily frequency, but it is very difficult to identify jumps with less frequently sampled data. For example, the filter is able to identify about 60 percent of the jumps at a daily frequency. The 40 percent of the jumps not identified are too small to identify. At a weekly frequency, less than a third of the jumps are correctly identified. Since volatility aggregates, weekly variance is roughly five times as large as daily variance and therefore only the largest jumps can be identified at the weekly frequency.

Finally, we address a long-standing issue in empirical finance: Is the time series of returns consistent with information embedded in option prices? While risk premiums drive a wedge between the objective measure ($\mathbb{P}$) and risk-neutral measure ($\mathbb{Q}$) parameters, a well-specified model should be consistent with both the time series of returns and the cross-section of option prices. To analyze this issue, we estimate volatility, jump times and jump sizes from only returns and from both returns and options. This will allow us to evaluate the information about volatility in both data sources and to evaluate the consistency between the two. We have two main findings.

First, we quantify how much information options contain regarding volatility. If options were observed without error, volatility could also be estimated without error, but this is not reasonable as even at-the-money options have large bid-ask spreads. We incorporate bid-ask spreads into our filtering approach, assuming the observed option prices are noisy signals of the true price. We quantify estimation risk via the posterior standard deviation of the filtering distribution. We find that even with these pricing errors that the estimation risk falls by as much as forty percent with the addition of option prices.

Second, we find that it is still difficult to reconcile the existing models with the information jointly in returns and option prices. In the models with jumps in returns and jumps in volatility, the option based volatility tends to be higher than that embedded in returns. Factor risk premia reduce spot volatility, but the two sources are still not wholly consistent. These results are consistent with either model misspecification in the volatility process or time-varying risk-premiums. The nature of the misspecification is consistent with a model where the long-run mean of volatility varies over time. This model has been suggested by Duffie, Pan, and Singleton (2000) but has not been analyzed to date. Time-varying risk premia are also able to reconcile the issues, as spot volatility is not required to move as much to explain option implied expectations.

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 Optimal Filtering of Jump-Diffusions

Before introducing the models and our filtering approach, we briefly discuss the general filtering problem. We consider the *optimal* nonlinear filtering problem. In our context, the optimal filter is abstractly given by $p\left(L_t|P_{1:t}\right)$ where $P_{1:t} = (P_1, ..., P_t)$ are the discretely observed prices and L_t are the latent variables. In continuous-time, the optimal filter is given by $p\left(L_t|\left\{P_s\right\}_{s\leq t}\right)$ (see, for example, Pugachev 1987).

These posterior distributions are *optimal nonlinear* filters, in the sense that each of the conditional moments obtained from these densities, for example $E\left[L_t|P_{1:t}\right]$, are the optimal estimates in the mean squared error sense. More generally, the optimal estimate of any function of the state is given by $E[f(L_t)|P_{1:t}]$ which again, is obtained by integrating the optimal filter against the function f. Moreover, these filters are generally nonlinear, in the sense that the expectation is not necessarily a linear function of the observed prices. Only in the specialized cases of linear, Gaussian models (in either continuous or discrete-time) are the optimal filters linear (e.g. the Kalman filter).

While there are numerous estimators of latent state variables (especially volatility), there is a strong preference against using optimal filters and in favor of suboptimal filters, such as the extended Kalman filter and other approximate filters. There is a good reason for this: historically, the computational burdens of computing optimal filters were overwhelming. For example, Pugachev (1987) argues that the practical application of the optimal filter is "a matter of great and often insuperable difficulty." Due to this, it is common to focus on simpler, computationally feasible suboptimal filters.

The particle filtering methodology introduced by Gordon, Salmon, and Smith (1993) provides a general, computationally attractive method for computing optimal filters. In this paper, we show how to adapt this methodology to continuous-time models and use it to address a number of empirical issues.

2.1 Model Specification

We assume that there are two types of prices observed: prices whose dynamics are directly modeled, S_t, and derivative prices, C_t. Together, we let $P_t = (S_t, C_t)$. There is also a vector of state variables, X_t, that impact the level or distribution of prices. As an example, S_t could be an equity index, X_t its stochastic variance, and C_t the price of an option written on that index. Alternatively, C_t could be a vector of yields and X_t are the state variables describing the spot rate evolution. In general, C_t, will be a function of the observed prices, the states, and any parameters, $C_t = C(S_t, X_t, \Theta)$.

We assume sufficient regularity so that the driving variables, S_t and X_t, jointly solve the stochastic differential equation with jumps,

$$dS_t = \mu^s(\Theta, S_t, X_t)\,dt + \sigma^s(\Theta, S_{t-}, X_{t-})\,dW_t^s + d\left(\sum\nolimits_{n=1}^{N_t^s} Z_n^s\right) \tag{1}$$

$$dX_t = \mu^x(\Theta, X_t)\,dt + \sigma^x(\Theta, X_{t-})\,dW_t^x + d\left(\sum\nolimits_{n=1}^{N_t^x} Z_n^x\right), \tag{2}$$

where Θ is a vector of parameters, $t-$ denotes the limit taken from the left, W_t^s and W_t^x are potentially correlated Brownian motions; N_t^s and N_t^x are point processes with predictable intensities $\lambda^s(\Theta, S_{t-}, X_{t-})$ and $\lambda^x(\Theta, X_{t-})$; τ_n^s and τ_n^x are the jump times; and Z_n^s and Z_n^x are the jump sizes with $\mathcal{F}_{\tau_{n-}}$ conditional distributions $\Pi^s\left(\Theta, S_{\tau_{n-}}, X_{\tau_{n}-}\right)$ and $\Pi^x(\Theta, X_{\tau_n-})$.[7] The interpretation of the process is as follows. At a jump time, τ_n^s or τ_n^x, there is a discontinuity in the sample path, $S_{\tau_n^s} - S_{\tau_n^s-} = Z_n^s$ and $X_{\tau_n^x} - X_{\tau_n^x-} = Z_n^x$, and between jump times, $\tau_n \leq t < \tau_{n+1}$, the prices and state variables diffuse. As is common in financial applications, we assume that the state variables are exogenous, in the sense that the characteristics of the stochastic differential equation for the states are independent of the prices, that is, μ^x, σ^x, λ^x, and Π^x do not depend on S_t. For simplicity, we assume that S_t and X_t are univariate, although our approach applies equally to multivariate models.

The specification in (1) and (2) covers nearly all of the cases of interest in economics and finance: multi-variate diffusion models, multi-variate jump-diffusion models (Duffie, Pan, and Singleton 2000) and continuous-time regime switching models, diffusion models in which the drift and diffusion are driven by a continuous-time, discrete state Markov chain (see, e.g., Dai, Singleton, and Yang 2005 and Landen 2000). We do not consider infinitely active Levy processes, but, as noted by Barndorff-Nielson and Shephard (2004, 2006b), our approach generalizes to this important class of processes.

[7] Examples of sufficient conditions for a well-defined strong solution include standard linear growth and Lipschitz restrictions on the drift, diffusion and jump intensity and integrability conditions on the jump sizes and initial values, see, for example, Jacod and Shiryaev (1987) and Gihkman and Skorohod (1972).

Derivative prices are given by

$$C_t = C(S_t, X_t, \Theta) = E^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds} f(S_T) | S_t, X_t, \Theta\right] \tag{3}$$

where $\mathbb{Q}$ is a risk-neutral probability measure and f is the payoff function. Due to potential state dependence in the drift coefficients, the diffusion coefficients, and $C(S_t, X_t, \Theta)$, the model is also non-linear in the state variables. The only assumptions that we require on the function C is that it can be evaluated at a pair of points, (x, s), thus $C(\cdot, \cdot, \Theta)$ does not need to be analytical. It could, for example, solve a partial differential equation or be given as a integral. Derivative prices offer an interesting challenge for filtering. Since theory implies a deterministic relationship between prices, parameters and states, derivative prices are potentially fully revealing of both the parameters and states. With enough derivative prices, there is a stochastic singularity as it is impossible to simultaneously fit all of the prices with a low dimensional parameter and state vector.

To circumvent this problem, it is common to assume that some or all of the derivative prices are observed with error: if $C(X_t, S_t)$ is the model-implied price, then it is common to assume that $C_t = C(S_t, X_t, \Theta) + \varepsilon_t$ or $C_t = C(S_t, X_t, \Theta) e^{\varepsilon_t}$ where ε_t is normally distributed. There are multiple justifications for pricing errors. First, there is a genuine concern with noisy price observations, such as those generated by the bid-ask spread or interpolated prices. In the case of options, estimates for the bid-ask spread for at-the-money, short maturity index options options vary from three to ten percent (see, Bakshi, Cao, and Chen, 1997 or George and Longstaff, 1993). In term structure models, it is common to use zero coupon yields interpolated from coupon bonds or interest rate swaps. Second, even if the derivative prices are observed without error, the pricing error is a convenient way to break the stochastic singularity and to recognize that our models are approximations to reality.

2.2 Optimal Filtering of Continuous-time Models

We assume that prices and states evolve continuously in time, but the prices are observed only discretely at a fixed frequency, normalized to unit length without loss of generality. From the researcher's perspective, the jump times, jump sizes and X_t are latent, although we label X_t as the "state" variables. As is often the case, it is possible to expand the dimension of the state vector to include the jump times and sizes and then, formally, the jump times and sizes are "states," although we avoid this formality. We focus on the filtering problem and therefore assume the parameters are known.

To develop the filtering approach, we write the model in its state space form. The two observation equations are (suppressing parametric dependence)

$$S_{t+1} = S_t + \int_t^{t+1} \mu^s \left(S_v, X_v\right) dv + \int_t^{t+1} \sigma^s \left(S_{v-}, X_{v-}\right) dW_v^p + \sum_{t<\tau_n^s \leq t+1} Z_n^s \tag{4}$$

and

$$C_t = C\left(S_t, X_t\right) + \varepsilon_t. \tag{5}$$

The state evolution is

$$X_{t+1} = X_t + \int_t^{t+1} \mu^x \left(X_v\right) dv + \int_t^{t+1} \sigma^x \left(X_{v-}\right) dW_v^x + \sum_{t<\tau_n^x \leq t+1} Z_n^x. \tag{6}$$

Due to the stochastic integrals and the jumps, it is clear that the shocks in the model are non-normal, both in the observation equations and the state evolution.

The optimal filtering problem is abstractly solved by the sequence of conditional densities, $p\left(L_t | P_{1:t}\right)$, for $t = 1, ...T$. By Bayes rule,

$$p\left(L_{t+1} | P_{1:t+1}\right) = \frac{p\left(P_{t+1} | L_{t+1}, P_t\right) p\left(L_{t+1} | P_{1:t}\right)}{p\left(P_{t+1} | P_{1:t}\right)}, \tag{7}$$

where $p\left(P_{t+1} | L_{t+1}, P_t\right)$ is the likelihood (the distribution of observed prices conditional on latent states and past prices), $p\left(L_{t+1} | P_{1:t}\right)$ is predictive distribution of latent states, and

$p\left(P_{t+1}|P_{1:t}\right)$ is the predictive distribution of prices. Note that these densities are a function $P_{1:t}$, and are therefore t-dimensional.

Computing these densities is difficult as the none of the distributions are analytical. For example,

$$p\left(L_{t+1}|P_{1:t}\right)=\int p\left(L_{t+1}|L_t\right)p\left(L_t|P_{1:t}\right)dL_t, \tag{8}$$

is a high-dimensional integral. Even though the building blocks, Brownian motions, point processes and the jump sizes have known distributions, $p\left(P_{t+1}|P_t,X_t\right)$ and $p\left(L_{t+1}|L_t\right)$ are not known analytically. This is due to the complicated mixture structure of the model: there are multiplicative mixtures, $\left(\sigma^x\left(X_{v-}\right)dW_v^x\right)$, random sums of random variables $\left(\sum Z_n^x\right)$ and additive mixtures (both from $\int_t^{t+1}\sigma^x\left(X_{v-}\right)dW_v^x$, a limit of sums, and $\int_t^{t+1}\sigma^x\left(X_{v-}\right)dW_v^x+\sum Z_n^x$). Together, these factors make analytic treatment of the filtering problem impossible.

To solve the filtering problem in this setting, integration and simulation problems must be addressed. First, suppose that $p\left(P_{t+1}|L_{t+1},P_t\right)$ and $p\left(L_{t+1}|L_t\right)$ could be evaluated directly. Then, the filtering problem is one of updating the current filtering density, $p\left(L_t|P_{1:t}\right)$, to the next period's, $p\left(L_{t+1}|P_{1:t+1}\right)$. This is a problem of estimating the integral, $\int p\left(L_{t+1}|L_t\right)p\left(L_t|P_{1:t}\right)dL_t$. We solve this problem by approximating the filtering density, $p\left(L_t|P_{1:t}\right)$, with a discrete probability distribution, which is the essence of the particle filter. Second, to evaluate the $p\left(P_{t+1}|L_{t+1},P_t\right)$ and $p\left(L_{t+1}|L_t\right)$ when the prices and latent variables arise from continuous-time models, we use standard time-discretization methods and simulation.

The next section discusses the discrete-time particle filter and shows how to adapt the discrete-time particle filter to continuous-time jump-diffusions and models with non-analytic observations equations.

2.3 The Particle Filter

The particle filter requires only two assumptions: (A1) that the state evolution can be exactly simulated, that is, that exact samples from $p(L_{t+1}|L_t)$ can be drawn; and (A2) the likelihood can be exactly evaluated as a function of L_t and P_t. Under these assumptions, the particle filter delivers an estimate $p^N(L_t|P_{1:t})$ of the true filtering density, $p(L_t|P_{1:t})$. As the number of particles, N, increases, the approximation converges to the true filtering distribution under certain regularity conditions which we discuss below. We now describe the mechanics of the particle filter, drawing on the discussion in Doucet, de Freitas and Gordon (2001) and Pitt and Shephard (1999).

The particle filter approximates the filtering density by a discrete probability distribution, that is, $p(L_t|P_{1:t})$ is approximated by a set of particles, $\left\{L_t^{(i)}\right\}_{i=1}^N$ with probability $\left\{\pi_t^{(i)}\right\}_{i=1}^N$,

$$p^N(L_t|P_{1:t}) = \sum_{i=1}^N \delta_{L_t^{(i)}} \pi_t^{(i)}$$

where δ is the Dirac function. Once the distribution is discretized, integrals become sums. For example, the estimate of the predictive distribution is

$$p^N(L_{t+1}|P_{1:t}) = \sum_{i=1}^N p\left(L_{t+1}|L_t^{(i)}\right)\pi_t^{(i)} \approx \int p(L_{t+1}|L_t)\, p(L_t|P_{1:t})\, dL_t$$

and the filtering density at time $t+1$ is defined via the recursion:

$$p^N(L_{t+1}|P_{1:t+1}) \propto p(P_{t+1}|L_{t+1}, P_t) \sum_{i=1}^N p\left(L_{t+1}|L_t^{(i)}\right)\pi_t^{(i)}.$$

Thus, once $p(L_t|P_{1:t})$ is approximated by a set of particles, $p^N(L_{t+1}|P_{1:t})$ and $p^N(L_{t+1}|P_{1:t+1})$ can be computed.

The previous paragraph showed that the transition from a particle representation of $p^N(L_t|P_{1:t})$ to $p^N(L_{t+1}|P_{1:t+1})$ is straightforward, but one steps remains, that is, to generate new particles, $\left\{L_{t+1}^{(i)}\right\}_{i=1}^N$, and probabilities, $\left\{\pi_{t+1}^{(i)}\right\}_{i=1}^N$ that approximate $p^N(L_{t+1}|P_{1:t+1})$.

There are a number of different approaches that can be used to update or propagate the particles: the weighted bootstrap (also known as the sampling/importance sampling (SIR) algorithm), rejection sampling and MCMC methods. We follow the literature and use the weighted bootstrap/SIR algorithm (Smith and Gelfand 1992, Gordon, Salmond and Smith 1993) to propagate the particles for its generality and simplicity.

To understand the mechanics of the weighted bootstrap, we view $p^N\left(L_{t+1}|P_{1:t}\right)$ as the prior and $p\left(P_{t+1}|L_{t+1},P_t\right)$ as the likelihood:

$$p^N\left(L_{t+1}|P_{1:t+1}\right) \propto \underbrace{p\left(P_{t+1}|L_{t+1},P_t\right)}_{Likelihood} \underbrace{\sum\nolimits_{i=1}^{N} p\left(L_{t+1}|L_t^{(i)}\right)\pi_t^{(i)}}_{Prior}.$$

Updating the particles is similar to computing the posterior distribution given draws from the prior. Generating new states is straightforward as (A1) implies that we can directly simulate the state evolution. To do this, propagate the states forward by drawing $L_{t+1}^{(i)} \sim p\left(L_{t+1}|L_t^{(i)}\right)$. Given the updated states, $\left\{L_{t+1}^{(i)}\right\}_{i=1}^{N}$, the weighted bootstrap then re-samples these states $\left\{L_{t+1}^{(i)}\right\}_{i=1}^{N}$ with weights $\pi_{t+1}^{i} \propto p\left(P_{t+1}|L_{t+1}^{(i)},P_t\right)$. For clarity, we state the algorithm in steps:

1. Given $\left\{L_t^{(i)},\pi_t^{(i)}\right\}_{i=1}^{N}$, simulate the latent state vector forward using the state evolution equation. That is, draw $L_{t+1}^{(i)} \sim p\left(L_{t+1}|L_t^{(i)}\right)$.
2. Evaluate the likelihood function at the new state, $p\left(P_{t+1}|L_{t+1}^{(i)},P_t\right)$, and set
$$\pi_{t+1}^{(i)} = \frac{p\left(P_{t+1}|L_{t+1}^{(i)},P_t\right)}{\sum\nolimits_{i=1}^{N} p\left(P_{t+1}|L_{t+1}^{(i)},P_t\right)}.$$
3. Finally, re-sample N particles with replacement from the multinomial distribution of $\left\{L_{t+1}^{(i)}\right\}_{i=1}^{N}$ with probabilities $\left\{\pi_{t+1}^{(i)}\right\}_{i=1}^{N}$.

Gordon, Salmond, and Smith (1993), using an argument from Smith and Gelfand (1992), show that these re-sampled draws provide a sample from the approximated filtering density, $p^N(L_{t+1}|P_{1:t+1})$. An advantage of the weighted bootstrap is that by sampling with replacement according to the probabilities $\pi_{t+1}^{(i)}$, the procedure selectively and over time eliminates states with very low probability by disproportionately re-sampling states with higher probability. Thus the algorithm propagates high likelihood states forward while discarding low likelihood states. Given the computationally simplicity, it is straightforward to program and the algorithm runs quickly. This implies that it is possible to estimate the densities with high precision (large N) for large T.

Another advantage of the particle filter is that it can be easily adapted and improved. One approach, the auxiliary particle filter, is a straightforward extension of the particle filter and is described in Pitt and Shephard (1999). Auxiliary particle filters are particularly useful when dealing with outliers and low probabilities states, as the traditional particle filter can have difficulties sampling the tails. We have found it to be extremely useful in models with rare jumps and we use it below.

The particle filter (and its auxiliary extensions) has been applied to a wide-variety of discrete-time models in econometrics and statistics. Pitt and Shephard (1999) and Chib, Nardari and Shephard (2002) use particle filtering in discrete time stochastic volatility models to construct likelihood functions for estimation or testing. Barndorff-Nielson and Shephard (2004) outline how to use particles to filter volatility in a discrete-time Levy driven stochastic volatility model. Concurrently to our work, Pitt (2003) uses particle filtering and data augmentation to construct likelihood functions for maximum likelihood parameter estimation in diffusion based stochastic volatility models and Durham and Gallant (2002) use a filtering algorithm that resembles the particle filter for estimation of stochastic volatility models.

2.4 Adapting the particle filter to continuous-time models

The previous section shows that the particle filter effectively imposes only two requirements on the state space model: it must be possible to simulate the latent state evolution and evaluate the likelihood function as a function of the latent states and observables. These restrictions are extremely mild, as both of these issues have been of central importance for all of asset pricing. Kloeden and Platen (2001) provides textbook discussions of simulation methods for stochastic differential equations, and Liu and Li (2000) and Mikulevicius and Platen (1988) consider the case of jump-diffusions in detail. Evaluating the likelihood is a combination of simulation and pricing. In the latter case, there is a large literature on closed-form, near-closed form and approximation methods for pricing.

To simulate the state variables forward and evaluate the likelihood function, we use time-discretized solutions to the stochastic differential equations. Assuming that prices are observed at times t and $t+1$, we simulate an additional $M-1$ points between those observations via the Euler scheme

$$
\begin{aligned}
S_{t+\frac{j+1}{M}} &= S_{t+\frac{j}{M}} + \mu^s\left(S_{t+\frac{j}{M}}, X_{t+\frac{j}{M}}\right)M^{-1} + \sigma^s\left(S_{t+\frac{j}{M}}, X_{t+\frac{j}{M}}\right)\varepsilon^s_{t+\frac{j+1}{M}} + Z^s_{t+\frac{j+1}{M}} J^s_{t+\frac{j+1}{M}} \\
X^{(i)}_{t+\frac{j}{M}} &= X^{(i)}_{t+\frac{j-1}{M}} + \mu^x\left(X_{t+\frac{j}{M}}\right)M^{-1} + \sigma^x\left(X_{t+\frac{j}{M}}\right)\varepsilon^x_{t+\frac{j+1}{M}} + Z^x_{t+\frac{j+1}{M}} J^x_{t+\frac{j+1}{M}}.
\end{aligned}
$$

where $j = 1, ..., M-1$, ε^s_t and ε^x_t are mean zero, jointly normally distributed (with constant correlation ρ) with common variance jM^{-1}, J^s_t and J^x_t are Bernoulli random variables with respective intensities $\lambda^s\left(S_{t-M^{-1}}, X_{t-M^{-1}}\right)M^{-1}$ and $\lambda^x\left(X_{t-M^{-1}}\right)M^{-1}$. The jump size distribution is exactly the same as in the continuous-time specification. These additional simulated states are the key to our filtering approach.

While we use an Euler scheme discretization for the jump-diffusion, there are other strong and weak discretization schemes that can be used. The only requirement is the that drift and diffusion coefficients satisfy the assumptions (e.g., differentiability) that justify the various schemes. For the jump times, we use a Bernoulli discretization. If the jumps are

Poisson (constant arrival intensity), the inter-arrival times can be exactly simulated which implies there is no discretization bias in the jump component. When the jump intensity is state dependent, the Bernoulli approximation is straightforward, but there are other algorithms available (see, e.g., Glasserman and Merener 2003).

The time discretization simulates augmented prices and state variables between times t and $t+1$. We collect the augmented values in the following vectors: $X_{t+1}^M = \left(X_t, ..., X_{t+\frac{M-1}{M}}\right)$, $S_{t+1}^M = \left(S_{t+\frac{1}{M}}, ..., S_{t+\frac{M-1}{M}}\right)$, $Z_{t+1}^{k,M} = \left(Z_{t+M^{-1}}^k, ..., Z_{t+1}^k\right)$, and $J_{t+1}^{k,M} = \left(J_{t+M^{-1}}^k, ..., J_{t+1}^k\right)$ where $k = s, x$. These vectors contain the simulated values of the variables in between times t and $t+1$. The entire matrix of latent variables is

$$L_{t+1}^M = \left(X_{t+1}^M, S_{t+1}^M, Z_{t+1}^{s,M}, Z_{t+1}^{x,M}, J_{t+1}^{s,M}, J_{t+1}^{x,M}\right).$$

Note that the latent variables include augmented prices between times t and $t+1$ in S_t^M and that the states are simulated up to one-discretization interval before the next observation. It is on this quantity that we define the particle filter. Note, that conditional on L_{t+1}^M and S_t, the likelihood is conditionally Gaussian.

Given the time-discretization, we now modify the particle filter to handle the continuous-time specification. Conditional on $\left\{\left(L_t^M\right)^{(i)}\right\}_{i=1}^N$, the first stage involves simulating the state variables forward. To do this, first generate jump times, jump sizes and discretized Brownian increments. All of these draws are straightforward as the distributions are conditionally independent. Next, the jump times, jump sizes and Brownian increments feed through the Euler scheme to generate $\left(X_{t+1}^M\right)^{(i)}$ and $\left(S_{t+1}^M\right)^{(i)}$. This provides the propagated state vector, $\left\{\left(L_{t+1}^M\right)^{(i)}\right\}_{i=1}^N$. Finally, these updated states are re-sampled with the appropriate probabilities. The Appendix provides the details of the algorithm.

2.5 Relation to existing literature

Our approach is related to a number of different literatures and in this sub-section, we briefly connect our work to prior related work.

First, our approach complements the recent literature on "realized volatility," see, for example, the review article by Andersen, Bollerslev, and Diebold (2004). The original focus was on using high frequency data to estimate volatility (Andersen, Bollerslev, Diebold and Labys 2001, Aït-Sahalia, Mykland, and Zhang 2005), but recently, Barndorff-Nielson and Shephard (2004, 2006a,b) have derived the limiting behavior of fractional moments and show that this approach can be used to estimate jumps in returns. There is a recent focus on estimating total volatility and separating stochastic volatility from jumps (see, e.g., Andersen, Bollerslev, and Diebold (2005), Barndorff-Nielson and Shephard (2004, 2006a,b), and Huang and Tauchen (2005)).

It is important to understand the differences between our approach and the existing shrinking interval estimators. First, our approaches is firmly parametric while the realized volatility literature is nonparametric. In many financial applications such as option pricing, term structure modeling and optimal portfolio allocations, explicit parametric models are used for applications, and our filtering approach applies in these cases. Second, our approach neither requires nor precludes the use of high frequency data. Third, our approach can be used to filter any latent state variable, not just those that affect the second moment. Thus, for example, we can estimate both time-varying volatility and time-varying expected returns. Finally, our approach provides a method to coherently combine the information in different sources. This allows us to combine the information from different sources, such as option prices and equity prices, to estimate volatility.

Second, Bates (2005) develops a filtering approach that works in models with exponential affine conditional characteristic functions. Bates uses deterministic, analytical approximations to filtering distributions, whereas our approach use discrete particle approximations. Bates (2005) focusses on parameter estimation, as he uses his filtering approach to compute likelihood functions, but does also report the filtered volatility estimates and compares to implied volatility. As Bates' approach relies on analytical approximations, standard gradient algorithms can be used to optimize the likelihood. Our particle approach does not result in differentiable likelihood functions. Our approach for filtering is more general than Bates as we do not rely on closed-form characteristic functions. Thus it applies in non-linear models. Our approach also applies in settings with multiple observed prices. For example, this allows us to jointly filter volatility from returns and options.

Finally, our approach builds on a large literature that develops and applies particle filtering methods in finance and economics. There are now a number of papers applying particle filtering methods to various discrete-time models: Kim, Shephard, and Chib (1998) apply the particle filter to a discrete-time log-stochastic volatility model to construct likelihood functions; Omori, Chib, Shephard and Nakajima (2005) apply the particle filter to a discrete-time stochastic volatility model with leverage effects; Calvet, Fisher, and Thompson (2004) apply the particle filter to a discrete-time switching-Markov model; and in macroeconomics, Fernandez-Villaverde and Rubio-Ramirez (2005) construct likelihood functions with particle filters for discrete-time general equilibrium models. Elerian, Shephard, and Chib (2001) compute marginal likelihoods in continuous-time models using an Euler approximation. Our particle filtering approach provides an alternative for computing likelihoods for model comparison.

2.5.1 Convergence of Particle Filters

Our filtering approach relies on two approximations: discretizing the stochastic differential equation and approximating the filtering distribution with particles. While there are extensive literatures analyzing convergence of each of these components, it is not straightforward to combine the asymptotics. In this section, we briefly discuss the convergence properties of our algorithms.

Our algorithm shares many properties with standard Monte Carlo estimates of expectations of smooth functions of the prices (or states) in continuous-time models. There, sample paths are discretized and expectations are approximated by a sample average across N simulations. Here, we also discretize the sample path at frequency M but the filtering density is approximated by a discrete set of N particles. Since the density can be viewed as an expectation of an indicator function, the approaches are quite similar.

When estimating expectations in continuous-time models using Monte Carlo and time discretizations, there is a natural trade-off between reducing discretization bias (large M) and Monte Carlo error (large N). In the case of estimating expectations of smooth functions, there are standard results on the "efficient" asymptotic trade-off. For example, a standard Euler discretization would imply that N should grow at a rate proportional to M^2, thus cutting the discretization interval in half would quadruple the number of simulations (see, Duffie and Glynn 1995). These results cannot be directly applied to our case because the traditional Monte Carlo literature addresses the convergence of expectations of smooth functions. Here, we are interested instead in pointwise convergence of filtering densities.

First, consider the convergence of the particle filter, ignoring discretization bias (by either assuming that we can exactly simulate from the state transition or that true model is generated by the $M = 1$ Euler discretization of the continuous-time model). In this case, the particle filter has well known convergence properties summarized in Crisan and Doucet (2002). To focus issues, consider only the estimation of the state variable at time t, X_t. If we denote the particle approximation to the filtering density for a fixed N as $p^N(X_t|P_{1:t})$, 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 $\lim_{N\to\infty} p^N(X_t|P_{1:t}) = p(X_t|P_{1:t})$, where $p(X_t|P_{1:t})$ 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 (Crisan and Doucet 2002).

Second, consider the more general case with $M > 1$. If we let $p^{M,N}(X_t|P_{1:t})$ 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 behavior of the filter for large M and N. From the previous results, we know that if the true model is given by a fixed M, the particle filter is consistent. This implies that the difficult step is proving 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 there are numerous results proving convergence of approximations of 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, which is neither smooth nor continuous.

For diffusions, there is a remarkable result in Bally and Talay (1996) proving pointwise convergence of $p^M(X_t|X_{t-1})$ to $p(X_t|X_{t-1})$. 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) combine the Bally and Talay (1996) with standard limiting results to show that an interacting particle system filter is consistent provided that M increases slower than N (for example, $M = \sqrt{N}$).[8] That N grows faster than M is intuitive and it is similar to the results cited above on Monte Carlo estimation of expectations.

If an analog to Bally and Talay (1996) existed for jump-diffusions, the arguments in Jacod, Del Moral and Protter (2002) would apply to derive the limiting distribution of the particle filter in the case of jump-diffusions. However, to our knowledge, there is no analog

[8]The algorithms in Jacod and Del Moral (2001) and Jacod, Del Moral and Protter (2002) do not perform any resampling nor can they adapt (as in Pitt and Shephard 1999) resulting in poor performance in practice.

to Bally and Talay (1996) for jump-diffusion processes. Proving pointwise convergence of the densities for jump-diffusions is an is open problem in stochastic analysis, although it has been conjectured that the Bally and Talay (1996) approach carries over to jump-diffusions by Hausenblas (2002).

3 Empirical Results

In this section, we first analyze the performance of the optimal filter in the context of the double-jump model of Duffie, Pan, and Singleton (2000). Section 3.1 introduces the model, Section 3.2 provides simulation based evidence for the SV and SVJ models, and Section 3.3 analyzes the informational content of equity returns and options regarding volatility.

3.1 Stochastic volatility with jumps in prices and volatility

We assume that the equity index (S_t) and its stochastic variance (V_t) jointly solve

$$d\log(S_t) = (r_t + \eta_v V_t - V_t/2)\,dt + \sqrt{V_{t-}}dW_t^s + d\left(\sum\nolimits_{j=1}^{N_t} Z_j^s\right)$$
$$dV_t = \kappa_v(\theta_v - V_t)\,dt + \sigma_v\sqrt{V_{t-}}dW_t^v + d\left(\sum\nolimits_{j=1}^{N_t} Z_j^v\right)$$

where W_t^s and W_t^v are Brownian motions, η_v is the volatility risk premium, $Z_j^s = \mu_s + \rho_s Z_j^v + \sigma_s\varepsilon_j$, $\varepsilon_j \sim N(0,1)$, $Z_j^v \sim \exp(\mu_v)$ and $N_t \sim Poi(\lambda t)$. We let SV denote the special case with $\lambda = 0$, SVJ the case with $\mu_v = 0$ and SVCJ denotes the general model. Throughout, we set $\rho_s = 0$, as this parameter is difficult to estimate (Eraker, Johannes and Polson (2003) and Chernov, Gallant, Ghysels and Tauchen (2003) find it negative, but insignificant). This also makes the filtering problem more difficult because there is a lower signal-to-noise ratio between jumps in volatility and returns.

This model has received significant attention as a model of equity index returns. Broadie, Chernov, and Johannes (2005), Chernov, Gallant, Ghysels and Tauchen (2003), Eraker

(2004), and Eraker, Johannes and Polson (2003) find support for this specification. This model also provides closed-form (up to the numerical solution of an ordinary differential equation) solutions for optimal portfolio weights and option prices. Second, this model contains a very general factor structure. It includes an unobserved diffusive state (V_t), unobserved jumps in the observed state (Z_j^s) and unobserved jumps in the unobserved state $\left(Z_j^v\right)$. Because of this complicated factor structure, the model provides a rigorous hurdle for our filtering approach.

3.2 Simulation based results

3.2.1 SV Model

To analyze the performance of our filtering algorithm as a function of the degree of data augmentation (M) and the sampling frequency of returns (daily, weekly, monthly), we simulated 100 sample paths from the continuous-time model using the Euler scheme and 100 data points per day. We sample the resulting process daily, weekly or monthly and compute the optimal filter for $M = 1, 2, 5, 10, 25$, and 100.[9] Since it is impossible to compute the true filter, we regard $M = 100$ as the true filter, in the sense that it is devoid of discretization error.

We summarize the filter's performance with two sets of statistics. The first is the root-mean-squared (RMSE) and mean-absolute error (MAE) between the approximate filtered means for a given M and the true filter, thus we compare the difference between $E^M\left[V_t|S_{1:t}\right]$ and $E\left[V_t|S_{1:t}\right]$. This is consistent with the literature and the fact that $E\left[V_t|S_{1:t}\right]$ is the minimum variance estimator of V_t. We also report the RMSE and MAE between the filtered means and the true simulated volatilities. These results are generally of less interest (as

[9] As an example of the computing time, the filter (N=10,000, M=10) took about 3 minutes of CPU time on a Xeon 1.8MHz processor in Linux with code written in C.

the optimal filtering distribution is our target), but they do provide a intuitive sense of how discretization bias and sampling frequency effects the accuracy of volatility estimates.

Tables 1 and 2 summarize the performance for two different parameter configurations, for daily, weekly, and monthly data, and for different degrees of data-augmentation. Table 1 reports the filtering results using the following "base case" for the parameters: $\theta_v = 0.9$, $\kappa_v = 0.02$, $\sigma_v = 0.15$, and $\rho = 0$. The units are in daily variances. For example, $\theta_v = 0.9$ corresponds to about 15 percent annualized ($\sqrt{\theta_v \cdot 252}$). The volatility parameters are consistent with the estimates in Eraker, Johannes, and Polson (2003). We constrain $\rho = 0$ to reduce the signal to noise ratio, making it more difficult to estimate volatility. Table 2 reports filtering results for different parameters that generate high and volatility variance paths ($\kappa_v = 0.08$, $\theta_v = 1.8$, and $\sigma_v = 0.30$).

Table 1 documents that for a fixed sampling interval (daily, weekly, or monthly), increasing M reduces the discretization bias, improving the accuracy of the filter. For monthly data, the RMSE and MAE fall drastically as M increases: they are roughly halved when M increases from 1 to 2, from 2 to 5 and from 5 to 10. This occurs because the conditional distribution of volatility increments, $p(V_{t+1}|V_t)$, is non-central χ^2 for all sampling intervals, but is very close to normal for high frequencies (daily data) but very non-normal for lower frequencies. In fact, the non-normalities increase and for typical parameterizations peak at frequencies close to the monthly range (Das and Sundaram 1999).

While data augmentation improves the performance for all sampling frequencies, the effects are extremely modest for the daily sampling frequency. The RMSE for daily data falls from 0.25 for $M = 1$ to 0.20 for $M = 2$. The results are similar for MSE, indicating that discretization bias is not a first order effect for data sampled daily. This is consistent with papers that analyze parameter estimation in continuous-time models and document via simulations that any effect of discretization bias on parameter estimates is swamped by sampling variation, at least for the sample sizes that we encounter in practice (see,

Table 1: Simulation results for the SV model using the parameters $\theta_v = 0.9$, $\kappa_v = 0.02$, $\sigma_v = 0.15$, and $\rho = 0$. R_1 (R_2) and M_1 (M_2) are RMSE and MAE errors between the filtering density for a given M and the true filtering density (or the true volatilities). The numbers are multiplied by 10 and throughout we use $N = 10,000$.

	Monthly				Weekly				Daily			
M	R_1	R_2	M_1	M_2	R_1	R_2	M_1	M_2	R_1	R_2	M_1	M_2
1	2.76	7.28	2.47	5.86	0.93	5.44	0.73	4.15	0.25	4.14	0.19	3.07
2	1.08	6.84	0.95	5.27	0.30	5.40	0.20	4.05	0.20	4.15	0.14	3.06
5	0.41	6.76	0.33	5.08	0.23	5.40	0.14	4.04	0.21	4.15	0.14	3.06
10	0.22	6.76	0.18	5.04	0.20	5.40	0.13	4.04	0.19	4.14	0.13	3.06
25	0.17	6.76	0.13	5.02	0.19	5.40	0.13	4.04	0.19	4.15	0.13	3.06

e.g., Pritsker 1998). This formalizes the intuition from Merton (1980) who showed that, in a constant volatility model, volatility estimates are improved by increasing the sampling frequency. This is why it is common to use daily or even intradaily data to learn about volatility (see Andersen, Bollerslev and Diebold, 2004, and Barndorff-Nielson and Shephard, 2004, for further results on the use of ultra high frequency data).

The results are sensitive to parameter values, in large part due to the fact that the degree of non-normality in $p(V_{t+1}|V_t)$ varies with the parameter values, especially κ_v and σ_v. When these parameters are small, there is little discretization bias in approximating the solution to the stochastic with the $M = 1$ Euler scheme. To quantify this affect, Table 2 considers a set of parameters with higher average volatility ($\theta_v = 1.8$), faster mean-reversion ($\kappa_v = 0.08$), and higher volatility of volatility ($\sigma_v = 0.30$). These parameters are consistent with volatility processes for individual stocks (see Bakshi and Cao 2004). Table 2 indicates, as expected, that there are larger gains to data augmentation and, again, the

Table 2: Simulation results for the SV model using the parameters $\kappa_v = 0.08$, $\theta_v = 1.8$, and $\sigma_v = 0.30$. R_1 (R_2) and M_1 (M_2) are RMSE and MAE errors between the filtering density for a given M and the true filtering density (or the true volatilities). The numbers are multiplied by 10 and throughout we use $N = 10,000$.

	Monthly				Weekly				Daily			
M	R_1	R_2	M_1	M_2	R_1	R_2	M_1	M_2	R_1	R_2	M_1	M_2
1	6.44	12.39	6.28	10.47	2.17	10.03	1.82	7.95	0.81	8.41	0.62	6.46
2	3.02	11.01	2.89	8.97	0.52	9.85	0.43	7.70	0.32	8.37	0.24	6.41
5	0.80	10.62	0.72	8.36	0.26	9.84	0.21	7.67	0.28	8.38	0.21	6.41
10	0.30	10.60	0.25	8.27	0.24	9.85	0.19	7.67	0.29	8.39	0.22	6.41
25	0.22	10.60	0.17	8.25	0.24	9.85	0.18	7.67	0.28	8.37	0.21	6.40

gains are much greater for less frequent sampling. The results show that even a modest amount of data augmentation (M=2) with daily data drastically reduces the errors.

Figures 1 and 2 provide a graphical depiction of the performance of the particle filtering algorithm. The top panels of Figures 1 and 2 display simulated returns, the middle panels display the true volatilities (dots) and posterior means for M=1, 2, 5, and 25. The bottom panel displays the difference between the true filtered posterior mean and the simulations for M=1, 2, 5, and 25. Figure 1 samples daily while Figure 2 samples monthly.

There are a number of noticeable results. First, errors at both frequencies and for all degrees of data augmentation are greater when volatility is higher. Second, the filtered estimates typically under-estimate volatility when volatility is high and over-estimate volatility when it is low. This is clearly the pattern in Figures 1 and 2. This occurs because the estimates are smoothed expectations based on past returns and volatility is mean-reverting.

Third, economically, there is no differences across the filtered volatility for various degrees of data augmentation when the data is sampled daily. In fact, in the figure it is impossible to see the different estimates in the middle panel as the largest differences are on the order of 0.05 when V_t is over 1.5. In Figure 2, with monthly sampling, data augmentation is economically important. For example, the large errors for $M = 1$, are drastically reduced via data augmentation. For example, around data points 375, the true variance is very reasonably high around four (30 percent annualized volatility) and the filtered estimate for $M = 1$ is more than 1 variance unit lower than the other filtered estimates with $M > 1$ estimates. This shows how important data-augmentation is with infrequently sampled data.

Figure 3 displays a smoothed version of the posterior distribution $p(V_t|S_{1:t})$ as we vary the degree of data augmentation. Here, the data is sampled monthly and the figure displays how the shape of the filtering distribution changes with the degree of data augmentation. The distribution $p(V_t|S_{1:t})$ is close to normal for $M = 1$, but clearly becomes increasingly non-normal (positively skewed) as the degree of data augmentation increases. In the case of $M = 1$, we do not fill in any data and essentially assume the conditional distribution of log returns and volatility increments are conditionally normal. When M is increased, the model generates non-normal conditional distribution for returns and volatility which directly appear in $p(V_t|S_{1:t})$ via equations (7) and (8). Although this figure shows an average volatility day, the effect is even stronger during extremely high or low volatility periods, when the affect of mean-reversion is stronger. There is a similar effect on the filtered distribution of returns. Although not included, the kurtosis increases with data augmentation as the approximations to the conditional distributions are more accurate.

3.2.2 Incorporating jumps

This section considers a model with jumps in prices and stochastic volatility, in order to understand how the filter performs in separating jumps from stochastic volatility. The

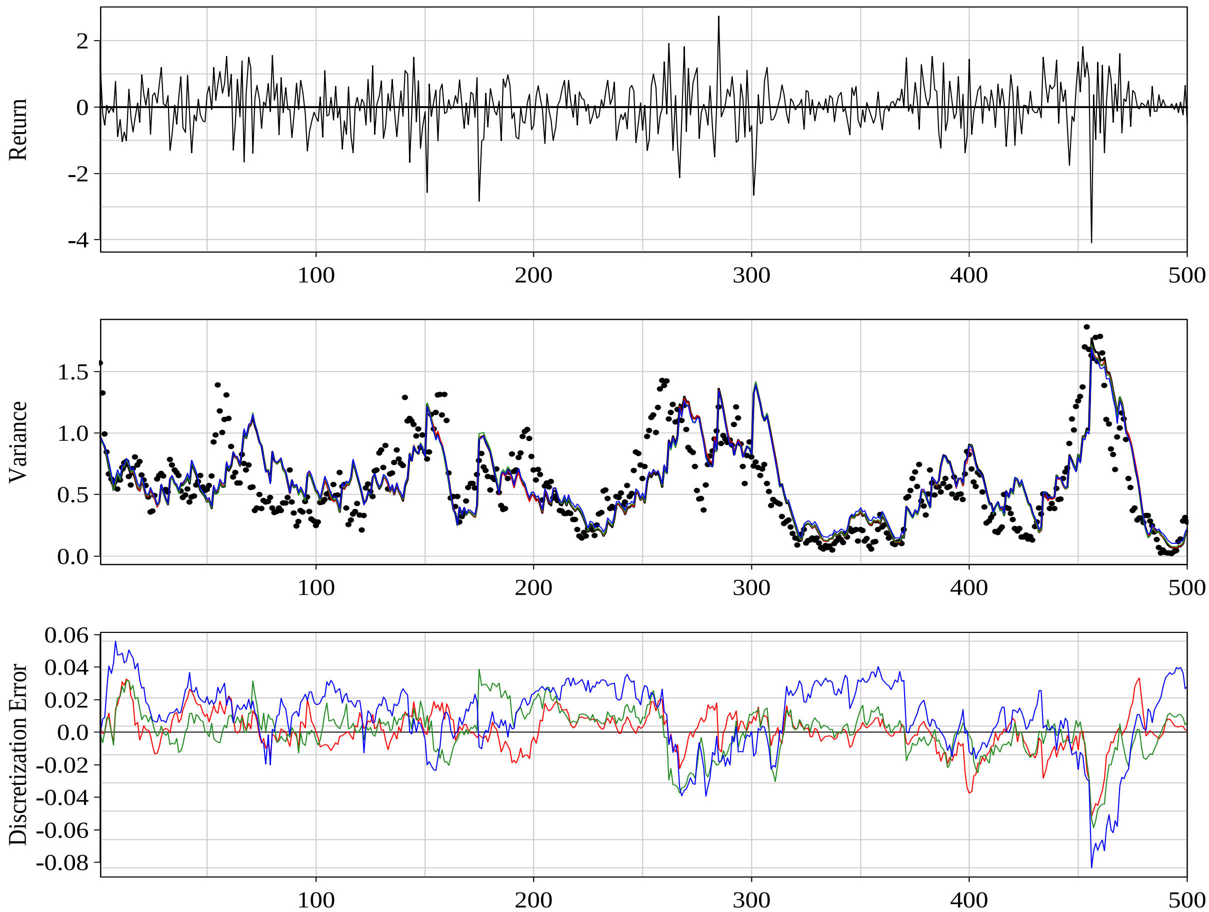

Figure 1: A representative simulated sample path of daily returns (top panel), posterior means of the filtered volatility distribution for M=1, 2, 5, 100 (the true volatilities are the dots), and discretization error for the difference between the posterior means and the true distribution for M=1, 2, 5, 100.

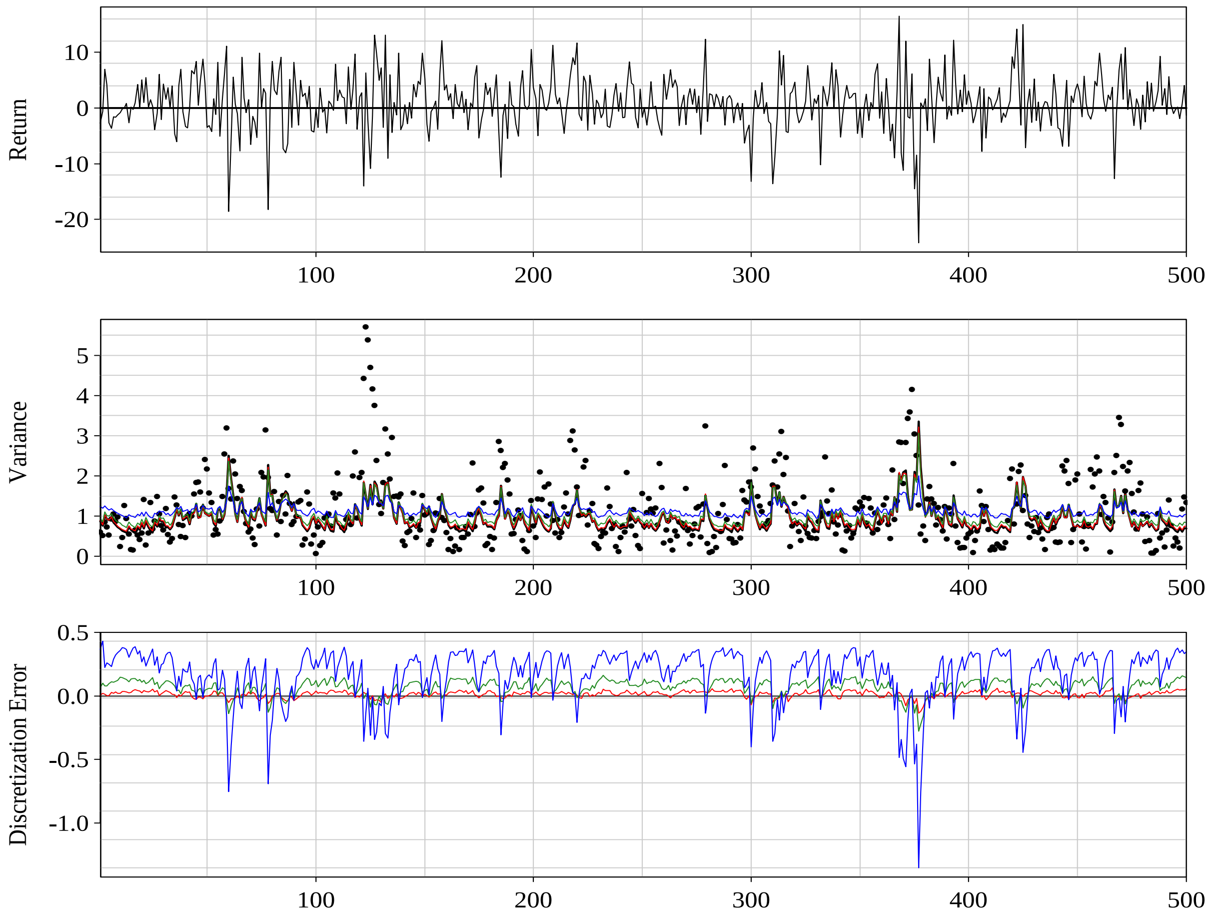

Figure 2: A representative simulated sample path of daily returns (top panel), posterior means of the filtered volatility distribution for M=1, 2, and 5 (the true volatilities are the dots), and discretization error for the difference between the posterior means and the true distribution for M=1, 2, and 5.

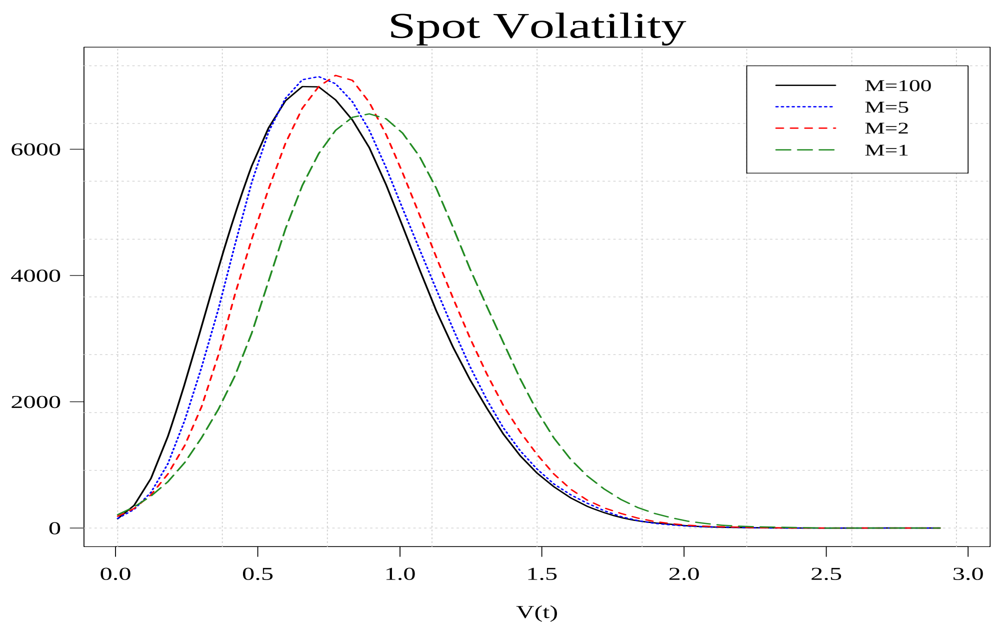

Figure 3: Filtering densities for the stochastic volatility model as a function of the degree of data augmentation (M). The densities were smoothed using kernels.

SVJ model adds an additional layer of complexity because the "volatility" of returns is now generated by two components: a slow-moving stochastic volatility component and infrequent, but large jumps. This complicates filtering because large movements could be generated by either high volatility or jumps. The parameters in the SVJ model are again based on Eraker, Johannes, and Polson (2003): $\mu = 0.05$, $\theta_v = 0.8152$, $\kappa_v = 0.02$, $\sigma_v = 0.10$, $\mu_y = -2.5$, $\sigma_y = 4.0$ and $\lambda_y = 0.006$.

Table 3 summarizes the filtering results for volatility, jump sizes, and jump times for daily and weekly frequencies. The results on jump estimation were so poor at the monthly frequency, there are not reported. For the variances and jump sizes, the first two panels provide results provide the RMSE and MSE distances between the approximate filtered posterior means for a given M and either the true filter or the true state, as in the previous section. The volatility filtering results are qualitatively similar to those in the previous section. Since jumps capture the largest movements, σ_v and θ_v are lower, which makes volatility more slowly moving, reducing the benefits of data-augmentation.

Turning to the filtered jump sizes, the results indicate that there is little benefit to data augmentation at either the daily or weekly frequency, thus R_1 and M_1 decrease very slowly. Similarly, there is no increase in the precision of the jump size estimate vis-a-vis the true jump sizes as augmentation increases. Since jumps are rare (about $0.006 \cdot 252 = 1.6$ per year), it is extremely unlikely that more than one jump would occur over either the daily or weekly frequency, and thus there are few gains to data augmentation. The jumps are far more precisely estimated using daily data as all of the error metrics are much smaller using daily data than weekly data. This is not at all surprising.

To quantify jump time estimates, we use two statistics to quantify the types of errors that could be made. If the filtered posterior probability of a jump is greater than 50 percent, we assume that a jump occurred over that sampling period. The first statistic quantifies how data augmentation affects jump time estimates. This statistic, labeled Hit_1, gives the percentage of jump times identified by the approximate filter that were identified by the true (M=100) filter. For example, a number of 90 for $M = 2$ indicates that of the total jump times identifies by the true filter, 90 percent of these jump times were identified by the approximate filter filling in one additional data point between observations. This quantifies the discretization bias. The second statistics, Hit_2, gives the percentage of actual or simulated jumps that were correctly identified by the approximate filter. This shows how accurately one can identify true jump times.

The results indicate that it is very difficult to estimate jump times at a weekly frequency. A comparison of the distance measures to the truth (R_2 and M_2) for daily and weekly frequencies show that RMSE (MSE) for weekly data is about five (12) times as large. Hit_2 indicates that with daily data, about 60 percent of the jump times are correctly identified.

This should not be surprising. Since $Z_j \sim N(2.5, 4^2)$, just over 40 percent of the jump sizes are between ±2.5 percent. Since daily diffusive volatility is about 0.9 ($\sqrt{0.8}$), more than 40 percent of the jumps are not "significant" as they are less than a three standard deviation diffusive shock. In addition, volatility changes over time, and thus it is not surprising that even with daily data, only about 60 percent of the jumps are identified.[10]

The ability to correctly identify jumps rapidly decreases for longer sampling intervals. For weekly data, only 30 percent of the true jump times are correctly identified. Via the aggregation in the stochastic differential equation, total diffusive return volatility over a week is roughly $\sqrt{5} \approx 2.25$ times larger than daily volatility. If daily volatility is, for example, about one percent, a three percent move is a statistical outlier. For weekly data, this requires an almost seven percent move, showing why far fewer jumps are correctly identified using weekly data. This indicates that even when the parameters are known, it is very difficult to identify jumps are weekly. Although most papers estimating models with jumps, even many of the earliest ones (e.g., Beckers 1981, Jarrow and Rosenfeld 1984) use daily data, there are papers that use less frequently sampled data. For example, Andersen, Benzoni, and Lund (2003) use weekly data to estimate a term structure model with jumps while Das and Uppal (2004) use monthly data to estimate a jump-diffusion model on equity returns. The results here indicate that it is likely very difficult to identify jump components at these frequencies.

Figure 4 provides a graphical view of the filtering results for a randomly selected simulation of 2000 days for the SVJ model with filtered estimates using $M = 10$ and $N = 10000$. 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

[10] This raises a subtle specification issue. The normally distributed jump specification implies that many of the jumps are going to be small. An alternative specification which constrains all jump movements to be large might be preferred.

(dots) and the filtered jump size estimates. The confidence bands on the volatility state are typically about 2 percent on each side of the posterior mean. This estimation risk has always been a major concern, given the important role that volatility estimates play in finance applications. In fact, Merton (1980) argues that the "Most important direction is to develop accurate variance estimation models which take into account of the errors in variance estimates" (p. 355). We discuss this in greater detail below.

Our algorithm does an excellent job identifying the large jumps, but it misses many of the smaller jumps that are less than about 2 percent in absolute value. . The missed jumps are too small to be identified as the algorithm cannot separate them from normal, day-to-day movements. These results are important for practical applications, as the results show that the algorithm is effective in identifying important (i.e., reasonably large sized jumps) movements.

The results for the SVCJ model are similar to those for the SVJ model and are not reported. At daily frequencies, it is possible to identify both return and volatility jumps as their occurrence in coincidental. In effect, the jump in returns signals that volatility has increased. If the occurrence is not coincidental, as is the case in one of the double-jump models in Duffie, Pan and Singleton (2000), then it is possible to estimate return jumps, but estimates of volatility jumps are not reliable.

3.3 The Information in Returns and Options

Most approaches for estimating volatility and jumps rely exclusively on returns, ignoring the information embedded in option prices. In principle, options are a rich source of information regarding volatility, which explains the common use of Black-Scholes implied volatility as a volatility proxy for practitioners. In contrast to this is the common academic approach of only using returns to estimate volatility.

In this section, we consider volatility estimates using returns and, additionally, option

Table 3: Simulation results for the SVJ model. The summary statistics for variance estimation are the same as in the previous tables. Hit_1 reports the percentage of jumps (as identified by the true filter) that are identified by the approximate filter. Hit_2 identifies the percentage of actual jumps that are identified by the approximate filter.

	Weekly				Daily			
M	R_1	R_2	M_1	M_2	R_1	R_2	M_1	M_2
				Variances				
1	0.020	0.365	0.016	0.282	0.020	0.281	0.015	0.214
2	0.016	0.364	0.012	0.282	0.020	0.281	0.014	0.214
5	0.015	0.364	0.012	0.282	0.020	0.281	0.014	0.214
10	0.015	0.364	0.012	0.282	0.021	0.281	0.015	0.214
25	0.015	0.364	0.012	0.282	0.019	0.281	0.014	0.214
				Jump Sizes				
1	0.016	0.588	0.005	0.132	0.005	0.120	0.001	0.011
2	0.014	0.587	0.005	0.131	0.004	0.120	0.001	0.011
5	0.013	0.587	0.005	0.131	0.004	0.120	0.001	0.011
10	0.013	0.587	0.005	0.131	0.004	0.120	0.001	0.011
25	0.013	0.587	0.005	0.131	0.004	0.120	0.001	0.011

	Jump Times			
	HIT_1	HIT_2	HIT_1	HIT_2
1	98.6	29.5	98.6	60.2
2	99.1	29.5	98.6	60.2
5	99.1	29.5	100	61.1
10	99.1	29.5	100	61.1
25	99.5	29.6	100	61.1

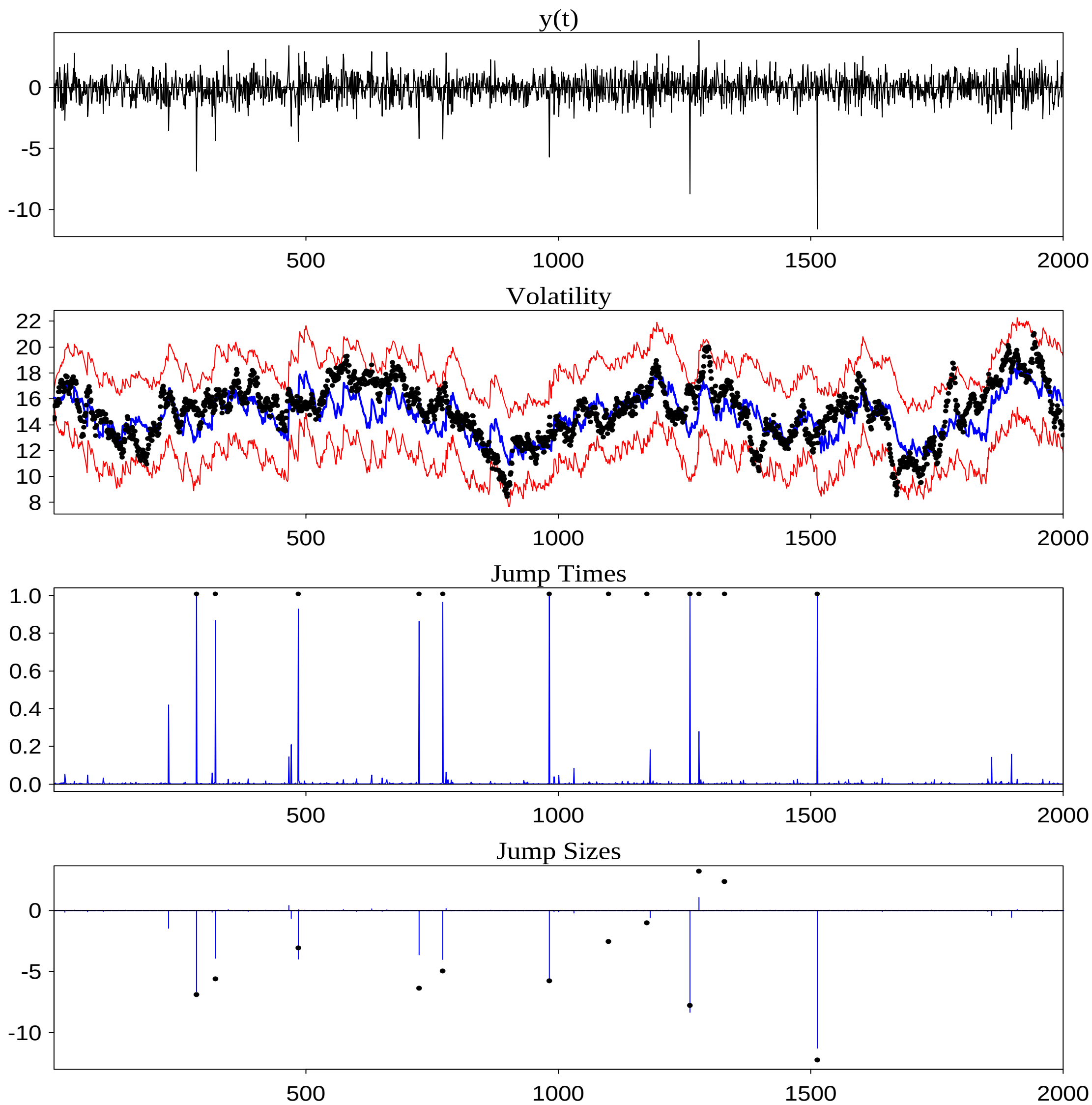

Figure 4: Performance of the filtering algorithm using simulated daily data. The top panel displays returns, the second panel displays the simulated volatilities (dots) and the $(10, 50, 90)$ quantiles of the filtering distribution, the third panel displays filtered jump times (true jumps displayed as a dot), and the fourth panel displays the filtered jumps sizes (true jump sizes are dots). Prices were simulated from an Euler scheme with 100 time steps per day and the algorithm was run for M=10 and N=10000.

prices. Our goal is to quantify the informativeness of options regarding volatility and also to examine whether the informational content of the two sources is consistent, as a good model would indicate.

To understand why option prices are, in principle, so informative about volatility, it is useful to recall the arguments in Hull and White (1987). Hull and White (1987) indicate that if there is stochastic volatility, then under certain conditions, the variance implied from the Black-Scholes model (Black-Scholes implied volatility) at time t for an an-the-money option that expires at time T, denoted, $\sigma_{t,T}^2$, is the average expected variance to maturity. If we assume there are no jumps or leverage effects, then their result implies that

$$\sigma_{t,T}^2 = (T-t)^{-1} E_t^{\mathbb{Q}} \left[\int_t^T V_s ds \right],$$

where $\mathbb{Q}$ is the equivalent martingale measure. Since option prices are one-to-one with implied volatility, this implies that option prices (at least at-the-money ones) are determined by expected variance over the life of the option contract.

Assuming a specific parametric model, the right-hand-side can be computed analytically. For example, in the case of square-root stochastic variance with jumps in variance (but not in prices), the Black-Scholes implied volatility relates the underlying parameters and

$$\sigma_{t,T}^2 = (T-t)^{-1} E_t^{\mathbb{Q}} \left[\int_t^T V_s ds \right] = \left(\theta_v^{\mathbb{Q}} + \lambda^{\mathbb{Q}} \mu_v^{\mathbb{Q}}\right) + \frac{V_t - \left(\theta_v^{\mathbb{Q}} + \lambda^{\mathbb{Q}} \mu_v^{\mathbb{Q}}\right)}{\kappa_v^{\mathbb{Q}} (T-t)} \left(e^{-\kappa_v^{\mathbb{Q}}(T-t)} - 1\right).$$

This relation does not hold exactly with jumps in prices, however, it does provide a number of general points. For short-dated options, $(T-t)$ is very small and $\frac{e^{-\kappa_v^{\mathbb{Q}}(T-t)}-1}{\kappa_v^{\mathbb{Q}}(T-t)}$ is close to one. This implies that $\sigma_{t,T}^2 \approx V_t$, which, in large part, is the reason why it is common to state that options are so informative about volatility. Second, the informational content of implied volatility in the context of a formal model crucially depends on model specification and accurate estimates of risk-neutral parameters. Misspecification or poorly estimated risk premia could results in directionally biased estimates of V_t based on options. Finally, index option prices generally contain a relatively large bid-ask spread relative to futures or individual stocks, as noted by Bakshi, Cao, and Chen (1997) or George and Longstaff (1993). This implies that while option prices are informative, they made be quite noisy as a measure of volatility.

We filter jumps and volatilities from the following system:

$$\begin{aligned} \log\left(C_t^{Mar}\right) &= \log\left(C^{Mol}\left(S_t, V_t, \Theta^{\mathbb{P}}, \Theta^{\mathbb{Q}}\right)\right) + \varepsilon_t. \\ d\log\left(S_t\right) &= \left(r_t + \eta_v V_t - V_t/2\right)dt + \sqrt{V_{t-}}dW_t^s + d\left(\sum\nolimits_{j=1}^{N_t} Z_j^s\right) \\ dV_t &= \kappa_v\left(\theta_v - V_t\right)dt + \sigma_v\sqrt{V_{t-}}dW_t^v + d\left(\sum\nolimits_{j=1}^{N_t} Z_j^v\right), \end{aligned}$$

where $\Theta^{\mathbb{P}}$ are the structural parameters of the system, $\Theta^{\mathbb{Q}}$ are the risk premiums, and ε_t is the pricing error which captures bid-ask spreads. We assume that $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$ and assume $\sigma_\varepsilon = 0.10$ (ten percent). In general, the higher the bid-ask the less informative option prices are about volatility, which implies our pricing error is conservative and will not bias our findings.

We focus on two issues. The first is the general informativeness of the two data sources. We filter volatilities using only returns and then using returns and options jointly and compare the average posterior standard deviations. That is, we compare the average posterior standard deviations of $p\left(V_t|P_{1:t}\right)$ and $p\left(V_t|P_{1:t}, C_{1:t}\right)$. Since options are informative about volatility, we expect the posterior standard deviations to be smaller with options. One advantage of our approach is that we are able to numerically quantify the information content. Second, we compare the filtering distributions with and without options to analyze the extent to which the information in returns is consistent with the information in option prices. The existing literature typically compares Black-Scholes implied volatility to future realized volatility in order to understand the predictive content of implied volatility. This literature is firmly non-parametric in the sense that specific models or risk premium estimates are not considered. Here, we are interested in understanding the informational content of the two sources in the context of a model that has been shown to fit both returns and options well.

To implement the filter, we S&P 500 index returns and options on the S&P 500 index futures contract. We use the at-the-money call option contract with maturity greater than one week. The options are American and thus it important to account for the possibility of early exercise. We use the American adjustment feature in Broadie, Chernov, and Johannes (2005) which uses an extremely accurate binomial Black-Scholes adjustment. The sample period is from 1987 to 2003. We use $N = 50,000$ and $M = 10$. We estimate the model using only objective measure parameters and also estimate the model using the risk premium estimated in Broadie, Chernov, and Johannes (2005). BCJ (2005) find a pure stochastic volatility model (with no jumps) to be so misspecified that we do not consider it separately here.

The results are in Table 4 and Figures (5) and (6) which display the filtering distributions for the SVCJ model. Regarding informativeness, Table 4 indicates that the posterior standard deviation falls drastically in each model when options are added. The decrease is greatest when we incorporate risk premia, as the posterior standard deviation is roughly 40 percent lower in the SVJ and SVCJ models. This decrease in estimation uncertainty occurs despite the fact that we assumed a rather large pricing error variance. These results show the precision of the information embedded in options regarding volatility and justifies the common perception that options are highly informative about volatility.

Regarding consistency, the results are not as strong. First, assuming no risk premia we see from Table 4 that the upper 95 percent tail of the filtering distribution using only returns is typically lower than the lower 5 percent tail of the filtering distribution using options and returns. This is most easily seen in Figure (5) where the green lines (which represent the $(5, 50, 95)$ percent quantiles of the filtering distribution using options and returns) are generally higher than those obtained only from returns. Whale the two series tend to move together (spikes at similar points, e.g.), the properties tend to be very different.

Table 4: Filtering results using returns and returns and option prices for the SVJ and SVCJ models.

Model	Risk Premiums	$std\,(V_t \vert P_{1:t})$	$std\,(V_t \vert P_{1:t}, C_{1:t})$	$Prob(P95 > Q05)$
SV	No	0.354	0.235	0.061
SVJ	No	0.298	0.228	0.144
	Yes	0.298	0.175	0.560
SVCJ	No	0.384	0.256	0.117
	Yes	0.384	0.233	0.274

A few examples provide the necessary intuition. In late 1990, the option based estimates of V_t were higher than those based on returns. This occurred after Iraq's invasion of Kuwait and imply the option based information anticipated larger moves that were subsequently observed. Next, note that the level of volatility from the two sources were broadly consistent in the low volatility period from 1993 to 1996. From 1997 onward, the estimates based on options are substantially higher than those based only on returns. Finally, the estimates using options have more short-lived spikes up and down. This could, in part, be due to data issues related to changing option contracts upon expiration (moving from a one-week to five-week option).

Figure 5 incorporates jump risk premia based on the estimates in Broadie, Chernov, and Johannes (2005). The jump risk premia reduce μ_s (from about -4 percent to -6 percent) and increase σ_s (from about 3 percent to 7 percent) and μ_v (from 1.5 to 7). Figure 6 displays the results incorporating risk premia which have the net result of increasing risk neutral expected variance and thereby generally decreasing option based estimates of V_t. The two data sources are now generally consistent from 1993 to 1996, however, the options based information still implies a higher volatility for the pre 1993 period and the post 1997

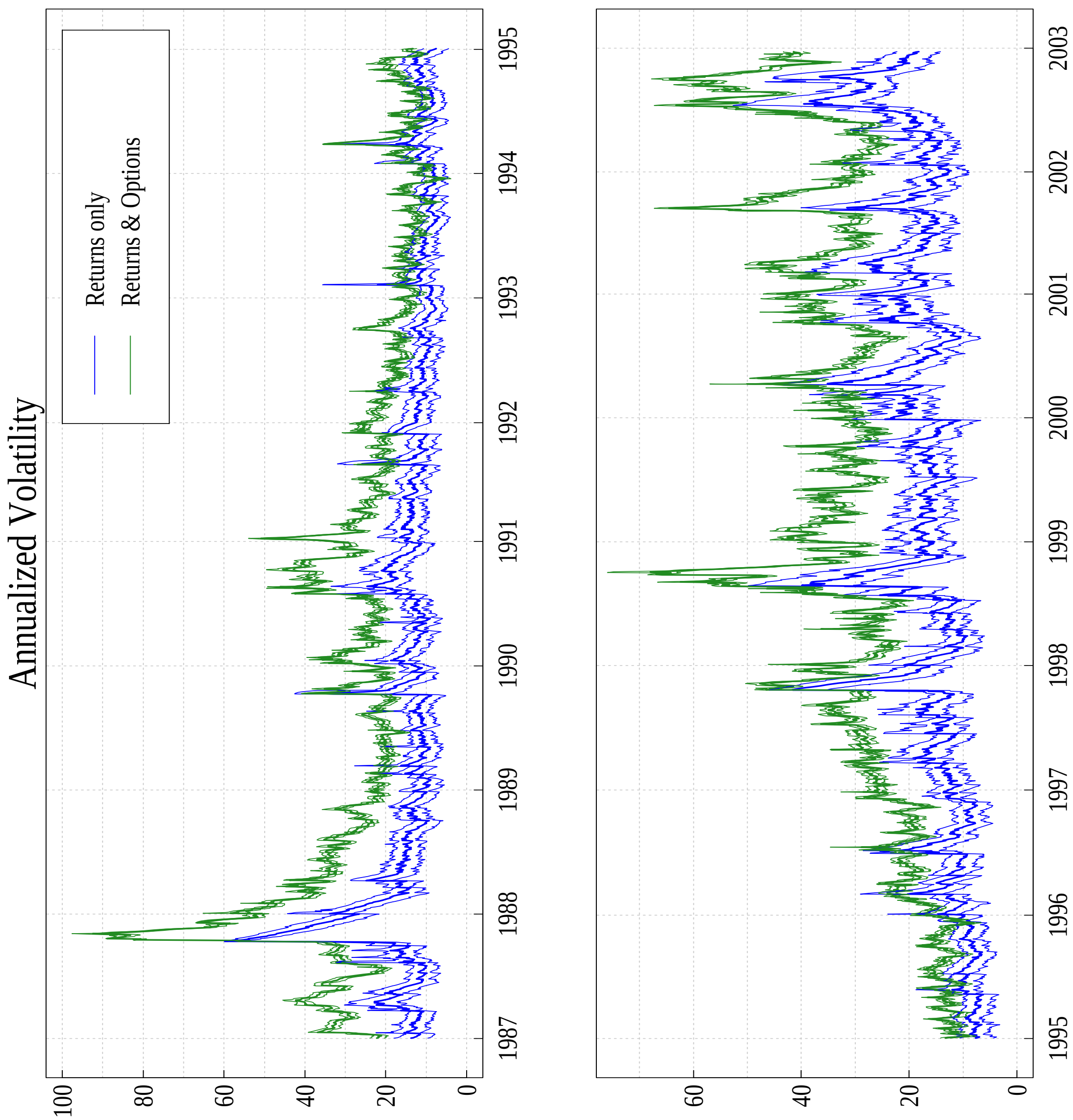

Figure 5: Filtered volatility using only returns data and both returns and options data. For each we include the posterior means and a $(5, 95)$ confidence band. The darker lines are using only returns (generally the higher of the two sets of lines) and the lighter lines are using both returns and options. We assume no risk premia.

period, indicating that the two sources are not generally consistent.

The main explanation for the difference is model misspecification, in terms of the specification of the stochastic evolution under the objective measure, $\mathbb{P}$, or time-varying risk premia. The main problem is that actual volatility (whether under $\mathbb{P}$ (returns) or $\mathbb{Q}$ (options)) is generally high for long periods of time (1987 to 1992 and 1997 to 2003) and low for long periods of time (1992 to 1996). In the model specified above, the long run mean of V_t is constant (θ_v in the SV and SVJ models, $\theta_v + \mu_v \lambda / \kappa_v$). It is unlikely that V_t would remain above or below its long run mean for an extended period of time. Duffie, Pan, and Singleton (2000) introduced a model with a time-varying long-run diffusive variance mean which could capture this feature. We find this to be a plausible explanation. Second, time-varying risk premia provide an alternative explanation. This implies that the risk-neutral parameters generating option prices vary over time, due to, for example, time-varying risk aversion or business cycle risk. While this could generate the required effects, the does not explain why these parameters vary across time, and is somewhat unsettling. Moreover, this explanation would not explain why volatility estimates based only on time series display a similar pattern of alternating low and high periods.

4 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 and index option returns.

In future work, we plan a number of methodological and empirical extensions. Methodologically, the logical net step is to consider sequentially learning about state variables and

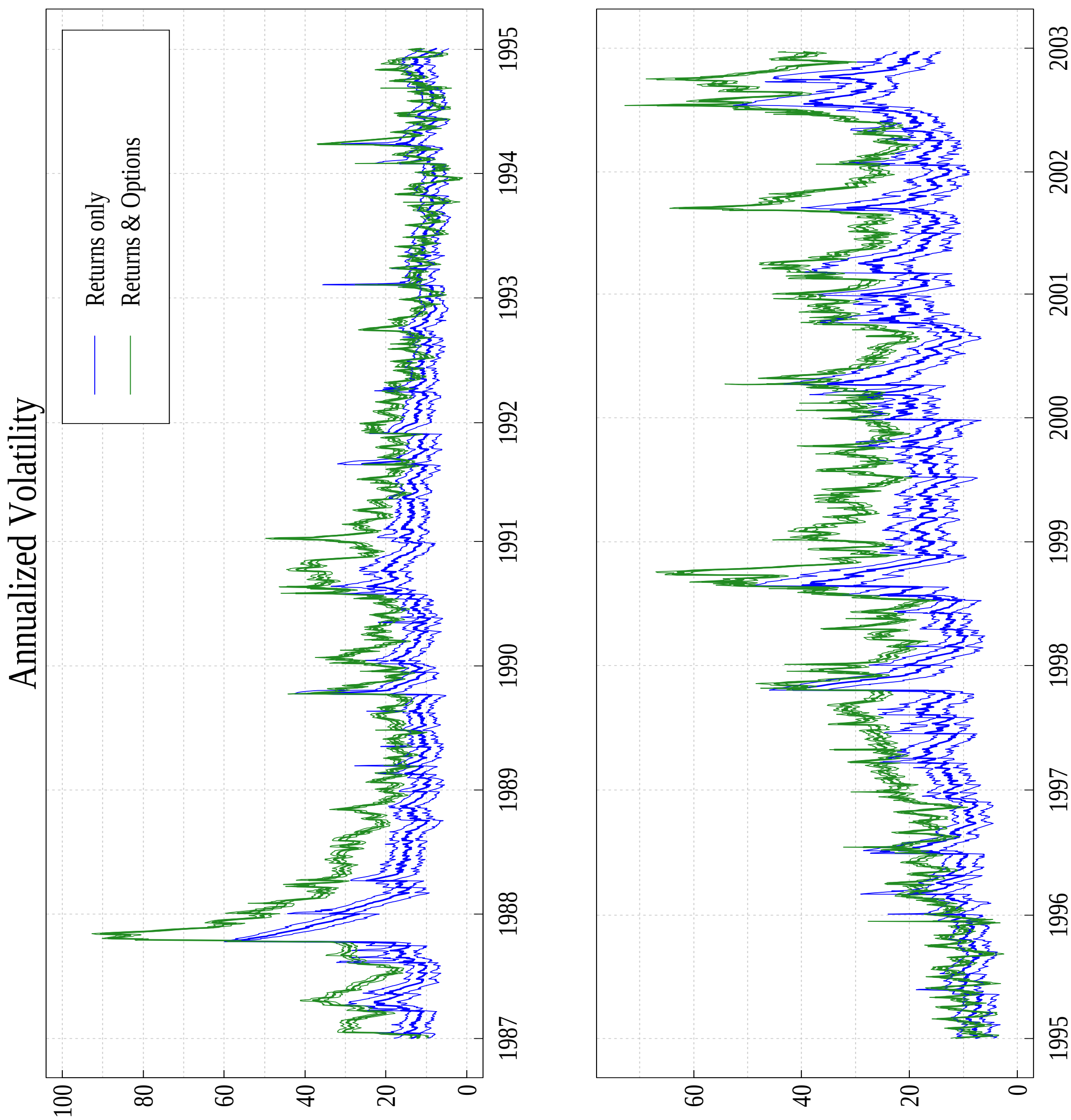

Figure 6: Filtered volatility using only returns data and both returns and options data. For each we include the posterior means and a (5, 95) confidence band. The darker lines are using only returns (generally the higher of the two sets of lines) and the lighter lines are using both returns and options. We assume the risk premia estimated in Broadie, Chernov, adn Johannes (2005).

parameters. Johannes, Polson, and Stroud (2004) is a first step in this direction. Another interesting extension is to apply particle filtering to continuous-time models with small jumps, such as those driven by Levy processes (see, e.g., Barndorff-Nielson and Shephard, 2006b). Empirically, an interesting extension would be to compare the accuracy of volatility and jump estimates obtained from our filtering methodology to those obtained from high-frequency data. Finally, the empirical results in the final section indicate that the SVJ and SVCJ model are misspecified, as options and returns do not provide the same information about volatility. It would be interesting to examine more flexible classes of models with time-varying risk premia or long-run volatility levels.

Appendix: Full Particle Filtering Algorithm for Jump-Diffusions

For simplicity, we first describe the full algorithm in the case of no derivative prices and then discuss how to adapt the procedure to deal with derivatives.

1. Given $\left\{ \left(L_{t+1}^{M}\right)^{(i)}, \pi_{t+1}^{(i)} \right\}_{i=1}^{N}$, simulate the latent state vector and latent prices forward. This requires the following steps. For $j = 1, ..., M-1$, conditional on $S_{t+(j-1)M^{-1}}^{(i)}$ and $X_{t+(j-1)M^{-1}}^{(i)}$:

   (a) Sample Brownian increments

   $$\left(\left(\varepsilon_{t+jM^{-1}}^{s}\right)^{(i)}, \left(\varepsilon_{t+jM^{-1}}^{x}\right)^{(i)} \right) \sim N\left(0, \frac{J}{M} M^{-1} \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right) \right);$$

   (b) Sample jump sizes from their conditional distributions

   $$\begin{aligned} \left(Z_{t+jM^{-1}}^{s}\right)^{(i)} &\sim \Pi^{s}\left(S_{t+\frac{(j-1)}{M}}^{(i)}, X_{t+\frac{(j-1)}{M}}^{(i)}\right) \\ \left(Z_{t+jM^{-1}}^{x}\right)^{(i)} &\sim \Pi^{x}\left(S_{t+\frac{(j-1)}{M}}^{(i)}, X_{t+\frac{(j-1)}{M}}^{(i)}\right); \end{aligned}$$

   (c) Sample jump times

   $$\begin{aligned} J_{t+\frac{j}{M}}^{s,(i)} &\sim Ber\left(\lambda^{s}\left(S_{t+\frac{(j-1)}{M}}^{(i)}, X_{t+\frac{(j-1)}{M}}^{(i)}\right) M^{-1}\right) \\ J_{t+\frac{j}{M}}^{x,(i)} &\sim Ber\left(\lambda^{x}\left(S_{t+\frac{(j-1)}{M}}^{(i)}, X_{t+\frac{(j-1)}{M}}^{(i)}\right) M^{-1}\right); \end{aligned}$$

   (d) Using the simulated shocks, simulate states and prices forward:

   $$S_{t+\frac{j}{M}}^{(i)} = S_{t+\frac{j-1}{M}}^{(i)} + \mu^{s}\left(t + \frac{(j-1)}{M}\right) M^{-1} + \sigma^{s}\left(t + \frac{(j-1)}{M}\right) \varepsilon_{t+\frac{j}{M}}^{s,(i)} + Z_{t+\frac{j}{M}}^{s,(i)} J_{t+\frac{j}{M}}^{s,(i)}$$

   $$X_{t+\frac{j}{M}}^{(i)} = X_{t+\frac{j-1}{M}}^{(i)} + \mu^{x}\left(X_{t+\frac{j-1}{M}}^{(i)}\right) M^{-1} + \sigma^{x}\left(X_{t+\frac{j-1}{M}}^{(i)}\right) \varepsilon_{t+\frac{j}{M}}^{x,(i)} + Z_{t+\frac{j}{M}}^{x,(i)} J_{t+\frac{j}{M}}^{x,(i)}.$$

where

$$\mu^s\left(t+\frac{(j-1)}{M}\right)=\mu^s\left(S^{(i)}_{t+\frac{j-1}{M}},X^{(i)}_{t+\frac{j-1}{M}}\right)$$

and

$$\sigma^s\left(t+\frac{(j-1)}{M}\right)=\sigma^s\left(S^{(i)}_{t+\frac{j-1}{M}},X^{(i)}_{t+\frac{j-1}{M}}\right)$$

2. Collect the new simulated prices and states into

$$\left(L^M_{t+1}\right)^{(i)}=\left(\left(X^M_{t+1}\right)^{(i)},\left(S^M_{t+1}\right)^{(i)},\left(Z^{s,M}_{t+1}\right)^{(i)},\left(Z^{x,M}_{t+1}\right)^{(i)},\left(J^{s,M}_{t+1}\right)^{(i)},\left(J^{x,M}_{t+1}\right)^{(i)}\right).$$

3. Evaluate the likelihood function at the new state, $p\left(P_{t+1}|\left(L^M_{t+1}\right)^{(i)}\right)$, and set

$$\pi^{(i)}_{t+1}=\frac{p\left(P_{t+1}|\left(L^M_{t+1}\right)^{(i)},P_t\right)}{\sum_{i=1}^{N}p\left(P_{t+1}|\left(L^M_{t+1}\right)^{(i)},P_t\right)}.$$

4. Re-sample N particles with replacement from the distribution of $\left\{\left(L^M_{t+1}\right)^{(i)}\right\}_{i=1}^{N}$ with probabilities $\left\{\pi^{(i)}_{t+1}\right\}_{i=1}^{N}$.

Adding derivative prices adds only one additional complication. To evaluate the derivative prices, we need to compute $C\left(X^{(i)}_{t+1},S_{t+1}\right)$, since S_{t+1} is observed. L^M_{t+1} contains state variables only up to time $\left(t+1-\frac{M-1}{M}\right)$, and therefore we must add an additional simulation step to simulate $X^{(i)}_{t+1}$ conditional on $X^{(i)}_{t+1-\frac{M-1}{M}}$. Then, given $C\left(X^{(i)}_{t+1},S_{t+1}\right)$, the likelihood function in step 3 used to compute $\pi^{(i)}_{t+1}$ is just a bivariate likelihood function. The simplicity and flexibility of the particle filtering approach is clear: even though the derivative prices are not known analytically, the filter still applies as we need only evaluate the likelihood.

References

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.

Aït-Sahalia, Y., 2004a, Closed-Form Likelihood Expansions for Multivariate Diffusions, working paper, Princeton University.

Aït-Sahalia, Y., 2004b, Disentangling Diffusion from Jumps, *Journal of Financial Economics*, 74, 487-528.

Aït-Sahalia, Y. L. Hansen, J. Scheinkman, 2005, Discretely-sampled diffusions. L.P. Hansen and Y. Aït-Sahalia (eds.), *Handbook of Financial Econometrics*, Amsterdam: North-Holland, forthcoming.

Aït-Sahalia, Y., J. Cocho-Diaz, and T. Hurd, 2006, Portfolio Choice with a Large Number of Assets: Jumps and Diversification, working paper.

Aït-Sahalia, Y. and R. Kimmel, 2005a, Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions, working paper.

Aït-Sahalia, Y. and R. Kimmel, 2005b, Maximum Likelihood Estimation of Stochastic Volatility Models, forthcoming, *Journal of Financial Economics.*

Aït-Sahalia, Y., P. Myklund, and L. Zhang, 2005, A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, *Journal of the American Statistical Association* 100, 1394-1411.

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., 2004, Parametric and Nonparametric Volatility Measurement. L.P. Hansen and Y. Aït-Sahalia (eds.), *Handbook of Financial Econometrics*, Amsterdam: North-Holland, forthcoming.

Andersen, T.G., Bollerslev, T. and Diebold, F., 2005, Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility, working paper.

Andersen, T. Bollerslev, T., Diebold, F.X. and Labys, P., 2001, The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association* 96, 42-55.

Anderson, B. and J. Moore, 2004, *Optimal Filtering*, Dover.

Bakshi, G. and C. Cao, 2004, Risk-Neutral Kurtosis, Jumps, and Option Pricing: Evidence from Most Actively Traded Firms, working paper.

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

Bally, V, and D. Talay, 1996, The law of the Euler scheme for stochastic differential equations (II) : convergence rate of the density, *Monte Carlo Methods and Applications*, 93-128.

Barndorff-Nielson, Ole and Neil Shephard, 2004, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics*, 2, 1-48.

Barndorff-Nielson, Ole and Neil Shephard, 2006a, Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation, *Journal of Financial Econometrics* 4, 1-30.

Barndorff-Nielson, Ole and Neil Shephard, 2006b, Impact of jumps on returns and realised variances: econometric analysis of time-deformed Levy processes, forthcoming, *Journal of Econometrics*.

Bates, D., 2001, Post-'87 Crash fears in S&P 500 futures options, *Journal of Econometrics* 94, 181-238.

Bates, D., 2005, Maximum Likelihood Estimation of Latent Affine Processes, forthcoming, *Review of Financial Studies.*

Beckers, S., 1981, A note on estimating the parameters of the diffusion-jump model of stock returns, *Journal of Financial and Quantitative Analysis* 16, 127-140.

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.

Broadie, M, M. Chernov, and M. Johannes, 2005, Model Specification and Risk Premiums: Evidence from Futures Options, forthcoming, *Journal of Finance.*

Calvet, L., A Fisher, and S. Thompson, 2004, Volatility Comovement: A Multifrequency Approach, forthcoming, *Journal of Econometrics.*

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

Chernov, M., E. Ghysels, A. Gallant, and George Tauchen, 2003, Alternative models for stock price dynamics. *Journal of Econometrics* 116, 225-257.

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

Chib, S., M. Pitt, and N. Shephard, 2005, Likelihood based inference for diffusion driven models, working paper.

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

Crisan D. and A. Doucet A., 2002, A Survey of Theoretical Results on Particle Filtering for Practitioners, forthcoming, *IEEE Trans. Signal Processing.*

Dai, Q., Singleton, K., and W. Yang, 2005, Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields, working paper.

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

Das, S., 2002, The Surprise Element: Jumps in Interest Rates, *Journal of Econometrics* 106, 27-65.

Das, S, and R. Uppal, 2004, International portfolio choice with systemic risk, *Journal of Finance* 59, 2809-2834.

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, 2002, The Monte Carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields,* 120, 346-368.

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

Duffie, D. and P. Glynn, 1995, Efficient Monte Carlo Estimation of Security Prices, *Annals of Applied Probability* 5, 897-905.

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

Duffie, D. and K. Singleton, 1993, Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica* 61, 929–952.

Durham, G. and R. Gallant, 2002, Numerical techniques for maximum likelihood Estimation of Continuous-time Diffusion Processes. *Journal of Business and Economic Statistics* 20, 297-316

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., 2002, Do equity prices and volatility jump? Reconciling evidence from spot and option prices, *Journal of Finance* 59, p. 1367-1403

Eraker, B., M. Johannes and N. Polson, 2003, The impact of jumps in volatility and returns." *Journal of Finance* 59, 227-260.

Fernandez-Villaverde, J. and J. Rubio-Ramerez, 2005, Estimating Macroeconomic Models: A Likelihood Approach, working paper.

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.

Gallant, A. Ronald, and George Tauchen, 2002, Simulated Score Methods and Indirect Inference for Continuous-time Models, L.P. Hansen and Y. Aït-Sahalia (eds.), *Handbook of Financial Econometrics*, Amsterdam: North-Holland, forthcoming.

George, T. and F. Longstaff, 1993, Bid-Ask Spreads and Trading Activity in the S&P 100 Index Options Market, *Journal of Financial and Quantitative Analysis* 28, 381-397.

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

Glasserman, P. and N. Merener, Numerical Solution of Jump-Diffusion LIBOR Market Models, *Finance and Stochastics* 7, 1-27.

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

Hull, J. and A. White, 1987, The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance* 42, 281-300

Hausenblas, E. 2002, Error analysis for approximation of stochastic differential equations driven by Poisson random measures, *SIAM Journal of Numerical Analysis* 40, 87-113.

Huang, X. and G. Tauchen, 2005, The Relative Contribution of Jumps to Total Price Variance, forthcoming, *Journal of Financial Econometrics* 3, 456-499;

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

Jarrow, R. and E. Rosenfeld, 1984, Jump Risks and the Intertemporal Capital Asset Pricing Model, *Journal of Business,* 7, 337-51.

Johannes, M., 2004, The Statistical and Economic Role of Jumps in Interest Rates, *Journal of Finance* 59, 227-260.

Johannes, M., N. Polson and J. Stroud, 2004, Sequential Optimal Portfolio Performance: Market and Volatility Timing. Working paper, Columbia University.

Jones, C., 1998, Bayesian Estimation of Continuous-Time Finance Models, working paper, USC.

Kim, S. N. Shephard, and S. Chib, 1998, Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies* 65, 361-93

Kloeden, P. and E. Platen, 2001, *Numerical Solution of Stochastic Differential Equations*, Springer.

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, 2000, *Statistics of Random Processes*, Springer-Verlag.

Liu, J., F. Longstaff, and J. Pan, 2002, Dynamic asset allocation with event risk. *Journal of Finance*, 58, 231–259.

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., 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.

Omori, Y., S. Chib, N. Shephard, and K. Nakajima, 2005, Stochastic volatility with leverage: fast likelihood inference, working paper.

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

Pan, J., and J. Liu, 2003, Dynamic Derivative Strategies, *Journal of Financial Economics* 69, pages 401–430.

Piazzesi, M., 2005, Bond yields and the Federal Reserve, forthcoming, *Journal of Political Economy.*

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.

Pritsker, M., 1998, Nonparametric Density Estimation and Tests of Continuous-Time Interest Rate Models, *The Review of Financial Studies* 11, 449-87.

Pugachev, V., 1987, *Stochastic differential systems: analysis and filtering*, Wiley, New York.

Smith, A. and A. Gelfand, 1992, Bayesian statistics without tears: A sampling/resampling perspective, *American Statistican* 46, 84-88.